Ca' Foscari
University
of Venice

Master's Degree programme

in

Global Development and Entrepreneurship

LM-56

Final Thesis

Data Valuation

The role of data in corporate valuation

Supervisor
Ch. Prof. Nicola Chiaranda

Graduand
Federico Casagrande
Matriculation Number 846876

Academic Year
2020 / 2021

Contents

List of Graphs/Tables

Summary

Introduction **7**
Chapter 1 - Big data economy **9**
1.1 Data 9
1.2 Data types and characteristics 13
1.3.1 Data as economic goods 16
1.3.2 Economic multiplier 19
1.3.3 Price elasticity 20
1.3.4 Capital 20
1.3.5 Impact of economics on data 21
1.4 Data vs Information 21
Chapter 2 - The cost of data **25**
2.1 Data generation 25
2.2 Validation and storage 30
2.3 Analysis 32
2.4 Providing access to data and communicating it 33
2.5 Building a data competent organization 33
2.6 Hidden costs 36
2.7 Data and the economy of scale 37
Chapter 3 - Business models **39**
3.1.1 The economics of networks 39
3.1.2 Economics of learning 39
3.1.3 Data and circular economy 41
3.2.1 Data monetization 42
3.2.2 Using data to improve business decisions 43
3.2.3 Using data to improve business operations 43
3.2.4 How to improve customer offering with data 44
3.2.5 Selling data and insights 46
3.3.1 Measuring the impact of data on business models 48
3.3.2 Value creation 52
3.3.3 Big data business model maturity index 55
3.4 Competitive differentiation 56
3.5 Bloomberg analysis 57
3.6 Successful data business models 58
Chapter 4 - Data valuation **61**
4.1.1 The five Vs of Big Data 61
4.1.2 Key-value drivers 62
4.1.3 Asset and Business Valuation 63
4.1.4 Intangible Assets valuation 69
4.1.5 Consequences of capitalized operating expenses 71
4.1.6 “The dark side of valuation” 73
4.2.1 M&A in the economy of data: the S&P Global and IHS Markit case 76

4.2.2 IPO in the economy of data: Snowflake and the underpricing 79
4.3 Valuation 80
4.4 Financial statement analysis 88
4.4.1 Non-GAAP and Goodwill: IBM 88
4.4.2 Amazon 89
4.4.3 Bitcoin as a digital asset: Tesla 91
4.5 The investor's point of view 92
4.5.1 FCF forecast 92
4.5.2 Market valuation 93
4.5.3 Are listed data companies mature firms? 95
4.6 Valuation conclusions 96
Chapter 5 - Alternative data valuation 99
5.1 The creditor's point of view 99
5.2 The rating agency's point of view 103
5.3 Data valuation evolution 104
5.3.1 Legal evolution 105
5.3.2 Technological changes 106
5.4 Does Value Investing still work? 108
Conclusion 115
References 117
Consulted Websites 119

List of Graphs/Tables

Table 1.1.1 Large data firms

Table 1.1.2 Data firms vs non-digital firms

Graph 1.1.3 Economic impact of data

Graph 1.1.4 Impact of D&A on sectors

Graph 1.2 Data types

Graph 1.4.1 Data value chain

Graph 1.4.2 Value of information

Graph 2.1.1 Data type-source

Graph 2.1.2 Sales of smart home devices

Graph 2.1.3 Cost of GDPR compliance

Graph 2.7 Marginal costs of digital assets

Graph 3.1.2 Economic value of digital assets

Graph 3.1.3 Data and circular economy

Graph 3.2.1 Data monetization

Graph 3.2.4.1 Data use cases

Graph 3.2.4.2 Fast-growing organizations

Graph 3.2.5 Revenue per user

Graph 3.3.1.1 The cost of poor data quality

Graph 3.3.1.2 Impact of data on productivity

Graph 3.3.1.3 Impact of data on ROA

Table 3.3.2.1 V2F and V2C

Table 3.3.2.2 V2F and V2C 2

Graph 3.3.2.3 Extra value

Graph 3.3.2.4 Customer Lifetime Value

Graph 3.3.3 Data business model maturity

Graph 3.4 Impact of data usage

Table 3.5 Data revenue by company size

Graph 4.1.3 Valuing a Business or Asset

Graph 4.1.3.2 Value range

Graph 4.1.3.3 Degree of intelligence and data value

Graph 4.1.3.4 Data management and value

Graph 4.2.1 S&P Global and HIS Markit M&A

Table 4.3.1 Dataset Summary

Table 4.3.2 Profitability: Dataset and Market

Table 4.3.3 Profitability: Information Industry and Market

Table 4.3.4 Profitability growth

Table 4.3.5 Cost structure: Information Industry and Market

Table 4.3.6 Price and Value

Table 4.3.7 Cost of capital

Table 4.3.8 Growth rate

Table 4.4.2 Amazon: Financial statement

Graph 4.5.2 Porter's Five Forces

Graph 5.1.1 Altman Z-score S&P 500 companies

Table 5.1.2 S&P 500: Top 10 constituents

Graph 5.1.3 S&P 500: Sector breakdown

Graph 5.2.1 S&P corporate rating methodology 1

Graph 5.2.2 S&P corporate rating methodology 2

Table 5.3.1 Data usage barriers

Graph 5.3.2 Blockchain

Graph 5.4.1 The rise of intangibles

Graph 5.4.2 Growth vs value investing

Introduction

"The most valuable commodity I know of is information."[1]

Gordon Gekko, Wall Street (1987 film)

Data is shaping the future of humanity. It is facilitating research, making societies more productive, and it is helping the efficacy of public policies as well as boosting the productivity of private businesses.

According to Bernard Marr [2] less than 0,5% of all data is analyzed or used. However, data is revolutionizing the way we work and it is the companies that view data as a strategic asset that will survive and thrive.

The worldwide big data market size revenue has increased significantly from 7,6 billion US dollars in 2011 to over 50 billion in 2020. The most reliable predictions indicate that the market size is expected to double within 7 years[3].

Companies, governments, academic institutions, and citizens have already started using data in their everyday activities. Data is embedded in IoT devices like smartphones and modern cars that interact with data warehouses. The growing volume of data has created a digital economy of big data technologies and services that have been exploited by innovative firms.

These economic actors are investing billions to unlock the potential of big data because, in the next few years, the IT function's success will depend on how effectively it enables the organization to unleash the power of big data. At the same time, data storage, processing, and transfer cost have fallen to near zero. That is why big data is the heart of new business models.

Data is not considered an asset by current accounting standards[4][5][6][7][8]. However, it is important to know the value of this asset because it may be very beneficial for firms and it must be evaluated in the case of M&As.

[1] Wall Street - Movie Quotes - Rotten Tomatoes
[2] Data Strategy 2017
[3] Source: Statista
[4] https://www.gartner.com/smarterwithgartner/why-and-how-to-value-your-information-as-an-asset/
[5] https://blogs.gartner.com/andrew_white/2019/09/04/consequence-valuing-data/

In this thesis, I have focused on the impact of data on corporate decisions and activities. To find the value of data, there are two possible approaches. Analyzing the financial statements of all the companies, trying to extract from them the value of big data.

Alternatively, one could only study the statements of big data firms and their business models to figure out how they evaluate big data. I have used the latter (inductive approach) because I think my analysis can be more efficient and data-based.

In Chapter 1, I have described the data Industry and the main characteristics of data.

Afterwards, I have analyzed the business model of data firms. In Chapter 2 I focused on the cost structure of these companies and in Chapter 3 I have illustrated how these firms use data to be profitable.

Chapter 4 is the core of this Thesis. In this section, I have reported the most common corporate valuation tools and I have applied these methodologies to data companies. I have also described some peculiar cases of IPO, M&A, and accounting procedures that are helpful to identify the value of data from financial statements. Moreover, I have analyzed the competitive position of data firms in the market.

Finally, I have offered an alternative valuation of data firms from both the point of view of creditors and rating agencies.

[6] https://datafloq.com/read/data-valuation-worth-zero-or-trillions/1916

[7] https://www.acutus-ca.com/wp-content/uploads/2019/07/Should%e2%80%9cData%e2%80%9dBeRecognisedasanAssetonTheBalanceSheet.pdf

[8] https://www.acutus-ca.com/2019/07/31/should-data-be-recognised-as-an-asset-on-the-balance-sheet/

Chapter 1 - Big data economy

1.1 Data

In the last few decades, technological advancements have radically changed the way companies operate. Digital assets are becoming more and more important and their proper use is now crucial for corporate success.

Firms that own a significant quantity of digital assets or that operate in a digital market are very hard to evaluate. The first difficulty is that of using a pricing method that suits digital assets such as data. These goods have very peculiar characteristics that make them different from others. Using traditional valuation methods may not be appropriate when it comes to assets that do not lose value with usage (for instance). Besides, digital firms may have unique and never seen business models that produce externalities that are difficult to measure.

For example, in 2014, Aswath Damodaran, professor of corporate finance at NYU criticized the Uber valuation made by Benchmark Capital that he claimed to have overvalued the American company. According to Damodaran, the value of Uber at that time was nearly $5,9 billion. The investment fund thought that the overall value of the firm was $17 billion. The main reason why these two valuations outcomes are considerably different is that Damodaran did not consider the network effect that Uber triggers. He used traditional valuation techniques for a very innovative firm (Parker et al. 2016).

Network effects, born in the last few years, mean the impact that the users of services have on the value created for each user (Parker et al. 2016). In other words, digital firms allow users to be both customers and suppliers at the same time. This is the case of Facebook where users upload information and access it. So, the distinction between buyers and sellers is blurred in the digital economy.

The network effect can be positive if the platform makes the sharing of information, data, and activities efficient or negative, in case platforms are poorly managed. In the digital economy, the platform effect is the largest source of value creation and competitive advantage for a firm (Parker et al. 2016). The platform effect is based on the existence of data and information to share. Assessing the value created by this

effect would be impossible without knowing the value of information shared on the platform. That is why it is necessary to create new valuation techniques.

The birth of a new digital economy has created a huge set of opportunities to be exploited. The presence of digital assets helps firms to make better decisions as well as boosting labour productivity. Amongst these assets, there is one that stands out for utility and versatility: Data.

Data does not only improve the decision-making of firms but also allows policymakers and scientists to run more efficient policies and research. This is helping, for instance, to track progress on sustainable development goals and the effectiveness of public spending (UN Data Economy 2019). Considering the rising environmental concerns and the median age of the world population, it is undeniable that maximizing the utility of our actions will be crucial in the future.

In 2014, PWC had estimated that there were 4,4 zettabytes of data in existence, as many stars in the visible universe (PWC 2014). One-third of this data is probably generated by enterprises and the remaining is mostly generated by consumers that interact with their services.

The MIT indicated in the 2016 report "The rise of Data capital" (MIT Technology Review 2016) that 84% of the market value of S&P500 companies comes from intangible assets such as data and software. In the US only, the possible value of these assets was nearly $8 trillion.

Companies like Google, Amazon, Uber, and Netflix realized that data is an extremely valuable raw material that can create new kinds of value. Data must be analyzed and turned into something more useful than the data itself. You must turn it into information. That is why economists agree on considering data as the "new oil" (Demchenko et al. 2018) This new commodity is one of the main reasons why the new digital economy had a disruptive effect on the market.

Google, for example, processes over 70.000 searches every second. If these searches were to be converted into hard copies, they would generate 1 billion 200-page books per second (UN 2019 Data Economy).

The 2019 UN report on the data economy indicates the biggest data companies in the world.

Table 1.1.1 Large data firms

Niche areas	Dominant firms
Search engines	Google
Social media/Messaging	Facebook, WhatsApp, WeChat
Share economy platforms	Uber, Airbnb
Content and service provider	Netflix, Venmo, Expedia
Retailer	Amazon, eBay, Alibaba
Operating systems	Microsoft, Apple, Google
Data hardware	Apple, Samsung, Cisco

The industries in which they operate are diverse: search engines, retail, share platforms, and so on. Table 1.1.2 (UN 2019) also compares some of these data companies to other large firms that do not operate in the digital economy. Digital firms have a noticeably higher Price-Earning ratio. This indicates that investors are willing to pay the shares of digital firms more than traditional firms'.

Table 1.1.2 Data firms vs non-digital firms

	Total number of employees	Total revenue (bn, $)	Total market valuation (bn, $)	Price-Earning ratio
Apple	132,000	266.0	868.8	17.20
Amazon	566,000	178.0	560.0	190.16
Facebook	25,105	40.7	508.9	32.74
Google	85,050	89.4	720.8	58.65
Microsoft	134,944	110.0	570.0	57.34
Exxon	69,600	237.0	348.6	17.52
Johnson and Johnson	134,000	76.0	375.0	22.13
Proctor and Gamble	95,000	65.0	226.8	23.93
Royal Dutch Shell	92,000	305.0	271.9	12.51
Walmart	1,500,000	500.3	289.8	22.69

Besides, the impact of data on the economy as a whole is remarkable. In the EU only, the direct, indirect, and induced impact on the economy is estimated at over 300 billion euros in 2017 (UN 2019). The direct impact is only 19% of the total. Externalities and the spillover effect have a much larger impact on the economy.

Finally, data is the core of all digital technologies such as AI, Machine Learning, Blockchain, IoT, and all internet-based services. If the digital economy grows, the demand for data will follow.

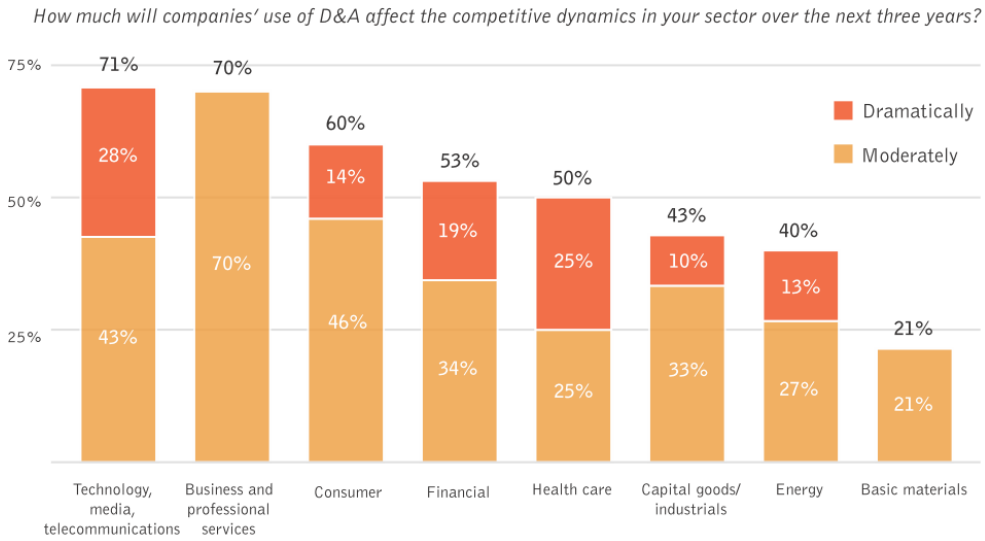
Graph 1.1.3 Economic impact of data

Sector by sector view of Data & Analytics (KPMG 2015)

Companies from every sector have something to gain from Data and Analytics. KPMG indicates the industries in which data recently had the most disruptive effect (KPMG 2015). Graph 1.1.4 illustrates whether the use of D&A has affected moderately or dramatically the industries from 2015 to 2018. Considering that the growth of the data economy has just begun, the disruptive effect is likely to continue in the future.

The impact of D&A is particularly high in Technology, media, and telecommunication. Also, Business, Finance, and healthcare benefit a lot from data.

Graph 1.1.4 Impact of D&A on sectors



1.2 Data types and characteristics

Data is a unique economic good. First of all, it is not fungible (MIT Technology Review 2016). There are different streams of information that are different to compare and evaluate. This is why it is really hard to find the exact value of a dataset. Also, the value could be different for two different firms since they would probably have different analytics skills. The real value of data is not data itself but the information that you can extract from it. With other commodities, like a barrel of oil, one unit can be substituted for another. Secondly, it is non-rivalrous (MIT Technology Review 2016). Data can be used by more than a person (or computer) at a time and it is not consumed in the usage process. This makes data being similar to other intangible assets such as patents. Besides, data is an experience good (MIT Technology Review 2016): it has no intrinsic value. Its value increases after being used and analyzed properly.

Data can be classified according to many different parameters. The first categorization that I use is Dimension and Data types (GSMA 2018).

Personal Data

This data is identifiable with an individual and it may be public or private. Companies and public organizations (ex: Hospitals) have always collected personal data such as name, address, health records, and bank details. The collector is in charge of protecting the privacy of the people by storing and using the data according to the guidelines prescribed (General Data Protection Regulation). This data can be grouped by how it was created or shared (GSMA 2018).

- Volunteered: data explicitly shared by the people. It may be uploaded on social media and websites
- Observed: data collected in a non-intrusive way on the basis of behaviors (ex: location used to share content)
- Inferred: information created by analyzing already existing data (ex: credit ranking, next purchase suggestions)

A subcategory needs to be mentioned: pseudonymous data. “This is defined as personal data that has been subjected to technological measures such as hashing or encryption such that it no longer directly identifies an individual without the use of additional information” (GSMA 2018). This data does not identify with a person

without the use of additional information. The European General Data Protection Regulation (GDPR) considers it as a type of personal data. Pseudonymity is a safeguard that reduces privacy risks. There is also personal data that is available more publicly such as data shared on social media, public registers, tax payments in countries like Norway (GSMA 2018).

Non-personal Data

Machines also generate a lot of data that is fully anonymized. Engine performance, transport ticket sales, energy usage, shipping records are some examples of data belonging to this category. In this case, no privacy concerns seem to be relevant (GSMA 2018).

Two characteristics affect personal and non-personal data. Timeliness and Format (GSMA 2018).

Timeliness

Timeliness is the time dimension of data because it may be recorded at a precise point in time (GSMA 2018). More up-to-date data is usually more valuable because it is more informative (ex: financial indicators). Also, users may be willing to pay to access data earlier than others because "The most valuable commodity is information"[9]. Getting to know something before the others (competitors) do so could make you stand out in a business. Some other data might be more dynamic and generate value over time (ex: traffic flow). The difference between static and dynamic data is timeliness. In other words, how data is measured in respect of time (GSMA 2018).

Format

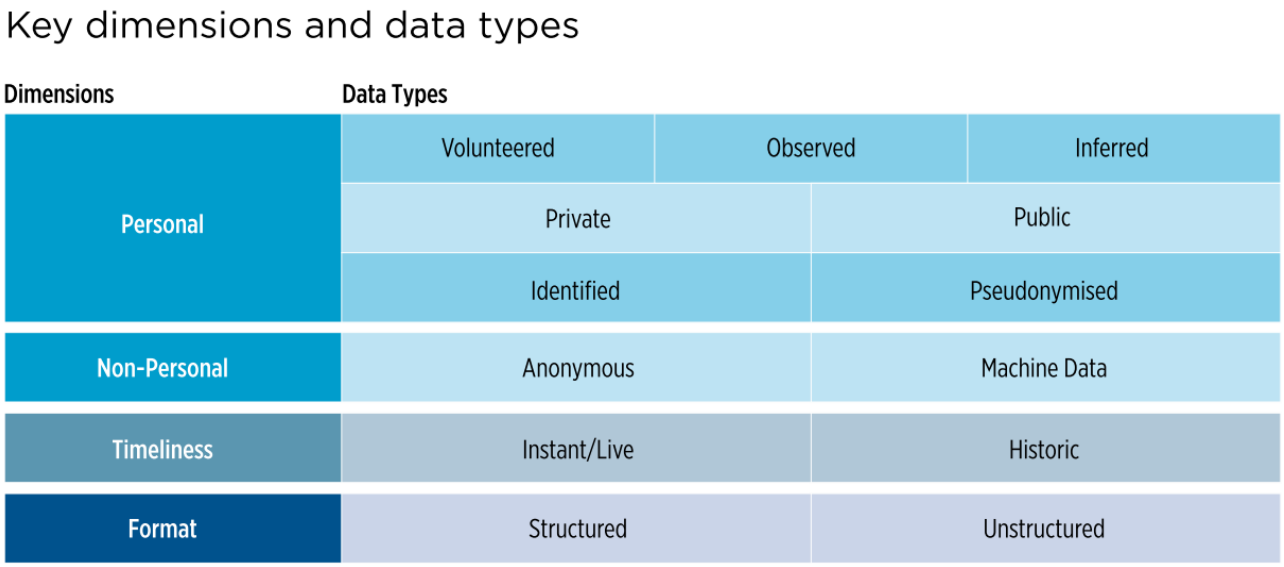
Another important factor is understanding whether data is structured or not (GSMA 2018). Structured data is easy to identify because they are organized into a database. Unstructured data, on the other hand, are harder to organize and search since it is not stored in one only file or database. More than 80% of all data is unstructured. The latter can become structured after a deep organizational process. Many companies have started developing services that organize this data so that it can be analyzed and

[9] Regulating the internet giants - The world's most valuable resource is no longer oil, but data | Leaders | The Economist

they can extract value from it. The main challenges in this process (GSMA 2018) involve:

- Relevance: Artificial Intelligence is powerful and accurate but not infallible
- Volume: the speed at which this type of data is increasing is remarkable. Most companies cannot afford to deal with such a large quantity of data because it requires a high investment in labor and infrastructure
- Quality: unstructured data is unverified until someone analyses it. For example, a mistyped search could lead to a stream of irrelevant adverts
- Usability: unstructured data must be verified and analyzed before one can use it. This is costly and it takes time.

Graph 1.2 Data types



Another categorization is suggested by PWC in its 2019 report "Putting a value on data". PWC distinguishes between the data source and data category.

In the first case, data can be

- Authored: created with a human process (ex: photos)
- User-provided: provided by users without any expectations (ex: social media)
- Captured: recorded from real events (ex: financial transactions)
- Derived: generated combining and processing other data (ex: credit scores)

Data categories are (PWC 2019)

- Master: describe things that are critical to a firm's operations

- Transactional: describe transactions
- Reference: information used for categorizing data
- Metadata: characteristics that make data easier to use
- Unstructured: data lacking syntax to describe objects and attributes

PWC also suggests a size-based categorization of data (PWC Data drive innovation 2014).

Big Data is the most famous category of data. It regards voluminous datasets that embed diverse and complex pieces of information. Little data is similar to Big Data but on a smaller scale.

Internal enterprise data is collected by an organization about its own process. It may be digital and it includes both qualitative and quantitative information. It can also be anonymous.

Open data is a source for governments and institutions that need as much information as possible to make decisions. The adjective “open” refers to the fact that this data can be used by anyone without restrictions (PWC 2014).

My data is securely held data about individuals (PWC 2014). Access to this data is regulated by strict rules (ex: health records).

1.3.1 Data as economic goods

A 2018 research made by the University of Amsterdam (Demchenko et al. 2018) suggests that data needs to be Findable, Accessible, Interoperable, Reusable (FAIR data principles). Also, data needs governance and management by companies and institutions. Data can also be traded (like all economic goods) (Demchenko et al. 2018) considering that:

- IoT tools that produce data can also add value to data by producing secondary and complementary data, for instance
- Personal data can also be traded to improve market research and services development
- Data exploration can be traded even if that data is somehow only a raw material

- Existing datasets' value increases if they can be offered in the market

The trade of data facilitates the recognition of data as an economic good. Therefore, the industrial and commoditized data has some further economic principles known as STREAM properties (Sovereign, Trusted, Reusable, Exchangeable, Actionable, Measurable) (Demchenko et al. 2018).

Quality, value, traceability, branding, authenticity, are also important to allow data commoditization (Demchenko et al. 2018).

Data sovereignty

It allows companies and data owners to control their data (Demchenko et al. 2018). These economic actors can allow or not allow other parties to use the data they have collected and/or created. In any case, if others want to use or access this data, they must have the explicit consent of data owners.

Of course, this principle contrasts with open data (data that is accessible to anyone) stored on publicly accessible clouds. Modern clouds let the owners decide who to authorize the access to so that this principle can be respected.

Trusted data

Data must be trusted and verifiable because it is used in the decision-making process. One should not only verify the truthfulness of data but also the models used to collect/create it and as much info as possible on the data provider. The trustworthiness of data is, in fact, ensured by the reputation of the provider.

Data reusability

Data should be reusable multiple times (Demchenko et al. 2018). Data reusability can create multiple opportunities for economic actors such as SMEs and individuals. Unlike other assets, the value of data may not deteriorate when it is engaged in economic activities (amortization). Sometimes data even acquire value with usage because it becomes more complete and informative. In other cases, the value may decrease because data becomes obsolete and not usable anymore (Demchenko et al. 2018).

Data exchangeability

Producers and consumers can interact in a normal exchange of goods. Data can be exchanged for other goods or money even though the pricing of data is still a complicated matter. Pricing models have not been fully developed yet because some problems still stand (privacy, deterioration,…).

Data actionability

After purchasing data, one should be able to use it in a process (economic activity, research,…) (Demchenko et al. 2018). Some of the most common uses of data include customer experience, logistics, and user feedback. The artificial intelligence used by most firms widens the number of activities in which data can be used.

Data measurability

First of all, data can be evaluated as an economic good to trade. Secondly, data can be measured as a part of an economic process (Demchenko et al. 2018). In other words, you could assess the impact (and consequently, the value) of data in the economic activities of firms or research activities run by institutions. In the first case, the evaluation is quite hard to make because effective data valuation models have not been created yet. Assessing the impact of data on corporate activities seems to be easier. You could use a backward approach: starting from the financial statements of firms, it is possible to figure out what the value of data is (or, at least, an estimate of abnormal earnings generated by the presence of data).

Bill Schmarzo, CTO of Dell, and Mouwafac Sidaoui, Associate Professor of IT at USFCA, have applied some economic concepts to data, to understand this asset more deeply[10]. They claim that the importance of data has changed over the years. Initially, data was considered as the exhaust from the operating activities of the firms. The accumulation of this "waste" has created an enormous quantity of data: Big Data. Companies have realized that this good can become an asset if one can use it properly.

[10] Applying Economic Concepts To Big Data To Determine The Financial Value Of The Organization's Data And Analytics, And Understanding The Ramifications On The Organizations' Financial Statements And IT Operations And Business Strategies (delltechnologies.com)

Scarcity

Scarcity means limitations. The scarcity of resources is one of the founding principles of economics (Schmarzo, Sidaoui 2017). Companies and organizations have limited resources to use, which is why it is crucial to use them efficiently, taking into account the budget constraint and priorities. Therefore, these agents need to acquire and develop analytical skills. How these organizations prioritize their analytic resources (capital and labor) will be essential in determining their success or failure.

Postponement theory

Some may choose to postpone a decision. One may seek to gain more information or delay the decision to obtain better terms. The need for postponing a decision is due to insufficient data that does not let you assess risks and costs adequately (Schmarzo, Sidaoui 2017).

Efficiency

Organizations need to create metrics to measure the effectiveness of operations and performance. They can also use metrics to make a predictive analysis so they can quantify potential discrepancies. This is a never-ending mechanism because Artificial Intelligence allows companies to store data coming from previous operations and learn from them (Machine Learning). This eventually increases the efficiency by preventing repeating the same mistakes made in the past. The interaction between data and AI is crucial (Schmarzo, Sidaoui 2017).

1.3.2 Economic multiplier

A multiplier (in economics) is an economic effect that makes the increase of income larger than the initial amount of spending. The multiplier effect indicates how investments in acquiring new data and developing data analytics tools can generate a return higher than the initial investment (Schmarzo, Sidaoui 2017).

For example, in the 1980s, some consumer packaged good manufacturers like P&G, Unilever, and Kraft based their strategies on audit data (Schmarzo, Sidaoui 2017). They have been able to use the same data for multiple operations such as demand planning,

supply chain optimization, pricing, and trade promotion. So they paid for data once but the value of data was multiplied several times, creating a multiplier effect.

1.3.3 Price elasticity

This concept refers to the percentage of increase/decrease in demand for a good due to an increase/decrease in price. It is expressed by the formula:

$$\eta = (Q\,1 - Q\,0) \div (Q\,1 - Q\,0) \,/\, (P1 - P\,0) \div (P\,1 + P\,0)$$

where:

η= price elasticity of demand

$Q\,0$= quantity demanded at price $P\,0$

$Q\,1$= quantity demanded at price $P1$

Data assets are inelastic (Schmarzo, Sidaoui 2017). An increase in price does not significantly affect customers' behavior. The inelasticity is due to the return on investment (ROI) generated by these resources. Their quantity demanded only increases when firms use them more proficiently.

1.3.4 Capital

Adam Smith defined capital as "a part of a man's stock which he expects to afford him revenue" (Schmarzo, Sidaoui 2017). This definition is relevant since data may provide a revenue stream in four ways (Schmarzo, Sidaoui 2017):

- Optimization of a business process (ex: increasing customer retention rate)
- Reducing exposure to risk through management compliance, security, and so on
- Discovering new revenue opportunities identifying unmet markets
- Delivering a better customer experience that increases customer satisfaction

1.3.5 Impact of economics on data

Data as an asset

The most traditional assets are, for instance, stock, bonds, vehicles, and buildings. Regardless of their tangibility, they suffer from two limitations :

1) They cannot be used simultaneously for different uses (ex: an airplane can only fly one route at a time)
2) They depreciate with usage

Data does not suffer from the aforesaid limitations (Schmarzo, Sidaoui 2017). The same piece of data can be used for multiple operations at the same time. Besides, usage may increase the value of data by making it more complete or informative. Hence, data is an interesting asset in which to invest.

Data as a currency

Currencies are used as a medium of exchange in transactions. They represent a quantifiable financial value. Data does not suffer from the two transaction limitations that I stated above (Schmarzo, Sidaoui 2017). This makes them potential currencies.

For example, national governments may require citizens to provide personal information (age, place of birth,…) to access some public services (healthcare). By doing this, they can manage the data and try to turn it into useful information on public utility.

Data as monetization

Most companies realized that one does not monetize the data by selling it but it is more profitable to sell insights obtained with data analysis (ex: customer purchase behavior, new market demands, …) (Schmarzo, Sidaoui 2017).

1.4 Data vs Information

Turning data into digital intelligence became crucial for firms. The ability to collect, store, analyze data brings a strong comparative advantage. In other words, you need to transform data into information.

The information allows making better decisions, improving efficiency, and performance. This asset is the result of data processing and analysis. Data is, in fact, a

raw material that you need to put into an operating activity to extract something useful from it. Therefore, you need plant and equipment (hardware and software) such as AI tools, computers, and of course workers to make data useful (University of Melbourne 1999).

This economic process is called Data Value Chain (GSMA 2018). Since the whole digital economy relies on the existence of data, and what matters is the right data, not all data, acquisition and analytics is a crucial process (GSMA 2018). The four steps of this data value chain are:

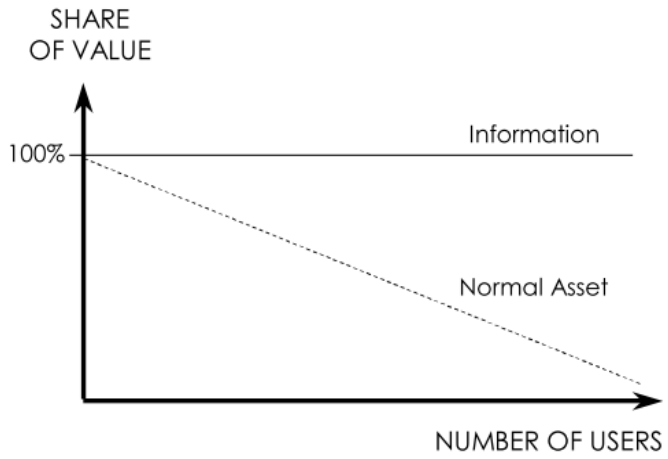
1) Generation: recording and capturing data
2) Collection: collecting, validating, and storing data
3) Analytics: processing and analysis to generate knowledge (information)
4) Exchange: putting the outputs to use, internally or by trading them

Graph 1.4.1 Data value chain

The information has a cost and a value. Seven laws characterize it as an economic good (Demchenko et al. 2018).

1) Information is infinitely shareable. It can be shared without a consequent loss of value for any parties. A good example of this “law” is the World Wide Web. The information that you can find on the internet is available for everyone at the same time. While normal assets lose a share of value with an increase in the number of users, information’s value remains constant. Information can even experience an increase of value because the larger the number of users, the larger the utility extracted from it. However, replication does not add value because it does not produce new information.

Graph 1.4.2 Value of information



2) The value of information increases with use. As I said before, most resources depreciated with usage. The information does not because it does not really have value itself. The value of unused information is actually negative (so information becomes a liability) because organizations bear storage and acquisition costs anyway. Unused information is a waste. It is important to know what information exists, where it is located, how to use it. Also, organizations need information literacy, skills required to extract value from information (after being aware of its existence).
3) Information is perishable. Like other assets, information may depreciate over time. For example, if a customer has changed their address, the old piece of information is not useful anymore. That is why you may need to verify that a specific piece of information has been collected recently. If not, you should be careful using it in your operations.
4) The value of information increases with accuracy. Inaccurate or out-of-date information can be very costly. If you use it in your economic activities you can make operational errors and base your decision-making on the wrong information. If information accuracy falls below a certain level, information becomes a liability because it is misinformation.
5) The value of information increases when combined with other information. This process is called information integration. For example, customers and sales

information can be put together into a business plan to produce a more complete and reliable strategy.

6) More is not necessarily better. In most cases, the more resources you have (ex: money) the better. In the case of information, quality seems to be more important than quantity. With the development of IT services and tools, the lack of information is not a problem anymore. The overabundance of information can be an issue instead. Information overload makes comprehension more difficult. The value of information increases until the quantity obtained reaches the optimal limit. However, decision-makers' confidence keeps increasing if the quantity crosses the point of overload. The feeling they might have is: “I have more information to analyze, then my model will be more complete/accurate”.
7) Information is not depletable. Information, unlike other assets, is self-generating: the more you use it, the more you have it. The original information used in the operation remains and you add the new information, which is the output of your analysis. Information is not a scarce resource.

To conclude, data (and consequently information) is probably the least managed asset. It is still not well understood and it is not easy to figure out how to use it properly. The fact that there are not enough satisfying data valuation models makes data management even more challenging. In the following chapters, I describe how the most successful digital firms have managed data and how they turned it into a very profitable asset. This could help to set up some general rules on how to evaluate and manage it.

Chapter 2 - The cost of data

As stated in the previous chapter, data is a raw material. It is the input of economic, social, and human activity. The outcome (information) is generated through data analysis which is a costly process. Also, acquiring and storing data is not free. You can get data from an internal process or external companies/organizations. In both cases, you must bear some costs that need to be included in the valuation model. It is important to figure out how much the data costs to be able to evaluate it. Besides, one of the main challenges companies face is finding a cost-effective data strategy that allows you to use your data to get useful insights about your business and, generally speaking, reality. (Oracle 2013).

In this chapter, I have described the data value chain, meaning the process that starts from the data acquisition and terminates with the obtainment of the outcome which is, in this case, information.

2.1 Data generation

The first step of the value chain is data generation (GSMA 2018).

There are many different ways to acquire data. One can, for instance, take surveys, capturing details when selling products, asking customers some questions, and so on (Marr 2017). You can also record transaction details or analyze the e-mails as well as pictures (ex: multimedia firms).

On the other hand, when acquiring data from external parties, you just get some information that other companies/organizations have collected and analyzed. Of course, when you acquire data from other companies (external data) you can require customization of data to only receive information useful for your business. This customization that data service providers offer is called “getting the job done” and it is one of the most useful value propositions of a business model (Osterwalder, Pigneur 2010).

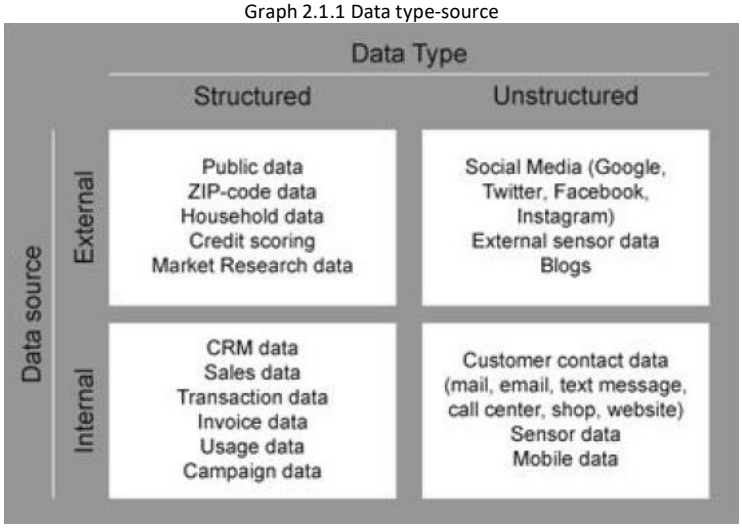
It is possible to get data from specialized companies like Amazon, IBM, Google, or even from institutions. For example, in 2013, the US government pledged to make all government data free and accessible (Marr 2017). The availability of this data could

help researchers and companies to make more effective valuations and create more efficient business models.

Internal and external data can also be combined. Hence, every firm can have a unique set of information and business model. There are several options for accessing data. However, what matters the most is not getting as much data as you can but it would be more appropriate to only get the data that actually helps to reach your strategic goals. You don't want all data, you want the right data.

Graph 2.1.1 shows two different categorizations (Verhoef et al. 2016). What matters is not only the data source (where you get your data) but also the data type, meaning what kind of data you get. In other words, regardless of the data source, you could get easy-to-analyze structured data (which is however usually less informative) or unstructured data. The latter is certainly harder and more costly to analyze (it requires someone to "translate" a dataset into readable and usable information) but it could produce a more useful outcome (Marr 2017).

Graph 2.1.1 Data type-source



Internal data is usually cheaper to get (Marr 2017). Not only internal data is cheaper but it has another advantage: you do not need to rely on third-party information to develop your business model or to make decisions. According to Marr, the biggest downside of internal data is the need to bear some costs to maintain and secure it. These costs include both labor and capital. Therefore, these are both fixed (clouds where you store data, analysis of a dataset) and variable (the analysis may depend on

customer's requests). Also, internal data may not produce enough useful information so you could also need to acquire external data (and pay for it).

The strategy adopted by Netflix could be quite profitable. The firm started collecting both structured and unstructured external data (through the distribution service). After collecting a sufficient amount of external data, Netflix has started producing its own content to be able to collect the same type of data internally. The combination of the two different strategies (adopted at different stages of the company) seems to be very profitable (Marr 2017). This case suggests that one could consider the combination of these two different collection strategies at the same time or starting a business relying on external sources and then creating an internal data collection process.

Getting data from third parties reduces the risks of bearing some fixed costs but it makes other threats emerge. In case you don't get free external data (like government data), your company must rely on information provided by other firms. Therefore, you often have to pay for access and you will somehow be dependent on other firms' activities. Also, you should constantly weigh the risk of not getting the data (what will the consequences be for your firm?), the risk of keeping the "subscription", and the possibility of producing the same data internally (How much would that cost?), if possible (Marr 2017).

Of course, there are many advantages to external data regardless of the type (social media data, Google Trends data, economic and weather data).

First of all, this data is usually collected and analyzed by some of the largest companies in the world. These firms have the capability, know-how, infrastructure to collect a huge amount of data, store it, and analyze it. In addition, these activities represent their core business. They can select the insights that your business needs so you will only get useful data. As I stated before, you don't want all data, you want the right data. In fact, having too much (useless) data to store and analyze could drain resources that you could employ more profitably. Most SMEs do not have enough resources to deal with both the data business and their core business (Marr 2017).

This is the reason why external data is usually more useful. The thing is that if you get structured internal data (ex: transaction and usage data), this will probably give you little information that will not be enough to base your decisions on. Unstructured data

needs a deep and costly analysis so that you have to create a specialized business unit to deal with it.

When you get data from third parties, you usually get insights that are useful for your business. These may come from structured data or analyzed unstructured data. The cost to bear if you want to get it is justified by the high utility that this data carries into the firm. The cost refers to data science expertise (analysts that you need to hire), specialized tools (data analytics software, computers), and the creation of a data-driven infrastructure[11]. I have described these aspects later in the chapter.

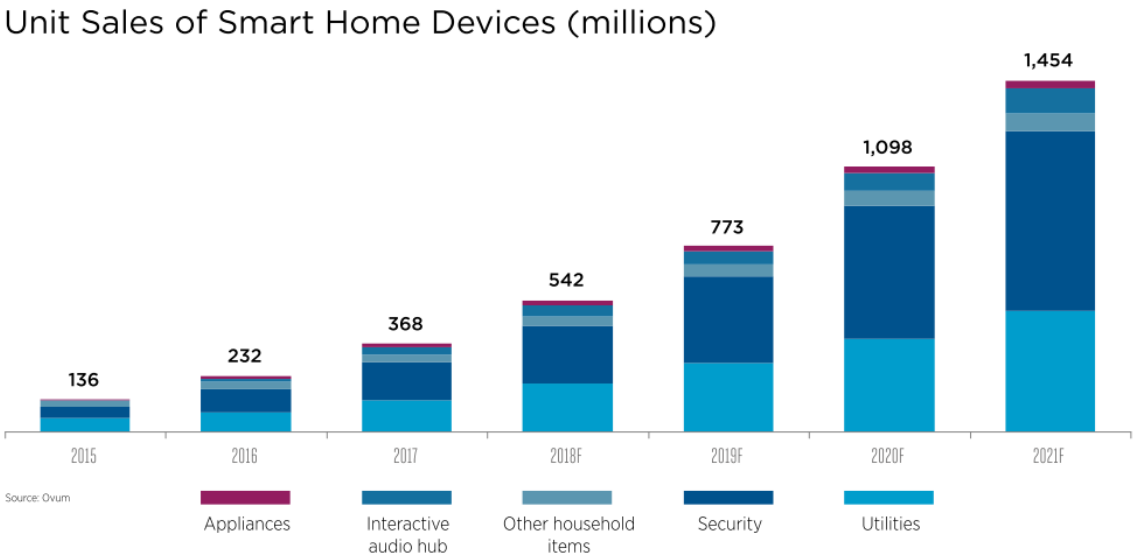
The California-based firm Apixio, for example, harnessed external data successfully (Marr 2017). It has started collecting all health records stored in doctors' databases, hospital records, government records, and so on. After doing that, it could analyze this data and made it available to patients and clinicians so that they could use it to have more individualized care.

If the data that you want does not exist yet, you could use some technological tools such as IoT devices that can collect it. For example, the agricultural company Springg exploited the potential of some special devices that can analyze the soil and extract data. After, they collected it and analyzed it so that farmers could benefit from the information produced (Marr 2017). Even if you do need to invest some money in devices and in finding innovative ways to collect data, being first to market with the data could give you a significant competitive advantage over competitors.

Another example of IoT tools that capture data is smart home devices. The bar chart below illustrates that the sale of these tools has increased significantly over time (GSMA 2018). As a consequence, the amount of data collected by these devices improves and then the storage, analysis costs will follow up.

[11] Structured vs. Unstructured Data: A Complete Guide - Talend

Graph 2.1.2 Sales of smart home devices



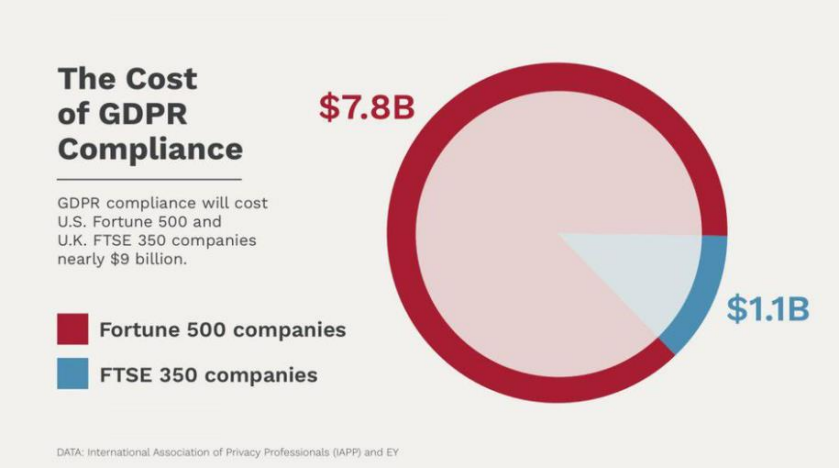
When talking about data acquisition, one must take into account that data can be collected if the service providers gain explicit consent from users. Even though for machines it is unproblematic and relatively easy to collect the data, the data generation is based on human inputs (GSMA 2018). Consumers use a free service and agree to their data to be shared with the provider. This mechanism allows users to benefit from a free service (they do not pay with money, they pay with data) and service providers to get the data they need. It is a win-win business model. However, this requires the respect of regulations such as the GDPR that entails compliance costs[12][13] and limitations. According to the EU, millions of SMEs are not GDPR compliant[14]. Some of these firms are concerned about the possibility that authorities will impose penalties for that, since, unlike big firms, they could not afford to hire lawyers. Others have gladly become compliant because they don't believe the GDPR would slow down business growth. If that is true, it means the data business is so profitable that one would not mind bearing consultancy and technology costs to get into it.

Forbes reported that the GDPR costs for US Fortune 500 and UK FTSE 350 companies is 9 $billion (2018)[15]. Also, it suggests that some firms dump a huge amount of data rather than becoming compliant. Mark Zuckerberg has himself admitted that the

[12] Is GDPR worth the cost? (computerweekly.com)
[13] • GDPR: implementation cost by sector 2018 Statistic | Statista
[14] Millions of small businesses aren't GDPR compliant, our survey finds - GDPR.eu
[15] The GDPR Racket: Who's Making Money From This $9bn Business Shakedown (forbes.com)

respect of these regulations is costing Facebook million of dollars. The compliance is particularly costly if one wants to do business in Europe, where regulations are more strict. Finally, small firms are likely to be hit hardest for two reasons. First, most compliance costs are fixed costs. So you need to generate a notable revenue to pay back the large initial investment. Secondly, the regulations are so vague that good lawyers and advisors could help companies to bypass them. Of course, only big firms can afford to rely on their competence to do that by incurring additional costs.

Graph 2.1.3 Cost of GDPR compliance



To sum up, data can be collected internally by creating a suitable infrastructure or by relying on an external data company to capture data on your behalf. If you decide to get external data, you can buy it from data companies without changing the infrastructure. Alternatively, you could get data in both ways (Marr 2017).

2.2 Validation and storage

Once the data is generated and collected, it must be validated (GSMA 2018). This data is to be put into a dataset that can be analyzed. Also, it can be integrated with other data to make a more complex analysis. Before doing this, you must make sure that the data has been collected correctly by verifying the functioning of the IoT devices employed. After, you can store the data.

Data storage is the activity of organizing the data that you have previously generated and collected in an organized and efficient way. This organization should make accessibility and analysis easy.

This activity is carried out thanks to the interaction between hardware and software. The new technologies make storing and securing the data quite cheap since they allow firms to keep the data in a cloud rather than in a physical space. The advantage of clouds is that they prevent the loss of data due to disasters because data can be recovered if something bad occurs. Besides, a dataset stored in a cloud can be easily distributed across the storage network (GSMA 2018).

Hard disks and in-house servers are cost-effective solutions that seem perfect for a small business that does not probably need to store a large amount of data (Marr 2017).

If data is your core business, or you need to store a large amount of data, you need a more sophisticated way to store it (Marr 2017). Sophisticated does not necessarily mean expensive, given that storage systems are designed to run on cheap hardware. Also, open-source systems can be set up for free if you know how to do it, and if you have time to develop an in-house storage system. Otherwise, a paid-for solution is more suitable (Marr 2017).

Securing data is part of data storage. In fact, data is protected by regulations such as the GDPR, that force service provider to make sure data is protected at any time. Some argue that storing data in hard disks is safer because it cannot be accessed anywhere via the internet (Marr 2017). On the other hand, data stored in clouds can only be accessed by some authorized individuals, and nowadays clouds are very safe. Also, hard disks do not allow big firms to have a flexible corporate structure (data cannot be accessed from different corporate branches and IoT devices cannot be interconnected) and they make expanding the business too expensive (you have to buy new hard disks instead of just expanding the cloud storage capacity). Besides, storing the data into clouds allows computers to only perform small parts of the overall computing task (Marr 2017). Hence, the hardware requirement is smaller and cheaper.

The data can be stored in warehouses (ordered archive which is structured and fixed) or data lakes (the data is just poured in without a structure). The advantage of a warehouse is an organization that has proven to be successful (Marr 2017). Data lakes allow firms to shape data into whatever form is needed, making the corporate activity more flexible (Marr 2017).

2.3 Analysis

As I stated in Chapter 1, data is a raw material that is to be transformed into useful applications. In other words, you must turn the data into insights and information before you can use it as a profitable asset. Analytics is what extracts value from data. What analytics you apply depends on your strategic objectives and what works for one business may not work for yours (Marr 2017).

Nowadays, companies can work with any kind of data. Even unstructured data has become quite convenient to store and analyze.

First of all, the data needs to be in a suitable format (GSMA 2018). For example, a series of temperatures is raw data. Creating a pattern of temperature over a period would be information. Raw data is not very informative itself so you need to process it. Also, data mining may occur. This process establishes links between different pieces of data (ex: how likely a customer that bought product A will also buy B, C, D) and it discovers patterns automatically (GSMA 2018). The growth of unstructured data fostered the development of data mining techniques that are possible thanks to Artificial Intelligence (GSMA 2018). AI can analyze photos, videos, texts collected from social media (for instance) and help to figure out consumer preferences.

There are many different types of data analytics. Some of the most important are Video, Text, Voice, Image, Voice, Data mining, Business Experiments, Correlation, Scenario, Regression, and Forecasting (Marr 2017). These are all about trying to make a hypothesis around data and get a meaningful result. If the outcome of analytics is a mere correlation/description, data is not useful or it has not been analyzed properly (GSMA 2018).

More advanced analytics are increasingly becoming the core of many business models. These are machine learning, deep learning, and cognitive computing (Marr 2017). Machine Learning uses advanced techniques that do not repeat the same algorithm over and over. The algorithm improves itself by learning from the feedback of previous calculations and improves its performance (GSMA 2018). These techniques are so effective that they can even anticipate human actions to prevent people from making mistakes (GSMA 2018).

The AI acts quite independently but some sort of human action is required such as instructions and interventions if something happens during the process (Marr 2017).

The best way to make analytics more cost-effective is by combining different types of data and analytics. By doing this, you can get better insights and you can verify them (Marr 2017).

2.4 Providing access to data and communicating it

Once data is analyzed, the company is ready to providing access to data to its customers, both internal and external. Insights must be easily accessible and understandable so that customers get real value from them. People across the company should also be able to use the data you have previously analyzed.

Therefore, the company has to establish some data hubs to give people access to its data (Marr 2017).

Companies like Walmart and Citibank have created large data hubs to allow their employees to access data so that their productivity increases (Marr 2017).

Other firms have chosen another business model. John Deere's, for instance, allows farmers to access data gathered by their machines as well as data collected by other farmers worldwide (Marr 2017). This business model makes information sharing extremely useful because not only the service provider generates and offers data but also customers themselves can do it (Marr 2017). This way, John Deere's can offer a larger amount of data than the one it has collected.

Managing data is also crucial for corporates' success. When data is not properly managed, it can become meaningless and consequently valueless (Marr 2017).

If data is out of date, incorrectly categorized/organized, or misused, no firm can generate value with such a problematic good.

Finally, data must be communicated to the people and machines that need it.

You can opt for traditional reporting or complex and easy-to-understand communication. IoT devices can help with that but what matters the most is communicating data to targeted users (Marr 2017).

2.5 Building a data competent organization

In order to do these activities, it is vital to set up a series of competencies, skills, and interactions in the firm (Verhoef et al. 2016).

There are two ways of developing data skills. The first one consists of building in-house competencies. Alternatively, you could opt for outsourcing.

Some of the required investments to build up a data-driven firm are storage, servers, cloud infrastructure, analytics, and applications (Verhoef et al. 2016).

Being a data-driven firm means that you do not want to deliver only what you are asked for but also to become an advisor and an initiator for your customers (Verhoef et al. 2016).

The main obstacle to the development of data skills is that there is a shortage of these in the labor market (Verhoef et al. 2016). Therefore, companies have to struggle to get the best data scientists by offering them high salaries. As a consequence, only large firms are likely to hire these competent workers because they are in a condition to offer better terms (Marr 2017). Hence, some firms are trying to overcome this problem with creative solutions such as crowdsourcing data analysis.

The five competencies that you must build up internally are:

1) Business skills
2) Analytical skills
3) Computer science skills
4) Statistics and mathematics skills
5) Creativity

You want to build teams with the right blend of these skills to produce a better outcome (Marr 2017). In particular, if data is your core business, hiring data scientists has proven to be a worthy investment. When hiring new people, you want to attract workers with a solid background in analytics. Later on, you will have to spend some time and resources to train them in the tools that your company uses (Marr 2017). That is why you also need to set up an intelligent recruiting staff and an efficient recruiting process (Marr 2017). Besides, you should be aware of the fact that most candidates did not have the opportunity to learn analytics applications at school. So, a training period must be taken into account (Marr 2017).

Rather than hiring new people, you could decide to train your existing staff. This alternative does not necessarily mean a big financial investment but it does imply an

investment in time (Marr 2017). You could rely on the free courses provided by a large number of universities on websites like Coursera (Marr 2017).

In addition, some large firms may consider hiring a chief data officer (CDO). This position should be held by someone that has both technical and business skills. Too much technical CDO would focus on tools and data analysis excessively but too few technical skills do not make it suitable for this role. A good CDO should have these qualities (Marr 2017):

1) High-level vision to understand the business strategy
2) Implementation skills, to implement the corporate strategy at every level within the firm
3) Awareness of the importance of accuracy, security, and privacy
4) Good at identifying business opportunities
5) Data-driven
6) Responsible for recognizing the value in monetizing data (see Chapter 3)

To sum up, the right people should ideally have four types of capabilities (Verhoef et al. 2016).

1) Analytical. For instance, conceptual and numerical skills, statistical modeling.
2) Business sense. Being passionate about the business, industry-specific knowledge, organization, leadership, problem-solving.
3) Data and tools. Some examples include R, Matlab, SQL, Hadoop.
4) Communication and visualization. Translating insights into decisions, storytelling skills, visual art design.

Small firms could consider a CDO as unnecessary or possible and decide to rely on an external consultant or onto the internal management (Marr, 2017).

When building up data competencies internally is not possible/convenient, you could outsource your data analysis activity (Marr 2017). There are more and more data service providers in the market. Each of them can even specialize in a specific industry (ex: data for retail online).

Industry-specific providers are becoming an interesting alternative to big providers (Facebook, Amazon, IBM) because they can offer a personalized service. Although big

firms usually have larger datasets, they might not be the best option for you (Marr 2017). When choosing the ideal provider, you should consider the 5 data skills that I mentioned above. Your data supplier should also possess them. To make sure your provider gives you the best possible service, it is recommended to set up a data strategy first, so that the provider can personalize the service properly (Marr 2017).

If you do not want to lock your business with one provider only, crowdsourcing data analysis could be a good option. Kaggle is one of the most successful data analysis platforms[16]. It acts as a middleman: firms bring the data, set a problem to solve, and compensation. Data scientists offer the best solution. It is a good alternative for those that want to access external data skills and solutions (Marr 2017).

2.6 Hidden costs

What I previously described are the main activities (and costs) that one must bear to set up a data firm. However, some other costs may influence the success of the data business. These additional costs might not be monetary costs but they are equally important in the business model.

According to the data management company Precisely, there are 5 hidden costs of Big Data[17].

- Inefficient data integration. As I said before, data must be transformed and integrated. If the process is redundant, too manual, or inefficient, you waste staff time and unnecessary infrastructure.
- Networking data costs. If you move terabytes of data, the networking devices that you must use can be quite expensive.
- Unnecessary backups. If you do too many backups, you are not acting with a cost-effective approach. The backup activity should be optimized[18].
- Low-quality data. If data contains many errors and/or it is not informative, storing and analyzing it produces useless costs.

[16] Kaggle: Your Machine Learning and Data Science Community
[17] 5 Hidden Costs of Big Data - Make sure you're not missing these (precisely.com)
[18] Data Backup vs. Disaster Recovery: Yes, There's a Big Difference (precisely.com)

- Uninformed employees. The human capital required for data business should be properly educated and trained. Otherwise, a lot of data quality errors and unnecessary activities may occur.

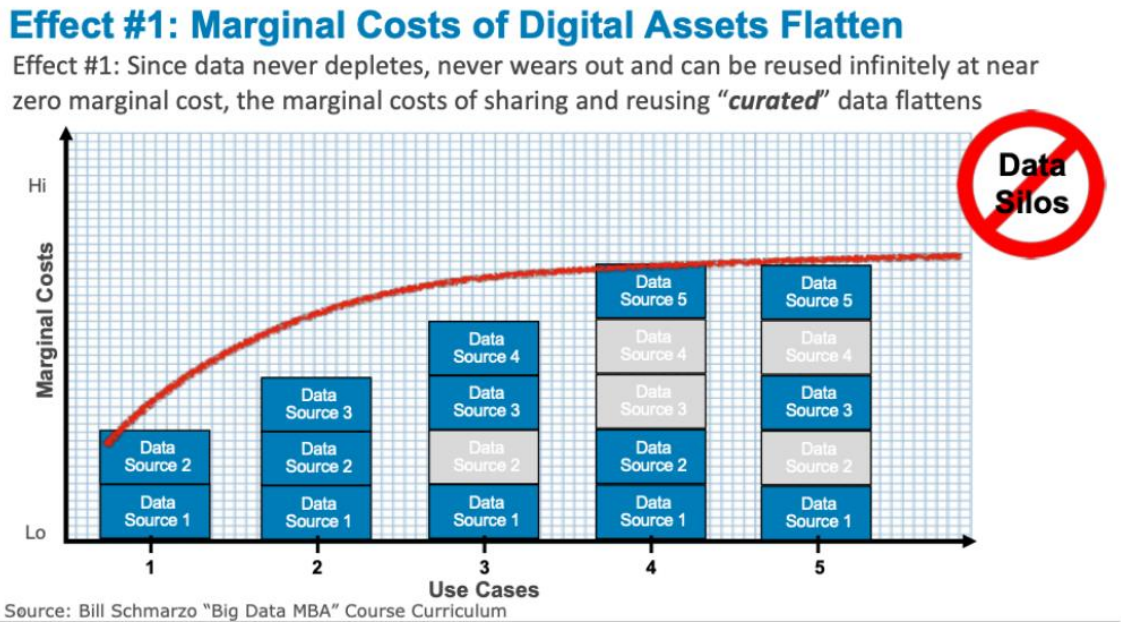
2.7 Data and the economy of scale

The digital economy is based on user interactions in networks (Parker et al. 2016).

The larger the number of users, the larger the value that can be generated through the network. Data follows the same pattern[19].

Besides, Bill Schmarzo indicates how the marginal cost of digital assets diminishes when production increases[20].

Graph 2.7 Marginal costs of digital assets



Also, the value of these assets increases with the increase in usage and interactions. Therefore, Schmarzo suggests that a firm should store data into data lakes so that they can use it across multiple cases. He describes this characteristic of digital assets as a new form of the economy of scale, namely "Economy of Learning".

This new economic model is even more powerful than the economy of scale because not only the marginal cost decreases (leading to an increase of margin) but the

[19] https://doi.org/10.1007/s40844-019-00157-x

[20] (1) In Digital Transformation, Economies of Learning More Powerful than Economies of Scale | LinkedIn

economic value of goods produces increases in usage. Therefore, the size of a digital business is one of the vital factors in assessing its profitability. Additionally, not only the asset value increases with usage but it accelerates even more with analytics[21] because this process increases its accuracy and complexity. I have described the profitability of big data more accurately in Chapter 4.
Hence, considering the cost structure of data, the whole data business model rests on an economy of scale principle: the bigger, the better.

To summarize, before even starting a data business (or turning a business into a data business), you must consider all of the costs that the data value chain implicates. If these costs are too big, your strategy may not be profitable. In other words, if the corporate strategy is not cost-effective, data is not a valuable asset (Marr 2017).
The real cost of data is the operation, management, and integration costs[22]. You must also consider storage and security costs, albeit storing the data is becoming cheaper and cheaper. The whole business lays on an economy of learning principle.
Finally, you must consider hidden costs such as opportunity costs and inefficiency.

[21] (1) In Digital Transformation, Economies of Learning More Powerful than Economies of Scale | LinkedIn
[22] The Big Cost Of Big Data (forbes.com)

Chapter 3 - Business models

In this chapter, I have described the multiple business models that a company can opt for to benefit from data.

Firms can use the data for their own use or trading. In other words, they can improve corporate efficiency by using the data or they can sell data/insights to other companies (GSMA 2018).

3.1.1 The economics of networks

Before the explosion of IoT, data visualization was the most common business model of data. Companies used to offer data visualization to their customers. Nowadays, the most common and profitable business model is that of making data and IoT devices interact (Faroukhi et al. 2020).

Considering the initial investments required to enter the data industry, the main concern companies have is trying to identify an efficient data monetization approach (Faroukhi et al. 2020).

The newest big data business models rely upon the sharing of information on complex ecosystems (Faroukhi et al. 2020) based on the interaction of multiple agents on network platforms (Parker et al. 2016).

Data business models are characterized by the network effect in which machines play a crucial role (Wagner 2020). AI boosts the utility of the economics of networks by automatizing the “learning by using” model (Wagner 2020).

When developing a data monetization strategy, you should consider that data is a valuable asset that you can collect by letting the users interact on platforms. Hence, your monetization policy should not lead people to avoid the platform and make the platform being profitable at the same time (Parker et al. 2016).

3.1.2 Economics of learning

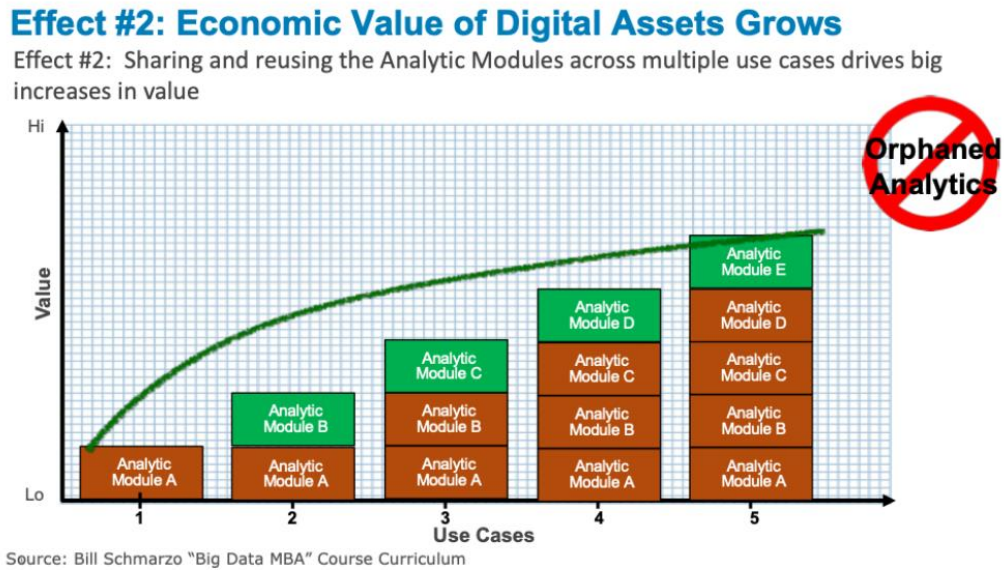
As I explained in the previous chapter, data firms mostly bear the fixed costs necessary to build up their infrastructure and analyze data. Therefore, these companies are likely to be profitable only when they generate large revenue. Once a sufficient quantity of

revenue is generated, they benefit from the economy of scale, meaning a saving in marginal costs due to the increased production[23].

Besides, the former Dell Chief Technology Officer Bill Schmarzo claims that the transformation of digital assets causes the "Economics of Learning" effect[24]. Not only the business becomes more profitable because marginal costs decrease with the increase of production but the value of the output also increases. This happens because data is both non-rivalrous (it is not consumed in the usage process) and an experience good, so its value is generated by analytics and usage (MIT Technology Review, 2016).

Schmarzo also underlines the importance of proper data storage. Data lakes, unlike silos, are collaborative platforms that facilitate data usage and sharing across organizations. Silos remain under the control of a single corporate department. Since the value of data is linked to the economics of learning, the larger the number of users, the better.

Graph 3.1.2 Economic value of digital assets



[23] Economies of Scale Definition (investopedia.com)

[24] In Digital Transformation, Economies of Learning More Powerful than Economies of Scale | LinkedIn

3.1.3 Data and circular economy

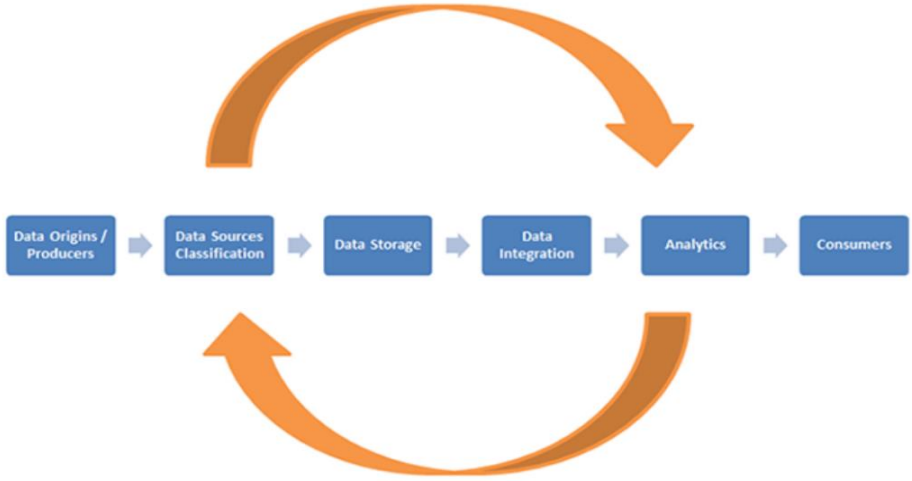
In addition, the data business model has a strong connection with the circular economy (Geissdoerfer et al. 2020).

First of all, data collection and analytics make the supply chain more efficient and prevents firms from wasting resources[25]. It also allows cities to have better resource management and be less polluted[26].

Secondly, one of the most important aspects of the circular economy is the need for dematerializing (Geissdoerfer et al. 2020). Data analytics reduces the need for hardware, it offers a service instead of a product and helps to rationalize the demand (Geissdoerfer et al. 2020). The higher the rationality of goods/services demanded, the lower the waste.

Finally, the data value chain has many similarities with the circular economy. As I mentioned before, data can be re-used multiple times to increase its value. Therefore, the data business model has a feedback loop in the value chain, from data classification to analytics. This pattern is somehow similar to the circular economy's in which products do not easily become waste and can be reused multiple times to feed the analytics process. In a normal circular organization, optimized materials reduce the concept of waste. In the data business, like Schmarzo claims, efficient storage can do that.

Graph 3.1.3 Data and circular economy



[25] Big data and circular economy - The revolution will be circular | Reuters Events | Sustainable Business
[26] Big Data serving the circular economy | Living Circular (veolia.com)

3.2.1 Data monetization

The last step of the data value chain[27] is the exchange of data (GSMA 2018).

It consists of using the data internally or externally by turning it into a valuable asset.

Big data firms usually have higher benefits from data monetization because once processed and analyzed, the data can be used multiple times (Faroukhi et al. 2020).

All of the firms can monetize data. Normal firms can improve business decisions and operations. Data firms use data as a core business asset (Marr 2017).

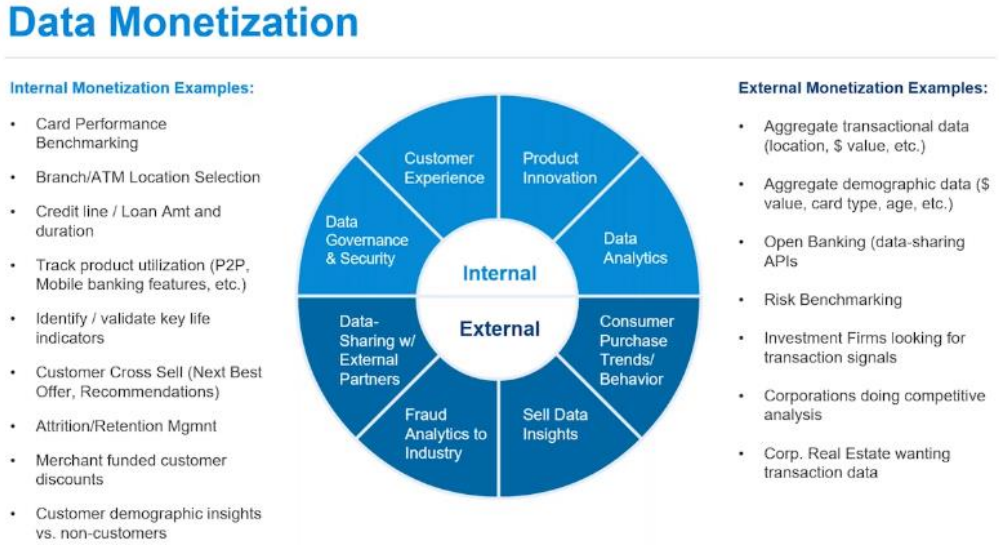
Data can be used in internal or external activities.

Internally, data can provide useful insights, help the firm to deliver better services, improve the customer experience, and offer better products.

There are several external applications such as the sharing of data to partners and the sales of data/insights[28].

These two different monetization methods are also called explicit monetization when you sell your data, and implicit when you use your data to enhance your data-based products (Faroukhi et al. 2020). During the 2000s, monetization was mostly implicit while nowadays the rise of IoT made explicit monetization gaining popularity (Faroukhi et al. 2020).

Graph 3.2.1 Data monetization



[27] After data generation, collection, analytics (see Chapter 2). The data value chain is the business of collecting and exploiting all types of data (see Chapter 2).

[28] Transform Your Payments Data into Revenue: ARM Insight Talks Safe Synthetic Data Monetization | PaymentsJournal

3.2.2 Using data to improve business decisions

Data can provide useful insights to make better business decisions. You can have a deeper understanding of your customer's needs and the market thanks to data analytics. It can also help you to ask yourself useful questions about your business (Marr 2017). According to Marr, there are 4 categories of questions that data helps to answer.

1) Questions related to customers, market, and competition. For example, data informs you about customer behaviors so you can create better segmentation and pricing.
2) Questions related to finance. You may know what your key sales are, profit trends, the most and least products/services, and cost-saving opportunities.
3) Questions related to internal operations. These are related to the supply chain, the suitability of products and machinery, the efficiency of operations in different business units.
4) Questions about your people. Since employees are a crucial asset (Marr 2017), you should know their core competencies, efficiency, productivity. Also, you may want to know if your recruitment channel is successful and whether there are reasons why employees leave the firm.

You could allow all of your people to access data in order to improve their awareness. Considering that data can be interpreted in different ways, you need to communicate it properly with suitable data visualization tools such as Tableau (Marr 2017).

The evidence shows that the impact of quality data on the decision-making of firms is remarkable. Good data can increase the benefits realization by 10% by making a project's assumptions and predictions more accurate (Gartner 2011).

3.2.3 Using data to improve business operations

Data can also be used to improve corporate efficiency. Data-enhanced operations are less dependent on human labor because they rely on machines. The interaction between sensors, IoT devices, and machine learning is crucial. Data can mainly improve everyday processes and customer offerings (Marr 2017).

1) Manufacturing process. Sensors installed into manufacturing equipment can capture data that helps to monitor the efficiency of machines and the product's performance. The feedback is then received by computers that analyze it. Maintenance teams can fix a problem before the machines have one (Marr 2017).
2) Warehousing and distribution. Supermarkets use cameras and sensors to monitor the quality of their fresh food so that they never sell rotten fruits and vegetables. Large retail companies such as Amazon can stock their products more efficiently to reduce shipping costs (Marr 2017).
3) Business process. Insurance and credit card companies can detect fraud if they receive a warning in case of unusual purchases and operations (Marr 2017). In the last 20 years, IT has been used to automate many business processes, each year to lower costs. Data quality helps to maximize process efficiency (Gartner 2011).
4) Sales and marketing. 58% of Chief Marketing Officers claim that search engine optimization and marketing is where big data had the largest impact on corporate marketing programs[29]. Data can help you to develop personalized recommendations for customers (ex: ads on social media) and dynamic pricing (Marr 2017).

3.2.4 How to improve customer offering with data

Firms can create new customized products/services or improve the existing ones thanks to data analytics. For instance, the agricultural manufacturer John Deere allows farmers to interact in a platform where they can share information about their own fields (Marr 2017). Hence, the other farmers can access this data to decide where and when it is better to plant[30].

In other cases, data can be used to offer better products or to create new technologies. Hypersurfaces, for example, is a technology that relies on AI that is

[29] Ten Ways Big Data Is Revolutionizing Marketing And Sales (forbes.com)
[30] 5 Minute 'Big Data' Case Study: John Deere | LinkedIn

capable of capturing data with sensors and transform them into sounds or digital commands[31].

Graph 3.2.4.1 illustrates the most common use cases for big data. As I stated above, analyzing customers' behaviors, improving operations, preventing frauds, offering better products/services, and warehouse optimization are the main uses for big data.

Graph 3.2.4.1 Data use cases

The advisory firm McKinsey & Companies says that fast-growing sales organizations use data analytics more effectively, while others still struggle doing it[32].

Graph 3.2.4.2 shows a correlation between data usage and sales growth rate. Even though this does not imply causation, it is still to be noted in light of the possible data uses that I previously explained.

[31] Core Technology — Hypersurfaces
[32] Unlocking the power of data in sales (mckinsey.com)

Graph 3.2.4.2 Fast-growing organizations

The research and advisory company Gartner claims that there is a good correlation between the use of IT systems and labor productivity (Gartner 2011). However, there are large variations across companies. The main reason why this occurs is poor data quality. Poor data quality may cause a 10% loss of savings related to the use of IT systems (Gartner 2011). Hence, what matters is good data, not all data.

3.2.5 Selling data and insights

Customers' expectations and technological advancements (ex: IoT) have transformed the supply chain into an interconnected system based on collaboration, rather than control procedures (Faroukhi et al. 2020).

Hence, data sales often occur on platforms. One of the most challenging aspects of this monetization is to benefit from data analytics without destroying the network effect that you have worked so hard to create (Parker et al. 2016). There are four different ways of monetizing the value of the platform for the users (Parker et al. 2016).

1) Charging all users. Generally speaking, this mechanism would discourage participation but in some cases, membership quality can be guaranteed more easily this way (ex: exclusive clubs)

2) Charging one side while subsidizing another. This system works only if one category of users evaluates the opportunity to join the platform as very profitable.
3) Charging most users full price while subsidizing stars. It could be beneficial if some particular users (stars) have the capability of attracting many others.
4) Charging some users full price while subsidizing price-sensitive users. It is difficult to predict what the most price-sensitive users are.

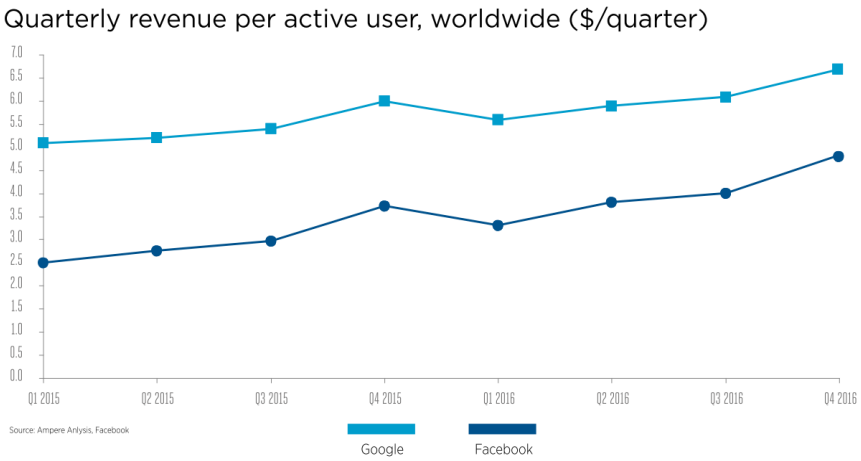
The Chinese company Alibaba, for example, adopted a membership fee as a monetization approach (Parker et al. 2016). It combined this strategy with incentives to salespeople who convinced others to join the platform (Parker et al. 2016).

Firms can sell both data and insights (GSMA 2018). The value of data is given by quantity, quality, and uniqueness. Another important value driver is the reputation of the data supplier because if you know where the data comes from, you know that you can trust the data (GSMA 2018). Data format also matters since the dataset must be machine-readable (GSMA 2018).

Trading insights, without the underlying data, might also be profitable given that companies seek information to base their corporate strategy on (GSMA 2018). Insights may help companies to make more accurate and suitable advertisements (GSMA 2018).

Graph 3.2.5 shows that large enterprises like Facebook and Google benefit a lot from this business because they have been able to increase the revenue generated by each (active) user thanks to the increase of online advertising policies (GSMA 2018).

Graph 3.2.5 Revenue per user



Before even trading your data, you should consider potential opportunities to create partnerships with third parties that might be interested in your data (Marr 2017). For example, a car manufacturer could partner with an insurance company to provide data on accident rates, average speed, and other information that could help insurance companies to lower their financial risk (Marr 2017). Since the data has not a fixed value but it also depends on who owns the data, creating the right partnership could make your data more valuable without the need for additional analytics (Marr 2017). In other words, you can have higher revenue and equal costs: this would make your firm more profitable.

The evidence shows that users are usually happy for companies to use and benefit from their data because this operation generates higher value for consumers (Marr 2017) and increases transparency[33]. This is evidenced by the increasing amount of user-generated data on social media (ex: over 300000 stories posted on Instagram in a minute)[34].

3.3.1 Measuring the impact of data on business models

Nobody has properly figured out how to assign value to information. However, data has a significant impact on a firm's performance. Even though companies have no obligations to indicate data in the Balance Sheet as an asset[35], it is crucial to quantify

[33]Contro il virus abbiamo un disperato bisogno di dati - Fabio Sabatini - Internazionale
[34]User-generated content - Statistics & Facts | Statista
[35] Why is Data Missing from the Balance Sheet? | CFO.University

the impact of data on companies. Otherwise, you cannot put a monetary value on data.

IBM also suggests that there are 4 types of different values.

1) Volume-based value. The higher the number of insights, the better.
2) Velocity-based value. The more customers you can get into your platform, the higher the value proposition that you can offer.
3) Variety-based value. The larger the variety of customers, the better-informed decisions you can make.
4) Veracity-based value. The most consolidated and consistent the data, the more likely you will make the right decisions.

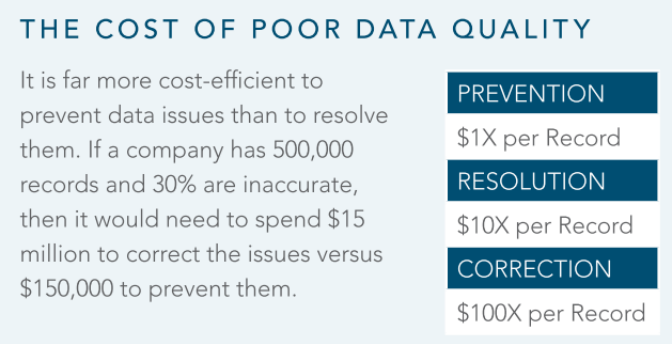
Before assessing the impact of data on firms, you need to identify some Key Performance Indicators (KPIs) that you think data is affecting[36]. Some examples of KPIs include cost-savings, cycle time, maintenance costs, production volume, capacity utilization, sales, revenue by channel, churn rate, customer lifetime value, acquisition retention costs, average fraud score, order approval rates, turnover rate, the average cost[37]. Measuring the impact has also proven to be good for sales growth, efficiency, investment decisions, and marketing building[38].

Since poor data quality drains a company on average $8.2 Million annually (inefficiency, missed sales, unrealized opportunities), and 88% of firms use data inefficiently, it is necessary to manage the data properly (DNB 2015). The data analytics company Dun&Bradstreet claims that good data management is a choice based on 5 tenets.

1) Data quality is a business issue, not an IT issue. IT plays a consultative role in recommending options to improve data quality.
2) Data governance is essential. Preventing issues in a dataset is 100 times cheaper than correcting them. Low data latency and duplicates in databases are indicators that show that a good data management policy has been adopted.

[36] A data-driven approach is a vital investment for organizations (villanova.edu)
[37] KPIs to Measure ROI from Data Analytics Initiatives | Tiempo Dev
[38] Why Measuring Impact Is Good For Business (forbes.com)

Graph 3.3.1.1 The cost of poor data quality



3) Data has good quality if it is accurate, complete, up-to-date. Also, it should not have conflicting information across the organization. High-quality data increases the revenue, helps to serve customers properly, reduces spending, and helps the company to have a competitive advantage. These factors translate to a higher cash flow.
4) Sales and marketing are linked to data quality. Deliverability, response rate, win rate improve when using high-quality data.
5) High-quality data boosts corporate's ROI by driving revenue growth, allowing informed decision-making, adding value-added to the business process, and improving financial profitability.

Moreover, the University of Texas at Austin analyzed the business impact of effective data. The study describes 150 Fortune 1000 firms (median firm: 36000 employees, $388000 in sales per employee) and lists three series of performance measures (Barua et al. 2012):

1) Financial impact
2) Customer-focused impact
3) Operational impact

In this study, five data attributes were considered: quality (accuracy, scope, timeliness, recency), usability (data must be processable), intelligence (what trends, patterns, recommendations they give), remote accessibility (authorized remote use), sales mobility (data that can be effectively sold/monetized).

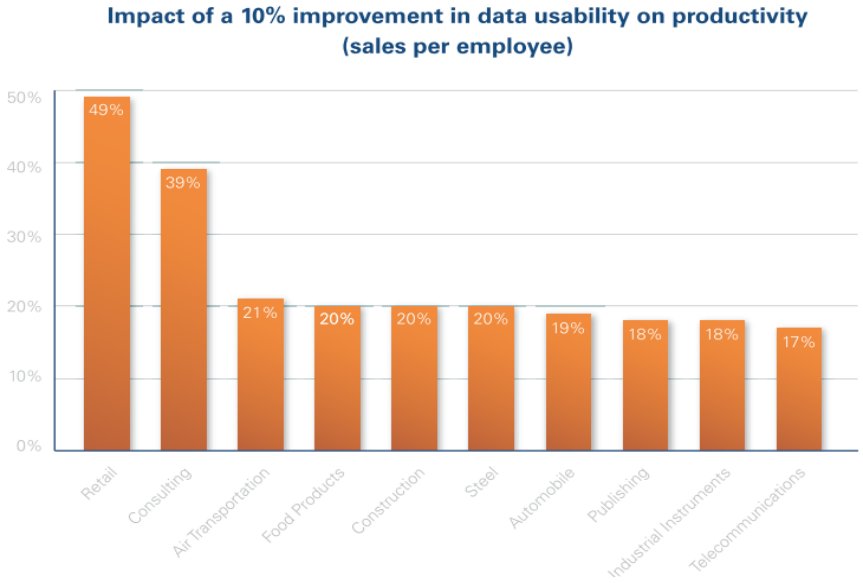
The most affected indicators are four[39], namely the productivity of employees, ROE, ROIC, ROA.

Impact on the productivity of employees

A 10% increase in data usability translates to $55900 additional sales per employee annually. This impact is more significant in some industries. For instance, the retail industry reports the highest sales increase (49%). As Graph 3.3.1.2 illustrates, the increase is also remarkable in the consulting industry (39%) while in the others it is between 17% and 21% (Barua et al. 2012).

Although some industries seem to benefit more from an increase in data usability, they all experience a substantial increase in their labor productivity when high-quality data is used (Barua et al. 2012).

Graph 3.3.1.2 Impact of data on productivity



Impact on Return on Equity (ROE)

A 10% increase in data usability means a 16% increase in ROE. Given that the median ROE of the firms analyzed is 21%, this is an outstanding result (Barua et al. 2012).

[39] Regarding the 150 Fortune 1000 firms (median firm: 36000 employees, $388000 in sales per employee) analyzed in the study

Impact on Return on Invested Capital (ROIC)

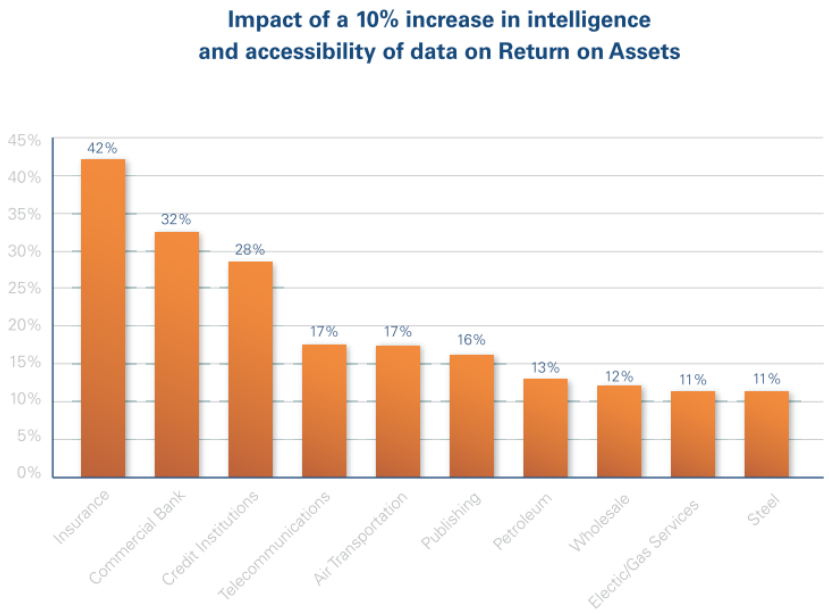
The median ROIC in the sample is 18%. The 10% increase in data usability makes ROIC grow by 1,4% (Barua et al. 2012). ROIC is computed as:

$$\frac{net\ income - dividends}{invested\ capital}$$

Impact on Return on Assets (ROA)

In this study, the median ROA is 10,25%. ROA increases by 0,7% with the aforesaid increase in data usability. Graph 3.3.1.3 shows that the increase is significantly higher in Insurance, Commercial Banks, and Credit institution firms. On average, ROA grows more in the financial industry (Barua et al. 2012).

Graph 3.3.1.3 Impact of data on ROA



3.3.2 Value creation

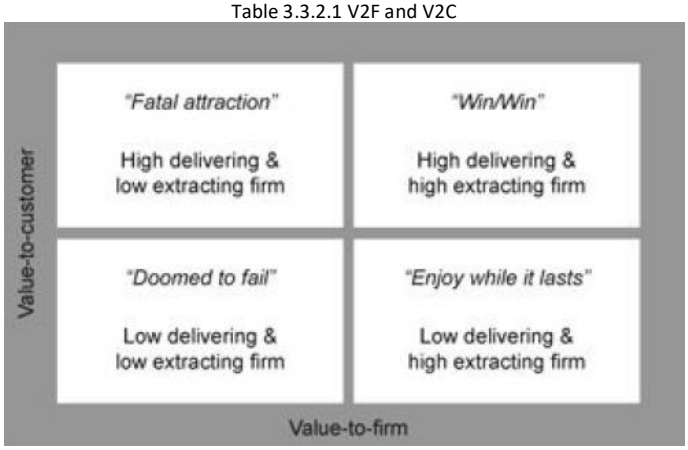
The ultimate goal of a big data strategy is value creation (Verhoef et al. 2016). The value can be value to the customer (V2C) or value to the firm (V2F).

Firms can capture value only if they have been able to deliver a satisfying value proposition to customers before (Verhoef et al. 2016).

This creates a win-win strategy. Typically, firms tend to care about the value to the firm only. The fact that V2F is dependent on V2C creates a natural incentive that makes firms more customer-driven (Verhoef et al. 2016).

As Table 3.3.2.1 explains, the win-win strategy holds only if V2F and V2C coincide. On the contrary, if one of the two values goes missing, one of the two parties will lose the incentive and the business strategy will come to an end (Verhoef et al. 2016).

Table 3.3.2.1 V2F and V2C



A third value exists. It is called Value to Society (V2S) and it is a value created by the firm, not for a single targeted customer but the society as a whole. Companies like P&G use a tactical approach and aim to show that their value proposition has several advantages for the whole community. By doing this, more people will be inclined to become customers and will consequently get V2C. Hence, V2S can be seen as a driver of V2C (Verhoef et al. 2016).

V2F and V2C can be measured and classified into 3 different categories: the market, brand, customer (Verhoef et al. 2016).

Table 3.3.2.2 reports some of the most important value metrics. For the customer, these include the product itself, the attractiveness of the brand, and satisfaction.

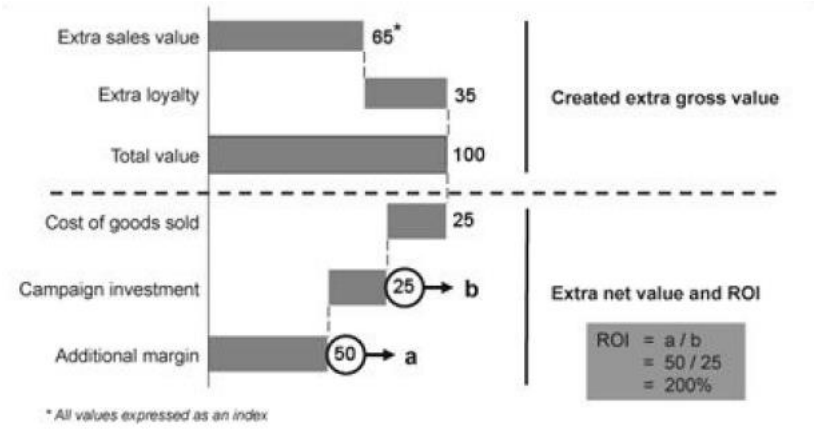
On the other hand, firms extract value from V2C if they increase their market, penetrate it with their brand, and if they can maintain the customer forever (Verhoef et al. 2016).

Table 3.3.2.2 V2F and V2C 2

	Value-to-customer ☺☺☺	Value-to-firm €€€
Market/ category	• Product awareness • Product attractiveness • Product uniqueness	• Market volume/size • Market growth • Number of competitors • Market concentration • Trial rate • Repeat volume
Brand	• Brand/advertising awareness • Brand association • Brand consideration • Brand preference • Brand linking • Brand likes/comments	• Brand penetration • Brand sales • Brand/market share • Brand repurchase • Brand equity
Customer	• Customer satisfaction • Net promotor score (NPS) • Customer Effort score • Reviews: volume & valence	• Customer lifetime value (and components) • Customer engagement value • Path to purchase • Marketing ROI

Considering that data analytics helps to make more efficient marketing campaigns and, as a consequence, to increase the number of customers (and their CLV), the impact of data can be measured as the extra value generated by its usage (Verhoef et al. 2016). One must also take into account the increase in expenditure that firms bear to analyze and use data, included in the Cost of goods sold. Hence, the impact (value) of data is given by the net value generated by its usage (Verhoef et al. 2016).

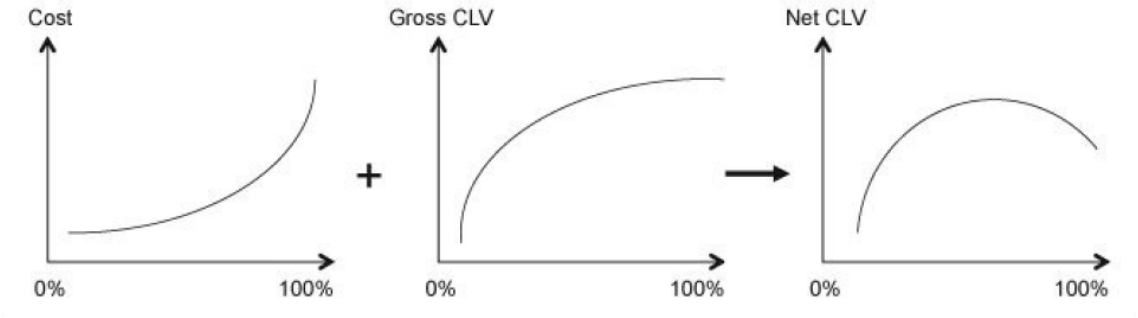
Graph 3.3.2.3 Extra value



The importance of data quality in business deserves another mention. The increase of data quality is costly, and these costs rise in a non-linear function with the increase of data quality. At the same time, the marginal CLV function may have a diminishing trend once a certain CLV is achieved. Therefore, it is crucial to assess the optimal level of data quality below the 100% score (Verhoef et al. 2016).

In other words, since the value of data is measured as a net increase of profitability generated by its usage (extra revenues – extra costs), improving the quality of your data might not be profitable. If the Net Present Value[40] of data usage is lower than 0, you should consider not enhancing data quality.

Graph 3.3.2.4 Customer Lifetime Value



Customer lifetime value (CLV)

As I mentioned before, data usage can help to understand customers' needs. A company that uses data efficiently is more likely to create personalized products/services, set an adequate price, and detect changes in customers' preferences. These aspects make data-driven firms attract more customers, as well as keeping the old ones, more easily than others (Verhoef et al. 2016). Since customers contribute to increasing the earnings of the firm, it is possible to compute the value of each customer. This is called Customer Lifetime Value (CLV) and it is computed as follows[41].

$$Margin \times \frac{Retention\ rate}{1 + Discount\ rate - Retention\ rate}$$

3.3.3 Big data business model maturity index

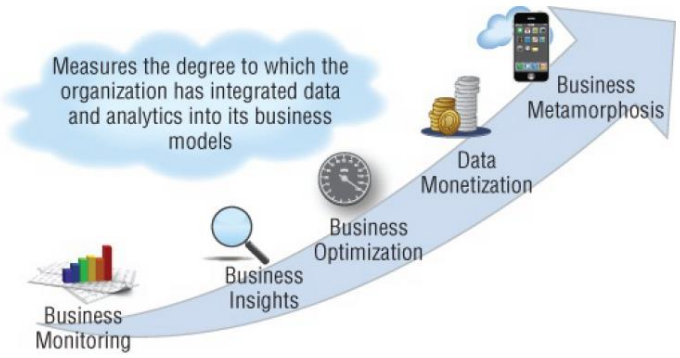
The Dean of Big Data Bill Schmarzo created the big data business model maturity index which is a tool that helps companies to measure "where they sit today with respect to their adoption of big data" (Schmarzo 2015).

[40] Net Present Value (NPV) (investopedia.com)
[41] Customer lifetime value - Wikipedia

The concept of this model is that most organizations tend to find themselves in one of the 5 steps, namely Business Monitoring, Business Insights, Business Optimization, Data Monetization, and Business Metamorphosis (Schmarzo 2015).

The first three phases are internal process optimization. Schmarzo claims the value of data emerges when monetization starts (Schmarzo 2015).

Graph 3.3.3 Data business model maturity



3.4 Competitive differentiation

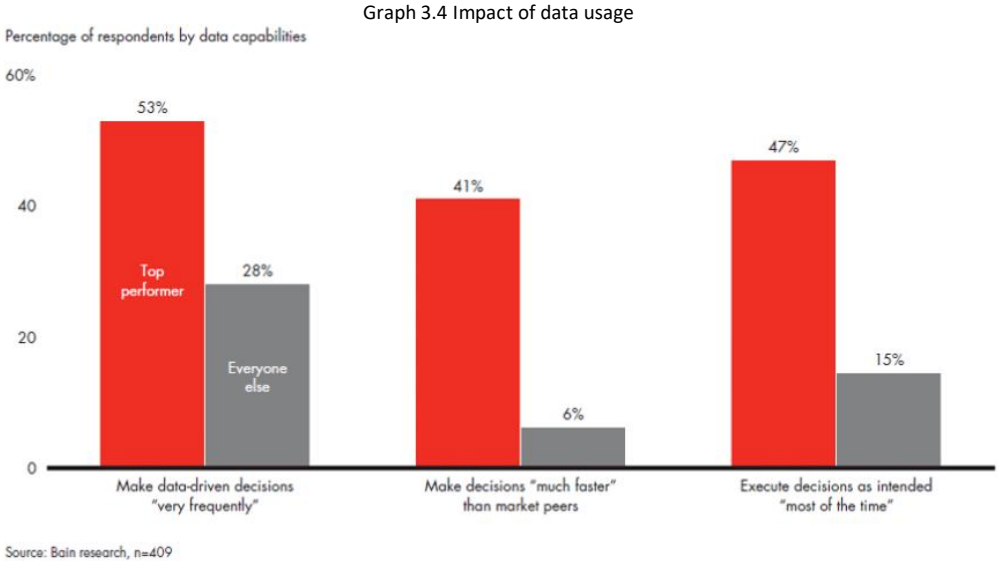
Competitive differentiation is achieved when a company offers products/services different from what competitors offer in terms of uniqueness (Schmarzo 2015). You are also different if you have different operational capabilities. Technology can make you stand out. For example, Google's differentiation strategy relies on Ad Serving, Apple on iTunes, Netflix on movie recommendations, Amazon on the supply chain. Big data usage helps companies to drive a change. However, it is not something that disrupts business models itself but rather a tool that companies can use to change their business models (Schmarzo 2015).

Bain & Company studied over 400 companies around the world, with revenues of more than 1 $Billion. Only 4% of these are good at analytics (Bain & Company 2013). Their analytical skills make them stand out because:

1) They are twice as likely to be top-performing in their industries
2) They are three times more likely to make decisions as intended
3) They make decisions five times quicker than the others

These indicators underline the importance of comparing indexes within the company and comparing them to some benchmarks such as the average industry performance (Bain & Company 2013). By doing this, you can figure out not only how your firm has improved with data usage but also how it but also how it is positioned in its market/industry thanks to data usage. As Bain & Company's study indicates, quality data makes firms more efficient and profitable. The improvement attributable to data usage is described in Graph 3.4 as the ability to make decisions faster and being able to execute them as intended.

Graph 3.4 Impact of data usage



3.5 Bloomberg analysis

The Bloomberg analyst Mandeep Singh claims that data can help to build business models that generate recurring sales to customers[42]. This is very beneficial for firms because having recurring (and stable) revenue helps them to predict what their future cash flow will be. This is particularly useful when using the DCF valuation approach[43]. Furthermore, companies like Microsoft have experienced a sharp increase in profit due to the implementation of a cloud-based business model[44]. Besides, the acquisition of

[42] Bloomberg Intelligence report 11/12/2017, Mandeep Singh
[43] Discounted Cash Flow (DCF) Definition (investopedia.com)
[44] Bloomberg Intelligence report 24/10/2016, Caitlin Noselli

LinkedIn, which is based on a network effect business model, supplied Microsoft with a considerable amount of useful data that can be integrated into its business model[45].
In addition, Bloomberg reports that about 65% of the $143 Billion big data market [46] comes from large enterprises. Small firms mostly use bundled offerings from the larger ones.

Table 3.5 Data revenue by company size

Big Data Revenue Split by Company Size ($B)

Vendor Revenue by Office Size ($B)

Employees	2017	2018	2021	CAGR 2016-21
10-99	$11	$12	$17	11%
100-499	$21	$24	$32	11%
500-999	$16	$18	$25	11%
1000+	$92	$103	$147	12%

Source: IDC

Bloomberg

3.6 Successful data business models

In this section, I have reported some of the most successful examples of data strategies and business models. These successful strategies have been described by Bernard Marr, in his 2016 book "Big Data in practice".

Netflix: big data improves decision-making

The screenwriter William Goldman once said: "Nobody knows the least goddam thing about what is or isn't going to work at the box office"[47]. Netflix proved him wrong.
Netflix started as a DVD-mailing business and later on changed its core business into a streaming website. This shift allowed the firm to collect a larger amount of data on customers' preferences to be used as fuel for its algorithm that helps to predict what movies would customers enjoy based on their previous activity. Happy customers are likely to renew the subscription and, so, to generate recurring sales (Marr 2016).
Secondly, Netflix has moved towards contents creation thanks to data usage.

[45] Bloomberg Intelligence report 05/05/2017, Andrew Eisenson
[46] The size of big data market is computed as the aggregate revenue of data firms (Statista). Data companies are those that use data as a core asset in their business model.
[47] https://www.bernardmarr.com/default.asp?contentID=1093

For instance, data had shown that subscribers appreciate movies directed by David Fincher and starring Kevin Spacey. That is why Netflix decided to commission the production of the tv show *House of Cards*, one of the most successful shows on the platform (Marr 2016).

Finally, Netflix also analyses the time customers spend using the service (Marr 2016). This helps the company to predict who is likely to cancel the subscription (lower usage, higher probability).

Facebook: how platforms work

Facebook has over 2 billion active users[48] so it can access a huge amount of data. This data is also very personal because the users themselves provide it by sharing information about their age, hobbies, interests, and so on (Marr 2016).

Other companies pay Facebook to advertise their products/services to users that match the demographic profiles of their customers. This system allows firms to undertake more efficient marketing campaigns (Marr 2016).

Facebook took advantage of the platform mechanism to create a huge database that has an enormous value for the advertisers (Marr 2016).

IBM: Machine learning

Computers are quick and infallible. However, they need inputs given by humans that can make mistakes. In the past, computers could only process the inputs that humans gave them. IBM figured that this flawed system needed to be fixed (Marr 2016).

Considering that computers can learn much more quickly than humans, they developed some useful algorithms that allow machines to teach themselves something so that they become really infallible. The fuel of this machine learning process is data (Marr 2016).

Thanks to the internet, IBM had collected a considerable amount of data and used it to give computers the largest possible dataset to extract information from. It also keeps the data up-to-date (encyclopedias, scientific studies, news, statistics) so the machines can update (Marr 2016).

Machine learning makes computers work at their full potential (Marr 2016).

[48] • Facebook MAU worldwide 2020 | Statista

Amazon: data in the retail industry

Amazon was originally an online bookshop. Later on, it became one of the largest online retail platforms and web services providers.

Amazon based its entire business model on a “recommendation engine” technology. In other words, Amazon knows what customers want, when they want it, basing its assumptions on their previous activity on the platform (Marr 2016).

The problem of a large online marketplace is that customers have a huge range of different goods to choose from. This could make them feel overwhelmed because customers become data-rich but insight-poor (Marr 2016).

Therefore, Amazon used recommendation technology to help customers choose. This system becomes more and more efficient over time because it relies on previous activity. Hence, customers also have the incentive to be more active. Besides, the larger the number of users the better, because Amazon can also recommend items purchased by similar users (Marr 2016).

Finally, Amazon launched the Amazon Prime streaming video, book rental, and it started selling big data to advertisers. This put it in competition with Netflix, Google, Facebook (Marr 2016).

Chapter 4 - Data valuation

"Customer data is more valuable than store walls." [49]

Luc Vandevelde, Managing Director, Promodes Carrefour

According to the advisory firm Gartner, the biggest challenge in the Big Data industry is determining how to get value from data[50]. This shows the importance of establishing a data valuation model.

In the first part of this chapter, I have described the key value drivers of data and the most useful valuation methods.

Secondly, I have reported some interesting cases of M&A and IPOs that show some peculiar characteristics of the Information Industry, the valuation, and the pricing of Big Data companies.

After, I have analyzed 151 Big Data firms. After doing that, I have presented the most relevant financial ratios and valuation multiples that I believe should be taken into account before starting a proper valuation of data companies. I have also presented some data companies' reports that show how the value of data could be viewed when analyzing corporate financial statements.

Finally, I have explained some factors that investors should take into account when analyzing data firms.

4.1.1 The five Vs of Big Data

There are five elements that transform Big Data into a business asset[51]. Volume, velocity, variety, veracity, and value[52].

- Volume. It is the base of the pyramid. Big Data has to be big by definition.
- Velocity. The speed with which information flows gives companies a competitive advantage. it might be more beneficial to have a lower amount of

[49] Valuing information as an asset, SAS 2010
[50] Gartner - Big Data Industry Insights 2015
[51] The five V's of big data | BBVA
[52] The 5 V's of big data - Watson Health Perspectives (ibm.com)

data that flows quickly than a lot of data at a low speed. The velocity of data could be a crucial aspect in some industries (ex: healthcare).

- Variety. Data can be obtained from different sources. The importance of one source over the others depends on the business model of the firm. Besides, knowing the data generator also matters (ex: who searched for a piece of information on Google).
- Veracity. It is equivalent to quality. Is data reliable? Is it clean? Is it informative?
- Value. This "V" is at the top of the pyramid. If the other "Vs" are not met, it is unlikely that the data will be used as a valuable asset. The data should allow companies and organizations to get a tangible benefit from its use.

These elements generate 5 different types of value. Hence, one must take into account all these different values before estimating the value of data. The total value could be considered as the aggregate value or, in other words, the sum of these values[53].

4.1.2 Key-value drivers

According to PWC, there are 8 value drivers of data (PWC 2019).

1) Exclusivity. If the dataset is unique, the data will probably be more valuable.
2) Timeliness. If the data is up to date (still useful), it could have value.
3) Accuracy. If the dataset contains a considerable amount of errors, the data is not accurate because it does not describe the "real world". Before using a dataset, you should make sure that it is clean. The provenance of data matters.
4) Completeness. This is important to make sure the dataset is not affected by any bias. Also, the wider the view it offers, the better.
5) Consistency.
6) Usage restrictions. The less restricted the use of the data, the higher its value.
7) Interoperability/accessibility. Consumers will choose the most accessible dataset. Also, consumers prefer data that can be used for multiple use cases.
8) Liabilities and risk. Potential risks reduce the value of the data.

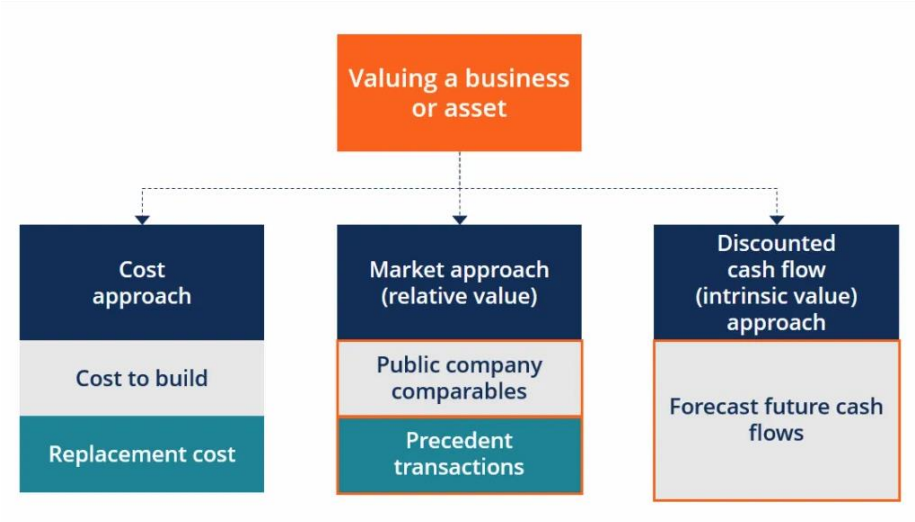
[53] How to measure the value of big data - Information Age (information-age.com)

4.1.3 Asset and Business Valuation

Typically, firms and assets are evaluated with one of the following methods[54]:

1) Cost approach
2) Market approach
3) Income approach (Discounted Cash Flow or DCF)

Graph 4.1.3 Valuing a Business or Asset



The cost approach considers the costs incurred in creating the data. The value is calculated as:

- The cost to produce/reproduce the data using similar methods (Reproduction cost approach)
- The cost of replicating the utility of the data without creating a copy of the data (Replacement cost approach).

This approach is useful to determine the base value of the data (initial value) and not to capture the future value that the data generates (IMDA & PDPC 2019).

The market approach considers the market value of similar datasets. It provides good evidence of the value because it proves that if a certain dataset was sold, it would generate some money (IMDA & PDPC 2019).

There are two different types of market approach[55]:

[54] Business Valuation Guides - Learn Important Valuation Concepts (corporatefinanceinstitute.com)
[55] Market Approach - Methods, Uses, Advantages and Disadvantages (corporatefinanceinstitute.com)

1) Public company comparables. Comparables must be in the same industry and similar characteristics like demand and supply factors, operational processes, and financial composition.
2) Precedent transactions. It involves the use of multiples based on transactions in the same industry. This method is valuable when a purchase/sale is under consideration. However, transactions may have happened in different market conditions.

The market approach is straightforward, it involves simple calculations, it uses real and public data, and it is not dependent on subjective predictions. Besides, it allows firms to easily justify the price of their assets. The difficulty of identifying comparables and the lack of publicly available data to use for this method are some of the most relevant disadvantages.

The Income approach considers the future cash flow that the data produces. The cash flow is produced for the owner or the right holder. It takes into account the growth rate of the cash flow. It is computed as a discounted cash flow (DCF) with the following formula[56]:

$$DCF = \frac{CF_1}{(1+r)^1} + \frac{CF_2}{(1+r)^2} + \frac{CF_3}{(1+r)^3} + \ldots + \frac{CF_n}{(1+r)^n} = \sum_{t=1}^{n} \frac{CF_t}{(1+r)^t}$$

This formula includes the cash flows that assets generate over time (CF), the time value of money (discount), and the cost of capital (r).

If the future cash flow can be estimated, this is a very convenient way to evaluate assets and companies. The DCF is the intrinsic value, because it does not consider any comparables, and it can be adjusted considering many different scenarios (ex: different growth rates).

The biggest drawback of this method is the fact that it entirely relies upon the ability to generate reliable estimates. A large number of assumptions are needed, as well as considerable forecasting skills.

[56] Discounted Cash Flow (DCF) - Overview, Calculation, Pros and Cons (corporatefinanceinstitute.com)

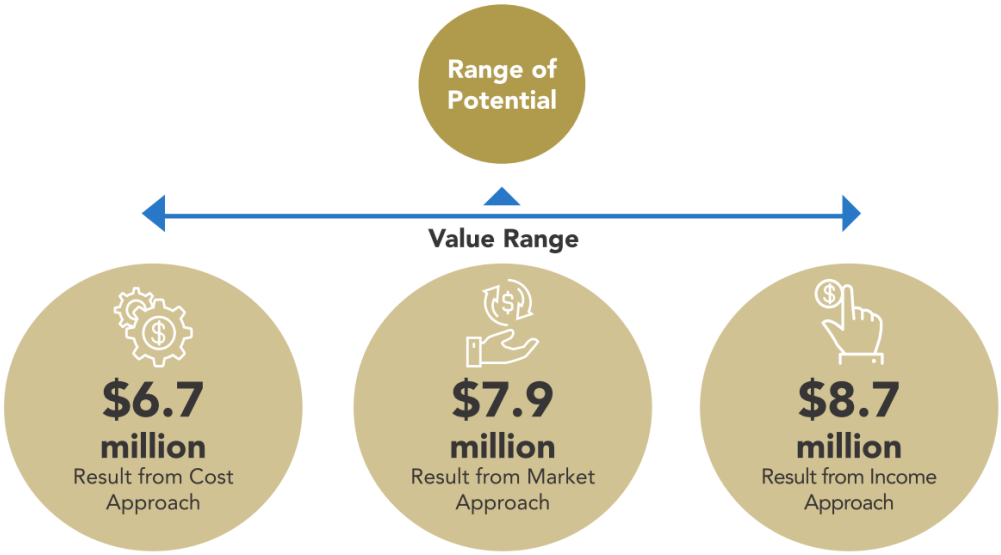
Hence, high-growth companies and innovative projects are examples where the DCF might not be the most appropriate approach to use.

The Personal Data Protection Commission of Singapore has illustrated an interesting example of how the value of data can be different when using the three different valuation methods (IMDA & PDPC 2019).

Graph 4.1.3.2 shows that the value range varies from $6,7 million when using the cost approach to $8,7 million as a result of the use of the income approach. These results all refer to the same sample company indicated in the 2019 report (IMDA & PDPC 2019).

Therefore, it is appropriate to apply more than one valuation approach in valuing the data, so that organizations can arrive at a more robust view of the real value (IMDA & PDPC 2019).

Graph 4.1.3.2 Value range



Gartner claims that choosing the appropriate valuation method is essential to grasp the value of the data. The Intrinsic Value regards the completeness, correctness, and exclusivity of the data. The Business Value considers whether the data is good or relevant for a specific purpose and the Performance Value is the measurement of how

much the data affects the key business drivers. Using a combination of all of the valuation methods is a good way to get the real value of the data[57].

Besides, another valid method is the asset-based approach. It is less commonly applied than the previous ones but it is still valid (Willamette 2018). This method requires more time (and costs) than any other method. It consists of finding the total value of the asset of a company. This is necessary when an asset acquisition will take place (Willamette 2018).

This approach is based on the value of all of the assets and liabilities of a company. The value can be:

- Fair value and Fair market value
- Investment value
- Owner value
- Use value
- User value

The value is likely to be different from the value recorded in the balance sheet because financial statements are presented in compliance with GAAP. GAAP-based balance sheets normally exclude internally generated assets such as Data (Willamette 2018). The asset-based approach evaluates all of the company's assets and all of the company's liabilities, so its application should be considered in every business valuation since every company owns assets (Willamette 2018). It is particularly applicable in the case of an asset purchase transaction, even though the lack of comparable traded companies in the subject industry may affect the analyst's ability to use this method. This issue does not undermine the utility of the asset-based approach (Willamette 2018).

The British analytics powerhouse SAS suggests that the value of the data is given by the degree of intelligence that a dataset embeds (SAS 2010).

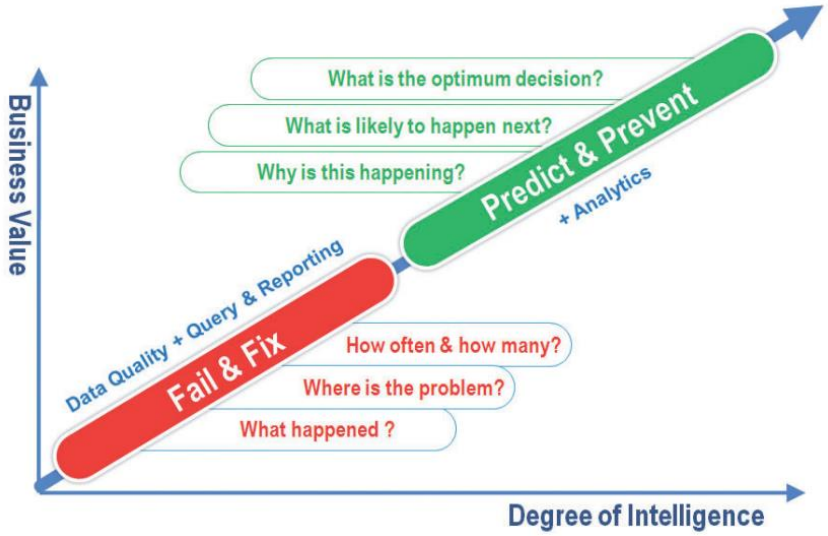
At first, the presence of data allows organizations to identify the problems and fix them. The first utility is represented by the capability of identifying what has occurred in the past (SAS 2010).

[57] Why and How to Measure the Value of Your Information Assets (gartner.com)

Most importantly, the data gives useful information that can be used to make more accurate predictions of what will happen in the future and to make sure that organizations can plan events, strategies, goals more efficiently, thanks to the data usage (SAS 2010).

The second characteristic, as indicated in Graph 4.1.3.3, is the one that makes the data extremely valuable. The capability to fix problems relies on the data quality and its graphical representation, while the latter is based on data analytics (SAS 2010).

Graph 4.1.3.3 Degree of intelligence and data value



It must be noted that information assets might be transitory and constantly need to be replenished with a flow of new data to preserve their functionality (SAS 2010). For instance, financial market information resources are updated daily or even by the second.

In fact, Noah Consulting states that data management has an effect on the value of the data asset (Noah Consulting 2013). Of course, good data management has a good effect and poor management has a negative effect. The "curated" data can also have a positive impact on the value of associated datasets[58] .

The valuation approach proposed by Noah Consulting is based on the need to set up a base value that can increase/decrease over time. The increase (or decrease) is named ROI (Noah Consulting 2013). The value that matters is the net value because generating data involves costs.

[58] (1) In Digital Transformation, Economies of Learning More Powerful than Economies of Scale | LinkedIn

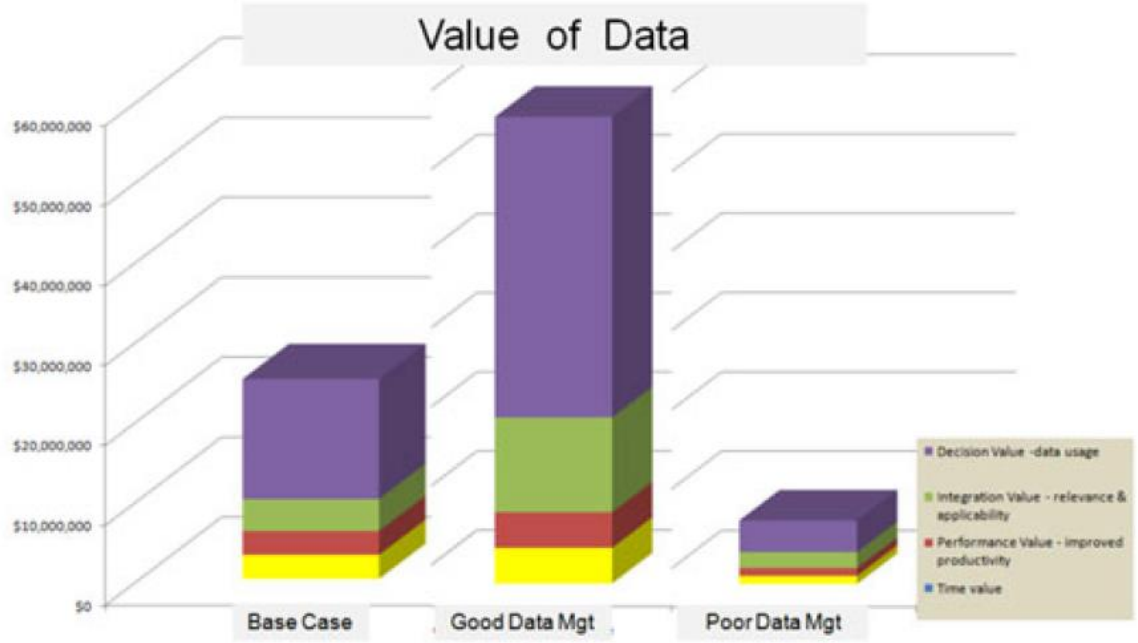
There are different types of value described in the study (Noah Consulting 2013):

- The time value of data. Normally, the value of data decreases over time because it may become out-of-date. However, some factors like the renewed interest in the data or updates can make its value increase again.
- Performance value of the data. This value comes from people's productivity. If the availability of information makes workers more productive, the increase in productivity is attributable to the data.
- Integration of data. The data can be integrated with other data to make larger and more useful datasets (Big Data).
- Value of decisions based on the data usage. Better decisions can be made when the data offers useful insights to firms and organizations.
- Value related to risks. Risks of making bad decisions and regulatory risks (ex: breaking the rules may cause firms to get fined) are examples of risks that have to be considered. If companies and organizations do not own/use some data, they might be inefficient.

In the case of good data management, the costs to produce the data and to make decisions based on data should go down. As for the value, good management enhances the value of the data. The opposite is true if the data is poorly managed (Noah Consulting 2013).

Graph 4.1.3.4 illustrates the results of the study carried by Noah Consulting. It shows the base value of the data, labeled as “Base Case”, the value in case of good data management and poor data management. The value is divided into decision value which is related to data usage, integration value, performance value, and time value. These affect the value of the data itself because the value of a dataset is related to its use (Noah Consulting 2013).

Graph 4.1.3.4 Data management and value



Given that data reduces the variance between optimal decisions and decision makers' decisions, the data obtained at no cost never make the decision-maker worse off. If the data has been purchased or produced with a costly process, the value of the data is the Net Value computed as the difference between the sum of costs and the sum of "benefits" obtained thanks to the availability of the data (Stander 2015).

4.1.4 Intangible Assets valuation

Firms that own a considerable amount of intangible assets have very peculiar characteristics (Damodaran 2009).

- Inconsistent accounting for investments made in intangible assets. Accounting principles separate capital expenses from operating expenses. In the case of tangible assets, the distinction is easy: expenses that produce benefits over many years are capital expenses, if they only affect the current year they are operating expenses. Most companies with high levels of intangible assets spend considerable amounts of money on training personnel and brand name advertising. These items produce uncertain benefits over time. Hence, they are

often accounted for as operating expenses. As a consequence, firms with intangible assets often report low CAPEX (Damodaran 2009).

- They generally borrow less money. Bankers are more incline towards lending money to firms that have a substantial amount of tangible assets. Also, many technology firms have just emerged from the growth phase of their life cycle (Damodaran 2009).
- Heavy use of equity options. Technology firms are at the beginning of their life cycle and they are dependent on retaining human capital (Damodaran 2009).

These characteristics have noteworthy consequences on the valuation. First of all, the value of assets is typically obtained by measurements such as earnings and book value. The flawed accounting treatment of intangibles makes these numbers unreliable. This affects both the DCF valuation method and the comparison between multiples of different companies (Damodaran 2009). Given the above, computing the return on equity and capital would not be particularly useful. The reinvestment made by firms with intangible assets is accounted as operating expenses. In addition, these firms bear another risk which is that of losing value very quickly. Since human capital plays an important role in these firms' investments, the mobility of the employees becomes an additional risk. Also, monitoring the value of intangibles is time-consuming and difficult (Damodaran 2009). Most analysts agree on considering the limitations coming from accounting rules more impactful for young firms with higher investments (Damodaran 2009). The impact of R&D spending as operation expenditure has an impact on the balance sheet because the value of the (intangible) assets created by research does not show up in the "Assets" section. Hence, R&D expenses must be capitalized. The amount to be capitalized is computed considering the amortizable life of assets. In the case of pharmaceutical companies, the amortizable life is quite long since the approval for new drugs takes time. R&D in software firms should be amortized in a shorter time because the products of this sector are created more quickly. The following formula equals the value of the research asset (Damodaran 2009).

$$\sum_{t=-(n-1)}^{t=0} R\,\&\,D_t \,\frac{(n+t)}{n}$$

The Adjusted Book Value of Equity then would become equal to the Book Value plus the value of the research asset. Of course, the Operating Income also needs to be readjusted otherwise, the capitalized R&D expenses would be counted twice. The Adjusted Operating Income is equal to the Operating Income plus R&D expenses minus the amortization of the research asset (Damodaran 2009). . Finally, the Adjusted Net Income would be equal to:

$$Net\ Income + R\&D\ expenses - Amortization\ of\ research\ asset$$

There are different types of capital expenses that are treated as operating expenses. For instance, advertising expenses are typically accounted as operating even though some companies like Coca-Cola or Gillette, which rely on their brand, could argue that these expenses should be treated as CAPEX (Damodaran 2009). Another notable case is that regarding recruiting and training expenses. In the case of new technology firms, such as Amazon, Selling General and Administrative expenses (SG&A) should be treated as CAPEX because they help to increase the brand name awareness and bring long-term customers (Damodaran 2009).

4.1.5 Consequences of capitalized operating expenses

If these expenses are capitalized, there will be a series of consequences on the corporate valuation (Damodaran 2009).

The first set of consequences regards the DCF Valuation.

- Earnings. The adjustments consist of adding the current year's expense and subtracting the amortization of past expenses. Hence, the effect on Earnings will be null if the expenses have remained stable, and positive if they have increased over time (Damodaran 2009).
- Reinvestment. Reinvestments increase or decrease by the same amount as earnings (Damodaran 2009).

- Free Cash Flow to Equity (FCFE). FCFE is computed by netting Reinvestment from Earnings. Therefore, there is no effect on FCFE because the two change by the same magnitude.
- Reinvestment Rate. If Earnings and Reinvestment increase (as a consequence of the capitalization), the reinvestment rate will increase.
- Capital invested. Unamortized prior year's expenses are treated as an asset. Hence, they add to the Equity of the firm. The effect of capitalization is higher if the amortizable life is long. This affects Pharmaceutical firms (long-run effect of R&D expenses) much more than Software companies, for which these capitalized expenses typically finish their utility sooner (Damodaran 2009).
- Return on Equity (ROE). The effect on ROE is uncertain since the capitalization of operating expenses affects both Earnings and the Capital invested. However, it can be understood that if the ROE increases it is because the return on these expenses (R&D, SG&A) is higher than the return of traditional investments (Damodaran 2009).
- Expected growth rate. As a general rule, the higher the reinvestment rate, the higher the growth rate. Anyway, the ROE must be considered (Damodaran 2009).

The effect on earnings of capitalizing operating expenses is greater at young firms because operating expenses such as R&D are presumably a large proportion of total expenses (Damodaran 2009). Besides, the capitalization effect can vary across firms in the same sector because the capitalization depends on the amortizable life of the expenses (Damodaran 2009). Ignoring these two things could lead to wrong analysis. Young firms and firms with a long amortizable life of R&D would seem overvalued (high PE and EV/EBITDA). It would be extremely useful to compute both P/(E+R&D) and P/(E+Net R&D) to perform a more accurate relative valuation (Damodaran 2009).
The Bank of Spain also suggests that the intangible assets generated internally through operating expenses must be made explicit (Banco de España 2007).
If M_j is the stock of intangibles (that will vary over time as the result of the depreciation), C_j the adjustment costs, x is the depreciation rate of the intangibles, and

f(C) is the flow of the intangibles produced with the adjustment costs incurred in the period (t), then:

$$M_{jt} = (1 - x_j) \cdot M_{j,t-1} + f(C_{j_t})$$

The study made by the Bank of Spain in 2007 has shown that the banks that high training expenses (which should be capitalized) are 10% more valuable than banks with lower training expenditure (Banco de España 2007). The compensations for this cost are higher investment rates and IT skills (Banco de España 2007).

4.1.6 "The dark side of valuation"

Typically, the Free Cash Flow to the Firm (FCFF), used in the DCF valuation approach, is computed using the following formula (Damodaran 2001):

$$EBIT\ (1 - tax\ rate) -$$
$$(CAPEX - Depreciation) -$$
$$Change\ in\ Non - cash\ Working\ Capital =$$
$$FCFF$$

In this formula, the Net Capex (CAPEX – Depreciation) and the increase in Non-cash Working Capital represent the reinvestments made to generate growth (Damodaran 2001). As I explained earlier, the fact that many R&D expenses are treated by technology firms as Operating Expenses rather than CAPEX may cause an inappropriate result of the valuation when using the DCF approach (Damodaran 2001). Besides, if a new technology firm has not generated a positive operating income yet, the FCFF is negative too. Also, even if a new technology firm has a positive operating income, the reinvestment needs could make the FCFF be negative. This should be taken into account when establishing a set of assumptions on which the DCF valuation is based (Damodaran 2001).

The expected growth rate must be determined to use the DCF approach. The growth of companies is mostly determined by the quality and quantity of their reinvestment rate, which multiplies the Return on Capital (Damodaran 2001). Given that

$$Reinvestment\ rate = \frac{(CAPEX - Depreciation + \Delta Non - cashWC)}{EBIT\ (1 - tax\ rate)}$$

And

$$Return\ on\ Capital = \frac{EBIT\ (1 - tax\ rate)}{Capital\ Invested}$$

The future cash flow that the firm generates can be estimated using the aforesaid indicators which tell a lot about a firm's profitability and potential to increase the profitability in the future (Damodaran 2001).

However, this may not be valid for technology firms such as Big Data companies. The fact of treating R&D investments as operating expenses may cause both the reinvestment rate and return on capital to be considerably different from what they actually are (Damodaran 2001). Secondly, the valuation regarding companies that report operating losses (such as Amazon in the early-2000s) needs to be adjusted when estimating the future cash flow, so their profitability should not be estimated as a growth on the current earnings otherwise the estimated values could not be positive (Damodaran 2001). Also, most technology firms rely on low Debt for their financing, the cost of capital of these firms probably coincides with the cost of equity. When they issue bonds to borrow money, they usually issue hybrid securities like convertible bonds. This happens when the firms become larger and stable. As a consequence, their Debt-to-Equity ratio and cost of capital can change significantly over time (Damodaran 2001).

As for the growth rate to indicate in the DCF valuation method, there is one way that could make the model less sensitive to the initial assumptions and, consequently, more accurate (Damodaran 2001).

Considering that the growth of technology firms strongly depends on the quality and quantity of their reinvestment rate, it is possible to substitute the growth rate with the Reinvestment rate multiplied by the Return on Capital (Damodaran 2001).

Hence, the Terminal Value (value of the firm beyond the forecasted period when future cash flows can be estimated[59]) is:

$$\frac{EBIT_{n+1}\,(1-t)(1-Reinvestment\ rate)}{Cost\ of\ Capital_n - (Reinvestment\ rate * Return\ on\ Capital)}$$

Even if the Terminal Value must be computed when performing a DCF valuation, one must also take into account that firms' survival is not granted (Damodaran 2001).

Since the DCF valuation that I presented relies on the assumption that the firm will continue to exist forever, it would be useful to insert in the analysis the probability that a company will cease to exist. That is why it is important to identify whether a firm is a young firm (beginning of its life cycle) or a mature firm. Young firms may have liquidity issues due to the low cash flow from operations and the high requirement in terms of investment that they need to grow (Damodaran 2001).

You can compute a ratio called Cash Burn Ratio which can measure the amount of time needed to run out of cash, based on the EBITDA (Damodaran 2001).

$$Cash\ Burn\ Ratio = \frac{Cash\ Balance}{EBITDA}$$

Also, the cash flow estimation should take into account the probability of survival of the firms. In other words, you should make explicit that the cash flow will be produced only if the firm survives (Damodaran 2001).

It is possible to do so by multiplying the cash flow by 1 minus the probability of distress. At the same time, you must add the distressed sale value multiplied by the probability of distress (Damodaran 2001).

This consideration is particularly important when the estimated cash flow is very optimistic. This can happen when the valuation is based on the enthusiasm that young

[59] Terminal Value (TV) Definition (investopedia.com)

growing firms can give. Of course, the probability of distress is also dependent on subjective considerations but adding a numerical measure of uncertainty can be considerably useful when making predictions (Damodaran 2001).

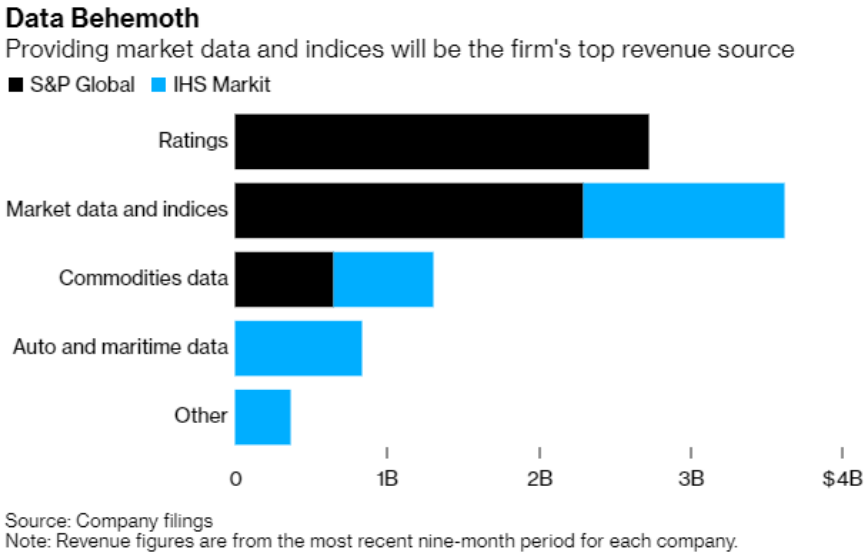
These observations are contained in Damodaran's 2001 book "The Dark Side of Valuation". Damodaran claims that the explosions of New Tech firms made the traditional valuation models ill-suited for these innovative firms. Therefore, he has examined some hidden and difficult valuation issues that may arise when evaluating New Tech companies. Since data companies are part of the "new economy", I think his observations are useful for my analysis.

4.2.1 M&A in the economy of data: the S&P Global and IHS Markit case

On November 30 2020 S&P Global and IHS Markit have announced a merger agreement worth $44 Billion[60].

The main reason why S&P decided to invest such a large amount of money is explained by Bloomberg [61].

Graph 4.2.1 S&P Global and HIS Markit M&A



As indicated in Graph 4.2.1, one of the key value drivers of IHS Markit is its ability to produce and collect data. Lance Uggla, Chairman and Chief Executive Officer of IHS Markit said: "Our highly complementary products will deliver a broader set of offerings

[60] S&P Global-IHS Markit merger, 2020's largest deal, shows value of financial data | S&P Global Market Intelligence (spglobal.com)
[61] S&P Global in Talks to Acquire IHS Markit for $44 Billion - Bloomberg

across multiple verticals for the benefit of our customers, employees, and shareholders. Our cultures are well aligned, and the combined company will provide greater career opportunities for employees. We look forward to bringing together our teams to realize the potential of this combination."[62]

In the same press release, IHS clarified that there are two sets of reasons why this merger has occurred.

The first one is strategic. This M&A operation will allow both companies to:

1) Increase the business scale and earnings to have additional financial flexibility.
2) Differentiation: the combined company will attract high-growth adjacencies such as ESG, climate transition, SMEs. The Total Addressable Market of these is estimated at $20 Billion.
3) Increase the customer value proposition by providing better insights.
4) Combine the expertise of two leader firms into one organization.

Additionally, IHS recognized that the M&A will also have financial benefits.

1) Enhanced growth. IHS estimated that the combined company will have a 6,5%/8% annual growth rate.
2) Higher profitability.
3) Maintenance of a strong credit profile to pursue further growth.
4) Higher Free Cash Flows and Payout Ratio.

Besides, the deal helped S&P Global to strengthen its position as a market data provider. Most data gathered by IHS is pricing data, reference data for various asset classes, valuation data, financial indices. This data is not publicly available and complex. Therefore, it is complementary to S&P's operating activities [63].

Another important aspect of this Merger is that regarding IHS's Balance Sheet. As reported by Seeking Alpha[64], the firm has a considerable amount of intangible assets ($4 Billion over $16 Billion of Total Assets). Also, Goodwill accounts for nearly 60% of Total Assets.

[62] S&P Global and IHS Markit to Merge in All-Stock Transaction Valuing IHS Markit at $44 Billion, Powering the Markets of the Future | IHS Markit

[63] Why Does S&P Global Want To Buy IHS Markit? (forbes.com)

[64] IHS Markit: A Great Add-On For S&P Global (NYSE:INFO) | Seeking Alpha

Another key metric for IHS is the Accounts receivable. Even if the Covid impacted tremendously the economy, no deterioration of this item occurred. Furthermore, the cash generation capacity of the firm makes its debt side stable and safe. In fact, the operating cash flow surpasses the interest expenses by 6 times.

As for the purchase itself, S&P Global has evaluated the equity of the target firm $40 Billion, which is a lot considering that IHS generates $1 Billion with operating activities[65]. However, S&P decided to pay with its shares (near a record high in November 2020) giving up one-third of itself. The S&P Global Market Cap at the time of the acquisition was around $76 Billion, which means that S&P purchased a firm worth more than half of S&P's value paying one-third of the value. The deal has been named "All-stock deal" for this reason[66].

S&P stated that the merger will make the firm generate over $5 Billion in Free Cash Flow annually, allowing it to invest over $1 Billion in technological advances[67].

FitchRatings has indicated that S&P needs an initial investment in synergies expenses to finalize the deal, driven by optimizations needed to make the two firms cooperate[68]. This is no surprise because, as I have reported in Chapter 2, most data firms require a considerable initial investment to build up the infrastructure and corporate competencies needed to make the firm work.

The risk of this operation could be represented by regulatory scrutinies that occur to prevent the birth of a monopoly[69] considering that the regulators have become more active in intervening in these cases lately[70] and they do not intend to stop doing it in the future[71]. The authorities started the antitrust scrutiny[72] as expected[73][74]. However, the two companies expect to officially close the deal in the second half of 2021[75].

[65] Final IHS Markit Annual Report 2020, pag. 58
[66] S&P Global to buy IHS Markit in all-stock $44 billion deal (businesstoday.in)
[67] S&P Global to buy IHS Markit for $44 billion, expanding data empire | Reuters
[68] Fitch Affirms S&P's 'A-' IDR Following IHS Markit Merger Announcement; Outlook Stable (fitchratings.com)
[69] S&P Global reportedly in talks to buy IHS Markit, as race for scale among data providers takes off | Fortune
[70] Big Tech & Antitrust | S&P Global (spglobal.com)
[71] Big Tech has an antitrust target on its back, and it is only going to get bigger - MarketWatch
[72] 8-K (sec.gov)
[73] S&P deal for IHS likely to draw antitrust scrutiny from Biden administration | Reuters
[74] S&P deal for IHS likely to draw antitrust scrutiny from Biden administration (yahoo.com)
[75] S&P Global and IHS Markit Provide Update on their Pending Merger (prnewswire.com)

4.2.2 IPO in the economy of data: Snowflake and the underpricing

The cloud company Snowflake became a public company after the successful September 16 2020 IPO[76]. The lead underwriters for the offering were Goldman Sachs, Morgan Stanley, J.P. Morgan, Allen & Co, and Citigroup[77]. This Initial Public Offering was the biggest regarding a software firm[78] as the price of Snowflake's shares doubled within the first day. The post-IPO rally boosted the price of Snowflake shares by 258%, leading to a company's value of $120 Billion, ten times more than the value of the Snowflake in February 2020, when private funding had happened[79].

The IPO made Snowflake more valuable than companies like Dell, Uber, and General Motors[80].

The outstanding result of this IPO confirmed the strong interest of large investors (ex: Warren Buffet) for data firms that grow at high rates[81]. In fact, despite the economic crisis caused by Covid-19[82], Snowflake's revenue has experienced impressive growth in the first half of 2020[83], even though its net loss nearly doubled[84].

Anyway, the investors did not seem concerned about Snowflake's financial statements. The firm started a partnership with Salesforce to be able to compete against Amazon, Microsoft, Alphabet (Google owner) in the cloud industry. Even Berkshire Hathaway, which usually prefers to invest in mature companies like Apple and Amazon, made a bet on Snowflake's success[85].

However, some analysts started questioning the success of this IPO, claiming that Snowflake has probably been mispriced[86].

Some investors bought Snowflake's shares at 150 times sales. In other words, it will take 150 years for Snowflake's revenues to "buy" the company at that price. Even if Snowflake experienced an impressive growth rate (which is expected to continue in the following years), it is unlikely that investors are happy about that. Hence, investors are probably confident about the fact that Snowflake (and the data market as a whole)

[76] Snowflake IPO: (SNOW) starts trading on the NYSE (cnbc.com)
[77] Snowflake IPO raises $3.36 billion in year's biggest U.S. listing | Reuters
[78] Snowflake shares more than double. It's the biggest software IPO ever - CNN
[79] Cloud-storage firm Snowflake is now more valuable than IBM after 258% post-IPO rally | Markets Insider (businessinsider.com)
[80] Snowflake (SNOW) Soars 130% in $3.36 Billion Record Software IPO Debut - Bloomberg
[81] Snowflake Priced Its IPO at $120. Here's Where the Stock Finished Its First Day of Trading. | Barron's (barrons.com)
[82] Covid-19 Recession: 10 Important Numbers That Sum Up America's Economic Crisis One Year Later (forbes.com)
[83] Snowflake stock soars in largest software IPO ever; Madrona among earlier venture investors - GeekWire
[84] Snowflake IPO: Snowflake IPO raises $3.36 billion in year's biggest US listing - The Economic Times (indiatimes.com)
[85] Snowflake shares more than double. It's the biggest software IPO ever - CNN
[86] Was Snowflake's IPO mispriced or just misunderstood? | TechCrunch

will grow considerably in the future and, at the same time, there is very little risk that it will be a dead-end in the near future[87].

Even before the IPO, it was well known that the IPO would have been a great success. The confidence was due to the estimated future growth rate of the company and the sector. Also, the track record of Snowflake's management was an encouraging factor[88]. Snowflake itself claimed that the IPO will make the company stand out within its industry because it will help to increase its visibility, transparency towards customers and partners, and a future expansion[89].

Since the data economy is based on the network effect, it is no surprise that Snowflake's management is optimistic about the future of the company. The IPO will help Snowflake to acquire new customers (and new data) that will contribute to the success of the firm with more information to analyze. Therefore, Snowflake's data will be more complete, accurate, and, consequently, more valuable (see Chapter 3).

The lesson that can be learned from Snowflake's IPO is that a considerable amount of data firms could probably be listed not at top price in the future so that investors are attracted by the opportunity of getting a good deal. The shares might be intentionally underpriced to boost demand[90]. It must be noted that the existing shareholders are not normally happy with underpricing because it reduces the liquidity that they can get during an IPO[91]. Considering that they give up some percentage of the equity to raise money[92], the underpricing is not to be taken as a conventional IPO strategy[93].

4.3 Valuation

In this section, I have reported my analysis of 151 firms that operate in the Information industry. I have analyzed some indicators that are useful to figure out where the value of data firms comes from, if they are overvalued, and how their value should be measured. All the data that I have used comes from the Merrill Lynch BofA database (last update: April 2021) and Koyfin (April 2021 data). I have also used the data

[87] What Happened When I Bought Snowflake on IPO Day | The Motley Fool
[88] 3 Reasons To Buy Into Snowflake's IPO (forbes.com)
[89] The Snowflake IPO - What does it mean? | Snowflake Blog
[90] Underpricing Definition (investopedia.com)
[91] Why Does a Company Decide to Go Public? (groww.in)
[92] The Benefits and Costs of Going Public | Torys LLP
[93] Investor Bulletin: Investing in an IPO (sec.gov)

contained on the website of Prof. Aswath Damodaran[94]. As reported on the website, the last data update was on January 2021.

I have reported 3 types of data. I have referred to these 3 as Dataset, Market, and Information Industry.

- Dataset. It refers to the 151 data companies contained in the Merrill Lynch BofA database. It is a narrow dataset that only includes data stocks in BofA's watchlist.
- Information Industry. 215 data companies contained in Damodaran's dataset. The word Information is used as a synonym of data. Hence, the Information Industry's values are to be considered as the average value of the data industry. This dataset is broader than the first one, considering that it contains 64 firms more.
- Market. Damodaran's dataset contains 44.394 firms of all industries and sectors. The Market's values are the average of the whole market. I have used this dataset as a benchmark to understand how data firms perform in relation to the market.

Table 4.3.1 Dataset Summary

Dataset Summary	
N. of Stocks	151
N. of Industries	14
N. of Countries	20
% Stocks USA	58%
Average Market Cap ($M)	130.762,76
Lowest Market Cap ($M)	817,64
Highest Market Cap ($M)	2.102.225,34
Market share	60%

As reported in table 4.3.1, I have analyzed 151 stocks belonging to 14 industries (or Big Data sub-sectors) and 20 countries. 58% of these stocks are American and they represent 60% of the Big Data market. I have computed the market share as a ratio of the total revenue of the 151 stocks over the revenue of the whole data industry[95].

[94] Damodaran Online: Home Page for Aswath Damodaran (nyu.edu)
[95] • Global Big Data market size 2011-2027 | Statista

Table 4.3.2 Profitability: Dataset and Market

	Dataset	Market
Gross Margin	53,35%	32,93%
ROE	13,50%	10,69%
Net Margin	11,29%	7,26%
EBITDA/Sales	27%	13,94%

The first aspect that I have analyzed is profitability.

The average Gross Margin of the companies that I have studied is considerably larger than the market's (53% vs 32%). The Gross Margin is computed as the difference between Revenues and COGS (cost of goods sold)[96]. Therefore, I think this indicator is useful to understand the profitability of a firm's business model (in this case, that of big data firms) rather than the profitability of the whole firm because it only considers the costs incurred in generating the revenue. As the results show (Table 4.3.2), big data firms have a very profitable business model.

Of course, the Net Margin is much lower but the ratio between the mean of the 151 companies and the benchmark (nearly 1,6) is approximately the same.

The operating activities of these companies are also more efficient, as proven by the higher EBITDA-To-Sales ratio (27% vs 13%). Investors that are not concerned about any aspect related to the capital structure, taxes, depreciation, and amortization might find this very encouraging.

From the point of view of an investor, it is also interesting to observe that the ROE is higher on data firms (13% vs 10%).

Table 4.3.3 Profitability: Information Industry and Market

	Information Industry	Market
ROE	27,33%	10,69%
ROE-COE	18,60%	2,11%
Gross Margin	51,89%	32,93%
Net Margin	17,28%	7,26%
Operating Margin	27,11%	9,69%
After Tax, lease, R&D Margin	22,31%	8,17%

Similar results can be obtained when comparing the Information industry values to those of the market, as Table 4.3.3 shows.

[96] Gross Margin Definition (investopedia.com)

In addition, data firms tend to have a much larger Operating Margin (27% vs 9% of the market). Besides, even when considering Taxes and other expenses such as leasing and R&D expenses, the Income generated by data firms remains considerably higher (22% vs 8%).

The last profitability indicator that I have reported is the difference between Return on Equity and Cost of Equity (ROE-COE). The reason why I think this indicator is relevant is that investors would probably be interested in knowing whether investing in data firms (and in any other firm) would make them better off. The fact that ROE-COE in the Information industry is higher than that of the market (18% vs 2%) explains why many investors decide to bet on data companies (see Paragraph 4.2.1).

Table 4.3.4 Profitability growth

Profitability growth (2020-2022)	Dataset
Sales	39%
EBITDA	44%
Net Income	55%

The stocks that I have analyzed will presumably experience significant growth in profitability between the next two years (Table 4.3.4).

According to Merrill Lynch, the average growth in Sales will be 39% (vs 25% of the whole data industry[97]).

The EBITDA (Earnings Before Interests Taxes Depreciation and Amortization) will also increase by 44% and the Net Income will grow by 55%.

The data contained in Table 4.3.3 shows not only that the industry will continue to grow but also that the increase in profitability of data firms will be higher than the increase in Revenue. Considering that data firms have to bear a significant amount of fixed costs and few variable costs (see Chapter 2), this is no surprise.

[97] • Global Big Data market size 2011-2027 | Statista

Table 4.3.5 Cost structure: Information Industry and Market

	Information Industry	Market
CAPEX/Sales	14,11%	3,97%
COGS/Sales	48,11%	67,07%
R&D/Sales	2,31%	1,87%
SG&A/Sales	21,30%	14,81%
Capex/Depreciation	58,71%	145,85%

As for the cost structure, Table 4.3.5 indicates some aspects that I have already reported when explaining the main costs that data companies bear (Chapter 2), as well as the accounting principles that regulate these costs (Chapter 4, Paragraphs between 4.1.4 and 4.1.6).

Data firms have higher Sales, General and Administrative costs, R&D expenses, and CAPEX (as a percentage of Sales) than the market benchmark.

SG&A expenses include overhead costs that are often fixed or semi-fixed. In particular, General&Administrative expenses do not include costs related to the sale of products/services but rather to the general ongoing of the business. Human resources costs are also included[98]. Considering that data firms bear a substantial initial investment in the infrastructure and in training the employees (see Chapter 2) and that fixed costs prevail in data firms, the SG&A/Sales ratio is likely to decrease when the company grows, as an effect of the economy of scale.

The CAPEX/Depreciation ratio is significantly lower in data firms. Even if this could theoretically indicate that these firms are in a declining phase (because the reinvestment on fixed assets is lower than the loss of value of existing FA[99]), the result should not be seen as concerning. First of all, an excess of reinvestment could burn cash too quickly, leading to further problems[100]. Besides, as I described before, most investments made by data firms are not accounted for as CAPEX. Operating expenses (ex: training expenses) must be capitalized. The traditional accounting procedures do not take this into account and, as a consequence, the traditional valuation methods might be inapplicable.

[98] SG&A Expense (Selling, General & Administrative) - Guide, Examples (corporatefinanceinstitute.com)
[99] Capital Expenditures to Depreciation Ratio (bizfluent.com)
[100] The capital cycle is something every investor should be aware of — UK Value Investor

Table 4.3.6 Price and Value

	Information Industry	Market	Dataset
PE ratio	39,07	37,04[101]	64,49
PEG ratio	2,32	1,20	NA
Price/Sales	7,49	1,50	10,00
P/B	6,28	1,90	11,69
EV/EBITDA	25,30	14,08	50,12
EV/Sales	8,12	2,28	10,26

Some argue that tech stocks are overpriced[102][103]. Table 4.3.6 shows that this might be false. The Price-to-Earnings ratio of the data industry is slightly higher than the market's (39 vs 37). The average PE of the dataset that I have analyzed is considerably higher. Considering that I made my analysis on the top 151 data stocks, meaning nearly 60% of the total data industry (see Table 4.3.1), one could infer that investors tend to overinvest only in the top data firms.

The price/Sales ratio clearly shows that data firms are overpriced in respect of the revenue. In other words, these firms should generate higher revenue to justify the price of their shares. Information companies are 5 times more expensive than how they should be if using this indicator. However, Table 4.3.2 indicates how data firms are way more profitable than the others. This implies that they need fewer revenues to generate the same Net Income. Also, Table 4.3.5 and Chapter 2 report that most costs that information companies bear are fixed or semi-fixed. Hence, they do not increase proportionally when Sales increase.

Even when comparing the Price to the Book Value (Assets-Liabilities), information companies appear as overpriced. In my opinion, this ratio is not so useful as it would be for industrial firms that have a substantial amount of Fixed Assets coming from CAPEX. As I mentioned before in this study, internally generated assets (such as data) are not to be booked in the Balance Sheet. Besides, investments made by data firms mostly come in the form of Operating Expenses. These should be capitalized when performing a corporate valuation because they are not typically indicated as Fixed Assets. The fact that this is one of the most important ratios used by the so-called

[101] P/Es & Yields on Major Indexes | Market Data Center | Barron's (barrons.com)
[102] These 3 Tech Stocks Are Absurdly Overvalued Right Now | The Motley Fool
[103] 3 Tech Stocks That Are Wildly Overpriced | Nasdaq

"Value Investors" (ex: Warren Buffett)[104] could be one of the reasons why Value Investing has not been as effective as it used to be before the digital age lately[105].

Out of the 151 stocks that I have analyzed, 87% of these report a P/B higher than the market. So this characteristic is shared by most firms of the dataset. It must be noted that the average P/B of the 151 companies is expected to decrease by 10% in the next two years.

Finally, the EV/EBITDA of the market is 14 and that of the Information industry 25. The mean EV/EBITDA of the 151 companies contained in my analysis is 50. Even if this could indicate that data companies are overvalued, it is natural that firms in new high-growth industries have higher EV/EBITDA because it is strongly influenced by the size of the firms[106]. Also, new firms' EBITDA could be lower than zero. For these reasons, even if the EBITDA multiple is one of the most used in corporate valuations, it gives a better understanding of value if one compares the multiples of two (or more) firms in the same industry[107].

Table 4.3.7 Cost of capital

	Information Industry	Market
Beta	1,01	0,94
Cost of Equity	5,70%	5,37%
E/(D+E)	91,44%	67,42%
Cost of Debt	3,00%	3,00%
Cost of Capital	5,40%	4,34%

Table 4.3.7 shows how data firms bear a slightly higher Cost of Equity. This could be partly explained by the higher volatility (Beta) of data firms in respect of the market[108]. Therefore, even if the Cost of Debt of data companies equals that of the market (3%), the Weighted Average Cost of Capital (WACC) is higher in the data industry.

As for the capital structure, there is a noticeable difference in terms of percentage of Equity on total capital (91% is the average in the data industry, 67% in the market).

This could indicate that most data companies are not mature firms yet. Equity financing is more suitable for newborn, high-growing firms while Debt financing is

[104] Book Value vs. Market Value: What's the Difference? (investopedia.com)
[105] Does value investing still work? | The Economist
[106] How do size, growth and margin affect valuations? | Grant Thornton
[107] What Is Considered a Healthy EV/EBITDA ? (investopedia.com)
[108] The cost of equity is computed using the Capital Asset Pricing Model (CAPM). It is positively correlated with the Beta

convenient if a company is able to generate enough Cash Flows needed to meet the Debt obligations[109].

For this reason, I have suggested that even some listed firms could be evaluated as startups (see section 4.5.3).

Table 4.3.8 Growth rate

	Information Industry	Market
ROE	14,35%	8,25%
Retention Ratio	67,59%	28,69%
Fundamental Growth	9,70%	2,37%
ROC	24,34%	6,05%
Reinvestment Rate	-7,08%	33,16%
Expected Growth in EBIT	-1,72%	2,01%

The last aspect that I have analyzed is the growth rate of data firms in respect of the market.

The expected growth rate of Return on Equity (ROE), Return on Capital (ROC), and Fundamental Growth[110] of data firms are substantially higher than the market.

This confirms the optimism in terms of the expected growth rate of the data industry (see Chapter 1) and the fact that data firms become more profitable when they grow[111].

The Retention Ratio[112] growth rate of data firms is also higher (67% vs 28%). This could indicate that most data firms need to retain their Earnings to sustain corporate investments. In my opinion, the reason why shareholders accept a lower payout ratio[113] is that they are confident about the growth rate of the companies. Also, they probably expect the firms to be more profitable in the future than they could be now. In addition, this indicator could show that data firms need more money to grow.

On the other hand, the Reinvestment[114] growth rate of data firms is negative (-7%) while the expected growth rate of the market is positive (2%). The reinvestment rate

[109] Debt vs. Equity Financing: Pros And Cons For Entrepreneurs (forbes.com)

[110] Growth in EPS

[111] In chapter 2 I have reported that data firms mostly bear fixed costs. Hence, they need a remarkable initial investment but not many further variable costs. Also, in Chapter 3, I have described that the increase in profitability is exponential. In platforms, the larger the number of users, the larger the utility of each user.

[112] Percentage of Earnings kept back in the business as retained earnings. A high value could indicate that the firm is investing a lot.

[113] Proportion of Earnings paid as dividends

[114] The reinvestment rate measures how much a firm is reinvesting past Earnings to generate future growth. Although this is a good place to start, it is not necessarily the best estimate of the future reinvestment rate. It is important to continue treating R&D expenses and operating lease expenses consistently. The reinvestment rate for a firm can be negative if its depreciation exceeds

for a firm can be negative if its depreciation exceeds its capital expenditures[115]. The fact that most technology firms account for investments as Operating Expenses and not CAPEX (see section 4.1.6) could affect this value significantly. Besides, the companies that have grown a lot in the last few years might not need to reinvest that much. Therefore, when evaluating a specific firm, it would be useful to compare this value to the industry average's, to figure out if the firm is reinvesting more or less than its comparables[116].

Finally, the expected EBIT growth rate of data companies is also negative (-1%), while the market's is positive (2%). The expected growth in EBIT is correlated to the Reinvestment rate[117]. Therefore, if the Reinvestment rate is negative, this indicator will also be negative.

4.4 Financial statement analysis

In this section, I have described some cases regarding how data and other digital assets are accounted for in companies' balance sheets. Also, I have shown how data produces value for companies. This information is helpful to figure out how to evaluate data by using a deductive method.

4.4.1 Non-GAAP and Goodwill: IBM

All listed companies are required to report the financial data according to the General Accepted Accounting Principles (GAAP)[118]. However, a considerable number of companies also report Non-GAAP items such as non-GAAP Earnings. Even if considering these non-GAAP items could be risky (companies might want to report data that gives a better view of the firm[119]), one must consider that the accounting staff of these firms has access to more information than external parties. Non-GAAP measurements may be useful when a company incurs a large one-time expense and

its capital expenditures. Firms that have over invested in capital equipment or working capital in the past may be able to live off past investment for a number of years, reinvesting little and generating higher cash flows for that period.

[115] The Fundamental Determinants of Growth (nyu.edu)

[116] The Fundamental Determinants of Growth (nyu.edu)

[117] Expected GrowthEBIT = Reinvestment Rate * Return on Capital

[118] Non-GAAP Reporting (cfainstitute.org)

[119] Understanding GAAP vs. Non-GAAP (investopedia.com)

internally generated assets (such as data). In any case, investors should be critical when analyzing non-GAAP Earnings[120].

When talking about data, another relevant accounting item is Goodwill. The IBM 2019 report shows that a consolidated custom in accountancy is to book purchased data assets as an increase in Goodwill[121]. For instance, IBM purchased the IT firm RedHat[122]. Consequently, IBM has reported an increase in its Goodwill and an increase in Debt[123].

The factors that increased IBM's value are[124]:

1) Faster Data Analytics
2) Higher portability by reducing the existing barriers in the previous infrastructure
3) Reduction of costs for the cloud through the use of RedHat solutions
4) Modernization of the data center useful for a large-size company like IBM

Therefore, IBM has reported an increase in Goodwill just because the acquisition of data and data infrastructure has been beneficial for the firm. This thing suggests that data does not have an absolute value but its value is strongly related to the buyer. If a specific dataset and data analytics infrastructure fit into the buyer's organization, the increase in value given by the acquisition could translate into an increase in Goodwill. Goodwill is to be accounted for as the difference between the price paid and the market value of the assets that are in the books of the target[125]. Otherwise, data does not have a value. Hence, one could conclude that the intrinsic value of data is zero.

4.4.2 Amazon

Amazon is composed of Amazon North America, Amazon International, and Amazon Web Services (AWS).

Amazon reports that AWS is more efficient than most data centers. This is true because[126]:

[120] Non-GAAP Earnings - Overview, Significance, Common Measures (corporatefinanceinstitute.com)
[121] IBM report 2019, pag.30
[122] Red Hat officially acquired by IBM
[123] IBM report 2019, pag.30
[124] Alleanze strategiche IBM | Red Hat
[125] IBM report 2019, pag. 67
[126] Amazon report 2019, pag.5

1) The higher utilization of the infrastructure makes the usage more efficient. Most data companies make servers work at 18% of their capacity to handle usage spikes. AWS is able to manage the maximum usage rate more efficiently.
2) AWS improved the energy usage efficiency

Amazon also claims that, despite the high utility of using its data to improve the retail services efficiency, it faces risks related to data loss and security breaches[127]. The risks are to let other parties misuse Amazon's data and to not using that data, meaning not being able to use costly information to improve the corporate efficiency. This would be a problem also because Amazon faces an additional risk, which is that of managing the Inventory through AI use. In these cases, data would somehow become a liability because its generation and maintenance involve spending money even if data is not used.

Amazon says that AWS requires "Technology and Content" expenses (R&D, maintenance, and investments in the infrastructure) that will increase over time but at a lower rate than they have raised in the past due to an increase in the estimated life of the initial investment (servers, infrastructure)[128]. This is the cost structure logic of most data companies (see Chapter 2).

Similarly to IBM, Amazon reports the acquisition of technologies, together with intangible assets that do not qualify for separate recognition, as an increase in Goodwill[129].

[127] Amazon report 2019, pag.10
[128] Amazon report 2019, pag. 27
[129] Amazon report 2019, pag. 55

Table 4.4.2[130] Amazon: Financial statement

	Year Ended December 31,		
	2017	2018	2019
North America			
Net sales	$ 106,110	$ 141,366	$ 170,773
Operating expenses	103,273	134,099	163,740
Operating income	$ 2,837	$ 7,267	$ 7,033
International			
Net sales	$ 54,297	$ 65,866	$ 74,723
Operating expenses	57,359	68,008	76,416
Operating income (loss)	$ (3,062)	$ (2,142)	$ (1,693)
AWS			
Net sales	$ 17,459	$ 25,655	$ 35,026
Operating expenses	13,128	18,359	25,825
Operating income	$ 4,331	$ 7,296	$ 9,201
Consolidated			
Net sales	$ 177,866	$ 232,887	$ 280,522
Operating expenses	173,760	220,466	265,981
Operating income	4,106	12,421	14,541
Total non-operating income (expense)	(300)	(1,160)	(565)
Provision for income taxes	(769)	(1,197)	(2,374)
Equity-method investment activity, net of tax	(4)	9	(14)
Net income	$ 3,033	$ 10,073	$ 11,588

Table 4.4.2 indicates how the profitability of Amazon differs when analyzing the different segments of the company.

Even if AWS accounts for 12% of Amazon's Net Sales, 63% of Operating Income comes from Web Services. Further considerations are needed:

1) As I mentioned before, Amazon is confident about the fact that the costs related to the AWS segment will leap less than they have precedently raised. AWS has doubled its revenue from 2017 to 2019. The upward trend is expected to continue. Hence, AWS profitability is likely to increase.
2) If the data analytics is effective, AWS growth will probably affect the Retail segments (North America and International) through the increase in operations efficiency. Therefore, the value of AWS should be computed as the Cash Flow produced by AWS itself and the additional CF produced by the retail segments that have benefited from AWS services.

4.4.3 Bitcoin as a digital asset: Tesla

In February 2021, Tesla has announced the purchase of $1,5 Billion of Bitcoin[131]. Later on, CEO Elon Musk said he would start accepting it as a payment method for Tesla's

[130] Amazon report 2019, pag. 67

products. Tesla later suspended vehicle purchases using Bitcoin due to environmental concerns[132]. This announcement made the Bitcoin price drop dramatically[133].

In April 2021, Musk sold $272 Billion of Bitcoin, boosting his company's profits[134]. On May 17 2021, Musk claimed that Tesla has not sold any Bitcoin[135].

The interesting aspect of this operation is the accounting of Bitcoins in the Balance Sheet.

Tesla stated that Bitcoin has been accounted for as a "Digital Asset"[136]. In the Balance Sheet, Digital Assets are to be booked as Intangibles-Goodwill. Also, these assets initially recorded as a cost, are to be remeasured on the consolidated Balance Sheet[137].

The event has been immediately taken as controversial because, even if the recording of Bitcoin as an Asset did not happen in the GAAP framework, it could become an accounting precedent[138]. In other words, if Tesla has recorded an extremely volatile Asset like Bitcoin[139], the value of internally generated assets such as data could potentially appear as Non-GAAP items in the future.

4.5 The investor's point of view

In this section, I have described how, in my opinion, investors should view data firms and what parameters they should take into account when evaluating them

4.5.1 FCF forecast

There are two possible ways to estimate the future Cash Flow of a firm.

1) One could estimate how the investments made will boost Revenues without involving a leap in costs. Considering the cost structure of data firms (see Chapter 2) their business models (see Chapter 3), and the empirical analysis (see sections 4.3 and 4.4), it is unlikely that the Gross Margin of these companies will decrease. In my view, the Gross Margin is a good measure of

[131] Tesla buys $1.5 billion in bitcoin, plans to accept it as payment (cnbc.com)
[132] Tesla will no longer accept Bitcoin over climate concerns, says Musk - BBC News
[133] Bitcoin (BTC) price falls after Tesla stops car purchases with crypto (cnbc.com)
[134] Tesla's bitcoin speculation helped boost profits by $101 million in Q1 (cnbc.com)
[135] 'Tesla has not sold any Bitcoin', says Elon Musk to end speculations on company's position (livemint.com)
[136] Inline XBRL Viewer (sec.gov)
[137] Tesla and bitcoin: the accounting | Financial Times (ft.com)
[138] Tesla Bitcoin Bet Exposes Limits of Crypto Accounting Rules (bloombergtax.com)
[139] Bitcoin price today, BTC live marketcap, chart, and info | CoinMarketCap

the profitability of a business model, because it only entails Revenues and Cost of Revenues (COGS). As for the Net Margin, it also considers other Expenses, Interests, and Taxes. Hence, it is a good measure for the profitability of a firm as a whole. This can include expenses not to be ascribed to the business model and financial mismanagement[140]. If investors are confident about their financial skills, they should be more concerned about a low Gross Margin than a low Net Margin.

2) Considering the newness of the data industry, the characteristics of the data business model, and the macroeconomic predictions, one could forecast a company's Cash Flow with a top-down approach. In other words, one could compute the current company's market share (Revenue/Aggregate Industry Revenue) and, assuming that the firm will have the same market share after X years, estimating the future Revenue. The formula that can be used is:

$$Revenue_T = Revenue_0 * (1 + g) * (1 + Increase\ in\ MS)$$

Where

$Revenue_T$ = Revenue at the end of the forecast

$Revenue_0$ = Initial Revenue/Current Revenue

g= growth of the industry in time interval = T

Increase in MS= increase in Market Share of the firm

Considering the data business model, the companies with a high Market Share (Revenue) are likely to consolidate their position[141] while smaller data firms need to do something that makes them stand out[142].

4.5.2 Market valuation

Considering the monopolizing attitude of the so-called FAANG stocks[143] , and the high initial fixed costs required to make the data business model work, a market analysis is also required.

[140] How does gross margin and net margin differ? (investopedia.com)
[141] Data Monopolists Like Google Are Threatening the Economy (hbr.org)
[142] Featured | Data Monopolies | SBI (strategicbusinessinsights.com)

Perhaps the most useful tool to analyze the attractiveness of a market is Porter's Five Forces analysis.

Graph 4.5.2 Porter's Five Forces

The five elements of this analysis are[144][145][146]:

1) Competitive Rivalry. In this phase of the data market, I personally think the degree of rivalry is quite high. Data firms know the importance of gaining a high market share in the data industry. The higher the number of participants in the platform, the higher the value of the output (see Chapter 3). Therefore they invest a considerable amount of money in purchasing data and AI companies that could help to the consolidation in the market (see section 4.2.1). in the future, I believe this "clash" for the data monopoly will alleviate.
2) The threat of New Entry. The high initial fixed costs required to make the data business model work is an enormous hurdle. However, the fact that data companies gained a significant market share in such a small time (see Chapter 1) could suggest that new firms may possibly replace the existing ones. The digital economy is evolving at an unprecedented pace. Hence, I believe the threat of new entry is moderately low.

[143] FAANG Stocks Definition (investopedia.com)
[144] Porter's 5 Forces Definition: Analyzing Businesses (investopedia.com)
[145] Porter's Five Forces Revisited: Are There Really Five Forces? (oxfordcollegeofprocurementandsupply.com)
[146] How Competitive Forces Shape Strategy (hbr.org)

3) The threat of Substitution. In the digital economy, no alternative asset can replace data as the core commodity. As I mentioned in Chapter 1, data is the "new oil". At the moment, this threat is low but new technologies to analyze and collect data could be invented in the future.
4) Power of suppliers. I do not see a big threat. In many cases, data firms are also computers producing firms (Apple, Microsoft, IBM). Also, the most valuable asset that data firms produce (data) is internally generated. Hence, I think this threat is relatively small.
5) Power of buyers. Considering that the economic buyer and the end-user are usually two different individuals (see Chapter 3). The buyers, that are sometimes big companies, have a certain power. So this threat is moderate. The power of the data business model (and the digital platforms) is to get data from the users, transform the data into information, and give information back to the users, so they can use it for their own business/activities. It is a win-win business model.

4.5.3 Are listed data companies mature firms?

The difficulty in evaluating data firms may be caused by the fact that most data firms are newborn companies.

Most startups are difficult to evaluate because it is almost impossible to estimate their future Revenues, Costs, and consequently the Cash Flow (Harvard Business Review 2013). Therefore, a new valuation methodology has become extremely popular in the USA: the Lean Startup.

The concept of this valuation methodology is that the value of a newborn business is not derived from a meaningless business plan based on untested hypotheses and random guesses (Harvard Business Review 2013). As Mike Tyson once said: "Everyone has a plan till they get punched in the mouth"[147]. Business plans rarely survive after the first contact with customers because a newborn business is not the small version of an existing one (Harvard Business Review 2013). Therefore, it is more useful to evaluate

[147] (2) Mike Tyson su Twitter: "Everyone has a plan till they get punched in the mouth. #miketyson #vintagetyson https://t.co/yJgHgqxrkK" / Twitter

startups by the quality of the hypotheses that are contained in the business model. Examples of these hypotheses include how to get customers, what value is delivered to customers, through which channels customers are reached, and so on[148].

The quality (and value) of a business is given by the quality of the business model. If everything makes sense, the business is likely to be profitable (Harvard Business Review 2013).

This valuation method could be a complementary tool to be used in data corporate valuations (Harvard Business Review 2013). Analyzing data companies' business models (Chapters 2 and 3) is crucial to test the hypotheses as well as to have a better understanding of their financial statements (section 4.3).

4.6 Valuation conclusions

From the point of view of the Equity holders[149], the data market as a whole is an interesting investment opportunity.

The industry is expected to grow considerably in the future (see Chapter 1), the business model of data firms is likely to show more benefits when the number of customers[150] will increase (see Chapter 3), and data companies are on average more profitable than the others (see section 4.3). However, the high initial investment (see Chapter 2) required to sustain the growth of these companies could be a drawback to consider. Also, the high amount of accounting adjustments needed to evaluate these firms properly[151] could be time-consuming and make investors indecisive on whether investing in these firms would be appropriate.

In my view, the main determinants of the value of data companies are:

- The expected growth rate of the Industry (Macroeconomic perspective)
- The innovative, profitable, and (apparently[152]) flawless business model
- The increasing investments in digitalization carried out by institutions[153]

[148] Business Model Canvas - Wikipedia
[149] Investors and existing shareholders
[150] The customers of data companies can be intended as users of a platform (see Chapter 3)
[151] See sections 4.1.4, 4.1.5, 4.1.6
[152] See section 4.5.2

- The connection between Big Data and the Green Economy, which is another growing sector[154]
- The lower leverage[155]
- The appeal in terms of Return on Equity[156]

On the other hand:

- Data firms are often overvalued [157], especially the US stocks[158]
- The considerable initial investment required to sustain high growth rates could make investors reluctant
- Data firms' Payout Ratio is lower than that of the Market (24% vs 45%)[159]
- The high series of accounting peculiarities make data firms difficult to evaluate. Hence, a good stock picking strategy is difficult to carry out.

Therefore, investors should:

- Take into account the average growth rate and WACC of the data industry. Considering that the industry as a whole is growing, they should not invest in firms with a lower expected growth rate and higher cost of capital. What makes data firms stand out is not the cost of capital, it is the expected growth rate. Also, even if data firms pay lower dividends, they are generally more profitable. Hence, future dividends paid are likely to be higher. In my opinion, the Gross Margin is a key indicator because it shows how profitable a business model is (see section 4.5.1)
- Consider investing in ETFs, which is a good way to match the market's performance without entering difficult stock picking strategies and risking selecting the wrong stock[160]
- Be willing to accept that investing in data firms is a long-term game. Income investors that seek a regular income coming from Dividends would be better

[153] Recovery plan for Europe | European Commission (europa.eu)
[154] Big data and natural environment. How does different data support different green strategies? - ScienceDirect
[155] Since Debt is senior in respect of Equity, shareholders have a higher probability to receive a Dividend
[156] ROE-COE is higher. See Table 4.3.2
[157] Goldman Sachs: Technology stocks are overvalued (cnbc.com)
[158] Is there a bubble in US tech stocks—and will it pop in 2021? — Quartz (qz.com)
[159] Damodaran Online: Home Page for Aswath Damodaran (nyu.edu)
[160] How to Invest in ETFs (Exchange-Traded Funds) | The Motley Fool

recuse themselves from the data industry because one of the key determinants of data firms' value is optimistic expectations

- Invest in US stocks if they do not mind overpaying stocks[161]. Consider purchasing non-US stocks to seek bargains. However, the US market is the leader of Big Data (see Table 4.3.1), so the probability of buying a good stock is higher when purchasing a US stock
- Studying the adjacent industries. Data is the “new oil”, so it is the commodity that makes all of the digital industries work. Understanding how each of these interacts with the data industry is a good way to figure out if a specific data company will be a market leader. Also, downward trends of the adjacent industries are a possible threat to data firms.

[161] From the traditional point of view

Chapter 5 - Alternative data valuation

So far, I have described how data (and data companies) can be evaluated from the point of view of the investors. In this chapter, I have reported some alternative valuations.

First of all, I have described how creditors may decide to lend money to data firms, considering that data is an asset protected by privacy rights.

After, I have reported the most common corporate rating systems. I have applied these ratings to data firms.

I have also described how legal and technological changes may affect corporate valuations in the data industry.

Finally, I have analyzed the role and the utility of the value investing methodology and whether it still works.

5.1 The creditor's point of view

In some legislations, such as the European Union, the authorities recognize a series of privacy rights to individuals. These rights make sure individuals have full control over the data they provide to organizations. The European General Data Protection Regulation (GDPR) ensures privacy rights such as[162]:

1) The right to be informed
2) The right to access
3) The right to rectification
4) The right to erasure
5) The right to restrict processing
6) The right to data portability
7) The right to object
8) Rights concerning automated decision making and profiling

As I already mentioned in Chapter 2, being data compliant can be very expensive. For the creditor, this might not be a problem, because creditors only lend money to data

[162] What is GDPR, the EU's new data protection law? - GDPR.eu

firms. This does not involve GDPR compliance costs per se. However, if the borrower defaults in his payments, the creditor becomes the owner of the collateral asset that the borrower had offered to the lender as security for the loan[163]. Typically, collateral assets include Real Estate assets and other tangible assets. Besides, collaterals can normally be resold more easily because they are not protected by privacy rights. Liquidity (how easily a collateral asset can be converted into cash) is one of the key determinants of a collateral's value[164].

Considering that the law ensures a lot of privacy rights to individuals, that Non-performing Loans (NPLs) are likely to increase[165] , and that being GDPR compliant is costly (see Chapter 2), banks may want to consider these aspects in their credit risk models. Even if data protection law does not prevent organizations from selling data (as long as they have the consent of data owners), buyers need to also have a lawful basis for owning and processing the data[166]. These aspects that make lending money to data firms complicated might be the reason why Information Services firms have a lower Debt-to-Equity ratio than the market (9,41% vs 62,56%)[167].

One of the most powerful tools to estimate the probability of bankruptcy of a firm is the Altman Z-score. The Z-score is computed as follows[168]:

$$1.2\frac{Working\ Capital}{Total\ Assets} + 1{,}4\frac{Retained\ Earnings}{Total\ Assets} + 3{,}3\frac{EBIT}{Total\ Assets} + 0{,}6\frac{MVE}{Total\ Liabilities} + 1\frac{Sales}{Total\ Assets}$$

The formula is composed of five ratios[169].

1) Working Capital/Total Assets. Working Capital is the difference between Current Assets and Current Liabilities. A negative ratio means that a firm will struggle to meet short-term financial obligations.
2) Retained Earnings/Total Assets. A high level could show that a company relies on its profitability for reinvesting, rather than borrowing.

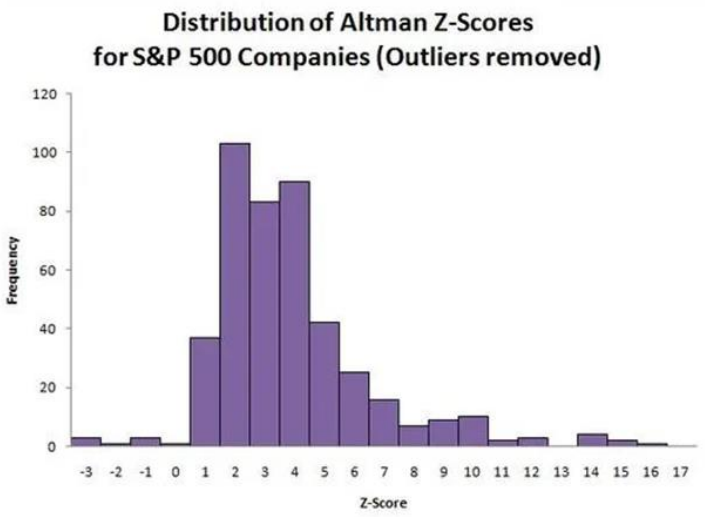
[163] Collateral - Definition, Types, Collateral vs. Security (corporatefinanceinstitute.com)
[164] The Risks of Asset-Based Loans (thebalancesmb.com)
[165] Non-performing loans (europa.eu)
[166] The value of data, policy implication , Bennett Institute for Public Policy of Cambridge, 2020
[167] Damodaran Online: Home Page for Aswath Damodaran (nyu.edu)
[168] Altman Z-Score Definition (investopedia.com)
[169] Altman's Z-Score Model - Overview, Formula, Interpretation (corporatefinanceinstitute.com)

3) Earnings Before Interests and Taxes/Total Assets. The ability to generate profits from operations (EBIT) demonstrates that a company can stay profitable and meet its financial obligations.
4) Market Value of Equity/Total Liabilities. A higher Market Capitalization means high investor confidence in a firm's financial stability.
5) Sales/Total Assets. The higher the ratio, the lower the investment needed to generate Revenues.

A score below 1,8 means that the company has a serious bankruptcy risk, while firms with a score above 3 are safe. Banks may consider readjusting the coefficients[170], taking into account the accounting peculiarities of the data firms (see Chapter 4). For example, a study reported by the Indonesian Bulletin of Electrical Engineering and Informatics shows how the default rate of the 23 listed technology firms analyzed is strongly correlated to the ratio for Market Value of Equity to Total Liabilities[171].

Graph 5.1.1 Altman Z-score S&P 500 companies

Altman Z-Scores for S&P 500 Companies (With Outliers Removed)	
Mean	3.37
Standard Error	0.13
Median	2.89
Mode	1.9
Standard Deviation	2.71
Sample Variance	7.33
Skewness	1.40
Range	23.86
Minimum	-7.93
Maximum	15.93



Graph 5.1.1 illustrates how S&P 500 Companies perform quite well in terms of Altman Z-score[172]. The minimum score of the dataset analyzed by Yahoo Finance is -7,93, the maximum is 15,93. The reported mean is 3,37.

[170] Examine the financial health of the listed technology companies in Malaysia using Altman's Z-score test (scitation.org)
[171] Analysis on the Performance of Technology Companies with Z-score Model | Weng Hoe | Bulletin of Electrical Engineering and Informatics (beei.org)
[172] Tech Companies Reach High Altman Z-Scores (yahoo.com)

This study is particularly relevant because some of the top 10 stocks of this index are companies for which data is a core asset (see Table 5.1.2)[173].

Table 5.1.2 S&P 500: Top 10 constituents

Top 10 Constituents By Index Weight

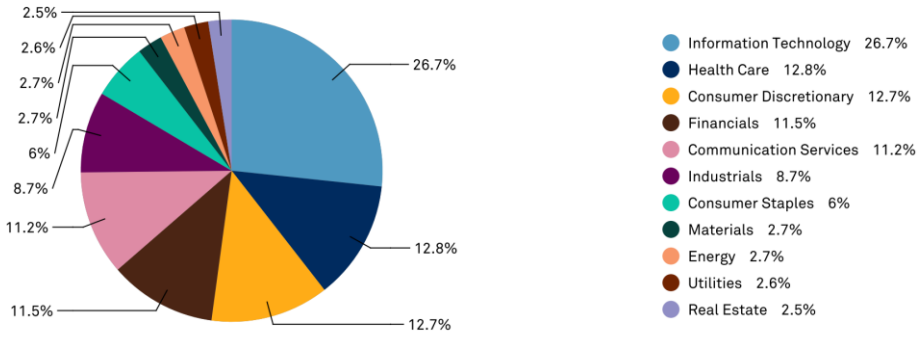
CONSTITUENT	SYMBOL	SECTOR*
Apple Inc.	AAPL	Information Technology
Microsoft Corp	MSFT	Information Technology
Amazon.com Inc	AMZN	Consumer Discretionary
Facebook Inc A	FB	Communication Services
Alphabet Inc A	GOOGL	Communication Services
Alphabet Inc C	GOOG	Communication Services
Tesla, Inc	TSLA	Consumer Discretionary
Berkshire Hathaway B	BRK.B	Financials
JP Morgan Chase & Co	JPM	Financials
Johnson & Johnson	JNJ	Health Care

*Based on GICS® sectors

In addition, the first sector of the index is Information Technology[174] (Graph 5.1.3).

Graph 5.1.3 S&P 500: Sector breakdown

Sector* Breakdown



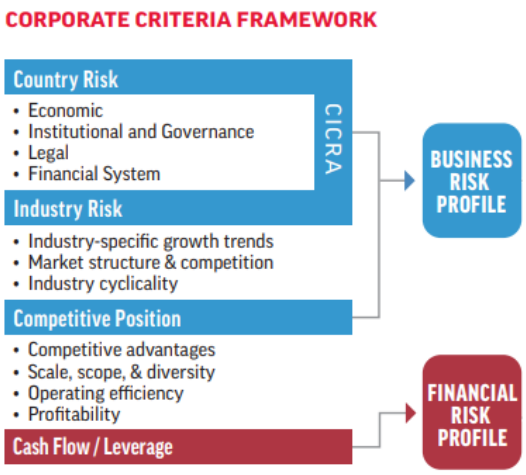
*Based on GICS® sectors

The weightings for each sector of the index are rounded to the nearest tenth of a percent; therefore, the aggregate weights for the index may not equal 100%.

[173] Apple, Microsoft, Amazon, Facebook, Alphabet
[174] Source: S&P 500 Fact Sheet

5.2 The rating agency's point of view

Graph 5.2.1 S&P corporate rating methodology 1

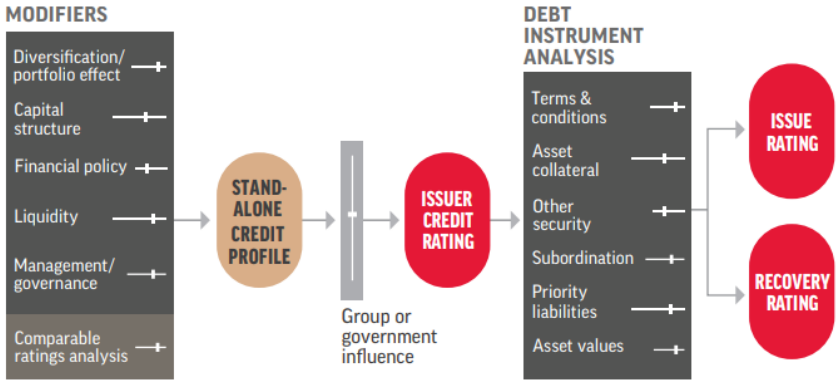


The first step of the S&P corporate rating methodology[175] is identifying the Country Risk, the Industry Risk, and the Competitive Position of the firm in the Industry. These three factors indicate the Business Risk Profile. The Cash Flow/Debt ratio indicates the Financial Risk Profile.

The second step consists of analyzing additional factors (capital structure, financial policy, liquidity) that could influence the Credit Profile. Once this procedure is completed, a Credit Rating is established.

The last assessment regards the Debt's conditions, collateral assets, subordination, and other aspects that could affect the credit risk. Finally, a rating is issued[176]. A recovery rating[177] is also indicated.

Graph 5.2.2 S&P corporate rating methodology 2



As for the Business Risk profile and the Financial Risk Profile of data companies, I can affirm that:

[175] 8fd4392a-4aae-4669-bd74-a9b86e18d781 (spratings.com)
[176] The rating goes from AAA (Extremely Strong) to D (Default)
[177] Rating for speculative-grade issuers

- The profile can be affected dramatically by antitrust procedures and restrictive regulations in terms of data privacy. Constraints on data usage may be imposed by institutions. These could make the profitability of data firms decrease because GDPR compliance costs would skyrocket and, in case of excessive limitations on the use/collection of data, the value of the information would decline. Considering the cost structure of these companies [178], if they cannot extract a minimum amount of valuable data from their customers[179], they cannot be profitable
- The Industry-specific growth trend is upward. However, the birth of new technologies to collect and analyze data could threaten the data firms severely
- The business model of data companies[180] is likely to make small/medium firms worse off in the competition. One key determinant of value for data firms is the business size[181]. Small firms that offer a competitive advantage to large companies should consider being purchased instead of engaging in an asymmetric competition[182]
- Data companies are more profitable and less leveraged than the other firms[183].

The Terms & Conditions under which data can be collected and used, as well as the difficulty in considering data as a suitable collateral asset for loans,[184] are to be considered. Finally, the value of the data owned by a specific company depends on the buyer of the data[185]. Hence, the asset value must be determined case by case.

5.3 Data valuation evolution

In my opinion, the methodologies to evaluate data companies may evolve in the future. The change in legal frameworks, the rotation of macroeconomic trends, and the technological changes may affect the validity of the analysis that I have reported so

[178] Significative initial investments and high fixed costs (see Chapter 2)
[179] Estimated as a Net Value
[180] See Chapter 2 (Costs) and Chapter 3 (Revenues and usiness model of data firms)
[181] See Chapter 3 and Chapter 4
[182] See Sections 4.2.1 and 4.4.1
[183] See Chapter 4
[184] See Section 5.1
[185] The value of the data depends on whether the buyer can use the data for its business. See Section 4.2.1

far. Also, these changes could be interpreted as opportunities or threats for data organizations and, consequently, might affect their value.

5.3.1 Legal evolution

Table 5.3.1 shows that there are some barriers (technical, legal, and behavioral) that data companies currently face in collecting, storing, analyzing, and using the data. Up until now, these did not prevent organizations from exploiting data as a business asset. As I already mentioned, the data business model is, so far, a win-win business. The customers can lend their data in exchange for services provided by the data firms. Both parties benefit from this "trade".

Table 5.3.1 Data usage barriers

	Technical barriers	Legal Barriers	Behavioural barriers
Collection	• Uniqueness of the data, or access to it • Supply side: economies of scale, scope, learning by doing, speed • Demand side: network effects and two-sided markets	• Data protection and privacy laws • Data ownership	• Exclusivity agreements • Access prices and conditions • Disabling data collecting software
Storage	• Storage costs	• Data protection and privacy laws	• Lock-in and switching costs
Synthesis and analysis	• Lack of interoperability (including a lack of standardisation) • Analytical tools		
Use	• Inability to locate and reach relevant consumers • Lack of interoperability (including a lack of standardisation)	• Data protection and privacy laws • Anti-discrimination laws	• Contractual limitations

Source: Rubinfeld and Gal (2017[37]); Gal and Rubinfeld (2019[26]); CMA (2016[93])

However, some privacy concerns arose (OECD 2020). Customers tend to underestimate the risks of breaches of privacy when accepting the terms & conditions for the use of digital services. In addition, firms could abuse their dominance by lowering the level of privacy offered to consumers. They are likely to agree to a lower level of privacy because they lack bargaining power and awareness about the risks of this behaviour (OECD 2020). Also, the ownership of data impacts the competitive advantage of a firm in respect of others in the same industry. Hence, a rising number

of Mergers between data companies might raise barriers to entry and raise rivals' costs (OECD 2020).

As a general rule, entrepreneurs always choose to operate in legal environments where they have the strongest rights. However, from an IP Law perspective, the American dominance in the data industry[186] can be explained by the competitive environment in which data organizations operate, regardless of the lower[187] IP protection (Boldrin and Levine 2008).

Therefore, the possible development that data firms should fear the most is that authorities could intervene by blocking the mergers or limiting the freedom to merge and form cartels that could allow organizations to dominate the data market (OECD 2020).

As for data protection, law remedies would not threaten data companies because data is already collected with explicit consent (OECD 2020).

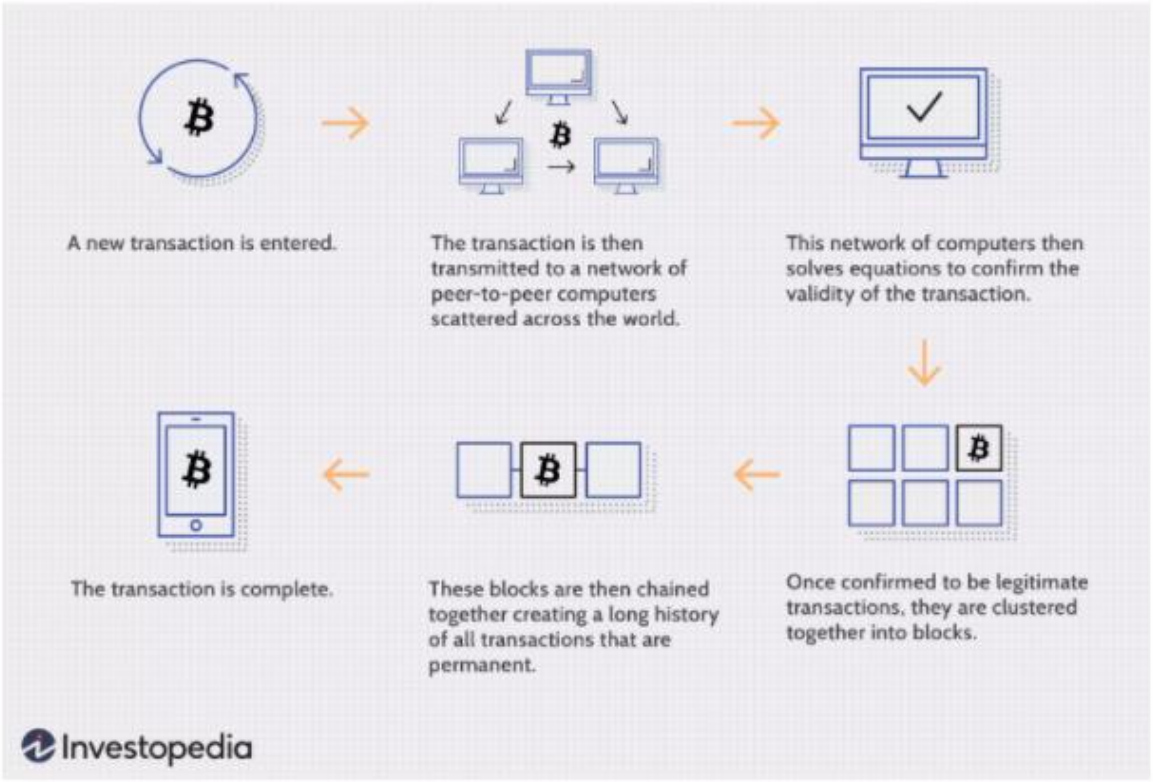
5.3.2 Technological changes

The birth of new technologies that collect, analyze, and allow easy storage of data can seriously threaten the data industry. In this industry, one of the most popular innovations is the Blockchain. The Blockchain is "a shared, immutable ledger that facilitates the process of recording transactions and tracking assets in a business network[188]".

Graph 5.3.2 shows the operation of a Blockchain.

[186] See section 4.3
[187] In respect of the European Union
[188] What is Blockchain Technology - IBM Blockchain | IBM

Graph 5.3.2 Blockchain



Blockchain-based cloud storage makes the data distribution chain decentralized, allowing any authorized peer subject to receive data updated in real-time, without the need for a particular infrastructure[189].

This technology can make cloud storage cheaper (there is no need for a centralized infrastructure), faster, and enhance its security[190]. On the other hand, it lacks regulation, so it has been used for illicit activities[191].

There are some projects, such as Sia[192], that aim to compete with existing storage solutions[193]. The peculiarity of this project is that instead of renting storage from a centralized provider, peers on Sia rent storage from each other[194]. This system can also be used for smart contracts, meaning programs stored on a blockchain that run when some conditions are met[195]. Sia's users can decide what data will be stored and at what price[196]. Also, it allows users to create a database of potential data hosts, and

[189] What Is Blockchain Technology? How Does It Work? | Built In
[190] How Can Blockchain Technology Improve Cloud Storage? - Alibaba Cloud Community
[191] Blockchain Definition: What You Need to Know (investopedia.com)
[192] Sia
[193] Sia whitepaper, pag. 1
[194] Sia whitepaper, pag.1
[195] What are smart contracts on blockchain | IBM
[196] Sia whitepaper, pag. 1

form contracts with reliable ones only[197]. Sia claims that this would facilitate the decentralization of storage[198].

Another Blockchain-based cloud storage project is the world-famous Ethereum. The value proposition offered by Ethereum similar to Sia's[199]. Ethereum's founders claim that the traditional online file storage companies offer unnecessarily expensive solutions[200]. Ethereum allows users to earn some money by "renting" their hard drives. Unused space is not paid for[201].

The rise of Blockchain-based solutions is likely to disrupt many industries[202]. Amongst these, the data industry. Although the increasing interest in the Blockchain could be a threat for data organizations, the interaction between Blockchain and Big Data could make the data industry benefit because this technology brings considerable value to the industry[203]. Besides, some data organizations, like IBM, have already embraced Blockchain to anticipate the future market trend[204].

Therefore, when evaluating a data firm, I recommend considering the disruption risk[205] in the model. The disruption risk is to be inserted not only at the industry level but, more importantly, as a firm-specific risk because the forecasting and anticipation skills differ from company to company.

5.4 Does Value Investing still work?

Tech stocks (such as data companies) are considered growth stocks[206]. These companies tend to be more expensive in terms of PE and Price-to-Sales ratio. Therefore, they should be evaluated using alternative metrics or characteristics like competitive advantage and business models[207]. I think some concepts typical of value investing valuations could be applied to data firms (albeit some claim that value investing is not useful anymore[208]) because some tech stocks are also value

[197] Sia whitepaper, pag. 4
[198] Sia whitepaper, pag. 4
[199] In particular, smart contracts
[200] Ethereum whitepaper, pag. 22
[201] Ethereum whitepaper, pag. 22
[202] 35 Top Blockchain Companies to Know 2021 | Built In
[203] Blockchain Companies are Using Big Data Like Never Before: Here's Why | Hacker Noon
[204] IBM Blockchain - Enterprise Blockchain Solutions & Services | IBM
[205] Threats to a company's long-term viability posed by more agile competitors, increased regulation, changes in consumer tastes and new technologies.
[206] Value vs. Growth Investing: Which Should You Buy? | The Motley Fool
[207] Best Growth Stocks to Buy in Q1 2021 | The Motley Fool
[208] Value Investing's Old Rules 'Have Outlived Their Usefulness,' Academics Argue | Institutional Investor

companies[209]. Besides, buying stocks below their intrinsic value is never a bad idea[210]. For these reasons, I have reported my analysis on value investing in this section. In particular, I described its main rules and applied them to the data industry to investigate whether it could be effective or not.

Value Investing is a strategy that consists of buying stocks that are undervalued by the market because they trade for less than their intrinsic value. The fundamental rules of this investment method are[211]:

- Buy businesses, not stocks. You should ignore trends in stock prices and focus on the fundamentals of the company. A compulsive "buy/sell" strategy is time-consuming and expensive in terms of commissions
- Love the business you buy into. You should try to know as much as you can about a business rather than judging a stock by hurried research. Some companies look different if analyzed beyond the basic ratios
- Invest in companies that you understand. The value of a firm is given by the cash flow that it will generate over time. If you do not understand a business, you can only make random guesses about the future cash flows of that firm. What investors should do is understanding a business so that they are able to forecast the cash flows properly
- Find well-managed companies because good management is an important value-added. The best managers ignore the market and focus on creating long-term shareholder value
- Don't stress over-diversification. Even if you diversify your portfolio, everything gets expensive at the end of a bull market. You should buy as many undervalued stocks as you can
- Be ready to invest in companies that are better than the ones you own. Wait for better times if there are no good investing opportunities
- The market only matters when you enter or exit a position. You want to hold the stocks as long as their fundamentals are strong, regardless of price changes. The longer you avoid capital gains[212] and transaction costs, the better

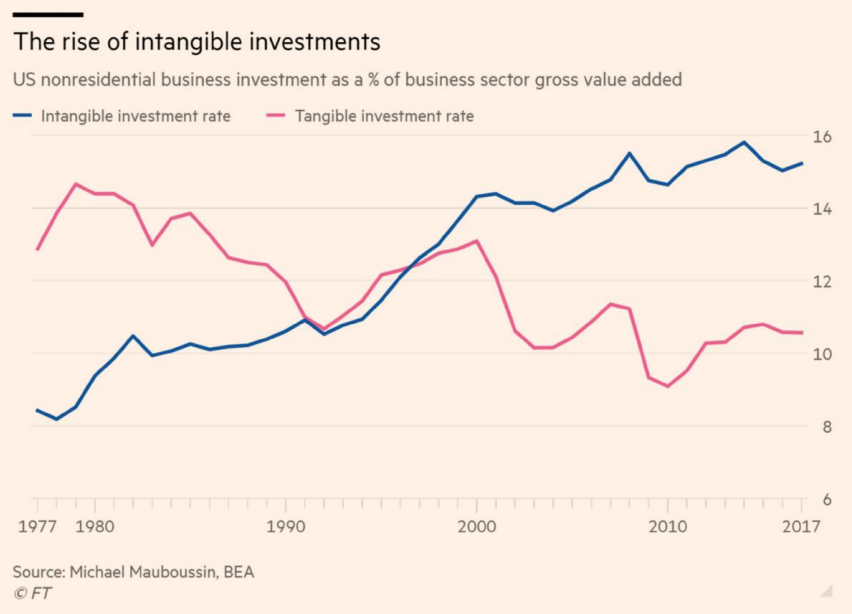
[209] Can value investors profit from tech companies? - The Value Perspective - Schroders
[210] https://www.cfauk.org/pi-listing/an-innovative-approach-to-value-investing#gsc.tab=0
[211] Strategies of Legendary Value Investors (investopedia.com)
[212] Sale price higher than purchase price

- The value investor, unlike the passive investor[213], seeks above-average companies. Therefore, the value investor's return will be considerably higher.

In recent times, some started questioning the validity of the Value Investing method applied to modern markets[214][215][216].

The main criticisms are related to the fact that this investing methodology has not been able to help to understand the value of tech stocks before they have become mainstream. Also, it appears that value investing was a great idea before it was popular. Now that most people invest as value investors, the return of this strategy has decreased[217].

Graph 5.4.1 The rise of intangibles

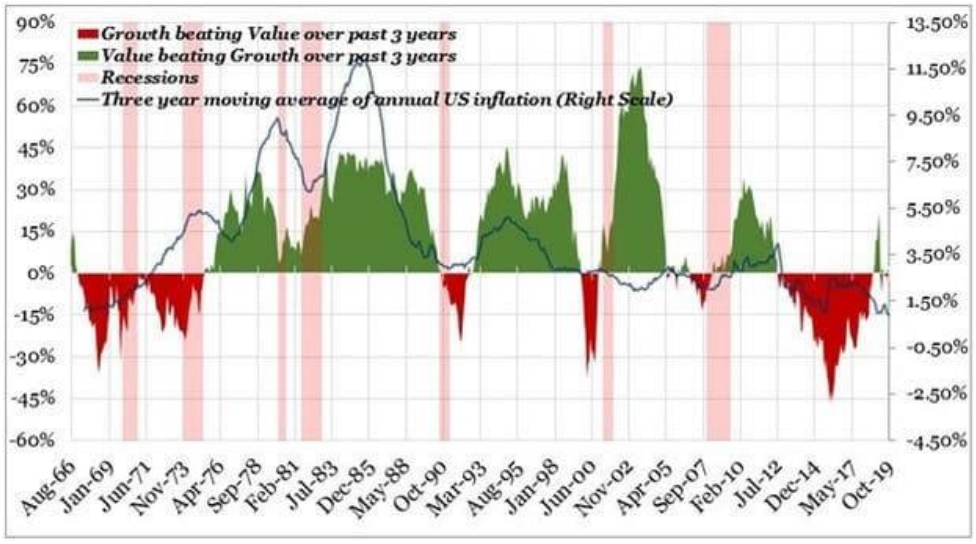


In addition, value investing is based on tangible assets but the meaning of Earnings and Book Value has changed considerably over time[218]. As Graph 5.4.1 shows, investments in intangibles have increased considerably since 1977. In the same time interval, the percentage of US investments in tangible assets has declined. This radical change may indicate the need for a new investing methodology[219].

[213] Passive investing is an investment strategy to maximize returns by minimizing buying and selling
[214] Does value investing still work? | The Economist
[215] Does value investing still make sense? (afr.com)
[216] Does Value Investing Still Work In The Internet Age? (brokenleginvesting.com)
[217] Is Value Investing Dead? Not So Fast, Says A Recent Study (forbes.com)
[218] Why Value Investing still works in markets, Financial Times, 18 November 2020
[219] Why Value Investing Sucks | Institutional Investor

Graph 5.4.2 illustrates the performance of the value investing strategy and the growth investing strategy[220]. The results indicate that value investing outperformed growth investing, with few exceptions, from the 1970s to 2012. However, the return of growth investing strategies has beaten that of value investing in the last decade.

Graph 5.4.2 Growth vs value investing



Annualized 3-Year average monthly U.S. value premium to P/B ratio based value and growth strategies.

Source: Meritocracy Capital Partners of Vancouver.

Despite the criticisms, the principles of value investing still hold. The fact that Earnings and Book Value have partially lost their ability to represent an economic value[221] does not mean that the whole value investing methodology is outdated.

Since the intangible investments are accounted for as expenses[222] and tangibles investments are assets[223], a company investing in intangibles will have lower Earnings and Book Value than a company investing the same amount of money in tangible assets[224]. Making some accounting adjustments and considerations on the peculiarity of data firms before evaluating them[225] could help to overcome this problem.

[220] Growth investing is an investment style and strategy that is focused on increasing an investor's capital. Growth investors typically invest in growth stocks—that is, young or small companies whose earnings are expected to increase at an above-average rate compared to their industry sector or the overall market.
[221] Is value investing dead?, Wealth & Investing - THE BUSINESS TIMES
[222] On the Income Statement
[223] On the Balance Sheet
[224] Why Value Investing still works in markets, Financial Times, 18 November 2020
[225] See Chapter 4

To conclude, I think the value investing method is still effective, and its principles are applicable to data firms, but it needs to be nourished with some additional considerations.

1) The base principle of value investing is correct[226]. One should not buy stocks that are overpriced. What has changed is not this principle but the way investors should measure value. In other words, the traditional valuation multiples are not suitable for data firms because many items are not accounted for in the financial statements the traditional way. Hence, some accounting adjustments are needed when evaluating data firms.
2) Since Balance Sheets and Income Statements do not describe the value of data companies correctly[227], one could also consider evaluating data firms on the basis of their business plans and business models' validity[228].
3) Trying to know the business you buy into is useful but one should be aware that technological changes have been faster and faster in the past. The fact that technological innovation follows an exponential pathway indicates that technological disruption will accelerate. Sometimes, even the experts cannot predict how fast this process can be[229]. Even a recently-born market like the data industry will risk being disrupted soon[230]. For this reason, I believe that an investor should try to know as much as possible about an industry. In particular, it is useful to understand the business model and the costs/revenues structure of the firms. After doing that, they should invest in the most appealing companies quickly, in order to benefit from a future increase in their market value. I do not think that spending too much time in this phase would be beneficial.
4) It is useful to insert a disruption risk index in corporate valuations. This index should take into account both legal and technological risks[231].
5) Investors should read the Non-GAAP items carefully[232]. Internally-generated assets (ex: data) are not considered as assets by current accounting

[226] Value investing still makes sense (and will always do) (mapfre.com)
[227] See Chapter 4
[228] See section 4.5.3
[229] The Speed of Disruption and Impact on Business - The Fourth Industrial Revolution Has Begun | Gen Re
[230] See section 5.3
[231] See section 5.3
[232] See section 4.4.1

standards[233] but it is important to know their value when purchasing a firm whose business strongly relies upon these assets.

6) Considering that the value of data is in its usage[234], investors should determine whether the data they intend to purchase can integrate well into their existing business[235].
7) Data is a technological asset. For this reason, a technological appraisal done by data experts should be included in the valuation.
8) Data is considered as the "new oil" or, in other words, the commodity that makes the digital economy work[236]. It is also useful to add a brief evaluation of data's adjacent industries to understand how the demand for data can evolve over time.

[233] See Chapter 1
[234] See Chapter 4
[235] See section 4.4.1
[236] See Chapter 1

Conclusion

My Thesis is a step-by-step analysis of the data industry and, of course, the most important data firms.

First of all, I have tried to understand the Industry as a whole, its main characteristics, and its predicted growth. Before even starting this, I spent some time learning what data is and how it can be classified.

After, I have analyzed the costs and revenue structure of each firm. It is particularly important to figure out that data companies must bear a considerable amount of fixed costs, like most digital companies, and that they benefit from the network effect, typical of digital platforms.

Besides, data firms are overpriced if analyzed with the traditional valuation methods. Also, these methods are not suitable for data firms because they are based on accounting indicators that are not representative of the real profitability of firms with intangible assets, such as data.

Therefore, it is important to develop a new valuation methodology. In Chapters 4 and 5 I described the rules that investors, creditors, and rating firms should follow to evaluate or rate data companies.

Since the financial statements only represent a partial view of the whole picture, it is crucial to investigate whether some alternative valuations can be made.

Further research can be made in assessing the role of competitive positioning and estimating the growth of the firms, rather than on the analysis of meaningless indicators. Non-GAAP items must be considered carefully when evaluating data firms. In addition, analyzing the quality of a data firm business model is crucial, because most data firms are young firms that will probably dominate the market in the future. So it is not always possible to represent the value of these firms in quantitative terms. Even when using quantitative metrics, investors should use those that suit young companies. For instance, it could be more appropriate to use the Free Cash Flow Yield instead of the Discounted Cash Flow, because data firms have some characteristics typical of startups.

As for multiples and ratios, investors should consider “inventing” new indicators that are more effective in capturing the real value (or profitability) of data firms. For

example, one may want to find the CAPEX (adjusted) which takes into account some capitalized operating expenses.

To conclude, one-size-fits-all valuations don't work as they used to because each sector (and each company in that sector) has its own peculiar characteristics. This is why I suggest creating an valuation methodology suitable for each sector and sub-sector.

As a final remark, this Thesis is not meant to be a final research on data valuation. The birth of the data Industry is quite recent and this Industry will probably evolve in the future. Hence, my research is to be considered as an initial attempt to establish a set of analyses to be made in order to analyze data firms properly.

References

Amazon (2019), Annual Report

Bain & Company (2013), The value of big data: how analytics differentiates winners

Banco de España (2007), How do intangible assets create economic value? An application to banks

Barua A. et al., (2012), Measuring the business impacts of effective data, The University of Texas at Austin

Bennett Institute for Public Policy of Cambridge (2020), The value of data: policy implication

Boldrin, M., & Levine, D. (2008). Against Intellectual Monopoly. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511510854

Damodaran A., (2001), The dark side of valuation

Damodaran A., (2009), Valuing companies with intangible assets

Demchenko Y. et al, (2018), Data as economic goods, ITU Journal, Special issue No. 2, 23 Nov. 2018

Dun&Bradstreet (2015), The big payback on quality data

Eisenson A. (2017), Bloomberg Intelligence report, 05/05/2017

Faroukhi A. et al., (2020), Big data monetization throughout big data value chain: a comprehensive review, Journal of Big Data, article number 3, 2020

Gartner (2011), Measuring the business value of data quality

Gartner (2015), Big Data Industry Insights

Geissdoerfer M. et al., (2020), Circular business models: a review, Journal of cleaner production 2020

GSMA (2018), The data value chain

Harvard Business Review (2013), Why the lean start-up changes everything

IBM (2019), Annual Report

IHS Markit (2020), Annual Report

IMDA & PDPC (2019), Guide to data valuation for data sharing

KPMG (2015), Data and analytics: a new driver of performance and valuation

Marr B., (2016), Big Data in practice, John Wiley & Sons Ltd, United Kingdom

Marr B., (2017), Data Strategy, Kogan Page, United States

MIT Technology Review (2016), The rise of data capital

Noah Consulting (2013), Quantitative value of data and data management

Noselli C. (2016), Bloomberg Intelligence report, 24/10/2016

OECD (2020), Consumer data rights and competition

Oracle (2013), Bringing the value of big data to the enterprise

Osterwalder A. and Pigneur Y., (2010), Business model generation, John Wiley & Sons Inc., New Jersey

Parker G. et al., (2016), Platform Revolution, W. W. Northon & Company, New York

PWC (2014), Deciding with data

PWC (2019), Putting a value on data

SAS (2010), Valuing information as an asset

Schmarzo B. and Sidaoui M., (2017), Applying economic concepts to big data

Schmarzo B., (2015), Big Data MBA, John Wiley & Sons Inc., Indianapolis

Singh M. (2017), Bloomberg Intelligence report, 11/12/2017

Stander B. (2015), The modern asset: big data and information valuation, Stellenbosch University

United Nations (2019), Data Economy: Radical transformation or dystopia?

United Nations (2019), Digital Economy Report

Verhoef P. et al., (2016), Creating value with big data analytics, Routledge, London and New York

Wagner D., (2020), Economic patterns in a world with artificial intelligence, Evolutionary and Institutional Economic Review 2020

Willamette (2018), Fundamentals of the Asset-Based business valuation approach

Consulted Websites

AIP Conference Proceedings. https://aip.scitation.org/

Alibaba Cloud. https://www.alibabacloud.com/

Barron's. https://www.barrons.com/

BBC. https://www.bbc.com/

BBVA. https://www.bbva.com/

BEEI. https://beei.org/

Bernard Marr. https://www.bernardmarr.com/

Bizfluent. https://bizfluent.com/

Bloomberg Tax. https://news.bloombergtax.com/

Bloomberg. https://www.bloomberg.com/

Broken Leg Investing. https://www.brokenleginvesting.com/

Builtin. https://builtin.com/

Business Today. https://www.businesstoday.in/

BusinessInsider. https://www.businessinsider.com/

CFA Institute. https://www.cfainstitute.org/

CFA Society United Kingdom. https://www.cfauk.org

CFO.University. https://cfo.university/

CNBC. https://www.cnbc.com/

CNN Business. https://edition.cnn.com/

CoinMarketCap. https://coinmarketcap.com/

ComputerWeekly. https://www.computerweekly.com/

Corporate Finance Institute. https://corporatefinanceinstitute.com/

Damodaran Online. http://pages.stern.nyu.edu/~adamodar/

Dell. https://infocus.delltechnologies.com/

Ethereum. https://ethereum.org/en/

European Central Bank. https://www.bankingsupervision.europa.eu/

European Commission. https://ec.europa.eu/

Financial Review. https://www.afr.com/

Financial Times. https://www.ft.com/

FitchRatings. https://www.fitchratings.com/

Forbes. https://www.forbes.com/

Fortune. https://fortune.com/

Gartner. https://www.gartner.com/

GDPR.EU. https://gdpr.eu/

GeekWire. https://www.geekwire.com/

Gen Re. https://www.genre.com/

Grant Thornton. https://www.grantthornton.co.uk/

Groww. https://groww.in/

Hackernoon. https://hackernoon.com/

Harvard Business Review. https://hbr.org/

HIS Markit. https://investor.ihsmarkit.com/

Hypersurfaces. https://www.hypersurfaces.com/

IBM. https://www.ibm.com/

Information Age. https://www.information-age.com/

Institutional Investor. https://www.institutionalinvestor.com/

Internazionale. https://www.internazionale.it/

Investopedia. https://www.investopedia.com/

Kaggle. https://www.kaggle.com/

LinkedIn. https://www.linkedin.com/

LivingCircular Veolia. https://www.livingcircular.veolia.com/

Mapfre. https://noticias.mapfre.com/

MarketWatch. https://www.marketwatch.com/

McKinsey. https://www.mckinsey.com/

Mint. https://www.livemint.com/

Nasdaq. https://www.nasdaq.com/

Oxford College of Procurement & Supply. www.oxfordcollegeofprocurementandsupply.com

Payments Journal. https://www.paymentsjournal.com/

PR Newswire. https://www.prnewswire.com/

Precisely. https://www.precisely.com/

Quartz. https://qz.com/

Red Hat. https://www.redhat.com/

Reuters Events. https://www.reutersevents.com/

Reuters. https://www.reuters.com/

S&P Global. https://www.spglobal.com/

S&P Ratings. https://www.spratings.com/

Schroders. https://www.schroders.com/

ScienceDirect. https://www.sciencedirect.com/

SEC. https://www.sec.gov/

Seeking Alpha. https://seekingalpha.com/

Sia. https://sia.tech/

Snowflake. https://www.snowflake.com/

Statista. https://www.statista.com/

Strategic Business Insights. http://www.strategicbusinessinsights.com/

Talend. https://www.talend.com/

TechCrunch. https://techcrunch.com/

The balance small business. https://www.thebalancesmb.com/

The Business Times. https://www.businesstimes.com.sg/

The Economic Times India. https://economictimes.indiatimes.com/

The Economist. https://www.economist.com/

The Motley Fool. https://www.fool.com/

Tiempo Development. https://www.tiempodev.com/

Torys. https://www.torys.com/

Twitter. https://twitter.com/

UK Value Investor. https://www.ukvalueinvestor.com/

Villanova University. https://taxandbusinessonline.villanova.edu/

Wikipedia. https://en.wikipedia.org/

Yahoo Finance. https://it.finance.yahoo.com/