

**Sentiment analysis of financial tweets: what impact  
on the share price of the companies concerned?**



# Acknowledgments

First, I would like to thank my promoter, Mr. Ashwin Ittoo, for his time and guidance.

I also want to express my gratitude to the members of the jury, Mr Cedric Gilain and Mr Nicolas Falmagne, who took the time to read this dissertation.

Also, I would like to acknowledge the Stack Overflow community.

Finally, I am very grateful to my family and friends for their support



# Table of contents

1	Introduction .....	1
2	Theoretical framework .....	5
2.1	Sentiment analysis: .....	5
2.1.1	Machine learning approach: .....	6
2.1.2	Lexicon-based approach .....	9
3	Literature review.....	11
4	Analysis .....	17
4.1	Companies selection .....	17
4.1.1	Ab Inbev .....	17
4.1.2	Galapagos .....	24
4.2	Data.....	30
4.3	Sentiment analysis .....	31
4.4	Correlation .....	32
5	Results & limitations .....	45
6	Ethical part.....	47
7	Conclusion.....	51
8	Annexe.....	I
9	Bibliographie .....	III



# 1 INTRODUCTION

---

The desire to know if stock markets are predictable has long attracted the interest of academic research and businesses. The first works on the subject were based on two well-known theories: random walk theory and the efficient market hypothesis (EMH). EMH suggests that the stock market is strongly influenced by new information as opposed to present and past prices (Fama, 1991; Fama, Fisher, Jensen, & Roll, 1969). Stock prices should therefore follow a random walk pattern given the unpredictability of the news; hence prices cannot be predicted with more than 50% accuracy (Walczak, 2001).

Although these theories suggest that news is not used to determine market prices, researchers are trying to demonstrate their usefulness and impact on the different variables.

In their study, Antweiler & Frank (2004) have shown that the Internet Stock message was not just noise but did contain financially relevant information. To do this, they analysed over 1,5 million messages from the 45 companies in the Dow Jones Industrial Average (DIJA) that they collected from Yahoo! Finance and Raging Bull. They found a correlation between Internet message board activity, stock volatility and trading volume. However, the effect on stock return has not yet been fully established.

Tetlock's 2007 study explored the interactions between media and stock market over a 16-year period spanning from 1984-1999 by utilising daily content from the Wall Street Journal's column, "Abreast of the market". Although he explains that the media is not a proxy for new information, he acknowledges that they are not "a sideshow with no relationship to asset markets".

Yahoo! and the Wall Street Journal (WSJ) are well-established sources in the financial field, but the flow of information has significantly increased since the arrival of microblogs such as Twitter.

This platform, which has millions of subscribers, has changed the way we express ourselves and spread information. Millions of tweets, messages posted on Twitter, are published every day.

In regards to opinions and feelings analysis, Pak & Paroubek (2010) state that tweets represent a rich, attractive source of data given that users mainly express emotions or facts.

Emotions play a key role in our daily choices and stock market decisions are also influenced by our emotional state, especially when it is negative. Gilbert & Karahalios (2010) used a Granger causal framework to find that increased expressions of anxiety, worry and fear in LiveJournal blogs predict downward pressure on the S&P 500 Index. Their article agrees with Tetlock's 2007 findings: high media pessimism is a predictor for decline in prices before a return to normal; and exaggerated levels in either extreme of pessimism forecast a high volume of market transactions. The main difference between these two studies lies in that Gilbert & Karahalios' dataset, comprising 20 million posts, remained raw; they did not filter it to solely contain financial information.

Negative emotional states drive markets. Case in point: the recent tweet posted by President of the United States of America, Donald Trump, evoking planned tariff increases for Chinese products plunged financial markets into a state of significant anxiety.



The impact was very significant and, on a global level, the market fell instantly.... \$1,36 trillion was wiped off global equities in less than a week<sup>1</sup>! The S&P 500 and the Stoxx Europe 600 respectively lost 2,2% and 3,4%. In both cases, this represents the largest The 45th President of the United States is well known for his use of Twitter as a means of communication. This social media platform affords instantaneous spread of his messages

---

<sup>1</sup> Source : <https://economictimes.indiatimes.com/markets/stocks/news/each-word-of-trumps-tariff-tweets-wiped-13-billion-off-stocks/articleshow/69238055.cms>



to more than 62 million subscribers; all whilst circumventing traditional, possibly censorious, large-scale media.

Donald Trump tweets regularly about specific companies or entities. December 2016's messages focused primarily on the aerospace industry. Some posts shook up the market; illustration:

“The F-35 program and cost is out of control. Billions of dollars can and will be saved on military (and other) purchases after January 20th.”<sup>2</sup>

This statement<sup>3</sup> refers to the F-35 model of Lockheed Martin, a competitor of Boeing which has committed to developing fighter aircraft for the US military.

The effects were soon felt: shares of the defense company fell as much as 5,4% intraday before closing down at 2,5% on a day when the S&P 500 slipped 0,1%<sup>4</sup>; estimated at a value of \$4 billion.

But Lockheed was not the only company affected. Shares of Boeing initially dropped 0,72% but by session close traded to a positive 0,43%. Shares of General Dynamics also fell 2,87% before rallying to close down at 0,94%. Prominent defense industry ETFs PowerShares Aerospace & Defense Portfolio and iShares U.S. Aerospace & Defense both sustained losses eclipsing 1% for the session<sup>3</sup>.

The effect of Trump's tweets has already been the subject of several studies. What emerges from the literature is that the President's message about specific companies can move stock prices and increase trading volume, volatility, and institutional investor attention (Ge, Kurov & Wolfe, 2017). Born, Clark & Myers (2017), similarly conclude. In addition, they state the majority of market reaction, be his tweets positive or negative, occurs the same trading day as the announcement; raising volume by 87% the first post-tweet trading day, 56% the consecutive, and lastly causes an increase of Google Search activity hits for the mentioned companies throughout the week of its publication.

---

<sup>2</sup> Source : <https://www.cnbc.com/2016/12/12/lockheed-martin-shares-drop-after-trump-says-f-35-program-too-expensive.html>

<sup>3</sup> Source : <https://www.cnbc.com/2016/12/12/lockheed-martin-shares-drop-after-trump-says-f-35-program-too-expensive.html>

<sup>4</sup> <https://www.wsj.com/graphics/trump-market-tweets/>

However colossal the impact of these tweets, their effect, according to Born & al., 2017; Ge & al., 2017; Rayarel, 2018 is ephemeral.

What relevance should be accorded to the impact of Trump's mighty yet short-lived tweets? Although he acts as a single individual, albeit with considerable influence, his postings alone cannot globally encompass investor opinion and reaction therein. Hence the importance of considering tweeting in a broader sense; as a variable on companies and markets. In this thesis we aim to shed light on both the themes of sentimental analysis of microblogging and its effect on the stock market.

In the case of market performance prediction of two specific Belgian businesses, should sentimental analysis be taken into consideration?

Is there a correlation between the polarity of tweets and the variation in the share price?

Our main contributions are as follows:

- A reliable method of sentimental analysis
- A complementary financial approach

This work is structured as follows. In the next section, we will start by an explanation of the sentiment analysis and its various tools in order to facilitate understanding.

In section 3, we will review the literature on the subject.

Section 4 is devoted to the analysis. First, we will describe the selected companies starting with a brief presentation followed by a financial analysis from an investor's point of view. Afterwards, we will explain how the dataset was formed and cleaned. Then, we will present the statistical approach chosen.

After that, the section 5 will be used to present the results obtained and limitations of our work.

The research question from an ethical perspective will be addressed in Section 8.

Finally, we will answer the questions raised in the introduction and propose some ideas for future work in the last section.

## 2 THEORETICAL FRAMEWORK

---

The purpose of this section is to explain the theoretical framework of sentimental analysis. We will start by defining it and then briefly present its different components.

### 2.1 SENTIMENT ANALYSIS:

According to Bing Liu (2012), sentiment analysis, also called opinion mining, is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written language towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. According to him, "the term emotional analysis may have first appeared in Nasukawa and Yi (2003), and the term opinion extraction appeared for the first time in Dave et al (2003)". Sentiment analysis is a field within Natural Language Processing (NLP).

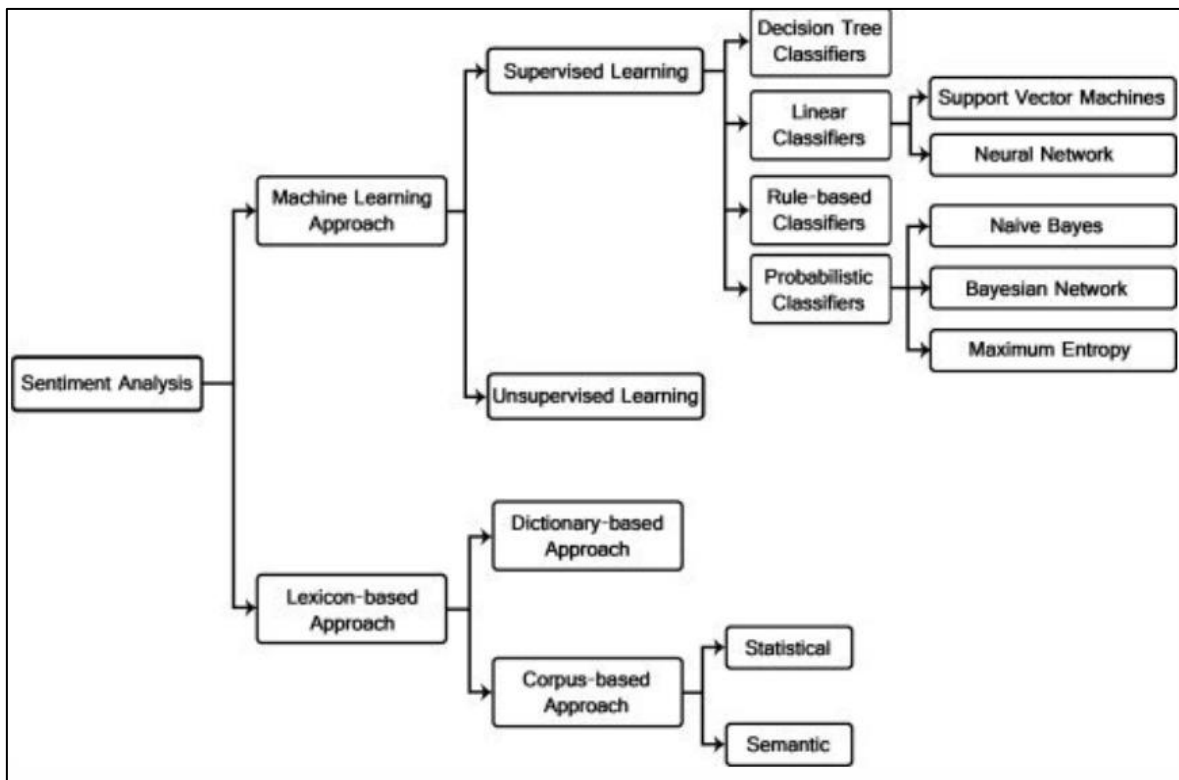
This is one of the most active areas of study in natural language processing given the continuous growth of available data.

Today, the advent of social networks and other discussion forums gives us access to a very wide range of textual information. These can be divided into two broad categories: facts and opinions. On the one hand, facts that are objective expressions of something. And on the other hand, opinions that are usually subjective expressions describing feelings towards a subject.

There are different types of sentiment analysis:

- Polarity: positive, negative and neutral
- Feelings and emotions (happiness, hope, fear, ...)
- Intent: what is the intent of the message (question, complaint, ...)
- Aspect-based analysis: when you want to know not only the polarity of the message but also what the message is about. Example: "The camera of the new Iphone is disappointing". This opinion is negative but discusses in particular the Iphone's camera.

The main approaches for the implementation of sentimental analysis are represented by this diagram (*Figure 1 - Sentiment Analysis*).



*Figure 1 - Sentiment Analysis*

### 2.1.1 Machine learning approach:

Machine Learning is an area of Artificial Intelligence that builds algorithms that allow computers to learn to perform tasks from data instead of being explicitly programmed.

In the case of sentimental analysis, the machine learning makes it possible to develop algorithms capable of extracting feelings from text data.

The two main categories of machine learning are supervised and unsupervised.

#### 2.1.1.1 Supervised learning:

The objective of supervised learning is to predict a defined outcome in an unknown situation using the knowledge acquired from the data. First, the desired output will be given, so that the machine knows what it has to achieve. In order to determine the path, the

algorithm will be "supervised" by receiving annotated data. The algorithm will then use this knowledge to predict the specific outcome of an unseen situation. Supervised machine learning is therefore used to model an input-output model based on labelled data that will be able to predict future data.

Supervised learning separated into two categories of algorithms:

- Regression: when you want to identifying real values (dollars, weight,...). Regression is used to estimate the cause-and-effect relationship between a dependent variable (production) and one or more independent variables (input). There are two types of regression: linear and logistic. The main difference between the two is that linear regression is used to predict values while logistic regression is used for classification tasks.
- Classification: when you want to sorting items into categories

In sentimental analysis, we are faced with a problem of classifying the text, which is the process of categorizing the text into organized groups.

Several algorithms can be used for text classification. These algorithms are called classifier. We will only present here the most used ones.

- **Naïve-Bayes:**

Naïve-Bayes is a set of classification algorithms based on Bayes' theorem that is based on a common principle: each classified element is independent of the value of any other element.

Bayes theorem provides a way of calculating posterior probability  $P(c|x)$  from  $P(c)$ ,  $P(x)$  and  $P(x|c)$ . Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

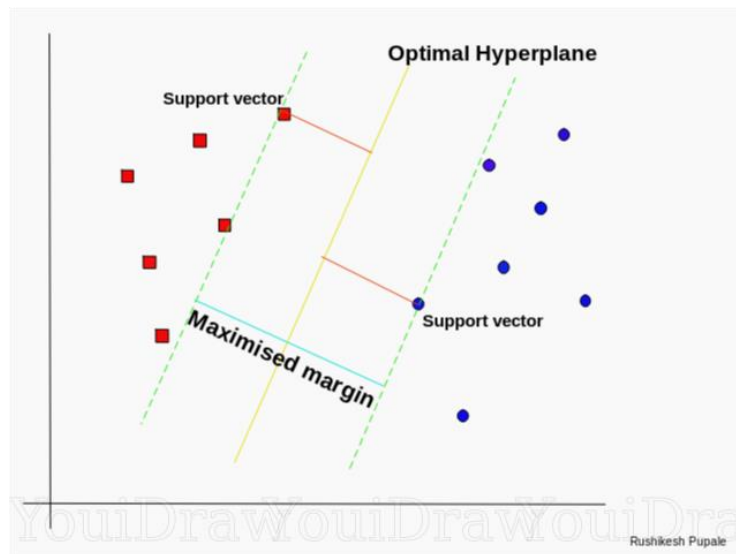
Above,

- $P(c|x)$  is the posterior probability of *class* ( $c$ , *target*) given *predictor* ( $x$ , *attributes*).
- $P(c)$  is the prior probability of *class*.
- $P(x|c)$  is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$  is the prior probability of

*predictor.*

- **Support Vector Machine:**

The objective of the SVM classifier is to find the best separating line (or hyperplane) between data of two classes (in the case of binary classification). In order to do that, the algorithm will find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes (as depicted in Figure 2 - Optimal Hyperplane using the SVM algorithm ). The future data points will then be classified based on their position relative to the optimal hyperplane.



*Figure 2 - Optimal Hyperplane using the SVM algorithm*

#### *2.1.1.2 Unsupervised learning:*

The unsupervised learning process differs significantly from its supervised counterpart because, here, the algorithm used by the machine is not "trained". The machine will make its own interpretations based on the raw data given to it. The objective will then be to group the unlabeled data according to patterns, similarities and differences without prior training.

Unsupervised learning separated into two categories of algorithms:

- **Clustering:** when you want to discover the inherent groupings in the data.
- **Association:** when you want to discover rules that describe large portions of your data

### **2.1.2 Lexicon-based approach**

This method is based on the use of corpus or dictionaries in which words are annotated with a polarity score. The text to be analysed thus obtains a score that represents the sum of the words that compose it. Lexicons containing dictionaries across different domains are available for sentimental analysis such as SentiWordNet for example. This approach has the advantage of not requiring automation data. On the other hand, lexicons do not contain all words and expressions.





### 3 LITERATURE REVIEW

---

The purpose of this section is to give the reader an overview of different studies conducted on the sentimental analysis of financial tweets as well as the methods applied to prove their impact on financial markets.

Although we have read a wide range of works, we will only place in this part those that have most influenced us in our journey.

One of the reference works in this field is the study of Bollen, Mao & Zeng (2011). These researchers measured the correlation between collective mood states derived from large-scale Twitter flows and the Dow Jones Industrial Average (DJIA) value over time.

In this study, they analyzed the text content of daily tweets with two mood tracking tools: Opinion Finder (OF) that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). They first carried out a training test over a short period of time to see if people's moods reacted to socio-cultural events such as election or thanksgiving. OF showed a clear correlation, the opinion was positive at the time of these events. On the other hand, not all moods reacted the same. GPOMS therefore gives a multi-dimensional appearance that OF does not have. Afterwards they analysed the link between the public mood states and the stock market and more particularly the DJIA closing values. They used a Granger causality analysis and Fuzzy neural network for their prediction. They obtained as a result an 86.7% direction accuracy (up and down) of the closing price changes and a reduction of the Mean Average Percentage Error (MAPE) by more than 6% which shows a clear correlation between the public mood analysed through twitter feeds and the DJIA. Despite a significant data set collected during 2008 (10M tweets with explicit indication of user mood), there are only 15 transaction dates in their test set. Although they have achieved a high degree of accuracy, such a short period may not be sufficient to conclude from the effectiveness of their method.

In the same year, Zhang, Fuehres & Gloor (2011) collected Twitter feeds for six months and measured collective hope and fear every day. To do this, they labelled each tweet with mood words such as "fear," "worry," "hope," etc. They then analyzed the correlation between these indices and the stock market indicators. They concluded that the percentage

of emotional tweet was positively correlated with the VIX and significantly negatively correlated with Dow Jones, NASDAQ and S&P 500. They did not attempt to use their model to predict the value of stock prices.

Since sentimental analysis has become an interesting subject of study, much work has followed the publication of Bollen & al (2011). All this research is based on the same idea: collecting tweets<sup>5</sup> that reflect the positive and negative mood of the public to determine if they have an influence on stock market fluctuations.

For example, Rao & Tushar (2012), targeted two indexes: the NASDAQ-100 and the DJIA, and the thirteen large-cap technology companies (ex.: Apple) over a period of more or less one year.

The analysis by company shows a very strong correlation between the two positive and negative moods and the price movement of securities or indices. According to them, the model is relevant because it is not limited to simple discussions where only certain companies are cited, but on the overall public sentiment towards an index or company.

They showed a high correlation (up to 0.88 for returns) and a better directional accuracy (up to 91%).

They consider their model to be more robust, through the use of an Expert Model Mining System (EMMS), they have arrived at a more accurate prediction system. The MAPE for the DJIA was reduced to 0.8 compared to 1.79 for other work. They also have shown that the best Time Window is monthly in Matrix Correlation.

Some researchers, such as Sprenger, Timm O. Tumasjan, Andranik Sandner, Philipp G., Welpe, Isabell M. (2014) have determined that the quality of a tweet does not necessarily guarantee its propagation (retweet). On the other hand, they found that users who give more relevant investment advice than average will have more impact because their messages reach a more active community (more followers and retweets).

By analyzing, as a dataset, 250,000 equity-related messages, they found a correlation between tweets sentiment, stock market performance, stock performance and trading

---

<sup>5</sup> i.e. all messages published on microblogs and Internet message board.

volume. They used the same method as Antweiler & Franck (2004), the naive Bayesian text classification method, to analyze and classify messages daily into three categories: Buy, sell or keep.

However, not all studies rely Twitter as dataset, Nguyen, Shirai & Velcin (2015) took messages on 18 Internet message boards each concerning a single company on Yahoo Finance for one year (2012-2013).

In addition to identifying the sentiment, they also determined the subject of the message (product, service, dividend, etc.) and thus created a pair of indices: "Topic-Sentiment".

Their conclusion is that it is more interesting to use these two indices together than each separately. They also determined that sometimes Sentiment analysis model was no more effective than Price Model analysis for stocks that do not fluctuate too much and whose price model is sufficient.

In their article, they used two methods: the JST and the Aspect-Based Model. In the end, their model gives only 54.41% accuracy in forecasting price and share movements, except for some where they are above 60%. Despite these results, they consider their model adequate for actions that are difficult to predict with past prices.

On the other hand, Ranco, Aleksovski, Caldarelli, Grčar & Mozetič, (2015) have linked their sentimental analysis to an event study, which is an analysis of abnormal return prices observed during external events. An abnormal performance is the difference between the actual performance and the performance that should have been achieved if the event had not occurred.

As for the data used, they collected 1.5 million tweets over 15 months on the 30 DJIA titles. They used a Granger test and a Pearson test, but it seems that the correlation is relatively low. As for text classification, they used a SVM.

Their analysis leads to the conclusion that the cumulative Twitter sentiment at events gives an indication of the market direction (Peak) and that the determined events involve a peak (e. g. Announcement of primary results). But it is important to note that they observed the same correlation when unexpected events occurred. They also state that when financial tweets and tweet feelings and volume are used, they find a statistically significant impact.

Unlike others, some researchers (Pagolu, Reddy, Panda, Majhi & Challa, 2016) have decided to analyze a single company: Microsoft. They collected 250,000 tweets over a one-year period. In order to take into account all the information about Microsoft's activity and not just the share, they filtered the tweets with terms like \$Msft, #microsoft, #windows, etc.

In this study, they analyzed the sentiment of tweets using two "textual representation": Word2Vec and N-gram. Their classifier has achieved an accuracy of 70.2% with Word2Vec and 70% with N-gram. In terms of correlation analysis, they also used two models: the Lib SVM which gives a result of 71.82% and the Logistic regression algorithm which gives a result of 69.1%. According to them, the ideal window size is reached when the value of feelings is three days higher than the value of the stock price.

In conclusion, they sought to demonstrate that the emotions and feelings of the audience on Twitter are reflected in price fluctuations. Their methodology allowed them to demonstrate that this correlation was strong.

Finally, in 2017, Oliveira, Cortez & Areal offered a very robust methodology with a data set containing nearly 31 million tweets over three years (2012-2015) and covering all equities traded on the US market.

This work analyzed the return forecasts of several indices as well as portfolios of different sizes and sectors of activity by focusing on the three stock market variables: volume, return and volatility. They used five different regression methods to predict stock price changes: Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF) and Ensemble Mean (EA). Overall, the model that has proven to be most effective is the SVM.

Their analysis of microblog sentiments and survey indices (AA II and II, USNC and SEMPIX) showed that they were performing well in the accuracy of the SP500 in small-cap portfolios and sectors such as HiTech, Energy and TeleCom. Small cap stocks seem to be more affected by sentiment because they are held by individual investors. Indeed, smaller stocks seem less attractive to professional investors.

All the studies we reviewed appear to agree on the value of microblogs in order to obtain predictions on different market variables (yield, volatility and volume) for financial analysis.

In addition, these data have advantages over traditional survey measures: daily publication, low cost, speed and targeting of specific companies.

We note that data collection differs somewhat from one study to another in terms of:

- The measurement of moods
- The number of companies/indices targeted;
- The duration;
- The tools used.

The emotional state of each human being influences their choices. The estimation of public emotions (moods) aims to provide, when they are sufficient or charged with specific topics (stocks or keywords), information on investors' feelings towards the stock markets. We encounter interactions between media content and stock market activity of varying duration.



## 4 ANALYSIS

---

### 4.1 COMPANIES SELECTION

In order to differentiate ourselves from the studies mentioned above, we wanted to select multinationals of Belgian origin, with large capitalization and which would have recently made the news. This last point is important for our data collection

After some research, we chose Anheuser-Busch Inbev (AB Inbev) & Galapagos.

We will start by introducing the company. Then we will give information about its shareholding and its stock. We will conclude with a brief financial analysis.

#### 4.1.1 Ab Inbev

##### Presentation

The world's largest brewer is well-known in our country given that its history has begun in Belgium more than 800 years ago. Over the years, the company has evolved into a group through mergers and acquisitions (Interbrew and Ambev merged to form InBev. Anheuser-Busch then joined, to create AB InBev). Today the group has extended its presence to more than 150 countries, but its headquarters is still located in Belgium, in Leuven.

The belgico-brazilian brewer's portfolio is very broad with more than 500 brands including Beck's, Budweiser, Corona and Jupiler to name a few. Almost one in four beers consumed in the world comes from AB Inbev (Reuters, 2019) which allows it to occupy the undisputed leadership position far ahead of its main competitor, Heineken NV.

However, the group's worldwide revenues decreased by 3.4% in 2018 compared to the previous year to reach a total amount of \$54,6 billion. This slowdown can be explained by stagnant sales in the US and a decline in EMEA (Europe Middle East & Africa). Furthermore, the change in mentality among consumers play a role in this decline, they have more awareness and want to make healthier choices which leads them to turn to drinks such as wine and sparkling water. Ab Inbev is trying to exploit this trend by diversifying its portfolio with less alcoholic beverages which already represent 8% of its total volume (Trefis Team, Great Speculations, 2019). The new strategy of the company is

the premiumization: an approach that consists in upgrading an existing product, offering it in a "luxury" version or declining it in premium forms. But the brewer still relies on his miracle formula to stay at the top of his market: synergies and cost savings achieved due to acquisition. However, this strategy, although effective so far, is starting to run out of steam; indeed, there is not much left to buy and AB Inbev must be careful not to violate the rules of fair competition as it has already done in the past<sup>6</sup>.

Thanks to the CEO, Carlos Brito, who has put the company's focus back on finding and adapting to new trends. For example, the company has acquired about a dozen craft brewers in a bid to hire more entrepreneurial talent, is ramping up spending on technology and recently partnered with Tilray Inc. to research cannabis-infused beverages (Buckley, 2019).

### Recent news

The financial sphere has talked a lot about Ab Inbev recently, Indeed, at the beginning of July, the company announced that it wanted to raise \$9.8 billion with the IPO of its Asian Pacific unit, Budweiser Brewing Company APAC Ltd.

Budweiser Brewing Company APAC, which holds a portfolio of more than 50 beer brands in the region, was to sell 1.6 billion primary shares at a unit price between HK\$40-HK\$47 (US \$5.13-\$6.02). It was to be the largest IPO of the year and was to be used to reduce the parent company's debt which reached \$102.5 billion after the acquisition at the end of 2016 of its British rival SABMiller for about \$100 billion. Budweiser APAC shares were to be listed on July 11 and start trading on July 19 but it never happened (AB Inbev prepares, 2019). The giant brewer has decided not to proceed with the project due to "market conditions". It must be understood that some major American investors were not interested or at least not at that price (Zhu, Franklin & Blenkinsop, 2019).

To remedy this failure, the Belgian-based brewer has decided to sell its Australian subsidiary, Carlton & United Breweries (CUB), to Japan's Asahi Groups Holdings for \$11.3 billion (Blenkinsop & Barbaglia. 2019) The markets reacted positively to the news with a 4.5% increase in AB Inbev's share price (Strauss, 2019).

---

<sup>6</sup> The company was fined €200 million by the EU for abusing its dominant position on the Belgian beer market between 2009 and 2016. Source : [https://europa.eu/rapid/press-release\\_IP-19-2488\\_en.htm](https://europa.eu/rapid/press-release_IP-19-2488_en.htm)



The group seems to be doing well in view of in view of recent announcements (25/07/19)<sup>7</sup>:

- +14% compared to the same period last year;
- Best quarterly results in 5 years in terms of volume.

### Shareholding & stock

The major shareholders of the company are a stichting<sup>8</sup> (foundation) of AB Inbev with 34% of voting rights, EPS Participation Sàrl with 6,66% (an affiliated company of EPS SA) and BRC Sàrl with 2,04%. The restricted shares are held by Altria Group Inc. with 9,46% and Bevco Lux Sàrl with 4,95%<sup>9</sup>.

- Total outstanding capital: 1,238,608,344.12 €
- Total number of outstanding shares: 2,019,241,973
- Total number of outstanding ordinary shares: 1,693,242,156
- Total number of outstanding restricted shares: 325,999,817
- Total number of outstanding subscription rights (each right entitles the holder to subscribe to one new ordinary share): 0
- Free float: = 83,85% (% of share that are not restricted)

Source : <https://www.ab-inbev.com/investors/corporate-governance/shareholder-structure.html>

AB Inbev's shares are traded on the Euronext (ABI), the Mexico (MEXBOL: ANB) and the South Africa (JSE: ANH) stock exchanges and with American Depositary Receipts<sup>10</sup> on the New York Stock Exchange (NYSE: BUD). In this analysis, we will focus on AB Inbev (BUD) shares that are traded on the New York Stock Exchange because the information available on Stocktwits refers to these shares/ADR.

---

<sup>7</sup> Source: <https://www.zonebourse.com /ANHEUSER-BUSCH-INBEV>

<sup>8</sup> “A **Stichting** (foundation) is a Dutch legal entity with limited liability, but no members or share capital, that exists for a specific purpose. This form of entity makes it possible to separate functions of ownership and control.” Source: <https://findwords.info/term/stichting>

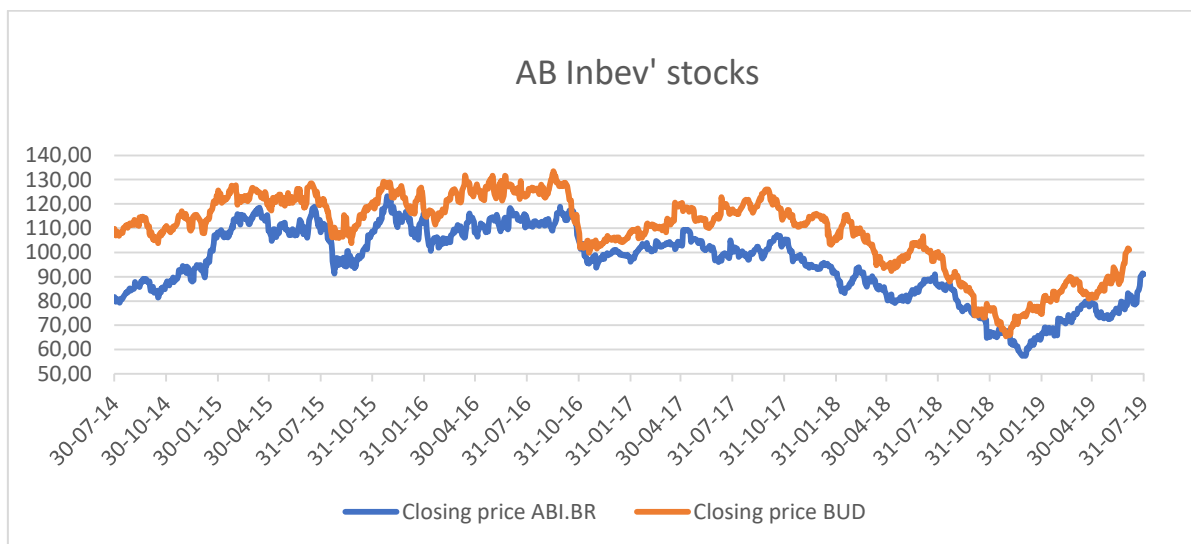
<sup>9</sup> See Annex N°1

<sup>10</sup> An American depositary receipt (ADR) is “a negotiable certificate issued by a U.S. depository bank representing a specified number of shares—or as little as one share—investment in a foreign company's stock. The ADR trades on markets in the U.S. as any stock would trade. ADRs represent a feasible, liquid way for U.S. investors to purchase stock in companies abroad.” Source: <https://www.investopedia.com/terms/a/adr.asp>

**AB Inbev****ABI.BR (EURONEXT)****BUD (NYSE)**

<i>Price (Closing on 31 July 2019)</i>	91,07€	\$101,58
<i>52 Week-range</i>	56,32 – 92,71€	\$64,55 - \$102,70
<i>Market Cap</i>	€168.938B	\$191.667B
<i>Return*</i>	0,02%	0,003%
<i>Volatility*</i>	1,40%	1,36%
<i>Price*</i>	95,59€	\$109,2
<i>Volume*</i>	1712030,8	1665723,83

Table 1 - \*5-Year average



As we can see, AB Inbev's share price fell sharply in 2018 before rising again. Its average return over 5 years is quite low as is its volatility. The difference between the two actions (BUD) and (ABI.BR) is justified by the differences in the exchange rate between the dollar and the euro.

#### 4.1.1.1 *Financial analysis*<sup>11</sup>

In this section, we will analyse the company's health and its competitiveness in relation to other major players in the sector through a small financial analysis based on standard ratios. These ratios are used by investors to value companies and to compare them to competitors (see tables next page).

Looking at a few "peer companies" selected, Anheuser Busch Inbev NV is the absolute market leader among large breweries and beer distributors, especially since the acquisition of SABMiller PLC in 2015 for 72,8 billion British pounds.

The company's sales amounted to 47,6 billion euros in 2018, more than twice as much as second major player Heineken NV, but the Belgian company shows a much more profitable structure than its peers.

In fact, if all peer companies show an EBITDA between 20% and 23% of revenue, AB Inbev's leading position seems to allow the company to gain more synergies and practice higher prices than its competitors.

AB Inbev is also known for owning many "premium" beer brands, which are more likely to generate higher margins than many regular "pils" beers.

AB Inbev's Current and Quick ratios are at the lower bound of the peer group. Due to its size, the company seems to have a large power by both its suppliers (long payment period for trade payables) and its clients (short cashing period for trade receivables). It should be noted that the company has to repay large amounts of debt and is used to paying dividends, preventing it from largely growing its cash position.

Due to the high level of leverage of the acquisition of SABMiller in 2015, AB Inbev has level of indebtedness clearly higher than its competitors. However, its capital structure may be considered as "safe", and no solvency risk seem to be threatening for the company.

From a valuation perspective, AB Inbev's valuation is in line with the valuation of comparable companies. Its EV/SALES is higher due to its better profitability and its cost

---

<sup>11</sup> The figures are based on Thomson Reuters and the annual report of AB Inbev

of equity is higher due to its high level of indebtedness, but the other valuation multiples and the WACC of the company are completely coherent with the peer group.

As a summary, the company's two key financial features are its size and the related high profitability, and its high level of indebtedness, linked to the leveraged acquisition of SABMiller in 2015, which is not yet fully repaid.

Company Name	Country	Identifier	Size indicators			Profitability			Liquidity	
			Market Cap (M€)	EV (M€)	Revenue (FY0, M€)	EBITDA /SALES*	EBIT /SALES*	Return On Equity*	Current Ratio (FY0)	Quick Ratio (FY0)
Anheuser Busch Inbev NV	Belgium	ABI.BR	151.860	251.050	47.623	39%	31%	12%	53%	41%
Heineken NV	Netherlands	HEIN.AS	55.430	72.707	22.471	23%	16%	16%	87%	69%
Carlsberg A/S	Denmark	CARLb.CO	18.492	21.343	8.370	21%	13%	10%	52%	40%
Molson Coors Brewing Co	USA	TAP	10.249	18.629	9.390	20%	12%	9%	64%	51%
Boston Beer Company Inc	USA	SAM	4.182	4.214	871	20%	14%	20%	192%	134%
Royal Unibrew A/S	Denmark	RBREW.CO	3.371	3.788	977	21%	17%	27%	64%	42%
<b>Peer group median (incl. Anheuser Busch Inbev NV)</b>			<b>14.370</b>	<b>19.986</b>	<b>8.880</b>	<b>21%</b>	<b>15%</b>	<b>14%</b>	<b>64%</b>	<b>46%</b>

\*4 year historical average

Company Name	Solvency		Valuation				Discount rate			
	Total assets /Total debt	Market Cap/EV	EV/SALES	EV/EBITDA	EV/EBIT	P/E	Unlevered Beta	Levered Beta	Cost Of Equity	WACC
Anheuser Busch Inbev NV	211%	60%	5,16	12,54	15,51	26,88	0,61	1,28	9,0%	5,9%
Heineken NV	281%	76%	3,14	12,57	18,45	29,60	0,43	0,75	6,4%	5,0%
Carlsberg A/S	491%	87%	2,55	12,01	17,32	26,36	0,59	0,82	6,5%	5,3%
Molson Coors Brewing Co	287%	55%	1,95	10,29	17,94	12,65	0,46	0,75	6,7%	5,1%
Boston Beer Company Inc	N/A	99%	4,25	23,95	32,65	43,41	0,71	0,70	6,4%	6,3%
Royal Unibrew A/S	302%	89%	3,84	16,57	20,84	24,17	0,27	0,47	4,5%	4,2%
<b>Peer group median</b>	<b>287%</b>	<b>81%</b>	<b>3,49</b>	<b>12,56</b>	<b>18,19</b>	<b>26,62</b>	<b>0,52</b>	<b>0,75</b>	<b>6,5%</b>	<b>5,2%</b>

#### 4.1.2 Galapagos<sup>12</sup>

Galapagos NV is a Belgian biotechnology company specialized in the discovery and development of small molecule medicines with novel modes of action. The company was founded in 1999 following a joint venture between Crucell and Tibotec. Since then, the company has continued to grow by raising funds through private placements and public offering. In addition, it has partnered with large pharmaceutical companies such as Servier and Gilead. Sales and acquisitions have also taken place to support this growth. All this for the purpose to focus on the development and execution of their extensive R&D pipeline. The business model of Galapagos is hybrid combining internal discovery programs with service activities.

Now, more than 800 employees work for the Galapagos group (including Fidelta). The headquarters is located at Mechelen but the group have facilities facilities in The Netherlands, France, Switzerland, the US and Croatia. Their flagship product is the filgotinib, an investigational compound for rheumatoid arthritis and other inflammatory diseases.

Belgian biotech does not intend to stop there. Its ambition is to become a leading global biopharmaceutical company thanks to the discovery, development and commercialization of high-quality medicines to fight disease such as rheumatoid arthritis, inflammatory bowel disease and fibrosis.

Galapagos operates in a highly competitive and volatile market which can rapidly change through mergers and acquisitions (M&A) and where breakthrough innovation are one of the most important sources of revenue.

The biggest companies in the biotech industry are (Hayes, 2018):

- Amgen with a market capitalization of \$121,61 billion;
- Gilead Sciences with a market cap of \$88,16 billion;
- Celgene with a market cap of \$53,85 billion;

---

<sup>12</sup> Source : <https://www.glp.com>

- Biogen with a market cap of \$60,85 billion.

However, the biotech and pharmaceutical markets are closely linked. Indeed, biotechnology researches are one of the major expenses of the pharmaceutical market, which is dominated by companies such as Pfizer, Roche, Johnson & Johnson and Sanofi (Ellis, 2019).

#### Recent events

The share price of Galapagos (GLPG) has gone up by more than 15% on July 15, 2019 following the announcement of an agreement signed with the American giant Gilead Sciences. This agreement includes an initial payment of \$3.95 billion and a direct equity investment of \$1.1 billion from Gilead. The American company's shares in Belgian biotech will therefore increase from 12.3% to 22% by subscribing for new shares at a price of €140.59 per share. The two companies also reviewed certain points concerning the development and commercialization of the filgotinib. Galapagos will have more application in the global strategy of the drug and will participate more broadly in the marketing of the product in Europe, which will improve its commercial presence.

#### Shareholding:

The major shareholders of the biotech are Gilead (12,33%), Van Herk Investments (9,81%), Wellington Management Group (6,25%), ...

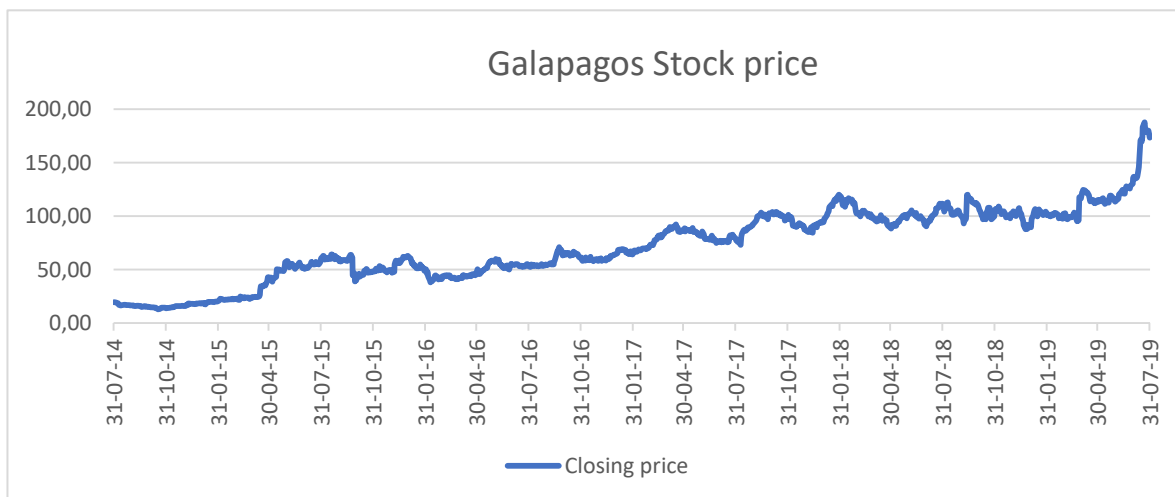
- Share capital = €296,534,760.91
- Outstanding shares = 54,823,101
- Outstanding warrants = 5,958,292
- Free float = 100%

Galapagos NV (ticker: GLPG) has been listed on Euronext Amsterdam and Brussels since 6 May 2005 and on the Nasdaq since 14 May 2015. The company is part of the Bel20, the AEX Index and the Nasdaq Biotechnology Index.

**Galapagos****GLPG (NYSE)**

<i>Price (Closing on 30 July 2019)</i>	\$179,99
<i>52 Week-range</i>	85.00 – 191.63
<i>Market Cap</i>	\$ 9.674B
<i>Return*</i>	0,21%
<i>Volatility*</i>	2,8%
<i>Price*</i>	\$ 71,93
<i>Volume*</i>	98863,38

Figure 3- \*5-Year average



Galapagos' stock has been on an upward trend for several years. Its return is quite high as is its volatility.



#### 4.1.2.1 *Financial Analysis*<sup>13</sup>

In this section, we will analyze the company in the same way as before (see the tables on the next page).

In order to analyse the key financial indicators of Galapagos NV, two categories of "peer" companies may be screened: first regional biotech companies (red frame), with a size and an often unprofitable business comparable to Galapagos NV, then the well-known "big pharma", that have ruled the pharma world for decades.

Looking at a few key profitability indicators, Galapagos NV has shown to remain unprofitable during the past years (negative ROE), but so do most young biotech/biopharma companies, which strive to have their products get through to commercialization.

In fact, looking at the regional companies screened, only mature companies Ucb SA (founded 90+ years ago) and Fagron NV (a previously Omega Pharma run business) have reached a positive EBITDA four years in a row.

However, the unprofitability of Galapagos NV does not prevent investors to believe in the future of the company, as shown by its skyrocketing valuation (EV/SALES multiple of 22,56) and by the recent 5 billion USD investment of Gilead Sciences Inc<sup>14</sup>.

With a current ratio (and quick ratio) of 604%, Galapagos seemed to have a doubtful and costly use of its assets as per last closing date (31/12/2018), but in the absence of financial indebtedness, its strong equity position (book value of 1,2 billion euros) and cash position (1,3 billion euros) have seemed to contribute to building the trust of investors in the company.

Finally, observing the mature "big pharma" companies seems coherent to understand the market in which Galapagos NV operates, especially in perspective of Galapagos NV's 74% CAGR (compound annual growth rate) in revenue during the last three years, and of Gilead Sciences Inc's latest investment. Such companies generate recurring cash flows, and

---

<sup>13</sup> the figures are based on Thomson Reuters and the annual report of Galapagos

<sup>14</sup> Online press release: <https://www.gilead.com/news-and-press/press-room/press-releases/2019/7/gilead-and-galapagos-enter-into-transformative-research-and-development-collaboration>

only allow analysts to forecast limited growth in the future. They also have a much more efficient financial structure with respect to both liquidity and solvency, with respectively current ratios between 120% and 157% and decent amounts of financial indebtedness.

It is therefore no surprise to see that Galapagos, which is absolutely debt free, shows among the highest levels of betas (1,55), required return on equity and WACC (both amounting to 10,6%), reflecting the perceived risk of the investment in the company. As stated above, the cash flows forecasted for the company still allow it to show one of the highest EV/SALES multiples among all the peers screened.

Source: Thomson Reuters

Company Name	Country	Identifier	Size indicators			Profitability			Liquidity	
			Market Cap (M€)	EV (M€)	Revenue (FY0, M€)	EBITDA /SALES*	EBIT /SALES*	Return On Equity*	Current Ratio (FY0)	Quick Ratio (FY0)
Galapagos NV	Belgium	GLPG.AS	8.425	7.320	318	-32%	-34%	-8%	604%	604%
Ucb SA	Belgium	UCB.BR	13.493	13.760	4.632	29%	21%	11%	132%	103%
argenx SE	Belgium	ARGX.BR	4.792	3.863	21	-177%	-178%	-39%	1429%	1429%
Fagron NV	Belgium	FAGRO.BR	1.068	1.327	472	21%	17%	42%	106%	67%
Mithra Pharmaceuticals SA	Belgium	MITRA.BR	1.000	1.058	58	-27%	-33%	-23%	426%	396%
Celyad SA	Belgium	CYAD.BR	128	80	3	-754%	-790%	-50%	536%	536%
MDxHealth SA	Belgium	MDXH.BR	70	47	25	-55%	-62%	-35%	417%	401%
Asit Biotech SA	Belgium	ASIT.BR	24	18	0	N/A	N/A	-189%	198%	198%
Johnson & Johnson	USA	JNJ	311.438	324.169	71.337	35%	28%	30%	147%	120%
Novartis AG	Switzerland	NOVN.S	207.668	225.683	46.356	30%	19%	15%	120%	97%
Pfizer Inc	USA	PFE	190.219	220.736	46.776	41%	29%	24%	157%	133%
Sanofi SA	France	SASY.PA	92.724	113.090	35.677	29%	20%	13%	141%	98%
Gilead Sciences Inc	USA	GILD.O	74.356	71.190	19.293	62%	57%	74%	338%	330%
Carl Zeiss Meditec AG	Germany	AFXG.DE	8.679	8.894	1.281	16%	14%	11%	447%	358%
BB Biotech AG	Switzerland	BION.S	3.261	260	5	88%	88%	5%	1526%	1526%
<b>Peer group median (incl. Galapagos NV)</b>			<b>8.425</b>	<b>7.320</b>	<b>472</b>	<b>25%</b>	<b>18%</b>	<b>11%</b>	<b>338%</b>	<b>330%</b>

\*4 year historical average

Company Name	Solvency		Valuation				Discount rate			
	Total assets /Total debt	Market Cap/EV	EV/SALES	EV/EBITDA	EV/EBIT	P/E	Unlevered Beta	Levered Beta	Cost Of Equity	WACC
Galapagos NV	N/A	115%	22,56	N/A	N/A	N/A	N/A	1,55	10,6%	10,6%
Ucb SA	701%	98%	2,94	10,42	13,49	20,02	0,58	0,69	5,7%	5,2%
argenx SE	N/A	124%	82,00	N/A	N/A	N/A	0,65	0,65	5,9%	5,9%
Fagron NV	207%	81%	2,81	14,26	18,06	25,05	0,60	1,37	9,5%	7,7%
Mithra Pharmaceuticals	203%	94%	18,29	62,12	74,60	N/A	0,86	0,86	6,7%	5,8%
Celyad SA	5729%	160%	25,58	N/A	N/A	N/A	2,08	2,08	13,6%	13,5%
MDxHealth SA	16009%	150%	1,83	N/A	N/A	N/A	0,96	0,96	7,2%	7,2%
Asit Biotech SA	549%	135%	0,00	N/A	N/A	N/A	N/A	0,50	4,6%	4,4%
Johnson & Johnson	502%	96%	4,43	11,68	15,07	21,73	0,47	0,68	6,2%	5,9%
Novartis AG	454%	92%	4,99	14,33	22,29	31,40	0,77	1,06	8,0%	7,3%
Pfizer Inc	382%	86%	4,57	10,47	14,06	18,58	0,40	0,62	5,9%	5,4%
Sanofi SA	451%	82%	3,08	10,69	16,59	25,95	0,57	0,79	6,7%	5,3%
Gilead Sciences Inc	233%	104%	3,54	8,32	9,38	14,03	0,58	1,10	8,9%	8,9%
Carl Zeiss Meditec AG	6258%	98%	6,67	39,93	39,93	67,86	0,44	0,45	4,6%	4,5%
<b>Peer group median (incl.</b>	<b>478%</b>	<b>98%</b>	<b>4,50</b>	<b>11,68</b>	<b>16,59</b>	<b>23,39</b>	<b>0,59</b>	<b>0,83</b>	<b>6,7%</b>	<b>5,9%</b>

## 4.2 DATA

Our analysis focuses on two companies: AB Inbev (BUD) and Galapagos NV (GLPG). Our work investigates the relationship between Stocktwits ([www.stocktwits.com](http://www.stocktwits.com)) messages and the price of the shares considered. Both datasets were collected over a period of 7 months, from January to July of this year (432 messages for BUD and 366 for GLPG). Details are presented below.

Market data was collected via Yahoo!Finance<sup>15</sup>. In this research, only the closing price was considered.

Messages published on Stocktwits (below we will call them tweet for ease of use) were extracted with the help of a Python script. The tweets were filtered using the company's stock cash-tag (e.g. "\$GLPG" for Galapagos NV) to obtain only the desired data. To the best of our knowledge, all the tweets were extracted.

For every tweet, the following data were included:

- Profile (the name of the user who published the message)
- Date of the tweet
- Symbols contained in the tweet (i.e. stock cash-tag)
- The message itself
- The number of likes
- The tagging of the message: bearish or bullish<sup>16</sup>

The text itself was then pre-processed to facilitate sentimental analysis. The mentions "@", symbols, emails, websites, special characters (emoticons) or icons were removed from the original text (see: Figure 4- Pre-process example).

---

<sup>15</sup> <https://finance.yahoo.com/>

<sup>16</sup> Stocktwits offers the possibility to tag messages posted with the mention Bullish (upward trend) and Bearish (downward trend).

Text	Pre-processed text
\$BUD holding. Another gap tomorrow tomorrow! Gotta get over 91 🤔 🤔 🤔 🤔 🤔 🤔	holding Another gap tomorrow Gotta get over 91

Figure 4- Pre-process example

We decided to use Stocktwits (a social media network for investors and traders with 2 million registered members) as a data source rather than Twitter for several reasons. Stocktwits is a platform created especially for investors to discuss markets (Pagolu & al., 2016). The information is supposed to be more relevant than the information found on Twitter.

As Nguyen & al (2015) explain, messages are messier because there is no consensus on hashtags (#AAPL, \$AAPL) and the high noise level makes it difficult to find a specific stock related post.

The negative aspect is that the amount of data is reduced. Another criticism that could be mentioned is the fact that while Stocktwits reflects the opinions of experts, it does not reflect public (non-professional) opinion. But as the literature mentions, non-professional traders are particularly sensitive to news ("noise traders"), which could have affected the solidity of this study (Antweiler & Frank, 2004).

Finally, tweets extraction is easier on Twitter. Stocktwits is much less flexible and we had to bypass the limits to be able to collect data.

### 4.3 SENTIMENT ANALYSIS

The analysis of processed tweets was done with TextBlob<sup>17</sup>, a python library for processing textual data. We chose this library because it offers an integrated sentimental analysis tool.

This tool gives a polarity score between (-1,0,1) for each processed tweet.

The score of each tweet per day was added together, represented by  $SA_d$ .

The daily polarity ( $P_d$ ) of the aggregated tweets is given by the following formula:

---

<sup>17</sup> <https://textblob.readthedocs.io/en/dev/index.html>

$$P_d = \frac{SA_d}{TW_d}$$

Where:  $TW_d$  = number of tweets per day

#### 4.4 CORRELATION

We have therefore determined two time series: the closing price of each company's shares and the daily polarity (sentiment) of the tweets. Noted that we only took the polarity of the days with which a closing price was associated

In order to determine the correlation between these two time series, we decided to use the vector autoregression (VAR) model. This model appears to be conducive to this type of study (Tetlock, 2007, Groß-Klußmann & Hautsch, 2011, Ranco & al., 2015).

This model is “one of the most successful, flexible and easy-to-use models for multivariate time series analysis” (Springer, 2006).

VAR analysis is used to test whether there is a causality between two variables that are currently unknown which one is the dependent and which one is the independent.

To build the analysis, some steps are required:

- Stationarity testing
- Lag optimum selection
- Granger causality test
- VAR model estimation
- Classical assumption testing

The approach is presented below.

VAR analysis of AB Inbev

1) Stationarity testing

Stationarity testing is very necessary to do for time series analysis, because data which is does not have stationary process will have higher chance to be spurious when added to a regression. In this research, tweets and closing price are being tested to determine their stationarity by using Phillips-Perrons (PP) test. AGG is a notation for tweet sentiment and CP is notation for Closing Price. D(AGG) shows the first difference form of tweet sentiment, also D(CP) shows the first difference form of closing price.

Table 1

Variable	T-statistic	Probability	Conclusion
AGG	-8.590	0.000*	Stationary
D(AGG)	-7.079	0.000*	Stationary
CP	-2.000	0.286	Not stationary
D(CP)	-8.964	0.000*	Stationary

\*significance level of 5%

By the table 1 above, the conclusion is closing price in level is not stationary in significance level of 5%. However, it is stationary in its first difference form. Also, tweet is stationary both in level and in first difference. Therefore, because both variables are stationary in first difference, we will use both variables in their first difference form.

## 2) Lag optimum selection

The next step is to determine the optimum lag for VAR model. In this research, we select the optimum lag based on several criterion such as LR, FPE, AIC, SC, and HQ.

Table 2

Lag	LR	FPE	AIC	SC	HQ
0	NA	1.220946	5.875377	5.938129	5.900385
1	23.81271	0.969594	5.644784	5.833041	5.719808
2	16.13592	0.853641	5.517080	5.830842*	5.642120*
3	7.686411	0.848433*	5.510208*	5.949475	5.685264
4	2.094550	0.917445	5.587070	6.151841	5.812141
5	1.341720	1.003897	5.675018	6.365294	5.950105
6	9.719852*	0.955551	5.622610	6.438390	5.947712
7	9.287278	0.912278	5.572074	6.513358	5.947192

Based on the table 2, the optimum lag of 2 is the best out of all lags tested by SC and HQ criterion. However, lag of 3 is the best out of all lags tested by FPE and AIC. Therefore, lag 2 will be used for Granger causality test and estimation of VAR model because at some point SC is usually more preferred than other criterion.

## 3) Granger causality test

Granger causality test is used to test whether there is causality between tweet sentiment and closing price.

Table 3

Pairwise Granger Causality Tests

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
CP does not Granger Cause AGG	79	22.9120	0.0043
AGG does not Granger Cause CP		15.7712	0.0134



Based on Table 3, the hypothesis of “Tweet sentiment does not Granger Cause closing price” has the probability of 0.0043, which means that with significance level of 5% the conclusion is to reject the null hypothesis. It also happens for the second hypothesis. It means that tweet sentiment is significantly causing the closing price, also closing price is significantly causing the tweet sentiment.

4) VAR model estimation

VAR model estimation used lag 2 as the maximum lag. Therefore, the result is as follows:

Table 4

	D(CP)		D(AGG)	
	Coefficient	t-Statistics	Coefficient	t-Statistics
D(CP(-1))	0.028	4.191*	0.004	6.372*
D(CP(-2))	0.0107	0.094	0.060	1.608
D(AGG(-1))	-0.331	2.036*	-0.686	6.534*
D(AGG(-2))	0.355	1.087	-0.434	-4.051
C	0.311	1.622	-0.030	-0.486
R-squared	0.558		0.381	
Adj. R-squared	0.413		0.347	

\*significance level of 5%

Based on Table 4, the first column shows the effect of variables to closing price. It is shown that closing price is significantly affected by closing price in one period before and tweet sentiment in one period before. The coefficient of 0.028 for closing price in one period before shows that if there is one unit rise of closing price in one period

before, then the current closing price will increase by 0.028 unit. While the coefficient of tweet sentiment in one period before is -0.331, which shows that if there is one unit rise of tweet sentiment (it means that there are more positive tweets than the negative tweets), then the current closing price will decrease by -0.331 unit.

Also, for the second column shows the effect of variables to tweet sentiment. It is shown that tweet sentiment is significantly affected by closing price in one period before and tweet sentiment in one period before. The coefficient of 0.004 for closing price in one period before shows that if there is one unit rise of closing price on one period before, then the current tweet sentiment will increase by 0.004 unit. While the coefficient of tweet sentiment in one period before is -0.686, which shows that if there is one unit rise of tweet sentiment (it means that there are more positive tweets than the negative tweets), then the current closing price will decrease by -0.686.

5) Classical assumption testing

There are 3 assumption that have to be tested, such as normality, homoscedasticity, and autocorrelation.

- Normality testing

Equation	Jarque-Bera	df	Prob.
1	2.686	2	0.3210
2	2.012	2	0.3657
Joint	4.698	4	0.3433

Based on the table above, it is shown that both equations are not violating the normality assumptions.

- Homoskedasticity

Chi-sq	df	Prob.
46.899	42	0.2785

Table above shows that the residuals are homoscedastic.

- Autocorrelation

Lags	LM-Stat	Prob
1	5.80721	0.1288
2	6.085038	0.1485
3	6.737142	0.1504
4	1.400248	0.8442

-

Table above shows that there are no autocorrelation in the model, which is good.

Additional graphs:

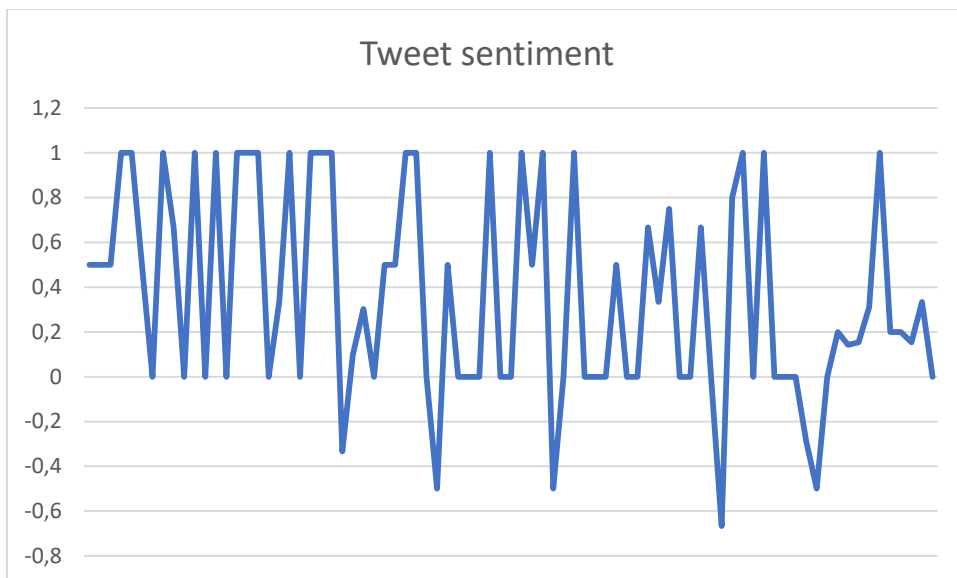
### Closing Price

Closing Price have tendency to increase by time, however it does not increase very significant. The plot shows that the closing price is increasing but not very fluctuative. The average value of closing price is 84.26.



### **Tweet sentiment**

Tweet sentiment is very fluctuative, however the average value is 0.33. It means that the average there are 1 positive tweet out of 3 tweets (in which the other tweet is not negative).



1) Stationarity testing

Stationarity testing is very necessary to do for time series analysis, because data which is does not have stationary process will have higher chance to be spurious when added to a regression. In this research, tweets and closing price are being tested to determine their stationarity by using Phillips-Perrons (PP) test. AGG is a notation for tweet sentiment and CP is notation for Closing Price. D(AGG) shows the first difference form of tweet sentiment, also D(CP) shows the first difference form of closing price.

Table 1

Variable	T-statistic	Probability	Conclusion
AGG	-9.009	0.000*	Stationary
D(AGG)	-43.988	0.001*	Stationary
CP	0.499	0.985	Not stationary
D(CP)	-9.357	0.000*	Stationary

\*significance level of 5%

By the table 1 above, the conclusion is closing price in level is not stationary in significance level of 5%. However, it is stationary in its first difference form. Also, tweet is stationary both in level and in first difference. Therefore, because both variables are stationary in first difference, we will use both variables in their first difference form.

2) Lag optimum selection

The next step is to determine the optimum lag for VAR model. In this research, we select the optimum lag based on several criterion such as LR, FPE, AIC, SC, and HQ.

Table 2

Lag	LR	FPE	AIC	SC	HQ
0	NA	25.49072	8.914065	8.976337	8.938906
1	14.63413	23.11297	8.816059	9.002875	8.890582
2	23.23669*	18.39469*	8.587404*	8.898764*	8.711609*
3	7.093052	18.44929	8.589645	9.025549	8.763533
4	6.612431	18.59122	8.596024	9.156472	8.819593
5	1.910227	20.13571	8.673811	9.358803	8.947063
6	8.457317	19.58739	8.643274	9.452811	8.966208
7	8.650608	18.92331	8.604762	9.538842	8.977378

Based on the table 2, the optimum lag of 2 is the best out of all lags tested. Therefore, lag 2 will be used for Granger causality test and estimation of VAR model.

### 3) Granger causality test

Granger causality test is used to test whether there is causality between tweet sentiment and closing price.

Table 3

Pairwise Granger Causality Tests

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
AGG does not Granger Cause CP	80	30.27515	0.0002
CP does not Granger Cause AGG		41.10941	0.0001

Based on Table 3, the hypothesis of “Tweet sentiment does not Granger Cause closing price” has the probability of 0.0002, which means that with significance level of 5% the conclusion is to reject the null hypothesis. It also happens for the second hypothesis. It means that tweet sentiment is significantly causing the closing price, also closing price is significantly causing the tweet sentiment.

#### 4) VAR model estimation

VAR model estimation used lag 2 as the maximum lag. Therefore, the result is as follows:

Table 4

	D(CP)		D(AGG)	
	Coefficient	t-Statistics	Coefficient	t-Statistics
D(CP(-1))	0.033	5.302*	0.003	8.355*
D(CP(-2))	-0.369	-3.346	-0.0002	-0.023
D(AGG(-1))	0.465	8.388*	0.586	5.464*
D(AGG(-2))	-0.005	-0.004	-0.417	-3.923
C	1.383	1.811	-0.0128	-0.188
R-squared	0.789		0.669	
Adj. R-squared	0.677		0.538	

\*significance level of 5%

Based on Table 4, the first column shows the effect of variables to closing price. It is shown that closing price is significantly affected by closing price in one period before and tweet sentiment in one period before. The coefficient of 0.033 for closing price in one period before shows that if there is one unit rise of closing price in one period before, then the current closing price will increase by 0.033 unit. While the coefficient of tweet sentiment in one period before is 0.465, which shows that if there is one unit rise of tweet sentiment (it means that there are more positive tweets than the negative tweets), then the current closing price will increase by 0.465 unit.

Also, for the second column shows the effect of variables to tweet sentiment. It is shown that tweet sentiment is significantly affected by closing price in one period before and tweet sentiment in one period before. The coefficient of 0.003 for closing price in one period before shows that if there is one unit rise of closing price on one period before, then the current tweet sentiment will increase by 0.003 unit. While the coefficient of tweet sentiment in one period before is 0.586, which shows that if there is one unit rise of tweet sentiment (it means that there are more positive tweets than the negative tweets), then the current closing price will increase by 0.586.

#### 5) Classical assumption testing

There are 3 assumption that have to be tested, such as normality, homoscedasticity, and autocorrelation.

##### - Normality testing

Equation	Jarque-Bera	df	Prob.
1	2.621	2	0.3219
2	2.266	2	0.3220
Joint	4.887	4	0.3220

Based on the table above, it is shown that both equations are not violating the normality assumptions.

##### - Homoskedasticity

Chi-sq	df	Prob.
4.47289	42	0.5621

Table above shows that the residuals are homoscedastic.



- Autocorrelation

Lags	LM-Stat	Prob
1	7.98238	0.0922
2	8.71752	0.0883

Table above shows that there are no autocorrelation in the model, which is good.

Additional graphs:

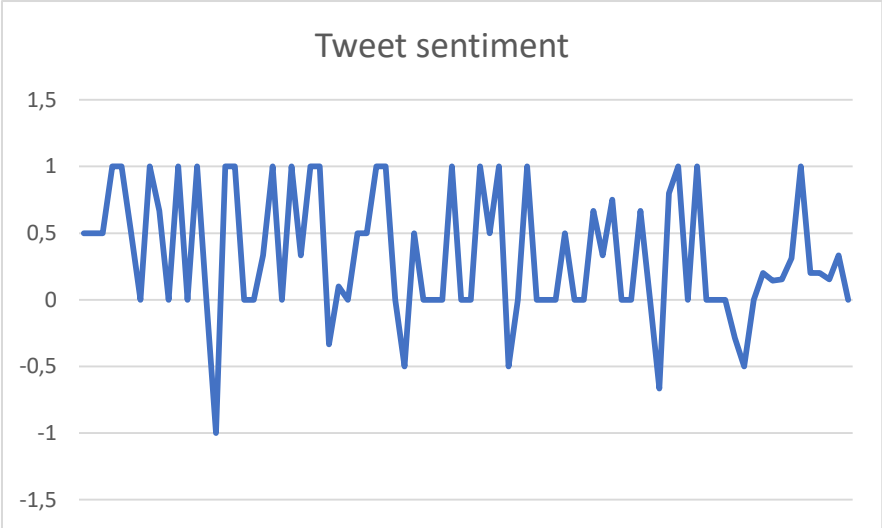
### Closing Price

Closing Price have tendency to increase by time, which means that it has positive trend. The average value of closing price is 119.1.



### Tweet sentiment

Tweet sentiment is very fluctuative, however the average value is 0.33. It means that the average there are 1 positive tweet out of 3 tweets (in which the other tweet is not negative).



## 5 RESULTS & LIMITATIONS

---

Our analysis shows that the model can be used (classical assumption).

**Our results** are as follows:

- AB Inbev: there is positive correlation from closing price in one period before, however tweet sentiment has negative correlation
- Galapagos: there is positive correlation between tweet sentiment and the closing price in both equations.

**The major limits** of our work:

- The biggest limit of our work is obviously in the size of our dataset. This is probably a little low to confirm the robustness of our study. We were unable to perform a training data set and a dataset test;
- We did not use a specific financial glossary such as that of Loughran & McDonald (2011). Indeed, TextBlob uses a "traditional" dictionary. However, many words are qualified as negative in general- purpose polarity lexicon and neutral in the financial field (e. g. liabilities);
- We only took into account the historical closing prices and feelings of Stocktwits' messages.



## 6 ETHICAL PART

---

This section will discuss how the subject of this thesis can be linked to a more general ethical reflection on the business world and the global economic system.

Today, social networks have transformed the way we communicate. Everyone, whether properly informed or not, can express themselves and take position on a subject. This opinion will be seen and eventually shared on a large scale and at high speed without prior verification.

The information thus transmitted may also have an impact on financial markets and listed companies. Indeed, investors discuss on financial forums the shares to be purchased, those from which they must sell, etc.

If some Internet users decide to use this means of communication, as well as forums and other platforms to convey their ideas with their own emotions, they may not always be aware of the impact that their spontaneous tweets could have.

Others, better informed, consciously use it to manipulate financial or other information. This is then a financial manipulation and therefore a breach of ethics. Indeed, ethics focuses on the analysis of the moral principles that determine "good" or "bad" in a given situation (Xhaufclair, 2017).

There are also hackers among Twitter users, such as those who, in 2013, hacked into a well-known source of information, the Associated Press (AP) and reported false information that President Obama and other White House employees were injured in an attack. As a result of this tweet, in a matter of minutes, the Dow Jones Industrial Average (DJIA) fell 143.5 points, and the S&P 500 temporarily lost an estimated value of US\$136 billion (Prigg, 2015) Although markets recovered very quickly, many investors went headlong and lost a lot of capital because of these artificially created market conditions. This illegal act is qualified as "wrong" according to ethical codes.

But what happens when an influential person takes a position on a subject that could have repercussions on the stock market and/or the economy by spreading his or her own emotions on a platform such as the Internet?

Could this be a case of market manipulation? Can we talk about an ethical breach?

This personal publication could indeed reflect a political or economic sentiment and lead to spontaneous reactions from uninformed Internet users.

Are we facing a case of ethical dilemma?

A situation represents an ethical dilemma if it raises moral questions. To determine this, the situation must represent a choice that will have consequences for others, have alternatives and be perceived as ethical by one or more stakeholders (Xhaufclair, 2017).

### 1. Identification of the ethical dilemma

Is it appropriate for people with access to sensitive financial data or with some power to disclose their opinions and personal information on the Internet?

Financial tweets can disrupt real information beyond economic and political issues. Information can distort the short-term vision and therefore lead to inappropriate reactions.

### 2. Description of the ethical dilemma

The choice to communicate on a platform such as Twitter, Stocktwits etc. sensitive information that will impact a company or the market will have consequences. Companies and investors may be affected in the short term.

Citizens will also be affected by this through immediate insecurity among those who read this tweet and may not seek correct or in-depth information from more reliable sources.

The choice to disclose this information can impact many people. The shareholders who have invested in the company, the competing companies. On a larger scale, the economy could be

weakened and thus affect the entire population, as in the trade war between the United States and China.

The ease and speed of information flow can lead to manipulation of markets and public opinion because the financial world is very reactive during trading hours.

Indeed, since information is measured or controlled at its release by only one person or company, it can cause chain reactions without allowing, in a relatively short period of time, to discern the true from the false and to be properly informed.

### 3. Identification of alternatives and their consequences

#### **Alternative**

"Information is power" J-E. Hoover

Any leader, whoever he is, has the opportunity to transmit in a neutral way by having concrete data in order not to provoke negative or other feelings and distort the public's free will.

#### **Consequence**

One of the consequences would be a safeguard on information that could move markets. Eventually, a policy that would not support the overflows of Twitter users. It could even sanction any manipulation that would influence economic transactions, knowing that the people affected are a public that does not necessarily have access to complete information.

### 4. Analysis of alternatives according to several ethical theories

#### **a) Consequentialism**

Consequentialist theories form a set of moral theories that support the view that it is the positive consequences of a given action that must form the basis of any moral judgment of that action.

Consequentialism establishes that some actions are better than others and that we must work to achieve them. We must act to promote better situations.

The ethical dilemma, which arises before the sharing of an opinion or sensitive information, must take into account all the consequences that will result from it.

According to consequentialism, only information with positive consequences should be shared on the Internet.

### **b) Deontology**

Each profession is governed by a set of rules, obligations and duties that are incumbent on a professional in the performance of his or her task.

Those who do so are expected to behave in a certain way. This set of rules and duties is called deontology.

Everyone in their profession is therefore subject to a more or less strict code of ethics.

Each publication, each sharing of information will be judged according to its compliance with this code of ethics.

### **c) Utilitarianism**

Utilitarianism focuses on optimizing the greatest possible happiness for the greatest number of people.

According to this perception, only the message that positively affects the majority will be published. The "tweet" should be of general use.

It will be necessary to act (or not act) to maximize happiness for a majority of people.



## 7 CONCLUSION

---

The financial analysis allowed us to understand the business of the companies analyzed as well as the difficulties of their market. In addition, we were able to observe the evolution of the action over time. The consultation of the ratios gave us a vision of the risks and opportunities that these companies offer to potential investors in relation to their peers.

These different criteria gave us an overview that could give us an opinion about these two companies.

Although we have based ourselves on serious analyses, this remains a personal opinion. Sentimental analysis therefore makes it possible to collect the opinions of other investors in order to determine the general mood of the financial sphere towards these companies.

In this study, we limited ourselves to determining the correlation between daily closing prices and daily polarity. This restriction does not allow us to identify any convincing links between the observations at the beginning of the analysis and the results obtained with the VAR.

Since the two variables are significantly correlated, we make assumptions to explain our results:

- Closing price is significantly affected by tweet sentiment because people will check how is the current sentiment through social media (i.e. Stocktwits). That's why any sentiment related to the price of share in Stocktwits can significantly affect people's decision, which in the end will affect the closing price
- Tweet sentiment is significantly affected by the closing price because the current stock price will make people react either right or wrong, and people will publish their feelings on social networks. This is why the closing price can affect how is the current tweet sentiment.

It seems interesting to us to continue this research by including more variables and a larger dataset. Future work could analyse all Bel20 companies by filtering Twitter<sup>18</sup> using the same

---

<sup>18</sup> the bel20 is not referenced on Stocktwits

methodology as Pagolu & al. (2016) and thus provide an overall assessment of these companies. It may also be relevant to compare them with each other.

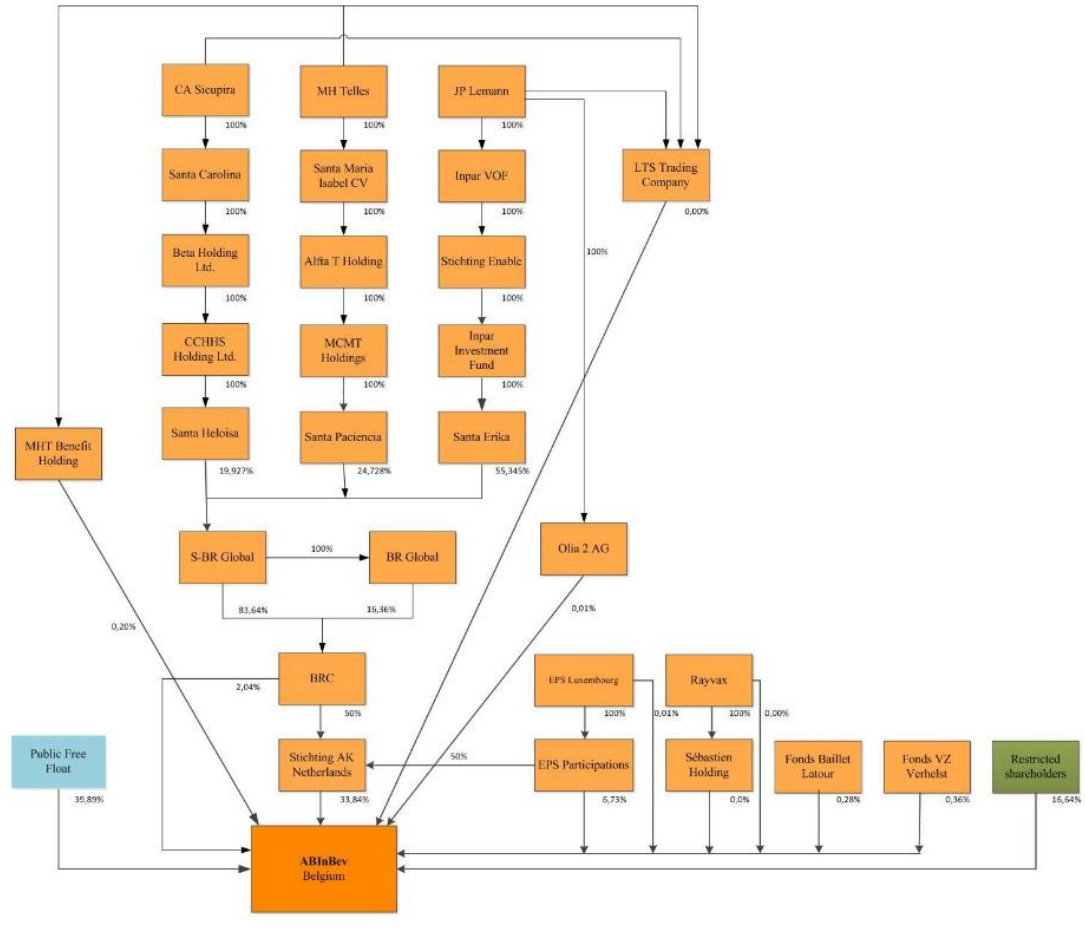
During our readings, we discovered a study that combined sentimental analysis with event study. This new angle seemed to us to offer an adequate added value. Indeed, during the collection of our data we noticed that on some dates large increases in terms of tweets volume, yields and trading volume were recorded. By placing them side by side with our "Recent news" section, we could see a clear correlation. These encouraging observations confirm the relevance of this approach. In order to verify the reliability of the model, two types of companies should be compared: those that are in the news (those that drive discussions on microblogs) and those that are not.

In conclusion, I must admit some shortcomings. The subject of this thesis being significantly different from my studies at HEC Liège (Audit), I acknowledge that I have experienced difficulties throughout this work which have affected the quality of my analys

# 8 ANNEX

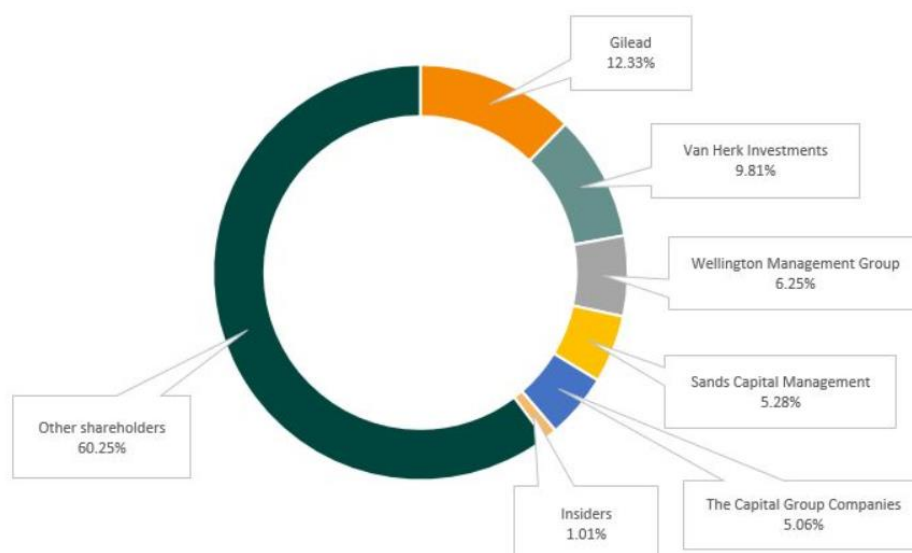
## Annex N°1: AB Inbev' Shareholding

### Shareholder Structure



Major shareholders	Number of Shares	% of voting rights <sup>(1)</sup>
<b>Holders of Ordinary Shares</b>		
<b>Stichting Anheuser-Busch InBev</b> , a stichting incorporated under Dutch law (the "Reference Shareholder")	663,074,832	33.89%
<b>EPS Participations Sàrl</b> , a company incorporated under Luxembourg law, affiliated to EPS, its parent company	130,257,459	6.66%
<b>EPS SA</b> , a company incorporated under Luxembourg law, affiliated to the Reference Shareholder that it jointly controls with BRC	99,999	0.01%
<b>BRC Sàrl</b> , a company incorporated under Luxembourg law, affiliated to the Reference Shareholder that it jointly controls with EPS	39,962,901	2.04%
<b>Rayvax Société d'Investissements SA</b> , a company incorporated under Belgian law	484,794	0.02%
<b>Sébastien Holding SA</b> , a company incorporated under Belgian law, affiliated to Rayvax, its parent company	10	0.00%
<b>Fonds Verhelst SPRL</b> , a company with a social purpose incorporated under Belgian law	0	0.00%
<b>Fonds Voorzitter Verhelst SPRL</b> , a company with a social purpose incorporated under Belgian law, affiliated to Fonds Verhelst SPRL with social purpose, that controls it	6,997,665	0.36%
<b>Stichting Fonds InBev-Baillet Latour</b> , a stichting incorporated under Dutch law	0	0.00%
<b>Fonds Baillet Latour SPRL</b> , a company with a social purpose incorporated under Belgian law, affiliated to Stichting Fonds InBev-Baillet Latour under Dutch law, that controls it	5,485,415	0.28%
<b>MHT Benefit Holding Company Ltd</b> , a company incorporated under the law of the Bahamas, acting in concert with Marcel Herrmann Telles within the meaning of Article 3, §2 of the Takeover Law	3,972,703	0.20%
<b>LTS Trading Company LLC</b> , a company incorporated under Delaware law, acting in concert with Marcel Herrmann Telles, Jorge Paulo Lemann and Carlos Alberto Sicupira within the meaning of Article 3, §2 of the Takeover Law	4,468	0.00%
<b>Olia 2 AG</b> , a company incorporated under Liechtenstein law, acting in concert with Jorge Paulo Lemann within the meaning of Article 3, §2 of the Takeover Law	259,000	0.01%
<b>Holders of Restricted Shares</b>		
<b>Altria Group Inc.</b> <sup>(2)</sup>	185,115,417	9.46%
<b>Bevco Lux Sàrl</b> <sup>(3)</sup>	96,862,718	4.95%

## Annex N°2 : Galapagos' Shareholding



## 9 BIBLIOGRAPHIE

---

- Adam, H. (2018). *The Biggest Biotech Companies*. Récupéré sur <https://www.investopedia.com/articles/investing/011515/biggest-biotech-companies.asp>
- Antweiler, W., & Frank, M. (2001). Information Content of Internet Stock Message Boards. *The Journal of Finance*.
- Bollen, J., Mao, H., & Xiaojun, Z. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 1-8.
- Gilbert, E., & Karahalios, K. (2010). Widespread Worry and the Stock Market. 58-65.
- Groß-Klußmann, A., & Hautsch, N. (2011). *Journal of empirical finance*. 321-340.
- Liu, B. (2012). *Sentiment Analysis(Introduction and Survey) and Opinion Mining*. Morgan.
- Loughran, T., & Mcdonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Distionaries, and 10-Ks. *Journal of Finance*, 35-65.
- Mark, P. (2015). *The tweet that cost \$139 BILLION: Researchers analyse impact of hacked message claiming President Obama had been injured by White House explosion / Daily Mail Online*. Récupéré sur <https://www.dailymail.co.uk/sciencetech/article-3090221/The-tweet-cost-139-BILLION-Researchers-analyse-impact-hacked-message-claiming-President-Obama-injured-White-House-explosion.html>
- Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, 125-144.
- Pagolu, V., Reddy, K., & Panda, G. (2016). Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *Scopes*.
- Pak, A., & Patrick, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining.

- Predicting Financial Markets: Comparing Survey, N. T. (2011). Mao, Huina; Counts, Scott, Bollen, Johan.
- Ranco, G., & Aleksovski, D. C. (2015). The effects of twitter sentiment on stock price returns. *Plos One*, 1-21.
- Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using Twitter sentiment analysis. *NSIT*, 1-5.
- Sprenger, T., Tumasjan, A., & P., S. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 926-957.
- Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of finance*, 1139-1168.
- Zhang, X., Fuehres, H., & Gloor, P. (2011). Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”. *Procedia*, 55-62.
- Zhu, J., & Joshua, F. P. (2019). *AB InBev pulls Budweiser listing, canceling year's largest IPO - Reuters*. Récupéré sur <https://www.reuters.com/article/us-abinbev-asia-ipo/ab-inbev-pulls-budweiser-listing-canceling-years-largest-ipo-idUSKCN1U70L8>

