LUISS

Dipartimento
di Impresa e Management

Cattedra di Marketing Big Data Analysis

Working-from-home jobs, during the Covid-19 pandemic – a textual analysis derived from consumer reviews

Prof. Alessio Maria Braccini

RELATORE

Prof. Serena Arima

CORRELATORE

Filippo Mazzola - Matr. 720191

CANDIDATO

Anno Accademico 2020/2021

Table of contents

“Working-from-home jobs during the Covid-19 pandemic: a textual analysis derived from consumer review 2
1. Introduction 3
2. Literature Review 5
3. Research Methodology and Data Analysis 7
3.1 The companies 8
3.2 Procedures 11
3.2.1 Data Scraping 12
3.2.2 Data Cleaning 12
3.2.3 Data Analysis 14
4. Results 19
4.1 Sentiment analysis 22
5. Regression analysis 27
5.1 Results 29
5.2 Spill-over effects 35
5.3 Pandemic effects 37
6. Conclusion 40
References 42
Summary 44

1. Introduction

The recent Covid-19 pandemic forced most people worldwide to adopt some changes in their lives. The virus affected the way people perceive some social aspects: from basic activities to hobbies, from teaching to learning methods, but also government and business operations. It is possible to look at two opposite extremes that are present during the peak of a pandemic. On the one hand, most people are scared and suspicious about interacting with others, increasing anxiety that brings to isolation and loneliness. On the other hand, some superficial people try to find more ways to mislead government restrictions; behaving in this way though, they increase virus circulation and postpone the end of an epidemic. Besides this, the most affected aspect is the work related one; and this is not only for hospitals or annexed jobs. In general, most jobs get quite reconverted from being physical to digital, working from home via computer, most likely.

During the first months (from March to June 2020) countries have differed regarding non-pharmaceutical measures. Some governments imposed rigid lockdowns: people couldn't go out from their home, with the only exception to supermarkets and pharmacies. So, the only workers' categories that could physically move to their job place were checkout workers at the supermarkets, pharmacists, truck drivers, doctors, policemen and politicians. All the other categories were stopped: they were obliged to remain at home. The biggest impact may take place inside either the educational environment or in sectors with the necessity of a human presence and interaction between clients and workers (i.e., bar, restaurant, pubs, shops, beauty centres, barber shops, massage centres and similar activities). This is so because researchers demonstrated that this type of virus can be transmitted by air or by touching each other.

Despite the health pandemic, types of jobs that allow working from home increase every year and the numbers are supposed to be even higher in the future due to technology. That is why investigating developments in the work-from-home industry is important. Firms can take advantage of customer opinions through their platform reviews. As a matter of fact, consumer behaviour represents a fundamental mean for managers to understand the actions they should take to be successful in the short, medium, and long term. To accomplish this purpose, it is necessary to comprehend how consumers think and feel about a specific product, or a how the market perception of a product category in general change over time. There are different ways to understand what characteristics people search in a product or service.

The study of consumer behaviour can be mainly discovered by the following means: surveys (both physical and digital); focus groups; depth interviews; online tracking. A novel and interesting method this thesis focuses on is the collection and analysis of consumers' reviews from internet sources. Comments are left mostly on the company web page, but also on blogs or third parties web sites, e.g. TripAdvisor. The main reason for customers to leave a review is that they want to express their opinion about a product or service, mainly to help other people in making a more accurate purchase choice. The fascinating aspect is that it is not only possible to discover what features they retain most important by looking at what aspect they talk about more often, but also to know the "direction" of their opinion (positive, negative or neutral). Obviously, in this way it is also possible to determine the general consumer preference, so what they think about the product, or the firm itself.

It is possible to classify the usefulness of this kind of data collection into four points: 1) increasing the understanding of customers; 2) identification of areas for development and growth; 3) forecasting the future; and 4) improving the personalization of items both on the substantial and the superficial aspects. For what regards the first listed point, it is very important to know your customers and to understand their preferences, as well as how do they evolve over time. Knowing what they like the most and their prospects about products, increases the likelihood of their level of satisfaction. The activity of analysing their opinions will generate, as output, a better comprehension of how a company can fill the gap between what they want and what they receive.

Second, the action of analysing data can make a firm noticing where it can be improved or, in other words, in which area it can expand its business. It can also suggest the sector to choose to diversify the offered products and services. For example, products divided per number of sales can indicate the most successful market to focus on and so indicating to enlarge the firm's market share in that sector by manufacturing other similar products or investing more on the existing ones. Moreover, the enlargement can concern the sale channels: online only, physical stores presence, where to locate the physical retail shops (for example, where most of satisfied consumers live).

Next, Forecasting the future prospect of the sector is needed since the market is always changing, especially in the today's technological era. Analysing data can assist firms in this intent. Observing trends about the most used channel to buy, the season most people bought some product, or they search that service after a particular event can help a company. It can contribute to make a company's profit reaches its peak, pushing up also customers' satisfaction at the same time.

Finally, the last point may improve the product's features that a firm yield to its clients and the communication related to it. Data can be used to know customers and their preferences. This knowledge can also be used as a mean to personalize the items and their marketing campaigns so that customers' willingness to buy and their satisfaction increase. Data analysis is also useful to facilitate the act of grouping customers into clusters and so to be more impactful on that cluster, by personalizing all the business activity for them.

This thesis research has several objectives: first it tries to identify the strong and weak points of a service, the reviews of their customers over time and across groups, and comparing the characteristics of similar companies; second, the thesis aims at investigating spillover effects in customer reviews. In theory, it could be that a period of bad reviews for a company may push clients to switch to another service offered by a competitor. If this is the case, this evidence would strengthen the idea that consumers do not only consider the reviews of one specific company before making a purchase, but they look at different services in the market and are influenced by the general sentiment. In other words, the review for a company's product affects the volume and sentiment of a product offered by a competitor firm. Of course, the necessary conditions are that these companies belong to the same sub-industry and that they offer a service that is more homogenous as possible. For identification purposes, in this way, we can isolate spillover effects of the reviews from any other inherent differences in the companies (e.g., service offered). Therefore, I decided to focus on the work-from-home freelance industry.

2. Literature Review

People spend a huge portion of their daily life in contact with disparate technological devices. They pass their time connected to an online platform, or at least with the possibility to be contacted by others through a web-based services. Because of these connections, the amount of data produced is so huge that data analysts must be careful to screen and select the right information to analyse and do not loose time with poor quality data or fake news.

Every action an individual makes on the internet usually provide useful information about their priorities, thoughts, beliefs, and their commercial inclinations. Just by surfing on the internet or sending messages, clients are leaving important details that companies can use to study the consumers' behaviour and any related theories.

Not all the researchers use the web scraping techniques to carry out their research. Unfortunately, it could represent a missed chance to increase their research quality and the output reliability. Basing a research on data can produce more accurate results and, consequently, it can improve the trustworthiness people release on that analysis. The reason behind this lack of data utilization can derive by an absence of knowledge about which method is the best to follow and the step to reproduce to accomplish this task. Nevertheless, web scraping can be a functional way to discover crucial insights about consumers' behaviour (Boegershausen et al. 2020).

In some cases, web scraping can also decrease time and cost for gathering data, being especially helpful for junior career scholars (Edelman 2012). In addition, the knowledge of this process can be particularly important to analysts who collaborate with restricted reserves' institutions, especially those who investigate costumers' behaviour, because web scraping can annul their initial disadvantages (Barnes et al. 2018). Web data scraper can be also useful to work with, as colleague, because of the realization of elaborated drafts, which can, for example, comprehend the collection of data from different web pages and compare them (Rafaeli, Ashtar, and Altman 2019).

The use of web scraping in research, mostly on those related to analyse consumers' behaviour, helps the data analyst to produce reliable, reproducible, and generalizable results. This method can also influence other researchers in the future to perform the same action so that everyone can benefit of the abundance of high quality research: it is an advantage for people everywhere.

Web scraping can also be used to extract information about people from all parts of the world, since they leave a digital, and not physical, track (Kosinski et al. 2016). Moreover, the more the data set characteristics are heterogeneous the more it augments the probability to be sure about the generalization of results (Maner 2016; Rad, Martingano, and Ginges 2018). Web scraped data can answer to several client research questions, like the relationship between company and customers (Wang and Chaudhry 2018); the effect of electronic word of mouth on sales (Hyrynsalmi et al. 2015); the nowcasts of food inflation (Macias and Stelmasiak 2019); and many others.

Internet can be a prosperous data source for analysing the consumers' behaviour, even if not all the researchers decide to take advantage from it, also because of a lack of capability. Specifically, big data analysis can find general norms applicable to an elevate number of subjects, that are not possible to discover through small amount of data. However, big data bring a series of challenges that are not easy to handle, like storage bottleneck, spurious correlation, and measurement errors (Fan et al. 2014). Moreover, about 80% of all potentially usable economic data are unstructured, that are difficult to be stored and so to be subsequently analysed to achieve useful information from them (Das and Kumar 2013).

In this sense, text mining is a technique used to extrapolate contents from textual data, like consumer reviews, comments, or general documents. It can identify, categorize, and decipher words as well as their context. Text mining is needed to automatically understand a large number of words and texts. Text mining includes the creation of a corpus (i.e., it basically consists of the union of all the words in the data set); the corpus cleaning (e.g., removing punctuation, stop-words and white space; transforming all the words in lower case letter; and stemming words); create a document term matrix and inspecting it (analysing the associate words, most frequent terms, tag clouds, and the sentiment score); and, last but not least, the corpus division in a series of topics. In particular, the sentiment analysis can be performed either at sentence level or at document level (or both). It will produce the polarity score (positive, negative or neutral) and its intensity (its minimum and maximum vary depending by the tools, method, lexicon and dictionary used).

Today, online users are even more important than in the past. They generate contents (UGC) useful for future consumers who will benefit from their comments about the proved service or a purchase. On the web, negative reviews can influence other users to express their opinions and to decide if it is a good idea to buy or not a product or service (Chatterjee 2001). In addition, online reviews can determine how much movies' daily box office performs well or not on the market, in terms of sales and revenues generated (Duan et al. 2008). Similarly, marketers studied the effect of online reviews on the amount of hotel room bookings, finding a notable influence of them on the final output (Ye et al. 2009). Moreover, there can exist also a differentiation among gender and age on the advantage received by online consumers' reviews. In fact, regarding travels in particular, female readers can take more information and generate more new ideas than males. For what regards the age, it depends on the reader conduct and intuition (Gretzel and Yoo 2008).

Other potential further research can include the integration of online and offline sources to collect the data and see the changes, if any, in the output (Berger et al. 2010; Datta et al. 2018).

3. Research Methodology and Data Analysis

The easiest, fastest, and most impactful way to participate and be noticed by others is to leave a comment in the dedicated space. With different extents, every people care about the judgement of others, because they give importance to their name. Among firms, there is no one that can permit to have a bad reputation. The reason behind that is explainable by the fact that if people have a negative opinion about a certain company, it cannot continue to survive. In a relative short period of time that firm will be expired because no one will buy anymore products or services it offers. Most people are not attracted by something considered as a poor quality, immoral, or simply not cool product or service. Through these comments it is possible to understand what common people think about a product/service, its defects, if any, and so the aspects in which it is deficient. At the same time, it is possible to know the positive aspects of an offers' range. It does not represent a less important aspect because the company can stress them to underline their superiority on competitors, during advertisements for example.

Therefore, companies care about comments that are left on social media, forums, and well-known digital marketplace, such as Amazon. Another different possibility is to find reviews on a dedicated website. One of the most reliable and famous sites is Sitejabber, which I have taken the raw data from, and will be useful for the analysis presented in this thesis.

Sitejabber is a digital platform where people can leave comments regarding numerous firms. It was founded in San Francisco, California. For more than ten years, it gave the chance to consumers to make more informed purchases and, at the same time, it makes companies able to monitor buyers' opinions. Sitejabber helps users to discover high-quality corporations and avoid fraud. Registered users have the chance to rate firms using either an overall system of rating stars or a criterion that comprehend the assignment of a single vote to different voices. They are service, value, shipping, returns and quality. In the past, Sitejabber has also obtain awards from different prestigious entities.

Companies monitoring their business can obtain a better understanding of their clients and so can undertake the right moves to increase their satisfaction. Hence, reading and analysing the people comments is fundamental to provide a better service to clients or possible clients. Currently, Sitejabber is used with this scope by more than 20.000 companies.

The limit of this web site is that it has not a dedicated space for a single product or service, but it divided its internal web pages by firms. Comments that are written by reviewers that are present in the same web pages can talk about different product or service, even if they should refer to something offered by the same company. Obviously, the number of different arguments discussed in the same page increased as the firm offers' range increases too.

However, on Sitejabber it is possible to also search firms by selecting a certain category. Among others, I decided to focus on the category called "Jobs" and in particular on the sub-category called "Work At Home". It comprehends the relative comments of online platforms where it is possible to find jobs. This kind of jobs are of special interest and relevance because they can only be performed at home. So, before the pandemic, they were perfect for someone who was searching for a second job as a source of increasing his or her monthly income. At the same time, it could represent a good option for some who preferred to work from home (or even from a desert island) respect to physically move to the office. Covid-19 changed this aspect from various preferences workers can manifest to a new standard they must adapt to. For this reason, it is interesting to see how workers' behaviour changed in the period before and after the pandemic outbreak. The fascinating aspects include the change in the number of comments left, their influence on others, the discussed topic and the rating stars' average, among others.

I have collected data from Sitejabber through the Google Chrome plug-in, Data Miner. Data Miner gives the possibility to download different rows, consisting on the consumers' reviews, and columns in an excel file. Columns consist of reviewers' name, text (i.e., the review itself), reviews' title, date when it was written, rating stars, and number of helpful votes. In this occasion I did not consider the photo, also because it is out of the scope of this thesis, but it is an interesting avenue for future research.

Going more into details, in this thesis I analyse the following five firms: "Freelancer"; "FlexJobs"; "UpWork"; "InboxDollars"; and "Fiverr". They all have a considerable number of reviews. Sitejabber also gives the possibility to see the ranking for each determined category, based on the average rating stars left by users. The selected companies are within the first six positions for the best performing firms within the sector, except for Fiverr.

It is interesting to discover the different characteristics associated to diverse performance levels and even the different reason behind them. Three, out of five, selected companies are base in the US (i.e., FlexJobs, UpWork, and InboxDollar), while Freelancer is based in Australia and Fiver in Israel. Their offered service consists of giving the possibility to some people to post a job offer (always jobs possible to be carried out in smart-work modality), and to other users to apply for that offer (obviously, they have to meet certain eligible criteria). They are online platforms that connect employers and possible employees.

3.1 The companies

Freelancer is an Australian based company with tens of million registered users and millions of published job offers. Furthermore, on Sitejabber, Freelancer has the greatest amount of reviews among the chosen firms, equal to 9,203 units, and an average rating score of 4.59 stars out of five. The latter value is the second highest in this category, after FlexJobs. This result represents a great success since it is more difficult for a company to maintain a high rating score with a large amount of comments. Moreover, thanks to the company's overview on Sitejabber, it is possible to discover that a recurrent indication made from contented clients regards support team, customer service and live chat. So, these strengths make Freelancer a distinguished company on Sitejabber. Next sections will analyse if these clients' references will have a confirmation both in the topic and the term frequency analysis. Furthermore, positive reviews in the last 12 months (even if it does not completely cover the analysis period) represent the 95.8% of total reviews left by all the users. More precisely, during the same past period, there have been 1,154 positive comments, 10 neutrals and 41 negative.

Consequently, in the last year, reviews have totally been 1,205, so about 100 for each month.

In addition, on Sitejabber it is possible to have the average rating stars' level and the number of reviews' growth during the past five years per company. There, we can see a quite stable average rating stars' level: only in 2017 it was slightly lower four, while during the rest of the period it seems to take a level of about 4.5. Instead, for what regard the number of reviews' growth, even if it obviously cannot decrease, for Freelancer it is also increased over the last five years. Last but not least, in 2018 Freelancer won the customer choice award.

The second firm to provide more information about is *FlexJobs*. It was founded in 2007 by Sara Sutton who was thinking to a different way to work, making life easier for workers. FlexJobs counts a total of 3,999 reviews associate to the important average score rating equal to 4.68 stars - the highest for this category. Through the Sitejabber offered overview, satisfied users had often used the terms "remote work", "great service" and "home position" inside their comment texts. This is always an interesting aspect to compare with the kind of analysis it is possible to observe later in this thesis. Moreover, the percentage of positive reviews during the last 12 months is equal to 95.2%. Within the last year, 3,512 reviews are positive, 164 of them are neutrals and just 14 are the negative ones.

It is possible to derive an important information from this data. The sum of the reviews left on Sitejabber during the last year is 3,690 for FlexJobs. It represents a huge amount in comparison to the total quantity of comments left for this company. It means that in the past years (without considering the last one) there have been written only 309 reviews, while in the last 12 months they have been more than eleven times that amount. This sudden increase can be explained by the introduction of the pandemic situation that obliged many people to work from their home. Looking at Sitejabber graphs, from 2019 it has been verified a huge change in the curve's slope: it is incredibly steeper compared to the previous years. It makes sense because, as expected, Covid-19 boosted surfing on internet with the scope to research this kind of platform. Instead, it is easy to understand that ratings always remained on a level higher than 4.5 over all the five years' duration period. However, it is dutiful to underline that it has been verified a small, almost imperceptible, decrease from 2017 to 2021. In fact, during 2017 the average rating was 4.76, while in 2021, as already expressed before, it is equal to 4.68.

The third company taken into consideration is *UpWork*, previously called Elance-oDesk till 2015. It was founded in 1999 in USA and its headquarters are located in Mountain View and San Francisco, California. The society complete name is UpWork Global Inc and it is listed in the US stock exchange under NASDAQ with the name of UPWK. UpWork has more than 12 million registered freelance workers and five million clients. Every year, on this platform there are published three million job offers, which total economic value is equal to about a billion US dollar. This makes UpWork the biggest online platform for freelance workers worldwide. On Sitejabber, users left 2,565 reviews about this firm and it has a consumer rating of 4.36. In their comments people used great platform, long term and full time as the most frequent expressions. Positive comments during the last 12 months, for this company, represents the 96%, the higher percentage among firms presented till now. Specifically, in absolute terms, the positive ones have been 1,161, neutrals have been just 8 and negative ones have been 41. For what regards historical average rating for the last five years, it slightly decreased from 2017 to 2019, while it had a much higher positive slope for the rest of the period taken into exam, especially from 2019 to 2020.

Almost the same steep increase can be found in the number of reviews' growth. In fact, a vertiginous comments increment from 2019 to 2020 occurred, continued with a lower extension from 2020 to 2021. Instead, during the three previous years (i.e., 2017, 2018 and 2019) the average maintained a quite flat trend. As before mentioned, the sudden increase can be caused by the Covid-19 virus. This consideration can also be verified by the topic and the most frequently terms used analysis, which could be find in the next chapters of this thesis.

The fourth company I am going to provide more information of is *InboxDollars*. It was founded in 2000 by the entrepreneur Daren Cotter. InboxDollars has millions of active users and since 2006 it has paid more than 60 million USD to its members. From its overview present on Sitejabber, it is obvious that despite its highly appreciated service, value and quality, InboxDollars has a scarce rating for what regards shipping and even more scarce for returns. Again, they will be all subjects for the analysis present in the next sections of this thesis.

Moreover, this firm has a total of 1,672 comments left by users on Sitejabber and an average rating score of 4.01 stars out of five. These last two value represent the lowest among the companies introduced till now. A confirm for the low consumer rating stars' level is constituted by the moderately low percentage of positive reviews during the last 12 months: it is 82.7%. One more time, it represents a lower value compared to the others here in this thesis. Expressed in absolute terms, it means that there have been 1,017 positive, 69 neutrals and 144 negative comments.

While, considering average ratings, it is possible to affirm that, as in the case of the previous firm (i.e., UpWork), there has been a marginally decrease during the three-year period between 2017 and 2019. At the same time, always similarly to UpWork, it has been registered a high remarkable growth from 2019 to 2020. The latter phenomenon is continued with a much less intensity also in the subsequent, and last, period of time. The number of reviews, also, manifested a similar trend: it shows a relevant rise from 2019 to 2020. However, in this case, the positive tendency continued, with almost the same rate, for the following time period, so from 2020 to 2021. Moreover, it can be said that a quite flat trend for the first three years was displayed (i.e., from 2017 to 2019).

Finally, *Fiverr* is an Israelite company based in Tel Aviv and founded in February 2010 by Kaufman and Wininger. Even if it collaborates with people from all around the world, it has not a good name among Sitejabber's users. In fact, Fiverr has a consumer rating of 1.98 stars out of five, indicating that most people who left a review was not satisfied with the provided service. Note that this firm has been specifically selected in order to obtain another point of view on this sector: so the reasons why an online platform of this kind shouldn't work properly. That is, I did not want to address the analysis just considering successful companies in the chosen industry.

Thanks to the Sitejabber's overview it is possible to know that reviewers most frequently mentioned customer service, credit card and low service quality problems. This phenomenon will always be verified during the analysis presented successfully in this thesis. Furthermore, people left only 1,058 comments about Fiverr on Sitejabber, it also represents the lowest amount among all the five selected company of this thesis.

Confirming what previously stated, positive reviews in the last 12 months are only the 17.9%. In absolute values, 10 have been the positive ones, just one has been neutral and 45 have been negative.

In addiction also from average ratings it can be possible to state the flat negative trend from five years ago till now. There was been a very soft pick in 2019, even if the line never overcome the level of two stars. It may depend on the relative small firm life cycle, and so by the moderate later entrance in the in this fast going market. Instead, by the number of reviews growth it is possible to underline a steeper increase during the first two years' period (i.e., from 2017 to 2018 and from 2018 to 2019). While, during the rest of the years, can be registered a flatter tendency. It has been registered the opposite for the other companies presented, maybe because of the pandemic explosion in the first months of 2020.

3.2 Procedures

The first operation has been the data scraping. At the beginning, I wanted to find the most efficient way to obtain the most possible homogeneous data, especially regarding the columns in the dataset.

Secondly, I cleaned the dataset to have the data ready to be analysed, because right after the data scraping it usually happen that they are not so organized, precise and structured in an optimal condition.

After that, I started the data analysis. The objects for this paragraph, successfully better illustrated, are multiple. Before anything else, giving a complete comprehension about the difference among the selected platforms; the reason why they differ for what regards the users' satisfaction level, the number of reviews left on Sitejabber and most the discussed arguments; successfully, it is interesting to study if the most frequent words presented on the Sitejabber's companies overview correspond to the most frequent also on R; then analysing if the rating stars scores matches those derived by the different dictionaries existing on R; after that also studying the numerous n-grams generated by R; finding the optimal number of topic in order to include all the reviews both for a single firm and also considering all the companies together.

3.2.1 Data Scraping

For the data scraping section, I used Data Miner, the Google Chrome extension. Data Miner allows users to create a dataset to insert all the data present in a web page. However, the data collected must be storable in an excel file: for example, a picture (like a user's profile picture) cannot be transformed in a content adoptable in excel. So, the content has usually to be written words, or other type of characters like numbers, punctuation character or other special characters. At the same time, also colours or visual aspects, in general, cannot obviously be reported. Nevertheless, I was able to extract the number of stars left for each comment, even if they were indicated by drawn yellow stars. So, countable items can be scraped even if they are not originally expressed in a character format (like numbers).

So, after the diverse operations to scrape the data, I collected eight columns. In particular, they are *Author*, containing the author name and the pointed surname or directly the nickname; *Review Number*, corresponding to the number of comments left for that author, helping to understand how much that user is available to left multiple reviews; *Number of Helpful Votes*, stating the number of people who consider that comment useful in order to know something more about that firm or product/service they are commenting to, or in the case they agree with that statement; *Title*, representing the summary, in few words, about what it is written in the entire review, or, in other cases, it may simply represent the starting words for that comments and the platform register those words as title; *Stars*, containing the number of stars, so the level of satisfaction for that company, product or service by that corresponding author; *Date*, simply corresponding to the date when people left the comment on Sitejabber; *Verified Purchase*, expressing if the purchase is certificated or not, so if the person writing the review has really bought the product/service or he/she has left it without a certainty about the purchase; *Text*, stating the review itself, the entire comment left from the user.

3.2.2 Data Cleaning

Before, performing the data cleaning functions, I run the libraries. In the case of this thesis, they are the following: readxl; stringi; stringr; plyr; dplyr; ggplot2; ngram; tm; wordcloud; dbscan; proxy; esquisse; ggthemes; tidytext; factoextra; NbClust; sentimentr; syuzhet; textdata; parallel; plotly; dplyr; lexRankr; and stm. They are fundamental to run functions.

However, after importing in R all the five scraped excel datasets through Data Miner, as before mentioned, I added a column where I inserted just the company name in each row, as a content, for each of the five firms. This action was made in order to recognize, in the future, to which company that comment belongs.

Right after that I created a unique data set, in which I unified all the datasets for the five companies taken into exam. I was able to perform this action through the function *rbind*, and R gave in return a data set of 2344 observations.

Subsequently, I started to perform the cleaning part. Hence, as a first operation of this kind I wanted to take just the unique rows, eliminating in this way all the rows that were repeated multiple times, leaving just one copy of them. It can happen during the data scraping that the Data Miner system duplicate the same row (so, the same comment, in this case), with the exact number and type of information, multiple times. My objective was obviously to maintain all the rows without any duplicates. This intuition was important because the number of rows in the data set decreased from 2344 to 2319. From this reduction is possible to know that R deleted 25 rows because of a useless repetition of data. This is important to avoid duplicates influence on the average sentiment score, the average rating stars' level, the topic analysis and all the other analysis aspects. In the specific, the filter deleted 25 rows coming from the Freelancer dataset.

The next operation performed was removing the word "review" from the Review Number column, because it was a detail that did not give me an important information. I used the function gsub to have the desired result: maintaining only numbers and eliminating the word "review". After that, I wanted to transform the content present in the review number's column (only numbers) in a numeric type. In this way, all the possible operation I had in mind to perform, I could do them knowing that R perceives that content as a number. Then, I performed the same operation for the term "helpful" present in the column "Number of Helpful Votes" and the terms "rating stars" present in the Stars' column.

The same applies for the act of transforming the entire content (only numbers) existent inside the Stars and Number of Helpful Votes' column in numeric. In the case of Stars column, R removed also the ".0" part present right after the number, because equal to every review and so not beneficial for making any type of comparison.

Now, it is possible to observe a more interesting phenomenon: transforming the perception of R about the date column. At the beginning, before the cleaning it was represented in the following way. The date structure was months name in full; day expressed in number followed by "th" (or st, nd and rd); and then the year always in number. It was a sort of mixed composition between numbers and letters. And, most of all, R couldn't recognize it as a date format. So, it was needed to change it. For this purpose, I was obliged to firstly divide the date column's content in two separate part: the first was composed by the month and the day, while the second only by the year. I recalled the new created column as "Month&Day" and "Year", respectively, and I added them to the data base.

I effectuated the same division operation for month and day in order to have them in other two separated columns. In this way I had three different columns containing the same content as before, but split, since I deleted the “Month&Day” column because it was not needed anymore. I had also another column with the original content as the last image present above, but then I changed it in the date format. I had also to substitute the month denomination in letters to numbers (i.e., January corresponds to 1 and etc.) Then, I deleted the not more useful columns: they are called “dd”, “mm” and “Date”. I deleted the old Date column and I rename another one with the same name, but the new one is expressed in date format, with the *as.Date* function.

Finally, in order to be more accurate, I removed the space in front of the Year column number and I transformed its content in a numeric format. These were the last performed actions for the data cleaning part. So, at the end, I had three different columns expressing the year, month and day of each comment and another separated column expressing the date all in one.

3.2.3 Data Analysis

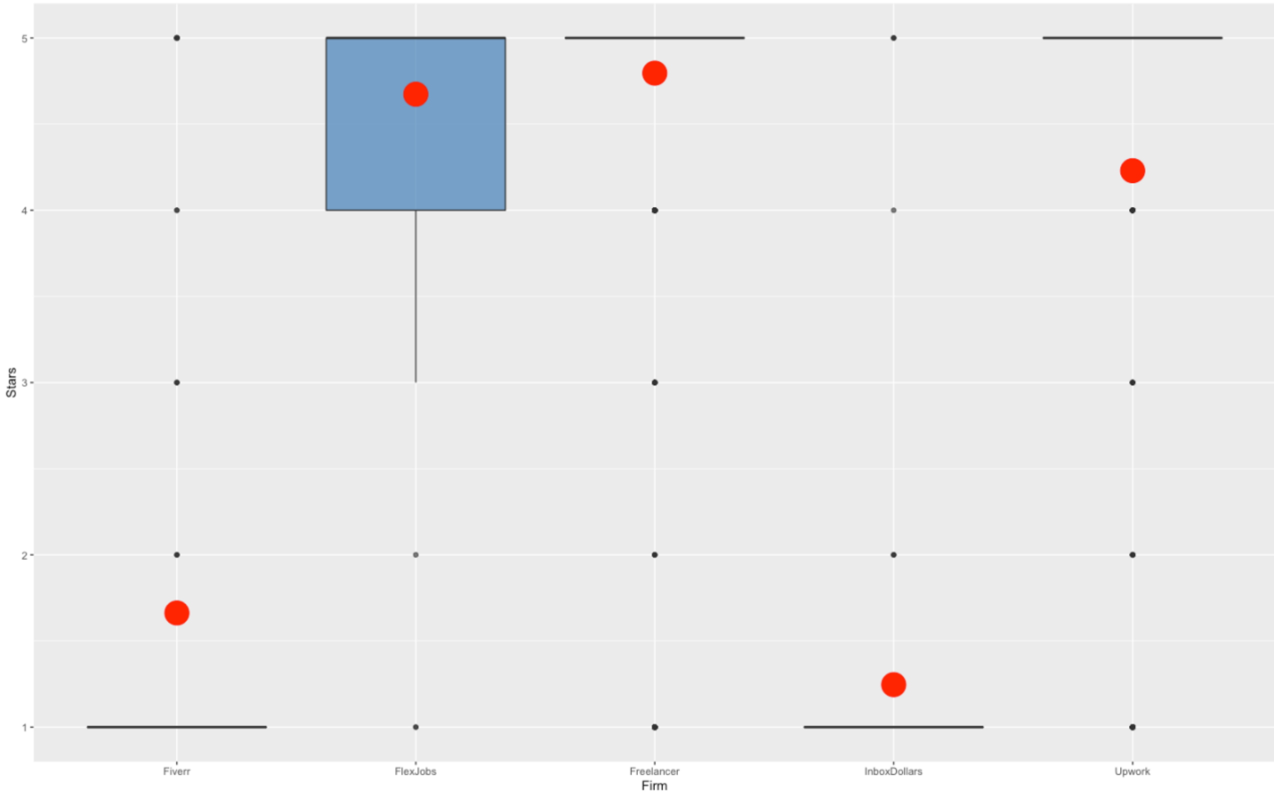
The first operations for the data analysis part were needed to obtain more information about the data set. For example, I wanted to know how many time each stars appear, so the number of comments with the same rating star level. In order to achieve this information, I used the *as.data.frame* function.

The most frequent rating star’s level is represented by five, that it is obviously the highest. In particular, remembering that 2319 is the total number of comments, 1820 reviewers left the maximum assessment level. An interesting observation is represented by the number of stars equal to one: its number is the second highest and it is equal to 244. At the same time, 210 times people assessed four as the experience vote for the five firms taken into exam. So, its frequency is higher than comments with rating both equal to three (26 times) and those equal to two (equal to just 19 counts).

From this evidence, it is possible to understand that most people left a positive judgement (stars equal to five is about 78%). However, a second insight is represented by the fact that central positions (two and three stars) are those with the lowest number of frequency. While, extreme ones (excluding five for a moment), so one and four stars, are those with a higher recurrence. It means that people want to leave an impactful review to let of stream, express their emotions and also to help future clients to make a better purchasing decision. Hence, if they had a negative experience, most of them assigned the minimum (e.g., one) rating stars’ level, and not a vote slightly better (like two stars).

However, in order to have a better and more rapid intuition, it is also possible to directly have a look to the histogram representing the same phenomenon: so the frequency of each rating star’s level in the data set.

Together with the histogram, I realized a box plot, expressing the rating stars’ level divided per each firm present in the data set (e.g., Fiverr; FlexJobs; Freelancer; InboxDollars; and Upwork).



In this way it is clear the difference among the different rating star's level per each firm, even because graphically it is more impactful. In the box plot present above, the horizontal lines represent the median, while the red points represent the average rating stars. Thanks to the box plot it is possible to note that Fiverr and InboxDollars have a median equal to one and a mean lower than two: in the case of Fiverr it is above 1.5 while the opposite is true for InboxDollars. However, all the other three firms have a median equal to five and an average rating stars above four. Upwork has a mean slightly higher than four and lower than 4.5. On the contrary, Freelancer and FlexJobs have a mean above 4.5 and, between the two (but also among all the five companies), Freelancer is the one with the highest mean: it is very close to five.
Nevertheless, red points and horizontal black lines all reflects the value reproduced by R, calculating exactly them with the appropriate function *summarize*.

Differently, the zero level of helpful votes has the higher frequency (equal to 1959) among all the other. It is very distant also from the second position in the ranking, occupied by the category one number of helpful votes with a frequency of 109, almost one twentieth. However, it is true that as the number of helpful votes increases, their frequency decreases, with the exception of few cases: 6, 10, 12, 13, 14, 20, 26, 36 and 37 number of helpful votes. However, generally they are inversely proportional. I realized the histogram also in this case.

The same approach was verified for the Review Number column content: there is a category with a huge frequency, enormously bigger than the others. Also in this case the most predictable outcome is the one with the higher frequency. In particular, people who left just one comment were 2207 out of 2319. The other categories are only twelve more and they are hugely distant from the first one. In the second position it is findable two as review number and it has a frequency of 72. All the others are linked to a number even lower.

The whole data rating star's mean has a mean equal to 4.44. It states that both the sample and the all data give in return a quite high mean, always considering that the possible maximum is 5 stars. Moreover, since in general the sample has to be the most representative as possible, this is the case in which the expected event is respected.

Another fascinating aspect is constituted by seeing the number of comments grouped per each month. It may depend by the possible increase or decrease in the market demand and supply; by a possible new company introduction or old company failure; but mostly by the pandemic outbreak. Remembering that the gathered data regards the period from September the first, 2019 to August the thirty-first, 2020, so a full year. The month with the higher number of review left is August, with 545 reviews. The second one is July with 305 reviews and the third one is June with 208. The surprising fact to observe is that the last month (i.e., that with the lowest number of comments) is the one located right after the pandemic beginning. In fact, the last one month in this ranking is April with 62 observations. For what regards the others, a distinction is needed: the first six positions are all occupied by months from September to November 2019 and from June to August 2020, so the first three and the last three months of the chosen period. While, the months in the middle of the period, so those from December 2019 to May 2020, occupy the last six positions in the ranking. This means that during the autumnal and summer periods people are more inclined to leave a comment. At the same time, very probably, they are more inclined to use the service the analysed platforms provide, too. This event can be due to the greater amount of free time people has during the summer months and the usual end of fixed-term contracts coinciding with the academic year's end. Then, if people did not find any suitable job during those months they continue to search during the subsequent months (i.e., the autumnal ones).

At the same time, taking into consideration the job particularity, another reason for searching jobs performable at home is that people may want to travel during summer and they desire to find something that simply needed only an internet connection. In this way people can discover new places and meet other people, but simultaneously they can earn some money useful also for the holiday itself. Of course, the already described represent possible motivations, but there can exist many others of them.

In addition, the month of May is situated in the seventh position, so right after the summer and autumnal months. This situation can be explained by the advantage some people take to search another job before the interested period. So, before their fixed-term contract conclusion or before their holiday period. In this way, when the summer comes they are yet ready to enjoy the good time.

However, the Covid-19 outbreak phenomenon does not seem to provoke any radical change. The only aspect to underline, as already said, it is the last position occupied by the month of April (the months right after the Covid-19 starting, in fact the official date provided by the World Health Organization is 11 March 2020). Moreover, March itself is also placed in one of the last positions with 112 number of comments. More precisely it is the third one from bottom.

The main reason explaining the few observations for April, March and May can be the jobs market stop for the months right after the Covid-19 beginning. So, people left less comments than in the other months because they had less experience to share with others. During that period (i.e., spring 2020), Governments obliged people to stay at home thinking that the situation would be finished within some months at maximum. Hence, people did not think about searching a job to perform from their home because they thought the subsequent months they would be able to freely go out and so also working outside their houses. That time, people who had a job and so they had to go daily to the office, they hadn't to go there anymore. At the same time, unemployed people may have thought finding a job where it was needed a physical presence because hoping to come back to "normality" soon again. The first times, there was not the general opinion that this pandemic would change the way of work, as at the time of writing (Summer 2021) it is quite diffused.

Moreover, Governments helped also with economic aid to unemployed people and to owners who were obliged to close their activities. So workers for those kind of sectors, may think to live with that amount of money without searching for another job, even because without going out the expenses decreased. By contrast, when people realized that the situation would not change soon (i.e., during the summer months), they start to search for jobs performable by home.

After that, I run the command useful to know the number of reviews left grouped not only per month but also per firm. So the same division encountered before is once again grouped also per company.

Ranking them in an increasing order, it is possible to discover that in the first position it is located the firm FlexJobs in August with 341 number of comments left. While the next four positions are all occupied by Freelancer in October, September, November and February with 188, 175, 154, and 107 reviews, respectively. After that, there are occupied by Upwork for the months of July and June with 107 and 100 observations, respectively. Freelancer is the firm with most counts, manifesting itself ten times out of the 16 rows. While, even if there are present other two firms (they are Flexjobs and Upwork, with two and four apparitions, respectively), the remaining two (i.e., InboxDollars and Fiverr) are not present at all.

However, for what regards months, it is possible to say that even if August is in the first position, it appears for the second and third time only in the 12th and 13th positions, for this special ranking. It is also true that these three observations summed together give a result similar to the total observation in August, without considering the firms (527 against 545).

Nevertheless, there is not a month present more than three times and the other two months with exactly three apparitions are June and July. In fact, they are third and second, respectively, in the ranking of number of comments just divided per months. So, that high position in the ranking is verified also here with a number of apparitions higher than the other months. Of course, these three manifestations are linked to the firms Freelancer, Upwork and FlexJobs, for both June and July.

Instead, April and May are the only two months that are not present for Freelancer. In particular, the month of April is not present at all within these first 16 rows. Indeed, April is the last month in the ranking of comments grouped just for months.

Finally, for what regards the spillover effect, it is possible to hypothesize that the huge number of reviews left on June and July on the Freelancer and Upwork dedicate pages, produced a positive effect on the month of August for FlexJobs.

Rather, then I run the codes to obtain the average rating star grouped by months. Here, July represent the month with the highest score (4.65), while August is situated in the second position with 4.58 stars. So, they inverted their positions respect to the comments per month ranking, where August was at the first place and July at the second one. Surprisingly, April has not only the lowest number of comments (as I described before) but also the worse average rating stars' level (3.45). It is even the most distant from the previous position. In fact, in the second last position there is March with 4.01 stars. Their difference is equal to 0.56 while the maximum distance among all the other positions in the ranking is 0.14 (it is among May and November, located in the 7th and 8th positions, respectively). This insight is useful to underline the great gap between April's average rating star and all the other ones.

In addition, March is located in the second last position, here (with 4.01 stars), when in the number of comments divided per months' ranking it was in the third last position. So, again, as in the case of April, a low level of average rating stars is linked to a low number of comments. On the contrary, this trend is also confirmed by the positive (or high) level of average rating star linked to the high number of comments left during a certain month (like August and July).

For the average rating score divided per months and also firms, instead, FlexJobs is located at the first place for the month of July with 5 stars. It is so because, looking at the number of comments for that firm in that month, reviewers left just one of them. Hence, it is obvious that if the unique reviewer has left a judgement of five stars, so also the final average will be equal to the same amount of stars. Rather, the next nine positions are occupied by Freelancer for nine different months of course. There is only one interruption due to Upwork for the month of July, located in the seventh position. So, the top ten ranking is composed by eight months linked to Freelancer, one to FlexJobs (the first place) and one for Upwork. The manifested months for Freelancer are May, January, September, October, March, July, December and February, respectively in the order they appear in the ranking. Only after March for Freelancer, it is the time of July for Upwork.

However, they are very close to each other: it is enough thinking about the difference between the first and the tenth is equal to 0.215. February for Freelancer, holding the tenth place, has an average of 4.785. Despite there is not a so huge difference in the values, they have a different number of comments, as I described before. Consequently, the incidence of each review is dissimilar for each month and firm analysed.

Furthermore, the minimum possible value to assign is 1 and it is also the average of some observations in this ranking. This means that all the reviews left in that month for that particular firm, assigned a value of 1. In the specific, they are 15 out of 53 total observations. They regards all the firms at issue, except for Freelancer. Going more into details, they are 7 months for InboxDollars; 4 months for Fiverr; 3 for Upwork; and just one for FlexJobs. The months in questions are quite heterogeneous among all the seasons.

In addition, it is curious to note that there is not a single position in this ranking, giving an average that is between 3 and 3.99. From a judgment of 4.52 there is a gap till another judgement of 2.66, from a position to the next one. This is a confirmation that the reviewers influence each other, because they follow the trend left by the previous writers for that firm in that month. If, instead, there would be a diverse judgement among them (so, reviews with assessment similar to 5 and other ones similar to 1), the average would be about 3. Since this result is not occurred they left similar evaluations for a certain firm during that month. It can also be the case that the firm at issue was performing particularly well or in an unsatisfactory way. If this case is verified, all the people had the same impression about the firm service in that period and so they left similar judgements. However, the probability that this case occurred is low, statistically speaking.

Then, I run also the string of code useful to have both the number of comments in each month and the average rating stars' level, in the same table. This operation is also useful in order to compare the two values: how the number of comments is in relation to the average rating stars. However, I have already express an opinion for this relation.

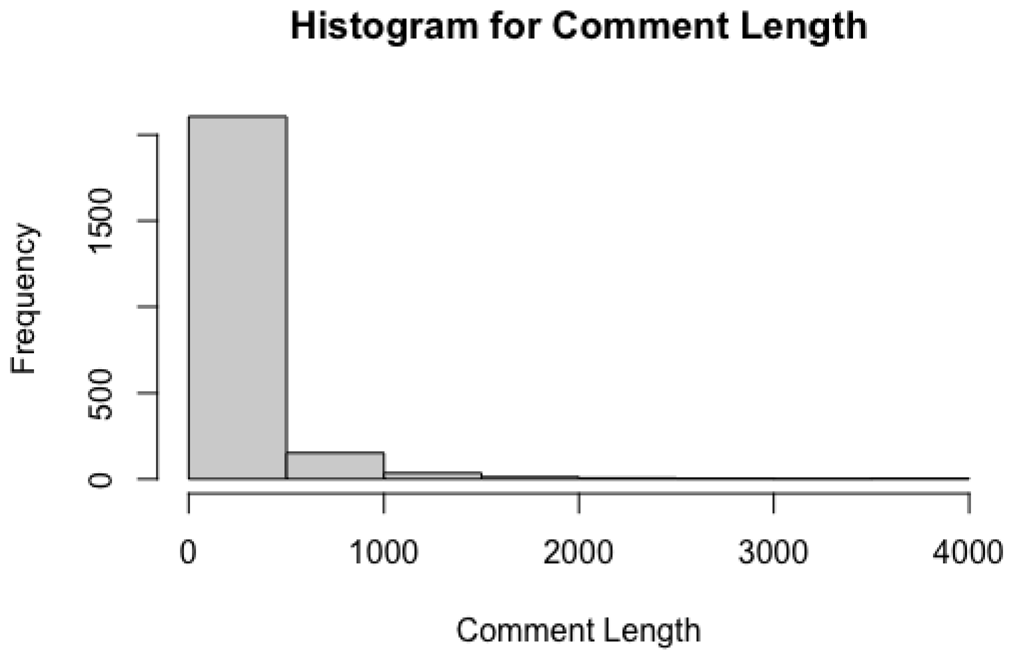
In addition, I have also plotted these observations in a graph. It is composed by the number of comments on the x-axis and the average rating stars' value on the y-axis. Through it, it is possible to see that there is just one month with a very high level of number of comments and also a high average rating stars: it is August. It is located in the upper-right side of the plot. While, on the opposite part of the graph (in the bottom-left side), there is just a month with a low level for both of the two variables: it is April. After that, I have also performed the commands to see in the same table the average rating stars and the number of comments' vales both divided per firm and per month. Also in this case, I realized the graph showing the various observations: they are 53.

4. Results

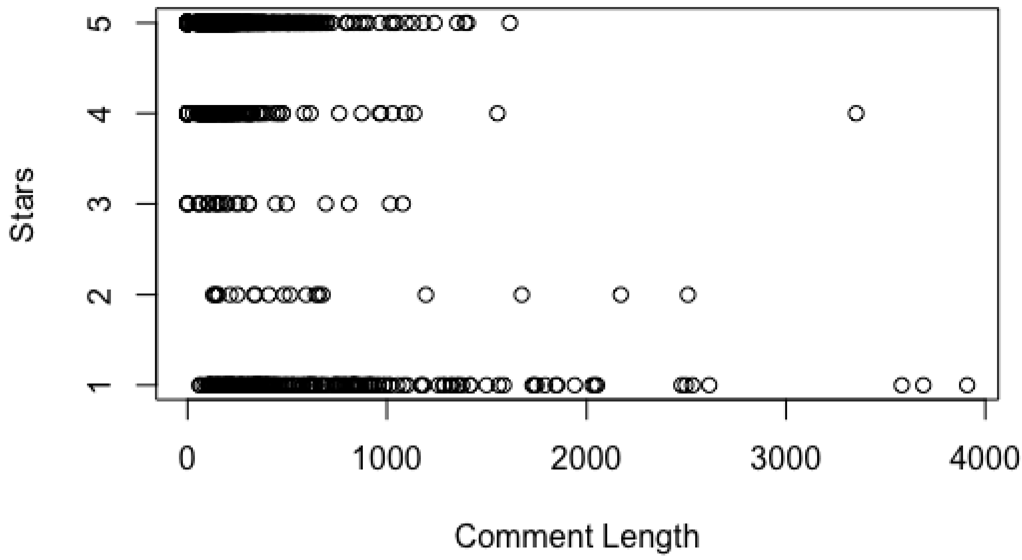
This section explores the data from a textual analysis point of view. Various strands of literature have leveraged textual and sentiment analysis approaches to understand consumer reviews (for some of the most recent papers see, e.g., Liang et al., 2015, Archak et al., 2011, Salehan and Kim 2016). This method allows to collect precious information on users' behaviour, exploiting potentially different aspects of the human being (e.g., culture, gender, ethnicity, etc.), which ultimately shed light on how people make sense of the world. Of course, one important step of the textual analysis is to make assumptions on the most likely intuitions that might be made about the text. This approach is used in many disciplines: from economics to management, from cultural study to philosophy, among others.

This thesis adopts the textual analysis approach to interpret and classify how consumer behavior in the work-from-home industry has changed with respect to the pandemic onset. As introduced in previous chapters, the idea is to focus on one unique industry to be able to isolate any substitution effects between firms. This is done by means of investigating user reviews left on the Sitejabber website about 5 firms.

The first step is to check how long the 2,319 comments are, in general. The histogram below presents the various bins in terms of length.



We can clearly see that, overall, comments are relatively short. The distribution of comment length has a long right tail, meaning that very long comments are rather rare. The average comment length is 243 characters, including spaces between words. This may suggest that users leave a short but concise comment after using the service. This is just a guess. Let's explore this intuition in more depth. The graph below shows the relationship between the length of a comment and the number of stars with which a consumer rates the service. One message we can derive from this relationship is that comments with high ratings have a shorter comment. They also have lower dispersion in the comment length. This suggests that consumers with negative reviews (1 or 2 stars) spend more time in describing any problem they might have encountered during the service received. On the other hand, positive attributes might not be explicitly typed in a comment. This may represent an aspect in which companies can improve. For example, firms may incentivize users to spend more time in writing their comment, even if (and especially so) they had a positive experience with the firm. One such way to achieve that is to reward consumers with bonus points to be spent on the platform later on, or to provide them with some gadgets or small tokens upon accurately commenting.

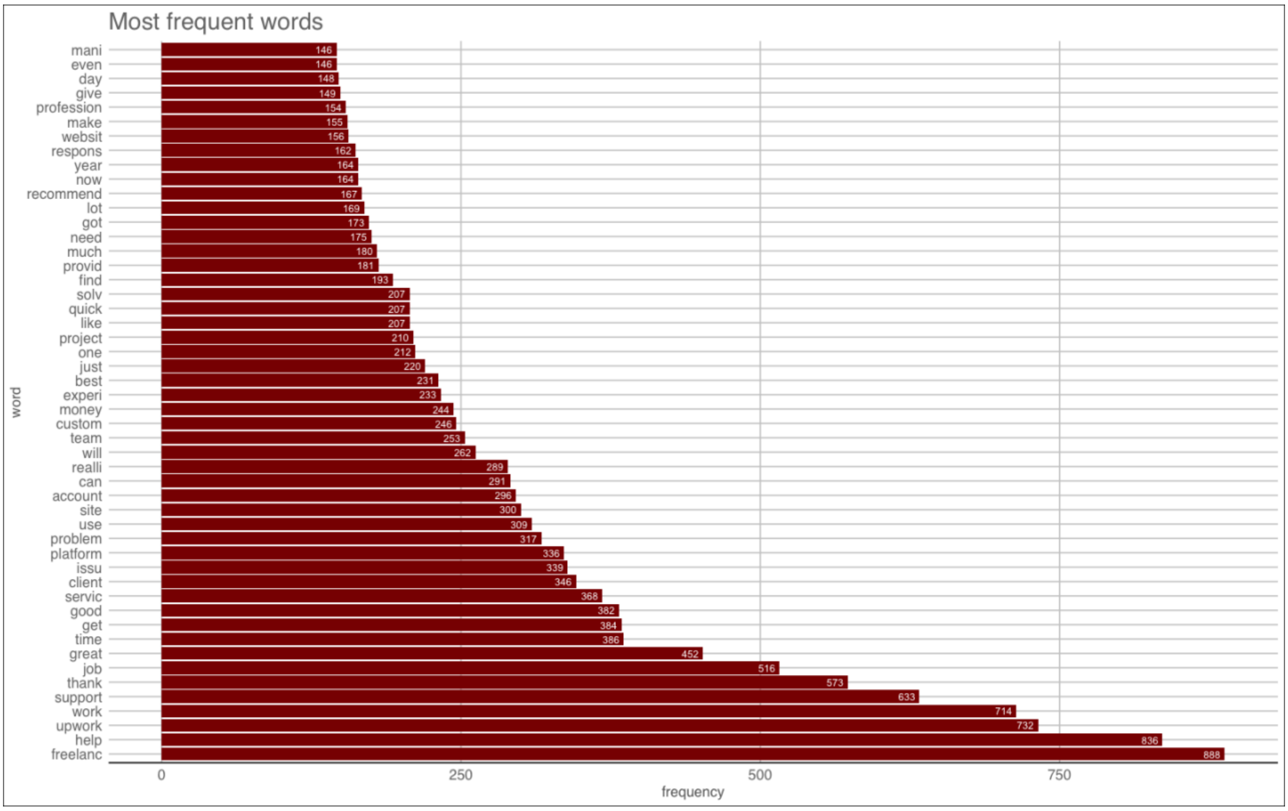


Next, an interesting dimension to explore is the word cloud. This allows to picture all words in one single image, exploiting also the commonality of words used across comments. The image below represents the word cloud of all comments released for the 5 companies under scrutiny.

We can see that a few words appear the ones to be most typed: support, quick, answer, fast, solve. These words relate to the concept of the quality of the service received, especially when it comes to post-sale service. Not surprisingly, something went unplanned for consumers which they had to solve along the way, either with the help of the platform managers and customer service or without it (this latter case includes either solving the issue alone or with the third-party user that demanded the job). In any case, the sentiment on post-sale service could be positive, as suggested by the terms in the word cloud that remind of favourable conditions.

This is an issue we will explore in details in the sentiment analysis later. One thing to note though is the fact that the word cloud above is a partial picture because it only includes a portion of the total words.

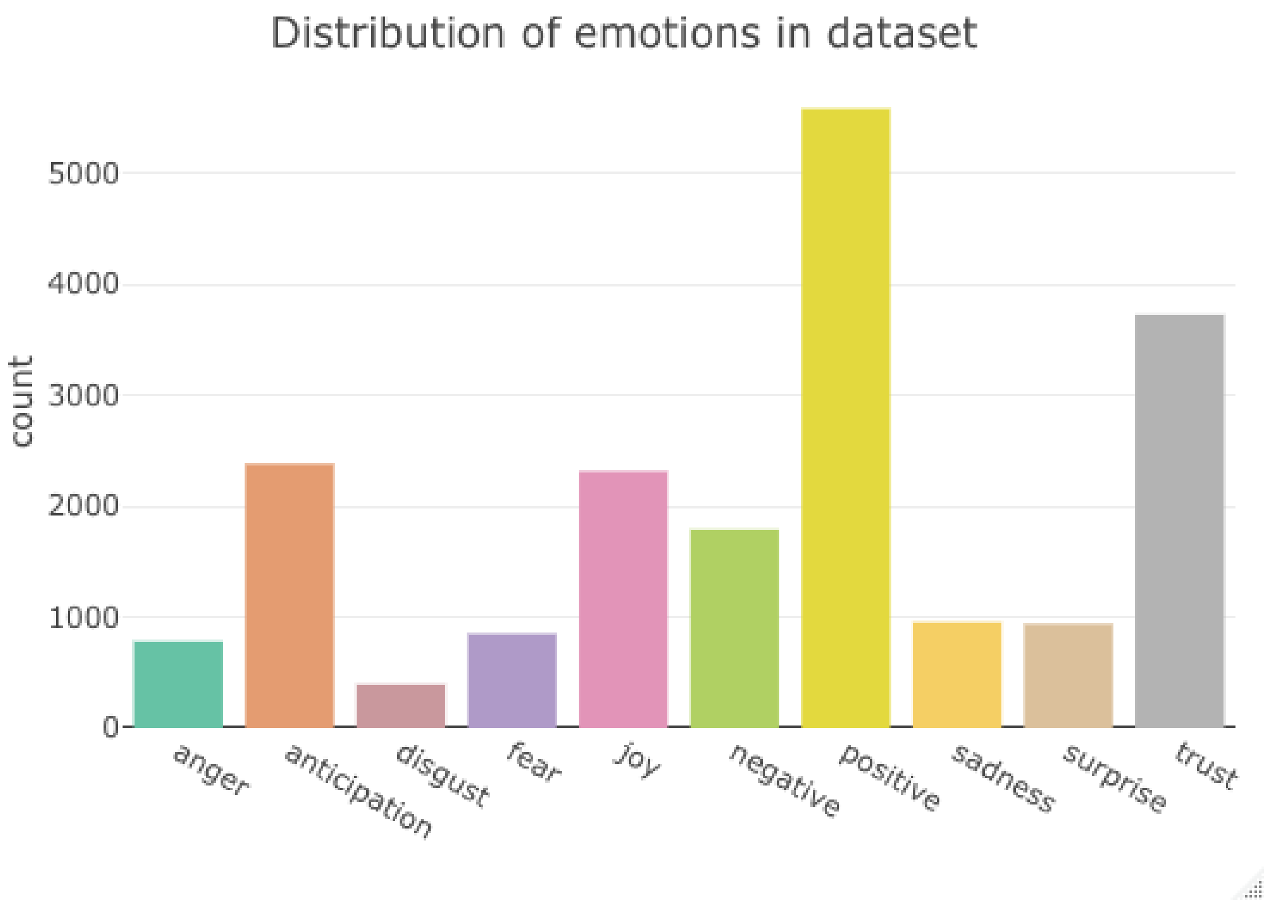
Looking at the same issue from a different angle but from a more comprehensive perspective could be done through an histogram of most frequent words. The graph below describes the distribution of number of times a word is typed among comments.



We can see that some of the most used words are the same as in the previous graph. For example, “support”, “problem”, “quick” are only some of the ones that appear tin the top 20. However, in this complete overview we can see that some other terms appear more systematically. For example, these are: “freelance”, “upwork”, “job”, “platform”, “account”, “profession”, etc. These terms are related more to the type of job itself. In fact, it seems that many users start describing carefully what their service consisted of to other users, likely to do a sort of advertisement as well. In fact, in this way freelancers could attract new demand and future job opportunities.

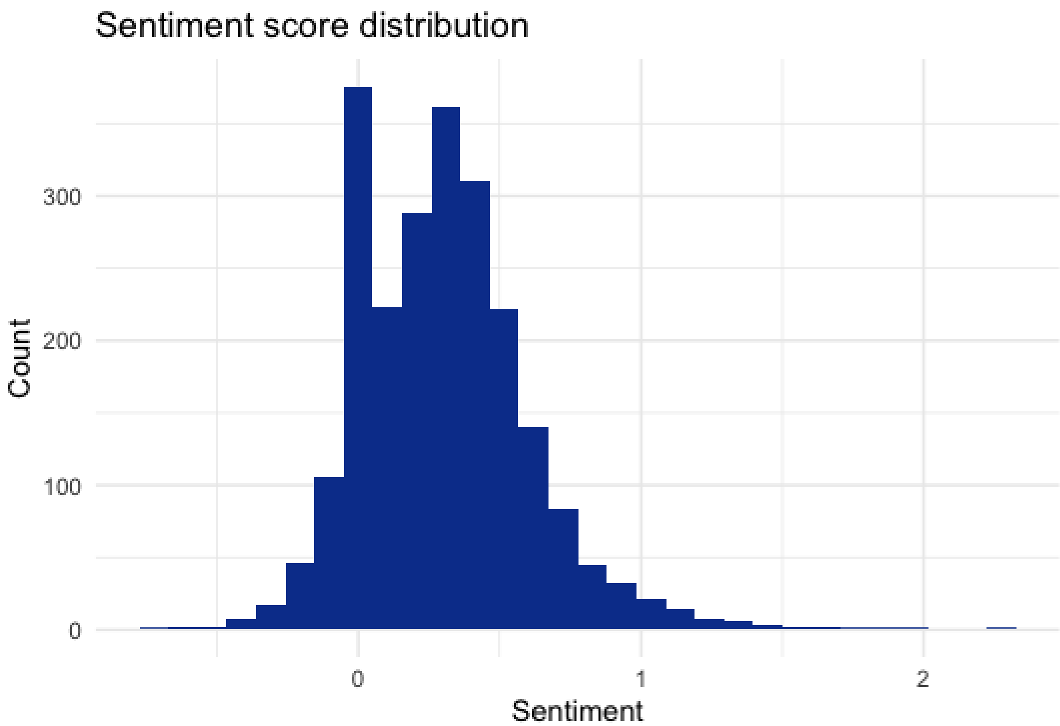
4.1 Sentiment analysis

In this section, we start exploring the data by a sentiment analysis point of view. We use four different dictionaries in R: syuzhet, afinn, bing, nrc. These sets differ with respect to the number of words and also by the weight they put on each single word. Moreover, the weight range for each word is different from each other: syuzhet and bing go from -1 to 1, while afinn and nrc from -5 to 5. This means that extremes are over-weighted in the latter with respect to the former cases. In particular, the graph below is based on nrc. This is because this type of dictionary already classifies the word in a comment by type of emotions. For example, more positive feelings are associated to the following categories: “positive”, “joy”, “surprise”, “trust”, "anticipation". On the other hand, negative feelings are: "anger", "disgust", "fear", and "sadness". Notably, a specific word may fall in multiple categories.

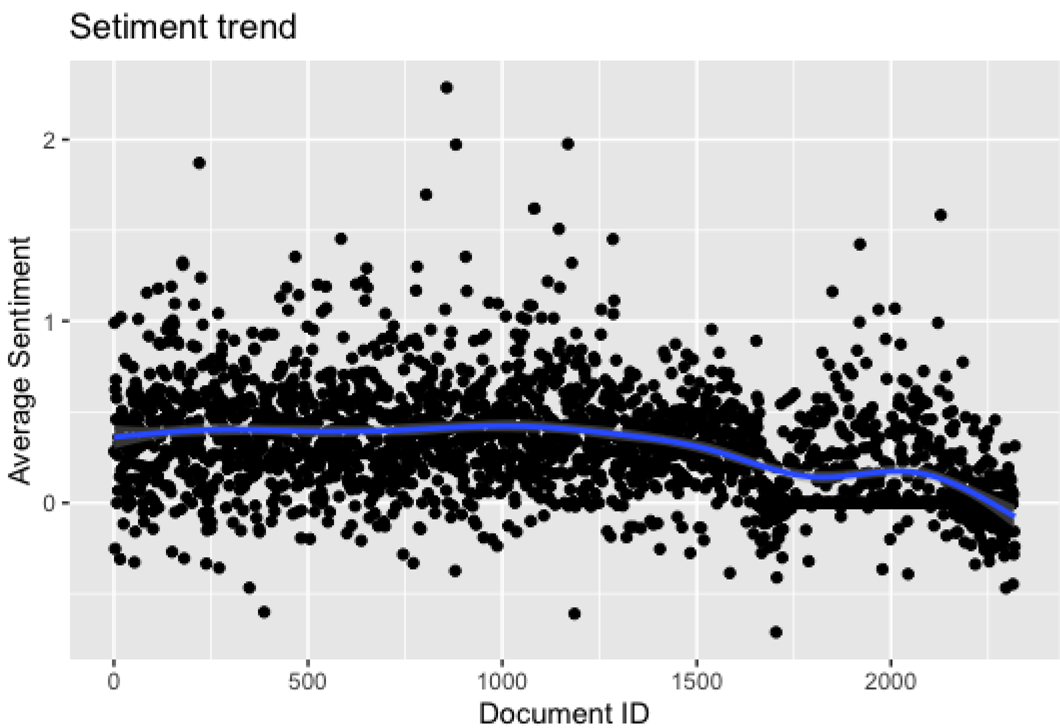


We can observe that users are overall users are satisfied with the service received, as the number of words in the more positive categories are more frequent than those in the negative ones. In particular, the "positive" bin is higher than the "negative" one, "joy" is higher than "sadness", "trust" is higher than "fear", etc.

One interesting attribute of the sentiment analysis is also that feelings can be mapped into numeric values. This is important to gauge the overall review from a quantitative perspective. In particular, the sentiment analysis here is performed at sentence level, through the *sentimentr* algorithm in R. The longer comments are first split into multiple sentences. Then, each sentence is analysed as a standalone phrase. One limitation with this approach is that it does not take into account the possibility that sentences within a comment might be linked to each other. On the other hand, it explores the comment more granularly, because it could be that a comment includes both positive and negative parts.

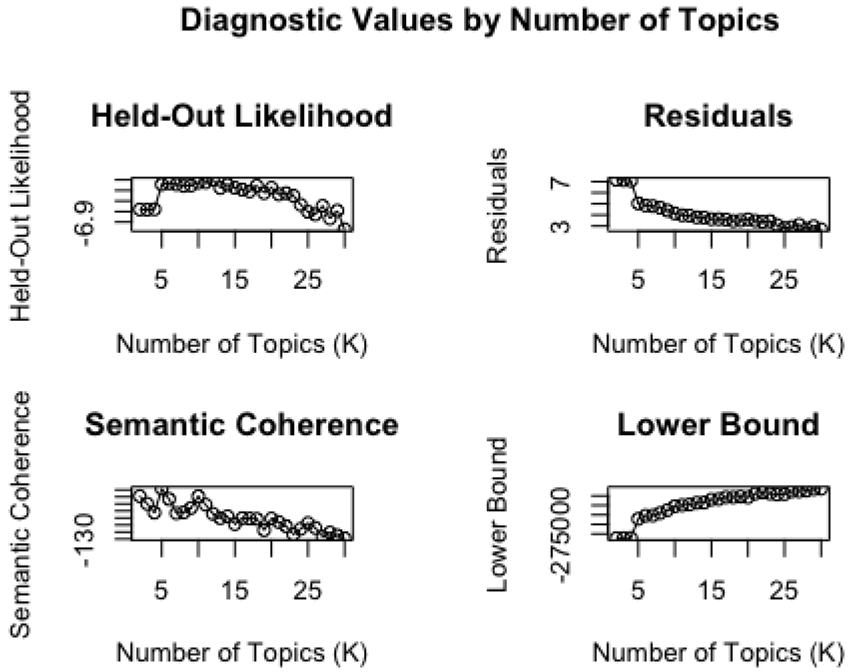


The distribution of the sentiment score across sentences is centered around . The shape reminds of a normal – or bell-shaped – distribution. The most common score is 0, which means that most sentences were neutral, i.e. neither negative nor positive. This distribution includes positive values mostly, and it reaches 2.3 as the highest number. Interestingly, the score varies across sentences, but also across comments. In the graph below, we group sentences within a comment ("Document ID", on the x-axis), and plot them with their respective score.



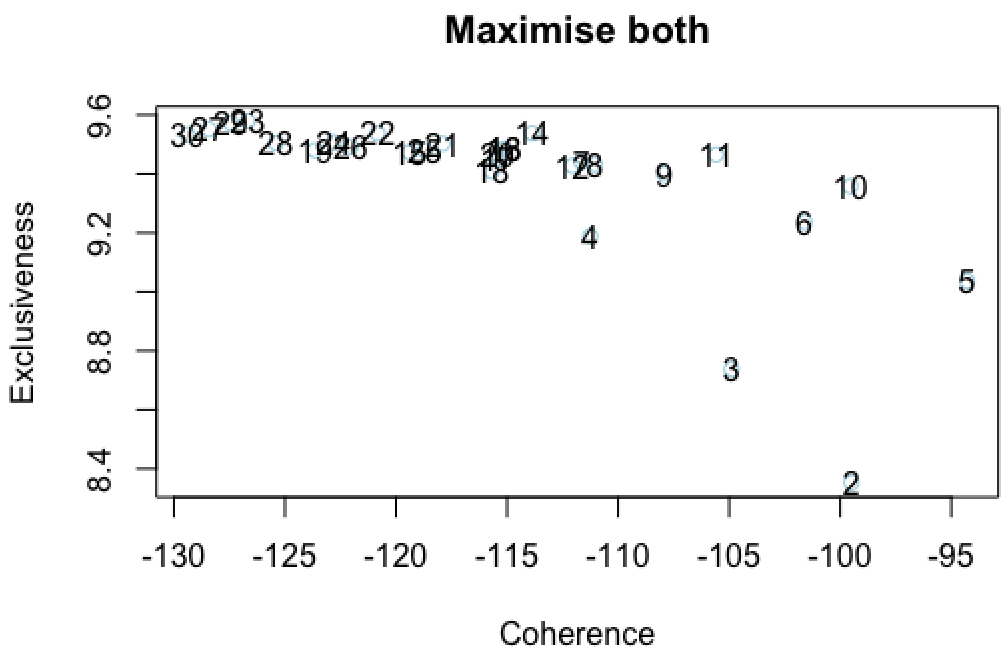
From this graph we can see that there is abundant variation in the sentence score within comment. This suggests that users include both positive and negative opinions in their review. The average score (blue line) is however well above 0, confirming what we have found in earlier analyses.

As a last approach in the textual analysis, we investigate the data from a topical perspective. That is, the adoption of the function stm.search allows to draw a topic distribution with the Structural Topic Model (STM). It is a mixture model, combining three different models, where the researcher must set *k* topics where each document can belong to (Roberts et al., 2014). The possible number of topics must be chosen, and the code is structure in such a way that this number must lie between 2 and 30. In fact, it is impossible that all reviews regard one unique argument. Below we can run a number of diagnostic tests to assist in the choice of *k.* The held-out likelihood and semantic coherence must be maximized; the residual dispersion must be minimized (Roberts et al., 2014).

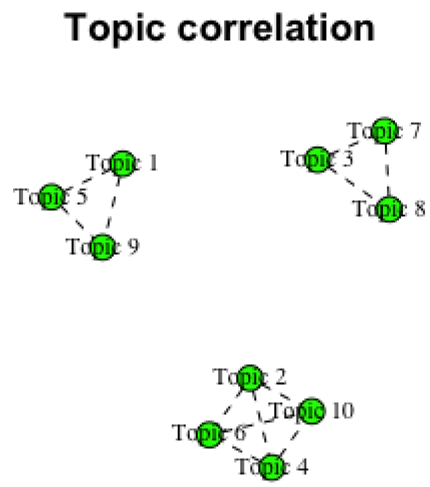


The held-out likelihood suggests that any number between 5 and 15 is maximized; from the semantic coeherence, 5 or 10. From the residuals, the higher the number of topics, the better. Therefore, to conclude, we could choose 5 or 10.

Another approach is to look at both coherence and held-out likelihood. The objective is to maximize the exclusiveness (y-axis) and coherence (x-axis) at the same time. This trade-off can be solved by looking at the most top-right number in the graph below.



Here we can see that the maximized number of topic is 10, which corresponds to a coherence value of -100 and exclusiveness of around 9.4. Another interesting aspect to note is that all small numbers of topics (say, below 6), show a relatively high value for coherence and low exclusiveness. This is due to the fact that each topic is coherent with all other topics because they have many word-keys in common, while on the other hand have very low exclusiveness because of their similarity. At the other extreme, we have the high exclusiveness-low coherence numbers, such as those above 23. These combinations have a high exclusiveness as they are very different from each other, while this exclusiveness is not very important because the topic in this setting is quite the same (work-from-home). Of course, the sub-topics are different from each other (e.g., price, customer service, user friendliness, etc.). That is why "10" could be considered as the maximized value. Each topic includes words that refer to a specific concept. In general, the topics can be grouped as follows: "platform usage", "problem issue", "problem resolution", "support quality", "website interface", "project", "service speed", "payment", "account", "experience". The relationship among topics is depicted below.



Interestingly, even though the topics are divided into 10, the macro topics are only three: topics 1, 5 and 9 belong to the "platform experience" macro group; the second big group is "remuneration and satisfaction", consisting of topics 3, 7 and 8; finally, topics 2, 4, 6 and 10 go together in the category "service support". All in all, this exercise suggests that it is important to split reviews into different topics to analyse comments from a macro perspective.

5. Regression analysis

In this last section before the conclusion, it is presented the relationships among the variables in the dataset in a more rigorous way. The aim of the regression analysis – estimated with OLS - is to investigate whether and how much a variable is determined by another one.

First, it is interesting to study how the four dictionaries (Syuzhet, Bing, Afinn and Nrc), correlates the variable *Stars*. After running the function ***cor*** in R they gave as output the following results: 0.27 for the correlation between Stars and Syuzhet; 0.36 for Stars and Bing; 0.44 for the dictionary Afinn; and 0.075 for Nrc. We can note that they are all positively correlated with Stars, so this suggests that the users' judgement (expressed in number of stars), about the experience with a determined firm, is in line with the assessment provided by the various dictionaries. Hence, calculating the score for positive and negative words, these four dictionaries left a total score like those ones left by users through stars. This is also a method to test the effective reliability of the used dictionaries. However, the numbers listed above also indicate the correlation intensity among the variable stars and all the dictionaries. Afinn is the one with the highest intensity compare to the other dictionaries. On the contrary, Nrc is the one with the lowest value and so the less correlated with Stars.

Going ahead, another interesting insight is represented by discovering the correlation between Stars and "Comment Length". In fact, it is -0.489. We can observe that the sign is negative and the number (in absolute value) is higher than that one representing the correlation between Stars and the dictionary Afinn (who is the highest among all the four dictionaries). So, it is relatively highly correlated, even if in negative terms.

As a prove of what has been just discovered, I cross the data running the string of code that give as result the correlation between the various dictionaries and the variable "Comment Length". It is shown that the dictionaries Syuzhet and Bing left respectively a value of 0.189 and 0.0108. So, they are positively correlated even if the intensity is not so high. While, for what regards Afinn, it indicates a negative value (being in line with what discovered before: a high correlation with the number of stars). It makes sense because if Afinn and Stars are high positively correlated and, at the same time, Stars and Comment Length are high

negatively correlated, so also Afinn and Comment Length must be high negatively correlated. In this case the intensity is not so high because the indicated value is -0.0468. In any case, it maintains the negative sign, hence the prove has been respected.
Same is true for the correlation between the dictionary Nrc and the variable Comment Length. Since this dictionary was slightly correlated with the number of stars (even if always in a positive way), it is logic to find a high and positive value when it is in relation with the variable indicating the number of characters. Indeed, it is 0.3408297, as could be predicted.

As anticipated at the beginning of this section, I run a OLS regression analysis. The baseline model takes the following form

$$NumberHelpfulVotes = ReviewNumber + Stars + CommentLength + Dictionaries + YearFE + MonthFE + DayFE + Cluster\_dtmFE + Cluster\_tfidfFE + FirmFE + \varepsilon$$

where the dependent variable Y is represented by the number of helpful votes each review receives ex post. The multivariate analysis includes several X, such as: ReviewNumber, Stars, CommentLength, and the four Dictionaries (i.e., Syuzhet, Bing, Afinn and Nrc). A detailed description of the variables is provided in section 3.2.1. Date, Cluster and Firm fixed effects are needed to control for unobservable variations. The analysis is performed at review-level, and ε represents the error term in the model.

In the four tables presented below, the dependent variable is the same: Number of helpful votes. It represents the output performed by users. It can be defined as the appreciation to a comment written by a person demonstrated by other users. So, it happens after the review has been written and posted. On the other hand, the variable Stars is determined by the writer at the same time in which he/she is writing the comment. It is decided before the action of posting the comment on the platform. Hence, it can be also used as an input because it can influence other users to declare that review useful or not. Of course, most of the motivation should depend by what it is written in the comment text, but also the number of stars can have an impact in theory.

There are 2319 observations in the baseline model. Below the estimated coefficients, the values in parenthesis represent the t-value. Together with the number of asterisks, it is useful to check if the coefficient value is statistically significant. It is so if the t-value is (in absolute terms) above 1.64 with one asterisk; if it is above 1.96 when two asterisks are present; and 2.56 with three asterisks.

5.1 Results

The first table below (**Table 1**) shows the effects of all the X variables listed on the first column on the left on the Number of helpful votes as Y. The different columns do not include always all the X variables but just those rows occupied by values. For convenience, I inserted numbers from one to seven in the first row. They are useful to differentiate the model versions.

Table 1

Y=HelpfulVotes	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ReviewNumber	2.025*** (95.803)	2.022*** (95.457)	2.023*** (95.709)	2.022*** (95.562)	2.025*** (95.426)	2.024*** (95.377)	2.034*** (95.423)
Stars	-1.073*** (-13.263)	-1.026*** (-11.851)	-1.020*** (-11.835)	-1.005*** (-11.432)	-1.075*** (-12.727)	-1.011*** (-11.446)	- 0.994*** (-11.215)
CommentLength	0.001*** (4.853)	0.001*** (5.095)	0.001*** (5.074)	0.001*** (5.107)	0.001*** (4.441)	0.001*** (4.299)	0.0003 (0.620)
SentimentSyuzhet		-0.068 (-1.553)				-0.048 (-0.536)	-0.070 (-0.782)
SentimentBing			-0.062 (-1.785)			-0.026 (-0.439)	0.009 (0.157)
SentimentAfinn				-0.031* (-1.975)		-0.025 (-0.911)	-0.001 (-0.038)
SentimentNrc					0.002 (0.078)	0.075 (1.486)	
Anger							0.655*** (4.147)
Anticipation							-0.172* (-2.000)
Disgust							-0.458* (-2.450)
Fear							-0.006 (-0.039)
Joy							-0.123 (-1.059)
Sadness							0.524** (3.162)
Surprise							0.155 (1.110)
Trust							-0.071

							(-0.896)
Negative							-0.254* (-1.976)
Positive							0.214** (3.084)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2319	2319	2319	2319	2319	2319	2319
Multiple R-squared	0.8334	0.8336	0.8336	0.8337	0.8334	0.8338	0.8369
Adjusted R-squared	0.8296	0.8297	0.8297	0.8298	0.8295	0.8297	0.8322
The independent variable (Y) is Number of Helpful Votes. Control Variables are: Firm (i.e., Freelancer, FlexJobs, Fiverr, InboxDollars, and Upwork); Year, Month, Day (indicated as Time); Cluster_dtm, and Cluster_tfidf (indicated as Cluster). The numbers between brackets represent t-value. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.							

It can be easily seen that in the first model version (Column 1) the independent variable included are only Review Number, Stars and Comment Length. They all are statistically significant.
In particular, the Review Number coefficient is 2.025 with a t-value equal to 95.803 and three asterisks. It can be interpreted as follows: each time the number of reviews posted increases by one, the Number of helpful votes increases by 2.025. This means that people consider more useful comments written by experienced users, who write comments often or at least not occasionally.

For what regards Stars, instead, the coefficient is -1.073 with a t-value of -13.263 and three asterisks, so it is statistically significant. Hence, each increase of one star makes the Number of helpful votes decreases by slightly more than one. This signifies that users think that comments with less stars (so those describing the negative aspects of a firm) are more useful than those with a high number of stars. Understandably, people who read the comments believe that reviews in which firms are described as perfect, do not reveal some alert to consider when the reader will experience a service for that firm. Those alerts can also make him/her decide to change firm or to have not an experience in general with the sector or industry. Moreover, in this case, the number of comments with a low level of rating stars is much lower than those with many stars, so also the rarity push people to consider that comment more useful than others.

Lastly, in this first model version, the Comment Length coefficient is 0.001 with a t-value equal to 4.853 and three asterisks. So, even if it is statistically significant, the impact of a character increase on the number of helpful votes is economically small. This makes sense because a character (neither a single word) inclusion or exclusion makes no difference in a review. At least a sentence can change the consideration about if a comment is useful or not, but we know that a sentence is composed by many characters.

In the second specification, instead, it is included also the Sentiment Syuzhet, leaving also the other already seen three variables in the model. We can observe that this sentiment is negatively correlated and it is not statistically significant, because its coefficient is -0.068 and its t-value is -1.553 without any asterisk. It means that each increase of one rating point makes the helpful votes decrease by 0.068. It is not a large decrease, but it is useful to understand that (as in the case of Stars) the more comments are positive the less people retain them useful.

The other three independent variables, instead, decreased (in absolute terms) their coefficient, except for Comment Length who did not change. Review Number moved from 2.025 to 2.022, while Stars from -1.073 to -1.026, so their intensity decreased but the sign remained the same. Also their t-value decreased and they maintain their characteristic to be statistically significant, as in the case of Comment Length, who slightly increased its t-value. This means that, in this case Stars and Review Number have a lower impact than before on the Number of helpful votes, respectively in negative and positive terms.

In the third model version, the "new" variable Sentiment Bing behaves similarly to the past one (i.e., Sentiment Syuzhet). They have a similar coefficient (-0.068 vs. -0.062) with a similar t-value (-1.553 vs -1.785) and only the latter is statistically significant. So, the comment here is similar to the one left for Sentiment Syuzhet: they are negatively correlated with the Number of helpful votes because (as also the variable Stars behaves) people find more interesting the comments in which writer criticizes a certain firm. Listing negative aspects is more useful than emphasize the positive ones.

The other three variable included, rather, remain very similar respect to the other two model version already seen. Consequently, also their behaviour and the reasons behind it didn't change.

In the fourth model version, Sentiment Afinn coefficient is equal to -0.031 and it is statistically significant since its t-value is -1.975. Its relation with the dependent variable is negative, even if its value is not so high (in absolute terms). So, increasing by one the Sentiment Afinn, the Number of helpful votes decreases by 0.031. Even here the same consideration of the other two already observed sentiments holds. Also for the other three variables included in this model version, the similarity with the past one is respected.

In the case of the fifth model version, the introduction of the Sentiment Nrc gives a positive coefficient of correlation (unique case among all the four dictionaries), in fact it is equal to 0.002. It is very low so its impact on the studied Y is not so relevant, but again its particularity is represented by its positivity. It means that increasing the Sentiment Nrc by one, the Number of helpful votes increases by 0.002. So, only in this case, people consider more interesting comments with a higher sentiment. However, it is not statistically significant because of its t-value (i.e., 0.078). The other three variable present in this version register a slight increase because Review Number coefficient became equal to 2.025 and the Stars one now is -1.075 (the highest showed till now). Comment Length, instead, is remained always equal during this five model version: its coefficient is still 0.001.

In the sixth specification, all the four sentiment are included at the same time. Changes regards precisely those four variable (i.e., Sentiment Syuzhet, Sentiment Bing, Sentiment Afinn and Sentiment Nrc), while the three always present (i.e., Review Number, Stars and Comment Length) maintained a similar value to the past. Sentiment Syuzhet, Sentiment Bing and Sentiment Afinn decreased their value in absolute terms. Since they were (and they are also now) negative, they decreased their negativity, moving nearer to the zero. So their impact on Y is lower than before. It means that, in this case, their increase by one unit makes the Number of helpful votes decrease less than with the other model version. However, they are not statistically significant because of their low level of t-value. On the contrary, the Sentiment Nrc increased its coefficient, despite it was already the highest among the dictionaries in the past model versions. Now it is 0.075, not so huge but also in absolute terms it overcome the other sentiments' coefficients. Nevertheless, it is not statistically significant, as the others: its t-value is 1.486 and no asterisks are present.

For what regards the seventh and last model version (of this table 1), located on the last column, I substituted the variable Sentiment Nrc with all the sub-emotions that constitute it. They are Anger; Anticipation; Disgust; Fear; Joy; Sadness; Surprise; Trust; Negative; and Positive. I wanted to see more in details how the dictionary is correlated with the Number of helpful votes and if the other dictionaries (or independent variables, in general) would change.

The results are the following: Review Number, Comment Length and Stars (since it is less negative than before) increased their coefficients. Comment Length is the only one (among these three) that is not statistically significant anymore, since its t-value is now 0.620. For what regards the sentiments, instead, Sentiment Syuzhet is the only one that is decreased because now it measures -0.070. While the other two (since Sentiment Nrc is excluded) are increased: Sentiment Bing coefficient is 0.009 (it became positive) and Sentiment Afinn one is -0.001. They are still not statistically significant because their t-values are -0.782; 0.157; and -0.038 respectively for Sentiment Syuzhet; Sentiment Bing and Sentiment Afinn. As an additional confirmation, no one of them has asterisks.

Keeping into account emotions, Anger is the variable with the highest coefficient (it is 0.655) both if we consider it in absolute terms or not. It is one of the few emotions statistically significant. This indication confirms the yet known fact that people appreciate the most reviews where it is exhibited negative impressions about a firm. As a confirm of what just stated, the emotion with the second highest coefficient (0.524) is Sadness: another indicator of negative experience with a company. It is also statistically significant since t-value is 3.162. Other proves are embodied by the emotions Joy and Trust: they are positive emotion but in this case they show a negative coefficient, they are -0.123 and -0.071 respectively. Both of them are not statistically significant because their t-values are below the minimum threshold. On the other hand, all the other emotions listed demonstrate the contrary because "positive" emotions have a positive coefficient while the opposite is true for the negative ones. The coefficient are the following: -0.458 for Disgust (with -2.450 as t-value), -0.006 for Fear (with -0.039 as t-value); 0.155 for Surprise (with 1.110 as t-value); -0.254 for Negative (with -1.976 as t-value); and 0.214 for Positive (with 3.084 as t-value). In fact, Disgust (with one asterisk), Negative (with one asterisk) and Positive (with two asterisks) are the statistically significant ones.

A reason behind their lack of following the above mentioned theory, is that people prefer to listen to about some "negative" arguments and not all of them. Maybe people are scared or simply not interested to certain "negative" aspect writer were referred about a particular firm.

The second table (**Table 2**) is shown below and it exhibits the relationship between the same Y and X variables, yet for one firm at a time. This exercise is interesting because it evaluates whether the baseline results are driven by any specific firms or not. As before, it has the Number of helpful votes as Y variable, while the X variable are Review Number; Stars; Comment Length; Sentiment Syuzhet; Sentiment Bing; Sentiment Afinn; and Sentiment Nrc. In this case, the number of observations varies because in each column (and so in each model version) are inserted the data just for that firm. In the specific, inside the column one there are only comments released for Upwork (they are 440); in the second column only those for Freelancer (i.e., 1288); in the third only those for FlexJobs (435 observations); the fourth is dedicated at InboxDollars (they are only 73); and the last one to Fiverr (who contains 83 comments). Hence for each column I excluded reviews that were not left on that firm webpage.

Table 2

Y=Number of Helpful votes; X=all variables in the first column. All firms are included, even if one at the time.

	(1) Upwork	(2)Freelancer	(3) FlexJobs	(4) InboxDollars	(5)Fiverr
ReviewNumber	2.132*** (94.483)	1.796*** (90.299)	3.281*** (15.254)	1.768*** (4.478)	3.545 *** (5.579)
Stars	0.049 (0.321)	-0.484*** (-7.553)	-0.89*** (-4.415)	-0.761 (-1.033)	-1.167 (-1.266)
CommentLength	0.002*** (5.811)	0.003*** (11.845)	0.0006 (0.905)	-0.002 (-1.265)	-0.003 -1.462
SentimentSyuzhet	0.007 (0.053)	0.064 (1.105)	0.395 (1.905)	-0.248 (-0.391)	0.031 (0.043)
SentimentBing	0.010 (0.106)	0.028 (0.776)	0.130 (0.815)	-0.040 (-0.079)	-0.339 (-0.527)

SentimentAfinn	-0.025 (-0.589)	0.011 (0.695)	0.006 (0.084)	-0.093 (-0.423)	0.027 (0.095)
SentimentNrc	0.031 (0.443)	-0.099** (-2.956)	-0.302** (-2.706)	0.385 (1.081)	0.062 (0.130)
Time FE	Yes	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes	Yes
Obs.	440	1288	435	73	83
Multiple R-squared	0.932	0.8479	0.6695	0.7393	0.8124
Adjusted R-squared	0.9276	0.8431	0.6318	0.609	0.5474
The independent variamble (Y) is Number of Helpful Votes. Control Variables are: Year, Month, Day (indicated as Time); Cluster_dtm, and Cluster_tfidf (indicated as Cluster). The numbers between brackets represent t-value.					

Starting from the Review Number, as it is possible to see from the table above, the firm Fiverr has the highest coefficient (i.e., 3.545), even if the FlexJobs one (i.e. 3.281) is slightly below that amount. The other three firms present a coefficient equal to 2.132 for Upwork; 1.796 for Freelancer; and 1.768 for InboxDollars. All the five companies a t-value greater than the maximum threshold, so they are statistically significant. In any case, the coefficient values mean that increasing by one the review number, the number of helpful votes increases (because they all are positive values) by the coefficient amount. So, within the Fiverr comments, a change in the review number affect in a more sensible way the Number of helpful votes, respect to reviews regarding other firms. It has a higher impact. An effect that can be translated in a more consideration from users about the level of review number left by each writer. InboxDollars' users consider that variable less important in order to categorise a comment as useful.

The variable stars, instead, is interesting because the firm Upwork has a positive coefficient (i.e., 0.049), while all the other firms have a negative one. Fiverr has the most negative coefficient (equal to -1.167). Nevertheless, Freelancer (-0.484) and FlexJobs (-0.89) are the only two companies with a statistically significant value, in fact their t-value are -7.553 and -4.415 respectively, and they both hold three asterisks. So, only for Upwork, increasing the number of stars, it increases also the Number of helpful votes: they are directly proportional. Even if, in absolute value, Upwork has the lowest influence on the dependent variable. Instead, for all the other four firms, if the rating stars increases, the Y decreases. Fiverr has the highest impact in negative terms.

For what regards Comment Length, in general, all the firms have a really small coefficient level. They all are very near to zero. So, their impact on the Number of helpful votes is almost null. The only two companies with a statistically significant value are Upwork (coefficient equal to 0.002 and its t-value is 5.811) and Freelancer (the coefficient is 0.003 and its t-value correspond to 11.845).

Among the four sentiments the value to be underline because more relevant are the FlexJobs (0.395 of coefficient) and InboxDollars (-0.248) for Syuzhet; Fiverr (coefficient equal to -0.339) for Bing; none for Afinn because too small in absolute terms; and Flexjobs (-0.302) and InboxDollars (0.385 as coefficient) for Nrc. FlexJobs for Syuzhet has the highest value among all the dictionaries both in absolute value and not. This means that the more people write "positive" words (or at least those who Syuzhet consider to be positive) the more other users consider that comment useful. The same is true with all the other positive value for all the sentiment, but in this case the impact is higher. On the contrary, the opposite logic is applied for all the negative coefficients.

Then is due to emphasize that only two value are statistically significant and they are all "belonging" to the sentiment Nrc: the already seen value of FlexJobs (t-value of -2.706) and the Freelancer (-2.956 as t-value) value.

5.2 Spill-over effects

Next, we evaluate the spillovers effects across firms. This is done by calculating the average sentiment each firm except one receives in a given date, and study its effect (if any) on the firm reviews themselves. Table 3 shows the results.

Table 3

	(1) meanS	(2) meanQ	(3) meanT	(4) meanU	(5) all mean (S; Q; T; U)
ReviewNumber	2.027*** (85.317)	2.029*** (85.960)	2.031*** (86.030)	2.028*** (85.544)	2.032*** (86.151)
Stars	-1.047*** (-9.330)	-1.004*** (-8.978)	-1.014*** (-9.080)	-1.032*** (-9.207)	-0.998*** (-8.936)
CommentLength	0.001** (2.975)	0.001** (3.031)	0.001** (2.932)	0.001** (2.968)	0.001** (2.983)
SentimentSyuzhet	-0.023 (-0.200)	-0.016 (-0.145)	-0.006 (-0.060)	-0.014 (-0.122)	-0.013 (-0.119)
SentimentBing	-0.038 (-0.487)	-0.033 (-0.424)	-0.027 (-0.356)	-0.036 (-0.464)	-0.026 (-0.341)
SentimentAfinn	-0.029 (-0.834)	-0.024 (-0.683)	-0.028 (-0.800)	-0.028 (-0.815)	-0.023 (-0.678)
SentimentNrc	0.090 (1.418)	0.074 (1.182)	0.071 (1.137)	0.082 (1.301)	0.066 (1.059)

meanS	0.090* (1.699)				0.028 (0.347)
meanB		0.112*** (4.899)			0.108* (2.418)
meanA			0.269*** (4.977)		0.302** (2.763)
meanN				0.185** (3.053)	-0.351* (-2.343)
Firm FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes	Yes
Obs.	2319	2319	2319	2319	2319
Multiple R-squared	0.8442	0.8461	0.8462	0.8448	0.8471
Adjusted R-squared	0.8391	0.841	0.8411	0.8397	0.8417
The independent variamble (Y) is Number of Helpful Votes. Control Variables are: Firm (i.e., Freelancer, FlexJobs, Fiverr, InboxDollars, and Upwork); Year, Month, Day (indicated as Time); Cluster_dtm, and Cluster_tfidf (indicated as Cluster). The numbers between brackets represent t-value.					

From table 3, the model includes the mean value of each type of sentiment of other firms as X variables. In the first column, we can see that other firms' SentimentSyuzhet is weakly influencing the number of helpful votes a comment for a firm receives. This is because the coefficient is positive and statistically significant. In column 2, 3 and 4, I do the same exercise, while changing the type of Sentiment spillovers mean value. In these cases, we can see that the effects is stronger, both in statistical and in economic terms. This strengthens the idea that the helpfulness of reviews is influenced by the average sentiment in the industry, and that managers of firms should monitor that for their decisions. Finally, in column 5, we attach all mean sentiments to the model to assess which spillover types is most affecting Number of helpful votes. It is interesting to notice that other firms' SentimentBing and SentimentAfinn are positively influencing the reviews impact of a firm, while SentimentNrc is negatively doing so. Moreover, spillover effects appear even more important than the effect of the sentiment of the review itself.

5.3 Pandemic effects

Lastly, column 4 explores how the relationships between the variables in the baseline model changed in 2020 with respect to 2019. This is interesting because in 2020 Covid-19 hit this industry. Therefore, the hypothesis is that reviews for this type of services are affected by general restrictions government have taken. It can be either way: on the one hand, reviewers can be more affected by other firms' sentiment because more people join the platform. On the other hand, the effects can be smoothed as managers can anticipate these developments and start to reward reviews of their clients with some reward program.

Table 4 takes the baseline model and add interaction terms between each independent variable X and a dummy variable 2020. The latter takes value 1 if the review was left on the platform at some point in 2020, and zero if the comment is in 2019. In the first column, we can see that the marginal effect of the experience proxy (ReviewNumber) decreases in 2020 with respect to 2019, as indicated by the negative and statistically significant coefficient -1.152. This could be explained by the fact that with Covid-19 restrictions more people start joining the platform and can say what they think about a firm. Therefore, the effect of experience is not anymore that important.

The same can be said for *Stars*. The interaction term takes the opposite value with respect to the standalone one, indicating that there is a mitigation effect in 2020. The results for CommentLength are mixed and weak, suggesting that the length of the comment does not influence the importance of a review.

Finally, for what regards the sentiment values, we can state that 2020 has not changed the intensity or direction of the effect a review sentiment has on the number of helpful votes. This is generally true for all types of sentiments except for SentimentSyuzhet which had a stronger effect in 2020 with respect to 2019. Finally, the spillover effects does not appear to influence the Number of helpful votes differently in 2020 with respect to 2019. Exceptions are SentimentBing and SentimentAfinn, which are reinforced in 2020. This is in line with the idea that people start to research more before making a choice in a pandemic. More informed decisions are made because of more time available and a shift in the type of jobs carried out.

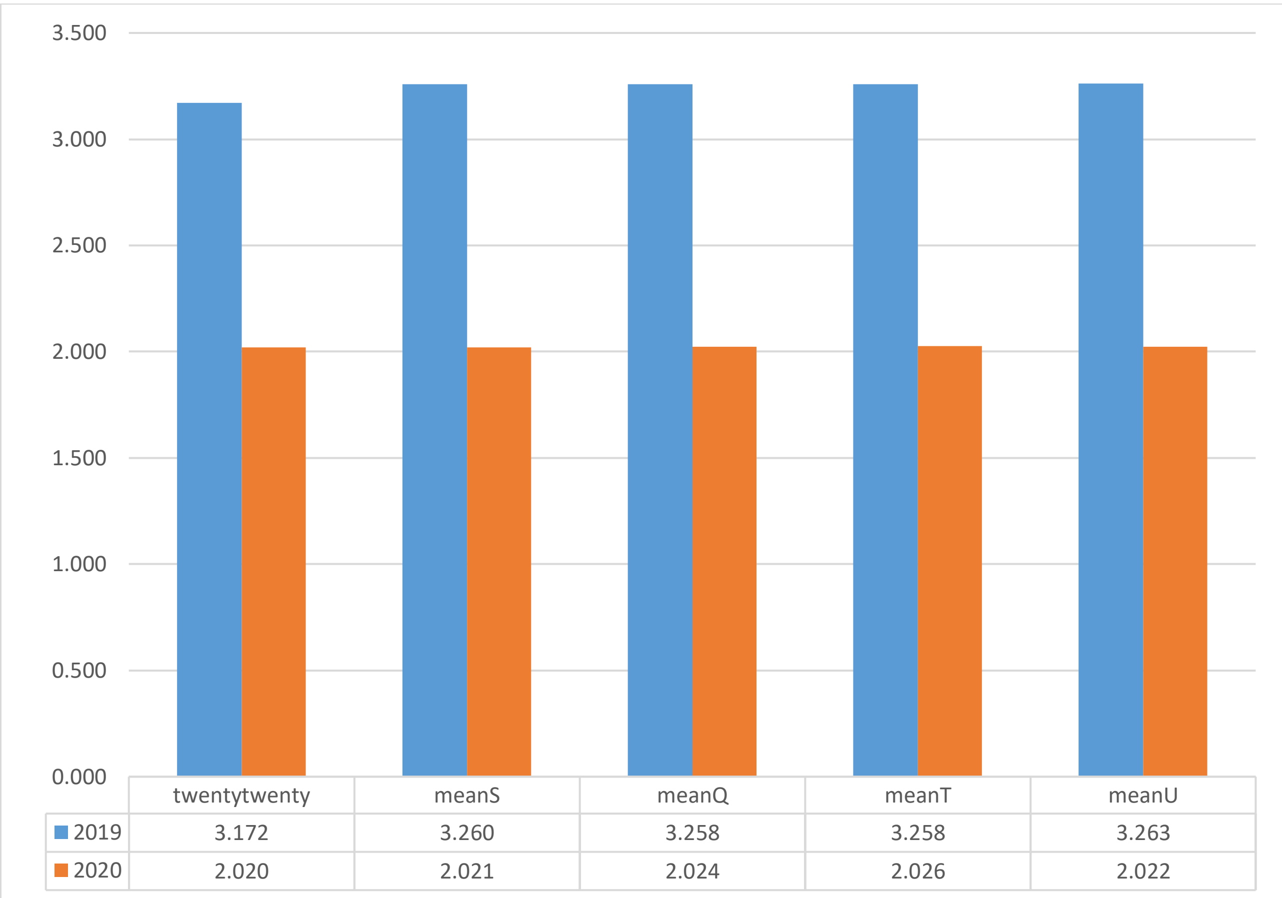
Table 4: Covid-19 effects.

	(1) **twentytwenty**	(2) meanS* **twentytwenty**	(3) meanQ* **twentytwenty**	(4) meanT* **twentytwenty**	(5) meanU* **twentytwenty**
ReviewNumber	3.172*** (14.210)	3.260*** (12.764)	3.258*** (12.817)	3.258*** (12.850)	3.263*** (12.796)
ReviewNumber X 2020	-1.152*** (-5.139)	-1.239*** (-4.833)	-1.234*** (-4.838)	-1.232*** (-4.838)	-1.241*** (-4.847)
Stars	-1.298*** (-10.054)	-1.443*** (-8.213)	-1.439*** (-8.096)	-1.447*** (-8.265)	-1.450*** (-8.266)
Stars X 2020	0.415** (2.718)	0.531** (2.692)	0.570** (2.844)	0.583** (2.955)	0.563** (2.849)

CommentLength	0.001** (2.696)	0.0006 (1.016)	0.0005 (0.977)	0.0005 (0.953)	0.0005 (0.973)
CommentLength X 2020	-0.0001 (-0.288)	0.0003 (0.432)	0.0004 (0.562)	0.0003 (0.532)	0.0003 (0.529)
SentimentSyuzhet	0.437** (-2.795)	-0.472* (-2.063)	-0.448 (-1.954)	-0.465* (-2.037)	-0.466* (-2.030)
SentimentSyuzhet X 2020	0.619** (3.244)	0.667* (2.521)	0.620* 2.340	0.656* (2.486)	0.658* (2.483)
SentimentBing	-0.026 (-0.256)	-0.185 (-1.126)	-0.192 (-1.169)	-0.189 (-1.154)	-0.189 (-1.146)
SentimentBing X 2020	-0.001 (-0.011)	0.181 (0.970)	0.195 (1.053)	0.203 (1.094)	0.189 (1.013)
SentimentAfinn	0.011 (0.254)	0.048 (0.668)	0.046 (0.640)	0.047 (0.664)	0.047 (0.655)
SentimentAfinn X 2020	-0.063 (-1.110)	-0.109 (-1.327)	-0.094 (-1.152)	-0.103 (-1.263)	-0.106 (-1.289)
SentimentNrc	0.197* (2.142)	0.316* (2.371)	0.308* (2.323)	0.316* (2.383)	0.317* (2.378)
SentimentNrc X 2020	-0.188 (-1.733)	-0.312* (-2.081)	-0.318* (-2.130)	-0.329* (-2.204)	-0.320* (-2.139)
meanS		0.0155 (0.183)			
meanS X 2020		0.110 (1.009)			
meanB			0.023 (0.599)		
meanB X 2020			0.106* (2.189)		
meanA				0.007 (0.087)	
meanA X 2020				0.367*** (3.334)	
mean					0.019 (0.216)
meanN X 2020					0.224 (1.821)

Firm FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Cluster FE	Yes	Yes	Yes	Yes	Yes
Obs.	2319	2319	2319	2319	2319
Multiple R-squared	0.8392	0.8505	0.8521	0.8528	0.851
Adjusted R-squared	0.8347	0.8448	0.8464	0.8472	0.8454
The independent variamble (Y) is Number of Helpful Votes. Control Variables are: Firm (i.e., Freelancer, FlexJobs, Fiverr, InboxDollars, and Upwork); Year, Month, Day (indicated as Time); Cluster_dtm, and Cluster_tfidf (indicated as Cluster). The numbers between brackets represent t-value.					

A way to summarize the data, present in the already shown tables, is creating a graph in order to represent them in a different and easiest way to read them. The below graph has been realized with this precise intent.



Here, there are represented the first two row of the fourth and last table. The rectangular in blue indicates the Review Number value in 2019 while the orange one are the Review Number coefficients in 2020. Their difference reflects the additional effect of twentytwenty; meanS; meanQ; meanT; or meanU on the dependent variable Y (e.g., the Number of Helpful Votes, in this case) in 2020 respect to 2019.
However, it is possible to say that the trend among the various means is quite stable both within the 2019 and within 2020. Rather, on average, the difference between the two years is not so small. 2020 has an always increasing trend, with the exception of the last mean (i.e., meanU). Instead, 2019 shows
an instable variation, because it increases, it decreases and then it increases once again.
Lastly, the values appear also at the bottom of the graph, together with the legend helping to rapidly understand the data.

6. Conclusion

This thesis explores how consumer behaviour evolves in the pandemic exploiting some data from the work-from-home industry. The data were scraped from Sitejabber through the Google Chrome plug-in, Data Miner. There is a large variation in terms of comments usefulness, length and timing across different firms. A multi-variate regression analysis suggests that the number of reviews and the length of the comment a worker leave positively affect the usefulness of her comments for other people. More negative or pessimistic reviews are more useful than positive ones. Feelings such as “Anger”, “Sadness” or “Disgust” are most affecting the number of helpful votes for a review. Finally, there are strong spillovers across firms in terms of general sentiment. When people leave a bad review for a firm, this influences what other people think about the comments of another firm. The Covid-19 pandemic in 2020 seems to have amplified these effects.

For what regards the regression analysis results, instead, it is possible to state, looking at the first table, that the main independent variable who mostly influence the dependent one (i.e., Number of Helpful Votes) is the Review Number, even considering all the independent variables as absolute values (even if Review Number has a positive impact on Y). This is true making an average of values present in the various model versions. In the second position of this particular ranking, there is the variable Stars, who has a negative impact on the Number of Helpful Votes). While, Comment Length is located in the last position with a very low coefficient (e.g., about 0.001 on average). Lastly, the different Sentiments, in general, have a middle level impact on the studied Y.

After that, always from the first table, it is intuitive that among the emotions who constitute the Sentiment_Nrc, Anger is the one who has the highest impact and it is statistically significant, like Sadness who occupies the second position among the others. Here we can note an interesting phenomenon: the first two impactful emotions can be categorized as negative one. So a conclusion can be that people give more importance to comments with a higher level presence of complaints, or, however, negative opinions about a specific firm.

Taking into consideration the second table, instead, it is possible to say that results are robust across firms, they do not depend on the firm taken under scrutiny.

Table 3 give us an insight about the spill over effects: they are present inside these five companies and so managers have to consider them in the decision making process. So, there is the evidence that the comments left about a firm impact people in general and then also other reviews for the other companies. This can be explained by the fact that the same people can write about two or more different firms (since he/she can buy two different products/services produced by different entities). At the same time, people can also influence his/her family members or friends talking badly about his/her experience.

From the fourth and last table, rather, during Covid-19 pandemic, the three main independent variables (i.e., Review Number, Stars and Comment Length) effect on the Number of Helpful Votes behave in different ways. In the case of Review Number and Comment Length the interaction term is positive so they impacted positively the Y, while Stars has a negative sign. However, for all these three variables, the coefficient is higher (considering all numbers as absolute values) in 2019 than in 2020. So, the pandemic situation mitigated their effect on the Number of Helpful Votes.

References

Boegershausen, J., Borah, A., Stephen, A. (2020). Fields of Gold: Web Scraping for Consumer Research. Marketing Science Institute Working Paper Series, Report No. 20-143.

Edelman, B. (2012). Using internet data for economic research. Journal of Economic Perspectives, 26(2), 189-206.

Barnes, C. M., Dang, C. T., Leavitt, K., Guarana, C. L., & Uhlmann, E. L. (2018). Archival data in micro-organizational research: A toolkit for moving to a broader set of topics. Journal of Management, 44(4), 1453-1478.

Rafaeli, A., Ashtar, S., & Altman, D. (2019). Digital traces: New data, resources, and tools for psychological-science research. Current Directions in Psychological Science, 28(6), 560-566.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. Psychological methods, 21(4), 493.

Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. Journal of Experimental Social Psychology, 66, 100-106.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. Proceedings of the National Academy of Sciences, 115(45), 11401-11405.

Wang, Y., & Chaudhry, A. (2018). When and how managers' responses to online reviews affect subsequent reviews. Journal of Marketing Research, 55(2), 163-177.Hyrynsalmi, S., Seppänen, M., Aarikka-Stenroos, L., Suominen, A., Järveläinen, J., & Harkke, V. (2015). Busting myths of electronic word of mouth: the relationship between customer ratings and the sales of mobile applications. Journal of theoretical and applied electronic commerce research, 10(2), 1-18.

Macias, P., & Stelmasiak, D. (2019). Food inflation nowcasting with web scraped data. Narodowy Bank Polski, Education & Publishing Department.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National science review, 1(2), 293-314.

Das, T. K., & Kumar, P. M. (2013). Big data analytics: A framework for unstructured data analysis. International Journal of Engineering Science & Technology, 5(1), 153.

Chatterjee, P. (2001). Online reviews: do consumers use them?.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. Decision support systems, 45(4), 1007-1016.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. International Journal of Hospitality Management, 28(1), 180-182.

Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. Information and communication technologies in tourism 2008, 35-46.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. Marketing science, 29(5), 815-827.

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. Marketing Science, 37(1), 5-21.

Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach. International Journal of Electronic Commerce, 20(2), 236-260.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. Management science, 57(8), 1485-1509.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems, 81, 30-40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural topic models for open-ended survey responses. American Journal of Political Science, 58(4), 1064-1082.

Summary

Despite the health pandemic, types of jobs that allow working from home increase every year and the numbers are supposed to be even higher in the future due to technology. That is why investigating developments in the work-from-home industry is important. Firms can take advantage of customer opinions through their platform reviews. As a matter of fact, consumer behaviour represents a fundamental mean for managers to understand the actions they should take to be successful in the short, medium, and long term. To accomplish this purpose, it is necessary to comprehend how consumers think and feel about a specific product, or a how the market perception of a product category in general change over time. There are different ways to understand what characteristics people search in a product or service.

This thesis research has several objectives: first it tries to identify the strong and weak points of a service, the reviews of their customers over time and across groups, and comparing the characteristics of similar companies; second, the thesis aims at investigating spillover effects in customer reviews. In theory, it could be that a period of bad reviews for a company may push clients to switch to another service offered by a competitor. If this is the case, this evidence would strengthen the idea that consumers do not only consider the reviews of one specific company before making a purchase, but they look at different services in the market and are influenced by the general sentiment. In other words, the review for a company's product affects the volume and sentiment of a product offered by a competitor firm. Of course, the necessary conditions are that these companies belong to the same sub-industry and that they offer a service that is more homogenous as possible. For identification purposes, in this way, we can isolate spillover effects of the reviews from any other inherent differences in the companies (e.g., service offered). Therefore, I decided to focus on the work-from-home freelance industry.

People spend a huge portion of their daily life in contact with disparate technological devices. They pass their time connected to an online platform, or at least with the possibility to be contacted by others through a web-based services. Because of these connections, the amount of data produced is so huge that data analysts must be careful to screen and select the right information to analyse and do not loose time with poor quality data or fake news.

Every action an individual makes on the internet usually provide useful information about their priorities, thoughts, beliefs, and their commercial inclinations. Just by surfing on the internet or sending messages, clients are leaving important details that companies can use to study the consumers' behaviour and any related theories.

Not all the researchers use the web scraping techniques to carry out their research. Unfortunately, it could represent a missed chance to increase their research quality and the output reliability. Basing a research on data can produce more accurate results and, consequently, it can improve the trustworthiness people release on that analysis. The reason behind this lack of data utilization can derive by an absence of knowledge about which method is the best to follow and the step to reproduce to accomplish this task. Nevertheless, web

scraping can be a functional way to discover crucial insights about consumers' behaviour (Boegershausen et al. 2020).

In this sense, text mining is a technique used to extrapolate contents from textual data, like consumer reviews, comments, or general documents. It can identify, categorize, and decipher words as well as their context. Text mining is needed to automatically understand a large number of words and texts. Text mining includes the creation of a corpus (i.e., it basically consists of the union of all the words in the data set); the corpus cleaning (e.g., removing punctuation, stop-words and white space; transforming all the words in lower case letter; and stemming words); create a document term matrix and inspecting it (analysing the associate words, most frequent terms, tag clouds, and the sentiment score); and, last but not least, the corpus division in a series of topics. In particular, the sentiment analysis can be performed either at sentence level or at document level (or both). It will produce the polarity score (positive, negative or neutral) and its intensity (its minimum and maximum vary depending by the tools, method, lexicon and dictionary used).

Today, online users are even more important than in the past. They generate contents (UGC) useful for future consumers who will benefit from their comments about the proved service or a purchase. On the web, negative reviews can influence other users to express their opinions and to decide if it is a good idea to buy or not a product or service (Chatterjee 2001). In addition, online reviews can determine how much movies' daily box office performs well or not on the market, in terms of sales and revenues generated (Duan et al. 2008). Similarly, marketers studied the effect of online reviews on the amount of hotel room bookings, finding a notable influence of them on the final output (Ye et al. 2009).

However, this thesis explores how consumer behaviour evolves in the pandemic exploiting some data from the work-from-home industry. The data were scraped from Sitejabber through the Google Chrome plug-in, Data Miner. There is a large variation in terms of comments usefulness, length and timing across different firms. A multi-variate regression analysis suggests that the number of reviews and the length of the comment a worker leave positively affect the usefulness of her comments for other people. More negative or pessimistic reviews are more useful than positive ones. Feelings such as "Anger", "Sadness" or "Disgust" are most affecting the number of helpful votes for a review. Finally, there are strong spillovers across firms in terms of general sentiment. When people leave a bad review for a firm, this influences what other people think about the comments of another firm. The Covid-19 pandemic in 2020 seems to have amplified these effects.

For what regards the regression analysis results, instead, it is possible to state, looking at the first table, that the main independent variable who mostly influence the dependent one (i.e., Number of Helpful Votes) is the Review Number, even considering all the independent variables as absolute values (even if Review Number has a positive impact on Y). This is true making an average of values present in the various model versions. In the second position of this particular ranking, there is the variable Stars, who has a negative impact on the

Number of Helpful Votes). While, Comment Length is located in the last position with a very low coefficient (e.g., about 0.001 on average). Lastly, the different Sentiments, in general, have a middle level impact on the studied Y.

After that, always from the first table, it is intuitive that among the emotions who constitute the Sentiment_Nrc, Anger is the one who has the highest impact and it is statistically significant, like Sadness who occupies the second position among the others. Here we can note an interesting phenomenon: the first two impactful emotions can be categorized as negative one. So a conclusion can be that people give more importance to comments with a higher level presence of complaints, or, however, negative opinions about a specific firm.

Taking into consideration the second table, instead, it is possible to say that results are robust across firms, they do not depend on the firm taken under scrutiny.

Table 3 give us an insight about the spill over effects: they are present inside these five companies and so managers have to consider them in the decision making process. So, there is the evidence that the comments left about a firm impact people in general and then also other reviews for the other companies. This can be explained by the fact that the same people can write about two or more different firms (since he/she can buy two different products/services produced by different entities). At the same time, people can also influence his/her family members or friends talking badly about his/her experience.

From the fourth and last table, rather, during Covid-19 pandemic, the three main independent variables (i.e., Review Number, Stars and Comment Length) effect on the Number of Helpful Votes behave in different ways. In the case of Review Number and Comment Length the interaction term is positive so they impacted positively the Y, while Stars has a negative sign. However, for all these three variables, the coefficient is higher (considering all numbers as absolute values) in 2019 than in 2020. So, the pandemic situation mitigated their effect on the Number of Helpful Votes.

Finally, other potential further research can include the integration of online and offline sources to collect the data and see the changes, if any, in the output (Berger et al. 2010; Datta et al. 2018).