SENTIMENT ANALYSIS:
CUSTOMERS PERSPECTIVES ABOUT
UBER vs. LYFT

Presented to:
DR. NOR HASLIZA MD SAAD

Presented by:

Radwa Khaled abdalhameed
radwakhaledabdelhamed@gmail.com

Table of Contents

1.0 Abstract

The first objective of this report is to analyze the main factors influencing customer satisfaction as well as service quality, and price. Without any one of these factors, the customers might not be fully satisfied with the services. The second objective is to examine the relationship between customer satisfaction and word-of-mouth perceptions in both companies, Uber and Lyft.

What we are talking about here is analyzing a piece of text to determine the sentiment behind it. By using some tools in RapidMiner, we did all these steps for both Uber and Lyft data. The first is data collection for both Uber and Lyft, which was collected from Twitter, about 200 raw.

second, cleaning by a few steps: Selecting attributes to handle missing values and unnecessary values such as the username or the ID, etc. and removing the white space.

Removing Duplicates using the "Remove Duplicate" Operator

Removing the Missing Values using the "Replace Missing Values" Operator

Third, conduct sentiment analysis to group tweets that are collected into either Positive or Negative,

as positive sentiment (compound score > 0), negative sentiment (compound score < 0), neutral sentiment (compound score == 0).

and finally, Frequency Analysis and Word Cloud which Frequency analysis concerned with the number of times a word appears in the document. While the Word Cloud propose visualization of the text representation

1.1 Introduction

Sentiment analysis helps With customer service and how it is important to understand customer feelings, With the data from social media extracted by RapidMiner, we can do a lot more in the future, such as: the company can create a new product; the company can follow the latest trends in the labor market and many more; get more information towards the product or services that they need from anywhere; and sentiment analysis, which can help to know what customer feedback will lead to the negative effect or positive effect of the company.

We have two companies that want to know the sentiment: Uber and Lyft.

Transportation network companies such as Uber and Lyft are very helpful for many people who have difficulty driving a car or using the regular transit system, such as older residents, people with disabilities, and normal people.

Both Uber and Lyft provide many services in different countries, which makes the competition strong.

Uber serves hundreds of cities in dozens of countries, including the U.S. and Canada, as well as other cities around the globe in Central and South America, Africa, Asia, Australia, New Zealand, and the E.U. Depending on your location, Uber may be your only alternative taxi option. In the U.S., the service lists over 300 cities, from Las Vegas, Chicago, and New York to Fargo, Pensacola, and Kalamazoo, It offers several classes of service whose availability varies by city, such as UberX, UberX Share, Uber Comfort, Uber Black, Uber Scooters, Uber WAV and Uber Eats.

Lyft's serves show in the cities where the ride service operates in Canada and the United States, where it serves all 50 states and the District of Columbia. The company offers several service classes, which vary by city, such as Original Lyft, Lyft XL, Lyft Lux, Lyft Black, and Lyft Black XL.

And there are drawbacks, including increased and unpredictable travel time and the social implications of sharing a confined space with strangers. To understand how travelers, drivers, and others understand and communicate about these services, we analyzed online commentary about Uber and Lyft on the popular micro-blogging site Twitter. We analyzed more than 1500 raw tweets collected for each of Uber and Lyft, "tweets" the Twitter is an American microblogging and social networking service connect people and allow to share their thoughts with a big audience about the services, coding for who tweets, the emotional tone of the tweets, and what aspects of the services are tweeted about.

We find that positive tweets are quite rare and are outnumbered by negative tweets.

However, the most common tone is humorous. As expected, there is considerable negative tweeting about service characteristics such as delays in trips. Drivers are insufficiently qualified. Account usability is hard for some people, It is not available for all areas in the same country, it is very expensive, and app usability has many problems. However, most tweets are about drivers and their behavior. Negative tweets about drivers outnumber positive ones.

We work on this report because it makes those companies determine and know what the customer's feelings are and how to improve their service by focusing on using safety as their strength.

1.2 Research Statement

Transportation is one of the important things in life. It helps us to meet our needs, and nowadays we have many transportation applications that make the mobility process easier.So it is important to focus on customer service and make sure of customer satisfaction. According to this report, we selected the two most important platform mobility. Lyft and Uber by using sentiment analysis tools to determine and understand the trends, react to customer opinion, and analyze the data about their service quality and customer satisfaction collected from social apps like Twitter and extract data from websites by using Data ScraperSo this paper is to produce a more comprehensive view of the sentiments of their consumers and provide for their needs in safety, cost, and responsiveness in all good ways.

1.3 Research Objectives

In general, this report uses Twitter and Data Scraper as data sources and RapidMiner as a medium for implementing the opinion mining process to address the following research question for each of the applications:

Q1: Is the perception of the user positive, negative, or neutral?

Q2: Is that the negative element by the user based on the inability to use the application or towards the service or towards the cost?

Q3: Are the drivers and cars qualified enough?

1.4 Research Questions

To achieve the objectives above, this report will attempt to answer the following:

RQ1: What is the sentiment about Uber and Lyft?

RQ2: What is the topic of discussion about Uber and Lyft?

RQ3: which has better service. Uber or Lyft?

RQ4: Is this customer's Uber and Lyft satisfaction or dissatisfaction?

RQ5: How are people responding to this advertisement/product release/news item?

RQ6: How have bloggers' attitudes towards the president changed since the election

1.5 Contribution:

The contribution to this research is as follows:

In this report, we analyze the sentiment of Twitter data using a lexicon-based approach.

- dictionary-based approach is to analyze the feeling of being positive or negative that a text conveys to the public.
- Pre-processing and analysis of the collected data are performed, followed by sentiment analysis.
- The experimental evaluation was conducted using a set of tweets as a data source to prove that it is worthwhile and to implement the opinion mining in other similar applications.

2.0 Literature Review

Business around the globe is realizing the importance of doing data analytics to stay competitive in the challenging market. Applying sentiment analysis has helped companies to monitor their business and stay alert to any issue to arise. analysis the reviews and tone embedded in the text that can be collected from different platforms. And according to the statistics ,46% of people have chosen to use social media to complain about the intended company. (Christensen, Peter, and Adam Osman.,2021). Any business put high dependence on their customers' satisfaction level, so extracting details from the analysis deemed to be valuable for suggestions and solutions to any related issues. Information on whether your service or product being complimented or criticized, brand monitoring helps to categorize the positive or negative mentions of your service online. And let you know the characteristics of your services and business from all over the world.

In the past there was a ride-sharing business, developed a business model that had been functioning in the same way for generations, in city in busy streets a person in need of a ride stood on a street corner and waved down a taxi. On quiet streets, or in towns without roving taxis, the person would phone a local car service and request a pickup. Now there is a business based on this and there are applications for that. Uber and Lyft a famous example, nowadays. (Kamais, C. E.,2019). Uber's E-hail services allow hiring a driver using a smartphone app at any time from almost any location. Software locates drivers near the client and offer a selection of options from the cheapest carpooling choice to luxury wheels. The price is set and paid in advance. (Kamais, C. E.,2019)

The cost of these services is affected by ''surge pricing ''; which mean that the price is affected by the demand, it revises the cost of its rides from hour to hour based on local demand, the more calls are made prices will rise up. Doing the analysis help us know that customer face issue with the on-demand ride services) Lu, A., Frazier, P., & Kislev, O.,2018). as transport costs can affect mobility in a different way across the population, based on a study assign a large price reduction on Uber in Egypt over a 3-month period, collect comprehensive data on participant mobility using Google Timeline. The outcomes are a 50% price reduction quadruples Uber usage and had a 42% increase in total travel. (Christensen, P., & Osman, A., 2021)

Uber and its competitors particularly Lyft, have completely transformed the personal transportation industry. Beside mobility, Ride-sharing services like Uber and Lyft have some characteristics. chasing down the taxi or calling and waiting for a car service, instead of that mobile applications user can ask for a car from any location and have it arrived in munities, in some applications you do not need to enter your location the know it. No cash changes hands as the passenger's credit card is linked to the application. A receipt is sent via email. (Wan, W. N. A. A. B., Mohamad, A. F. M. F., Shahib, N. S., Azmi, A., Kamal, S. B. M., & Abdullah, D. 2016). Provide a professional service as the drivers for Uber and Lyft use their own cars and they tend to keep them clean and well-maintained also drivers are polite and well-spoken. Unprofessional drivers are weeded out because passengers get to rate the driver’s performance. A consistently low rating will force a driver out of Uber or its competitors. (Wan, W. N. A. A. B., Mohamad, A. F. M. F., Shahib, N. S., Azmi, A., Kamal, S. B. M., & Abdullah, D. 2016)

Because passengers can rate the driver's performance, unskilled drivers are eliminated. A driver will be fired from Uber or Lyft if they receive consistently poor ratings.as analysis, emotion detection, and sentiment analysis for data from different platforms especially Twitter help improving the service offered and get a positive experience for ride-sharing customers. And for drivers the transaction is cashless, Thus the driver does not risk unpaid rides or the need to carry cash for change. And the same as the customers can rate their drivers, the drivers also can rate their clients. rude, aggressive, and disruptive passengers can be canceled and reports of unsafe behavior toward drivers can cause the deactivation of an account (Alemi, F., Circella, G., Mokhtarian, P., & Handy, S. ,2019). Sentiment analysis helps us to track other competitors and can assist in exploring how they are perceived in comparison to you. As for any industry, price competition may be destructive, Increasingly, Uber, Lyft they, engaged in a battle to provide the cheapest service. And with cheap piece and available cars customers used to take cars even for short distance. (Christensen, P., & Osman, A., 2021). Sentiment analysis will help us to clearer understanding of customer feedback, track it to understand the actions needed to prevent in the future and try to deliver a positive customer experience.

3.0 Methodology

3.1 Data Structure

Years ago, exactly before RapidMiner, people needed to take the advice of others or to read the ratios or reports of some specialists to know how the experience of using some products or services, But Now as a benefit of The power of the Internet, People can share their thoughts, Feelings or reacts about anything. So We can get Peoples' opinions and Experiences about any Product or Service easily using Some applications like RapidMiner to extract peoples' opinions and reactions to know whether they Positively, Negatively, or Neutrally use a specific product or service. This process of extracting peoples' opinions is called Sentiment analysis.

But because of the huge data that results from the Sentiment analysis Process, we need to go through 5 main Processes, including Data collection, Data Pre-processing, Sentiment Classification, Text Pre-processing, and Data Visualization which can be shown the Briefly in the Following Figure:

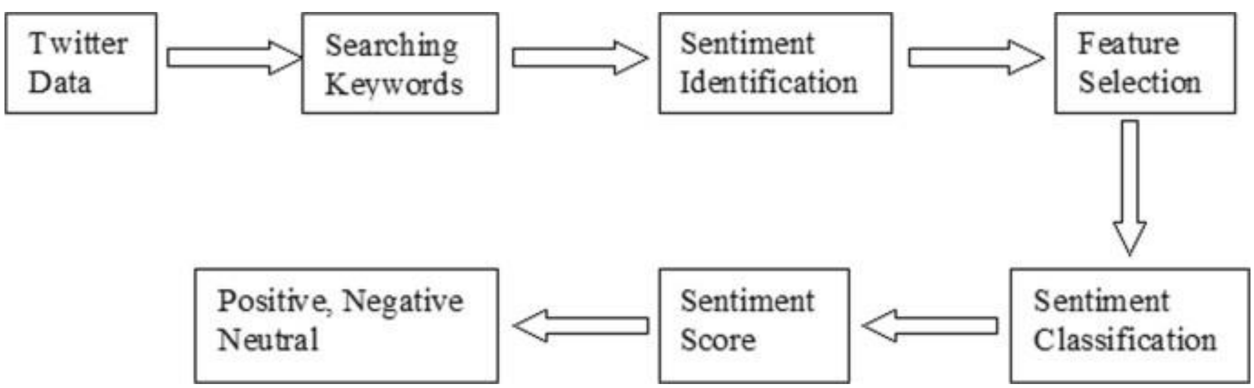


Figure 1 Flow chart of the Process of Sentiment Analysis (Chang, 2021)

3.2 Sentiment Approaches in RapidMiner

RapidMiner is a programming-free data analysis platform, which allows the user to design data analysis processes in a plug-and-play fashion by wiring operators (Ristoski et al., 2015). So, it is used for designing Processes and Visualize the results. RapidMiner is an open source, data science software, which allows users to perform data analysis tasks. It aims to make the process easier by offering an intuitive GUI interface (Rapidminer Inc., 2016). RapidMiner is a tool for conducting data mining workflows for various tasks, ranging from different areas of data mining applications to different parameter optimization schemes (Jovanovic et al., 2014). So we will use it our tool for this research.

3.3 Research Design

Firstly, We extracted the data set for Uber and Lyft brands, the date of extraction is July 13th 2022 but the date of data used is only from 15th July to 16th July for 2022 For Uber. And from 12th July to 15th July,2022.

3.3.1 Data Collection

Through this research, we collected data from Twitter; 500 raw tweets were collected about both Uber and Lyft to get the highest and most accurate data about the two brands so that each opinion which is represented in a tweet would be analyzed so we can get the best results. Thedata collection Process from RapidMiner is shown in the following Figures.

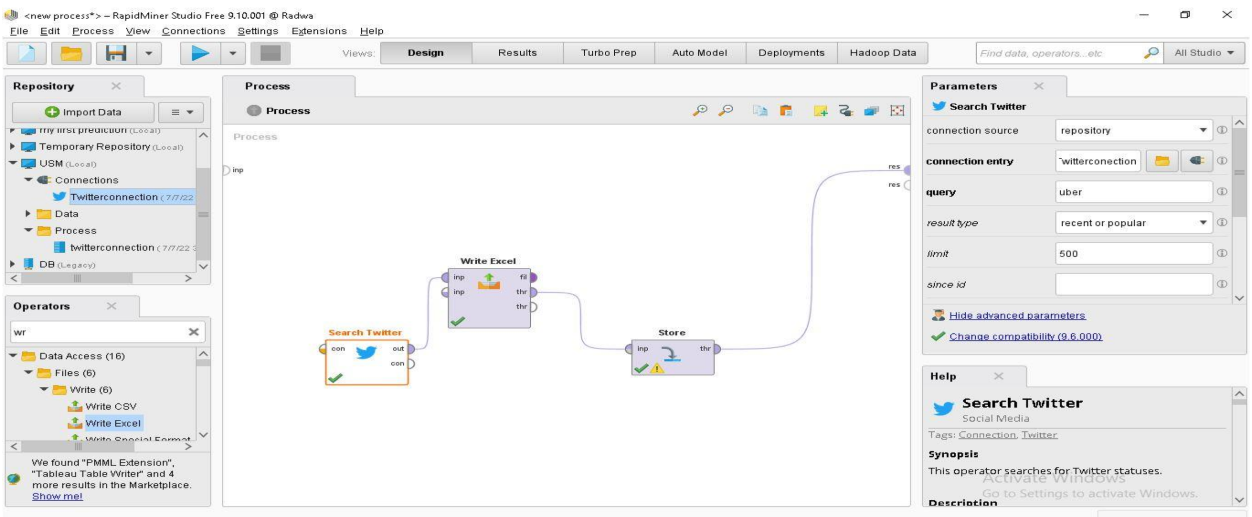
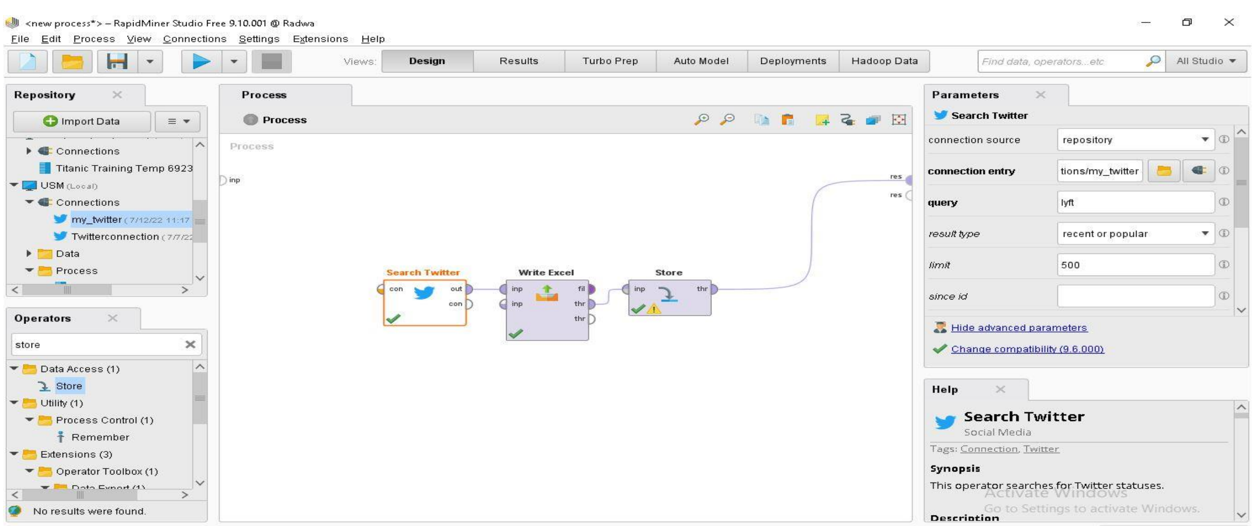


Figure 2 Data Collection about uber

Also, this Process is Done for Lyft Brand to get the data about it as shown in the next Figure:

Figure 3 Data collection about lyft



3.3.2 Data Cleaning for Uber

Raw data will not be used as effective as the cleaned data, So we will clean the raw data in the following steps;

Selecting attributes

because there are missing values and unnecessary values such as the username or the Id etc. So, we will explain these steps for Uber as an example in the following Figures.

Open in Turbo Prep Auto Model

Filter (1,000 / 1,000 examples): all

Row No.	Created-At	From-User	From-User-Id	To-User	To-U...	Lang...	Source	Text	Geo-L...	Geo-L...	Retwee
1	Jul 14, 2022 ...	Denise 'Holly...	2540948251	?	-1	en	<a href="http...	TIP: RENT A ...	?	?	38
2	Jul 15, 2022 ...	Leah Clark	15974916	?	-1	en	<a href="http:/...	Wow… @lyft i...	?	?	19
3	Jul 12, 2022 ...	Mark D. Levine	17143007	?	-1	en	<a href="http:/...	A reminder o...	?	?	476
4	Jul 15, 2022 ...	Scott Clodfelter	253047208	?	-1	en	<a href="http:/...	Hey @lyft pret...	?	?	0
5	Jul 15, 2022 ...	EPISODE 46:...	8983171147...	kathygr...	2114...	en	<a href="http...	@kathygriffin ...	?	?	0
6	Jul 15, 2022 ...	✕	1655797448	?	-1	en	<a href="http:/...	I hate smellin...	?	?	0
7	Jul 15, 2022 ...	Derek Whiten...	326243050	SuzAnn...	1219...	en	<a href="http:/...	@SuzAnneD...	?	?	0
8	Jul 15, 2022 ...	John grekko	1528678739...	Lucy45...	1483...	en	<a href="http...	@Lucy456Se...	?	?	0
9	Jul 15, 2022 ...	いしはらけん ...	2168734375	?	-1	ja	<a href="http:/...	RT @shin_s...	?	?	1
10	Jul 15, 2022 ...	Masaki NYC~...	2396331090	?	-1	ja	<a href="http:/...	lyftの運ちゃん...	?	?	0
11	Jul 15, 2022 ...	Margie Power	1477128668	BarbMc...	2009...	en	<a href="http...	@BarbMcQu...	?	?	0
12	Jul 15, 2022 ...	Separate_an...	1244372287...	?	-1	en	<a href="http...	RT @DGowa...	?	?	1
13	Jul 15, 2022 ...	Sen @ super...	1330428668...	?	-1	en	<a href="http:/...	RT @LeahCl...	?	?	19
14	Jul 15, 2022 ...	Howard Kush...	21632362	?	-1	en	<a href="http...	Have seen th...	?	?	0
15	Jul 15, 2022 ...	yarn yoshi	391620787	?	-1	en	<a href="http:/...	Sophia & I ar...	?	?	0
16	Jul 15, 2022 ...	Joseph Russo	1516881536...	VICEUK	1599...	en	<a href="http:/...	@VICEUK @...	?	?	0

ExampleSet (1,000 examples, 0 special attributes, 12 regular attributes)

Figure 4 Un clean data

Then we selected the necessary attributes we will use as shown in *Figure 5*

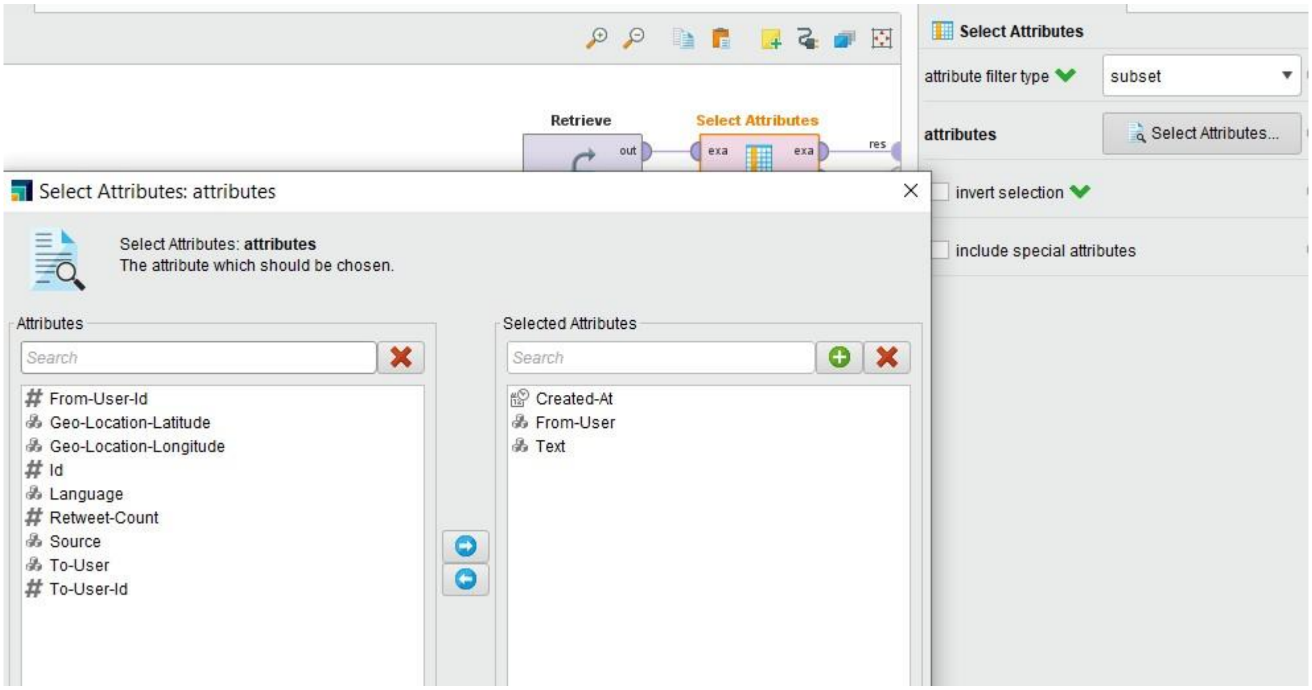


Figure 5 selecting attributes will be used

Then we need to remove unnecessary values and trim texts Which means removing the white space because it is unused as shown in the processes as follows;

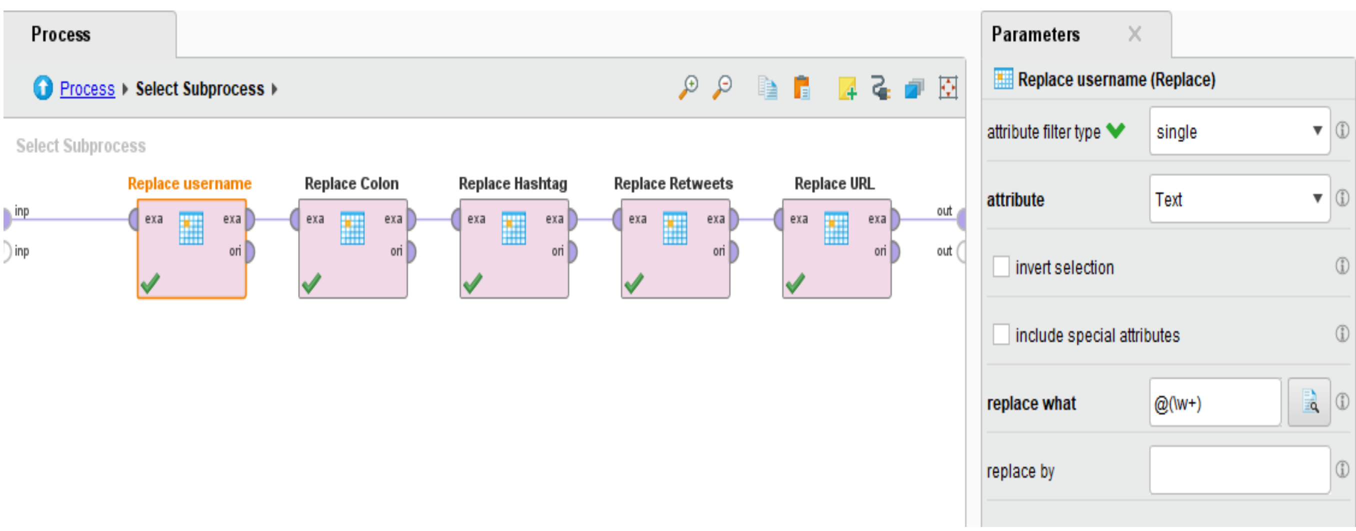


Figure 6 Selecting unusable attributes to remove them

And the Process as a whole is shown in the following Figure;

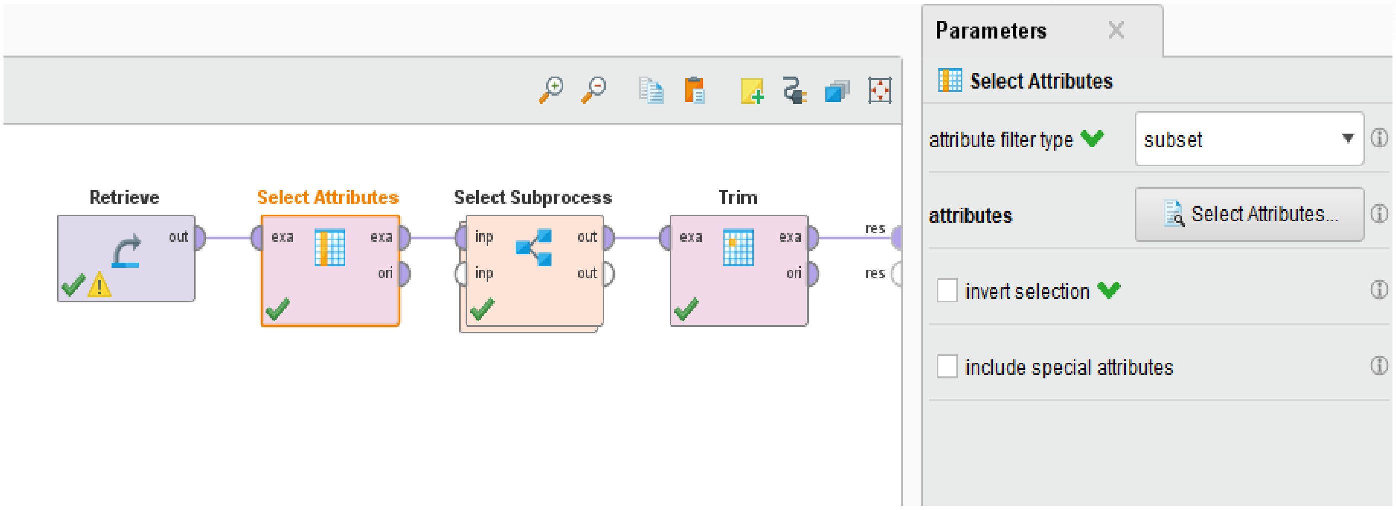


Figure 7 Process from selecting attributes then removing unusable values to trimming

Removing Duplicates

Now, Our data is nearly clean but we are facing that there are a lot of duplicates which leads to giving un accurate results so we need to remove them using the "Remove duplicate" Operator as shown in the following Process:

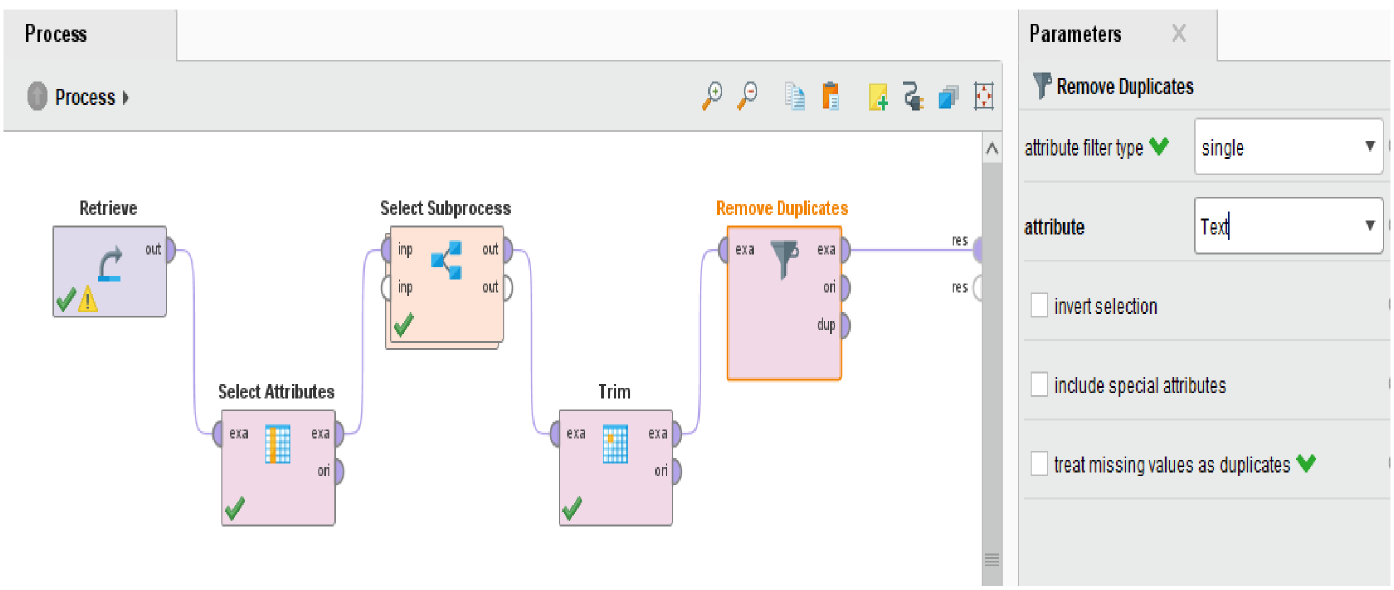
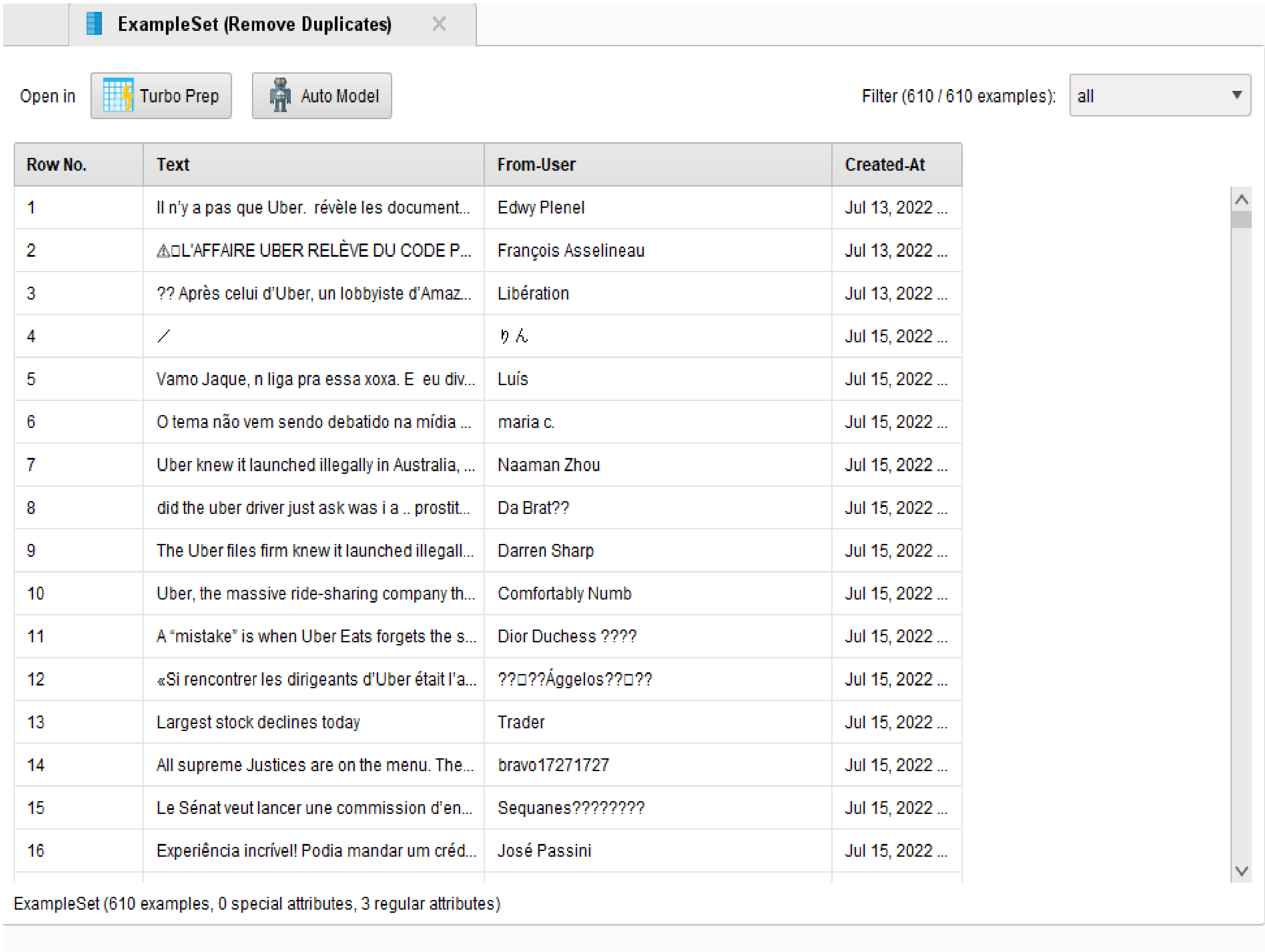


Figure 8 Process till removing duplicate

So, We can see the results in the following Figure that the data become 610 raw instead of 1,000 raw;



ExampleSet (Remove Duplicates)

Open in Turbo Prep | Auto Model | Filter (610 / 610 examples): all

Row No.	Text	From-User	Created-At
1	Il n'y a pas que Uber. révèle les document...	Edwy Plenel	Jul 13, 2022 ...
2	⚠□L'AFFAIRE UBER RELÈVE DU CODE P...	François Asselineau	Jul 13, 2022 ...
3	?? Après celui d'Uber, un lobbyiste d'Amaz...	Libération	Jul 13, 2022 ...
4	⟋	りん	Jul 15, 2022 ...
5	Vamo Jaque, n liga pra essa xoxa. E eu div...	Luís	Jul 15, 2022 ...
6	O tema não vem sendo debatido na mídia ...	maria c.	Jul 15, 2022 ...
7	Uber knew it launched illegally in Australia, ...	Naaman Zhou	Jul 15, 2022 ...
8	did the uber driver just ask was i a .. prostit...	Da Brat??	Jul 15, 2022 ...
9	The Uber files firm knew it launched illegall...	Darren Sharp	Jul 15, 2022 ...
10	Uber, the massive ride-sharing company th...	Comfortably Numb	Jul 15, 2022 ...
11	A "mistake" is when Uber Eats forgets the s...	Dior Duchess ????	Jul 15, 2022 ...
12	«Si rencontrer les dirigeants d'Uber était l'a...	??□??Águelos??□??	Jul 15, 2022 ...
13	Largest stock declines today	Trader	Jul 15, 2022 ...
14	All supreme Justices are on the menu. The...	bravo17271727	Jul 15, 2022 ...
15	Le Sénat veut lancer une commission d'en...	Sequanes????????	Jul 15, 2022 ...
16	Experiência incrível! Podia mandar um créd...	José Passini	Jul 15, 2022 ...

ExampleSet (610 examples, 0 special attributes, 3 regular attributes)

Figure 9 Data after Removing duplications

Removing The Missing Values

We need to remove Missing values to make the best use of our data and make sure that our data is completely clean. So, we used the "Replace Missing Values" Operator as shown in the Following Figure;

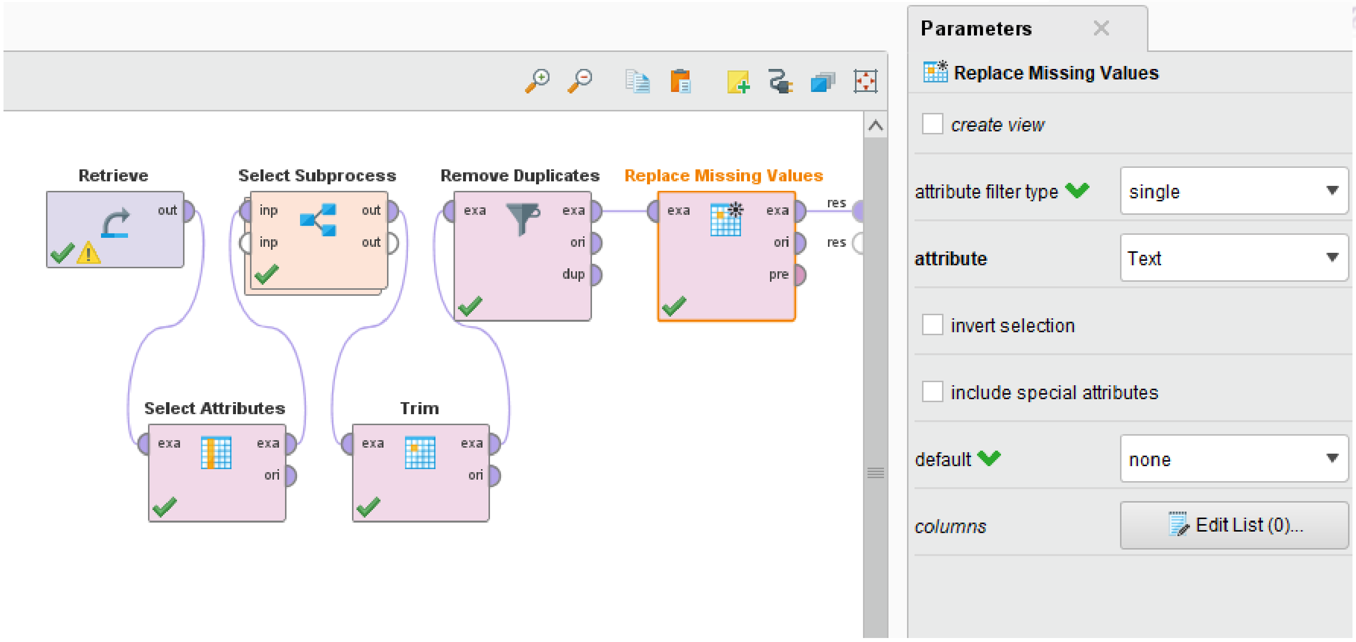


Figure 10 The Process until Replacing Missing Values

And The results in our data that contains missing Values which will be deleted in uber data are as follows;

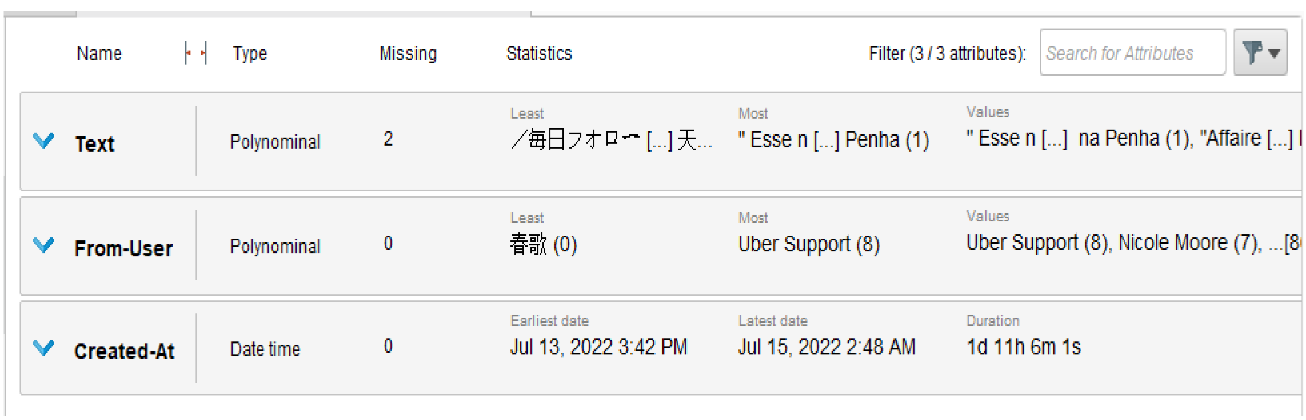


Figure 11 Data Containing Missing values

Finally, this our Process to clean the Data;

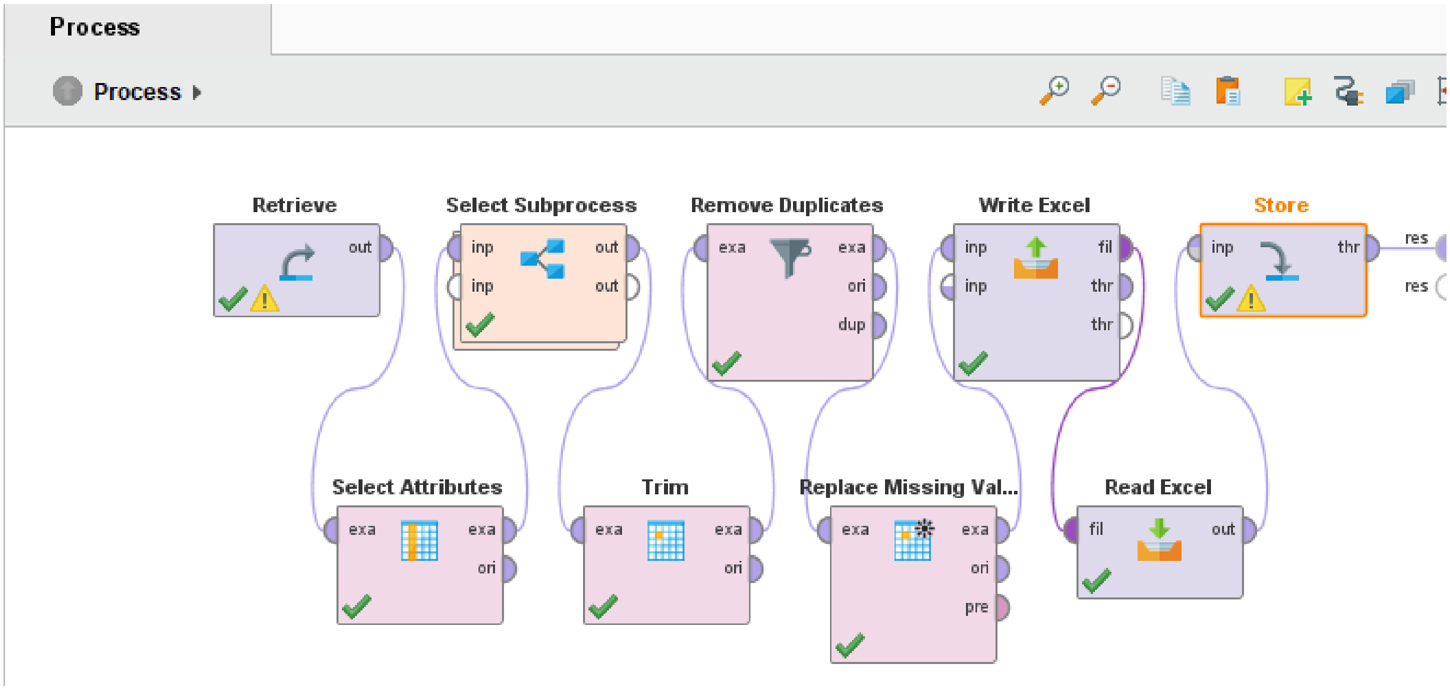


Figure 12 Process to clean the Uber Data

3.3.3 Data Cleaning For Lyft

The Same steps are made to make the data for Lyft Clean as shown in the following Figure.

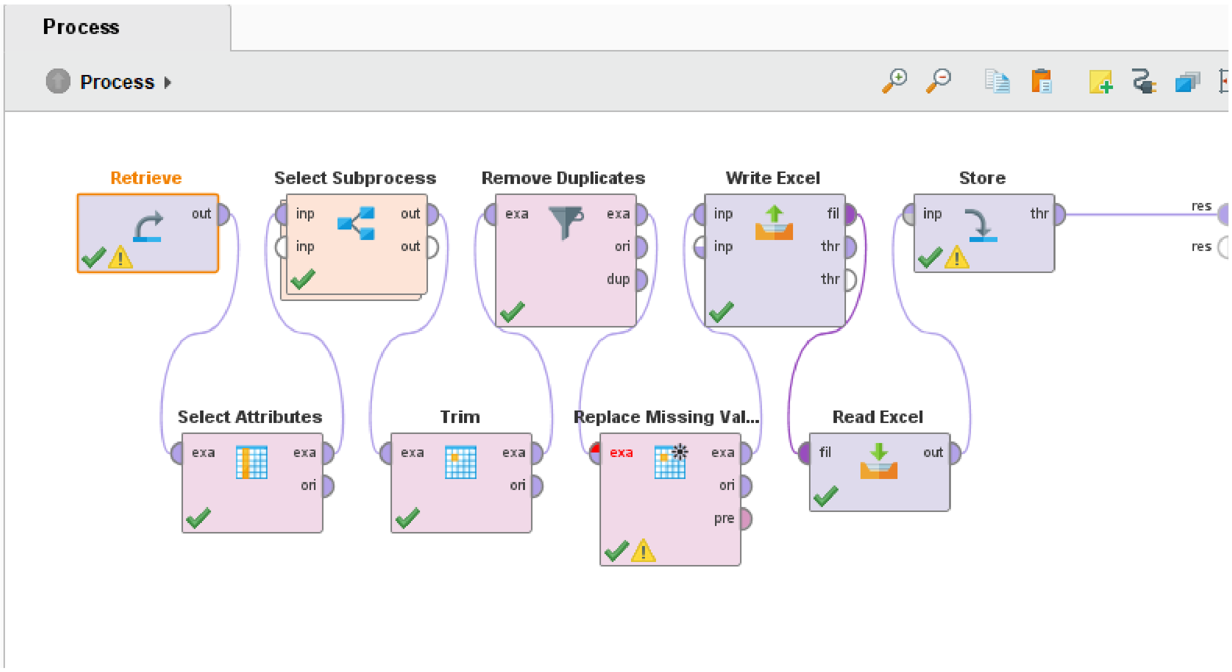
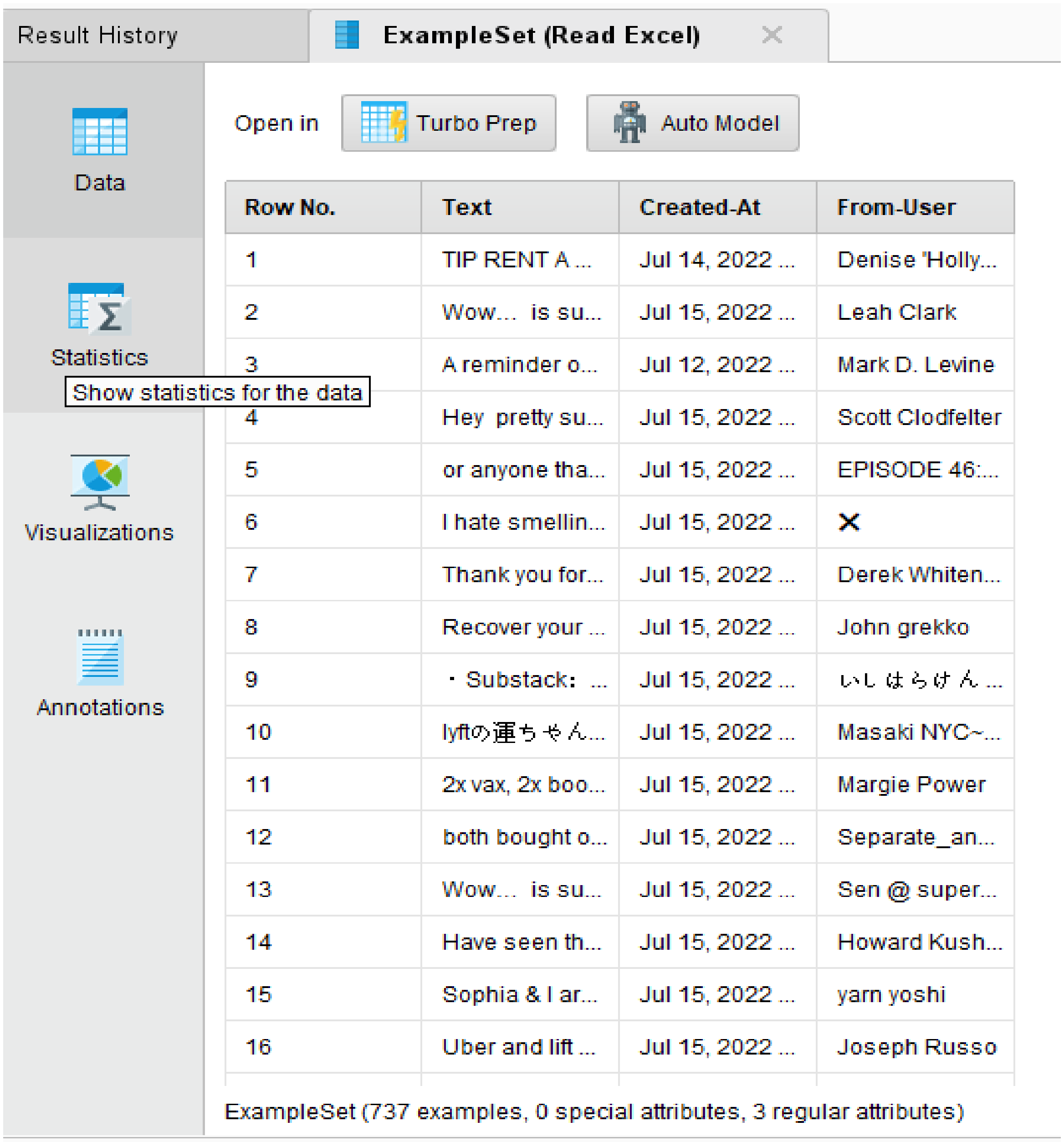


Figure 13 Process to clean Lyft Data

But when removing the duplicate Values, it results 737 raw instead of 1,000 raw as follows.



Row No.	Text	Created-At	From-User
1	TIP RENT A ...	Jul 14, 2022 ...	Denise 'Holly...
2	Wow... is su...	Jul 15, 2022 ...	Leah Clark
3	A reminder o...	Jul 12, 2022 ...	Mark D. Levine
4	Hey pretty su...	Jul 15, 2022 ...	Scott Clodfelter
5	or anyone tha...	Jul 15, 2022 ...	EPISODE 46:...
6	I hate smellin...	Jul 15, 2022 ...	✕
7	Thank you for...	Jul 15, 2022 ...	Derek Whiten...
8	Recover your ...	Jul 15, 2022 ...	John grekko
9	・Substack: ...	Jul 15, 2022 ...	いしはらけん ...
10	lyftの運ちゃん...	Jul 15, 2022 ...	Masaki NYC~...
11	2x vax, 2x boo...	Jul 15, 2022 ...	Margie Power
12	both bought o...	Jul 15, 2022 ...	Separate_an...
13	Wow... is su...	Jul 15, 2022 ...	Sen @ super...
14	Have seen th...	Jul 15, 2022 ...	Howard Kush...
15	Sophia & I ar...	Jul 15, 2022 ...	yarn yoshi
16	Uber and lift ...	Jul 15, 2022 ...	Joseph Russo

Figure 14 Cleaned Lyft Data

And we can make sure from the statistical view as follows.

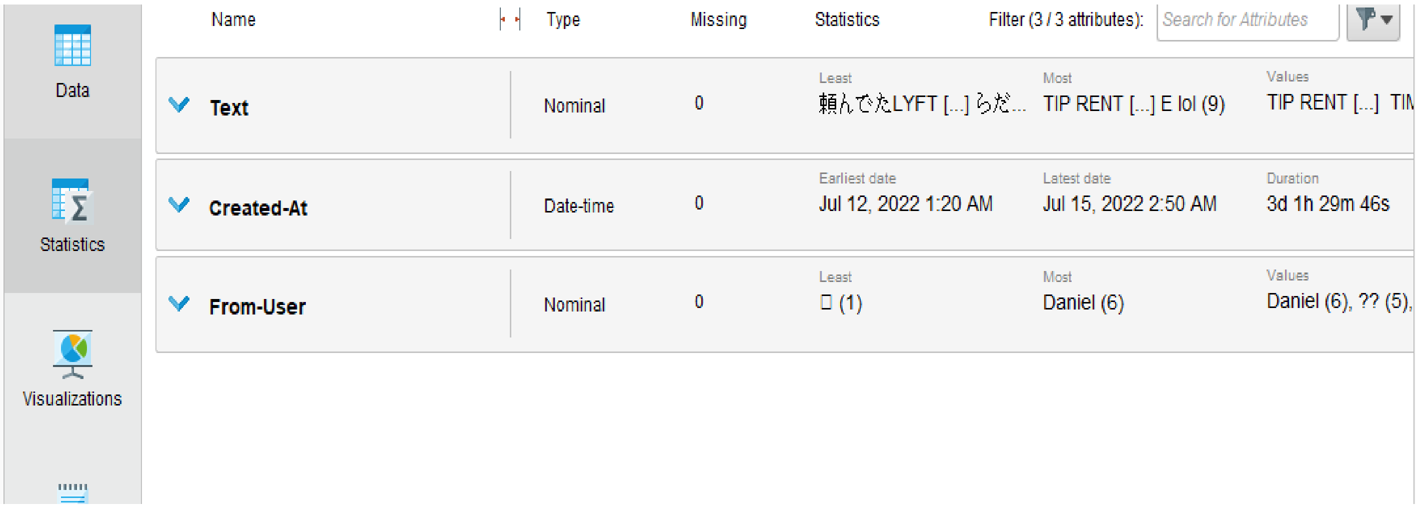


Figure 15 Missing Values in statistical View

3.3.4 Sentiment Analysis

After cleaning the data and revieing it carefully to make sure that there are not any errors. Now we need to make the sentiment analysis. The sentiment means grouping tweets that are collected into either Positive, Negative. So, we can use VADER to make the most appropriate sentiment analysis. VADER is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. The VADER algorithm outputs the polarity of sentiment to three classes (Positive, Negative, and Neutral) based on the compound score. The Compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1(most extreme negative) and +1 (most extreme positive).

positive sentiment: (compound score > 0)

negative sentiment: (compound score < 0)

neutral sentiment: (compound score == 0)

This process can be shown as follows:

We can Start Firstly by

3.3.4.1 Sentiment analysis for Uber

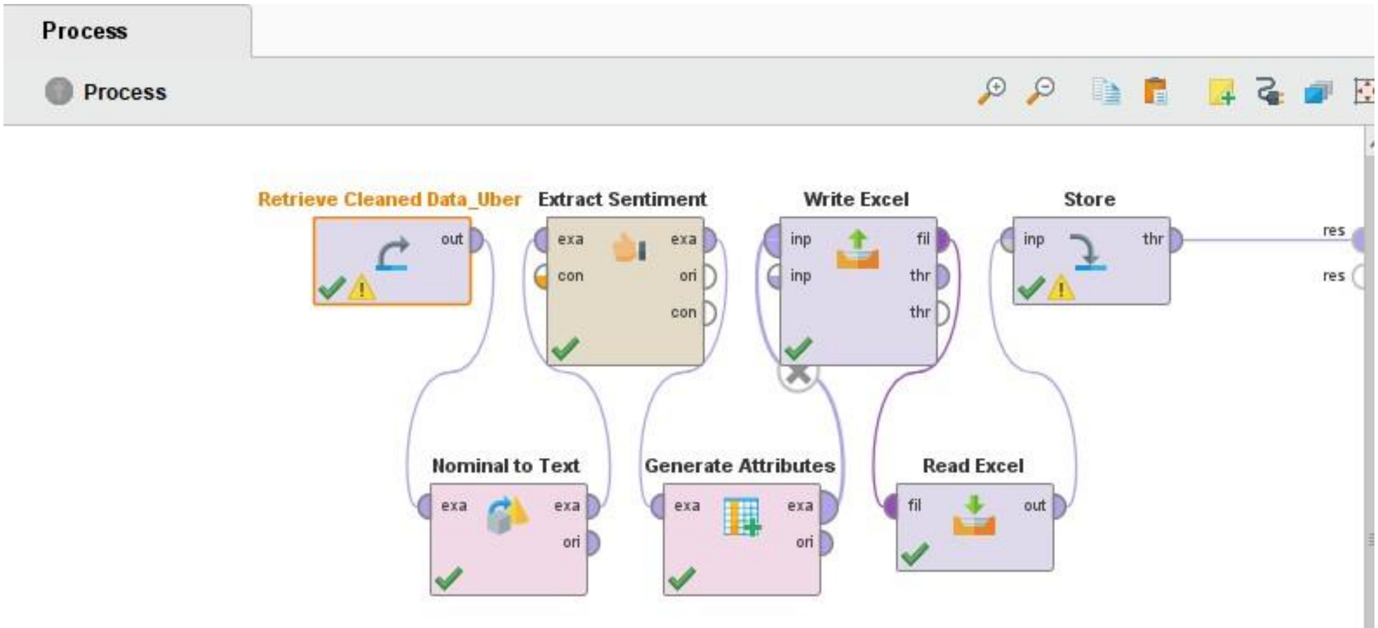


Figure 16 Process of Uber Sentiment analysis

Firstly, We need to generate and Specify the sentiment (Positive, Negative or Neutral) based on the score he result. For the generate attribute, function expressions be set as be shown below:

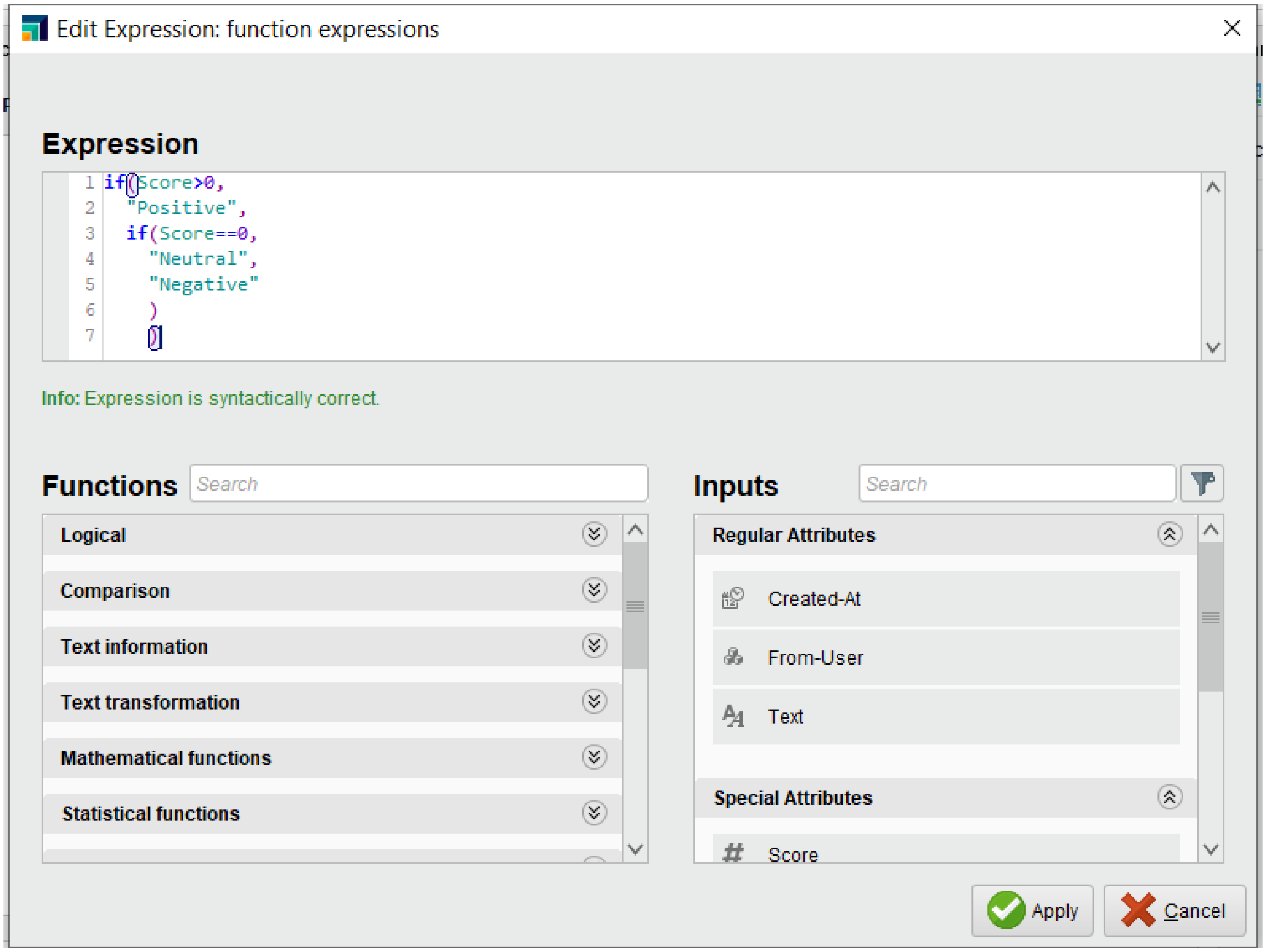


Figure 17 Function Expression of generate the attribute

And the result of sentiment analysis can be shown as the following figure:

Open in Turbo Prep Auto Model Filter (610 / 610 examples): all

R...	Text	From-User	Created-At	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Sentiment
2	⚠□L'AFFA...	François Ass...	Jul 13, 2022 ...	1.026	l (0.51) l (0.51)	0	1.026	53	55	Positive
3	?? Après c...	Libération	Jul 13, 2022 ...	0	?	0	0	20	20	Neutral
4	⁄	りん	Jul 15, 2022 ...	0.179	amazon (0.18)	0	0.179	22	23	Positive
5	Vamo Jaq...	Luís	Jul 15, 2022 ...	0	?	0	0	21	21	Neutral
6	O tema nã...	maria c.	Jul 15, 2022 ...	-0.308	no (-0.31)	0.308	0	24	25	Negative
7	Uber kne...	Naaman Zhou	Jul 15, 2022 ...	0.256	launched (0.1...	0	0.256	19	21	Positive
8	did the ub...	Da Brat??	Jul 15, 2022 ...	0	?	0	0	13	13	Neutral
9	The Uber f...	Darren Sharp	Jul 15, 2022 ...	0.128	launched (0.1...	0	0.128	18	19	Positive
10	Uber, the ...	Comfortably ...	Jul 15, 2022 ...	0.231	sharing (0.46...	0.231	0.462	20	22	Positive
11	A "mistake...	Dior Duches...	Jul 15, 2022 ...	0	?	0	0	23	23	Neutral
12	«Si renco...	??□??Águel...	Jul 15, 2022 ...	0	?	0	0	18	18	Neutral
13	Largest st...	Trader	Jul 15, 2022 ...	0	?	0	0	36	36	Neutral
14	All supre...	bravo172717...	Jul 15, 2022 ...	0.667	supreme (0.6...	0	0.667	13	14	Positive
15	Le Sénat v...	Sequanes??...	Jul 15, 2022 ...	0	?	0	0	22	22	Neutral
16	Experiênci...	José Passini	Jul 15, 2022 ...	0.667	hahaha (0.67)	0	0.667	16	17	Positive
17	Jajajajaj...	??✨	Jul 15, 2022 ...	0	?	0	0	16	16	Neutral

ExampleSet (610 examples, 0 special attributes, 10 regular attributes)

Figure 18 Sentiment analysis for Uber

But from the table which contains a lot of numbers, specially in 610 raw, it will be very difficult to get the final result to know the sentiment. So, to understand the sentiment easily we can visualize it in different ways like Pie chart and par chart as follows in the next figures;

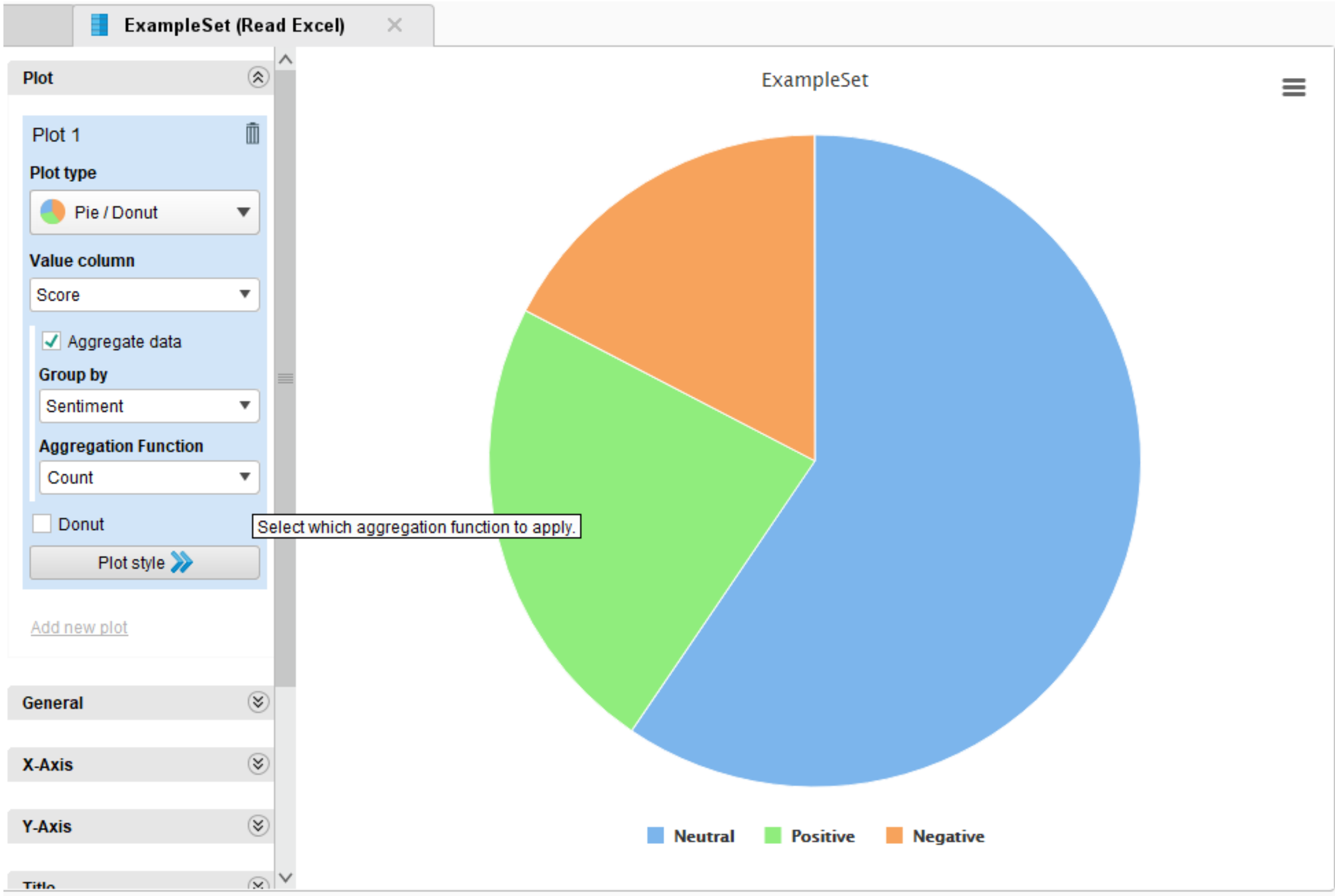
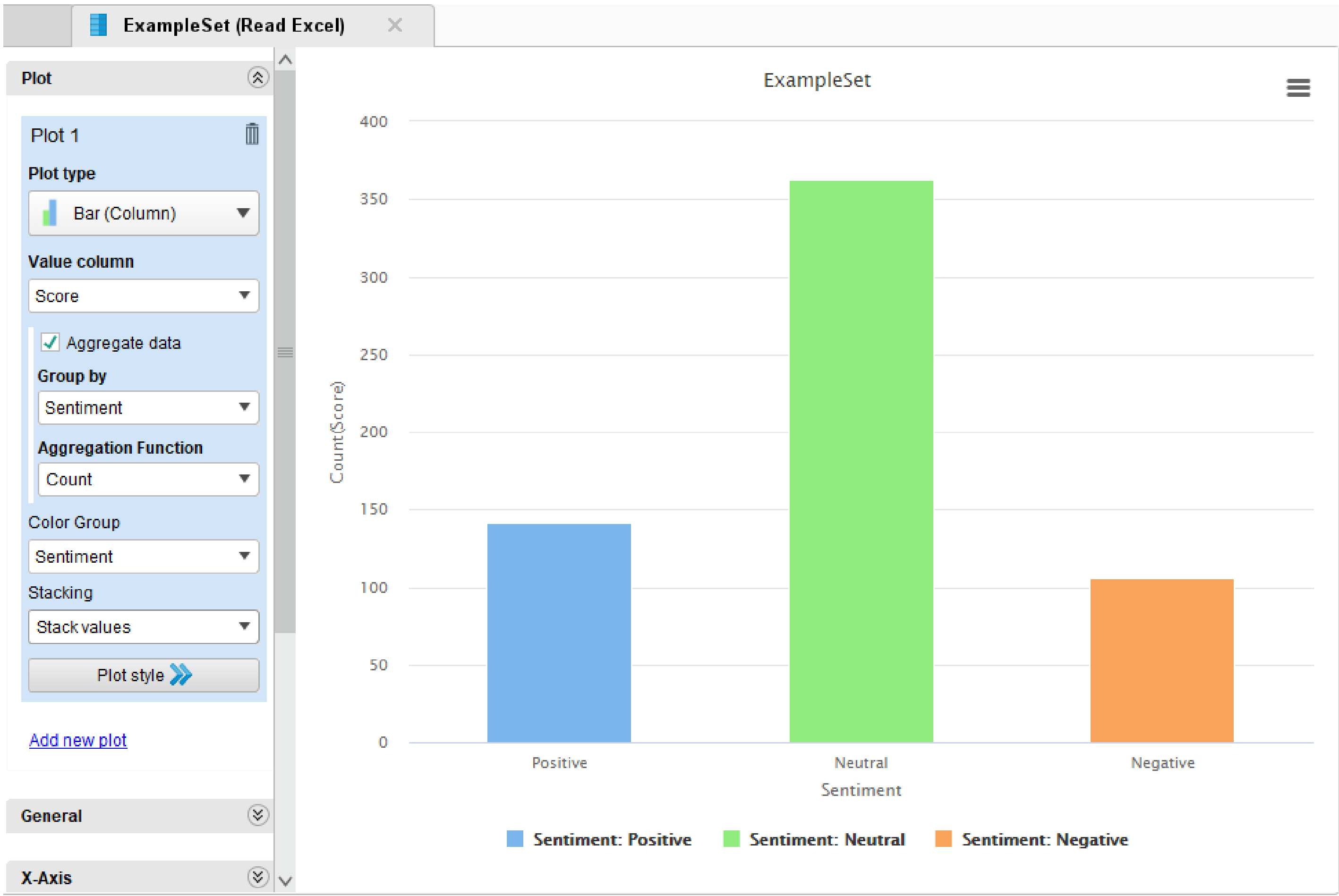


Figure 19 Pie chart showing Uber sentiment analyis

To make more Sure, We can visualize as Par chart

Figure 20 Par chart showing Uber Sentiment analysis



From these visualizations in the Pie chart and Par chart, We can get that the highest sentiment is Neutral sentiment then Positive and finally Negative sentiment.

We can best see the difference using Funnel chart as Follows;

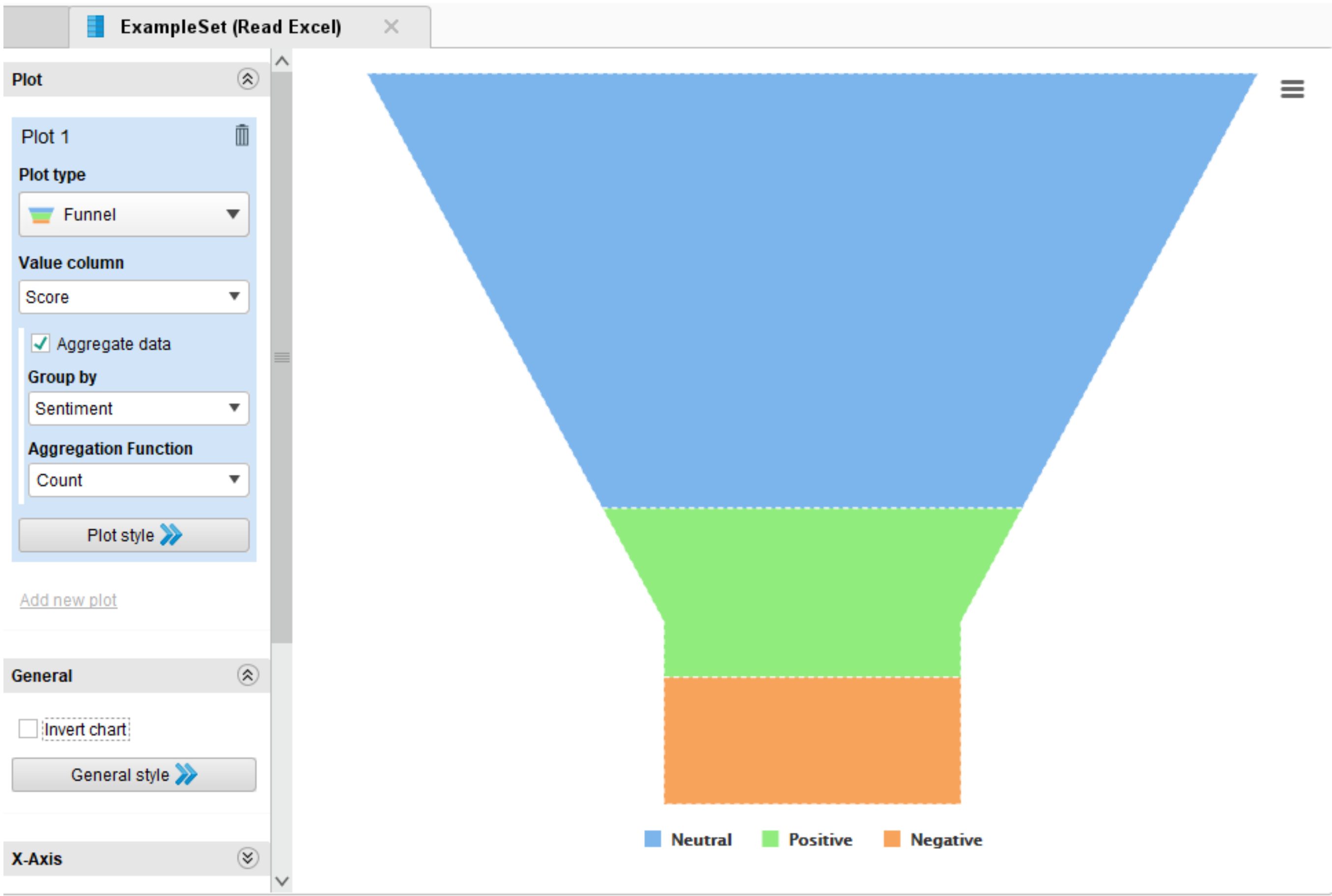


Figure 21 Funnel Chart for best showing Uber sentiment

3.3.4.2 Sentiment analysis For Lyft

We can make the same Process and steps to make lyft Sentiment analysis but the results will be different because the raw data are different

So, Firstly the Process as a whole as follows;

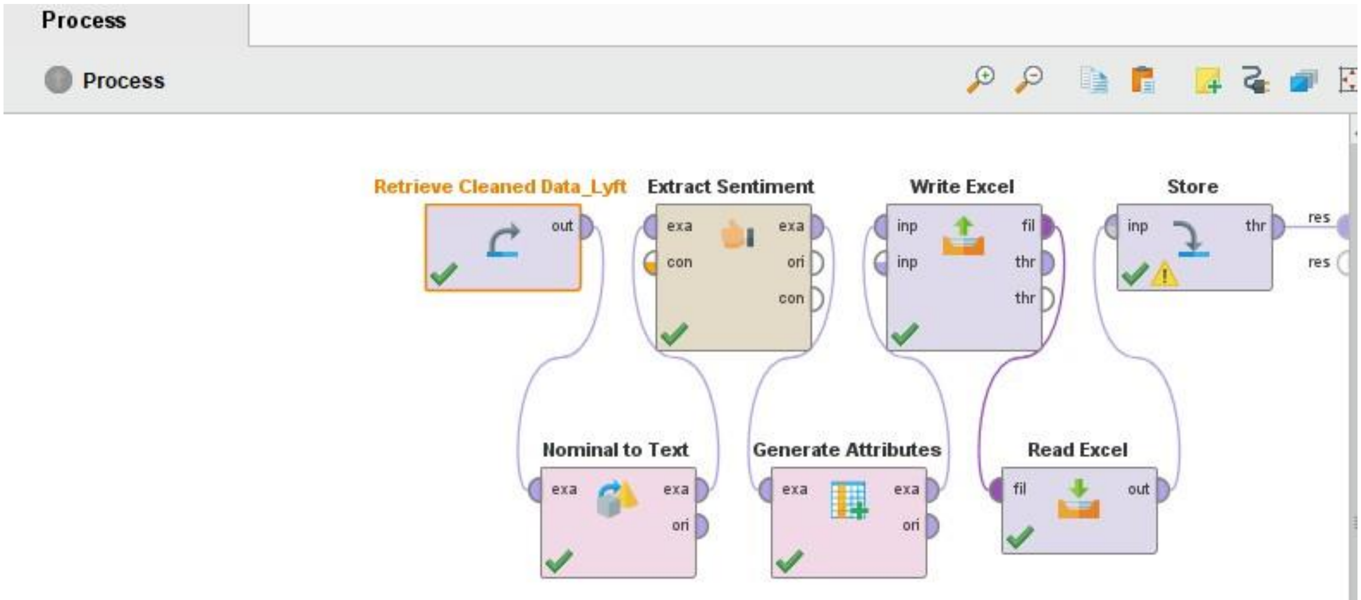


Figure 22 Sentiment analysis Process for Lyft

And the generator Expression would be as Follows;

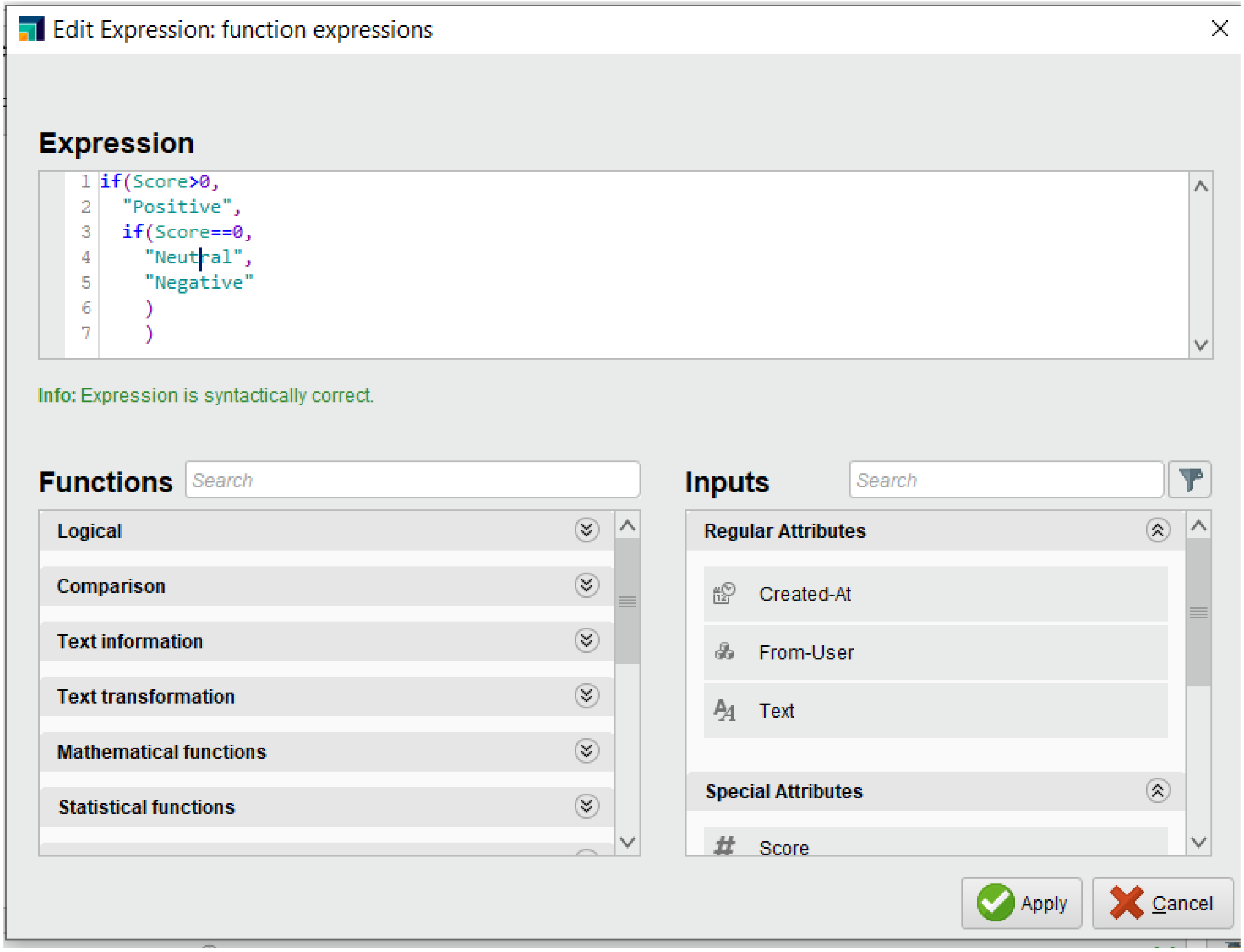


Figure 23 generator expression for Lyft

The Result of this process can be shown as follows;

Open in Turbo Prep Auto Model Filter (737 / 737 examples): all

Row ...	Text	Created-At	From-User	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Sentiment
1	TIP RENT...	Jul 14, 2022 ...	Denise 'Holly...	1.513	trust (0.59) lu...	0	1.513	47	50	Positive
2	Wow... is...	Jul 15, 2022 ...	Leah Clark	2.590	wonderful (0....	0	2.590	43	47	Positive
3	A reminde...	Jul 12, 2022 ...	Mark D. Levine	0	?	0	0	52	52	Neutral
4	Hey pretty...	Jul 15, 2022 ...	Scott Clodfelter	0.897	pretty (0.56) ...	0	0.897	25	27	Positive
5	or anyone...	Jul 15, 2022 ...	EPISODE 46:...	0.462	true (0.46)	0	0.462	23	24	Positive
6	I hate sm...	Jul 15, 2022 ...	✕	-0.564	hate (-0.69) li...	1.333	0.769	14	18	Negative
7	Thank you...	Jul 15, 2022 ...	Derek Whiten...	0.385	thank (0.38)	0	0.385	4	5	Positive
8	Recover y...	Jul 15, 2022 ...	John grekko	0	?	0	0	13	13	Neutral
9	・Substa...	Jul 15, 2022 ...	いしはらけん...	0	?	0	0	5	5	Neutral
10	lyftの運ち...	Jul 15, 2022 ...	Masaki NYC~...	0	?	0	0	5	5	Neutral
11	2x vax, 2x ...	Jul 15, 2022 ...	Margie Power	0.949	grateful (0.51...	0	0.949	48	50	Positive
12	both boug...	Jul 15, 2022 ...	Separate_an...	0	?	0	0	10	10	Neutral
13	Wow... is...	Jul 15, 2022 ...	Sen @ super...	1.897	wonderful (0....	0	1.897	20	23	Positive
14	Have see...	Jul 15, 2022 ...	Howard Kush...	0	?	0	0	28	28	Neutral
15	Sophia & I...	Jul 15, 2022 ...	yarn yoshi	0.538	amazing (0.7...	0.179	0.718	23	25	Positive
16	Uber and ...	Jul 15, 2022 ...	Joseph Russo	-0.359	screw (-0.10) ...	0.692	0.333	53	57	Negative

ExampleSet (737 examples, 0 special attributes, 10 regular attributes)

Figure 24 Results of sentiment analysis for Lyft

Finally, the Visualization using Par chart and Pie chart are as follows;

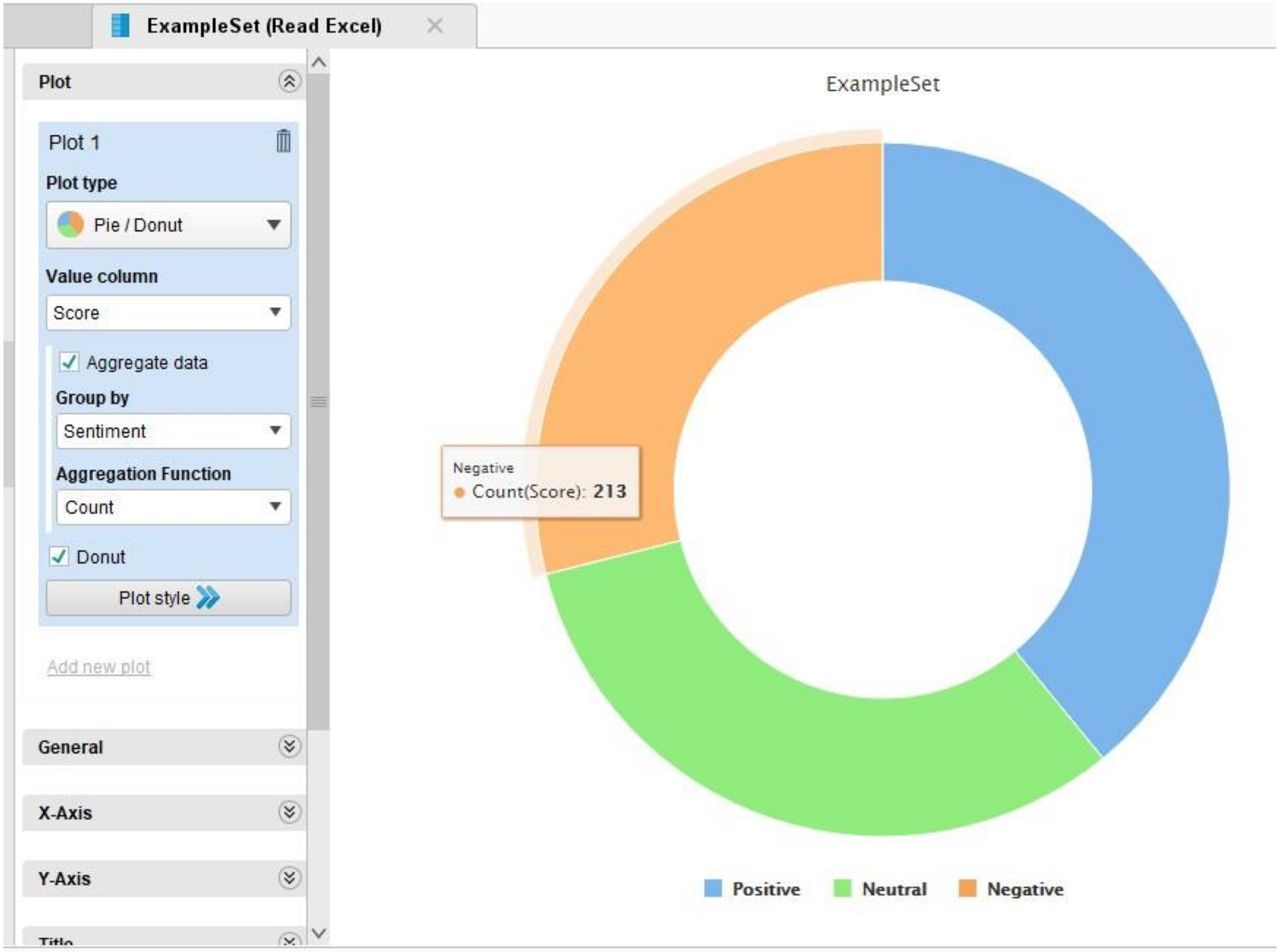


Figure 25 Pie chart showing Lyft sentiment analysis

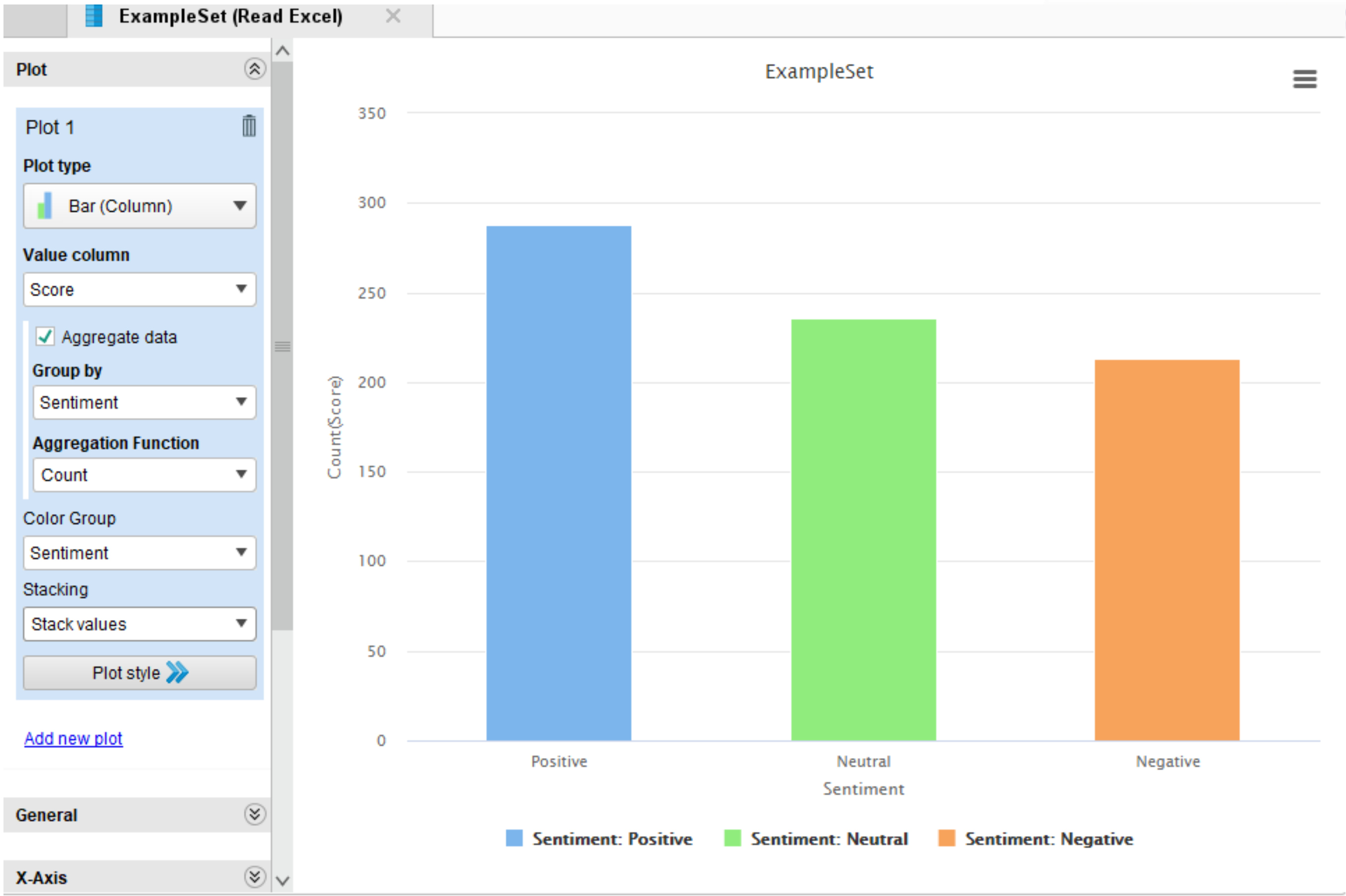


Figure 26 Par chart showing Lyft sentiment analysis

As we can see from the lyft sentiment analysis that the first class is Positive opinions, then Neutral Opinions to lastly Negative Opinions.

As well, Here in Lyft we can see the best showing for sentiment analysis using Funnel Chart;

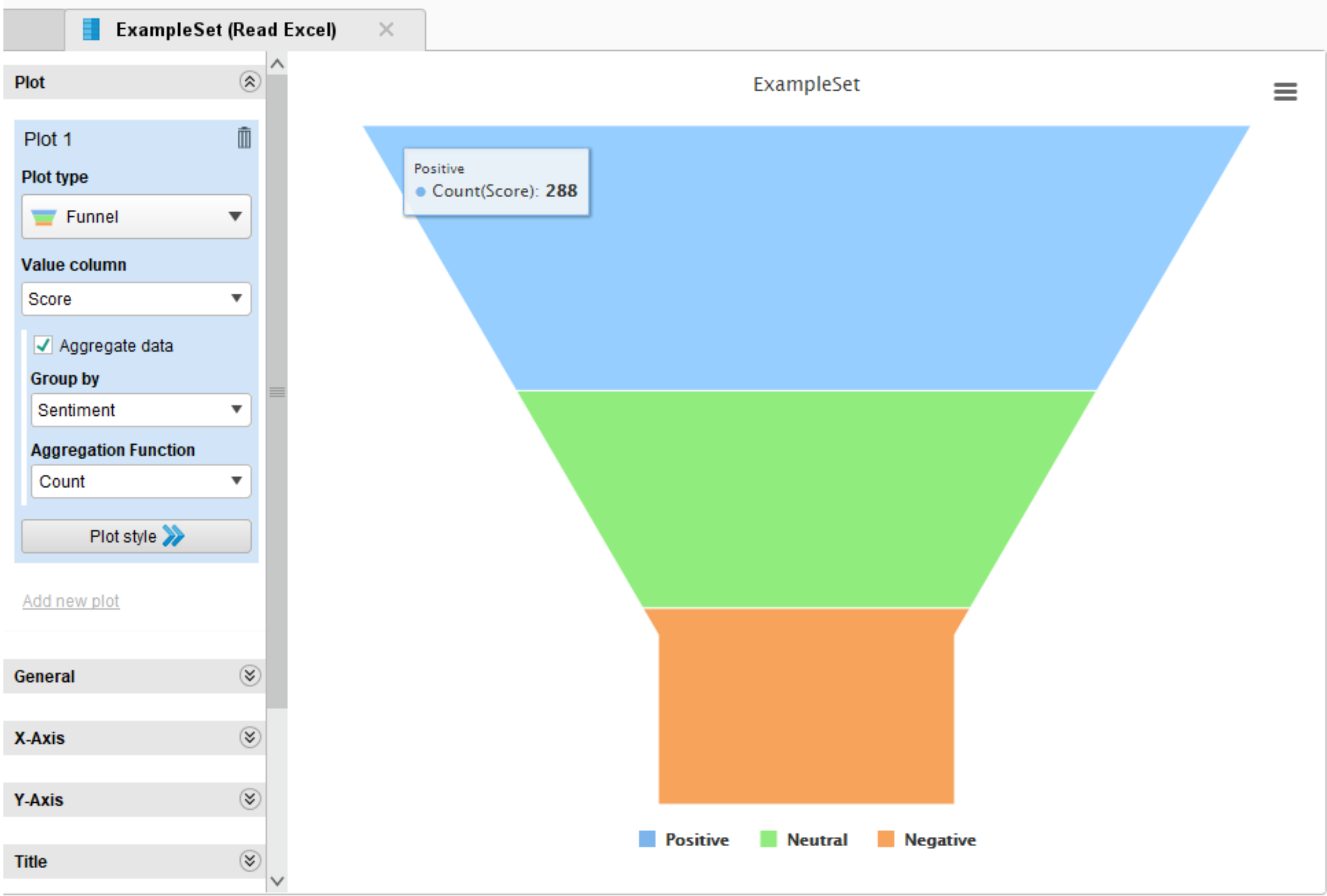


Figure 27 Funnel chart showing Lyft sentiment

3.3.5 Frequency Analysis and Word Cloud

Frequency analysis concerned with the number of times a word appears in the document. While the Word Cloud propose visualization of the text representation. The word cloud displays the words used in the dataset and the more frequently the word appeared, then the bigger the word is displayed in the word cloud. This process is important to know how the exact perspective of the customer towards every word that they have tweet and how-to sentiment it.

3.3.5.1 For Uber

We can show the Process as follows;

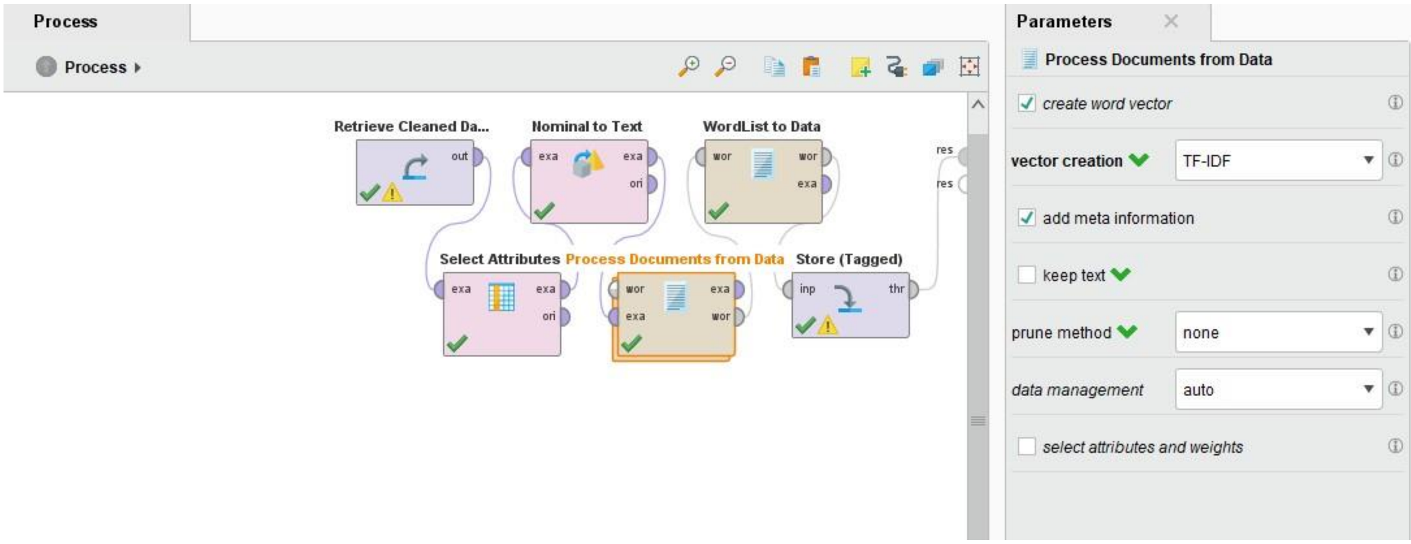


Figure 28 Frequency and Word cloud process for Uber

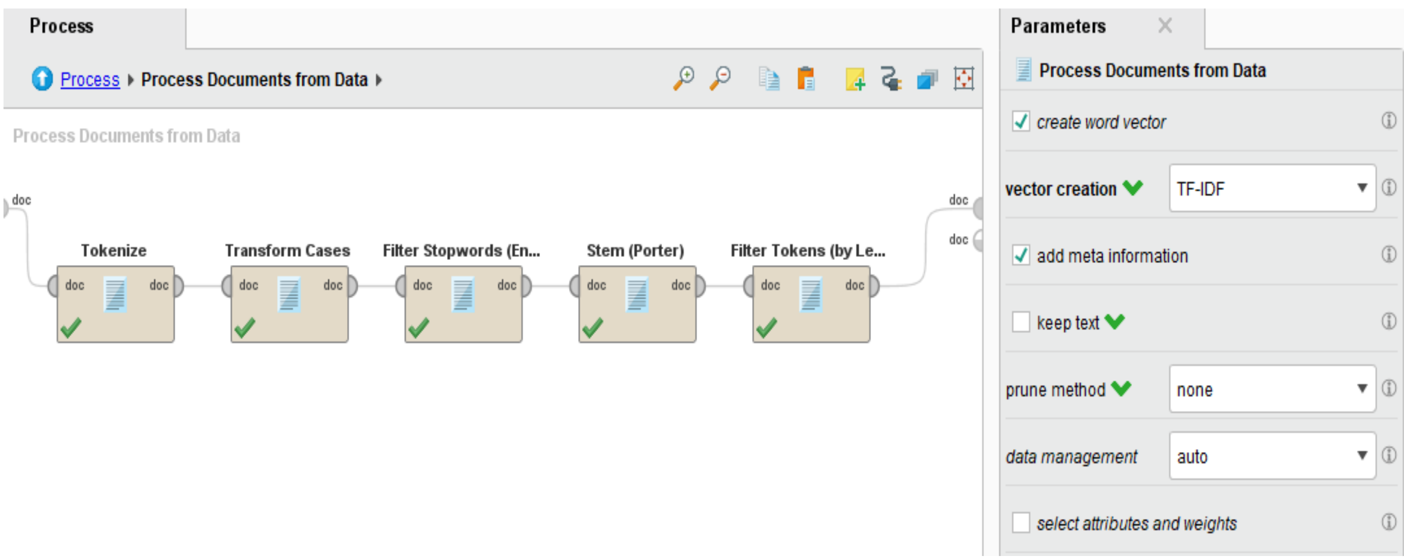


Figure 29 Subprocess to split the text

Both above figures show the process of generate list. *Figure 28* be use and this operator work to converts all nominal attributes to string attributes. Each nominal value is simply used as a string value of the new attribute. If the value is missing in the nominal attribute, the new value will also be missing. For the *Figure 29* shows subprocess that basically from operator named "Process documents from data". In this operator, which have some few subprocess which is tokenize; to split the text into a single as be mentioned earlier in this paper, transform cases; to standardize all words etc., Stem (Portal); reduce the length of the word, Filter Tokens (by length); will remove any too short or too long words.

And the results could be as follows;

WordList (Process Documents from Data)

Word	Attribute Name	Total Occurences	Document Occurenc...
aae	aae	1	1
aamiin	aamiin	1	1
abad	abad	1	1
abbiamo	abbiamo	1	1
abl	abl	2	2
abogada	abogada	1	1
abordado	abordado	1	1
absolu	absolu	2	2
abus	abus	1	1
acaba	acaba	1	1
acabei	acabei	1	1
academ	academ	1	1
accabl	accabl	2	2
accept	accept	3	3
acceptera	acceptera	1	1
access	access	2	2

Figure 30 Results of the Frequency and word cloud process for Uber

Tokenization is the process of splitting any sentences, statement or few words that be given to the tools into a word and every word will have the converter's own scoring which is how many times the word be used, the popularity of the words, and the sentiment of the word itself.

- Nominal to text operator will be used and this operator works to convert all nominal attributes to string attributes. Each nominal value is simply used as a string value of the new attribute. If the value is missing in the nominal attribute, the new value will also be missing.

- Process document from data operator have some subprocess

-Transform Cases: standardize word to lower case.

-Filter Stop words (English): removes English stop words such as the, is, and from a document.

-Stem (Porter): reduce the length of the words until a minimum length is reached.

-Filter Tokens (By Length): filters tokens based on their length

- Wordlist to Data operator to convert created word list into a data set.

- finally, we used the Store operator to save the dataset.

This process can be summarized in the following Figure;

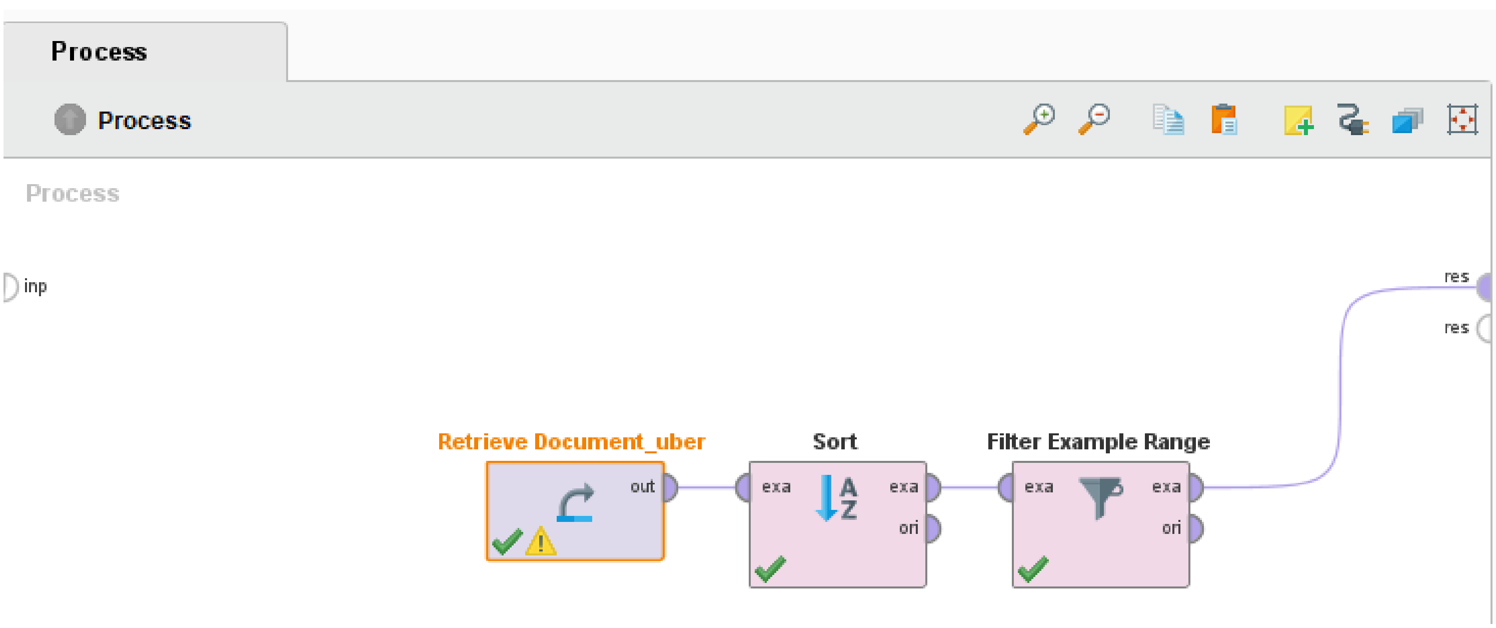


Figure 31 the rest of the Process to make word cloud

And The Result of it can be Shown as a table as Following;

Result History | ExampleSet (Filter Example Range)

Data | Statistics | Visualizations | Annotations

Row No.	word	in documents	total
1	uber	437	476
2	que	86	111
3	macron	90	90
4	est	67	84
5	qui	45	51
6	pour	42	48
7	driver	43	46
8	file	41	41
9	sur	38	40
10	uberfil	39	39
11	taxi	33	34
12	affair	23	28
13	vou	18	26
14	avec	21	24
15	mai	22	24
16	não	9	23

ExampleSet (25 examples, 0 special attributes, 3 regular attributes)

Figure 32 Data table to show the result of Frequency and Word Cloud analysis

Graphically we can see the results as following;

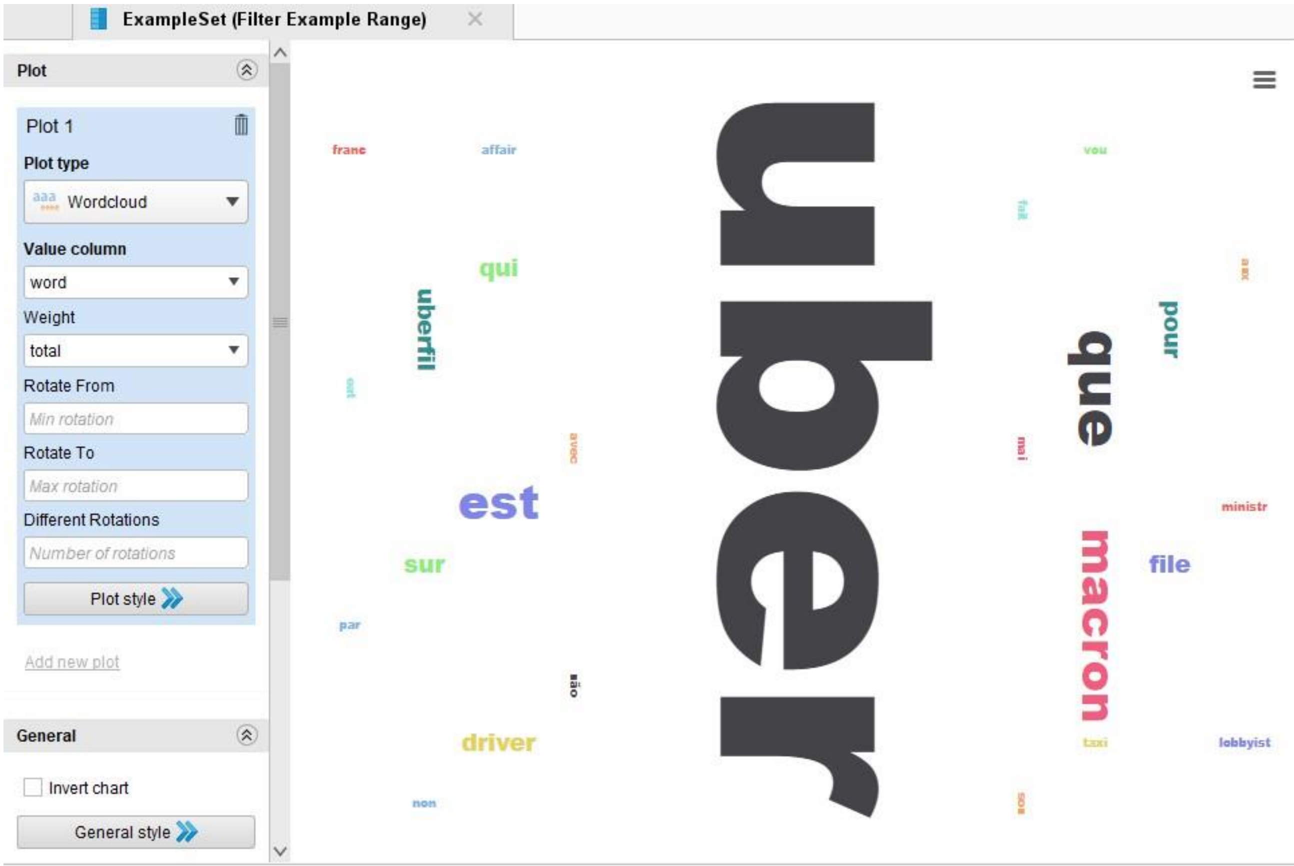


Figure 33 Graphically, The result of Frequency and Sentiment analysis for Uber

3.3.5.2 For Lyft

The Same Processes and subprocess are done For Lyft Brand except the data and the Results

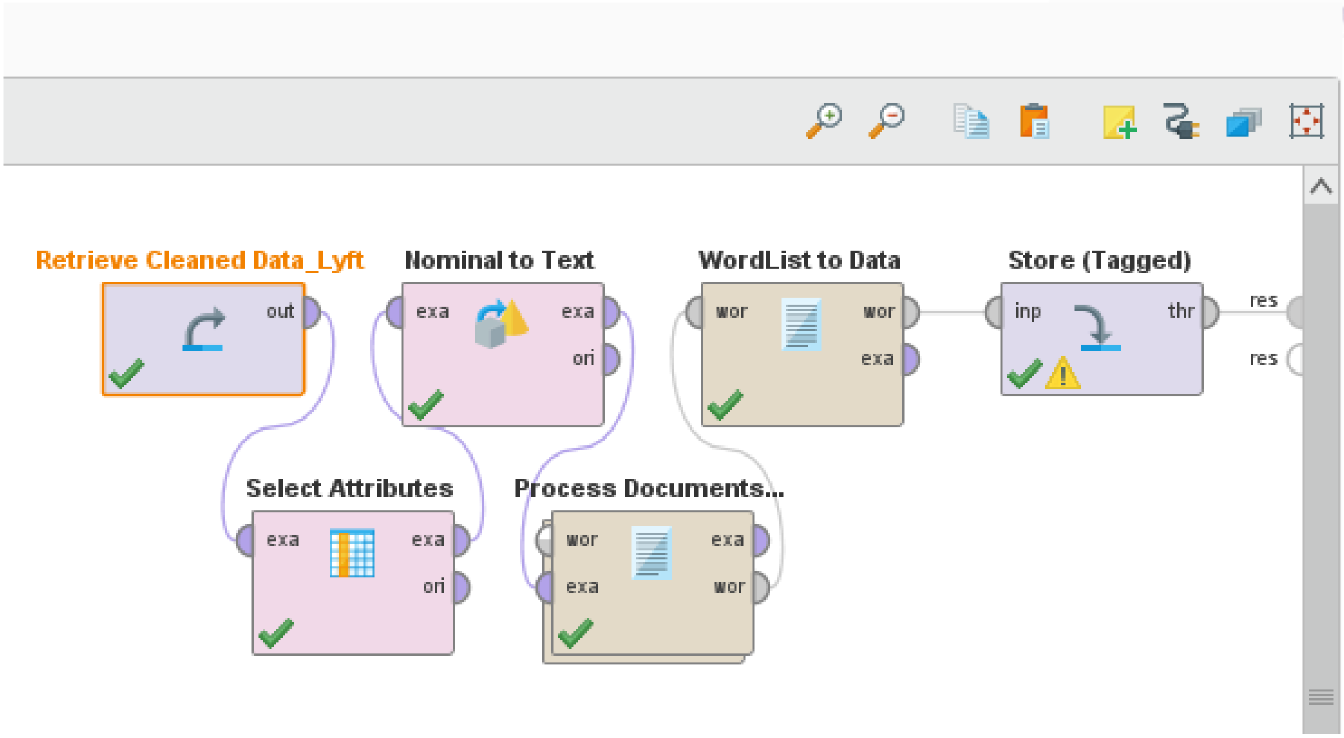


Figure 34 First Process for Frequency and Word Cloud Analysis

Also, For the Subprocess

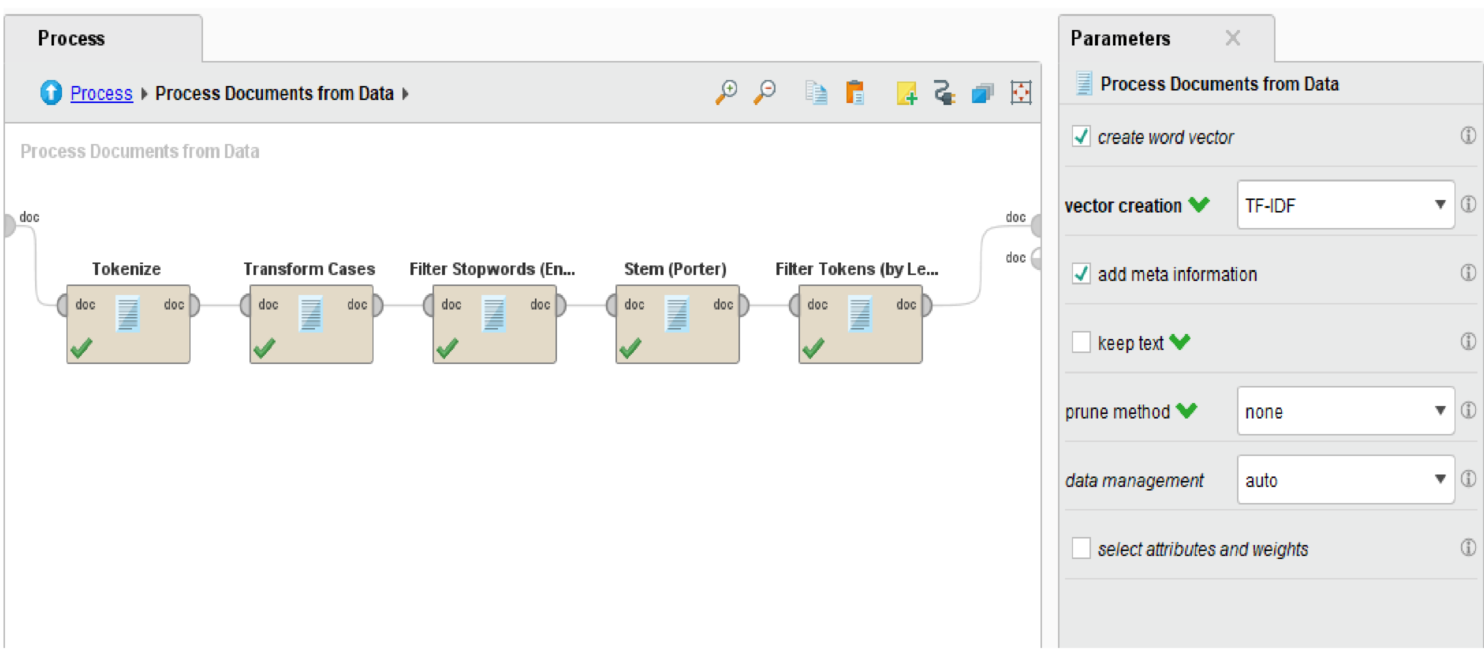


Figure 35 Subprocess To tokenize documentation

But the Results changed as follows;

WordList (Process Documents from Data)

Word	Attribute Name	Total Occurences	Document Occurences
aapl	aapl	2	2
aavad	aavad	1	1
abandon	abandon	1	1
abboouuttt	abboouuttt	1	1
abbv	abbv	1	1
abil	abil	3	3
abl	abl	8	8
abomi	abomi	1	1
abort	abort	1	1
absolut	absolut	3	3
abt	abt	1	1
abus	abus	3	3
acc	acc	1	1
accaduto	accaduto	1	1
accentur	accentur	2	2
accept	accept	7	6

Figure 36 Results for the First Process For Frequency and Word cloud analysis

So, we can continue the last Process to get the Word Cloud analysis;

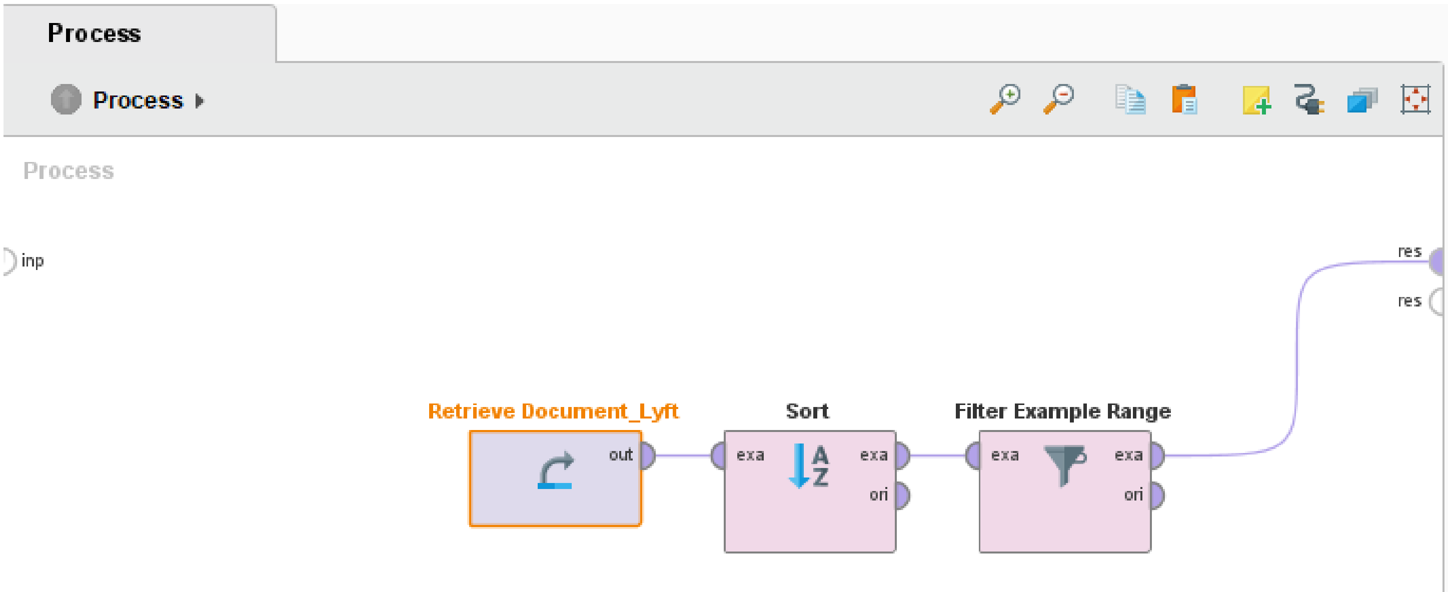


Figure 37 Final Process to get Lyft Word Cloud analysis

And the results would be;

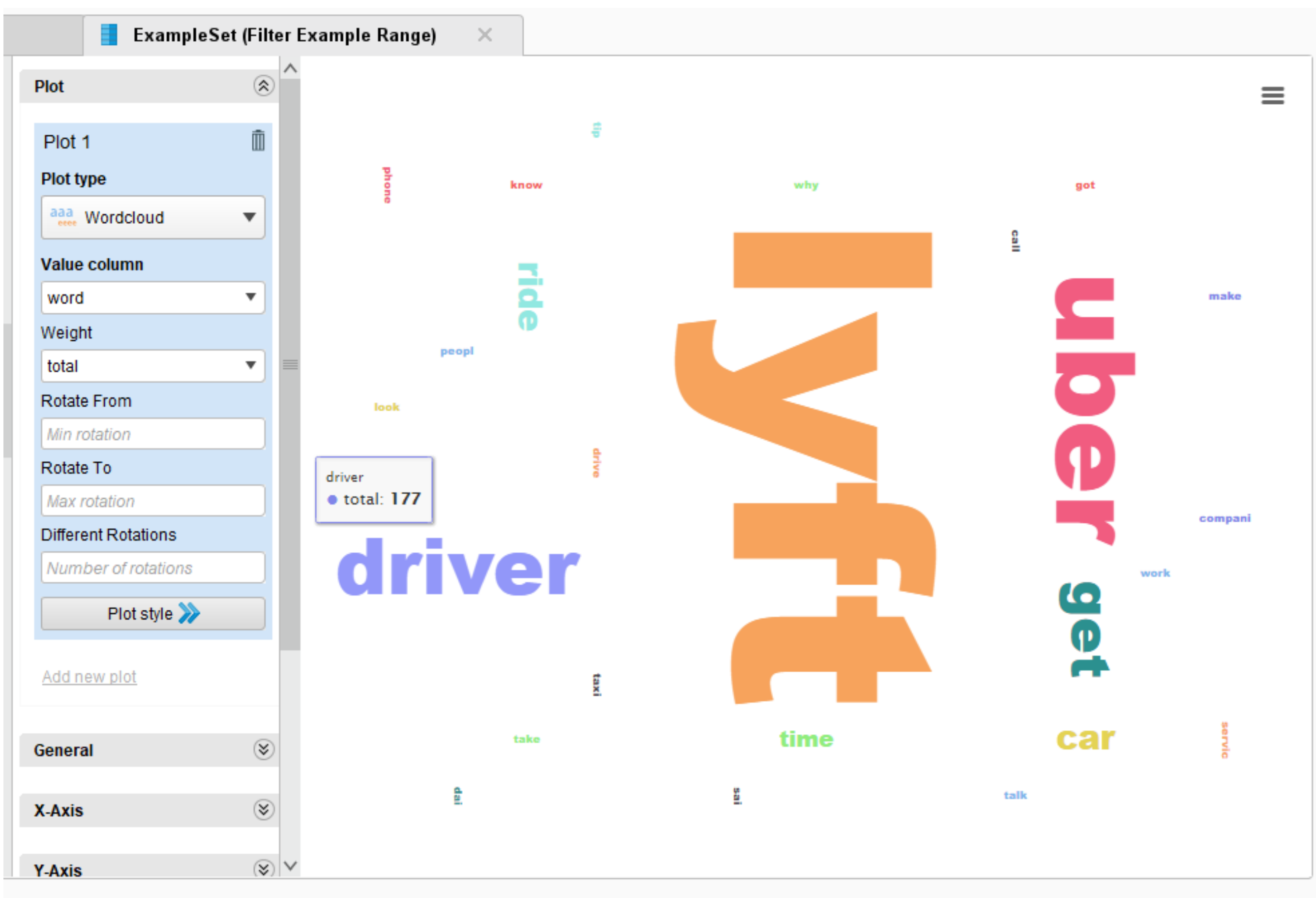
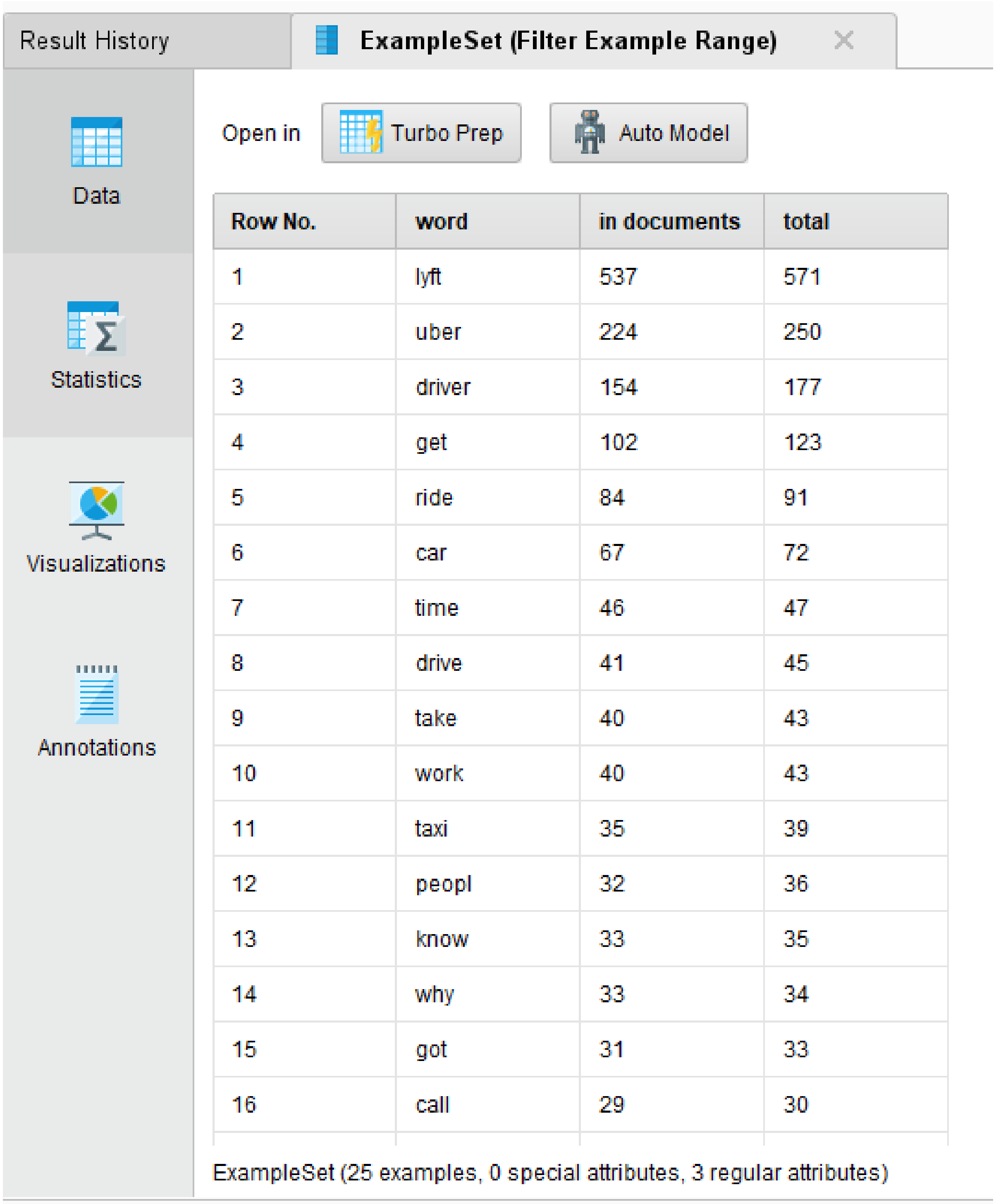


Figure 38 Word cloud Visualization for Lyft brand

But in the data view the results as follows;



Row No.	word	in documents	total
1	lyft	537	571
2	uber	224	250
3	driver	154	177
4	get	102	123
5	ride	84	91
6	car	67	72
7	time	46	47
8	drive	41	45
9	take	40	43
10	work	40	43
11	taxi	35	39
12	peopl	32	36
13	know	33	35
14	why	33	34
15	got	31	33
16	call	29	30

Figure 39 results of the Frequency and Word cloud analysis in the data view

4.0 Findings

We started with raw data that will be used in this paper as 1,000 raw data for Uber brand and also another 1,000 raw data for Lyft brand. But after data cleaning for both brands, Uber brand's raw data becomes 610 and Lyft brand's raw data becomes 737 raw data. After completing the sentiment analysis, that is text mining, two types of results were generated: <u>sentiment overview</u> and <u>word overview</u> for each brand. As far as general overview of perspective from customer towards each brand expressed in tweets are concerned, results are presented in figure shown below:

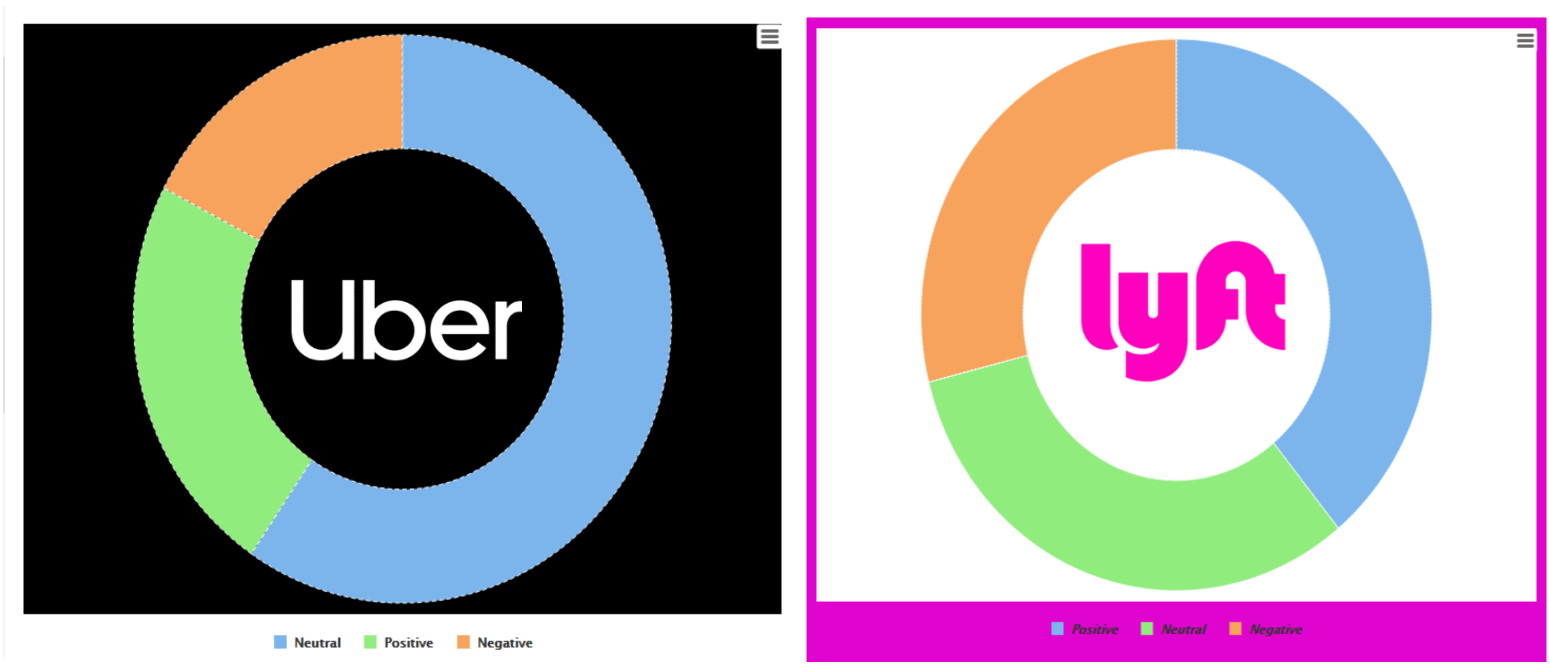


Figure 40 Sentiment analysis: Uber vs. Lyft

The chart above shows the overview of sentiment that be given by the users of Twitter about

both brands Uber and Lyft. On the left side of *Figure 17* show the pie chart for Uber while in other hand is pie chart for Lyft. From data that complete the sentiment analysis for Uber, the highest sentiment towards Uber is Neutral with (60%), while having Positive sentiment withscoring (23 %) and (17%) Negative.

On the other hand, For Lyft from the total of 737 data after cleaning process, the highest sentiment towards this brand is Positive with scoring 288 (39%), Neutral as much as 236 (32%) and Negative is 213 (29%) from the total data.

We can see the difference Between the most common Positive and Negative sentiment For Uber For example as Follows;

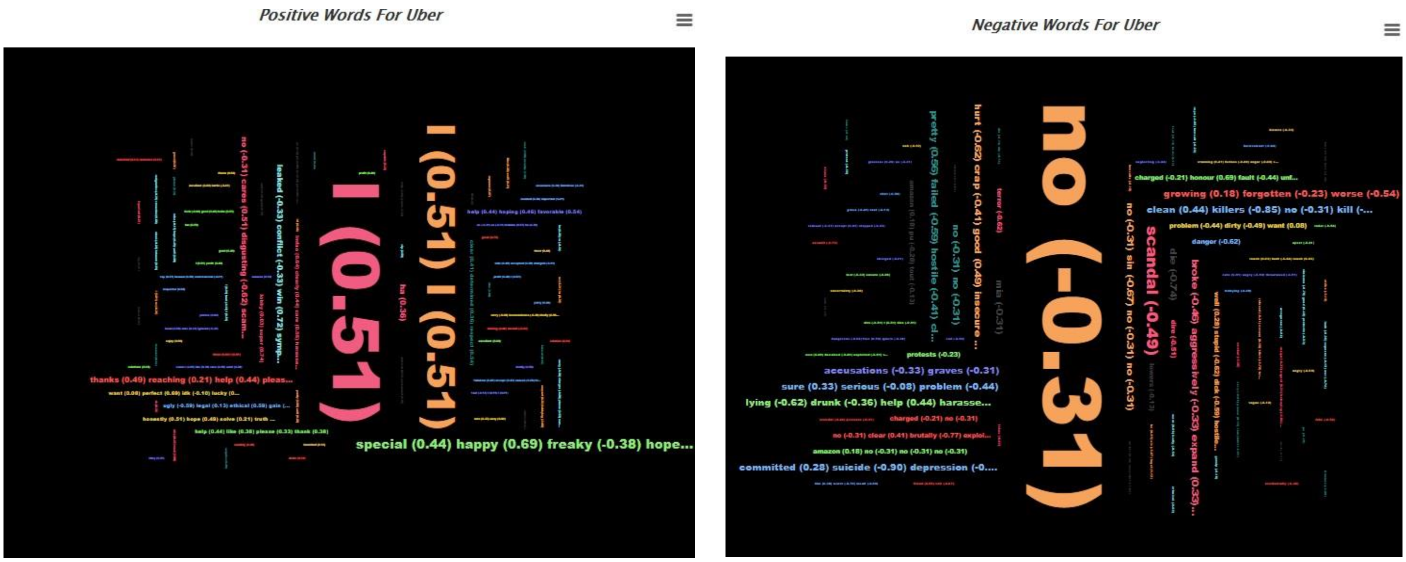


Figure 41 Positive vs. Negative Words For Uber

And The Same for Lyft Brand;

Figure 42 Positive vs. Negative Words For Lyft Brand

Both the figure above is showing the word cloud for the Uber and Lyft brands. In the left side, it is shown the positive sentiment with the scoring that can be said they are the words most frequently used for expressing the Positive talks on twitter, but the other side is vice versa.

These Word most used for Uber and Lyft Brands can be shown in the Following Tables;

Positive		Negative	
Word	**Total Occurrence**	**Word**	**Total Occurrence**
Good	71	No	62
Ha-ha	71	Killers	61
Rewarded	69	Pay	61
profit	63	hurt	58
lucky	63	Crap	58
Mercy	61	Insecure	58
pls	61	Block	58
brave	59	Broke	57
Nice	57	aggressively	57
Safe	56	Lying	56

Table 1 Positive vs. Negative Words Most expressed in Uber

And the same for Lyft as follows in the following Table;

Positive		Negative	
Word	**Occurrence**	**Word**	**Occurrence**
Great	90	Thieves	71
Glad	74	Charge	71
Friends	74	hell	71
Alert	74	cancel	71
Safe	74	Hate	66
Like	69	Wrong	66
Win	67	Significant	66
Help	67	Pressure	63
Well	67	Shit	62
Peace	67	No	62

Table 2 Positive vs. Negative Words Most expressed in Lyft

5.0 Discussion

comparing between the two major brands which are Uber and Lyft. We can answer two research question of this study which are:

Q1: What is the perception of user; is that positive, negative, or neutral?

Q2: Is that the negative element by user based towards the product of the brand or towards the

brand itself?

As a whole, We can find that the two brands are very related and we can conclude that Lyft brand has The most positive sentiment instead this brand is on Some states of USA and Canada while Uber is considered a worldwide brand. This Comparison can be shown as follows;

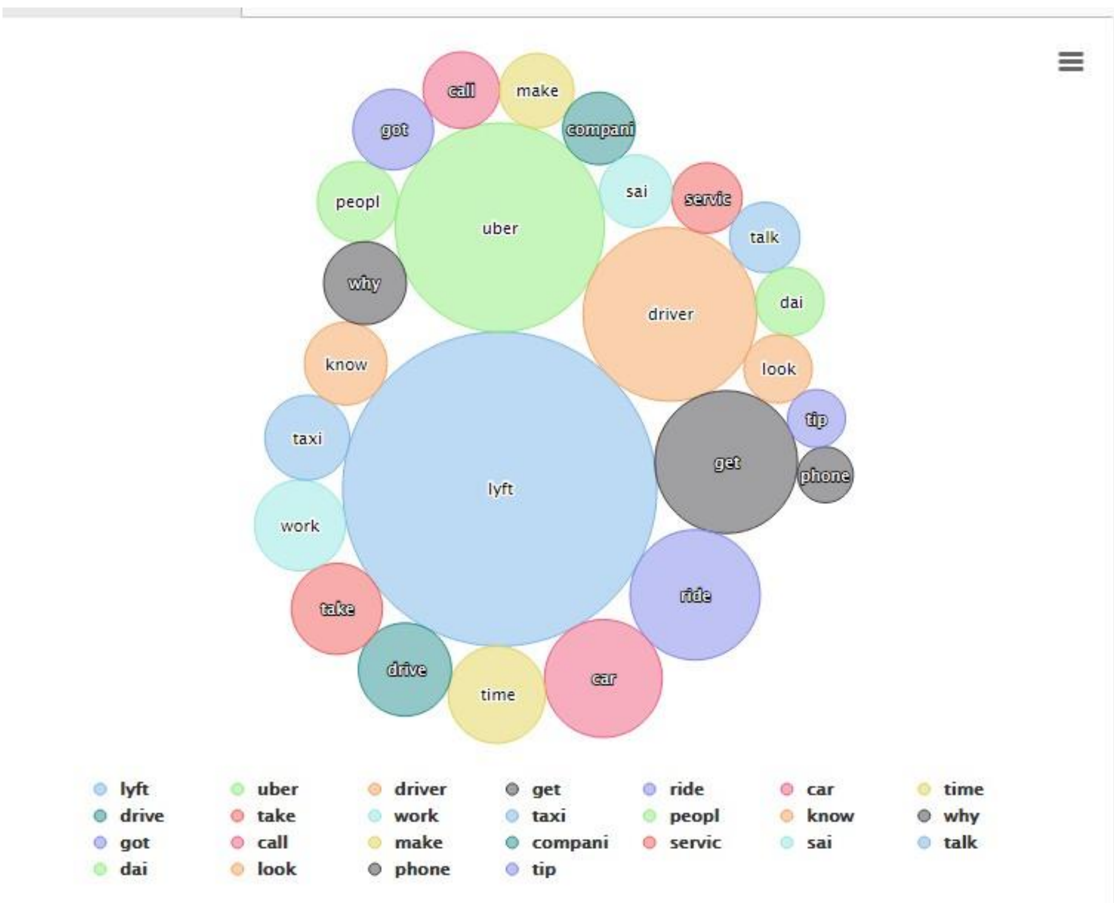


Figure 43 Comparison between Uber and Lyft and other Key words

5.1 Uber Sentiment

Now, We can answer the first research question which is What is the perception of user; is that positive, negative or neutral?

The answer can be summarized in the following graph;

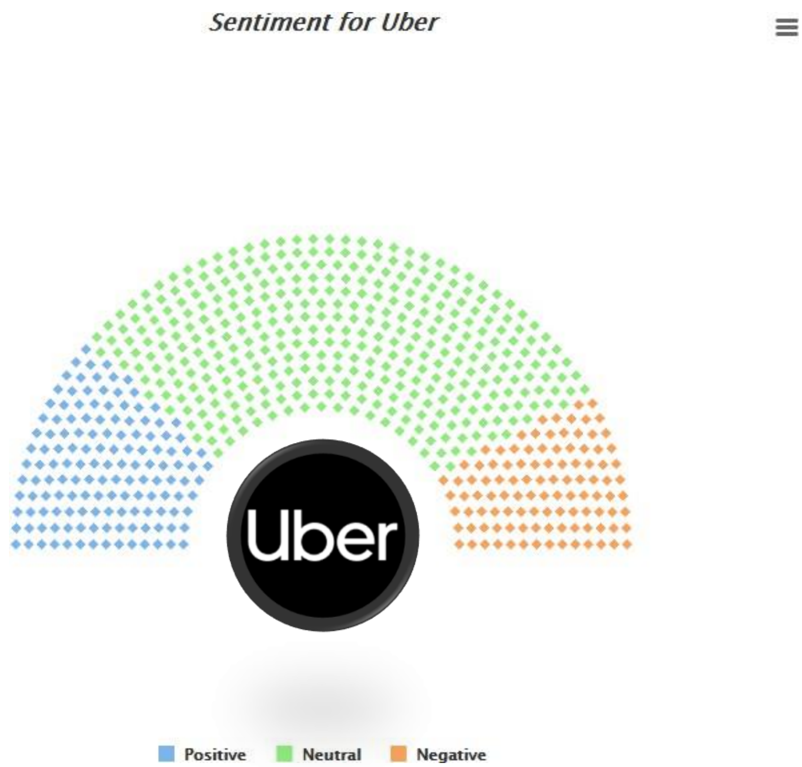


Figure 44 Sentiment analysis for Uber

As shown from this figure, the highest sentiment for Uber is Neutral, then Positive and Finally Negative sentiment.

5.2 Lyft Sentiment

Here We can also answer the first research question which is What is the perception of user; is that positive, negative or neutral?

The answer can be summarized in the following graph;

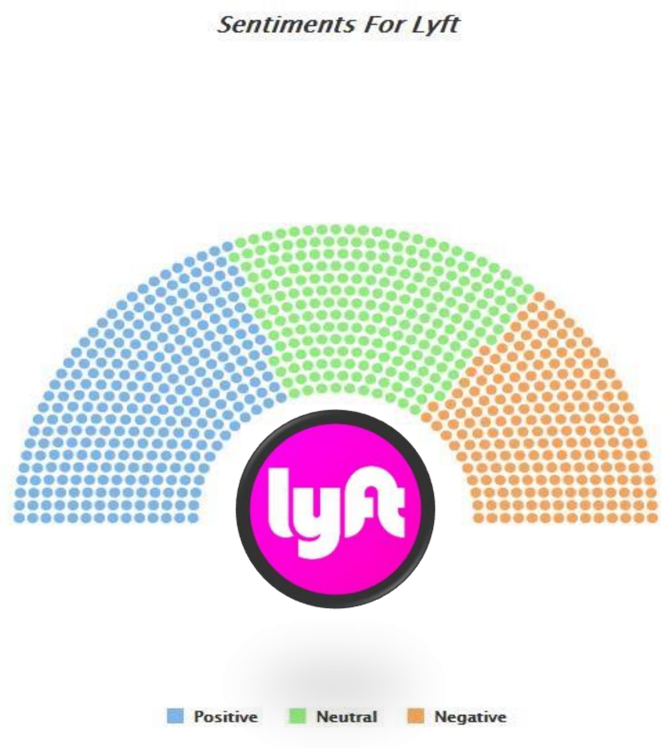


Figure 45 Sentiment Analysis for Lyft

But Here for Lyft, the Most sentiment is Positive, then Neutral to finally Negative Sentiment.

5.3 Limitation

RapidMiner consider one of the most popular data mining tools it's easy to use even if you don't have a programming skills ,java was used to write Rapid Miner and it include more than 500 operators with different approaches to point out connection in the data and you can choose from many data mining , text mining, web mining beside sentiment analysis despite its many advantages it has many limitations as it has low ability to train and test the dataset. Rapid Miner supports various machine learning algorithms like various logic programming algorithms and extremely randomized trees. But it has some limitations because it's difficult to use for a group of data scientist as it lacks a way of collaboration. Beside the visualization offered has limited ways of being modified (this compared with other tools) which conflicts with the data and knowledge needed to be visualized.

Rapid Miner supports various files format inputs and outputs like ASCII, csv, dat, dbms, xml, sap, pdf, weka, and images. But while doing analysis we face a problem Most of the time while dealing with big data, like having large number of examples and attributes, it takes a lot of time. The cumulative time increase when the user is optimizing manually different attributes based on the results. And it becomes difficult to manage hundreds of models available. Some versions of RapidMiner have Limitations! Even with the student version there is a limit of 10,000 rows of output, so when we try to do an analysis on a 15,000-point dataset it takes a very long time to process it, besides 2000 points will randomly be omitted.

Besides the previous limitations, there is some other limitations we face during the analysis, RapidMiner is about sentiment analysis, and most of its process is related to the Lexicon-based Approach particularly in using Vader. And contrast to analysis sentences, Vader is used to analyzing words thus its limitation her is processing the sarcastic sentences which it means that the words that be used not mean as what it is. So, when users use sentiment like “I feel like a God, people just looking for me when they need me’’ it mean that people only talk to the user when they need something. Thus, this sentiment cannot be analysis by Vader approach because some ofthe word such as ‘God’ and ‘need’ will have high score in the Vader approach.

And when we do our visualizations graphs, we notice that it has some limitations about the scoring words where analysis of the words that can be shown is limited to 500 only which mean that large dataset and projects with high volume cannot be processed with RapidMiner.

furthermore, we try during the project to extract data in Arabic for Uber and Lyft, but RapidMiner have limitation of language that can be extracted, tokenize, or even to process using Vader , although RapidMiner The user can request that RapidMiner add up some other language by giving them some dictionary information, according to their website, but it still takes a while for certain languages to be extracted using RapidMiner .

RapidMiner based on Lexicon-based Approach and that mean it cannot teach the tools to understand certain words directly, in certain languages there are terms that can be with difficult meaning, or one word could have many meanings, and it is difficult for RapidMiner to process because the lack of training process compared to the Machine Learning Approach.

6.0 Conclusion

Sentiment analysis facilitates a holistic analysis by studying the impact of a piece of text to determine the sentiment behind it. By using natural language processing, text analysis, statistics, and a series of algorithms to extract and identify the emotions of words into positive, negative, and neutral categories, which is an easy way to use and reduce cost and time.

We can collect data from Twitter or other social media networks or online sources.

We perform analysis with a specific tool called "RapidMiner" by analyzing several tools and processes.

From the results and analysis, the negative sentiment can help the company know their position well and also get more information about public opinion towards their services and company. They can also do analysis to get the customer feelings and feedback from both Uber and Lyft. From the analysis, we get that in Uber the highest sentiment is neutral sentiment, then positive and finally negative sentiment. and in Lyft sentiment analysis, the positive opinions are high, followed by neutral and the negative ones.

So, the Uber company should focus on improving their service, getting raid in on time, reducing trips' cost, and most importantly, customer safety.

References

Christensen, P., & Osman, A. (2021). The Demand for Mobility: Evidence from an Experiment with Uber Riders.

Alemi, F., Circella, G., Mokhtarian, P., & Handy, S. (2019). What drives the use of ridehailing in California? Ordered probit models of the usage frequency of Uber and Lyft. Transportation Research Part C: Emerging Technologies, 102, 233-248.

Wan, W. N. A. A. B., Mohamad, A. F. M. F., Shahib, N. S., Azmi, A., Kamal, S. B. M., & Abdullah, D. (2016). A framework of customer's intention to use Uber service in tourism destination. International Academic Research Journal of Business and Technology, 2(2), 102-106.

Mohamed, M. J., Rye, T., & Fonzone, A. (2020). The utilisation and user characteristics of Uber services in London. Transportation Planning and Technology, 43(4), 424-441.

Shokoohyar, S. (2018). Ride-sharing platforms from drivers' perspective: Evidence from Uber and Lyft drivers. International Journal of Data and Network Science, 2(4), 89-98.

Kamais, C. E. (2019). Emerging security risks of e-hail transport services: Focus on Uber taxi in Nairobi, Kenya. International Journal of Security, Privacy and Trust Management (IJSPTM) Vol, 8.

Grilo, I., Shy, O., & Thisse, J. F. (2001). Price competition when consumer behavior is characterized by conformity or vanity. Journal of public economics, 80(3), 385-408.

Lu, A., Frazier, P., & Kislev, O. (2018). Surge Pricing Moves Uber's Driver Partners. Available at SSRN 3180246.

https://www.investopedia.com/articles/investing/110614/taxi-industry-pros-cons-uber-and-other-ehail-apps.asp#citation-13

https://www.geeksforgeeks.org/what-is-sentiment-analysis/

https://www.insightsforprofessionals.com/marketing/customer-experience/sentiment-analysis-in-customer-service

https://www.commsights.com/benefits-of-sentiment-analysis-for-businesses/