

**DATA 2206 - 02**

**Project Analysis Report**





US Airlines Twitter Sentiment Analysis

.

**Group Members:**

Saju Thomas Padinjaremanchadikunnel Xavier (100808785)

Roshna Babu (100805012)

Neha Joseph (100827187)

**Contents**

[1. Executive Summary 2](#_Toc89602349)

[2. Problem statement: 2](#_Toc89602350)

[3. Dataset 2](#_Toc89602351)

[4. Analytics solutions and recommendations 2](#_Toc89602352)

[5. Problem description 2](#_Toc89602353)

[6. Data Description 3](#_Toc89602354)

[7. Data preparation details 3](#_Toc89602355)

[8. Data analysis solutions 5](#_Toc89602356)

[9. Main findings 5](#_Toc89602357)

[10. Conclusion 5](#_Toc89602358)

[11. Bibliography 6](#_Toc89602359)

[12. Appendix 6](#_Toc89602360)

# **Executive Summary**

In recent years, Twitter has become the important online customer service platform. As a result, a company's image on Twitter is critical, and this is especially true for airlines, given that many tweets are about travel. Responding to tweets, in fact, has been proved to generate income, drive higher satisfaction than other customer care channels, and, probably most crucially, satisfied Twitter users spread the word. In this project, we use tweets gathered from Twitter to learn about people’s flight experiences and give airline companies suggestions on how to make their trip more enjoyable.

The data set which we downloaded from Data.world contains about 15,000 tweets, collected from February 2015 on various airline reviews. Every review is labelled as either positive, negative or neutral. First, we want to build a model to perform sentiment analysis on the data set. Second, more interestingly, we want to assign a reason to each negative response, such as late flight, lost luggage, etc. Our goal is to assign a label to this unspecified group. By knowing every review’s negative reason, we can give specific suggestions to different airline companies on how to improve their service.

# **Problem statement:**

In current data set extracted from Data. World, we have tweets for 6 US airlines. We need to analyse how travellers in February 2015 expressed their feelings on Twitter i.e., whether the tweets are positive, negative or neutral. This work can be useful for the airline company to understand what are the problems to working on.

# **Dataset**

The dataset was obtained from [Home | data.world](https://data.world/). The dataset contains details about the tweets shared by customers about different US airlines which are, Virgin America, United airlines, US airways, American airways, South west airways, Delta airways. The data set contains data from the year 2015 and it also has people’s opinion and sentiments related to respective airlines.

# **Analytics solutions and recommendations**

From the twitter analysis of Us airlines we found that among the responses more than 60% of the tweets were negative in which United and US airways have the higher proportion of negative comments. The main reason for negative comments is customer service issues and late flight.

The goals of this project are:

1) do sentiment and opinion mining for six major U.S. airlines that performs on customer reviews so that the airlines can have fast and concise feedback

2) Make recommendations on the most important aspect of services they could improve given customers complains.

# **Problem description**

The digital connectivity bestows immense power to the customers in terms of vocalizing their thoughts, opinions, and reviews on a brand. The customer views expressed on Twitter, Facebook, and other online forums are forming the base of customer strategy for brands worldwide. Airlines can use Sentiment analysis or Opinion Mining of the posts shared on various social media platform to gain user insights and strategize their brands

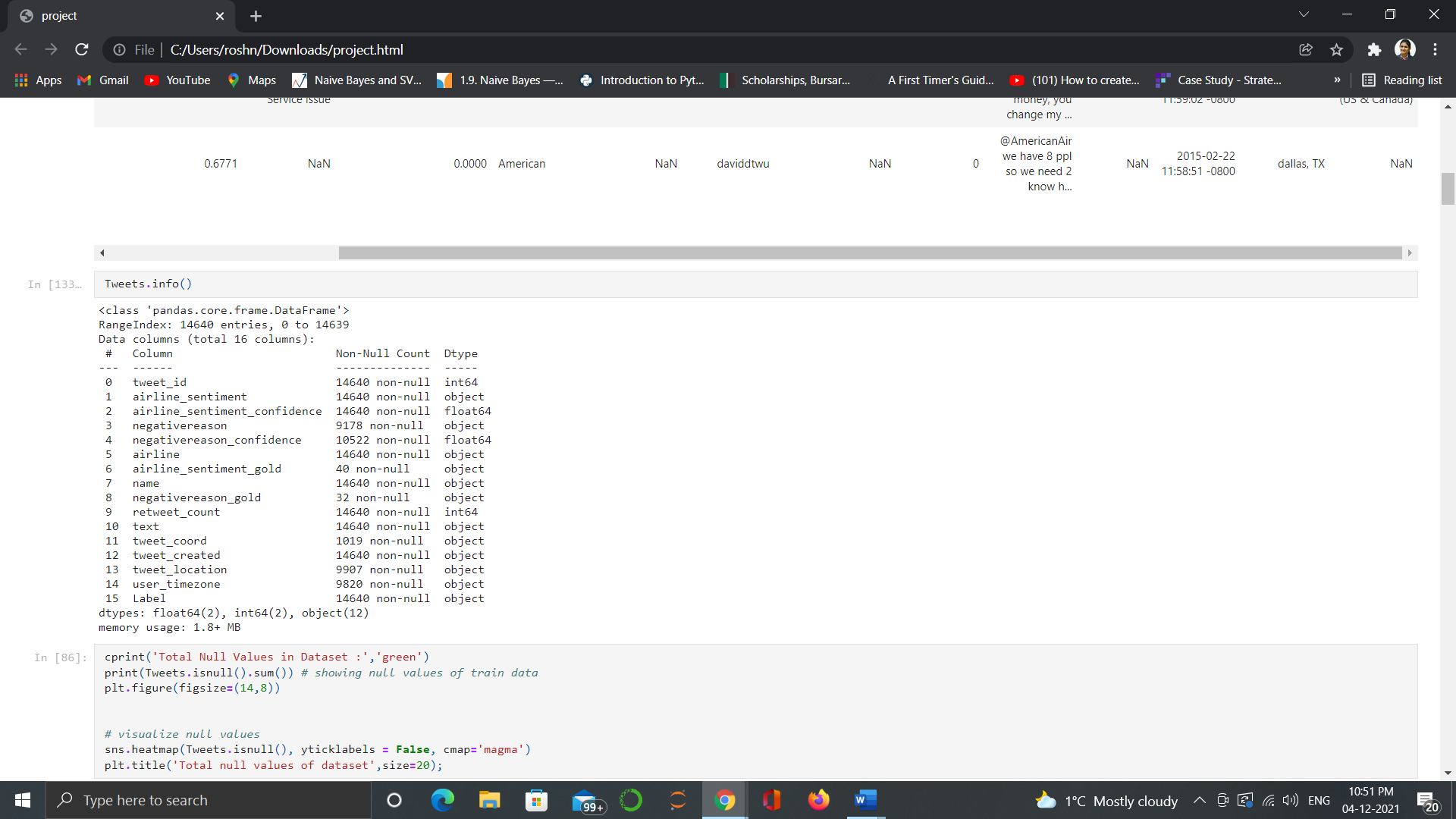
The current data set (which can be found on Data. World, we have tweets for 6 US airlines. We need to analyse how travellers in February 2015 expressed their feelings on Twitter i.e., whether the tweets are positive, negative or neutral.

# **Data Description**

The data set contains about 15,000 tweets, collected from February 2015 on various airline reviews. Every review is labelled as either positive, negative or neutral. First, we want to build a model to perform sentiment analysis on the data set. Second, more interestingly, we want to assign a reason to each negative response, such as late flight, lost luggage, etc.

The sentiment analysis labels are positive (20%), negative (60%), and neutral (20%). The negative reason labels are bad flight (7.45%), cancelled flight (9.62%), customer services issues (39.77%), damaged luggage (0.84%), flight attendant complaints (6.05%), flight booking problems (6.19%), late flights (1.99%), long lines (19.97%), and lost luggage (8.23%).

The below image shows the list of input and output variables.



# **Data preparation details**

In the pre-processing step, non-English word, symbols and website links are eliminated. Then the whole data set is randomly separated into training set (10000 samples, 70%) and test set (4640 samples, 30%).

**Key charts**

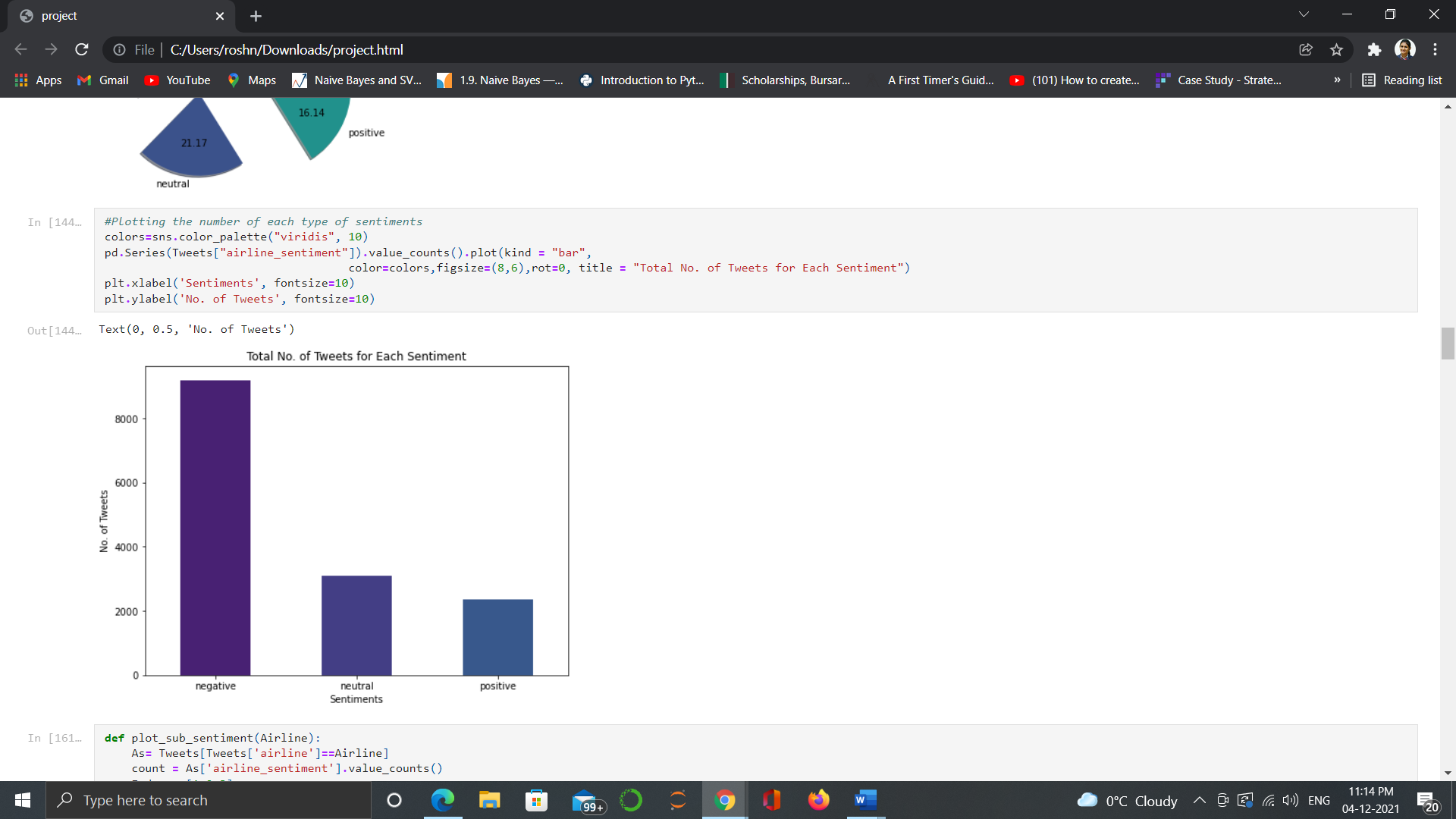
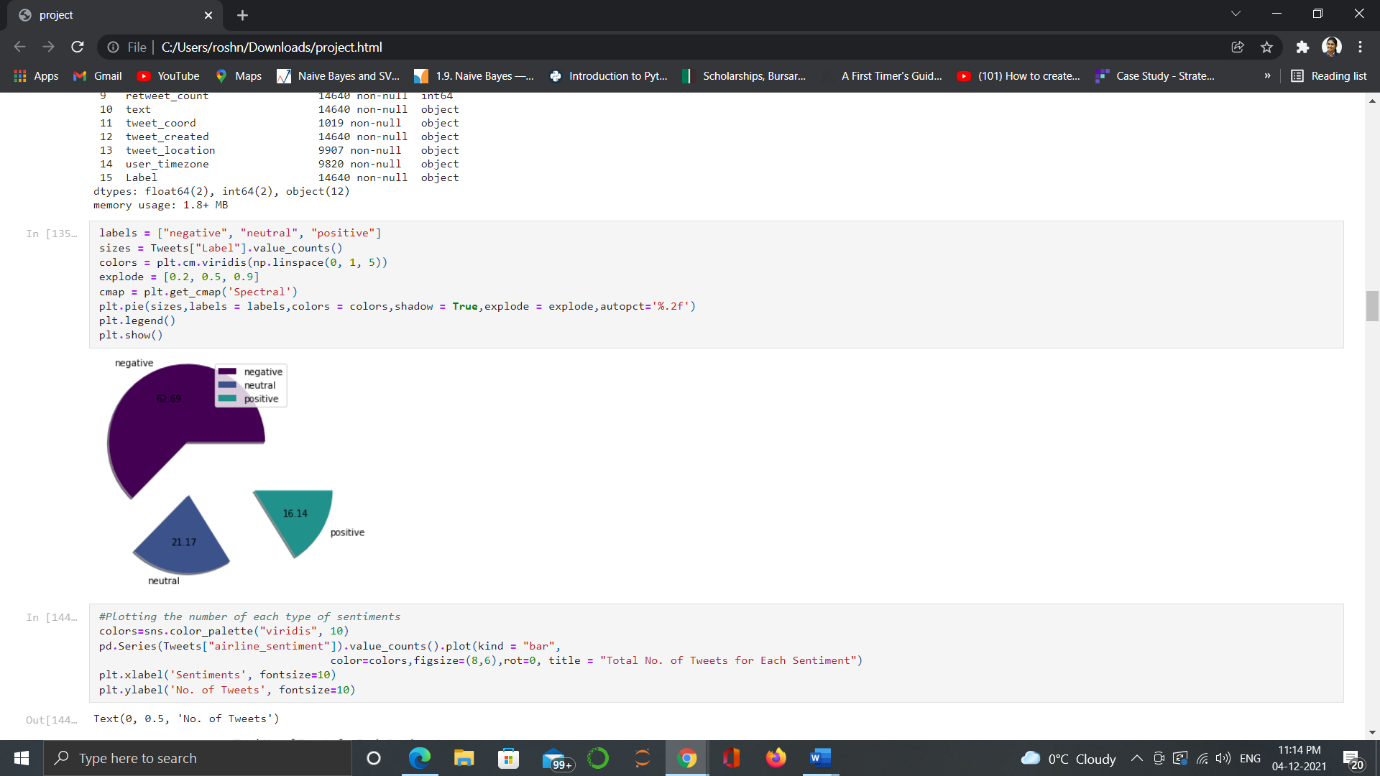


Chart 1: percentage of sentiments chart 2: total no. of tweets for each sentiment

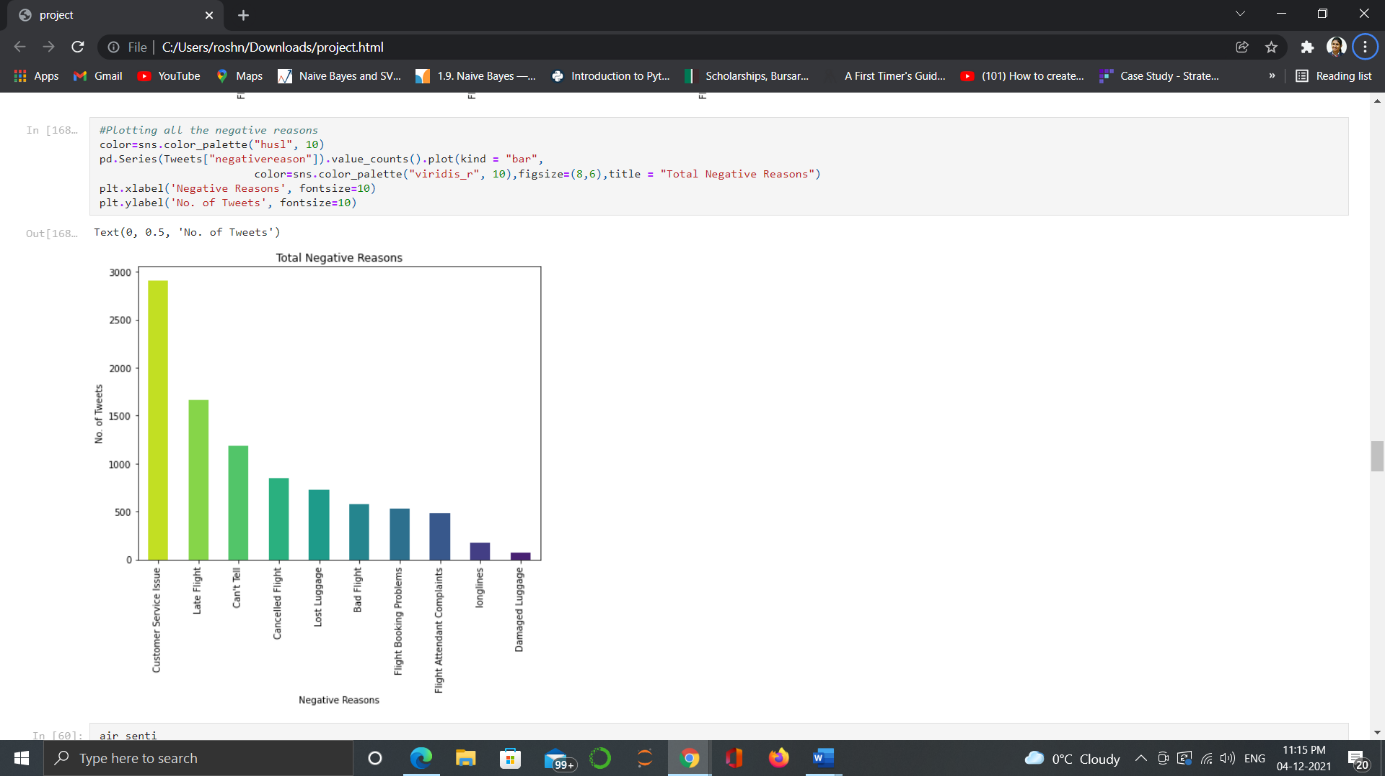


Chart 3: Total negative responses

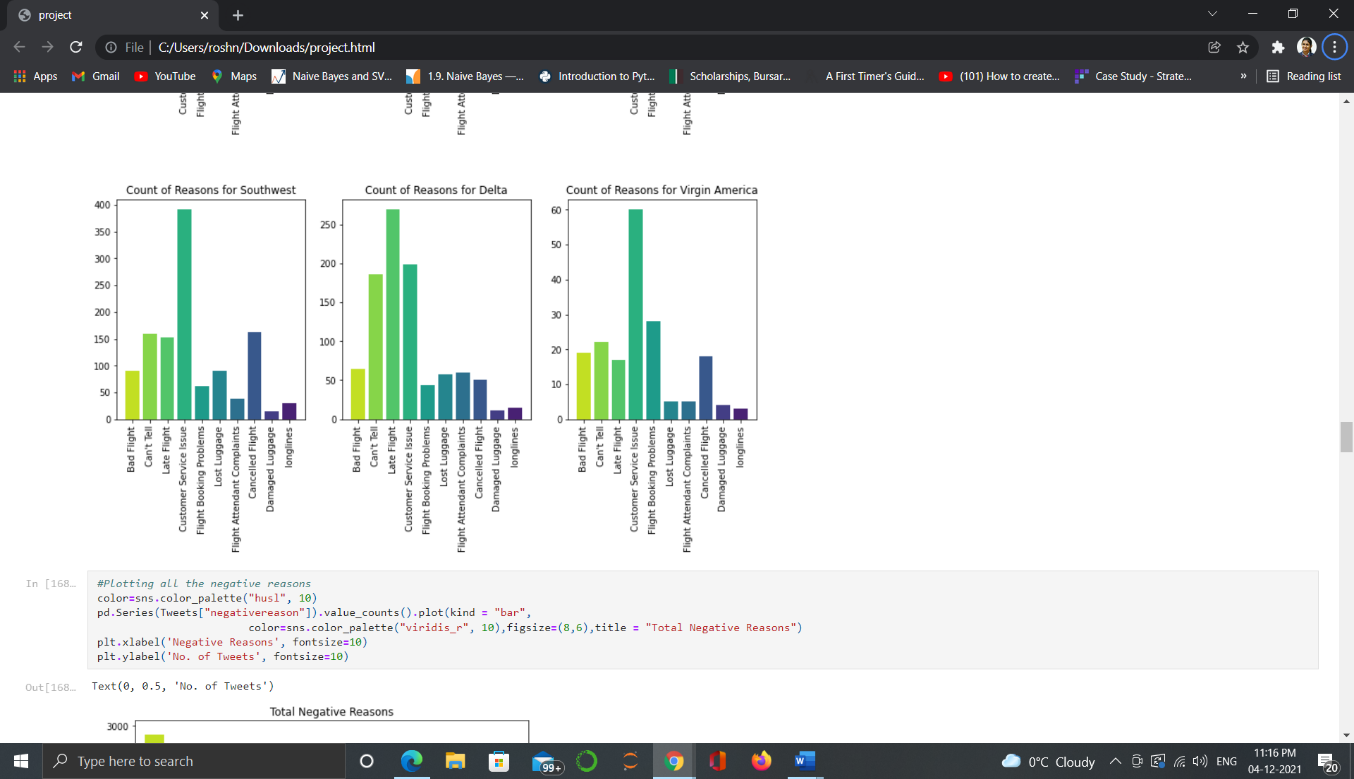
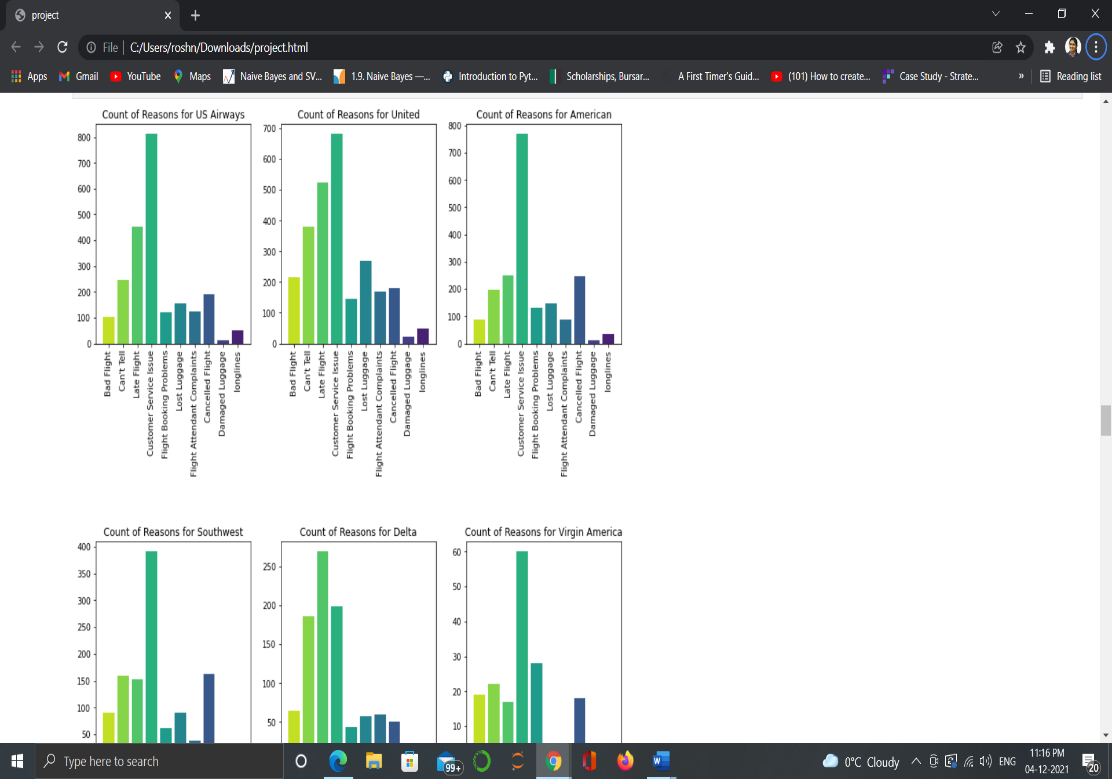


Chart 4: count of responses for each airline

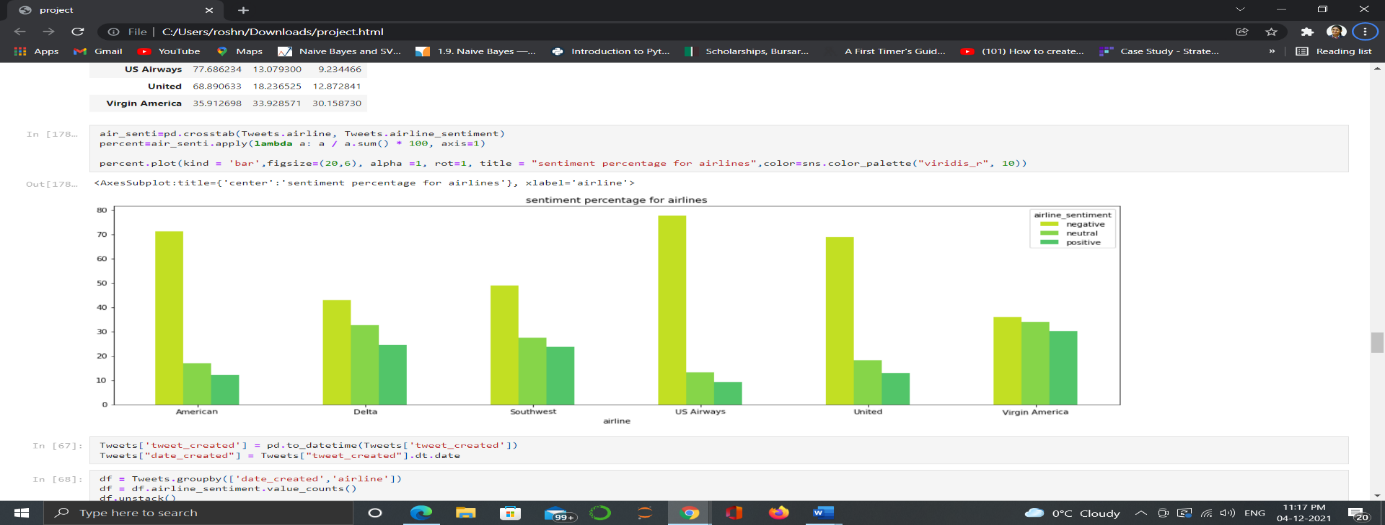
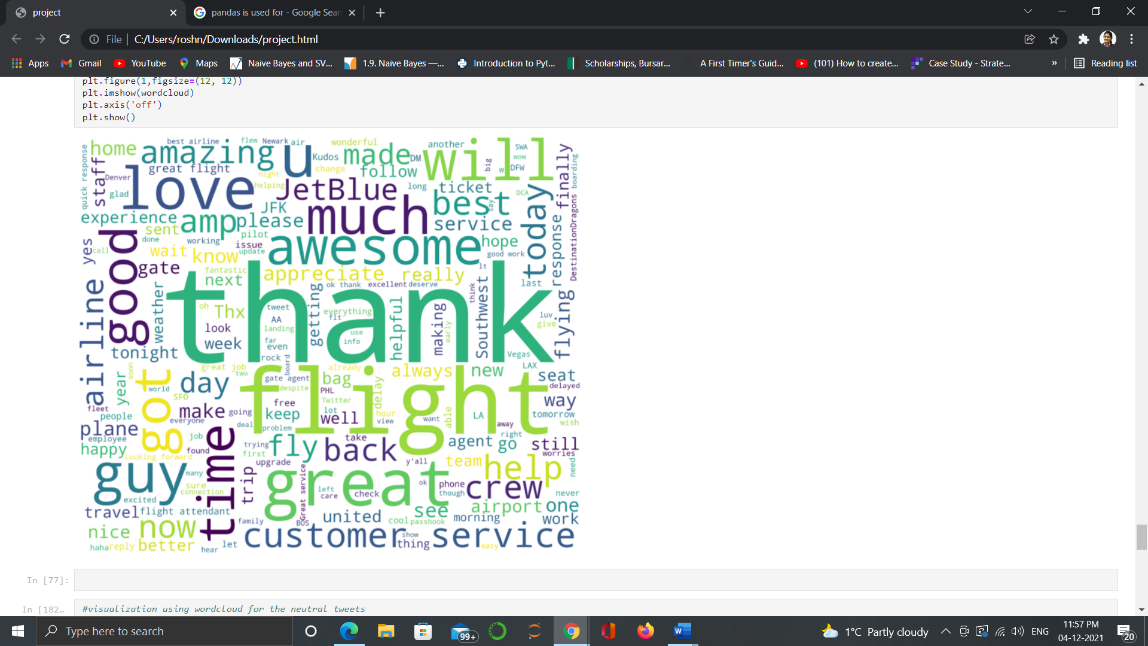
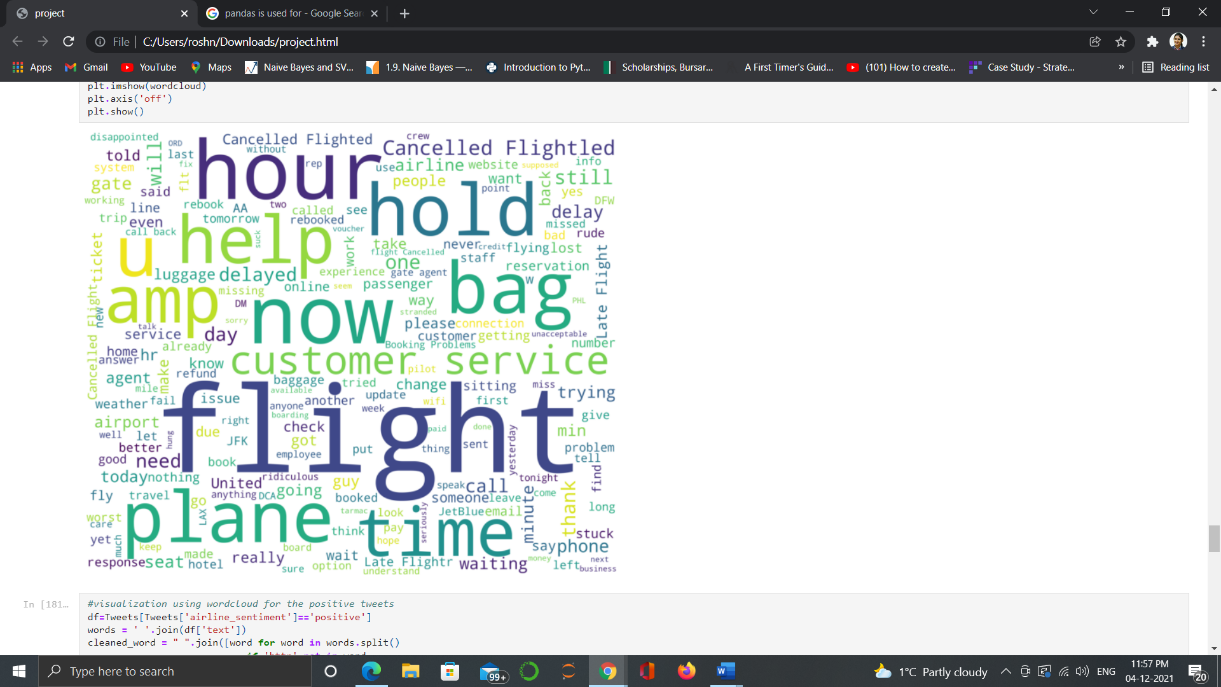


Chart 5: sentiment percentage for each airline

# **Data analysis solutions**

* For this work Python with: Pandas, numpy, seaborn, Matplotlib and NLTK is used.
* By using matplotlib library all the graphs are plotted.
* By using numpy library all the tables are created.
* By using NLTK library word cloud is created for most used positive and negative responses.



The meaningful words that can be spotted in the negative sentiments’ word cloud directly are “hold” and “customer service”. On the other hand, the meaningful words that can be spotted in the positive sentiments’ word cloud directly include “thank”, “awesome”, “good” and “great”.

# **Main findings**

* More than 60% of the tweets were negative.
* United had the greatest number of tweets, followed by US airways and American airlines
* United, US airways and American airlines have higher proportions of negative comments. Whereas, Southwest, Delta and Virgin America had lower proportion of negative comments.
* Two main reasons for negative comments were customer service issues and late flights.
* US Airways had the highest proportion of negative tweets due to customer service issues.
* The main reason for negative comments was customer service issue.

# **Conclusion**

Twitter reviews are much more straight forward, and thus most of the sentiments are expressed directly at the word level. It is possible to judge a twitter airline review’s sentiment only by identifying positive words in a review. Therefore, given the nature of our data set, the task can be solved at bag-of-word level well.

**Advantages**

* Companies can use sentiment extremity and opinion point acknowledgment to pick up a more profound comprehension and the general extent of estimations.
* These experiences can progress focused insight, enhance client benefit, accomplish better brand picture, and upgrade competitiveness.
* In aircraft service industry, it is hard to gather information about clients' input by polls, yet Twitter gives a sound information source to them to do client opinion examination.

**Limitations**

* The dataset which we choose has lots of null values hence we cannot rely on the dataset fully.
* There are several words which come quite often in tweet but those do not seem informative. Removing such words from tweet will give more enhanced result.

**Recommendation**

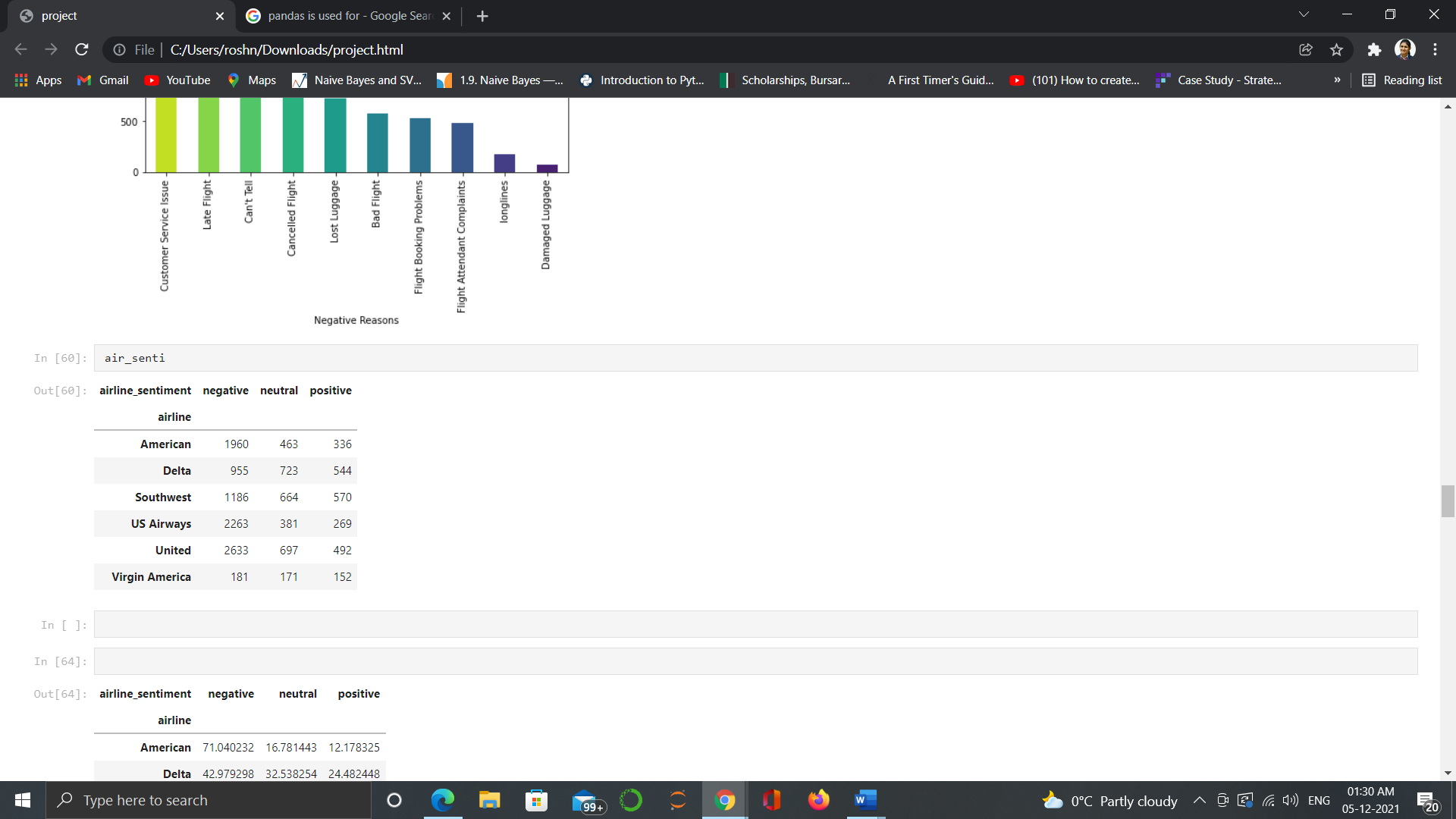
The majority of the customers has a negative review about the US airlines. Most of the people are not happy about the Customer services being provided by the airlines. Hence the airlines have to look forward to make certain changes in the customer services and other facilities and also have to be more cautious about the cancellation and delaying of the flights. It is recommended to do the opinion mining and sentiment analysis every now and then to understand their growth and customer support.

# **Bibliography**

* Dataset - <https://data.world/data-society/twitters-about-us-airline/workspace/file?filename=twitter-airline-sentiment.csv>
* <Bose, B. (n.d.). Twitter sentiment analysis - introduction and Techniques. Digital Vidya. Retrieved December 5, 2021, from https://www.digitalvidya.com/blog/twitter-sentiment-analysis-introduction-and-techniques/>

# **Appendix**

|  |  |  |
| --- | --- | --- |
| **AIRLINE** | **1st negative reason** | **2nd negative reason** |
| **Virgin America** | **Customer Service Issue** | **Flight Booking Problems** |
| **United Airlines** | **Customer Service Issue** | **Late Flight** |
| **Southwest** | **Customer Service Issue** | **Late Flight** |
| **Delta** | **Late Flight** | **Customer Service Issue** |
| **US Airways** | **Customer Service Issue** | **Late Flight** |
| **American Airline** | **Customer Service Issue** | **Late Flight** |

****