Twitter US Airline Sentiment

Shukai Wang
Xi'an Shiyou University

January 16, 2021

Table of content

Problem

Description

Description

Sentiment analysis work on the issues of each major US airline. Twitter counted the tweets related to airlines since February 2015, and analyzed whether the sentiments contained in these tweets were positive, neutral or negative.

- Analyze how travelers in February 2015 expressed their feelings on Twitter.

Data

Basic Information of Data

Table 1: Train

Attribute	Explanation
airline_sentiment	About the attitude of airlines.
text	The text content of the tweet.
airline	The name of the airline.
retweet_count	The number of reposts of the tweet.
tweet_created	The time the tweet was generated.

It also includes attributes such as tweet_id, airline_sentiment_confidence, negativereason, negativereason_confidence, airline_sentiment_gold, name, negativereason_gold, tweet_coord, tweet_location, user_timezone, etc.

Date Analysis

- Positive, negative and neutral emotions accounted for the number and proportion of the total number of people respectively. Among them, the number of negative attitudes is the largest and the number of positive attitudes is the least.

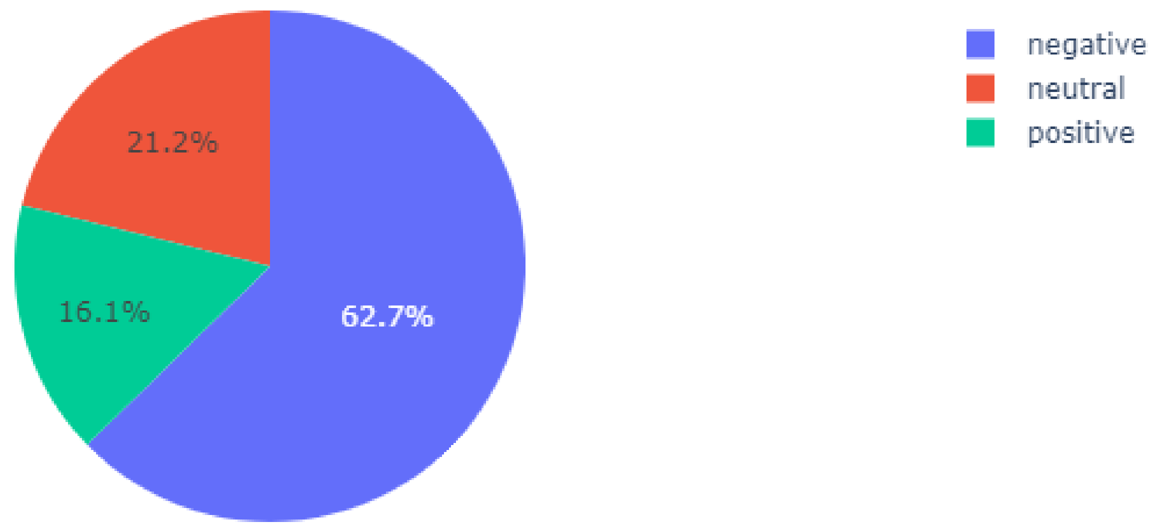

Figure 1: Sentiment Map

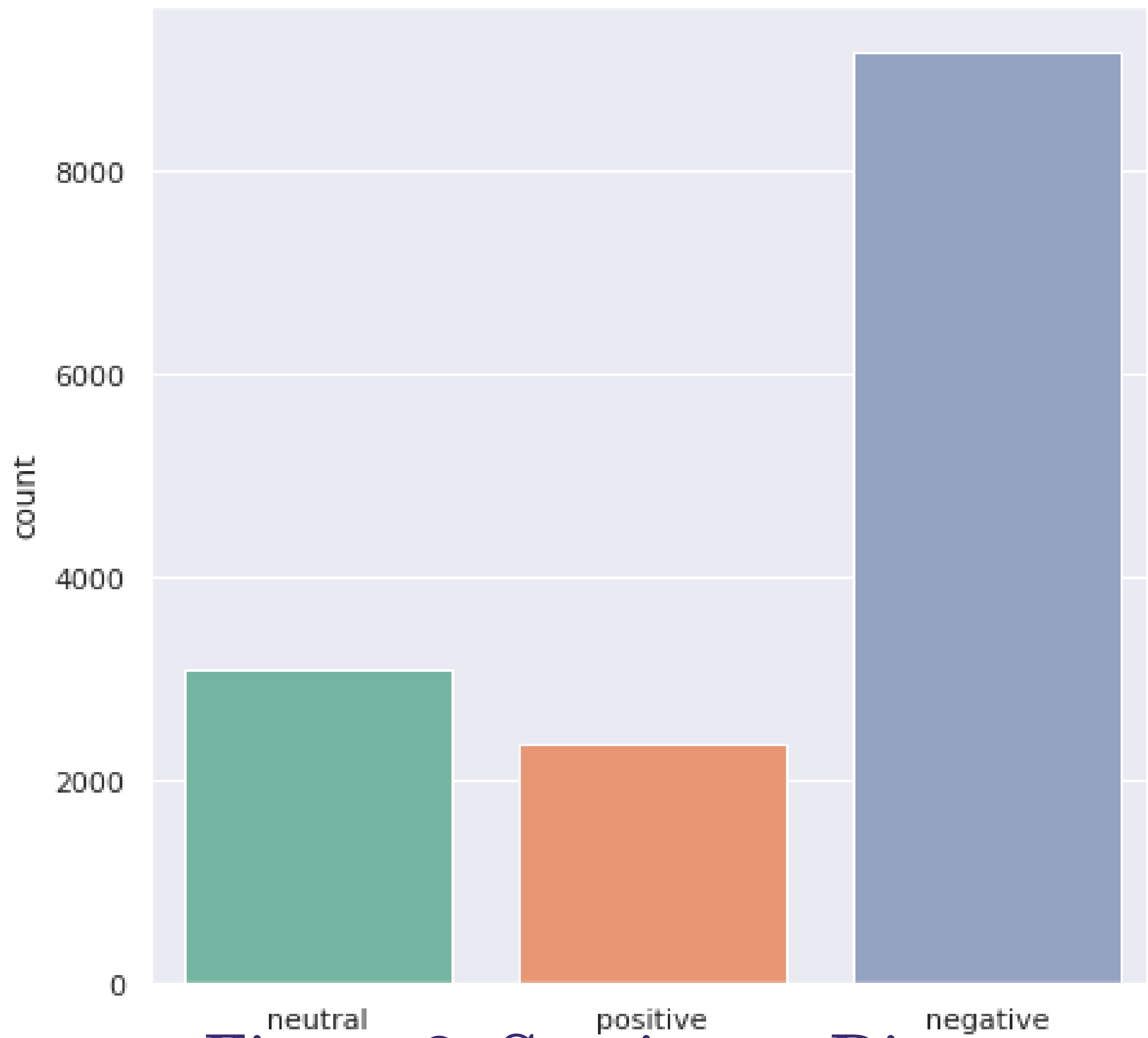

Figure 2: Sentiment Dist

Airline

- The percentage of each airline in all airlines.

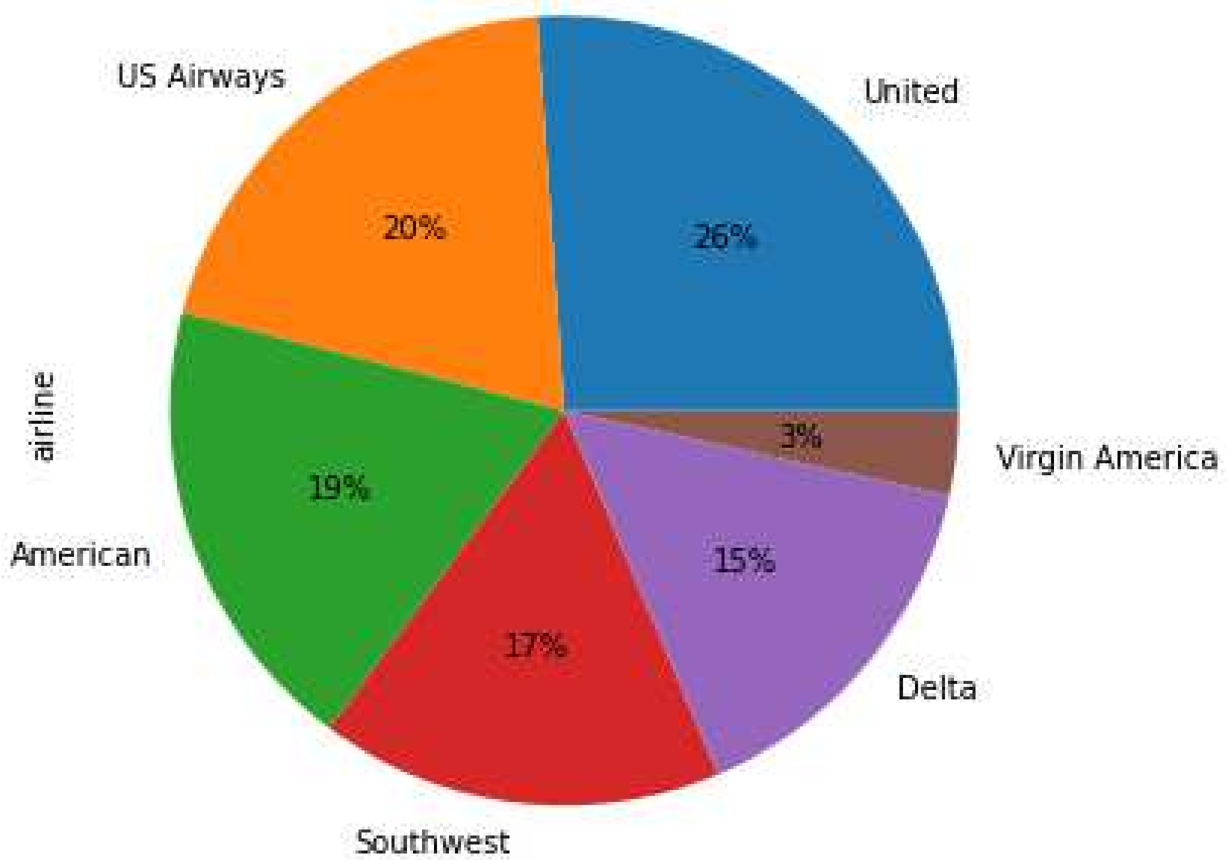

Figure 3: Airline Map

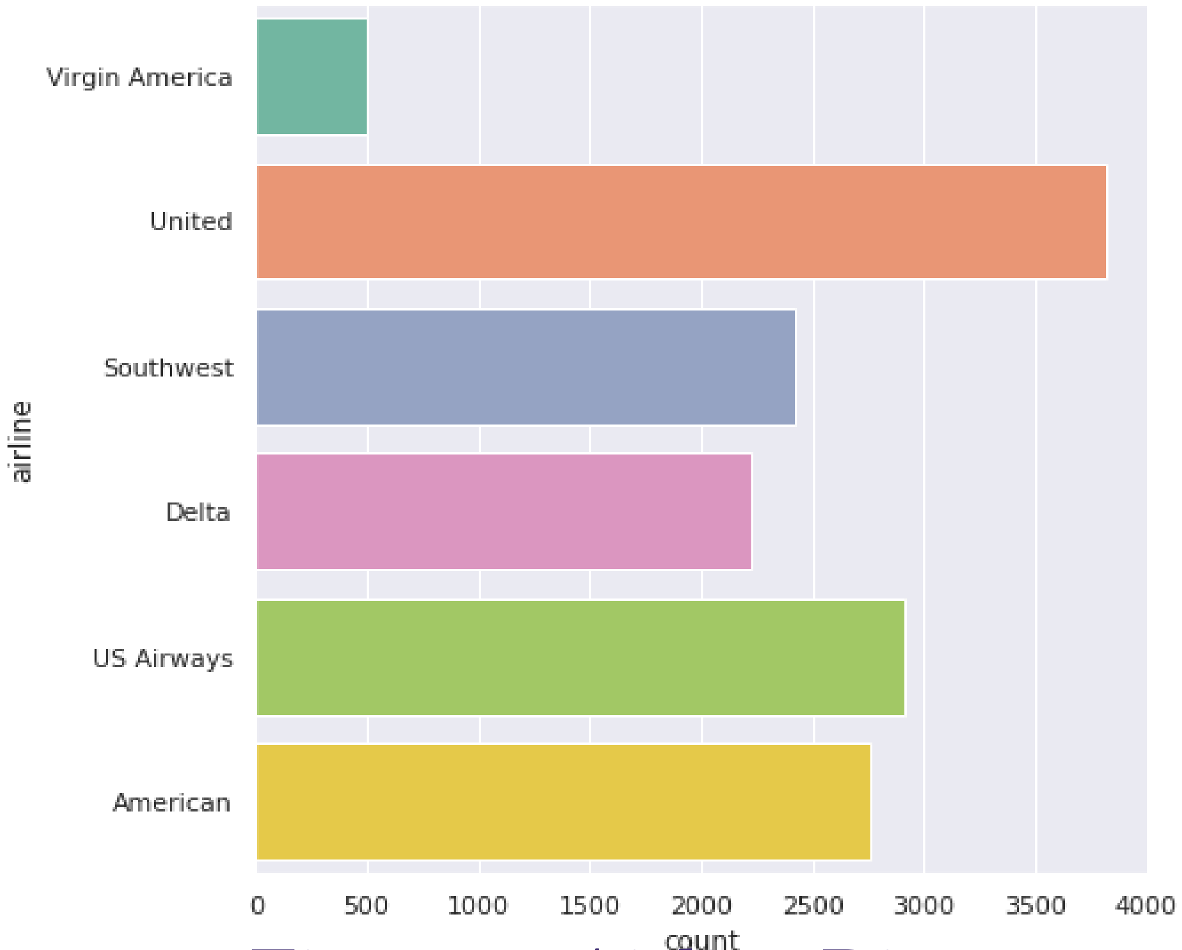

Figure 4: Airline Dist

- Distribution of attitudes of different airlines.

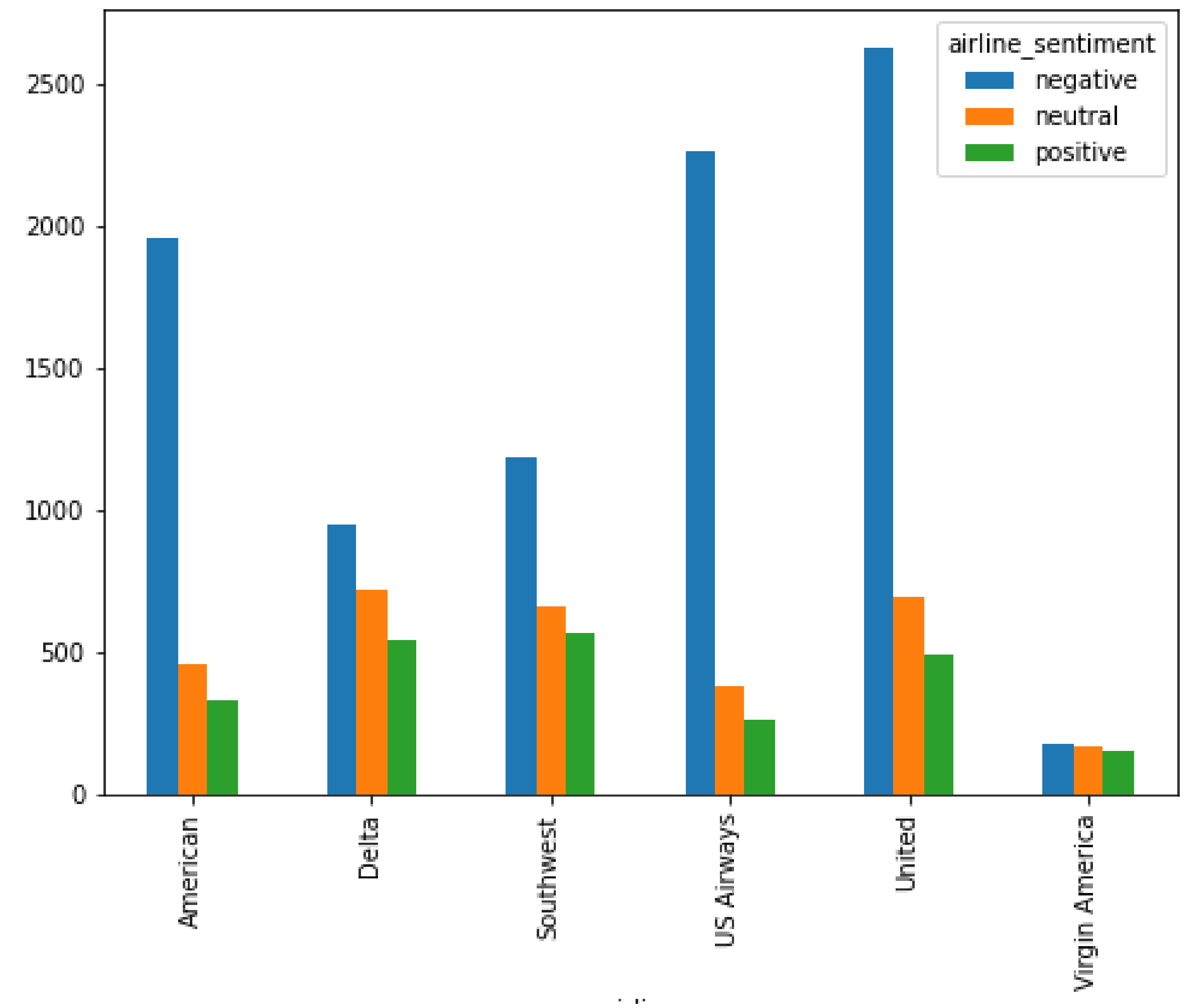

Figure 5: Airline Attitude

- The individual attitude distribution of each airline.

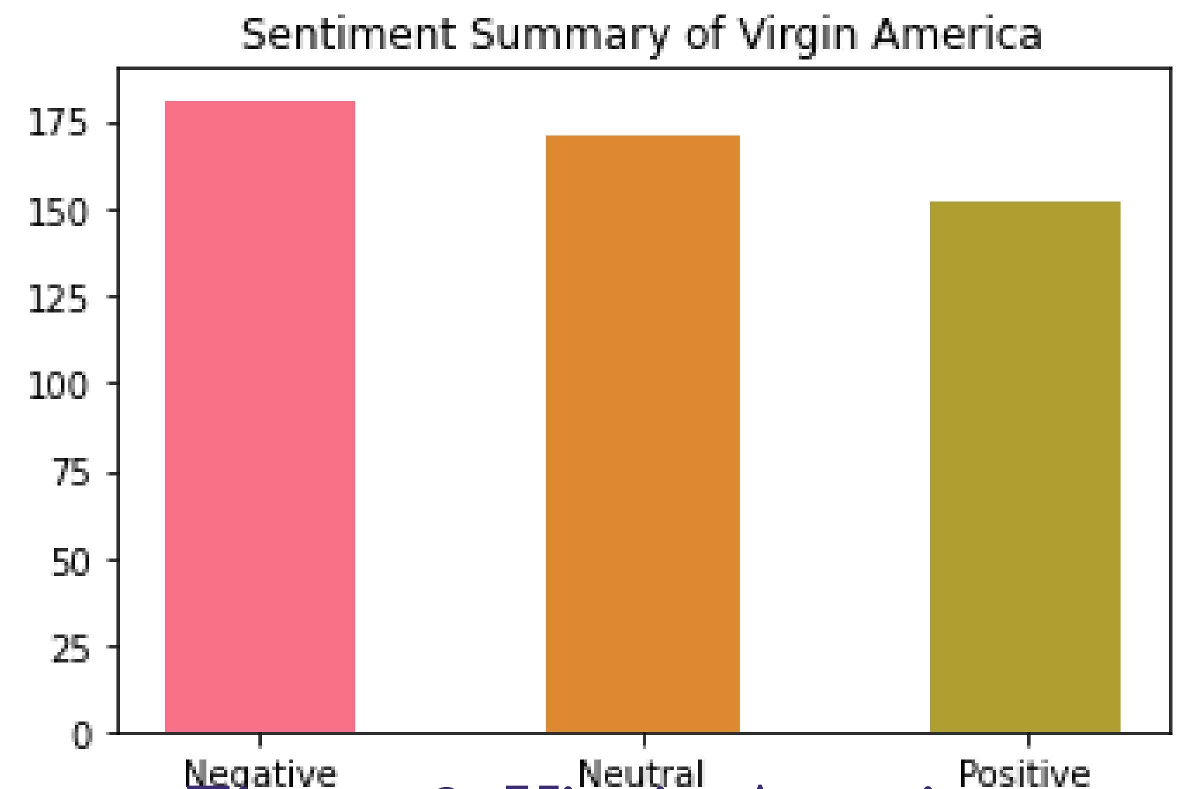

Figure 6: Virgin America

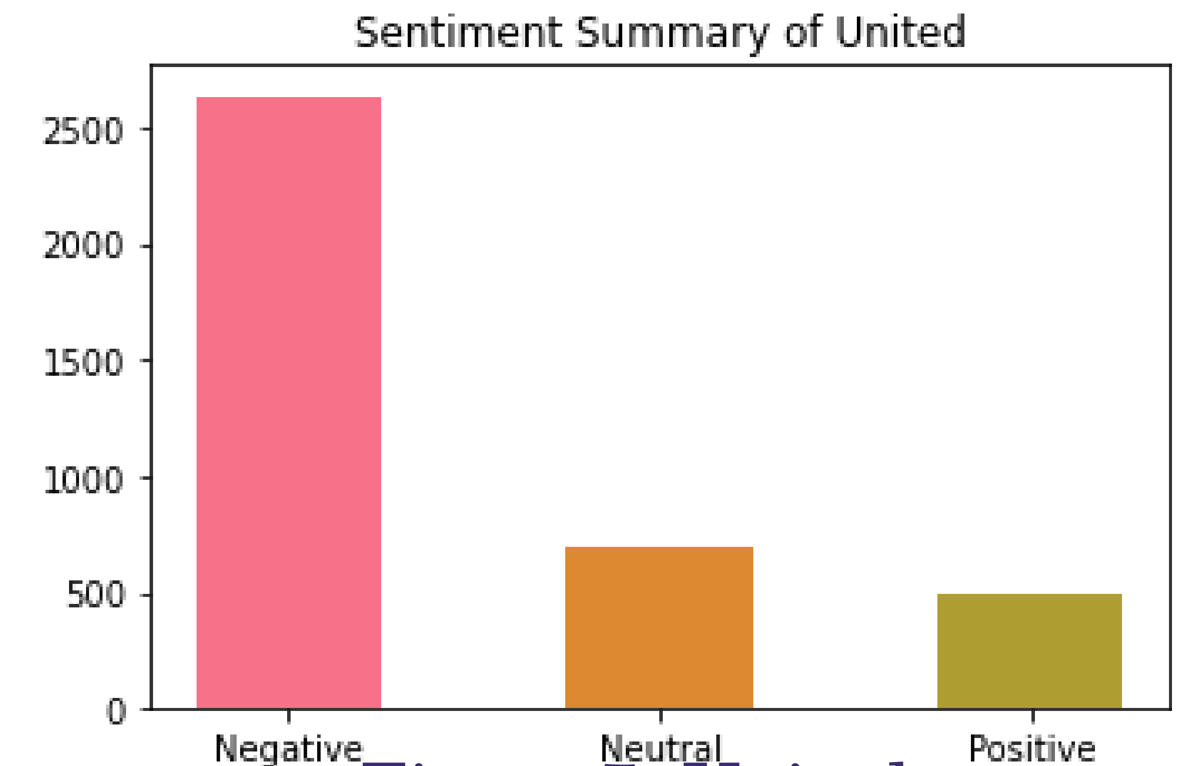

Figure 7: United

Airline Attitude

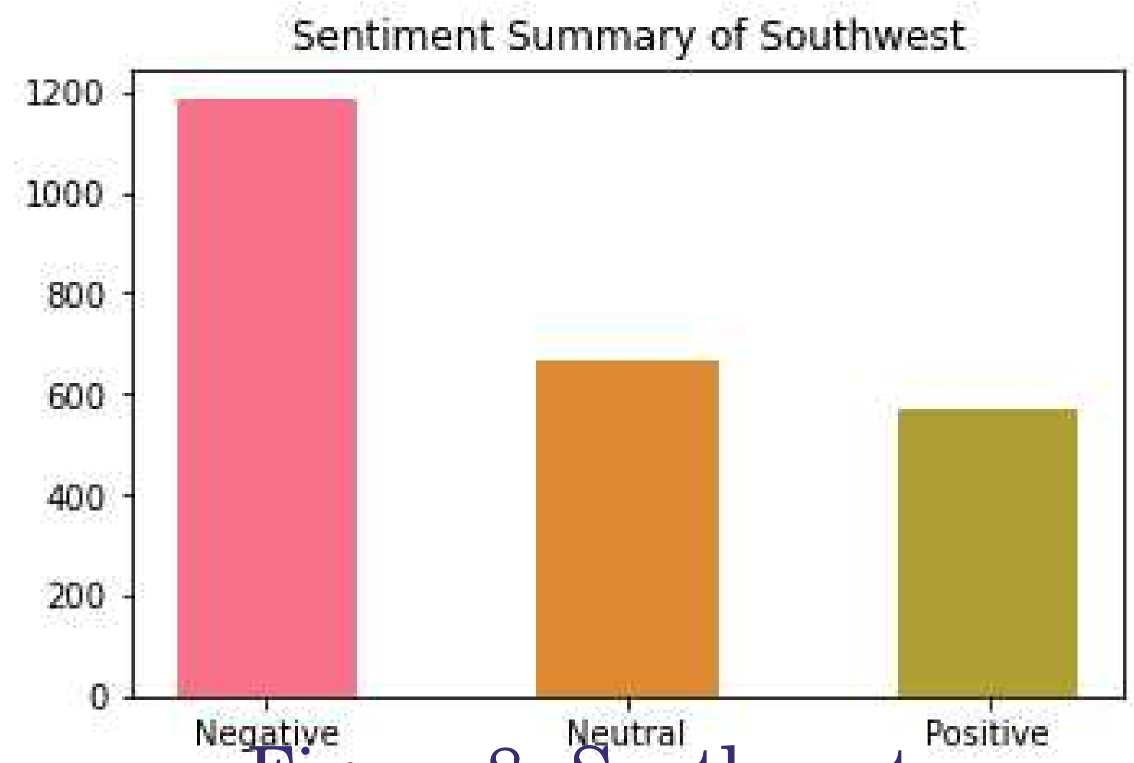

Figure 8: Southwest

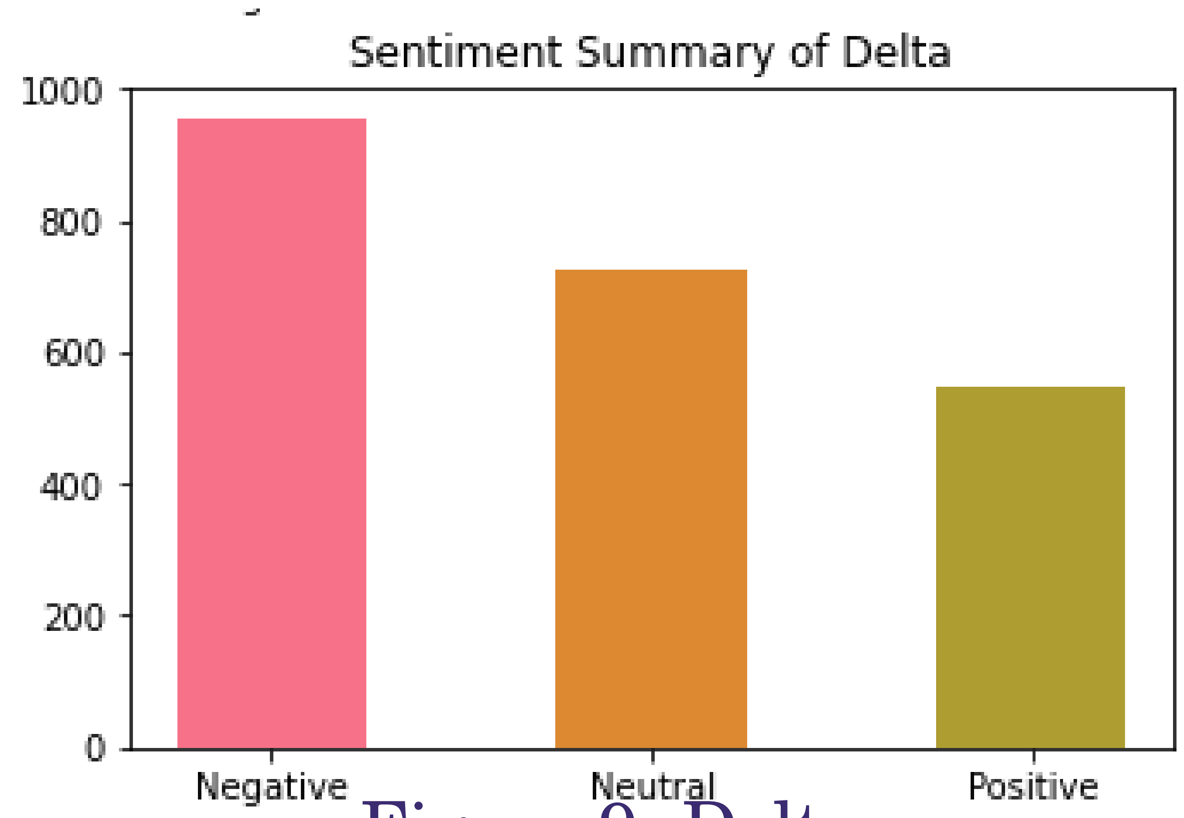

Figure 9: Delta

Airline Attitude

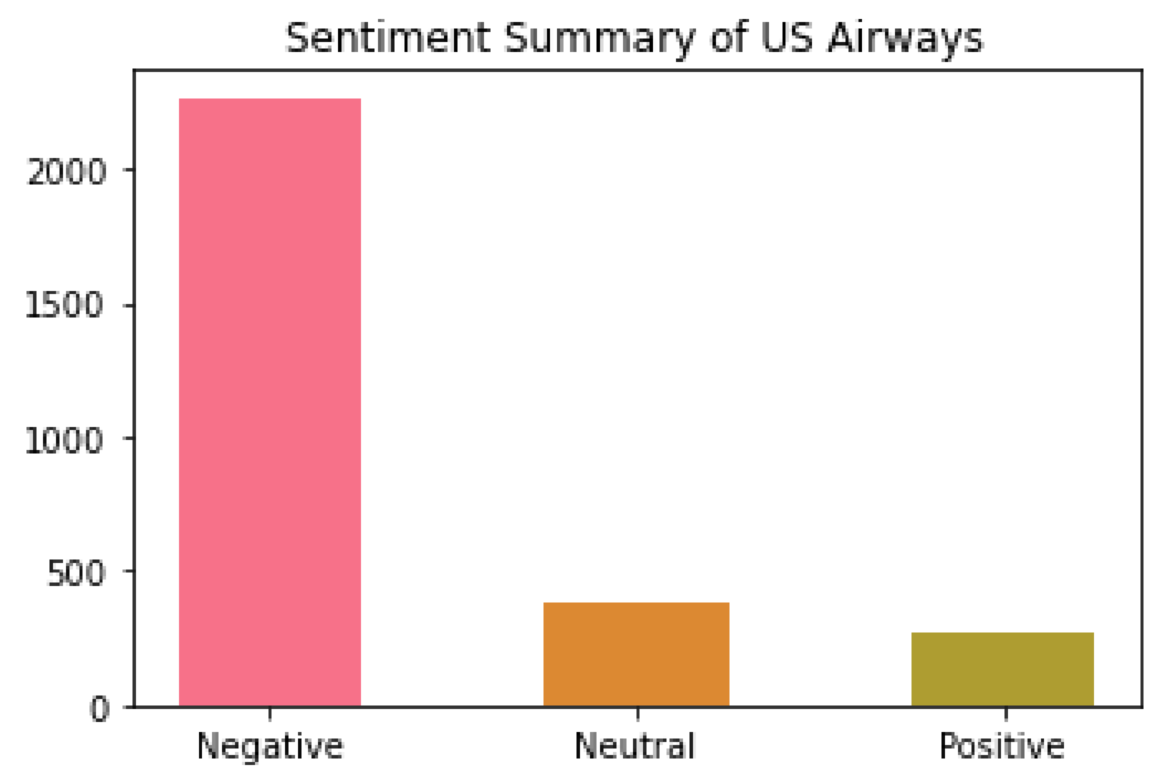

Figure 10: US Airways

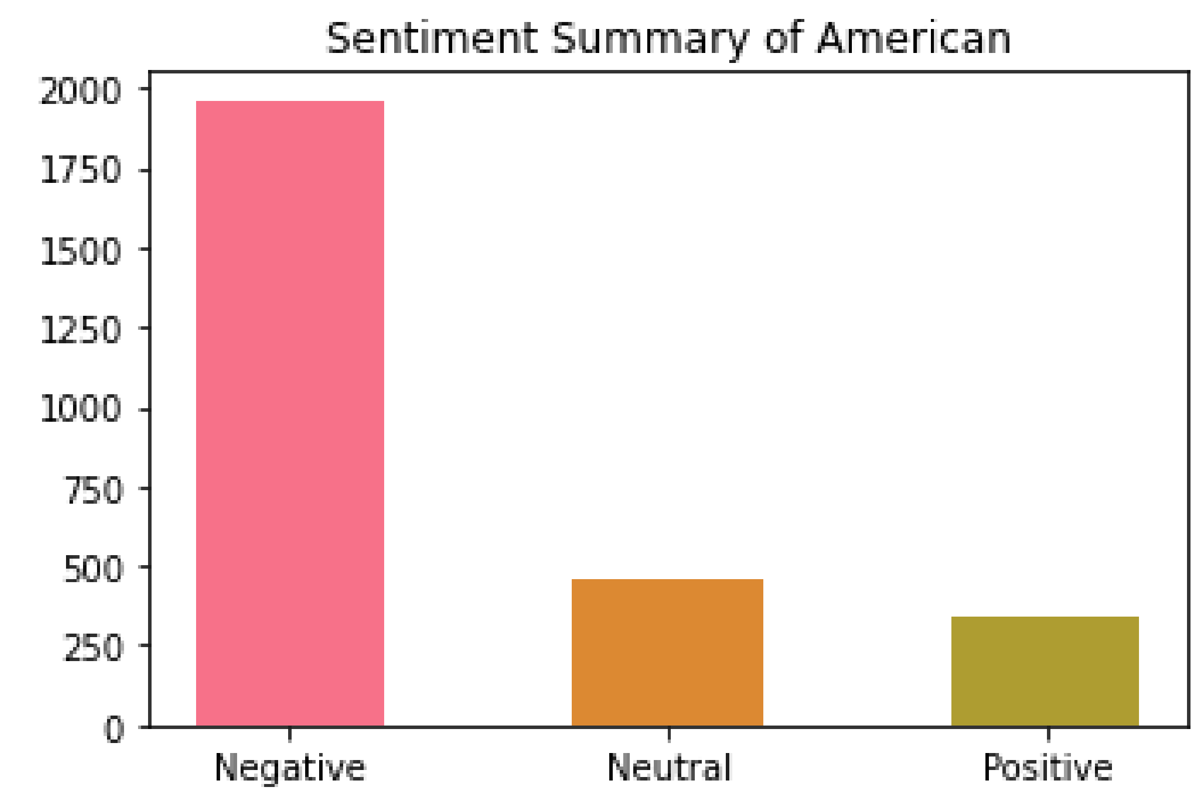

Figure 11: American

negative Word Cloud

- Words with negative emotions involve flight, canceled, hour, etc.

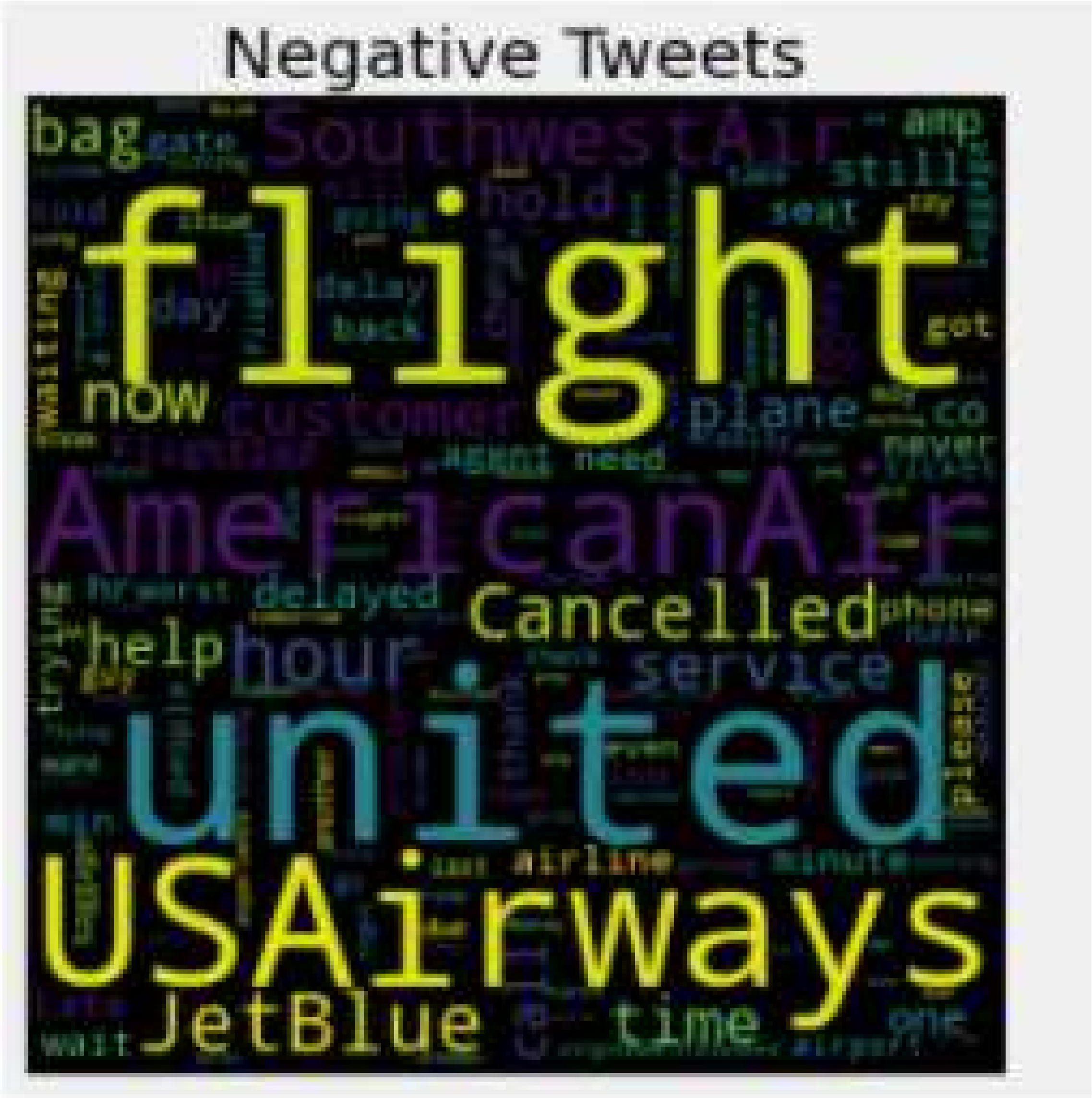

Figure 12: Negative Words

Positive Word Cloud

- Words with negative emotions include thanks, great, etc.

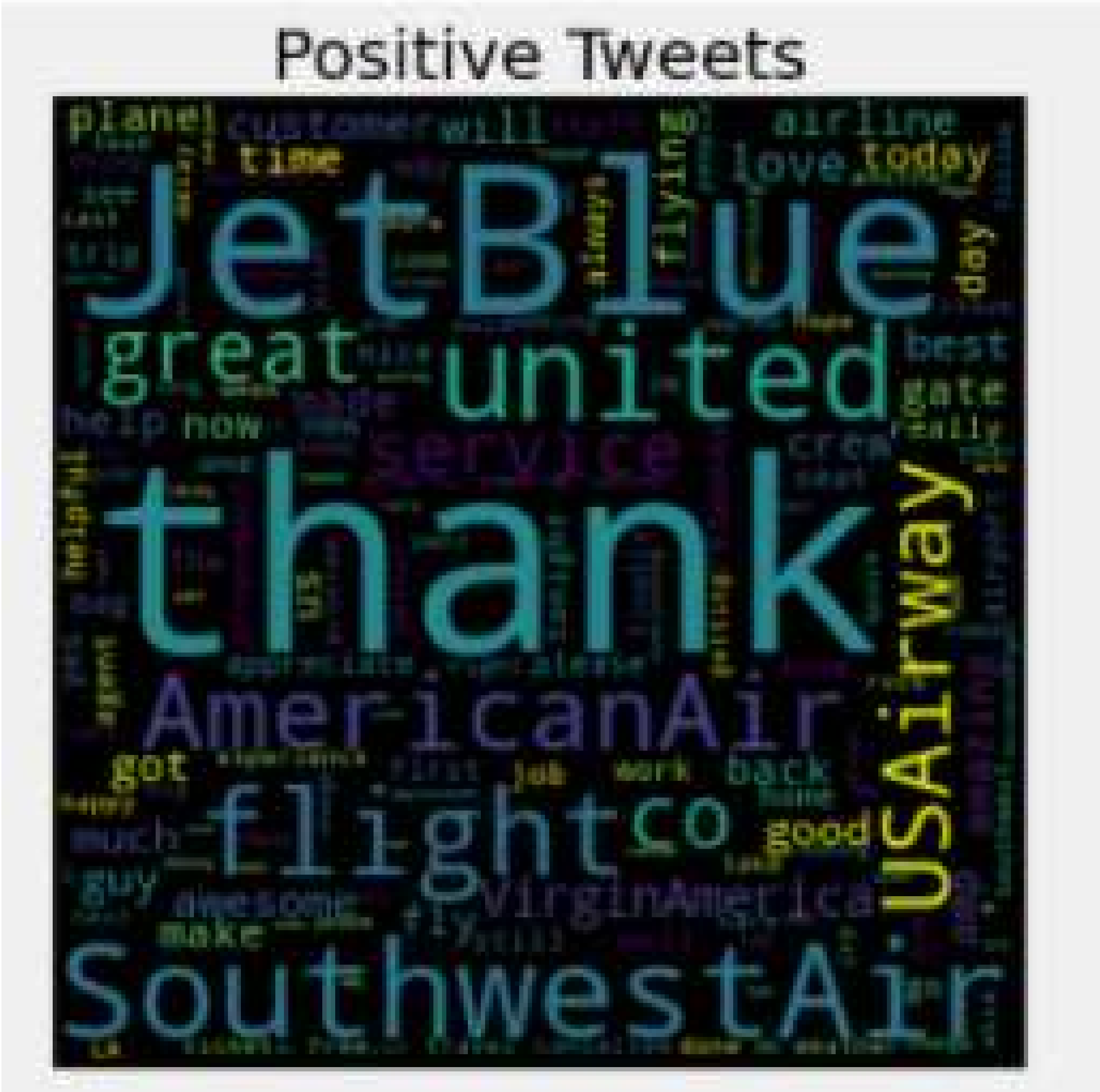

Figure 13: Positive Words

Neutral Word Cloud

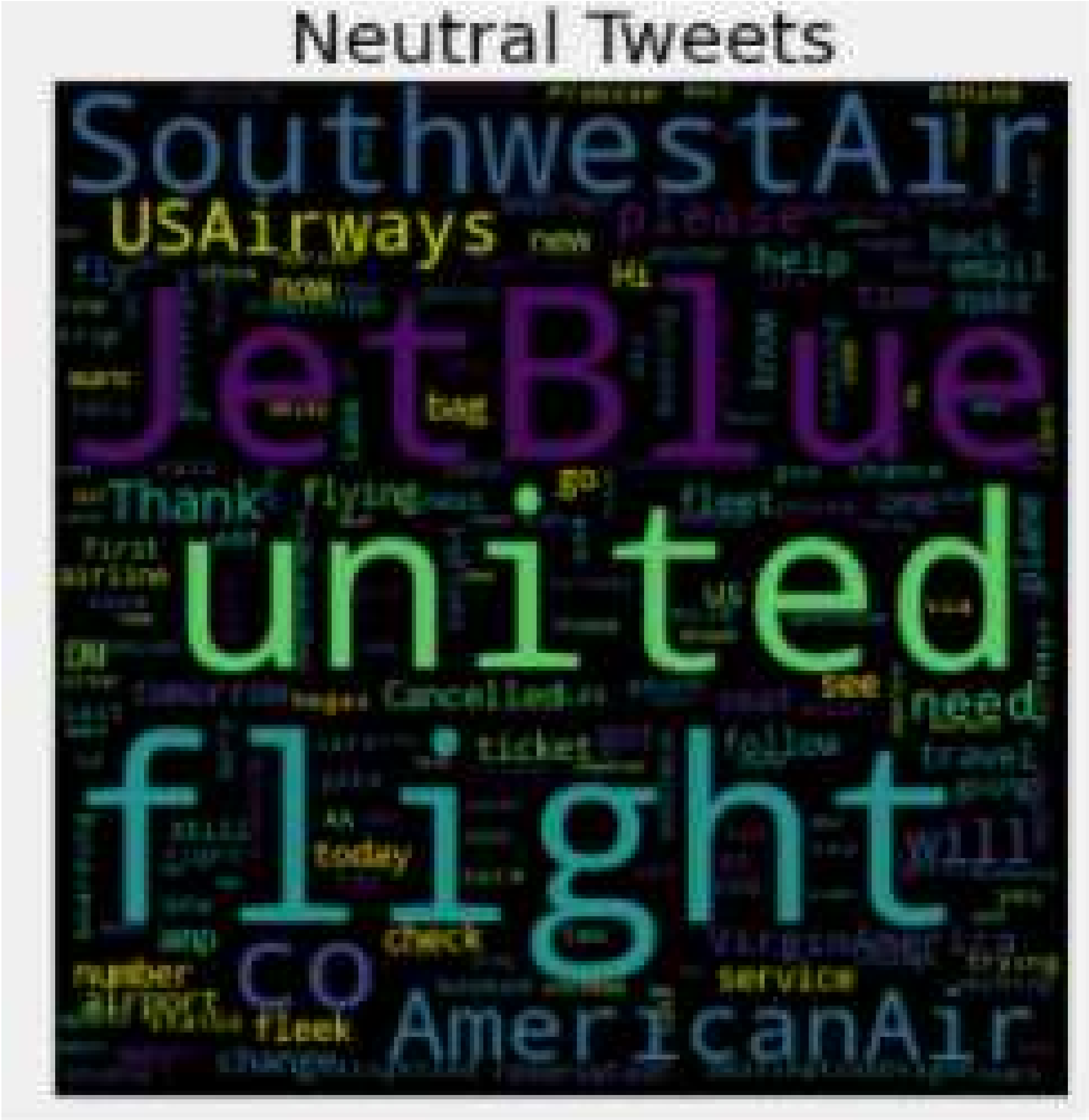

Figure 14: Neutral Words

- Perform statistical tactics on the causes of negative attitudes.

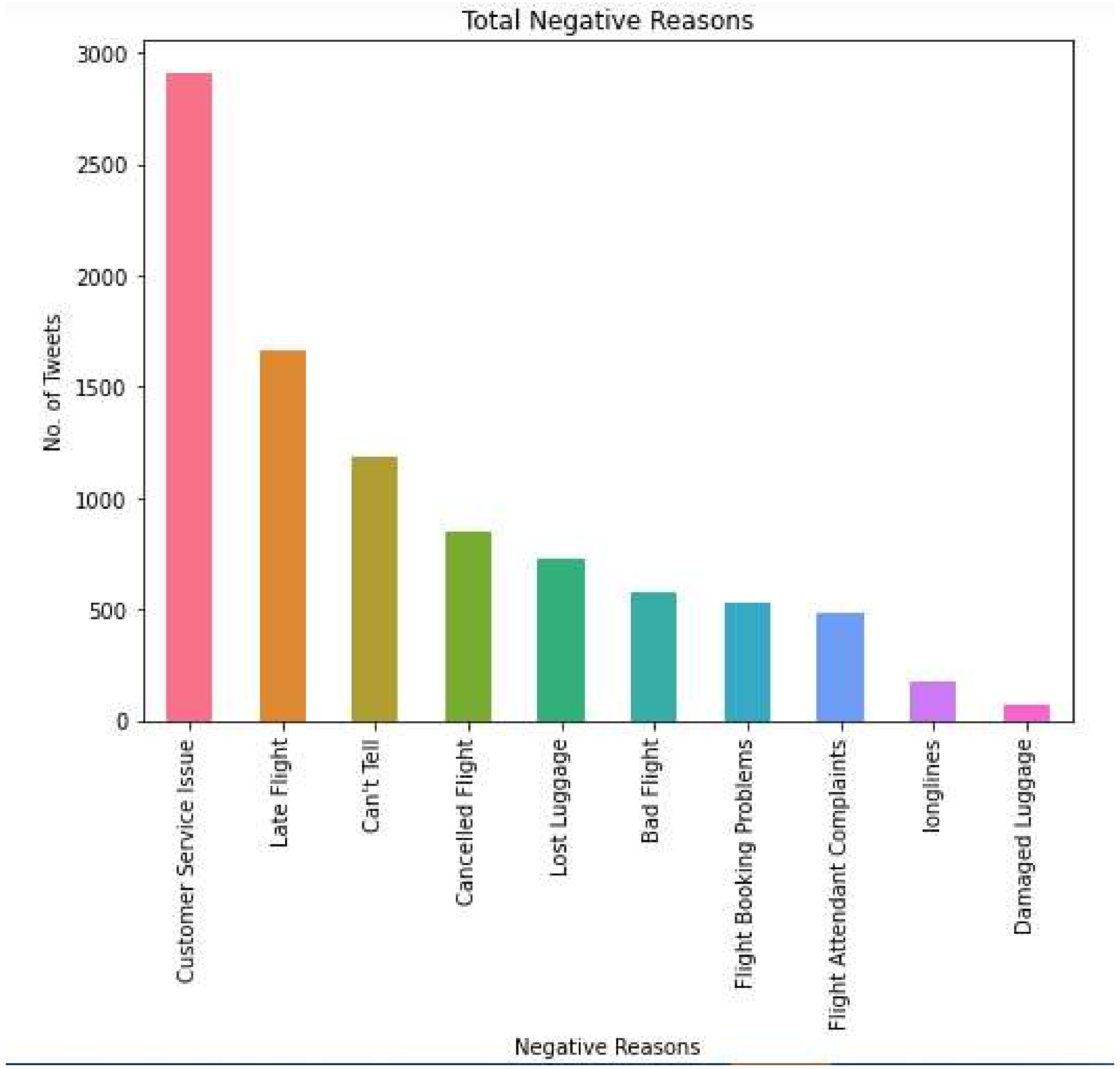

Figure 15: Total Negative Reasons

Data Processing

Data Processing

- Processing data independent of results.
  - For example:tweet_id,name,tweet_location,tweet_coord etc
- Normalization of training data and test data.
  - For example:airline_sentiment,airline etc
- Split train and test data.

Modeling

Model

- Random Forest
- Gradient Boosting
- LSTM RNN

Table 2: Comparative Results

Model	Accuracy
Random Forest	0.8135245901639344
Gradient Boosting	0.8265027322404371
LSTM RNN	0.8865852952003479

Conclusion

Conclusion

- On this issue, the new model is better than the traditional machine learning model.
- Pay attention to the adjustment of parameters when training the model.
- Other models can be used to further improve accuracy.