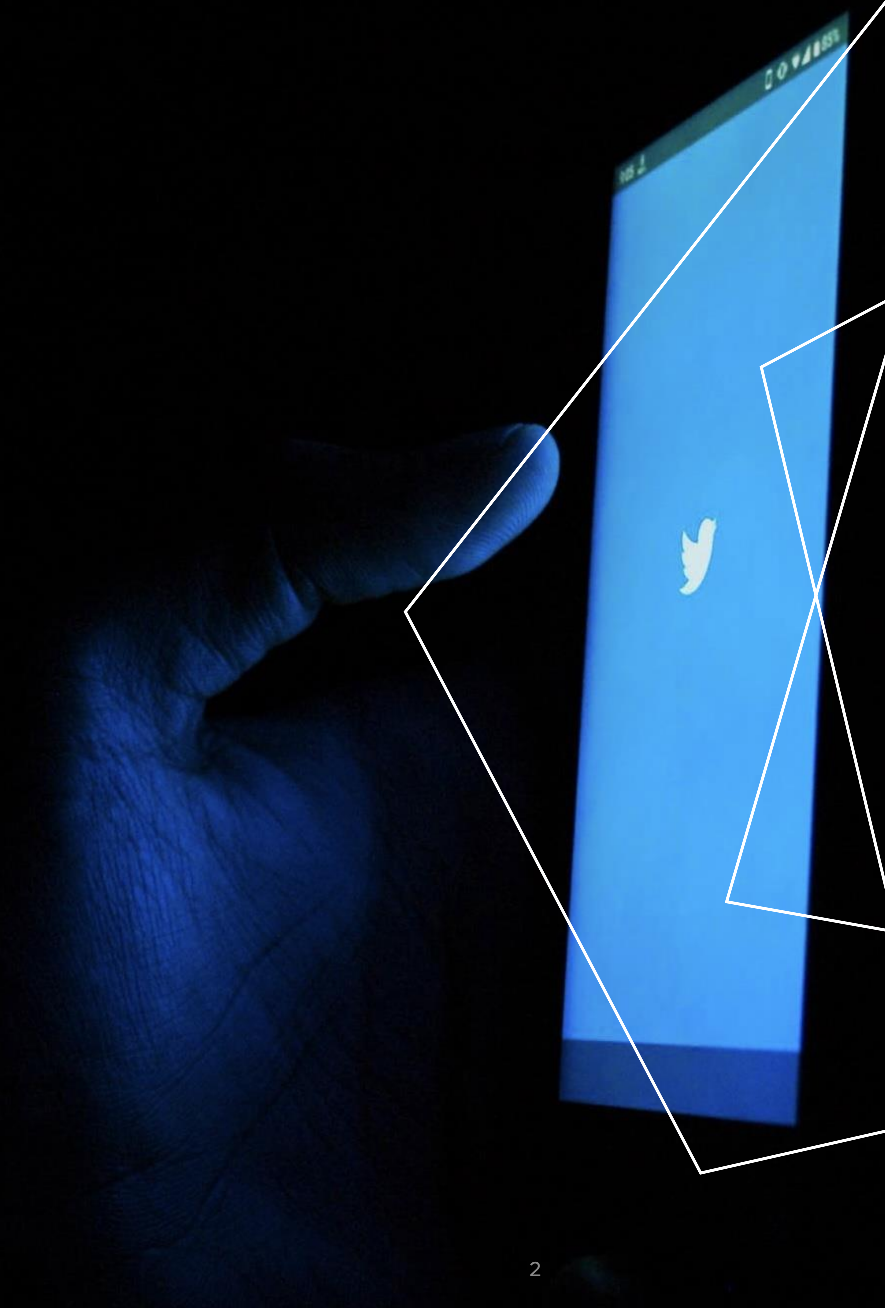
AIRLINE SENTIMENT ANALYSIS

By Yesim Cebeci

TABLE OF CONTENTS



BUSINESS UNDERSTANDING

Airline companies aim to use a model to identify unhappy customers and direct negative tweets to the proper channel.

One way to do this negative tweets can be identified using sentiment analysis and redirected to customer service with a bot.

By implementing this system , Airline Companies can improve customer satisfaction, increase brand loyalty by addressing negative feedback.

DATA UNDERSTANDING

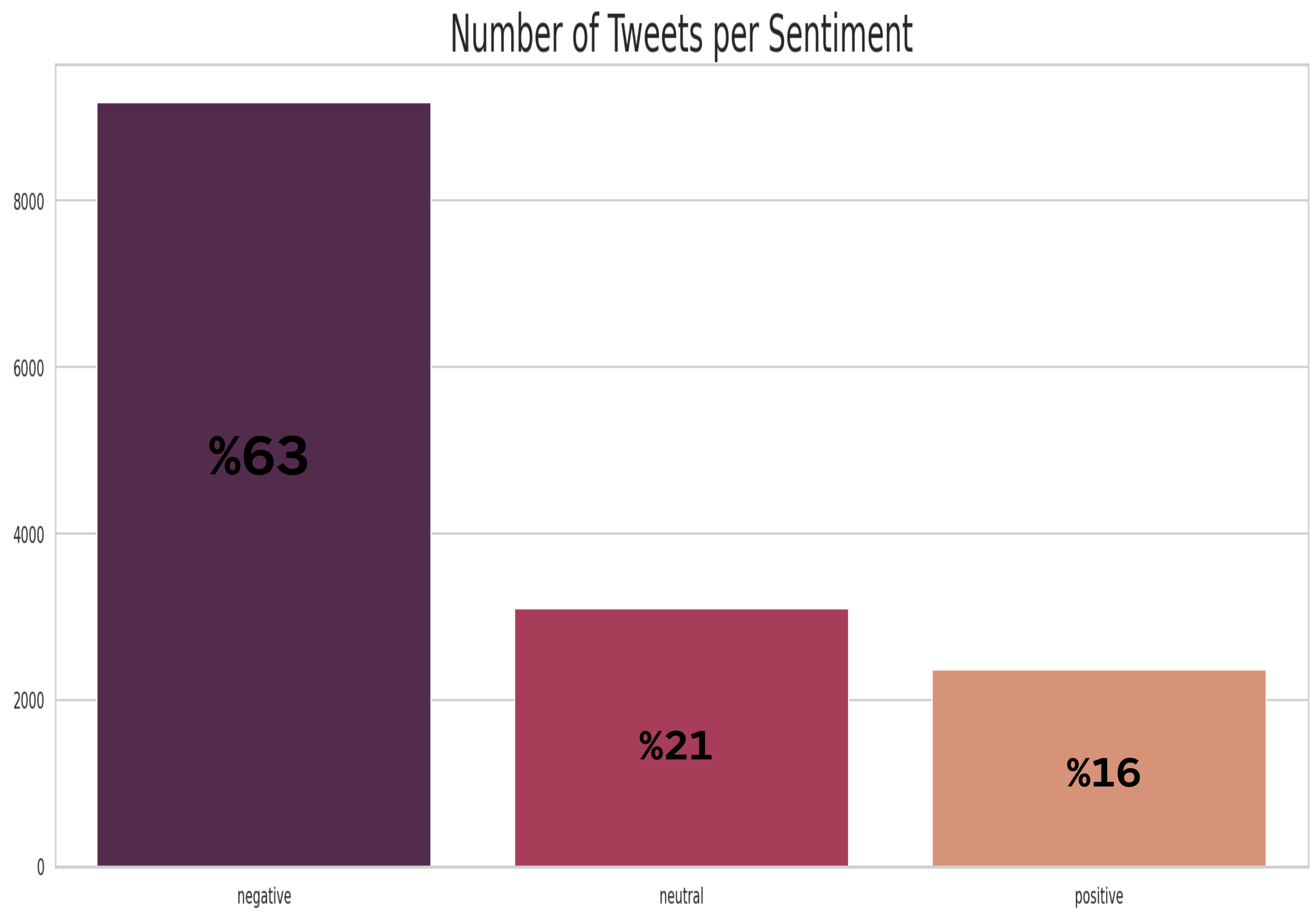
- The dataset used in this project is provided on Kaggle.

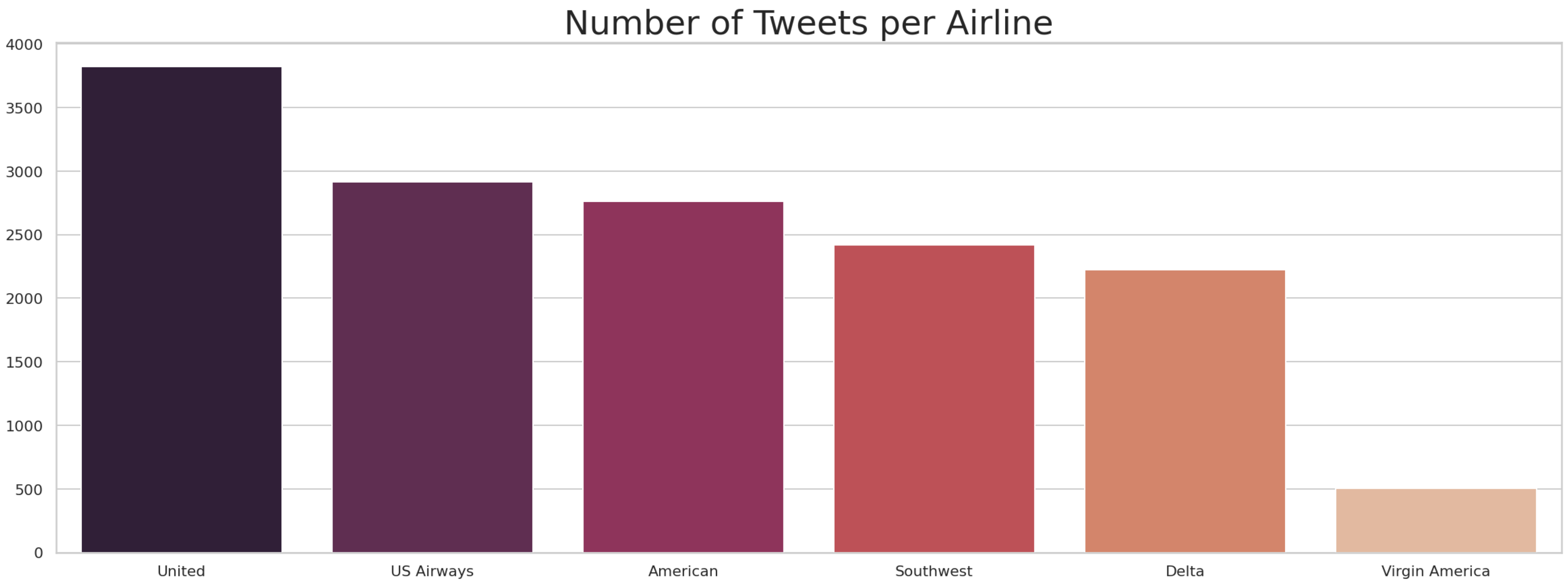
- This Twitter data was scraped on **February 2015**

- It contains tweets on six major United States (US) airlines: **United, Us Airways, American, Soutwest, Delta** and **Virgin America.**

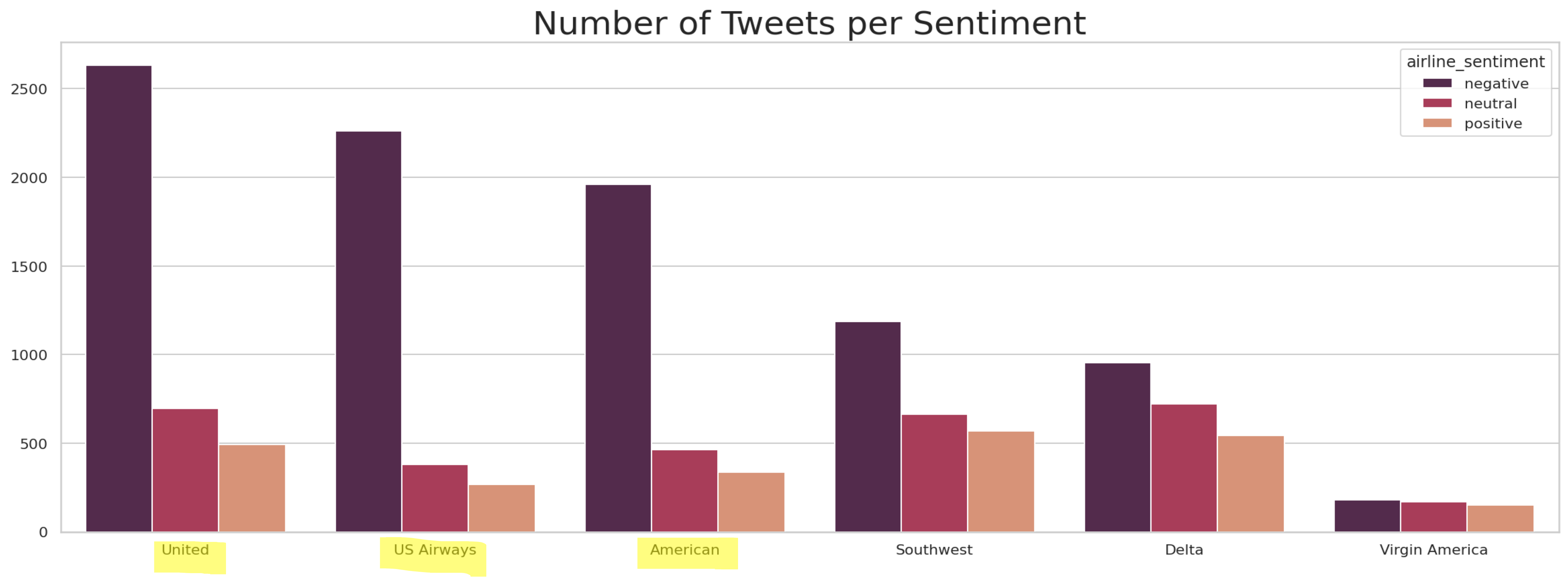
- Mainly focusing on "**airlines**" , "**airline_sentiment**" and "**text**" columns

- Tweets are overwhelmingly **negative**.
- This is an imbalanced dataset.



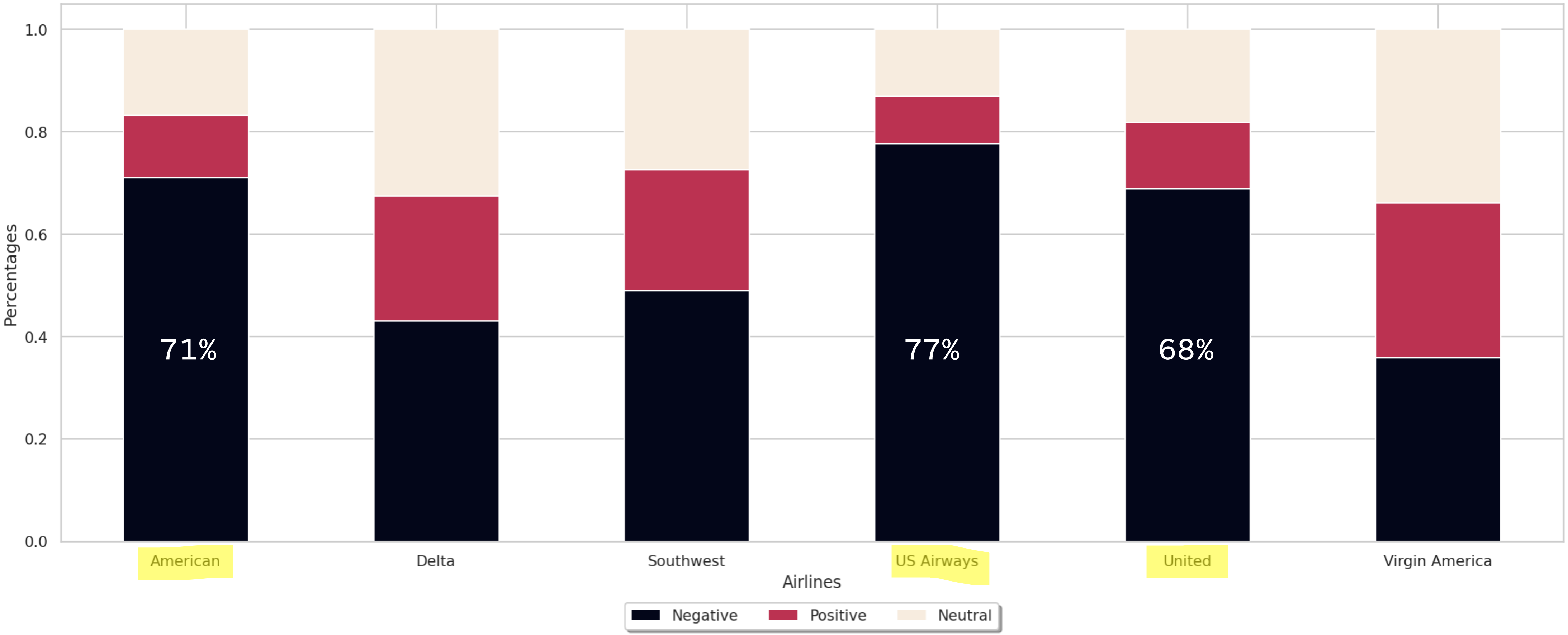


- Most of the tweets belongs to **United Airlines** and followed by **US Airways and American.**



- The number of negative sentiments, **United Airline** ranks the first, followed by **US Airways Airline** and **American Airline**

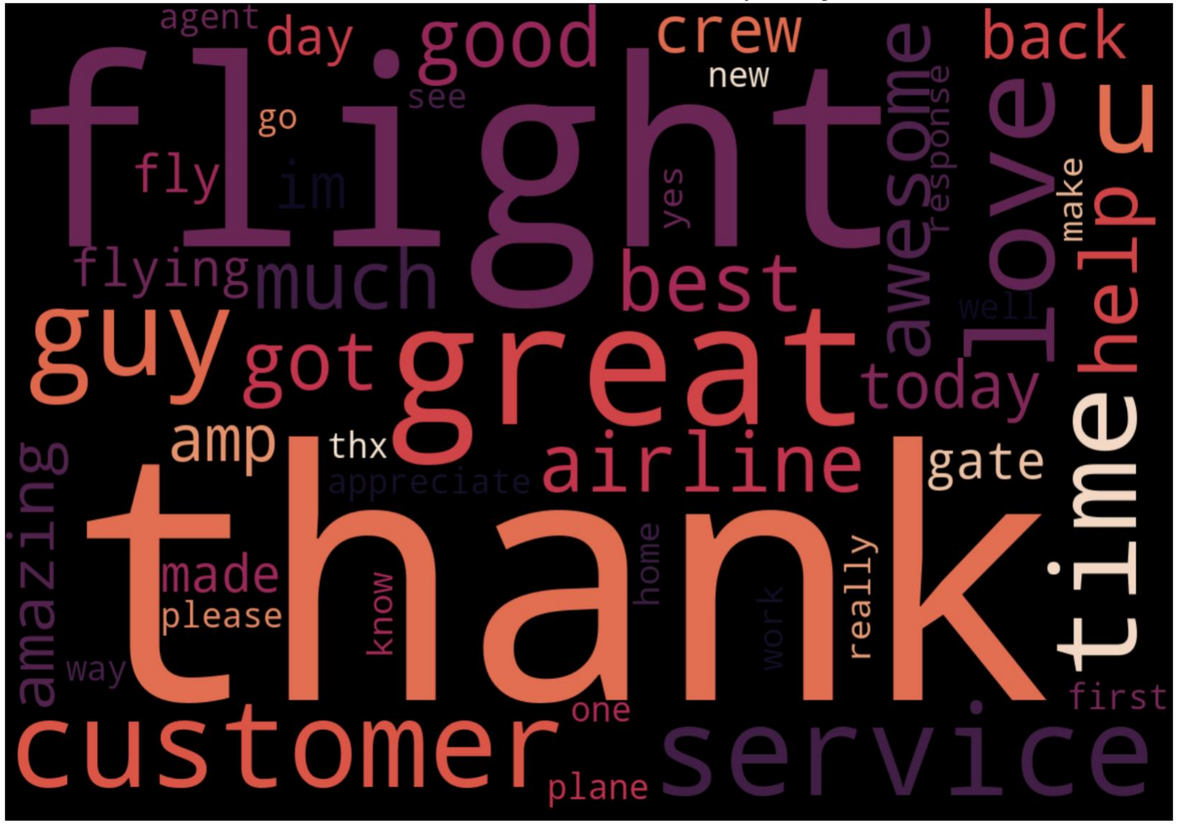
- The numbers of negative, neutral and positive sentiments for **Virgin America Airline** is fairly balanced.



- **US Airways** has the highest proportion of negative sentiments at **77%,** followed by **American Airlines** at **71%**.

- **Virgin America** has highest percentage positive tweets.

Positive Feedback Word Frequency

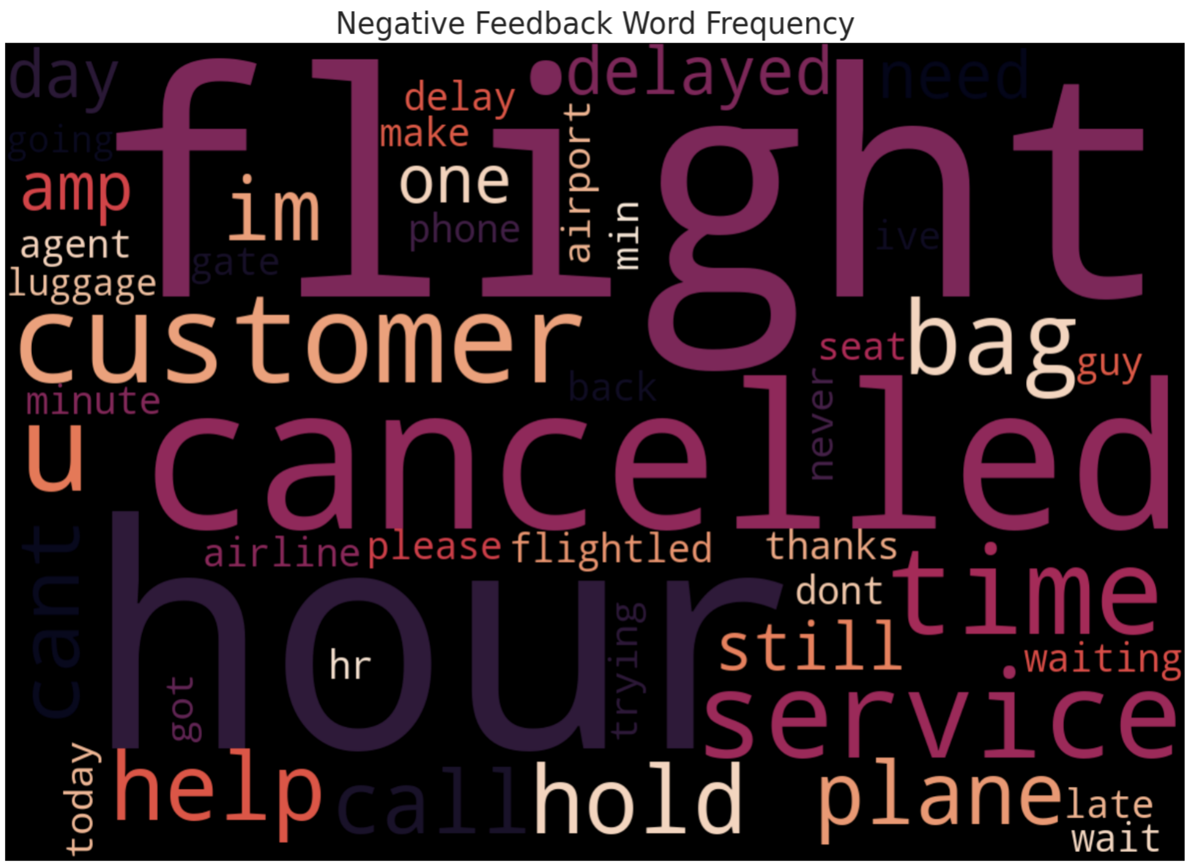


❖ Positive sentiments mostly consist of words such as **'thank,' 'flight,'** and **'great'**.

❖ This shows people tend to appreciate the airline on social media when they have positive flight experience

➢ Hey @united you've upgraded me on a 10 hour International flight. I forgive you :-) thank

➢ @JetBlue great flight on a brand new jet. Great seating. Beautiful plane. Big fan of this airline.



- ❖ Tweets related to **"flight"** and **"hour"** are causing the most negative tweets.
- ❖ **"canceled"**,**"delayed"**, **"customer"** and **"service"** have higher frequencies than other words

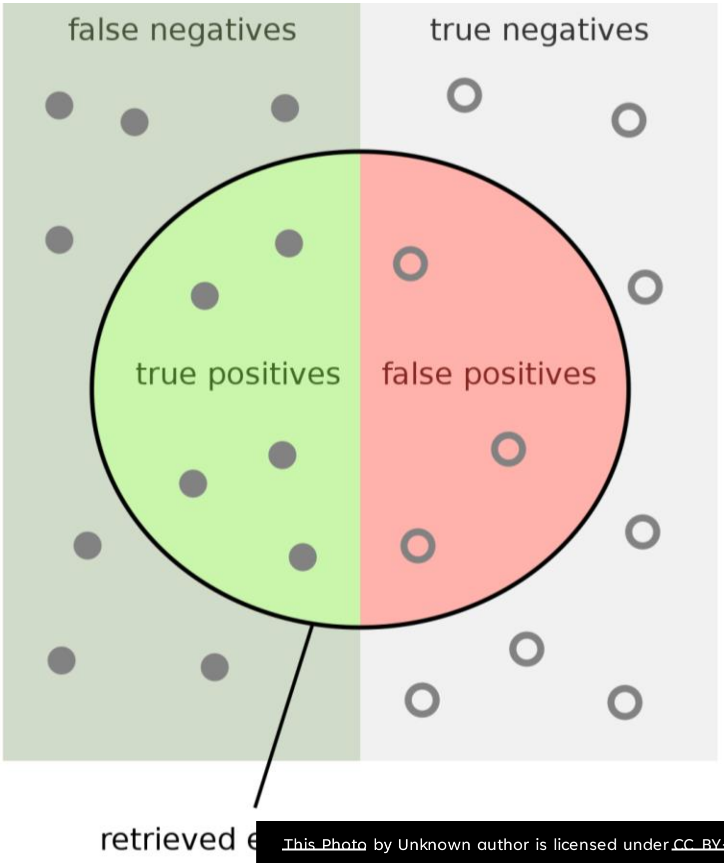
- ➢ @united 4840 WTF! Why can't you update your times in a timely manner? You've known the fight was more delayed! Ur service is awful!
- ➢ @united still no refund or word via DM. Please resolve this issue as your Cancelled Flightled flight was useless to my assistant's trip.

METRIC TO USE: F1-SCORE

- **False Positive:** Model predicts negative tweet but it's actually positive.

- **False Negative:** Model predicts positive tweet but it's actually negative.

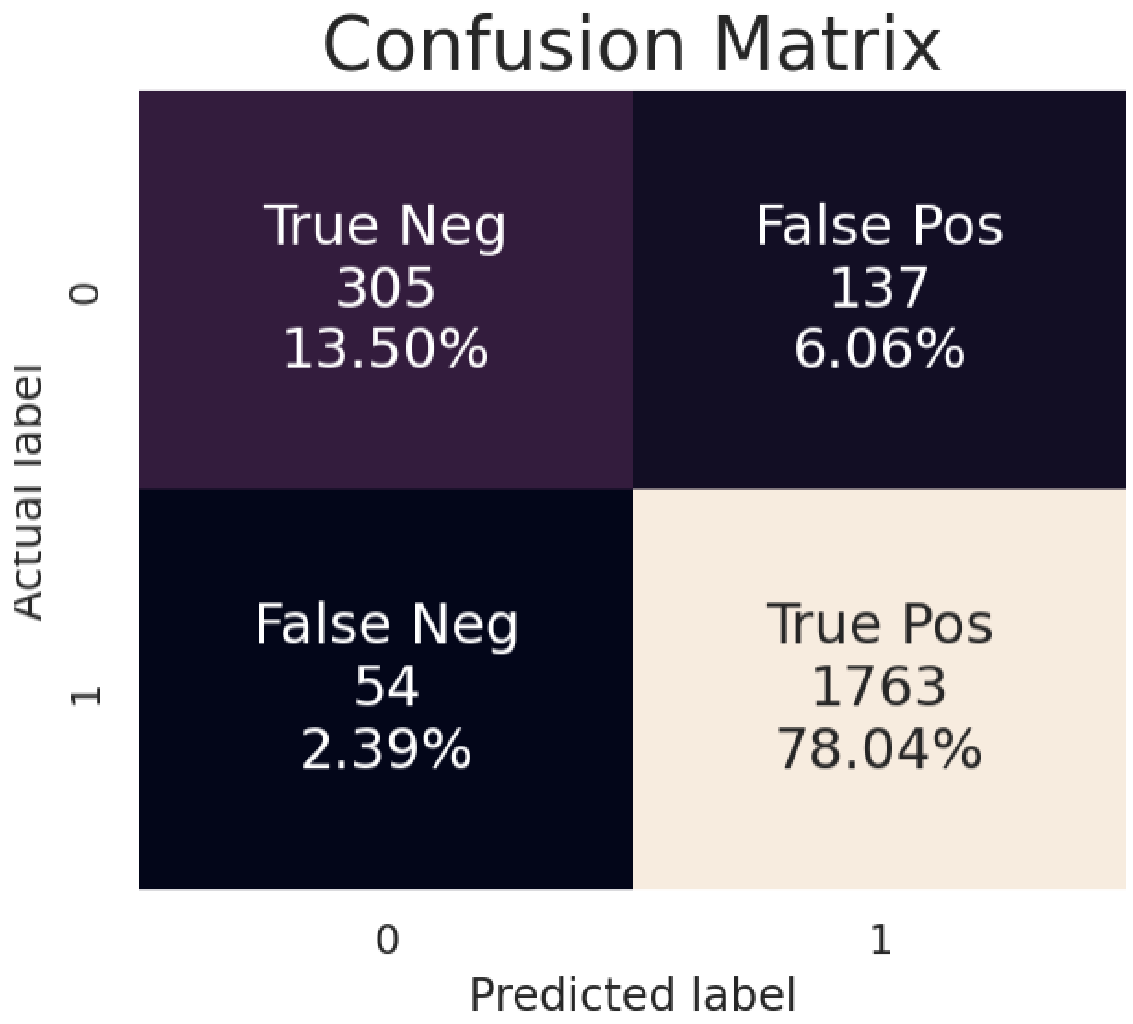
- To balance the downsides of False Positive and False Negative, F1-Score is used to find maximum negative sentiment tweets while avoiding incorrect customer service actions.



BEST PERFORMING MODEL

- Throughout the project ,Our goal was reducing False Negatives and False Positives

- Best F1 score were achieved through parameter tuning with GridSearch on Support Vector Classifier



CONCLUSION

- The model appears to be working effectively and making accurate predictions in general
- Interpreting sentiment in text can be difficult due to factors like sarcasm and irony

Tweets	Predicted Sentiment
"@united that's cool - now what?"	Positive
"counter agents at RDU deserve a medal. #thankyou"	Positive
" I'm rebooked now, but the line was 300 people deep. "	Negative
"@united @retailbagholder hahaha. At least they gave u a refund."	Negative

RECOMMENDATION

- Cost analysis of false positives vs. false negatives needed for next step in airlines.
- Bot offers personalized solutions based on negative sentiment reason. E.g. discounts or refunds for flight delays.
- This model can be used for airlines based on specific business needs

NEXT STEP

- Data limited to Feb. 2015 tweets. Collecting full year data would result in a stronger, more general model.

- Analysis shows negative tweet dominance. Future work aims for larger, balanced data for improved model performance.

- Sarcasm in tweets can skew model results. Further study needed.

THANK YOU

Yesim Cebeci

ysm.cebeci@gmail.com

https://github.com/yesimcebeci