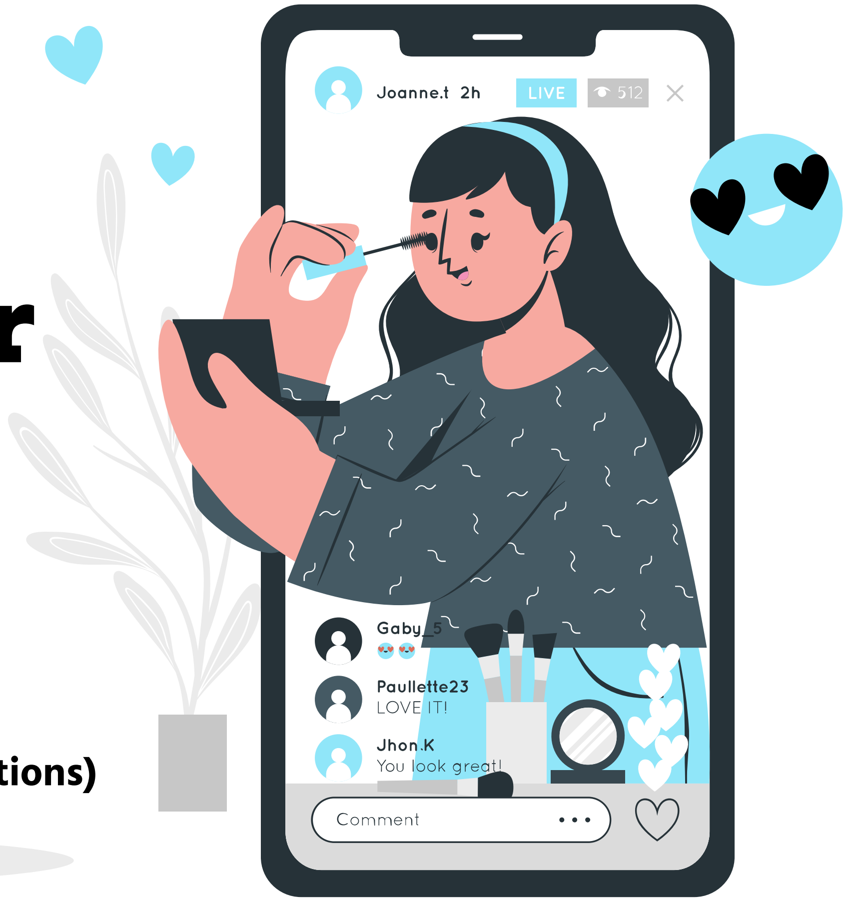
US Airline Travel Twitter Sentiment Analysis



Bassey Dominic Edet
Intern Data Scientist (Brainwave Matrix Solutions)
Basseydominic10@gmail.com

Overview

This twitter sentiment analysis is tilted towards the tweets on the topic; U.S Airline Travel, between 16th to 24th February 2015.

The data was extracted and analyzed for the over all sentiment scores (positive, negative, neutral) on the topic using Natural Language Processing libraries in Python while visualizing the trend over time with line chart using matplotlib and seaborn libraries in Python.

—Cont'd

Objectives

01

Analyze social media data (twitter) to understand public sentiment

02

Use NLP technique to preprocess text data

03

Use NLP technique to extract sentiment scores

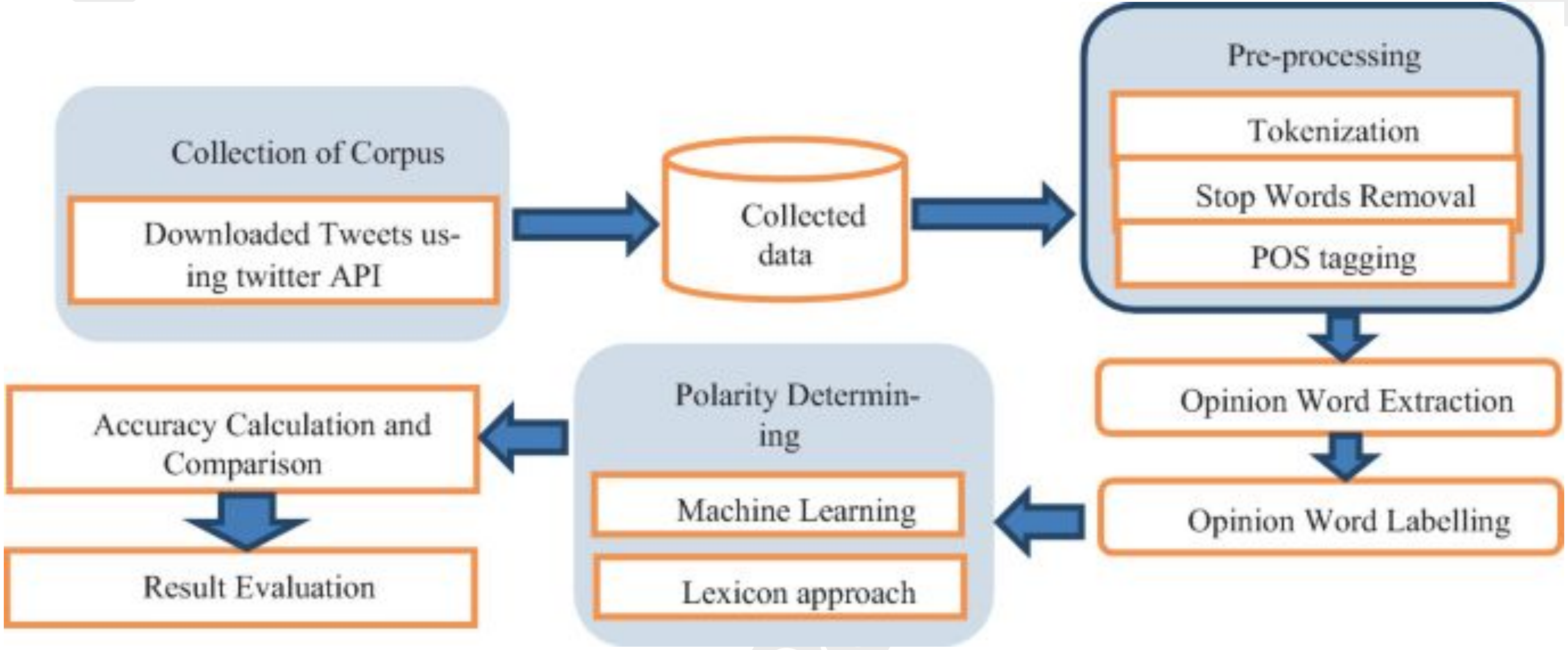
04

Visualize sentiment trends over time

Preview of Cleaned Dataset

airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	airline	text	tweet_created	cleaned_t
neutral	1.0000	Can't Tell	0.899923	Virgin America	@VirginAmerica What @dhepburn said.	2015-02-24 11:35:52-08:00	
positive	0.3486	Can't Tell	0.000000	Virgin America	@VirginAmerica plus you've added commercials t...	2015-02-24 11:15:59-08:00	plus yo ad commerc experie ta
neutral	0.6837	Can't Tell	0.899923	Virgin America	@VirginAmerica I didn't today... Must mean I n...	2015-02-24 11:15:48-08:00	didnt to must m need t another
negative	1.0000	Bad Flight	0.703300	Virgin America	@VirginAmerica it's really aggressive to blast...	2015-02-24 11:15:36-08:00	re aggres b obnoxi entertainm
negative	1.0000	Can't Tell	1.000000	Virgin America	@VirginAmerica and it's a really big bad thing...	2015-02-24 11:14:45-08:00	really big th

Steps in Twitter Sentiment Analysis



NLP Preprocessing Code Snippet

```python
# Step 17: Using re to clean tweet (NLP Preprocessing)
# Define a function to clean the tweet text
def clean_text(text):
    text = re.sub(r'http\S+', '', text)  # Remove URLs
    text = re.sub(r'@\w+', '', text)     # Remove mentions
    text = re.sub(r'#\w+', '', text)     # Remove hashtags
    text = re.sub(r'\d+', '', text)      # Remove numbers
    text = re.sub(r'[^\w\s]', '', text)  # Remove punctuation
    text = text.lower()                  # Convert to lowercase
    return text
```

✓ 0.0s

```python
# Step 18: Apply the cleaning function to the tweet column
df['cleaned_text'] = df['text'].apply(clean_text)
```

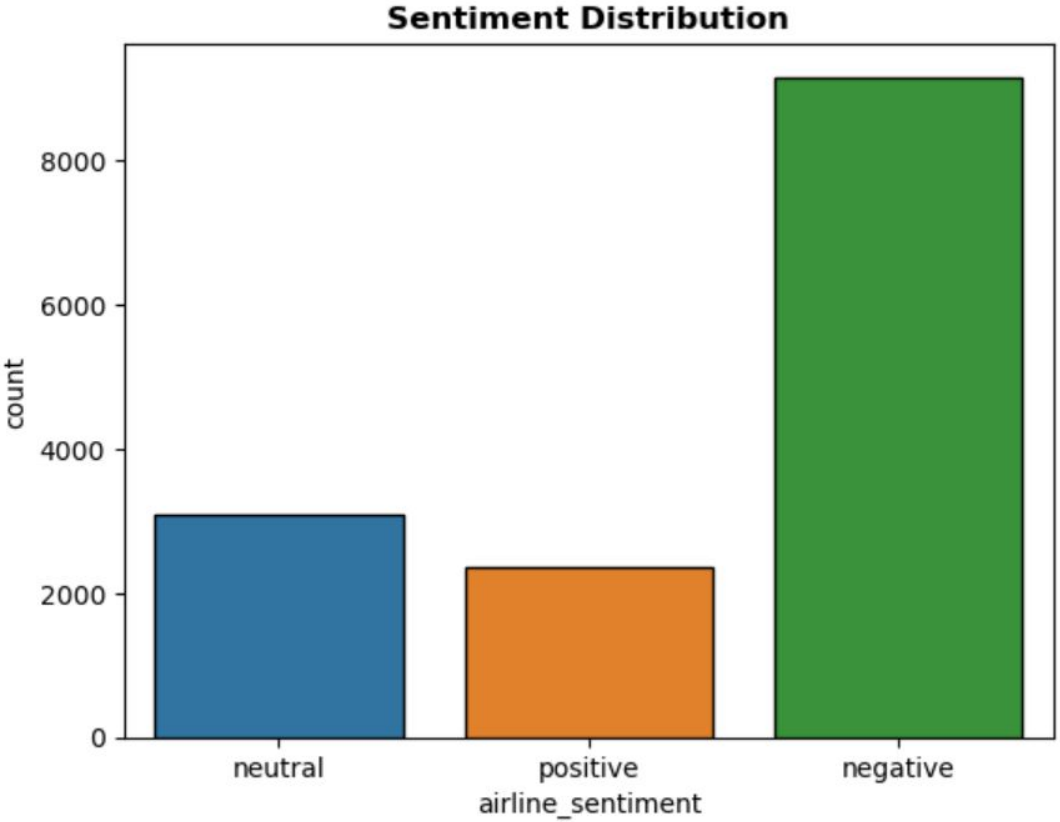
✓ 0.6s

```python
# Step 19: Retrieve English stopwords
stop_words = get_stopwords("en")
```

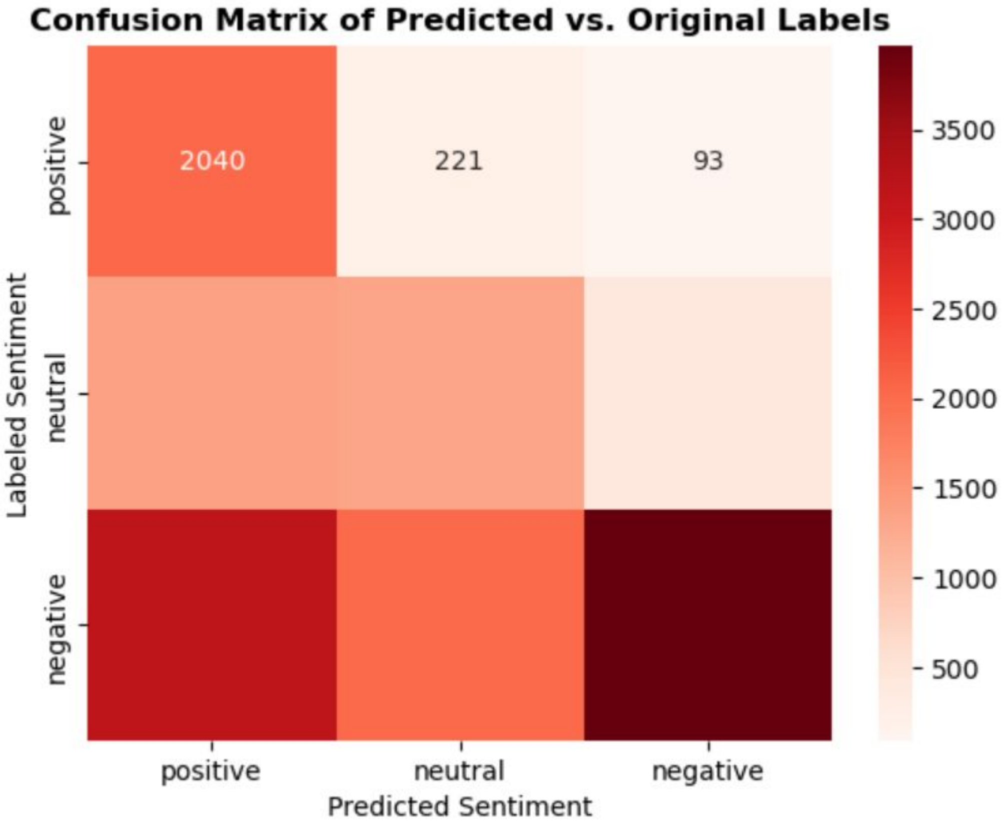
✓ 0.0s

```python
# Step 20
df['cleaned_text'] = df['cleaned_text'].apply(lambda x: ' '.join(word for word in x.split() if word not in stop_words))
```

Sentiment Distribution



Confusion Matrix (Predicted vs Original)



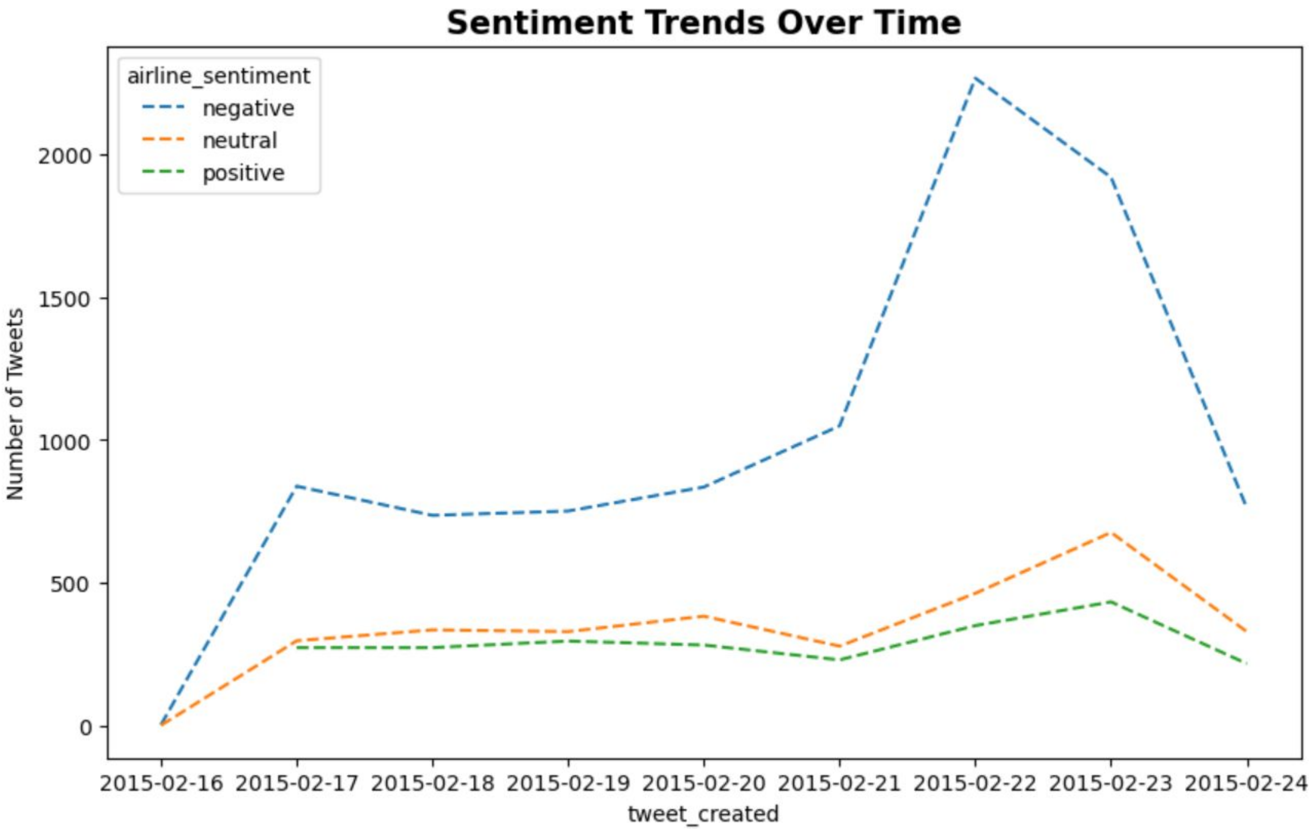
Confusion Matrix (Predicted vs Original)

The confusion matrix shows a comparison between predicted and actual sentiment labels for U.S. airline tweets.

About **2,040** tweets labeled as positive were correctly predicted as positive while **221** positive tweets were misclassified as neutral, and **93** were misclassified as negative.

Misclassification between neutral and other sentiments suggests that the model has difficulty differentiating between neutrality and positive/negative tones, which is common in sentiment analysis.

Sentiment Trends Over Time



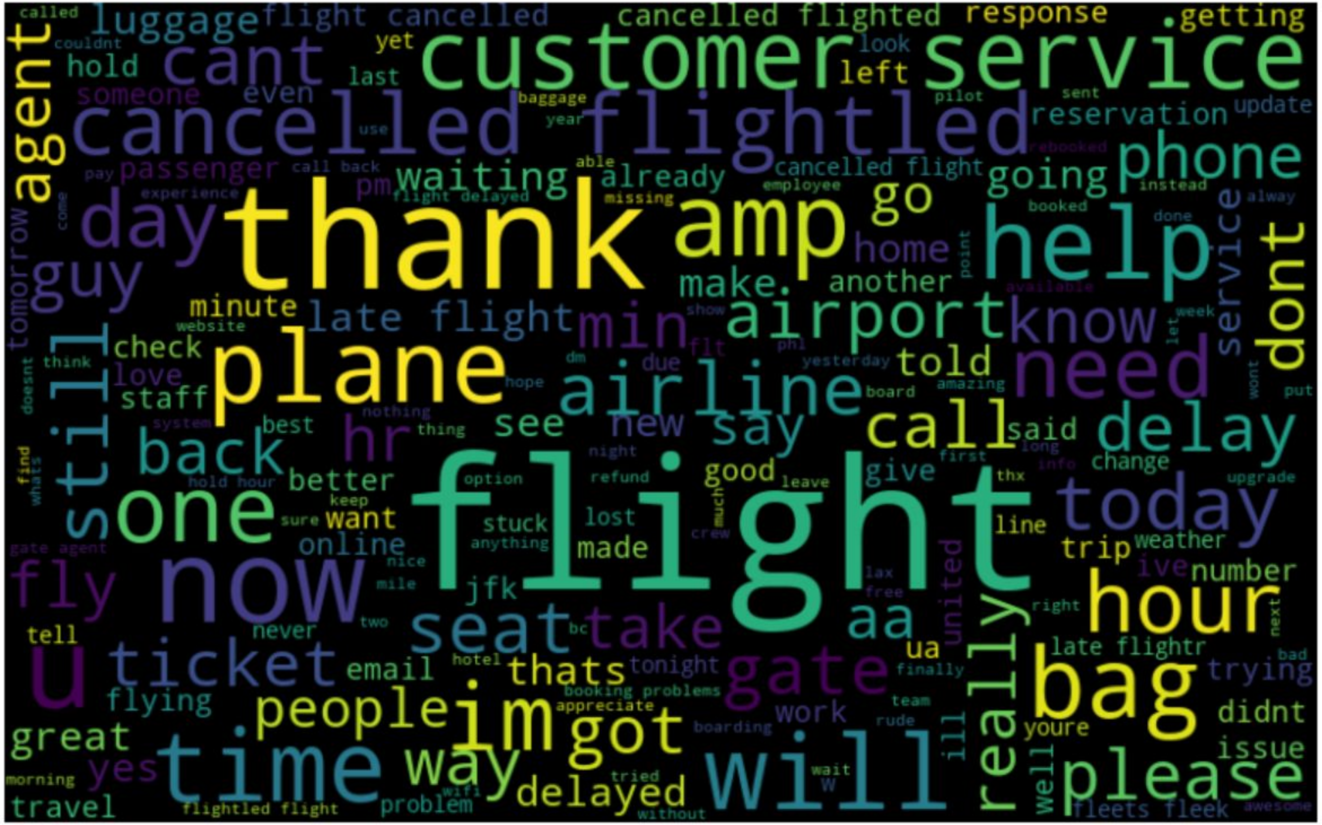
Sentiment Trends Over Time

The chart shows sentiment trends over time for U.S. airline-related tweets between February 16 and February 24, 2015. The sentiments are categorized as **negative**, **neutral**, and **positive**, with the number of tweets tracked daily.

Negative tweets were significantly higher than both neutral and positive sentiments throughout the period. A major spike in negative sentiment occurred around February 22-23, reaching over 2,000 tweets. Positive sentiment is consistently the lowest, with a slight rise but staying under 500 tweets daily.

The spike in negative sentiment around February 22-23 could indicate specific events like late/cancelled flights, lost luggage, customer service issues or other issues affecting airlines. This trend suggests that many passengers were dissatisfied with their airline experiences during this time.

Words Frequently Used in Tweets



Words Frequently Used in Tweets

This word cloud provides a visual representation of the most frequently used words in tweets related to U.S. airline travel, highlighting key terms from customer feedback.

Customer Service Concerns: Words like *"customer service," "agent," "help,"* and *"need"* are prominent, indicating that many users are reaching out for assistance and expressing concerns related to customer service.

Flight Issues: Words such as *"flight," "delay," "cancelled," "wait,"* and *"time"* suggest that flight delays and cancellations are common themes. This aligns with common customer frustrations in airline travel.

Appreciation and Gratitude: The word *"thank"* stands out, indicating that some users are appreciative of assistance when it's received. However, given the context, it's likely mixed in with other issues, such as delayed help or customer service.

Baggage and Reservation Problems: Words like *"bag," "luggage,"*

Recommendations

- Considering the low number of positive tweets, airlines might consider customer experience improvements to encourage more positive feedback. Proactive communication and updates during disruptions, combined with better service recovery efforts, could lead to increased positive sentiment.
- To balance the overwhelming negative sentiment, airlines could encourage satisfied customers to share positive experiences online. Loyalty programs or post-flight surveys with an option to share feedback on social media could help.