
Stock Analysis and Forecasting based on Airline Reviews

Milestone 2

By: Prajakta Gaydhani

Advisor: Dr. Carol Romanowski

Problem Statement

Opinion Mining of different customer reviews or feedback on major Airlines in the United States and analyzing the effect of customer's sentiments on the stock market of different airline companies.



Goal of the Project

- Correlation between customer reviews and stock price movement of different Airline Companies.
- Accurately forecast stock prices for different Airline Companies using customer reviews and stock market data.

Completed : Milestone 1

1. Collected Data

- Collected customer reviews and stock market data
- Airlines: American, JetBlue, Alaska, Delta, United
- Sources: Twitter, Skytrax, TripAdvisor and Yahoo! Finance
- Duration: Jan. 2010 - Sept. 2018

2. Cleaned Customer reviews :

- Removed rows containing non-english words
- Eliminated @usernames, #symbols, URLs, punctuations, special symbols
- Replaced contractions like should've with should have
- Removed whitespaces
- Repetitions like happppppy were converted to happy

Completed : Milestone 1

3. Corrected spelling using TextBlob NLP library in Python
4. Natural Language Processing on Customer reviews using NLTK library
 - Tokenization
 - Removed Stop-words
 - POS tagging
 - Lemmatization

Overview : Milestone 2

- Integrated cleaned customer reviews from sources
- Opinion Mining on customer reviews
- Visualized the results
- Integrated sentiments with stock market data

Opinion Mining - VADER¹

- **VADER (Valence Aware Dictionary and sEntiment Reasoner)**
 - Lexicon and rule based sentiment analysis tool
 - Open sourced under MIT license
 - Attuned to sentiments in social media
 - Can be used in combination with NLTK library
- **VADER Scoring**
 - Compound score by summing the valence scores of each word in the lexicon
 - Normalized between -1 (most extreme negative) and +1 (most extreme positive)

[1] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Opinion Mining - VADER

- **Example:**

Tweet : really bad, horrible flight

VADER Outcome: {'neg': 0.791, 'neu': 0.209, 'pos': 0.0, 'compound': -0.8211}

VADER would classify this tweet as negative

Considered only positive and negative tweets for analysis

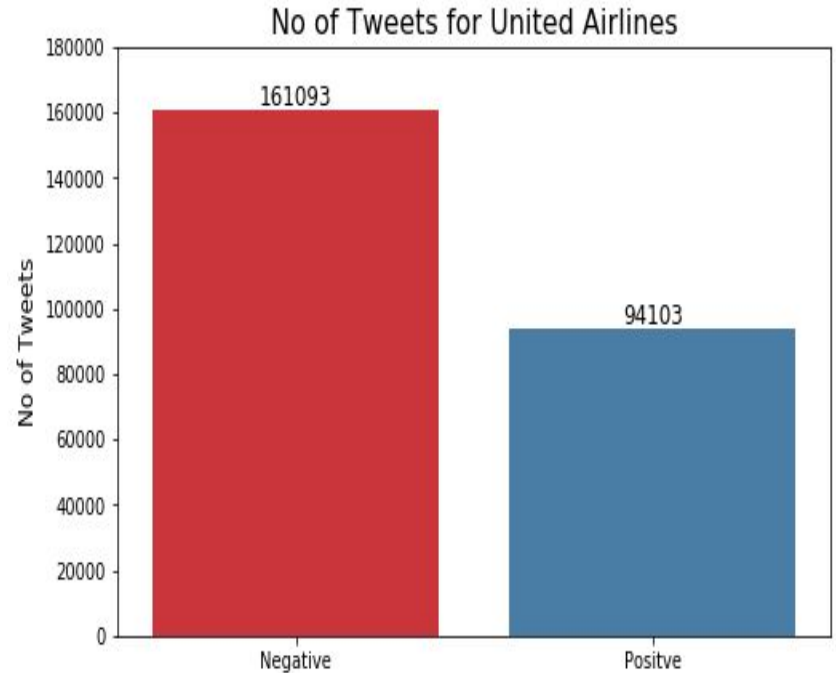
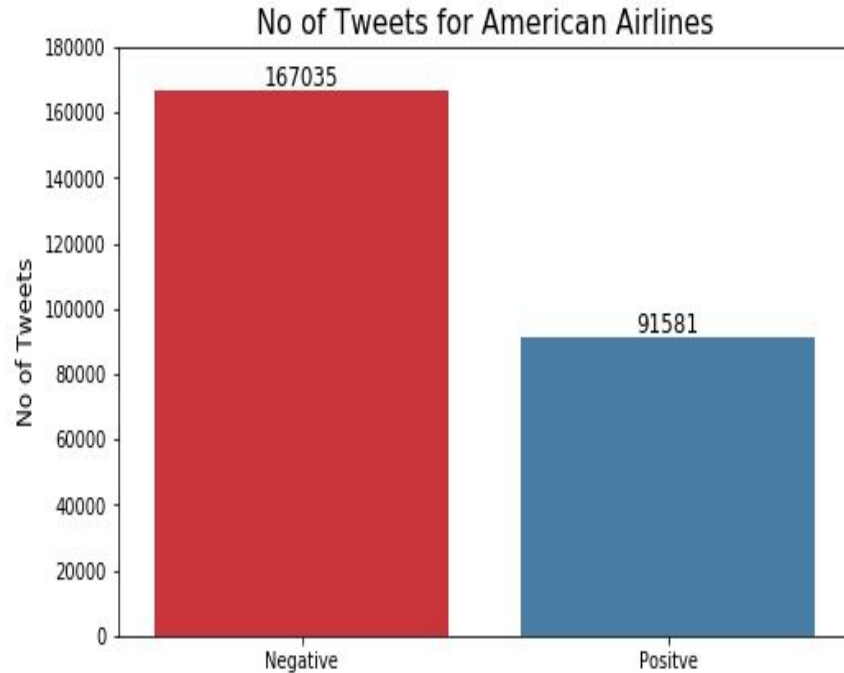
- **For classification**

- positive sentiment: compound score ≥ 0.5
- neutral sentiment: (compound score > -0.5) and (compound score < 0.5)
- negative sentiment: compound score ≤ -0.5

Insights

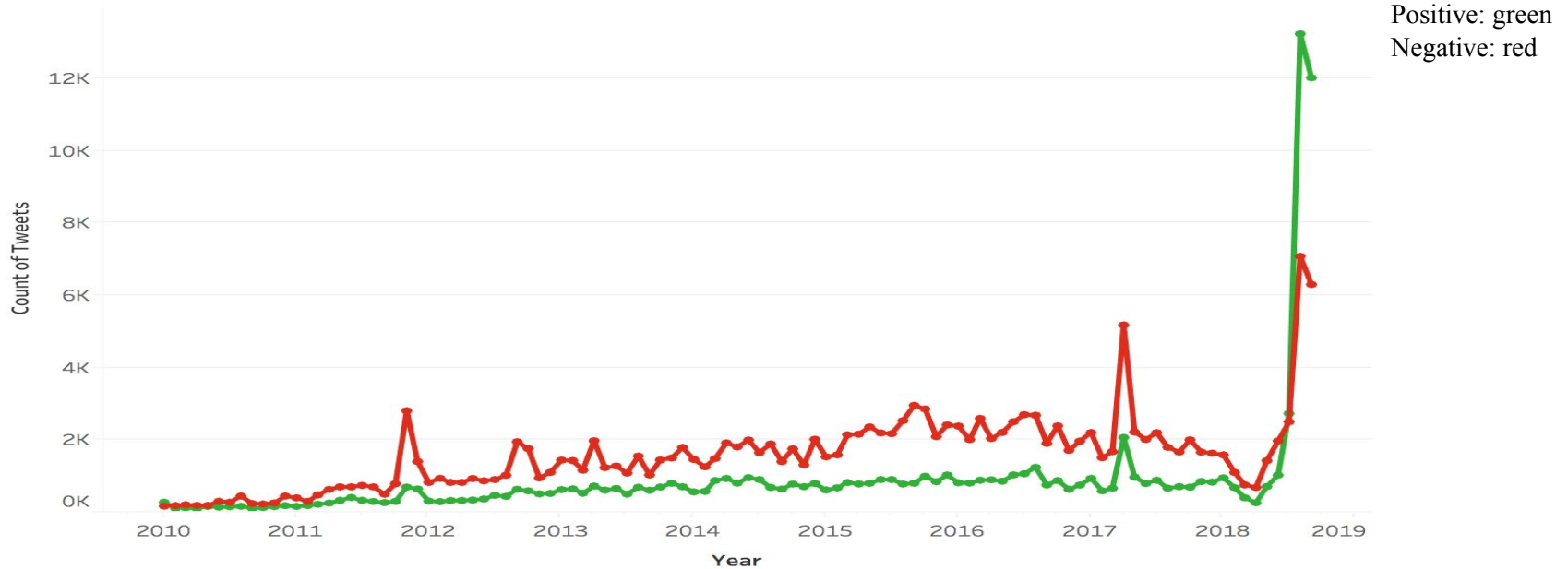
American Airlines and United Airlines

Sentiment Analysis



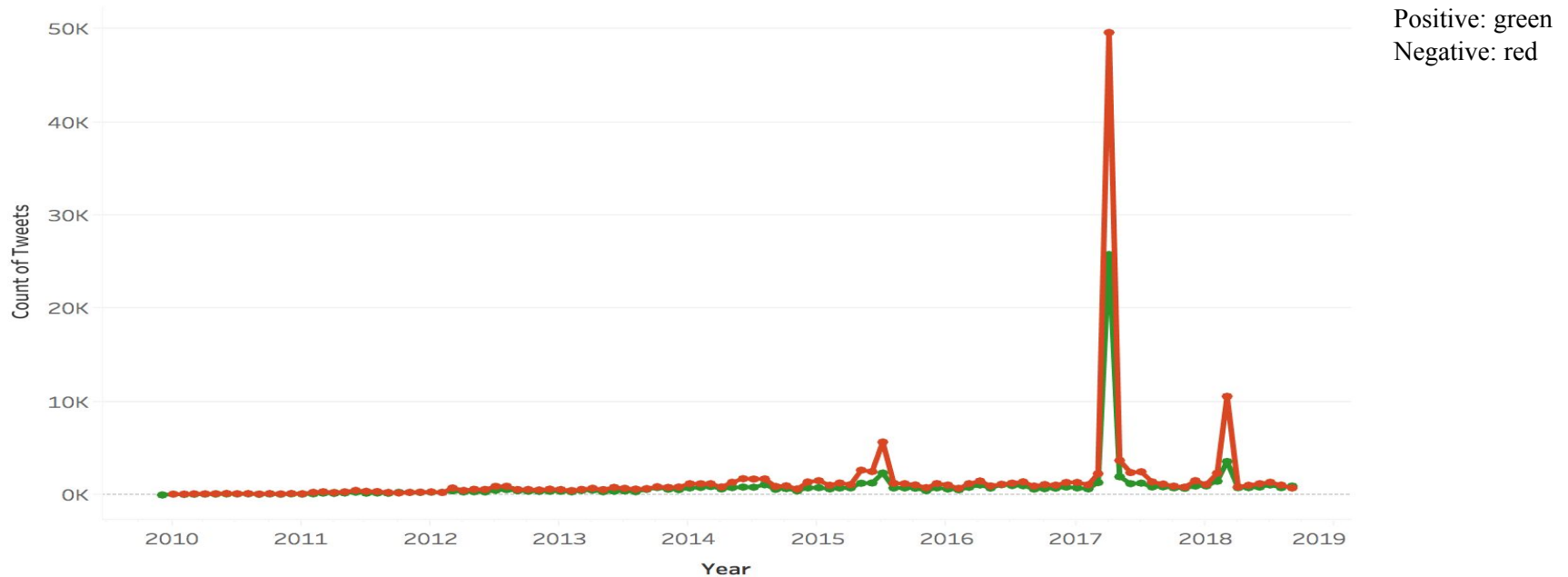
Sentiment Analysis - American Airlines

Number of Tweets (Positive/Negative)



Sentiment Analysis - United Airlines

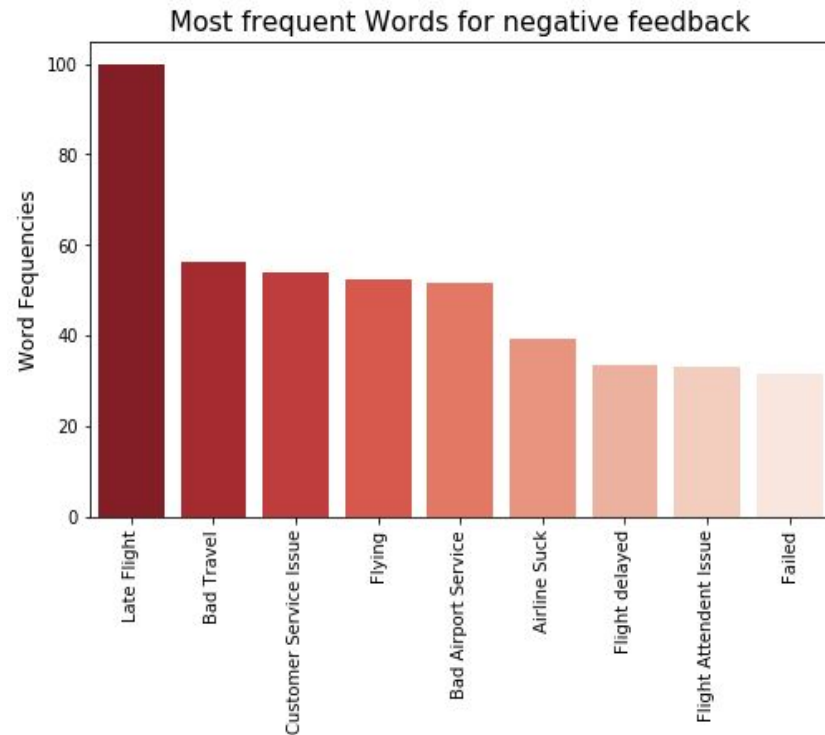
No of Tweets (Positive/Negative)



Word Cloud for Negative Sentiments



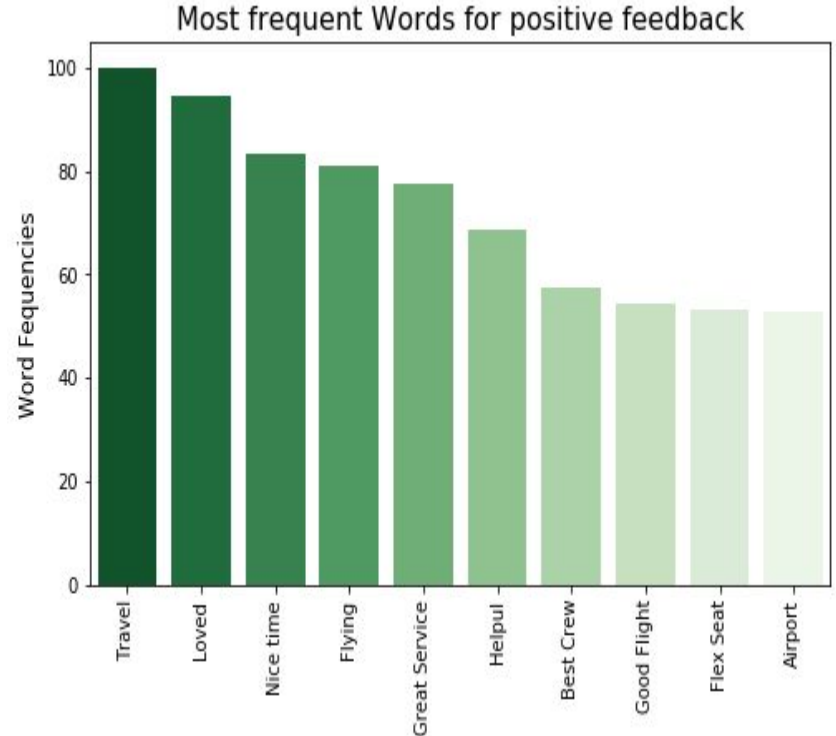
American Airlines



Word Cloud For Positive Sentiments



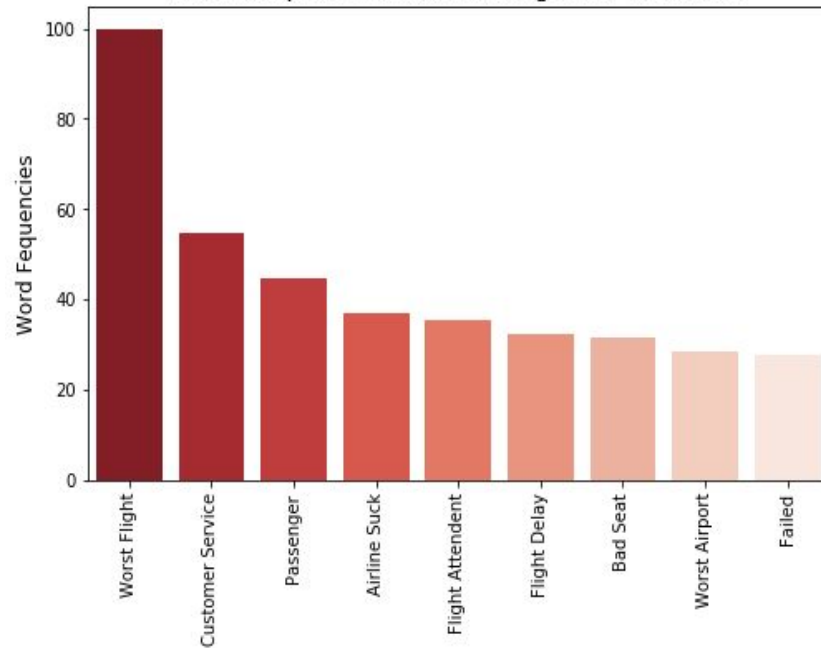
American Airlines



Word Cloud for Negative Sentiments

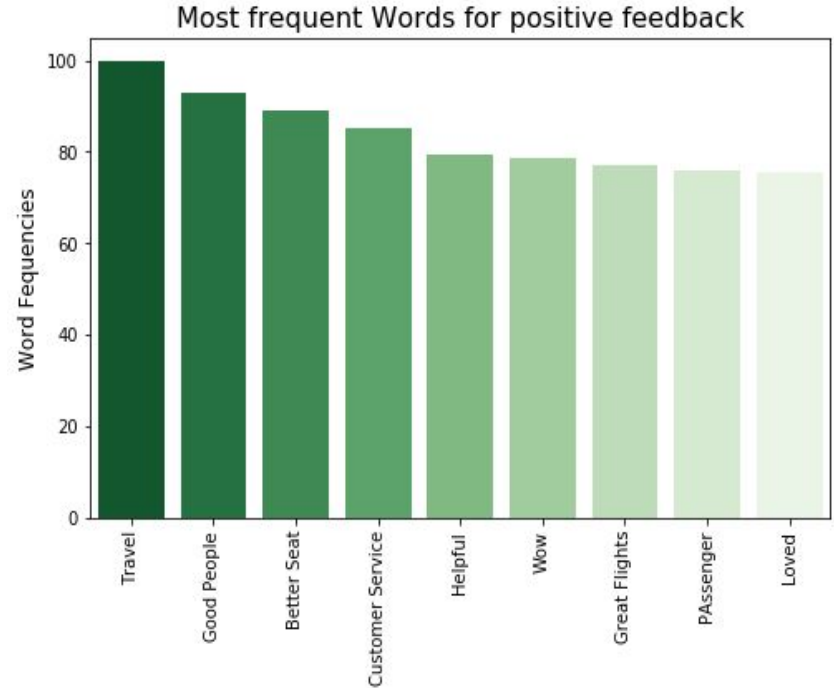


Most frequent Words for negative feedback



United Airlines

Word Cloud For Positive Sentiments



United Airlines

Stock data and polarity integration

- Integrated sentiments and stock market data based on Date
- Sample dataset

| Date | Pos Count | Neg Count | Total Count | Open | High | Low | Close | Adj Close | Volume |
|---------|-----------|-----------|-------------|-------|-------|-------|-------|-----------|---------|
| 9/27/18 | 10 | 21 | 31 | 21.41 | 21.42 | 19.10 | 19.30 | 18.54 | 961200 |
| 9/28/18 | 13 | 17 | 30 | 19.30 | 20.53 | 19.20 | 20.52 | 19.69 | 5747900 |
| 9/29/18 | 20 | 15 | 35 | 20.44 | 20.58 | 20.10 | 20.21 | 19.414 | 1078200 |

Milestone 3

- Smooth Stock market data
- Correlate sentiments and stock market movement
- Build machine learning and deep learning models
- Evaluate results

Thank You