Tweets Analysis of Top Domestic Airlines in USA: Alaska, JetBlue and Southwest

Sanzida Parvin

Introduction

Today, travel by air is the fastest means of commute, specially, for the long distance. Besides cost, comfortability is another major concern for selecting the airline by the travelers. One can find the best airline by seeing the review of a specific airline's website, howbeit, those reviews can be biased or filtered. People share their experience for a service in social media like Twitter, can be more authentic as people express their views instantly without being influenced. The objective of this project is to analyze the traveler's sentiment shared in Twitter for the three major domestic airlines in USA, Alaska, JetBlue and Southwest Airline. The outcomes of this analysis will be of great interest for both parties, travelers and airline companies. The travelers will be able to choose the right airline, at the same time, the airline companies can upgrade their services to reach the business goal by meeting the customer's satisfaction. In this project, the reactions of the travelers have been collected from Twitter, cleaned and analyzed by Python and R programming.

Data Collection and Analysis

The Twitter 'tweepy' package for Python and the R programming language were used to collect, parse, model, and analyze the data. The REST API of the tweepy package was used in Python to collect all the tweets related to the three airlines. This was done by querying tweet for 'AlaskaAir', 'Alaska Airlines', 'JetBlue', 'JetBlueAir', 'JetBlue Airline', 'SouthwestAir', 'Southwest Airlines' over the one-week time period. These querying words are frequently used to make a tweet related to those airlines. The tweets were cleaned by removing the retweets, videos, replies and duplicated tweets. The cleaned tweets were properly formatted for the sentiment analysis.

The 'textblob' package was used for sentiment analysis in python. After sentiment analysis, the result was saved as a 'csv file' to be used for analysis in R. Upon importing the data into R, a Chi-Squared Goodness of Fit Test was performed, which first assumes that each of the three airline's positive tweets are equally likely. At the significance level of 0.01, there is sufficient evidence to reject the null hypothesis, that there are not equal proportions of positive tweets about each airline. These proportions can be seen in Figure 1 and Figure 2. Since the Chi-Squared Test does indicate that the proportions are not equal, a one-sample test of proportions was performed to further confirm this hypothesis. This test compares the proportion of positive tweets at each of the airline to the equally likely probability of 1/3. According to this test, there is

enough evidence to claim that the Alaska airline got majority of positive tweets. The percentage of positive tweets to each of the airlines are shown in Table 1.

Table 1. Percentage of positive tweets to each of the airlines

Airlines	Number of Tweets	Positive Tweets	Percentage
Alaska	950	535	56.32%
JetBlue	1266	641	50.63%
Southwest	1576	848	53.81%

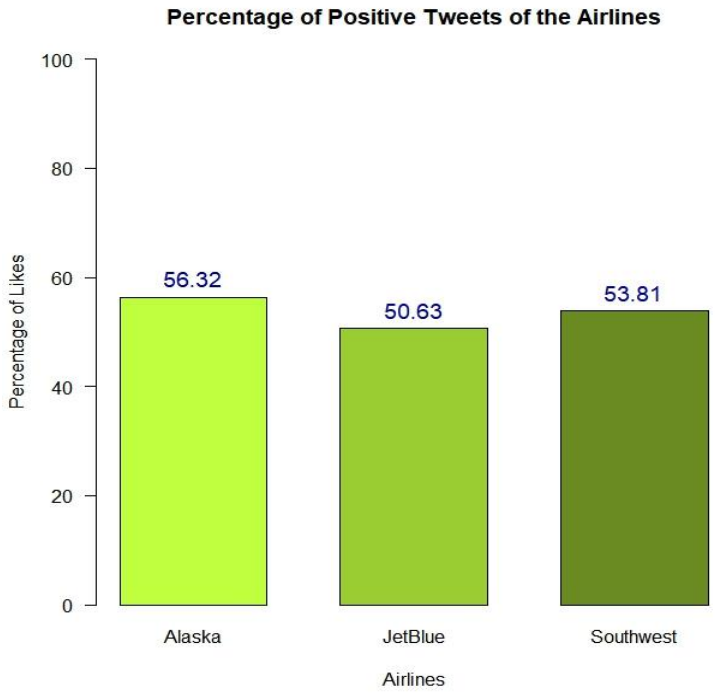

Figure 1: Percentage of positive tweets

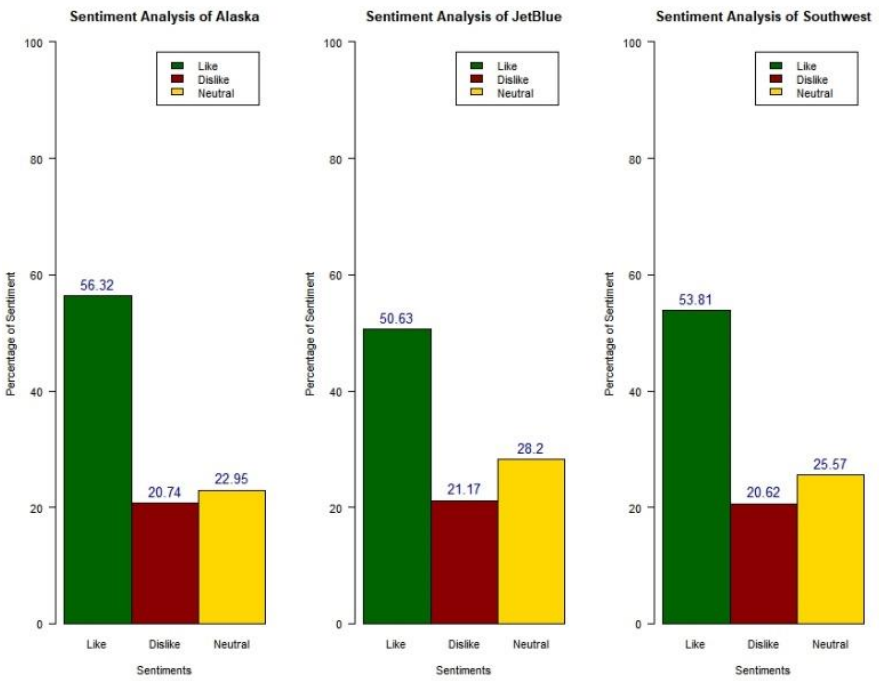

Figure 2: Percentage of all tweets

Conclusion

The view of the travelers for the three domestic airlines in USA, Alaska, JetBlue and Southwest, have been collected from Twitter, cleaned and analyzed by Python and R programming. The analysis clearly shows that, statistically, the frequency of positive tweets about Alaska Airline is greater than other two airlines.

References:

1. https://dev.to/rodolfoferro/sentiment-analysis-on-trumpss-tweets-using-python
2. https://stackoverflow.com/questions/42705314/getting-full-tweet-text-from-user-timeline-with-tweepy
3. https://stackoverflow.com/questions/12481430/how-to-display-the-frequency-at-the-top-of-each-factor-in-a-barplot-in-r