Nate Bitting July 12, 2015
PREDICT452 – Assignment 2
Twitter API – US Airline Flight Delay Analysis

**Executive Summary**

Delayed flights are inevitable when travelling by air in the US. The management of Southwest Airlines would like to do a competitive analysis for delayed flights leveraging Twitter data of customers tweeting about delayed flights on all of the major US Airlines. The US airlines used in this competitive analysis include Alaska, American, Delta, Frontier, JetBlue, Southwest, United, US Airways, and Virgin America. We will do an initial assessment to understand how many tweets regarding delayed flights occur over time as well as assess popularity of those tweets by leveraging the Twitter API.

**Methodology**

The tools leveraged for the data acquisition process from the Twitter API involved Python libraries including *twitter* (used to interact with the Twitter API) and *pandas* (used to convert JSON objects into tabular format). The fields we elected to use for this analysis included the the following:

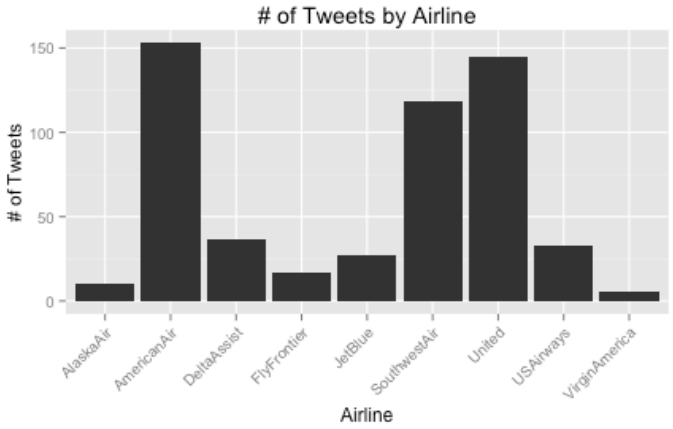
- *airline (name of airline)*
- *created_at* (date tweet was posted)
- *favorite_count* (number of times a tweet was favorited)
- *retweet_count* (number of times a tweet was retweeted)
- *user_name* (name of user who posted the tweet)
- *user_location* (location of the user, if available)
- *tweet_text* (actual text content of the tweet)
- *popularity* (calculated field *favorite_count + retweet_count*)

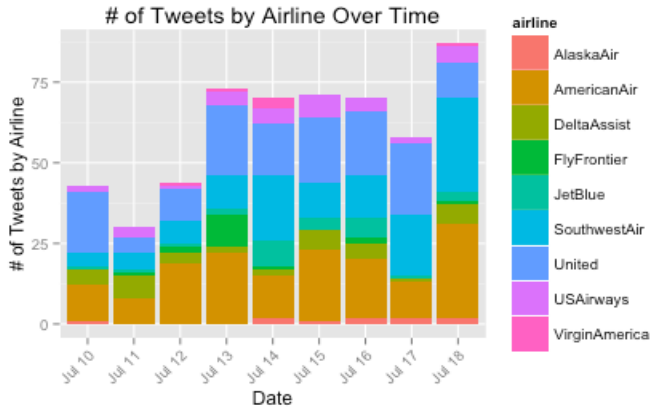
The query term used with the Twitter API was "@TwitterHandleOfAirline flight delay."

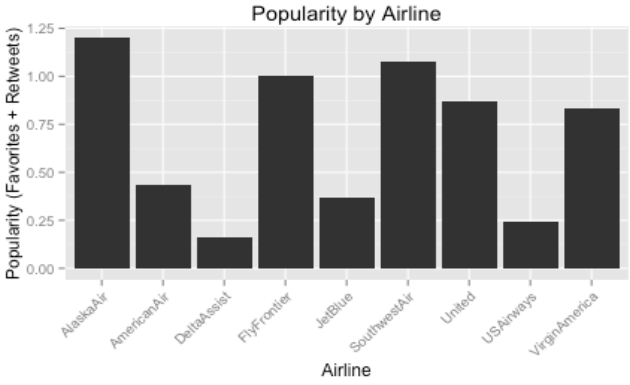
Nate Bitting July 12, 2015
PREDICT452 – Assignment 2
Twitter API – US Airline Flight Delay Analysis

**Exploratory Data Analysis**

The intent behind this analysis was to first identify which airlines had the highest number of tweets related to delayed flights and compare those results to Southwest to better understand our placement among the competition. As you can see in the chart to the right, American Airlines received the highest number of tweets related to flight delays followed by United and Southwest, respectively. We then wanted to explore if there was a trend on a particular day caused by some natural occurrence (i.e. weather) that affected all airlines. As you will notice in the 2nd chart to the right, July 18th experienced a slight spike in delayed flights with Southwest and American having the highest number of tweets. Lastly, we wanted to assess the popularity of tweets with regards to each airline as shown in the 3rd chart on the right. What was interesting to see was that although Alaska had the lowest number of tweets, they had the highest popularity. This is something we would like to explore further in a future analysis.




**Management Recommendation**

Based on what we have learned from this analysis, we clearly understand that our customers are communicating their sentiment with regards to delayed flights. We must ensure our customer service group is monitoring Twitter to ensure we can address the concerns of our customers as they are happening, not after

the fact. We also can see that although we are not the highest in number of tweets related to delayed flights, we must identify the root causes of the delays to avoid them in the future. As a future recommendation, I believe we should next explore the actual content of the text in the tweets themselves to have a better understanding of what customers are actually saying about the flight delays themselves. Did we take action to improve the overall sentiment? Are we providing full transparency to our customers as to the reasons for the delays? Does that improve the sentiment? All of these scenarios can be explored by conducting a follow-on analysis to this current assignment.