Harsha Mallajosyula Data Scientist University of California, Berkeley

INCLEMENT WEATHER AND AIRPORT DEPARTURE DELAYS

Introduction:

The scope of this analysis is to design a model that can predict Air Carrier departure delays out of New York City's John F. Kennedy International Airport. The analysis is limited to the years 2013 to 2015. Preliminary results from this analysis reveal promising insights into how machine learning can be applied to predict air carrier delays during inclement weather conditions.

New York City's John F. Kennedy (JFK) airport is one of the busiest airports in the World. In 2014, approximately 52 million passengers flew through the airport (Crain News, 2015). Nate Silver's analysis of 30 major US airports in 2015 ranks JFK 29th for outbound delays. Several other airports in the east coast also rank poorly in his analysis. Each year, New York City sees extreme weather patterns. Winters are cold, with many months experiencing sub-zero minimum temperatures. Summers are hot and humid. The North East region of the United States is also prone to seasonal thunderstorms and cyclones. Analyzing the effects of weather patterns on departure delays can help air carriers develop better forecasting models aimed at providing passengers with accurate information about flight delays.

Data Exploration:

The Bureau of Transportation Statistics maintains a record of on-time air carrier performance. The dataset is publicly available and records exist for the years 1987-2015. For the purpose of this analysis, air carrier performance data is limited to the years 2013-2015. Data for 2015 is available for the first ten months. Data for this study was also restricted to JFK. The dataset includes the *departure_delay_new* record, or outbound delays in minutes. At JFK, for the year 2013, the range for this record varies from 0 (indicating no outbound delays) to 1301 minutes (21.68 hours). In 2013, on an average, the months of June and July experienced the longest delays (22.51 and 25.48 minutes respective). Departure delays in October and November were shorter than rest of the year (at 7.71 and 7.9 minutes respectively). Average departure delays for the remaining months ranged between 11 to 17 minutes (Fig 2]. Departure delays for domestic air-carriers operating out of JFK are captured in Figure 1.

Historical daily weather data at JFK airport was obtained from the website Wunderground.com. Daily weather records include min, max and mean data for temperature, visibility, wind speed and cloud cover. The dataset also contains a categorical variable named *EVENT* that captures occurrences of inclement weather events such as snowfall, rain, fog, thunderstorm or a combination of any of these events. The weather dataset was merged with the air carrier performance dataset.

A preliminary exploratory analysis performed for the month of January does not reveal a strong correlation between minimum temperature and departure delays. Between Jan 20th and Jan 25th as temperatures fell, delays increased(Fig 3). A similar analysis performed on the variables minimum visibility (ranging from 0 to 10 feet) and departure delays shows a recognizable pattern: at high values of minimum visibility, departure delays were shorter (Fig 4).

Exploring the categorical variable *EVENT* provided interesting insights. In January of 2013, on an average, departure delays during a 'rain-snow' event were longer than for any other event or the monthly average (Fig 5). Similar analysis for the month of December is provided in Fig.6.

Machine Learning Analysis:

A supervised learning algorithm (Support Vector Machines) was selected to perform the machine-learning component of the analysis. Air carrier performance data for the year 2013 was used to train the algorithm. Data from 2014 was used to cross-validate (CV) and select the optimal parameters for the model. Finally, testing was done on the data for 2015. Features used to perform training, cross-validation and testing were weather parameters [minimum temperature, minimum visibility, maximum wind speed, cloud cover and occurrence of a weather event as defined by the categorical variable *EVENT*].

Training, testing and cross-validation labels were generated by coding up a label of '0' for a non-delay event (*departure_delay_new* =0) and a label of '1' for a delay event (*departure_delay_new>0)*. The cross-validation and testing labels were compared against predictions generated by the model to evaluate the model's accuracy. The highest accuracy score obtained after cross-validation is 0.641 (implying 64.1% of departure delays in the CV dataset were correctly classified by the model). The accuracy score obtained after applying the model with the cross-validated parameters to the test data is 0.625.

Conclusion and Limitations:

The accuracy score obtained through the machine-learning exercise is well within expectations. Several other features not included in the model's features list can contribute to air carrier departure delays including air traffic congestion, aircraft maintenance issues and security-related issues. However, weather attributes do have a big impact on departure delays. From this preliminary analysis of on-time performance data at JFK, airport authorities and air carriers at JFK can anticipate longer departure delays in case of a weather event and should develop passenger notification systems that are able to a relay delay times with great precision. One limitation of this study is that the design does not accommodate for weather conditions at destination cities, which can impact departure delay out of JFK.

Harsha Mallajosyula Data Scientist University of California, Berkeley

Appendix:

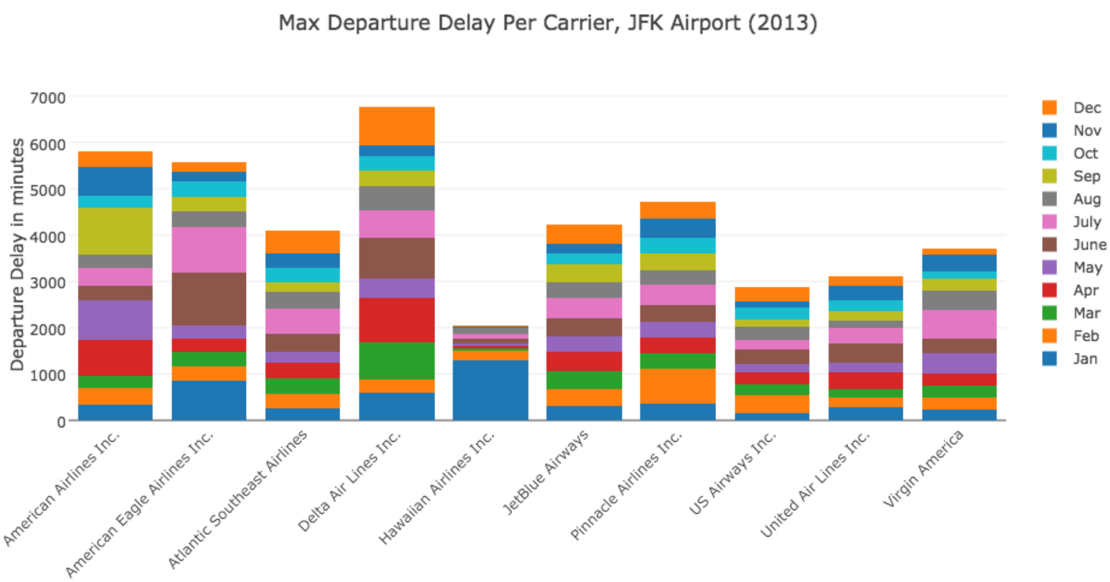

Figure-1 describes the cumulative departure delays in minutes for various air carriers operating out of JFK

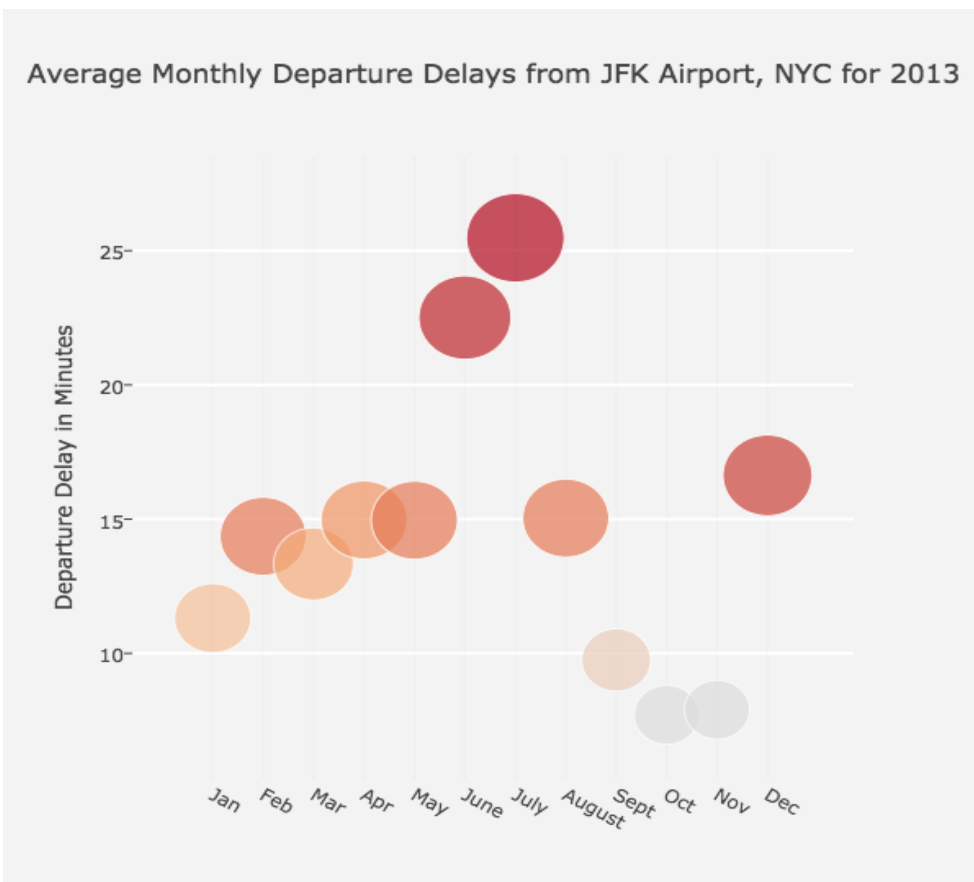

Figure-2 plots the average departure delay in minutes for each month of 2013 at JFK

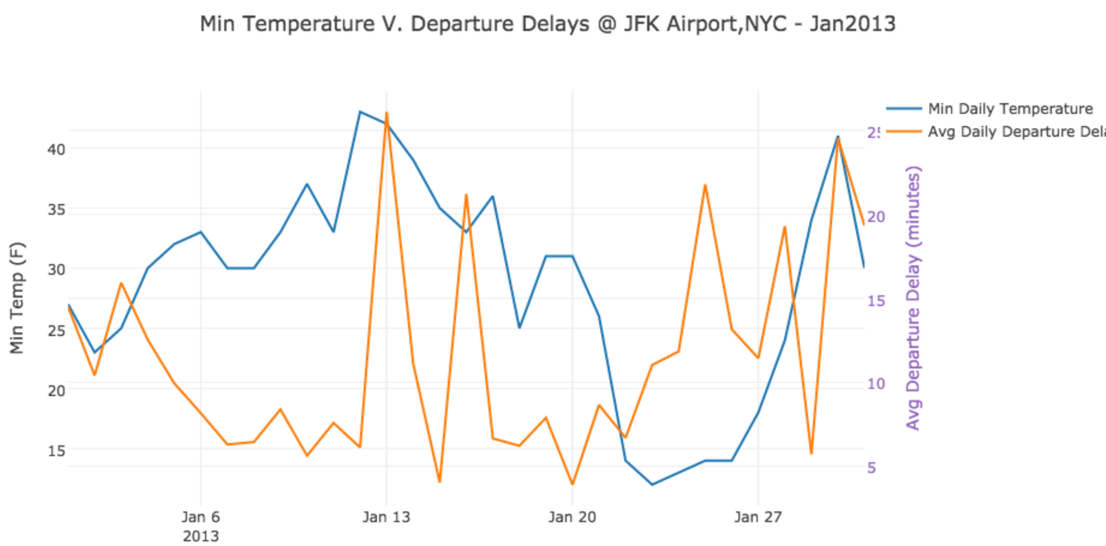

Figure-3 shows the Min Daily Temperature and Departure Delay variations in January 2013, at JFK. No clear patterns are recognizable. Between Jan 20-25 departure delays increased as temperatures fell.

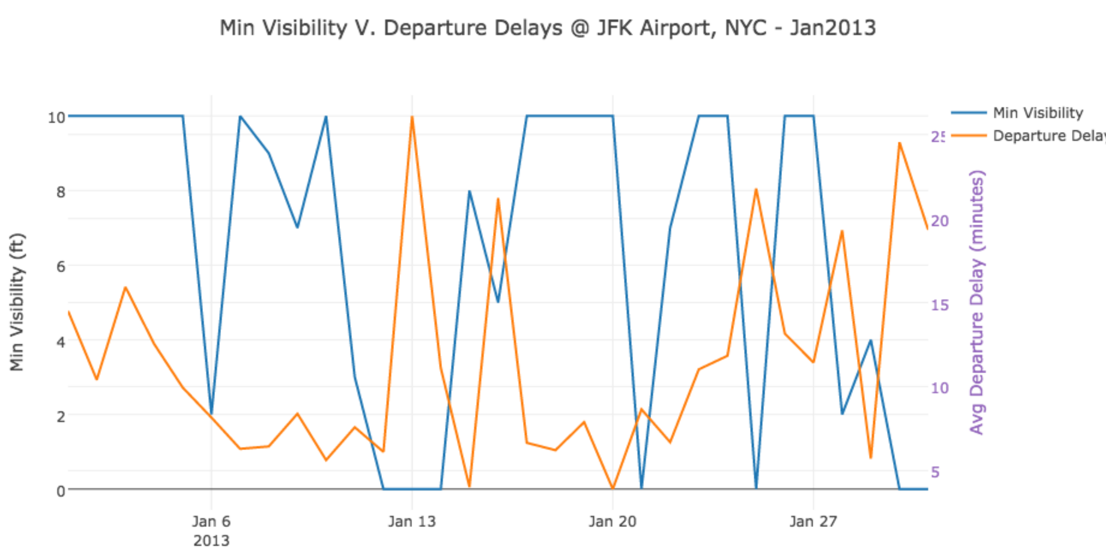

Figure-4 shows the Min Visibility and Departure Delay variations in January 2013, at JFK. Departure delays are shorter when min_visibility is at its maximum of 10 ft.

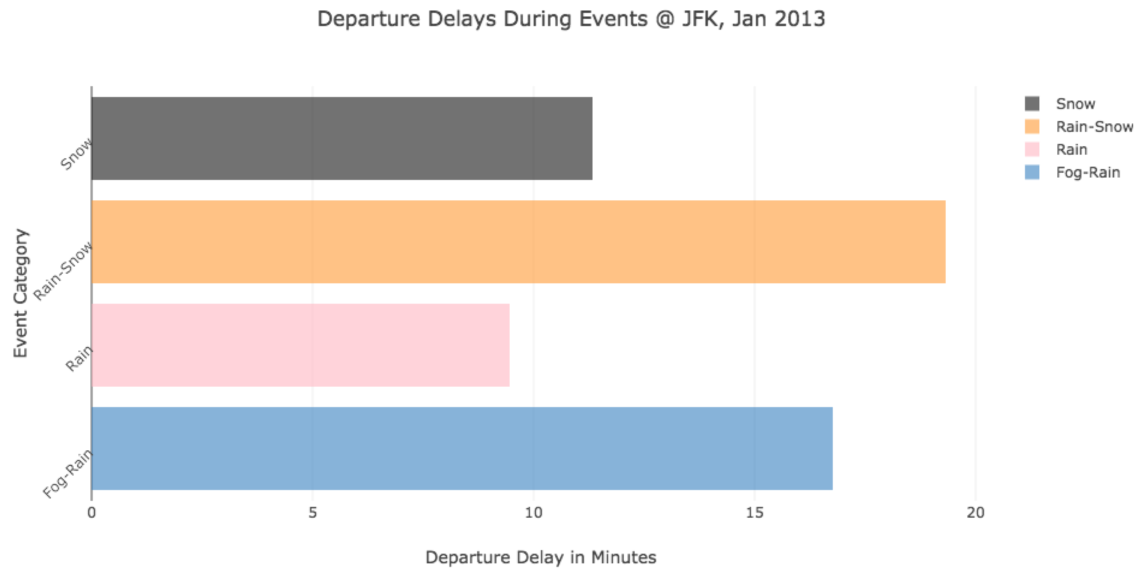

Figure-5 categorizes the departure delay experience in each of the available event categories. Data for January 2013 reveals that a traveller typically will encounter longer delays during a 'Rain-Snow' or a 'Fog-Rain' event.

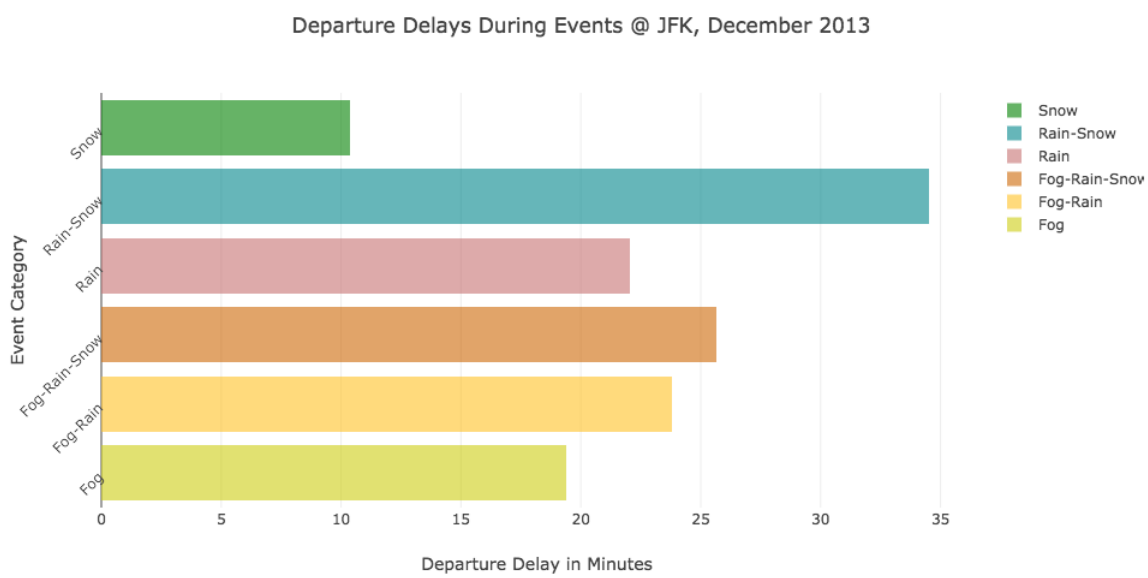

Figure-6 categorizes the departure delay experience in each of the available event categories. Data for December 2013 reveals that a traveller will encounter longer delays during a 'Rain-Snow' or a 'Fog-Rain-Snow' event.

References:

Ha, Yoona. "Local Airport Traffic Hits Record High."*http://www.crainsnewyork.com/*. 2015. Web. 2016.

"John F. Kennedy International Airport." *Wikipedia*. Web. 17 Jan. 2016.

"On Time: On Time Performance." Bureau of Transportation Statistics. Web. 17 Jan. 2016.

Silver, Nate. "The Best And Worst Airlines, Airports And Flights, Summer 2015 Update." *Fivethirtyeight.com*. 2015. Web. 2016.