**Predicting Upcoming Flight Delays**

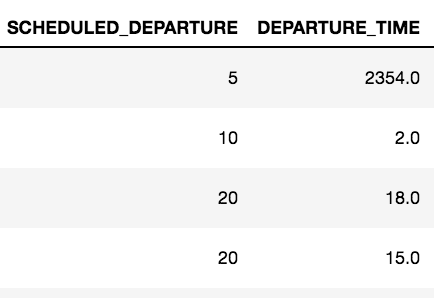
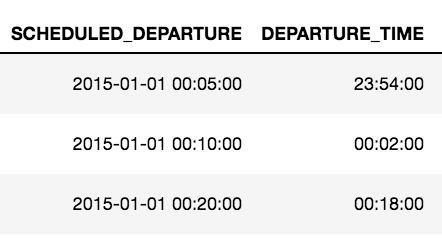
**Project Changes:**

I have not made too many changes to my project between my Delivery 2 to now. However, there are still a couple of changes I would like to note. For one, I decided to not use DarkSky API anymore. I found another API named World Weather API which seemed to be much better suited for my project. This API allowed me to pull historical weather data from January 1, 2015 to December 31, 2015 which matches the data in original flights data-frame. The API obtains many columns that will be helpful, including data in centimeters of snow, wind gust, cloud cover, wind direction degree, wind speed, etc.

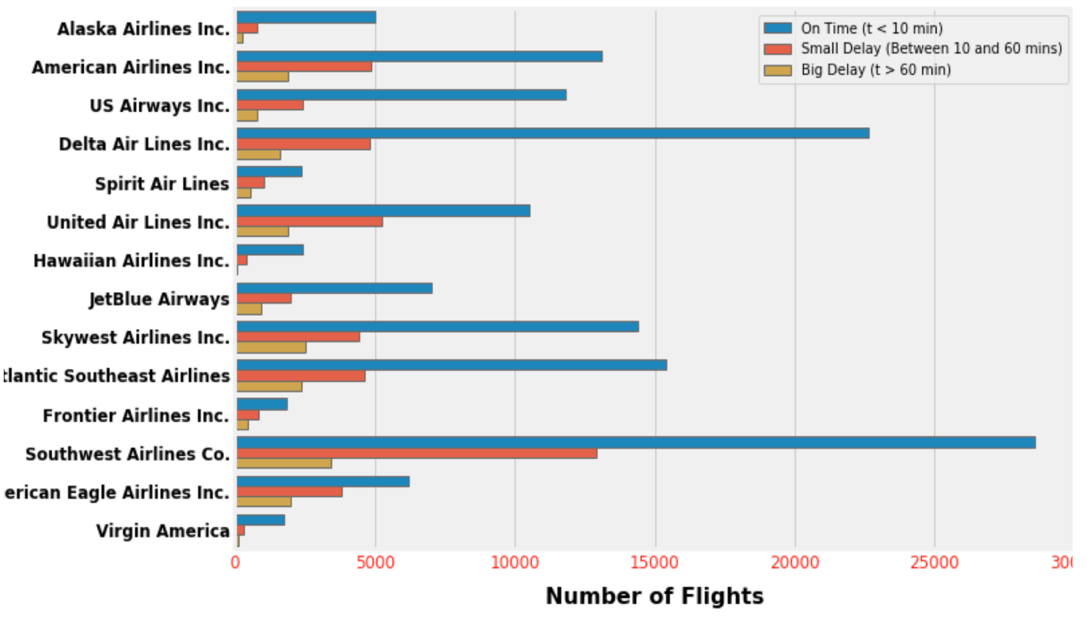
**Exploratory Data Analysis**

As mentioned in my previous Delivery, I am lucky because my dataset was already in pretty good shape. However, there was still some more data cleansing and Exploratory Data Analysis (EDA) to perform. For one, the date was split into three different columns (year, month, and day) which I then combined into “DATE.”. I also converted the “DATE” to a datetime.time. As a result, I also had to convert the date in the historical weather data-frame so that the two data-frames could match. I then joined the two data-framed based on the “DATE” column.

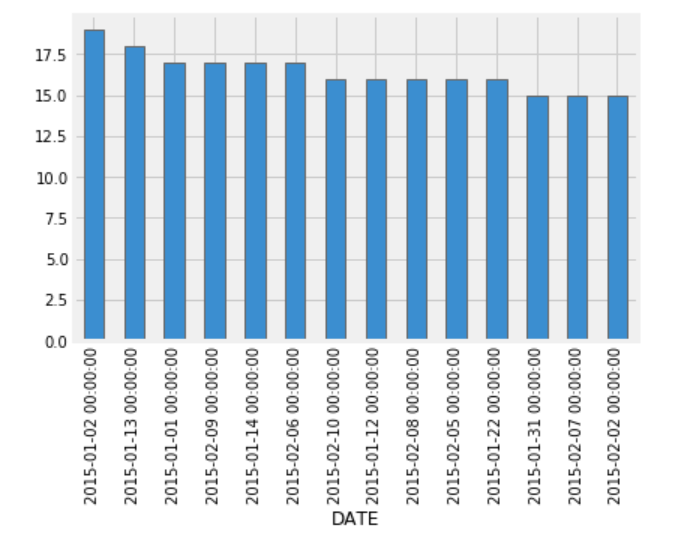
From the historical weather data-frame, I dropped many columns that did not seem necessary, such as: sunHour, uvIndex, uvIndex1, moonrise, moonset, moon illumination, max temperature, min temperature, sunset, sunrise, temperature, heat index, feels like, and humidity.

****

***Caption****: An example of how the times were formatted before (left) to how they are now formatted*



***Caption:*** *Shows the magnitude of delays based on each airline. On-time flights include delays less than 10 mins while big-delayed flights begin at 60 minutes*

****

**Caption:** *Shows wind gust levels in relation to various dates. Knowing that there are high levels on wind gust, I will be able to use wind gust in my ML algorithms*

**Preparation and Model Construction:**

Because I wanted to use certain variables for my model(s). I used label encoder. This allowed me to convert categorical variables into a form that can be used for my Machine Learning predications. I also created a matrix to then allow me to implement linear regression.

I am still working on my models. I completed a linear regression however; I still would like to visualize them and of course work on other models as well. Scikit-Learning. It changes categorical/text data into numbers. My Mean-Square-Error (MSE) resulted as 74.8 which I obviously was not pleased by.

**Reflection and Next Steps:**

I am well aware that my project is not yet complete. There is much more that needs to be done. After seeing my MSE, I know that I needs improvement. I will need to construct a better model. I plan to try a polynomial regression instead of just a linear regression in hopes of better results. Additionally, I plan to include the historical weather data in these models. For Delivery 3, I mostly performed EDA on the historical weather data, but I do plan to use them in order to create accurate flight predictions.

In my next and final deliverable there is a lot of work to be done. I plan to use other machine learning algorithms to compare the accuracy between my models. A few algorithms that I plan to play with are: Alternating Least Squares (ALS), Naïve Bayes, Random Forest, Decision Tree, etc.

**References:**

* <https://www.kaggle.com/usdot/flight-delays#airlines.csv>
* <https://www.worldweatheronline.com/developer/api/>