**Client Need Number One – A better understanding of the data from both the weather and the airport.**

While looking at both the weather and flight, I noticed that most of the delays were concentrated in three categories. The aircraft arriving late, an issue with the carrier, and NAS, which stands for National Aviation Systems, NAS delays are usually due to weather (53.6% of the time during December of 2019). Still, they can be caused by several other factors as well. Such as airport operations or air traffic control.

At HIX, delays from aircraft arriving late, the carrier and NAS made up over 96% of the delayed flights with the weather being less than 3.5% and security is well under 1%. I next focused my recommendations on the aircraft arriving late by airline, but I could not find anything significantly outside of the rations at which the aircraft arrived.

I then began to look at the percentage of flights that day vs. the days when delays occur, discovering that Monday was more likely to have flight delays than other days of the week, including weekends.

**Client Need Number Two – A model that predicts the chance that a departing flight will be delayed.**

After I had completed my exploratory data analysis of both the weather and the flight information, I tried four separate models, including Neural Networks, SVM Classifiers, Logistic Regression, and Random Forests.

Ultimately, settling on a random forest because it is explainable (which means actionable insights for the client) and achieved 84% accuracy.

**Client Need Number Three – High level summary of any insights/recommendations**

Something is going on with the team on Mondays; they have the highest percentage of delayed flights relative to any other day, while Saturday is the lowest. I’d recommend cross-training the team on Mondays with some of the high performers on the Saturdays.

Additionally, the longer the flight the more can be made up in the air. Whenever possible shorter flights should have shorter taxis, both when taking off and when landing.

I would also really look into the NAS delays, they make up 32% of the delayed flights but there is no specificity on what exactly happened when. I recommend collecting the specifics on the NAS delays over the next month and adding them to the model.

**Summary of Logic for Data Science Team in CA –**

For the model building I initially started with a Neural Network and SVM to optimize accuracy, but I realized that provided very little intuition to the client. I also choose very few features because if you use something like arrival delay, clearly the flight has already been delayed. Because of this I focused on key aspects of the weather, the days of the week, flight duration, taxi time, location they were going to and coming from.

I would also like to explore more options with weather factors, but time was running short.