*Predicting Upcoming Flight Delays*

The goal of my project is to predict upcoming flight delays based on previous delays and cancellations. My project can help airports and airlines prevent future delays and increase productivity as a result. Additionally, it can help customers know which airlines to use and/or avoid. I plan to use a dataset found on Kaggle, *“2015 Flight Delays and Cancellations.”*

The data was recorded by the U.S. Department of Transportations (DOT) Bureau of Transportation Statistics which tracks the on-time performance of domestic flights operated by large air carriers. The dataset includes information on a number of on-time, delayed, canceled, and diverted flights that were published in DOT’s monthly Air Travel Consumer Report. As of now, I only plan to use the three CSV files within the dataset. The first CSV, “airlines” is 359 Bytes and has two columns. The second CSV, “airports” is 24 KB and contains seven columns. Lastly, the third CSV, “flights” is 592.4 MB and contains 31 columns.

Before I can complete the analysis stage of the data, I plan to merge the datasets and clean up the datasets. There may be null values or unnecessary columns. Once my dataset is good to go, I plan to use machine learning algorithms, such as Alternating Least Squares (ALS) to test my data and find predicted flight delays. Although the analysis and testing aspects of my project are essential, I also want to ensure that my visualizations represent my results in an appropriate manner. I plan to play around with various visualization tools and techniques to see which one best represents my data as well as the results.