Predicting Chicago Flight Delays

DS 4002 Case Study- Ani Ponugoti

Prompt: Traveling can often be a source of stress for people, whether they're concerned about packing, being on time to the airport, or reaching their destination safely, among many other factors. Flight delays, which often happen with little notice, can add on to this existing stress. With the COVID-19 pandemic, there has been a shortage of airline employees in recent years, and flight delays and cancellations have become much more frequent. Officials at the busy Chicago O'Hare International Airport have noticed this happening on their own turf and have hired you as the lead data scientist to predict flight delays for three of their busiest and often delayed airlines– Spirit, Southwest, and Delta– to determine which airlines they may need touch base with to ensure they are departing as close to on time as possible. They have provided you with three datasets from the U.S. Department of Transportation containing 2022 flight departures from the Chicago O'Hare Airport via Spirit, Southwest, and Delta. A data dictionary and the first few observations of the Spirit dataset can be found in the paper case study folder. All three datasets can be found in the DATA folder of the Github repository.

Deliverable: Your task is to build a model of your choice to predict the number of minutes a flight will be delayed for each of the airlines– Spirit, Southwest, and Delta. You must use the data provided and use "Departure delay (Minutes)" as the response variable you are predicting with your choice of predictor variables. Format your results and insights into a presentation to give to the Chicago O'Hare International Airport management team. Your code, data, figures, and presentation must be uploaded to a Github repository. More specific details regarding this can be found in the rubric.