IST 718

Michael Morrey

Ryan Tervo

Sana Khan

Checkpoint 1

**Problem Statement:**

As a travel agency, it is a top priority to ensure the best customer satisfaction by providing the most reliable recommendations to our clients. One of the challenges we face when it comes to booking flights is recommending the best possible airlines according to client preferences. To assist our clients in booking flights on the most reliable airlines, we want to research which airlines provide the best service in different areas such as price, airplane delays and cancellations, and flight schedules. Additionally, to assist in our client recommendations, we want to build predictive models to predict whether a flight will be delayed or cancelled.

**The Data:**

Our main dataset was obtained from the Bureau of Transportation Statistics website. This dataset contains many columns relating to airline schedules and on-time performance. The date range of flights in this data set is January 2022 to November 2022, which is the most recent month we could obtain. We also obtained a dataset of airline names which was merged into the main dataset based on airline code. After this data was combined, this resulted in over 6 million rows and 100 columns. We plan on eliminating many unneeded columns and rows to reduce the size. These datasets can be downloaded at the following link:

<https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FGJ&QO_fu146_anzr=b0-gvzr>

To assist in pricing analysis, we will attempt to find a dataset relating to airline pricing that can be merged based on the airline. Lastly, we will attempt to find a historical airport weather dataset that can be possibly used to obtain predictor variables.

**The Team:**

Our team consists of Michael Morrey, Ryan Tervo, and Sana Khan. Each of us will play an equal role in the different sections of coding, such as data cleansing and exploratory analysis. As of right now, our tentative roles in this project are as follows:

* Michael Morrey: Project check-in write-ups, data cleaning, EDA, modeling, final written report
* Ryan Tervo: Obtaining datasets, data cleaning, EDA, modeling, final written report
* Sana Khan: Obtaining datasets, data cleaning, EDA, modeling, final written report