Daniel Mejia

DSC 680 – First Document

**Topic**:

For the second project, I will create a system that tries to estimate the price of flights to places in this file depending on the next features: distance, date, agency, flightType, from, and to.

**Business Problem**:

Creating a system that can predict prices of flights can be an advantage to recommend fliers flight on the best and cheapest months. This could allow a business to take advantages of prices in given months to specific destinations.

**Datasets**:

The data source will be from Kraggle. Only one file will be used flights.csv, the file contains the required data to make the prediction required to get the prices in each month. This will be a regression and not classification problem.

**Methods**:

Standard EDA will be used to explore and analyze the data. Also, a step further will be done by getting data important information to make recommendations. The data was explore using the provided function available to data frames such as describe, and similar methods.

**Ethical Considerations**:

This file does not contain personal data and ethically does not have any issue. However, the model can be bias depending on how is created. My model is only predicting prices, this will have no ethical issues. But if the model was created to recommend places depending on prices on a given month. Then, some destinations could be excluded and filter out by the model.

**Challenges/Issue**s:

The most challenging part has to do with the categorical values which will need to be converted into numerical values since machine models understand numbers better than words.

**References**:

The references will come from the documents obtain from Kraggle and the analysis obtained to get an answer. There will be another source to make the data more plausible with evidence.