Don Price

DSC 530: Data Analysis

Final Project Write-Up

**Modeling Local Flight Conditions for Small Aircraft**

The intent of this project is to model flight conditions for flying small aircraft in a local Chicago suburb. The number of monthly flights was obtained for a local airport in Palwaukee, IL. Flights counts only included local flights that are more indicative of flights taken by smaller aircraft that would be used during flight training. Ten years of weather data was obtained from the federal government databases for the local suburb.

Temperature was confirmed as the most important weather variable with average monthly temperatures explaining 47% of the change in monthly flights. Wind gusts and precipitation were not significant influences on flights, which was unexpected. Precipitation had an unusual correlation with more flights occurring during periods of higher precipitation. Presence of certain outliers likely caused this unusual result. Other weather conditions, such as pressure, did not significantly influence the ability to fly.

Flight data was only available monthly. Despite having daily and hourly weather data from the federal government, the data set was ultimately constructed monthly. This aggregation likely reduced the significance of weather, especially precipitation and wind gusts which tend to be more temporary (hourly or daily phenomenon). Had the model been constructed daily or hourly, I would expect the variables to be more significant in the model. This hypothesis assumes a pilot delays a flight during inclement weather, but they will resume flights when conditions are more favorable. These weather fluctuations are not likely captured when aggregating data to a month.

One challenge faced during the modeling was the seasonality and trend within the target variable, local flight counts. The difference in means test confirmed the more recent years had a mean number of flights that differed from earlier years. This may be due to population changes, airport expansion, pandemic impacts on flying, or just the interest in learning to fly. To correctly model the impact of weather, more research would be needed to model and remove any trend in the time series data. Seasonality was also present in the data as confirmed by my PMF scenario showing fewer flights in winter months compared to non-winter months. To correctly analyze my hypothesis, seasonality should not be removed as it’s at least partly influenced by the weather data. My limited knowledge of time series decomposition ultimately made this task unreachable for this project, but I intend to invest additional time after the course ends to revisit the time series chapters and ensure I fully understand this problem.

Overall, this was a great project that reinforced the class contents. I feel significantly more prepared, as we move into deeper research on algorithms and larger data sets, having these foundational components for EDA. I also have a greater appreciation for why we need to understand the data before putting it into a model. As mentioned above, I need to additional time learning time series analysis, but otherwise, I feel confident I understood class concepts.