Giselle Dos Santos

Professor Matthew Metzger

DSC530-T301 Data Exploration and Analysis

November 2024

**Aviation Market Analysis Summary**

**Statistical/Hypothetical Question**

The primary question guiding this analysis was: How do flight frequency, operational capacity, and route distances influence seat allocation and capacity utilization across flights in the Canadian aviation market? This question was explored through analyses focusing on variables like Departure Count, Operating Airline Capacity, Distance (km), Seats per Operation, and Operational Month. The goal was to understand how these factors relate to capacity adjustments, in relation to seat allocation based on demand.

**The outcome of the Exploratory Data Analysis (EDA)**

The Exploratory Data Analysis revealed that the Canadian aviation market is dominated by short-haul, small-capacity flights, with most operations involving fewer than 200 seats. Outliers, including long-haul flights, high-capacity aircraft, and frequent departures, were identified as key deviations requiring further analysis. A weak negative correlation between departure frequency and seat allocation suggests smaller aircraft are used on high-frequency routes. While the t-test indicated significant differences in seat allocation between high- and low-frequency routes, the regression analysis showed that departure count alone is not a strong predictor of seat allocation, emphasizing the need for additional variables to improve predictive modeling.

**Areas for Improvement in the Analysis**

One aspect that could have been improved is incorporating additional variables impacting seat allocation, such as load factor or flight duration. These variables would offer a deeper understanding of how efficiently seats are utilized and whether longer flight durations correlate with larger aircraft capacities. Additionally, segmenting data by airline or region might reveal specific operational strategies, as different airlines may adjust capacity based on demand, competition, and business model.

**Assumptions and Potential Challenges**

An underlying assumption in this analysis was that seat allocation is primarily influenced by demand and frequency (DepCount). However, this assumption overlooks operational constraints, such as aircraft type limitations or regulatory caps on certain routes, which can also affect capacity. This analysis also overlooked scope limits on airline fleets, which directly affect the fleet size and aircraft size that they can operate. A key challenge was ensuring consistency in variable names and handling missing data, affecting correlation results. Additionally, while the regression and correlation analyses provided some information, interpreting causation vs. correlation was a challenge, as external factors not captured in the dataset could influence results. Another challenge I had that directly affected my results was the process of data cleaning. If I was to redo this analysis, it would be beneficial to spend more time on this step.