Math 189 Project Proposal

Authors: Sean Perry, Sahana Narayanan, Yujin Lee, Beatrice Fernandez

Focus

Flight delays can hurt the economy and stranded thousands of very unhappy passengers for days, such as the case during the holiday season back in 2022. To ensure these issues are reduced in the future, we are interested in the design decisions behind the flight scheduling systems and how they can mitigate these potential delays. To do so, we will identify the variables associated with delays, see how different airlines/airports handle these variables, and compare the delays between airlines/airports to better understand how systems can handle delays. We will focus on the following research questions: Which airlines have successfully developed methods of preventing delays? What airports successfully reduce delays via their design? What is the probability of having a departure delay, given a departure airport and an airline? How quickly do airlines and airports reduce delays when they start occurring through a given day?

Data

The data for this project comes from the US Bureau of Transportation Statistics and can be obtained here: https://developer.ibm.com/exchanges/data/all/airline/. From this subset of the data, we will be focusing on records from the year 2000 onwards. The dataset contains many attributes that are relevant to our research questions such as departure and arrival airports, departure and arrival delays, cancellations, and types of delays.

Analysis

We want to focus on finding the most important variables behind delays and determining how these variables may differ between airports vs airlines, as well as investigating how well airports and airlines are at reducing the intensity of delays. Regression analysis and variable selection can help identify the factors in our dataset best at optimizing delay prediction. We will be investigating the relationship between airlines and departure delays, through the use of visualizations and hypothesis testing. We would then be able to distinguish which variables affect delays overall.

Expected Outcomes

Our expected outcomes would be being able to conclude which variables are associated with delays, and which airlines tend to be 'better' or 'worse' for flight delays. With some background literature about the operations into these airlines and the work we do in the project, we should be able to infer what system designs can help mitigate delay risk.