Tyler Trocchi, Jasmine Weekes, Jared Jones

MSA 8030

Team 5 - Flywise

Executive Summary

For this project, our group decided to work on the Flywise dataset. Especially as the holiday season is upon us, air travel has become an integral part of society. Air travel is often necessary in order to see family and friends, attend important events, experience other cultures and more. While all of these experiences are usually worth the price of a plane ticket, air travel itself can still be a hassle. Between booking early to avoid overpriced flights, planning your trip and itinerary, lots of work is done before you even step foot in the airport. Once your flight date arrives, even more pre-flight steps check-in and security are needed in order to make it to your gate. The last thing you should have to worry about after all of this preparation is your flight being delayed or even cancelled. We aim to combat this through investigating this business problem: how to minimize delays and predict cancellations through analytics and advanced modeling in order to increase customer satisfaction and decrease operational costs to the airline.

Flight delays and cancellations affect both the customer and the airline alike. From the customer's perspective, a delayed or cancelled flight could not only mean missing an important event, but also additional expenses from having to find alternative travel arrangements, additional food costs, and an overall loss of productivity from other commitments such as work. From the airline's perspective, it is estimated that flight delays and cancellations cost airlines upwards of $20 billion dollars every year in the United States alone. Beyond operational costs, extended delays can result in fines from governmental regulatory agencies and require airlines to compensate customers fairly. Specifically, a new law was passed in April 2024 from the

Department of Transportation, stating the airlines must automatically refund passengers for cancelled or "significantly delayed" flights without requiring the customer to initiate the process. A pattern of flight delays and cancellations also damages an airline's reputation, destroying consumer trust and leading them to pursue their flights with a different airline. With flight delays and cancellations affecting both sides of the air travel industry, our solution provides customers and airlines with ways to avoid extra financial burdens and other unwanted consequences.

The data we have will enable us to build, train and test a model that will be able to accurately predict flight delay durations and cancellations. Our dataset contains flight data from the year 2018, with 27 columns and over 7.2 million rows of flight data. Some of the key features we decided to focus on include: flight date, origin, and destination, departure time, and departure delay. In order to cut down on any unnecessary data that would create noise and decrease the model's effectiveness, we went through and counted up the null values for all the columns. From this we found that the features detailing the reason for the flight delay contained large amounts of null values, however they provide valuable data so we decided to not eliminate those columns.

We decided to further narrow down our very large dataset by establishing our role as external consultants for Delta Airlines, which gave us just under a million rows of data, about 13% of the original dataset. From this data, we determined that over 30% of these flights were delayed with an average delay of 33.6 minutes. In addition to this, we aimed to analyze the seasonality within the departure delay data. We found that the average flight delay duration peaks in the summer months of June, July and August, with additional peaks in November and January, likely due to the Thanksgiving, Christmas and New Year's holidays. The number of delays by month follows a similar pattern with a lower amount in the months of January and February, suggesting that there were fewer delays these months but that the delays were more significant, likely due to winter weather. Another element of the delays we decided to analyze was the time of day associated with these delays. Through this, we found that the number of delays in the early morning (12am-6am) were very low, with a large spike between 6am-8am and then a gradual decrease in the average frequency of delays throughout the rest of the day. Inversely, there is a spike in the length of delays during those very early morning hours, a significant decrease during 4am-6am, and then a steady increase in delay duration until it peaks again in the evening (6pm-8pm). This presents an opportunity for Delta to assess their operations, staffing, and other factors affecting delays. It would be wise to start with those early morning flights, as the data indicates that though there were fewer flights during this time, these flights on average were delayed by a longer amount of time compared to the later morning hours. If this trend can be minimized, customer satisfaction can be further enhanced as the flight schedules will more accurately represent the actual take-off times.

To aid Delta Airlines's efforts in accomplishing their business goals, we aim to build a neural network model that predicts departure delay durations and cancellations based on the key features mentioned earlier. This model will then be integrated into Delta's employee management system and flight app to proactively flag dates/times when flight delays are likely so that managers and operations can prepare accordingly and to notify customers if there is a likelihood of their flight being delayed and by how long. The motivation behind deciding to use a neural network model is because of its adaptability and ability to handle the non-linearity of the flight data. Initially, the model will be built based on this data from 2018, but once we are able to connect with the software engineering team at Delta Airlines, we can gain access to a larger amount of historical flight data and current flight information. Some important input features will include the time, route, and delay cause-related features, Typically, the data will be split 80/20 for training and testing respectively, ensuring the model can be evaluated on unseen data. Once the model is established, we will continue to work with Delta's engineers to ensure the model package integrates seamlessly into the airline's existing system structure. We anticipate this entire process taking up to 6 months to be completed.

This solution can be an important advancement for Delta Airlines in how they mitigate the challenges posed by flight delays and cancellations. By using this model, we have the potential to set a new standard in predictive analytics within the airline industry. Our collaboration with Delta will help the business meet real-world demands, and ensure the airline maintains its rank above its competitors. We anticipate that this initiative will significantly transform air travel for Delta, enhancing both passenger experience and the airline's operational success.

Works Cited

"Biden-Harris Administration Announces Final Rule Requiring Automatic Refunds for Airline Passengers." *Transportation.gov*, U.S. Department of Transportation, 15 Apr. 2024, https://www.transportation.gov/briefing-room/biden-harris-administration-announces-final-rule-requiring-automatic-refunds-airline. Accessed 9 Dec. 2024.