Project Final Report

Matthew Chi, Braden Hicks, Matteo Magcalas, Mohammud Ibrahim, Fahim Islam

Introduction

This project explores domestic flight delay and cancellation data from 2009 to 2018, aiming to uncover patterns and trends within the data, specifically what causes these delays/cancellations and how impactful each type of delay/cancellation was to airlines. To dive further into our topic of interest, we developed two Shiny apps and a total of eight variant sets of visualizations highlighting seasonality, causes of delays, and the relationship between airline reviews and performance. Throughout this project, we used datasets from Kaggle (Airline Delays, Airline Passenger Satisfaction), and divided our focus into three main questions for discussion.

Literature Review

We aim to explore multiple visualizations regarding customer satisfactions and factors that contribute to it. From drawing upon these sources through these visualizations, we hope to inspire our own designs to successfully accomplish our goals and learn tools to create effective visualizations that display the relationship between customer satisfaction and the variety of factors that contribute to it.

I. Customer Review Topic Visualizations

Figure 1: Treemap of the proportion of topics that show up most often in Tripadvisor reviews (Mehra, 2023). Includes 15,000+ reviews from the peak of the COVID-19 pandemic and before, illustrating the several factors that impact people the most on airlines.

Figure 1, from the Lufthansa Innovation Hub Analysis, is a helpful visualization of trends we can expect to see in our data. It gives us a general idea of some of the main factors that people felt strongly about regarding their flights. The article goes deeper into the three biggest factors (Flight Irregularities, Luggage, Check-in & Boarding), and as to why these factors might impact customer satisfaction on flights. This might approximate what we see in our data in relation to these indicators. However, we will be looking at additional factors such as Wi-Fi access and customer class. Since our theme of this visualization is customer satisfaction on flights, we can draw on this with our design by making another visualization which looks at relationships between factors inflight amenities, delays, luggage issues, and overall customer satisfaction.

II. Airline Visualizations

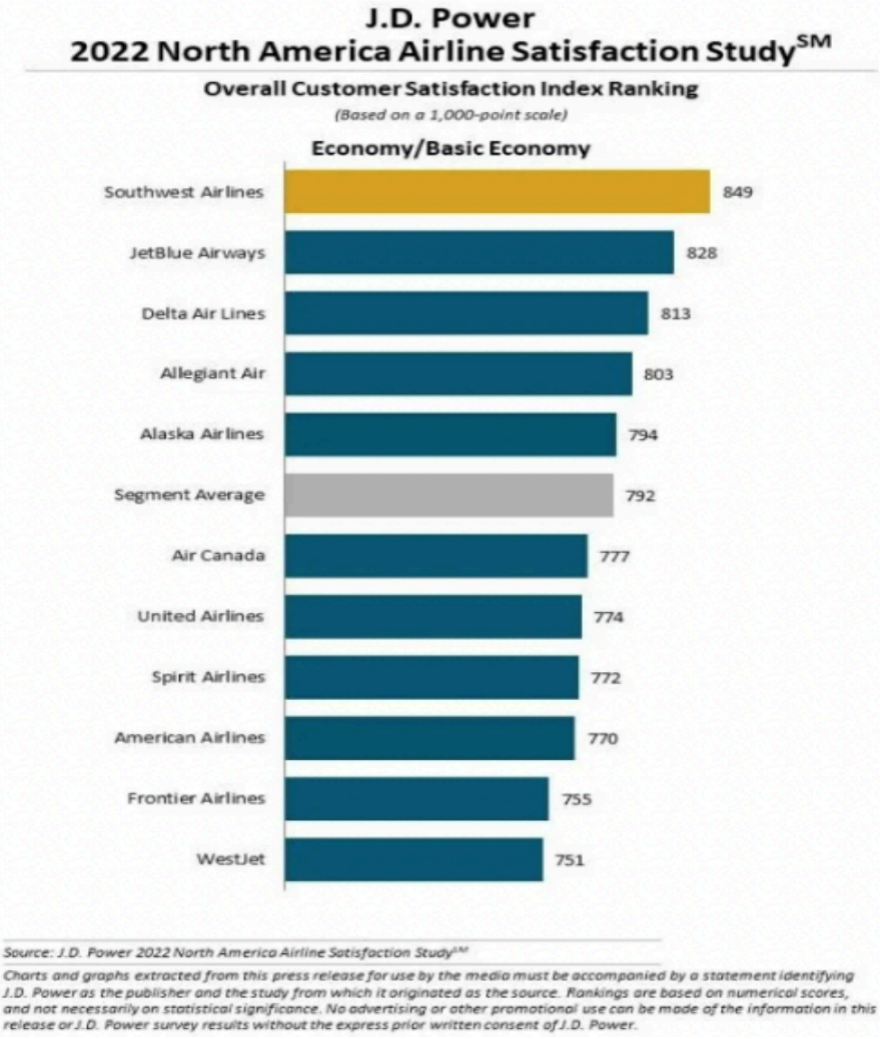


Figure 2: From JD Power. It relates overall customer satisfaction to airlines that serve the United States among economy-class customers, based on a 1000-point index.

The 2022 J.D. Power North America Airline Satisfaction Study (Figure 2) provides a detailed ranking of customer satisfaction across several North American airlines, focusing on Economy class experiences. The study covers key airlines such as Southwest Airlines, JetBlue Airways, Delta Air Lines, and others. The ranking chart uses a horizontal bar layout to compare satisfaction scores across airlines, making it easy to see which airlines are excelling (Southwest Airlines at 849) and which are lagging behind (WestJet at 751). Another visualization could focus on how different classes (Economy, Business, First Class) compare across airlines, perhaps through a combination of bar charts that showcase satisfaction by class and service component.

III. Flight Delay Visualizations

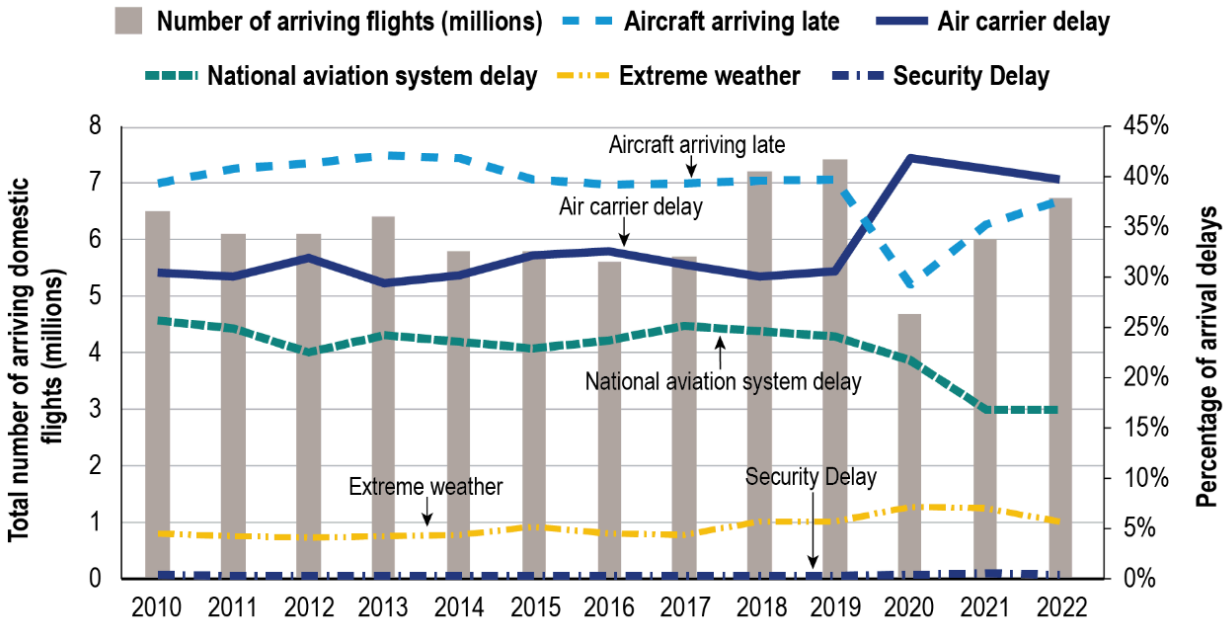


Figure 3: Percent of Flight Delay by Delay Cause (2010-2022), from the US Department of Transportation's Bureau of Transportation Statistics.

Figure 3 shows the percentage breakdown of flight delays by causes like carrier delays, late-arriving aircrafts, and extreme weather from 2010 to 2022. The data highlights how different delay factors fluctuate annually, offering insights into operational challenges and trends. For example, extreme weather delays appear to peak seasonally, aligning with known weather patterns (i.e. winter storm delays). We have been inspired by this approach and will take a similar path, analyzing and visualizing delay causes and their impacts on airlines.

IV. Interactive Data Dashboards

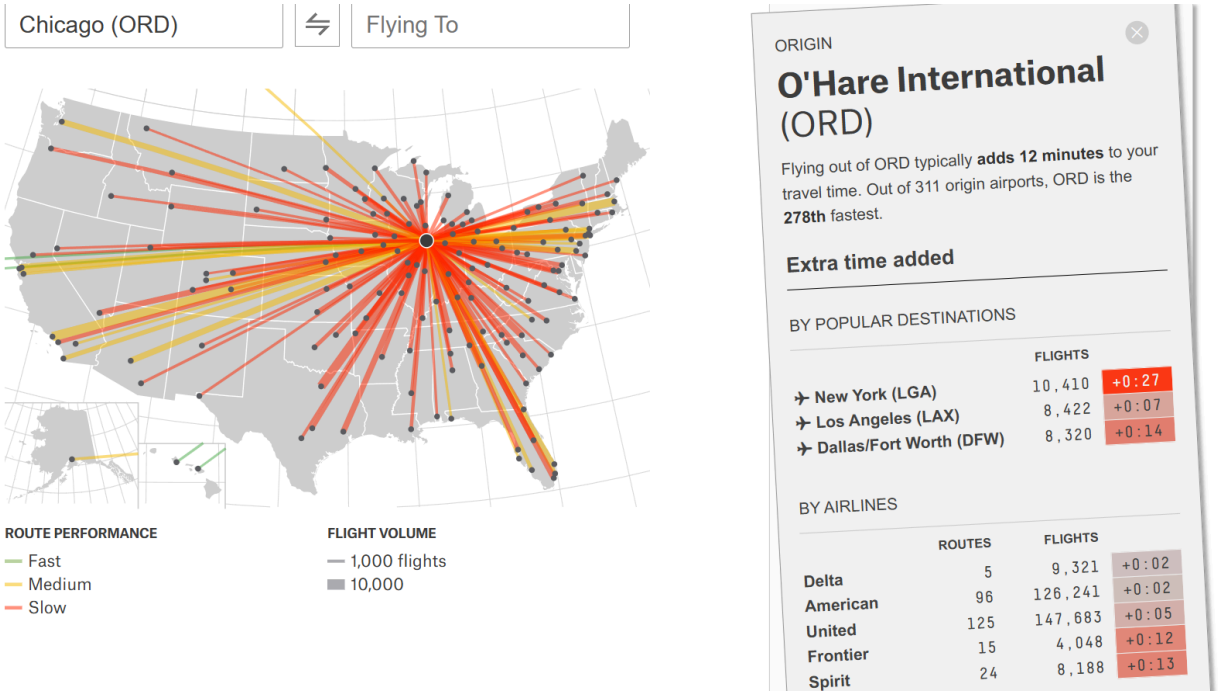


Figure 4: Interactive Flight Performance Dashboard by FiveThirtyEight (2015), analyzing 6 million flights to essentially figure out which airports, airlines and routes are most likely to get a passenger from their departure to their arrival on time, and which ones will experience delays.

The interactive dashboard (Figure 4, provided by FiveThirtyEight) visualizes airline performance through an intuitive, filterable interface. It allows the user to select the location they are departing from as well as where they are headed. It then lists the airlines that provide this service and which ones will deliver you the fastest. Figure 4 displays a screenshot of the dashboard with the departure being Chicago O'Hare, and it shows all the destinations that airlines provide departing O'Hare, as well as the amount of delay that comes with flying to certain destinations with certain airlines. We plan to also design an interactive Shiny app that enables users to filter and explore relationships between delay causes, airline performance and passenger satisfaction.

Design: Interface and Synthesis

Our objective was to present our findings in a way that would be accessible and engaging for readers of all backgrounds, regardless of their familiarity with complex statistical methods or previous knowledge on the topic of airline delays. The FiveThirtyEight dashboard included information on all departure delays, but we wanted to take it a step further and examine what caused *significant* departure delays. Significant delays, defined as a departure delay of 30 minutes or longer, are major inconveniences to travellers, whereas a delay of ten minutes is not necessarily as impactful. Drawing on insights from prior visualizations in the literature, such as

the Lufthansa treemap highlighting key satisfaction factors and the JD Power rankings of airlines, we ensured that our analysis of significant flight delays, which we will refer to solely as delays from now on, incorporated variables that have a proven impact on passenger experiences, like airline performance, delays, and inflight amenities. The image below depicts the negative relationship between the median departure delay and the average airline rating given by travellers. Knowing this, we sought to combine our background knowledge from the literature review with the data available at hand to determine what other factors play a role in departure delays.

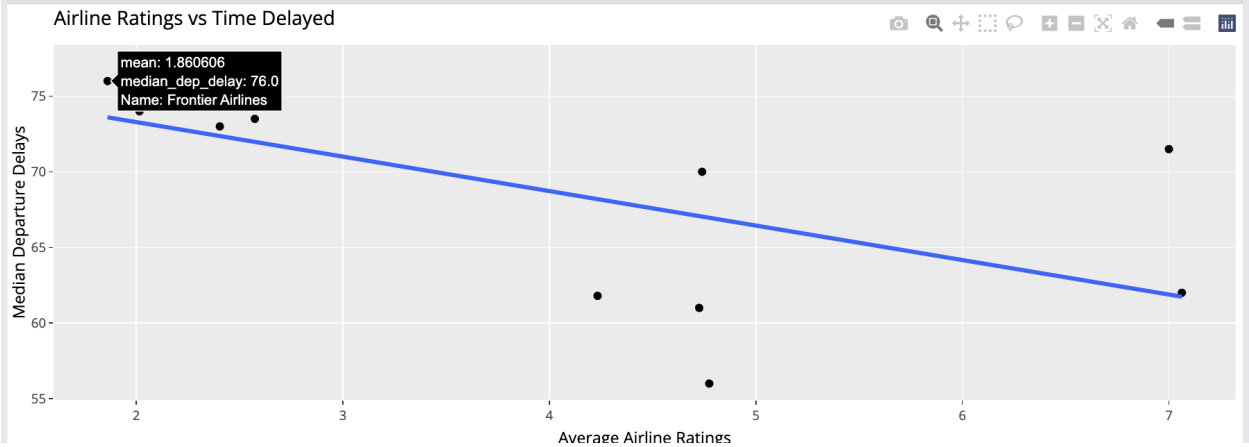


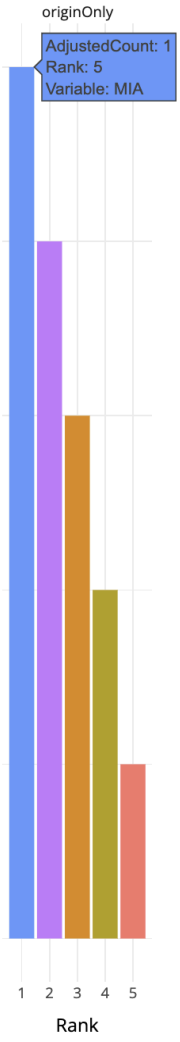
Figure 5: 2018 Airline Ratings vs. Time Delayed (Shiny App 1)

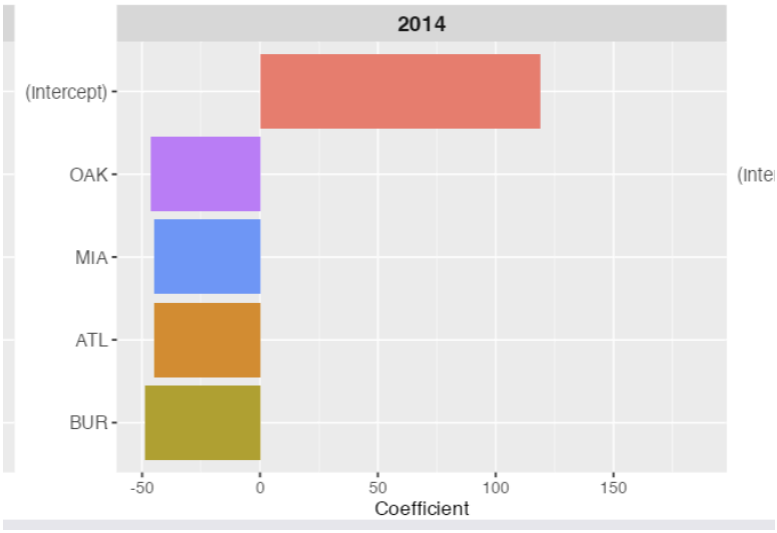
Tasked with finding the impact of various factors such as date, airline, and airport of origin, we decided to develop logistic regression models to analyze the influence of various factors on flight delays. These models enabled us to uncover patterns and relationships, such as how specific delay causes (like late-arriving aircraft or weather) align with operational challenges, much like the trends observed in the Bureau of Transportation Statistics report.

To display these insights, we developed two Shiny Apps inspired by examples like the FiveThirtyEight interactive dashboard. Our app visualizes the results interactively, allowing users to explore the impact of different factors at their own pace. For instance, they can select a specific date range, compare airlines, or focus on particular airports to see how these variables contribute to delays. This approach mirrors the user-centric interactivity of existing dashboards while incorporating additional variables. By translating complex statistical ideas into a clear, intuitive interface, our Shiny App bridges the gap between technical analysis and practical understanding, empowering users to explore the data independently.

Users also have the option to limit the amount of information on their screen. There are three logistic regression models, which determine the top five most impactful airlines, airports, and

alternative statistics (like date and in-air flight duration). They can be displayed simultaneously, displaying the maximum amount of data, or individually for a deeper analysis.





Figures 6 & 7: 2014 Origin of Flight Logistic Model (Shiny App 1)

The images above, taken from the primary Shiny App, show the origin of flight Logistic Model (2014 only). Miami International Airport, MIA, was ranked the number one most impactful airport in regards to the departure delays of a flight. The horizontal bar plot, second image, provides a more in-depth analysis of these rankings. A negative coefficient means the appearance of this airport, MIA, actually reduces the length of a flight delay. This is counterintuitive, as we, and presumably many others, believed that a larger airport would have more delays due to the high volume of flights, but in fact it is the opposite. This news is exciting, as it suggests that travellers do not have to avoid popular airlines during travel peaks in the year, in fact, it would

be recommended that they did fly through this airport. Larger airports offer many more flights, making travelling on a tight schedule much more flexible.

In the second Shiny App, delays are broken down into categories, like weather, security, and late aircraft, to name a few. These categories are tracked along airlines, helping readers determine which airlines struggled, or excelled, with types of delays.

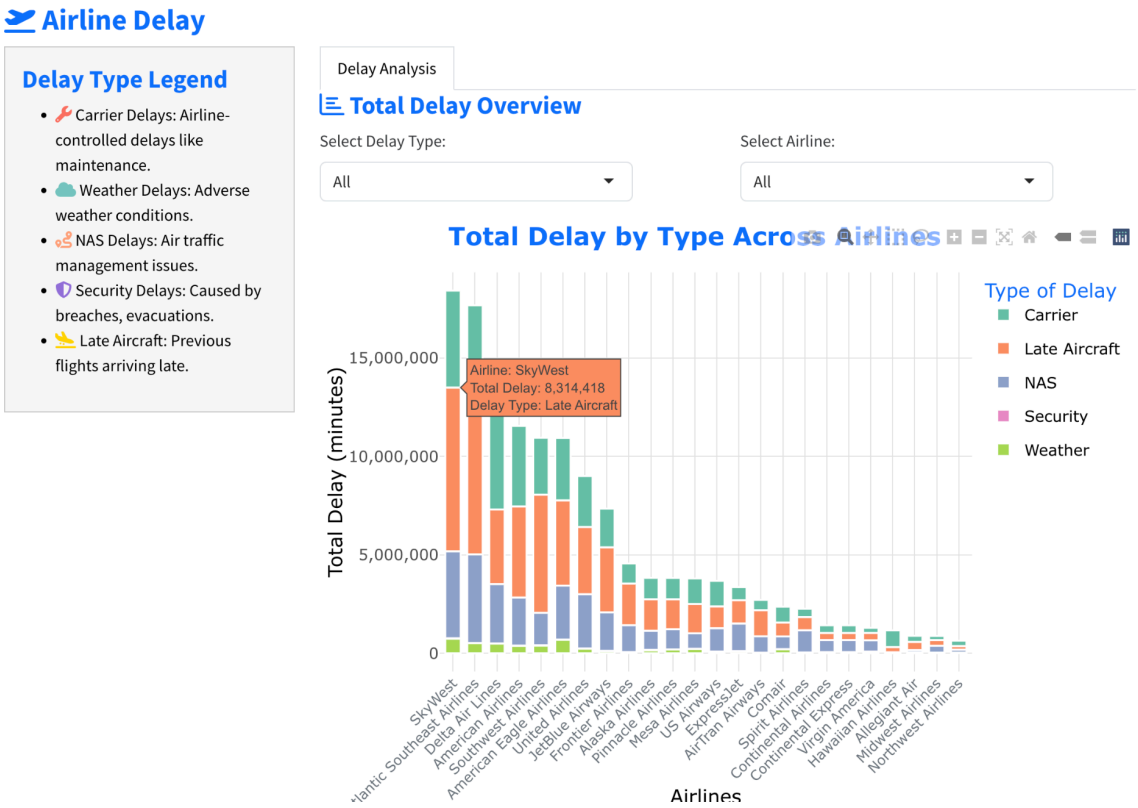


Figure 8: Total Delay Overview (Shiny App 2)

We see that this application successfully displays the types of delays at each airline, arranged from most to least delays, in an understandable manner. The popup contains more data, providing the possibility for an in-depth analysis of the application at the user's pace.

Conclusion

The goal of this project was to empower users to make informed decisions when planning flights, especially during high-traffic travel seasons like summer and winter. Our analysis revealed key insights into airline and airport performance, suggesting that some airlines and airports consistently handle high volumes of flights more efficiently than others. These findings can help travelers choose connecting flights at airports with better track records during busy periods, potentially reducing the likelihood of significant delays.

Additionally, other factors, such as flight duration (in-air), also influenced delay patterns, providing further considerations for trip planning. By highlighting these trends and making them accessible through our interactive Shiny App, we aim to equip users with actionable knowledge to optimize their travel experiences. Ultimately, we hope this project fosters a better understanding of flight delays and helps travelers navigate the complexities of air travel with greater confidence.

Works Cited

Mehra, S. (2023, December 14). A new deep dive into passenger frustration with airlines. TNMT. https://tnmt.com/passenger-frustration-with-airlines/.

North American Airline Passenger Satisfaction Declines. J.D. Power. (2022, May 11). https://www.jdpower.com/business/press-releases/2022-north-america-airline-satisfaction-study.

Bureau of Transportation Statistics. “Percent of Flight Delay by Delay Cause 2010-2022.” Percent of Flight Delay by Delay Cause 2010-2022 | Bureau of Transportation Statistics, Sept. 2023, www.bts.gov/browse-statistical-products-and-data/info-gallery/percent-flight-delay-delay-cause-2010-2022.

King, Ritchie, and Nate Silver. “Which Flight Will Get You There Fastest?” FiveThirtyEight, 24 Nov. 2015, projects.fivethirtyeight.com/flights/.