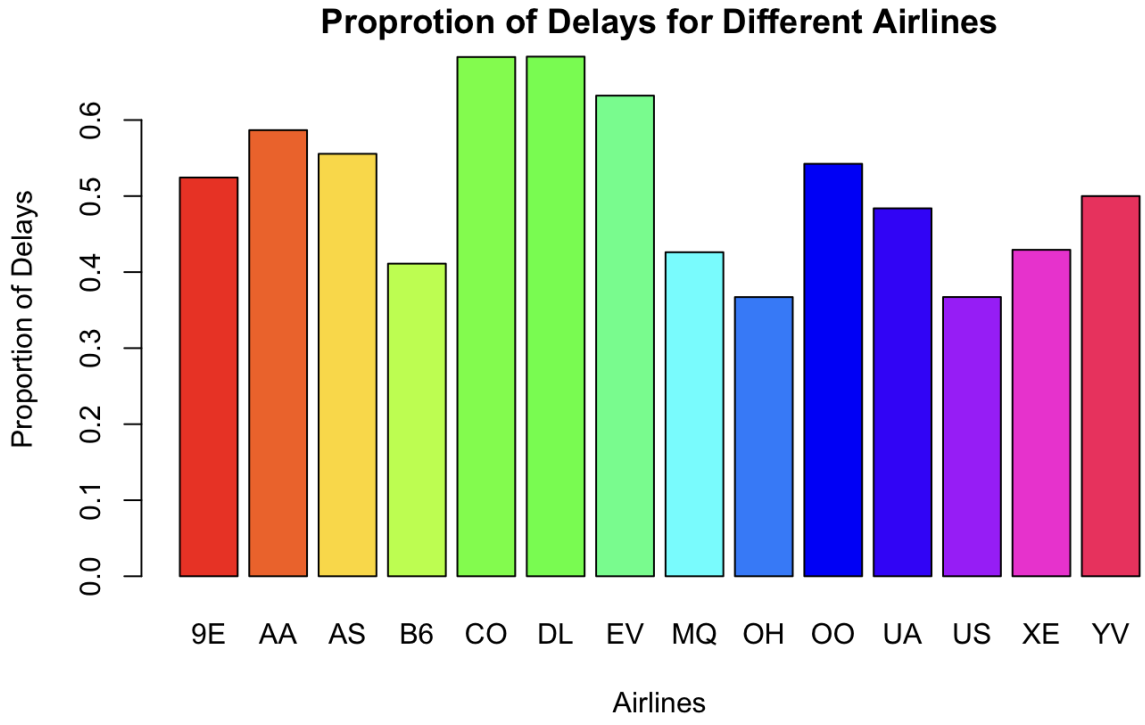
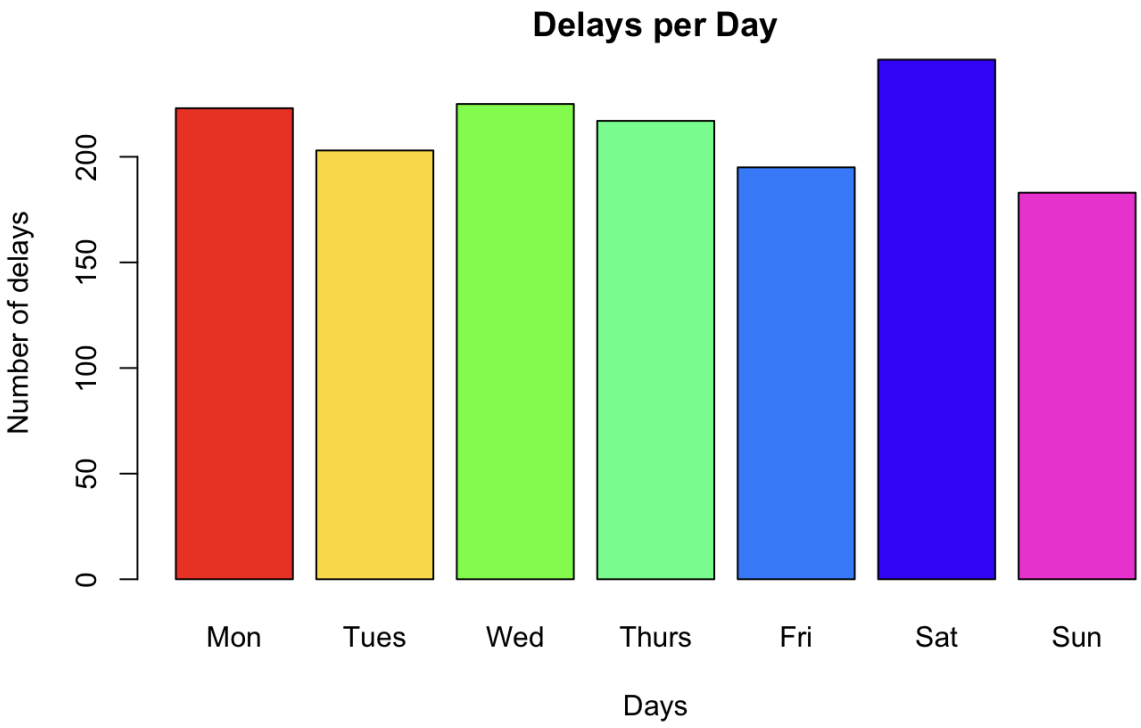
What Causes Delayed Departures At Chicago O'hare International Airport?

By: Rishi Mahadevan

Chicago O'hare International Airport is one of the busiest airports in the world, with over 54 million passengers in 2021. With such an influx of passengers, delays are inevitable to happen at O'hare. However, have you ever wondered how you can avoid those delays? I am interested in finding out how different airlines and other factors affect the number of airplane delays at Chicago O'hare International Airport. Therefore, I first downloaded an Airlines Delay Dataset. The original dataset consisted of 539,383 rows and 9 columns (4,854,447 data points). My first step was to create a new data frame and filter out Chicago O'hare as the departure airport because there were many other airports in the departure column. After filtering for Chicago O'hare Airport, I got rid of the id and flight columns as those were not necessary for my analysis. I also mutated two columns to represent the duration of the flight in hours instead of minutes, and the time of departure in a time format instead of minutes after 12am.

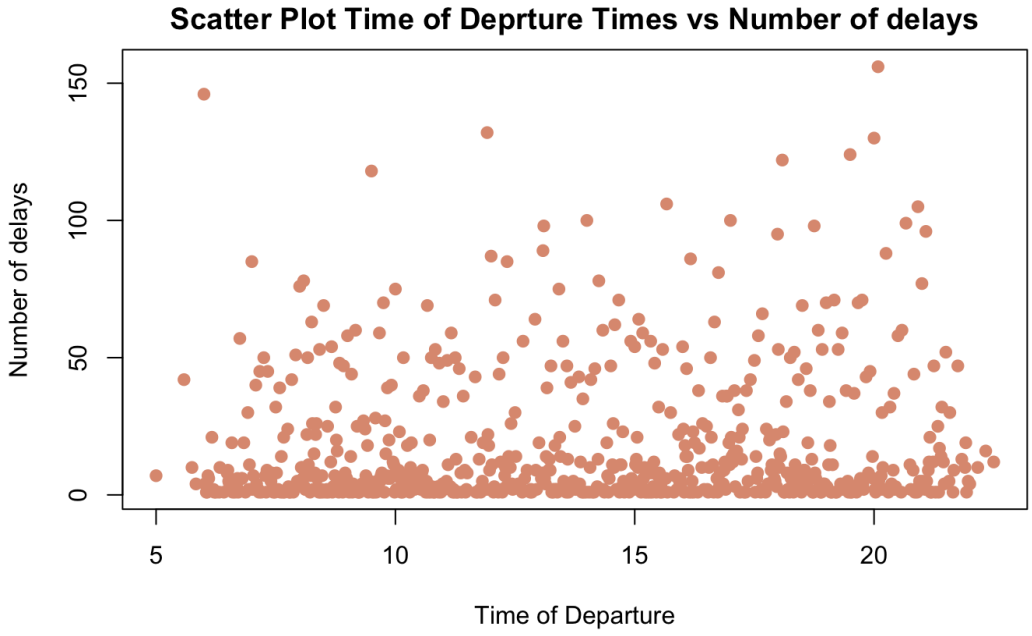


I first wanted to analyze the delays in terms of the airlines, to see which airlines have the most delays. I created a subset for each airline and found the average number of delays per flight for each airline, respectively. To do this I took the number of delayed flights and divided it by the total flights for each airline. I used a proportion to compare the airlines as each airline did not have the same number of total flights, so using the total number of delays for each flight would be inaccurate. The graph above shows the proportions and as you can see DL (Delta Airlines) has the highest delay per flight average. OH (PSA Airlines) has the lowest delay per flight average. Delta airlines is a more well-known airliner than PSA airlines, so they have a higher number of flights in comparison to PSA airlines. A higher number of flights increases the chance of a delay, meaning that it is essential to consider the airline when flying. A less popular airline will have fewer flights to operate, hence reducing the chance of a delay. The airline of operation can be the cause for a delay for a particular flight.



Next, I wanted to analyze how the day of departure could affect whether a flight is delayed. To do this I took a sample of the delays for each day of the week, as the totals were

not equal. Instead of finding a proportion I made a subset for each day and took a random sample for every subset. The samples were equal throughout, and the graph above shows the number of delays for each day of the week from the samples. According to the graph, Saturday appears to be the day with the most delays while Sundays have the least amount of delays. Also, consider that this is a sample and the number of flights varies each day.



The final part of my analysis was analyzing how the time of departure determined if a flight was delayed. I made a subset with all the airlines and filtered only for flights that were delayed. The scatter plot above shows the number of delays for each time of the day. The time is represented in army time, ranging from 5 am to 10 pm (5-22). As shown the time with the most delays is around 8 pm, and the least amount of delays occur around 5 am. However, more flights are in operation as the day progresses, so the delay of one flight could affect every flight after. This allowed me to conclude that the time of day matters if the flight is delayed or not.

The dataset I analyzed did not include external factors such as weather which plays a big role as well. I do believe, however, that the airline, day of departure, and time of departure

are also major reasons for delays. I learned a lot about how some important variables can affect the timeliness of any flight. For example, traveling by airlines with fewer flights in operation decreases any chance of a delay. Flying earlier in the day reduces the chance of a delay because fewer flights are flying, which means that the timeliness of your flight will most likely not be affected by previous flights. Based on my data the best way to minimize your chance of a delayed departure from Chicago O'hare Airport is to fly by PSA Airlines on a Sunday at around 5 am. You have the greatest chance of a delay if you fly by Delta Airlines on a Saturday around 8 pm. After breaking down the factors from my data set, I concluded that the airline, day of departure, and time of departure could affect whether a flight is delayed or not.