

Analysis of Delays in Flights - Visualisation Project Component 1

Maria Paul Thurkadayil

2024-12-08

Abstract

Airline delays cause significant losses for the aviation industry, impacting airports, airlines, and passengers. By analyzing the data it is found that around 45% of flights are delayed, with the highest frequency during midweek. The most and least delayed flights are those operated by “WN” Airlines and “HA”. Punctual flight performance is crucial for the aviation industry’s proper functioning, resulting in happier customers, increased profitability, improved efficiency, safety, and dependability.

Introduction

In the modern world, the aviation industry significantly influences the social and economic advancement of a nation or region. In today’s world of rapid change, the many amenities offered by airlines have begun to assume a critical role. It benefits our society by making it possible for people and products to be transported effectively. In addition, it promotes cross-cultural interaction, travel, and trade. The timely performance of flights in airports is of utmost importance for the proper functioning of the aviation industry. A punctual flight results in happier customers, greater profitability, improved airline efficiency, safety, and dependability. Cutting Down on flight delays can also lessen aviation’s carbon footprint and benefit the environment greatly.

Data Description

The data set contains information on whether the flights operated by different airlines from various airports are delayed or not. The data pertains to details about the airline, the flight number, the airports associated, the day of the week, the starting time and the duration(Length) of the scheduled flight. It contains 539383 rows and 9 columns. The link to the Kaggle page containing the data set can be found [here](#).

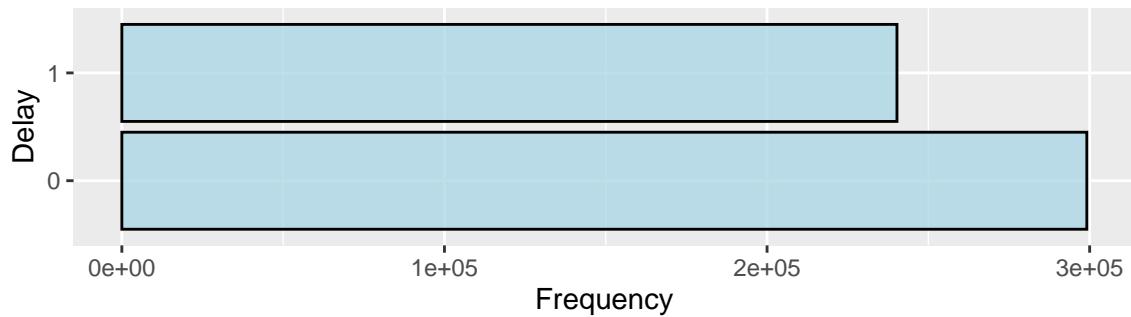
Table 1: Scheduled Departure of Flights

id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay
1	CO	269	SFO	IAH	3	15	205	1
2	US	1558	PHX	CLT	3	15	222	1
3	AA	2400	LAX	DFW	3	20	165	1
4	AA	2466	SFO	DFW	3	20	195	1
5	AS	108	ANC	SEA	3	30	202	0
6	CO	1094	LAX	IAH	3	30	181	1

Exploratory Data Analysis

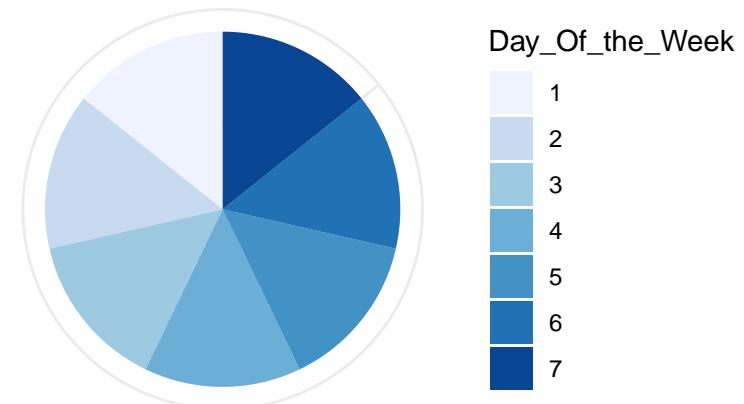
We can start the analysis by exploring the distribution of each feature across the data set and understanding their effect in the occurrence of delays.

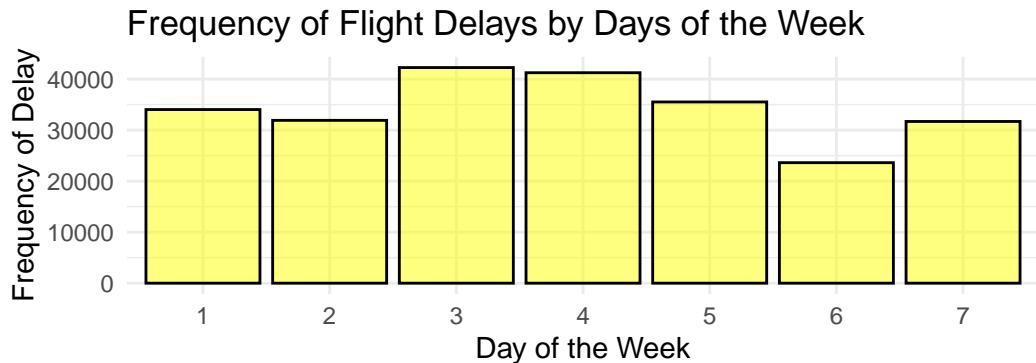
Frequency Distribution of Delay



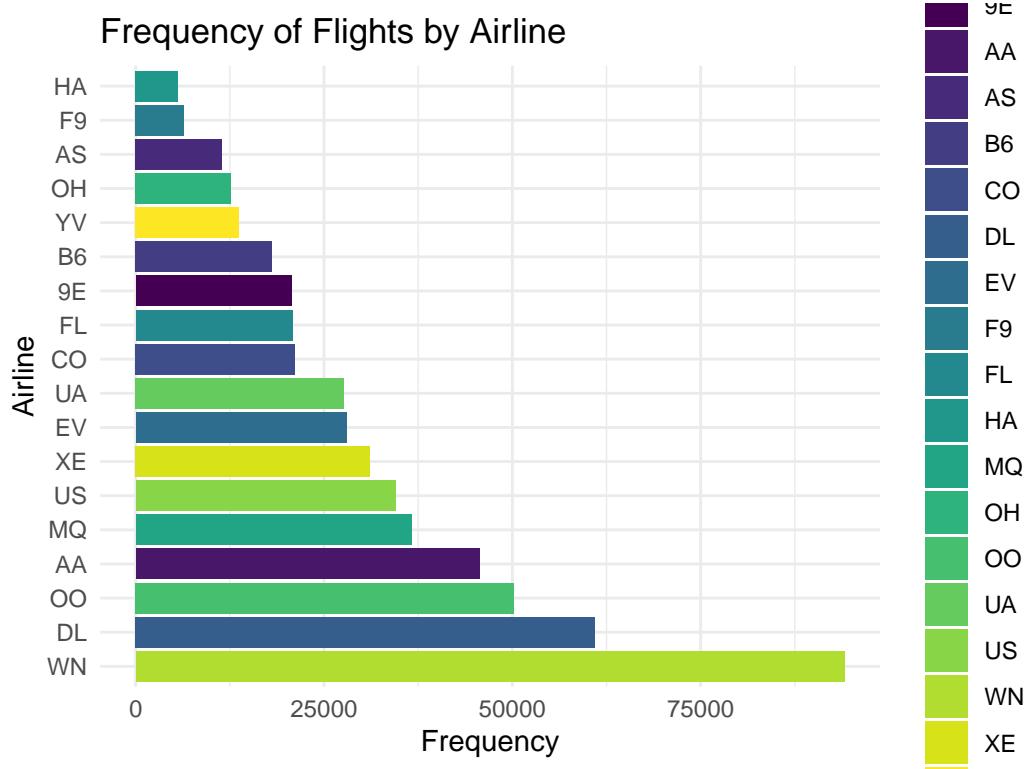
From the given data, It is observed that nearly 45% of the flights are delayed. This shows that the data is roughly balanced and the features can be evaluated to further understand their dependence on the occurrence of delays. The distribution of different features are analysed by plotting univariate and multivariate plots. Bar Plots are mostly employed to analyse the data as there are large number of unique values among features like Airlines and AirportFrom and AirportTo

Variation in Delay in Different Days of the Week



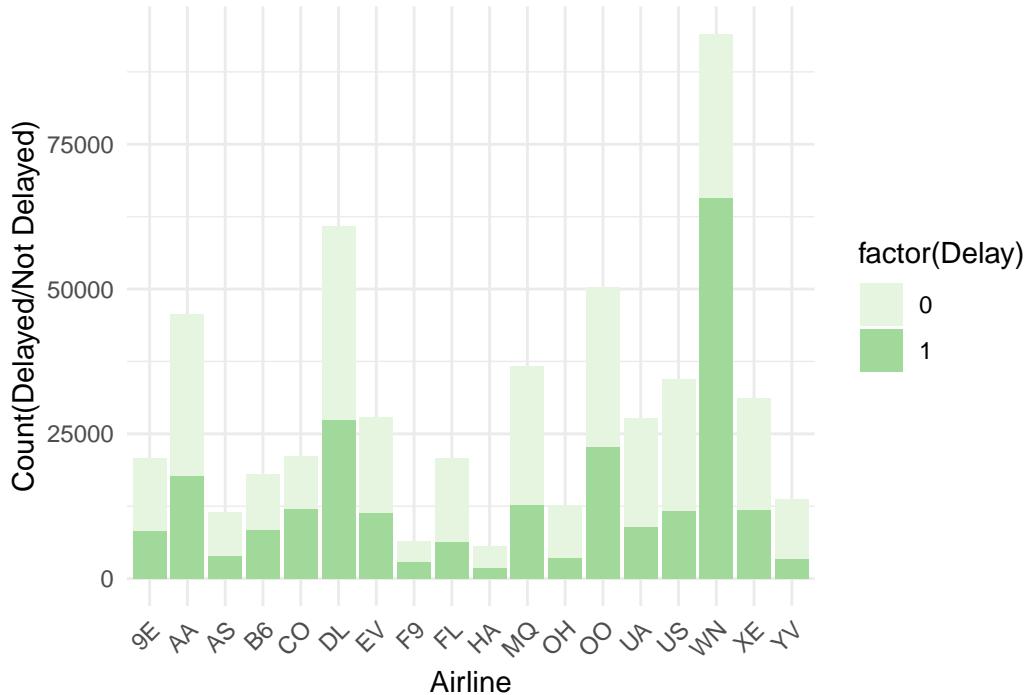


Most airline operations take place on the third to fifth day. The minimum delay is observed on Day 6(11%) and the higher proportion of delays are on Day 3 and 4 each with 17% . This shows that delays are more scattered in the middle of the week .

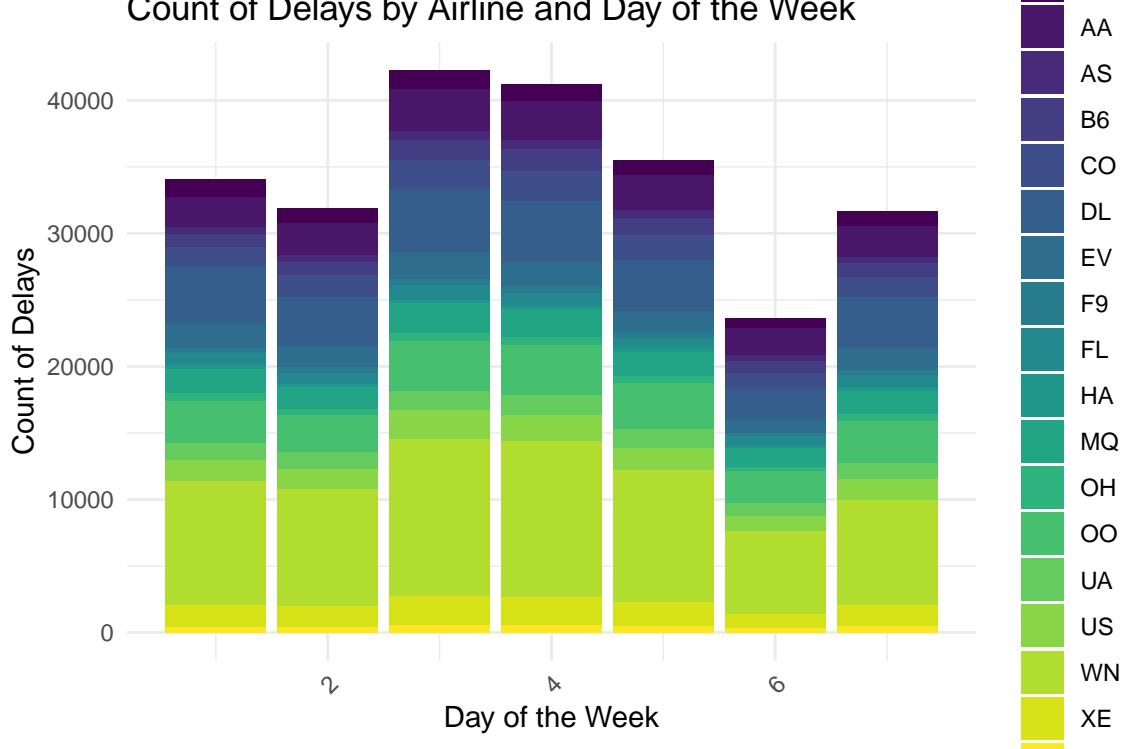


The above figure gives us the frequencies of flights for each airline. The largest number of flights is by WN and the least number of flights is by HA. There are large number of airlines in the given data and we can see that between 13-14% of the Flights are operated by WN. This shows the prominence of the given category in the flight data .

Distribution of Delay among different Airlines

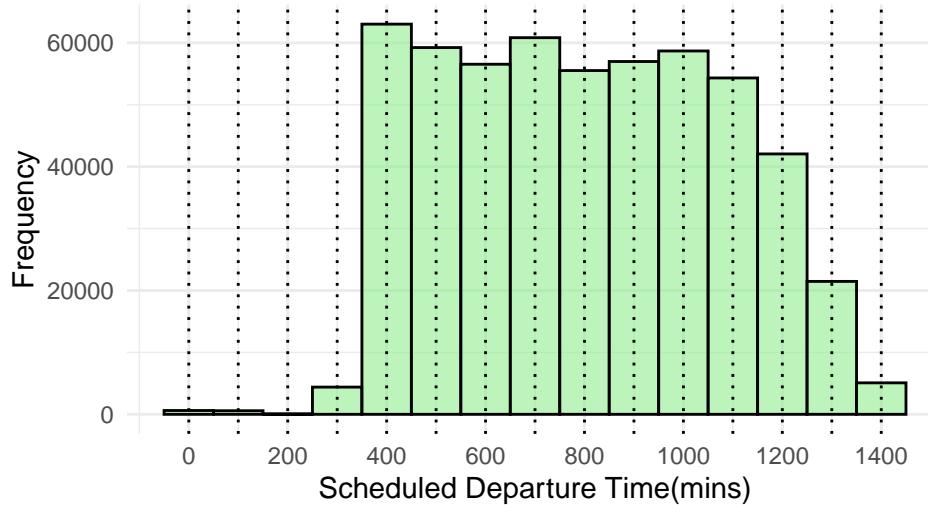


Count of Delays by Airline and Day of the Week



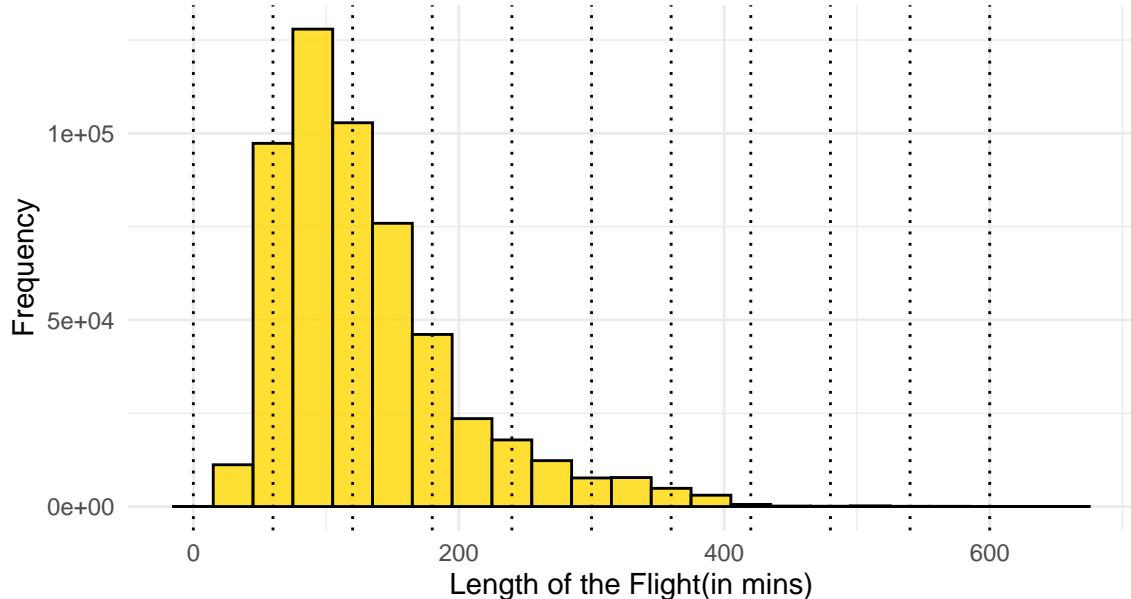
The flights operated by 'WN' Airlines are delayed the most and the least by 'HA'.The above graph depicts the variation of frequency of delays in different days of the week and by different airlines .

Distribution of Scheduled Departure Times

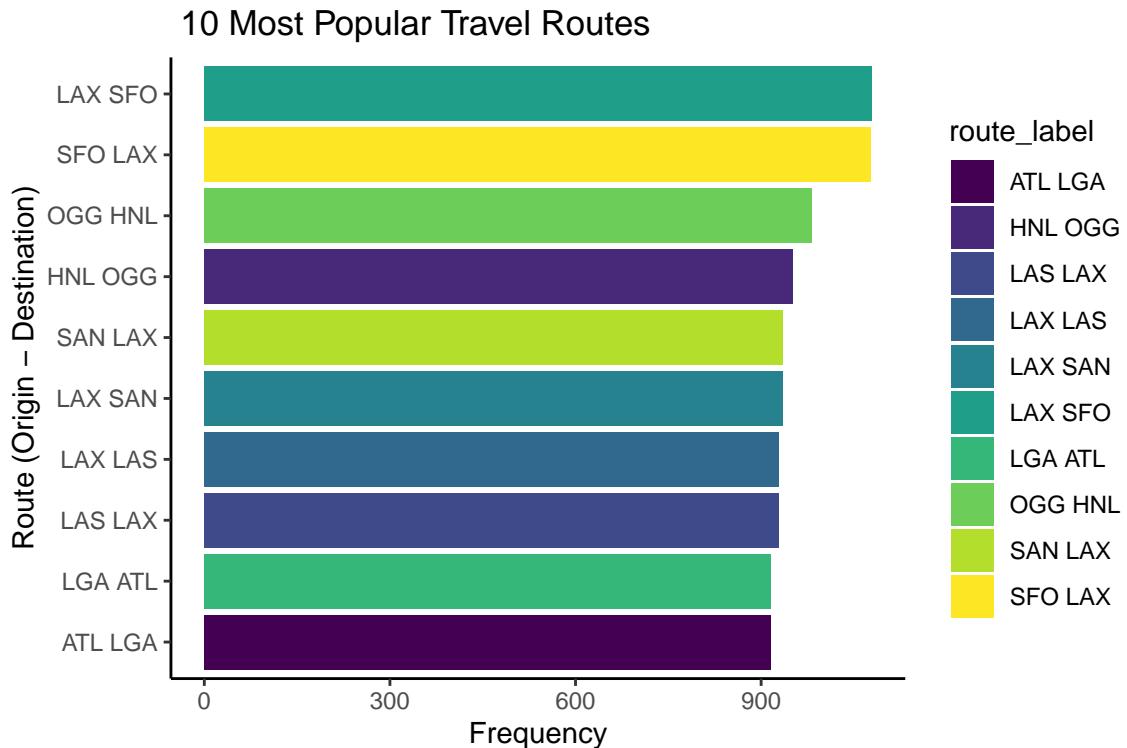


The chart shows a normal day's flights picking up around 4AM and we see a sharp increase in the amount of flights that take-off between then and the peak departure time of 8AM. From there, flights depart at a somewhat steady rate till 7PM, from which the amount of departures begin to trickle down till the end of the day.

Distribution of Length of the Flights



Flights of Shorter Duration(Length), that is less than 2 hours are higher in proportion in the given data .



It is interesting to note is that LAX to SFO(San Francisco International Airport) and its reverse flight were the most made trips with a combined 2156 flights. More than half of these flights were delayed though.

Results and Discussion

The major conclusions drawn from the EDA are :

- Roughly 45% of the flights are delayed.
- The midweek is when flight delays are at their highest.
- On the 3rd and 5th day, the majority of airlines are scheduled.
- Day 6 has the least amount of delays (11%) and Day 3 and Day 4 have the most delays (17% each).
- The most and least delayed flights are those run by “WN” Airlines and “HA” respectively
- The highest number of flights are during the time period between 350 and 400 mins.
- Significantly few flights are being operated in the initial hours of the day. Only a Small Number of Flights operate in the initial hours of the day.

Conclusion

Worldwide airline delays are one of the many issues that cause the aviation industry to suffer enormous losses. Airports, airlines, and passengers alike often bear the brunt of aircraft delays. These factors make it important to create the appropriate plans in order to maintain the aviation sector's smooth operation. From the data we can gain some insights on different factors that are contributing to the Flight Delay