

Analyzing U.S flight delays across major airports Using descriptive and analytical statistical tools and methods

This paper will analyze flight delays across major airports in the United States. Although flight delays have majorly improved over the years they are still a major cause of concern for travelers, throughout this study we will create statistical models to assess the current position of airline carriers, airports and major causes of delay. We will be using a dataset from the Bureau of Transportation Statistics. The dataset provides information on Airline On-Time Statistics and Delay causes for all major airports in the U.S from September 2018 to September 2019.

A. Background and Motivation

The transportation industry has been rapidly growing over the past few years, especially air travel. However, this growth has been plagued by flight delays and cancellations, leading to extreme losses to the airline industry. Flight delays are not only expensive to airlines but extremely inconvenient for travelers. Yet, this year it has been unavoidable. According to CNBC, American passengers in 2019 had a “rocky summer travel season”. U.S airline cancellation jumped from 1.7 % in 2018 to 2.4% in 2019. In addition to this, in June 2019 on- time arrivals fell down by 3.1% when compared to 2018 ^[1] According to the *Air Travel Consumer Report* issued in November 2019 by the United States department of transportation ^[2], flight delay has become the one of the main reasons why customers are dissatisfied with airlines. In the meantime, according to the research by Ball Et all., flight delays in the United States have resulted in a loss of around 4 billion dollars in terms of domestic GDP, there’s also an 8.3-billion-dollar loss incurred by airlines as well as a 16.7-billion-dollar cost to travelers ^[3]. Generally speaking, frequent flight delays have had a negative impact on travelers and airlines. As a consequence, we saw the need to look deeper into this data and further understand and analyze the current issue.

The aim of this paper is to:

- Identify the average flight delay times and delay times of individual airlines.
- Identify all major reasons behind flight delays and cancellations.
- Analyze the impact of particular causes on flight delays or cancellations.
- Identify coefficients for the flight delays/cancellation and economic factors such as profits of flight companies and overall GDP

B. Cause of delays

According to statistics provided by bureau of transportation, the causes of delays or cancellation can be divided into the following broad categories.

- Air Carrier: Because of the circumstances within airline’s control. For instance, fueling issues, baggage loading, airline maintenance and cleaning.
- Extreme weather: Forecasted or existing weather conditions that may result in potential hazard like hurricanes, blizzards and tornadoes.

- NAS (National Aviation System): Several factors can be attributed to NAS, such as heavy traffic volume, airport operations.
- Late-Arrival: The late arrival or departure of the previous flights, which may lead to the late arrival or departure of the current flights.
- Security consideration: Delays or cancellations due to security consideration, such as report on potential terrorist attack.

C. Chosen Data

a) Dataset

The dataset offers information about flight delays, each month for one year. It provides the carrier, the airport of departure, and number of flights. In addition, it informs if the flight was on time, if not it provides the reason for delay; due to air carrier, weather, National Aviation System, security, aircraft arriving late, cancelled or diverted. A flight is referred to as delayed when it arrives 15 minutes or later than scheduled. When multiple causes result in a delay, the cause is based on the highest number of delayed minutes it is responsible for.^[4]

Due to the economic losses that airline delays and cancellations may cause, we also focus on the U.S. Airline Financial Data and intend to make a comparison between airline delays and cancelation rate.^[5]

In addition, 2019 North America Airline Satisfaction Study also provides information about consumer satisfaction towards airlines companies^[6]. We will identify airlines which had the worst delays and those that had the least delays, and determine whether these figures are related to each other.

b) Tools incorporated

Python is an interpreted, high-level, general-purpose programming language. Python provides a wide variety of statistical including linear and logistic modeling, regression analysis, classical statistics tests, time-series analysis and graphical techniques, and is highly extensive. We will be using python to analyze figures and perform some statistics tests as well as regression analysis. We also intend to use python to visualize column charts, pie charts and histograms to summarize the data.

D. Representing and Analyzing the Data

a) Data Visualization and exploration

The first step of this study is to provide a graphical representation of carriers and flights at each of the major airports in the United States. This will provide a concise visualization of the distributions of airline carriers at each airport and how many flights each carrier has a month. This will be done by a frequency distribution graph of carriers and pie chart providing

percentage of flights of each. To further visualize this step we will highlight the most used airline carriers and most populated airports. This will be done by creating two scatter plots to illustrate the top 10 most populated airports and top 10 airlines across the U.S. To allow these graphs to be easily interpreted and analyzed we will be using python to showcase and assess our research.

The next step in this analysis is to take a look at seasonal patterns in flights across major airports. This will be done to assess any characteristics or patterns that occur of flight popularity. This will help us to understand if airport capacity contributes to any of the flight delays. We will use a scatter plot to represent these trends. It will include all major airports, months in the year and number of flights. This will help us identify various trends based on the peak and downfall information we receive. It will answer questions like “are delays more frequent during holiday seasons?”. Next, we will create a scatter chart to further investigate this. The x-axis will be average delay times. This will further lead us to information on when the most delays occurred, which seasons had the highest delay times and more. Next, we will look at average delay for each carrier, this will be done by creating boxplots and visualizing the distribution of delays. The same will be done for on time flights. We will be able to further analyze this using IQR. Based on this we can create a histogram illustrating departure delay using skewness and kurtosis.

Now that we have provided clear visualizations on the data available, we can begin further analyzing the delays and their causes. We will start by creating a pie chart for major airports providing a percentage of reasons of delay for each airport and an additional one for each airline. Based on this information and the location of the airport, we will be able to understand whether these factors are controllable by the airline or airport and what changes they can make to lower the occurrence of these delays.

b) Descriptive Statistical Measures

We will analyze the figures in the dataset individually and compare the durations of delays by comparing the mean, mode and median, which will help us to understand which reason will cause the longest duration of delay. We plan to adapt the figures to a distribution model and figure out whether it can apply to a pattern, and for a specific cause towards delays or cancelation, what will be the most common duration, and give advices on how to shorten the duration.

We will be using hypothetical analysis method to discuss the relationship between flight delays or cancellation with financial performance of flight companies. Hence, we put up the following hypothesis.

H_0 : The flight delays and cancellation will have a negative impact on profits of flight companies

H_1 : The flight delays and cancellation will not have a negative impact on profits of flight companies

Our assumption: Without considering other factors, if a company has higher flight delays and cancellation rate in 2019 compared to 2018, it will affect the profit of the flight company to a certain degree.

We also put up another hypothesis.

H_0 : The flight delays and cancellation will have a negative impact on customer satisfaction

H_1 : The flight delays and cancellation will not have a negative impact on customer satisfaction

Our assumption: Without considering other factors, if a flight company has higher flight delays and cancellation rate, it will be ranked in the latter position in terms of customer satisfaction.

E. Conclusion

The data visualizations and analysis done throughout this study will provide travellers vital information that will allow them to make informed decisions about the best time, airport and airline to travel. When individuals are informed, there will be a better distribution of flights as people viewing this information will try to avoid the days where airports are meeting maximum capacity or months where there are consistent weather barriers. Viewing information similar to what's provided above will allow travelers in the U.S to have an easier experience and may result in lower costs for airlines. At the moment this analysis does not incorporate various factors such as taxi times and weather patterns. Thus, it is not able to fully provide information on contributions to delays, but we will try to further incorporate this data by combining the present dataset with data sets on other relating factors that provide deeper insights into the problem.

[1] <https://www.cnbc.com/2019/08/15/more-us-airline-passengers-are-facing-canceled-and-oversold-flights.html>

[2] Air Travel Consumer Report, Issued: November 2019, <https://www.transportation.gov/sites/dot.gov/files/docs/resources/individuals/aviation-consumer-protection/357186/november-2019-atcrr.pdf>

[3] Ann, B.G. (2010). Flight delays cost \$32.9 billion, passengers foot half the bill. Berkeley News October 18, 2010.

[4] Airline On-time Statistics and Delay Causes, https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1

[5] Bureau of Transportation, Issued: September 2019, Second Quarter 2019 U.S. Airline Financial Data. <https://www.bts.gov/newsroom/second-quarter-2019-us-airline-financial-data>

[6] North American Airline Satisfaction Study, J.D. Power Studies[2019],
<https://www.jdpower.com/business/resource/jd-power-north-america-airline-satisfaction-study>