Airline Delay Analysis

08-2022
Data Analytics Program

Shey LeGras

Executive Summary

In 2021, The United States saw over 600 million travelers through the air. When traveling by air, preflight anxiety is something that many travelers face, wondering if they'll make their flight, will it be delayed, or will they reach their connecting location in time. In this analysis we will explore airline on time statistics and delay data to provide travelers with insights of their expected travel experience.

Data Source

The data is provided by the Office of Airline Information, Bureau of Transportation Statistics (BTS). The data is said to contain scheduled and actual departure and arrival times, as well as delay reasons reported by certified US carriers that account for at least one percent of domestic scheduled passenger revenue.to perform the analysis we will look at data from 2020 and 2021. The data can be found [here.](#)
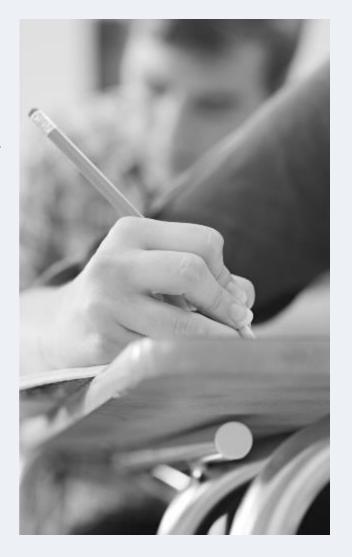
Data Limitations

The data does not contain data from all carriers, instead it contains data from carriers that make up at least one percent of passenger revenue. Airlines report the cause of delay in broad categories created by the Air Carrier On-Time Reporting Advisory Committee which can limit us knowing the exact cause of the delay. As the data is reported by the airline, we must consider room for human error as well as airlines providing false information to meet their own metrics for on time flights.

Data Cleaning

The data is provided by the Office of Airline Information, Bureau of Transportation Statistics (BTS). The data contains scheduled and actual departure and arrival times, as well as delay reasons reported by certified US carriers that account for at least one percent of domestic scheduled passenger revenue.

Data Profile

Column	Column Description	Datatype	Time-Variant
Year	Year data collected	Qualitative, Ordinal	No
Month	Month data collected	Qualitative, Ordinal	No
Carrier	Airline unique carrier code	Qualitative, Nominal	No
Carrier_name	Airline name	Qualitative, Nominal	No
Airport	Unique airport code	Qualitative, Nominal	No
Airport_name	Airport name, city, and state	Qualitative, Nominal	No
Arr_flights	Number of flights arriving at airport	Quantitative, Discrete	Yes
Arr_del15	Number of flights more than 15 minutes late	Quantitative, Discrete	Yes
Carrier_ct	Number of flights delayed due to carrier	Quantitative, Discrete	Yes
Weather_ct	Number of flights delayed due to weather	Quantitative, Discrete	Yes
Nas_ct	Numbe of flights delayed due to National Aviation System	Quantitative, Discrete	Yes
Security_ct	Number of flights delayed due to security	Quantitative, Discrete	Yes
Late_aircraft_ct	Number of flights delayed due to another flight being delayed	Quantitative, Discrete	Yes
Arr_cancelled	Number of cancelled flights	Quantitative, Discrete	Yes
Arr_diverted	Number of flights diverted	Quantitative, Discrete	Yes
Arr_delay	Total time (minutes) of delay	Quantitative, Continuous	Yes
Carrier_delay	Total time (minutes) of delay due to carrier	Quantitative, Continuous	Yes
Weather_delay	Total time (minutes) of delay due to weather	Quantitative, Continuous	Yes
Nas_delay	Total time (minutes) of delay due to NAS	Quantitative, Continuous	Yes
Security_delay	Total time (minutes) of delay due to security	Quantitative, Continuous	Yes
Late_aircraft_delay	Total time (minuets) of delay due to late aircraft	Quantitative, Continuous	Yes

Questions to Explore

What is the leading cause for airline delays?

Which airlines contribute the most to airline delays?

Do airport locations influence delay time?

How do airline delays vary throughout the year?