# 2020 Transportation Delay Metrics Report

## Camron Miene

## Introduction

The purpose of this report is to analyze the impact of precipitation on flight delays at Atlanta's Hartsfield-Jackson International Airport during the first four months of 2020. By examining the relationship between precipitation levels and various indicators of flight performance, such as arrival delays and cancellations, this study aims to shed light on the potential effects of inclement weather on air travel. This information may be useful for airport operators, airlines, and other stakeholders in the aviation industry as they seek to mitigate the impact of adverse weather conditions on their operations.

## Data

This project utilizes two primary datasets: The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) on flight delays, and NOAA climate data.

### Delays

To find and collect data on flight delays, I used BTS’s on-time performance tracker of domestic flights. The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS) tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled and diverted flights appears in DOT's monthly Air Travel Consumer Report, published about 30 days after the month's end, as well as in summary tables posted on this website. BTS began collecting details on the causes of flight delays in June 2003. Summary statistics and raw data are made available to the public at the time the Air Travel Consumer Report is released. Specifically, the data was extracted from the BTS's Airline On-Time Performance and Causes of Flight Delays data set, which are available at the following link: <https://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airline%20On-Time%20Performance%20Data&DB_Short_Name=On-Time>.

### NOAA

I used NOAA’s API to collect data from the National Oceanic and Atmospheric Administration (NOAA) for the Atlanta area for the first four months of 2020. The Global Summaries dataset, known as GSOM for Monthly, contains a monthly resolution of meteorological elements from 1763 to present with updates applied weekly. The major parameters are: monthly mean maximum, mean minimum and mean temperatures; monthly total precipitation and snowfall; departure from normal of the mean temperature and total precipitation; monthly heating and cooling degree days; number of days that temperatures and precipitation are above or below certain thresholds; extreme daily temperature and precipitation amounts; number of days with fog; and number of days with thunderstorms. The primary input data source is the Global Historical Climatology Network - Daily (GHCN-Daily) dataset.

[Web Services API (version 2) Documentation | Climate Data Online (CDO) | National Climatic Data Center (NCDC) (noaa.gov)](https://www.ncdc.noaa.gov/cdo-web/webservices/v2#datasets)

### Merged

The data includes information on flight performance for individual carriers at Hartsfield-Jackson International Airport for the months of January, February, March, and April 2020 along with precipitation data. I read in airline delay data from a CSV file then sorted the data to include only data from 2020, January through April, and Atlanta's Hartsfield-Jackson International Airport. I then wrote the filtered data to a CSV file and merged it with the weather data by month. The merged data was cleaned and transformed, with columns removed for ease of functionality. Finally, weather columns were renamed before writing the merged dataset to a CSV file.

The final data set includes the following fields:

*Table 1 Data dictionary*

|  |  |  |
| --- | --- | --- |
| Column | Type | Description |
| month | int | The month of the year (1-12) |
| year | int | Year of instance |
| carrier | categoric | Operating airline alpha numeric code |
| carrier\_name | categoric | Airline company name |
| airport | categoric | Operating airport alpha numeric code |
| airport\_name | categoric | Airport full name |
| arr\_flights | numeric | Arrival flights |
| arr\_del15 | numeric | arrival delayed 15 minutes |
| carrier\_ct | numeric | Airline count |
| weather\_ct | numeric | Weather delay count |
| nas\_ct | numeric | National Aviation System delay count |
| security\_ct | numeric | Security delay count |
| late\_aircraft\_ct | numeric | Late aircraft count |
| arr\_cancelled | numeric | Arrival cancelled |
| arr\_diverted | numeric | Arrival diverted |
| arr\_delay | numeric | Arrival delay in minutes |
| carrier\_delay | numeric | Airline delay in minutes |
| weather\_delay | numeric | Weather delay in minutes |
| nas\_delay | numeric | National Aviation System delay in minutes |
| security\_delay | numeric | Security delay in minutes |
| late\_aircraft\_delay | numeric | Late aircraft delay in minutes |
| precip | categoric | Type of weather measure |
| precip\_in | numeric | Amount of precipitation (inches) |

## Analysis

The purpose of this report is to analyze the impact of precipitation on flight delays at Atlanta's Hartsfield-Jackson International Airport during the first four months of 2020. By examining the relationship between precipitation levels and various indicators of flight performance.

### Summary

I created a summary table using psych package to provide descriptive statistics such as mean, standard deviation, median, etc. of the merged dataset. Table 2 displays the resulting statistics.

*Table 2 Merged Summary*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | mean | sd | median | mad | min | max | range | skew | kurtosis | se |
| arr\_flights | 1915.772 | 4389.726 | 365 | 465.5364 | 10 | 20669 | 20659 | 3.39599 | 10.7958 | 581.4334 |
| arr\_del15 | 185.2105 | 437.283 | 42 | 56.3388 | 0 | 2605 | 2605 | 4.01758 | 17.10594 | 57.91955 |
| carrier\_ct | 55.28456 | 108.9043 | 18.82 | 22.72826 | 0 | 501.37 | 501.37 | 3.027041 | 8.70833 | 14.42472 |
| weather\_ct | 8.289298 | 21.82713 | 0.95 | 1.40847 | 0 | 109.56 | 109.56 | 3.318999 | 10.39225 | 2.891073 |
| nas\_ct | 65.45754 | 177.4705 | 12.62 | 18.48802 | 0 | 1105.83 | 1105.83 | 4.480525 | 21.03136 | 23.50654 |
| security\_ct | 0.199474 | 0.581278 | 0 | 0 | 0 | 3.6 | 3.6 | 4.025449 | 18.75711 | 0.076992 |
| late\_aircraft\_ct | 55.97947 | 149.6535 | 8.31 | 12.32041 | 0 | 946.59 | 946.59 | 4.337817 | 20.81107 | 19.82209 |
| arr\_cancelled | 193.8421 | 709.5345 | 11 | 16.3086 | 0 | 4951 | 4951 | 5.552647 | 33.01996 | 93.98014 |
| arr\_diverted | 2.263158 | 5.452805 | 0 | 0 | 0 | 27 | 27 | 3.302276 | 11.16853 | 0.722242 |
| arr\_delay | 13507.19 | 33036.11 | 2488 | 3484.11 | 0 | 203424 | 203424 | 4.121441 | 18.60789 | 4375.74 |
| carrier\_delay | 5289.772 | 11622.01 | 1156 | 1433.674 | 0 | 53274 | 53274 | 3.0901 | 8.967699 | 1539.373 |
| weather\_delay | 1004.632 | 2746.446 | 36 | 53.3736 | 0 | 12690 | 12690 | 3.203809 | 9.231188 | 363.7757 |
| nas\_delay | 3014.965 | 8606.541 | 512 | 726.474 | 0 | 56818 | 56818 | 4.901375 | 25.76135 | 1139.964 |
| security\_delay | 10.63158 | 31.83082 | 0 | 0 | 0 | 172 | 172 | 3.409475 | 11.85904 | 4.216095 |
| late\_aircraft\_delay | 4187.193 | 12483.89 | 546 | 809.4996 | 0 | 85064 | 85064 | 5.081225 | 28.57279 | 1653.532 |
| precip\* | 1 | 0 | 1 | 0 | 1 | 1 | 0 | NA | NA | 0 |
| precip\_in | 242.0649 | 156.6561 | 175.7 | 6.96822 | 115.3 | 511 | 395.7 | 1.069588 | -0.74251 | 20.74961 |

*\*precip only includes one variable*

*\*Plots were built for functionality in RStudio with interactive aspects*

### Number of Flights by Airline:

Chart

Description automatically generated

### Relationship Between the Number of Flights Arrived and the Number of Delays:

A picture containing chart

Description automatically generated

### Correlation Between Delays and Precipitation:

Chart, bar chart

Description automatically generated

## Limitations

I was initially limited by the original project direction of using Twitter’s API to compare flight delay data to twitter mentions to view if there is an increase of decrease in an airline’s reputation. Once settled on comparing flight delay data with precipitation data I was limited by the cohesion of the datasets.

## Appendix 1

One of the API’s I tried to in corporate was Uber’s and the resulting data I was going to use to see if there was any correlation between flight delays and increased Uber travel time. The following figure is of an Uber API query I created of travel time from Atlanta's Hartsfield-Jackson International Airport dating 2020-01-01 to 2020-03-31.

Map

Description automatically generated

[Developers | Uber](https://developer.uber.com/docs/riders/references/api)

## Appendix 2

I also implemented the use of Airlabs.co API to provide UpToDate flight information to a json file via Python.

[AirLabs Data API](https://airlabs.co/)