**Aviation Analysis System (AAS)**

*Understanding USA flight activity in February 2020*

Chart, scatter chart

Description automatically generated

Developed by Team 7:

* Clara Yuan (21803572)
* Georgia-Rose Collins (20763699)
* Kristy Gray (23123867)
* Shayne Bates (22563009)

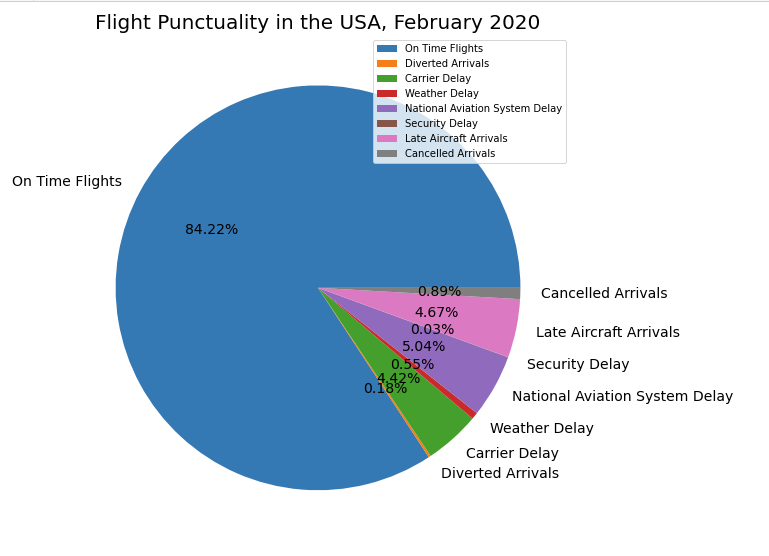
# **Automating the exploration of USA Flight behaviours**

There are over 3000 commercial airports in the USA, and in February of 2020, **574,268 flights** were scheduled to travel. It takes effective management and coordination to ensure that the maximum number of flights travel within intended timeframes. Passengers expect flights to be punctual and will be more likely to fly if the on-time arrival rate is high. Setting KPIs and achieving these expectations require specific goals focusing on increasing on-time arrivals. Appropriate parties require a deep understanding of current flight statistics, delays' causes, and potential location influences. This report details the scripts used to code and innovate the Aviation Analysis System (AAS) to automate this process.

## **Introducing the Aviation Analysis System (AAS)**

The AAS is a combination of easy-to-use functions, which enables the end-user to identify operation flight issues. Examples of AAS functionality:

* A user can establish that in February 2020:
  + the **total number of delayed flights in the USA was 84, 616**
    - with a **total time delay of 5, 819, 054 minutes**
  + the airport with the largest number of delayed flights was **Hartsfield-Jackson Atlanta International with 4609 flights**
    - this airport had the **longest delay time of 352, 569 minutes**
    - this airport was located at **longitude -84.42694444, latitude 33.6404444**
  + a bottleneck occurred at **Dallas / Fort Worth International airport,** with the highest number of delayed flights in Texas **(3838 delayed flights)**
* Delays can be further investigated by **visualising the proportions of each delay cause,** using a pie-chart (fig 1. Flight Punctuality in the USA, February 2020)



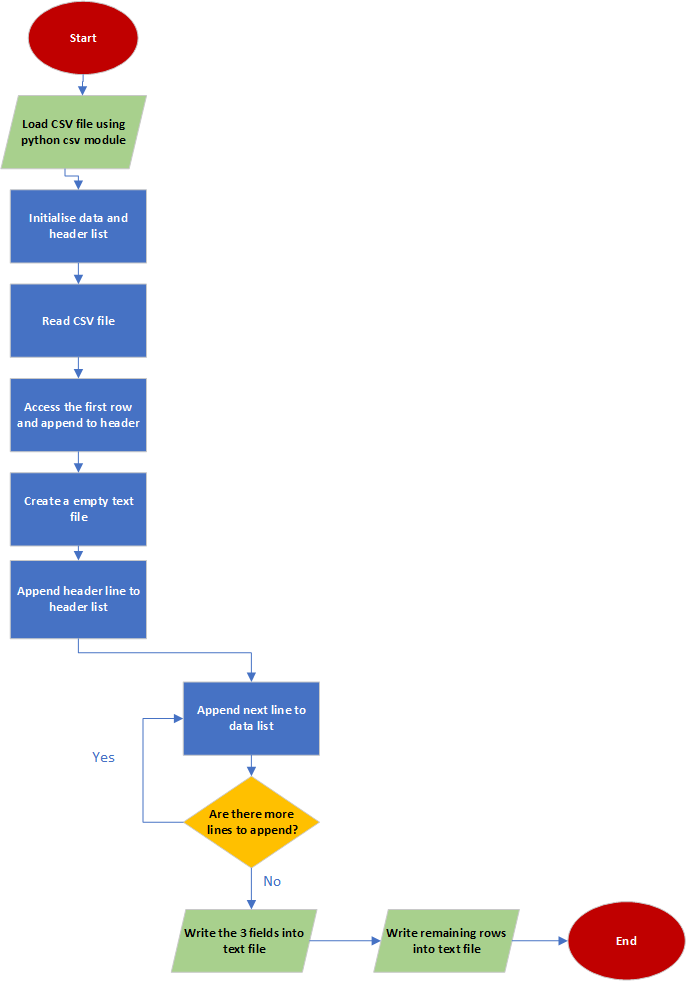
**Figure 1 – AAS generated pie-chart: Flight Punctuality in the USA, February 2020**

The Aviation Analysis System (ASS) gives the commercial aviation industry continuous access to user-friendly code, enabling them to assess issues and put steps into place to mitigate them.

## **Importing USA airport data from a CSV file**

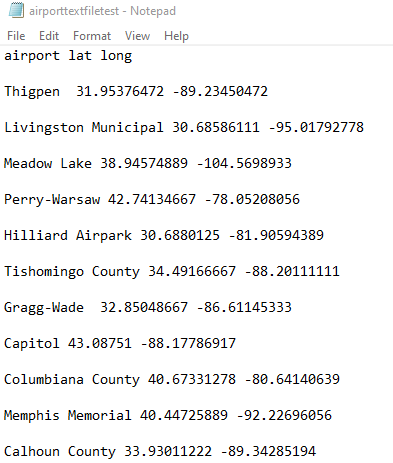
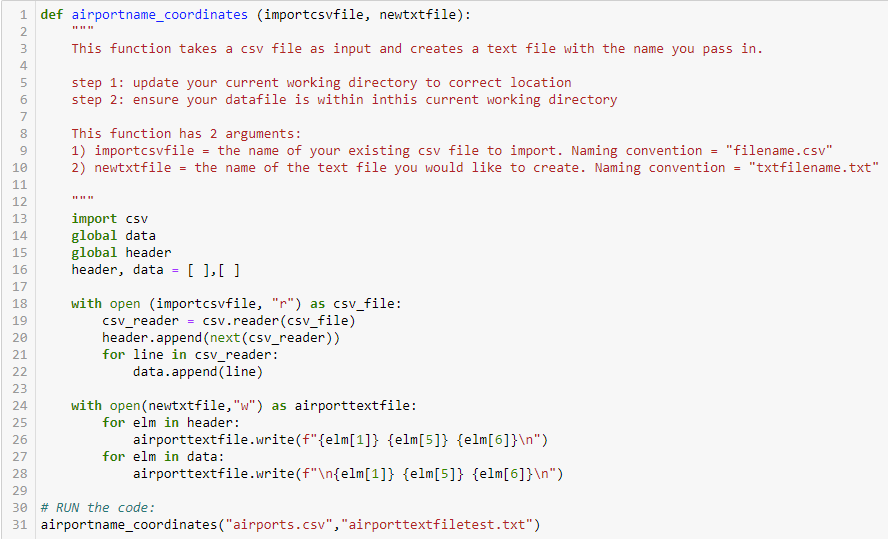
We commenced our study by flowcharting an algorithm (fig. 2) to establish how to code the extraction of data from a comma-separated (CSV) file. The algorithm extracts airports and location coordinates from the CSV, then and writes them to a text file.

**Figure 2 - Part A, Question 1 – Flowcharting an algorithm to extract coordinates of airports**



Using the algorithm, the AAS script airportname\_coordinates (fig. 3) was created to import/read a CSV file, export data, and write to text file.

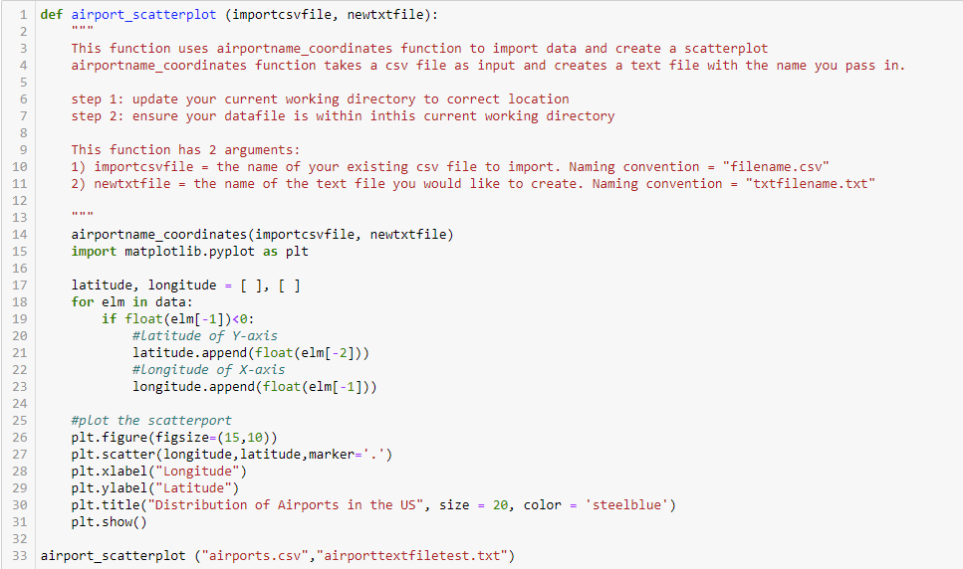
**Figure 3 - Part A - Question 2 - AAS Script airportname\_coordinates to import, extract and write data to a textfile**



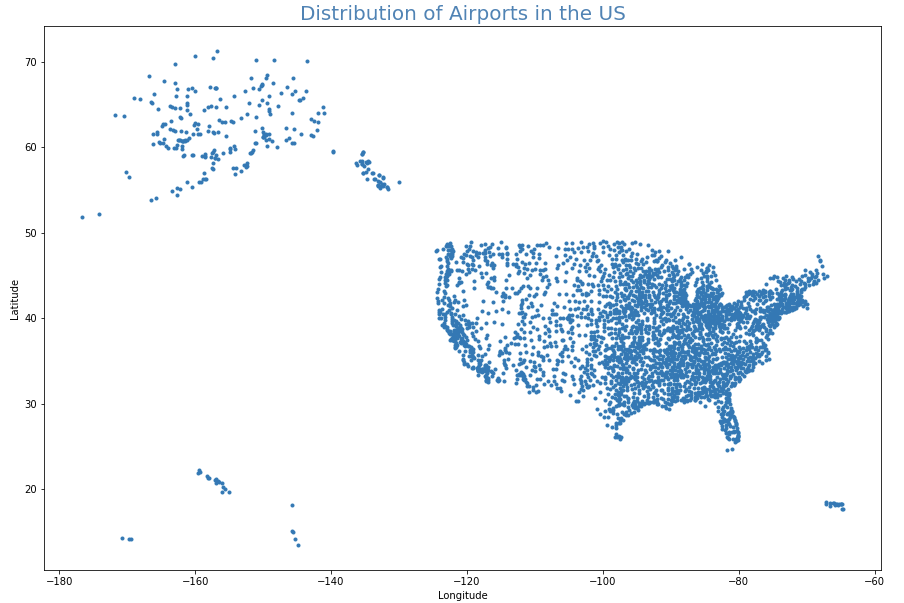
## **Understanding the distribution of airports in the USA**

The AAS script airport\_scatterplot (fig. 4) transfers all commercial airport locations in the USA to a scatterplot (fig. 5), utilising the data extracted from the CSV in the airportname\_coordinates (fig. 3) script.

# 



**Figure 4 - Part A, Question 3 – AAS Script airport\_scatterplot create a Scatterplot for USA Airports**



**Figure 5 - Part A, Question 3 – Visualisation of data from AAS script airport\_scatterplot**

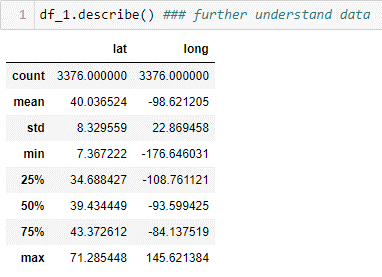
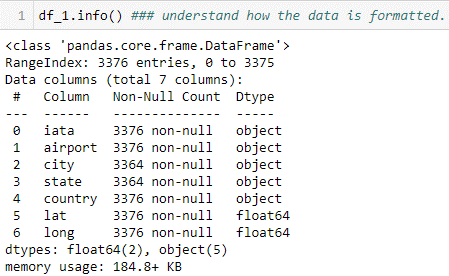
## **Preparing the airport data *from two datasets* for analysis**

USA flight data was analysed from February 2020 by importing and processing two data sets (fig. 6) into the AAS. Within the system, for efficiency the commonly used **‘airline\_delay\_causes’ CSV file was defined as df\_1** and the generic ‘**airports’ CSV was defined as df\_2.**



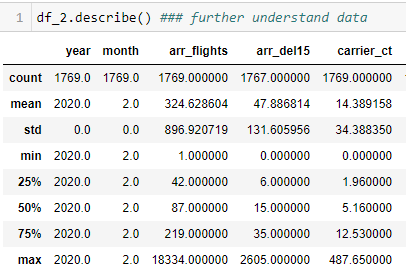
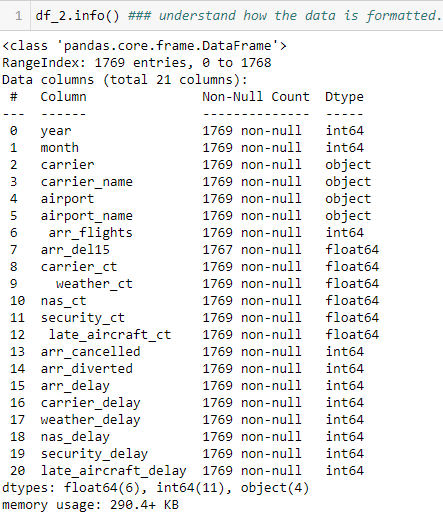
**Figure 6 - Part B - Importing the datasets and constructing dataframes into the AAS**

A rigorous data integrity check was completed to explore the structure and to understand the organisation of df\_1 in the AAS (fig. 7)



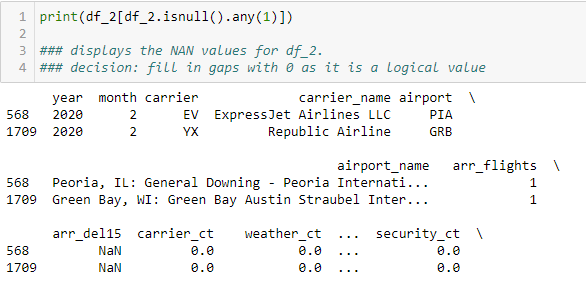
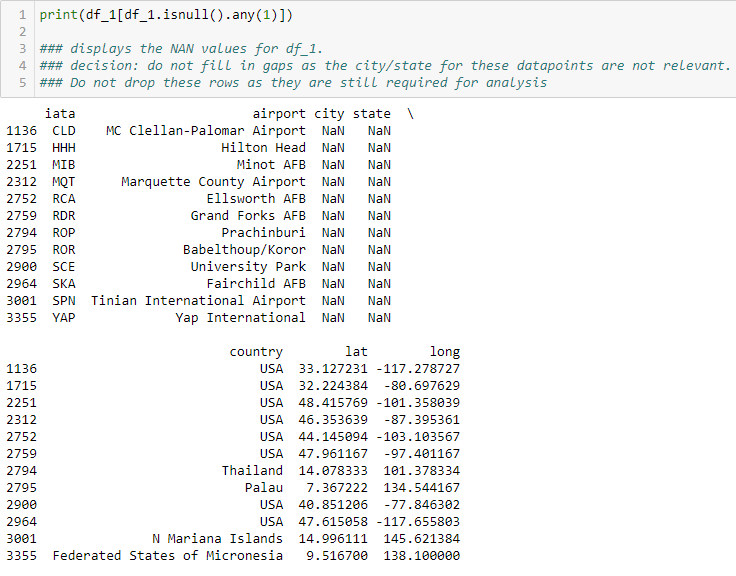
**Figure 7 - Part B - Exploring df\_1 and understanding how data is structured within the AAS**

The same approach was used for df\_2 (fig. 8).



**Figure 8 – Part B – Exploring df\_2 and understanding how data is structured within the AAS**

Once understood, the missing values within each of the datasets were identified (fig. 9).



**Figure 9 - Part B - Identifying missing values – Excerpts of output**

The specific values missing in df\_1 do not impact this study, so these were not adjusted or removed. Analysis of the data missing in df\_2 revealed that these values logically should be = 0 (due to the breakdown of the data). These missing values were filled with 0 (fig. 10).



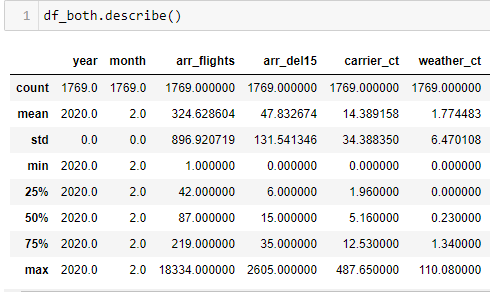
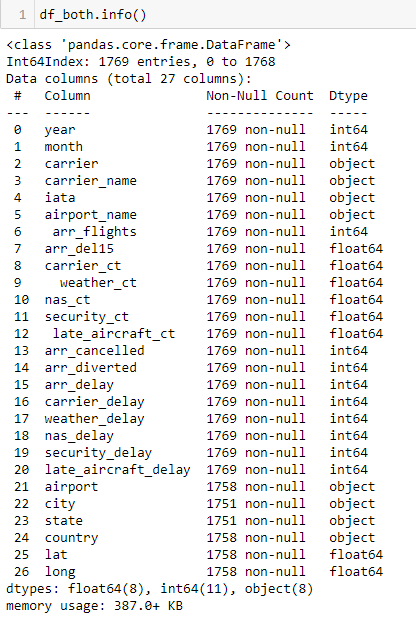
**Figure 10 - Part B - Cleaning data by filling NaN values**

For effective utilisation, the datasets were merged to consolidate information, defined as df\_both (fig. 11).



**Figure 11 - Part B - Merging dataframes df\_1 and df\_2**

The merged dataset df\_both was explored (fig. 12) and analysis of this dataset commenced.



**Figure 12 - Part B – Exploring merged dataframe df\_both**

# **Processing the merged data**

Understanding and exploring the behaviours of flights in the USA required establishing a specific set of questions to create efficient scripts to automate this process. The merged dataset df\_both was used to test these scripts.

## **Questions presented of the data and AAS scripts created to process these questions:**

### **What is the total number of flights in the USA?**

Including cancelled, diverted, delayed and on-time.

Text

Description automatically generated with medium confidence

### **What is the total number of delayed flights in the USA?**

Graphical user interface

Description automatically generated

### **What is the total delayed time (in minutes) of flights in the USA?**

Graphical user interface

Description automatically generated with medium confidence

### **What is the airport with the largest number of delayed flights?**

A picture containing text

Description automatically generated

### **What are the coordinates of the airport with the highest delayed time?**

Text

Description automatically generated

### **What is the airport** **in Texas that has the highest number of delayed flights?**

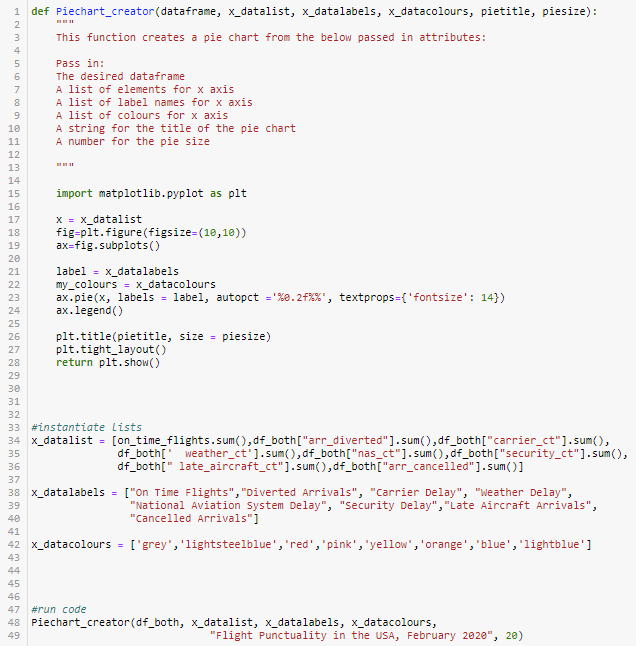
Text

Description automatically generated

### **What is the percentage breakdown of?**

* + On-time flights Text

    Description automatically generated
  + Delayed flights (over 15 minutes late)
    - air-carrier delays
    - weather delays
    - National Aviation System (NAS) delays
    - security delays
    - aircraft arriving late
  + Cancelled flights
  + Diverted flights



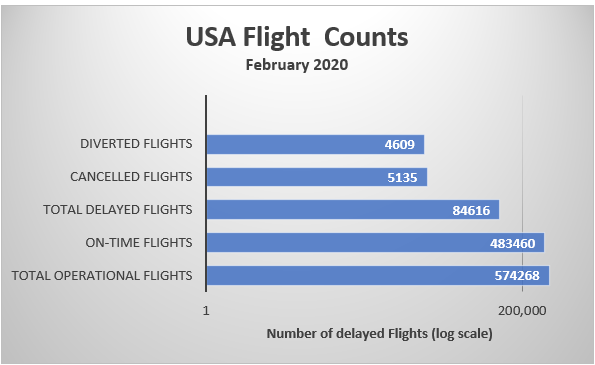
# **Analysing the results**

Establishing benchmarks based on historical data, coupled with collecting data via the ASS, will assist the aviation industry in establishing acceptable parameters. Users will be able to analyse data regularly to determine if airports and states are performing according to the benchmarks. Utilising the AAS, February 2020 data reveals:

## **Questions 1, 2, 3**

In February 2020, the USA had a total of 574,268 flights, which included on-time, cancelled, diverted, and delayed flights (fig. 13)

**Figure 13 - Bar Chart/Log Scale of total Flights in the USA**

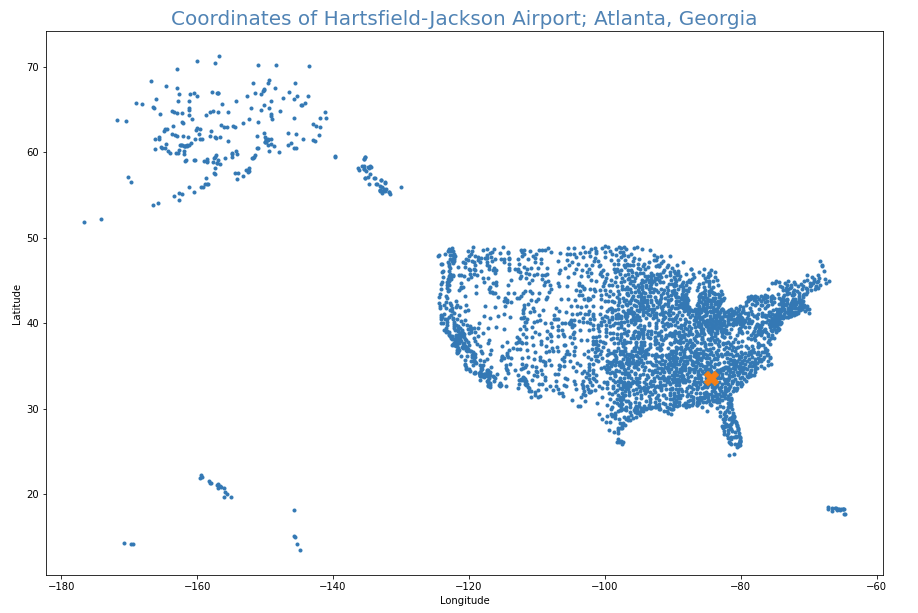


14.73% of these flights (84,616) were delayed, which were caused by carrier, weather, NAS, Security, and late arrival of aircraft.

Total flight delay time was 5,819,054 minutes, which equates to over 4,000 days’ worth of delayed flights.

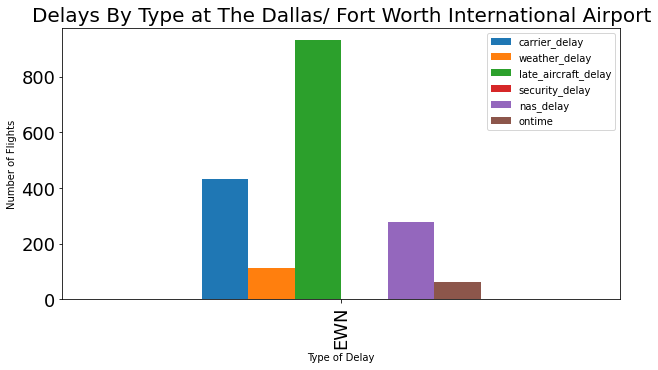
## **Questions 4, 5**

The airport with the largest number of delayed flights was Hartsfield-Jackson International in Atlanta, Georgia. This airport had a total of 4609 flights, accounting for a little over 5% of all delayed flights in the USA. Hartsfield-Jackson Airport (fig. 14) also had the highest delayed time at 352,569 minutes (5,876.15 hours), representing 6% of the total delayed time in the USA.



**Figure 14 - Part B - Question 5 -   
Coordinates of the airport with the highest delay time = X   
Hartsfield-Jackson International (coordinates: -84.4269444 longitude x 33.6404444 latitude).**

## **Question 6**



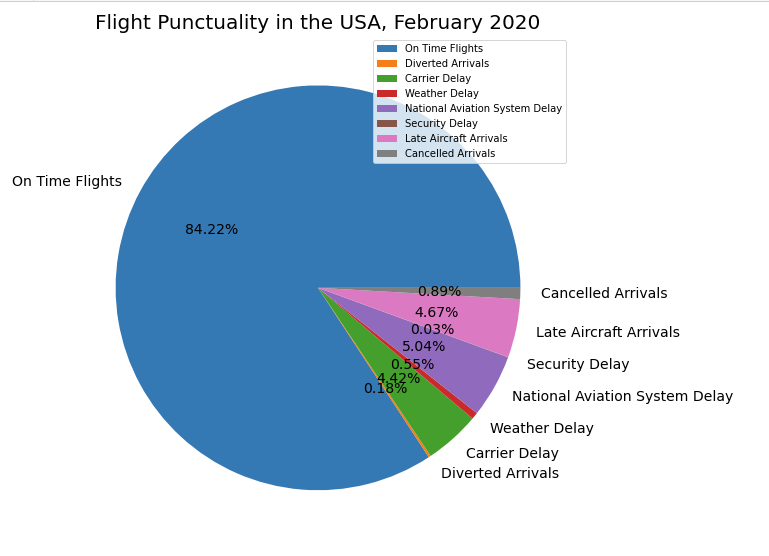
**Figure 15 - Part B - Question 6 - Dallas/Fort Worth International Airport in Texas breakdown of delayed flights**

Dallas / Fort Worth International Airport in Texas has a total of 3,838 delayed flights, which is 4.5% of the total delayed flights and 0.6% of all flights in the USA. Late aircraft was the primary cause of delay at this airport (fig. 15).

## **Question 7**

84.22% of flights were on time, 14.71% of all flights in the USA were delayed, 0.89% were cancelled, and 0.18% were diverted (fig. 16). Delays in flights were caused by several factors (fig. 16), with National Aviation System Delay making up 5.04% of flights, late aircraft delays at 4.67% and carrier delays were 4.42%. The analysis suggests this is where improvements can be made. Weather and Security only had a negligible impact on delays, with 0.55% and 0.03%, respectively

**Figure 16 - Part B - Question 7 – Pie chart of flight punctuality in the USA**



# **Leveraging the Aviation Analysis System**

Passenger satisfaction is crucial in ensuring success in the airline industry. Flight delays, cancellations, and diversions form negative experiences for passengers, deterring them from travelling by aircraft. In February 2020, the on-time arrival rate for USA flights was approximately 21 in 25 (around 84%). In that month 84,616 flights were delayed, mostly caused by Late Aircraft Arrivals, National Aviation System (NAS) Delay and Carrier Delay. To improve passenger satisfaction (which increases patronage), on-time arrival rate should be increased. We recommend these Carriers, Aircrafts and the NAS utilise our Aviation Analysis System (AAS) to monitor performance, set benchmarks to compare to and incentivise their teams accordingly, to improve the national on-time arrival rate.