JFK AIRPORT FLIGHTS DATA ANALYSIS

ATHALIA SETIAWAN

ABSTRACT

The data used was from Kaggle.com, titled "Flight Take Off Data – JFK Airport". It contained 28820 flight details taking off from New York's JFK (John F. Kennedy) Airport from November 2019 until January 2020, with a total of 23 features. For this research, the objective was to find what time of day has the least departure delays and the most departures at JFK Airport. It started with preparing the data, then made histograms to better understand the flight trends, and presented the final analysis in a heatmap. It turned out that there were more delays at night compared to daytime and that there were more flights in the morning and afternoon.

MOTIVATION

- The objective is to find *what time of day has the least delays from JFK Airport and when are the busiest hours*.
- This insight could be valuable to **travellers** so they can book flights at better times to avoid delays. It could also be useful for **airport managers** as they would be able to focus on the times at which flights are often delayed in winter season.
- Travellers might see a trend that they were not expecting in airplane departures, which would become interesting insight for them.

DATASET(S)

- This data set is called "**Flight Take Off Data – JFK Airport"**. It contains 28820 records of *flight details taking off form JFK Airport between November 2019 and January 2020*, with a total of 23 features.

- The features include month, date, day of the week, aircraft details, destination, departure delay, distance of the flight, weather condition details, and number of other airplanes scheduled for arrival and departure at the time. However, the features that were taken for analysis were departure delay and actual departure time.

DATA PREPARATION AND CLEANING

There was no data cleaning to be done as there were no null values.

Part of the data was copied into another data frame containing just the departure time and departure delay *(defined as departing more than 15 minutes later than scheduled departure time*) of all flights on a specific day. Two more data frames were made to group the departure time by hours of the day: one data frame for total number of flights in each hour, and another data frame for number of delayed flights in each hour.

A function was then created to ease the process of preparing separate data frames of multiple days.

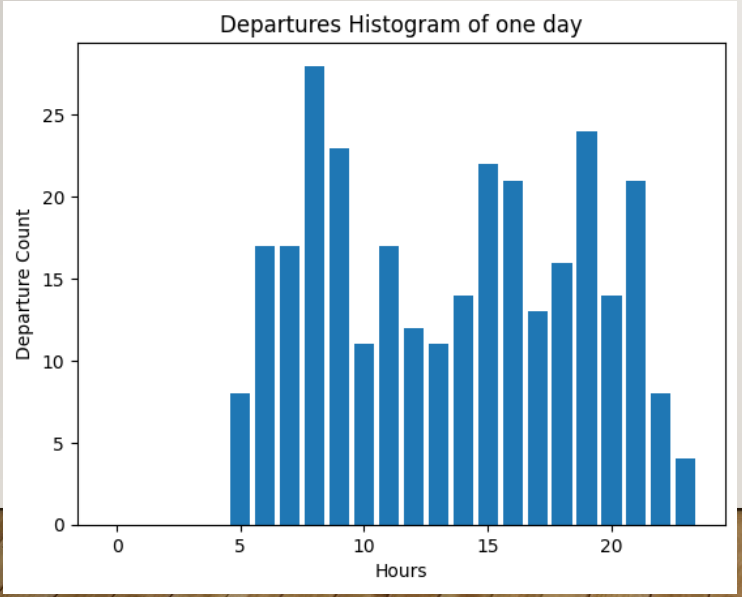
RESEARCH QUESTION(S)

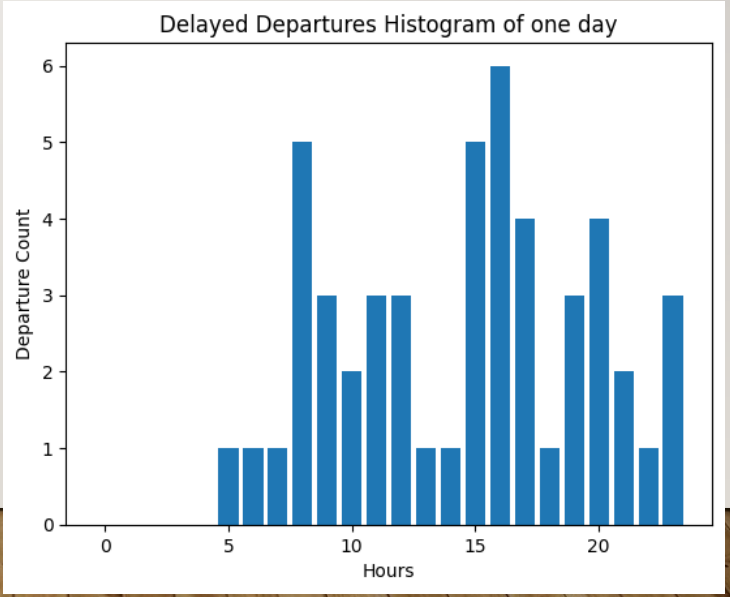
WHAT TIME OF THE DAY HAS THE LEAST DELAYS?

WHAT TIME OF THE DAY HAS THE MOST FLIGHT DEPARTURES?

METHOD

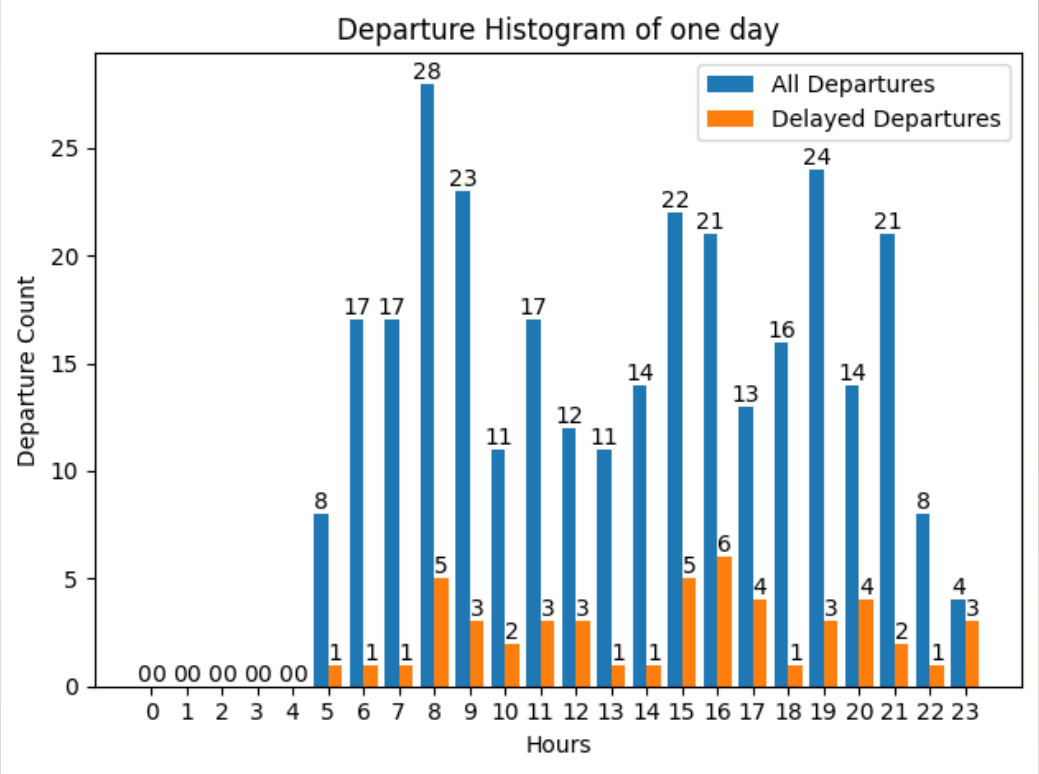
1. Histogram: to visualize the number of departures in each hour of one day and another histogram to visualize the number of delayed departures in each hour of one day. Making it for multiple days improved the understanding of trends of when there are more departures and delayed departures during the day.





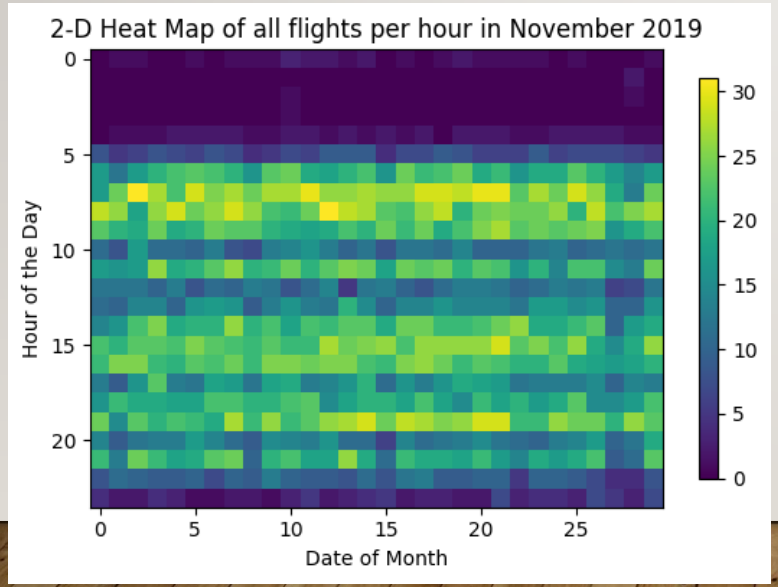
METHOD (CONT.)

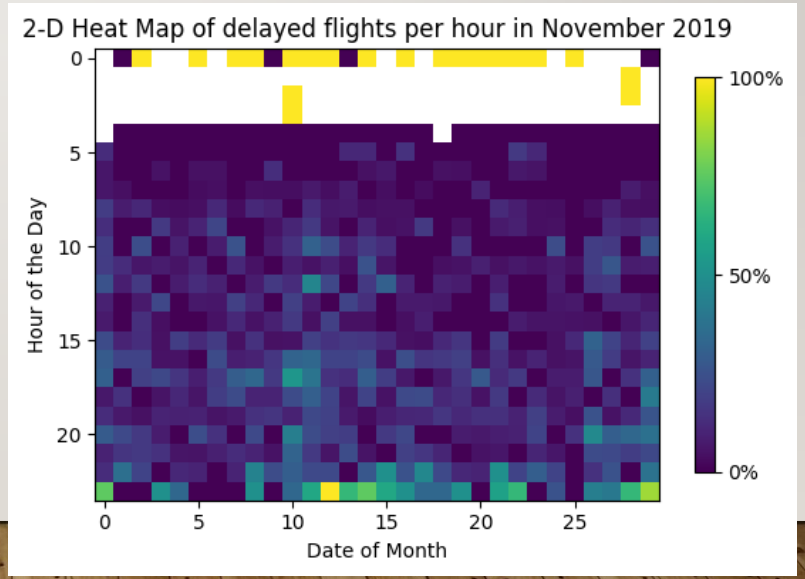
2. Grouped bar chart of the data: to compare the total departures and delayed departures better.



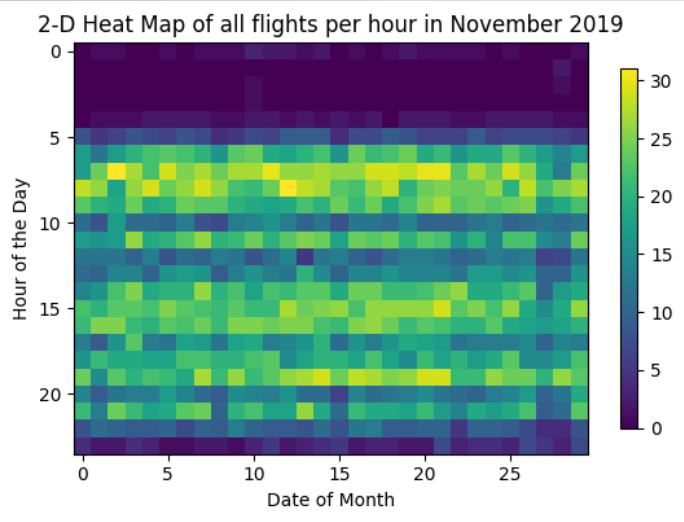
METHOD (CONT.)

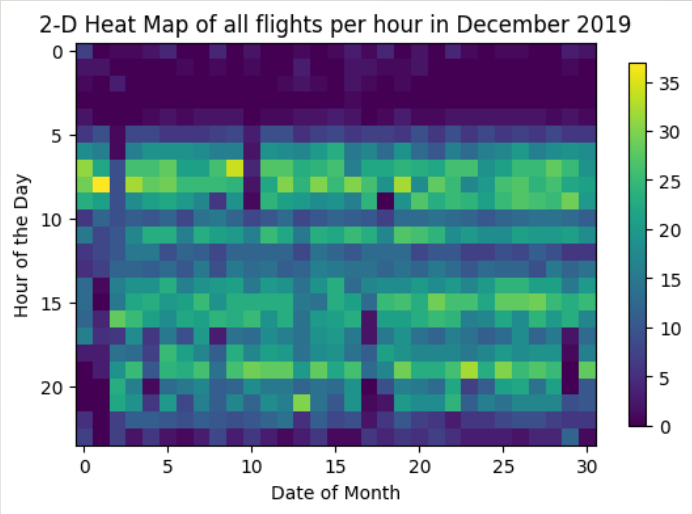
3. Heatmap: show the delayed departures in every hour of the day across one month. This step uses colors to display the busy times in an appealing and easy to interpret way, which are the advantages of a heatmap.

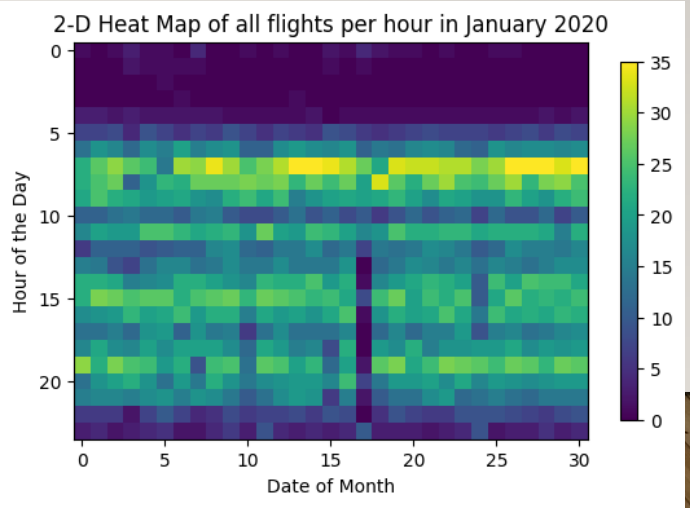




FINDINGS 1

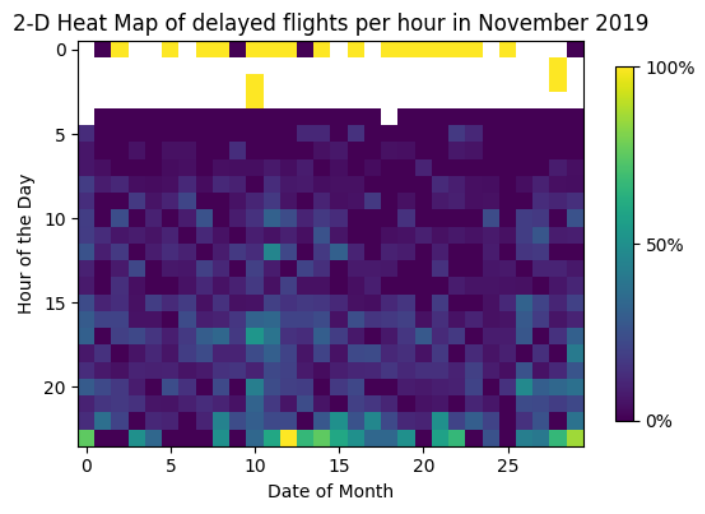


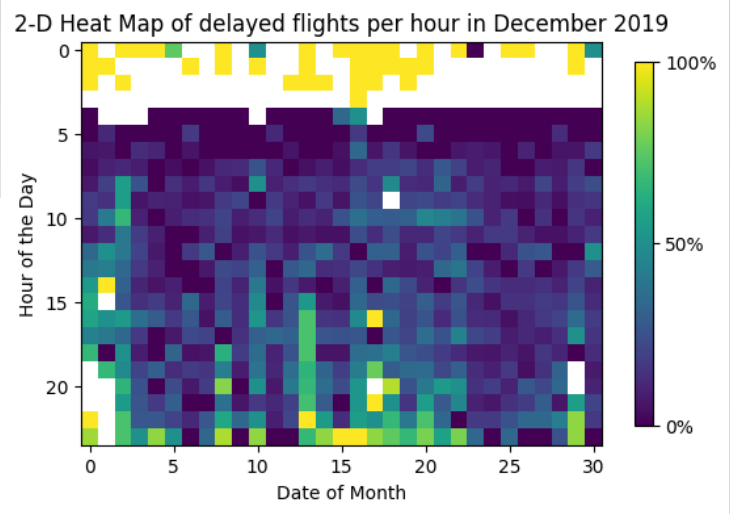


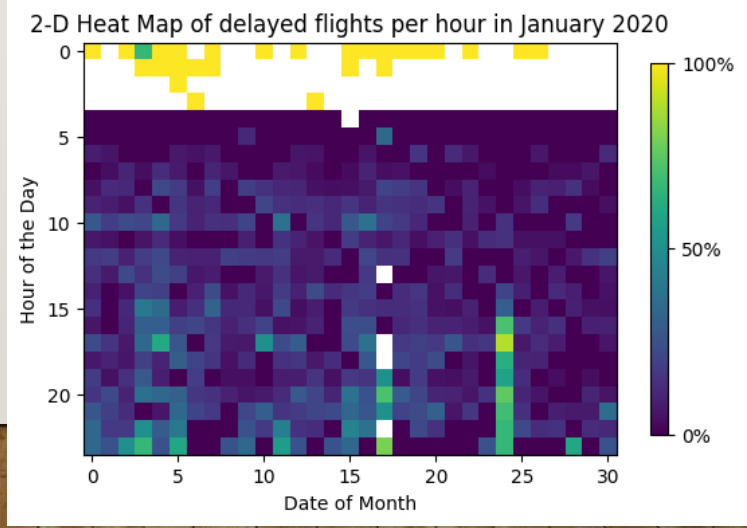


1. Very little flights from 11pm to 4am
2. Peak flights is in the morning at 7am to 8am
3. There are more flights overall in November 2019, compared to December and January

FINDINGS 2







1. Flights at midnight to 2am are very likely to be delayed
2. Flights get more delayed towards the evening
3. December has more delayed flights compared to other months
4. Flights before 9am are less likely to be delayed

LIMITATIONS

The data was limited to November 2019, December 2019 and January 2020 flights, so the trend might not be the same for other months of the year.

CONCLUSIONS

What time of the day has the least delays?

Generally, the least departure delays are between 5am and 10am. Starting at 3pm, there will likely be higher chance of departure delays.

What time of the day has the most flight departures?

Generally, the most flight departures are between 6am and 9am. The time period with the second most flight departures would be between 2pm and 8pm.

ACKNOWLEDGEMENTS

Data set was taken from https://www.kaggle.com/datasets/deepankurk/flight-take-off-data-jfk-airport.

Thank you to my parents for giving me useful feedback on my presentation.

REFERENCES

The definition for a departure delay was taken from OAG.com, which stated that "exactly 15 minutes after the scheduled time is late."