Northwestern

MSiA411 Data Visualization Project

MSiA411 Group6
Licheng Du, Linyue Zhang, Xingyu Wang, Yuwen Meng
Winter 2023

Executive Summary

- Business Problem
  - Investigate into flight delay issues from and to New York City
- Business Objective
  - Draw public’s attention to the seriousness of the issue
  - Layout preliminary analysis that could be used as guidelines for future researches and policies

Problem Definition & Sources Selected

- Why Reducing Flight Delay?
  - Recent investigation suggested that flight delay caused $28 billion lost in 2018 (including direct cost to airlines and passengers, lost demand, and indirect costs)
- Our Dataset:
  - 2013 NYC flights data (data source: https://github.com/tidyverse/nycflights13)
  - Database Structure:
    - Flights Table, including all flights' data in 2013 from/to NYC
    - Airplane Table, including all airplane model's data
    - Airport Table, including all airport data
    - Weather Table, including all weather conditions through 2013
  - We chose flights table as our primary data source as it covered all flight records throughout the year, while using airplane table and airport table as supplemental data in order to analysis on specific variables that may influence flight delay; we did not select weather table because we believe that bad weather causing delay is a common sense and such delay can hardly be intervened (unlike other factors).

https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/

Problem Breakdown

- For all flights departing from NYC:
  - To what destination (airport/state level) delay occurred the most?
  - What airline caused most delay (average delayed time)?
  - In what month of the year and what hour of the day does the delay occurred the most?
  - Does characteristics of the plane affect delay (plane age, seats, manufacturer)?

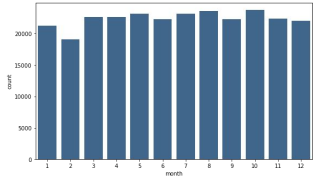
Data Transformation

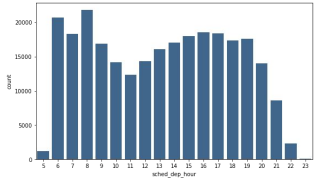
Data Transformations

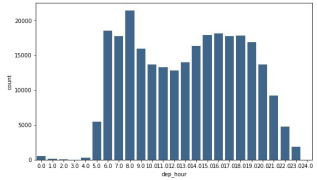
- Joined Flights table with Airports, Planes, and states name
- Add an indicator for delayed flights
- Dropped unnecessary columns and NAs
- Created new variables
  - Plane Age = 2013 - plane_year
  - Dep_date, Dep_hour, Arr_hour, Sched_arr_hour
  - Delayed flights proportion by destination and carrier

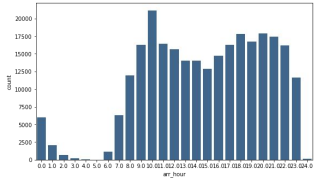
EDA

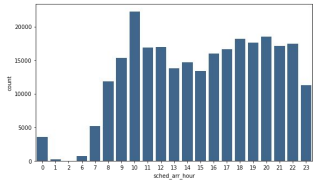
- Explore data values and range
- Explore feature distributions

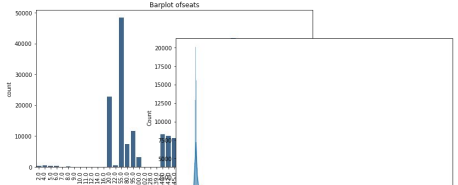


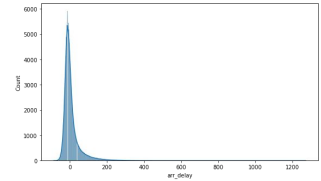


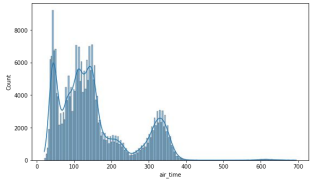


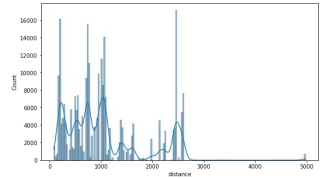


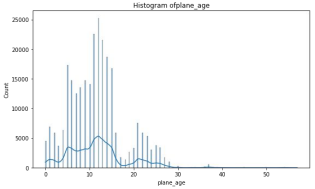













```
Carriers counts: 16
Origins: ['EWR' 'LGA' 'JFK']
Destinations counts: 99
Airplanes counts: 3246
Manufacturer counts: 28
Model counts: 121
Engine counts: 6
Date range: 2013-01-01 00:00:00 to 2013-12-31 00:00:00
```

Insights-Geography

- All flights in the dataset depart from NYC to other cities across U.S. Most flights have destinations in the east-to-middle part of the country, and that is also where most delays occur.
- Lowa, Oklahoma, and Alabama are top three destination states that have longest average delay time, approximately over an hour.
- Alaska, Montana, and Wyoming are bottom three destination states with the shortest average delay time, ranging from 16 to 40 min.

Insights-Time

- The percentage of flights delayed (in terms of landing time) in 12 months did not vary significantly
- However, the average duration of delayed time maximized during summer and minimized during early winter (October, November), this quite counter intuitive since people usually think delay is related to bad weather in the winter months.
- As for hour in the day, we found a trend that as hour increases, the proportion of flights delayed also increases (potentially due to build-up traffic jam)
- Similarly, the average duration of delayed time increases as hour in the day increases

Insights-Carrier+Manufacturer

- Carriers
  - Carriers with higher arrival delay percentage also have higher departure delay percentage
  - HA(Hawaiian Airlines) and OO(SkyWest) carriers have the longest delay time, US(US Airways) and UA(United Airlines) carriers have the shortest delay time.
- Manufacturer
  - The average delayed flight percentage for all manufacturers are 25%.
  - Planes made by Aguasta Spa & Robinson Helicopter Co have the longest delay time.
  - Planes made by Avions Marcel Dassau and Leblanc Glennt have the shortest delay time.

Insights-Interactive Exploration

- Investigate **Carriers** delayed flights with **geographical** information
  - United Airline (UA): large proportion of flights have less delayed time, Missouri has the largest average delay time
  - JetBlue (B6): large proportion of flights are heavily delayed
- Investigate delayed flights in **destination** with **seasonal** and **hourly** trend
  - CAE: 63.17 delayed minutes on average, 67% flights delayed
  - Delayed time are generally long throughout the year, probably
  - Delayed flights peak at 5-6pm departure time

Recommendation & Next Steps

- Recommendation
  - Passengers who fly with lower cost carriers needed to be noticed that the delayed time is longer
  - For passengers, choose planes from manufacturers with relatively low delayed flights
  -
- Next Step: include more dimensions for analysis and expand
  - Add data relevant to the reason for delay
  - Seasonal delay pattern for different section of airport network
  - Draft and conduct relevant policies and regularizations