

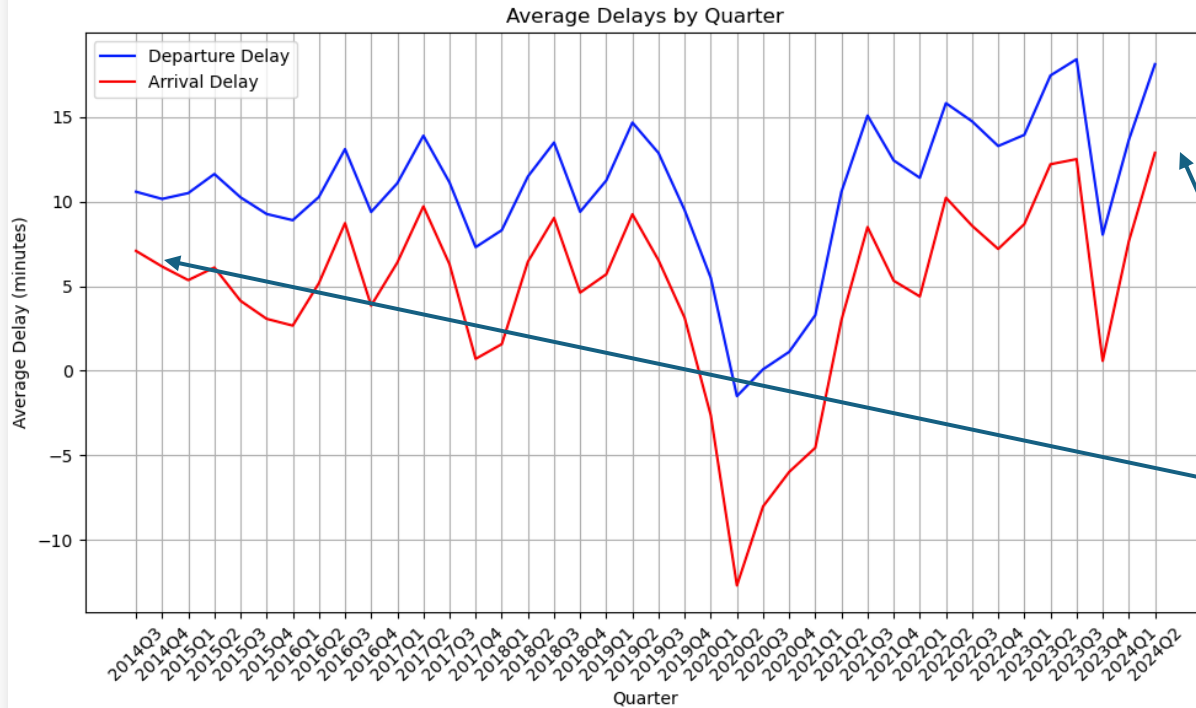
Aviation On Time Performance in the United States

Team 07

-
- USDOT Open Data
 - 13 millions Flight Records
 - From 2014-07 to 2024-06



Average Delays by Quarter



- Arrival Delay follows the trend of the Departure Delay
- Delay Dropped to Negative during COVID
 - Pilots Lay-off
- Average Arrival Delay Increased from ~7min in 2014 to ~12min in 2024 (70% Up)

Guiding Questions (Updated)

What are the most recurring reasons for flight delays and cancellations?



We aimed to discover the most frequent causes of flight delays and provide an analysis by

1. Airport
2. Airline
3. Routes
4. But only **Weather Data** turns out Promising

Airlines with a Hub and Spoke business model have higher flight delays and cancellations than those with a Point-Point model?



Very difficult to compare **Hub and Spoke** Routes
Instead, we shifted to comparing the difference in time delay between **Origin and Destination City**



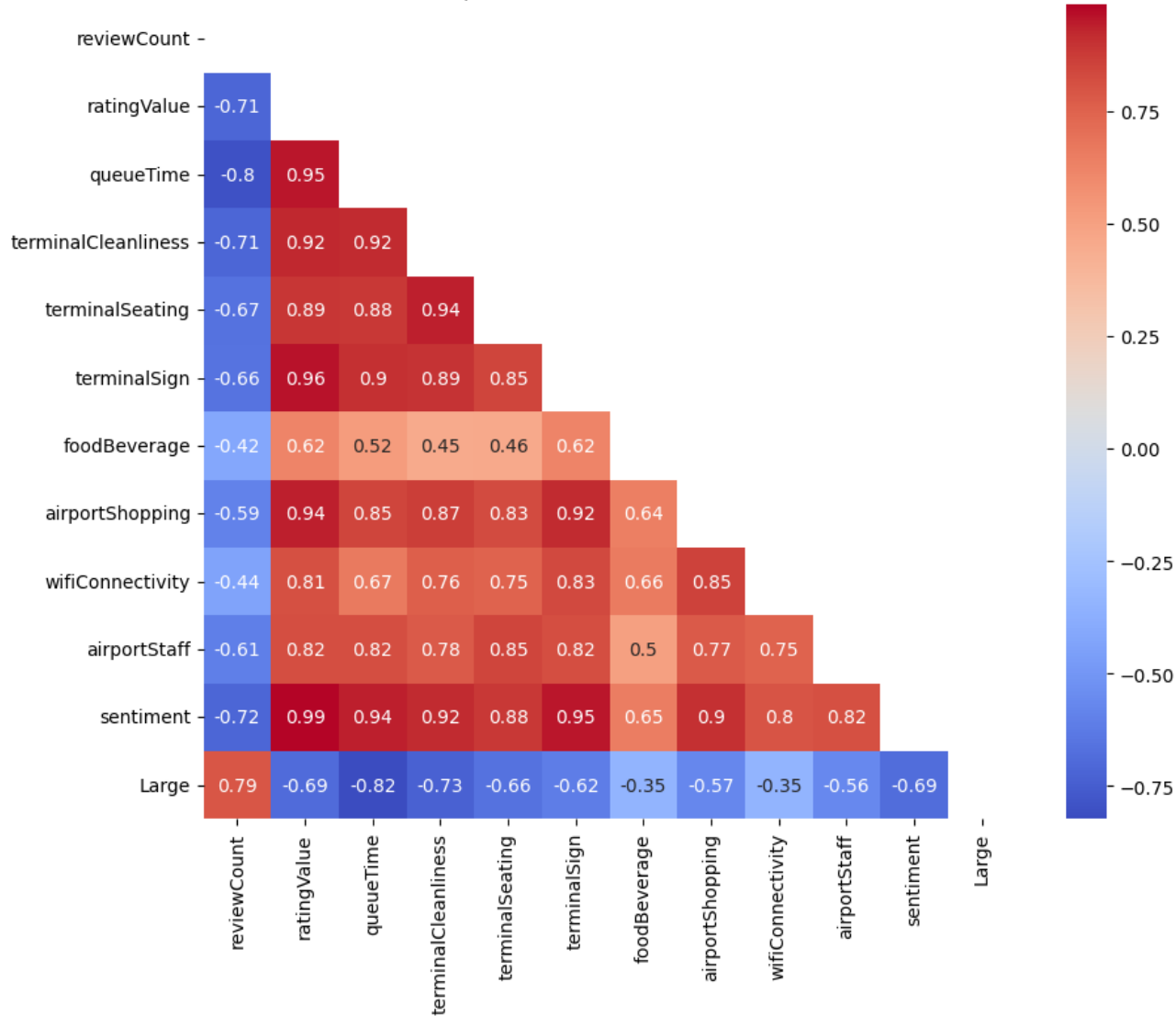


Airport On-Time Performance

- 24 Airports
 - 14 Large Hubs
 - 10 Medium
- Their respective Reviews from Skytrax



Airport Review Score Correlation



Airport Review Quality

Hypothesis

- Queue Time, Staff could lead to Delay

Obstacle:

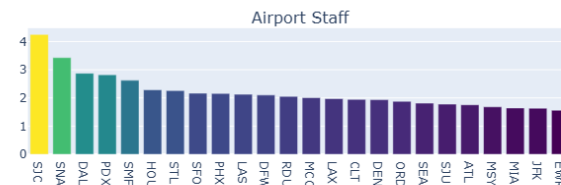
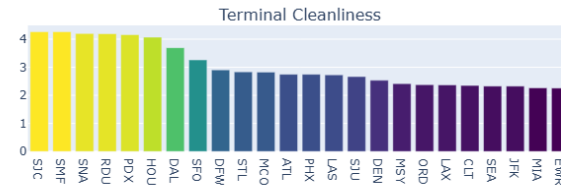
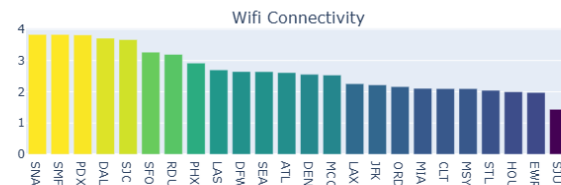
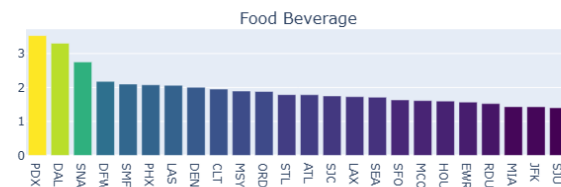
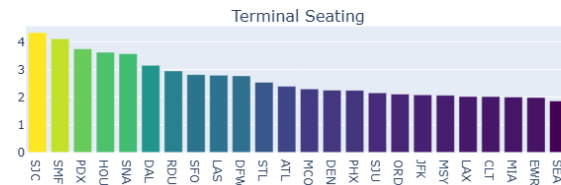
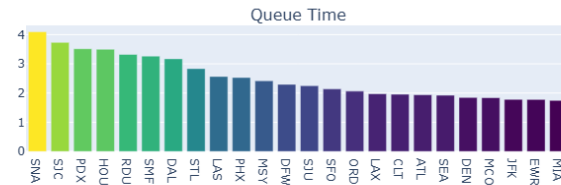
- Typical Bias in Review, More Review Lower Score
- The Busier the Airport, the Lower the Score Overall (except Food)

Future Action:

- Gather Official Statistics about Performance

Airport Review Ranking

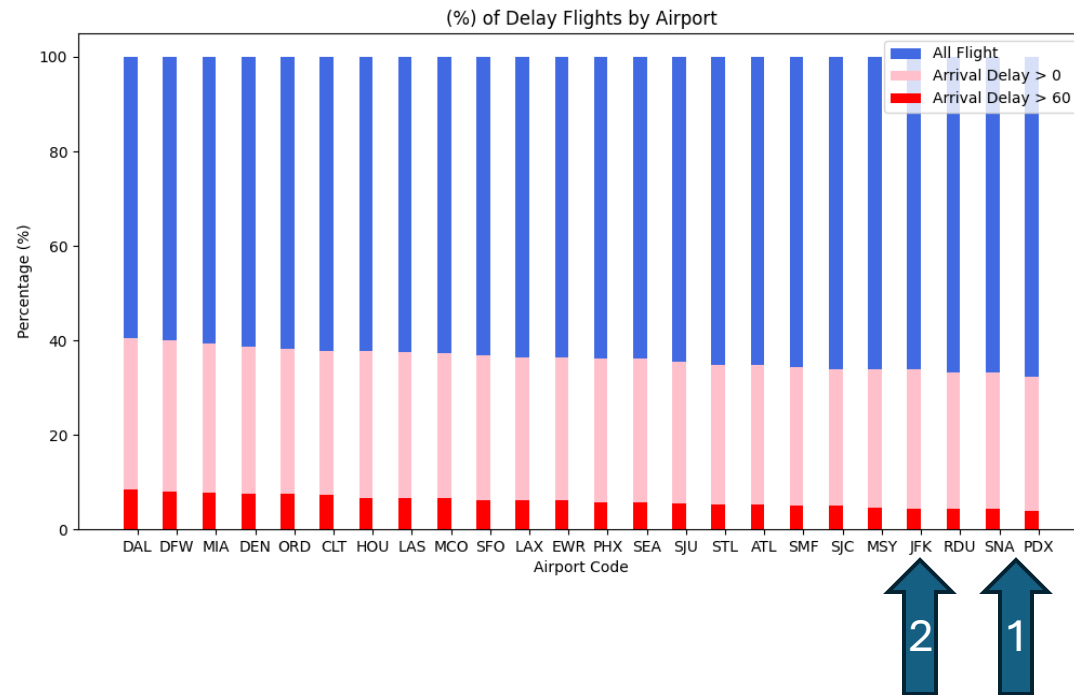
Individual Scoring Attributes of Airports



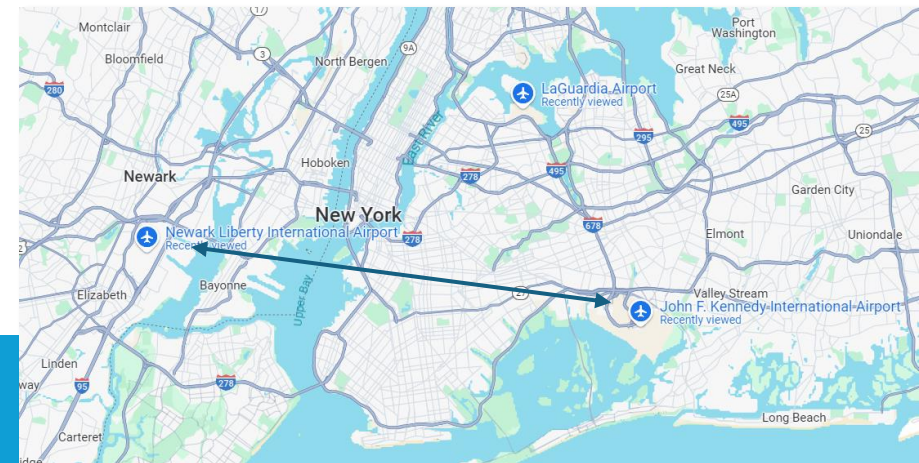
Top Airports

- PDX / SJC / SNA
- Medium Airports

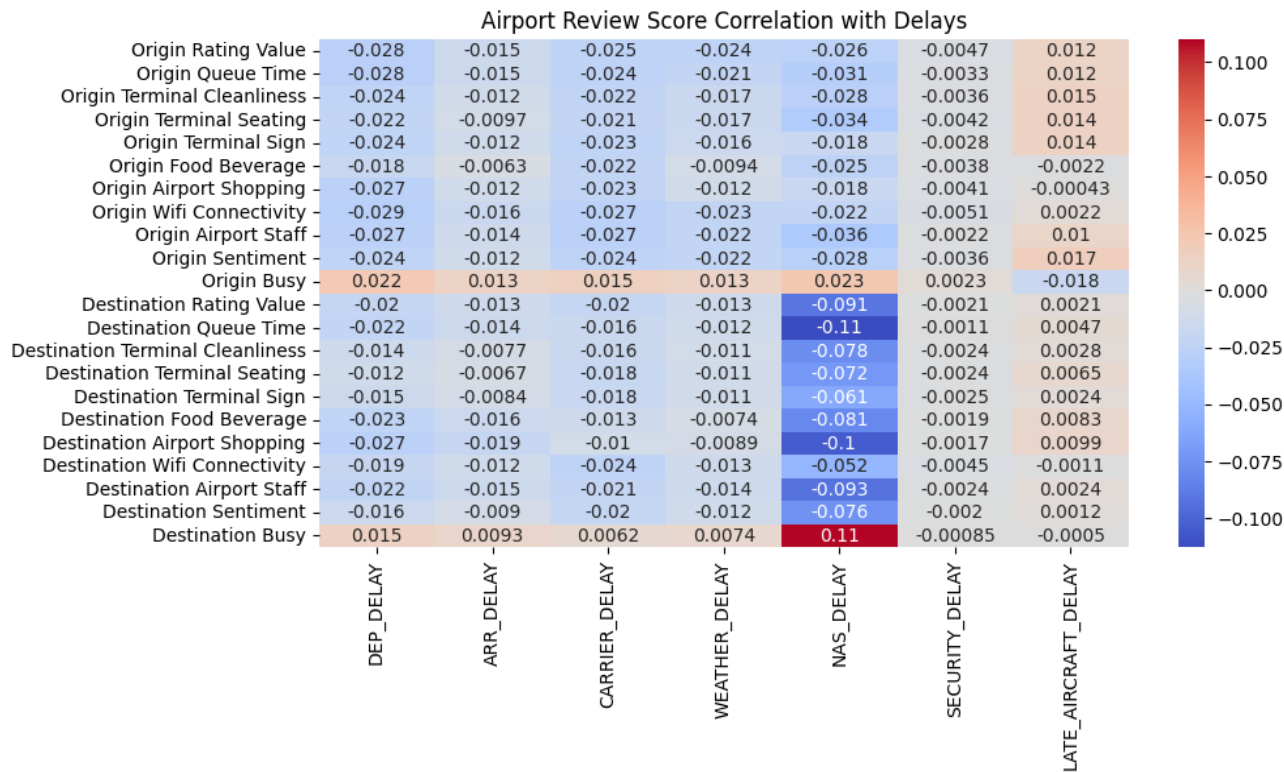
Origin Airport Arrival Delay %



1. Flight from SNA & PDX are both Well-Reviewed and On Time
2. JFK Better than EWR in NY



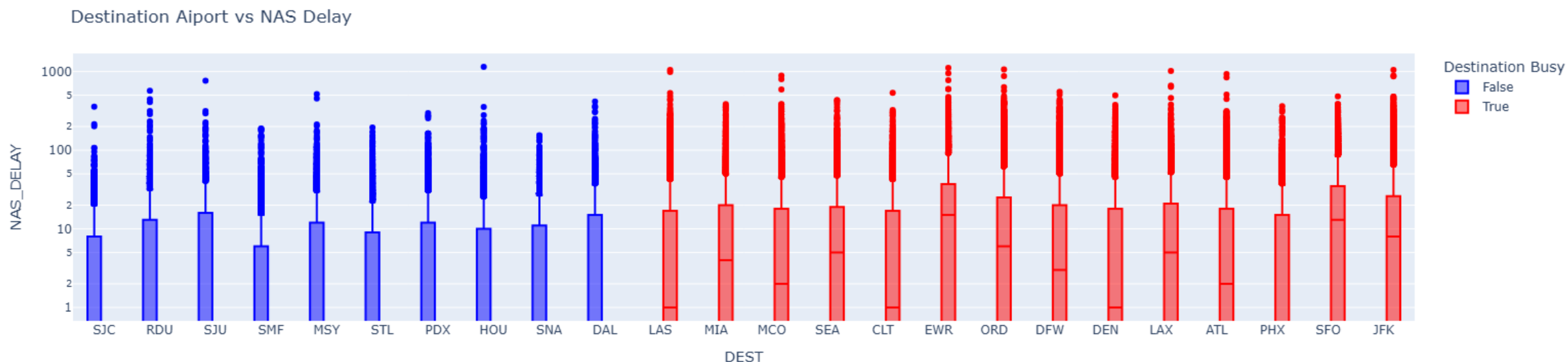
Airport Review Correlation with Delays



- Departure & Arrival Delay
 - Carrier
 - Weather
 - National Airspace System (NAS)
 - Security
 - Late Aircraft
- NAS Delay Correlated with Destination Airport
 - Bigger the Airport > More NAS Delay
 - Heavy traffic volume/ Air traffic control

Destination Airport vs Delay

- Small Airports Destination has a Median NAS Delay of ~0min
- Big Hubs Destination has a Median NAS Delay of ~2 to 10min



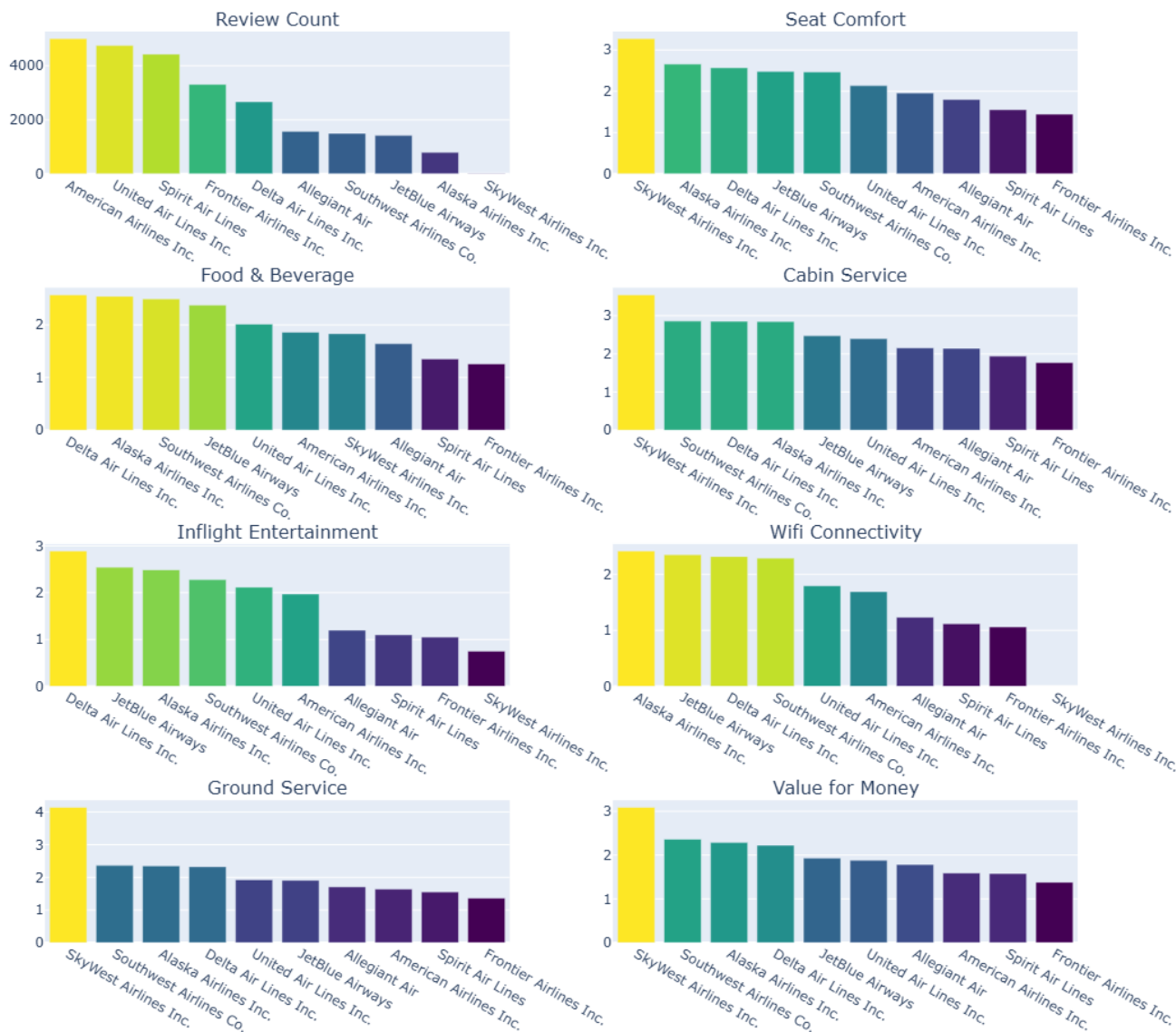


Airline On-Time Performance

- 12 American Airlines
- Their respective Reviews from Skytrax



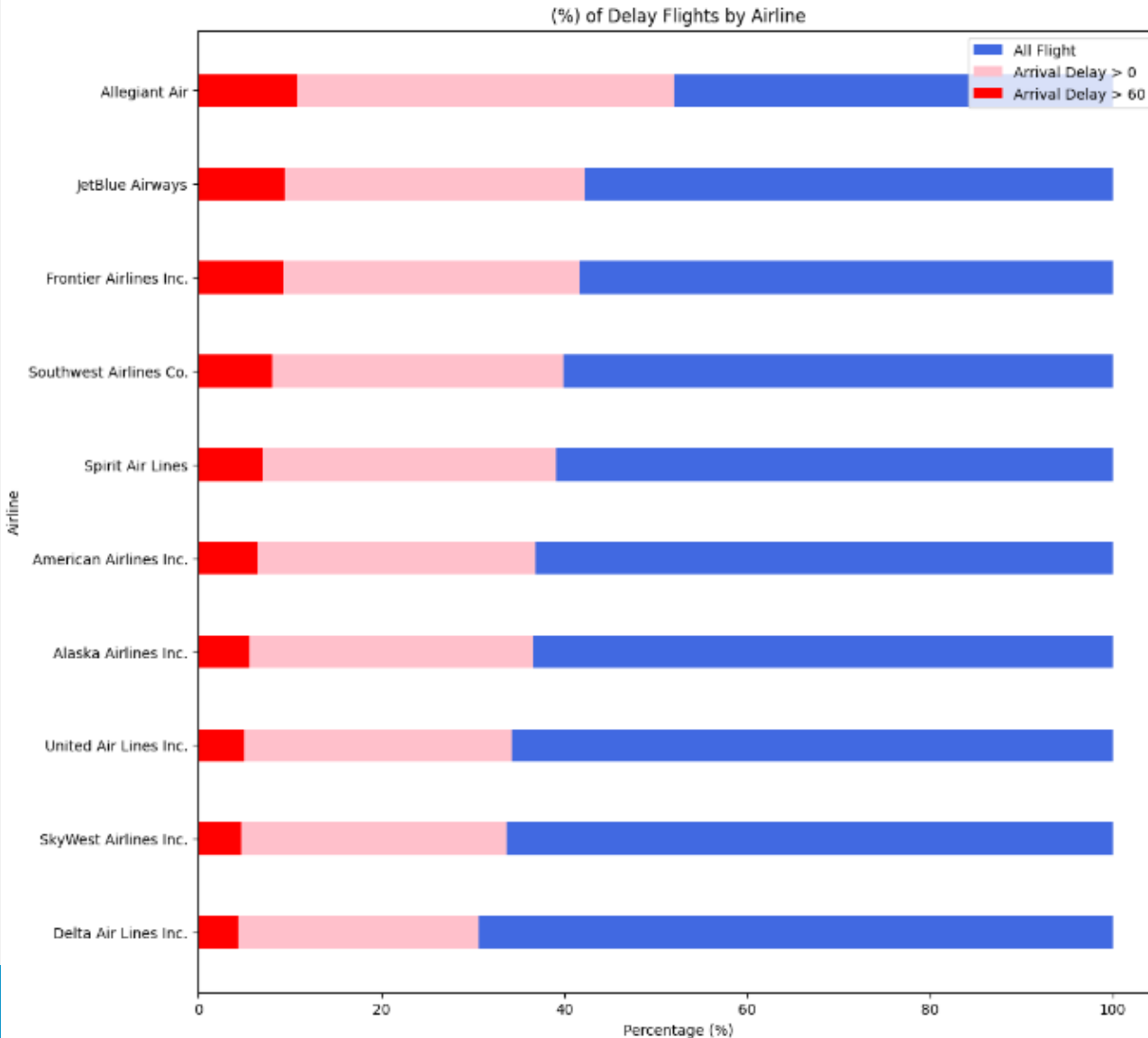
Individual Scoring Attributes of Airline



Airline Review Ranking

Top Airlines

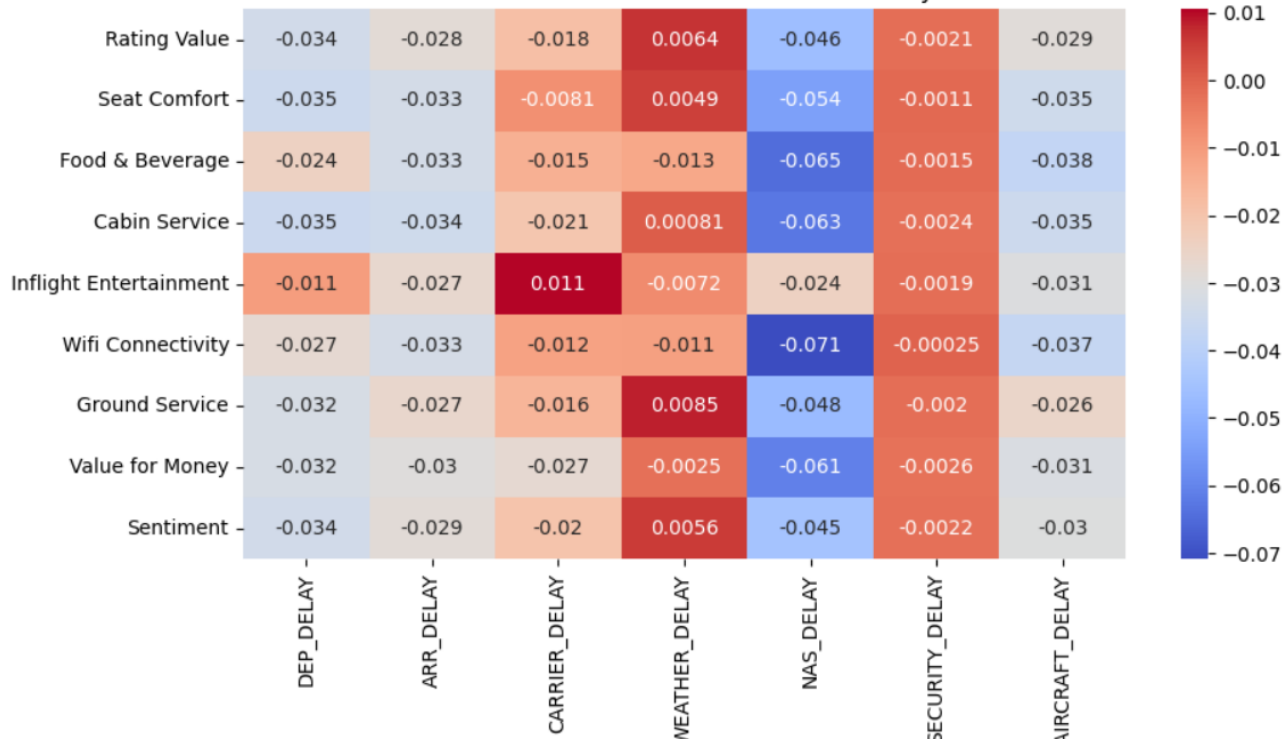
- ~5,000 Reviews per Airlines
- Skywest appears well, but only got 41 Reviews
- Delta seems to be the Winner of Rating Overall
- Results more varied than Airport Reviews



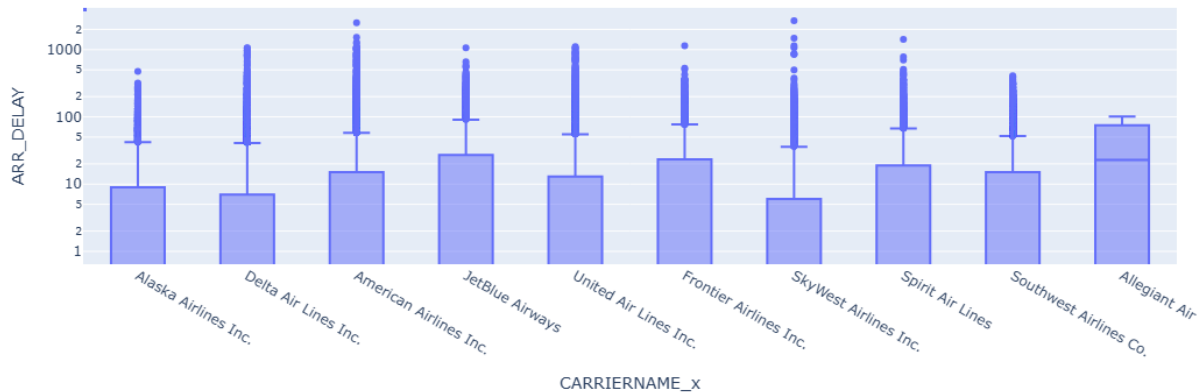
Airline Arrival Delay %

- Delta is great for On-Time Performance
- Allegiant is disproportionately Bad

Airline Review Score Correlation with Delays



Airline vs Arrival Delay



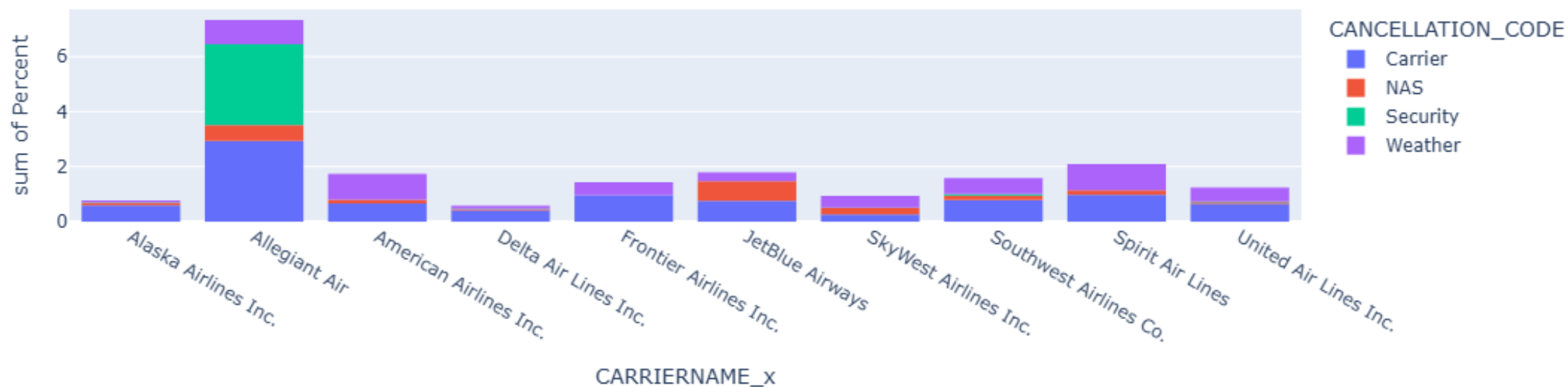
Airline Review Correlation with Delays

- Level of Service seems to reduce National Airspace System (NAS) Delay
- More Expensive Airlines > Less NAS Delay
- All Airlines have a Median Arrival Delay of 0
 - Except Allegiant > ~20min

Airline Cancellation

- JetBlue reports NAS Cancellation much more than others
- Allegiant cancels ~7% of their Flights, and >2% is for Security Reason

Airline Cancellation Code

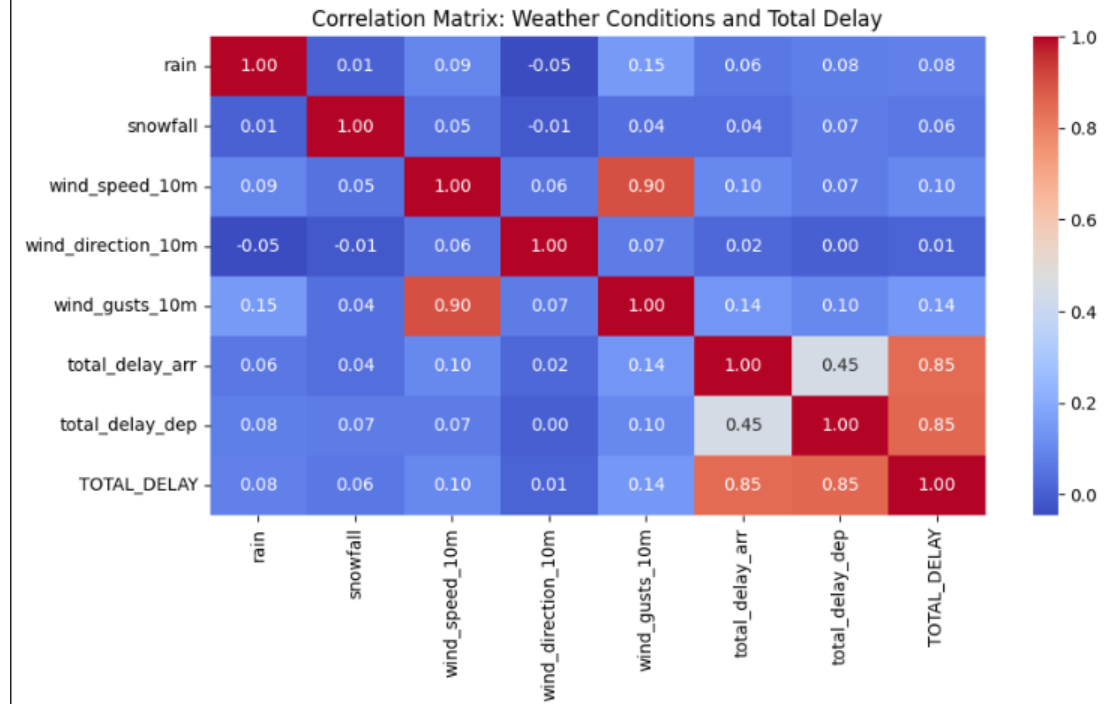
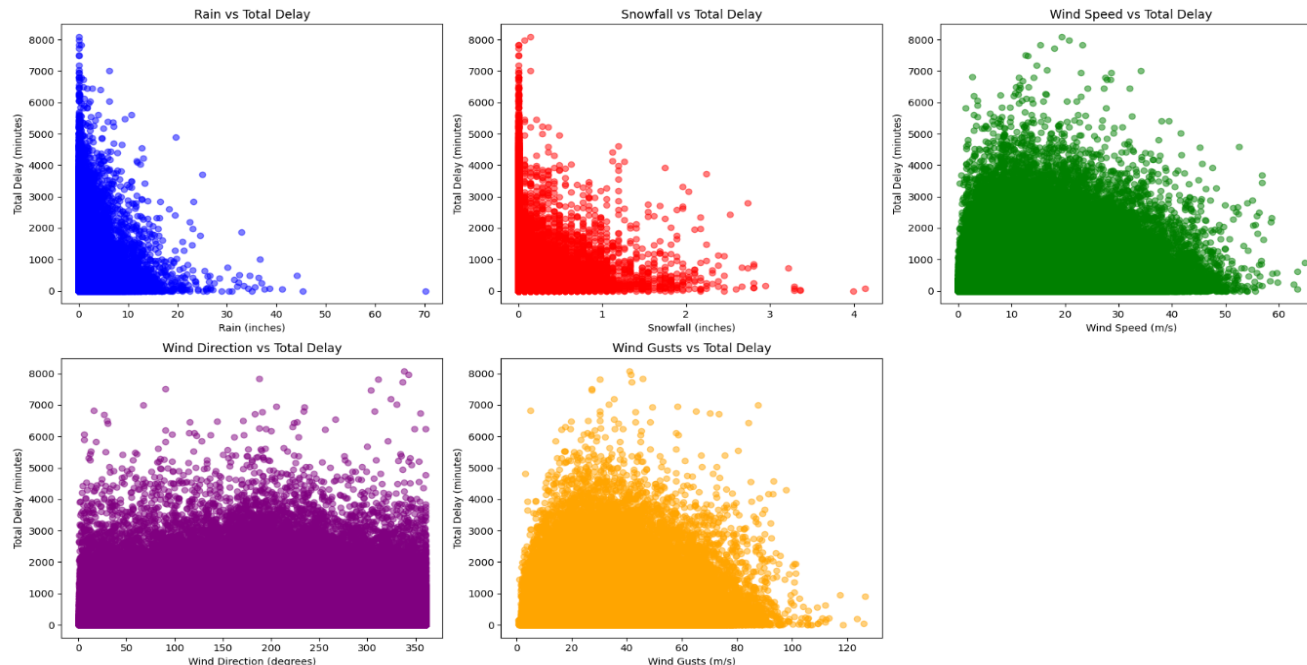


Airport Operation Disruptions



Relationship between Weather Conditions and Delays at Airports

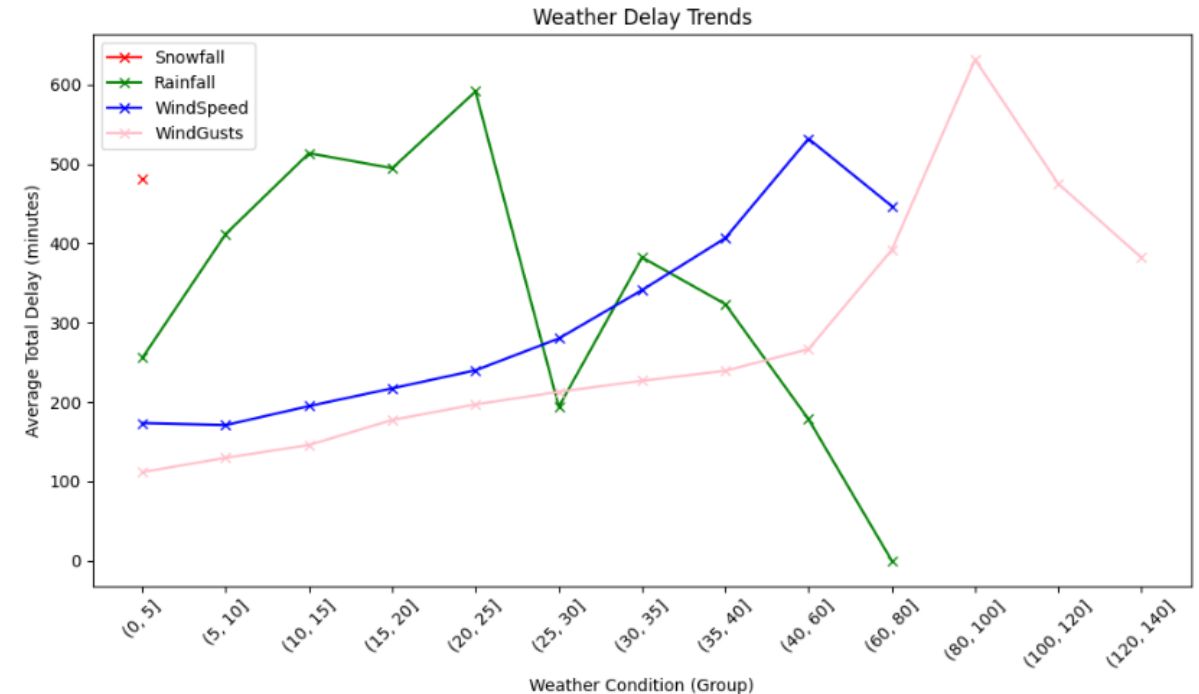
- Weak Relationship between hourly weather conditions and ARR and DEP
- Wind Speeds and Gusts have a moderate impact on delays, relatively stronger on ARR delays

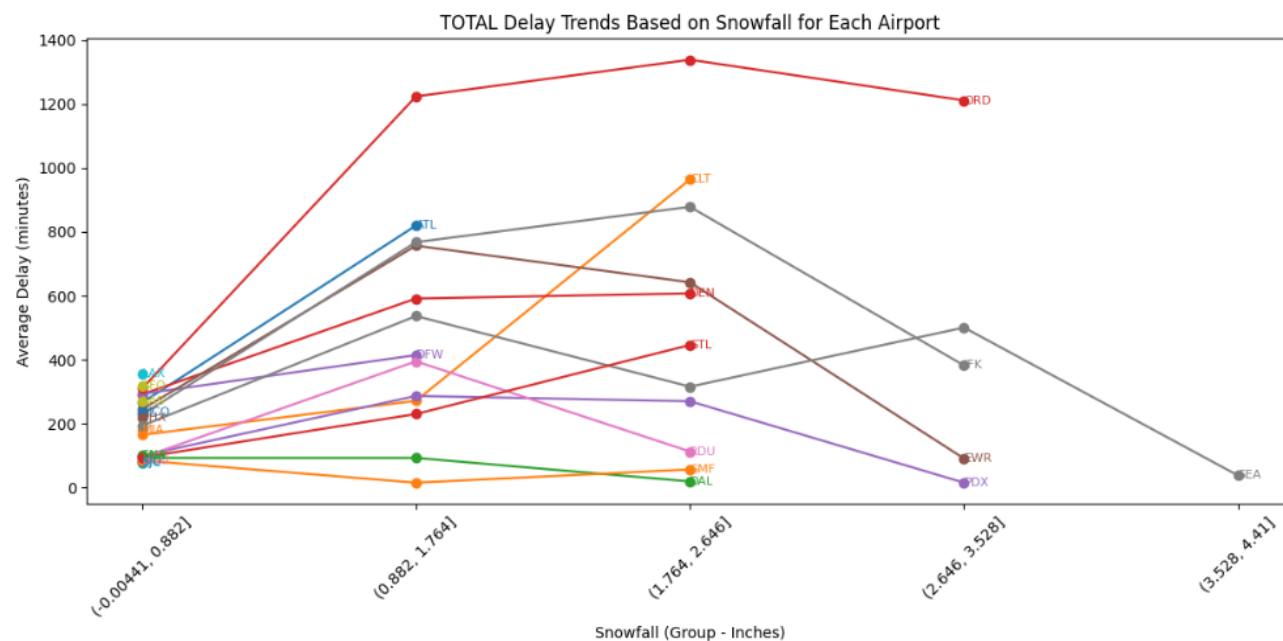
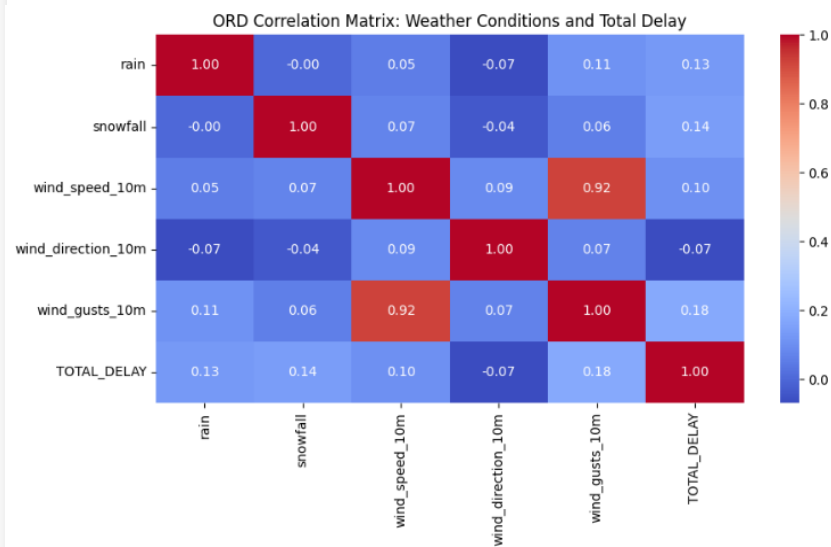
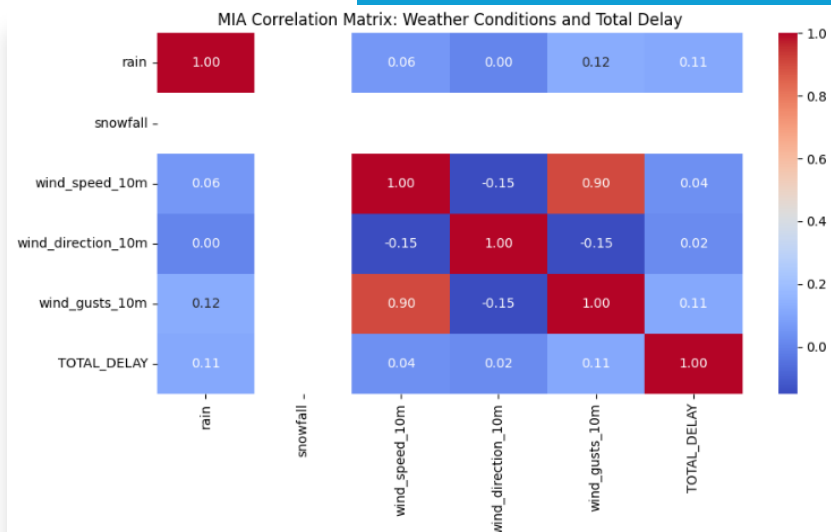


- Weather Conditions play a small role in the occurrence of delays
- Airline Operations , as well as infrastructure and human capital could be contributing to a larger extent

Weather Conditions and Delays Trends

- Increase in Wind Speed and Gusts are strong drivers of delay
- Drop after a certain threshold could mean that fewer data points exist at that range
- Snowfall does not occur at all airports
- Few extreme rainfall events

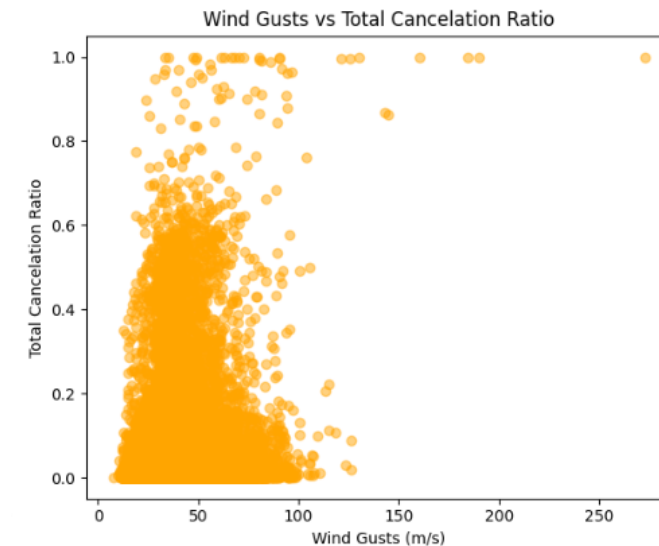
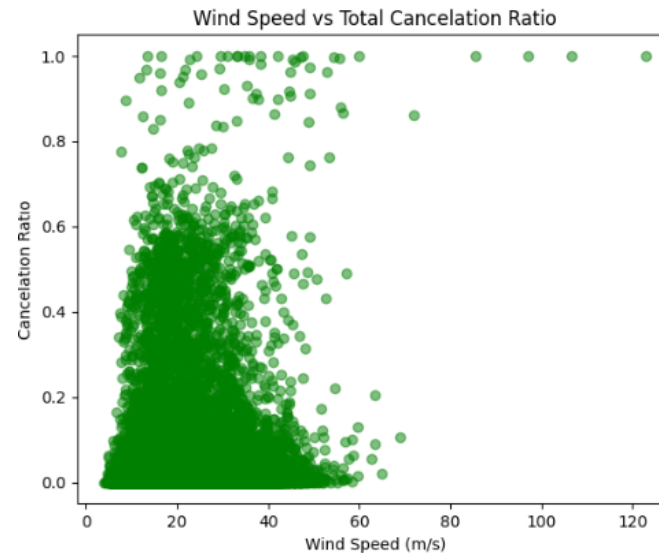
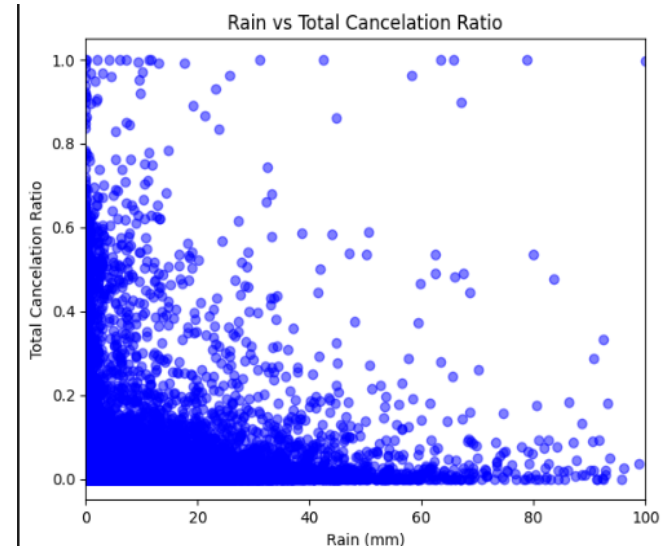
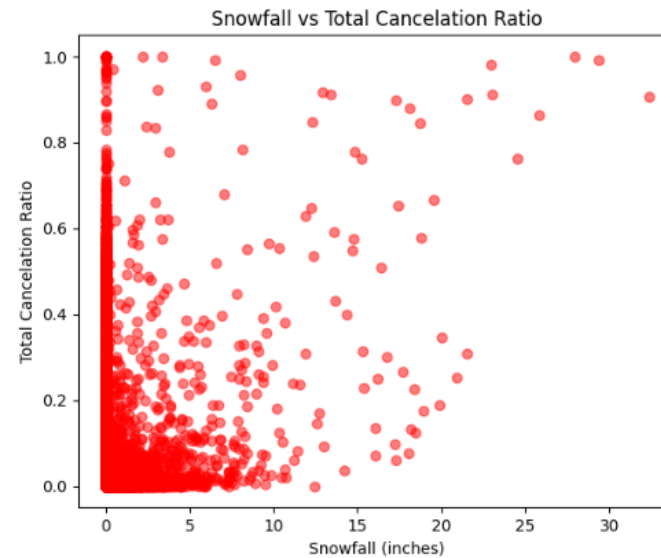




Weather Conditions and Delay Trends

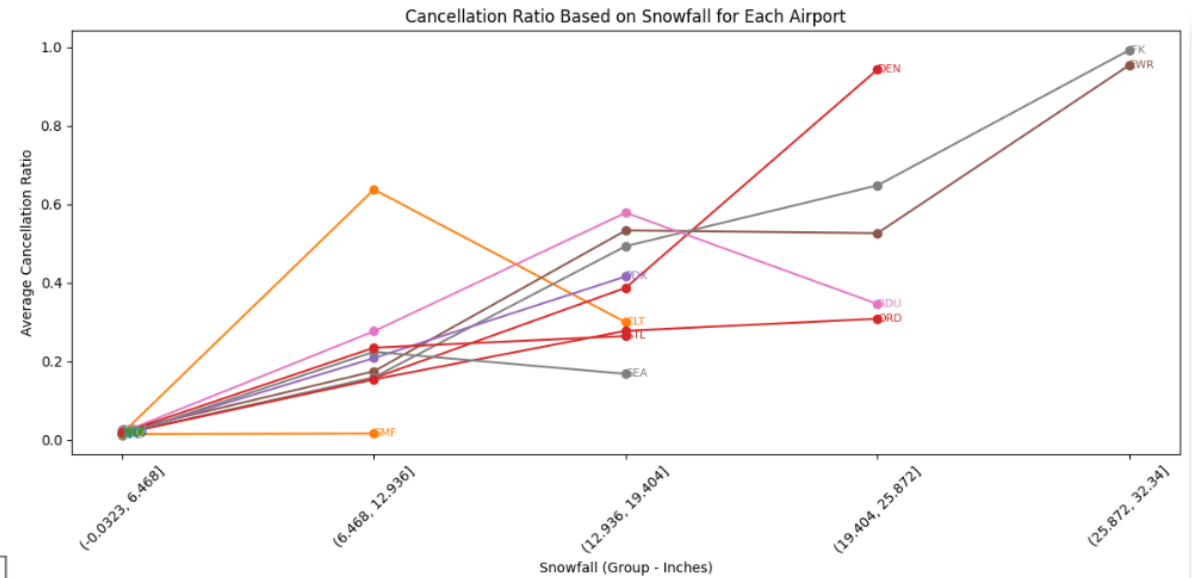
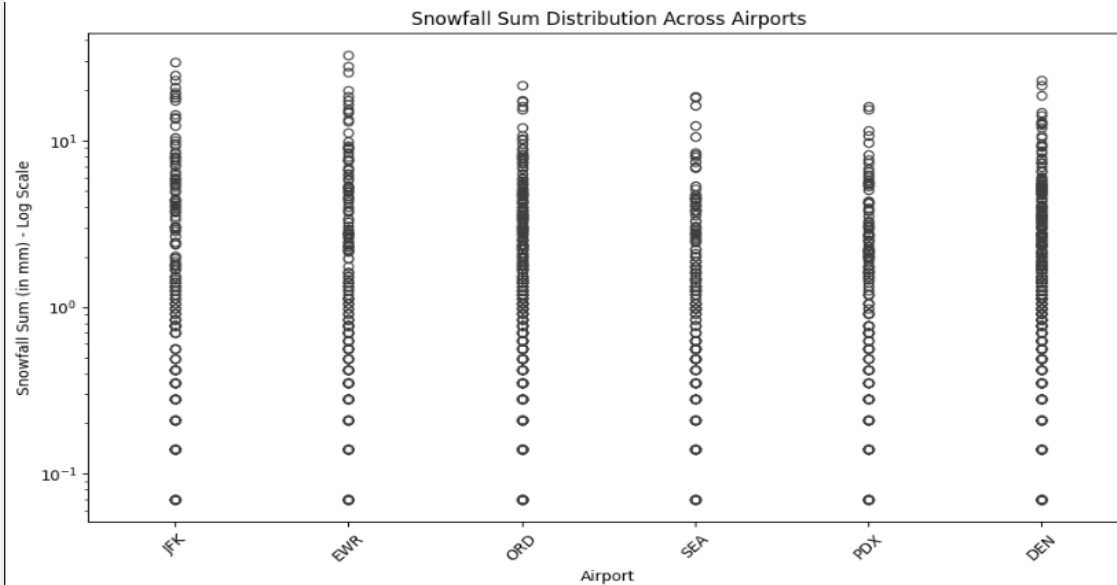
Weather Conditions and Cancellations

- Snowfall has a significant impact on flight cancellations. In particular, heavy snowfall (more than 5 inches) consistently leads to a higher cancellation ratio.
- There is no strong upward trend that indicates a clear, linear correlation between rainfall and cancellations.



Different Sensitivity to Snowfall:

- **JFK/EWR**(New York/New Jersey) and **DEN** (Denver) show a steep increase in cancellation ratio with higher snowfall.



Snowfall Distribution:

- Moderate snowfalls are very common at these airports, which correlate to higher frequency of flight cancellations

Weather Conditions and Cancellation Trends

Regression Analysis



Ridge Regression

- Polynomial Features, given Non-Linear relationships observed
- Using Standard Scalar (0,1) and regularization to help prevent overfitting

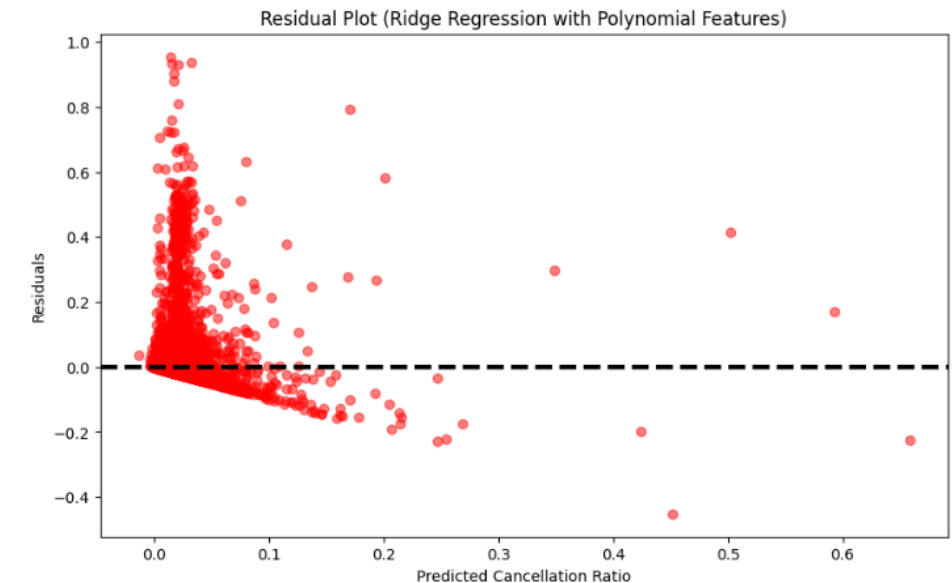
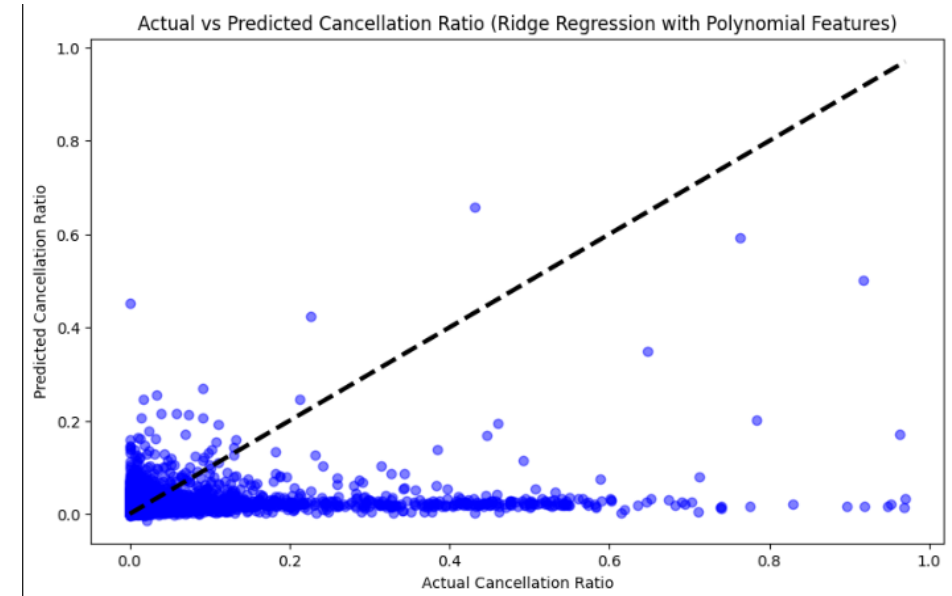
```
poly = PolynomialFeatures(degree=2, interaction_only=False, include_bias=False)
X_train_poly = poly.fit_transform(X_train_scaled)
X_test_poly = poly.transform(X_test_scaled)
```

✓ 0.0s

MSE: 3.46%

R-Squared: 0.049

	Feature	Coefficient
4	snowfall_sum	0.025039
3	rain_sum	-0.017247
21	weather_code rain_sum	0.014123
22	weather_code snowfall_sum	-0.009677
2	weather_code	0.006926

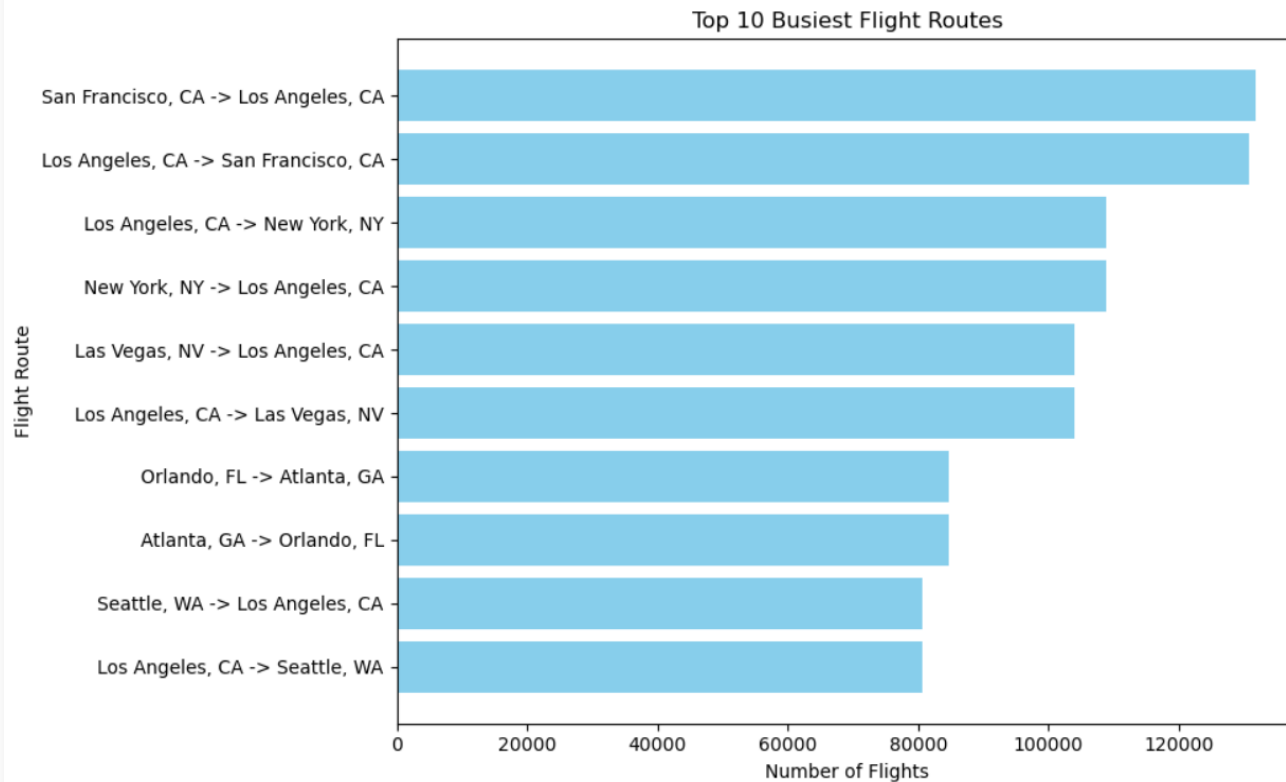




Routes



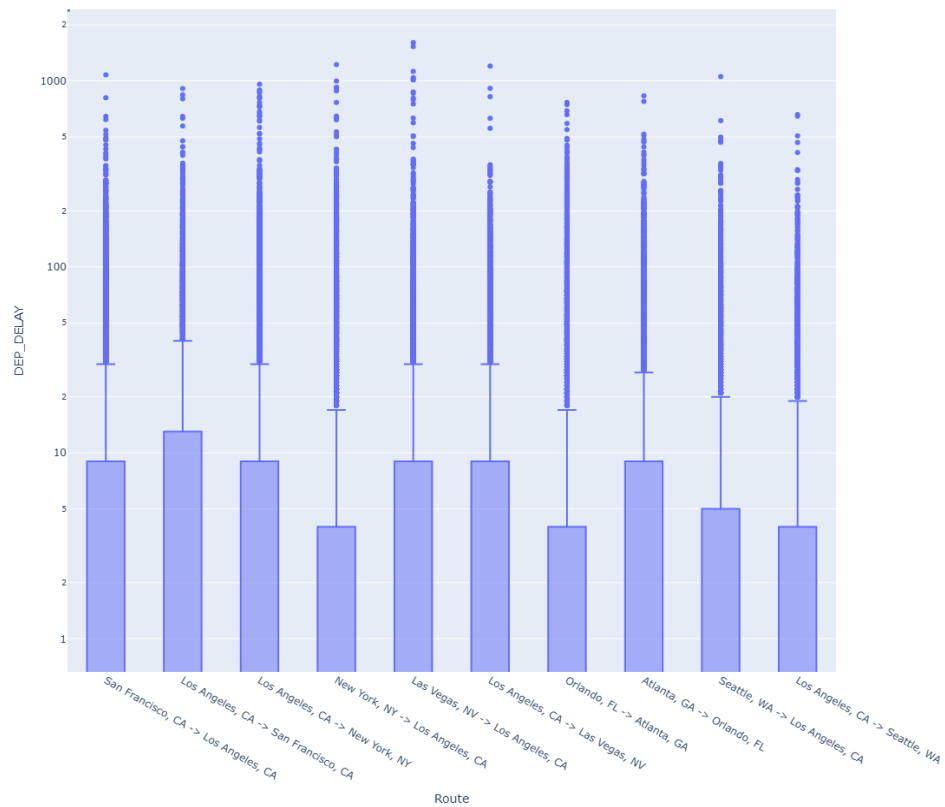
Top 10 Busy Routes



Hypothesis

- Busier Routes -> More Delays

Aiport Departure Delay

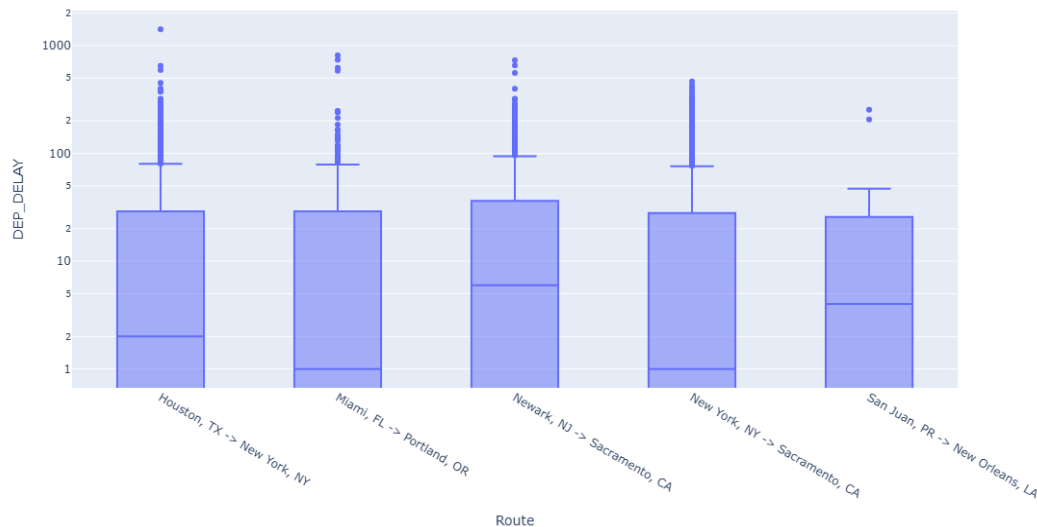


Top 10 Busy Routes

Result

- Top 10 Busiest Routes all have median Delay of 0 min
- The Correlation is 0.4%
- i.e. SF to LA, >30 Flights per Day
- i.e. Houston -> NY, ~1 Flight per Day
- 5 min Delay Median
- >30 min 25% of the Time

Route Departure Delay



Alternative Route Algorithm ??

```
for row in Each Row of Flight_Records:
    Baseline = row.Arrival_Time - row.Departure_Time + row.Arrival_Delay

    Optional_Flights = Filter(Flight_Records by Flight_Date +/- 1 Day and (Same Destination or Same Origin))

    Optional_Flights = Optional_Flights JOIN Optional_Flights ON Mid_Point

    Optional_Flights = Filter(Optional_Flights Transit_Time >= 60 min)

    Optional_Flights.New_Time = Optional_Flights.Arrival_Time_2nd \
        - Optional_Flights.Departure_Time_1st \
        + Optional_Flights.Arrival_Delay_2nd

    Possible_Suggestion = Filter(Optional_Flights by New_Time < Baseline)

return the Mean, Median, SD of Possible_Suggestion
```

Optimizing Arrival



Optimizing Delay

Route	Alternative Arr Delay Median	Alternative Arr Delay Mean	Example Route Reliability (%)
Houston, TX -> New York, NY	-9.5	-7.283299	97.255575
Miami, FL -> Portland, OR	-9.0	-8.605536	99.029126
New York, NY -> Sacramento, CA	-5.0	-4.071500	94.410786
Newark, NJ -> Sacramento, CA	-5.0	-3.935500	97.435492
San Juan, PR -> New Orleans, LA	-5.0	-2.579310	99.315068

Test Case: Top 5 Most Delayed Routes

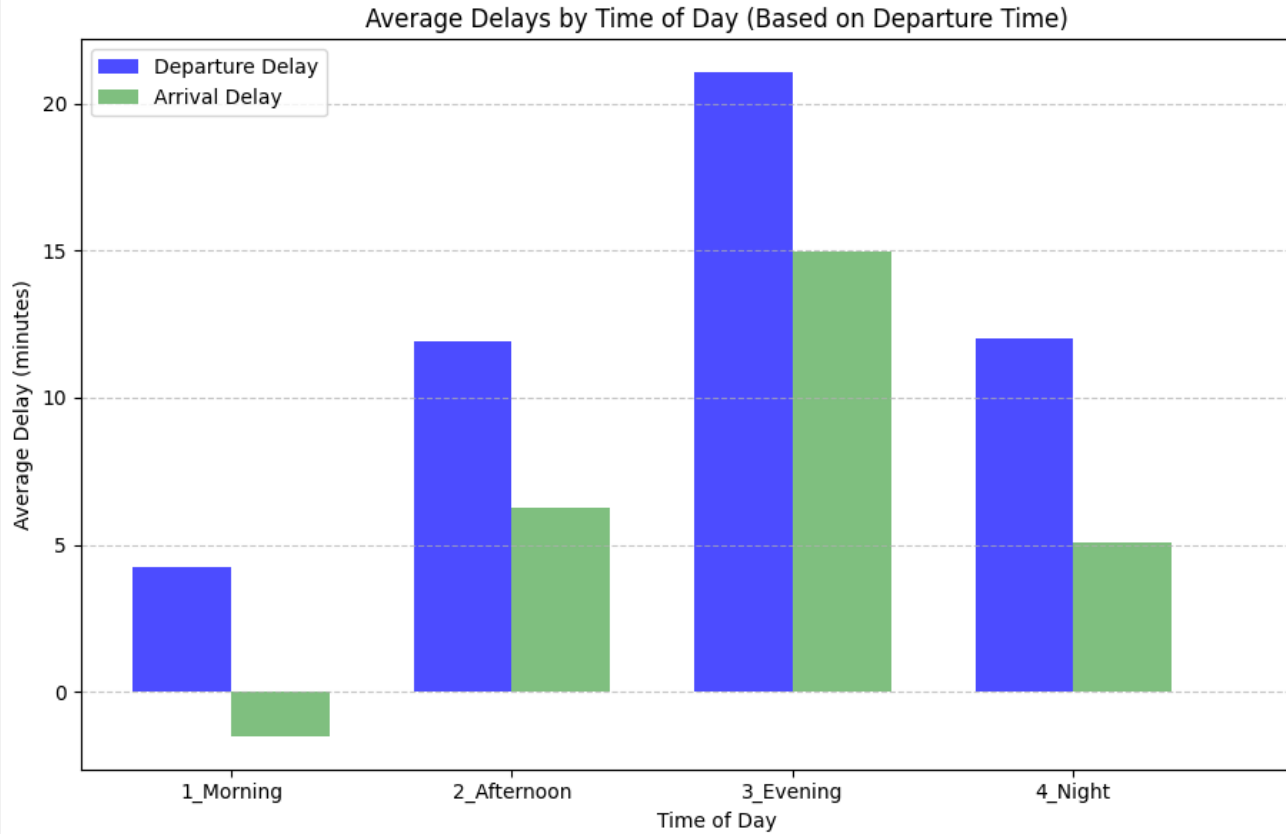
Optimizing Arrival Time

- Found Better Path in 0.25% of the Flights

Optimizing Delays Only

- Houston -> NY : Median Delay of 2min
- Houston -> New Orlean: Median Delay of -8min
- New Orlean -> NY: Median Delay of -4min

Miscellaneous Tips



When to Fly

- Morning (6am to 12pm)
- Avoid Evening Flights (6pm to 0am)

Obstacles and Improvements

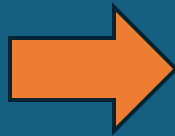


OBSTACLES

- **Difficult to find data regarding uncontrollable events such as:**
 - System Faults
 - Airport Closures/Incidents
- **Low Variance in Airport/Airline Reviews**
 - The data that we collected does not provide an explanation on the relationship between perception and on-time performance

IMPROVEMENT

- **Use Carrier specific data:**
 - Employee Turnover Rate
- **Labour Strike Frequencies**
- **Time Sequence Data**
 - Flight before/ after
 - Weather condition look-back



Ability to have
accurate
regression
models



End

