NYC Flights Analysis: Final

Chloe Lavery, Shirley Lee, Zach Nguyen,
Mihir Ranjan, Rhea Verma, Krishna Vijayendra

Project Summary

Hypothesis Questions:

a. What pattern(s) can be observed regarding flight punctuality in NYC?
b. Which of NYC's airports were most operationally efficient in 2013, and how has that changed to 2023? Which NYC airports have become more or less operationally efficient over the past decade?
c. How is bad weather at the original and destination airports affecting flight delays and on-time arrival rates? How much does the weather at either the origin or destination forecast flight delays?
d. Are certain geographic areas or types of carriers linked with better or worse flight arrival times and operational efficiency? What geographic or airport characteristics are linked with consistently superior on-time performance?
e. Do air-traffic disruptions (e.g. higher average departure / arrival delays) predict short-term equity returns (or volatility) for the major carriers?

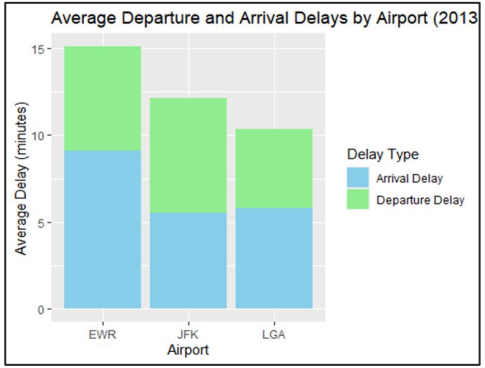
EDA: FLIGHT DELAY PATTERNS/VOLUME TRENDS

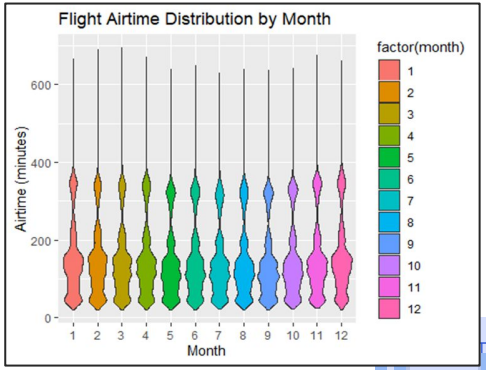
Areas of Interest:

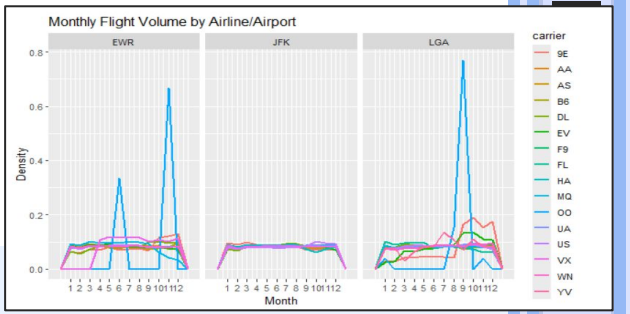
- Notable differences in **NYC Airports for delays**
- Differences in **Flight Airtime by month**
- If the three airports had significant differences in **flight volume by time of year**

Results:

- With an ANOVA table and Tukey test we determined there was a **significant difference in in both arrival and departure delays** for all three airports at 95% confidence
- **No significance** when comparing **airtime by month**
- JFK was normalized when looking at flight volume, the other airports had spikes at unexpected times
    - Unsure what caused these spikes, possibilities include school season







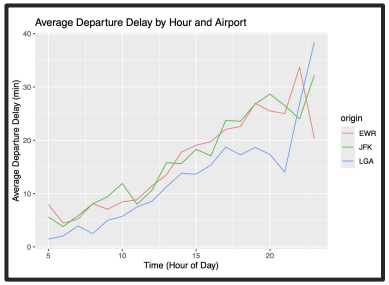
PATTERNS IN FLIGHT PUNCTUALITY

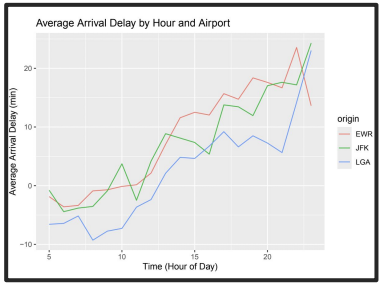
Hypotheses:

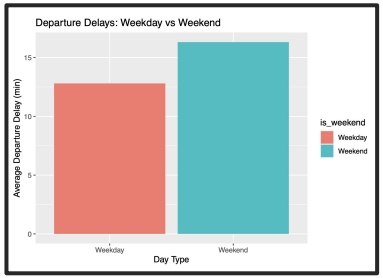
- Early **morning flights are more punctual than evening** and night flights, with peak delays in the evening.
- **JFK Airport** has the **highest average delays** due to high traffic and international flights.
- **Weekend flights have more delays than weekday flights** due to increased volume.
- Would likely a **longer flight airtime** distribution **during winter** months (December, January, February) due to turbulent weather conditions.

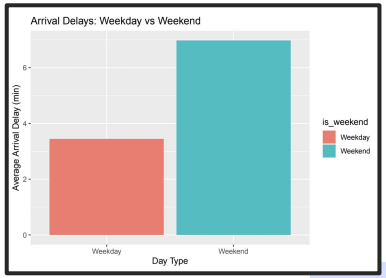
Conclusions:

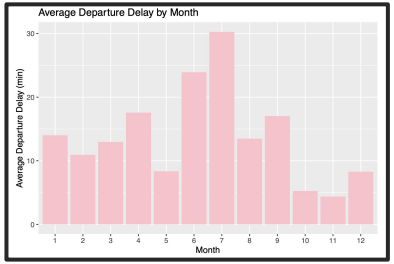
- Flights are **most punctual in the early morning**; delays increase throughout the day.
- **LaGuardia** generally experiences **fewer delays** than **JFK and Newark.**
- **Departure delays are greater** than arrival delays.
- **Weekends show lower punctuality** than weekdays.
- Delays **peak in summer**, especially July; fall (Aug-Nov) is the most reliable season.

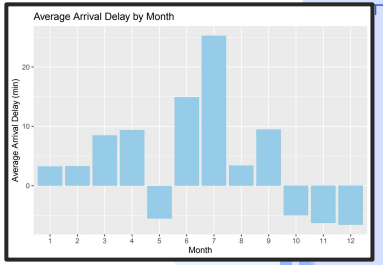








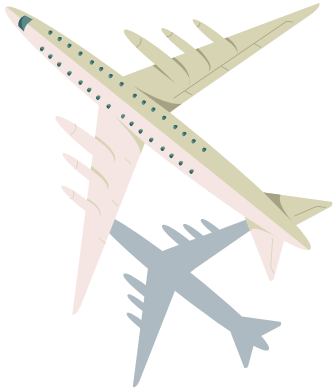




2013 VS. 2023 DELAY ANALYSIS

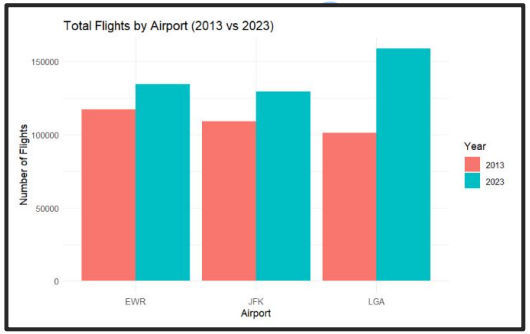
Hypotheses:

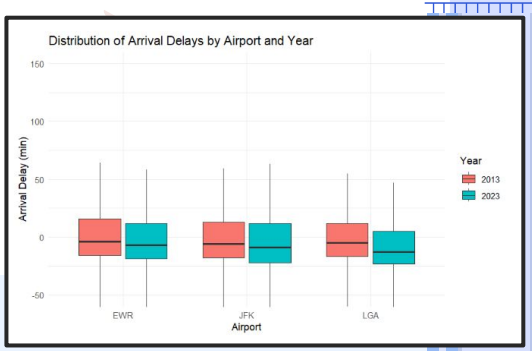
- **Flight volume is expected to slightly decrease**, matching the slight decline in New York City population statistics.
- Arrival and departure delays are **expected to trend similarly** across time and airports.
- **Efficiency (in terms of reduced delays) may improve post-COVID** due to operational reforms despite flight volume.



Conclusions:

- **Flight volume** significantly increased across all three airports.
- **Arrival delays** overall decreased from 2013 to 2023, especially at LGA.
- **Departure delays showed increases** across all airports.
- While there were some **notable improvements and increases**, no consistent decade-long trend in overall efficiency emerged.





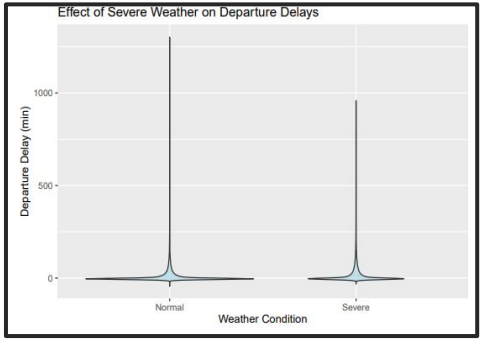
WEATHER & FLIGHT PUNCTUALITY

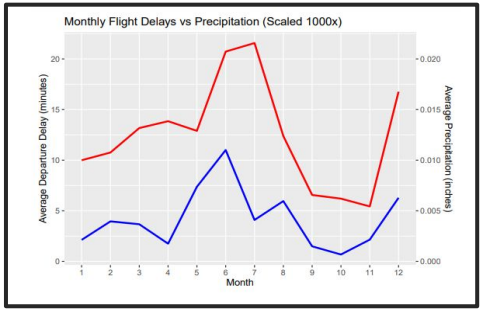
Hypotheses:

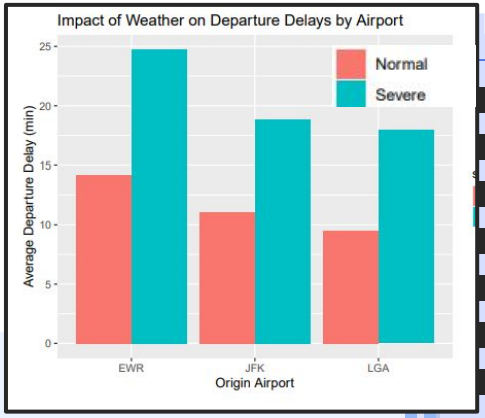
- Weather conditions impact the departure and arrival times of flights, with **more severe weather (wind, rain, etc.) causing greater delays**.
- There is a **monthly trend in delays**, correlating with seasonal weather.
- **Certain airports are more resilient to delays** in severe weather as they have less weather variability.
- Different **weather conditions/variables (wind, rain, etc.) can be used to predict** flight delays.

Conclusions:

- Weather conditions do impact departure/arrival times, however, **other factors may have a greater influence on departure/arrival times** as delays during normal weather show more variability.
- There is a **monthly trend in delays** that correlates with seasonal (“bad”) weather, but external factors, like increased summer travel volume, also influence delays.
- Certain airports are more resilient to delays in severe weather, **with LGA showing the lowest delays among NYC airports**, while EWR consistently has the highest. Other operational factors may contribute.
- Weather variables are statistically significant predictors of departure delays, **with precipitation having the strongest impact**, but the linear regression model explains only a small portion of delay variance, indicating additional influencing factors.







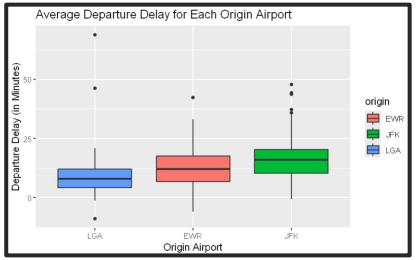
FLIGHTS PUNCTUALITY BY LOCATION/CARRIER

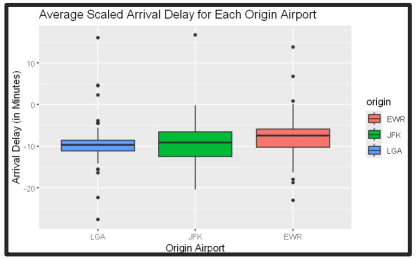
Hypotheses:

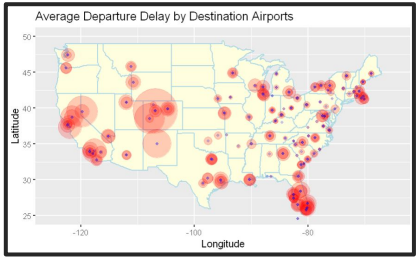
- **International airports** will have **more departure delays** than **domestic airport**.
- **Longer flight distances** tend to lead to **more arrival delays**.
- **Budget airlines** will have **more departure delays**.
- **Arrival delays** are affected by **departure delays** and **air time**, but not too much by carrier or airport locations.

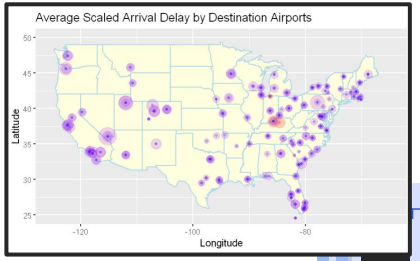
Conclusions:

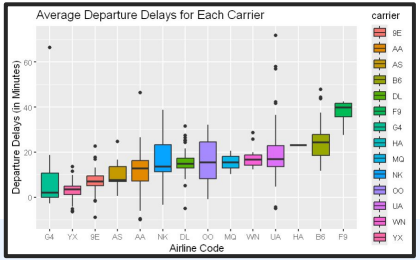
- **JFK** and **EWR** have **more departure delays** than LGA.
- Airports on the **West Coast** and **South/Southeastern States** seem to have **more departure delays**.
- **F9** and **B6** (both are **budget airlines**) have the **worst departure delays**; however, **G4** (also a **budget airline**), have the **least departure delays**.
- After **normalizing** out the departure delays, **arrival delays did not vary much** among airports and carriers.
- In general, **international airports** and **budget airlines** have **more departure delays**.
- **Arrival delays** after normalization tend to be **negative** for most airports and carriers.

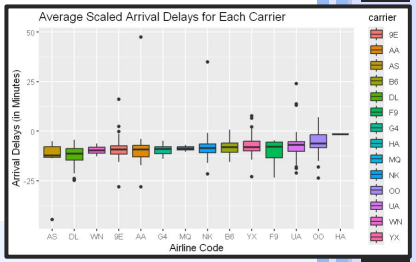












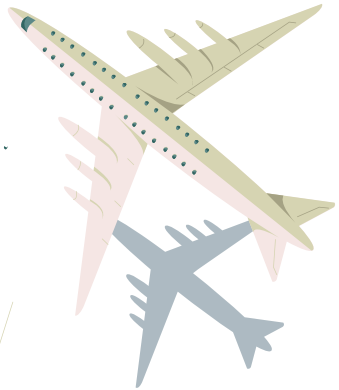
AIRLINE DISRUPTIONS AND VOLATILITY

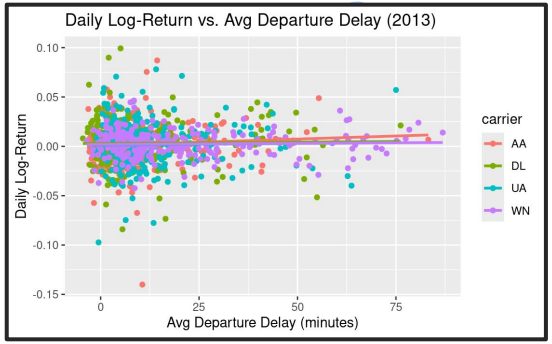
Hypotheses:

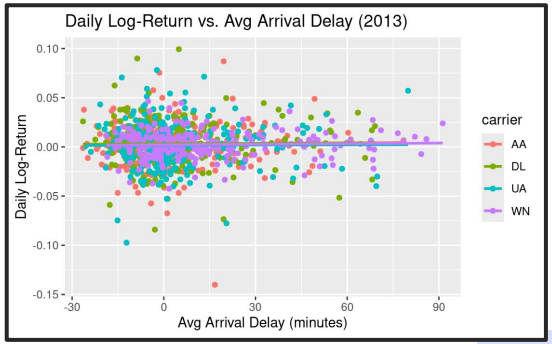
- Expected that **more airline disruptions, the carrier stock return is lowered** and vice versa.

Conclusions:

- Scatterplot(s)
  - Almost no slope
  - Widely scattered plot
  - No obvious visual pattern
- Regression Summary
  - P-values for each coefficient is >0.05
  - Multiple R-squared value is around 0.00217








```
Call:
glm(formula = pct_return ~ avg_dep_delay + avg_arr_delay + factor(carrier),
    family = gaussian, data = daily_panel)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.202e-03  1.622e-03   0.741    0.459
avg_dep_delay       1.330e-04  1.178e-04   1.129    0.259
avg_arr_delay      -7.639e-05  9.071e-05  -0.842    0.400
factor(carrier)DL   9.331e-04  1.892e-03   0.493    0.622
factor(carrier)UA  -7.567e-04  1.898e-03  -0.399    0.690
factor(carrier)WN -5.032e-04  1.959e-03  -0.257    0.797

(Dispersion parameter for gaussian family taken to be 0.0004489638)

    Null deviance: 0.44904  on 1003  degrees of freedom
Residual deviance: 0.44807  on  998  degrees of freedom
AIC: -4882.2

Number of Fisher Scoring iterations: 2

`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
[1] 0.0004528667
```

Project Conclusion

Conclusion:

a. Flight Punctuality Analysis: Flights tend to depart with more delay than they arrive, with delays increasing later in the day. Weekends and summer see more delays than the weekdays and winter.
b. 2013 vs. 2023 Analysis: There does seem to be a slight increase in departure delay coinciding with a slight decrease in arrival delay across all three airports. Some results, like arrival delay in LGA or departure delay in JFK, were notable, but there isn't really a consistent trend that suggests that airports have improved their efficiency regarding delays after 10 years.
c. Weather Analysis: Weather trends do have an impact on flight delays, however, other factors may be better predictors of flight trends. Although weather variables are statistically significant predictors of departure delays, the linear regression model explains only a small portion of delay variance.
d. Geographic/Carrier Analysis: International airports are more likely to have longer departure delays due to more potential air traffic. Arrival delays did not differ significantly among airports and carriers.
e. Volatility Analysis: There was almost no correlation between airport delays and stock prices with a very low R-sq. Furthermore, the MSE for 10-fold CV is 0.0004529, so the model is not overfitting.

Team Member Contribution

Chloe Lavery: Weather and Flight Punctuality Analysis, Project Summary, Project Results, Final Report Editing
Shirley Lee: Comparing Flights by Carrier/Location Graphs, Project Summary, Project Results, Final Report Editing
Zach Nguyen: 2013 vs. 2023 Graphs/Analysis, Final Report Compiling, Project Summary, Project Results
Mihir Ranjan: Effect of Volatility based on Delays, Project Summary, Project Results
Rhea Verma: Patterns in Flight Punctuality Analysis, Project Summary, Project Results, Final Report Compiling
Krishna Vijayendra: EDA Analysis Graphs, Commentary, Project Summary, Project Results

Thank you!