FLIGHT DELAY PREDICTION

MACHINE LEARNING

PILAR ESCRIG
VALERIA VALEVASHNIKOVA
ROCÍO ROMO
IVÁN SIMÓN

INTRO

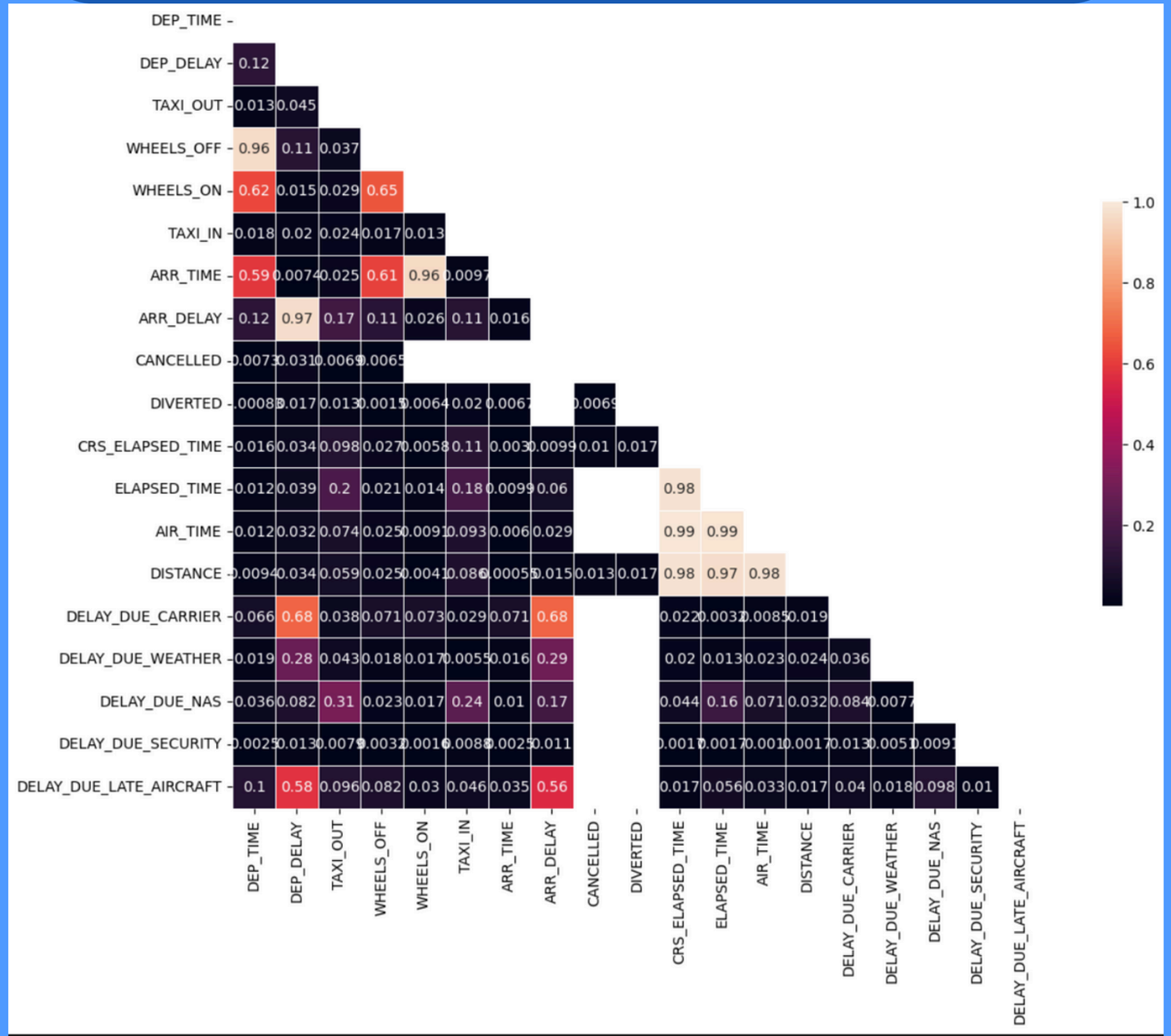
- Explore and analyse USA flights over a 12 months period
- Predict wether a flight will be delayed/cancelled based on different metrics.

DATASET SELECTION

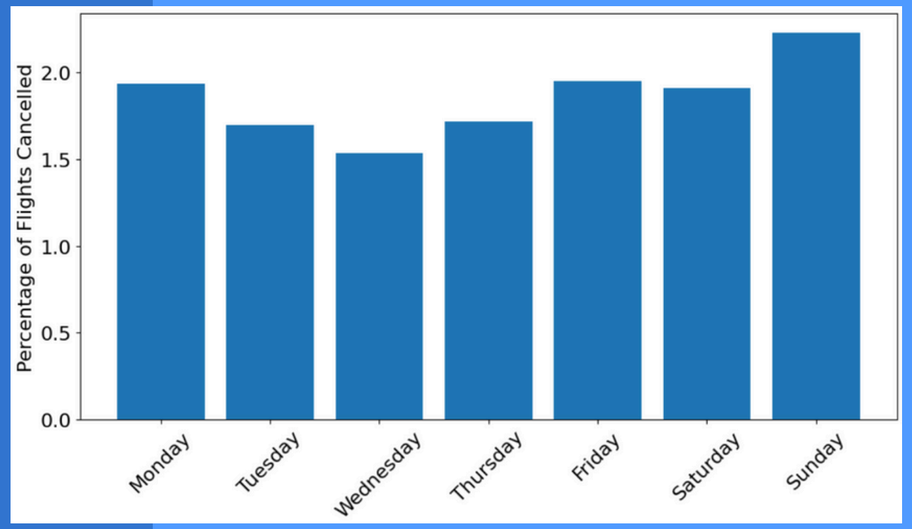
- Kaggle
- Flight Delay and Cancellation Dataset (2019-2023)

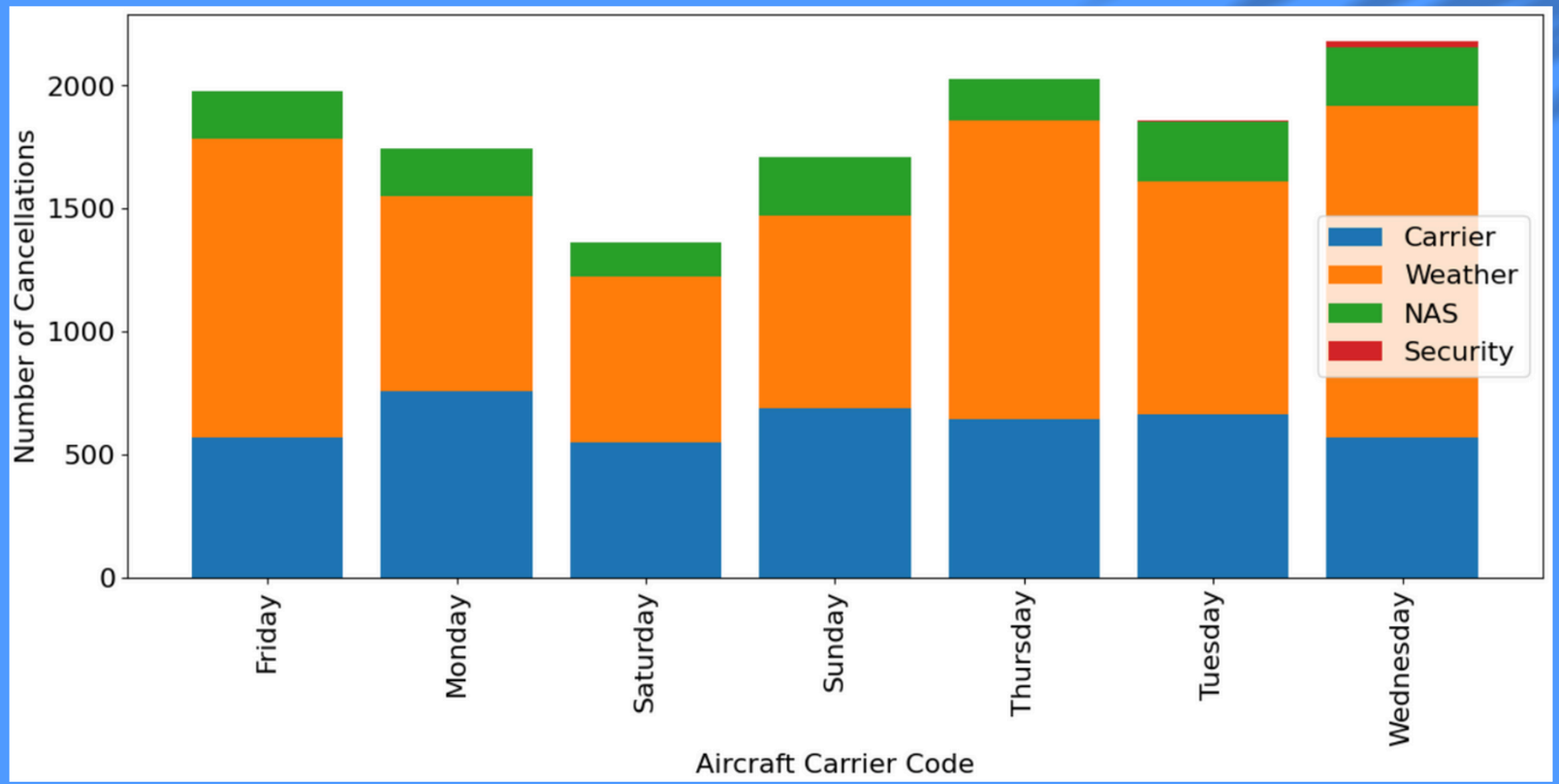
SEPTEMBER 2022 - AUGUST 2023

FEATURE ENGINEERING

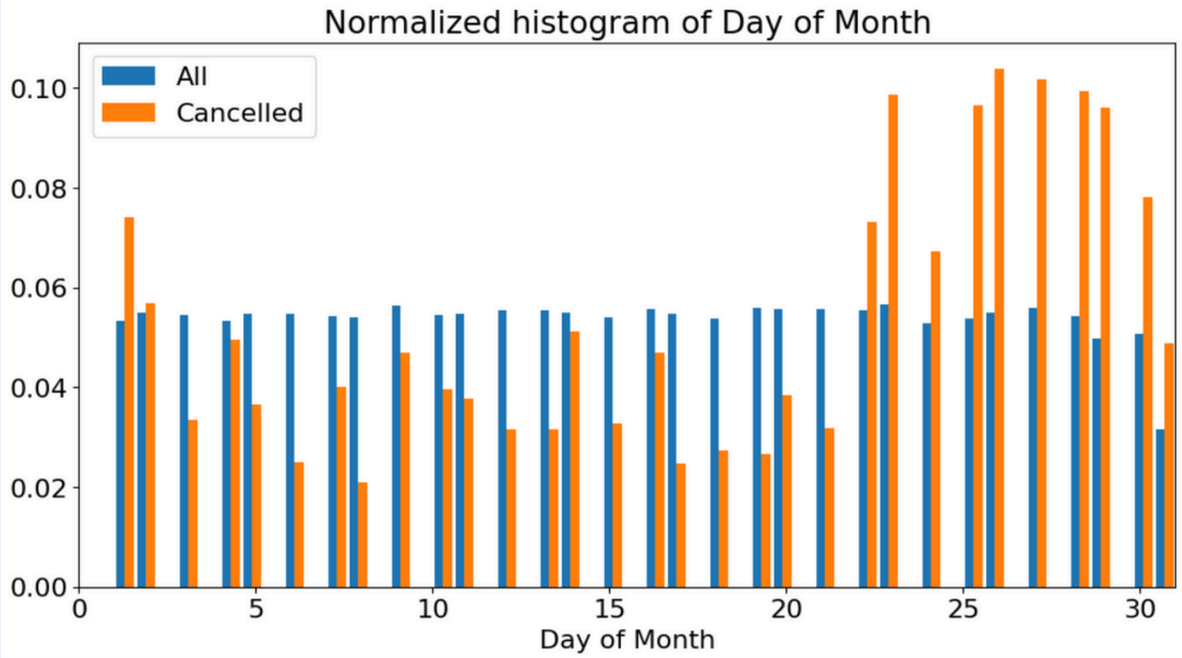


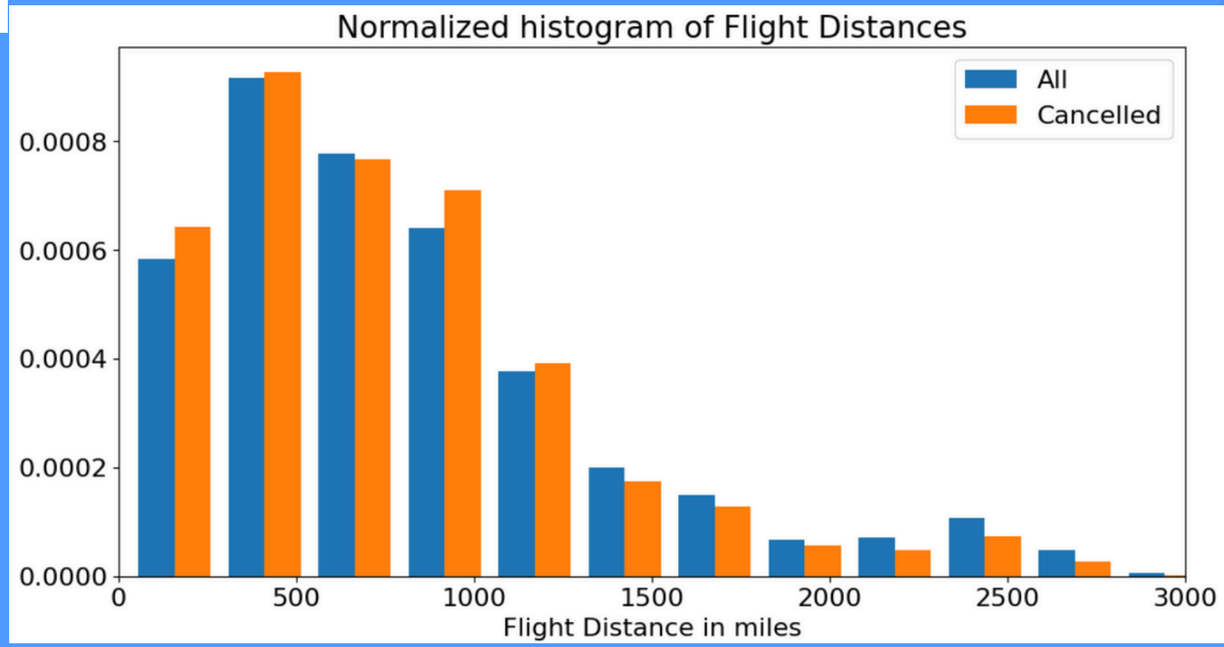
CANCELATIONS DAY OF WEEK



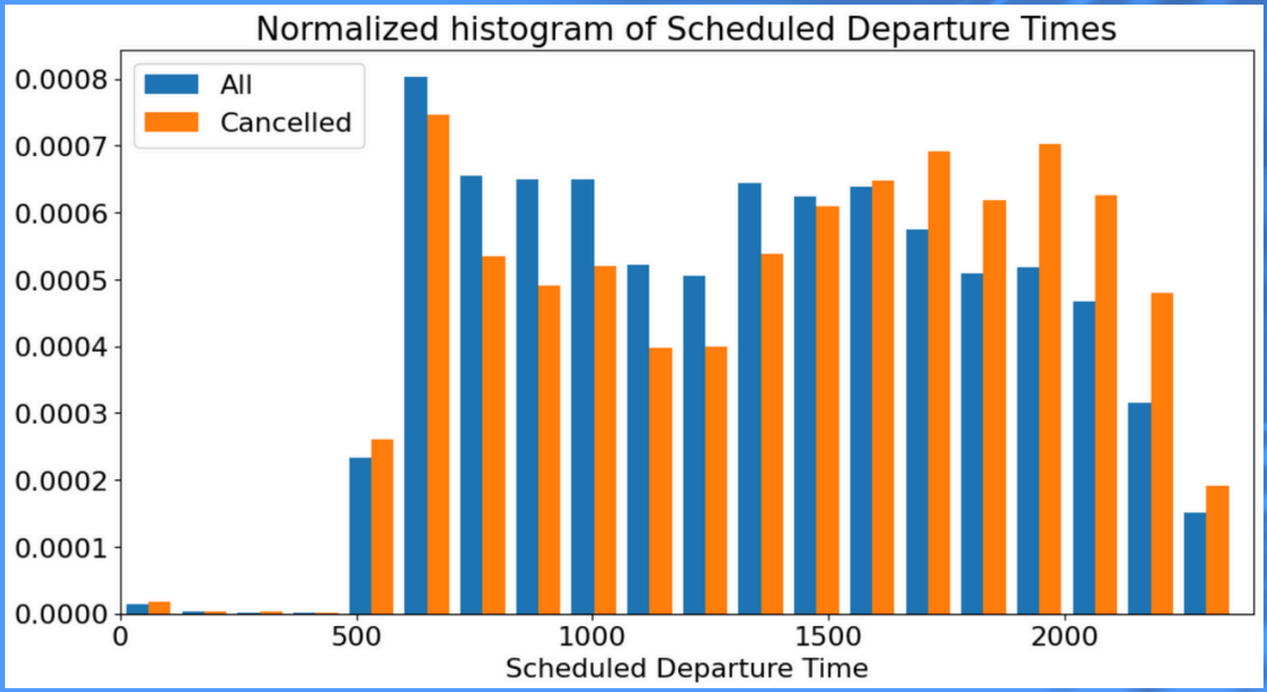


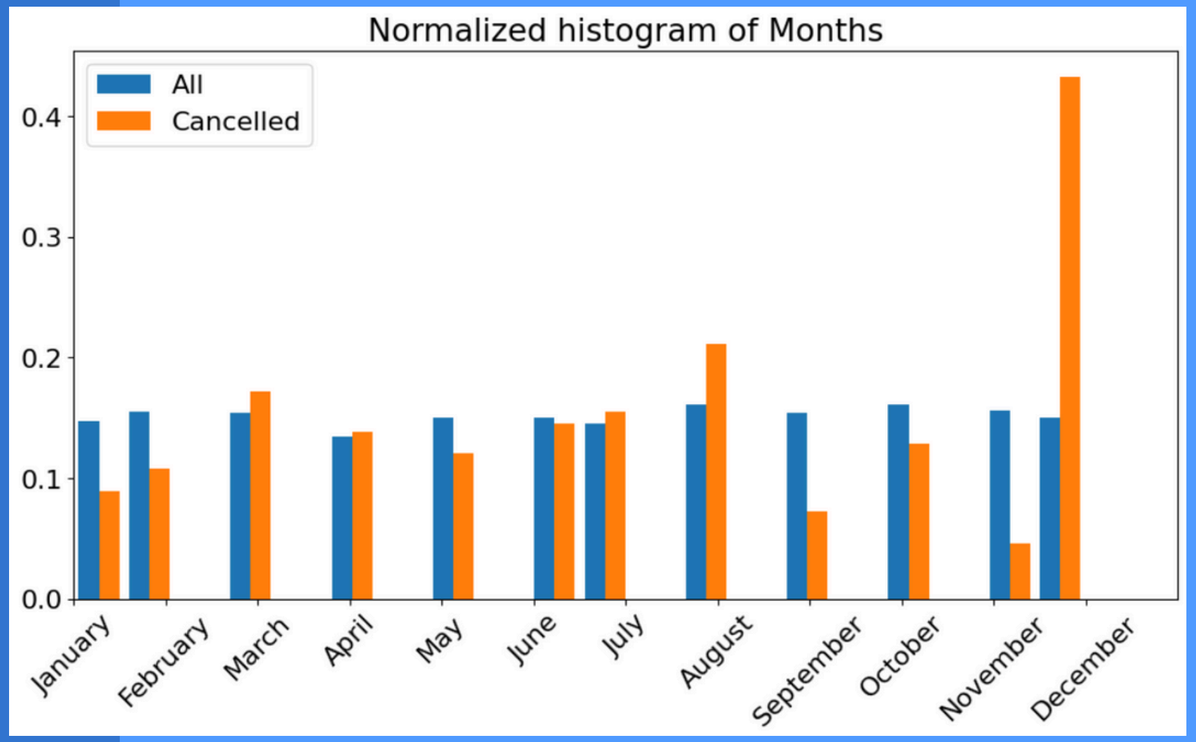
CANCELATIONS DAYS OF MONTH AND DISTANCE



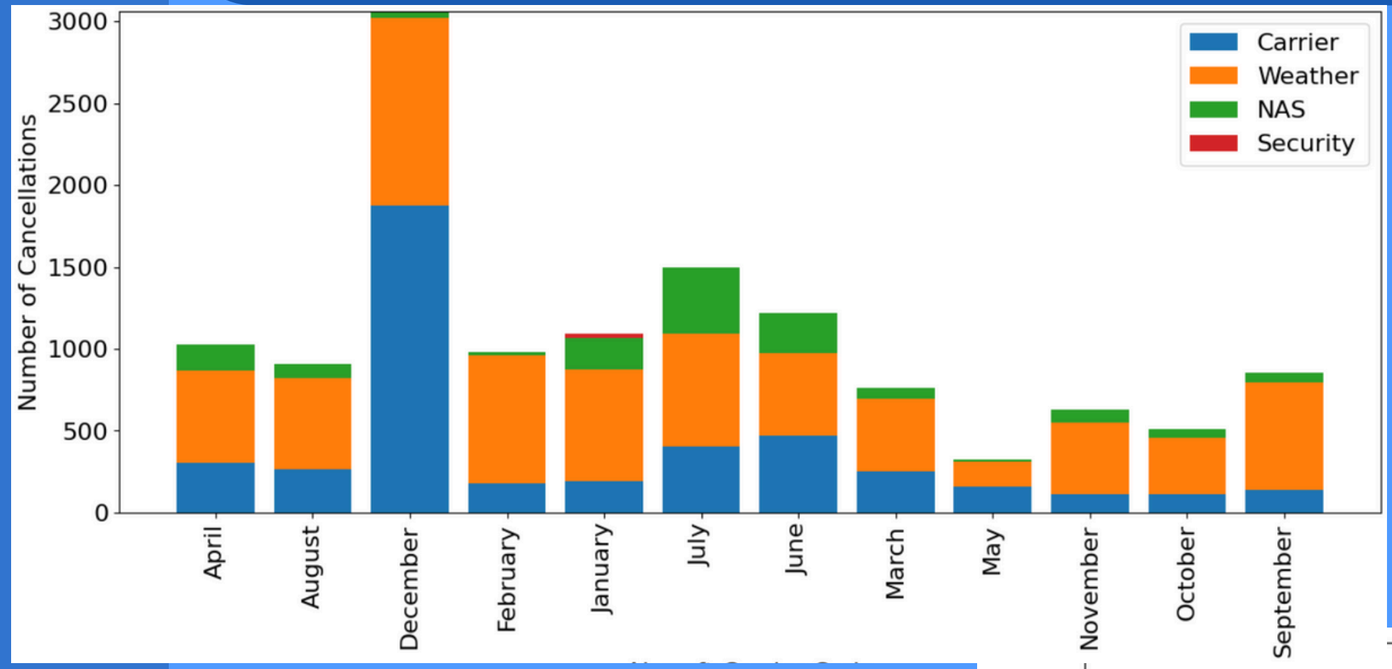


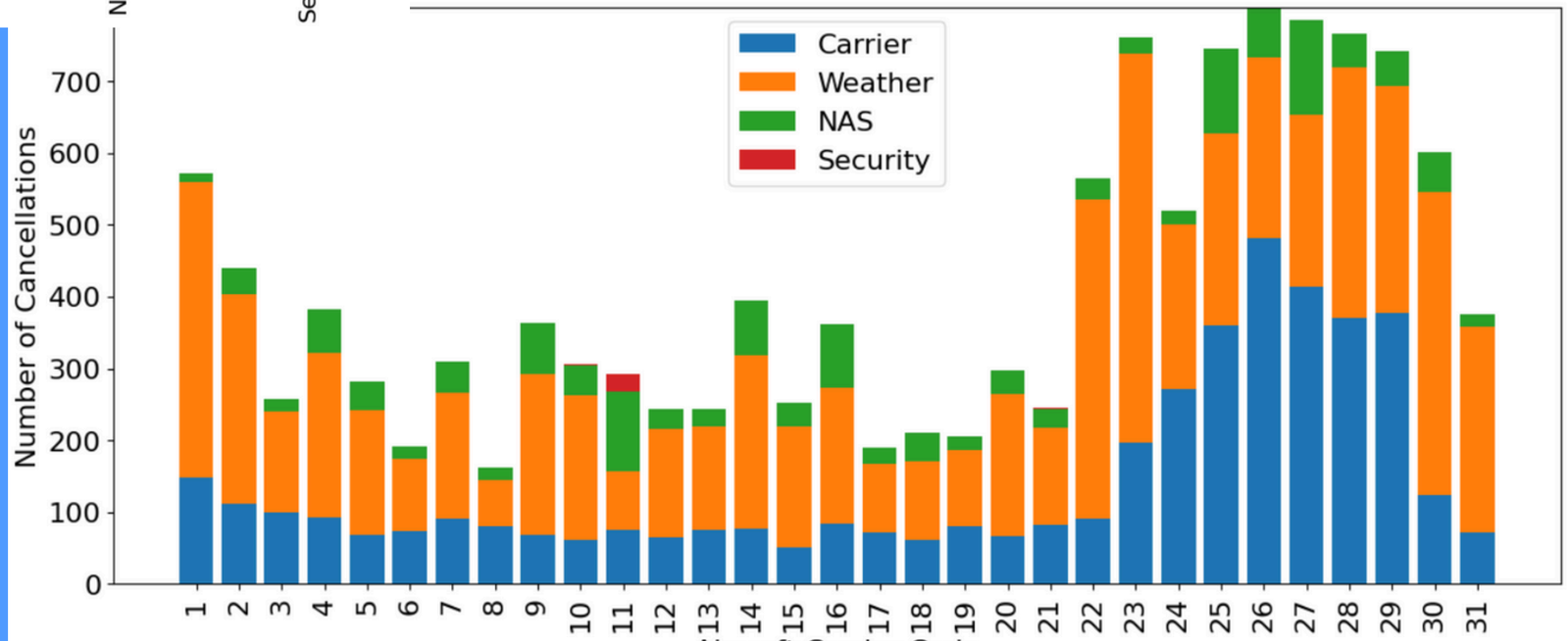
CANCELATIONS



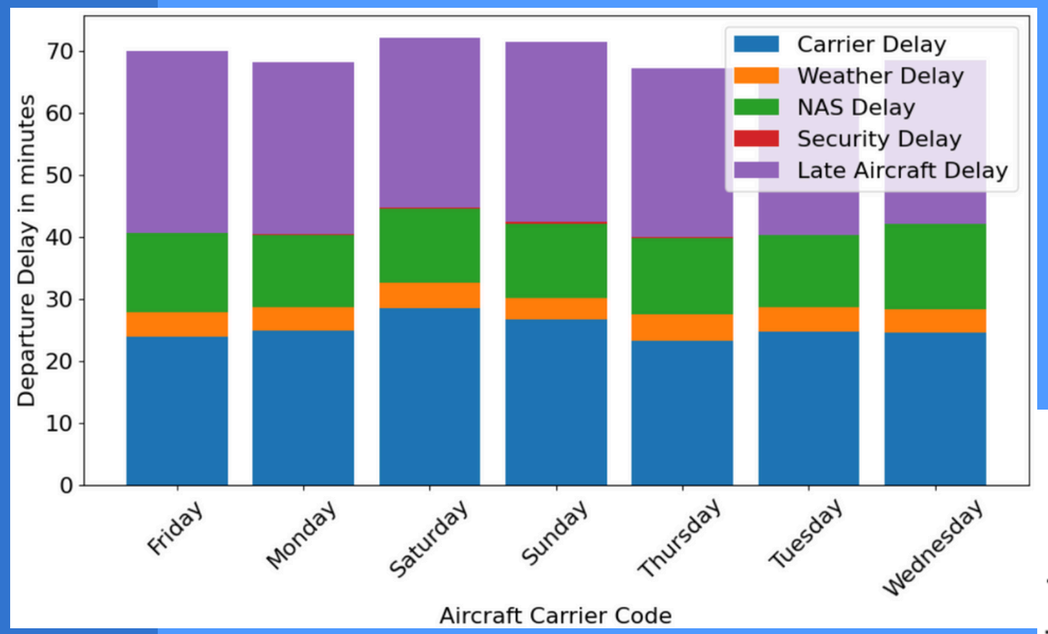


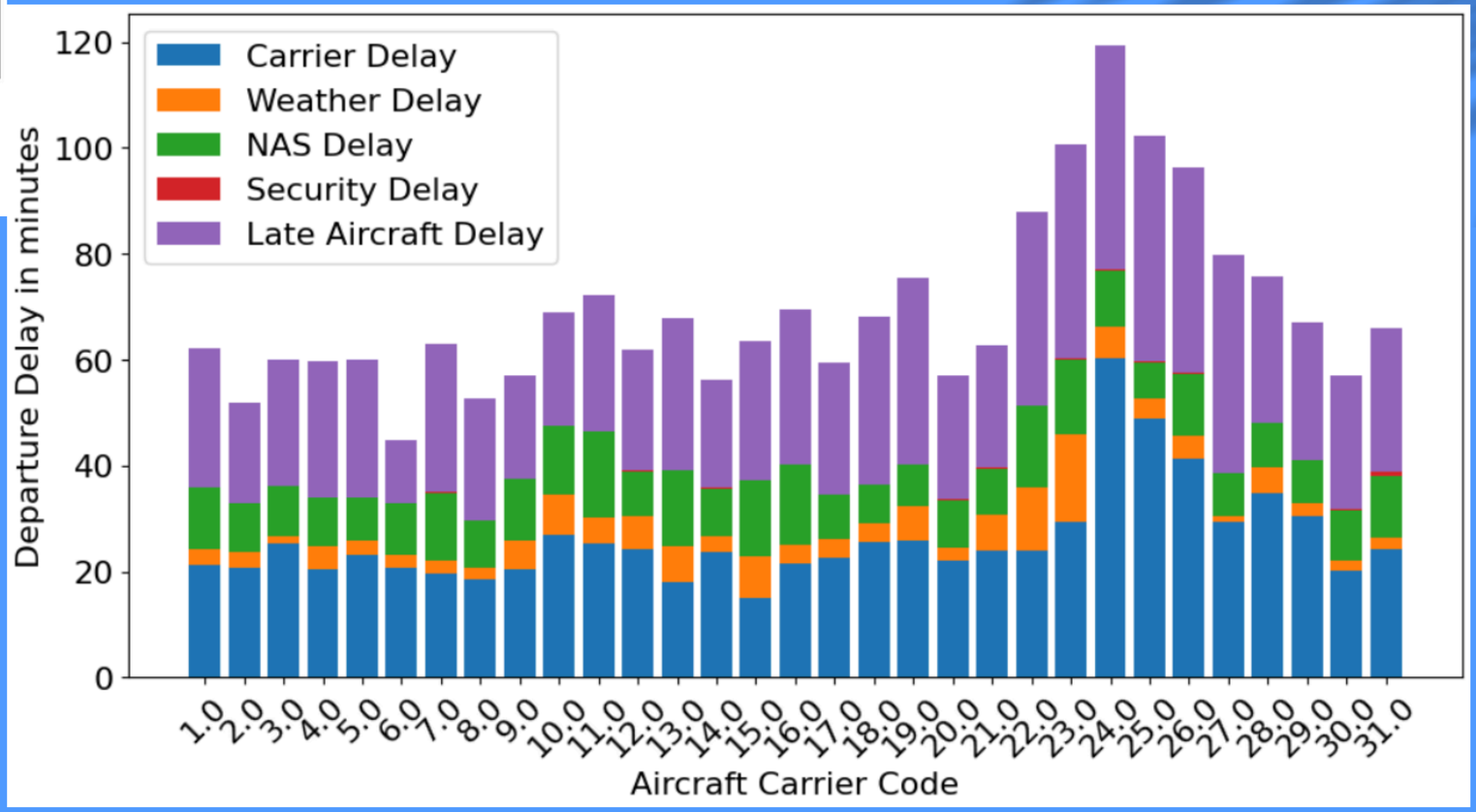
CANCELATIONS DAYS OF DECEMBER





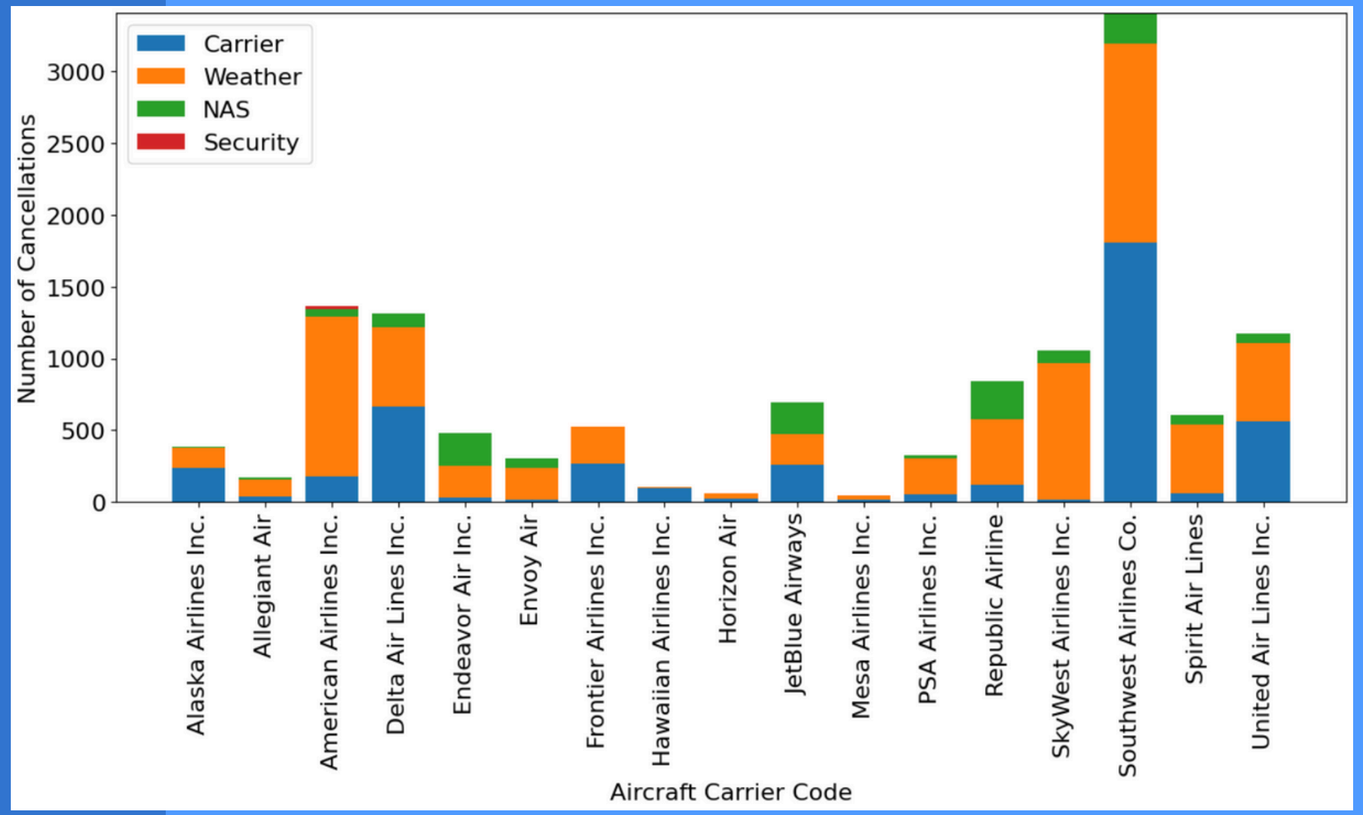
DELAY


Carrier Delay
Weather Delay
NAS Delay
Security Delay
Late Aircraft Delay
Departure Delay in minutes
0
10
20
30
40
50
60
70
Friday
Monday
Saturday
Sunday
Thursday
Tuesday
Wednesday
Aircraft Carrier Code

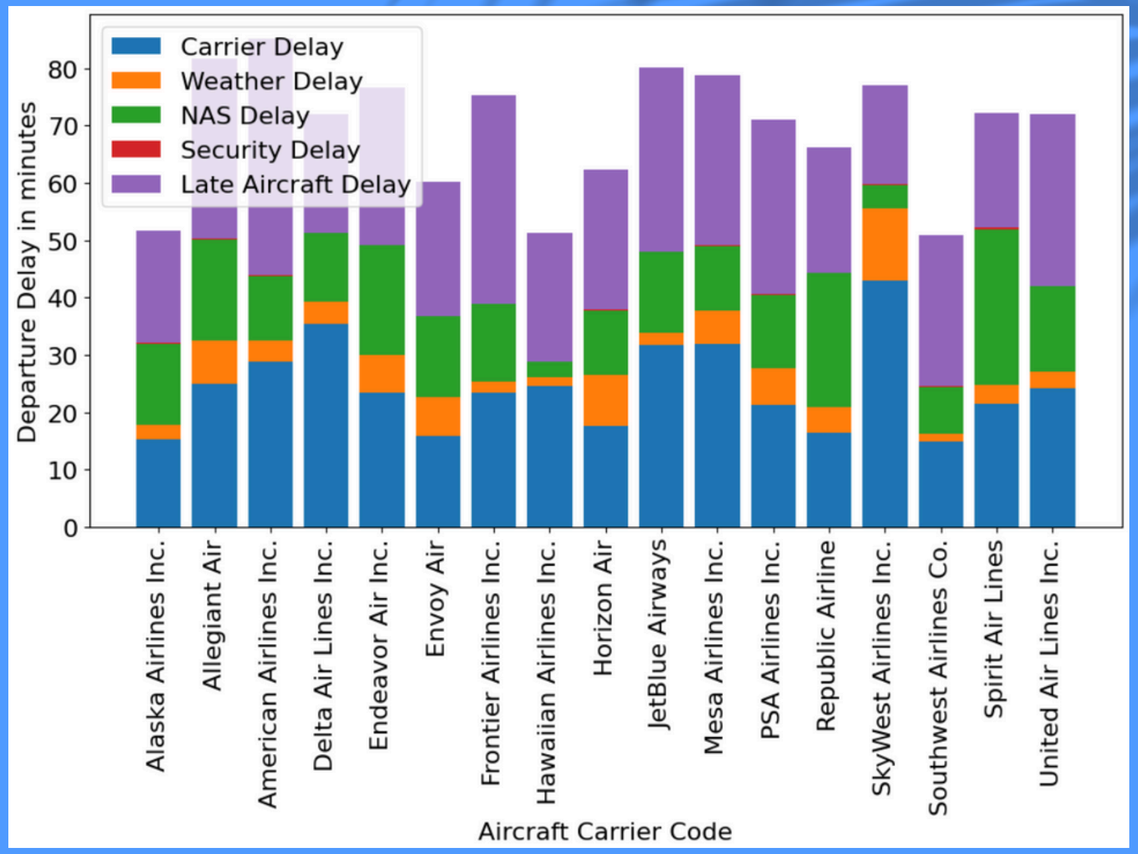

Carrier Delay
Weather Delay
NAS Delay
Security Delay
Late Aircraft Delay
Departure Delay in minutes
0
20
40
60
80
100
120
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
24.0
25.0
26.0
27.0
28.0
29.0
30.0
31.0
Aircraft Carrier Code

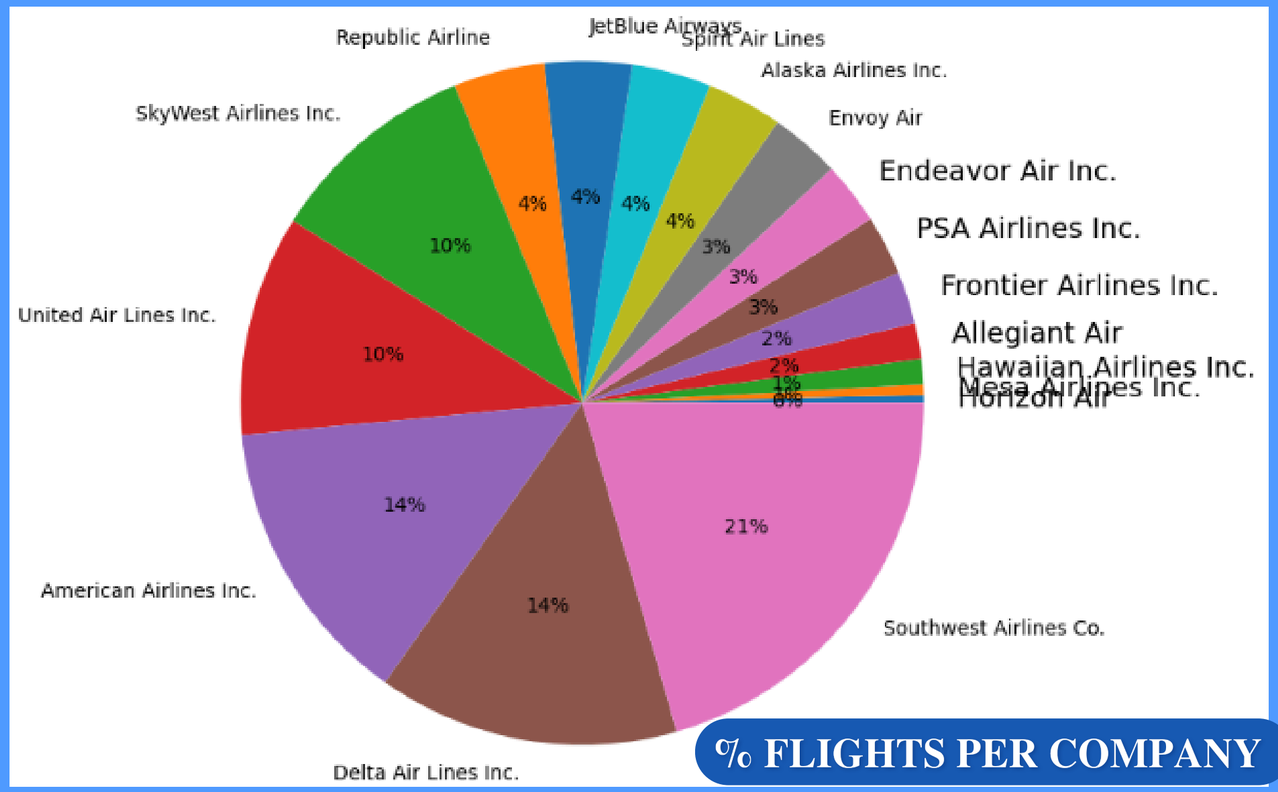
AIRLINE

CANCELATION

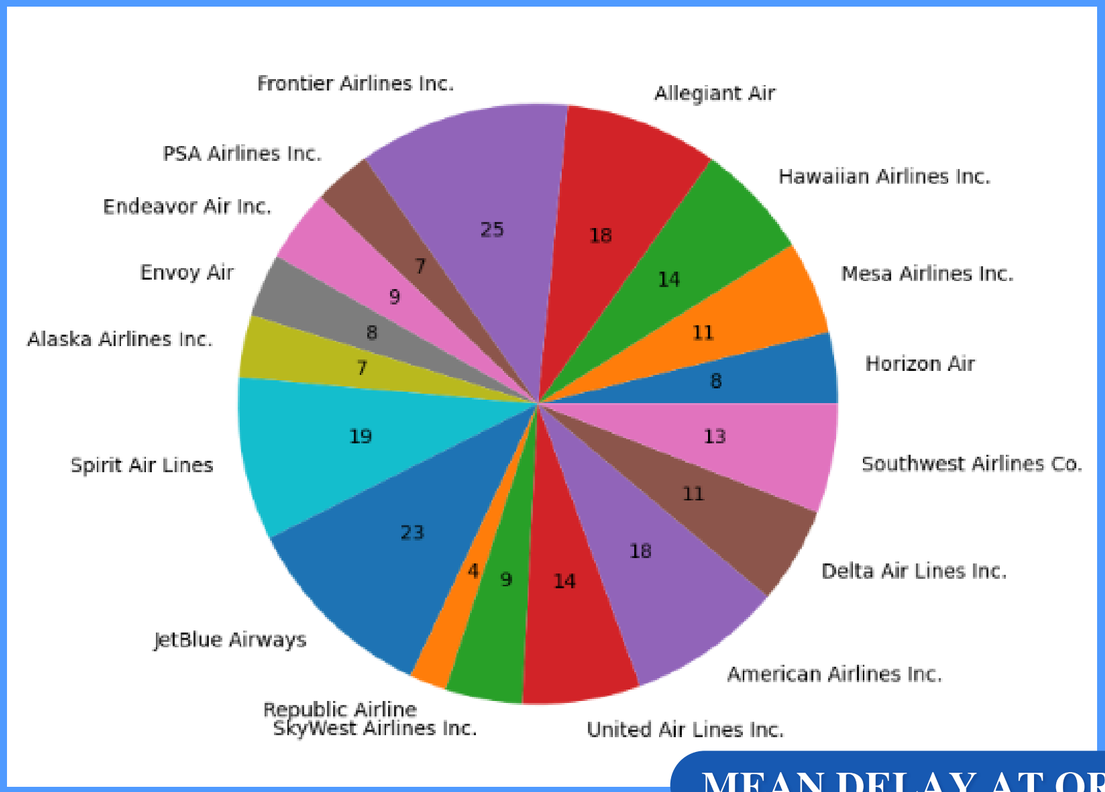


DELAY





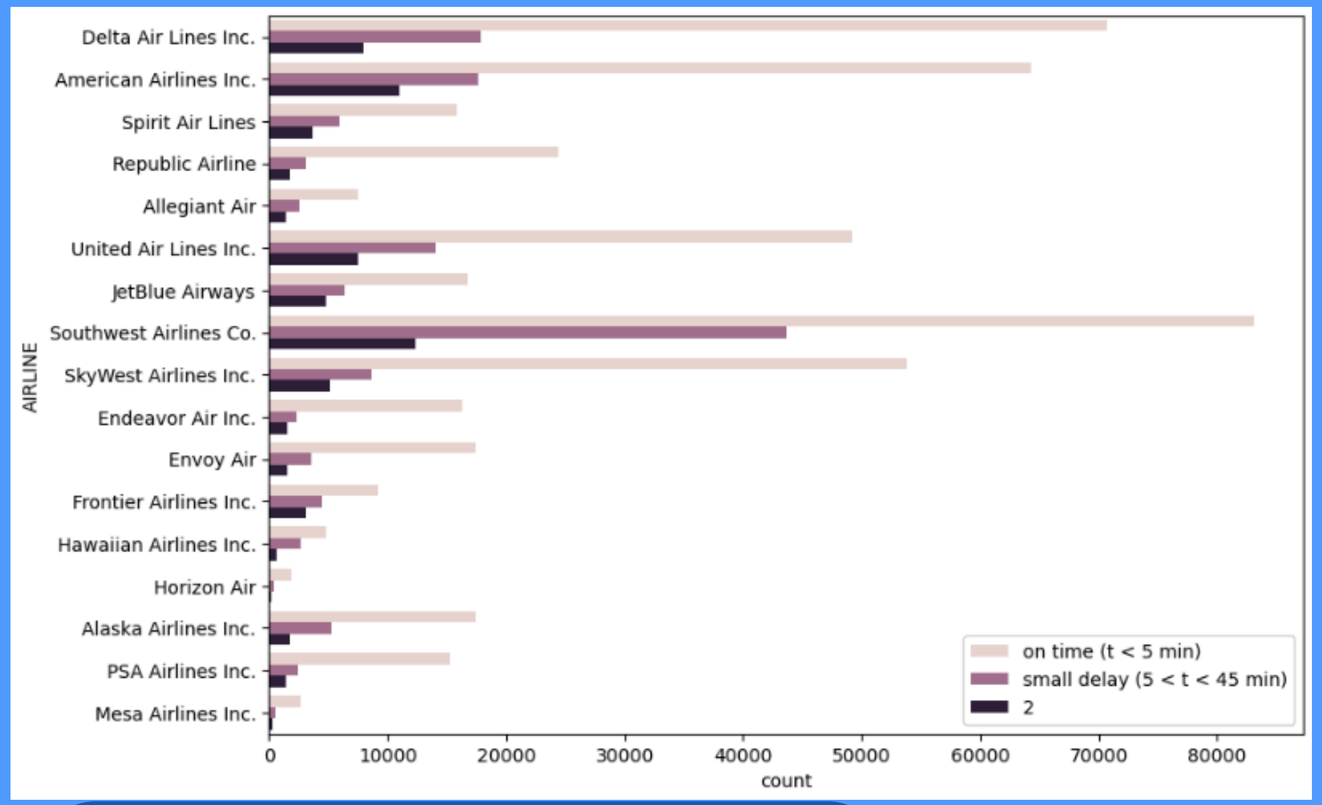
% FLIGHTS PER COMPANY



MEAN DELAY AT ORIGIN

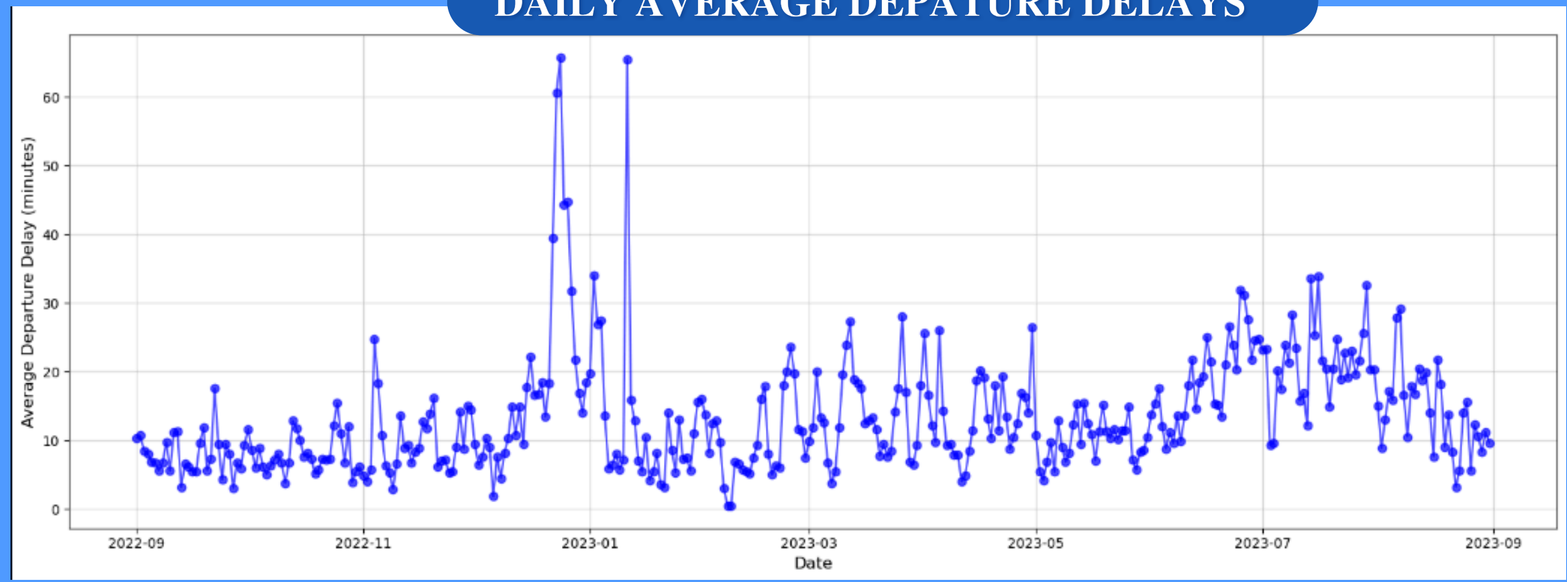
DEPARTURE DELAY

DELAY MAGNITUDE



COUNTING THE NUMBER OF DELAYS AS A FUNCTION OF TIME

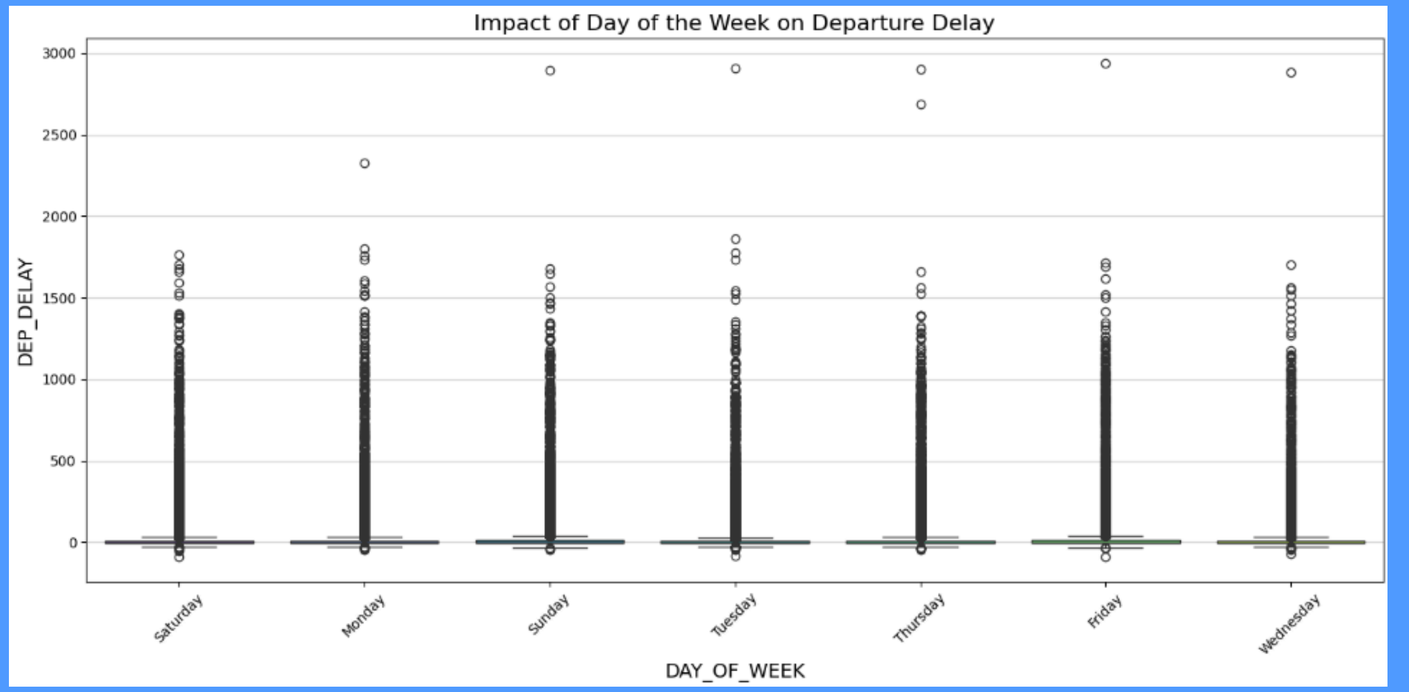
DAILY AVERAGE DEPATURE DELAYS

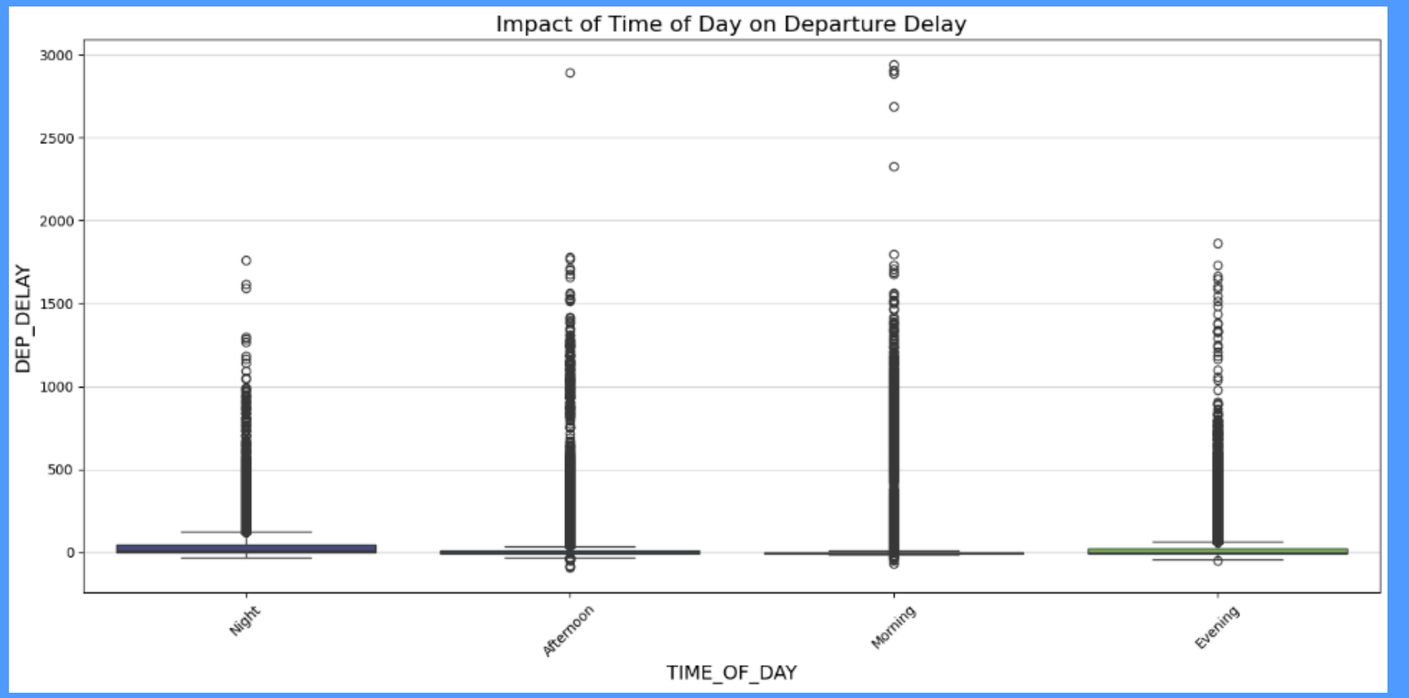


DELAY INFORMATION

The day with the fewest delays is Monday, they increase throughout the week until they reach a peak on Saturdays and begin to decrease on Sundays, with the lowest peak of the week being Mondays.

Most delays happen in the morning, and as the day goes on, they decrease.





FEATURE ENGINEERING

KNN NO NORMALIZED

```
knn.score(X_test_norm, y_test)
✓ 1m 0.2s

0.9082972931929555
```

KNN NORMALIZED

```
knn = KNeighborsRegressor(n_neighbors=10)
knn.fit(X_train_reduced, y_train)

knn.score(X_test_reduced, y_test)
✓ 50.1s

0.9269747148646493
```

MODEL BUILDING

MODEL 1

```
bagging_reg = BaggingRegressor(DecisionTreeRegressor(max_depth=20),
                               n_estimators=100,
                               random_state=42)
```

```
bagging_reg.fit(X_train_norm, y_train)
```

BaggingRegressor

estimator: DecisionTreeRegressor

DecisionTreeRegressor

```
pred = bagging_reg.predict(X_test_norm)
bag_acc = bagging_reg.score(X_test_norm, y_test)

print(f"MAE: { mean_absolute_error(pred, y_test): .2f}")
print(f"RMSE: { mean_squared_error(pred, y_test): .2f}")
print(f"R2 score: { bagging_reg.score(X_test_norm, y_test): .2f}")
```

```
MAE:  7.21
RMSE:  102.97
R2 score:  0.97
```

MODEL 2

```
forest = RandomForestRegressor(n_estimators=100,
                               random_state=42)
```

```
forest.fit(X_train_norm, y_train)
```

RandomForestRegressor

RandomForestRegressor(random_state=42)

```
pred = forest.predict(X_test_norm)
forest_acc = forest.score(X_test_norm, y_test)

print(f"MAE: { mean_absolute_error(pred, y_test): .2f}")
print(f"RMSE: { mean_squared_error(pred, y_test): .2f}")
print(f"R2 score: { forest.score(X_test_norm, y_test): .2f}")
```

```
MAE:  6.94
RMSE:  97.44
R2 score:  0.97
```

```
Bagging Accuracy: 0.9710040023054798
Random Forest Accuracy: 0.972561693125157
```

HYPERPARAMETER TUNING & MODEL OPTIMIZATION

GRID SEARCH

```
Mejor puntuación: 0.9634065124014897
MAE:  7.63
RMSE:  118.67
R2 score:  0.97
```

RANDOM SEARCH

```
Mejor puntuación: 0.9638021297043807
MAE:  7.55
RMSE:  117.13
R2 score:  0.97
```

KEY FINDINGS

- Cancelations and delays in December
- Cancelations and delays because of the airline
- Delays in the morning and fridays

REAL WORLD APPLICATIONS

- Airlines
- Airports
- Frequent travelers

CHALLENGES & LEARNINGS

Study in more detail the reasons for cancellation and delay

Help airlines detect the biggest problems they have and try to solve them

Be prepare to prevent cancelations and delays in special periods

PILAR ESCRIG
VALERIA VALEVASHNIKOVA
ROCÍO ROMO
IVÁN SIMÓN

THANK YOU