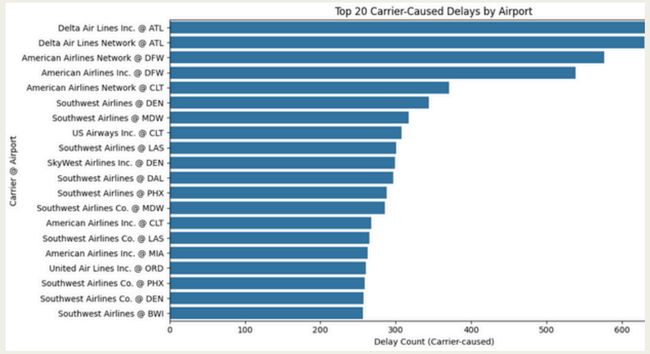
Optimizing Air Travel: Flight Delay Analysis and Prediction

Shourya Aggarwal
B.Tech Industrial Engineering
23119043

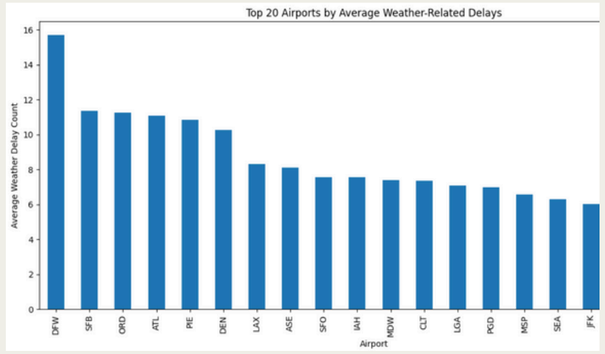
EXECTIVE SUMMARY

- There are about 300-500 NaN values which comprise about 0.3% of the total data.
- Delta Airlines @ATL airport had the most number of delay counts for a single month.
- DFW airport was most impacted by the weather-related delays for a single month.
- DFW airport caused the greatest number of security-related delays in a single month, thereby highlighting strict security protocols.
- Overall seasonality did not have much impact since the weather-related delays across all the months were more or less the same.
- Defined custom features as a part of feature engineering to build a stronger model.

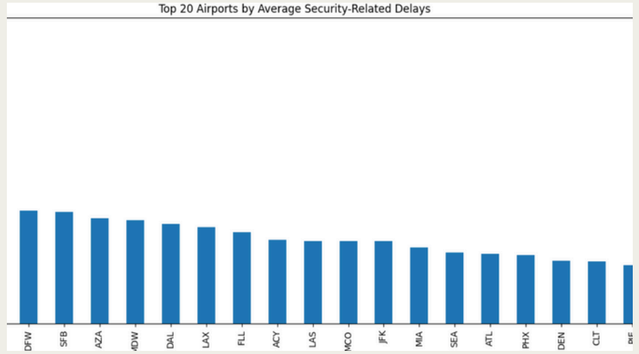
E D A



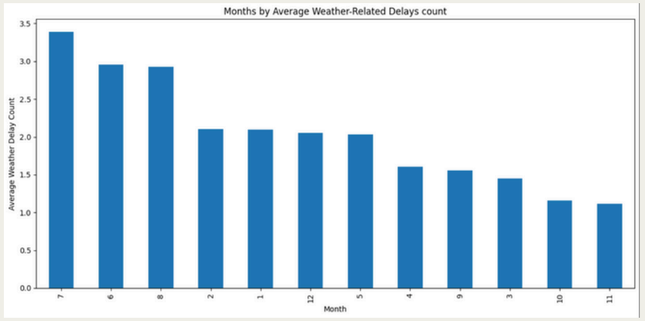
Top 20 carrier caused delays at specific airports



Top 20 Airports with most weather related delays in a months



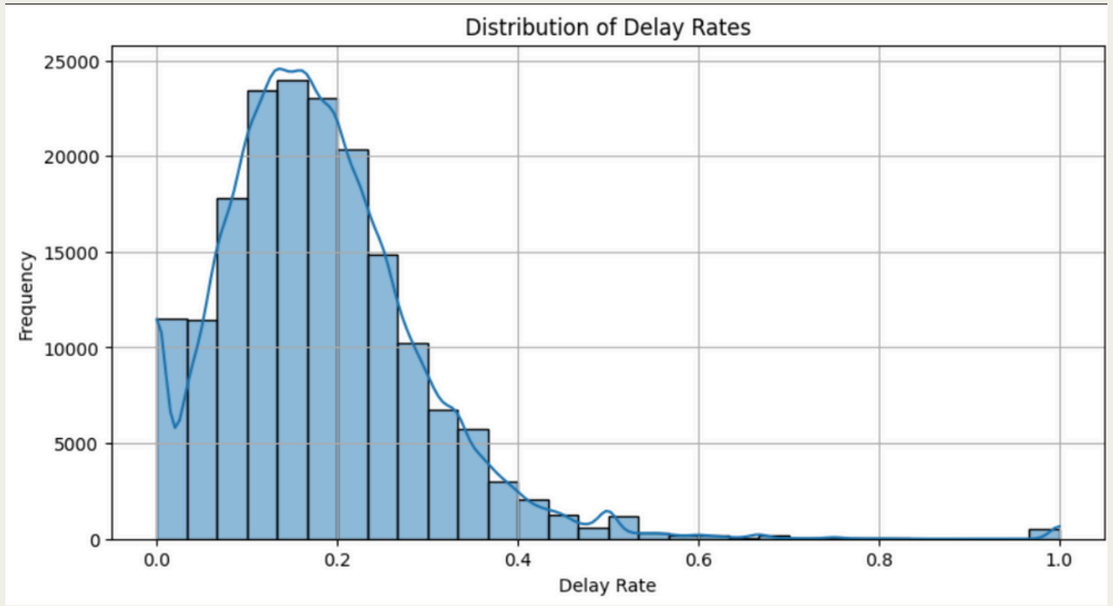
Top 20 Aiprort by security related delay counts

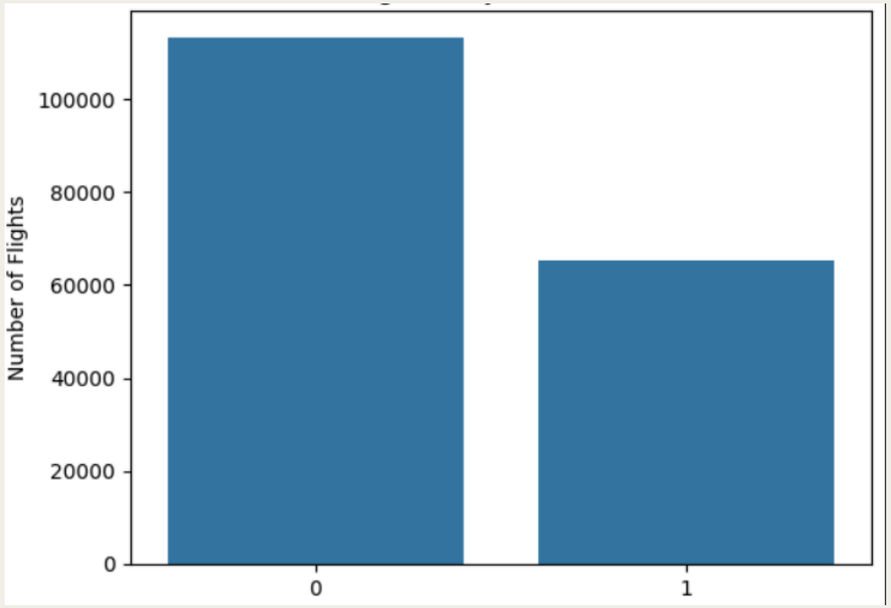


Months by average Weather related delays

The following set of features was custom engineered and added to the dataframe for better classification and regression of the delay.

E D A





To predict the classification of a Delay, a new target variable called delay rate was constructed, which was defined as all the flights delayed more than 15 minutes divided by the total number of flights. The following graph is a skewed Gaussian Distribution, and thus from the graph, a value of 0.2 was selected as a threshold delay rate.

RESULTS

Classification- XGBoost classifer

	PRECISION	RECALL	F1-SCORE	SUPPORT
0	0.87	0.82	0.85	22674
1	0.72	0.80	0.76	13076

Two models, namely RandomForest and XGBoost, were utilized for the classification and regression tasks, in which the XGBoost performance was superior.

Regression- XGBoost Regressor

- Mean Absolute Error: 4.52 minutes

OAI-OPERATIONAL ADJUSTABILITY INDEX

- The Operational Adjustability Index (OAI) is a custom metric designed to prioritize controllable delay components, allowing airlines and airports to focus on delays that can be operationally reduced.

- OAI =
  - 2.0 × Carrier Delay
  - 2.0 × Late Aircraft Delay
  - 1.5 × NAS Delay
  - 1.0 × Weather Delay
  - 0.5 × Security Delay

- Model Performance:
  - MAE: 333.5 minutes
  - R^2 Score: 0.996
  - (Using a Random Forest Regressor)

SHAP values could not be evaluated due to Kernel crash.