Airline

AIRLINE DELAYS

TEAM:

Ashley, Gopi, Kiena, Tracy, Zainab

BOARDING PASS

- **FLIGHT**
  B345
- **GATE**
  D8
- **SEAT**
  29E

Airline

Project Overview

25%- 30%

| DELAYS

Every year, approximately 25-30% of flights are delayed

$28B

| COST

Delays in flights cost passengers and the operations approximately more than $28 billion

Purpose

01 Which airline has the most delayed flights?

02 Which routes have the most delayed occurrences?

03 Will flying time impact delay occurrences?

04 Which date of the week will have the most delay? Weekend vs Weekday

05 Which airport (departure/arrival) is the worst?

Airline

Problem Statement

NOTE:

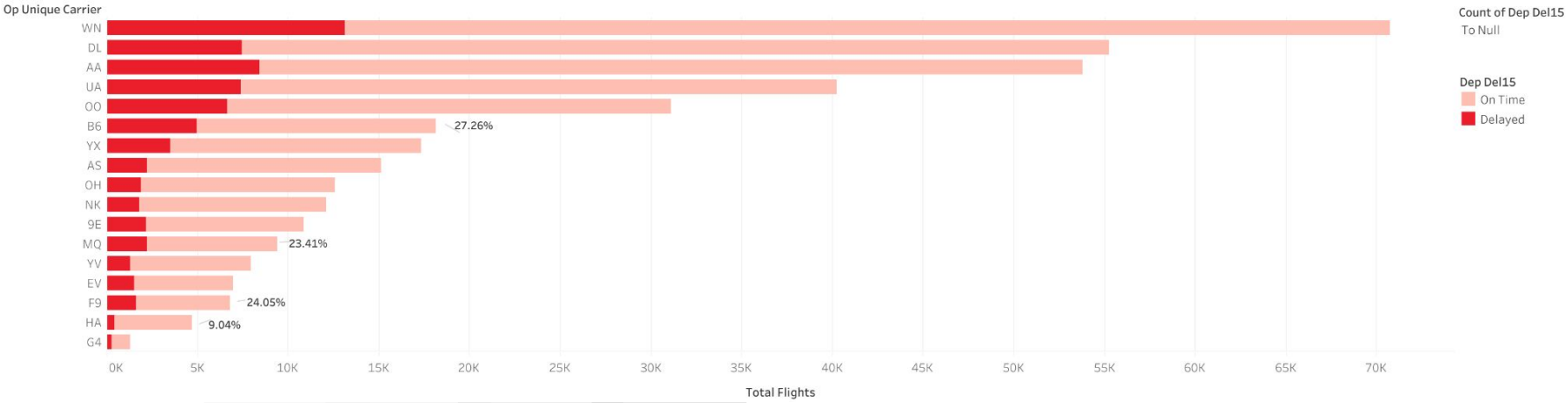
- Predicting Flight Delays
- Identify Key Factors causing Delay

BOARDING PASS

- **FLIGHT**
  B345
- **GATE**
  D8
- **SEAT**
  29E

Airline

Percentage of Delayed Flights by Airline



Airline

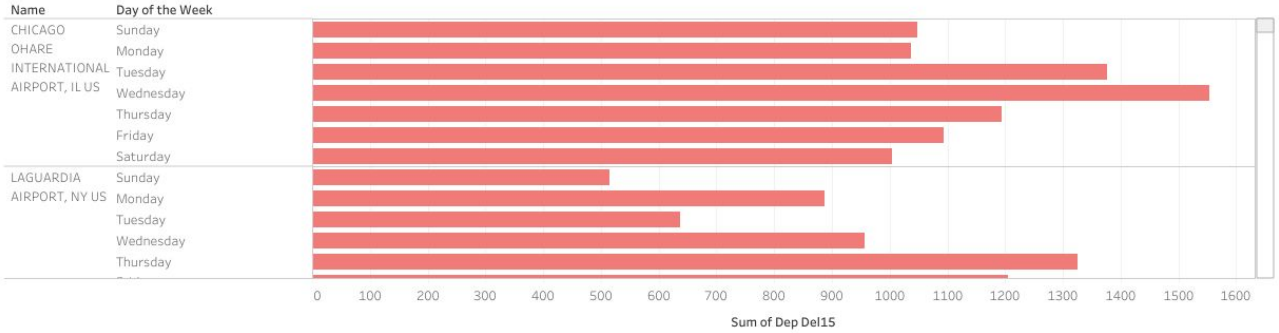
Flight Delays by Day of the Week and Airport

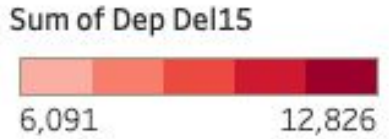
Total Number of Delayed Flights vs Day of the Week

Day of the Week	
Sunday	8,491
Monday	9,271
Tuesday	9,881
Wednesday	11,277
Thursday	12,826
Friday	9,023
Saturday	6,091

Total Delay Occurrences by Airport

Delay Occurrences of each Airport by the Day of the Week



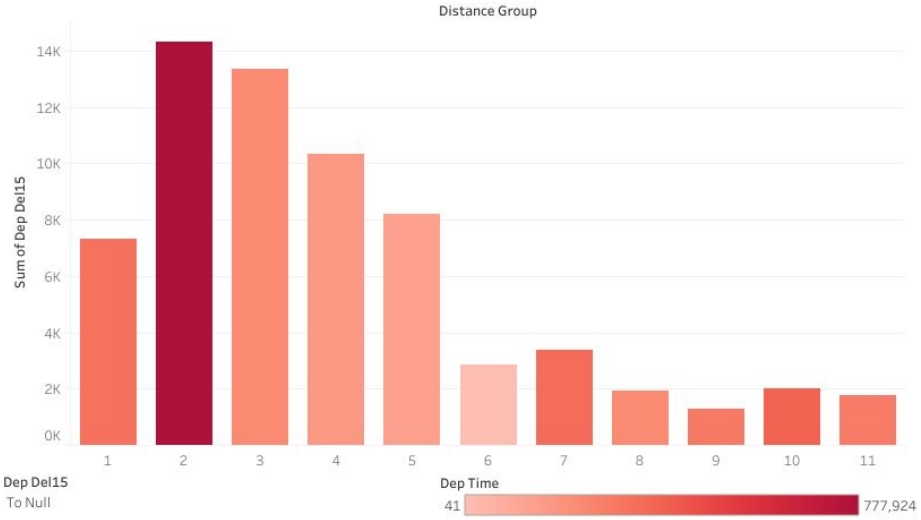


Airline

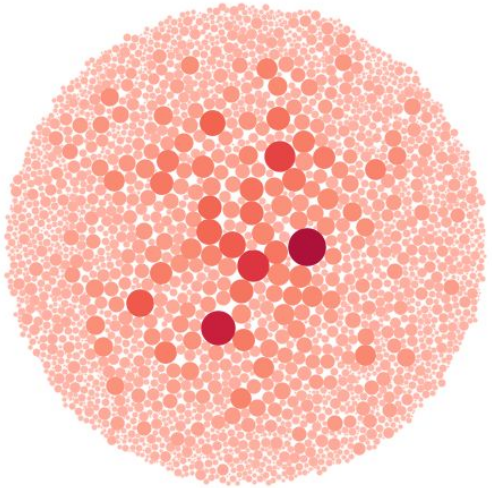
Flight Delays and Route Distances

Dep Time
41 777,924

Delay Occurences vs Distance Group



Routes with the Most Delays



Airline

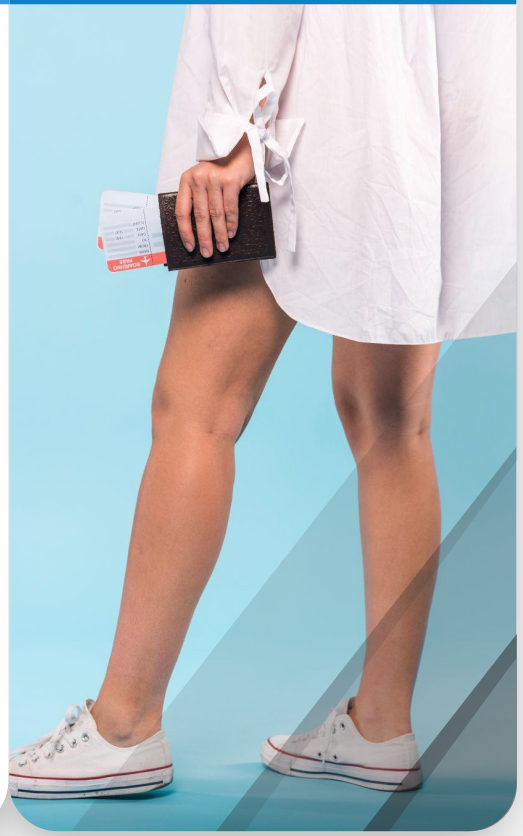
BOARDING PASS

Machine Learning Model

NOTE:

Supervised Learning

- Logistic Regression Model
- Balanced Random Forest Classifier
- Easy Ensemble AdaBoost Classifier



Machine Learning Log Regression (Logistic Regression Model)

In [47]:
```
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, y_pred))
```

```
0.8239913320531662
```

In [49]:
```
# Print the imbalanced classification report
print(classification_report_imbalanced(y_test, y_pred))
```

	pre	rec	spe	f1	geo	iba	sup
0	0.83	1.00	0.03	0.90	0.16	0.03	76582
1	0.67	0.03	1.00	0.05	0.16	0.02	16635
avg / total	0.80	0.82	0.20	0.75	0.16	0.03	93217

Constraints of the current model and planning for the future

CONSTRAINTS

Current Data set solely deals with the month of January. Given more time we would have included the full 12 months in the year.

PROPOSAL FOR THE FUTURE

1. Trying new datasets to include other routings, including international flights.
2. Test the Machine Learning model with other datasets.