

FLIGHT DELAY PREDICTION

Machine Learning - Python

01

INTRODUCTION

02

OBJECTIVE

03

PROBLEM STATEMENT

04

**DATA EXPLORATION AND
VARIABLES SELECTION**

05

COMPARISON OF MODEL ACCURACY

06

CONCLUSION

AGENDA



INTRODUCTION

- Flight delays cause airline carriers millions of dollars and inconveniences passengers, resulting in a loss of demand
- Airline companies add additional flights during the holiday season or before starting of any festival for better service.
- The flight status depends on some factors like weather conditions, Destination, origin, distance etc.
- Compare Classification and Regression models to predict flight status, avoid delays and ensure smooth running of flight operations

OBJECTIVES

- Evaluate the dependent variables.
- Predict the flight status as accurately as possible using Naïve Bayes (NB) model, Classification and Regression Tree (CART) model, logistic Regression model.
- Model training based on historical data.
- Select the best model by comparing the results.

PROBLEM STATEMENT

- Flight delay is a major issue in the aviation industry.
- Delays in air travel can be very costly to both passengers and airlines.
- It could make them late to their booked occasions or miss a corresponding flight.
- Passengers may not generally be entitled for a refund when a postponement happens.

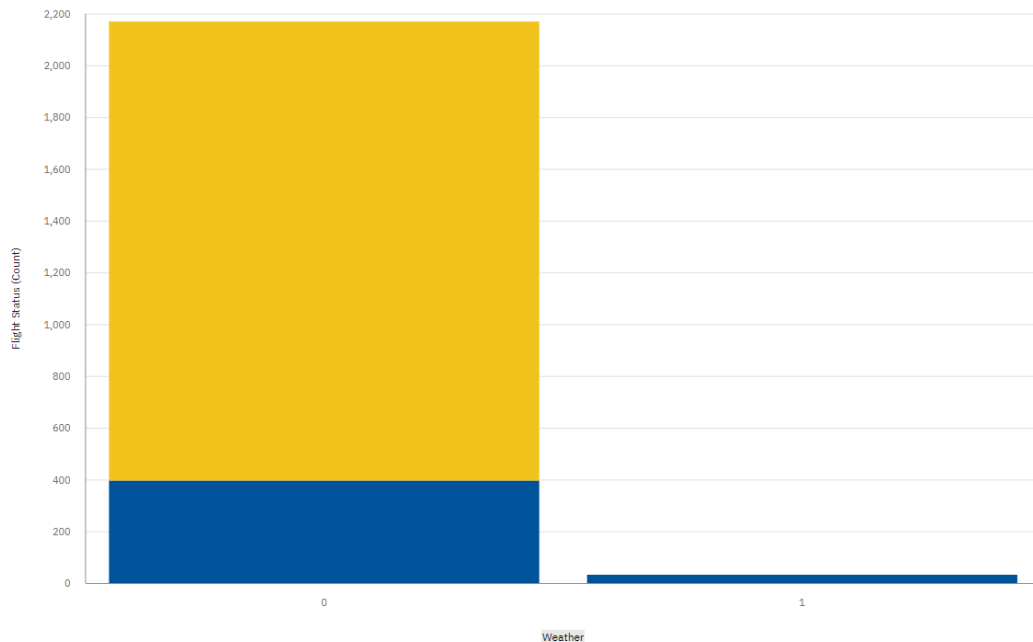


DATA EXPLORATION

FLIGHT STATUS BASED ON WEATHER CONDITIONS

Flight Status by Weather

Flight Status
● delayed ● ontime



 **NORMAL WEATHER**

 **BAD WEATHER**

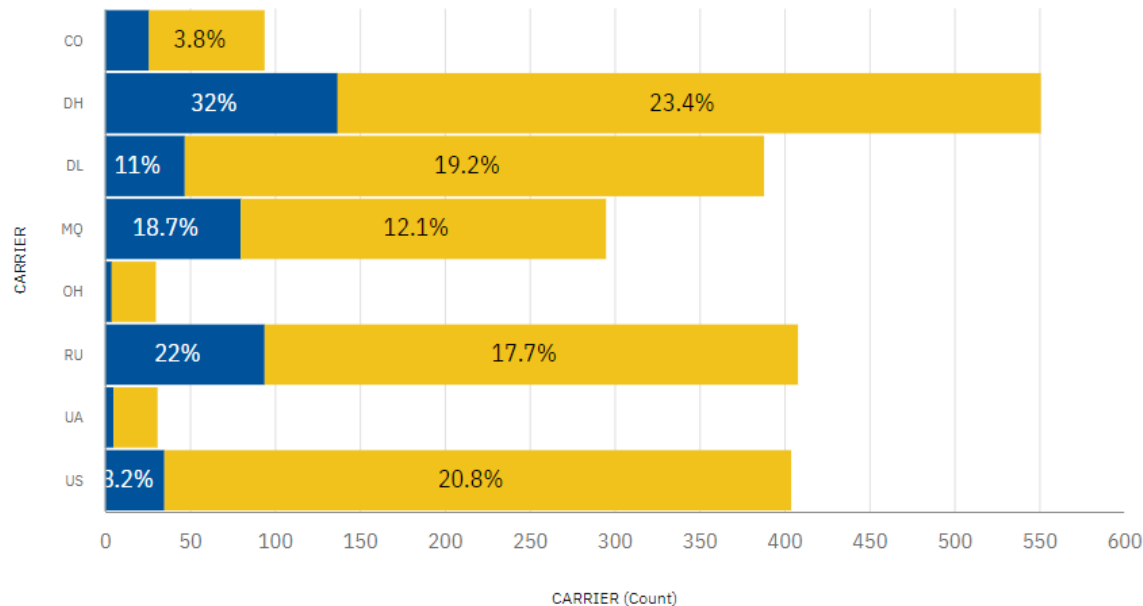
There is always a delay in
flight when weather
conditions are not normal.

FLIGHT DELAY FOR CARRIERS

CARRIER by Flight Status

Flight Status

● delayed ● ontime



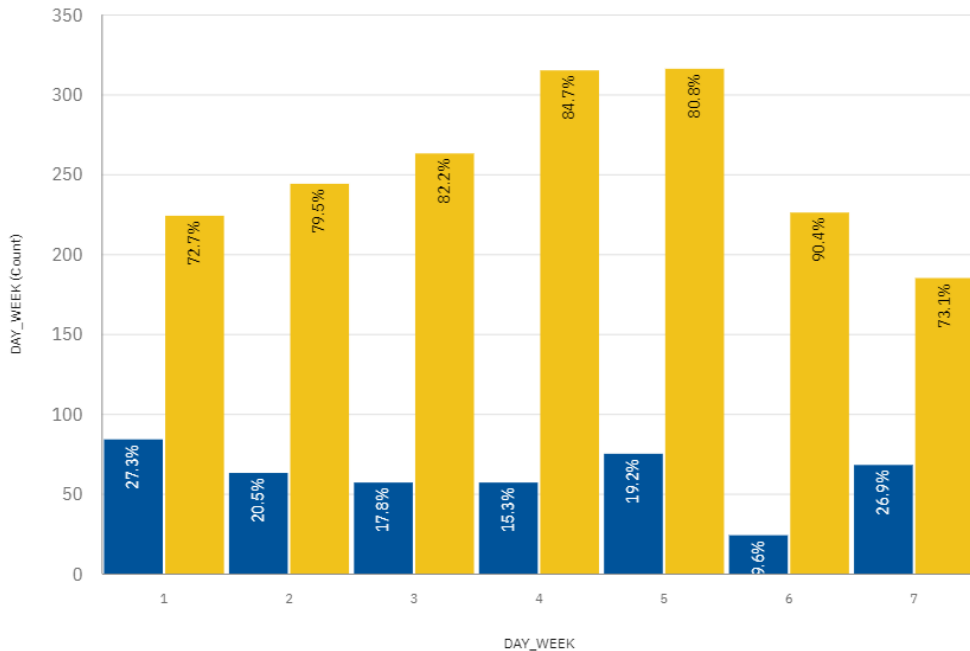
Some carriers are prone to having more delays than the others.

TREND ANALYSIS FOR DAY OF THE WEEK

Flight Status by Day of the Week

Flight Status

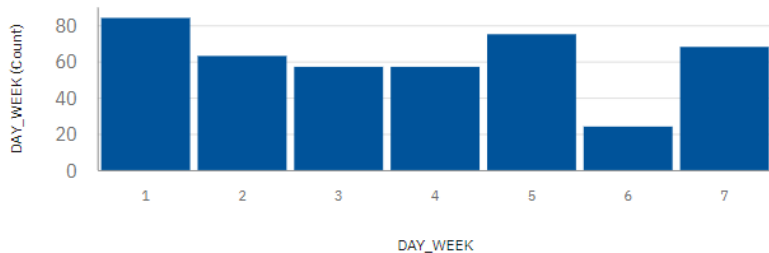
● delayed ● ontime



Flight Delay by Day of the Week

Flight Status

● delayed



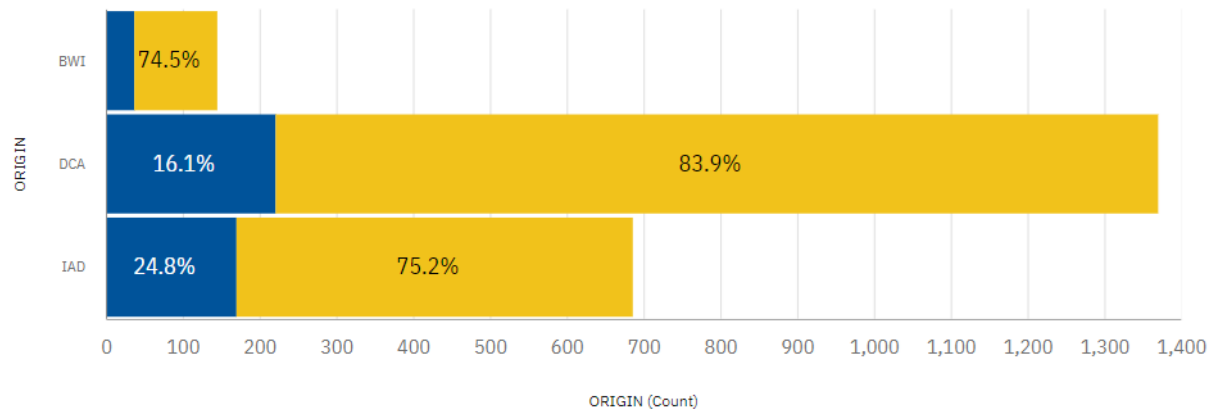
There is no trend in on-time delay ratio for day of the week.

ANALYSIS BASED ON ORIGIN

Flight Status by Origin

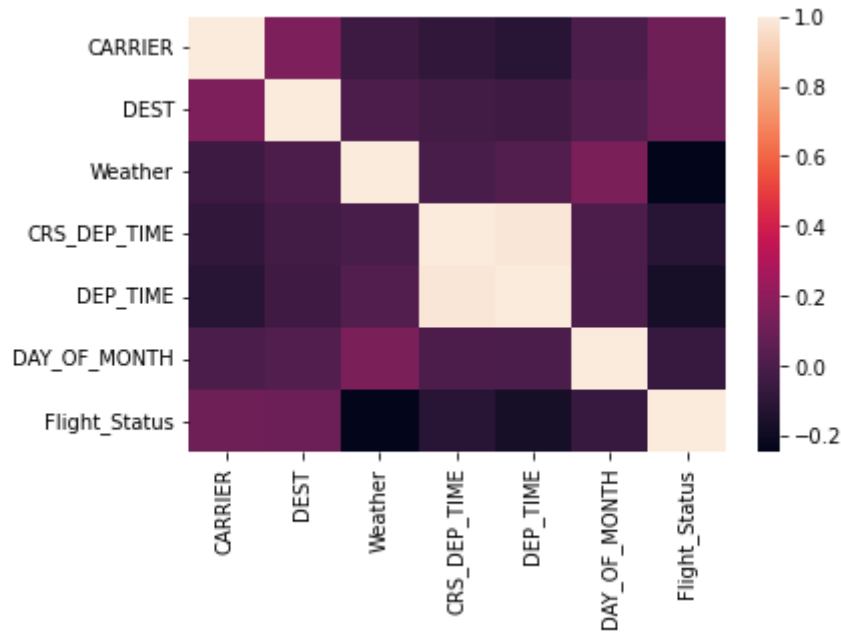
Flight Status

● delayed ● ontime



There is not much difference in percentage count for delayed or ontime flights based on origin.

VARIABLE SELECTION



Correlation among the variables selected based on the analysis and trend depiction.

COMPARISON OF MODEL ACCURACY

| Naïve Bayes model | Cart algorithm | Logistic Regression Model |
|-------------------|----------------------|---------------------------|
| 88.83% | 87.30% | 89.05% |
| Generative model | Discriminative model | Discriminative model |

- Logistic Regression is observed to yield the highest accuracy in the prediction of flight delays

CONCLUSION

- Logistic Regression is the most appropriate model to predict flight delays
- This model will alert carriers of probable flight delays
- As next steps, further analysis can be done to predict various categories of delays:
 - Weather related
 - Maintenance related
 - Shortage of crew
 - Air traffic related



THANK YOU

