

# Phase 2 Report

## Flight Delay Prediction Using Machine Learning

### 1. Introduction

This project explores how weather conditions impact flight delays at JFK Airport during January 2023. By integrating flight data with historical weather records, patterns and correlations were analyzed to support predictive insights.

### 2. Data Collection

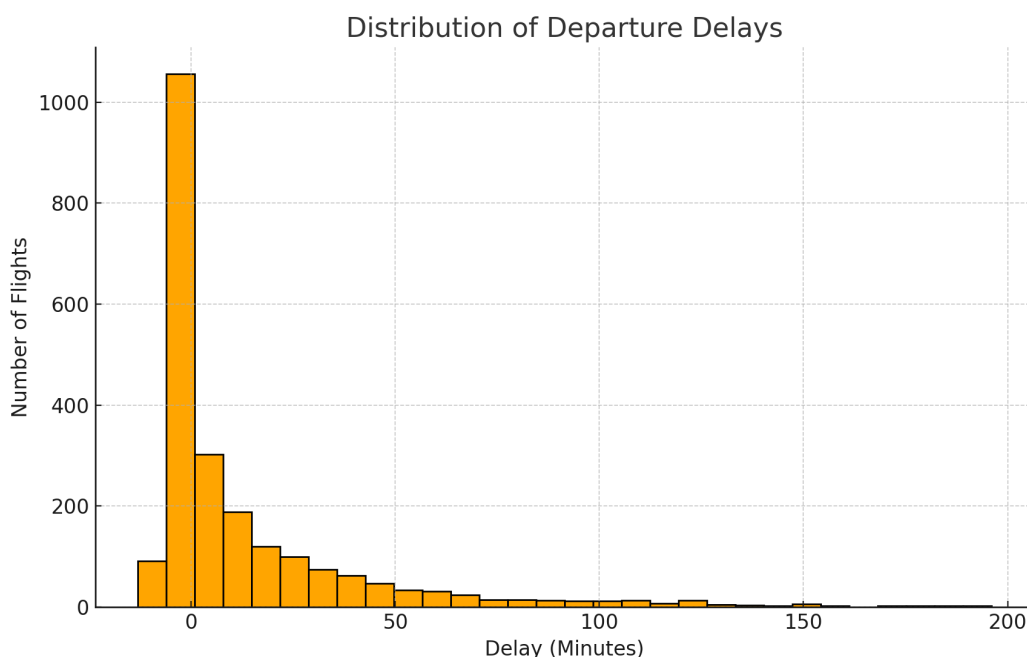
- Flight Data: 2,294 Delta Airlines departures from BTS.
- Weather Data: 31 daily records from NOAA.

Datasets were merged by date to associate each flight with corresponding weather conditions.

### 3. Exploratory Data Analysis (EDA)

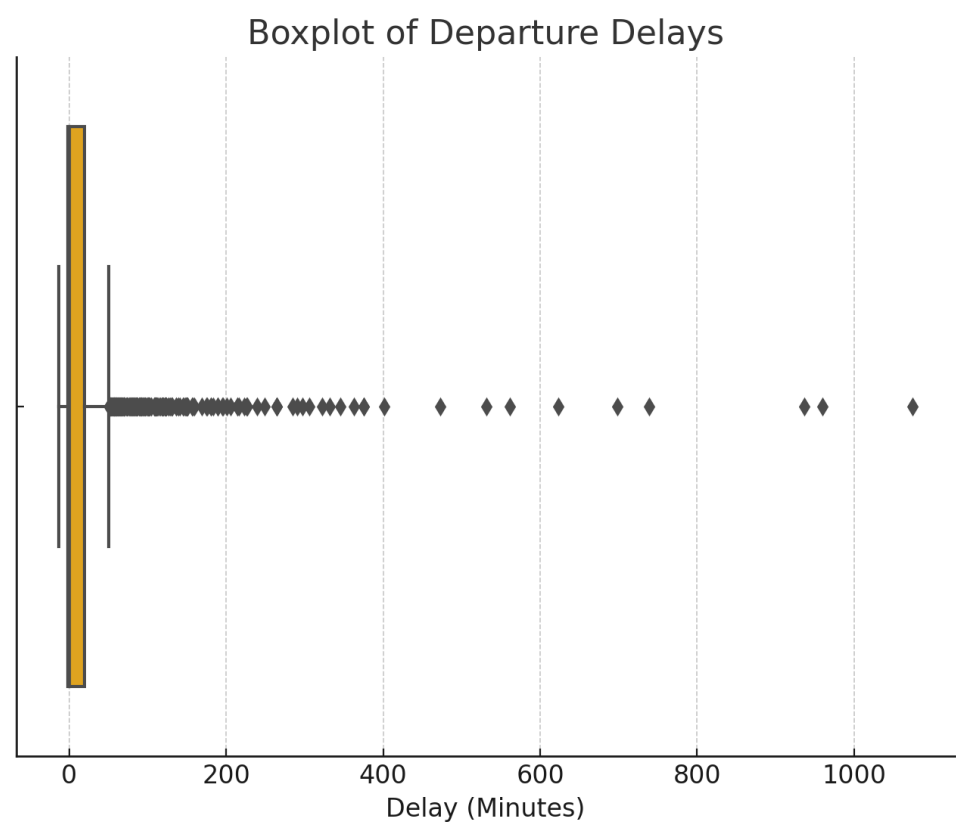
Key patterns were identified regarding delay distributions and weather conditions. Below are selected visualizations:

#### 3.1 Distribution of Departure Delays



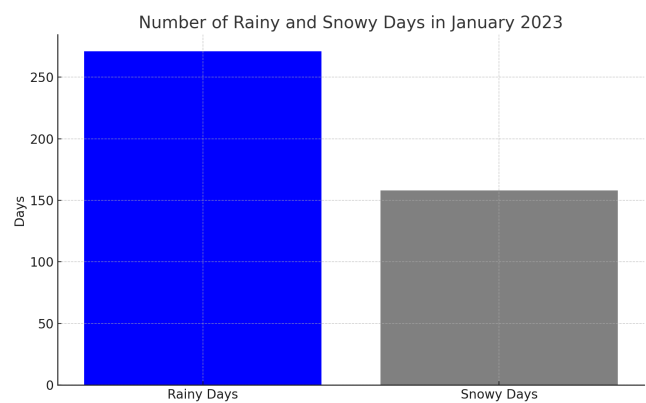
Most delays are minor, with a few extreme cases exceeding 100 minutes.

3.2 Boxplot of Departure Delays



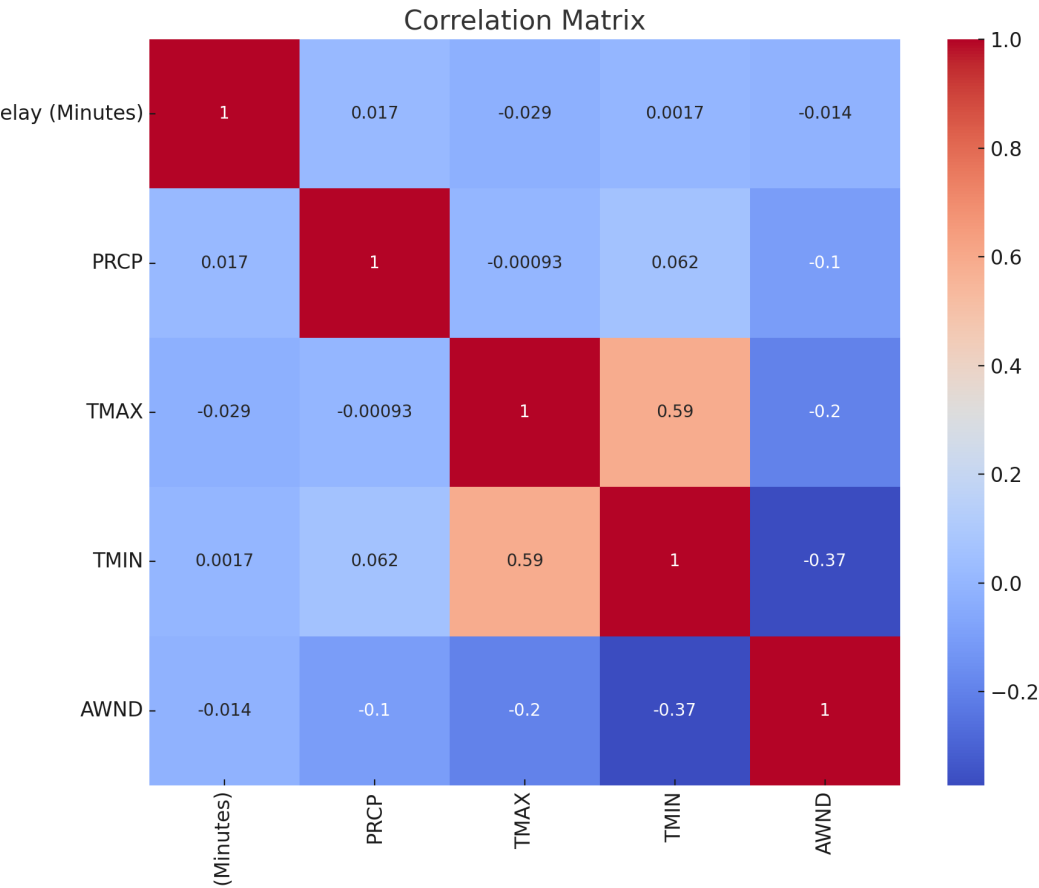
Outliers indicate occasional severe delays, though most flights remain within typical ranges.

3.3 Rainy and Snowy Days



Weather data shows several rainy days and fewer snowy days impacting operations.

3.4 Correlation Matrix



Weak correlations observed between weather factors and delays, suggesting limited linear influence.

-----

## 4. Hypothesis Testing

Objective: Assess if rainy days significantly impact departure delays.

### Defined Hypotheses:

- Null Hypothesis ( $H_0$ ): No significant difference in delays between rainy and clear days.
- Alternative Hypothesis ( $H_1$ ): Delays are longer on rainy days.
- Significance Level:  $\alpha = 0.05$
- Tests Used:
  - \* Two-Sample T-Test
  - \* Pearson Correlation
  - \* Spearman Correlation

### Results:

T-Test p-value: 0.838 -> Fail to reject  $H_0$

Pearson r: 0.017,  $p = 0.418$  -> No linear correlation

Spearman rho: 0.054,  $p = 0.0097$  -> Weak but significant monotonic correlation

Conclusion: Rain has a minor non-linear impact on delays.

-----

## 5. Conclusion

Weather conditions, particularly precipitation, show a weak influence on flight delays. Future work should expand data scope and apply predictive models.