Group 14 Ice Cream - STAT 167 Final Project Report

Lydia Niu, Alexis Castaneda, Aparna Petluri, Gracelynne Mohan, Jenny Zhang, Zoe Shum

2025-06-05

Introduction

Handling the heavy air traffic above New York, the John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA), and Newark Liberty International Airport (EWR) manage immense flight operations throughout the year. Given a dataset regarding the dynamic flight patterns between these three airports, we hope to explore the various factors that may affect flight delays and flight volume - amount of flights.

The overall question we want to answer is: "What factors influence flight volume and does this affect delay patterns across New York City's major airports (JFK, LGA, and EWR)?" We hypothesize that weather conditions, such as temperature or wind speed, play a significant role in the occurrence of delays among the three airports. By analyzing the nycflights13 dataset, we aim to provide valuable insight to travelers, as well as airline and airport administrators, as to flight frequency and delays in the New York metropolitan area.

Coherent Questions

To help answer our main question, we aim to answer the following:

1. Which months and seasons experience the highest flight volumes at each airport?
2. How do average delays vary by month and season?
3. Are delays more frequent/severe during specific weather conditions?
4. Are there significant differences in average delays/ flight volume across the 3 airports and across different seasons?
5. What relationship, if any, exists between busy days (high flight volume), weather conditions (temperature and wind speed) and flight delays?
6. Which airport is most affected by flight delays due to weather conditions among JFK, LGA, and EWR?

Data Description

For this project, we will be focusing our analysis on the flights, airports, and weather datasets found in the nycflights13 tidyverse package.

Flights: all flights that departed from NYC in 2013. Size: 336,776 x 19

Weather: hourly meteorological data for each airport. Size: 26,115 x 15

As part of our data preprocessing, we removed all canceled flights and filtered out any entries with missing values related to weather variables. We also excluded data from December 31, which was missing wind speed and temperature information. After cleaning the data, we created new data frames containing calculated statistics to support a more structured analysis. These steps allowed us to explore the correlations between environmental factors and operational efficiency at New York City's airports.

Visualization