**DATA 5300: An Analysis of United Airlines Departure Delays in 2013**

**Team Members**

## Dan Nguyen, Jennifer Poling, Paul Skentzos

**Introduction**

For an airline, efficiency and customer satisfaction are two sides of the same coin. A key factor in both is on-time performance. When a flight is delayed, it can cause a ripple effect, frustrating passengers, and costing the airline resources.

This report investigates departure delays for United Airlines (carrier code UA) flights from the three major New York City area airports (Newark, JFK, and LaGuardia) in 2013. Our goal is to understand what factors contribute to delays (temperature, wind speed, precipitation, visibility, etc.). By identifying patterns in these delays, United Airlines can be better prepared to manage operations and improve the travel experience for its customers.

Using a public dataset of all 2013 NYC flights, we analyzed how departure delays are related to:

* Time of day
* Time of year (season)
* Temperature
* Wind speed
* Precipitation (rain, snow, etc.)
* Visibility

**Data Methodology**

To conduct this analysis, we used two datasets from the nycflights13 package (this data is part of the R Studio system and is easily loaded as such: library(nycflights13). One data set contains all 336,776 flights from 2013 and the another contains detailed hourly weather data for the same period (2013).

Our preparation involved several steps to make the data easy to analyze and understand:

1. **Focus on United:** We filtered the main dataset to look at only the 57,979 flights operated by United Airlines.
2. **Define "Late":** We created two simple categories to better understand delays. We marked a flight as "late" if it departed more than 0 minutes after its scheduled time, and "very late" if it departed more than 30 minutes after its scheduled time.
3. **Merge Weather Data:** We combined the flight records with the hourly weather reports, matching them by the airport and the time of the scheduled departure.
4. **Create Categories:** To make analysis easier, we grouped some data into new, clear categories:
   * **Season:** We grouped months into "Winter" (Dec, Jan, Feb), "Spring" (Mar, Apr, May), "Summer" (Jun, Jul, Aug), and "Fall" (Sep, Oct, Nov).
   * **Time of Day:** We grouped scheduled departure hours into "Early" (5-7 am), "Morning" (8-10 am), "Midday" (11 am-1 pm), "Afternoon" (2-4 pm), "Evening" (5-7 pm), and "Night" (8-11 pm).
   * **Visibility:** We categorized hourly visibility as "Poor" (less than 3 miles), "Fair" (3 to 5 miles), or "Good" (more than 5 miles).
   * **Precipitation:** We created a simple "Yes/No" category for whether any precipitation was recorded in an hour.
   * **Temperature:** We created bins to compare temperature ranges.
   * **Wind Speed:** We created bins to compare wind speed ranges.
   * **Wind Gusts:** We created a simple “Yes/No” category for whether there was a wind gust at departure.

**Results**

We explored each factor using visualizations and statistical tests (specifically, permutation tests) to see if the patterns we observed were statistically significant, meaning they are very unlikely to be due to random chance.

**Time of Day and Time of Year**

Our initial exploratory chart, which shows departure delays by time of day and season, reveals a clear and consistent pattern.

* **Time of Day:** Delays are lowest in the "Early" morning hours and steadily increase as the day goes on, peaking in the "Evening" and "Night." This suggests that small delays early in the day may and often do cascade and compound, leading to much larger delays by the evening.
* **Time of Year:** The "Summer" months (June, July, August) consistently show longer delays across all times of day compared to other seasons. This is likely due to a combination of higher passenger volume and more frequent weather-related disruptions (like thunderstorms).

A graph of different seasons

AI-generated content may be incorrect.

Figure Departure Delay by Time of Day

When we ran a permutation test on the "late" rate by time of day, we found a statistically significant difference (p = 0.0001). This appears to support the idea that flights delays increase as the we get later in the day.

**Precipitation**

Not surprisingly, weather plays a role. We compared the "late rate" (the percentage of flights that were late or greater than 30 minutes late) on days with any precipitation to days with no precipitation.

The presence of rain, snow, or hail significantly increases the likelihood of a flight being delayed (p = 0.0001). This was indicated by our p value showing that there was a statistically significant difference, showing that weather does factor into flight delays.

A graph of a comparison of a normal and a normal condition

AI-generated content may be incorrect.

Figure : Precipitation Effect on Delays

A graph showing a number of different colored squares

AI-generated content may be incorrect.

Figure : Proportions of Late Departures by Precipitations

**Visibility**

We conducted a similar test on visibility, comparing the late rate for "Poor," "Fair," and "Good" visibility conditions.

* **Finding:** Visibility has a clear, significant impact (p = 0.0001). The worse the visibility, the higher the chance of a delay:
  + **Poor Visibility (<3 miles):** 57.9% of flights were late.
  + **Fair Visibility (3-5 miles):** 54.7% of flights were late.
  + **Good Visibility (>5 miles):** 46.2% of flights were late.

A graph showing different colored squares

AI-generated content may be incorrect.

Figure : Proportion of Late Departures by Visibility

**Temperature and Wind Speed**

We also analyzed the effects of temperature and wind speed.

* **Finding:** While our statistical tests did find a detectable relationship for both temperature and wind speed (p = 0.0001), the actual effect was very modest. Compared to the time of day, season, and precipitation, these factors appear to be much less important drivers of delays.

A graph of a temperature

AI-generated content may be incorrect.

Figure : Temperature Delays

A graph of a wind speed

AI-generated content may be incorrect.

Figure : Wind Speed Delays

**Discussion**

Based on our analysis of 2013 data, the most powerful predictors of departure delays for United Airlines in the NYC area are time of day and time of year. Delays are significantly more likely in the evening and during the summer. Bad weather, specifically precipitation (including rain, snow, etc.) and poor visibility, also reliably increases the rate of delays.

Our study shows that while factors like temperature and wind speed are statistically related to delays, their impact is small. For improving efficiency, efforts should be focused on the bigger factors. While temperature and wind speed have effects on aircraft, modern aircraft are designed to handle wide ranges of temperatures and wind speeds. For example, these delays would come into play during approaches as air traffic controllers appropriately space out aircraft to account for wind events.

For future work, it would be valuable to investigate the cause of the compounding evening delays. Are they due to the network-wide effects of "knock-on" delays from earlier in the day (both might be factors, but do “knock-on” delays effect the customer experience more)? Could strategic scheduling adjustments or resource allocation during peak summer evenings help reduce this congestion? Answering these questions could provide actionable insights for United to improve on-time performance and customer satisfaction.