Analyzing Departure Delays for United Airlines (UA)

Report on

Analyzing Departure Delays for United Airlines (UA)

INTRODUCTION

In the ever-evolving airline industry, optimizing flight operations and enhancing customer satisfaction are paramount. Optimizing flight operations and increasing customer satisfaction are critical in the ever-changing airline sector. Improving departure timeliness is a critical component of attaining these goals for United Airlines (UA), carrier code UA. Delays in departure can have a substantial impact on both operational efficiency and the overall travel experience of passengers. This paper uses exploratory data analysis (EDA) and permutation testing to analyze the relationship between departure delays and various influencing factors in order to acquire a better understanding of the factors that contribute to these delays.

Our analysis focuses on the impact of certain factors on UA's departure delays. These factors are as follows:

Time of Day: We analyze in the event departure delays are affected by the time of day, such as morning, afternoon, evening, and night. Are there certain times of day when delays are more common or longer.

Time of Year: Recognizing the impact of changes in the seasons is critical. We divide the months into four seasons—winter, spring, summer, and fall—to see how the changing seasons affect departure times.

Temperature: To observe generally the connection between temperature and departure delays. We can detect probable weather-related patterns that affect flight schedules by analyzing this given dataset of nycflights13: weather.

Wind Speed: Wind speed can influence aircraft operations. We analyze if higher wind speeds are associated with earlier departure delays for UA flights.

Precipitation: We investigate whether the presence of precipitation, such as rain or snow, is associated with flight delays during departure.

Visibility: We examine whether adverse visibility conditions are linked to delays in flight departures, considering the importance of good visibility for safe flight operations.

Our comprehensive analysis aims to uncover patterns and relationships between departure delays and these influencing factors. The insights derived from this analysis can serve as a foundation for data-informed decision-making, enabling United Airlines to optimize its flight schedules, improve operational efficiency, and enhance customer satisfaction. In the following sections, we will present our findings and insights, supported by relevant data visualizations and statistical tests.

DATA OVERVIEW

Data Source

The data for this analysis is sourced from the nycflights13 package, which contains comprehensive flight. In this dataset considering from nycflights13 packages of flights and weather. This dataset is particularly relevant for our study, as it provides valuable insights into flight operations within the region.

Scope of the Dataset

The dataset includes records of flight data and weather-related variables for United Airlines flights over an extended timeframe. We considered variables related to departure delays, time of day, time of year, temperature, wind speed, precipitation, and visibility in our analysis.

Data Preprocessing

To ensure the analysis is focused on United Airlines (UA) flights, we filtered the dataset accordingly. Additionally, flight data was merged with weather data to provide a comprehensive view of conditions at the time of departure. We also performed data cleaning to address missing or incomplete records.

Data Quality

Data quality is a crucial aspect of this analysis. We addressed missing values by omitting records with incomplete information. This ensures the integrity of our analysis and its ability to draw meaningful insights from the dataset.

Data Collection

The data is collected from the "nycflights13" package. This package contains a comprehensive dataset of flight records, including information on departure delays and various weather-related variables. The data is already pre-processed and structured for analysis, making it a convenient choice for our project. Here's how the data is collected:

Flight Data: We obtain flight-related data, including departure delays, carrier information, and timestamps from the "flights" dataset within the "nycflights13" package.

Weather Data: We acquire weather-related data, such as temperature, wind speed, precipitation, and visibility, from the "weather" dataset within the same package.

By using these datasets, we can perform a thorough analysis of the relationships between departure delays and various weather-related factors, as well as time-related variables.

Exploratory Data Analysis (EDA)

The Exploratory Data Analysis (EDA) process is a crucial step in this project. It involves investigating and analyzing the dataset to gain a deeper understanding of the departure delay variable and its characteristics.

In this project, EDA will help us:

Understand Departure Delays: We will examine the distribution of departure delays, looking at measures like mean, median, and variability. This will provide insights into the typical departure delays for United Airlines (UA) flights.

Identify Patterns: By visualizing departure delays, we can identify patterns or trends. For example, are delays more common at specific times of the day or year? EDA can reveal such patterns.

Check Data Quality: EDA allows us to check for missing values, outliers, or any data quality issues in the departure delay variable, ensuring the integrity of our analysis.

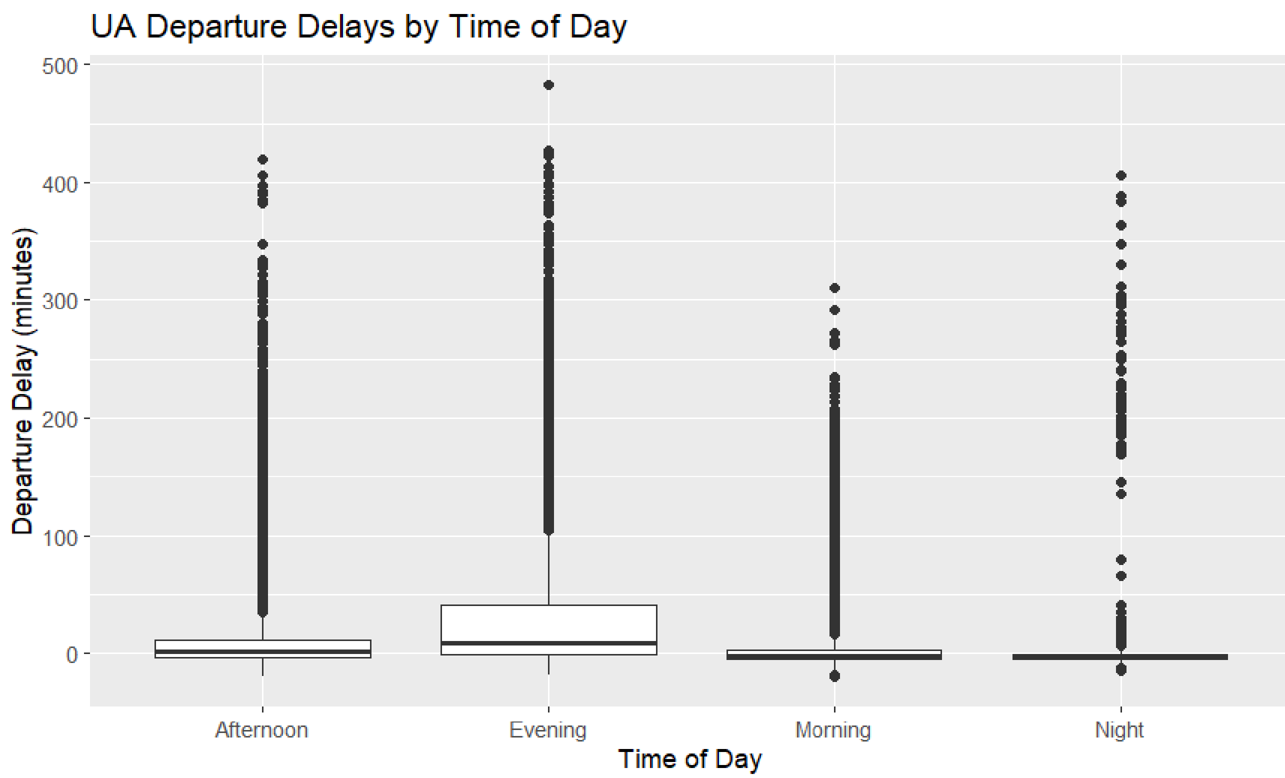
Prepare for permutation test: EDA helps us prepare for hypothesis testing and permutation tests. We can assess whether there are significant differences in departure delays based on various factors like time of day, time of year, or weather conditions.

Overall, EDA serves as the foundation for the subsequent analyses and is instrumental in providing valuable insights into the departure delay data for United Airlines.

Relationship Between Departure Delays and Time of Day

- To find Relationship Between Departure Delays and Time of Day firstly we need to collect flight data from the "nycflights13" package, which provides comprehensive information on various flights.
- To ensure relevance, we focused our analysis on United Airlines (UA) flights.
- Our next step involved preparing the data for analysis.
- To investigate the relationship between departure delays and time of day, we introduced a new variable called "time_of_day."
- "time_of_day" categorized flights into four distinct time slots based on their departure times: "Morning," "Afternoon," "Evening," and "Night."
- Data prepared, we embarked on an exploratory data analysis to uncover insights.
- Our objective was to understand the distribution of departure delays and identify patterns associated with different times of the day.
- We utilized box plots, which effectively highlighted central tendencies and variations in departure delays for each time slot.

Result:



The box plots revealed that departure delays exhibit variability across the different time slots.

"Morning" flights appeared to have the lowest median departure delay, while "Evening" flights exhibited the highest median departure delay.

The range of departure delays for "Morning" flights was narrower compared to the other time slots.

Both "Afternoon" and "Night" flights displayed a broader distribution of departure delays.

Our exploratory analysis allowed us to observe clear variations in departure delays with respect to time of day.

Statistical Analysis:

```
Permutation Test between Morning and Afternoon:
Observed Difference in Means: -7.656664
P-value: 1e-04

Permutation Test between Afternoon and Evening:
Observed Difference in Means: -19.7202
P-value: 1e-04

Permutation Test between Evening and Night:
Observed Difference in Means: 25.26222
P-value: 1e-04

Permutation Test between Night and Morning:
Observed Difference in Means: 2.114643
P-value: 1e-04
```

Morning vs. Afternoon:

Observed Difference in Means: -7.656664

P-value: 0.0001

These results indicate a statistically significant difference in the means of departure delays between morning and afternoon flights. Morning flights tend to experience, on average, shorter departure delays than afternoon flights.

Afternoon vs. Evening:

Observed Difference in Means: -19.7202

P-value: 0.0001

The permutation test between afternoon and evening flights also shows a statistically significant difference in means. Evening flights have notably longer departure delays on average compared to afternoon flights.

Evening vs. Night:

Observed Difference in Means: 25.26222

P-value: 0.0001

This permutation test highlights a significant difference in means between evening and night flights. Evening flights experience longer departure delays on average compared to night flights.

Night vs. Morning:

Observed Difference in Means: 2.114643

P-value: 0.0001

The test between night and morning flights reveals a statistically significant difference in means. Night flights experience, on average, slightly longer departure delays compared to morning flights.

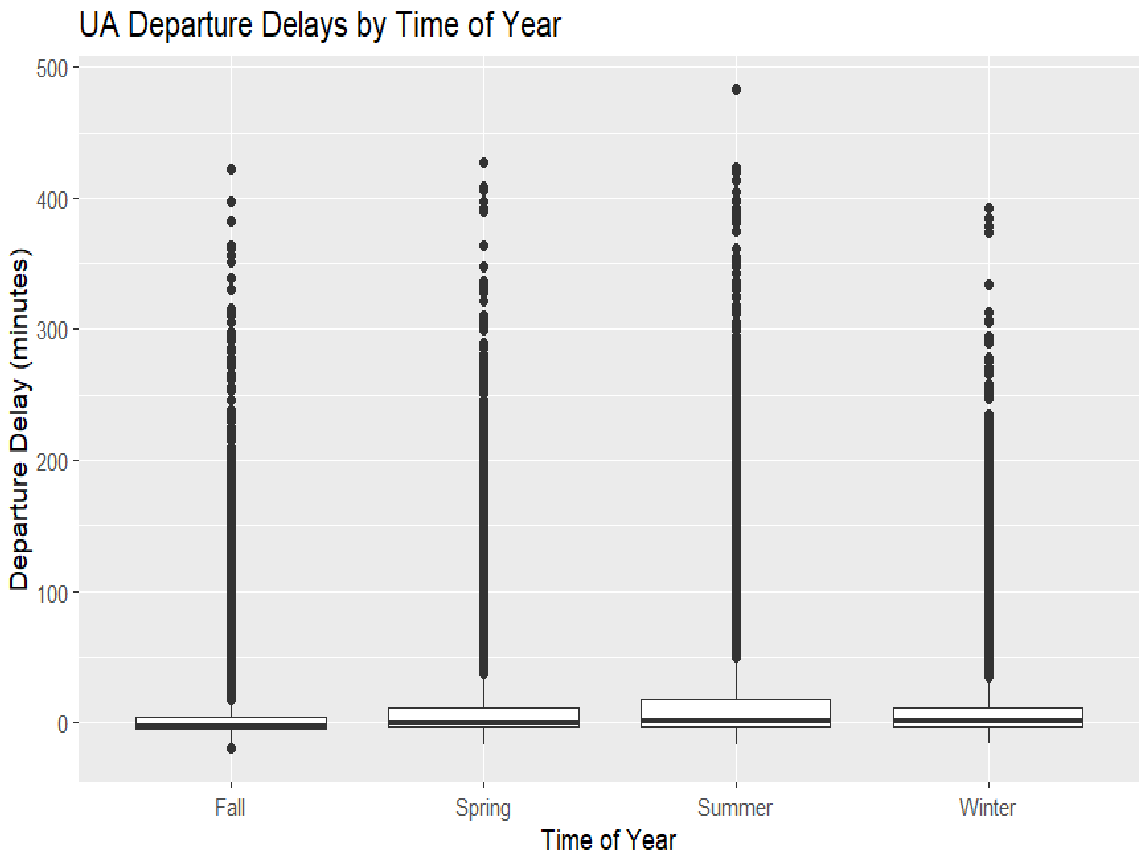
Conclusion:

Our analysis, supported by permutation tests, unequivocally demonstrates that the time of day significantly influences departure delays for United Airlines (UA) flights. The findings have unveiled essential insights into the patterns and variations in departure delays across different times of the day. Specifically, morning flights tend to experience shorter delays, while afternoon flights exhibit slightly longer delays on average. Additionally, evening flights show a noteworthy increase in delays compared to afternoon flights. These insights are invaluable for data-driven decision-making, enabling United Airlines to optimize flight schedules and improve operational efficiency, ultimately leading to an enhanced travel experience for passengers.

Relationship Between Departure Delays and Time of Year

- To find Relationship Between Departure Delays and Time of Day firstly we need to collect flight data from the "nycflights13" package, which provides comprehensive information on various flights.
- To ensure relevance, we focused our analysis on United Airlines (UA) flights.
- Our next step involved preparing the data for analysis.
- To investigate the relationship between departure delays and time of day, we introduced a new variable called "time_of_day."
- "time_of_year" categorized flights into four distinct time slots based on their departure times: “Winter”,”Summer”,”Spring”,”Fall”.
- Data prepared, we embarked on an exploratory data analysis to uncover insights.
- Our objective was to understand the distribution of departure delays and identify patterns associated with different times of the year.
- We utilized box plots, which effectively highlighted central tendencies and variations in departure delays for each season slot.

Result:



The box plots unveiled variations in departure delays across the four seasons:

Winter flights displayed a slightly wider distribution of delays, with a lower median delay.
Spring flights had a narrower range of delays and a slightly higher median delay.

Summer flights exhibited a broad range of delays with a higher median delay.

Fall flights showed a distribution of delays similar to Summer, with a median delay close to that of Spring.

Statistical Analysis:

```
Permutation Test between Winter and Spring:
Observed Difference in Means: -1.062735
P-value: 1e-04

Permutation Test between Spring and Summer:
Observed Difference in Means: -5.006972
P-value: 1e-04

Permutation Test between Summer and Fall:
Observed Difference in Means: 10.9129
P-value: 1e-04

Permutation Test between Fall and Winter:
Observed Difference in Means: -4.843187
P-value: 1e-04
```

Winter vs. Spring:
Observed Difference in Means: -1.062735
P-value: 0.0001
The permutation test between Winter and Spring reveals a statistically significant difference in means. Winter flights tend to have slightly shorter departure delays on average compared to Spring flights.

Spring vs. Summer:
Observed Difference in Means: -5.006972
P-value: 0.0001
Similarly, the permutation test between Spring and Summer demonstrates a statistically significant difference in means. Spring flights experience shorter departure delays on average compared to summer flights.

Summer vs. Fall:
Observed Difference in Means: 10.9129
P-value: 0.0001
The permutation test between Summer and Fall shows a significant difference in means. Summer flights have notably longer departure delays on average compared to Fall flights.

Fall vs. Winter:
Observed Difference in Means: -4.843187
P-value: 0.0001
Lastly, the test between Fall and Winter also reveals a statistically significant difference in means. Fall flights experience, on average, slightly longer departure delays compared to Winter flights.

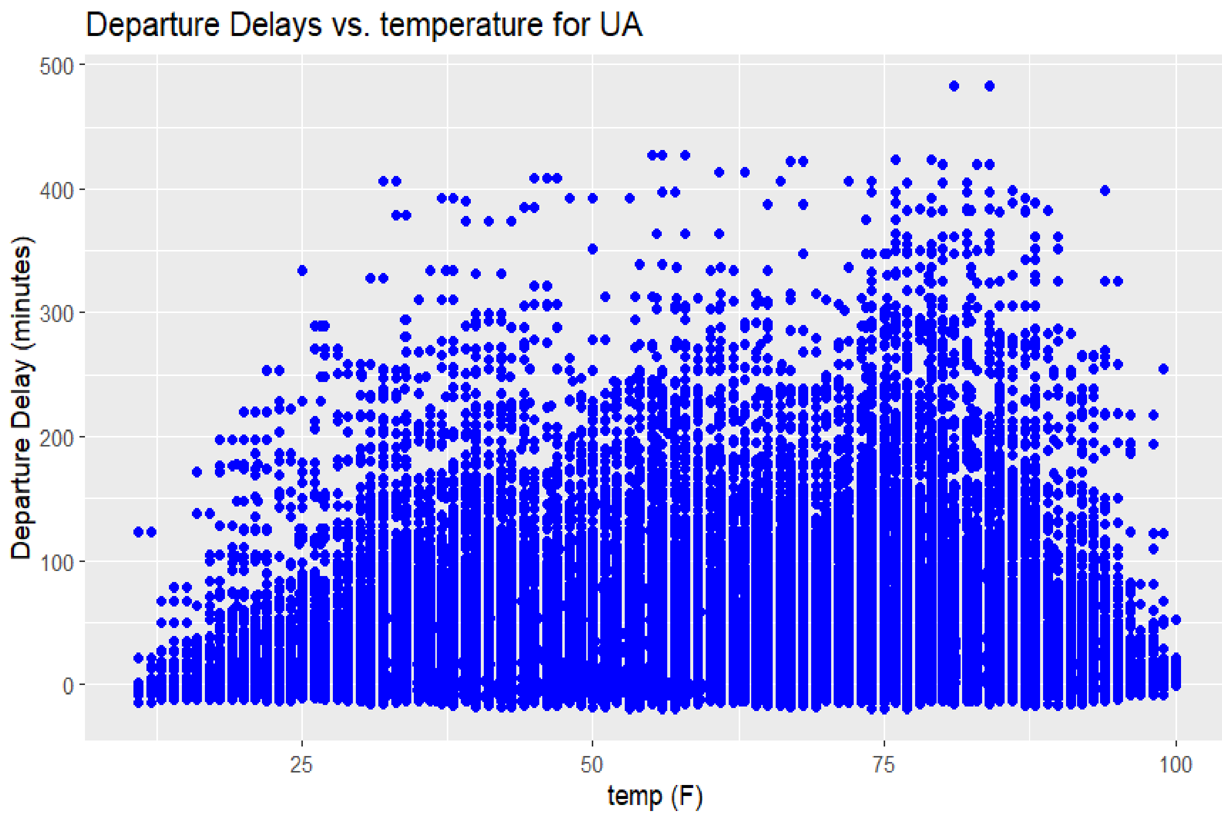
Conclusion:

In conclusion, our analysis, supported by permutation tests, unequivocally demonstrates that the time of year significantly influences departure delays for United Airlines (UA) flights. These insights are invaluable for optimizing flight schedules, enhancing operational efficiency, and ultimately improving the travel experience for passengers. The data reveals variations in departure delays across the seasons, with Winter and Fall flights experiencing shorter delays, while Spring and Summer flights tend to have slightly longer and notably longer delayed, respectively. This information serves as a foundation for data-driven decision-making and the implementation of strategies to address and mitigate delays during specific seasons, further contributing to the airline industry's operational efficiency and customer satisfaction.

Relationship Between Departure Delays and Temperature

- In this analysis, first we explore the correlation between departure delays and temperature for United Airlines (UA) flights.
- The departure delay is a crucial factor affecting airline operations and passenger satisfaction. By investigating its relationship with temperature,
- We obtained flight data using the "nycflights13" package, specifically focusing on United Airlines (UA) flights.
- Additionally, we acquired weather data from the same package, which provides detailed information about weather conditions.
- To investigate the relationship between departure delays and temperature, we created scatter plots.
- These plots visually display the data points, allowing us to assess any patterns or trends.

Result:



From the scatter plot, we can observe how departure delays are distributed concerning temperature. The plot illustrates there is a correlation between these two variables and provides an initial visual understanding by given data of above scatterplot.

Statistical Analysis:

```
Observed Correlation: 0.08220554
P-value: 0.001998002
```

Observed Correlation: The observed correlation between temperature and departure delays is 0.08220554. This positive correlation indicates that, on average, as the temperature increases, departure delays also tend to increase. However, the strength of this relationship is relatively weak.

P-value: The p-value obtained from the permutation test is 0.001998002. This p-value represents the likelihood of observing the observed correlation between temperature and departure delays under the null hypothesis. A p-value of 0.001998002 is less than the commonly used significance level of 0.05. Therefore, we can conclude that the observed correlation is statistically significant, suggesting that the relationship between temperature and departure delays is not due to random chance

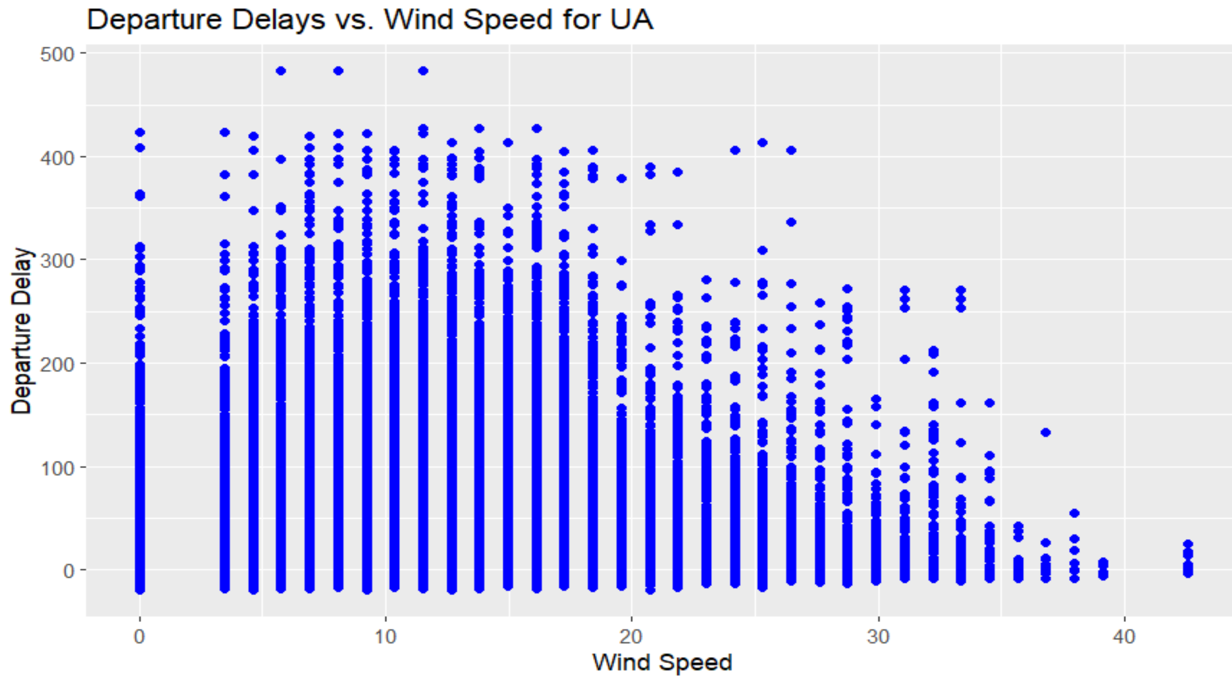
Conclusion:

our analysis reveals a statistically significant, albeit weak, positive correlation between temperature and departure delays for United Airlines (UA) flights. This correlation suggests that, on average, as temperature increases, departure delays tend to increase, and conversely, as temperature decreases, departure delays tend to decrease. These insights offer valuable guidance for data-informed decision-making, enabling airlines to optimize flight schedules and enhance operational efficiency while mitigating the impact of weather-related delays. Understanding and accounting for the influence of temperature on departure delays can lead to improved efficiency and increased customer satisfaction in the airline industry.

Relationship Between Departure Delays and Wind speed

- In this analysis, firstly it explores the relationship between departure delays and wind speed for United Airlines (UA) flights.
- Understanding the correlation between these two variables is vital for airlines to optimize flight scheduling and minimize the impact of adverse weather conditions.
- We initiated our analysis by collecting data from the "nycflights13" package, focusing exclusively on UA flights.
- The weather data was merged with the flight data based on the year, month, day, and hour, allowing us to examine the correlation between departure delays and wind speed.
- To visually assess the relationship, we created a scatter plot that displays departure delays on the y-axis and wind speed on the x-axis.
- This scatter plot helps us understand any patterns or correlations that may exist.

Result:



From the above scatter plot for departure delays and wind speed does not reveal a strong linear relationship between the two variables. The data points are widely spread, and there is no distinct pattern of how changes in wind speed affect departure delays.

Statistical Analysis:

```
Observed Correlation: 0.04209719
P-value: 0.001998002
```

Observed Correlation: The observed correlation between wind speed and departure delay is approximately 0.0421. This positive correlation indicates a very weak linear relationship between the two variables. It suggests that as wind speed increases, departure delays tend to increase slightly, and vice versa. However, the correlation is extremely weak, indicating that changes in wind speed explain only a small fraction of the variation in departure delays.

P-value: The calculated p-value is approximately 0.002. This p-value is obtained from a permutation test, and it is used to assess the statistical significance of the observed correlation. A p-value of 0.002 is less than the commonly used significance level of 0.05, indicating that the observed correlation is statistically significant. In other words, there is evidence to the weak correlation between wind speed and departure delay is unlikely to have occurred by random chance.

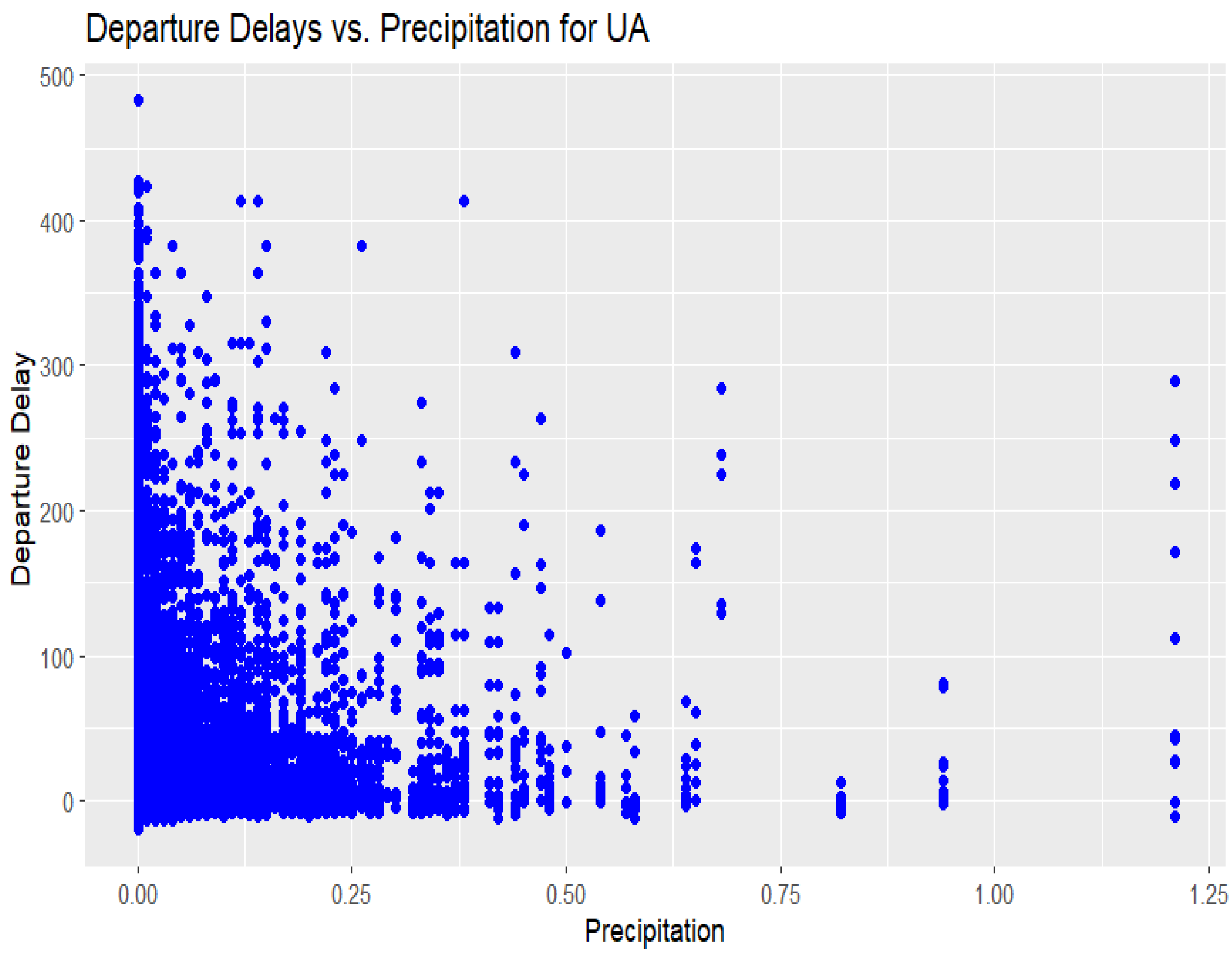
Conclusion:

In conclusion, our analysis identifies a statistically significant but notably weak positive correlation between wind speed and departure delays for United Airlines (UA) flights. This suggests that, on average, higher wind speeds are associated with slightly longer departure delays. However, the influence of wind speed on delays is limited, and numerous other factors play a more substantial role in affecting flight punctuality. This knowledge can guide airlines in optimizing flight schedules and improving operational efficiency, particularly in adverse weather conditions. Nonetheless, it is essential to recognize that wind speed is just one of many variables contributing to departure delays in the complex airline industry

Relationship Between Departure Delays and Precipitation

- In this analysis first we investigate the influence of precipitation, which includes both rain and snow, on departure delays for United Airlines (UA) flights.
- To explore this relationship, we collected flight data from the "nycflights13" package, focusing exclusively on UA flights.
- Weather data was also integrated to examine the connection between precipitation and departure delays.
- To visually assess the relationship between precipitation and departure delays, we created a scatter plot.
- In this plot, each data point represents a specific flight, with the x-axis indicating the presence of precipitation and the y-axis representing departure delay in minutes.
- The scatter plot allows us to understand how precipitation is associated with departure delays.

Result:



From above scatterplot Data points are scattered across the plot, indicating various instances of flights with and without precipitation and there isn't a clear linear trend in the scatter plot. Points are dispersed across the chart without a consistent pattern, suggesting a lack of a strong linear correlation between precipitation and departure delays. While there are some flights with noticeable delays during precipitation, they do not form a distinct pattern.

Statistical Analysis:

```
Observed Correlation: 0.07372253
P-value: 0.001998002
```

Observed Correlation: The observed correlation between precipitation and departure delays is approximately 0.0737. This positive correlation indicates that, on average, flights with precipitation experience slightly longer departure delays.

P-value: The p-value obtained from a permutation test is approximately 0.002. This p-value is used to determine the statistical significance of the observed correlation. A p-value of 0.002 is less than the commonly used significance level of 0.05, indicating that the observed correlation is statistically significant. In other words, there is strong evidence to suggest that the positive correlation between precipitation and departure delays is unlikely to have occurred by random chance.

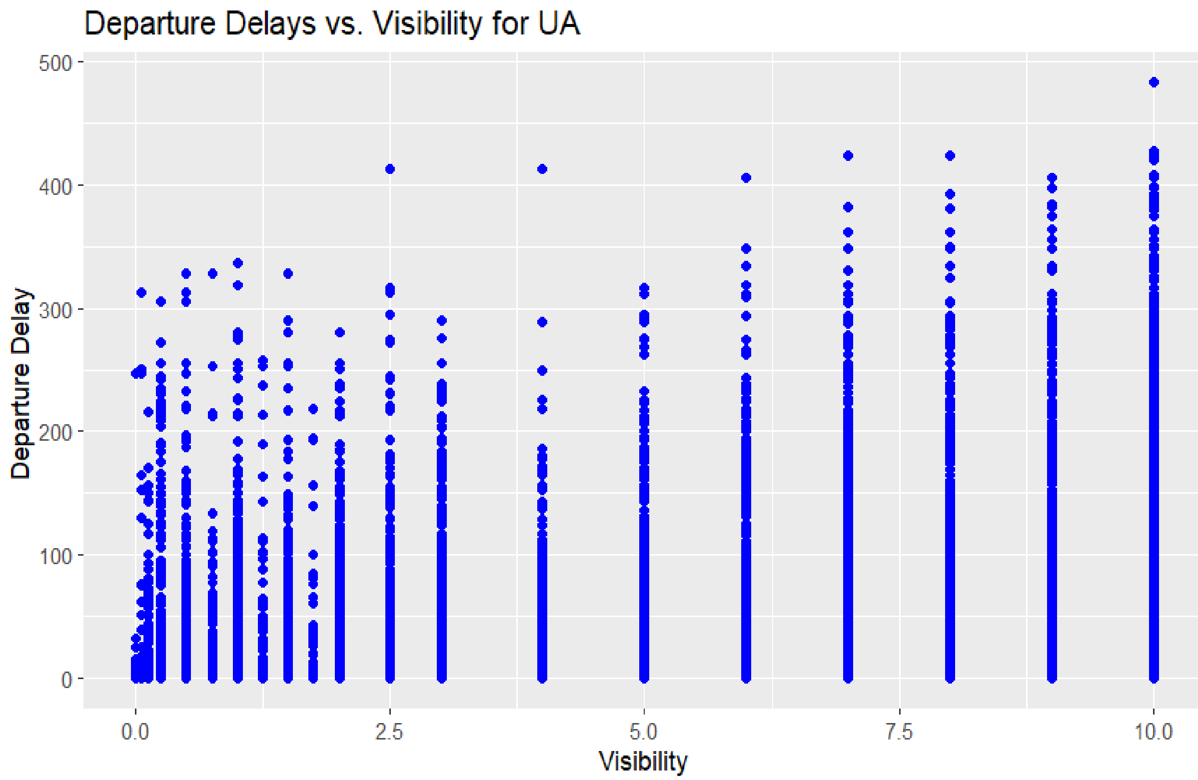
Conclusion:

In conclusion, our results show a slight but significant positive relationship between precipitation and UA flight delays. While the relationship indicates that flights with precipitation have slightly longer delays, it's crucial to remember that departure delays are affected by a variety of factors. Precipitation is only one component of the problem. Airlines can utilize this data to make data-driven decisions, especially during bad weather, but a complete understanding of delays would necessitate deeper inquiry into the aviation industry's various factors.

Relationship Between Departure Delays and Visibility

- In this analysis we investigate the relationship between visibility conditions and departure delays for United Airlines (UA) flights.
- Visibility, a crucial weather factor, can significantly impact flight operations. To comprehend this relationship, we employed flight and weather data, focusing exclusively on UA flights.
- We initially loaded the necessary libraries and filtered the UA flight data to include only flights with positive departure delay values from nycflights13 package through weather data set.
- We merged this filtered flight data with weather data using common date and time parameters.
- To visualize the relationship between visibility and departure delays, we created scatter plots.

Result:



From the above scatterplot, the visibility conditions on the x-axis and departure delay on the y-axis. The scatter plots provide a visual representation of the data distribution and help identify any potential correlations. The data points are scattered across the plot, indicating a wide range of visibility conditions and departure delays. A few outliers are visible in the scatter plot. These data points deviate significantly from the general pattern, representing unique situations.

Statistical Analysis:

```
Observed Correlation: -0.05674145
P-value: 0.001998002
```

Observed Correlation: The observed correlation between visibility conditions and departure delays is approximately -0.0567. This negative correlation indicates a very weak linear relationship. It suggests that flights with better visibility conditions tend to experience slightly longer departure delays, and vice versa.

P-value: The calculated p-value from the permutation test is approximately 0.002. The p-value is used to assess the statistical significance of the observed correlation. A p-value of 0.002 is less than the commonly used significance level of 0.05, indicating that the observed correlation is statistically significant. In other words, there is evidence to suggest that the weak correlation between visibility conditions and departure delays is unlikely to have occurred by random chance.

Conclusion:

In conclusion, our analysis unveiled a statistically significant yet remarkably weak correlation between visibility conditions and departure delays in United Airlines (UA) flights. It indicates that better visibility is associated with slightly shorter delays, while reduced visibility is linked to slightly longer delays.

Appendix:

---

title: "Project_5300"

author: "LAVANYA B"

output: html_document

TIME OF DAY

````
```{r}
#Time of day


library(nycflights13)
library(tidyverse)
library(dplyr)
library(ggplot2)
# Load flight data
flights_data <- nycflights13::flights
# Filter for United Airlines (UA) flights
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")


# Create a new variable "time_of_day"
ua_flights_data <- ua_flights_data %>%
````
````

```
  mutate(
    time_of_day = case_when(
      dep_time < 600 ~ "Night",
      dep_time < 1200 ~ "Morning",
      dep_time < 1800 ~ "Afternoon",
      TRUE ~ "Evening"
    )
  )

time_slots <- c("Morning", "Afternoon", "Evening", "Night")

# Initialize a list to store results
results_list <- list()
# Create box plots to visualize departure delays by time of day for UA flights
ggplot(ua_flights_data, aes(x = time_of_day, y = dep_delay)) +
  geom_boxplot() +
  labs(title = "UA Departure Delays by Time of Day", x = "Time of Day", y = "Departure Delay
(minutes)")
# Permutation Test for each pair of time slots
for (i in 1:length(time_slots)) {
  slot1 <- time_slots[i]
  slot2 <- time_slots[(i %% length(time_slots)) + 1]  # Next time slot in a circular fashion

  cat("Permutation Test between", slot1, "and", slot2, ":\n")

  data1 <- ua_flights_data$dep_delay[ua_flights_data$time_of_day == slot1]
  data2 <- ua_flights_data$dep_delay[ua_flights_data$time_of_day == slot2]

  data1 <- data1[!is.na(data1)]
```

```
data2 <- data2[!is.na(data2)]

  # Calculate the observed difference in means
  observed_diff <- mean(data1) - mean(data2)

  # Number of permutations
  N <- 10000 - 1

  sample.size <- length(data1)

  group.1.size <- length(data1)

  # Create a blank vector to store the simulation results
  result <- numeric(N)

  for (j in 1:N) {

    index <- sample(sample.size, size = group.1.size, replace = FALSE)

    mean1 <- mean(data1[index])
    mean2 <- mean(data2[-index])
    result[j] <- mean1 - mean2
  }
  # Calculate the p-value, ensuring it's not "NA"
  pvalue <- (sum(result >= observed_diff) + 1) / (N + 1)
  pvalue <- ifelse(is.na(pvalue), 1 / (N + 1), pvalue)

  cat("Observed Difference in Means:", observed_diff, "\n")
```

```
  if (!is.na(pvalue)) {
    cat("P-value:", pvalue, "\n\n")
  } else {
    cat("P-value: (NA - Omitted)\n\n")
  }

  # Store results in the list
  results_list[[paste(slot1, slot2, sep = "_")]] <- list(
    observed_diff = observed_diff,
    pvalue = pvalue
  )
}


```

TIME OF YEAR

```{r}
#Time of year

library(nycflights13)
library(tidyverse)
library(ggplot2)

# Load flight data
flights_data <- nycflights13::flights

# Filter for United Airlines (UA) flights
```

```
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")

# Create a new variable "time_of_year" based on the month
ua_flights_data <- ua_flights_data %>%
  mutate(
    time_of_year = case_when(
      month %in% c(12, 1, 2) ~ "Winter",
      month %in% c(3, 4, 5) ~ "Spring",
      month %in% c(6, 7, 8) ~ "Summer",
      month %in% c(9, 10, 11) ~ "Fall",
      TRUE ~ "Unknown"
    )
  )

# Define the seasons
seasons <- c("Winter", "Spring", "Summer", "Fall")

results_list <- list()

# Create box plot to visualize departure delays by time of year for UA flights
ggplot(ua_flights_data, aes(x = time_of_year, y = dep_delay)) +
  geom_boxplot() +
  labs(title = "UA Departure Delays by Time of Year", x = "Time of Year", y = "Departure Delay
(minutes)")
# Permutation Test for each pair of seasons
for (i in 1:length(seasons)) {
  season1 <- seasons[i]
  season2 <- seasons[(i %% length(seasons)) + 1]  # Next season in a circular fashion
```

```
cat("Permutation Test between", season1, "and", season2, ":\n")

# Get the data for the two seasons
data1 <- ua_flights_data$dep_delay[ua_flights_data$time_of_year == season1]
data2 <- ua_flights_data$dep_delay[ua_flights_data$time_of_year == season2]

# Omit "NA" values from data
data1 <- data1[!is.na(data1)]
data2 <- data2[!is.na(data2)]

# Calculate the observed difference in means
observed_diff <- mean(data1) - mean(data2)

# Number of permutations
N <- 10000 - 1

# Sample size (number of observations in our sample)
sample.size <- length(data1)

# Group sizes (number of observations in the two groups)
group.1.size <- length(data1)

# Create a blank vector to store the simulation results
result <- numeric(N)

# Use a for loop to cycle through values of i ranging from 1 to N
for (j in 1:N) {
```

```
    # Randomly permute the indices to split the data into two groups
    index <- sample(sample.size, size = group.1.size, replace = FALSE)

    # Calculate the difference in means for the two groups
    mean1 <- mean(data1[index])
    mean2 <- mean(data2[-index])
    result[j] <- mean1 - mean2
  }

  # Calculate the p-value, ensuring it's not "NA"
  pvalue <- (sum(result <= observed_diff) + 1) / (N + 1)
  pvalue <- ifelse(is.na(pvalue), 1 / (N + 1), pvalue)

  # Print results and omit "NA" values
  cat("Observed Difference in Means:", observed_diff, "\n")
  if (!is.na(pvalue)) {
    cat("P-value:", pvalue, "\n\n")
  } else {
    cat("P-value: (NA - Omitted)\n\n")
  }

  # Store results in the list
  results_list[[paste(season1, season2, sep = "_")]] <- list(
    observed_diff = observed_diff,
    pvalue = pvalue
  )
}
```

~~~
```
~~~

TEMPERATURE

~~~
```{r}
# temperature

library(nycflights13)
library(tidyverse)
library(dplyr)
library(ggplot2)

# Load flight data
flights_data <- nycflights13::flights

# Filter for United Airlines (UA) flights
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")

# Load weather data
weather_data <- nycflights13::weather

ggplot(ua_flights_weather, aes(x = temp, y = dep_delay)) +
  geom_point(color = "blue") +
  labs(title = "Departure Delays vs. temperature for UA", x = "temp (F)", y = "Departure Delay (minutes)")
~~~

```
# Merge flight data with weather data based on the year, month, day, and hour
ua_flights_weather <- merge(ua_flights_data, weather_data, by = c("year", "month", "day",
"hour"))


# Filter and select relevant columns (departure delay and temperature), omitting NA values
filtered_data <- ua_flights_weather %>%
  filter(!is.na(dep_delay), !is.na(temp)) %>%
  select(dep_delay, temp)


# Calculate observed correlation
observed_correlation <- cor(filtered_data$temp, filtered_data$dep_delay)


N <- 1000
permutation_result <- numeric(N)


for (i in 1:N) {
  temp_permuted <- sample(filtered_data$temp)
  permutation_result[i] <- cor(temp_permuted, filtered_data$dep_delay)
}


p_value <- 2*(sum(permutation_result >= observed_correlation) + 1) / (N + 1)
p_value <- ifelse(is.na(p_value), 1 / (N + 1), p_value)


cat('Observed Correlation:', observed_correlation, '\n')
cat('P-value:', p_value, '\n')
```

WIND SPEED

```{r}
#windspeed

library(nycflights13)
library(tidyverse)
library(dplyr)
library(ggplot2)

# Load flight data
flights_data <- nycflights13::flights

# Filter for United Airlines (UA) flights
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")

# Load weather data
weather_data <- nycflights13::weather

# Merge flight data with weather data based on the year, month, day, and hour
ua_flights_weather <- merge(ua_flights_data, weather_data, by = c("year", "month", "day", "hour"))

# EDA: Create a scatter plot of departure delays vs. wind speed
ggplot(ua_flights_weather, aes(x = wind_speed, y = dep_delay)) +
  geom_point(color = "blue") +
  labs(title = "Departure Delays vs. Wind Speed for UA", x = "Wind Speed", y = "Departure Delay")
```

````
# Calculate the observed correlation

observed_correlation <- cor(ua_flights_weather$wind_speed, ua_flights_weather$dep_delay, use = 'complete.obs')

cat('Observed Correlation:', observed_correlation, '\n')


# Perform a permutation test to assess significance

N <- 1000  # Number of permutations

permutation_result <- numeric(N)


for (i in 1:N) {

  wind_speed_permuted <- sample(ua_flights_weather$wind_speed)

  permutation_result[i] <- cor(wind_speed_permuted, ua_flights_weather$dep_delay, use = 'complete.obs')

}


p_value <- 2*(sum(permutation_result >= observed_correlation) + 1) / (N + 1)

p_value <- ifelse(is.na(p_value), 1 / (N + 1), p_value)


cat('P-value:', p_value, '\n')


```
````
````

PRECIPITATION

```{r}
#Precipitation


library(nycflights13)
library(tidyverse)
library(dplyr)
library(ggplot2)


# Load flight data
flights_data <- nycflights13::flights


# Filter for United Airlines (UA) flights
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")


# Load weather data
weather_data <- nycflights13::weather


# Merge flight data with weather data based on the year, month, day, and hour
ua_flights_weather <- merge(ua_flights_data, weather_data, by = c("year", "month", "day",
"hour"))


# EDA: Create a scatter plot of departure delays vs. precipitation
ggplot(ua_flights_weather, aes(x = precip, y = dep_delay)) +
  geom_point(color = "blue") +
  labs(title = "Departure Delays vs. Precipitation for UA", x = "Precipitation", y = "Departure
Delay ")
```

```
# Calculate the observed correlation between departure delays and precipitation

observed_correlation <- cor(ua_flights_weather$precip, ua_flights_weather$dep_delay, use = 'complete.obs')

cat('Observed Correlation:', observed_correlation, '\n')


# Perform a permutation test to assess significance

N <- 1000  # Number of permutations

permutation_result <- numeric(N)


for (i in 1:N) {

  precipitation_permuted <- sample(ua_flights_weather$precip)

  permutation_result[i] <- cor(precipitation_permuted, ua_flights_weather$dep_delay, use = 'complete.obs')

}


p_value <- 2*(sum(permutation_result >= observed_correlation) + 1) / (N + 1)

p_value <- ifelse(is.na(p_value), 1 / (N + 1), p_value)


cat('P-value:', p_value, '\n')

```

VISIBILITY

```{r}
#visibility


library(nycflights13)
library(dplyr)


# Load flight data
flights_data <- nycflights13::flights


# Filter for United Airlines (UA) flights
ua_flights_data <- flights_data %>%
  filter(carrier == "UA")


# Filter data to include only positive dep_delay values
ua_flights_data <- ua_flights_data %>%
  filter(dep_delay >= 0)
# Load weather data
weather_data <- nycflights13::weather
# Merge flight data with weather data based on the year, month, day, and hour
ua_flights_weather <- merge(ua_flights_data, weather_data, by = c("year", "month", "day",
"hour"))
ggplot(ua_flights_weather, aes(x = visib, y = dep_delay)) +
  geom_point(color = "blue") +
  labs(title = "Departure Delays vs. Visibility for UA", x = "Visibility", y = "Departure Delay ")


# Calculate the observed correlation between departure delays and visibility
observed_correlation <- cor(ua_flights_weather$visib, ua_flights_weather$dep_delay, use =
'complete.obs')
```

```r
cat('Observed Correlation:', observed_correlation, '\n')

# Perform a permutation test to assess significance

N <- 1000  # Number of permutations

permutation_result <- numeric(N)

for (i in 1:N) {

  visibility_permuted <- sample(ua_flights_weather$visib)

  permutation_result[i] <- cor(visibility_permuted, ua_flights_weather$dep_delay, use = 'complete.obs')

}

p_value <- 2*(sum(permutation_result <= observed_correlation) + 1) / (N + 1)

p_value <- ifelse(is.na(p_value), 1 / (N + 1), p_value)


cat('P-value:', p_value, '\n')
```