Project1

Ting-Yu, Lin

Introduction

This report analyzes departure delays for United Airlines (carrier code UA) using data from the nycflights13 package. The goal is to improve efficiency and customer satisfaction by understanding the factors influencing departure delays. We will explore the relationship between departure delays and various factors, including time of day, time of year, temperature, wind speed, precipitation, and visibility.

Time of Day

We will examine the relationship between departure delays and the time of day. This analysis will help us identify if there is a delay pattern throughout the day.

Table 1 displays the cumulative departure delay times for each hour, ranging from 5 a.m. to 11 p.m. Notably, the highest total departure delay is observed at 5 p.m.

hour <dbl>	total_dep_delay <dbl>
5	1803
6	14493
7	14369
8	21246
9	18816
10	21491
11	17335
12	24596
13	37915
14	49315
15	73497
16	44888
17	93544
18	88258
19	65536
20	89420
21	24842
22	349
23	185

Table1

Figure 1 presents a bar plot illustrating the average departure delay for each hour, spanning from 5 a.m. to 11 p.m. There is a notable trend of increasing mean departure delay as the hours approach midnight.

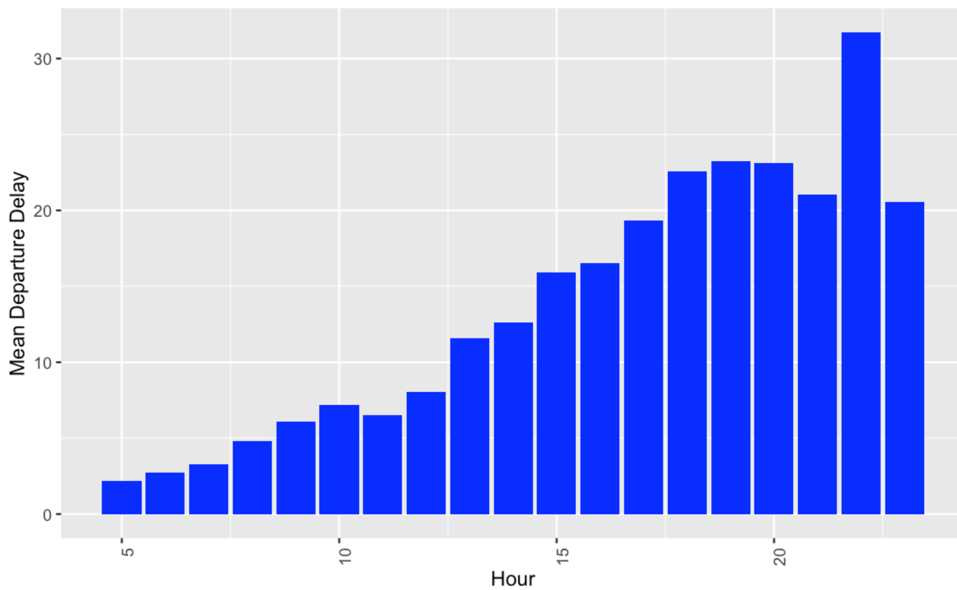

Figure 1

To delve deeper into this time-dependent phenomenon, we conducted a permutation test to compare the departure delay times between the evening till midnight (18:00 to 00:00) and morning (05:00 to 12:00) periods. The results revealed a significant difference in their means: 18.066. The associated p-value is 2e-04. This indicates that flights experience longer departure delays during the evening till midnight hours compared to the morning period. One plausible inference is that reduced visibility during the evening hours might contribute to these delays.

Time of Year

The next factor we will investigate is the time of year.

Table 2 presents the cumulative departure delay times for each month, with a noteworthy observation: the highest total departure delay is recorded in July.

month <int>	total_dep_delay <dbl>
1	38342
2	32125
3	57583
4	68149
5	60194
6	99503
7	100526
8	63439
9	32063
10	33491
11	30829
12	85654

Table

In Figure 2 of our analysis report, we display a bar plot illustrating the average departure delay for each month. It's worth noting that there is a notable peak in departure delays during the summer months. However, during the autumn and winter seasons, departure delays appear to be comparatively less severe. Nevertheless, a noteworthy observation is the emergence of a peak in departure delays in December. This pattern may be attributed to the peak travel season during summer, as well as the holiday rush in December, resulting in higher passenger volumes and increased airport congestion, which can contribute to flight delays.

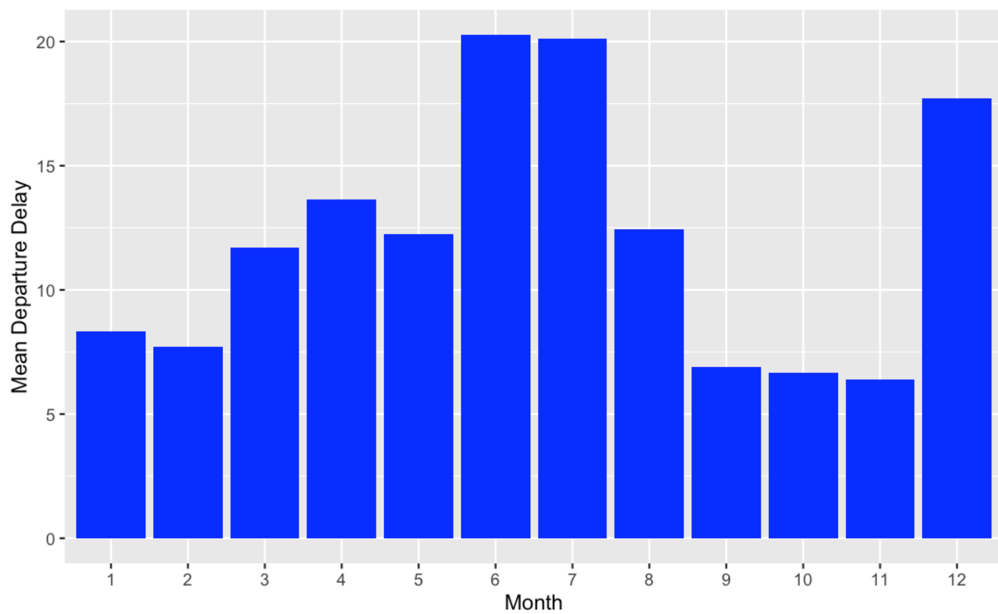

Figure 2

In the analysis, we performed a permutation test to compare departure delay times between the summer months (June to August) and the winter months (December to February). The results yielded a difference in their means, with a mean difference of 6.070. This statistical analysis, along with the associated p-value of 2e-04, highlights that flights tend to encounter longer departure delays during the summer season compared to the winter period.

Temperature

The temperature might impact flight operations, affecting equipment and crew performance. So, we merge the ua_flights dataset and the weather dataset by time hour to examine the relationship between temperature and delays using a point plot.

In Figure 3, we can observe the correlation between temperature and the mean departure delay. When the temperature is below 60°F, there is no pronounced strong relationship between temperature and delay. However, when the temperature exceeds 60°F, there is a positive correlation between temperature and departure delay.

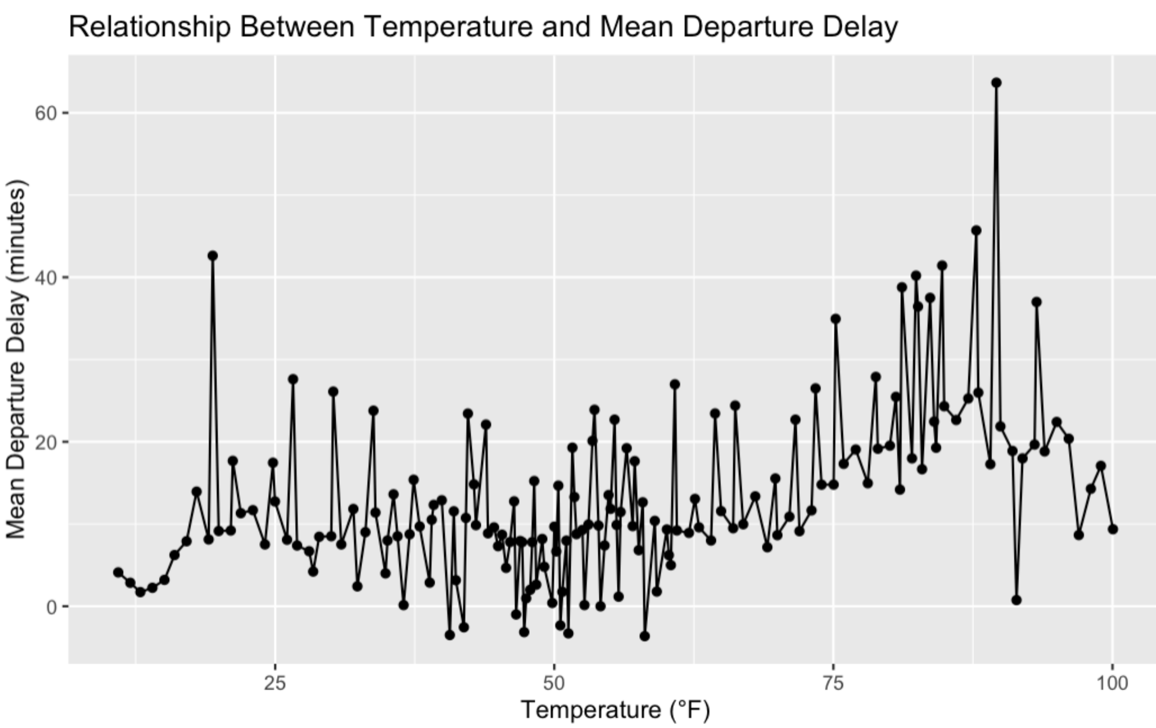

Figure 3

Wind Speed

Wind speed is another important meteorological factor that could influence departure delays.

In Figure 4, we can discern the correlation between wind speed and the mean departure delay. As the wind speed increases, there is a corresponding rise in the mean departure delay. Notably, when the wind speed reaches around 35 units, the mean departure delay reaches its peak. Subsequently, as the wind speed continues to rise, the mean departure delay decreases.

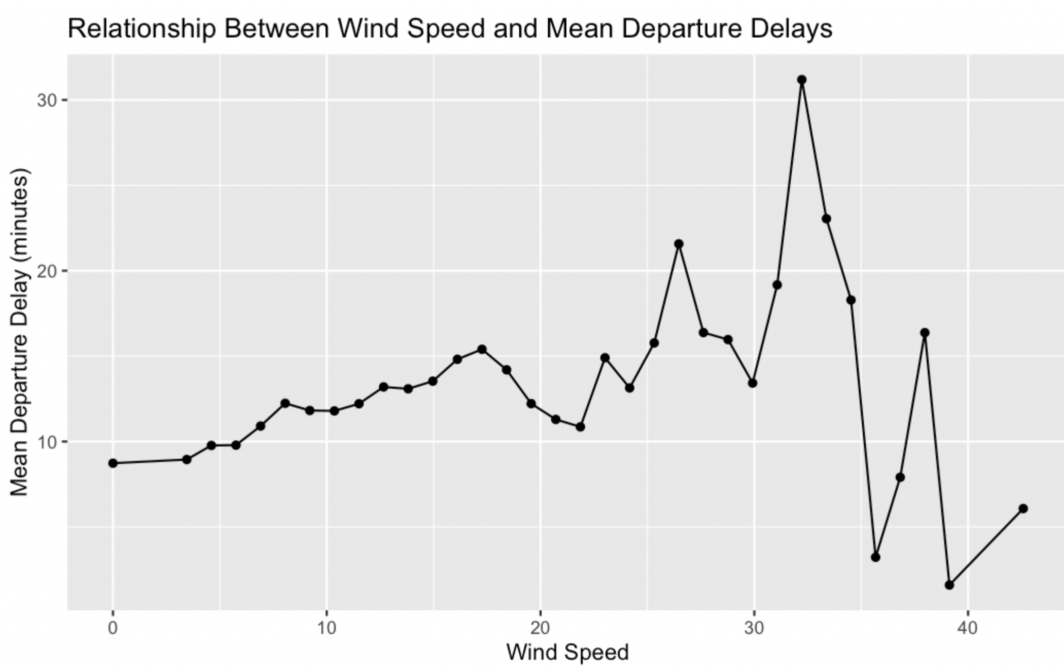

Figure 4

Precipitation

Precipitation has a significant impact on the normal operation of flights.

In Figure 5 of our analysis, we can observe the correlation between precipitation and the mean departure delay. When the precipitation is less than 0.5 units, there isn't a significant difference in the mean departure delay. However, when the precipitation exceeds 0.5 units, there is a notable variation in the mean departure delay, reaching a peak at around 0.65 units.

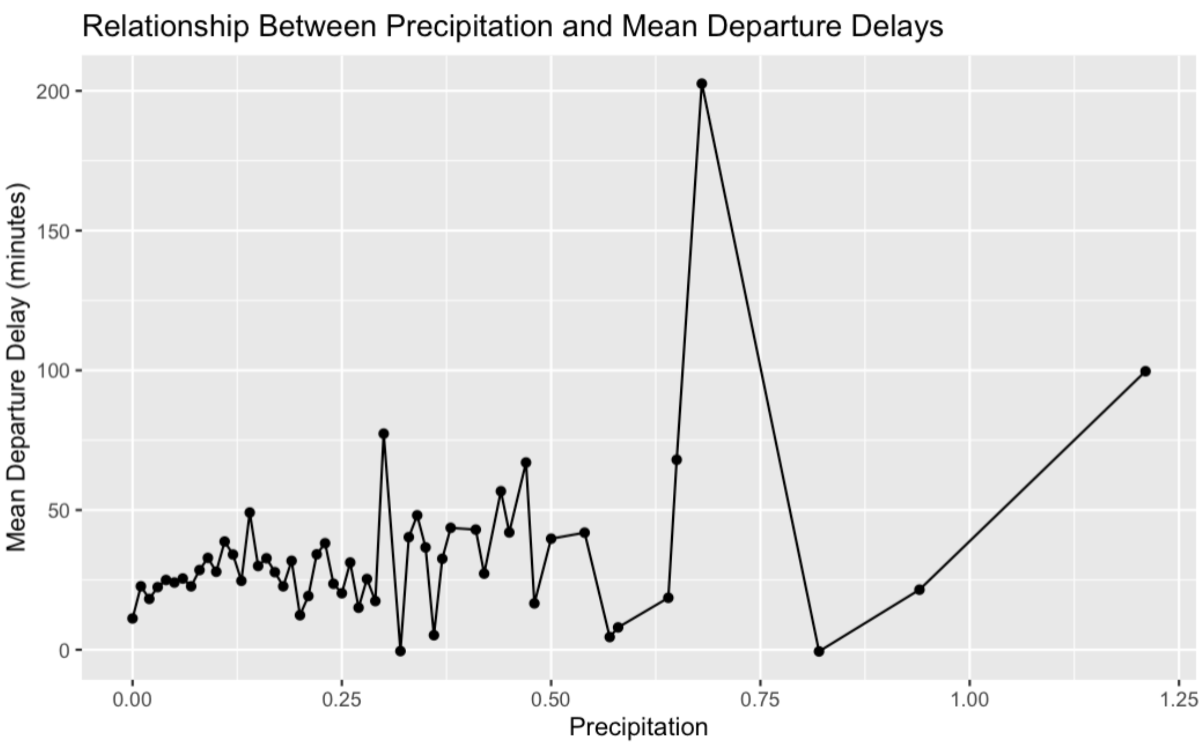

Figure 5

Visibility

In Figure 6 of our analysis, we can observe the correlation between visibility and the mean departure delay. There seems to be a stronger relationship between visibility and delay. Delays are higher when visibility is less than 2.5 miles.

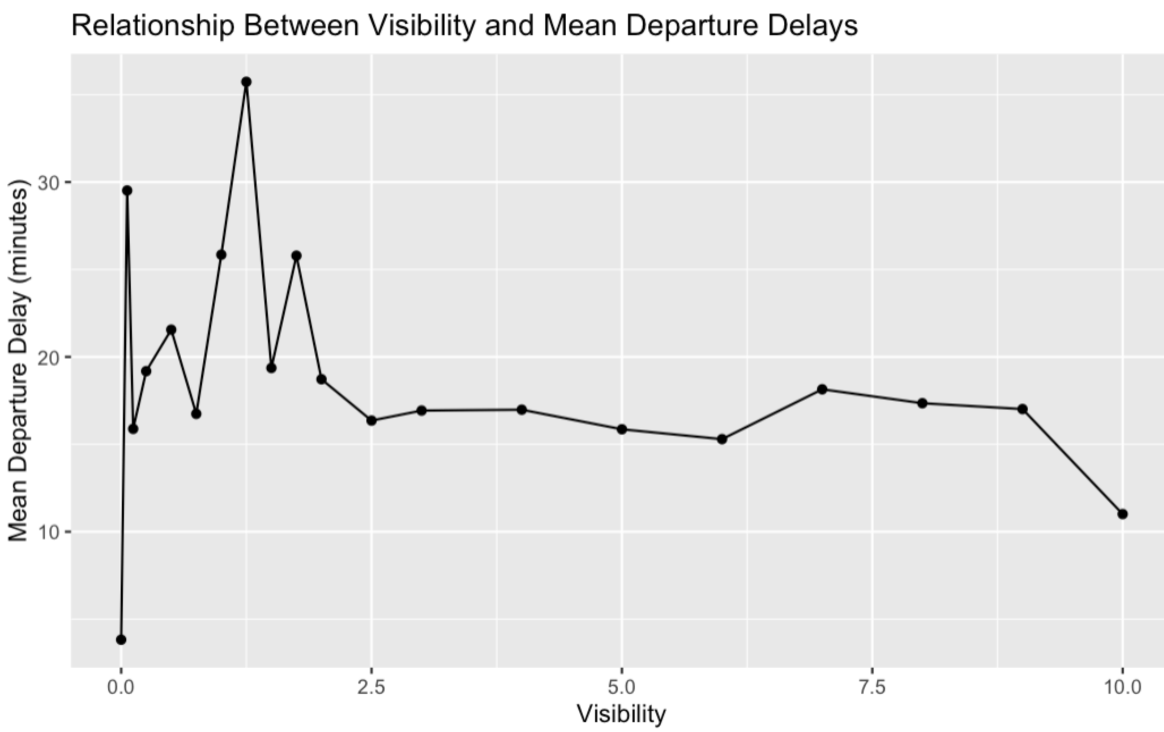

Figure 6

Conclusion

This report will provide a comprehensive analysis of departure delays for United Airlines, addressing the relationship with time of day, time of year, temperature, wind speed, precipitation, and visibility. By understanding the factors influencing departure delays, United Airlines can make informed decisions to enhance efficiency and improve customer satisfaction.

Code

The dataset only includes carrier UA

```{r}
ua_flights <- flights%>%
  filter(carrier=='UA')
print(ua_flights)
```

Table1

```{r}
ua_flights_day <- ua_flights %>%
  group_by(hour) %>%
  summarize(total_dep_delay = sum(dep_delay, na.rm = TRUE))
print(ua_flights_day)
```

Figure 1

```{r}
ua_flights_day <- ua_flights%>%
  group_by(hour)%>%
  summarize(mean_dep_delay = mean(dep_delay, na.rm = TRUE))
ggplot(data = ua_flights_day, mapping = aes(x = hour, y = mean_dep_delay)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Hour", y = "Mean Departure Delay")
```

Permutation test in Time of Day

```{r}
observed <- mean(ua_flights$dep_delay[ua_flights$hour >= 18 & ua_flights$hour <= 24], na.rm = TRUE) -
        mean(ua_flights$dep_delay[ua_flights$hour < 12 & ua_flights$hour > 5], na.rm = TRUE)

N <- 10^4 - 1
sample.size <- nrow(ua_flights)
group.1.size <- sum(ua_flights$hour >= 18 & ua_flights$hour <= 24)
result <- numeric(N)

for (i in 1:N) {
  index <- sample(sample.size, size = group.1.size, replace = FALSE)
  group1 <- ua_flights$dep_delay[index]
  group2 <- ua_flights$dep_delay[-index]
  result[i] = mean(group1, na.rm = TRUE) - mean(group2, na.rm = TRUE)
}

pvalue <- 2 * (sum(result >= observed) + 1) / (N + 1)
```

Table 2

```{r}
ua_flights_year <- ua_flights %>%
  group_by(month) %>%
  summarize(total_dep_delay = sum(dep_delay, na.rm = TRUE))
print(ua_flights_year)
```

Figure 2

```{r}
ua_flights_year <- ua_flights %>%
  group_by(month) %>%
  summarize(mean_dep_delay = mean(dep_delay, na.rm = TRUE))
ggplot(data = ua_flights_year, mapping = aes(x = factor(month), y = mean_dep_delay)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Month", y = "Mean Departure Delay")
```

Permutation test in Time of Year

```{r}
observed <- mean(ua_flights$dep_delay[ua_flights$month >= 6 & ua_flights$month <= 8], na.rm
= TRUE) -
       mean(ua_flights$dep_delay[ua_flights$month == 12 | ua_flights$month <= 2], na.rm =
TRUE)

N <- 10^4 - 1
sample.size <- nrow(ua_flights)
group.1.size <- sum(ua_flights$month >= 6 & ua_flights$month <= 8)
result <- numeric(N)

for (i in 1:N) {
  index <- sample(sample.size, size = group.1.size, replace = FALSE)
  group1 <- ua_flights$dep_delay[index]
  group2 <- ua_flights$dep_delay[-index]
  result[i] = mean(group1, na.rm = TRUE) - mean(group2, na.rm = TRUE)
}

pvalue <- 2 * (sum(result >= observed) + 1) / (N + 1)
```

Flights_weather dataset

```{r}
flights_weather <- merge(ua_flights, weather, by = "time_hour")
flights_weather
```

Figure 3

```{r}
temperature <- flights_weather %>%
  group_by(temp) %>%
  summarise(delay_mean = mean(dep_delay, na.rm = TRUE))

ggplot(data = temperature, aes(x = temp, y = delay_mean)) +
  geom_line() +
  geom_point() +
  labs(x = "Temperature(℉)", y = "Mean Departure Delay (minutes)") +
  ggtitle("Relationship Between temperature and Mean Departure Delays")
```

Figure 4

```{r}
wind_speed <- flights_weather %>%
  group_by(wind_speed) %>%
  summarise(delay_mean = mean(dep_delay, na.rm = TRUE))

ggplot(data = wind_speed, aes(x = wind_speed, y = delay_mean)) +
  geom_line() +
  geom_point() +
  labs(x = "Wind Speed", y = "Mean Departure Delay (minutes)") +
  ggtitle("Relationship Between Wind Speed and Mean Departure Delays")
```

Figure 5

```{r}
prec <- flights_weather %>%
  group_by(precip) %>%
  summarise(delay = mean(dep_delay, na.rm = TRUE))

ggplot(data = prec, aes(x = precip, y = delay)) +
  geom_line() +
  geom_point()+
  labs(x = "Precipitation", y = "Mean Departure Delay (minutes)") +
  ggtitle("Relationship Between Precipitation and Mean Departure Delays")
```

Figure6

```{r}
visibility<-flights_weather %>%
  #mutate(visib_cat = cut_interval(visib, n = 10)) %>%
  group_by(visib) %>%
  summarise(dep_delay = mean(dep_delay, na.rm = TRUE))

ggplot(data=visibility, aes(x = visib, y = dep_delay)) +
  geom_point()+
  geom_line()+
  labs(x = "Visibility", y = "Mean Departure Delay (minutes)") +
  ggtitle("Relationship Between Visibility and Mean Departure Delays")
```