

# Group 5: Alaska Airlines

Derek Atiawu, Mills Becouvarakis, Mauni Branch, & Sarah Wozniak

2024-09-25

## Key Findings

Alaska Airlines is the worst performing airline in June 2023 with 52 minutes of average delays. The worst route was LAX to SFO with 216 average departure delays and 242 average arrival delays. The best route was LIH-PHX with -9.47 average departure delays and -15.5 average arrival delays. These negative averages mean this route experienced more flights leaving early than flights leaving late. We also looked at flight delays by day of the month and found no significant relationship between these two factors.

This table shows the average amount of time Alaska Airlines had for delays in the month of June. The second table shows how many flights were delayed compared to on-time flights.

```
# A tibble: 1 × 2
  OP_UNIQUE_CARRIER TotalDelay
  <chr>                <dbl>
1 AA                  52.0
```

```
# A tibble: 4 × 3
  dep_status arr_status flight_count
  <chr>      <chr>      <int>
1 Delayed   Delayed      33204
2 Delayed   On-Time      9006
3 On-Time   Delayed      6224
4 On-Time   On-Time     34856
```

The following table shows the routes that had the most delays. The tables below show the top 5 worst and best routes, respectively.

```
# A tibble: 945 × 4
  ORIGIN DEST total_dep_delays total_arr_delays
  <chr>  <chr>      <int>      <int>
1 DFW    LAX         303        245
2 DFW    MIA         240        193
3 DFW    PHX         233        183
4 LAX    DFW         218        226
5 DFW    LAS         209        179
6 DFW    ORD         207        183
7 CLT    MCO         203        173
8 DFW    MCO         203        180
9 DFW    SAT         200        179
10 DFW   AUS         198        170
# i 935 more rows
```

```
# A tibble: 5 × 4
  ORIGIN DEST mean_dep_delays mean_arr_delays
  <chr>  <chr>          <dbl>          <dbl>
1 LAX    SFO            216.            242.
2 MEM    PHL            148             143
3 ORD    CID            145.            142.
4 BHM    DFW            116.            115.
5 STS    DFW            107.            106.

# A tibble: 5 × 4
  ORIGIN DEST mean_dep_delays mean_arr_delays
  <chr>  <chr>          <dbl>          <dbl>
1 STS    DFW            107.            106.
2 BHM    DFW            116.            115.
3 ORD    CID            145.            142.
4 MEM    PHL            148             143
5 LAX    SFO            216.            242.
```

The following table shows Alaska Airlines delays compared to other airlines.

```
# A tibble: 21 × 2
  OP_UNIQUE_CARRIER average_delay
  <chr>          <dbl>
1 AA            52.0
2 F9            51.2
3 ZW            44.8
4 B6            37.4
5 YV            34.0
6 G4            31.0
7 NK            30.8
8 MQ            29.7
9 OO            27.2
10 UA           27.2
# i 11 more rows
```

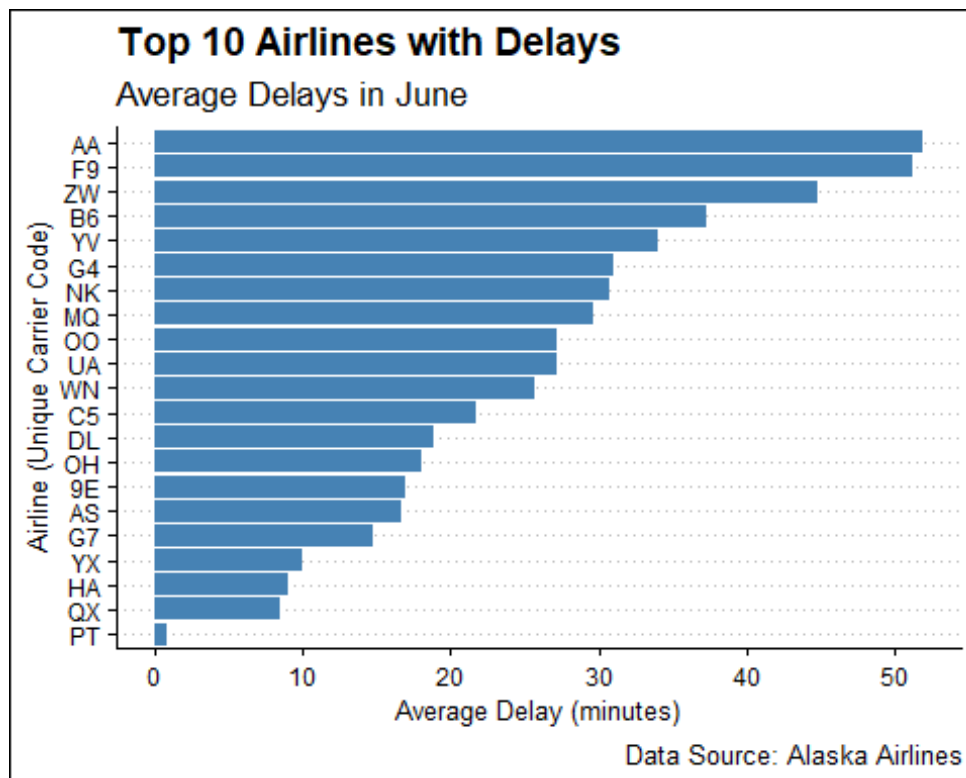
## Summary Table

This table summarizes key statistics for Alaska Airlines.

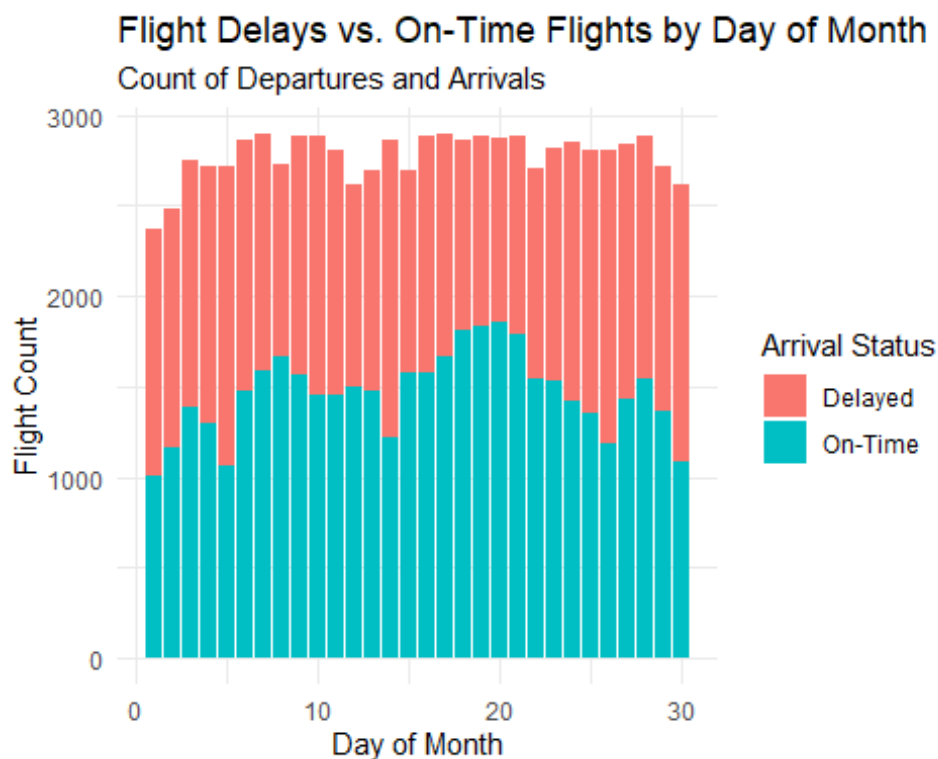
```
# A tibble: 1 × 6
  total_flights avg_dep_delay avg_arr_delay on_time_flights delayed_flights
  <int>          <dbl>          <dbl>          <int>          <int>
1    83290      28.9          23.0         34856         48434
# i 1 more variable: perc_delayed_flights <dbl>
```

## Visualizations

This is a bar chart representing the top 10 airlines with Delays in the month of June. Alaska Airlines is at the top with an average of 52 minutes.



This is a stacked bar chart representing Alaska delays vs on-time flights for the month of June.



## Business Insights

### Data Insights

Compared to other airlines, Alaska Airlines performed poorly in terms of flight status in June of 2024. It ranked at the very top of the list for high numbers of delayed flights with 52 minutes of average delays.

Alaska Airlines focuses mainly on the Western U.S. states. Compared to its competitors, it is a smaller airline, and its flights are often shorter and domestic rather than international. The airline's worst route for both departure and arrival delays is Los Angeles International Airport (LAX) to San Francisco International Airport (SFO). Although the flight is only 337 miles, the airports are two of the busiest in the U.S., and experience frequent congestion and overbooking. Alaska Airline's best route, however, is Lihue Airport (LIH) to Phoenix Sky Harbor Airport (PHX) with the fewest departure and arrival delays. This route likely performs well due to the few numbers of airlines flying in and out of Kauai, HI compared with the extreme traffic seen in the California airports.

### Potential Implications

Because Alaska Airlines focuses on the West, customers in the region likely depend on the airline for shorter flights within a state and to nearby states. The high number of delays will decrease customer satisfaction and negatively impact the airline's reliability. In response, we will likely see customers transferring to larger airlines viewed as more dependable.

### Recommendation

To improve Alaska Airlines' flight status compared to other larger airlines, we recommend that it can focus on targeting smaller airports like Lihue Airport (LIH) that may be out of the way for the larger, more dominant airlines to target with their higher number in international flights. Alaska Airlines should use its smaller fleet size to its advantage, capitalizing on these lower traffic routes to bolster its reputation as a dependable airline in the U.S. West and Pacific.

## Function Report

Our group found that `group_by()`, `mutate()`, and `ggplot()` were the most important tidyverse functions that we used to do our analysis. In our code, `group_by()` was used to group data by variables like `OP_UNIQUE_CARRIER`, `ORIGIN`, and `DEST`. We used `mutate()` to create new columns like `dep_status` and `arr_status` to classify flights as "Delayed" or "On-Time". Finally, `ggplot()` was used to create the bar chart visualizations showing flight delays by day as well as the top 10 airlines with delays.

## Appendix

This is the code used to achieve the results presented throughout the document.

```

####Drop NAs and filter by Alaska Airline
june_1 %>% filter(OP_UNIQUE_CARRIER == "AA") %>% drop_na() ->alaska_1
june_2 %>% filter(OP_UNIQUE_CARRIER == "AA") %>% drop_na() ->alaska_2

####Average Delays
average_aa_delays <- alaska_1 %>% group_by(OP_UNIQUE_CARRIER) %>%
  summarise(TotalDelay = mean(DEP_DELAY+ARR_DELAY))
print(average_aa_delays)

# A tibble: 1 × 2
  OP_UNIQUE_CARRIER TotalDelay
  <chr>                <dbl>
1 AA                    52.0

####Flights Delays vs Flights On-Time
flight_status <- alaska_1 %>% mutate(dep_status =ifelse(DEP_DELAY > 0,
"Delayed", "On-Time"),arr_status =ifelse(ARR_DELAY > 0, "Delayed", "On-
Time")) %>%
  group_by(dep_status,arr_status) %>%
  summarise(flight_count = n(), .groups = 'drop')
print(flight_status)

# A tibble: 4 × 3
  dep_status arr_status flight_count
  <chr>      <chr>      <int>
1 Delayed   Delayed      33204
2 Delayed   On-Time        9006
3 On-Time   Delayed       6224
4 On-Time   On-Time       34856

####Routes and Airports Most Delays
routes <- alaska_1 %>% group_by(ORIGIN, DEST) %>%
  summarise(total_dep_delays = sum(DEP_DELAY > 0),
            total_arr_delays = sum(ARR_DELAY > 0), .groups = 'drop') %>%
  arrange(desc(total_dep_delays))
print(routes)

# A tibble: 945 × 4
  ORIGIN DEST total_dep_delays total_arr_delays
  <chr>  <chr>      <int>      <int>
1 DFW    LAX        303        245
2 DFW    MIA        240        193
3 DFW    PHX        233        183
4 LAX    DFW        218        226
5 DFW    LAS        209        179
6 DFW    ORD        207        183
7 CLT    MCO        203        173
8 DFW    MCO        203        180
9 DFW    SAT        200        179
10 DFW   AUS        198        170
# i 935 more rows

```

```
####Worst Route
```

```
worst_route <- alaska_1 %>% group_by(ORIGIN, DEST) %>%  
  summarise(mean_dep_delays = mean(DEP_DELAY),  
            mean_arr_delays = mean(ARR_DELAY), .groups = 'drop') %>%  
  arrange(desc(mean_dep_delays))  
print(worst_route)
```

```
# A tibble: 945 × 4
```

|    | ORIGIN | DEST  | mean_dep_delays | mean_arr_delays |
|----|--------|-------|-----------------|-----------------|
|    | <chr>  | <chr> | <dbl>           | <dbl>           |
| 1  | LAX    | SFO   | 216.            | 242.            |
| 2  | MEM    | PHL   | 148             | 143             |
| 3  | ORD    | CID   | 145.            | 142.            |
| 4  | PDX    | CLT   | 117.            | 93.9            |
| 5  | BHM    | DFW   | 116.            | 115.            |
| 6  | SNA    | MIA   | 107.            | 93.6            |
| 7  | SDF    | MIA   | 107.            | 103.            |
| 8  | STS    | DFW   | 107.            | 106.            |
| 9  | DCA    | ATL   | 105.            | 90.3            |
| 10 | TPA    | LAX   | 104             | 85.3            |

```
# i 935 more rows
```

```
####Best Route
```

```
best_route <- alaska_1 %>% group_by(ORIGIN, DEST) %>%  
  summarise(mean_dep_delays = mean(DEP_DELAY), mean_arr_delays =  
mean(ARR_DELAY), .groups = 'drop') %>% arrange(mean_dep_delays)  
print(best_route)
```

```
# A tibble: 945 × 4
```

|    | ORIGIN | DEST  | mean_dep_delays | mean_arr_delays |
|----|--------|-------|-----------------|-----------------|
|    | <chr>  | <chr> | <dbl>           | <dbl>           |
| 1  | LIH    | PHX   | -9.47           | -15.5           |
| 2  | PSP    | PHX   | -7.38           | -7.38           |
| 3  | DCA    | EYW   | -7              | -28             |
| 4  | SAV    | ORD   | -7              | -13.8           |
| 5  | LGA    | DTW   | -6.33           | -23             |
| 6  | CHS    | ORD   | -5.5            | -14.8           |
| 7  | ORD    | CHS   | -5.5            | -12             |
| 8  | PHL    | PVD   | -5.5            | -8.5            |
| 9  | PVD    | PHL   | -5.4            | -11.4           |
| 10 | LGA    | BOS   | -5.3            | -12.1           |

```
# i 935 more rows
```

```
####Delays compared to other airline
```

```
airline_delays <- june_1 %>% group_by(OP_UNIQUE_CARRIER) %>%  
  summarise(average_delay = mean(DEP_DELAY+ARR_DELAY, na.rm=TRUE)) %>%  
  arrange(desc(average_delay))  
print(airline_delays)
```

```
# A tibble: 21 × 2
```

| OP_UNIQUE_CARRIER | average_delay |
|-------------------|---------------|
|-------------------|---------------|

|    | <chr> | <dbl> |
|----|-------|-------|
| 1  | AA    | 52.0  |
| 2  | F9    | 51.2  |
| 3  | ZW    | 44.8  |
| 4  | B6    | 37.4  |
| 5  | YV    | 34.0  |
| 6  | G4    | 31.0  |
| 7  | NK    | 30.8  |
| 8  | MQ    | 29.7  |
| 9  | 00    | 27.2  |
| 10 | UA    | 27.2  |

# i 11 more rows