Identifying Flight Duration Trends in Air Travel

Elvin Abdullayev

24-01-2024

1. Preparing the data

Load the flight, airline, and airport data to begin your analysis.

2. Complex data joining

Join the flights, airlines, and airports data frames together for a comprehensive dataset.

```
complex_join <- flights %>%
  left_join(airlines, by = "carrier") %>%
  left_join(airports, by = c("dest" = "faa")) %>%
  rename(airline_name = name.x, airport_name = name.y)

print(complex_join)
```

```
## # A tibble: 218,802 x 27
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>    <dbl>          <dbl>
##  1  2022     7     1        9           2129       160      118           2312
##  2  2022     7     1       12           1940       272      315           2253
##  3  2022     7     1       21           2120       181      140           2240
##  4  2022     7     1       21           2159       142      225             21
##  5  2022     7     1       22           2140       162      310             53
##  6  2022     7     1       23           2110       193      203           2259
##  7  2022     7     1       23           2100       203       NA              3
##  8  2022     7     1       39           1457       582      135           1626
##  9  2022     7     1       44           2155       169      134           2308
## 10  2022     7     1       57           1700       477      159           1829
## # i 218,792 more rows
## # i 19 more variables: arr_delay <dbl>, carrier <chr>, flight <dbl>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dttm>, airline_name <chr>,
## #   airport_name <chr>, lat <dbl>, lon <dbl>, alt <dbl>, tz <dbl>, dst <chr>,
## #   tzone <chr>
```

3. Data transformation

Transform the data to include flight duration for each flight.

```
transformed_data <- complex_join %>%
  mutate(flight_duration = air_time / 60)
print(transformed_data)
```

```
## # A tibble: 218,802 x 28
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <dbl> <dbl> <dbl>    <dbl>          <dbl>     <dbl>    <dbl>          <dbl>
##  1  2022     7     1        9           2129       160      118           2312
##  2  2022     7     1       12           1940       272      315           2253
##  3  2022     7     1       21           2120       181      140           2240
##  4  2022     7     1       21           2159       142      225             21
##  5  2022     7     1       22           2140       162      310             53
##  6  2022     7     1       23           2110       193      203           2259
##  7  2022     7     1       23           2100       203       NA              3
##  8  2022     7     1       39           1457       582      135           1626
##  9  2022     7     1       44           2155       169      134           2308
## 10  2022     7     1       57           1700       477      159           1829
## # i 218,792 more rows
## # i 20 more variables: arr_delay <dbl>, carrier <chr>, flight <dbl>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dttm>, airline_name <chr>,
## #   airport_name <chr>, lat <dbl>, lon <dbl>, alt <dbl>, tz <dbl>, dst <chr>,
## #   tzone <chr>, flight_duration <dbl>
```

4. Further data analysis

Determine the average flight duration and number of flights for each airline and airport combination.

```
analysis_result <- transformed_data %>%
  group_by(airline_name, airport_name) %>%
  summarise(avg_flight_duration = mean(flight_duration), count = n()) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'airline_name'. You can override using the
## `.groups` argument.
```

```
print(analysis_result)
```

```
## # A tibble: 359 x 4
##    airline_name         airport_name                    avg_flight_duration count
##    <chr>                <chr>                                         <dbl> <int>
##  1 Alaska Airlines Inc. Los Angeles International Air~                   NA   519
##  2 Alaska Airlines Inc. Portland International Airport                   NA   362
##  3 Alaska Airlines Inc. San Diego International Airpo~                   NA   546
##  4 Alaska Airlines Inc. San Francisco International A~                   NA  1304
##  5 Alaska Airlines Inc. Seattle Tacoma International ~                   NA  1273
##  6 Allegiant Air        Asheville Regional Airport                       NA   100
##  7 Allegiant Air        Cincinnati Northern Kentucky ~                   NA    58
##  8 Allegiant Air        Des Moines International Airp~                   NA    54
##  9 Allegiant Air        Destin-Ft Walton Beach Airport                 2.34    54
## 10 Allegiant Air        Gerald R. Ford International ~                 1.46    21
## # i 349 more rows
```

5. Finding the most frequent flight destination

Analyze which airline and city have the most flights from NYC.

```
frequent <- analysis_result %>%
  arrange(desc(count))

print(frequent)
```

```
## # A tibble: 359 x 4
##    airline_name           airport_name                 avg_flight_duration count
##    <chr>                  <chr>                                      <dbl> <int>
##  1 Delta Air Lines Inc.   Hartsfield Jackson Atlanta ~                 NA  5264
##  2 JetBlue Airways        General Edward Lawrence Log~                 NA  4524
##  3 American Airlines Inc. Miami International Airport                  NA  4301
##  4 Republic Airline       General Edward Lawrence Log~                 NA  3957
##  5 American Airlines Inc. Chicago O'Hare Internationa~                 NA  3905
##  6 American Airlines Inc. Charlotte Douglas Internati~                 NA  3823
##  7 Republic Airline       Ronald Reagan Washington Na~                 NA  3809
##  8 American Airlines Inc. Dallas Fort Worth Internati~                 NA  3659
##  9 JetBlue Airways        Orlando International Airpo~                 NA  3353
## 10 Republic Airline       John Glenn Columbus Interna~                 NA  3274
## # i 349 more rows
```

6. Determining the longest flight duration

Identify which airline and airport have the longest average flight duration from NYC.

```
longest <- analysis_result %>%
  arrange(desc(avg_flight_duration))

print(longest)
```

```
## # A tibble: 359 x 4
##    airline_name           airport_name                 avg_flight_duration count
##    <chr>                  <chr>                                      <dbl> <int>
##  1 Delta Air Lines Inc.   Daniel K Inouye Internation~                10.7     15
##  2 United Air Lines Inc.  Kahului Airport                             10.2     54
##  3 United Air Lines Inc.  Reno Tahoe International Ai~                5.12     4
##  4 American Airlines Inc. Glacier Park International ~                4.60    10
##  5 American Airlines Inc. Montrose Regional Airport                   4.34     3
##  6 American Airlines Inc. Jackson Hole Airport                        4.34    13
##  7 Delta Air Lines Inc.   Gallatin Field                              4.29    10
##  8 United Air Lines Inc.  Montrose Regional Airport                   4.20     8
##  9 United Air Lines Inc.  Yampa Valley Airport                        4.11     5
## 10 SkyWest Airlines Inc.  George Bush Intercontinenta~                3.77    11
## # i 349 more rows
```

7. Discovering the least common destination

Find out the least common destination airport for flights departing from JFK.

```r
transformed_data %>%
  filter(origin == "JFK") %>%
  group_by(airport_name) %>%
  summarize(count = n()) %>%
  arrange(count)
```

```
## # A tibble: 66 x 2
##    airport_name                                    count
##    <chr>                                           <int>
##  1 Eagle County Regional Airport                      17
##  2 Gallatin Field                                     48
##  3 Palm Springs International Airport                 59
##  4 Barnstable Municipal Boardman Polando Field        88
##  5 Norman Y. Mineta San Jose International Airport    92
##  6 Albuquerque International Sunport                 121
##  7 Reno Tahoe International Airport                  123
##  8 San Antonio International Airport                 183
##  9 John Wayne Airport-Orange County Airport          184
## 10 Ontario International Airport                     184
## # i 56 more rows
```

```r
least <- "Eagle County Regional Airport"
```