5 questions ask about flights dataset.

Naphon Seeluang

2023-09-22

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3      v readr     2.1.4
## v forcats   1.0.0      v stringr   1.5.0
## v ggplot2   3.4.3      v tibble    3.2.1
## v lubridate 1.9.2      v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(nycflights13)
```

5 Questions about flights and solution

1. Which carrier most flight in sep and dec

```r
flights %>%
  filter(month %in% c(11, 12) ) %>%
  count(carrier) %>%
  arrange(-n) %>%
  head(5) %>%
  left_join(airlines, by = "carrier")
```

```
## # A tibble: 5 x 3
##   carrier     n name
##   <chr>   <int> <chr>
## 1 UA       9785 United Air Lines Inc.
## 2 B6       9030 JetBlue Airways
## 3 EV       8778 ExpressJet Airlines Inc.
## 4 DL       7942 Delta Air Lines Inc.
## 5 AA       5282 American Airlines Inc.
```

2. Which flight most arrival delay of two or more hours

```r
flights %>%
  select(origin, dest, arr_delay) %>%
  filter(arr_delay >= 120) %>%
  arrange(-arr_delay) %>%
  head(10)
```

```
## # A tibble: 10 x 3
##    origin dest  arr_delay
##    <chr>  <chr>     <dbl>
##  1 JFK    HNL        1272
##  2 JFK    CMH        1127
##  3 EWR    ORD        1109
##  4 JFK    SFO        1007
##  5 JFK    CVG         989
##  6 JFK    TPA         931
##  7 LGA    MSP         915
##  8 LGA    ATL         895
##  9 EWR    MIA         878
## 10 EWR    ORD         875
```

3. Which flight never leave rate but arrival more than two hours

```r
flights %>%
    select(origin, dest, matches("^(dep|arr)_delay$")) %>%
    filter(dep_delay <= 0, arr_delay > 120) %>%
    arrange(dep_delay, -arr_delay) %>%
    head(20)
```

```
## # A tibble: 20 x 4
##    origin dest  dep_delay arr_delay
##    <chr>  <chr>     <dbl>     <dbl>
##  1 LGA    ATL         -11       121
##  2 LGA    MSP          -9       145
##  3 JFK    IAD          -9       132
##  4 LGA    BOS          -8       123
##  5 LGA    BNA          -7       126
##  6 EWR    DFW          -5       143
##  7 LGA    DTW          -5       124
##  8 JFK    SFO          -5       121
##  9 LGA    ORD          -4       124
## 10 JFK    SFO          -3       140
## 11 JFK    AUS          -3       138
## 12 LGA    DEN          -3       132
## 13 LGA    CLE          -3       127
## 14 LGA    DCA          -3       124
## 15 LGA    DCA          -3       124
## 16 JFK    SJU          -3       122
## 17 JFK    LAX          -2       194
## 18 LGA    DFW          -2       179
## 19 JFK    BUF          -2       142
## 20 JFK    LAX          -2       132
```

4. Which fastest flight in month between 7 and 9?

```r
flights %>%
    select(month, origin, dest, carrier, distance, air_time) %>%
    mutate(fastest_flight = distance/air_time) %>%
    filter(between(month,7,9)) %>%
    arrange(fastest_flight) %>%
    head(10)
```

```
## # A tibble: 10 x 7
##     month origin dest  carrier distance air_time fastest_flight
##     <int> <chr>  <chr> <chr>      <dbl>    <dbl>          <dbl>
##  1      8 JFK    PHL   9E            94       61           1.54
##  2      7 LGA    PHL   US            96       57           1.68
##  3      7 LGA    PHL   US            96       52           1.85
##  4      8 LGA    PHL   US            96       52           1.85
##  5      7 LGA    PHL   US            96       50           1.92
##  6      7 LGA    PHL   US            96       50           1.92
##  7      8 LGA    PHL   US            96       50           1.92
##  8      7 JFK    BOS   AA           187       96           1.95
##  9      8 JFK    PHL   9E            94       48           1.96
## 10      7 LGA    PHL   US            96       48           2
```

5. Which destination most flight ?

```r
flights %>%
  group_by(dest) %>%
  summarise(max_flight = max(flight)) %>%
  arrange(-max_flight) %>%
  head(10)
```

```
## # A tibble: 10 x 2
##     dest  max_flight
##     <chr>      <int>
##  1 ORD         8500
##  2 IAD         6181
##  3 SDF         6181
##  4 MSP         6168
##  5 ROC         6167
##  6 CVG         6140
##  7 DCA         6138
##  8 CLE         6120
##  9 BNA         6114
## 10 BUF         6103
```