6. R ÖDEVİ

Kevser Bahadır

2023-05-19

paketleri yüklüyoruz.

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(nycflights13)

## Warning: package 'nycflights13' was built under R version 4.2.3

df <- flights

Datayı inceleyelim.

head(df)

## # A tibble: 6 × 19  
## year month day dep\_time sched\_dep…¹ dep\_d…² arr\_t…³ sched…⁴ arr\_d…⁵ carrier  
## <int> <int> <int> <int> <int> <dbl> <int> <int> <dbl> <chr>   
## 1 2013 1 1 517 515 2 830 819 11 UA   
## 2 2013 1 1 533 529 4 850 830 20 UA   
## 3 2013 1 1 542 540 2 923 850 33 AA   
## 4 2013 1 1 544 545 -1 1004 1022 -18 B6   
## 5 2013 1 1 554 600 -6 812 837 -25 DL   
## 6 2013 1 1 554 558 -4 740 728 12 UA   
## # … with 9 more variables: flight <int>, tailnum <chr>, origin <chr>,  
## # dest <chr>, air\_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>,  
## # time\_hour <dttm>, and abbreviated variable names ¹​sched\_dep\_time,  
## # ²​dep\_delay, ³​arr\_time, ⁴​sched\_arr\_time, ⁵​arr\_delay

str(df)

## tibble [336,776 × 19] (S3: tbl\_df/tbl/data.frame)  
## $ year : int [1:336776] 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 ...  
## $ month : int [1:336776] 1 1 1 1 1 1 1 1 1 1 ...  
## $ day : int [1:336776] 1 1 1 1 1 1 1 1 1 1 ...  
## $ dep\_time : int [1:336776] 517 533 542 544 554 554 555 557 557 558 ...  
## $ sched\_dep\_time: int [1:336776] 515 529 540 545 600 558 600 600 600 600 ...  
## $ dep\_delay : num [1:336776] 2 4 2 -1 -6 -4 -5 -3 -3 -2 ...  
## $ arr\_time : int [1:336776] 830 850 923 1004 812 740 913 709 838 753 ...  
## $ sched\_arr\_time: int [1:336776] 819 830 850 1022 837 728 854 723 846 745 ...  
## $ arr\_delay : num [1:336776] 11 20 33 -18 -25 12 19 -14 -8 8 ...  
## $ carrier : chr [1:336776] "UA" "UA" "AA" "B6" ...  
## $ flight : int [1:336776] 1545 1714 1141 725 461 1696 507 5708 79 301 ...  
## $ tailnum : chr [1:336776] "N14228" "N24211" "N619AA" "N804JB" ...  
## $ origin : chr [1:336776] "EWR" "LGA" "JFK" "JFK" ...  
## $ dest : chr [1:336776] "IAH" "IAH" "MIA" "BQN" ...  
## $ air\_time : num [1:336776] 227 227 160 183 116 150 158 53 140 138 ...  
## $ distance : num [1:336776] 1400 1416 1089 1576 762 ...  
## $ hour : num [1:336776] 5 5 5 5 6 5 6 6 6 6 ...  
## $ minute : num [1:336776] 15 29 40 45 0 58 0 0 0 0 ...  
## $ time\_hour : POSIXct[1:336776], format: "2013-01-01 05:00:00" "2013-01-01 05:00:00" ...

summary(df)

## year month day dep\_time sched\_dep\_time  
## Min. :2013 Min. : 1.000 Min. : 1.00 Min. : 1 Min. : 106   
## 1st Qu.:2013 1st Qu.: 4.000 1st Qu.: 8.00 1st Qu.: 907 1st Qu.: 906   
## Median :2013 Median : 7.000 Median :16.00 Median :1401 Median :1359   
## Mean :2013 Mean : 6.549 Mean :15.71 Mean :1349 Mean :1344   
## 3rd Qu.:2013 3rd Qu.:10.000 3rd Qu.:23.00 3rd Qu.:1744 3rd Qu.:1729   
## Max. :2013 Max. :12.000 Max. :31.00 Max. :2400 Max. :2359   
## NA's :8255   
## dep\_delay arr\_time sched\_arr\_time arr\_delay   
## Min. : -43.00 Min. : 1 Min. : 1 Min. : -86.000   
## 1st Qu.: -5.00 1st Qu.:1104 1st Qu.:1124 1st Qu.: -17.000   
## Median : -2.00 Median :1535 Median :1556 Median : -5.000   
## Mean : 12.64 Mean :1502 Mean :1536 Mean : 6.895   
## 3rd Qu.: 11.00 3rd Qu.:1940 3rd Qu.:1945 3rd Qu.: 14.000   
## Max. :1301.00 Max. :2400 Max. :2359 Max. :1272.000   
## NA's :8255 NA's :8713 NA's :9430   
## carrier flight tailnum origin   
## Length:336776 Min. : 1 Length:336776 Length:336776   
## Class :character 1st Qu.: 553 Class :character Class :character   
## Mode :character Median :1496 Mode :character Mode :character   
## Mean :1972   
## 3rd Qu.:3465   
## Max. :8500   
##   
## dest air\_time distance hour   
## Length:336776 Min. : 20.0 Min. : 17 Min. : 1.00   
## Class :character 1st Qu.: 82.0 1st Qu.: 502 1st Qu.: 9.00   
## Mode :character Median :129.0 Median : 872 Median :13.00   
## Mean :150.7 Mean :1040 Mean :13.18   
## 3rd Qu.:192.0 3rd Qu.:1389 3rd Qu.:17.00   
## Max. :695.0 Max. :4983 Max. :23.00   
## NA's :9430   
## minute time\_hour   
## Min. : 0.00 Min. :2013-01-01 05:00:00.00   
## 1st Qu.: 8.00 1st Qu.:2013-04-04 13:00:00.00   
## Median :29.00 Median :2013-07-03 10:00:00.00   
## Mean :26.23 Mean :2013-07-03 05:22:54.64   
## 3rd Qu.:44.00 3rd Qu.:2013-10-01 07:00:00.00   
## Max. :59.00 Max. :2013-12-31 23:00:00.00   
##

1. 1 Ocak’taki tüm uçuşları seçiniz.

df %>%  
 select(flight, month) %>%  
 filter(month==1)

## # A tibble: 27,004 × 2  
## flight month  
## <int> <int>  
## 1 1545 1  
## 2 1714 1  
## 3 1141 1  
## 4 725 1  
## 5 461 1  
## 6 1696 1  
## 7 507 1  
## 8 5708 1  
## 9 79 1  
## 10 301 1  
## # … with 26,994 more rows

1. Yaz dönemindeki tüm uçuşları seçiniz. Haziran 6. ay

df %>%  
 select(flight, month) %>%   
 filter(between(month,6,8))

## # A tibble: 86,995 × 2  
## flight month  
## <int> <int>  
## 1 739 6  
## 2 1431 6  
## 3 1686 6  
## 4 1451 6  
## 5 725 6  
## 6 701 6  
## 7 540 6  
## 8 707 6  
## 9 1911 6  
## 10 5716 6  
## # … with 86,985 more rows

1. Kasım ve Aralık aylarında kalkan tüm uçuşları seçiniz.

df %>%  
 select(flight,month) %>%  
 filter(month %in% c(11,12))

## # A tibble: 55,403 × 2  
## flight month  
## <int> <int>  
## 1 745 11  
## 2 1816 11  
## 3 1895 11  
## 4 1714 11  
## 5 2243 11  
## 6 303 11  
## 7 2167 11  
## 8 2134 11  
## 9 563 11  
## 10 731 11  
## # … with 55,393 more rows

1. Varış gecikmesi 2 veya daha fazla saat olan uçuşları bulun. (Veride dk cinsindendir.)

df %>%  
 select(flight,dep\_delay) %>%  
 filter(dep\_delay >= 2)

## # A tibble: 120,382 × 2  
## flight dep\_delay  
## <int> <dbl>  
## 1 1545 2  
## 2 1714 4  
## 3 1141 2  
## 4 3768 8  
## 5 303 11  
## 6 135 3  
## 7 1837 13  
## 8 4144 24  
## 9 1701 8  
## 10 671 2  
## # … with 120,372 more rows

1. Houston’a uçan tüm uçuşları bulunuz. (HOU)

df %>%  
 select(flight,dest) %>%  
 filter(dest == "HOU")

## # A tibble: 2,115 × 2  
## flight dest   
## <int> <chr>  
## 1 625 HOU   
## 2 2596 HOU   
## 3 1066 HOU   
## 4 629 HOU   
## 5 20 HOU   
## 6 625 HOU   
## 7 3294 HOU   
## 8 2529 HOU   
## 9 629 HOU   
## 10 20 HOU   
## # … with 2,105 more rows

1. American veya Delta tarafından gerçekleştirilen tüm uçuşları bulunuz. (şunu sor)

df %>%  
 select(flight,carrier) %>%  
 filter(carrier == "UA" |  
 carrier == "DL")

## # A tibble: 106,775 × 2  
## flight carrier  
## <int> <chr>   
## 1 1545 UA   
## 2 1714 UA   
## 3 461 DL   
## 4 1696 UA   
## 5 194 UA   
## 6 1124 UA   
## 7 1187 UA   
## 8 1919 DL   
## 9 1743 DL   
## 10 1077 UA   
## # … with 106,765 more rows

1. İki saatten fazla varış zamanı geciken ancak geç kalkmayan (tam zamanında kalkış yapan) uçuşları bulunuz.

a <- df %>%  
 select(flight, arr\_delay , dep\_delay ) %>%  
 filter( arr\_delay >= 120 &  
 dep\_delay == 0 )  
print(a)

## # A tibble: 3 × 3  
## flight arr\_delay dep\_delay  
## <int> <dbl> <dbl>  
## 1 5181 130 0  
## 2 4626 128 0  
## 3 1057 140 0