

## Bursa Uludağ Üniversitesi

Bilgisayar Mühendisliği Bölümü

R Programlama ve Makine Öğrenmesi Uygulamaları Dersi Ödev H1O1

Beyzagül ŞAHİN 032290087

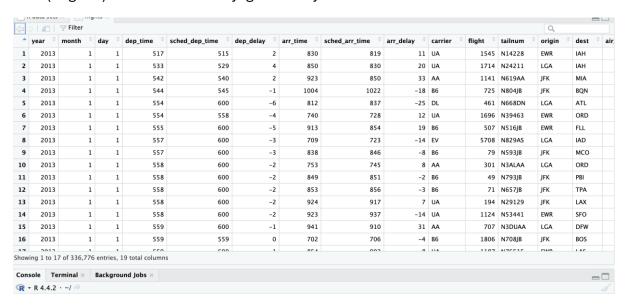
Alttaki kod ile gerekli olan paketleri yükledik.

Install.packages(c('corrr', 'tidyverse', 'nycflights13', 'Lahman', 'ggplot2'))

ls("package:nycflights13") Paket içini listeler.

```
> ls("package:nycflights13")
[[1] "airlines" "airports" "flights" "planes" "weather"
> View(flights)
> |
```

View('flights') komutu ile de veriyi görüntülüyoruz.



Left\_join() fonksiyonu, verisetlerini birleştirmeyi sağlar, Filter fonksiyonu ile de uçuş süresi en yüksek olan uçuşları gösterir.

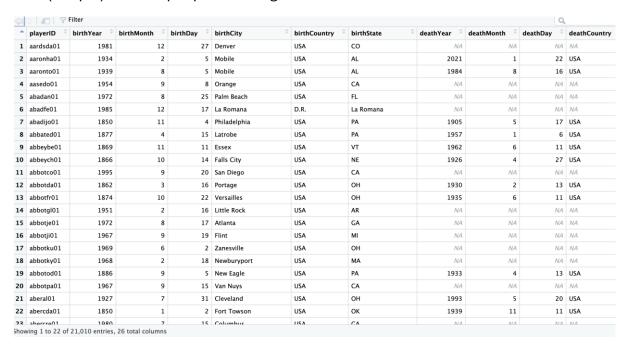
```
> data2<- left_join(weather,flights,by=c("origin","year","month","day","hour"))</pre>
> data2 %>% select(air_time,temp,wind_speed,humid) %>%
   correlate()
Correlation computed with
Method: 'pearson'
• Missing treated using: 'pairwise.complete.obs'
# A tibble: 4 \times 5
 term
            air_time temp wind_speed
  <chr>
             <dbl> <dbl> <dbl>
                                          <dbl>
1 air_time NA -0.036<u>7</u>
                               0.026<u>3</u> 0.040<u>5</u>
2 temp -0.0367 NA -0.140 0.0374
3 wind_speed 0.026<u>3</u> -0.140 NA
                                       -0.187
4 humid 0.040<u>5</u> 0.037<u>4</u> -0.187 NA
```

Right\_join fonksiyonu da veri birleştirmek için kullanılır. Filter fonksiyonu, bu kodda kalkış gecikmesi olan uçuşların filtrelenmesini sağlar. Groupby fonksiyonu da her hava yolu için gruplama yapar ve gecikmelerin ortalama ve medyanını hesaplar.

```
> data3<- right_join(airlines,flights,by="carrier")</pre>
> data3 %>% filter(dep_delay>0) %>% na.omit() %>% group_by(name) %>% sun
# A tibble: 16 \times 3
   name
                                 mean median
   <chr>
                                <dbl>
                                       <db1>
 1 AirTran Airways Corporation 40.6
                                        16
 2 Alaska Airlines Inc.
                                 31.5
                                        12
                                 37.2
 3 American Airlines Inc.
                                        16
 4 Delta Air Lines Inc.
                                 37.3
                                        16
 5 Endeavor Air Inc.
                                 48.5
                                        26
                                 44.7
 6 Envoy Air
                                        27
 7 ExpressJet Airlines Inc.
                                 50.2
                                        31
 8 Frontier Airlines Inc.
                                 45.2
                                        18
 9 Hawaiian Airlines Inc.
                                 44.8
                                         5
10 JetBlue Airways
                                 39.7
                                        20
11 Mesa Airlines Inc.
                                 52.9
                                        29.5
12 SkyWest Airlines Inc.
                                 58
                                        40
13 Southwest Airlines Co.
                                 34.8
                                        15
14 US Airways Inc.
                                 32.9
                                        16
15 United Air Lines Inc.
                                 29.8
                                        12
16 Virgin America
                                 34.2
                                        10
```

```
> flights %>% filter(dep_delay>0) %>% summarise(mean=mean(dep_delay))
# A tibble: 1 x 1
   mean
   <dbl>
1 39.4
```

View(People): Bu kod people verisini gösterir.



Oyuncuların ödüllerinin sayısı hesaplanır ve en çok ödül alan oyuncu bulunur.

Verilen aralıktaki oyuncuları listeler.

```
> People %>% mutate(BMI=weight/(height^2)*703) %>% filter(BMI>=25 & BMI<29.9) %>% nrow()
[1] 9719
```

En yüksek maaş alan oyuncunun bilgilerini gösterir.

```
> slice(data2,which.max(data2$salary)) #max salary
yearID teamID lgID playerID salary num.of.aw
1 2009 NYA AL rodrial01 33000000 31
```

En fazla ödül alan oyuncunun bilgilerini gösterir.

```
> slice(data2, which.max(data2$num.of.aw)) #max num of awards
   yearID teamID lgID playerID salary num.of.aw
      1986
              PIT
                     NL bondsba01 60000
 1
                                                   47
> table(AwardsPlayers$awardID) %>% as.data.frame() %>% arrange(desc(Freq))
                                   Var1 Freq
1
                           TSN All-Star 1525
2
            Baseball Magazine All-Star 1520
3
                             Gold Glove 1204
4
                         Silver Slugger 792
5
                  Most Valuable Player
                                         208
6
                    Rookie of the Year
                                         154
7
               TSN Pitcher of the Year 151
8
                        Cy Young Award 126
9
            Reliever of the Year Award
                                          94
                TSN Player of the Year
10
                                          92
11 TSN Major League Player of the Year
                                          89
               TSN Fireman of the Year
12
                                          88
13
                       Babe Ruth Award
                                          78
14
                      World Series MVP
                                          71
15
             Lou Gehrig Memorial Award
                                          69
16
                     All-Star Game MVP
                                          62
17
                           Hutch Award
                                          55
18
                Roberto Clemente Award
                                          54
19
                      Hank Aaron Award
                                          50
20
                               NLCS MVP
                                          49
21
                               ALCS MVP
                                          43
22
                 Pitching Triple Crown
                                          39
23
                                          36
           Comeback Player of the Year
> diamonds %>% mutate(t=x^2-sqrt(y)+(1/z)) %>% filter(t==min(t)) %>% select(depth)
# A tibble: 1 \times 1
 depth
  <db1>
1 62.6
```

Filtrelenme işlemiyle belirli bir renk ve berraklık seviyesindeki en düşük fiyatı bulur.

```
> diamonds %>% mutate(discount = case_when(
     cut=="Fair" ~ price*0.01,
     cut=="Good" ~ price*0.02,
     cut=="Very Good" ~ price*0.025,
     cut=="Premium" ~ price*0.03,
     cut=="Ideal" ~ price*0.03,
+ ),new.price=price-discount) %>% filter(color=="E" & clarity=="SI2" & new.price==min(new.price))
# A tibble: 1 \times 12
 carat cut color clarity depth table price
                                            х у
                                                        z discount new.price
 <dbl> <ord> <ord> <ord> <dbl> <int> <dbl> <dbl> <dbl> <dbl>
                                                           <db1>
                                                                       <db1>
1 0.23 Ideal E
                  SI2
                          61.5
                                  55 326 3.95 3.98 2.43
                                                              9.78
                                                                        316.
> diamonds %>% mutate(discount= case_when(
     cut=="Fair" ~ price*0.1,
     cut=="Good" ~ price*0.12,
     cut=="Very Good" ~ price*0.15,
     cut=="Premium" ~ price*0.18,
     cut=="Ideal" ~ price*0.18,
+ ),new.price=price-discount) %>% filter(color=="E"& clarity=="SI1") %>% arrange(new.price)
# A tibble: 2,426 × 12
                 color clarity depth table price
   carat cut
                                                          z discount new.price
                                                    У
                 <ord> <ord> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
   <dbl> <ord>
 1 0.21 Premium E
                    SI1
                              59.8
                                      61 326 3.89 3.84 2.31
                                                                  58.7
                                                                           267.
                               62
 2 0.26 Very Good E
                      SI1
                                      54 384 4.08 4.11 2.54
                                                                  57.6
                                                                           326.
                                      60 373 4.12 4.15 2.59
                 E
                      SI1
                               62.6
                                                                  44.8
 3 0.28 Good
                                                                           328.
                                      57 407 4.2 4.23 2.58
58 471 4.32 4.35 2.72
 4 0.27 Very Good E
                      SI1
                               61.2
                                                                  61.0
                                                                           346.
 5 0.31 Premium E
                      SI1
                               62.7
                                                                           386.
                                      56 486 4.01 3.99 2.5
 6 0.24 Ideal
                 Ε
                       SI1
                               62.5
                                                                  87.5
                                                                           399.
                                      58 499 4.27 4.29 2.66
 7 0.3 Ideal
                 Ε
                      SI1
                               62.1
                                                                  89.8
                                                                           409.
 8 0.3 Ideal
                               61.1
                                      57 499 4.3 4.34 2.64
                                                                  89.8
                                                                           409.
                Ε
                      SI1
                                      57 499 4.27 4.29 2.67
 9 0.3 Ideal
                Е
                       SI1
                               62.4
                                                                  89.8
                                                                           409.
10 0.3 Ideal
                Ε
                                      54 499 4.32 4.35 2.67
                                                                  89.8
                               61.6
                                                                           409.
# i 2,416 more rows
# i Use `print(n = ...)` to see more rows
 > cl<-diamonds$clarity %>% as.factor()
 > cl %>% levels()
 [1] "I1"
                                              "VS1" "VVS2" "VVS1" "IF"
                 "SI2"
                            "SI1"
                                    "VS2"
> cl %>% nlevels()
Γ17 Q
```

Elmasları kesim türüne göre gruplar ve her kesim türü için ortalama fiyatı hesaplar.

```
> diamonds %>% group_by(cut) %>% summarise(mean.pr=mean(price))
# A tibble: 5 \times 2
  cut
             mean.pr
  <ord>
                <dbl>
1 Fair
                4359.
2 Good
                3929.
3 Very Good
                <u>3</u>982.
4 Premium
                <u>4</u>584.
5 Ideal
                3458.
```