

Veri manipölasyonu

ilke

2022-07-27

```
install.packages("tidyverse")
```

```
## Installing package into '/home/ilke/R/x86_64-pc-linux-gnu-library/3.4'  
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## — Attaching packages
```

```
## —  
## tidyverse 1.3.2 —
```

```
## ✓ ggplot2 3.3.6   ✓ purrr  0.3.4  
## ✓ tibble  3.1.7   ✓ dplyr  1.0.9  
## ✓ tidyr   1.2.0   ✓ stringr 1.4.0  
## ✓ readr   2.1.2   ✓ forcats 0.5.1  
## — Conflicts — tidyverse_conflicts() —  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag()   masks stats::lag()
```

```
data()  
View(iris)  
head(iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1      5.1      3.5      1.4      0.2 setosa  
## 2      4.9      3.0      1.4      0.2 setosa  
## 3      4.7      3.2      1.3      0.2 setosa  
## 4      4.6      3.1      1.5      0.2 setosa  
## 5      5.0      3.6      1.4      0.2 setosa  
## 6      5.4      3.9      1.7      0.4 setosa
```

```
irisdata<- iris
```

alt veri kümesi çekmek için select()

Kayıp gözleme sahip olan satırları devre dışı bırakmak için na.omit() fonksiyonunu

mutate() fonksiyonu ilgili gözlemler üzerinde değişiklik yapmamızı sağlar

group_by() gruplama yapar

summarise() fonksiyonu ile raporlanır.

arrange() fonksiyonu satırları/gözlemleri içeriklerine göre sıralamak için kullanılır. Böylece veri tabanınızı hem istenilen kriterde sıralamış, hem de sadece verinin üst kısmına bakarak değişkenler arasındaki ilişkiyi görme şansını elde etmiş olursunuz.

Petal.Length>0.1 için çanak yaprak ve taç yaprak uzunluğu ortalamaları türlere göre raporlandı.(tür adetleri de rapor edildi)

```
irisdata %>%  
  select(Sepal.Length,Petal.Length,Species) %>%  
  filter(Petal.Length>0.1) %>%  
  group_by(Species) %>%  
  summarise(n = n(),  
            sepalort=mean(Sepal.Length),  
            petalort=mean(Petal.Length))
```

```
## # A tibble: 3 × 4  
##   Species      n sepalort petalort  
##   <fct>    <int>   <dbl>   <dbl>  
## 1 setosa     50    5.01    1.46  
## 2 versicolor 50    5.94    4.26  
## 3 virginica  50    6.59    5.55
```

HairEyeColor veri setine mutate() ile oran bilgisi eklendi

```
data<- data.frame(HairEyeColor)
```

```
veri1<- mutate(data,  
  oran = Freq / sum(Freq))
```

```
veri1
```

```
##   Hair Eye Sex Freq   oran  
## 1 Black Brown Male 32 0.054054054  
## 2 Brown Brown Male 53 0.089527027  
## 3 Red Brown Male 10 0.016891892  
## 4 Blond Brown Male 3 0.005067568  
## 5 Black Blue Male 11 0.018581081  
## 6 Brown Blue Male 50 0.084459459  
## 7 Red Blue Male 10 0.016891892  
## 8 Blond Blue Male 30 0.050675676  
## 9 Black Hazel Male 10 0.016891892  
## 10 Brown Hazel Male 25 0.042229730  
## 11 Red Hazel Male 7 0.011824324  
## 12 Blond Hazel Male 5 0.008445946  
## 13 Black Green Male 3 0.005067568  
## 14 Brown Green Male 15 0.025337838  
## 15 Red Green Male 7 0.011824324  
## 16 Blond Green Male 8 0.013513514  
## 17 Black Brown Female 36 0.060810811  
## 18 Brown Brown Female 66 0.111486486  
## 19 Red Brown Female 16 0.027027027  
## 20 Blond Brown Female 4 0.006756757  
## 21 Black Blue Female 9 0.015202703  
## 22 Brown Blue Female 34 0.057432432  
## 23 Red Blue Female 7 0.011824324  
## 24 Blond Blue Female 64 0.108108108  
## 25 Black Hazel Female 5 0.008445946  
## 26 Brown Hazel Female 29 0.048986486  
## 27 Red Hazel Female 7 0.011824324  
## 28 Blond Hazel Female 5 0.008445946  
## 29 Black Green Female 2 0.003378378  
## 30 Brown Green Female 14 0.023648649  
## 31 Red Green Female 7 0.011824324  
## 32 Blond Green Female 8 0.013513514
```

irisdata veri setinden Sepal kelimesi içeren değişkenleri iris2 alt veri setine atadık.

head ile iris2 alt veri setinin 6 satırını önizledik.

```
iris2<- irisdata %>%  
  select(contains("Sepal"))
```

```
head(iris2,6)
```

```
##   Sepal.Length Sepal.Width  
## 1      5.1      3.5  
## 2      4.9      3.0  
## 3      4.7      3.2  
## 4      4.6      3.1  
## 5      5.0      3.6  
## 6      5.4      3.9
```

irisdata veri seti, türlere göre gruplama yapılarak kaçar tane olduğu iris_group'a atandı.

```
iris_group<- irisdata %>% group_by(Species) %>%  
  summarise(n = n())
```

```
head(iris_group)
```

```
## # A tibble: 3 × 2  
##   Species     n  
##   <fct>   <int>  
## 1 setosa    50  
## 2 versicolor 50  
## 3 virginica 50
```

irisdata veri seti Sepal.Length için kaçar tane olduğu ve toplamaları

```
iris_sepal_length<- irisdata %>% group_by(Sepal.Length) %>%  
  summarise(n = n(),  
            topla= sum(Sepal.Length))
```

```
head(iris_sepal_length, 5)
```

```
## # A tibble: 5 × 3  
##   Sepal.Length  n topla  
##   <dbl> <int> <dbl>  
## 1     4.3     1   4.3  
## 2     4.4     3  13.2  
## 3     4.5     1   4.5  
## 4     4.6     4  18.4  
## 5     4.7     2   9.4
```