

# NBA, FIFA ve HARRY POTTER GRAFİKLEŞTİRME ve YORUMLAMA

Melih Gündüz

27.11.2022

## ÖZET

Bu raporda nba, fifa ve harry potter veri setleri kullanılmıştır. Nba veri setinde pozisyonlara göre oyuncu sayıları, atılan sayıların oyuncuların oynadıkları pozisyonlara göre oranı ve atılan sayıların, oyuncu pozisyonlarına göre yaş gruplandırması oranı görselleştirilmiştir. Fifa veri setinde oyuncu sayılarının en iyi oynadıkları pozisyona göre oranı, oyuncu sayılarının kontrat sürelerine göre oranı ve oyuncu sayılarının en iyi oynadıkları pozisyona ve kullandıkları ayaklarına göre oranı görselleştirilmiştir. Harry Potter veri setinde karakterlerin cinsiyetlerine göre oranı, karakterlerin evlerine göre oranı ve karakterlerin evlerine ve muggle olması ve ya diğerler durumlarına göre oranı görselleştirilmiştir.

Gerekli veri paketleri indirilmiştir ve kütüphaneye çağrılmıştır.

```
install.packages("ggplot2")
install.packages("tidyverse")
install.packages("dplyr")
install.packages("MetBrewer")
install.packages("plotrix")
install.packages("ggmosaic")
install.packages("treemapify")
library(ggplot2)
library(tidyverse)
library(dplyr)
library(MetBrewer)
library(plotrix)
library(ggmosaic)
library(treemapify)
```

# 1) 2022-2023 NBA OYUNCU İSTATİSTİKLERİ

Nba veri setinde oyuncular, pozisyon, takım, serbest atışlar, yaş, vb. veriler gözlenmiştir.

```
NBA <- read_delim("NBA.csv", delim = ";",  
  escape_double = FALSE, trim_ws = TRUE)
```

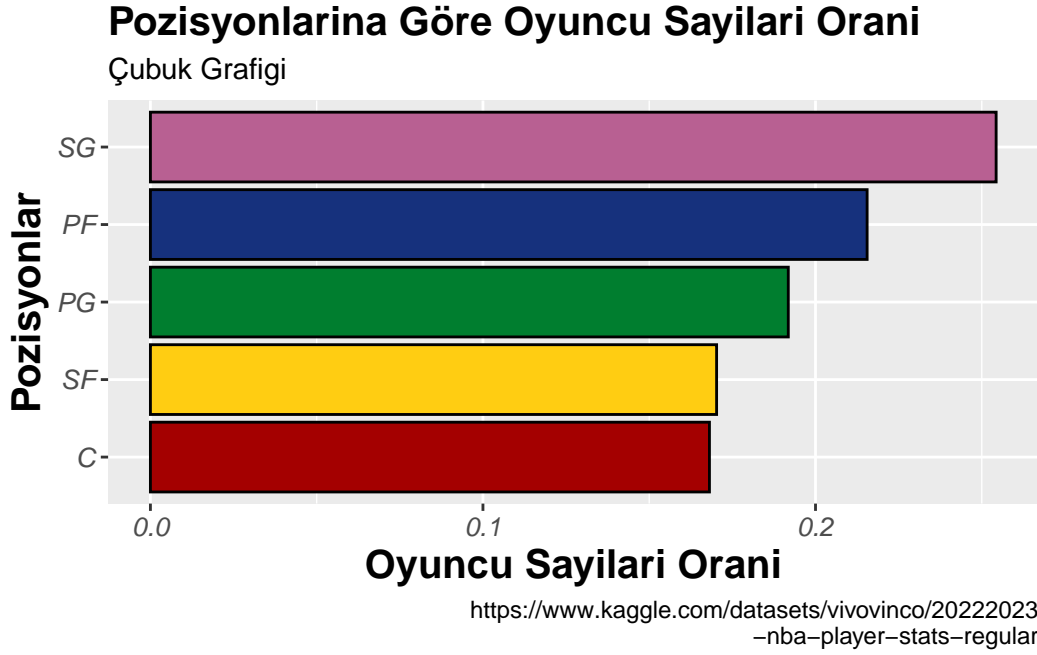
## 1.1) NBA OYUNCU POZİSYONLARINA GÖRE SAYILARI

Burada ki nba veri setinde pozisyon verisi alınarak oranı hesaplanmıştır.

```
nba <- NBA %>%  
  group_by(Pos) %>%  
  summarise(sayı=n()) %>%  
  mutate(oran= sayı/ nrow(NBA))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(nba, aes(reorder(x=nba$Pos, +nba$oran), y=nba$oran, fill=nba$Pos))+  
  geom_bar(stat="identity", color="black")+  
  labs(x="Pozisyonlar",  
    y="Oyuncu Sayilari Orani",  
    title="Pozisyonlarına Göre Oyuncu Sayilari Orani",  
    subtitle = "Çubuk Grafigi",  
    caption = "https://www.kaggle.com/datasets/vivovinco/20222023  
-nba-player-stats-regular")+  
  scale_fill_manual(values=met.brewer("Austria"))+  
  theme(legend.position = "none",  
    plot.title = element_text(size = 15, face="bold"),  
    axis.text = element_text(size=10, face="italic"),  
    axis.title = element_text(size=15, face = "bold"))+  
  coord_flip()
```



## YORUM

Yukarıdaki grafikte nba veri setinde pozisyonlarına göre oyuncu sayıları oranı grafiği çizilmiştir. Beş pozisyondan oluşmaktadır. Çubuk grafiği ile çizilmiştir. “SG” pozisyonu en yüksek orana sahip pozisyon olarak görülmüştür. “C” pozisyonu en düşük orana sahip pozisyon olarak görülmüştür. “C” pozisyonu ile “SF” pozisyonu arasında oran farkı diğerlerine göre daha az görülmüştür.

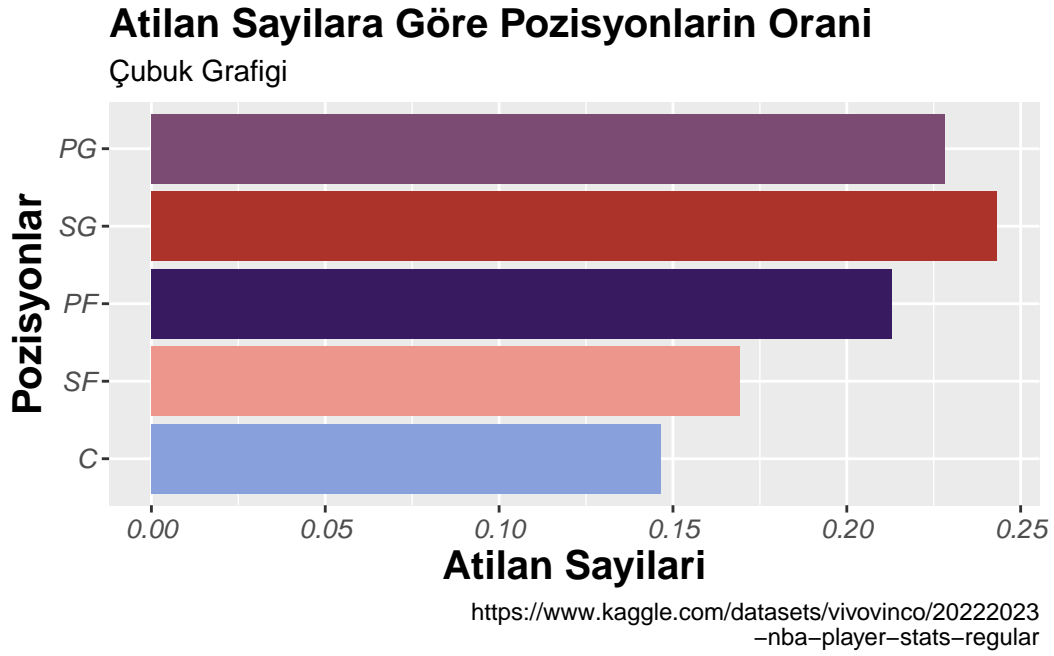
## 1.2) ATILAN SAYILARIN OYUNCULARIN OYNADIKLARI POZİSYONA GÖRE ORANI

Burada ki nba veri setinde pozisyon ve atılan sayılar verileri alınarak oranı hesaplanmıştır.

```
nba1 <- NBA %>%  
  group_by(Pos ,FG) %>%  
  summarise(sayı=n()) %>%  
  mutate(oran= sayı/ nrow(NBA))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(nba1, aes(x= nba1$FG/sum(nba1$FG),
                 y= reorder(nba1$Pos,+nba1$FG/sum(nba1$FG)), fill= nba1$Pos))+
  geom_bar(stat="identity")+
  labs(y="Pozisyonlar",
       x="Atılan Sayilari",
       title="Atılan Sayılara Göre Pozisyonların Oranı",
       subtitle = "Çubuk Grafiği",
       caption = "https://www.kaggle.com/datasets/vivovinco/20222023
-nba-player-stats-regular")+
  scale_fill_manual(values=met.brewer("Archambault"))+
  theme(legend.position = "none",
        plot.title = element_text(size = 15, face="bold"),
        axis.text = element_text(size=10, face="italic"),
        axis.title = element_text(size=15, face = "bold"))
```



## YORUM

Yukarıdaki grafikte nba veri setinde atılan sayıların oyuncuların oynadıkları pozisyonlarına göre oranı grafiği çizilmiştir. Beş pozisyonlu oluşmaktadır. Çubuk grafiği ile çizilmiştir. “SG” pozisyonu en fazla sayı atan olarak görülmüştür. “C” pozisyonu en az sayı atan olarak görülmüştür.

### 1.3) ATILAN SAYILARIN, OYUNCU POZİSYONLARI VE YAŞ GRUBUNA GÖRE ORANI

Burada ana veri setinden başka ana veri seti oluşturulmuştur. Daha sonrasında yaş aralıklarını belirtilerek bir tablo oluşturulmuştur.

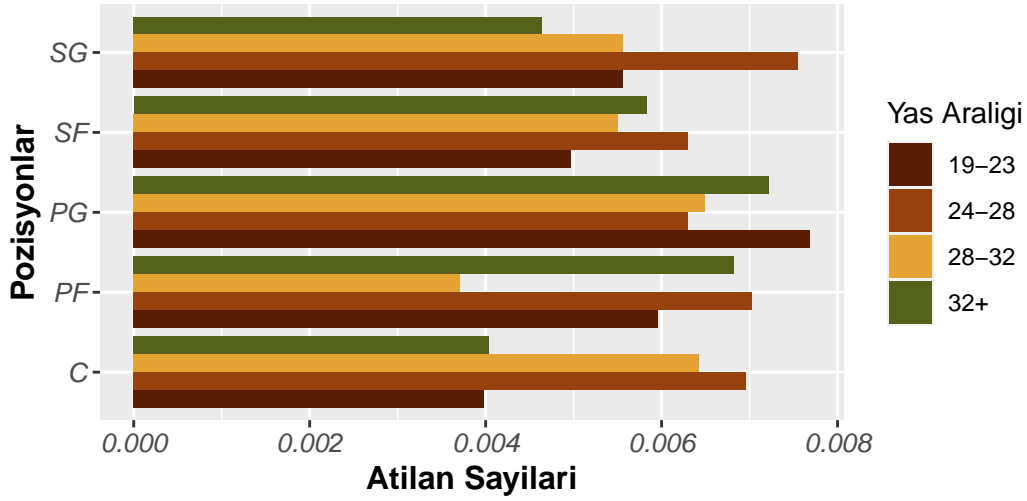
```
nba2 <- within(NBA, {  
  Age[19 <= Age & Age <=23] <- "19-23"  
  Age[24 <= Age & Age <=28] <- "24-28"  
  Age[28 <= Age & Age <=32] <- "28-32"  
  Age[32 <= Age ] <- "32+"  
})
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(nba2, aes(x= nba2$Pos, y=nba2$FG/sum(nba2$FG), fill=nba2$Age))+  
  geom_bar(stat="identity", position = "dodge")+  
  labs(x="Pozisyonlar",  
       y="Atılan Sayıları",  
       fill="Yaş Aralığı",  
       title="Atılan Sayıların Pozisyonlarına ve Yaş Aralığına Göre Oranı",  
       subtitle = "Çubuk Grafiği",  
       caption = "https://www.kaggle.com/datasets/vivovinco/20222023  
-nba-player-stats-regular")+  
  scale_fill_manual(values=met.brewer("Degas"))+  
  theme(plot.title = element_text(size = 13, face="bold"),  
        axis.text = element_text(size=10, face="italic"),  
        axis.title = element_text(size=12, face = "bold"))+  
  coord_flip()
```

## Atilan Sayıların Pozisyonlarına ve Yaş Aralığına Göre Oran

Çubuk Grafiği



<https://www.kaggle.com/datasets/vivovinco/20222023-nba-player-stats-regular>

### YORUM

Yukarıdaki grafikte nba veri setinde atılan sayıların pozisyonlarına ve yaş aralığına göre oranı grafiği çizilmiştir. Beş pozisyondan oluşmaktadır. Çubuk grafiği ile çizilmiştir ama gruplandırılmış yapılmıştır. “PG” pozisyonunda “19-23” yaş grubu en fazla sayı atan olarak görülmüştür. “PF” pozisyonunda “28-32” yaş grubu en az sayı atan olarak görülmüştür. “PG” pozisyonuna baktığımızda tüm yaş grupları diğer pozisyonlara göre daha fazla olduğu görülmektedir.

## 2) FIFA 23 OYUNCULARI

Fifa veri setinde oyuncu ismi, ülkeler, pozisyonları, değerleri, oynadıkları kulüpleri, kontratları vb. veriler gözlenmiştir.

```
library(readr)
fifa <- read_csv("fifa.csv")
```

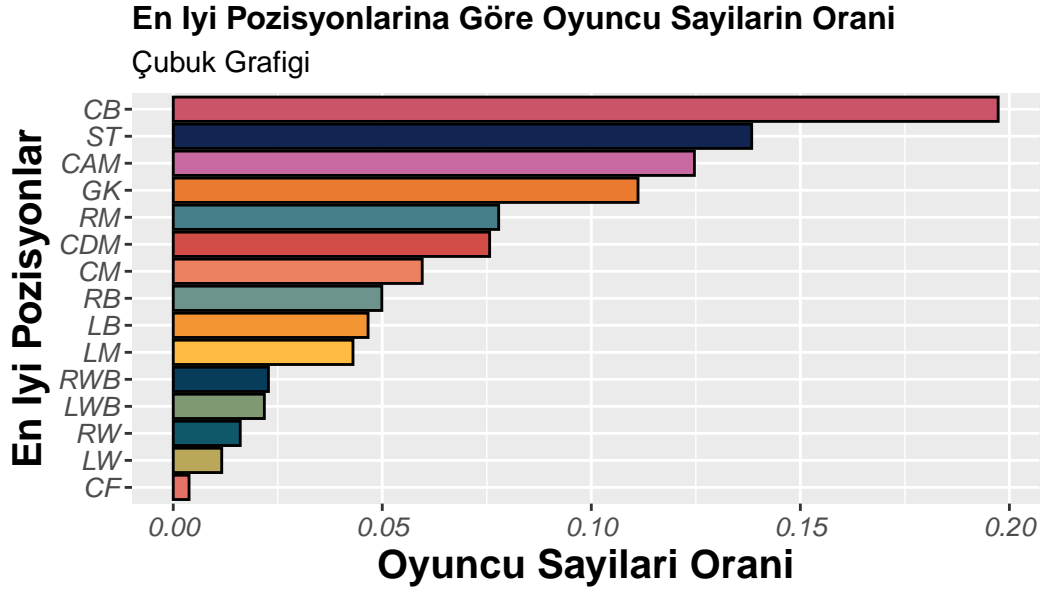
### 2.1) OYUNCU SAYILARININ EN İYİ POZİYONA GÖRE ORANLARI

Burada ki fifa veri setinde en iyi pozisyon verisi alınarak oranı hesaplanmıştır.

```
fifa1 <- fifa %>%
  group_by(`Best Position`) %>%
  summarise(sayi = n()) %>%
  mutate(oran = sayi / nrow(fifa))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(fifa1, aes(x=reorder(fifa1$`Best Position`, +fifa1$oran), y=fifa1$oran,
                        fill=fifa1$`Best Position`))+
  geom_bar(stat="identity", color="black")+
  labs(x="En İyi Pozisyonlar",
       y="Oyuncu Sayilari Orani",
       title="En İyi Pozisyonlarına Göre Oyuncu Sayilarin Orani",
       subtitle = "Çubuk Grafigi",
       caption = "https://www.kaggle.com/datasets/
/1898257a1915cbd22dd226f16c66dc218dfb65fd6f062755f4d0243bb8365a99")+
  scale_fill_manual(values=met.brewer("Cross", n=15))+
  theme(legend.position = "none",
        plot.title = element_text(size = 12, face="bold"),
        axis.text = element_text(size=10, face="italic"),
        axis.title = element_text(size=15, face = "bold"))+
  coord_flip()
```



<https://www.kaggle.com/datasets/1898257a1915cbd22dd226f16c66dc218dfb65fd6f062755f4d0243bb8365a99>

## YORUM

Yukarıdaki grafikte fifa veri setinde oyuncuların oynadıkları en iyi pozisyona göre oranı grafiği çizilmiştir. On beş pozisyondan oluşmaktadır. Çubuk grafiği ile çizilmiştir. “CB” pozisyonu en yüksek orana sahip pozisyon olarak görülmüştür. “CF” pozisyonu en düşük orana sahip pozisyon olarak görülmüştür. “RWB” pozisyonu ile “LWB” pozisyonu arasında oran farkı diğerlerine göre daha az görülmüştür.



## 2.2) OYUNCU SAYILARININ KONTRAT SÜRELERİNE GÖRE ORANLARI

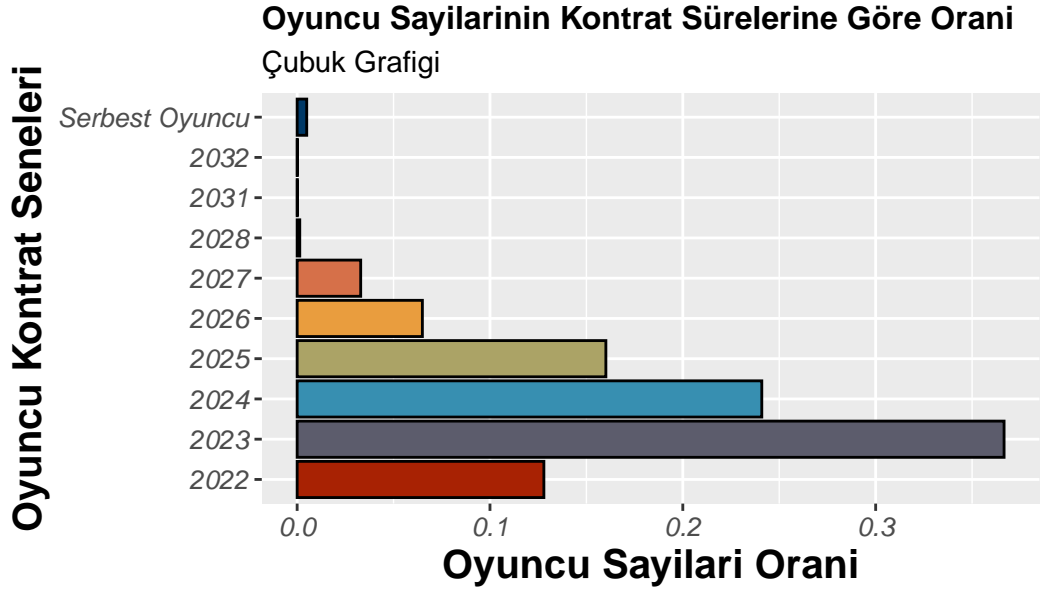
Burada ki fifa veri setinde kontrat süreleri verisi alınarak oranı hesaplanmıştır.

```
fifa2 <- fifa %>%
  group_by(`Contract Until`) %>%
  summarise(sayi = n()) %>%
  mutate(oran = sayi / sum(sayi))

fifa2 <- fifa2 %>%
  add_column(contrat = if_else(fifa2$`Contract Until` == "-",
                              "Serbest Oyuncu", fifa2$`Contract Until`))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(fifa2, aes(x=fifa2$contrat,
                  y=fifa2$oran, fill= fifa2$contrat))+
  geom_bar(stat="identity", color="black")+
  labs(x="Oyuncu Kontrat Seneleri",
       y="Oyuncu Sayilari Orani",
       title="Oyuncu Sayilarinin Kontrat Sürelerine Göre Orani",
       subtitle = "Çubuk Grafigi",
       caption = "https://www.kaggle.com/datasets/1898257a1915cbd22dd226f16c66dc218dfb65fd6f062755f4d0243bb8365a99")+
  scale_fill_manual(values=met.brewer("Juarez", n=10))+
  theme(legend.position = "none",
        plot.title = element_text(size = 12, face="bold"),
        axis.text = element_text(size=10, face="italic"),
        axis.title = element_text(size=15, face = "bold"))+
  coord_flip()
```



<https://www.kaggle.com/datasets/1898257a1915cbd22dd226f16c66dc218dfb65fd6f062755f4d0243bb8365a99>

## YORUM

Yukarıdaki grafikte fifa veri setinde oyuncuların sayılarının kontrat süresine göre oranı grafiği çizilmiştir. Dokuz tane kontrat süresi gözükmemektedir ve bunlar harici bir tane “serbest oyuncu” görülmüştür. “2023” yılında en fazla kontrat süresi görülmüştür. “2032” ve “2031” yıllarında en az kontrat süresi görülmüştür. “Serbest Oyuncu” da grafikte belli bir oranda görülmektedir.

## 2.3) OYUNCU SAYILARININ EN İYİ OYNADIKLARI POZİSYONA VE KULLANDIKLARI AYAKLARINA GÖRE ORANLARI

Burada ki nba veri setinde en iyi pozisyon ile oyuncu sayıları verileri alınarak oranı hesaplanmıştır.

```
fifa3 <- fifa %>%
  group_by(`Best Position`, `Preferred Foot`) %>%
  summarise(sayi = n()) %>%
  mutate(oran = sayi / nrow(fifa))
```

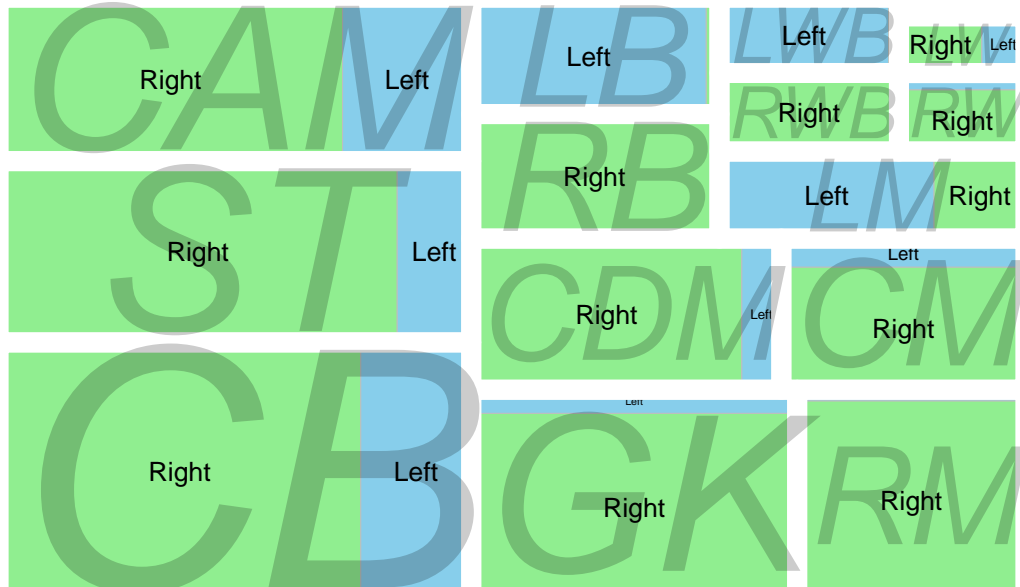
Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```

ggplot(fifa3, aes(area=fifa3$oran,
                  fill= fifa3$`Preferred Foot`,
                  label=fifa3$`Preferred Foot`,
                  subgroup=fifa3$`Best Position`))+
  geom_treemap()+
  scale_fill_manual(values = c("skyblue","lightgreen"))+
  geom_treemap_text(colour="black",
                   place="centre",
                   size= 10)+
  labs(title= "Oyuncu Sayilarin En Iyi Pozisyonlari ve Kullanilan Ayagina Göre Orani")+
  geom_treemap_subgroup_border(colour="white",
                              size=10)+
  geom_treemap_subgroup_text(place="centre",
                             grow=TRUE,
                             alpha=0.20,
                             colour="black",
                             fontface="italic")+
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face="bold"))

```

#### Oyuncu Sayilarin En Iyi Pozisyonlari ve Kullanilan Ayagina Göre Orani



## YORUM

Yukarıdaki grafikte fifa veri setinde oyuncuların oynadıkları en iyi pozisyona ve kullandıkları ayaklarına göre oranı grafiği çizilmiştir. Onbeş pozisyondan oluşmaktadır. Ağaç grafiği ile çizilmiştir. “CAM”, “ST” VE “CB” pozisyonları oranları diğer pozisyonlara göre daha fazla olduğu görülmüştür. “CF” pozisyonu sağ üstte kalan küçük bir alanda olduğu ve en az orana sahip olduğu görülmüştür. “RB” pozisyonunda sağ ayak kullanım oranı sol ayak kullanımına göre daha fazla olduğu görülmüştür. “LB” pozisyonunda sol ayak kullanımı oranı sağ ayak kullanımına göre daha fazla olduğu görülmüştür. Tüm pozisyonlarda sağ ayak ve sol ayak kullanımı görülmüştür.

## 3) HARRY POTTER

Harry Potter veri setinde isim, cinsiyet, meslek, ev vb. veriler gözlenmiştir.

```
library(readr)
harry <- read_delim("harry.csv", delim = ";",
  escape_double = FALSE, trim_ws = TRUE)
```

### 3.1) KARAKTERLERİN CİNSİYETLERİNE GÖRE ORANLARI

Burada ki harry potter veri setinde cinsiyet verisi alınarak oranı hesaplanmıştır.

```
harry1 <- harry %>%
  drop_na(Gender)%>%
  group_by(Gender) %>%
  summarise(sayi = n()) %>%
  mutate(oran = sayi / sum(sayi))
```

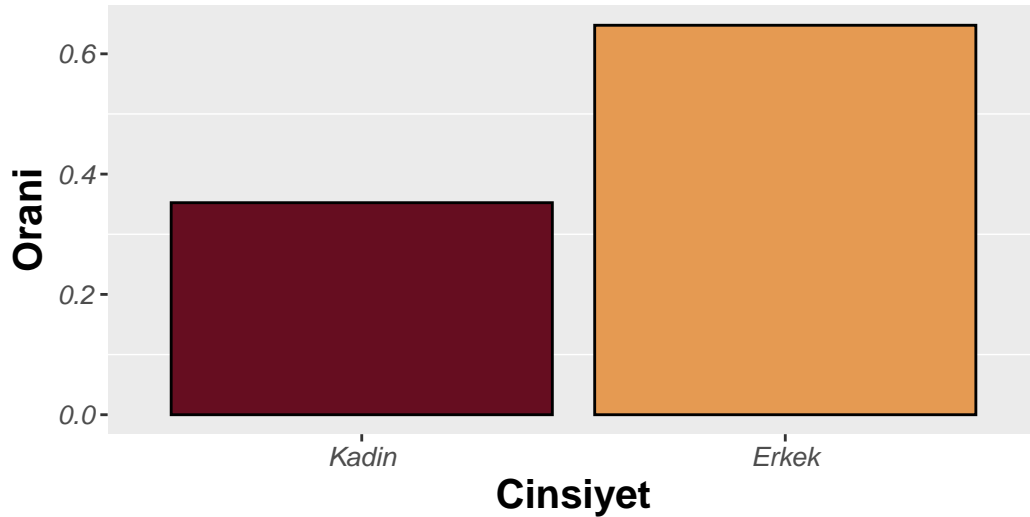
Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(harry1, aes(x=Gender, y=oran, fill=Gender))+
  geom_bar(stat="identity", color="black")+
  labs(x="Cinsiyet",
    y="Orani",
    title="Karakterin Cinsiyetlerine Orani",
    subtitle = "Çubuk Grafiği",
    caption = "https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset")+
  scale_fill_manual(values=met.brewer("Navajo"))+
```

```
scale_x_discrete(labels = c("Kadin", "Erkek"))+
theme(legend.position = "none",
      plot.title = element_text(size = 15, face = "bold"),
      panel.grid.major = element_blank(),
      axis.text = element_text(size = 10, face = "italic"),
      axis.title = element_text(size = 15, face = "bold"))
```

## Karakterin Cinsiyetlerine Orani

Çubuk Grafiği



<https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset>

## YORUM

Yukarıdaki grafikte harry potter veri setinde karakterlerin cinsiyetlerine göre oranı grafiği çizilmiştir. Erkek oranı kadınlara göre daha fazla görülmüştür. Çubuk grafiği ile çizilmiştir.

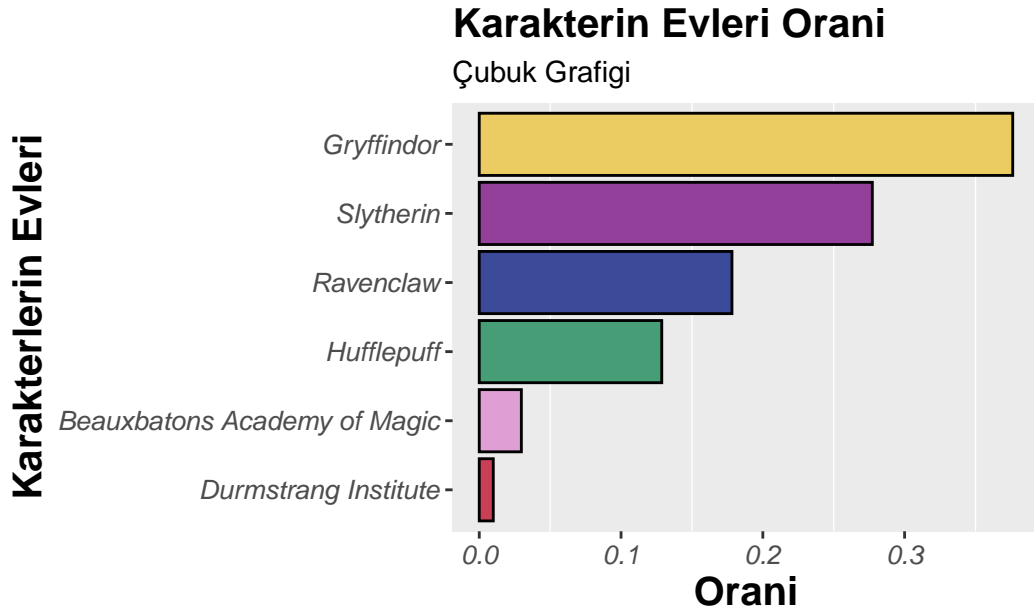
## 3.2) KARAKTERLERİN EVLERİNE GÖRE ORANLARI

Burada ki harry potter veri setinde ev verisi alınarak oranı hesaplanmıştır.

```
harry2 <- harry %>%
  drop_na(House)%>%
  group_by(House) %>%
  summarise(sayi = n()) %>%
  mutate(oran = sayi / sum(sayi))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(harry2, aes(x=reorder(House, +oran), y=oran, fill=House))+  
  geom_bar(stat="identity", color="black")+  
  labs(x= "Karakterlerin Evleri",  
       y="Orani",  
       title = "Karakterin Evleri Orani",  
       subtitle = "Çubuk Grafiği",  
       caption = "https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset")+  
  scale_fill_manual(values=met.brewer("Klimt", n=6), labels("Erkek", "Kadın"))+  
  theme(legend.position = "none",  
        plot.title = element_text(size = 15, face="bold"),  
        panel.grid.major = element_blank(),  
        axis.text = element_text(size=10, face="italic"),  
        axis.title = element_text(size=15, face = "bold"))+  
  coord_flip()
```



<https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset>

## YORUM

Yukarıdaki grafikte harry potter veri setinde karakterin evlerine göre oranı grafiği çizilmiştir. Altı tane ev çeşitti görülmektedir. Çubuk grafiği ile çizilmiştir. “Durmstrang Institute” evi en düşük orana sahip olarak görülmüştür. “Gryffindor” evi en yüksek orana sahip olarak görülmüştür.

### 3.3) KARAKTERLERİN EVLERİNE VE MUGGLE OLMASI VEYA DİĞERLERİNİN DURUMLARINA GÖRE ORANLARI

Burada ki harry potter veri setinde karakterlerin evleri ve muggle durumuna göre olan veriler alınarak oranı hesaplanmıştır.

```
harry <- harry %>%
  add_column(Muggle= if_else(harry$`Blood status` == "Muggle-born",
                             "Muggle olan", "Muggle olmayan"))

harry3 <- harry %>%
  group_by(House, Muggle) %>%
  drop_na(House, Muggle) %>%
  summarise(sayi= n())

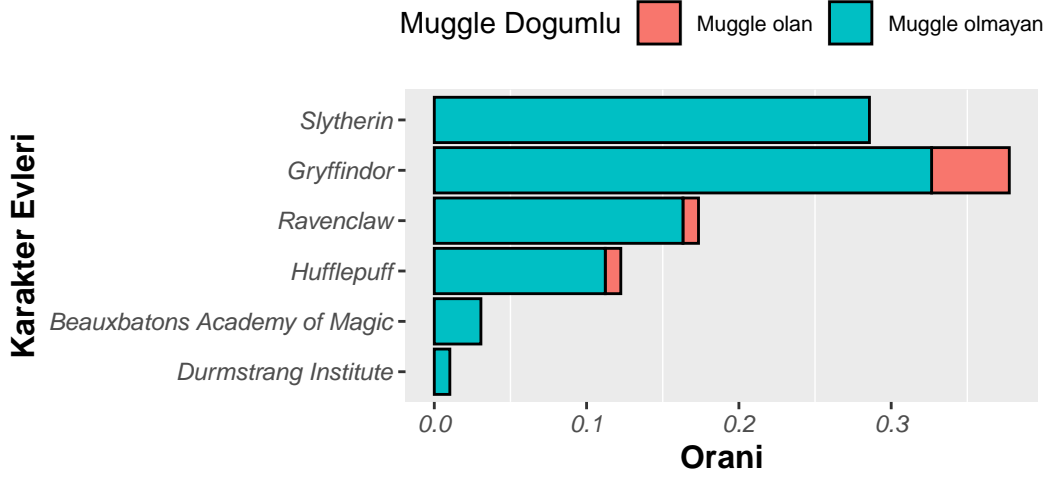
harry3 <- harry3 %>%
  mutate(oran= sayi / sum(harry3$sayi))
```

Burada “ggplot” yardımıyla uygun grafikleştirme yapılmıştır.

```
ggplot(harry3, aes(x= reorder(House, +oran), y= oran, fill= Muggle))+
  geom_bar(stat="identity", color="black")+
  labs(x="Karakter Evleri",
       y= "Orani",
       fill="Muggle Dogumlu",
       title="Karakterlerin Evlerine ve Muggle Olması
veya Olmaması Durumlarına Göre Orani",
       subtitle = "Çubuk grafiği",
       caption = "https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset")+
  theme(legend.position = "top",
        legend.text = element_text(size=8),
        plot.title = element_text(size = 11, face="bold"),
        panel.grid.major = element_blank(),
        axis.text = element_text(size=9, face="italic"),
        axis.title = element_text(size=12, face = "bold"))+
  coord_flip()
```

### Karakterlerin Evlerine ve Muggle Olması veya Olmaması Durumlarına Göre Oranı

Çubuk grafiği



<https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset>

### YORUM

Yukarıdaki grafikte harry potter veri setinde karakterlerin evlerine ve muggle olması ve olmamasına göre oranı grafiği çizilmiştir. Çubuk grafiği ile çizilmiştir. Muggle olan “Gryffindor”, “Hufflepuff” ve “Ravenclaw” bu evlerde görülmüştür. Muggle olan “Hufflepuff” ile “Ravenclaw” evlerin oranları eşit görülmüştür. Muggle olmayan tüm evlerde görülmüştür. Muggle olmayan “Durmstrang Institute” evi en düşük orana sahip olarak görülmüştür. Muggle olmayan “Gryffindor” evi en fazla orana sahip olarak görülmüştür.