

3 Farklı Veri Setinin İncelenip Görselleştirilmesi

Ahmet Batuhan Özdoğan

19.11.2022

Özet

Bu ödevde Kaggle sitesinden alınmış 3 farklı veri setinin görselleştirme ve yorumlama çalışmaları bulunmaktadır. Birinci veri seti Breaking Bad dizisi ,İkinci veri seti Friends dizisi ve Üçüncü veri seti ise En çok satan kitaplar olarak kullanılmıştır

Gerekli Kütüphanelerin yüklenmesi

```
install.packages("DALEX")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("gridExtra")
install.packages("readxl")
install.packages("tidyverse")
install.packages("ggforce")
install.packages("ggribes")
library(ggribes)
library(ggforce)
library(tidyverse)
library(gridExtra)
library(ggplot2)
library(DALEX)
library(dplyr)
library(readr)
```

Birinci Veri

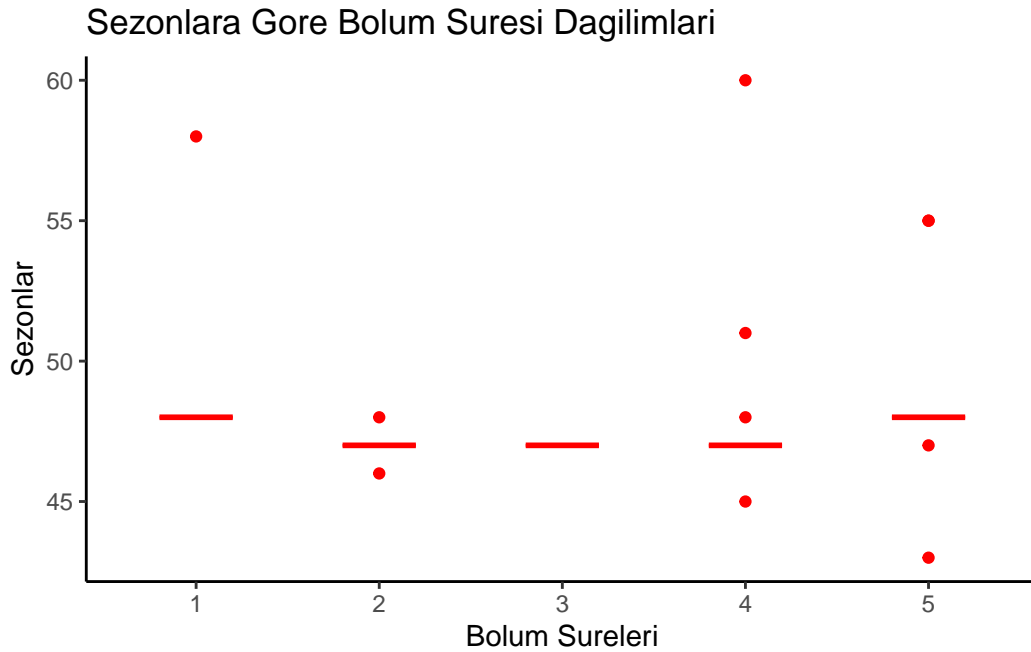
Veri Setinin Hakkında

Bu veri setimizde Breaking bad adlı dizimizin incelemesi ve görselleştirmesi bulunmaktadır

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

Breaking Bad Dizisinin Sezonlara Göre Bölüm Süresi Dağılımları

```
breaking_bad$Season <- factor(breaking_bad$Season, ordered = TRUE,
levels = c("1" , "2" , "3" , "4" , "5"))
ggplot(breaking_bad, aes(x = Season , y = Duration_mins)) +
geom_boxplot(color = "red" , width = 0.4 ) +
labs(y = "Sezonlar",
x = "Bolum Sureleri",
title = "Sezonlara Gore Bolum Suresi Dagilimleri") +
theme_classic()
```

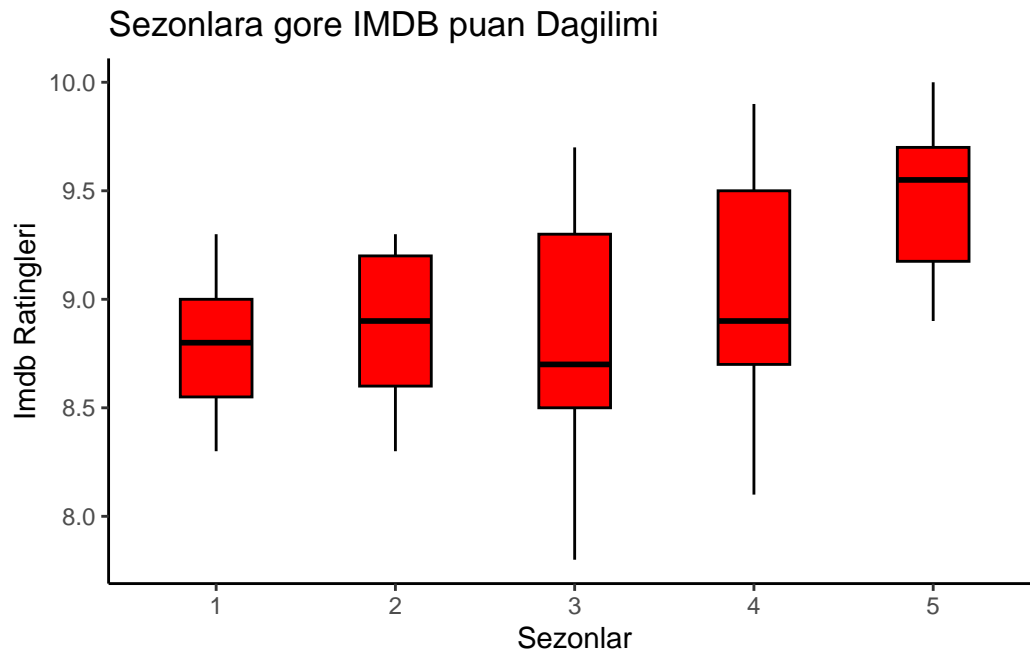


Rapor

Grafiğe Baktığımızda, bölüm süreleri arasında çok süre farkı gözlemlenmiyoruz , genel olarak baktığımızda her sezonda bölümlerin ortalama 45 dk olduğu gözlemlenmektedir

Breaking Bad Dizisinin Sezonlara göre IMDB puan Dagilimi

```
breaking_bad$Season <- factor(breaking_bad$Season, ordered = TRUE, levels = c("1", "2",  
breaking_bad %>%  
  ggplot(aes(x = Season, y = Rating_IMDB)) +  
  geom_boxplot(color = "black", fill = "red" , width = 0.4 ) +  
  labs(x="Sezonlar",  
        y= "Imdb Ratingleri",  
        title = "Sezonlara göre IMDB puan Dagilimi") +  
  theme_classic()
```

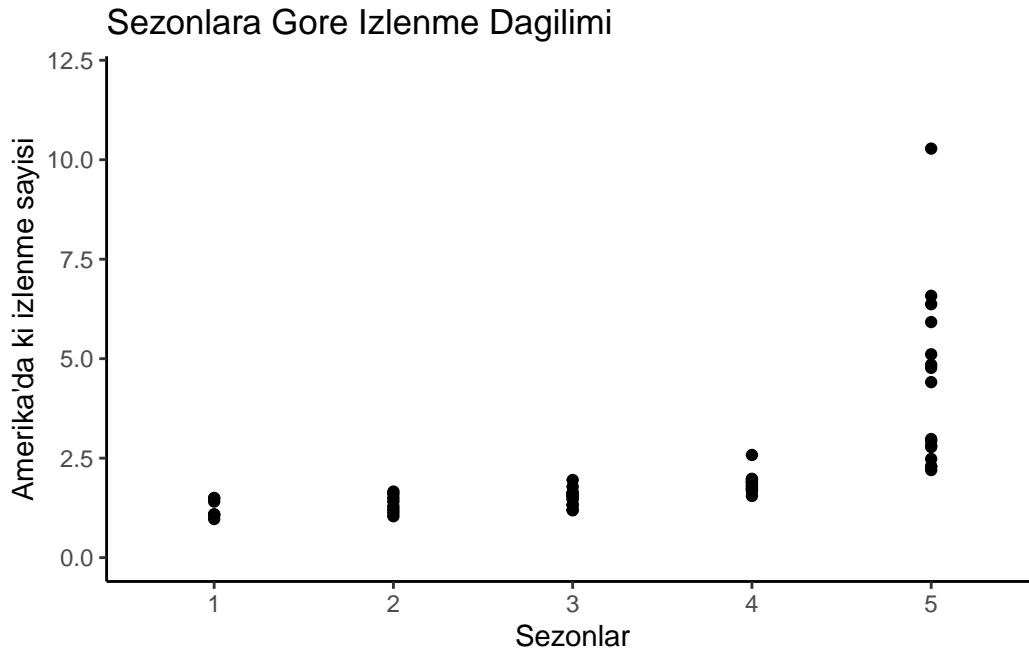


Rapor

Grafikteki sonuçlara göre, reytinglerin genel olarak 5 sezon boyunca yükselerek gittiğini söyleyebiliriz En düşük reyting değerinin 3. sezonda verildiğini ve en yüksek değer ise 5. sezonda elde edildiğini söyleyebiliriz

Sezonlara Göre İzlenme Dağılımı Grafiği

```
breaking_bad %>%  
ggplot(aes(x = Season, y = as.numeric(breaking_bad$`U.S. viewers_million`))) +  
geom_point() +  
labs(x="Sezonlar",  
y= "Amerika'da ki izlenme sayisi",  
title = "Sezonlara Gore Izlenme Dagilimi") +  
ylim(0,12)+  
theme_classic()
```



Rapor

Yukarıdaki Grafiğe göre izlenme sayısının en düşük olduğu sezon 1. sezon olup en yüksek olduğu sezon ise 5. sezon olarak gözükmemektedir izlenme sayısındaki en düşük değişimin ise 2. sezonda olduğu söylenebilmektedir

İkinci Veri

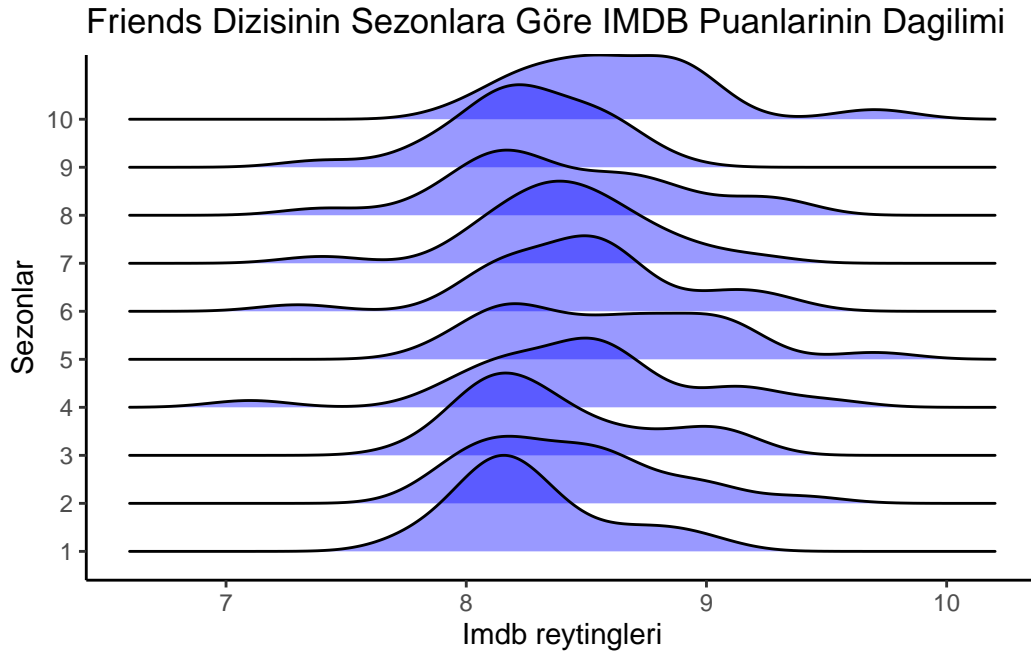
Veri Setinin Hakkında

Bu veri Setimiz ise Friends dizisi hakkında olup verilerin incelenmesi ve görselleştirilmesinden oluşmaktadır

```
library(readr)
fri <- read_csv("friends_imdb.csv")
```

Friends Dizisinin Sezonlara Göre IMDB Puanlarının Dagilimi grafigi

```
ggplot(fri, aes(x=imdb_rating, y=as.factor(season))) +
  geom_density_ridges(alpha=0.4 , scale=2 , fill="Blue")+
  labs(x = "Imdb reytingleri",
       y = "Sezonlar",
       title = "Friends Dizisinin Sezonlara Göre IMDB Puanlarının Dagilimi") +
  theme_classic()
```

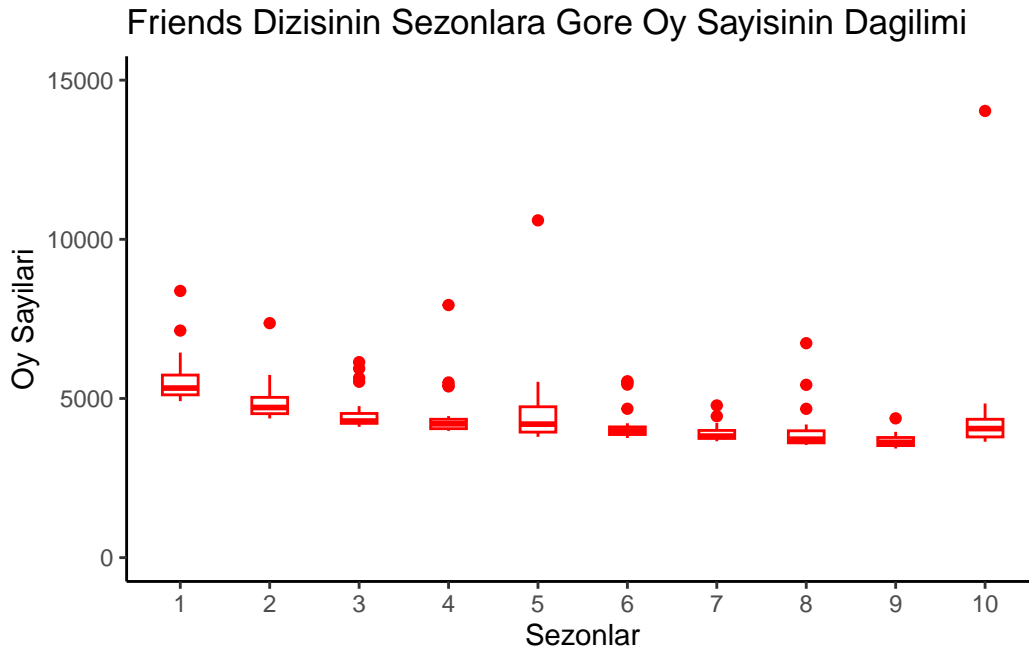


rapor

Friends adlı dizimizin sezonlarına göre IMDB puanlarına bakacak olursak en asimetrik değişim gösteren sezon 8. sezon olmuştur

Friends Dizisinin Sezonlara Gore Oy Sayisinin Dagilimi grafigi

```
ggplot(fri, aes(x=as.factor(season) ,y=total_votes))+  
  geom_boxplot(color = "red" , width = 0.4 ) +  
  labs(y = "Oy Sayilari",  
       x = "Sezonlar",  
       title = "Friends Dizisinin Sezonlara Gore Oy Sayisinin Dagilimi") +  
  ylim(1,15000)+  
  theme_classic()
```



Rapor

Friends Dizisinin Sezonlara Gore Oy Sayisinin Dagilimi grafigine bakilacak olursa en yüksek puan farklılığının 10. sezonda gerçekleştiğini söyleyebiliriz

Üçüncü veri

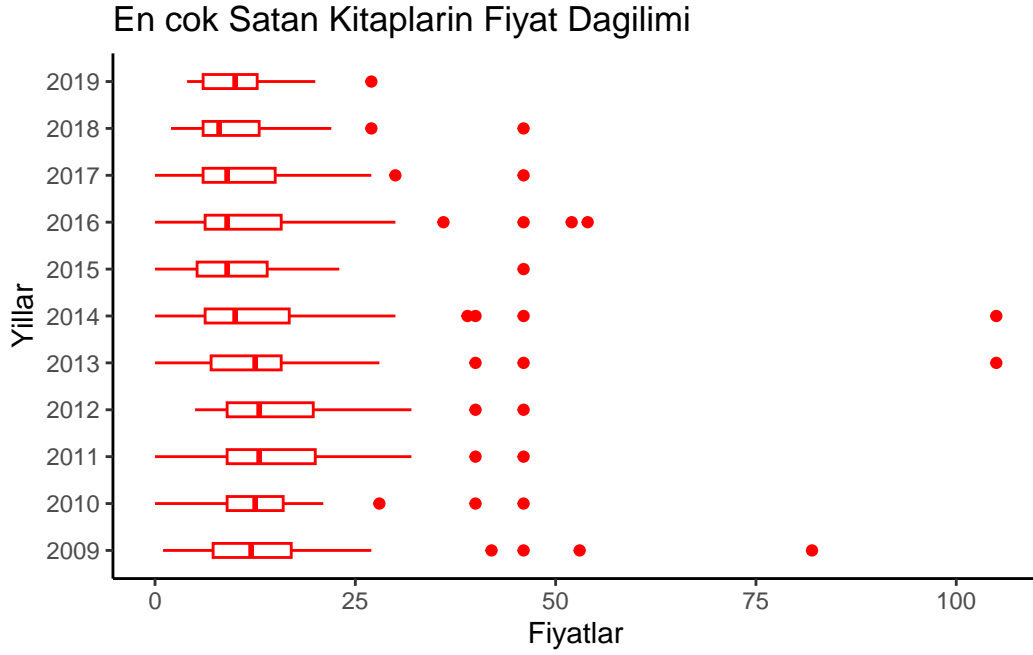
Veri Setinin Hakkında

Bu veri seti 2009-2019 yılları arasında Amazon'da en çok satan 50 kitabı içermektedir.

```
library(readr)
best <- read_csv("bestsellers with categories.csv")
```

En çok Satan Kitapların Fiyat Dağılımı Grafiği

```
best$Year <- factor(best$Year,
ordered = TRUE,
levels = c("2009" , "2010" , "2011" ,
"2012" , "2013" , "2014" ,
"2015" , "2016" , "2017" ,
"2018" , "2019"))
ggplot(best, aes(x = Price, y = Year)) +
geom_boxplot(color = "red" , width = 0.3 ) +
labs(y = "Yıllar",
x = "Fiyatlar",
title = "En çok Satan Kitapların Fiyat Dağılımı" ) +
theme_classic()
```

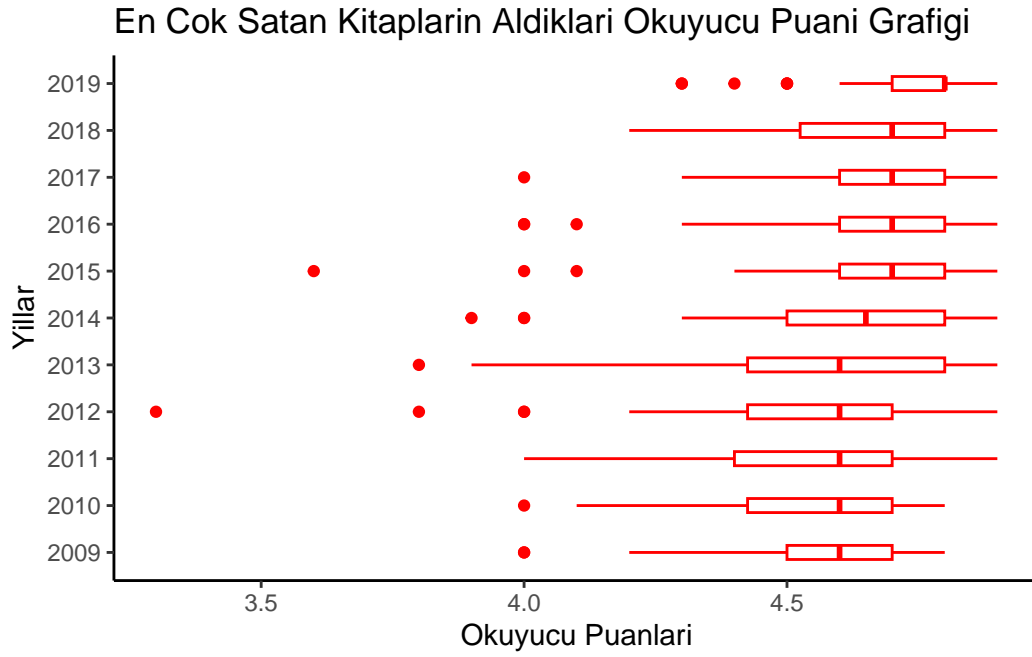


Rapor

En çok Satan Kitapların Fiyat Dağılımı Grafiğine bakacak olursak Fiyat dağılımı açısından en asimetrik olduğu yılların 2009,2013 ve 2014 yılları olduğunu söyleyebiliriz

En çok Satan Kitapların Aldıkları Okuyucu Puanı Grafiği

```
ggplot(best, aes(x = `User Rating`, y = Year)) +  
  geom_boxplot(color = "red" , width = 0.3 ) +  
  labs(y = "Yıllar",  
x = "Okuyucu Puanlari",  
title = "En Çok Satan Kitapların Aldıkları Okuyucu Puanı Grafiği " ) +  
  theme_classic()
```

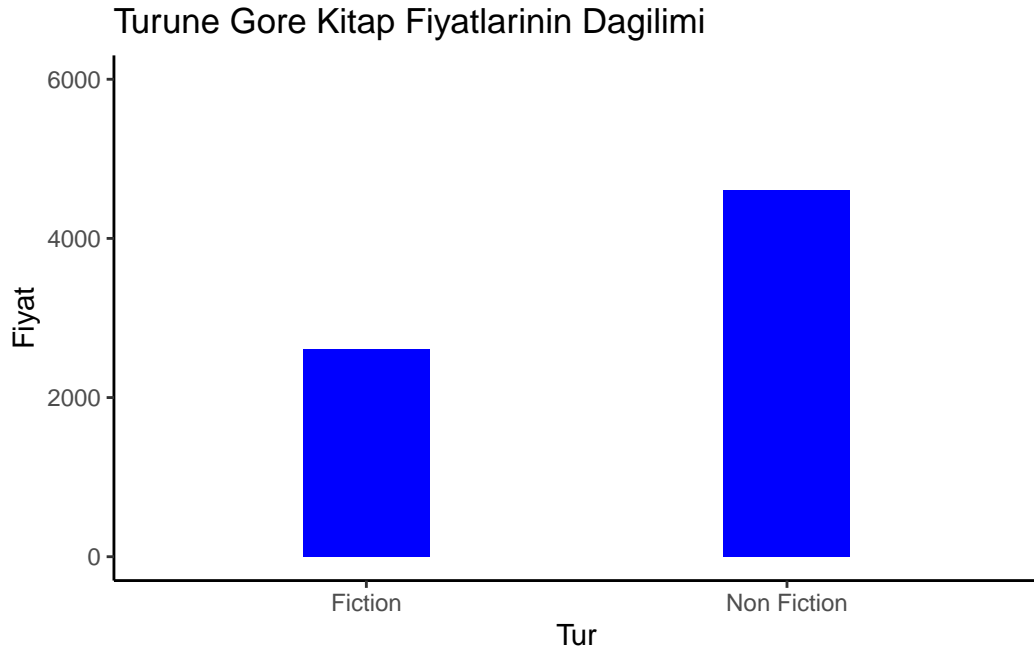



Rapor

En çok Satan Kitapların Aldıkları Okuyucu Puanı Grafiğine bakacak olursak genel olarak okuyucuların 4,5 puan verdiği söylenebilir

Türüne Göre Kitap Fiyatlarının Dağılımları Grafiği

```
ggplot(best, aes(x = Genre , y = Price)) +  
  geom_bar(stat="identity",fill="blue", width = 0.3)+  
  labs( x = "Tur" ,  
        y = "Fiyat",  
        title = "Turune Gore Kitap Fiyatlarinin Dagilimi" ) +  
  ylim(0,6000) +  
  theme_classic()
```



Rapor

Türüne Göre Kitap Fiyatlarının Dağılımları Grafiğine bakıcak olursak kurgusal olmayan kitapların genel olarak daha pahalı fiyatlara satıldığı söylenmektedir