

# **En Çok Satan Kitaplar, Marvel Filmleri ve Avengers Karikatürleri Grafik ve Yorumları**

Zeynep Afra SEZER

## **ÖZET**

Bu raporda 2009-2019 yılları arasında Amazonun en çok satan kitapları, Marvel filmleri ve Avengers karikatürleri veri setleri incelenip yorumlandırılmıştır. En çok satan kitaplar veri seti; fiyat dağılımlarına, aldıkları okuyucu puanlarına ve türlerine göre grafiklendirilmiştir. Marvel filmleri veri seti; filmlerin IMDB puanlarına ve süre dağılımlarına göre incelenmiştir. Avengers karikatürleri veri seti de; güncel karakter olup olmama durumlarına ve bu durumun cinsiyete göre etkisi incelenip grafiklendirilmiştir.

Bu rapor için kullanılan kütüphaneler;

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

## EN ÇOK SATAN KİTAPLAR

Bu veri setinde 2009-2019 yılları arasında Amazon'un en çok satan 50 kitabının veri kümesi verilmiştir. Veriler kurgu ve kurgusal olmayan olarak kategorize edilmiştir.

### 1.1 Yıllara göre en çok satan kitapların fiyat dağılımının değişimi

```
library(readr)
books <- read_excel("books.xlsx")
```

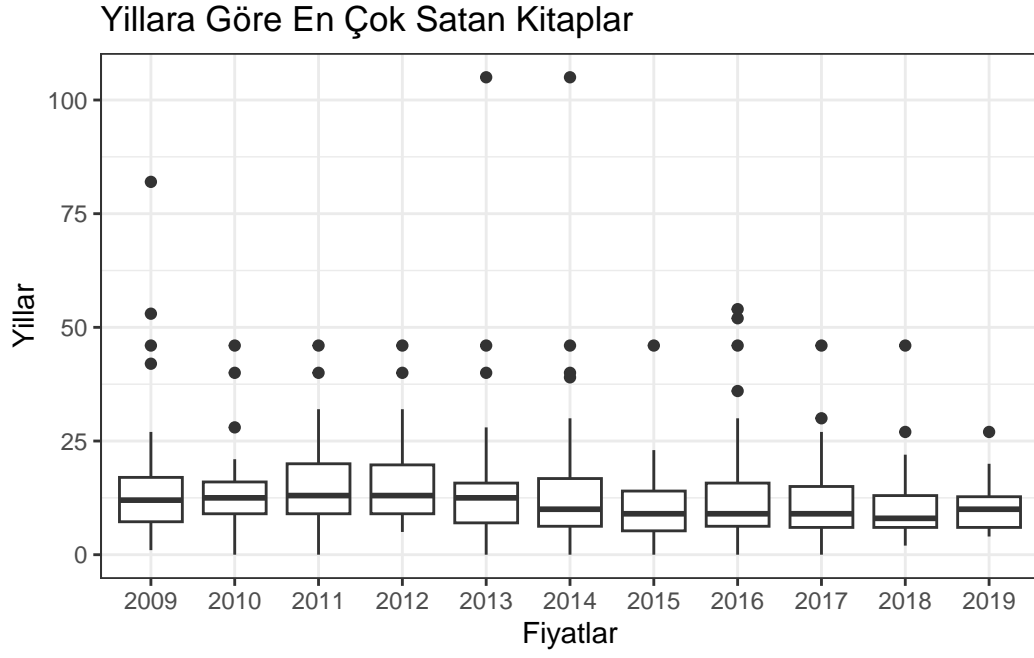
```
library(readr)
class(books$Year)
```

```
[1] "numeric"
```

```
library(readr)
books$Year <- as.factor(books$Year)
class(books$Year)
```

```
[1] "factor"
```

```
library(readr)
ggplot(books, aes(x= Year, y = Price)) +
  theme_bw() +
  geom_boxplot() +
  labs(x="Fiyatlar", y= "Yıllar",
       title = "Yıllara Göre En Çok Satan Kitaplar")
```

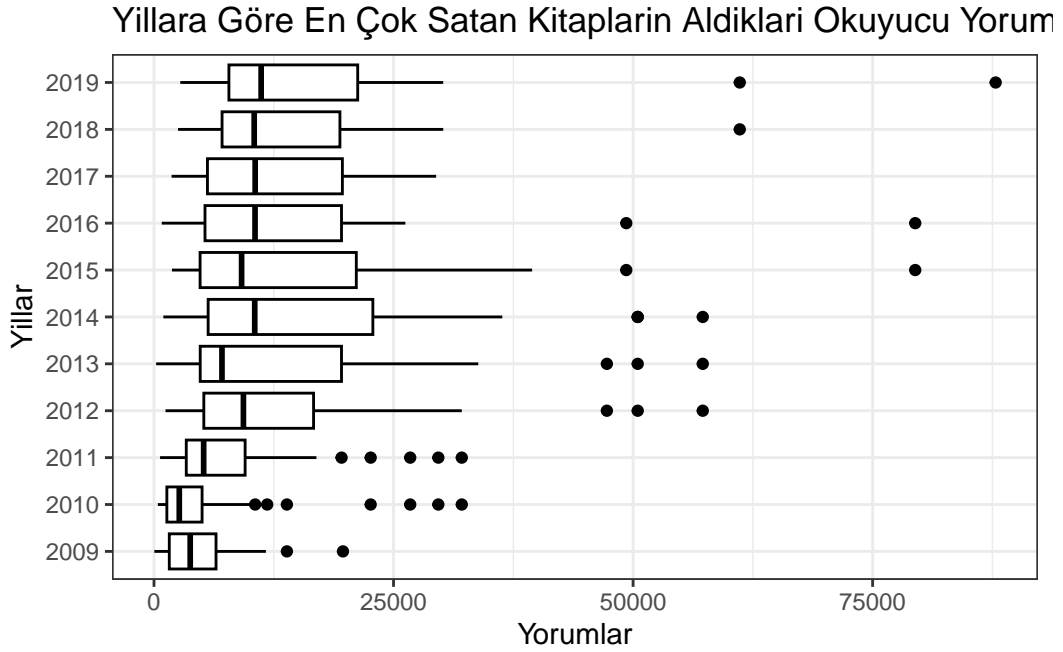


## YORUM

En az medyan 2018 yılındadır. 2013 ve 2014 yıllarında en yüksek aykırı değerler gözlenmiştir. En yayık veri bilgisi 2011 ve 2016 yılına aittir. Fiyatlar ağırlık olarak 0-50 değerler arasında gözlenmiştir. 100 değerinin üstündeki fiyat değerleri sadece 2011 ve 2016 yılında aykırı değerler olarak gözlenmiştir.

## 1.2 Yıllara göre en çok satan kitapların aldıkları okuyucu puanı dağılımının değişimi

```
library(readr)
ggplot(books, aes(x = Reviews, y = as.factor(Year))) +
  theme_bw() +
  labs(x = "Reviews", y = "Year") +
  geom_boxplot(color = "black")+
  labs(x="Yorumlar", y= "Yıllar",
       title = "Yıllara Göre En Çok Satan Kitapların Aldıkları Okuyucu Yorumları")
```



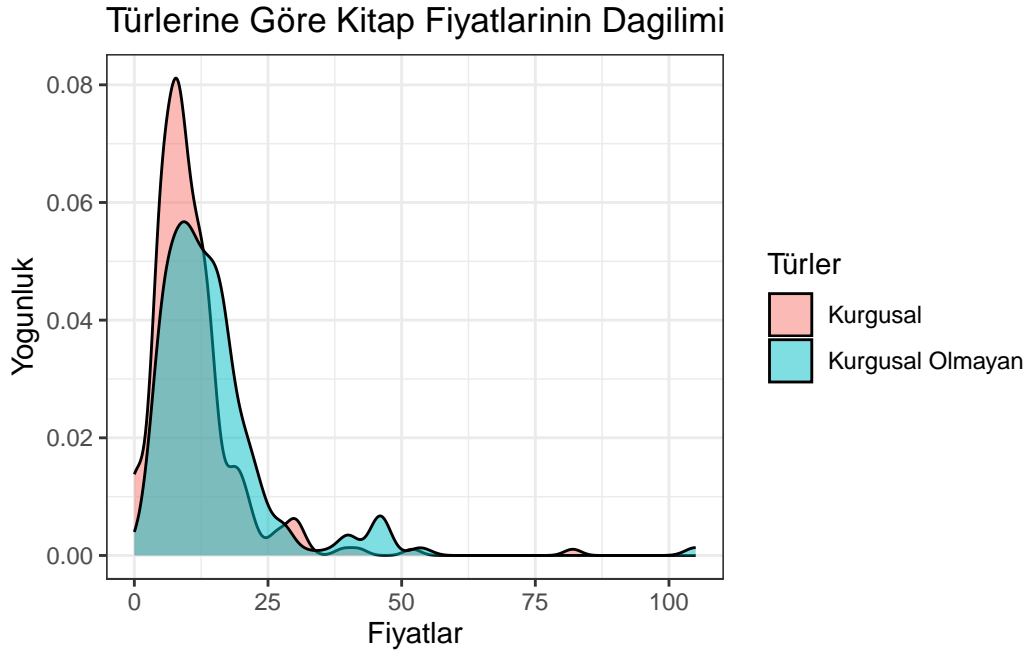
## YORUM

2009-2019 yılları arasında en çok okur puanı alan yıl 2015 yılıdır. En az veriye sahip yıl ise 2010 yılıdır. En az okur puanı bilgisi 2009-2011 yılları arasındadır Aynı zamanda en düşük medyana sahip yıl 2010 yılıdır. 7500 okur puanı değerinin üzerinde olan yıllar; 2015, 2016 ve 2019 yıllarıdır.

### 1.3 Türüne göre kitap fiyatlarının dağılımı

```
library(readr)
a <- books %>%
drop_na() %>%
filter(Genre %in% c("Non Fiction", "Fiction")) %>%
  group_by(Year, Genre) %>%
    summarise(Price)

ggplot (a, aes(x = a$Price, fill = a$Genre)) +
  theme_bw() +
  labs(x = "Price", y = "Density", fill = "Genre") +
  geom_density(alpha = 0.5) +
  labs(x="Fiyatlar", y= "Yogunluk", fill= "Türler",
    title= "Türlerine Göre Kitap Fiyatlarinin Dagilimi")+
  scale_fill_discrete(labels=c("Kurgusal","Kurgusal Olmayan"))
```



### YORUM

Kurgusal kitapların fiyatlarının ortalaması kurgusal olmayan kitapların fiyatlarının ortalamasından daha fazladır. Kurgu türü kitaplar 0-25 arası fiyat değerinde yoğunlaşırken, kurgu

dışı kitaplar ise 0-50 arası fiyat değerinde yoğunlaşmıştır. İki dağılım da sağa çarpıktır. Kurgusal kitapların dağılımında yaklaşık 80 değerinde aykırı değer vardır. Kurgusal olmayan kitapların dağılımında ise yaklaşık 110 değerinde aykırı değer vardır.

## MARVEL FİLMLERİ

Bu veri setinde 1944 - 2020 yılları arasındaki tüm Marvel filmlerinin verileri verilmiştir.

### 2.1 2000 yılı öncesi-2000 yılı ve sonrası Marvel filmlerinin IMDB puanlarının dağılımı

```
library(readr)
marvel <- read_excel("mdc.xlsx")

library(readr)
marvel_1 <- data.frame(marvel$year)

marvel_1$group <- as.factor(ifelse(marvel_1$marvel.year<2000, "2000 yili öncesi", "2000 yi

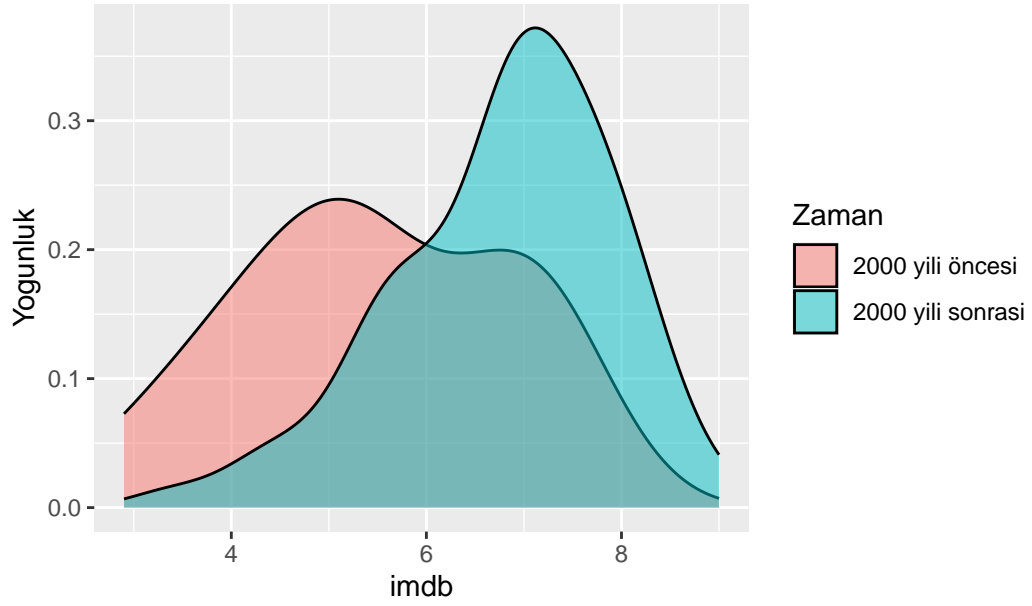
marvel_2 <- data.frame(marvel$imdb_rating, marvel_1)

marvel_2 <- marvel_2[-4,]

ggplot(marvel_2, aes( x = marvel_2$marvel.imdb_rating,
                      group =marvel_2$group, fill = marvel_2$group)) +
  theme(plot.title=element_text(size=12)) +
  geom_density(alpha = 0.5) +
labs(x="imdb", y= "Yogunluk", fill= "Zaman",
      title= "2000 yili öncesi-2000 yili ve sonrası Marvel filmlerinin IMDB puanlari")
```



2000 yılı öncesi–2000 yılı ve sonrası Marvel filmlerinin IMDB puanlar



## YORUM

2000 yılı öncesinde daha yayık bir veri dağılımı gözlenmiştir. 2000 yılı sonrasında puanların, 2000 yılı öncesindeki puanlara göre daha değişken olduğu gözlenmiştir. 2000 yılı sonrasındaki filmlerin imdb puanı dağılım grafiği sola çarpıktır.

## 2.2 2000 yılı öncesi-2000 yılı ve sonrası Marvel filmlerinin süre dağılımları

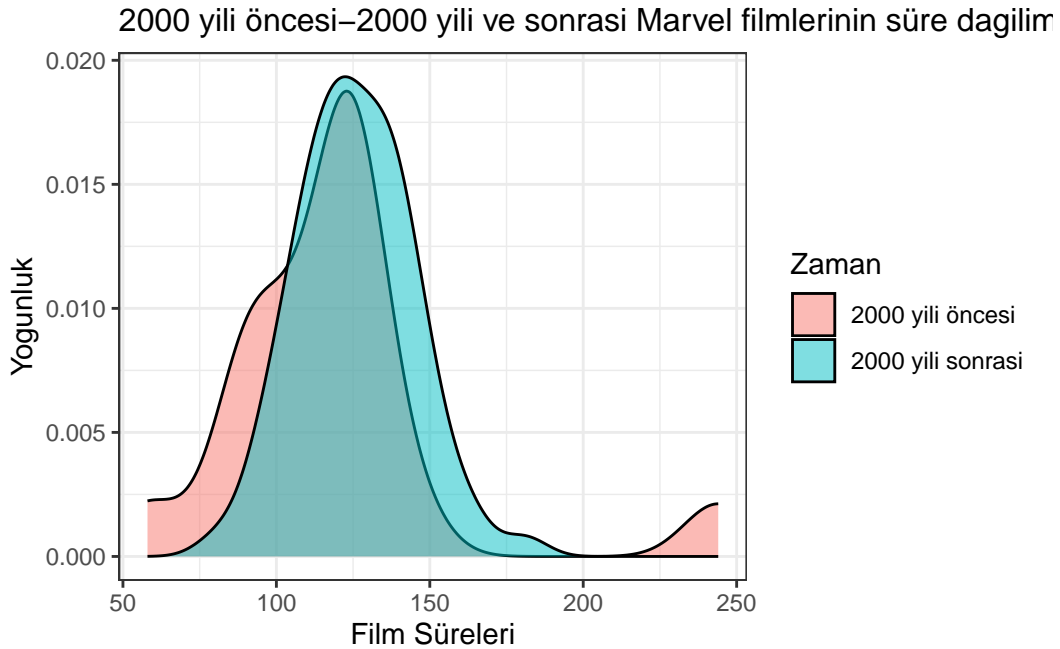
```
library(readr)
marvel_1 <- data.frame(marvel$year)

marvel_1$group <- as.factor(ifelse(marvel_1$marvel.year<2000, "2000 yılı öncesi", "2000 yılı sonrası"))

marvel_2 <- data.frame(marvel$runtime, marvel_1)

marvel_2 <- marvel_2[4,]

ggplot(marvel_2, aes( x = marvel_2$marvel.runtime,
                      group =marvel_2$group, fill = marvel_2$group)) +
  theme_bw() +
  theme(plot.title=element_text(size=12)) +
  geom_density(alpha = 0.5) +
  labs(x="Film Süreleri", y= "Yogunluk", fill= "Zaman",
       title= "2000 yılı öncesi-2000 yılı ve sonrası Marvel filmlerinin süre dağılımları")
```



## YORUM

2000 yılı öncesi ve 2000 yılı sonrasında süre dağılımları birbirine oldukça yakındır. Genel yorum yapacak olursak 2000 yılı öncesi ve sonrası Marvel filimlerinin süreleri çok fazla artış veya azalış göstermemiştir.

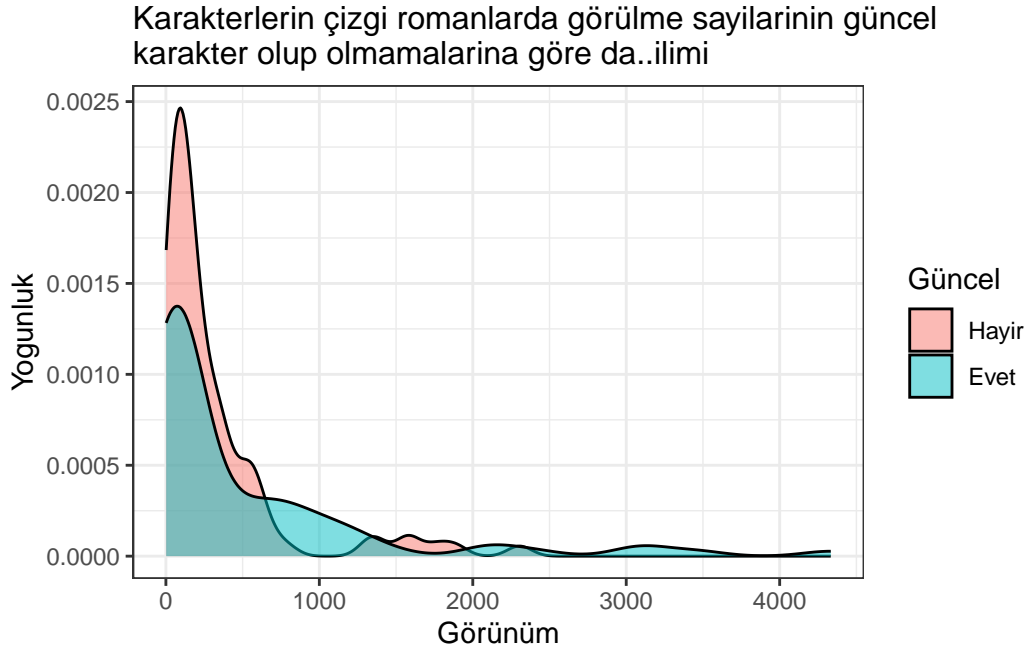
## AVENGERS KARİKATÜR KARAKTERLERİ

Bu veri setinde Marvel çizgi roman karakterlerinin Avengers'a katıldıkları zaman ile 30 Nisan 2015 arasındaki ölümlerini ayrıntılı verileri verilmiştir.

### 3.1 Karakterlerin çizgi romanlarda görülme sayılarının güncel karakter olup olmamalarına göre dağılımı

```
library(readr)
avengers <- read_excel("avengers.xlsx")

library(readr)
ggplot(avengers, aes( x = avengers$Appearances,
                      fill = avengers$`Current?`)) +
  geom_density(alpha = 0.5) +
  theme_bw() +
  theme(plot.title=element_text(size=12)) +
  labs(x="Görünüm", y= "Yogunluk", fill= "Güncel",
       title= "Karakterlerin çizgi romanlarda görülme sayılarının güncel
karakter olup olmamalarına göre dağılımı") +
  scale_fill_discrete(labels=c("Hayir","Evet"))
```



## YORUM

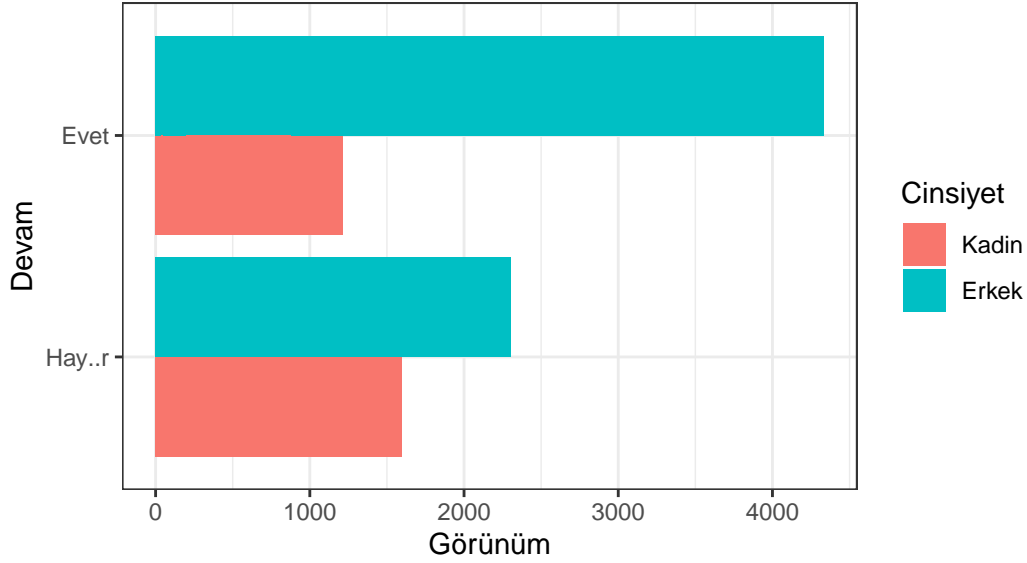
Güncel karakterlerin ekranda görünüm süresi güncel olmayan karakterlere göre daha fazladır. Güncel ve güncel olmayan karakterlerin ekranda görünüm süreleri yoğun olarak 0-1000 değeri arasındadır. Güncel olmayan karakterlerin ekranda görünüm yoğunluğu güncel olan karakterlere göre daha fazladır.

## 3.2 Karakterlerin çizgi romanlarda görülme sayılarının hem güncel karakter olup olmamaları durumu hem de cinsiyetlerine göre dağılımı

```
library(readr)
ggplot(avengers, aes(x= avengers$Appearances, y= avengers$`Current?`,
                     fill = avengers$Gender))+
  geom_col(position = "dodge") +
  labs(x="Görünüm", y= "Devam", fill= "Cinsiyet",
       title= "Karakterlerin çizgi romanlarda görülme sayılarının hem güncel
karakter olup olmamaları durumu hem de cinsiyetlerine göre dağılımı") +
  scale_fill_discrete(labels=c("Kadin", "Erkek")) +
```

```
scale_y_discrete(labels = c("Hayır", "Evet"))+  
theme(plot.title=element_text(size=8)) +  
theme_bw()
```

Karakterlerin çizgi romanlarda görülme sayılarının hem güncel karakter olup olmamaları durumu hem de cinsiyetlerine göre



## YORUM

Erkek karakterlerin çizgi romanlarda görünme sayıları her iki durumda da kadın karakterlerden daha fazladır. Güncel kadın karakterlerin karikatürlerde görünme sayısı güncel olmayan kadın karakterlerin karikatürlerde görünüm sayısından daha azdır. Güncel erkek karakterlerin karikatürlerde görünüm sayısı güncel olmayan erkek karakterlerin görünüm sayısından daha fazladır.