

# Breaking Bad, Veri Bilimci Maaşları ve En Çok Satan Kitaplar Veri Setlerinde Veri Görselleştirme

AUTHOR  
Sercan Öncü

PUBLISHED  
January 1, 2022

## ÖZET

Bu raporda Breaking Bad dizisi, En çok satan kitaplar ve Veri bilimci maaşları veri setleri incelenmiştir. Breaking Bad dizisi sezonlara göre bölüm süresi, reyting punları ve izlenme sayısı dağılımları görselleştirilmiştir. En çok satan kitaplar yıllarına ve türlerine göre okuyu puanları ve fiyat dağılımları görselleştirilmiştir. Veri bilimci maaşları firma büyüklüğü, deneyim ve uzaktan çalışma yüzdesine göre yıllık ve aylık maaşları görselleştirilmiştir.

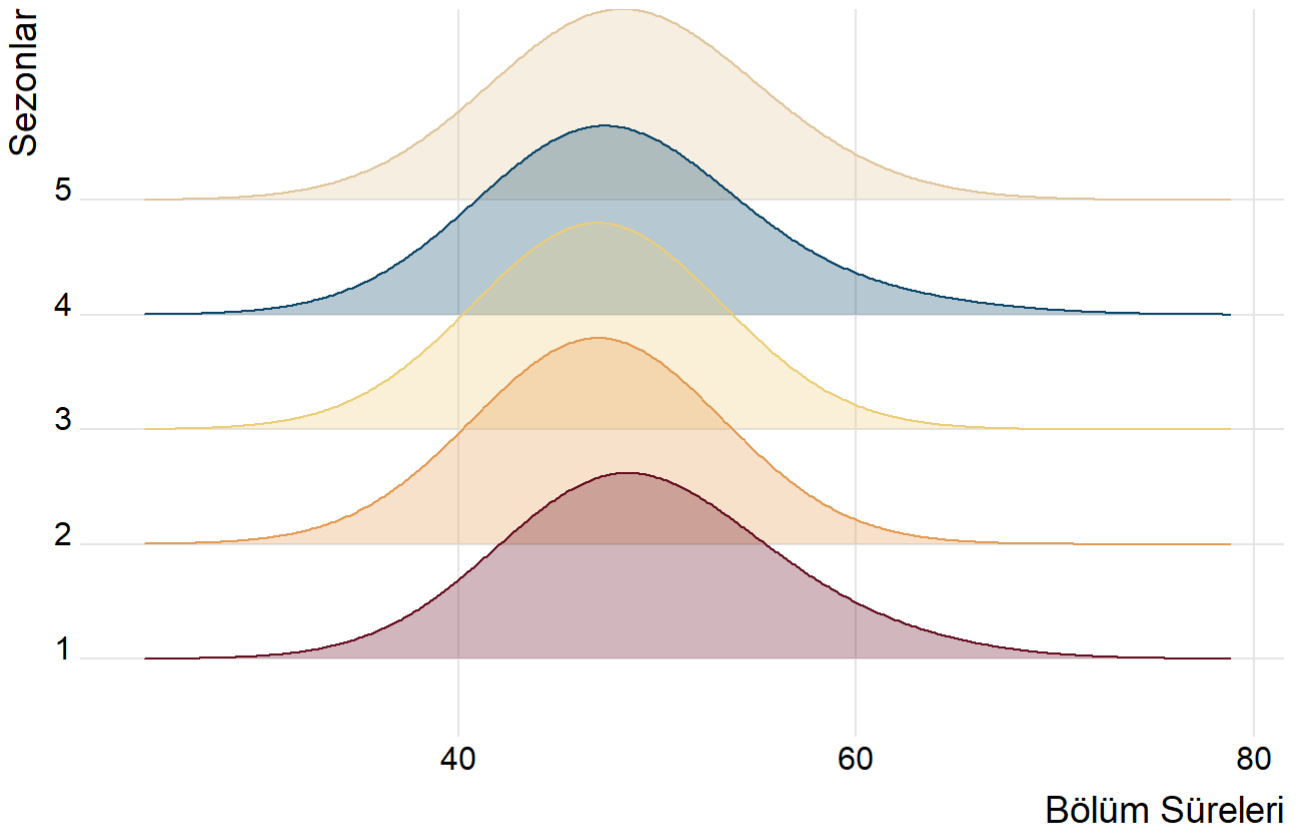
## 1. Breaking Bad

Bu veri seti sezon, bölüm, bölümlerin imdb puanı, bölümlerin süresi, Amerika'daki izlenme sayısı vb. başlıklar altında Breaking Bad dizisi için veriler içermektedir.

### 1.1 Sezonlara göre bölüm süresi dağılımları

```
breaking_bad <- read_csv("C:/Users/serca/Downloads/archive/breaking_bad.csv")
breaking_bad$Season <- factor(breaking_bad$Season, ordered = TRUE, levels = c("1", "2", "3", "4", "5" ))
breaking_bad %>%
  ggplot(aes(y = Season, x = Duration_mins, fill = Season, color = Season)) +
    geom_density_ridges(alpha = 0.3) +
    theme_ridges() +
    theme(legend.position = "none")+
    labs(x="Bölüm Süreleri",
         y= "Sezonlar",
         caption = "Veri Kaynagi: https://www.kaggle.com",
         title = "Sezonlara Göre Bölüm Sureleri Dagilimi")+
    scale_fill_met_d("Navajo")+
    scale_color_met_d("Navajo")
```

## Sezonlara Göre Bölüm Süreleri Dağılımı

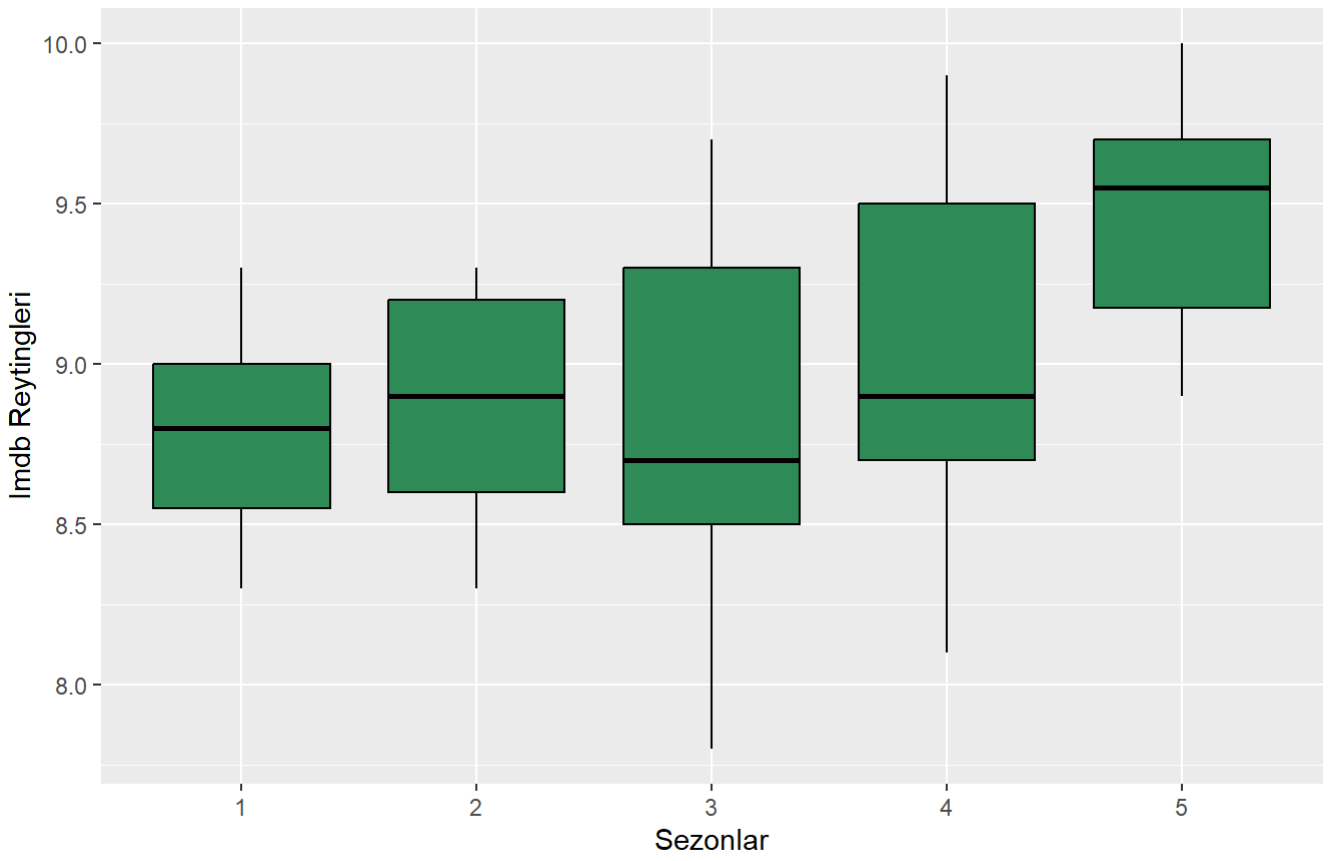


Yukarıdaki grafikte sezonlara göre bölüm süresi dağılımları incelenmiştir. Bölüm sürelerindeki en düşük değişim 3. sezonda ve en yüksek değişim 4. sezondadır. Bölüm sürelerinin ortalama uzunluğuna bakıldığında 1. ve 5. sezonlar daha uzun ortalamaya sahip olduğu görülmektedir.

## 1.2 Sezonlara göre reyting dağılımları

```
breaking_bad %>%  
  ggplot(aes(x = Season, y = Rating_IMDB)) +  
  geom_boxplot(color = "black", fill = "seagreen" ) +  
  labs(x="Sezonlar",  
       y= "Imdb Reytingleri",  
       caption = "Veri Kaynagi: https://www.kaggle.com",  
       title = "Sezonlara Göre Imdb Reytingleri Dağılımı")
```

## Sezonlara Göre Imdb Reytingleri Dagilimi



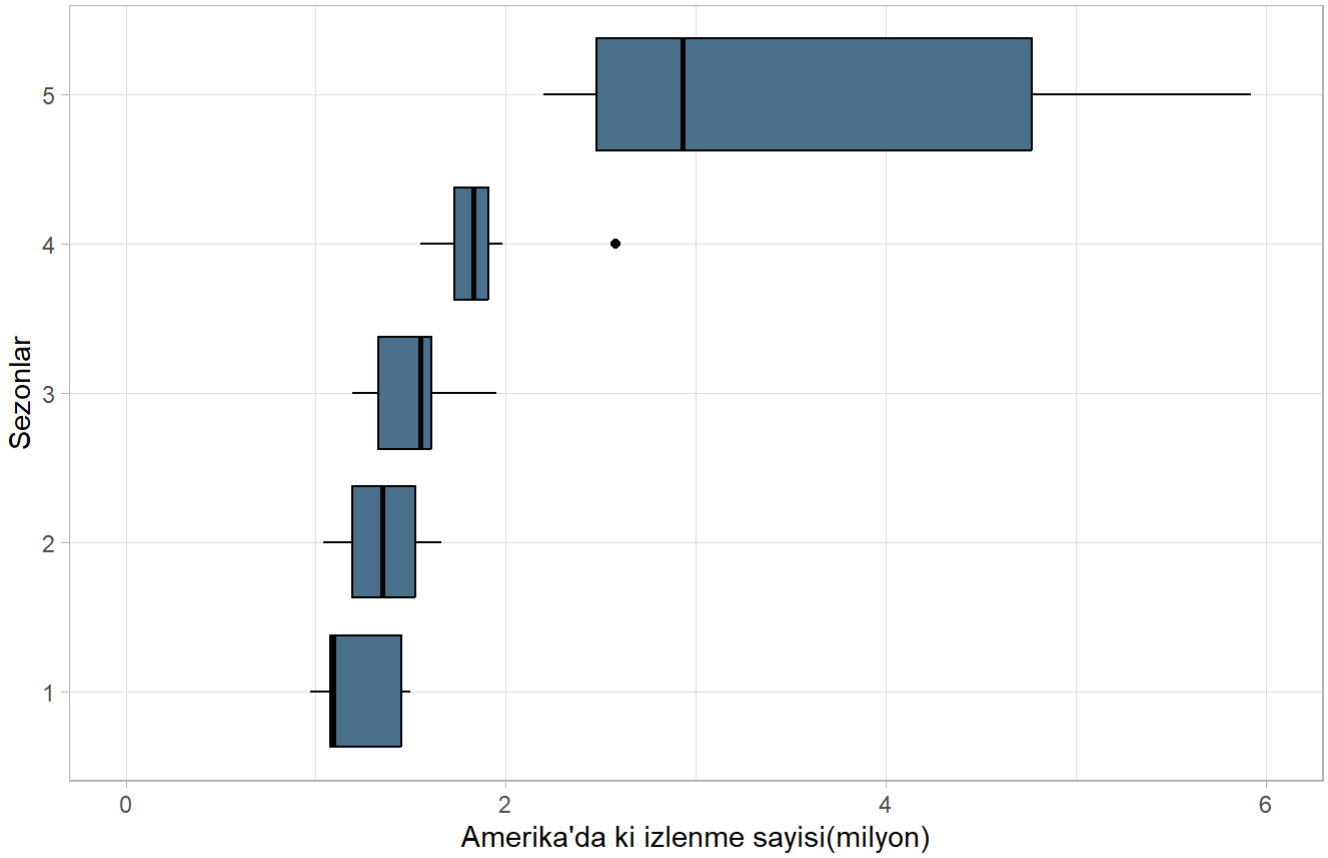
Veri Kaynagi: <https://www.kaggle.com>

Yukarıdaki grafikte sezonlara göre rating dağılımları incelenmiştir. Bölümlerin imdb puanları sezonlar ilerledikçe yükselmiştir. Bölümlerin imdb puanlarındaki değişim en yüksek 3. sezonda ve en düşük değişim ise 1. sezondadır. 5. sezonda dizinin hikayesi daha güzel işlenmiş olabilir.

## 1.3 Sezonlara göre izlenme sayısı dağılımları

```
breaking_bad %>%
  ggplot(aes(x = Season, y = as.numeric(`U.S. viewers_million`))) +
  geom_boxplot(color = "black", fill = "skyblue4") +
  labs(x="Sezonlar",
       y= "Amerika'da ki izlenme sayisi(milyon)",
       caption = "Veri Kaynagi: https://www.kaggle.com",
       title = "Sezonlara Göre Izlenme Sureleri Dagilimi")+
  coord_flip()+
  ylim(0,6)+
  theme_light()
```

## Sezonlara Göre İzlenme Sureleri Dagilimi



Veri Kaynagi: <https://www.kaggle.com>

Yukarıdaki grafikte sezonlara göre izlenme sayısı dağılımları incelenmiştir. En az izlenme ilk sezonda iken en yüksek izlenme 5. sezondadır. En düşük değişim 4. sezonda ve en yüksek değişim 5. sezonda olmuştur. Bölümlerin izlenmesi sezonlar ilerledikçe artmıştır. 5. sezon final sezonu olduğundan en yüksek izlenme sayısına sahip olmuş olabilir.

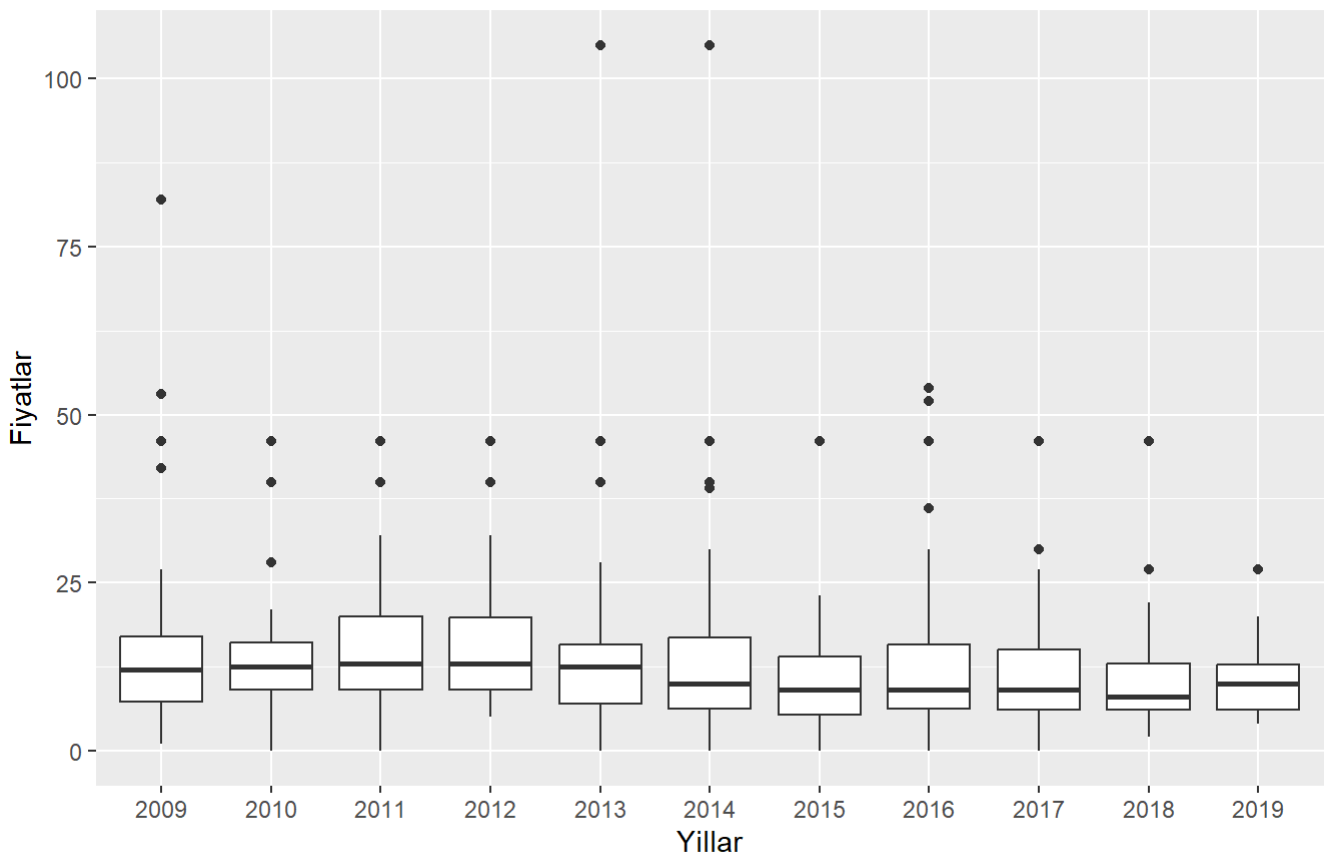
## 2. En çok satan kitaplar

Bu veri seti kullanıcı puanları, inceleme sayıları, fiyat, yıl, tür vb. başlıklar altında En çok satan kitaplar için veriler içermektedir.

### 2.1 Yıllara göre en çok satan kitapların fiyat dağılımı

```
bestsellers <- read_csv("C:/Users/serca/Downloads/archive (1)/bestsellers.csv")
bestsellers %>%
  ggplot(aes(x = as.factor(Year), y = Price)) +
  geom_boxplot() +
  labs(x="Yıllar",
       y= "Fiyatlar",
       caption = "Veri Kaynagi: https://www.kaggle.com",
       title = "Yıllara Göre Kitap Fiyatları Dağılımı")
```

## Yıllara Göre Kitap Fiyatları Dağılımı

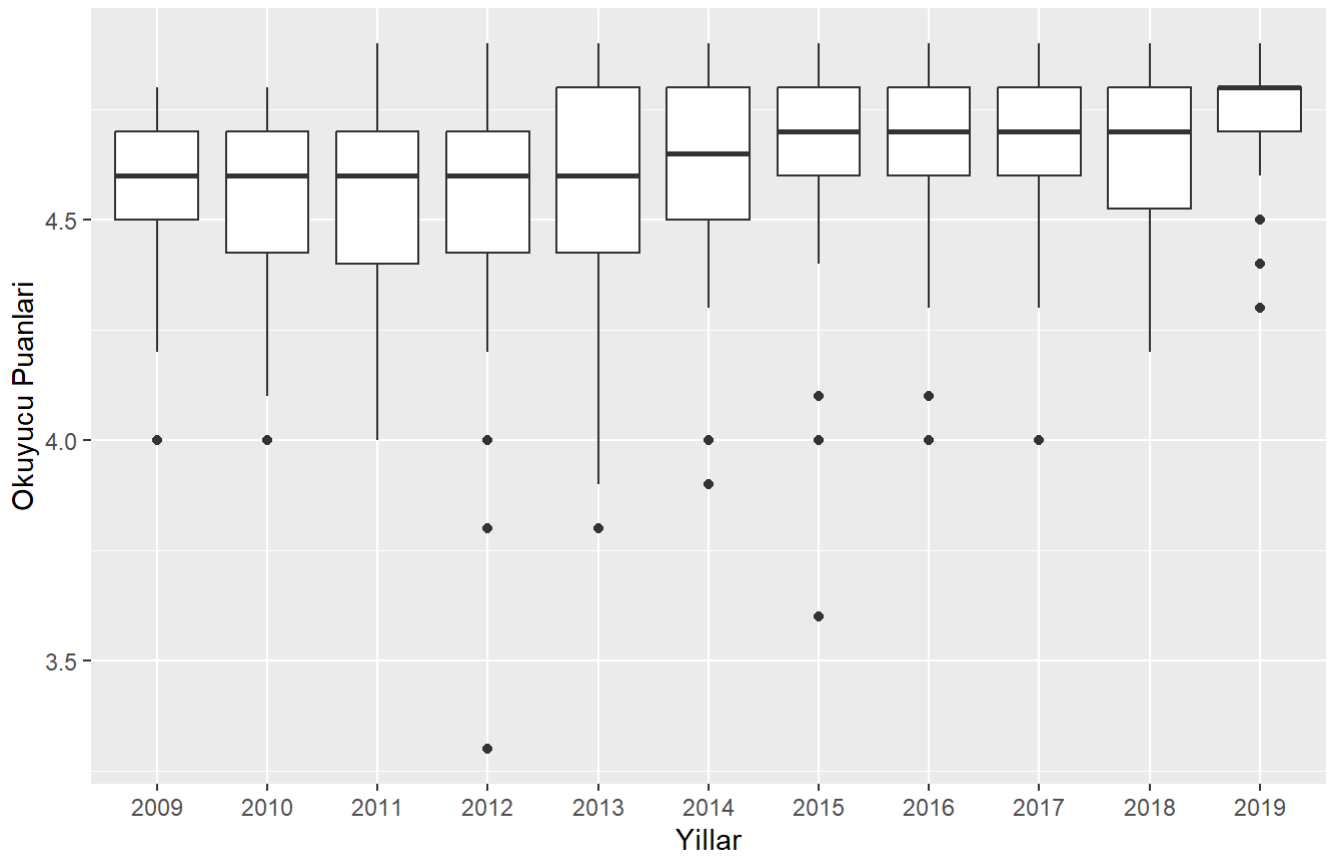


Yukarıdaki grafikte yıllara göre en çok satan kitapların fiyat dağılımı incelenmiştir. Yıllara göre kitap fiyatlarındaki en düşük 2019 yılında olduğu görülmüştür. En yüksek değişim ise 2011 yılında görülmüştür.

## 2.2 Yıllara göre en çok satan kitapların aldıkları okuyucu puanı dağılımı

```
bestsellers %>%  
ggplot(aes(x = as.factor(Year), y = `User Rating`)) +  
geom_boxplot() +  
labs(x="Yıllar",  
y="Okuyucu Puanları",  
caption = "Veri Kaynagi: https://www.kaggle.com",  
title = "Yıllara Göre Okuyucu Puanları Dağılımı")
```

## Yıllara Göre Okuyucu Puanlari Dagilimi



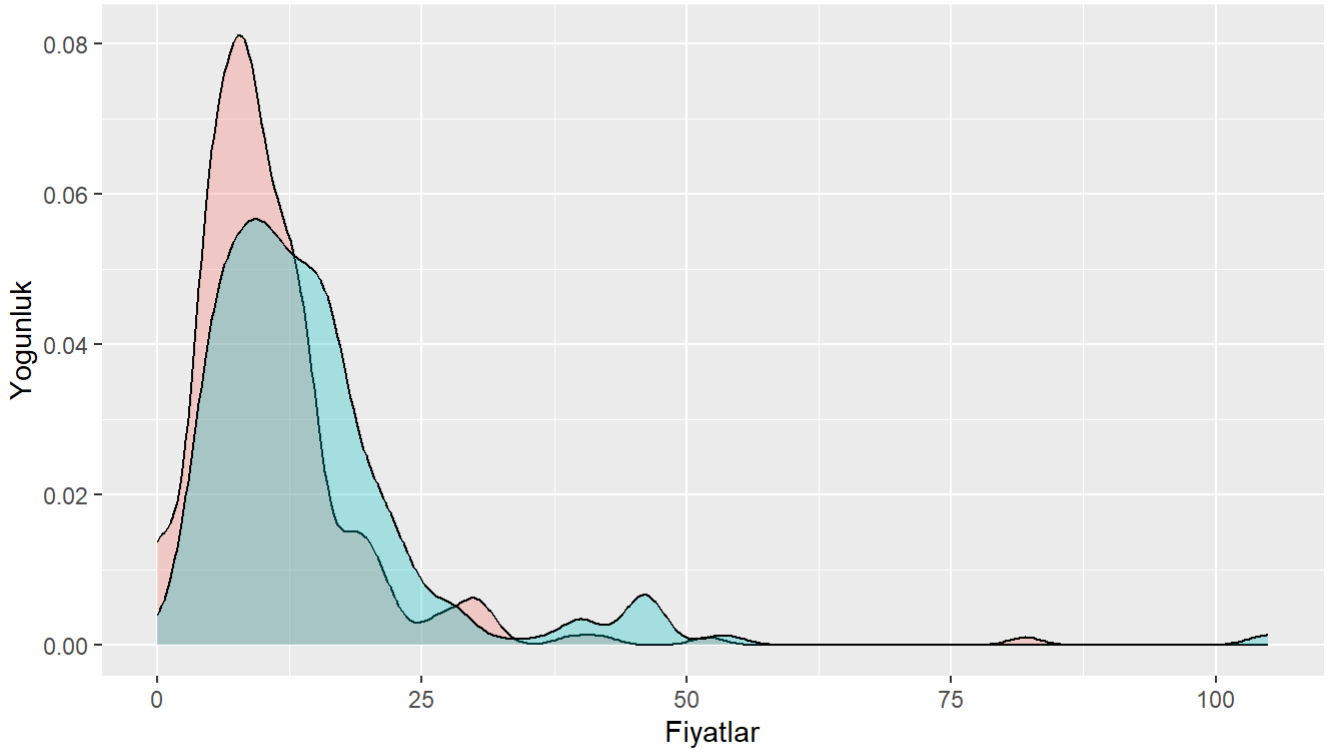
Yukarıdaki grafikte yıllara göre en çok satan kitapların aldıkları okuyucu puanı dağılımı incelenmiştir. Okuyu puanları yıllar ilerledikçe yükselmiştir ve en yüksek okuyu puanı 2019 yılında olduğu görülmüştür. 2013 yılında değişim en yüksek iken 2019 yılında değişim en azdır.

## 2.3 Türüne göre kitap fiyatlarının dağılımı

```
ggplot(bestsellers, aes(x = Price,  
                        fill = Genre)) +  
  geom_density(alpha = 0.3)+  
  theme(legend.position = "top")+  
  labs(title = "Türlere Göre Kitapların Yogunluk Dagilimi",  
        x = "Fiyatlar",  
        y = "Yogunluk",  
        fill="Türler",  
        caption = "Veri Kaynagi: https://www.kaggle.com") +  
  scale_fill_discrete(labels = c("Kurgusal",  
                                "Kurgusal Olmayan"))
```

## Türlere Göre Kitapların Yoğunluk Dağılımı

Türler Kurgusal Kurgusal Olmayan



Veri Kaynagi: <https://www.kaggle.com>

Yukarıdaki grafikte türüne göre kitap fiyatlarının dağılımı görselleştirilmiştir. Kurgusal olmayan türdeki kitapların fiyatının ortalamasının kurgusal türdeki kitapların fiyatından yüksek olduğu görülmüştür. Her iki türde de dağılım sağa çarpıktır. Her iki türde de fiyatlar 13-14 civarında çoğunluktadır.

## 3. Veri bilimci maaşları

Bu veri seti pozisyon, deneyim, maaş, şirket büyüklüğü vb. başlıklar altında Veri bilimci maaşları için veriler içermektedir.

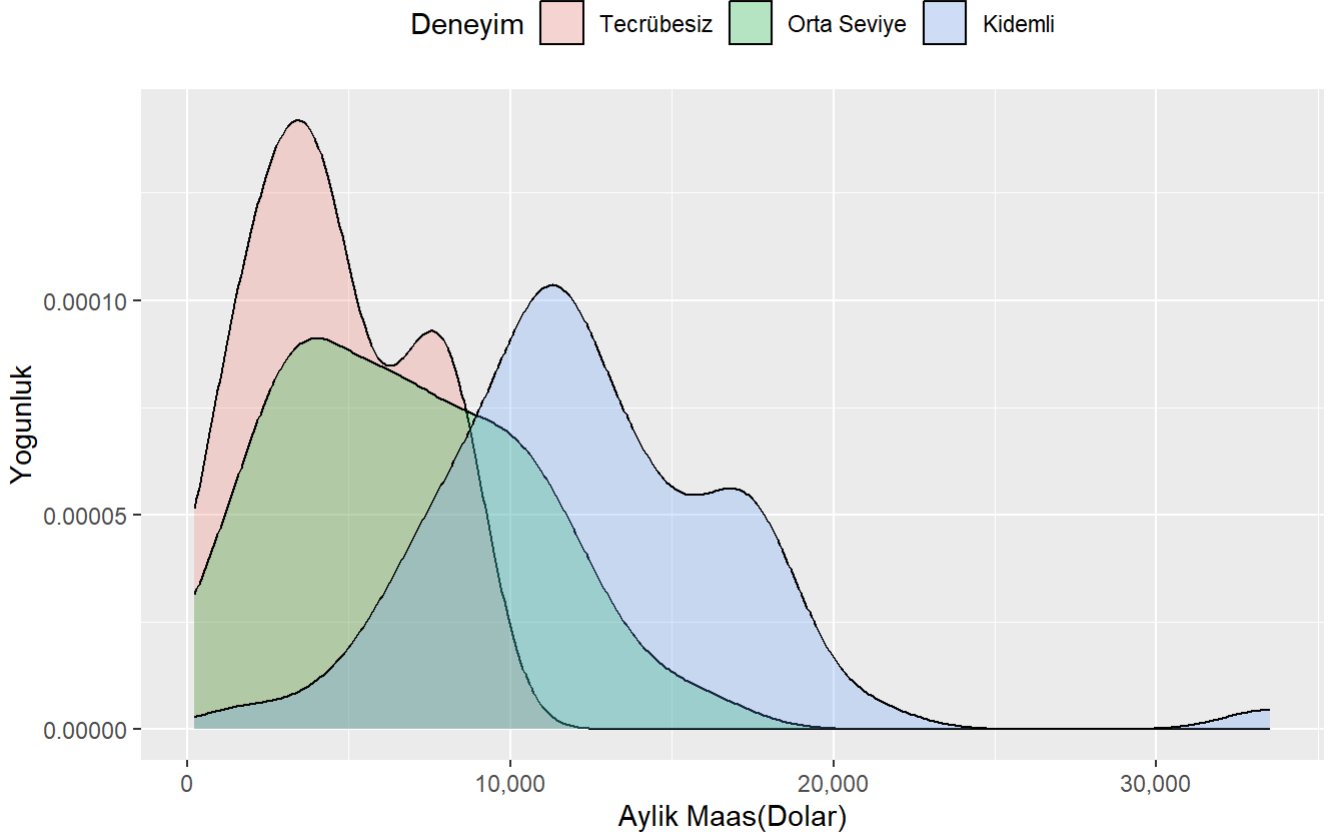
### 3.1 Tecrübe düzeylerine göre veri bilimi pozisyonları aylık maaşlarının dolar

karşılığı bazında dağılımı

```
Data <- read_csv(
  "C:/Users/serca/Downloads/archive (3)/Data_Science_Fields_Salary_Categorization.csv")
data2 <- filter(
  Data, Designation == 'Data Scientist')
ggplot(data2, aes(x = Salary_In_Rupees/978,
  fill = Experience ))+
  geom_density(alpha = 0.25)+
  scale_y_continuous(labels = scales::comma)+
  scale_x_continuous(labels = scales::comma)+
  theme(legend.position = "top")+
  labs(title = "Tecrübeye Göre Veri Bilimci Maas Dagilimi",
  x = "Aylık Maas(Dolar)",
  y = "Yoğunluk",
  fill = "Deneyim",
```

```
caption = "Veri Kaynagi: https://www.kaggle.com")+  
scale_fill_discrete(labels = c("Tecrübesiz", "Orta Seviye", "Kıdemli"))
```

### Tecrübeye Göre Veri Bilimci Maas Dagilimi



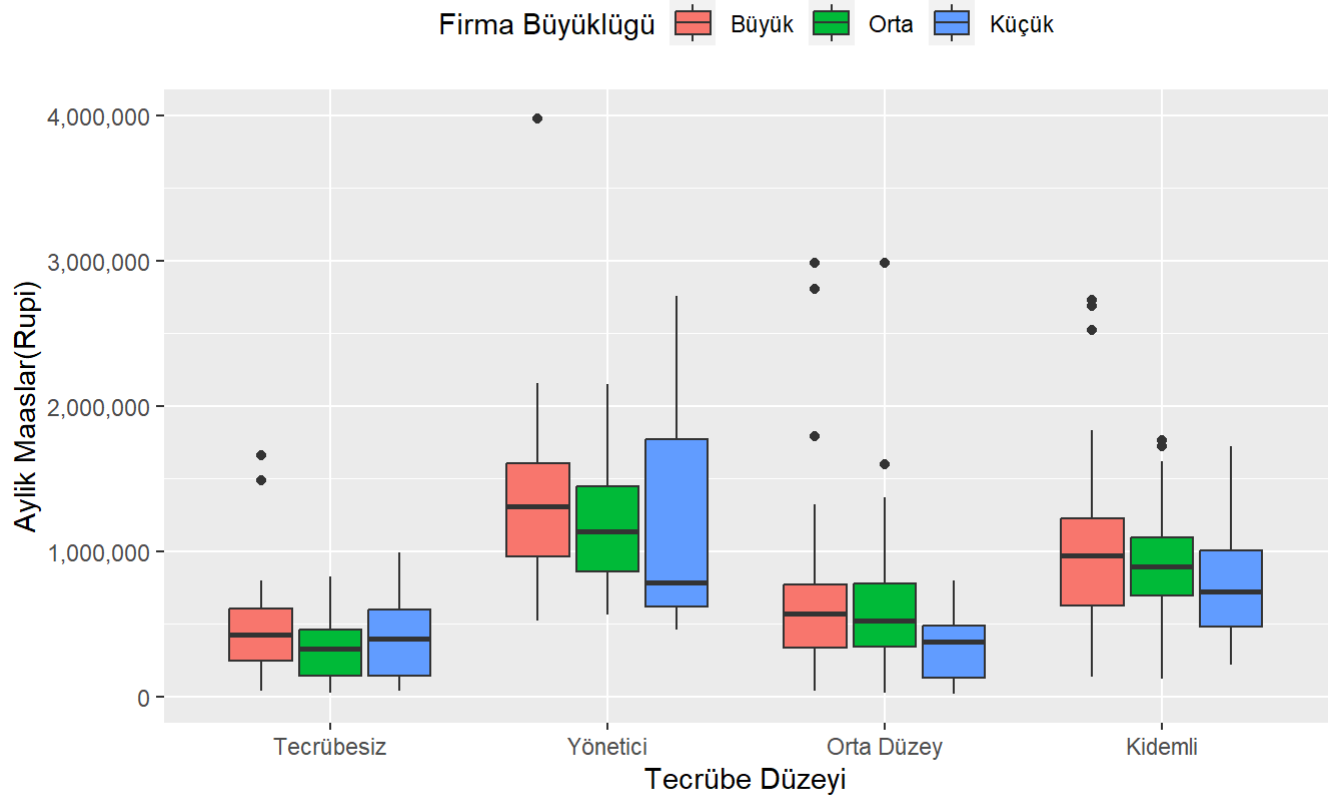
Yukarıdaki grafikte tecrübe düzeylerine göre aylık bazda veri bilimci maaş dağılımı incelenmiştir. Bu grafiğe göre, tecrübesiz olan kişilerin ortalama maaşları diğer tecrübe düzeylerine göre aşağıda kalmıştır. Tecrübe düzeyi artıkça maaş artışı olduğu rahatlıkla söylenebilir.

## 3.2 Firma büyüklüğüne ve tecrübe düzeyine göre aylık maaş tutarı dağılımları

```
ggplot(Data, aes(x = Experience ,  
                 y = Salary_In_Rupees/12,  
                 fill = Company_Size )) +  
geom_boxplot()+  
scale_y_continuous(labels = scales::comma) +  
labs(title = "Tecrübe Düzeyi ve Firma Büyüklüğüne Göre  
Maas Dagilimi",  
      x="Tecrübe Düzeyi",  
      y= "Aylık Maaslar(Rupi)",  
      fill = "Firma Büyüklüğü",  
      caption = "Veri Kaynagi: https://www.kaggle.com")+  
scale_fill_discrete(labels= c("Büyük", "Orta", "Küçük"))+  
theme(legend.position = "top")+  
scale_x_discrete(labels = c("Tecrübesiz", "Yönetici", "Orta Düzey", "Kıdemli" ))
```



## Tecrübe Düzeyi ve Firma Büyüklüğüne Göre Maas Dagilimi



Veri Kaynagi: <https://www.kaggle.com>

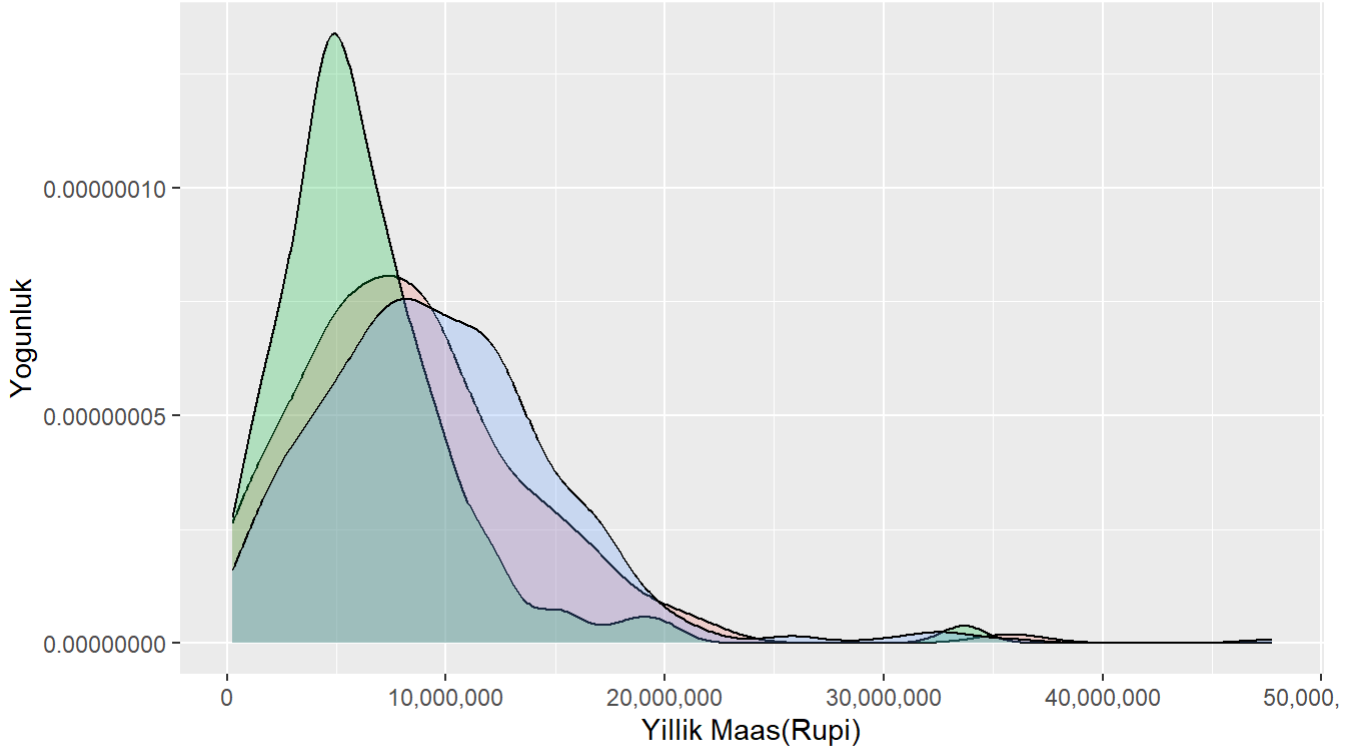
Yukarıdaki grafikte firma büyüklüğüne ve tecrübe düzeyine göre aylık maaş tutarı dağılımları incelenmiştir. En yüksek maaşlar genelde “Yönetici” deneyime sahip büyük şirketlerde çalışan kişilerdedir. “Orta Düzey” deneyime sahip kişiler en düşük maaşları küçük firmalarda almaktadır. En yüksek değişim “Yönetici” deneyime sahip küçük firma çalışanlarındadır ve en düşüğü ise “Tecrübesiz” deneyime sahip olup büyük firmalarda çalışan kişilerdedir. Bu grafiğe göre tecrübesiz kişilerin küçük şirketleri tercih etmesi maaş bakımından mantıklı bir tercih olur.

## 3.3 Uzaktan çalışma yüzdesi sistemine göre yıllık maaş ücreti dağılımları

```
ggplot(Data, aes(x = Data$Salary_In_Rupees,  
  fill = as.factor(Data$Remote_Working_Ratio),  
  ))+  
  geom_density(alpha = 0.25)+  
  scale_y_continuous(labels = scales::comma)+  
  scale_x_continuous(labels = scales::comma)+  
  theme(legend.position = "top")+  
  labs(title = "Uzaktan Calisma Yüzdeligine Göre Maas Dagilimi",  
    x = "Yillik Maas(Rupi)",  
    y= "Yogunluk",  
    fill = "Uzaktan Calisma Yuzdeligi",  
    caption = "Veri Kaynagi: https://www.kaggle.com")
```

## Uzaktan Calisma Yuzdeligine Gre Maas Dagilimi

Uzaktan Calisma Yuzdeligi 0 50 100



Yukarıdaki grafikte uzaktan alıřma yzdeliđine bađlı olarak yıllık maař dađılımını incelenmiřtir. En dřk maař ortalaması grafikte grldđ gibi %50 uzaktan alıřma oranı olan kiřilerdedir. %100 ve %0 uzaktan alıřma oranı olan kiřiler arasında maařlar az da olsa %100 uzaktan alıřan kiřilerde ykseklik gstermiřtir.