

# **Avengers,Marvel Filmleri ve Veri Bilimci Maařları**

Hüseyin KAYAR

17.11.2022

## **Özet**

Bu raporda Kaggle'dan alınan Avengers,Marvel Filmleri ve Veri Bilimci Maařları veri setlerini istenilen türe göre dağılımı bulunmuřtır. Bu dağılımlara çeřitli grafiklere dökülmüş ve yorumlanmıştır.

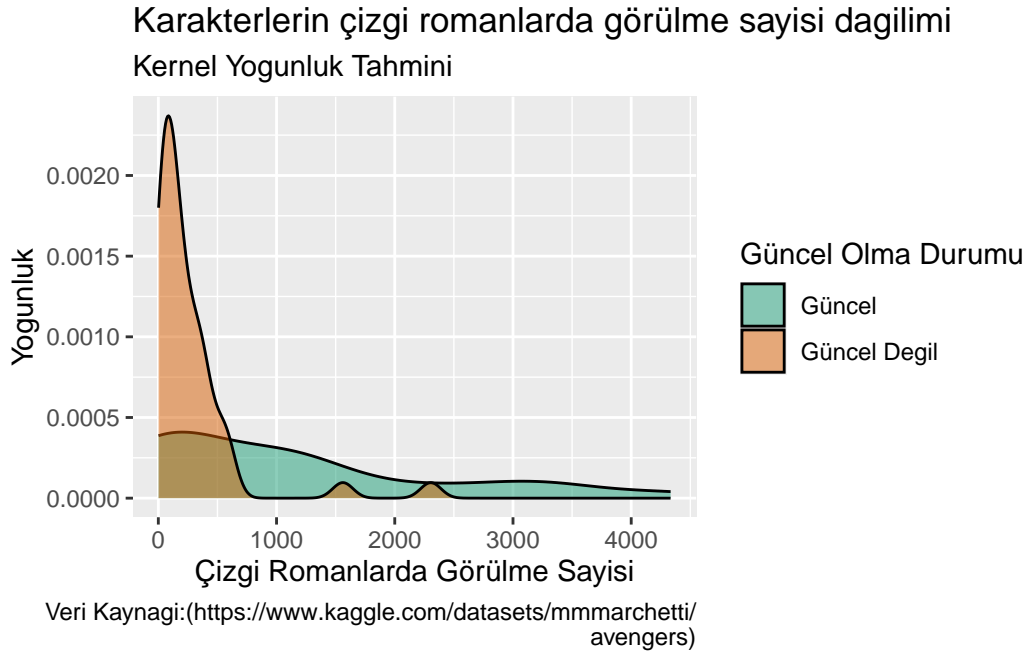
## **1-Avengers**

### **1.1 Avengers Veri Seti Tanıtımı**

Bu veri setinde popüler zamanların en sevilen film serisi olan Avengers (Yenilmezler) serisinin çizgi romanları incelenmiştir. İnceleme sonuçlarında elde edilen verileri 10 Farklı katagoriye ayrılmıştır. Bu raporda incelenmiş olan katagoriler aşağıdaki gibi grafiğe dökülmüřtür.

## 1.2 Karakterlerin çizgi romanlarda görülme sayısı dağılımı grafiği

```
ggplot(avengerData_changed, aes(x=Appearances, fill=Current.)) +  
  geom_density(alpha=0.5)+  
  labs(x = "Çizgi Romanlarda Görülme Sayisi",  
       y = "Yogunluk",  
       title = "Karakterlerin çizgi romanlarda görülme sayisi dagilimi",  
       subtitle = "Kernel Yogunluk Tahmini",  
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/mmmarchetti/avengers)")+  
  scale_fill_brewer(palette = "Dark2") +  
  guides(fill=guide_legend(title="Güncel Olma Durumu"))
```

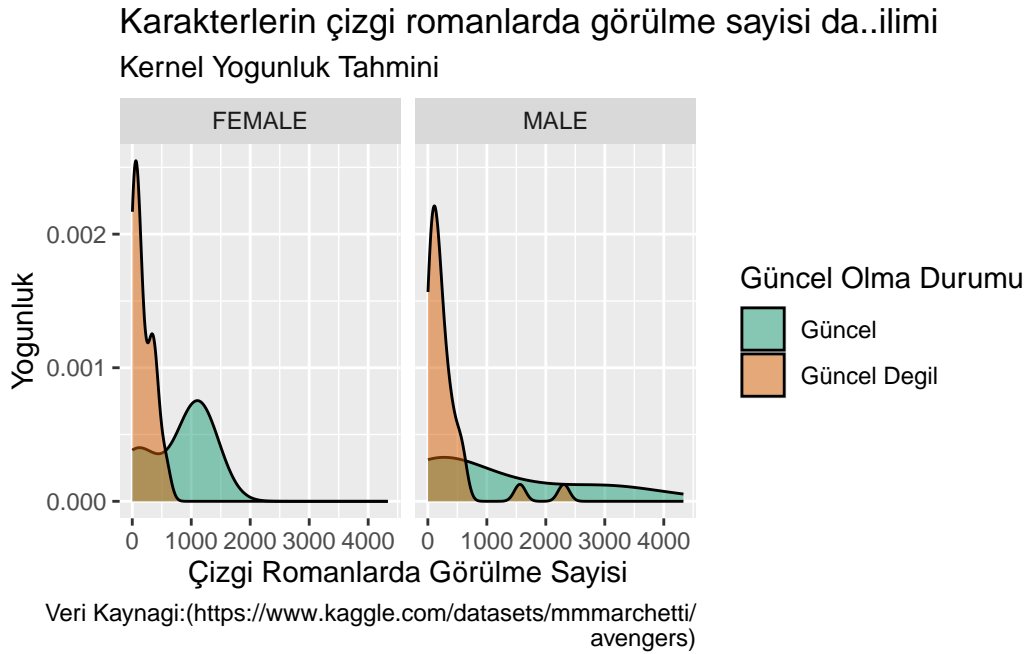


### 1.2.1 Grafik Yorumu:

Çizilmiş olan grafikte çok bariz bir şekilde görülmüş olan ilk şey çizgi romanlarda az sayıda ismi geçmiş karakterlerin bilinirliğinin az olmasıdır. Yapılacak olan çıkarım ise bir karakterin bilinirliği ile o karakterin isminin geçmiş olduğu romanlar sayısının doğru orantılı olmasıdır. Bir başka gözlenen olay ise karakterlerin çizgi romanlarda görülme sayılarının 2000 civarı olduğu yerlerde karakterin bilinirliğinin oranı %50'ye yakın olmasıdır.

### 1.3 Karakterlerin çizgi romanlarda görülme sayıları-güncel karakter olma durumu-cinsiyetleri grafiği

```
ggplot(avengerData_changed, aes(x=Appearances, fill=Current.)) +  
  geom_density(alpha=0.5)+  
  facet_wrap(~avengerData$Gender) +  
  labs(x = "Çizgi Romanlarda Görülme Sayısı",  
       y = "Yogunluk",  
       title = "Karakterlerin çizgi romanlarda görülme sayısı dağılımı",  
       subtitle = "Kernel Yogunluk Tahmini",  
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/mmmarchetti/avengers))+  
  scale_fill_brewer(palette = "Dark2") +  
  guides(fill=guide_legend(title="Güncel Olma Durumu"))
```



#### 1.3.1 Grafik Yorumu:

Çizilmiş olan grafikte güncel olma durumunun cinsiyetten daha önemli olduğu gözlemlenmiştir. Aynı zamanda kadınların tekrar etme sayılarının erkeklerden daha az olduğu gözlemlenmiştir. Kadınların çizgi romanlarda tekrar etme sayılarının çoğunlukla 1000'de yığıldığı gözlemlenmiştir.

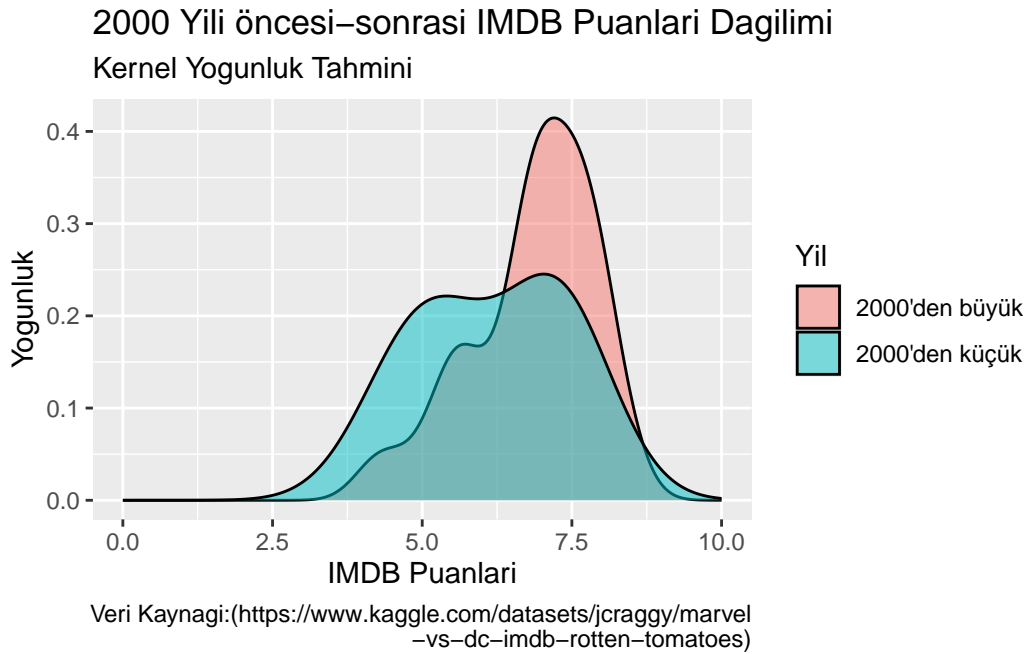
## 2- Marvel Filmleri

### 2.1 Marvel vs Dc Filmleri Veri Seti Tanıtımı

Bu veri setinde insanları ikiye bölen bir tartışma konusu olan iki büyük evrenin yanı sıra Marvel ve Dc filmleri olarak veriler toplanmış ve bu veriler 9 farklı kategoriye göre ayrılmıştır. Biz ise aşağıdaki şekilde 2 farklı grafikte bu veri grubunun Marvel filmlerin incelemiş bulunmaktayız.

### 2.2 2000 Yılı öncesi-sonrası IMDB Puanları Dağılımı

```
ggplot(marvelData, aes(x=imdb_rating, fill=year)) +  
  geom_density(alpha=0.5)+  
  lims(x = c(0, 10))+  
  labs(x = "IMDB Puanlari",  
       y = "Yogunluk",  
       title = "2000 Yili öncesi-sonrasi IMDB Puanlari Dagilimi",  
       subtitle = "Kernel Yogunluk Tahmini",  
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/jcraggy/marvel-vs-dc-imdb-rotten-tomatoes))+  
  guides(fill=guide_legend(title="Yil"))
```

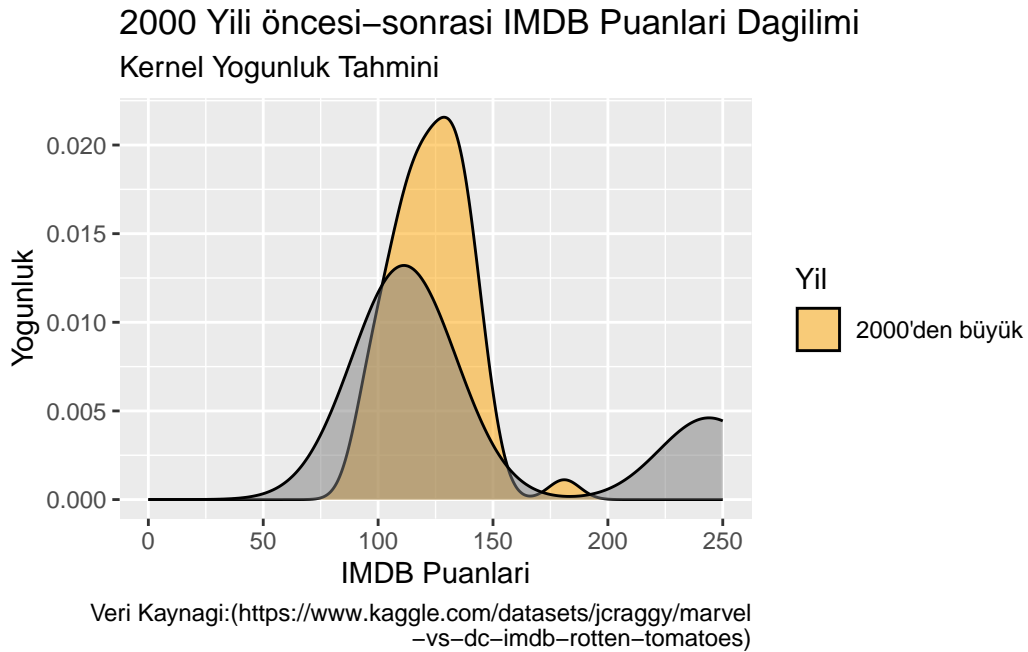


### 2.2.1 Grafik Yorumu:

Bu grafikte Marvel filmlerinin 2000 yılı öncesi ve sonrası IMDB puanı olarak gayet yüksek puanlar aldığı gözlemlenmiştir. Marvel filmleri 2000 yılı öncesi ve sonrası aldığı IMDB puanları çoğunlukla 7.5 puan olarak gözlemlenmiştir. 2000 yılı öncesi gözlemlenen IMDB puanları 2000 yılı sonrası puanlara göre daha düşüktür.

### 2.3 2000 Yılı öncesi-sonrası Süre Dağılımları

```
ggplot(marvelData, aes(x=runtime, fill=year)) +  
geom_density(alpha=0.5)+  
lims(x = c(0, 250))+  
labs(x = "IMDB Puanlari",  
      y = "Yogunluk",  
      title = "2000 Yili öncesi-sonrasi IMDB Puanlari Dagilimi",  
      subtitle = "Kernel Yogunluk Tahmini",  
      caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/jcraggy/marvel-vs-dc-imdb-rotten-tomatoes))+  
guides(fill=guide_legend(title="Yil"))+  
scale_fill_manual(values=c("2000'den büyük"="orange", "2000'den  
küçük"="darkgreen"))
```



### 2.3.1 Grafik Yorumu:

Bu grafikte Marvel filmlerinin süresinin çoğunlukla 120 dakika olarak gözlemlenmiştir. 2000’li yıllardan önce uzun süreli film denemeleri gözlemlenmiştir ancak 2000’li yıllardan sonra böyle bir veri alınmamıştır.

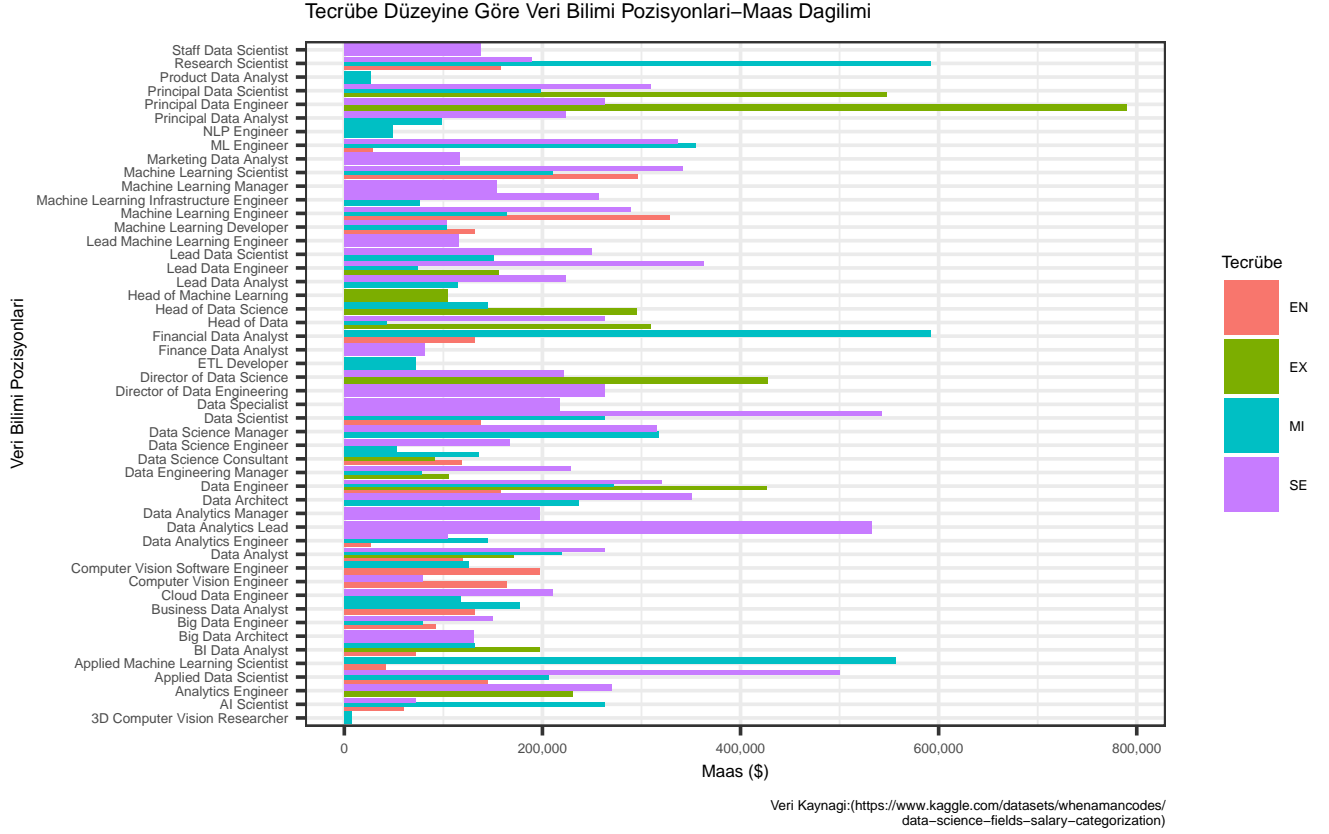
## 3-Veri Bilimci Maaşları

### 3.1 Veri Bilimci Maaşları Veri Seti Tanıtımı

Bu veri setinde 2020-2022 yılları arasında veri bilimci maaşları incelenmiştir. 9 ayrı kategoride incelenmiş olan bu veride bizim incelemiş olduğumuz veriler aşağıdaki şekilde grafikleştirilmiştir.

### 3.2 Tecrübe düzeyine göre veri bilimi pozisyonları-Maaş dağılımı grafiği

```
ggplot(DataScience_filtered, aes(x = Salary_In_RupeesDolar, y = Designation,
                                fill = Experience)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_x_continuous(labels = comma)+
  labs(x = "Maas ($)",
       y = "Veri Bilimi Pozisyonlari",
       fill = "Tecrübe",
       title = "Tecrübe Düzeyine Göre Veri Bilimi Pozisyonlari-Maas Dagilimi",
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/whenamancodes/
data-science-fields-salary-categorization)") +
  theme_bw()+
  theme(text = element_text(size = 5))
```

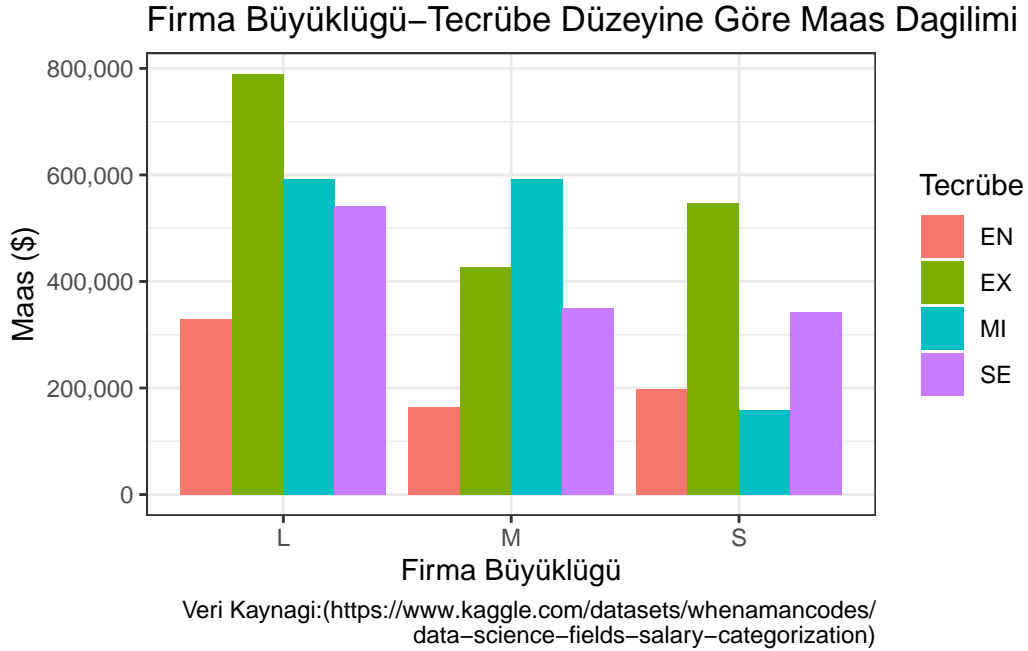


### 3.2.1 Grafik Yorumu:

Bu grafikte en yüksek maaş geliriine sahip olan kişinin yönetici(EX) olduğu gözlemlenmiştir. En düşük maaş skalası ise giriş seviyesi veri bilimci (EN) olduğu görülmüştür. Veri bilimi işçilerinin çoğunluk kısmının uzman düzeyde(SE) olduğu gözlenmiştir.

### 3.3 Firma büyüklüğü-Tecrübe düzeyine göre Maaş dağılımı grafiği

```
ggplot(DataScience_filtered, aes(x =Company_Size , y = Salary_In_RupeesDolar,
                                fill = Experience)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(labels = comma)+
  labs(x = "Firma Büyüklüğü",
       y = "Maas ($)",
       fill = "Tecrübe",
       title = "Firma Büyüklüğü-Tecrübe Düzeyine Göre Maas Dagilimi",
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/whenamancodes/data-science-fields-salary-categorization")) +
  theme_bw()
```



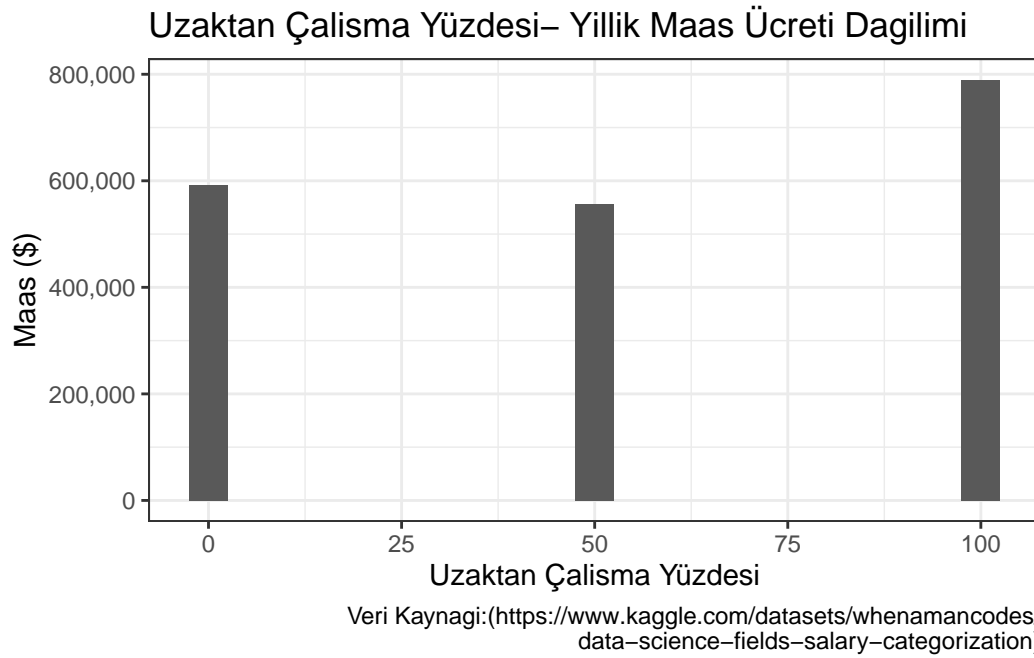
#### 3.3.1 Grafik Yorumu:

Bu grafikte firma boyutunun maaş üzerinde çok fazla etkisi olduğunun, ancak giriş seviyesi veri bilimci(EN) maaşlarının bundan çok fazla etkilenmediğini gözlemlenmiştir. Veride ilginç olan bir nokta olarak yönetici(EX) konumunda çalışan kişilerin orta büyüklükte firma seçmelerinin doğru olmadığı görülmüştür. Tecrübesiz, yani giriş seviyesi veri bilimciler (EN) için en uygun firma olarak büyük(L) firmalara girmeleri önerilmiştir.



### 3.4 Uzaktan çalışma yüzdesi- Yıllık maaş ücreti dağılımı

```
ggplot(DataScience_filtered, aes(x =Remote_Working_Ratio ,  
                                y =Salary_In_RupeesDolar, width=5)) +  
  geom_bar(stat = "identity", position = "dodge", binwidth = 0.1) +  
    scale_y_continuous(labels = comma)+  
  labs(x = "Uzaktan Çalışma Yüzdesi",  
       y = "Maas ($)",  
       title = "Uzaktan Çalışma Yüzdesi- Yıllık Maas Ücreti Dağılımı",  
       caption = "Veri Kaynagi:(https://www.kaggle.com/datasets/whenamancodes/data-science-fields-salary-categorization)") +  
  theme_bw()
```



#### 3.4.1 Grafik Yorumu:

Verilmiş olan grafikte gözlemlendiği üzere, %100 uzaktan çalışan isanlarının yıllık gelirinin daha yüksek olduğu çıkarımına varılmıştır. Bununla ters orantılı olarak en düşük yıllık maaş gelirinin hibrit(%50) çalışan işçilerin olduğu görülmüştür.