

Verilen Veri Setlerinde Oransal Kıyaslamalar

Hüseyin KAYAR

24.11.2022

Özet

Bu raporda Kaggle'dan alınan Avengers, Marvel Filmleri ve Veri Bilimci Maaşları veri setlerini istenilen türe göre oranları bulunmuştur. Bu oranların çeşitli 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 dökülmüştür.

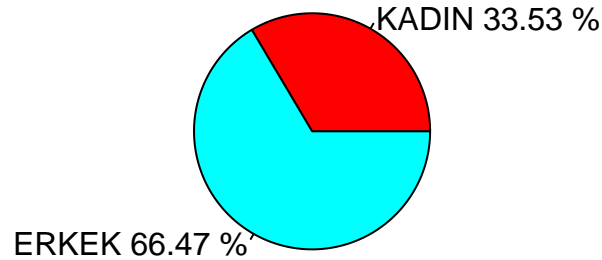
1.2 Karakterlerin Çizgi Romanlardaki Kadın-Erkek Oranı Grafiği

```
avengerPieGender <- avengers %>% group_by(Gender) %>%  
  summarise(Percent = 100 * n() / nrow( avengers ))
```

```
avengerPieGenderConverted <- within(avengerPieGender, {  
  Gender[Gender == "MALE"] <- "ERKEK"  
  Gender[Gender == "FEMALE"] <- "KADIN"  
})
```

```
pie(x = as.double(avengerPieGenderConverted$Percent),  
    label=paste(avengerPieGenderConverted$Gender,  
                format(round(avengerPieGender$Percent, 2), nsmall = 2), "%"), col=rainbow(length(avengerPieGenderConverted$Gender)),  
    main="Çizgi Romanlardaki Kadın-Erkek Oranı" )
```

Çizgi Romanlardaki Kadın–Erkek Oranı

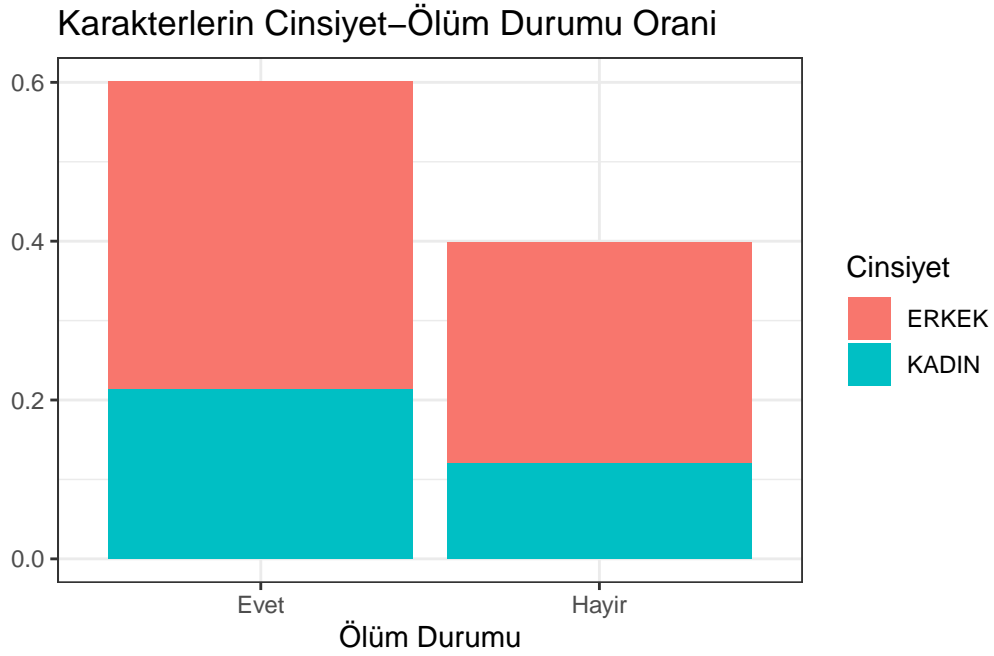


1.2.1 Grafik Yorumu:

Pasta grafiğini incelediğimizde çizgi romandaki erkek karakter oranının kadın karakter oranının 2 katına eşit olduğunu gözlemliyoruz. Çizgi roman karakterlerinin çoğunluğunu erkekler oluşturmaktadır.

1.3 Karakterlerin Cinsiyetlerine ve Ölüm Durumlarına (Yalnızca Death1) Göre Oranını Grafiği

```
ggplot(avengerMosaicConverted, aes(fill = Gender,  
                                   y = Number,  
                                   x = Death)) +  
  labs(title = "Karakterlerin Cinsiyet-Ölüm Durumu Oranı", y="", x="Ölüm Durumu",  
        fill="Cinsiyet")+  
  geom_bar(position = "stack",  
            stat = "identity")+  
  scale_x_discrete(labels = c("Evet", "Hayir")) +  
  theme_bw()
```



1.3.1 Grafik Yorumu:

Bu grafikte çizgi romandaki erkek karakterlerin ölüm oranlarının daha fazla olduğu açıkça görülmektedir. Kadınların bulunma oranları erkeklerden daha az olduğunu bu grafikte de görüntülemekteyiz.

2- Marvel Filmleri

Filmlerin türlerine göre oranlarına veri görselleştirme yöntemleriyle araştırınız. (Her bir türü grafikte ayrı ayrı kullanınız, oranları iki veya daha fazla tür üzerinden etiketlemeyiniz.)

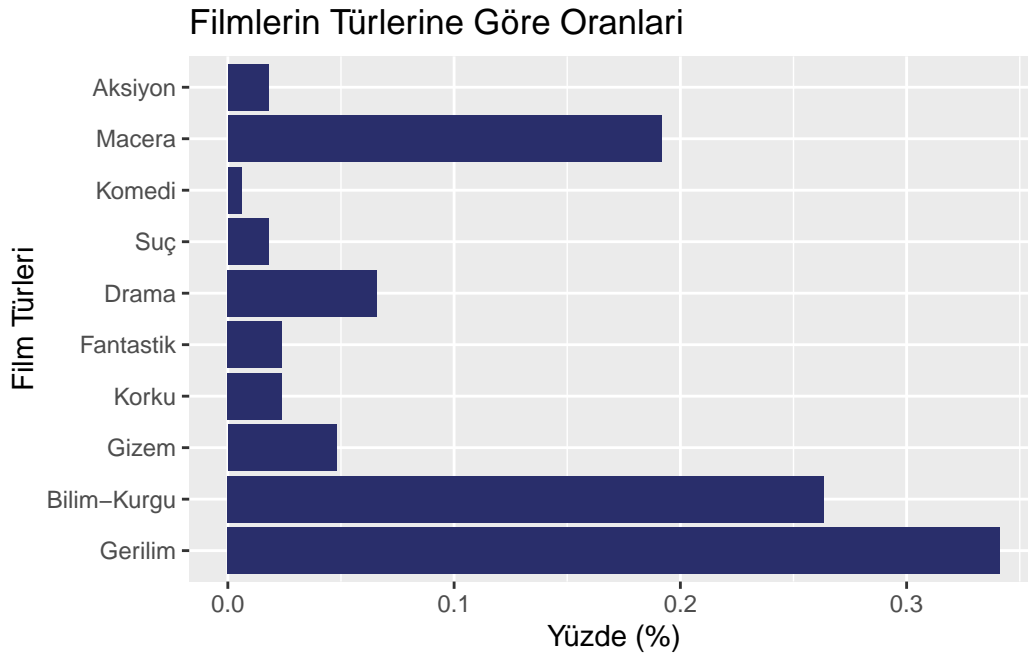
Filmlerin türlerine ve uzunluklarına (iki gruba ayırınız: 120 dakika altı ve üstü) göre oranlarına veri görselleştirme yöntemleriyle araştırınız. (Her bir türü grafikte ayrı ayrı kullanınız, oranları iki veya daha fazla tür üzerinden etiketlemeyiniz.)

2.1 Marvel vs Dc Filmleri Veri Seti Tanıtımı

Bu veri setinde insanları ikiye bölen bir tartışma konusu olan iki büyük evrenin yanı 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 Film-Tür Oran Grafiği

```
ggplot(marvelOnlyMovieTypes, aes(
  y = percent,
  x = genre)) +
geom_bar(position = "stack",
  stat = "identity",
  fill="#292e6b")+
labs(x = "Film Türleri",
  y = "Yüzde (%)",
  title = "Filmlerin Türlerine Göre Oranları")+
scale_x_discrete(labels = c("Gerilim","Bilim-Kurgu","Gizem","Korku",
  "Fantastik","Drama","Suç","Komedi","Macera","Aksiyon"))+
coord_flip()
```

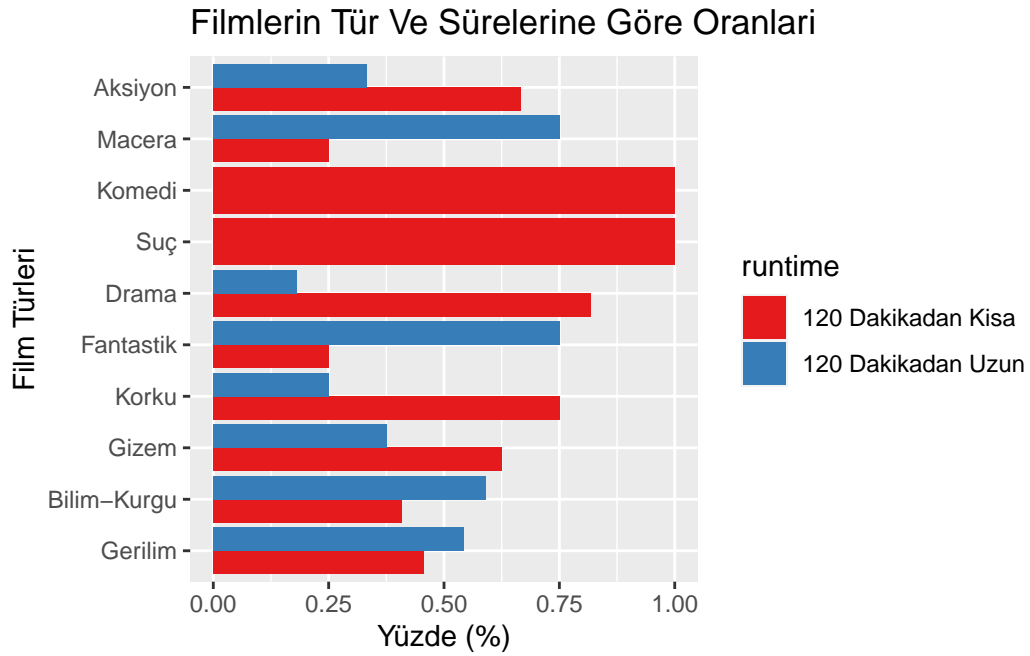


2.2.1 Grafik Yorumu:

Marvel filmlerin bir çoğunu aksiyon türünde çekmiştir. En az çekilmiş film türü ise gizem olarak görülmüştür. Genel olarak Marvel 3 türde film çekmiştir bunlar Bilim-Kurgu, Macera ve Aksiyondur.

2.3 Filmlerin Tür-Süre Oranları

```
ggplot(marvelOnlyMovieTypesWithRuntime, aes(fill = runtime,  
      y = percent,  
      x = genre)) +  
  geom_bar(position = "dodge",  
    stat = "identity")+  
  labs(x = "Film Türleri",  
    y = "Yüzde (%)",  
    title = "Filmlerin Tür Ve Sürelerine Göre Oranları")+  
  scale_x_discrete(labels = c("Gerilim", "Bilim-Kurgu", "Gizem", "Korku",  
    "Fantastik", "Drama", "Suç", "Komedi", "Macera", "Aksiyon"))+  
  coord_flip()+  
  scale_fill_brewer(palette="Set1")
```



2.3.1 Grafik Yorumu:

Bu grafikte uzun süreli filmlerin türleri kurgu ve drama olarak gözlemlenmektedir. Gizem ve korku türlerinde hiç uzun süreli film yapılmamıştır. Uzun ve kısa olarak birbirlerine en yakın yüzdeye sahip film türü aksiyondur.

3-Veri Bilimci Maaşları

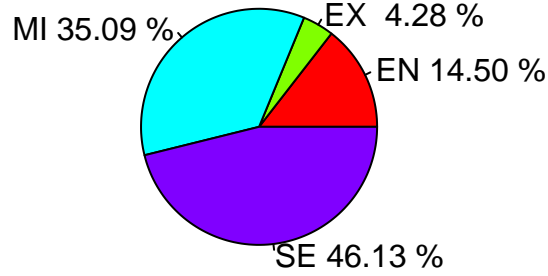
3.1 Veri Bilimci Maaşları Veri Seti Tanıtımı

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

3.2 Tecrübe Düzeylerine Göre Veri Bilimci Oranları

```
pie(x = as.double(dataScienceExperiencePercent$Percent), label=paste(dataScienceExperience
```

Tecrübe Düzeylerine Göre Veri Bilimci Oranları

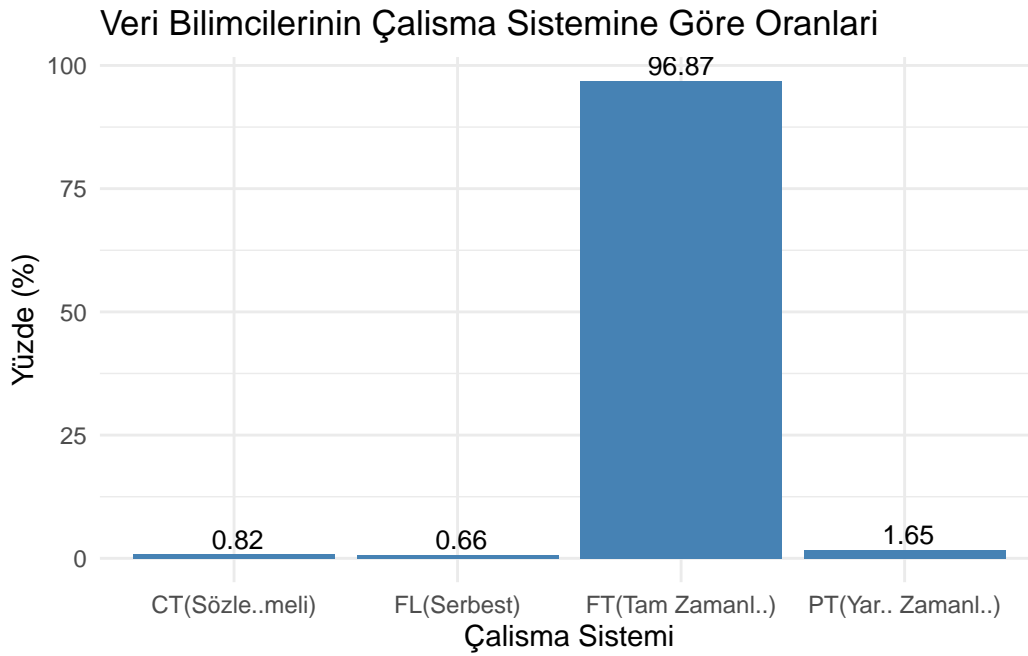


3.2.1 Grafik Yorumu:

Pasta grafiğimizde en büyük kısma sahip veri bilimci tecrübesini oluşturanların üst düzey veri bilimciler(SE - Senior level) olduğu görünmektedir. Bu kısımda en küçük yeri ise uzman veri bilimcilerin(EX-Expert) kapladığını görmekteyiz.

3.3 Veri Bilimcilerinin Çalışma Sistemine Göre Oranları

```
ggplot(data=dataScienceEmploymentStatusPercent,
       aes(x=Employment_Status, y=Percent)) +
  geom_bar(stat="identity", fill="steelblue")+
  labs(x = "Çalışma Sistemi",
       y = "Yüzde (%)",
       title = "Veri Bilimcilerinin Çalışma Sistemine Göre Oranları") +
  geom_text(aes(label=
format(round(dataScienceEmploymentStatusPercent$Percent, 2),
nsmall = 2)), vjust=-0.3, size=3.5)+
  scale_x_discrete(labels = c("CT(Sözleşmeli)", "FL(Serbest)",
                             "FT(Tam Zamanlı)", "PT(Yarı Zamanlı)"))+
  theme_minimal()
```



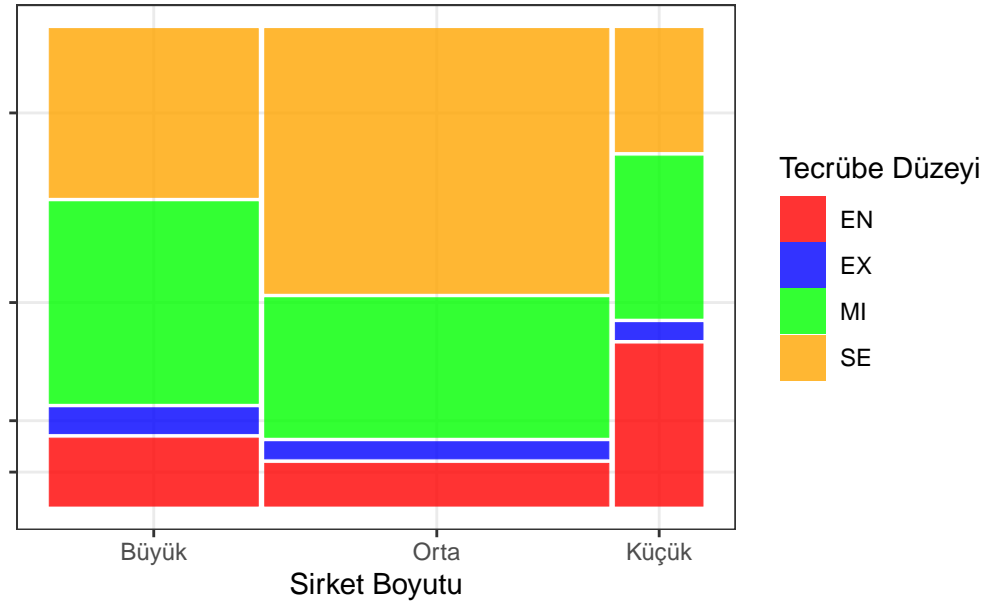
3.3.1 Grafik Yorumu:

Grafikte gözlemlenen en yüksek çalışma sistemi FT(Tam Zamanlı) olarak belirlenmiştir. CT(Sözleşmeli) ve FL(Serbest) çalışanların birbirine çok yakın bir yüzdelik olduğu gözlemlenmiştir. Sistemde bulunan çalışan grubunun neredeyse tamamı FT(Tam zamanlı) çalışmaktadır.

3.4 Firma Büyüklüğüne Ve Tecrübe Düzeyine Göre Veri Bilimci Oranları

```
ggplot(dataScienceMosaic) +  
  geom_mosaic(aes(x = product(Company_Size),  
                  weight = Number,  
                  fill = Experience)) +  
  labs(x = "Şirket Boyutu",  
       y = "",  
       fill = "Tecrübe Düzeyi",  
       title = "Firma Büyüklüğü-Tecrübe Düzeyi Oranı Grafiği") +  
  scale_x_productlist(labels = c("Büyük", "Orta", "Küçük")) +  
  scale_y_productlist(labels = c("", "", "", "")) +  
  scale_fill_manual(values = c("red", "blue", "green", "orange")) +  
  theme_bw()
```

Firma Büyüklüğü-Tecrübe Düzeyi Oranı Grafiği



3.4.1 Grafik Yorumu:

Bu grafikte görüleceği üzere tecrübeli yani yönetici (EX- Executive level /Director) pozisyonunda olan veri bilimcilerin en çok büyük boyutlu şirketlerde bulundukları gözlenmiştir. En tecrübesiz (EN - Entry level / Junior) seviyede veri bilimcilerin ise küçük boyutlu şirketlerde çalıştığı görülmüştür.