Çoklu Veri Setlerinin Oranları

YAZAR YAYINLANMA TARİHİ Emre Söylemez 27 Kasım 2022

ÖZET

Bu raporda önceden belirlenmiş olan 3 veri setine ait veri görselleştirme çalışmaları yer almaktadır. Öncelikle veri setleri tanıtılmış, sonrasında ise bu veri setleri ile alakalı olan araştırma sorularına ilişkin her biri ayrı bölümde yer alacak şekilde grafikler oluşturulmuş ve yorumlanmıştır.

1.Game of Thrones Veri Seti Bu veri setinde Game of Thrones dizisine ait birçok veri bulunmaktadır. 18 farklı değişken ve bu değişkenler için 73 farklı gözlem yer almaktadır. Bu raporda incelenecek olan değişkenler ise

şunlardır:

- 1.Başrol oyuncuları • 2.Başrol oyuncuları
- 3.Başrol oyuncuları Reyting

Başrol Oyuncularının Oranları

1.1. Başrol (Star1) oyuncularının oranlarını veri görselleştirme yöntemleriyle araştırınız.

library(ggplot2) library(dplyr) library(readr) got_ep <- read_csv("GOT_episodes_v4.csv")</pre> got <- got_ep %>% group_by(Star_1) %>% summarise(Episode_n = n()) %>% mutate(Percentage = (Episode_n/sum(Episode_n))*100) ggplot(got, aes(reorder(Star_1,+Percentage),Percentage))+ geom_bar(stat = "identity",fill="red")+ labs(x="Başrol Oyuncuları", y="Yüzde", title = "Başrol Oyuncularının Oranları", subtitle = "İlk Sıradaki Başrol Oyuncularının Bütün Bölümlere Göre Oranları")+ theme_bw()

İlk Sıradaki Başrol Oyuncularının Bütün Bölümlere Göre Oranları 80 60 20 Nikolaj Coster-Waldau Sean Bean Emilia Clarke Kit Harington Peter Dinklage Başrol Oyuncuları Yukarıda ilk sırada ismi geçen başrol oyuncularının bütün bölümlere göre yüzdesel oranını gösteren çubuk grafiği kullanılmıştır. Peter Dinklage adlı oyuncunun %80'e çok yakın bir değerde ve diğer

oyunculara göre açık ara farkla **en yüksek** orana sahip olduğu görülmektedir. Buradan Peter Dinklage adlı oyuncunun dizinin büyük bir kısmında rol aldığı çıkarımı yapılabilir. **Not:** Sadece ilk sırada girilen oyuncu isimleri olduğu için diğer oyuncuların dizinin sadece çok az bir kısmında rol aldığı gibi bir çıkarım, eldeki veriler dahilinde **hatalı** bir çıkarım olur. 1.2. Başrol (Star2) oyuncularının oranlarını reyting değelerlerine göre (iki gruba ayırınız: 8 puan altı ve üstü) veri görselleştirme yöntemleriyle araştırınız.

got1 <- got_ep %>% group_by(Star_2,Rating) %>% mutate(Rating= ifelse(Rating>8, "8.0 puan üstü", "8.0 puan altı")) %>% summarise(Episode_n = n()) %>%

mutate(Percentage = (Episode_n/sum(Episode_n))*100)

25 -

Emilia

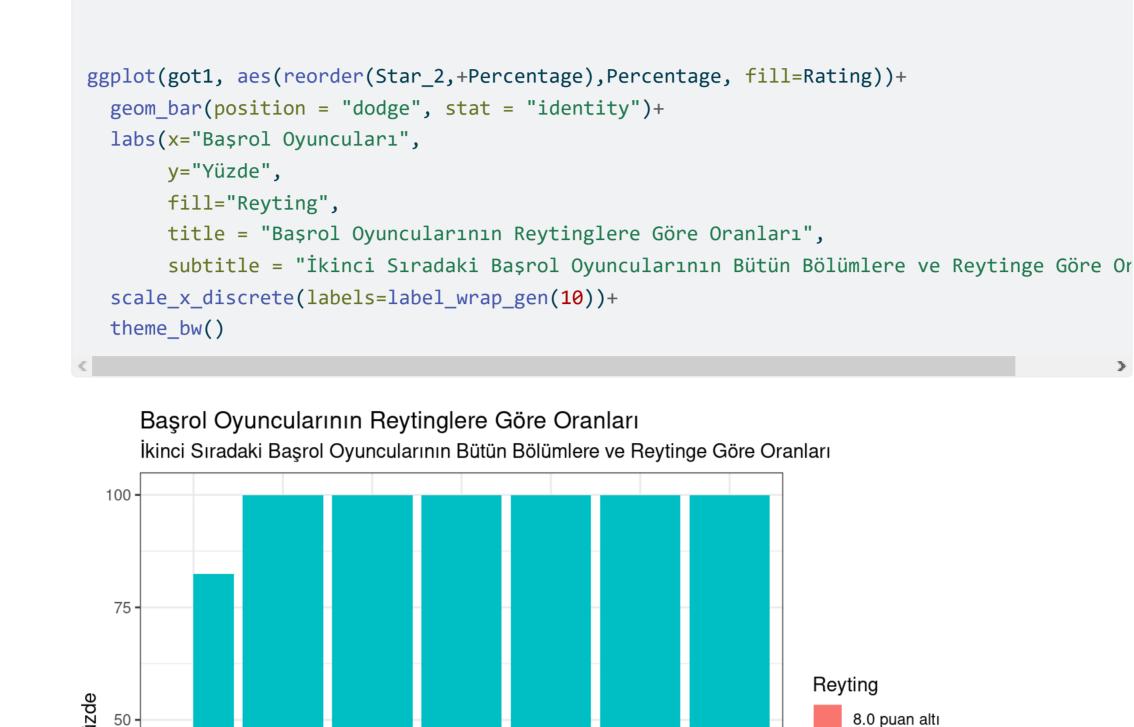
yöntemleriyle araştırınız.

John

Bradley Harington Headey Başrol Oyuncuları

Nikolaj

Coster-Waldau Clarke



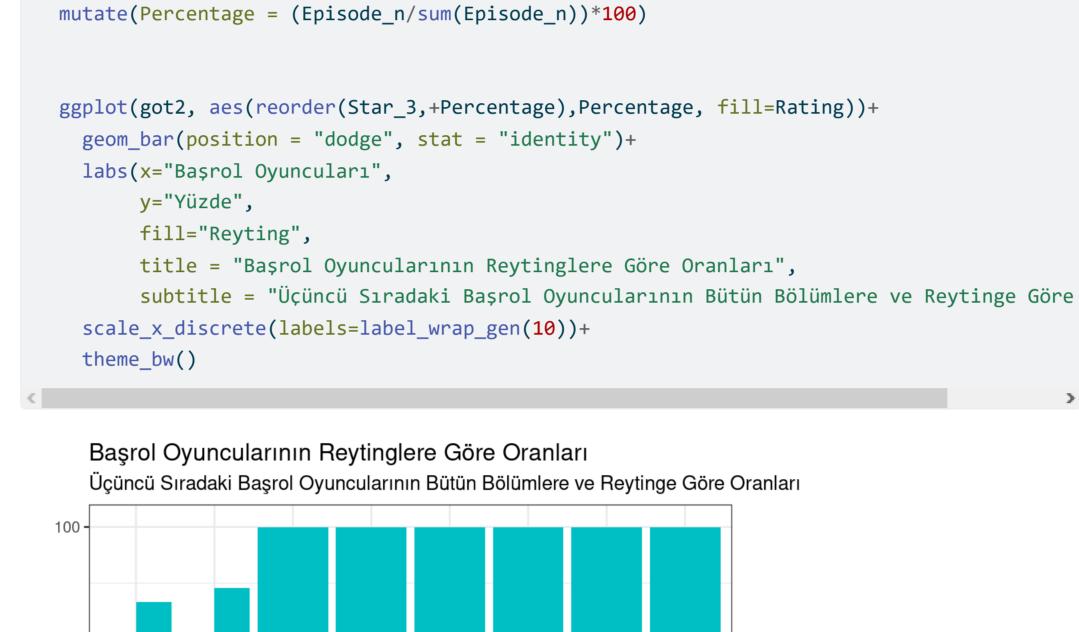
8.0 puan üstü

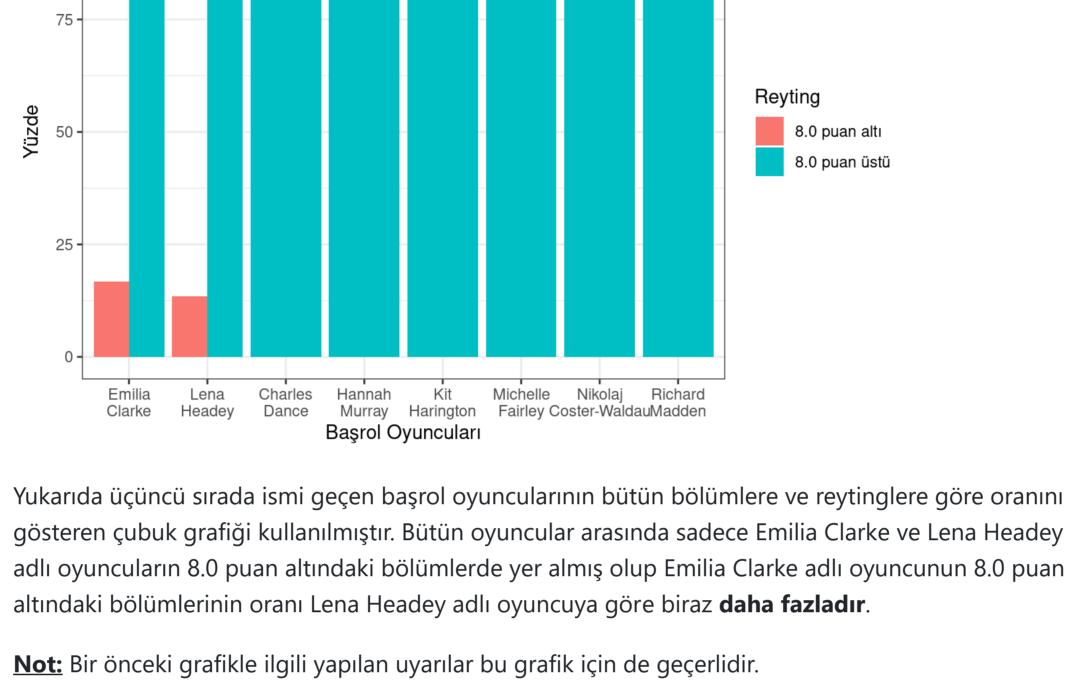
gösteren çubuk grafiği kullanılmıştır. Nikolaj Coster-Waldau adlı oyuncunun 8.0 puan altındaki bölümlerde olan **tek** oyuncu olduğu görülmektedir. **Not:** Bu veride sadece ikinci sırada ismi yazılan oyuncular olduğu için diğer oyuncuların dizi boyunca oynadıkları tüm bölümler 8.0 puan üstüdür şeklinde bir çıkarım yapılması mevcut veriler dahilinde hatalı bir çıkarım olur. Ayrıca Nikolaj Coster-Waldau adlı oyuncunun sadece ikinci sırada isminin yazıldığı bölümler değerlendirildiğinden bu oyuncunun oynadığı bölümlerin diğer oyunculara göre kötü olabileceği ve diğer oyunculardan daha kötü bir oyuncu olabileceği çıkarımı da oldukça **hatalı** bir çıkarım olur. 1.3. Başrol (Star3) oyuncularının oranlarını reyting değelerlerine göre (iki gruba ayırınız: 8 puan altı ve üstü) veri görselleştirme

Yukarıda ikinci sırada ismi geçen başrol oyuncularının bütün bölümlere ve reytinglere göre oranını

Mark Addy Michelle

got2 <- got_ep %>% group_by(Star_3,Rating) %>% mutate(Rating= ifelse(Rating>8, "8.0 puan üstü", "8.0 puan altı")) %>% summarise(Episode_n = n()) %>% mutate(Percentage = (Episode_n/sum(Episode_n))*100)





2.Breaking Bad Veri Seti Bu veri setinde Breaking Bad dizisine ait birçok veri bulunmaktadır. 10 değişken ve bu değişkenler için 62 farklı gözlem yer almaktadır. Bu raporda incelenecek olan değişkenler ise şunlardır:

- Reyting 2.1. Bölümlerin yazar sayısına göre oranlarını veri görselleştirme yöntemleriyle araştırınız.

mutate(Percentage = (Episode_n/sum(Episode_n))*100)

ggplot(brb1, aes(reorder(Writer_n, +Percentage), Percentage))+

library(stringr)

20

summarise(Episode_n = n()) %>%

breaking_bad1 <- read_csv("breaking_bad1.csv")</pre> brb <- breaking_bad1</pre> brb\$Writer_n <- lengths(strsplit(breaking_bad1\$`Written by`, ","))</pre> brb1 <- brb %>% group_by(Writer_n) %>%

```
geom_bar(stat = "identity",fill="red")+
labs(x="Yazar Sayısı",
     y="Yüzde",
     title = "Yazar Sayıları Oranları",
     subtitle = "Yazar Sayılarının Bütün Bölümlere Göre Oranları")+
theme_bw()
  Yazar Sayıları Oranları
  Yazar Sayılarının Bütün Bölümlere Göre Oranları
```

ayırınız: 9 puan altı ve üstü) göre oranlarını veri görselleştirme yöntemleriyle araştırınız. brb2 <- brb %>% mutate(Rating= ifelse(Rating_IMDB>9, "9.0 puan üstü", "9.0 puan altı")) %>% group_by(Writer_n,Rating) %>% summarise(Episode_n= n()) %>% mutate(Percentage= (Episode_n/sum(Episode_n))*100)

ggplot(brb2, aes(reorder(Writer_n, +Percentage), Percentage, fill=Rating))+

geom_bar(stat = "identity", position = "dodge")+

title = "Yazar Sayıları Oranları",

labs(x="Yazar Sayısı",

Yazar Sayıları Oranları

pokedex1 <- read_csv("pokedex1.csv")</pre>

pokedex2<- pokedex1

pkm <- pokedex2 %>%

group_by(type) %>%

summarise(Number = n()) %>%

y="Yüzde",

theme_bw()

olmasına karşın üç yazarlı bölümlerin biraz daha az tercih edildiği görülmektedir.

Yazar Sayısı

Yukarıda yazar sayılarının bütün bölümlere göre oranını gösteren çubuk grafiği kullanılmıştır.

Bölümlerin en fazla üç yazar tarafından yazıldığı bu dizide, bölümlerin %60'ın üzerinde bir oranla iki

yazar tarafından yazıldığı görülmektedir. Üç yazarlı ve tek yazarlı bölümlerin oranı birbirine yakın

2.2. Bölümlerin yazar sayısına ve reyting puanlarına (iki gruba

```
Yazar Sayılarının Bütün Bölümlere ve Reytinge Göre Oranları
60 -
40 -
                                                                                    Rating
                                                                                         9.0 puan altı
                                                                                         9.0 puan üstü
20 -
                                   Yazar Sayısı
```

Yukarıda yazar sayılarının ve reyting puanlarının bütün bölümlere göre oranını gösteren çubuk

görülmektedir. İki yazar ve üç yazar tarafından yazılan bölümlerin reytingleri birbirine çok yakın

olmakla birlikte **en düşük** reyting puanı üç yazar tarafından yazılan bölümlerde görülmektedir.

görülmektedir. En yüksek reyting oranının bir yazar tarafından yazılan bölümlerde olduğu

grafiği kullanılmıştır. Yazar sayısı arttıkça 9.0 reyting puanı altındaki bölümlerin artışına dair bir trend

subtitle = "Yazar Sayılarının Bütün Bölümlere ve Reytinge Göre Oranları")+

3. Pokemon Veri Seti Bu veri setinde Pokemon çizgi filmine ait birçok veri bulunmaktadır. 14 değişken ve bu değişkenler için 801 farklı gözlem yer almaktadır. Bu raporda incelenecek olan değişkenler ise şunlardır: - Jenerasyon - Pokemon Türü - Efsane 3.1. Tiplerine göre Pokemonların oranlarını veri görselleştirme yöntemleriyle araştırınız.

levels = c(1,2,3,4,5,6,7))

levels = c(0,1))

pokedex1\$generation<- factor(pokedex1\$generation, ordered = TRUE,</pre>

pokedex2\$is_legendary<- factor(pokedex1\$is_legendary,</pre>

mutate(Percentage = (Number/sum(Number))*100)

ggplot(pkm, aes(reorder(type, +Percentage), Percentage))+ geom_bar(stat = "identity", fill="red")+ labs(x="Pokemon Tipi", y="Yüzde", title = "Pokemon Tipi Oranları",

subtitle = "Pokemon Tiplerinin Bütün Pokemonlara Oranları")+

theme_bw() Pokemon Tipi Oranları Pokemon Tiplerinin Bütün Pokemonlara Oranları flying fairy ice steel dragon ghost fighting dark groundpoisonelectric rock fire psychic bug grass normal water Pokemon Tipi Yukarıda Pokemon tiplerinin bütün Pokemonlara olan oranlarını gösteren çubuk grafiği kullanılmıştır. 18 pokemon türünün görüldüğü bu grafikte **en yüksek** orana sahip Pokemon türü **Su** pokemonu, **en**

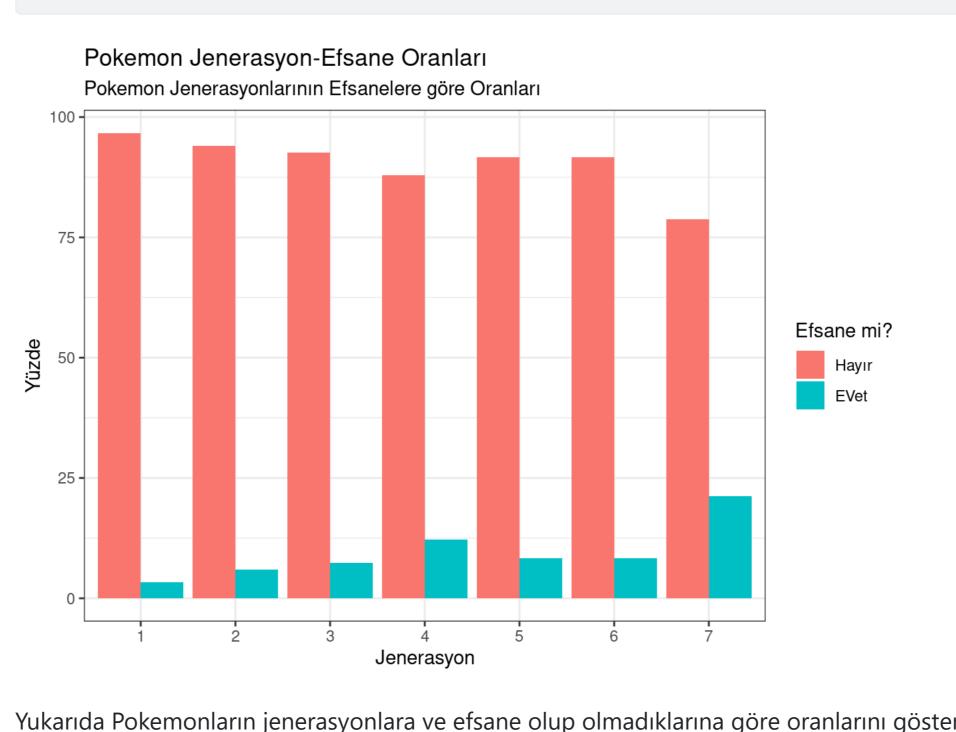
pkm1 <- pokedex2 %>% group_by(generation,is_legendary) %>% summarise(Number = n()) %>% mutate(Percentage = (Number/sum(Number))*100)

oranlarını veri görselleştirme yöntemleriyle araştırınız.

3.2. Jenerasyon ve efsane olup olmadıklarına göre Pokemonların

düşük orana sahip Pokemon türü ise uçan Pokemon türüdür.

```
ggplot(pkm1, aes(generation, Percentage, fill=is_legendary))+
 geom_bar(stat = "identity",position = "dodge")+
 labs(x="Jenerasyon",
      y="Yüzde",
      fill="Efsane mi?",
      title = "Pokemon Jenerasyon-Efsane Oranları",
      subtitle = "Pokemon Jenerasyonlarının Efsanelere göre Oranları")+
 scale_fill_discrete(labels=c("Hayır","EVet"))+
 theme_bw()
    Pokemon Jenerasyon-Efsane Oranları
    Pokemon Jenerasyonlarının Efsanelere göre Oranları
```



Yukarıda Pokemonların jenerasyonlara ve efsane olup olmadıklarına göre oranlarını gösteren çubuk grafiği kullanılmıştır. **En düşük** efsane Pokemon birinci jenerasyonda, **en yüksek** efsane Pokemon oranı ise yedinci jenerasyonda görülmektedir. Efsane pokemonlar ilk dört jenerasyonda artış trendinde iken, beşinci ve altıncı jenerasyonlarda dördüncü jenerasyona göre daha az efsane Pokemon görülmektedir.