Çoklu Veri Seti Dağılımları

YAZAR YAYINLANMA TARİHİ Emre Söylemez 17 Kasım 2022

ÖZET

Yazar

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 Dizisi 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: Sezon • Bölüm süresi Reyting

belirleyiniz. library(ggplot2) library(readr) got_ep <- read_csv("GOT_episodes_v4.csv")</pre> got_ep\$Season<- factor(got_ep\$Season, ordered=TRUE,</pre> levels=c(1,2,3,4,5,6,7,8)) ggplot(got_ep, aes(x=Season, y=Duration))+ geom_boxplot()+ labs(x="Sezon",

1.1.Sezonlara göre bölüm süresi dağılımlarını inceleyiniz. Bölüm

sürelerindeki en yüksek değişimin gözlendiği sezonu

y="Bölüm Süresi", title = "Game of Thrones Dizisi Sezonlara Göre Bölüm Süresi Dağılımı", subtitle = "Kutu Grafiği")+ theme_bw() Game of Thrones Dizisi Sezonlara Göre Bölüm Süresi Dağılımı Kutu Grafiği

Bölüm Süresi 60 50 Sezon Yukarıda sezonlara göre bölüm sürelerinin dağılımını içeren kutu grafiği kullanılmıştır. 2. Sezondan itibaren bölüm süreleri değişimi ile alakalı artış trendi görülmekte olup en yüksek değişimin 8.

Sezonda olduğu görülmektedir.

1. Sezonda gerçekleşmiştir.

subtitle = "Ridgeline Grafiği")+

Ridgeline Grafiği

theme_ridges(font_size = 12)

Not: Son 2 sezonda bölüm sayısındaki azalmanın bu değişimdeki artışa etkisini göz ardı etmemek gerekir. 1.2. Sezonlara göre rating dağılımlarını araştırınız. Rating değişiminin en düşük olduğu sezonu belirleyiniz.

library(ggridges) ggplot(got_ep, aes(y=Season, x=Rating))+ geom_density_ridges2()+

labs(y="Sezon", x="Reyting", title = "Game of Thrones Dizisi Sezonlara Göre Reyting Dağılımı", subtitle = "Ridgeline Grafiği")+ theme_ridges()

Game of Thrones Dizisi Sezonlara Göre Reyting Dağılımı Ridgeline Grafiği

1.3.Birinci yazara göre rating dağılımını araştırınız. Hangi yazarın senaryosunun daha yüksek rating aldığını belirtiniz. ggplot(got_ep, aes(y=Writer_1, x=Rating))+ geom_density_ridges2()+ labs(y="Birinci Yazar", x="Bölüm Reytingi",

title = "Game of Thrones Dizisi Birinci Yazara Göre Bölüm Süresi Dağılımı",

Game of Thrones Dizisi Birinci Yazara Göre Bölüm Süresi Dağılımı

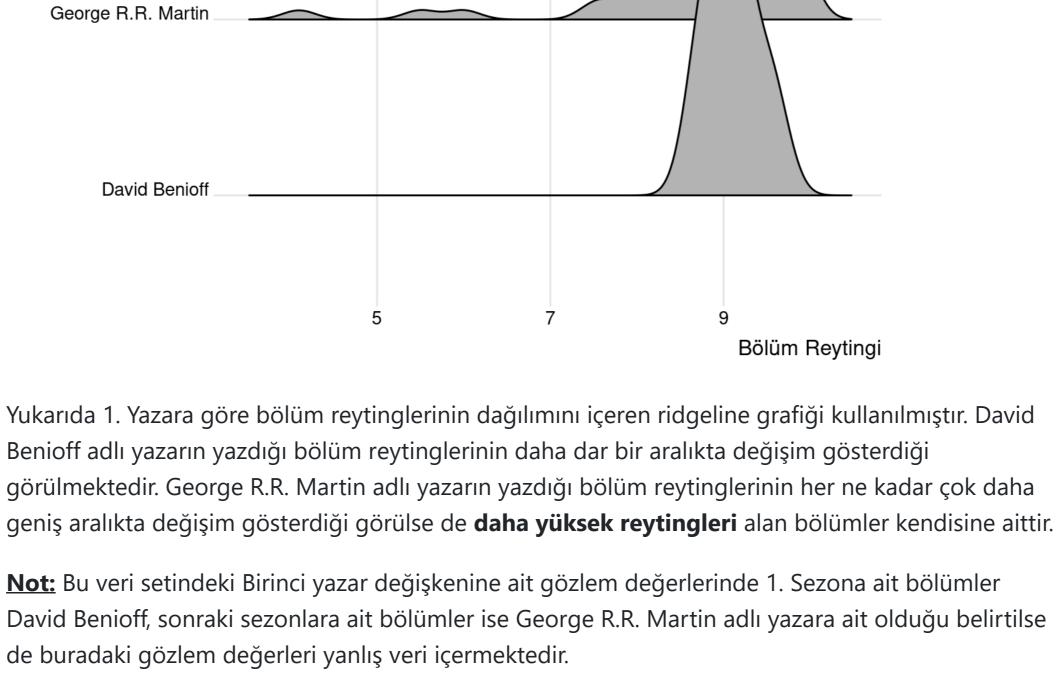
Yukarıda sezonlara göre bölüm reytinglerinin dağılımını içeren ridgeline grafiği kullanılmıştır. En

sezonların reytinglerinin benzer değişim aralıklarında seyrettiği görülmektedir. **En düşük değişim** ise

yüksek değişiminin 8.sezonda olduğunun görüldüğü grafikte 1-4 arası sezonların ve 5-7 arası

Reyting

George R.R. Martin



2.Breaking Bad Dizisi Veri Seti Bu veri setinde Breaking Bad dizisine ait birçok veri bulunmaktadır. 10 değişken ve bu değişkenler

George R.R. Martin adlı yazarın ilk 4 sezonda her sezon birer bölüm olmak üzere yazdığı bölümler

için 62 farklı gözlem yer almaktadır. Bu raporda incelenecek olan değişkenler ise şunlardır: Sezon Bölüm süresi

2.1.Sezonlara göre bölüm süresi dağılımlarını inceleyiniz. Bölüm

sürelerindeki en yüksek değişimin gözlendiği sezonu belirleyiniz. breaking_bad1 <- read_csv("breaking_bad1.csv")</pre>

şunlardır:

• The Pointy End

The Bear and the Maiden Fair

• The Lion and the Rose

• İzlenme sayısı(milyon)

labs(x="Sezon",

y="Bölüm Süresi",

Ridgeline Grafiği

Ridgeline Grafiği

Sezon

Blackwater

Reyting

breaking_bad1\$Season<- factor(breaking_bad1\$Season, ordered = TRUE,</pre> levels=c(1,2,3,4,5)) ggplot(breaking_bad1, aes(x=Season, y=Duration_mins))+ geom_boxplot()+

subtitle = "Kutu Grafiği")+ theme_bw() Breaking Bad Dizisi Sezonlara Göre Bölüm Süresi Dağılımı Kutu Grafiği 55

title = "Breaking Bad Dizisi Sezonlara Göre Bölüm Süresi Dağılımı",

Sezon Yukarıda sezonlara göre bölüm sürelerinin dağılımını içeren kutu grafiği kullanılmıştır. Bu grafikte en dikkat çeken durum, her sezon için bölümlerin genellikle belirli bir süreye sadık kalınarak çekildiği görülmektedir. Grafikte her sezon için bazı uç değerler görülmekte olup, bu uç değerler en çok 4. Sezonda görüldüğü için **en yüksek değişimin** gözlendiği sezon da bu sezon olmaktadır. 2.2.Sezonlara göre rating dağılımlarını araştırınız. Rating değişiminin en düşük olduğu sezonu belirleyiniz. ggplot(breaking_bad1, aes(y=Season, x=Rating_IMDB))+ geom_density_ridges2()+ labs(y="Sezon", x="Bölüm Reytingi", title = "Breaking Bad Dizisi Sezonlara Göre Bölüm Reytingi Dağılımı", subtitle = "Ridgeline Grafiği")+ theme_ridges() Breaking Bad Dizisi Sezonlara Göre Bölüm Reytingi Dağılımı

Yukarıda sezonlara göre bölüm reytinglerinin dağılımını içeren ridgeline grafiği kullanılmıştır. Grafikte en yüksek reytinglerin 5. Sezonda olduğu görülmektedir. 3. ve 4. Sezonlar dizi için reyting değişiminin en yüksek olduğu sezonlar olduğu görülmesine karşın reytinglerin yüksek olduğu dikkat çekicidir. Reyting değişimi 1. Sezonda **en düşük** olarak görülmektedir. 2.3. Sezonlara göre izlenme sayısı dağılımlarını araştırınız. İzlenme sayısı değişiminin en düşük olduğu sezonu belirleyiniz. ggplot(breaking_bad1, aes(y=Season, x=`U.S. viewers_million`))+ geom_density_ridges2()+ labs(y="Sezon", x="İzlenme Sayısı(milyon)", title = "Breaking Bad Dizisi Sezonlara Göre Bölüm Başına İzlenme Sayısı Dağılımı", subtitle = "Ridgeline Grafiği")+ theme_ridges(font_size = 12) Breaking Bad Dizisi Sezonlara Göre Bölüm Başına İzlenme Sayısı Dağılımı

10

Bölüm Reytingi

İzlenme Sayısı(milyon)

0.53(milyon) iken 2. Sezonda ise 0.62(milyon) şeklindedir. Buna göre izlenme sayısındaki değişim 1. Sezonda **en düşüktür. Not:** 2. Sezonda 5 bölüme ait veri olmadığı göz önünde bulundurulmalıdır. 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 Atak gücü • Hız Efsane 3.1.Jenerasyonlarına göre pokemonların atak güçlerinin dağılımını araştırınız. pokedex1 <- read_csv("pokedex1.csv")</pre>

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

Yukarıda sezonlara göre izlenme sayısı(milyon) dağılımını içeren ridgeline grafiği kulanılmıştır.

sayısı değişiminin düşüklüğü bakımından birbirlerine çok yakın olsalar da 1. Sezonda fark

İzlenme sayısındaki en yüksek değişimin 5. Sezonda olduğu görülmektedir. 1. ve 2. Sezonlar izlenme

50

Pokemonların Jenerasyonlara Göre Atak Gücü Dağılımı

title = "Pokemonların Jenerasyonlara Göre Atak Gücü Dağılımı",

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

ggplot(pokedex1, aes(x=attack, y=generation))+

subtitle = "Ridgeline Grafiği")+

geom_density_ridges2()+

x="Atak gücü",

labs(y="Jenerasyon",

Ridgeline Grafiği

labs(y="Jenerasyon",

x="Atak gücü",

fill="Efsane",

ve 7. Jenerasyonda görülmektedir.

jenerasyonu belirleyiniz.

geçememiş olduğu görülmektedir.

theme_bw()

subtitle = "Kutu Grafiği")+

theme_ridges()

Yukarıda jenerasyonlarına göre pokemonların atak güçlerinin dağılımını içeren ridgeline grafiği kullanılmıştır. Grafikte pokemonların atak güçlerindeki **en yüksek değişimin** 7. Jenerasyonda olduğu görülmektedir. 1. Jenerasyon ise atak güçlerinin en düşük olduğu jenerasyondur. Genel olarak her jenerasyonda ortalama güçte yoğunluk olduğu görülse de 7. Jenerasyon **en güçlü** olarak görülmektedir. 3.2.Jenerasyon ve efsane olup olmadıklarına göre atak güçlerinin dağılımını araştırınız. pokedex2<- pokedex1 pokedex2\$is_legendary<- factor(pokedex1\$is_legendary,</pre> levels = c(0,1)) ggplot(pokedex2)+ geom_density_ridges2(aes(x=attack,

group=interaction(generation,is_legendary),

title = "Pokemonların Jenerasyonlara ve Efsane Durumuna Göre Atak Gücü Dağılımı",

fill=is_legendary), alpha=0.7)+

y=generation,

100

150

200

Atak gücü

subtitle = "Ridgeline Grafiği")+ theme_ridges(font_size = 12) Pokemonların Jenerasyonlara ve Efsane Durumuna Göre Atak Gücü Dağılımı Ridgeline Grafiği Efsane 100 50 150 200 Atak gücü Yukarıda jenerasyon ve efsane olup olmadıklarına göre atak güçlerinin dağılımını içeren ridgeline

ggplot(pokedex1, aes(y=speed, x=generation))+ geom_boxplot()+ labs(x="Jenerasyon", y="Hız", title = "Pokemonların Jenerasyonlara Göre Hız Dağılımı",

grafiği kullanılmıştır. Grafikte efsane olmayan pokemonlar "0", efsane olan pokemonlar ise "1" olarak

ikili sisteme göre kodlanmıştır. Her jenerasyon için efsane pokemonların, diğer Pokemonlardan **daha**

güçlü oldukları görülmektedir. Efsane pokemonların atak güçleri arasındaki en yüksek değişimler 3.

3.3.Jenerasyonlarına göre hızlarının dağılımını araştırınız. En hızlı

Pokemonların Jenerasyonlara Göre Hız Dağılımı Kutu Grafiği 150

Jenerasyon Yukarıda jenerasyonlarına göre hızlarının dağılımını içeren kutu grafiği kullanılmıştır. Grafikte 3. Jenerasyonun **en hızlı** olduğu görülmektedir. 4. ve 5. Jenerasyonlar genel Pokemon yoğunluğu açısından daha fazla hızlı Pokemon içermesine rağmen en hızlı Pokemon olarak 3. Jenerasyonu