İlişkilerin ve Zaman Serilerinin Görselleştirilmesi

YAZAR YAYINLANMA TARİHİ Emre Söylemez 4 Aralık 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

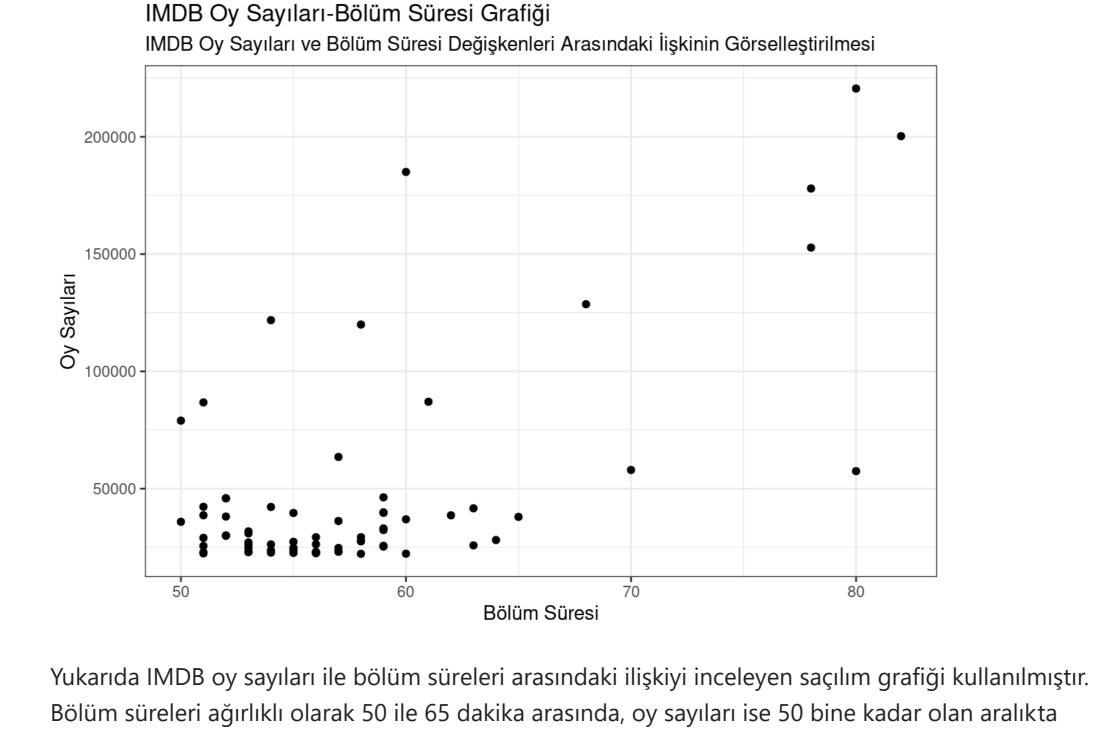
• 2.Başrol

şunlardır:

- Bölüm süresi Oylar
- yöntemleriyle araştırınız.

1.1. Başrol (Star1) oyuncularının oranlarını veri görselleştirme

```
library(ggplot2)
library(readr)
got_ep <- read_csv("GOT_episodes_v4.csv")</pre>
ggplot(got_ep, aes(Duration, Votes))+
  geom_point()+
  labs(title = "IMDB Oy Sayıları-Bölüm Süresi Grafiği",
       subtitle = "IMDB Oy Sayıları ve Bölüm Süresi Değişkenleri Arasındaki İişkinin Görse]
       x="Bölüm Süresi",
       y="Oy Sayıları")+
  theme_bw()
```



yoğunlaştığı görülmektedir. Bu süre aralığında 50 bin ile 200 bin arasında oy sayısı olan bölümler olsa da genele bakıldığında sayıca azdır. Bölüm süreleri 70 dakikanın üzerine çıktığında, oyların çoğu 150 bin üzerinde seyretmektedir. Bu bölümler genele bakıldığında sayıca çok az olduğu için aralarında bir ilişki olduğunu söylemek **hatalı** bir çıkarım olur. Not: Bölüm sürelerinin ve oylarının yüksek olduğu bölümlerin dizinin son sezonunda olduğu ve bu dizinin dünya çapında çok geniş kitlelerce takip edildiği düşünüldüğünde bu durumun da oy

sayılarında artışa etkili olabileceği gözden kaçırılmamalıdır. 1.2. IMDB oy sayıları ve bölüm süreleri arasındaki ilişkiyi, yardımcı oyuncunun (Star2) Nikolaj Coster-Waldau olup olmaması

durumuna göre veri görselleştirme yöntemleriyle araştırınız.

Nikolaj Coster-Waldau'nun yardımcı oyuncu olarak oynayıp

oynamamasının, oy sayıları ve bölüm süreleri arasındaki ilişki

üzerinde bir etkisi var mıdır? ggplot(got_ep, aes(Duration, Votes, color= ifelse(Star_2=="Nikolaj Coster-Waldau", "Evet", geom_point(size=3, alpha=0.6)+ labs(title = "IMDB Oy Sayıları, Bölüm Süresi ve 2.Başrol Oyuncu Grafiği", subtitle = "Oy Sayıları ve Bölüm Süresinin, 2.Başrolün Nikolaj Coster-Waldau Olup Ol x="Oy Sayıları", y="Bölüm Süresi", color="Oyuncu Nikolaj Coster-Waldau mu?")+ theme_bw()+

```
IMDB Oy Sayıları, Bölüm Süresi ve 2.Başrol Oyuncu Grafiği
         Oy Sayıları ve Bölüm Süresinin, 2.Başrolün Nikolaj Coster-Waldau Olup Olmaması Arasındak
   200000
Bölüm Süresi
100000
    50000
                                                             70
                                               Oy Sayıları
```

theme(legend.position = "bottom")

arasındaki ilişkiyi inceleyen saçılım grafiği kullanılmıştır. Nikolaj Coster-Waldau adlı oyuncunun 2. başrol oyuncu olarak genellikle 50-65 dakika arasında ve 50 bine kadar oylama yapılan bölümlerde

Oyuncu Nikolaj Coster-Waldau mu?

Evet

Hayır

Yukarıda IMDB oy sayıları ile bölüm süresinin, ikinci başrolün Nikolaj Coster-Waldau olup olmaması

oynadığı görülmektedir. 70 dakika üzerindeki bölümlerde ise **sadece** kendisi 2. başrol oyuncu olarak

oynamıştır ve bu bölümlerin çoğu 150 binin üzerinde oy sayısına ulaşmıştır. Bu bölümler sayıca çok az olduğu için aralarında bir ilişki olduğunu söylemek **hatalı** bir çıkarım olur. **Not:** Bir üst grafikteki uyarılar bu grafik için de geçerlidir. Ayrıca kendisinin 2. başrol oyuncu olarak yazılmadığı bölümlerde oynamadığına dair bir çıkarım yapamayacağımız için bölüm süreleri ve oy sayıları arasındaki ilişki olduğunu **söyleyemeyiz**.

2.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:

• Tür

pokedex2 <- read_csv("pokedex1.csv")</pre>

mi etkilemektedir?

geom_smooth(se=FALSE)+

x="Atak",

y="H1z")+

geom_point()+

araştırınız.

library(priceR)

library(tidyr)

kayıpları görülmektedir.

currency <- merge(usd, eur, by="date")</pre>

ggplot(currency1, aes(date, value, color=currency)) +

ggplot(pokedex2, aes(attack, speed))+

 Atak gücü • Hız

```
2.1. Pokemonların hızları ve atak güçleri arasındaki ilişkiyi veri
görselleştirme yöntemleriyle araştırınız. Hızlarının artması atak
güçlerini nasıl etkilemektedir?
```

ggplot(pokedex2, aes(attack, speed))+ geom_point()+ geom smooth(se=FALSE)+ labs(title = "Pokemonların Hız ve Atak Güçleri Arasındaki İlişki",

```
y="H1z")+
theme_bw()
   Pokemonların Hız ve Atak Güçleri Arasındaki İlişki
```

150 Atak Yukarıda Pokemonların hızları ve atak güçleri arasındaki ilişkiyi inceleyen saçılım grafiği çizilmiştir. Grafikte trendin yukarı yönlü olduğu görülmektedir. Hızları yükseldikçe atak güçleri de yükselen Pokemonlara dair gözlem değerleri görülmektedir. Hız ve atak güçleri arasında pozitif yönlü bir korelasyon görülmekte olup **orta** şiddettedir. 2.2. Pokemonların hızları ve atak güçleri arasındaki ilişkiyi, tiplerine göre veri görselleştirme yöntemleriyle araştırınız. Farklı tip pokemonların hızlarının artması atak güçlerini aynı düzeyde

```
facet_wrap(~type, ncol = 6)+
theme_bw()
   Pokemonların Hız ve Atak Güçleri ile Türleri Arasındaki İlişki
                                                                 fairy
                                                                              fighting
        bug
100
                      flying
                                                  grass
                                                                ground
                                                                               ice
```

labs(title = "Pokemonların Hız ve Atak Güçleri ile Türleri Arasındaki İlişki",

normal poison water 0 50 100 150 0 50 100 150 Yukarıda Pokemonların hızları ile atak güçleri arasındaki ilişkiyi türlerine göre inceleyen saçılım grafikleri çizilmiştir. Grafiklere bakıldığında genellikle trend yukarı yönlü görülmekle birlikte bazı türlerde dalgalanmalar da görülmektedir. Flying türü için çok az sayıda gözlem değeri bulunmaktadır. Bazı türlerde korelasyon daha kuvvetli iken bazı türlerde daha düşük olduğu görülmektedir. Sonuç olarak her tür bu durumdan farklı şekilde etkilenmektedir. 3. Döviz Kurları Veri Seti Bu veri setinde PriceR paketi kullanılarak bazı ülkelerin para birimlerinin değerlerinin değişimleri son 10 yıl baz alınarak incelenmiştir.

currency <- merge(currency, gbp, by="date")</pre> colnames(currency) <- c("date", "USD", "EUR", "GBP")</pre> currency1<- currency %>% pivot_longer(USD:GBP, names_to = "currency", values_to = "value")

3.1. Son 10 yılda, Türk lirasının, Euro, Amerikan doları ve Pound

usd <- historical_exchange_rates("TRY", to="USD", start_date = "2012-01-01", end_date = "20</pre>

eur <- historical_exchange_rates("TRY", to="EUR", start_date = "2012-01-01", end_date = "20</pre>

gbp <- historical_exchange_rates("TRY", to="GBP", start_date = "2012-01-01", end_date = "20</pre>

karşılıklarının değişimini veri görselleştirme yöntemleriyle

```
geom_line() +
  labs(x="Tarih", y="Kur Değerleri", title=" Son 10 yılda Dolar, Euro ve Sterlin Birimlerini
        color="Para Birimleri") +
  scale_color_discrete(labels=c("Euro", "Sterlin", "Dolar"))+
  theme_bw()
      Son 10 yılda Dolar, Euro ve Sterlin Birimlerinin Türk Lirasına Karşı Değeri
  0.5
  0.4 -
Kur Değerleri
                                                                        Para Birimleri
                                                                         Dolar
  0.2 -
  0.1
```

2016 2012 2014 2018 2020 2022 Tarih Yukarıda Türk lirasının, Euro, Amerikan doları ve Pound karşılıklarının son 10 yıldaki değişimini gösteren zaman serisi grafiği kullanılmıştır. Türk lirasının son 10 yılda bu üç para birimine karşı

düzenli olarak **değer kaybettiği** görülmektedir. 2014, 2018 ve 2021 yıllarında çok sert değer

3.2. Seçtiğiniz 3 farklı döviz kurunun Amerikan doları

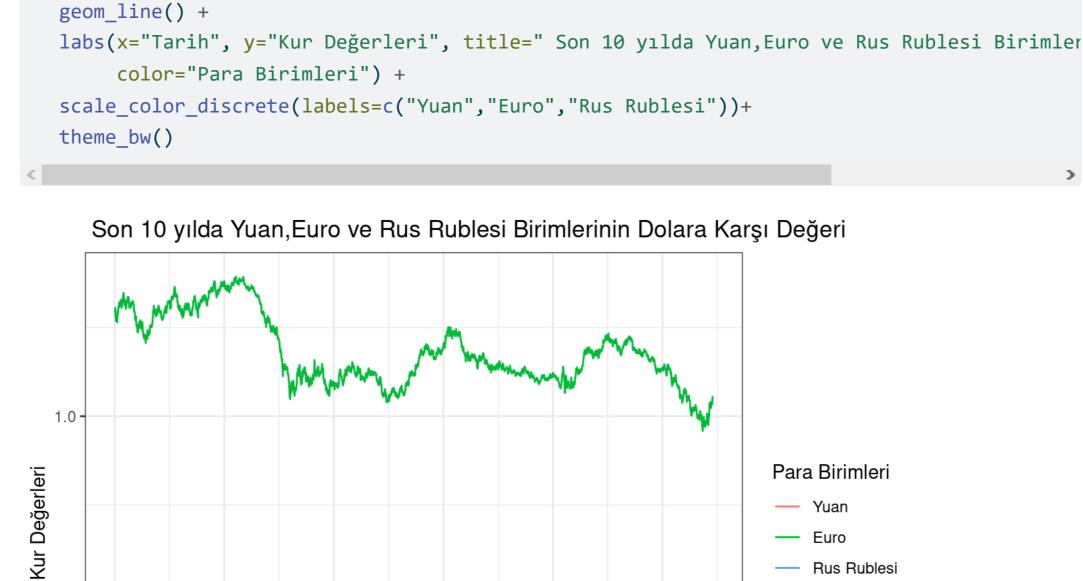
karşılıklarını veri görselleştirme yöntemleriyle araştırınız.

cny <- historical_exchange_rates("CNY", to="USD", start_date = "2012-01-01", end_date = "20</pre>

eur1 <- historical_exchange_rates("EUR", to="USD", start_date = "2012-01-01", end_date = "2</pre>

rub <- historical_exchange_rates("RUB", to="USD", start_date = "2012-01-01", end_date = "20</pre>

currency2 <- merge(cny, eur1, by="date")</pre> currency2 <- merge(currency2, rub, by="date")</pre> colnames(currency2) <- c("date", "CNY", "EUR", "RUB")</pre> currency2<- currency2 %>% pivot_longer(CNY:RUB, names_to = "currency", values_to = "value") ggplot(currency2, aes(date, value, color=currency)) + geom_line() + color="Para Birimleri") +



Euro

Rus Rublesi

2020 2016 2018 2022 2012 2014 Tarih Yukarıda Yuan, Euro ve Rus rublesinin son 10 yılda Amerikan dolarına karşı değişimini gösteren zaman serisi grafiği kullanılmıştır. Euronun, Yuan ve Rus rublesine karşı açık ara değerli bir para birimi olduğu görülmektedir. Ayrıca Euro 2022 yılının sonlarında yaşadığı geçici düşüş dışında, dalgalanmalar görülmesine rağmen Amerikan dolarından daha değerli kalabilmeyi başardığı görülmektedir. Yuan ve Rus rublesi ise Amerikan dolarına karşı çok değersiz görülmesine rağmen

stabil bir şekilde kaldığı görülmektedir. Grafikte Amerikan dolarına karşı en değersiz paranın Rus

rublesi olduğu görülmektedir.