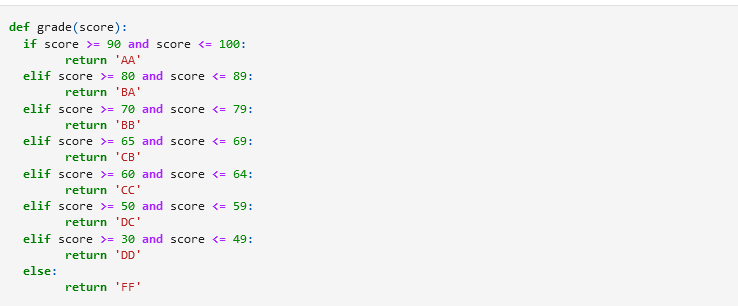
**PYTHON AI BOOTCAMP**

**Grup 10**

**1.Ödev**

Grup olarak kendimize matematik dersini seçtik. Geçme ve kalma olayını da kodda gösterdik. *Şekil 1.1*



*Şekil 1.1*

Bilgilerin girilmesi ve tutulması için başlarda Excel’den çekelim diye konuşmuştuk ama sonrasında bilgiler girileceği için Shell üzerinden bilgi alınmasını daha doğru bulduk. Bunu bir while döngüsüne sokup etkileşimli bir şekle çevirdik. *Şekil 1.2*

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Şekil 1.2*

Sonrasında bu kodu yazmak için ne tür bir veri tipi kullanacağımıza karar verdik. Yani herkes kendi istediği veri tipinde yazdı. En sonunda tek bir veri tipi seçtik.

Öğrencilerin geçti veya geçmedi diye etiketlenmesi için sınav notlarını kullandık.

En sonunda switch-case yapısı ile veri girdisi veya çıktısı alarak kodu dinamik hale getirdik.

Dataframe’i oluşturmak içinse Pandas kütüphanesini dahil ettik ve sonrasında built-in function ile de kolayca excele çevirdik.

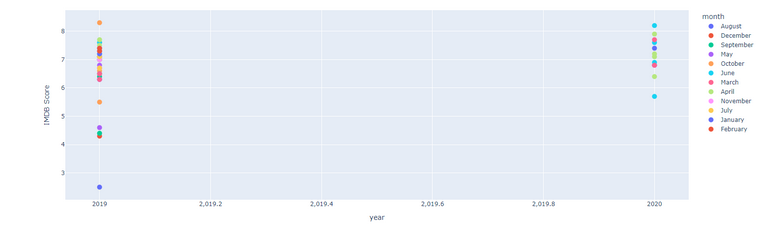
**2.Ödev**

Kaggle’dan aldığımız dataset ile beraber üzerinde işlemler yapmaya başladık. Soru soru ödevi yaptık.

* 1.soru için uzun soluklu film deyimi için öncelikle filmlerin dakikalarının medyan değerinden ne kadar uzak veya yakın olmasına göre değerlendirdik. Bu değeri seçerken direkt medyan değerini değil de medyan değerinden biraz daha fazla olan bir değeri alarak sorgunun verimini artırmış olduk. *Şekil 2.1*



*Şekil 2.1*

* 2. soru belirli bir tarih aralığındaki belgesel tipinde çekilmiş olan filmlerin IMDB değerlerini görselleştirmek. *Şekil 2.2*

*Şekil 2.2*

* 3.soru İngilizce dilinde çekilmiş en yüksek IMDB değerine sahip olan filmler. *Şekil 2.3*

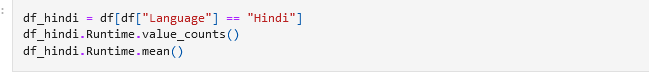
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Şekil 2.3*

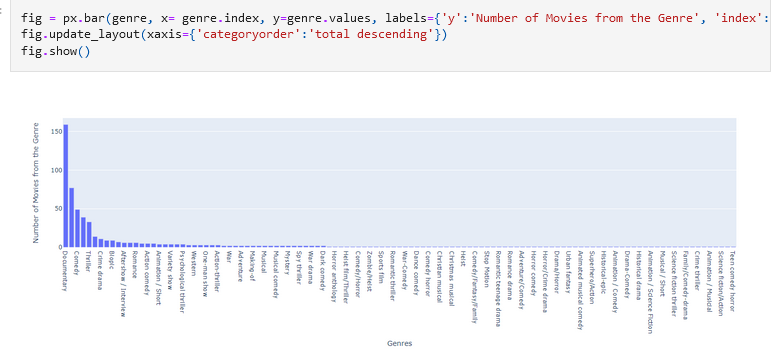
* 4.soru Hindu dilinde çekilmiş filmlerin ortalama süresi ne kadardır diye sorulmuş.

*Şekil 2.4*



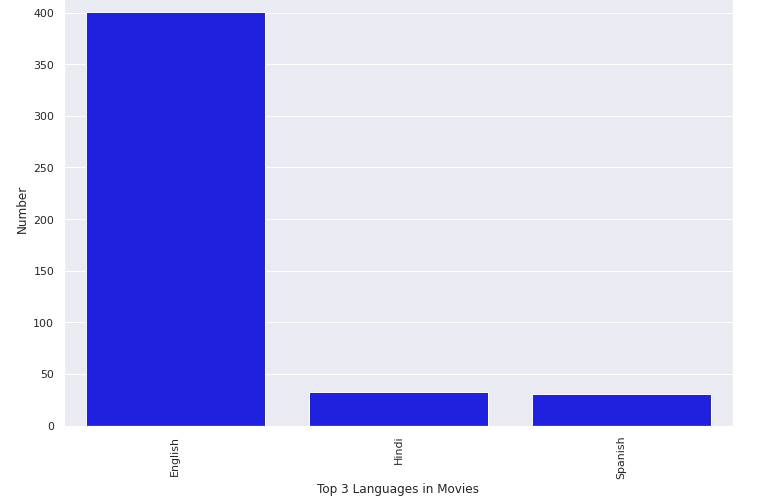
*Şekil 2.4*

* 5.soru kategori sütununda kaç değer vardır ve bunlar nedir? *Şekil 2.5*



*Şekil 2.5*

* 6.soru dataset içindeki en çok kullanılan 3 dil nedir ? *Şekil 2.6*



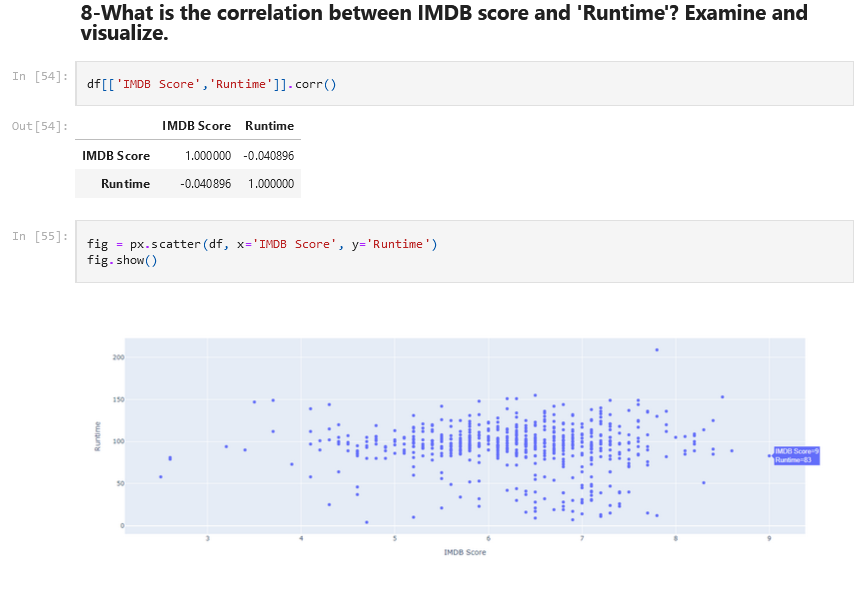
*Şekil 2.6*

* 7.soru en yüksek IMDB değerine sahip ilk 10 film. *Şekil 2.7*

metin içeren bir resim

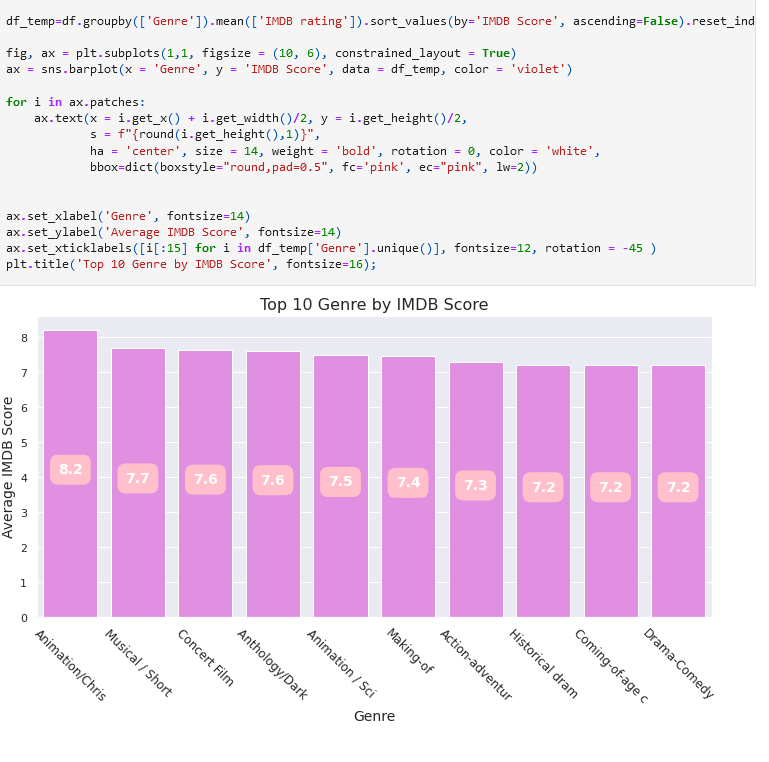
Açıklama otomatik olarak oluşturuldu

* 8. soru için IMDB değerleri ve uzunluk arasındaki ilişki nedir diye sorulmuş. *Şekil 2.8*



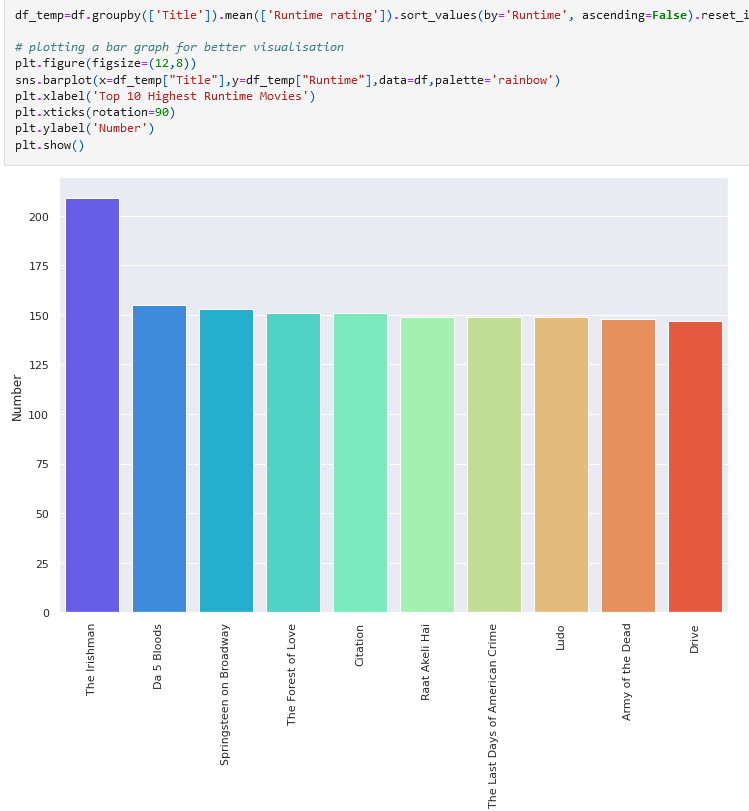
*Şekil 2.8*

* 9.soru IMDB skoruna göre ilk 10 kategori. *Şekil 2.9*

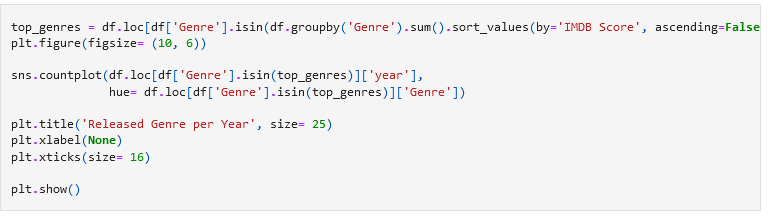


*Şekil 2.9*

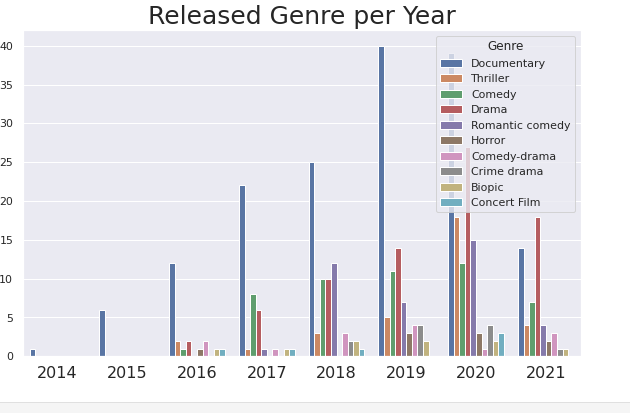
* 10.soru en uzun süreye sahip ilk 10 film. *Şekil 2.10*



* 11.soru yıllara göre hangi yılda en çok film yayınlanmıştır. *Şekil 2.11 ve Şekil 2.12*



*Şekil 2.11*

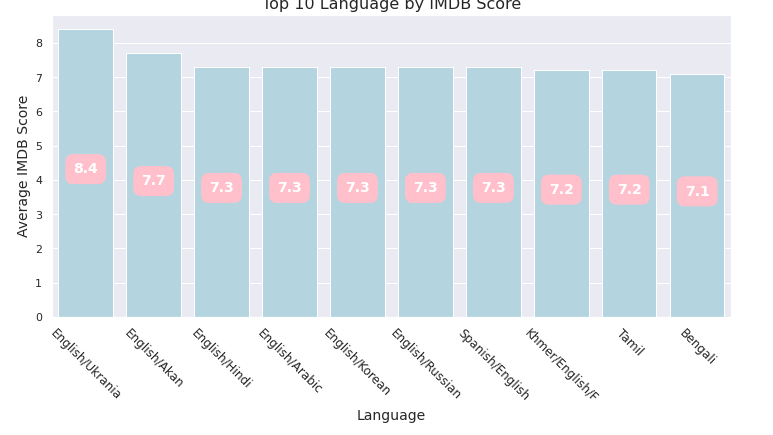
**

*Şekil 2.12*

* 12.soru hangi dildeki filmler en az IMDB ortalamasına sahiptir. *Şekil 2.13 ve Şekil 2.14*

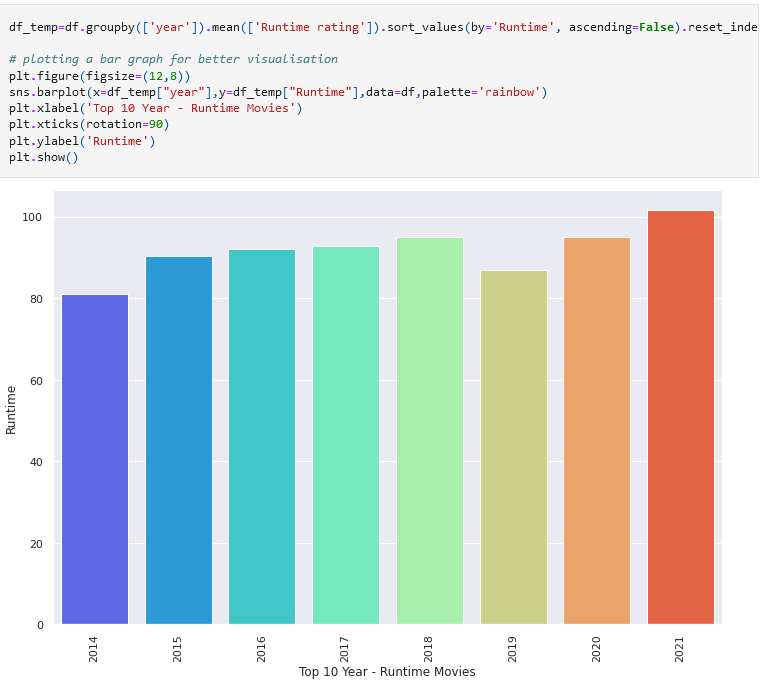


*Şekil 2.13*

**

*Şekil 2.14*

* 13.soru hangi yıl toplam dakikalarda daha fazladır. *Şekil 2.15*



*Şekil 2.15*

* 14.soru her bir dilin en fazla kullanıldığı kategori nedir? *Şekil 2.16*

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Şekil 2.16*

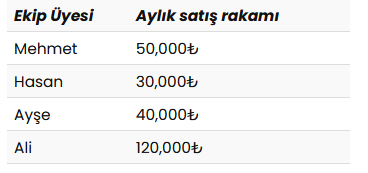
* 15.soru veri setinde Outlier veri var mıdır ? Outlier veriyi açıklayınız.

Bu veri setinde outlier veri vardır fakat istenilen veya hesaplanan değerlere göre değişkenlik gösterir. Bu yüzden bu değerleri hesaplanırken katmamak daha verimli sonuç elde etmemizi sağlar.

Bir diğer açıklama diğer gözlemlerden kayda değer derecede uzak olan gözleme aykırı veya uç değer denir. Aykırı gözlemler, veri setinin geri kalanından farklı davranır ve bu nedenle dikkat çeker.

Aykırı değerlerin ne kadar yanıltıcı olabileceğine dair şöyle bir örnek verebiliriz:

Diyelim ki, bir satış ekibini yönetiyorsunuz ve ekibinizin aylık satış rakamlarını inceliyorsunuz. *Şekil 2.17*

**

*Şekil 2.17*

Ekibinizin aylık satış ortalaması, 60,000 lira. Bu değere bakarak, ekibinizin satış hedefini artırırsanız, ciddi bir sorunla karşılaşacaksınız. Ali dışında hiç kimse 60,000 lira bile satış yapmıyor. Bu veri setinde, aykırı bir gözlem olan Ali’nin satış rakamları ekibin performansı hakkında yanlış çıkarımlar yaparak hatalı kararlar almanıza sebep olacaktır.

Aykırı değerler çok farklı sebeplerden kaynaklanabilir. Örnek olarak:

1. Veri girişi veya ölçüm sırasındaki hatalar.
2. Verideki bozulmalar
3. Gerçekten aykırı bir performans gösterimi (Ör: Cristiano Ronaldo, Michael Jordan vs.) veri setinde aykırı değerlere sebep olabilir.

Her aykırı değerin mutlaka hata olmadığının altını bir kez daha çizmek gerekir. Bazı durumlarda, aykırı gözlemler, veri setindeki yüksek varyansın (değişkenliğin) habercisidir.

Aykırı değerler nasıl tespit edilir? Aykırı değer analizi (outlier detection) nasıl yapılır?

Aykırı gözlemlerin tespitinde kullanılacak yöntemin seçimi, verinin hacmi, dağılımı ve değişken sayısına bağlı olarak değişmektedir.

Aykırı değerleri tespit etmek için kullanılacak yöntemler değişken sayısına göre tek değişkenli, çok değişkenli, dağılımına göre parametrik (bir dağıl veya parametrik olmayan testler (herhangi bir dağılımdan bağımsız) ve istatistiksel veya grafiksel yöntemler olarak farklı gruplara ayrılabilir.

Daha fazlası için lütfen [bakınız](https://ravenfo.com/2021/02/11/aykiri-deger-analizi/).