

Filmlerin Rating Değerlerini Tahmin Etme Çalışması

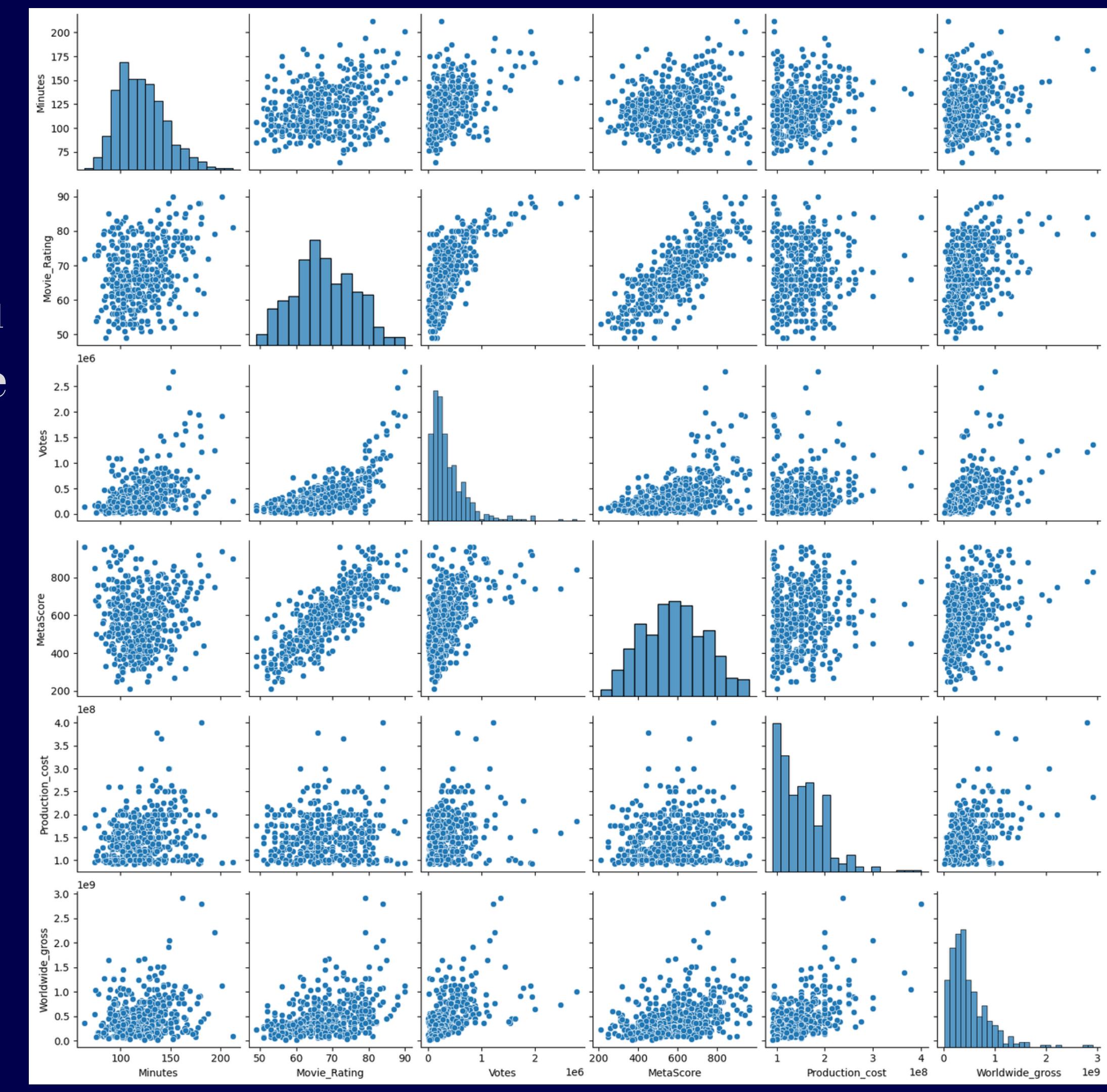
07.04.2024

Bu sunumda her bir film için farklı özellikleri göz önüne alarak rating değerlerini Regresyon Modelleri ve Validasyon yöntemleri kullanarak tahmin etmeye çalışacağız

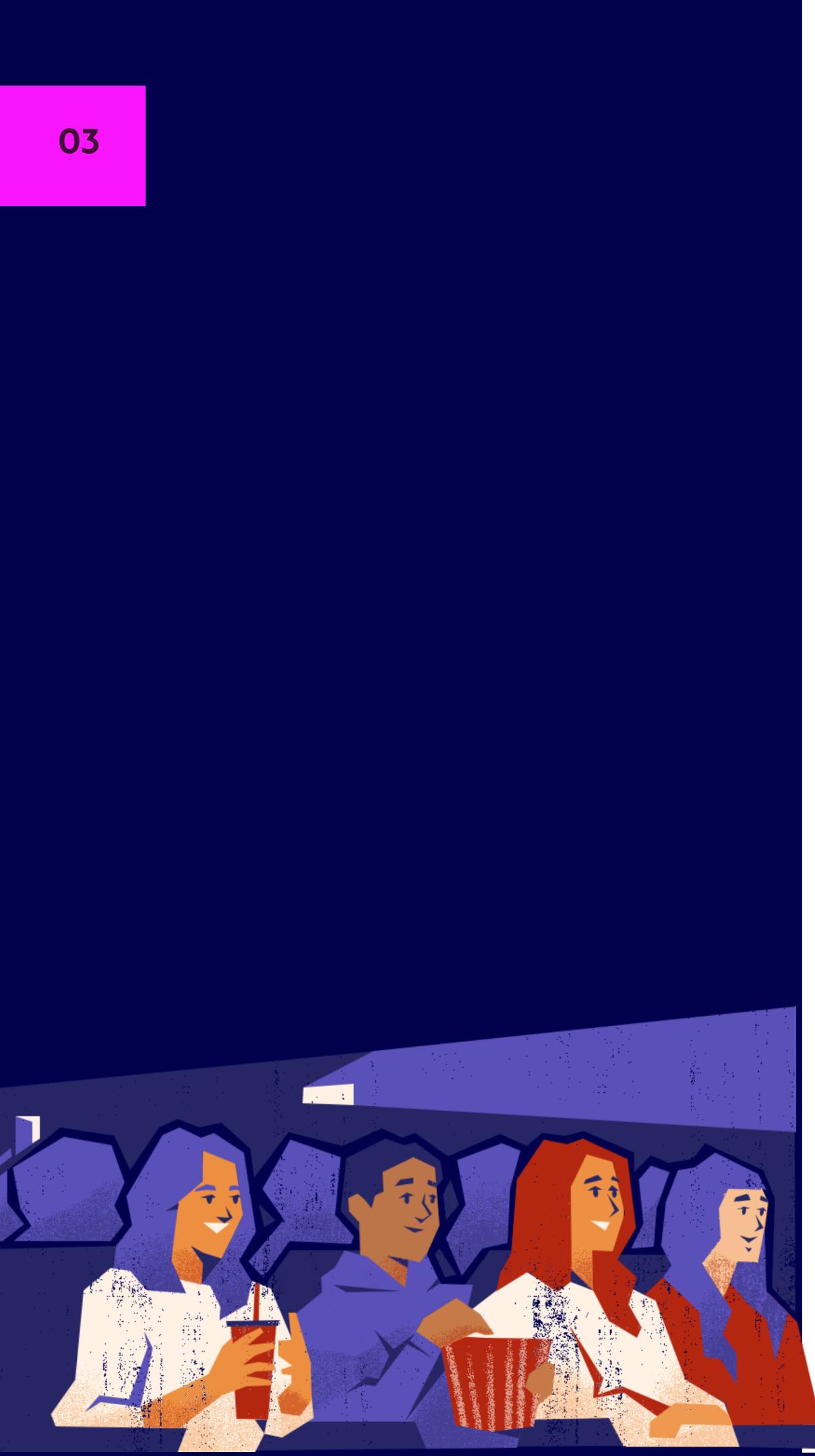


02

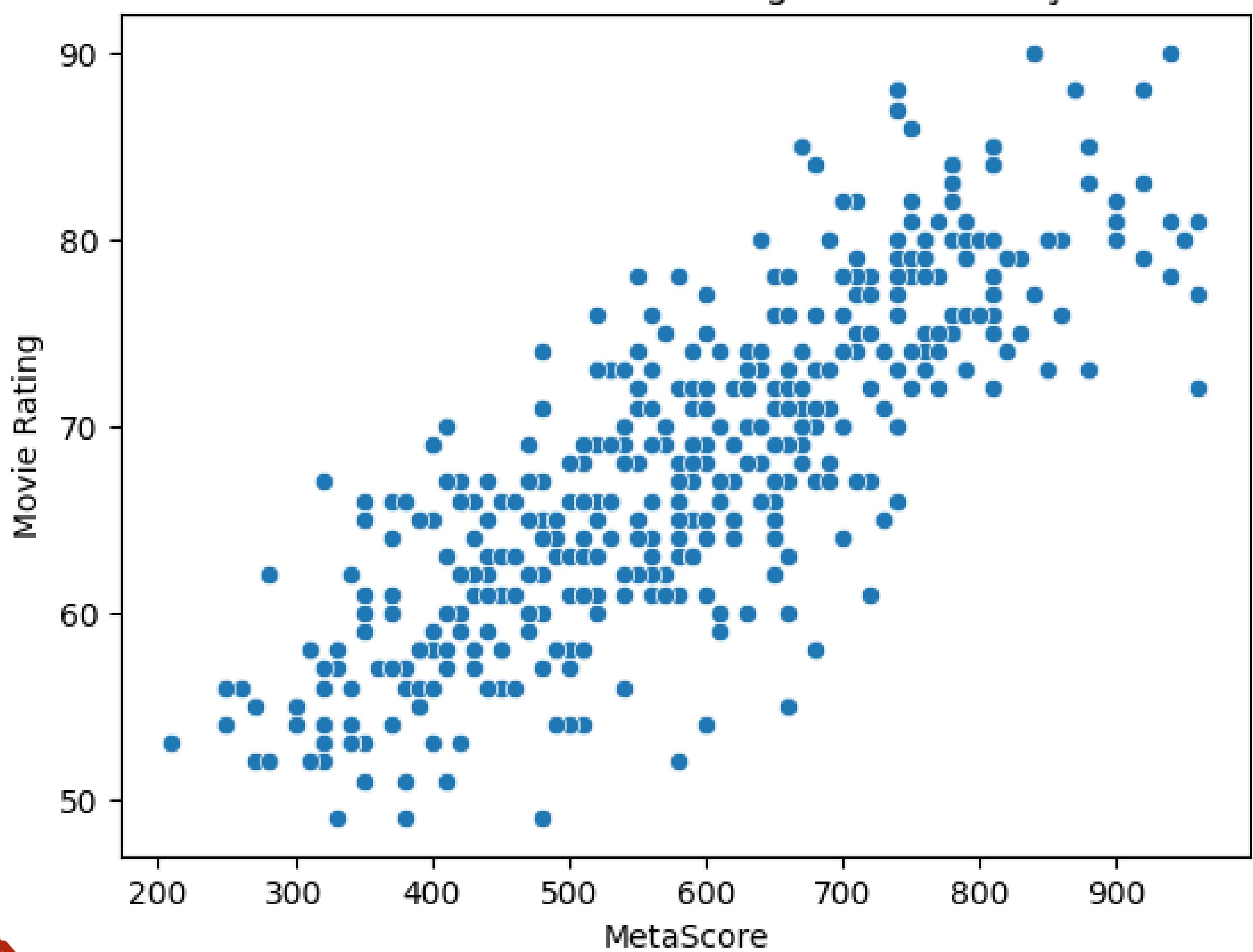
Veri setimizi incelediğimizde
verilerimizin linear olarak dağıldığı
alanların Movie Rating ve Metascore
arasında olduğunu ve bu alanlar
arasında ki ilişkiyi ön plana
almalıyız



MetaScore ile Movie Rating Arasındaki İlişki

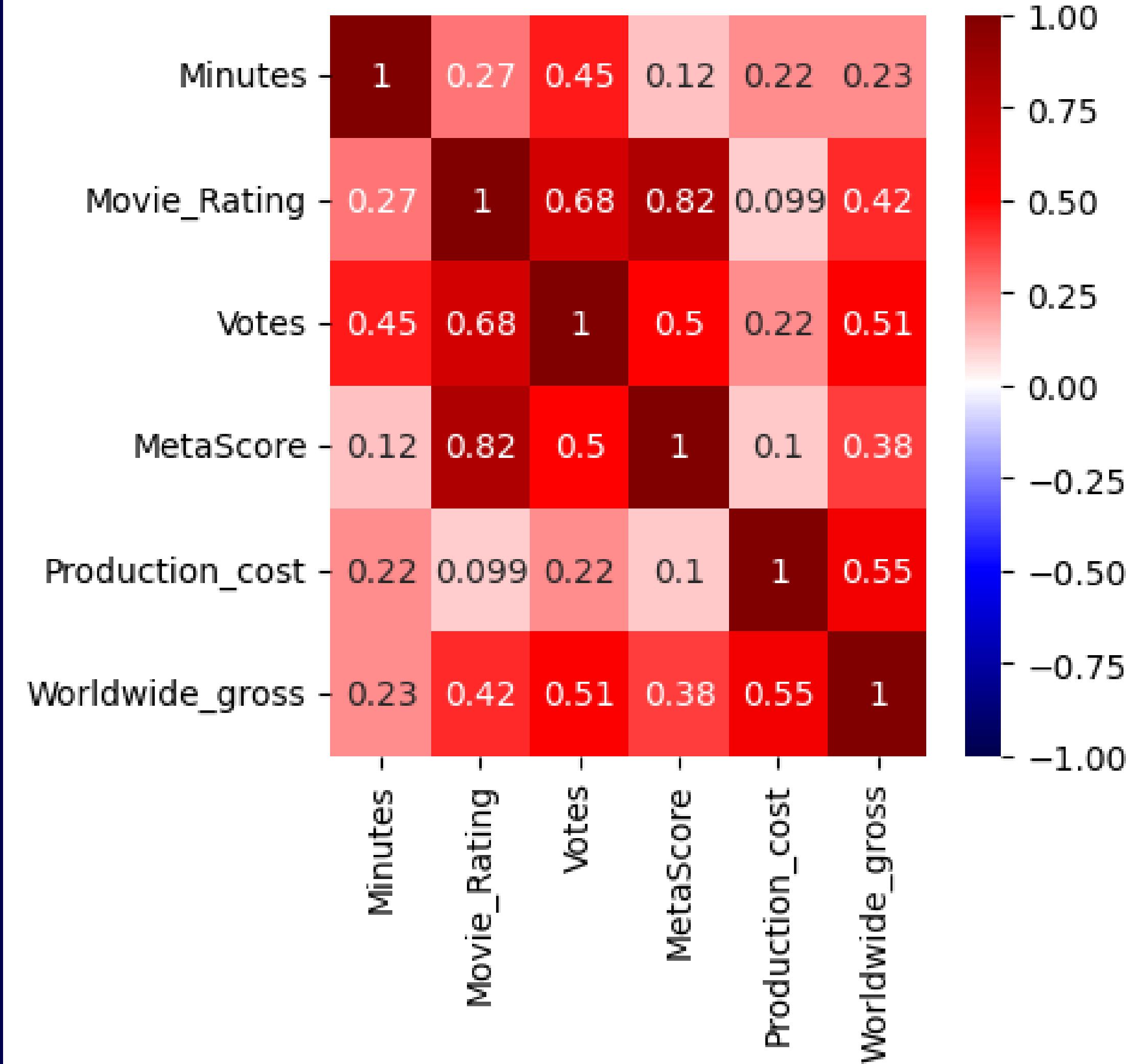


03



04

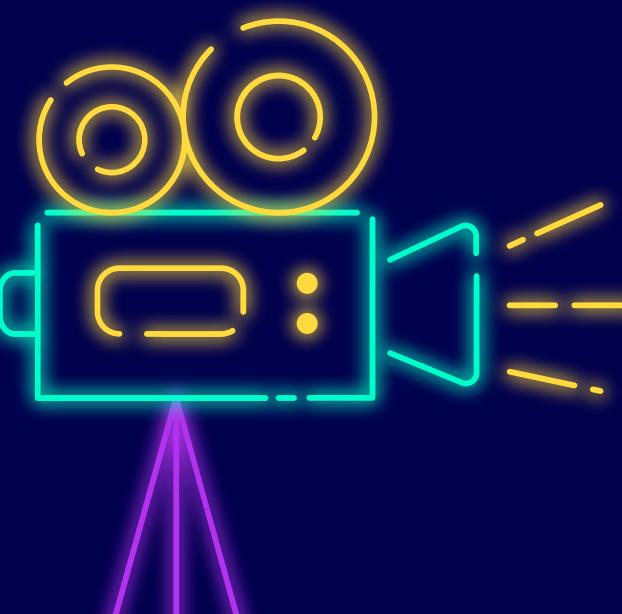
Heatmap ile ısı haritasına bakarak hangi featureslerin seçilebilirliği ve aralarında kurulabilecek ilişkiler hakkında bize bilgi verebilir. Burada da Metascore ve Rating arasında 0.82lik bir değer ortaya çıktı



```
x = df.drop(columns=['Movie_Rating', 'Production_cost', 'Worldwide_gross', 'Minutes'])
X= sm.add_constant(x)
y = df['Movie_Rating']
new_model =sm.OLS(y, X).fit()
new_model.summary()
```

```
new_model2 = smf.ols('Movie_Rating ~ MetaScore + Votes' , data=df).fit()
new_model2.summary()
```

| | | | |
|----------------|--------------|-----------------|-------|
| Dep. Variable: | Movie_Rating | R-squared: | 0.764 |
| Model: | OLS | Adj. R-squared: | 0.763 |



Veri setimizi eğitirken 2 farklı eğitim sonucunda R^2 değerinde 0.764 gibi bir sonuç ile karşılayıyoruz

Şimdi verimizi parçalara ayırarak eğitilmiş veri alanı dışından öğrenmemizi test ederek bir sonuç elde etmeye çalışalım



Test için veri ayırma



```
x_train, x_test, Y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

Validasyon için veri ayırma



```
x_train, x_val, y_train, y_val = train_test_split(x_train, Y_train, test_size=0.25, random_state=42)
```

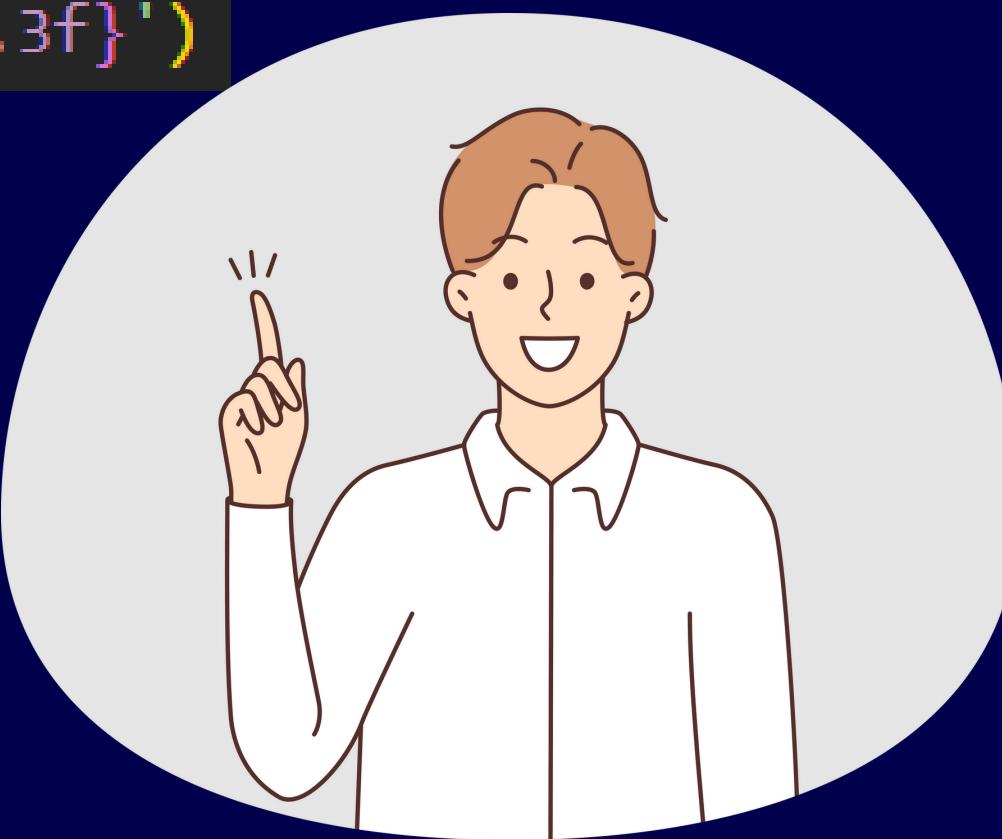
Lineer, Ridge ve Polinomal Regresyon modelleri ile train, test ve val değerleri üstünden yapılan çalışmalarda R² değerleri farklılık göstermektedir.

```
lm.fit(x_train, y_train)
print(f'Linear Regression R2: {lm.score(x_val, y_val):.3f}')

lm_reg.fit(x_train_scaled, y_train)
print(f'Ridge Regression R2: {lm_reg.score(x_val_scaled, y_val):.3f}')

lm_poly.fit(x_train_poly, y_train)
print(f'Polynomial Regression R2: {lm_poly.score(x_val_poly, y_val):.3f}')
```

```
Linear Regression R2: 0.806
Ridge Regression R2: 0.806
Polynomial Regression R2: 0.824
```



Son olarak projemizde Cross Validation kullanarak train kısmının eğitiminde iyileştirme yapabiliriz

| Model | Cross Validated Score (Mean) | Cross Validated Score (Std) |
|-----------------------|------------------------------|-----------------------------|
| Ridge Regression | 0.760425 | 0.05629 |
| Polynomial Regression | 0.760402 | 0.05654 |

En yüksek R^2 regresyon değerleri göz önüne alınarak veri seti içinde cross validation skoru için R^2 değeri en yüksek Polinomal Regresyonun Cross Validation sonrası Ridgeden daha düşük bir skora sahip olduğunu ama standart sapmasının daha yüksek olduğunu görebiliriz.

Sonuç olarak



Metascore verisi ile Movie Rating değerini %76 doğruluk oranında tahmin edebiliriz.

Tahmin yöntemi olarak Ridge veya Polinomal Regresyon kullanılabilir

Polinomal Regresyonda 2. dereceden fonksiyonun değeri hesaba katılmalıdır.

Öğrenmenin verimli olması için Cross Validation kullanılması önemlidir

