

Metal Müzik Puanları, Puanlar-İnceleme İlişkisi

Bu başlık altında Amazon sitesi üzerinde yapılmış olan en iyi metal müzik oylaması ve bu oylamanın inceleme sayısı ile ilişkisinin veri analizini yapacağız.

- 1) Öncelikle kullanacağız python kütüphaneleri eklemekle başlayalım. Numpy kütüphanesi bilimsel hesaplamaları hızlıca yapmamızı sağlayan kütüphanedir. Seaborn kütüphanesi veri görselleştirme konusunda bize yardımcı olacak olan kütüphanedir. Pandas kütüphanesi de veri işler, veri yükler, veri analiz eder.

```
import numpy as np
import seaborn as sns
import pandas as pd
```

- 2) Hangi dataset üzerinden çalışacağımızı belirledik.

```
df=pd.read_csv("metal.csv")
```

- 3) Datasetin ilk 10 çerçevesini görüntüleyelim.

```
df.head(10)
```

	artist	asin	media	review_count	star_rating	title	year
0	Motörhead	B000A14OI8	Audio CD,MP3 Music	136	4.9	BBC Live & in Session	2008
1	Inquisition	B08KMFYQMV	Audio CD,MP3 Music,Vinyl	134	4.9	Black Mass for a Mass Grave	2020
2	King Buffalo	B09JVM31QY	Vinyl,MP3 Music,Audio CD	130	4.9	Dead Star	2022
3	Helloween	B09RKMZDKR	Vinyl,MP3 Music	126	4.9	Helloween Blue/White Marbled	2022
4	Scar Symmetry	B00190KZBO	Audio CD,MP3 Music	125	4.9	Holographic Universe	2008
5	Blind Guardian	B07FGQRF93	Audio CD,Vinyl	124	4.9	Somewhere Far Beyond remixed 2012 / Remastered...	2018
6	Graveyard	B007VMSF6G	Vinyl,MP3 Music,Audio CD	123	4.9	Graveyard	2012
7	Unleash the Archers	B06XYTMFBL	Vinyl,MP3 Music,Audio CD	124	4.9	Apex	2017
8	King Diamond	B08FS37YN6	Audio CD,MP3 Music,Vinyl,Audio, Cassette	123	4.9	The Dark Sides	2020
9	Five Finger Death Punch	B07W5CHDP4	Vinyl	122	4.9	Wrong Side Of Heaven V1	2019

4) Datasetinin kaç sütun ve kaç satırdan oluştuğunu görmek istedim. Datasetimiz 6380 satır ve 7 sütundan oluşmaktadır.

```
df.shape
```

```
(6380, 7)
```

5) Değişkenlerin hangi tipte olduğunu ve bellek yönetimini görelim.

```
df.info
```

```
<bound method DataFrame.info of
0      Motörhead      B000A140I8      Audio CD,MP3 Music      136
1      Inquisition      B08KMFYQMV      Audio CD,MP3 Music,Vinyl      134
2      King Buffalo      B09JVM31QY      Vinyl,MP3 Music,Audio CD      130
3      Helloween      B09RKMZDKR      Vinyl,MP3 Music      126
4      Scar Symmetry      B00190KZBO      Audio CD,MP3 Music      125
...      ...      ...      ...      ...
6375      Whitesnake      B000006TNQ      Audio CD,Audio, Cassette      92
6376      Bowie, David      B00004Y7WV      Audio CD      92
6377      Atheist      B000AL8VNW      Audio CD,MP3 Music,Vinyl      92
6378      Converge      B002N1AE8W      Audio CD,MP3 Music,Vinyl      92
6379      FM      B00THJPIB6      Vinyl,MP3 Music,Audio CD      92

      star_rating      title      year
0      4.9      BBC Live & in Session      2008
1      4.9      Black Mass for a Mass Grave      2020
2      4.9      Dead Star      2022
3      4.9      Helloween Blue/White Marbled      2022
4      4.9      Holographic Universe      2008
...      ...      ...      ...
6375      4.7      Restless Heart      1998
6376      4.7      Bowie at Beeb: Best Of Of BBC Radio 68-72      2000
6377      4.7      Piece of Time      2005
6378      4.7      Axe To Fall      2009
6379      4.7      Heroes And Villains      2017

[6380 rows x 7 columns]>
```

6) df.describe() koduyla datasetimizin istatistiksel değerlerini görüntüleyelim.

```
[ ] df.describe()
```

	review_count	star_rating
count	6380.000000	6380.000000
mean	485.902978	4.765752
std	944.222457	0.067387
min	18.000000	4.700000
25%	104.750000	4.700000
50%	208.000000	4.800000
75%	496.000000	4.800000
max	28547.000000	5.000000

7) Eksik değerleri görmek için bir kod yazalım. Görüldüğü üzere bu datasetinde “year” değişkeninin 250, “artist” değişkeninin 20 eksik değeri var.

```
[ ] df.isna().sum()
```

```
artist      20
asin         0
media        0
review_count 0
star_rating  0
title        0
year        250
dtype: int64
```

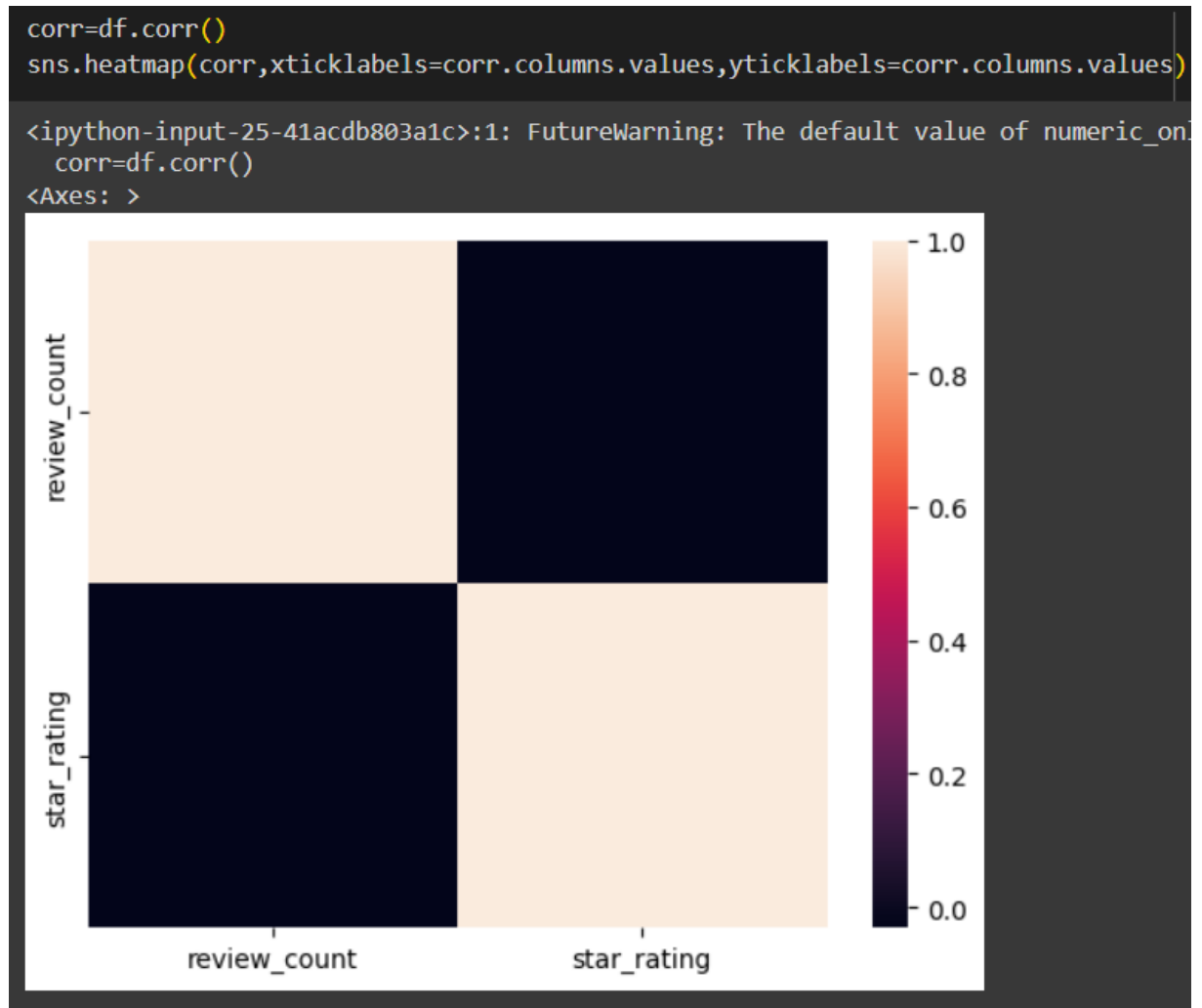
8) Koherasyon matrisi çizdirelim.

```
df.corr()
```

```
<ipython-input-22-2f6f6606aa2c>:1: FutureWarning
df.corr()
```

	review_count	star_rating
review_count	1.000000	-0.030487
star_rating	-0.030487	1.000000

9) Koherasyon katsayılarını daha iyi okumak için ısı haritası çizdirelim.



10) Şimdi datasetimizdeki benzersiz değerleri görüntüleyelim. 2073 tane artist, 56 tane yıl, 4 adet yıldız puanı bulunmaktadır.

```
df["artist"].unique()

array(['Motörhead', 'Inquisition', 'King Buffalo', ..., 'Strangeways',
      'Hammer King', 'Bowie, David'], dtype=object)

[ ] df["artist"].nunique()

2073

df["year"].unique()

array(['2008', '2020', '2022', '2018', '2012', '2017', '2019', '2010',
      '2011', '2009', '2014', '2021', '2016', '2006', '2001', '1997',
      nan, '2002', '2003', '1975', '1994', '1987', ' ', et al.', '1992',
      '2013', '2015', '2005', '1999', '2004', '2000', '2007', '1995',
      '1998', '1990', '1989', '2023', '1979', '1991', '1993', '1996',
      '1988', '1974', '1976', '1970', '1971', ' | ', '1980', '1983',
      '1977', '1973', '1984', '1978', '1986', '1981', '1985', '1972',
      '1982'], dtype=object)

[ ] df["year"].nunique()

56

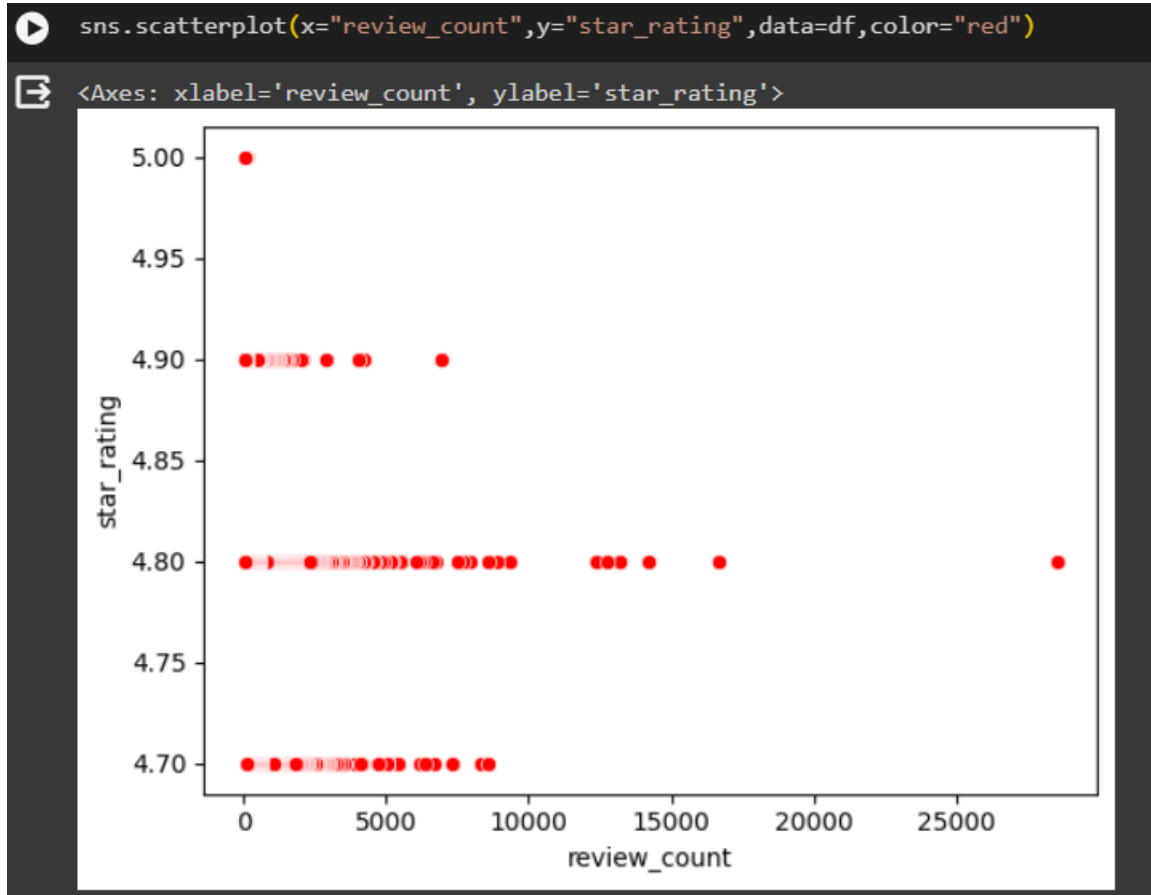
[ ] df["star_rating"].unique()

array([4.9, 5. , 4.8, 4.7])

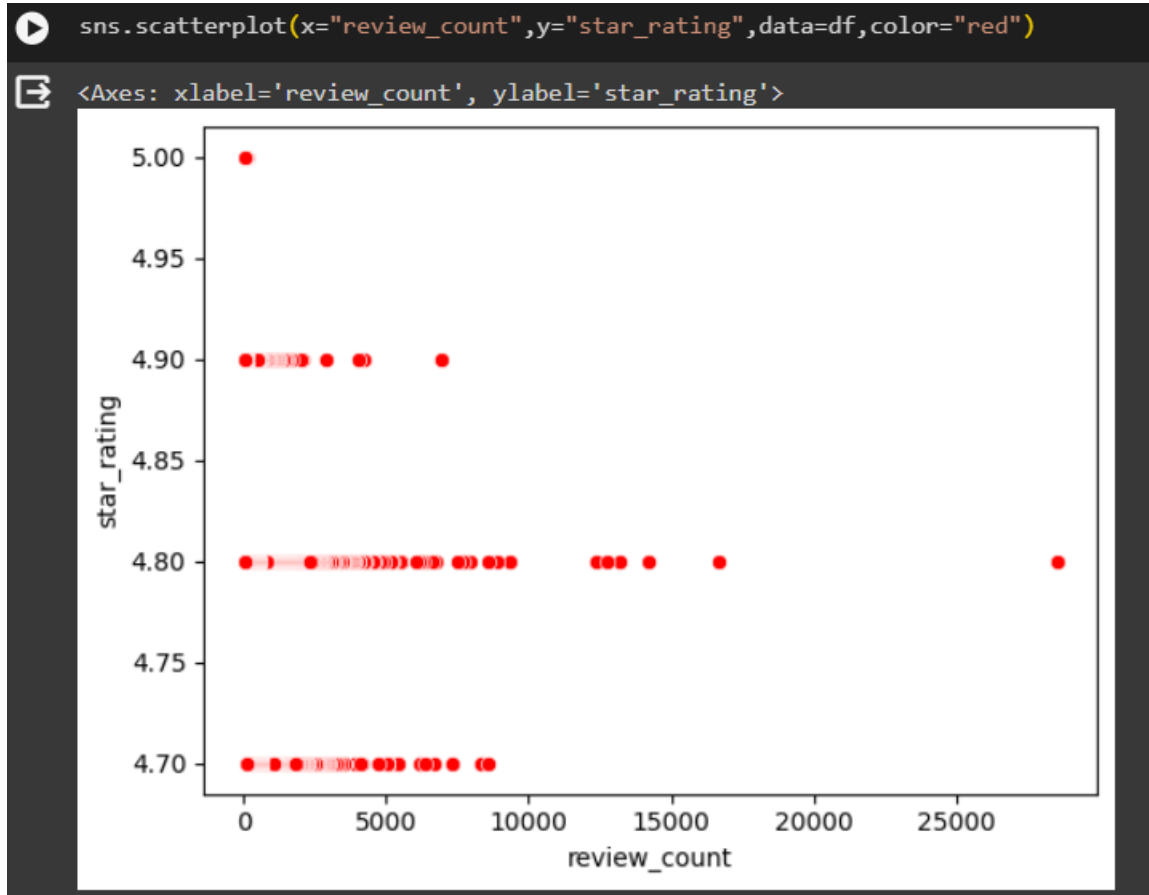
[ ] df["review_count"].unique()

array([136, 134, 130, ..., 19, 27, 18])
```

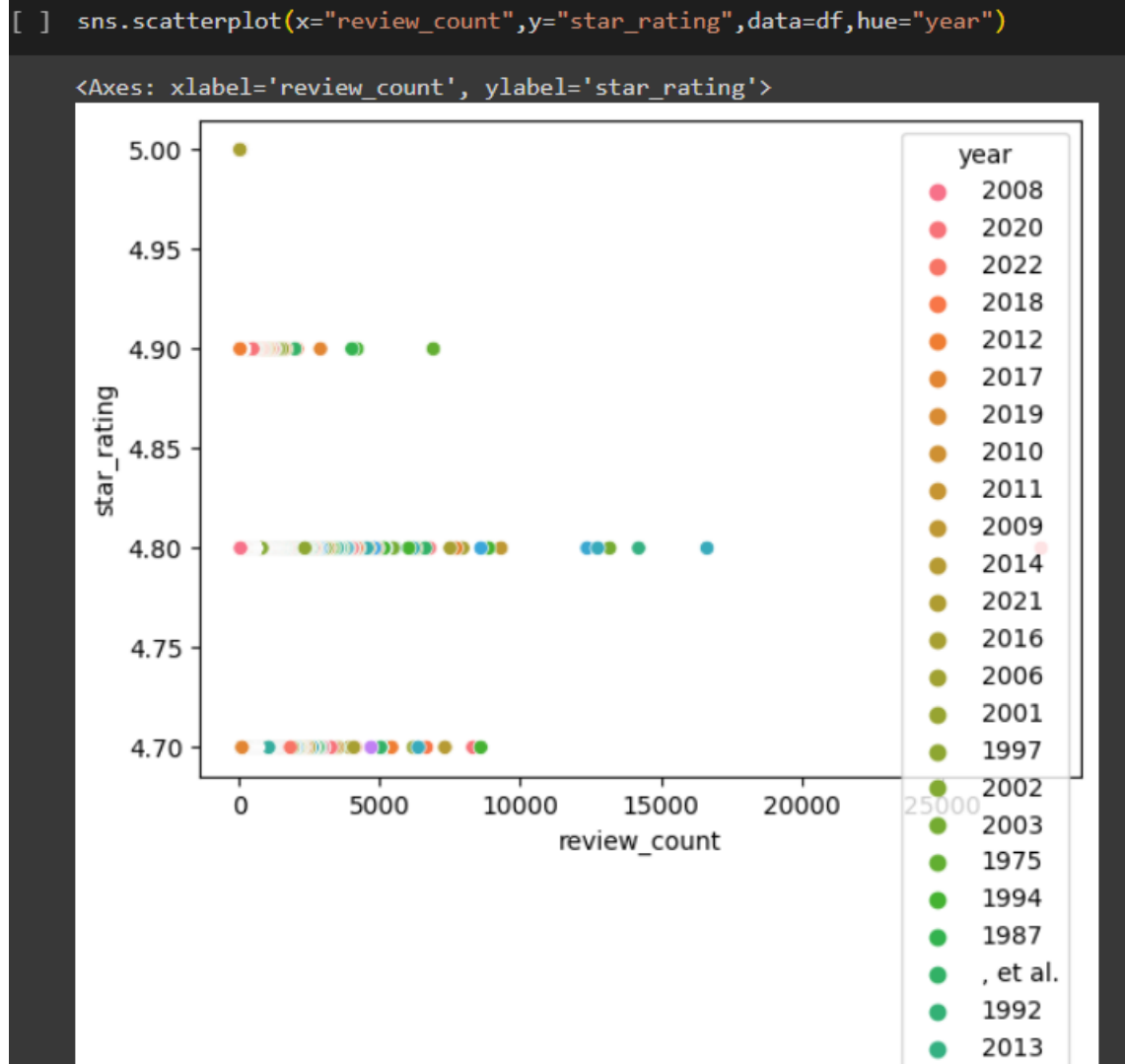
11) Şimdi de x eksenini review_count, y eksenini star_rating değişkenleri olarak belirleyip dağılım grafiğine bakalım. Bu grafikte 4.7, 4.9 yıldız verenlerin daha çok 0-10.000 civarlarında, 4.8 yıldız verenlerin 0-27.000 civarlarında, 5 yıldız verenlerin ise 0-1000 civarlarında olduğu söylenebilir.



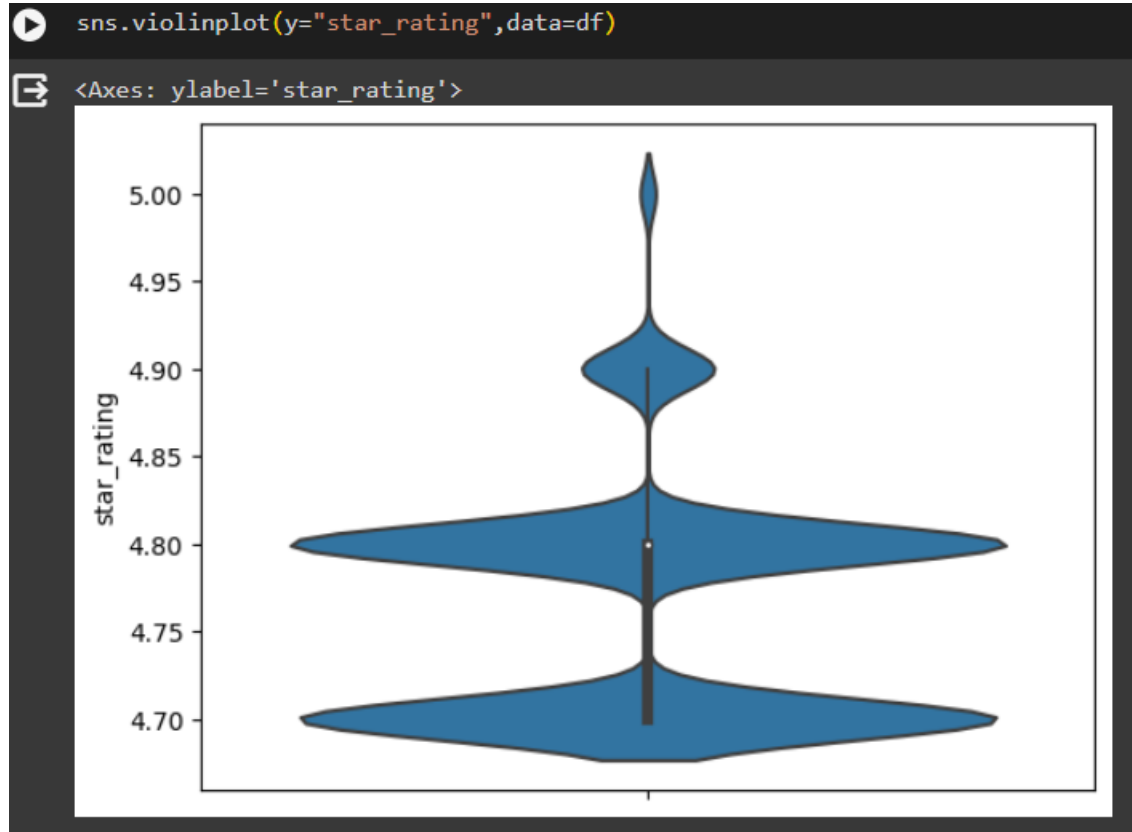
12) Frekansları incelemek için jointplot kullanabiliriz. En çok verilen yıldızın 4.7 ve 4.8 olduğunu söyleyebiliriz.



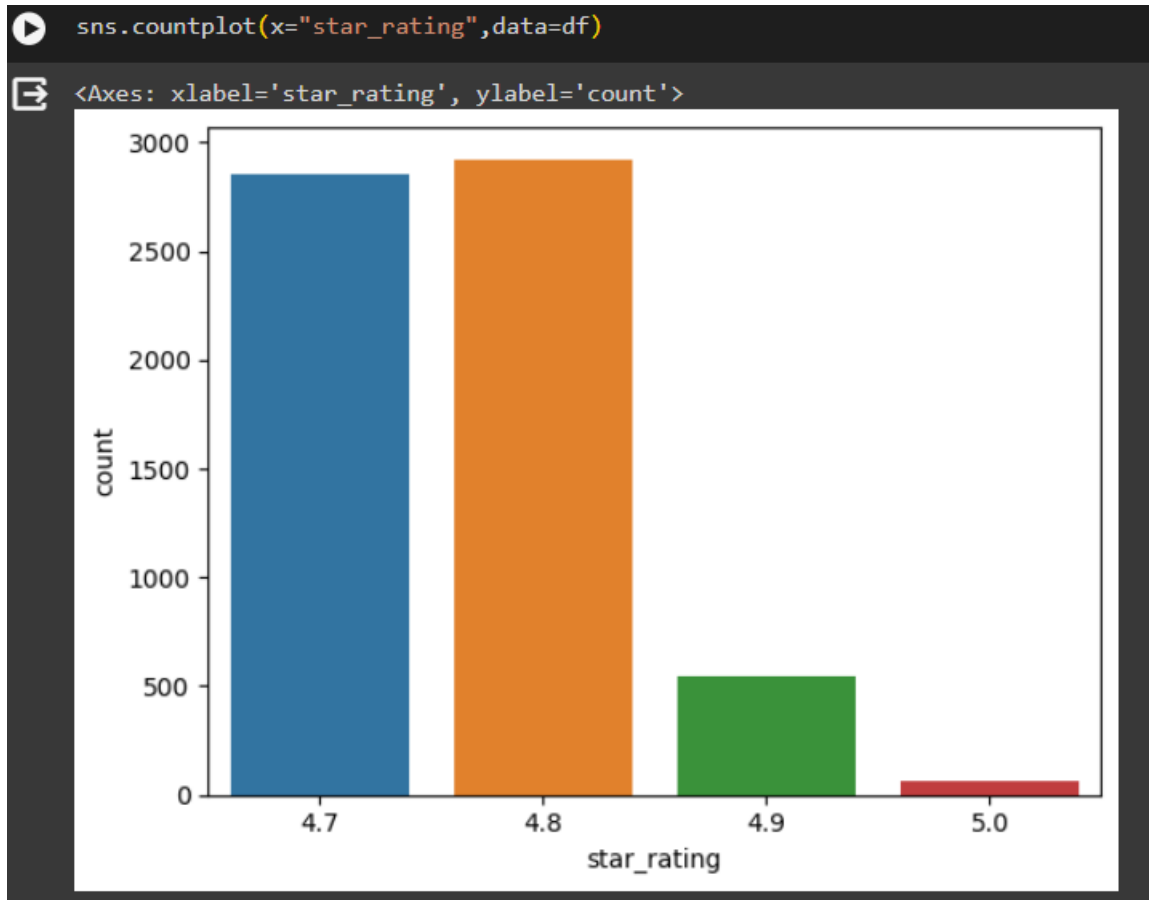
13) Burada da yıllara göre yıldız puanları ve inceleme sayısı arasındaki ilişkiyi görebiliriz.



14) Şimdi de keman grafiği çizdirelim. Bu grafik bize 4.7 ve 4.8 yıldız puanı değerleri arasında bir yoğunluk olduğunu göstermektedir.



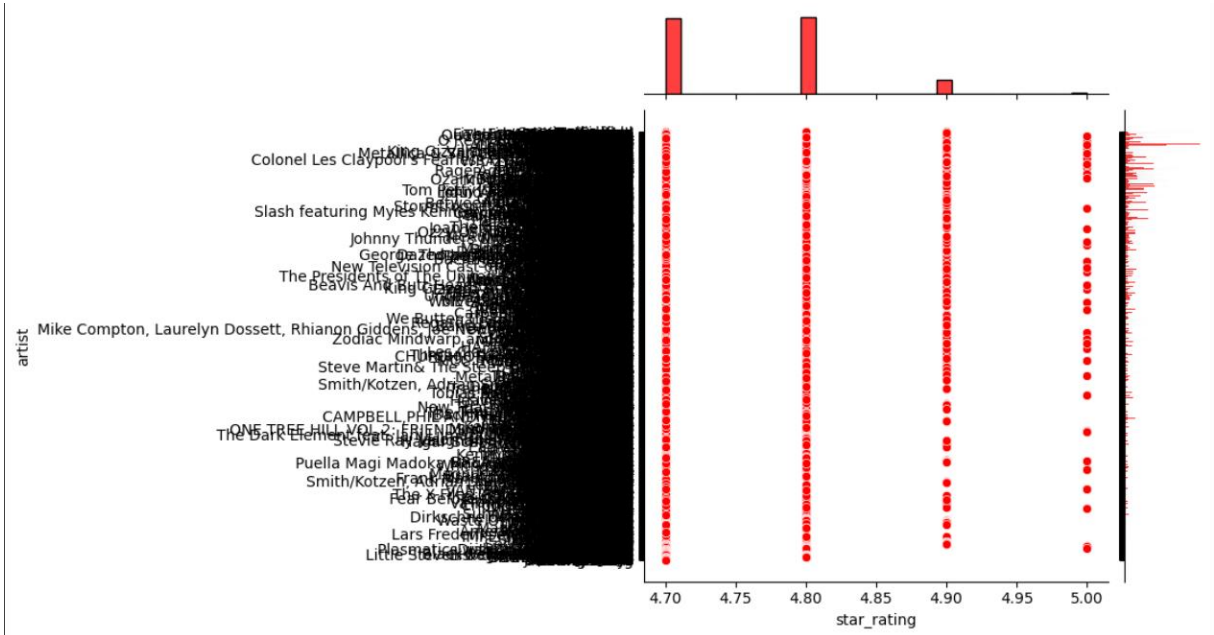
15) Şimdi de bu söylenenleri istatistiksel olarak bir tabloya dökelim.



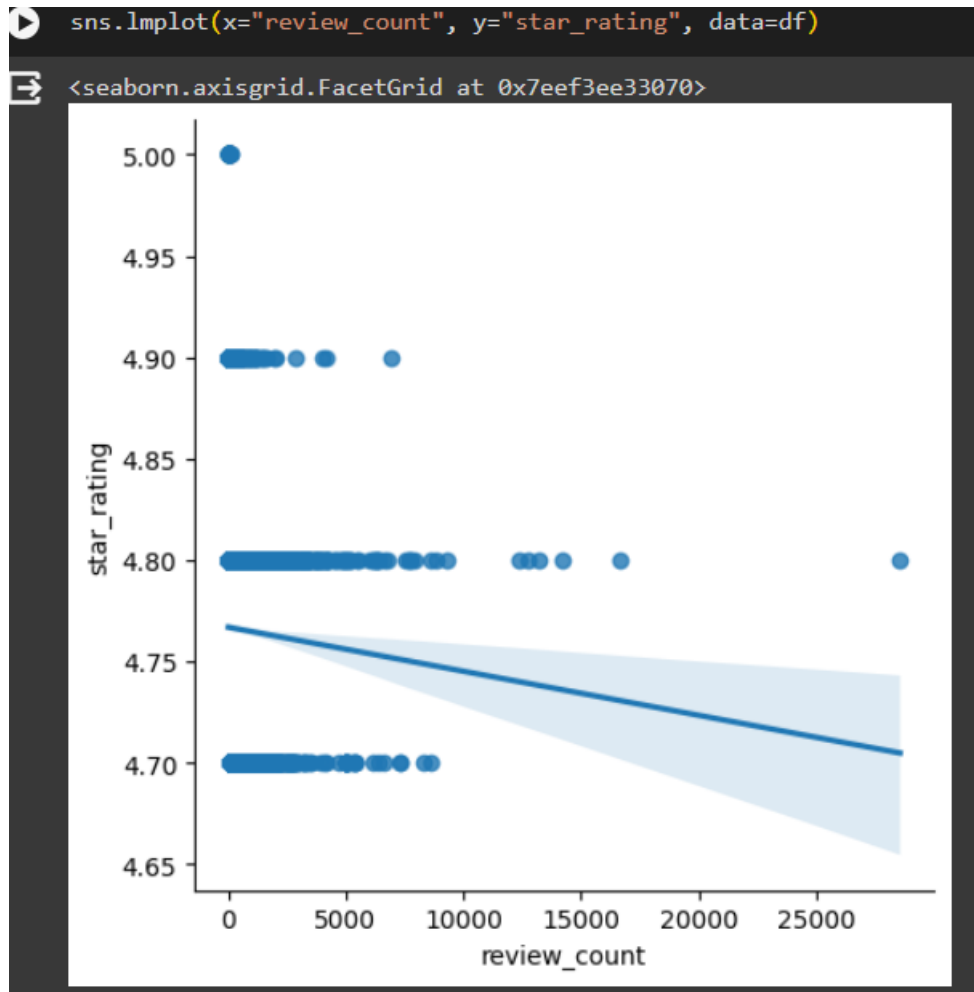
16) Burada da listenin içindeki şarkıların ortalaması alınarak (bir metal grubunun birden fazla şarkısı listede varsa) grupların başarı sıralaması görelim.

!!!Çok fazla grup olduğu için temiz bir grafik ortaya çıkmamaktadır.

```
sns.jointplot(x=df["star_rating"],y=df["artist"],color="red")
```



17) sns.lmplot sayesinde inceleme sayısı ve yıldız puanları arasındaki ilişkiye bakalım.
Bu ilişki zayıf bir ilişkidir.



18) Şimdi de yıldız puanına göre bir gruplandırma işlemi yapalım, değişken değerlerimizin ortalamasını görüntüleyelim.

