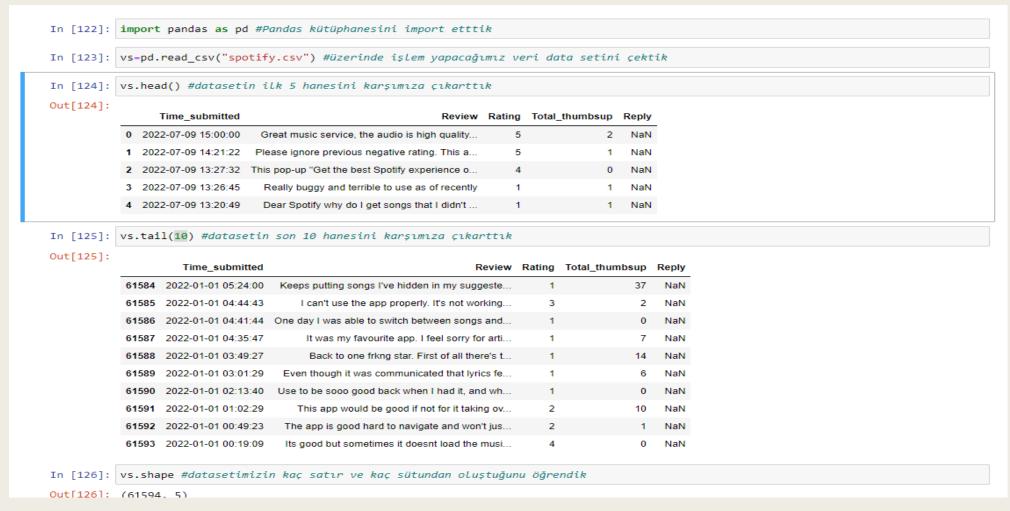
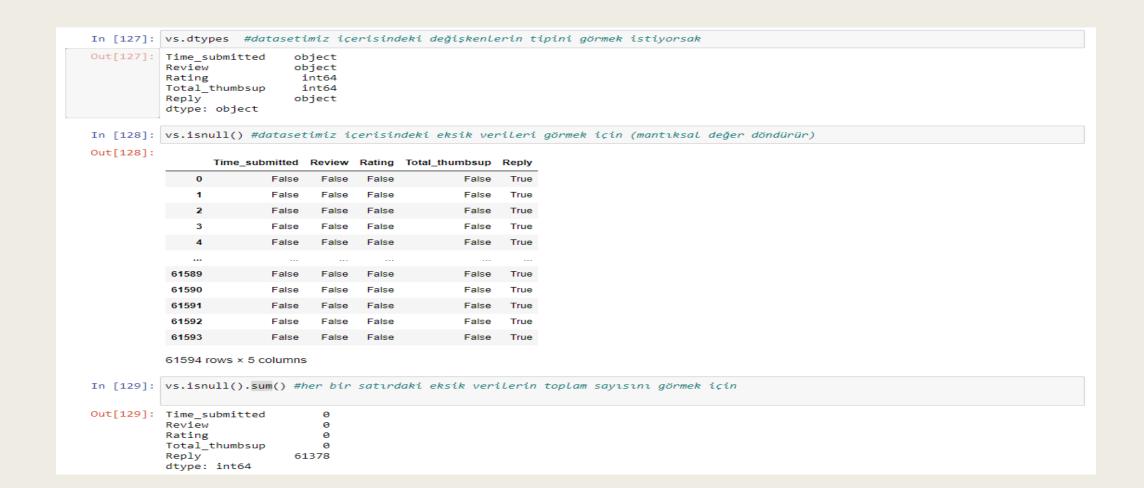
VERI MADENCILIĞI PANDAS İLE VERI ANALIZI

BATUHAN ASLAN 191001037

- Pandas kütüphanesini import ediyoruz ve «pd» olarak adlandırıyoruz.
- Veri setimizi okutup vs değişkenine atıyoruz.
- vs.head komutuyla ilk 5 satırı; vs.tail komutuyla parametreye 10 son 10 haneyi çıktı alıyoruz.
- vs.shape komutu ile veri setimizin kaç satır ve kaç sütun olduğunu öğreniyoruz.

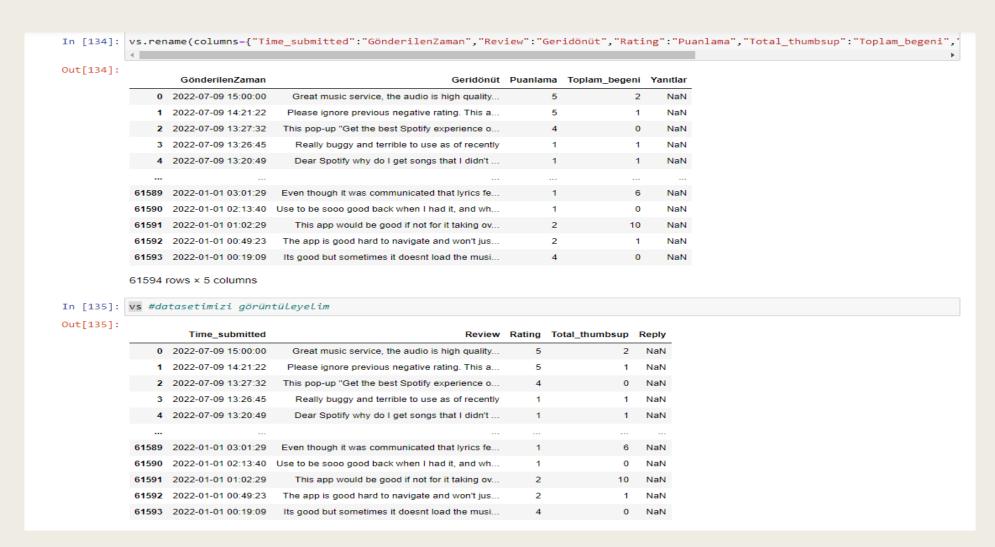


- vs.dtypes komutu ile değişken tiplerini öğrenebiliriz.
- vs.isnull() komutu ile boş yer olup olmadığını boolean değişken tipiyle öğrenebiliriz.
- vs.isnull().sum() komutu ile bu boş alanların hangi sütunda toplam kaç tane olduğunu öğrenebiliriz.

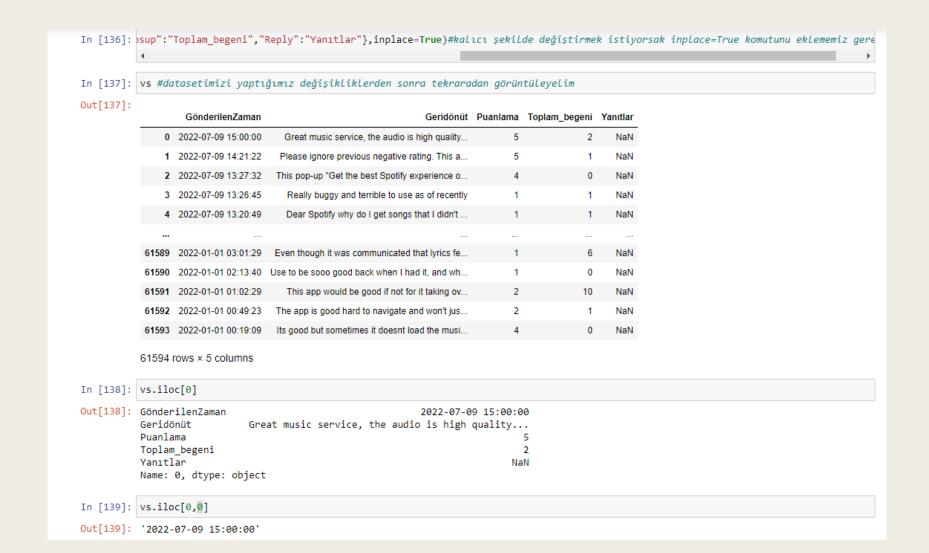


```
In [130]: vs.Review.head() #dataframedeki Review sütununa ulaştık ve ilk 5 satırını gördük
Out[130]: 0
                Great music service, the audio is high quality...
                Please ignore previous negative rating. This a...
               This pop-up "Get the best Spotify experience o...
                  Really buggy and terrible to use as of recently
                Dear Spotify why do I get songs that I didn't ...
           Name: Review, dtype: object
In [131]: vs.columns #sütun indexleri karşımıza gelmiş oldu
Out[131]: Index(['Time_submitted', 'Review', 'Rating', 'Total_thumbsup', 'Reply'], dtype='object')
In [132]: vs["Total thumbsup"].head() #total thumbsup sütunununun ilk 5 satırını görmüş olduk
Out[132]: 0
                1
           Name: Total thumbsup, dtype: int64
In [133]: vs[["Review", "Rating"]].head() #herhangi istediğimiz iki sütünun ilk 5 er satırlarını yan yana ekrana getirebiliyoruz
Out[133]:
                                             Review Rating
                Great music service, the audio is high quality...
               Please ignore previous negative rating. This a...
            2 This pop-up "Get the best Spotify experience o...
                Really buggy and terrible to use as of recently
                Dear Spotify why do I get songs that I didn't ...
```

- Veri setimizin sütunlarını değiştirmek istersek bu şekilde değiştirebiliyoruz fakat bu sadece bir seferlik değişikliğe yani kalıcı olmamak kaydıyla değişim sağlıyor.
- Tekrardan vs komutu ile görüntülediğimizde değişikliğin olmadığını görüyoruz.
- Eğer kalıcılık peşindeysek bunu da bir sonraki slayttaki gibi yapmamız gerekiyor.



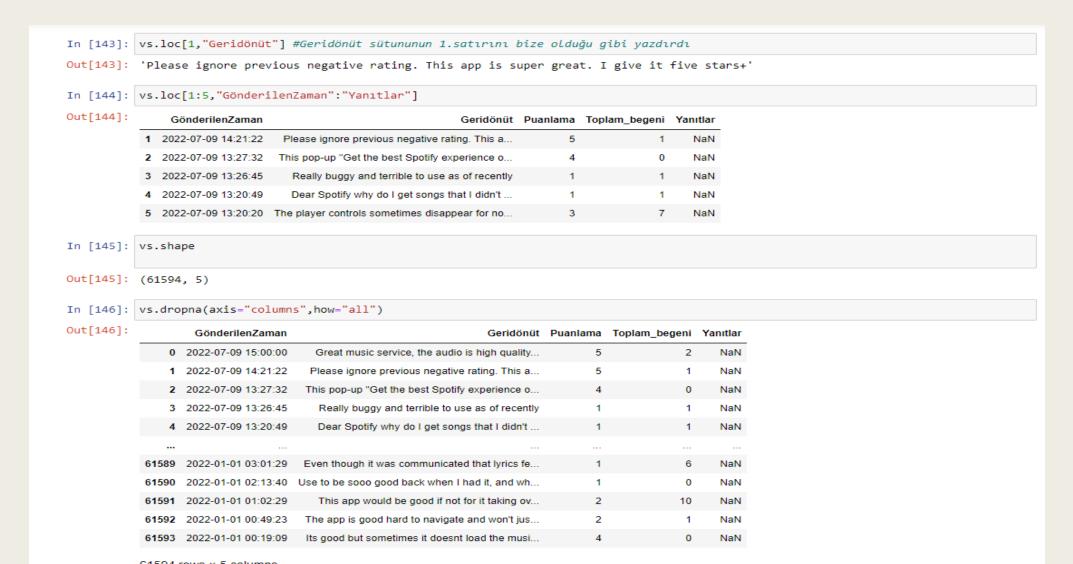
 vs.iloc komuyla içerisinde index olarak sayı girip o değere ait satır ve sütündaki verileri bizim önümüze getiriyor.



vs.iloc komutunda yine birden fazla satır veya sütundaki verileri görmek istiyorsak aşağıdaki gibi işlemler yapabiliriz.

	vs.iloc[0,[1,3,4]] #birden fazla sütun seçmek istersek yandaki gibi girebiliyoruız								
Out[140]:	Top Yar	ridönüt Great music service, plam_begeni nıtlar me: 0, dtype: object	the audio	is	high qu	uality 2 NaN			
n [141]:	vs.	.iloc[0:5,[1,3,4]] #5.indexe kadar	olan satırl	ar .	yazıldı	,			
Out[141]:		Geridönüt	Toplam_bege	ni	Yanıtlar				
	0	Great music service, the audio is high quality		2	NaN				
	1	Please ignore previous negative rating. This a		1	NaN				
	2	This pop-up "Get the best Spotify experience o		0	NaN				
	3	Really buggy and terrible to use as of recently		1	NaN				
	4	Dear Spotify why do I get songs that I didn't		1	NaN				
in [142]:	vs.	.iloc[0:5,1:4] #bu şekilde de satır	larda da ayı	nı	işlemi	uygulayabil	iriz		
Out[142]:		Geridönüt	Puanlama T	opla	am_beger	ni			
	0	Great music service, the audio is high quality	5			2			
	1	Please ignore previous negative rating. This a	5			1			
	2	This pop-up "Get the best Spotify experience o	4			0			
	3	Really buggy and terrible to use as of recently	1			1			

- vs.loc komutu ise 2 parametre alır ve hangi sütunun kaçıncı satırındaki veriyi istiyorsak ona göre işlemlerine devam eder.
- vs.shape komutu kullanarak verisetimiz içerisinde kaç satır ver kaç sütun olduğunu görebiliriz.



- vs.dropna komutu ile birlikte shape komutu veri seti içerisinde boşluk olarak gözüken değerlerdeki sütunları silmeye dayalı komuttur.
- HOW kısmına all yazarsak o satırın bütün halini ele alarak boşluk arar fakat any yazarsak herhangi bir yerde boşluk olması o sütunun silinmesi için kayda değerdir.

```
In [147]: vs.dropna(axis="columns",how="all").shape #sütun sayısı değişmedi eksik veri olmadığını anlıyoruz
Out[147]: (61594, 5)
In [148]: vs.dropna(axis="columns",how="any").shape #all yerine any yazarsak herhangi bir sütunda 1 tane eksik veri varsa o sütun silinir
Out[148]: (61594, 4)
In [149]: vs.Puanlama.value_counts() #PUANLAMA olarak 5 puan üzerinden hangisinden kaçar tane puan verilmiş onu analiz ediyoruz
Out[149]: 5
               22095
               17653
                7842
                7118
                6886
          Name: Puanlama, dtype: int64
In [150]: #En fazla sayıda verilen puanlamanın 5 puan olduğu yukarıda gözüktüğü gibidir.
In [151]: vs[vs.Puanlama==5].Geridönüt.value counts() #5 puan verenlerin Geridönütlerinin ne oldukları gözüktüğü gibidir.
Out[151]: Amazing music app
          Great music selection
          Very good music app
          Good sound quality
          Too many ads
          Very good app easy to use and has some ads
          amazing fabulous excellent Spotify relax me and my mind very nice fabulous app
          This app is really at a good position for music. Keep this process going! And please make updates. I have a small problem.. Wel
          1.. When I save a song in my playlist it doesn't go there. This is what my problem was so I started to use it from the liked so
          ngs menu. But I highly recommend this app for your chill, moody or any songs you like. With head phones it has the best quality
          to listen. overall I rate it a 5 star. I hope this reaches over a trillion downloads! Thank you for your time! 🛞
          I love Spotify but I hate ads
          It selects music based on your taste makes best "mix tape" which makes things easier. Thank you for your effort and dedication
```

```
In [152]: #5 puan vermiş kullanıcıların geridönütlerinin beğeni sayılarını ele alalım
In [156]: vs.GönderilenZaman.str.slice(0,4) #Hangi satırdaki yorumun hangi tarihte yapıldığına dair yılları aldığımız kısım.
Out[156]: 0
                   2022
                   2022
          2
                   2022
                   2022
                   2022
                   . . .
          61589
                   2022
          61590
                   2022
          61591
                   2022
          61592
                   2022
                   2022
          61593
          Name: GönderilenZaman, Length: 61594, dtype: object
 In [41]: vs.GönderilenZaman.str.slice(0,4).value counts() #hangi tarihteki yorumların kaç tane olduğunun toplam sayısını görüntüleme.
Out[41]: 2022
                  61594
          Name: GönderilenZaman, dtype: int64
In [157]: vs.dtypes #datatiplerini gösterme komutu
Out[157]: GönderilenZaman
                             object
          Geridönüt
                             object
          Puanlama
                              int64
          Toplam begeni
                              int64
          Yanıtlar
                             object
          dtype: object
 In [43]: vs["GönderilenZaman"]=pd.to datetime(vs.GönderilenZaman) #GönderilenZaman sütunumun değişken tipini datetime olarak değiştiriyoru
 In [44]: vs.dtypes #değişken tiplerini görüntülüyoruz
Out[44]: GönderilenZaman
                             datetime64[ns]
          Geridönüt
                                     object
          Puanlama
                                      int64
          Toplam_begeni
                                      int64
          Yanıtlar
                                     object
          dtype: object
```

```
In [45]: vs.GönderilenZaman.dt.year #değişkenimizi datetime yaptığımız için dt.year koduyla yıllara ulaşabiliyoruz.
Out[45]: 0
                  2022
                  2022
         2
                  2022
                  2022
                  2022
                  . . .
         61589
                  2022
         61590
                  2022
                 2022
         61591
         61592
                 2022
                 2022
         61593
         Name: GönderilenZaman, Length: 61594, dtype: int64
In [46]: vs.GönderilenZaman.dt.month #değişkenimizi datetime yaptığımız için dt.month koduyla aylara ulaşabiliyoruz
Out[46]: 0
         2
         61589
         61590
         61591 1
         61592
         61593
         Name: GönderilenZaman, Length: 61594, dtype: int64
In [48]: vs.Toplam_begeni.value_counts() #Toplam begeni alındığındaki toplam like sayısı
Out[48]: 0
                 42112
                  8360
                  2968
         2
                  1392
                   871
         4550
                    1
         4791
         713
                     1
         1297
                     1
         136
         Name: Toplam_begeni, Length: 532, dtype: int64
```

```
In [159]: vs.Toplam_begeni.mean() #Toplambegeni yüzdesini alma durumu
Out[159]: 7.649381433256486
In [50]: vs.groupby(vs.GönderilenZaman.dt.hour).Toplam_begeni.mean() #gün içerisindeki saatlere göre ortalama begeni sayyısını alma
Out[50]: GönderilenZaman
                 5.480966
                 7.484000
                 9.374297
                7.274706
                 7.737656
                 8.987698
                 8.182106
                9.706654
                11.192326
                 8.470015
                 8.517833
                 6.645014
          11
                 6.427363
          12
          13
                 6.992852
          14
                10.629987
          15
                 8.793189
          16
                 4.999695
          17
                 5.278981
                 8.211234
                 8.363393
                 8.652353
                 5.140644
          21
                 5.796882
          22
                 6.051101
          Name: Toplam_begeni, dtype: float64
```

Grafik bilgisi için matplotlib kütüphanesini yükleyip importluyoruz.

```
!pip install matplotlib
In [51]:
         Requirement already satisfied: matplotlib in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages
         (3.5.2)
         Requirement already satisfied: fonttools>=4.22.0 in c:\users\batuhan aslan\appdata\local\programs\python\python\10\lib\site-pac
         kages (from matplotlib) (4.34.4)
         Requirement already satisfied: cycler>=0.10 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages
         (from matplotlib) (0.11.0)
         Requirement already satisfied: numpy>=1.17 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages
         (from matplotlib) (1.23.1)
         Requirement already satisfied: packaging>=20.0 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packa
         ges (from matplotlib) (21.3)
         Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-pac
         kages (from matplotlib) (1.4.4)
         Requirement already satisfied: pillow>=6.2.0 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-package
         s (from matplotlib) (9.2.0)
         Requirement already satisfied: pyparsing>=2.2.1 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-pack
         ages (from matplotlib) (3.0.9)
         Requirement already satisfied: python-dateutil>=2.7 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-
         packages (from matplotlib) (2.8.2)
         Requirement already satisfied: six>=1.5 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages (fr
         om python-dateutil>=2.7->matplotlib) (1.16.0)
         import matplotlib #matplotlib kütüphnesini import ediyoruz
In []: %matplotlib inline #cizdireceğimiz arafiği satır içinde gösteriyoruz
```

 Plot() fonksiyonunu kullanarak saatleri gruplayarak, kullanıcıların hangi saatlerde ortalama kaç beğeni yaptığını gösteren kodun grafiği aşağıdaki gibidir.

