

VERİ MADENCİLİĞİ PANDAS İLE VERİ ANALİZİ

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.

```
In [122]: import pandas as pd #Pandas kütüphanesini import ettik
```

```
In [123]: vs=pd.read_csv("spotify.csv") #üzerinde işlem yapacağımız veri data setini çektik
```

```
In [124]: vs.head() #datasetin ilk 5 hanesini karşımıza çıkarttık
```

Out[124]:

	Time_submitted	Review	Rating	Total_thumbsup	Reply
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN

```
In [125]: vs.tail(10) #datasetin son 10 hanesini karşımıza çıkarttık
```

Out[125]:

	Time_submitted	Review	Rating	Total_thumbsup	Reply
61584	2022-01-01 05:24:00	Keeps putting songs I've hidden in my suggeste...	1	37	NaN
61585	2022-01-01 04:44:43	I can't use the app properly. It's not working...	3	2	NaN
61586	2022-01-01 04:41:44	One day I was able to switch between songs and...	1	0	NaN
61587	2022-01-01 04:35:47	It was my favourite app. I feel sorry for arti...	1	7	NaN
61588	2022-01-01 03:49:27	Back to one frkng star. First of all there's t...	1	14	NaN
61589	2022-01-01 03:01:29	Even though it was communicated that lyrics fe...	1	6	NaN
61590	2022-01-01 02:13:40	Use to be sooo good back when I had it, and wh...	1	0	NaN
61591	2022-01-01 01:02:29	This app would be good if not for it taking ov...	2	10	NaN
61592	2022-01-01 00:49:23	The app is good hard to navigate and won't jus...	2	1	NaN
61593	2022-01-01 00:19:09	Its good but sometimes it doesnt load the musi...	4	0	NaN

```
In [126]: vs.shape #datasetimizin kaç satır ve kaç sütundan oluştuğunu öğrendik
```

Out[126]: (61594, 5)

- 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 [127]: vs.dtypes #datasetimiz içerisindeki değişkenlerin tipini görmek istiyorsak
```

```
Out[127]: Time_submitted    object  
Review                  object  
Rating                  int64  
Total_thumbsup          int64  
Reply                   object  
dtype: object
```

```
In [128]: vs.isnull() #datasetimiz içerisindeki eksik verileri görmek için (mantıksal değer döndürür)
```

```
Out[128]:
```

	Time_submitted	Review	Rating	Total_thumbsup	Reply
0	False	False	False	False	True
1	False	False	False	False	True
2	False	False	False	False	True
3	False	False	False	False	True
4	False	False	False	False	True
...
61589	False	False	False	False	True
61590	False	False	False	False	True
61591	False	False	False	False	True
61592	False	False	False	False	True
61593	False	False	False	False	True

61594 rows x 5 columns

```
In [129]: vs.isnull().sum() #her bir satırdaki eksik verilerin toplam sayısını görmek için
```

```
Out[129]: Time_submitted    0  
Review                  0  
Rating                  0  
Total_thumbsup          0  
Reply                   61378  
dtype: int64
```

```
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...
          1    Please ignore previous negative rating. This a...
          2    This pop-up "Get the best Spotify experience o...
          3    Really buggy and terrible to use as of recently
          4    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 sütununun ilk 5 satırını görmüş olduk
```

```
Out[132]: 0    2
          1    1
          2    0
          3    1
          4    1
          Name: Total_thumbsup, dtype: int64
```

```
In [133]: vs[["Review","Rating"]].head() #herhangi istediğimiz iki sütunun ilk 5 er satırlarını yan yana ekrana getirebiliyoruz
```

```
Out[133]:
```

	Review	Rating
0	Great music service, the audio is high quality...	5
1	Please ignore previous negative rating. This a...	5
2	This pop-up "Get the best Spotify experience o...	4
3	Really buggy and terrible to use as of recently	1
4	Dear Spotify why do I get songs that I didn't ...	1

- 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 peşindeyse bunu da bir sonraki slayttaki gibi yapmamız gerekiyor.

```
In [134]: vs.rename(columns={"Time_submitted": "GönderilenZaman", "Review": "Geridönüt", "Rating": "Puanlama", "Total_thumbsup": "Toplam_begeni",
```

Out[134]:

	GönderilenZaman	Geridönüt	Puanlama	Toplam_begeni	Yanıtlar
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
...
61589	2022-01-01 03:01:29	Even though it was communicated that lyrics fe...	1	6	NaN
61590	2022-01-01 02:13:40	Use to be sooo good back when I had it, and wh...	1	0	NaN
61591	2022-01-01 01:02:29	This app would be good if not for it taking ov...	2	10	NaN
61592	2022-01-01 00:49:23	The app is good hard to navigate and won't jus...	2	1	NaN
61593	2022-01-01 00:19:09	Its good but sometimes it doesnt load the musi...	4	0	NaN

61594 rows x 5 columns

```
In [135]: vs #datasetimizi görüntüleyelim
```

Out[135]:

	Time_submitted	Review	Rating	Total_thumbsup	Reply
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
...
61589	2022-01-01 03:01:29	Even though it was communicated that lyrics fe...	1	6	NaN
61590	2022-01-01 02:13:40	Use to be sooo good back when I had it, and wh...	1	0	NaN
61591	2022-01-01 01:02:29	This app would be good if not for it taking ov...	2	10	NaN
61592	2022-01-01 00:49:23	The app is good hard to navigate and won't jus...	2	1	NaN
61593	2022-01-01 00:19:09	Its good but sometimes it doesnt load the musi...	4	0	NaN

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

```
In [136]: vs["sup":"Toplam_begeni","Reply":"Yanıtlar"],inplace=True)#kalıcı şekilde değiştirmek istiyorsak inplace=True komutunu eklememiz gere
```

```
In [137]: vs #datasetimizi yaptığımız değişikliklerden sonra tekraradan görüntüleyelim
```

Out[137]:

	GönderilenZaman	Geridönüt	Puanlama	Toplam_begeni	Yanıtlar
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
...
61589	2022-01-01 03:01:29	Even though it was communicated that lyrics fe...	1	6	NaN
61590	2022-01-01 02:13:40	Use to be sooo good back when I had it, and wh...	1	0	NaN
61591	2022-01-01 01:02:29	This app would be good if not for it taking ov...	2	10	NaN
61592	2022-01-01 00:49:23	The app is good hard to navigate and won't jus...	2	1	NaN
61593	2022-01-01 00:19:09	Its good but sometimes it doesnt load the musi...	4	0	NaN

61594 rows x 5 columns

```
In [138]: vs.iloc[0]
```

Out[138]:

GönderilenZaman	2022-07-09 15:00:00
Geridönüt	Great music service, the audio is high quality...
Puanlama	5
Toplam_begeni	2
Yanıtlar	NaN

Name: 0, dtype: object

```
In [139]: vs.iloc[0,0]
```

Out[139]: '2022-07-09 15:00:00'

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

```
In [140]: vs.iloc[0,[1,3,4]] #birden fazla sütun seçmek istersek yandaki gibi girebiliyoruz
```

```
Out[140]: Geridönüt      Great music service, the audio is high quality...
          Toplam_begeni      2
          Yanıtlar      NaN
          Name: 0, dtype: object
```

```
In [141]: vs.iloc[0:5,[1,3,4]] #5.indexe kadar olan satırlar yazıldı
```

```
Out[141]:
```

	Geridönüt	Toplam_begeni	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 satırlarda da aynı işlemi uygulayabiliriz
```

```
Out[142]:
```

	Geridönüt	Puanlama	Toplam_begeni
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
4	Dear Spotify why do I get songs that I didn't ...	1	1

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

```
In [143]: vs.loc[1,"Geridönüt"] #Geridönüt sütununun 1.satırını bize olduğu gibi yazdırdı
```

```
Out[143]: 'Please ignore previous negative rating. This app is super great. I give it five stars+'
```

```
In [144]: vs.loc[1:5,"GönderilenZaman":"Yanıtlar"]
```

```
Out[144]:
```

	GönderilenZaman	Geridönüt	Puanlama	Toplam_begeni	Yanıtlar
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
5	2022-07-09 13:20:20	The player controls sometimes disappear for no...	3	7	NaN

```
In [145]: vs.shape
```

```
Out[145]: (61594, 5)
```

```
In [146]: vs.dropna(axis="columns",how="all")
```

```
Out[146]:
```

	GönderilenZaman	Geridönüt	Puanlama	Toplam_begeni	Yanıtlar
0	2022-07-09 15:00:00	Great music service, the audio is high quality...	5	2	NaN
1	2022-07-09 14:21:22	Please ignore previous negative rating. This a...	5	1	NaN
2	2022-07-09 13:27:32	This pop-up "Get the best Spotify experience o...	4	0	NaN
3	2022-07-09 13:26:45	Really buggy and terrible to use as of recently	1	1	NaN
4	2022-07-09 13:20:49	Dear Spotify why do I get songs that I didn't ...	1	1	NaN
...
61589	2022-01-01 03:01:29	Even though it was communicated that lyrics fe...	1	6	NaN
61590	2022-01-01 02:13:40	Use to be sooo good back when I had it, and wh...	1	0	NaN
61591	2022-01-01 01:02:29	This app would be good if not for it taking ov...	2	10	NaN
61592	2022-01-01 00:49:23	The app is good hard to navigate and won't jus...	2	1	NaN
61593	2022-01-01 00:19:09	Its good but sometimes it doesnt load the musi...	4	0	NaN

61594 rows x 5 columns

- 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
          1    17653
          4     7842
          2     7118
          3     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
          9
          Great music selection
          8
          Very good music app
          7
          Good sound quality
          6
          Too many ads
          5

          ..
          Very good app easy to use and has some ads
          1
          amazing fabulous excellent Spotify relax me and my mind very nice fabulous app
          1
          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! 🤔 1
          I love Spotify but I hate ads
          1
          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
          1      2022
          2      2022
          3      2022
          4      2022
          ...
          61589   2022
          61590   2022
          61591   2022
          61592   2022
          61593   2022
          Name: GönderilenZaman, Length: 61594, dtype: object
```

```
In [41]: vs.GönderilenZaman.str.slice(0,4).value_counts() #hangi tarihteki yorumların kaç tane olduğunu 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 değiştiriyor
```

```
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  
1      2022  
2      2022  
3      2022  
4      2022  
...  
61589   2022  
61590   2022  
61591   2022  
61592   2022  
61593   2022  
Name: GönderilenZaman, Length: 61594, dtype: int64
```

```
In [46]: vs.GönderilenZaman.dt.month #değişkenimizi datetime yaptığımız için dt.month koduyla aylara ulaşabiliyoruz
```

```
Out[46]: 0      7  
1      7  
2      7  
3      7  
4      7  
..  
61589   1  
61590   1  
61591   1  
61592   1  
61593   1  
Name: GönderilenZaman, Length: 61594, dtype: int64
```

```
In [48]: vs.Toplam_begeni.value_counts() #Toplam beğeni alındığındaki toplam like sayısı
```

```
Out[48]: 0      42112  
1      8360  
2      2968  
3      1392  
4       871  
...  
4550     1  
4791     1  
713      1  
1297     1  
136      1  
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 beğeni sayısını alma
```

```
Out[50]: GönderilenZaman
0      5.480966
1      7.484000
2      9.374297
3      7.274706
4      7.737656
5      8.987698
6      8.182106
7      9.706654
8     11.192326
9      8.470015
10     8.517833
11     6.645014
12     6.427363
13     6.992852
14    10.629987
15     8.793189
16     4.999695
17     5.278981
18     8.211234
19     8.363393
20     8.652353
21     5.140644
22     5.796882
23     6.051101
Name: Toplam_begeni, dtype: float64
```

- Grafik bilgisi için matplotlib kütüphanesini yükleyip importluyoruz.

```
In [51]: !pip install matplotlib
```

```
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\python310\lib\site-packages (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-packages (from matplotlib) (21.3)  
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (1.4.4)  
Requirement already satisfied: pillow>=6.2.0 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages (from matplotlib) (9.2.0)  
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\batuhan aslan\appdata\local\programs\python\python310\lib\site-packages (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 (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

```
In [54]: import matplotlib #matplotlib kütüphanesini import ediyoruz
```

```
In [ ]: %matplotlib inline #çizdireceğimiz grafiği satır içinde gösteriyoruz
```

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

```
In [58]: vs.groupby(vs.GönderilenZaman.dt.hour).Toplam_begeni.mean().plot() #plot fonksiyonunu kullanarak grafiğimizi çizdiriyoruz
```

```
Out[58]: <AxesSubplot:xlabel='GönderilenZaman'>
```

