```
from google.colab import drive
drive.mount('/content/drive')

    Mounted at /content/drive


import pandas as pd
import numpy as np
import time



import pandas as pd
import time

start_time = time.time()

listening_csv_path = '/content/drive/MyDrive/dataset/listenings.csv'
listening_df = pd.read_csv(listening_csv_path, error_bad_lines=False)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)

    Skipping line 14591532: expected 7 fields, saw 8
    Skipping line 14591533: expected 7 fields, saw 8
    Skipping line 14591534: expected 7 fields, saw 8
    Skipping line 14591563: expected 7 fields, saw 8
    Skipping line 14592371: expected 7 fields, saw 8
    Skipping line 14592884: expected 7 fields, saw 8
    Skipping line 14594027: expected 7 fields, saw 11
    Skipping line 14597140: expected 7 fields, saw 8
    Skipping line 14597167: expected 7 fields, saw 8
    Skipping line 14597451: expected 7 fields, saw 8
    Skipping line 14599043: expected 7 fields, saw 9
    Skipping line 14599207: expected 7 fields, saw 9
    Skipping line 14601967: expected 7 fields, saw 8
    Skipping line 14604319: expected 7 fields, saw 11
    Skipping line 14605825: expected 7 fields, saw 9
    Skipping line 14606782: expected 7 fields, saw 8
    Skipping line 14606783: expected 7 fields, saw 8
    Skipping line 14606784: expected 7 fields, saw 8
    Skipping line 14606786: expected 7 fields, saw 8
    Skipping line 14606787: expected 7 fields, saw 8
    Skipping line 14606920: expected 7 fields, saw 8
    Skipping line 14606921: expected 7 fields, saw 8
    Skipping line 14606922: expected 7 fields, saw 8
    Skipping line 14606924: expected 7 fields, saw 8
    Skipping line 14606925: expected 7 fields, saw 8

    Skipping line 14609093: expected 7 fields, saw 8
    Skipping line 14609581: expected 7 fields, saw 9
    Skipping line 14610038: expected 7 fields, saw 8
    Skipping line 14611642: expected 7 fields, saw 8
    Skipping line 14612465: expected 7 fields, saw 8
    Skipping line 14612466: expected 7 fields, saw 8
    Skipping line 14612467: expected 7 fields, saw 8
    Skipping line 14612468: expected 7 fields, saw 8
    Skipping line 14612473: expected 7 fields, saw 8
    Skipping line 14612475: expected 7 fields, saw 8
    Skipping line 14612479: expected 7 fields, saw 8
    Skipping line 14612482: expected 7 fields, saw 8
    Skipping line 14613046: expected 7 fields, saw 8
    Skipping line 14613062: expected 7 fields, saw 8
    Skipping line 14613065: expected 7 fields, saw 9
    Skipping line 14613077: expected 7 fields, saw 8
    Skipping line 14613113: expected 7 fields, saw 9
    Skipping line 14613117: expected 7 fields, saw 8
    Skipping line 14613119: expected 7 fields, saw 11
    Skipping line 14613130: expected 7 fields, saw 8
    Skipping line 14613149: expected 7 fields, saw 8
    Skipping line 14613412: expected 7 fields, saw 8
    Skipping line 14614686: expected 7 fields, saw 9
    Skipping line 14614699: expected 7 fields, saw 9
    Skipping line 14615554: expected 7 fields, saw 8
    Skipping line 14616320: expected 7 fields, saw 8
    Skipping line 14617095: expected 7 fields, saw 8
    Skipping line 14619531: expected 7 fields, saw 8
    Skipping line 14619735: expected 7 fields, saw 9
    Skipping line 14619765: expected 7 fields, saw 8

    42.33872842788696


import pandas as pd
import time

# let's check the data:
```

```python
start_time = time.time()


print(listening_df)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
                    user_id          date                          track  \
    0             000Silenced  1299680100000                      Price Tag
    1             000Silenced  1299679920000  Price Tag (Acoustic Version)
    2             000Silenced  1299679440000      Be Mine! (Ballad Version)
    3             000Silenced  1299679200000                       Acapella
    4             000Silenced  1299675660000              I'm Not Invisible
    ...                  ...           ...                             ...
    14166904         zzzzita  1323117120000               Purple Light Dub
    14166905         zzzzita  1323117000000            Pretty Looks Version
    14166906         zzzzita  1323116820000              Ting A Ling Dub Wise
    14166907         zzzzita  1323116640000          You've Turned Away Dub
    14166908         zzzzita  1323116460000           Love Won't Come Dubbing

                   artist                                              album
    0             Jessie J                                         Who You Are
    1             Jessie J                                           Price Tag
    2               Robyn                                             Be Mine!
    3               Kelis                                             Acapella
    4            The Tease                                    I'm Not Invisible
    ...               ...                                               ...
    14166904  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
    14166905  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
    14166906  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
    14166907  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
    14166908  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...

    [13276607 rows x 5 columns]
    0.011649370193481445
```

```python
# drop the null rows:

start_time = time.time()


listening_df = listening_df.dropna()

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
    12.903310537338257
```

```python
# let's check the dataset again:

start_time = time.time()


print(listening_df)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
                    user_id          date                          track  \
    0             000Silenced  1299680100000                      Price Tag
    1             000Silenced  1299679920000  Price Tag (Acoustic Version)
    2             000Silenced  1299679440000      Be Mine! (Ballad Version)
    3             000Silenced  1299679200000                       Acapella
    4             000Silenced  1299675660000              I'm Not Invisible
    ...                  ...           ...                             ...
    14166904         zzzzita  1323117120000               Purple Light Dub
    14166905         zzzzita  1323117000000            Pretty Looks Version
    14166906         zzzzita  1323116820000              Ting A Ling Dub Wise
    14166907         zzzzita  1323116640000          You've Turned Away Dub
    14166908         zzzzita  1323116460000           Love Won't Come Dubbing

                   artist                                              album
    0             Jessie J                                         Who You Are
    1             Jessie J                                           Price Tag
    2               Robyn                                             Be Mine!
    3               Kelis                                             Acapella
    4            The Tease                                    I'm Not Invisible
    ...               ...                                               ...
    14166904  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
    14166905  The Heptones   The Heptones Dub Dictionary: The Backbone Insi...
```

```
      14166906  The Heptones  The Heptones Dub Dictionary: The Backbone Insi...
      14166907  The Heptones  The Heptones Dub Dictionary: The Backbone Insi...
      14166908  The Heptones  The Heptones Dub Dictionary: The Backbone Insi...

      [13276607 rows x 5 columns]
      0.009487628936767578


print(listening_df.info())

      <class 'pandas.core.frame.DataFrame'>
      Int64Index: 13276607 entries, 0 to 14166908
      Data columns (total 5 columns):
       #   Column   Dtype
      ---  ------   -----
       0   user_id  object
       1   date     int64
       2   track    object
       3   artist   object
       4   album    object
      dtypes: int64(1), object(4)
      memory usage: 607.8+ MB
      None


# let's see the shape of our dataframe:

start_time = time.time()

shape = listening_df.shape
print(shape)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)

      (13276607, 5)
      0.0020542144775390625


# select two columns: track and artist

start_time = time.time()


q0 = listening_df[['artist', 'track']]
print(q0)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)

                      artist                        track
      0             Jessie J                    Price Tag
      1             Jessie J  Price Tag (Acoustic Version)
      2               Robyn      Be Mine! (Ballad Version)
      3               Kelis                     Acapella
      4             The Tease              I'm Not Invisible
      ...                ...                          ...
      14166904  The Heptones             Purple Light Dub
      14166905  The Heptones          Pretty Looks Version
      14166906  The Heptones            Ting A Ling Dub Wise
      14166907  The Heptones           You've Turned Away Dub
      14166908  The Heptones       Love Won't Come Dubbing

      [13276607 rows x 2 columns]
      0.6459553241729736


# Let's find all of the records of those users who have listened to Rihanna

start_time = time.time()


q1 = listening_df[listening_df['artist'] == 'Rihanna']
print(q1)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)

                   user_id          date  \
      6           000Silenced  1294498440000
      28          000Silenced  1285438440000
      1623        00williamsl  1361485800000
      1635        00williamsl  1361485800000
      1654        00williamsl  1361048640000
```

```
...              ...              ...
14159310         ZueK666  1334970780000
14159311         ZueK666  1334970600000
14162708         Zwadzio  1446562980000
14162788         Zwadzio  1446562980000
14166430     zzzombiebones  1374354060000

                                              track    artist  \
6          Don't Stop The Music (Wideboys Bassline Mix)  Rihanna
28                                          Disturbia  Rihanna
1623                                Hatin On The Club  Rihanna
1635                                Hatin On The Club  Rihanna
1654                                      Complicated  Rihanna
...                                               ...      ...
14159310                                      Should I  Rihanna
14159311                        Pon De Replay (Remix)  Rihanna
14162708                            Where Have You Been  Rihanna
14162788                            Where Have You Been  Rihanna
14166430                                          S&M  Rihanna

                                     album
6                      Addicted 2 Bassline CD1
28            Good Girl Gone Bad: The Remixes
1623                                    Random
1635                                    Random
1654                                      Loud
...                                        ...
14159310      Music Of The Sun [Bonus Track]
14159311      Music Of The Sun [Bonus Track]
14162708      Talk That Talk (Deluxe Edition)
14162788      Talk That Talk (Deluxe Edition)
14166430                                  Loud

[41422 rows x 5 columns]
1.067509651184082
```

```python
# Let's find top 10 users who are fan of Rihanna

start_time = time.time()


q2 = listening_df[listening_df['artist'] == 'Rihanna'].groupby('user_id').size().reset_index(name='count').sort_values(by='count', ascen
print(q2)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
              user_id  count
7851           thiessu    179
4502        eyessetkyle    166
3102               adxx    164
6313      missnumberthree    156
4921      helloiamnatalie    128
6594               nmjnb    124
160              AndyyyA    123
223            BIGBANG186    121
6320           mixedvibes    120
2173      Pinky_von_Pout    115
1.902695894241333
```

```python
# Top 10 famous track

start_time = time.time()


q3 = listening_df.groupby(['artist', 'track']).size().reset_index(name='count').sort_values(by='count', ascending=False).head(10)
print(q3)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
                artist            track  count
884334     Justin Bieber            Sorry   3381
144100     Arctic Monkeys  Do I Wanna Know?   2862
1870435              Zayn        PILLOWTALK   2701
1648225        The Killers     Mr Brightside   2685
1357410            Rihanna              Work   2643
196192            Bastille           Pompeii   2606
1143458     Mumford & Sons  Little Lion Man   2514
884208      Justin Bieber     Love Yourself   2481
884396      Justin Bieber  What Do You Mean?   2480
1143517     Mumford & Sons          The Cave   2478
16.40806818008423
```

```python
# find top 10 famous tracks of Rihanna

start_time = time.time()


q4 = listening_df[listening_df['artist'] == 'Rihanna'].groupby(['artist', 'track']).size().reset_index(name='count').sort_values(by='cou
print(q4)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
         artist                          track  count
    931  Rihanna                           Work   2643
    408  Rihanna        Only Girl (In the World)   1747
    832  Rihanna  We Found Love (feat Calvin Harris)   1573
    629  Rihanna                            S&M   1304
    569  Rihanna                        Rude Boy   1301
    127  Rihanna                        Diamonds   1217
    297  Rihanna                  Kiss it Better    938
    889  Rihanna             Where Have You Been    843
     89  Rihanna          Cheers (Drink to That)    697
    364  Rihanna                       Needed Me    675
    1.0711939334869385
```

```python
# find top 10 famous albums

start_time = time.time()

q5 = listening_df.groupby(['artist', 'album']).size().reset_index(name='count').sort_values(by='count', ascending=False).head(10)
print(q5)

end_time = time.time()
execution_time = end_time - start_time
print(execution_time)
```

```
                     artist              album  count
    494296       Kanye West  The Life Of Pablo  22278
    940552           The xx                 xx  14131
    91664     Arctic Monkeys                 AM  14088
    1021188            alt-J   An Awesome Wave  13613
    628878    Mumford & Sons      Sigh No More  13521
    391411            Grimes        Art Angels  11628
    345310  Florence + the Machine           Lungs  11341
    54501             Adele                 21  11198
    527507    Lana Del Rey       Born to Die   9984
    90887       Arcade Fire        The Suburbs   9793
    14.446802139282227
```