

Analysis on Listening music in Different Age groups

The Data is a comprehensive collection of anonymized data that offers valuable insights into the behavior patterns and preferences of Spotify users. The data tells about how they are using the Spotify, what influence them to listen the music, time slot for listening the music. By listening the music whether the user's mood changes or not. What type of music and podcast are most likely to listen.

Understanding the Dataset

The Dataset have 520 rows and 18 columns. Each column is described below.

- Age : Age of each User
- Gender : Gender of each User
- Spotify_usage_period : How long have you been using Spotify?
- Hours per day : How much time have they listening music in a day?
- Spotify_listening_device : Which of the following devices do you primarily use to listen to Spotify?
- Preferred_listening_content : What do you prefer to listen more?
- Fav_music_genre : What genre(s) of music do you enjoy the most?
- Music_time_slot : What is your favourite time slot to listen to music?
- Music_influence_mood : When it comes to listening to music, which of the following moods or situations most strongly influences your choice of music?
- Music_lis_frequency : When do you listen to music more often?
- Music_expl_method : How do you discover new music on Spotify?
- Music_recc_rating : How do you rate the spotify music recommendations?
- Pod_lis_frequency : How often do you listen to Podcast?
- Fav_pod_genre : What genre(s) of Podcast do you enjoy the most?
- Pod_host_preference : Are you more inclined to listen to podcasts from unknown personalities, or do you prefer podcasts hosted by well-known individuals?
- Pod_variety_satisfaction : Are you satisfied with the variety and availability of podcasts on Spotify?
- Music_effects : Whether it is helpful to change your mood?
- Unnamed 17 : unwanted column.

Data Preparation and Cleaning

Data Cleaning is an important step and time-consuming step in data analysis. Formatting issues (e.g., rows and columns merged), missing data, duplicated rows, spelling mistakes, and so on could all be present. These challenges could make data analysis complex. Here we will address some of those.

In [2]:

```
1 # importing libraries
2
3 import numpy as np
4 import pandas as pd
5 import seaborn as sns
6 import matplotlib.pyplot as plt
7 from sklearn.preprocessing import LabelEncoder
8 import warnings
9 warnings.filterwarnings('ignore') # if there are any warnings due to version mismatch
```

Load the Dataset

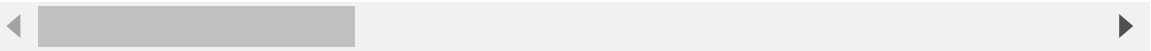
In [3]:

```
1 ds= pd.read_csv('Music_dataset.csv')
2 ds
```

Out[3]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
1	63.0	Male	More than 2 years	1.5	Computer or laptop	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
3	61.0	Female	1 year to 2 years	2.5	Smartphone, Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
...
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1558	56.0	Female	6 months to 1 year	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1560 rows × 7 columns



Some Basic Functions

In [4]:

```
1 # shape :- to find the number of columns and number of rows.
2
3 ds.shape
```

Out[4]:

(1560, 17)

In [5]:

```
1 #columns :- To find the column names.
2
3 ds.columns
```

Out[5]:

```
Index(['Age', 'Gender', 'spotify_usage_period', 'Hours per day',
       'spotify_listening_device', 'preferred_listening_content',
       'fav_music_genre', 'music_time_slot', 'music_Influencial_mood',
       'music_lis_frequency', 'music_expl_method', 'music_recc_rating',
       'pod_lis_frequency', 'fav_pod_genre', 'pod_host_preference',
       'pod_variety_satisfaction', 'Music effects'],
      dtype='object')
```

In [6]:

```
1 # head() :- to display the first five rows in the dataset.
2
3 ds.head()
```

Out[6]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening_cc
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	Pe
1	63.0	Male	More than 2 years	1.5	Computer or laptop	Pe
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	Pe
3	61.0	Female	1 year to 2 years	2.5	Smartphone, Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	

In [7]:

```

1 # tail() :- to display the last five rows in the dataset
2
3 ds.tail()

```

Out[7]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1558	56.0	Female	6 months to 1 year	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

In [8]:

```

1 # info() :- gives the details about the datatype and non-Null Count
2
3 ds.info()

```

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1560 entries, 0 to 1559

Data columns (total 17 columns):

#	Column	Non-Null Count	Dtype
0	Age	1557 non-null	float64
1	Gender	1560 non-null	object
2	spotify_usage_period	1560 non-null	object
3	Hours per day	1560 non-null	float64
4	spotify_listening_device	1560 non-null	object
5	preferred_listening_content	1560 non-null	object
6	fav_music_genre	1560 non-null	object
7	music_time_slot	1560 non-null	object
8	music_Influencial_mood	1560 non-null	object
9	music_lis_frequency	1560 non-null	object
10	music_expl_method	1560 non-null	object
11	music_recc_rating	1560 non-null	int64
12	pod_lis_frequency	1560 non-null	object
13	fav_pod_genre	1560 non-null	object
14	pod_host_preference	1560 non-null	object
15	pod_variety_satisfaction	1560 non-null	object
16	Music effects	1542 non-null	object

dtypes: float64(2), int64(1), object(14)

memory usage: 207.3+ KB

In [9]:

```
1 # value_counts() :- It counts the number of times a particular item repeated in a sp  
2  
3 ds['Music effects'].value_counts()
```

Out[9]:

```
Improve      1155  
No effect    354  
Worsen       33  
Name: Music effects, dtype: int64
```

In [10]:

```
1 ds['Gender'].value_counts()
```

Out[10]:

```
Female      1173  
Male        342  
Others       45  
Name: Gender, dtype: int64
```

In [11]:

```
1 ds['music_Influencial_mood'].value_counts()
```

Out[11]:

```
Relaxation and stress relief
585
Uplifting and motivational
201
Sadness or melancholy
165
Relaxation and stress relief, Uplifting and motivational
132
Relaxation and stress relief, Uplifting and motivational, Sadness or mela
ncholy, Social gatherings or parties    105
Relaxation and stress relief, Sadness or melancholy
99
Relaxation and stress relief, Uplifting and motivational, Sadness or mela
ncholy                                66
Social gatherings or parties
48
Relaxation and stress relief, Uplifting and motivational, Social gatherin
gs or parties                        42
Relaxation and stress relief, Social gatherings or parties
39
Uplifting and motivational, Sadness or melancholy
36
Relaxation and stress relief, Sadness or melancholy, Social gatherings or
parties                             24
Uplifting and motivational, Social gatherings or parties
12
Sadness or melancholy, Social gatherings or parties
3
Uplifting and motivational, Sadness or melancholy, Social gatherings or p
arties                             3
Name: music_Influencial_mood, dtype: int64
```

In [12]:

```
1 ds['fav_music_genre'].value_counts()
```

Out[12]:

```
Melody                777
classical             261
Pop                   255
Rap                   165
Electronic/Dance      48
All                   18
Rock                  12
Kpop                  12
Classical & melody, dance    6
Old songs              3
trending songs random    3
Name: fav_music_genre, dtype: int64
```

In [13]:

```
1 ds['music_recc_rating'].value_counts()
```

Out[13]:

```
3    570
4    522
5    258
2    168
1     42
```

Name: music_recc_rating, dtype: int64

In [14]:

```
1 ds['pod_lis_frequency'].value_counts()
```

Out[14]:

```
Rarely          603
Never           390
Once a week     273
Several times a week 234
Daily           60
```

Name: pod_lis_frequency, dtype: int64

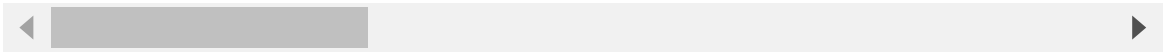
In [15]:

```
1 ds.rename(columns={'music_lis_frequency':'frequently_lis_music'},inplace=True)
2 ds
```

Out[15]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
1	63.0	Male	More than 2 years	1.5	Computer or laptop	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
3	61.0	Female	1 year to 2 years	2.5	Smartphone, Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
...	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1558	56.0	Female	6 months to 1 year	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1560 rows × 17 columns



Data Cleaning

In [16]:

```
1 # isnull() :- It shows the null values present in the columns as True(if it is null)
2 # sum() :- It add the count of the null values present in the columns
3
4 ds.isnull().sum()
```

Out[16]:

Age	3
Gender	0
spotify_usage_period	0
Hours per day	0
spotify_listening_device	0
preferred_listening_content	0
fav_music_genre	0
music_time_slot	0
music_Influencial_mood	0
frequently_lis_music	0
music_expl_method	0
music_recc_rating	0
pod_lis_frequency	0
fav_pod_genre	0
pod_host_preference	0
pod_variety_satisfaction	0
Music effects	18
dtype: int64	

In [17]:

```
1 # It is used to find the mean of remaining data
2
3 ds['Age'].mean().round(0)
```

Out[17]:

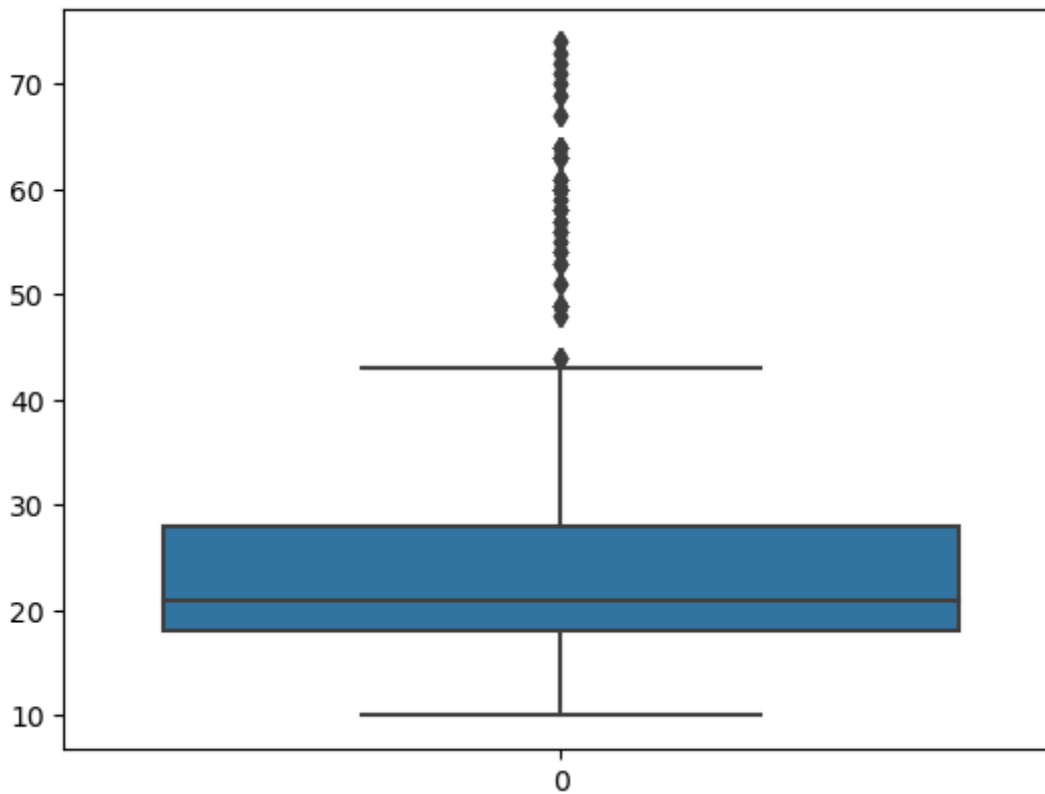
25.0

In [18]:

```
1 # Lets draw the Box plot for Age column to find the Outliers
2
3 sns.boxplot(ds['Age'])
```

Out[18]:

<Axes: >



In [19]:

```
1 # Cleaning the Outliers from the data
2 ds1 = ds[ds['Age']<=43]
3 ds1
```

Out[19]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 7 columns

◀

▶

In [20]:

```
1 ds1.shape
```

Out[20]:

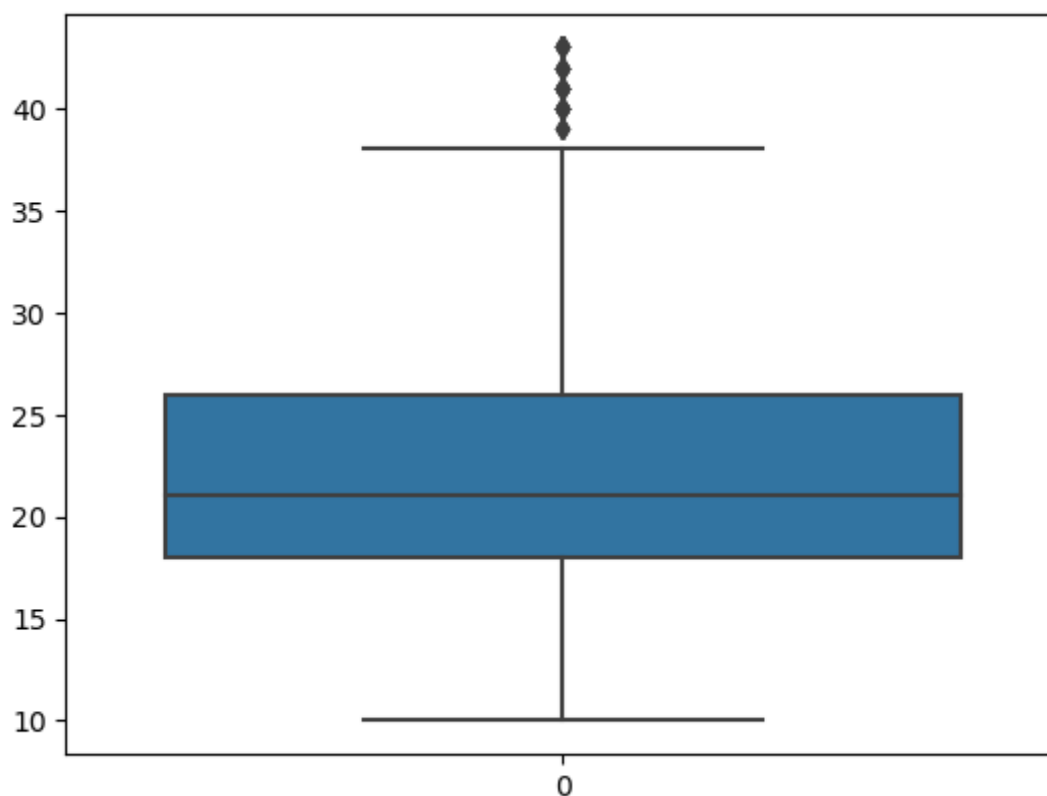
(1434, 7)

In [21]:

```
1 # after cleaning the Outliers(i.e age greater than 43), Lets draw the boxplot
2
3 sns.boxplot(ds1['Age'])
```

Out[21]:

<Axes: >



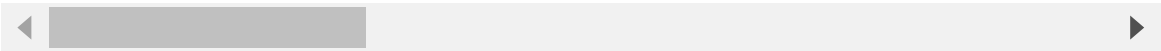
In [22]:

```
1 # to fill null values
2
3 ds1['Age'].fillna(ds1['Age'].mean().round(0),inplace=True)
4 ds1
```

Out[22]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 7 columns



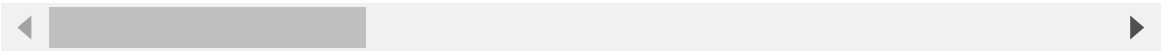
In [23]:

```
1 ds1['Music effects'].fillna('Improve',inplace=True)
2 ds1
```

Out[23]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 17 columns



In [24]:

```
1 ds1.isnull().sum()
```

Out[24]:

```
Age                0
Gender             0
spotify_usage_period  0
Hours per day      0
spotify_listening_device  0
preferred_listening_content  0
fav_music_genre    0
music_time_slot    0
music_Influencial_mood  0
frequently_lis_music  0
music_expl_method  0
music_recc_rating   0
pod_lis_frequency   0
fav_pod_genre       0
pod_host_preference  0
pod_variety_satisfaction  0
Music effects       0
dtype: int64
```

In [25]:

```
1 ds1.shape
```

Out[25]:

```
(1434, 17)
```

In [26]:

```
1 ds1
```

Out[26]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 17 columns



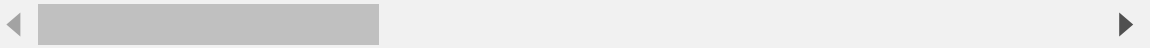
In [27]:

```
1 ds1.drop(['pod_host_preference'],axis=1,inplace=True)
2 ds1
```

Out[27]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



In [28]:

```
1 ds1.shape
```

Out[28]:

(1434, 16)

In [29]:

```
1 ds1.isnull().sum()
```

Out[29]:

Age	0
Gender	0
spotify_usage_period	0
Hours per day	0
spotify_listening_device	0
preferred_listening_content	0
fav_music_genre	0
music_time_slot	0
music_Influencial_mood	0
frequently_lis_music	0
music_expl_method	0
music_recc_rating	0
pod_lis_frequency	0
fav_pod_genre	0
pod_variety_satisfaction	0
Music effects	0

dtype: int64

Data Filtering

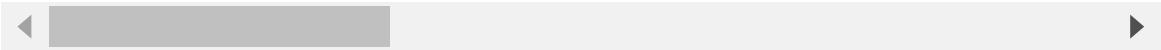
In [30]:

```
1 ds2=ds1[ds1["Age"] < 15]
2 ds2
```

Out[30]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
26	14.0	Male	6 months to 1 year	12.0	Smartphone, Smart speakers or voice assistants	
86	14.0	Male	6 months to 1 year	7.0	Computer or laptop	
127	13.0	Female	6 months to 1 year	2.0	Smartphone	
230	14.0	Male	1 year to 2 years	2.0	Smartphone	
278	14.0	Female	6 months to 1 year	4.0	Smart speakers or voice assistants	
...
1476	14.0	Female	Less than 6 months	9.0	Smartphone	
1482	13.0	Female	More than 2 years	6.0	Smartphone, Computer or laptop	
1488	14.0	Female	6 months to 1 year	2.0	Smartphone	
1527	14.0	Female	1 year to 2 years	2.0	Computer or laptop, Smart speakers or voice as...	
1536	14.0	Male	6 months to 1 year	7.0	Computer or laptop	

75 rows × 16 columns



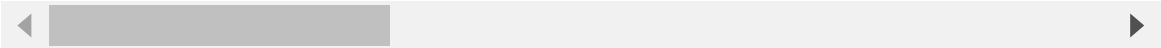
In [31]:

```
1 ds3=ds1[ds1.Gender=="Male"]
2 ds3
```

Out[31]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
5	18.0	Male	More than 2 years	5.0	Smartphone	
21	17.0	Male	More than 2 years	4.0	Smartphone	
24	16.0	Male	More than 2 years	3.0	Smartphone, Computer or laptop	
26	14.0	Male	6 months to 1 year	12.0	Smartphone, Smart speakers or voice assistants	
27	18.0	Male	Less than 6 months	6.0	Smartphone	
...
1536	14.0	Male	6 months to 1 year	7.0	Computer or laptop	
1537	20.0	Male	1 year to 2 years	3.0	Smart speakers or voice assistants	
1541	24.0	Male	6 months to 1 year	2.0	Smart speakers or voice assistants	
1544	19.0	Male	More than 2 years	2.0	Smart speakers or voice assistants	
1546	28.0	Male	Less than 6 months	8.0	Smart speakers or voice assistants	

324 rows × 16 columns



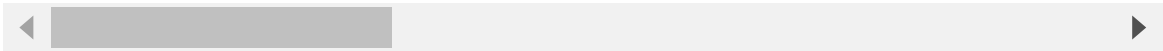
In [32]:

```
1 ds4=ds1[ds1.music_recc_rating==3]
2 ds4
```

Out[32]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
10	18.0	Female	More than 2 years	3.0	Smartphone	
11	19.0	Female	1 year to 2 years	8.0	Smartphone	
...
1544	19.0	Male	More than 2 years	2.0	Smart speakers or voice assistants	
1545	25.0	Female	1 year to 2 years	4.0	Wearable devices	
1547	27.0	Female	Less than 6 months	4.0	Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

516 rows × 16 columns



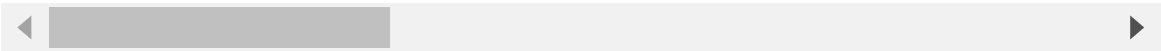
In [33]:

```
1 ds5=ds1[ds1.spotify_listening_device=="Smartphone"]
2 ds5
```

Out[33]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
7	21.0	Female	Less than 6 months	1.0	Smartphone	
8	19.0	Female	Less than 6 months	6.0	Smartphone	
...
1503	27.0	Female	1 year to 2 years	1.0	Smartphone	
1540	29.0	Female	6 months to 1 year	8.0	Smartphone	
1549	25.0	Female	More than 2 years	5.0	Smartphone	
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1557	19.0	Female	More than 2 years	8.0	Smartphone	

837 rows × 16 columns



Data Sorting

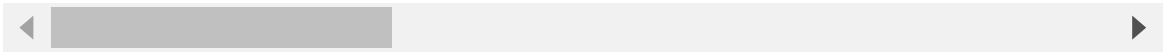
In [34]:

```
1 ds6=ds1.copy()
2 ds6
```

Out[34]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



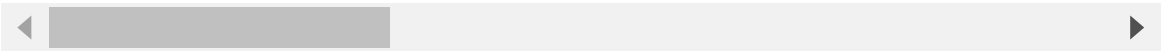
In [35]:

```
1 ds6.sort_values(by='Age',inplace=True)
2 ds6
```

Out[35]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
369	10.0	Female	6 months to 1 year	2.0	Smartphone	
1409	10.0	Female	6 months to 1 year	2.0	Smartphone	
889	10.0	Female	6 months to 1 year	2.0	Smartphone	
384	12.0	Female	Less than 6 months	2.0	Smartphone	
931	12.0	Female	6 months to 1 year	0.5	Smartphone	
...
1031	43.0	Female	More than 2 years	1.0	Smartphone	
1551	43.0	Female	More than 2 years	1.0	Smartphone	
887	43.0	Female	Less than 6 months	1.0	Smartphone	
1407	43.0	Female	Less than 6 months	1.0	Smartphone	
91	43.0	Female	Less than 6 months	6.0	Smartphone	

1434 rows × 16 columns



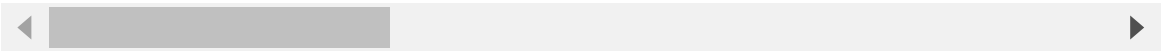
In [36]:

```
1 ds7=ds1.copy()
2 ds7
```

Out[36]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



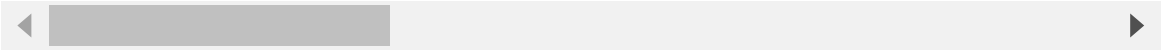
In [37]:

```
1 ds7.sort_values(by='music_recc_rating', inplace=True)
2 ds7
```

Out[37]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
200	25.0	Male	Less than 6 months	3.0	Smartphone	
533	19.0	Female	More than 2 years	2.0	Smartphone	
421	14.0	Female	Less than 6 months	5.0	Smartphone	
391	18.0	Female	More than 2 years	5.0	Smartphone	
1461	14.0	Female	Less than 6 months	5.0	Smartphone	
...
157	28.0	Male	Less than 6 months	4.0	Smartphone	
1258	20.0	Female	More than 2 years	3.0	Computer or laptop	
400	17.0	Female	1 year to 2 years	2.0	Smartphone	
1254	19.0	Male	More than 2 years	6.0	Smartphone	
767	25.0	Female	More than 2 years	2.0	Smartphone	

1434 rows × 16 columns



Aggregations

In [38]:

```
1 ds1
```

Out[38]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



In [39]:

```
1 ds1.describe()
```

Out[39]:

	Age	Hours per day	music_recc_rating
count	1434.000000	1434.000000	1434.000000
mean	22.472803	3.569979	3.520921
std	6.771901	2.993647	0.982709
min	10.000000	0.000000	1.000000
25%	18.000000	2.000000	3.000000
50%	21.000000	3.000000	4.000000
75%	26.000000	5.000000	4.000000
max	43.000000	24.000000	5.000000

Replace method

In [40]:

```
1 le=LabelEncoder()
```

In [41]:

```
1 ds1['preferred_listening_content']=le.fit_transform(ds1['preferred_listening_content'])
2 # 0 for Podcast
3 # 1 for Music
```

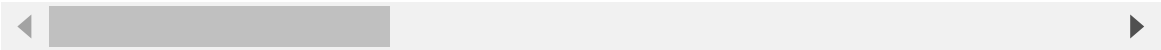
In [42]:

```
1 ds1
```

Out[42]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



In [43]:

```
1 ds1.describe()
```

Out[43]:

	Age	Hours per day	preferred_listening_content	music_recc_rating
count	1434.000000	1434.000000	1434.000000	1434.000000
mean	22.472803	3.569979	0.200837	3.520921
std	6.771901	2.993647	0.400766	0.982709
min	10.000000	0.000000	0.000000	1.000000
25%	18.000000	2.000000	0.000000	3.000000
50%	21.000000	3.000000	0.000000	4.000000
75%	26.000000	5.000000	0.000000	4.000000
max	43.000000	24.000000	1.000000	5.000000

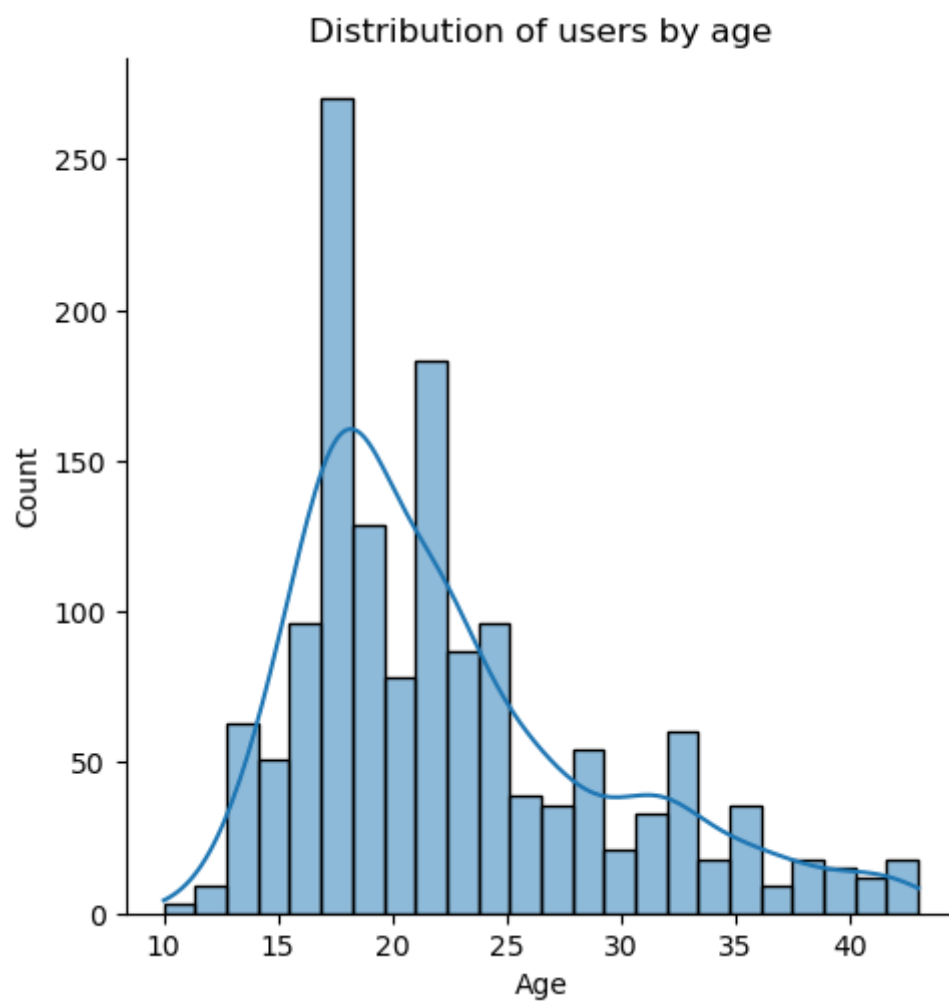
Data Visualization

In [44]:

```
1 data=ds1
2
3 sns.displot(data,x='Age',kde=True)
4 plt.title("Distribution of users by age")
```

Out[44]:

Text(0.5, 1.0, 'Distribution of users by age')

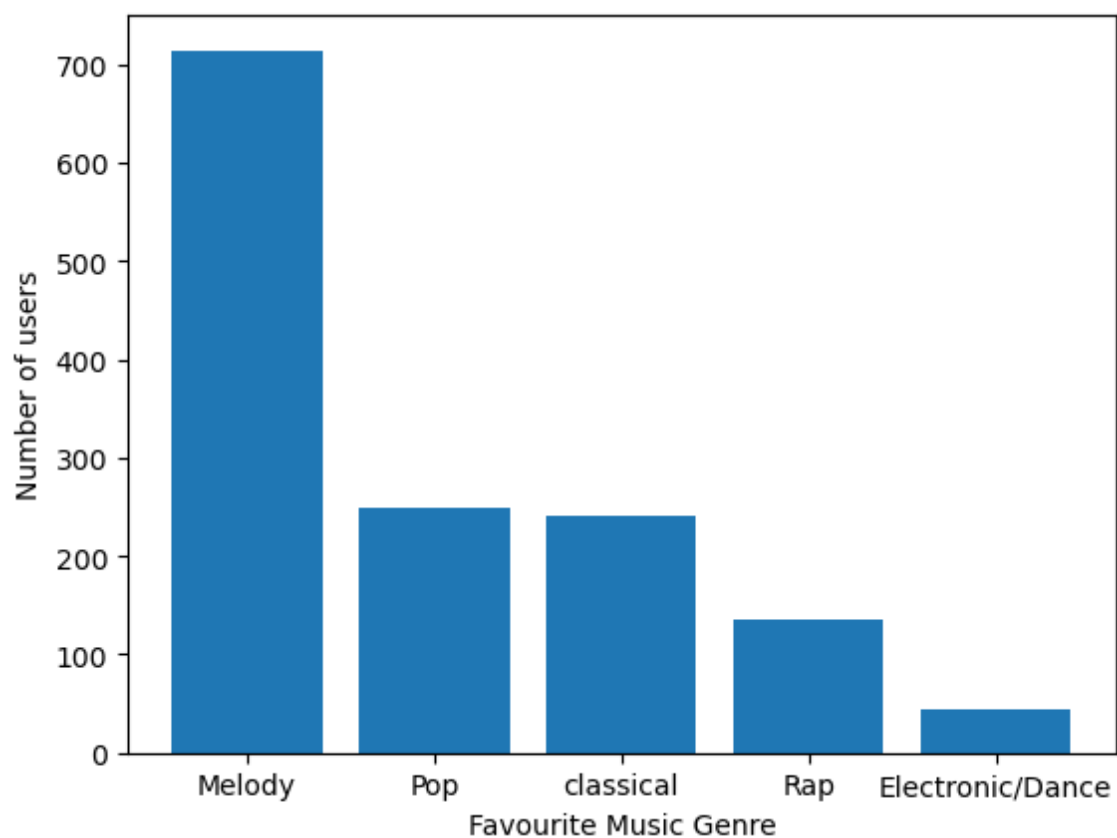


In [45]:

```
1 l1=ds1['fav_music_genre'].value_counts().head().keys()
2 l2=ds1['fav_music_genre'].value_counts().head()
3
4 plt.xlabel('Favourite Music Genre')
5 plt.ylabel('Number of users')
6 plt.bar(l1,l2)
```

Out[45]:

<BarContainer object of 5 artists>

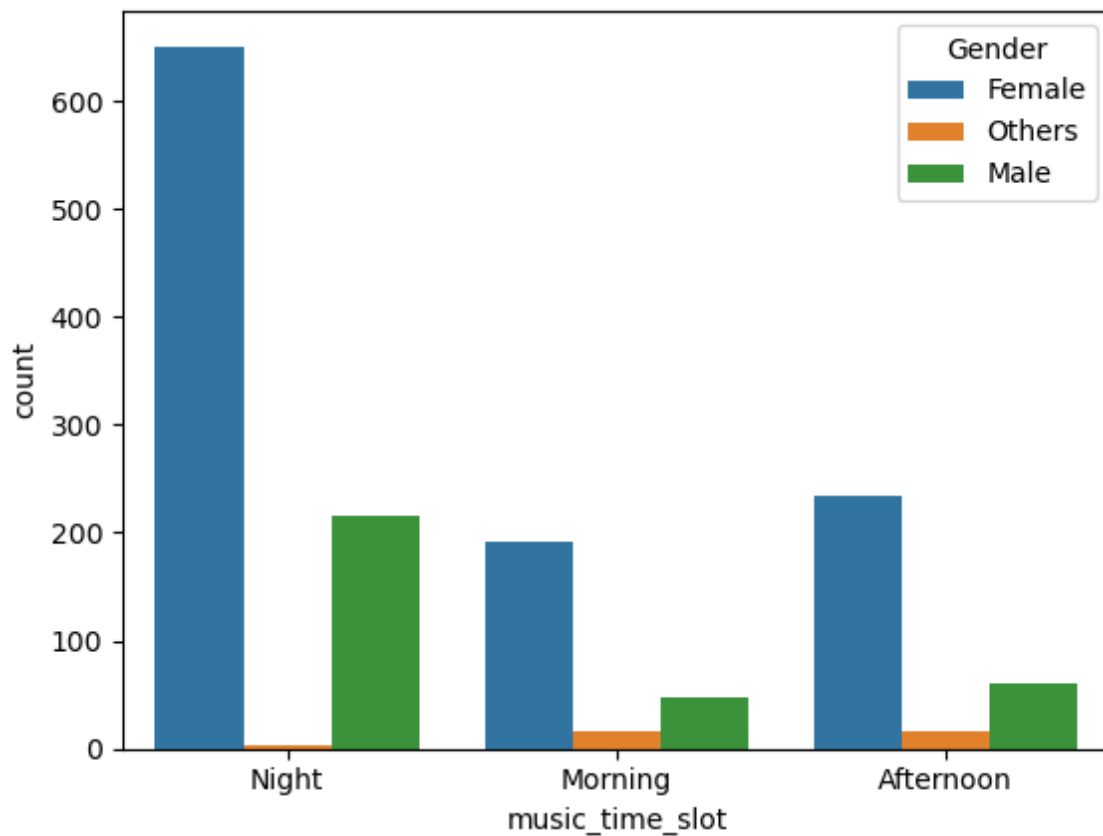


In [46]:

```
1 data=ds1
2 sns.countplot(data,x='music_time_slot',hue='Gender')
```

Out[46]:

<Axes: xlabel='music_time_slot', ylabel='count'>

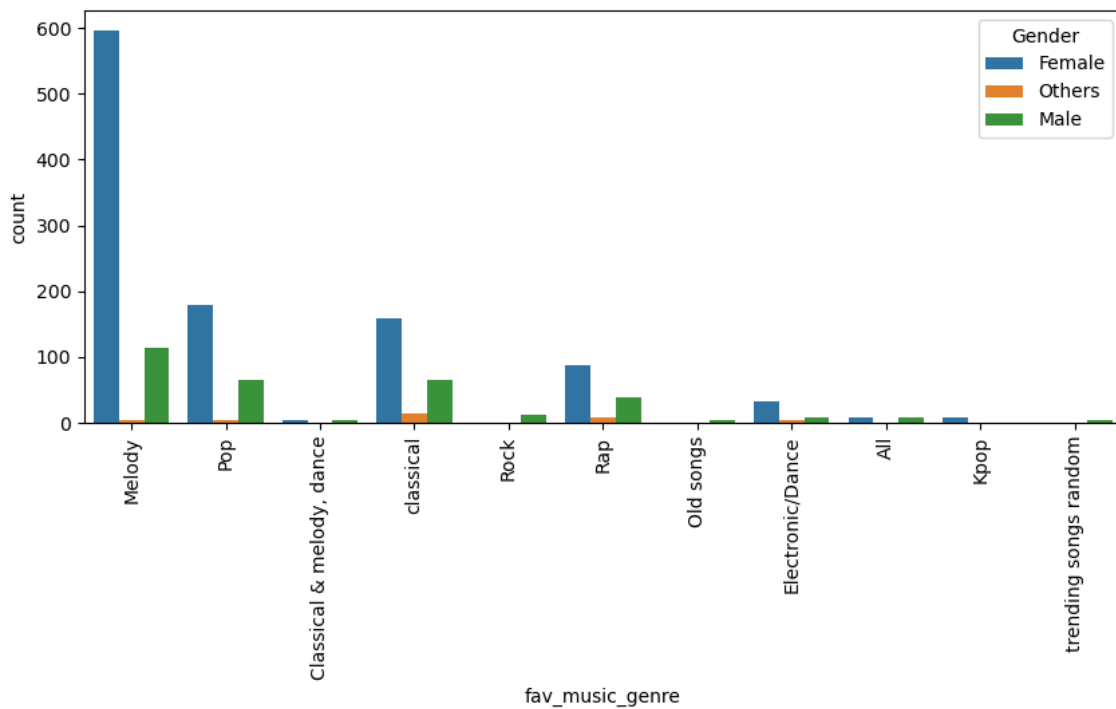


In [47]:

```
1 data=ds1
2 pt=(10,4)
3 plt.figure(figsize=pt)
4 sns.countplot(data,x='fav_music_genre',hue='Gender')
5 plt.xticks(rotation = 90)
```

Out[47]:

```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10]),
 [Text(0, 0, 'Melody'),
  Text(1, 0, 'Pop'),
  Text(2, 0, 'Classical & melody, dance'),
  Text(3, 0, 'classical'),
  Text(4, 0, 'Rock'),
  Text(5, 0, 'Rap'),
  Text(6, 0, 'Old songs'),
  Text(7, 0, 'Electronic/Dance'),
  Text(8, 0, 'All'),
  Text(9, 0, 'Kpop'),
  Text(10, 0, 'trending songs random')])
```

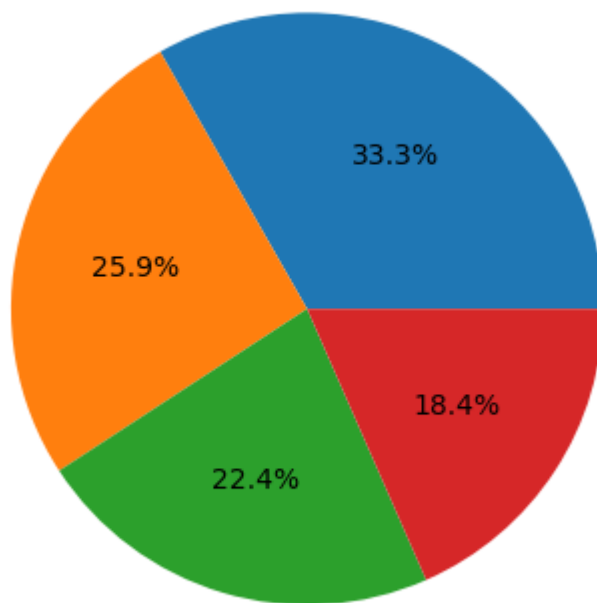


In [48]:

```
1 l3 = ds1['spotify_usage_period'].value_counts().head().keys()
2 l4 = ds1['spotify_usage_period'].value_counts().head()
3
4 plt.pie(l4,autopct='%1.1f%%')
```

Out[48]:

```
([<matplotlib.patches.Wedge at 0x286793e49d0>,
 <matplotlib.patches.Wedge at 0x286793e48e0>,
 <matplotlib.patches.Wedge at 0x286793e56c0>,
 <matplotlib.patches.Wedge at 0x286793e5d50>],
 [Text(0.5520857289507021, 0.9514206997374884, ''),
 Text(-1.0693532116452458, 0.25784434983919774, ''),
 Text(-0.3136875986236319, -1.0543244711519026, ''),
 Text(0.9210911298814088, -0.6013244801717206, '')],
 [Text(0.30113767033674654, 0.5189567453113573, '33.3%'),
 Text(-0.5832835699883158, 0.1406423726395624, '25.9%'),
 Text(-0.171102326521981, -0.5750860751737651, '22.4%'),
 Text(0.5024133435716774, -0.3279951710027566, '18.4%')])
```



In [49]:

```

1 l5 = ds1['music_Influencial_mood'].value_counts().head(7).keys()
2 l6 = ds1['music_Influencial_mood'].value_counts().head(7)
3 plt.pie(l6,labels=l5,autopct='%1.1f%%')

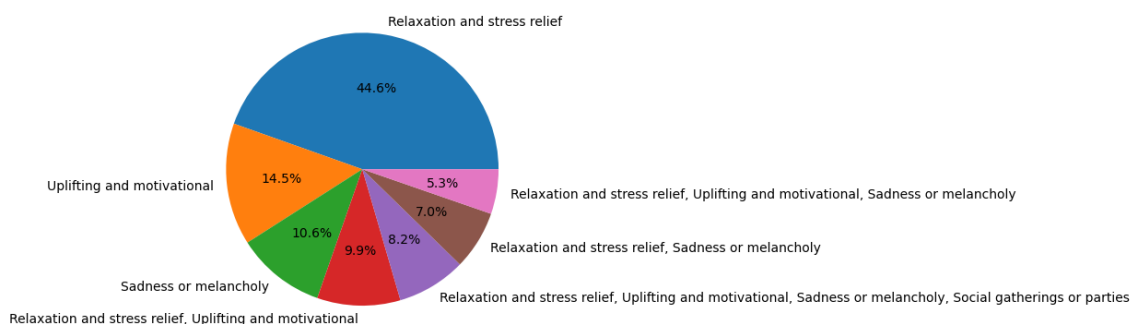
```

Out[49]:

```

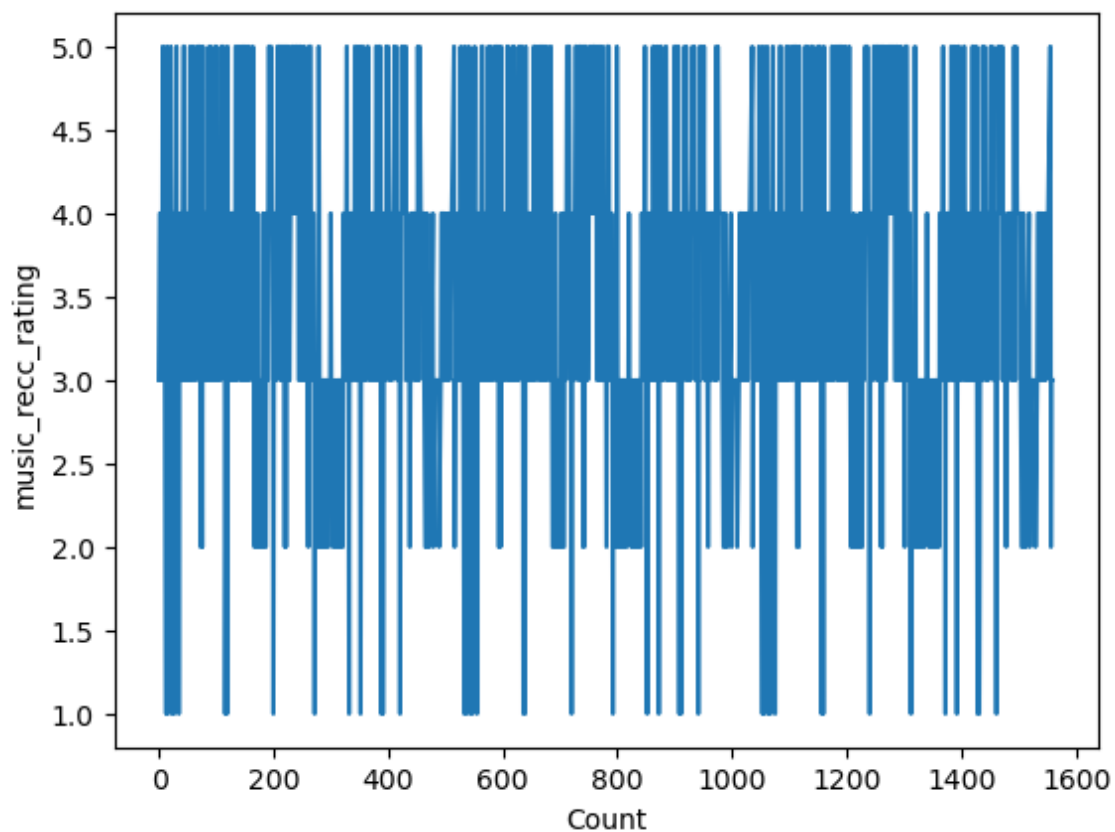
([<matplotlib.patches.Wedge at 0x2867944c370>,
 <matplotlib.patches.Wedge at 0x2867944c280>,
 <matplotlib.patches.Wedge at 0x2867944cf40>,
 <matplotlib.patches.Wedge at 0x2867944d5d0>,
 <matplotlib.patches.Wedge at 0x2867944dc60>,
 <matplotlib.patches.Wedge at 0x2867944e2f0>,
 <matplotlib.patches.Wedge at 0x2867944e980>],
 [Text(0.18645540811561334, 1.0840822758372357, 'Relaxation and stress re
lief'),
 Text(-1.092915940369366, -0.1246384663197701, 'Uplifting and motivation
al'),
 Text(-0.6830450111386719, -0.8622351841339878, 'Sadness or melanchol
y'),
 Text(-0.029141492485161364, -1.0996139201627713, 'Relaxation and stress
relief, Uplifting and motivational'),
 Text(0.5667401365619347, -0.9427648792831433, 'Relaxation and stress re
lief, Uplifting and motivational, Sadness or melancholy, Social gathering
s or parties'),
 Text(0.9362688075624267, -0.5774086247932496, 'Relaxation and stress re
lief, Sadness or melancholy'),
 Text(1.0847802449939465, -0.1823508159314718, 'Relaxation and stress re
lief, Uplifting and motivational, Sadness or melancholy')],
 [Text(0.10170294988124363, 0.5913176050021285, '44.6%'),
 Text(-0.5961359674741996, -0.06798461799260186, '14.5%'),
 Text(-0.37257000607563917, -0.47031010043672056, '10.6%'),
 Text(-0.01589535953736074, -0.5997894109978751, '9.9%'),
 Text(0.309130983579237, -0.5142353886998964, '8.2%'),
 Text(0.5106920768522327, -0.3149501589781361, '7.0%'),
 Text(0.5916983154512435, -0.09946408141716642, '5.3%')])

```



In [50]:

```
1 ds1['music_recc_rating'].plot()  
2 plt.ylabel('music_recc_rating')  
3 plt.xlabel('Count')  
4 plt.show()
```

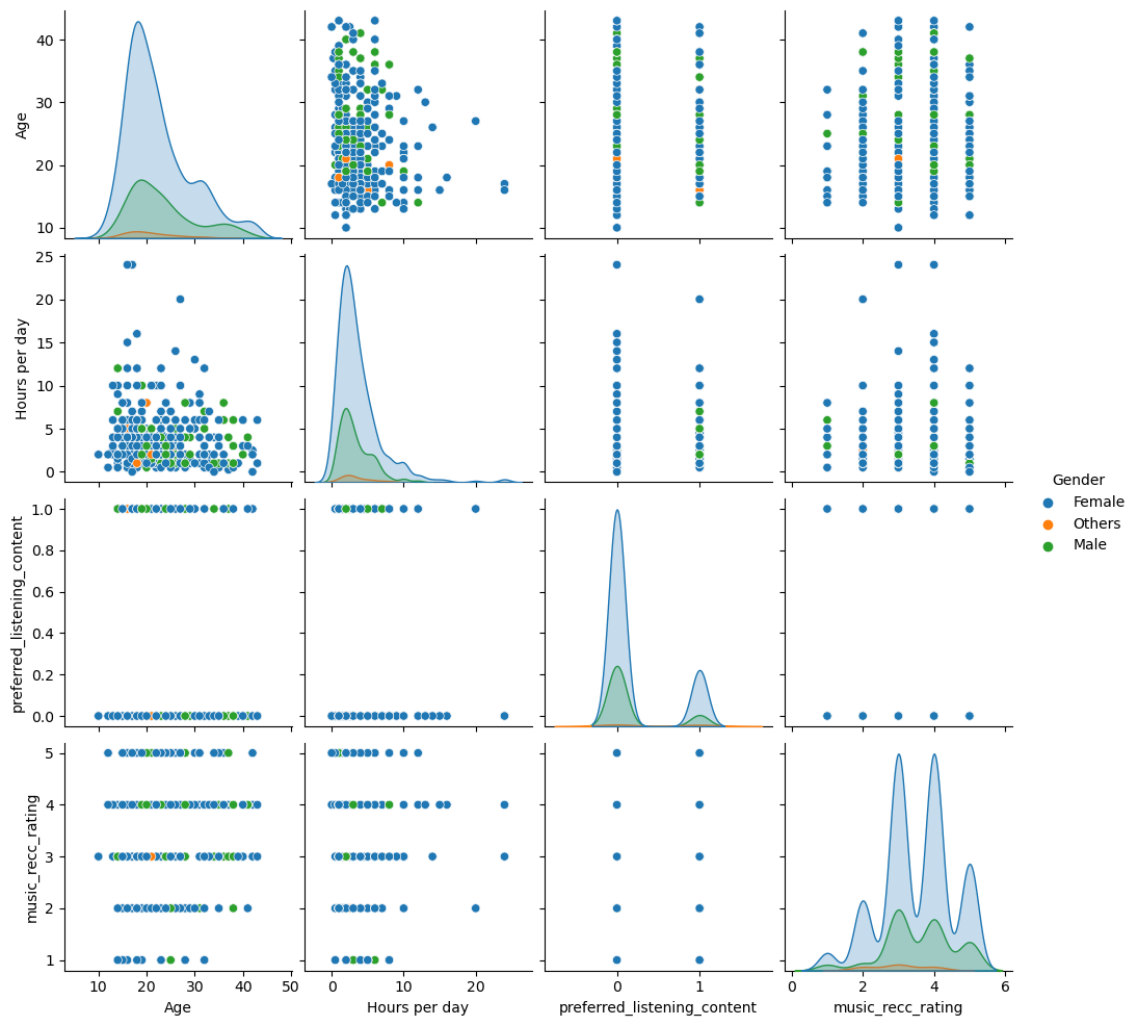


In [51]:

```
1 data=ds1
2 sns.pairplot(data,hue="Gender")
```

Out[51]:

<seaborn.axisgrid.PairGrid at 0x28679490d90>

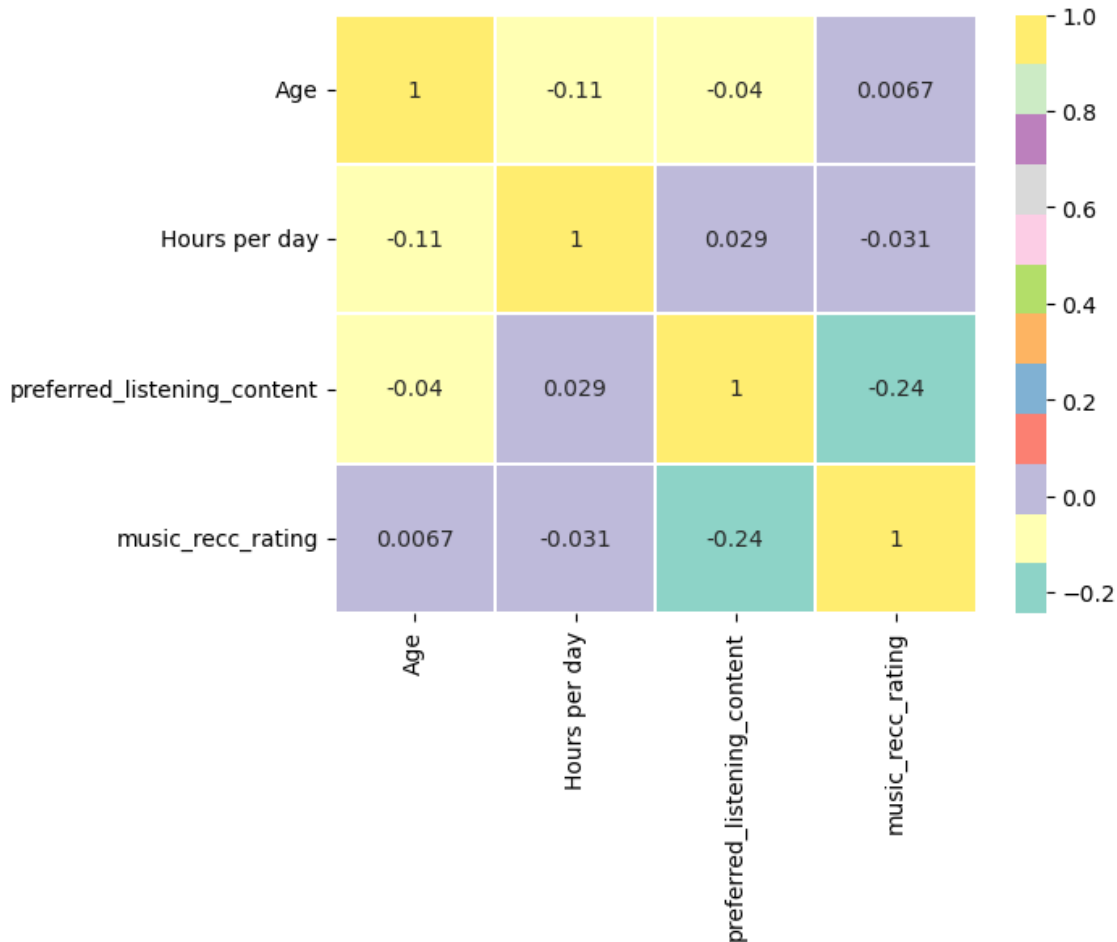


In [52]:

```
1 correlations=ds1.corr()
2 sns.heatmap(correlations, annot=True, cmap="Set3", linewidths=.20)
```

Out[52]:

<Axes: >



Some of the queries listed below

1. Which gender is spending more time in listening the music?

In [53]:

```
1 ds['Gender'].value_counts()
```

Out[53]:

```
Female    1173
Male       342
Others      45
Name: Gender, dtype: int64
```

Female users are more when compared to male and other users.

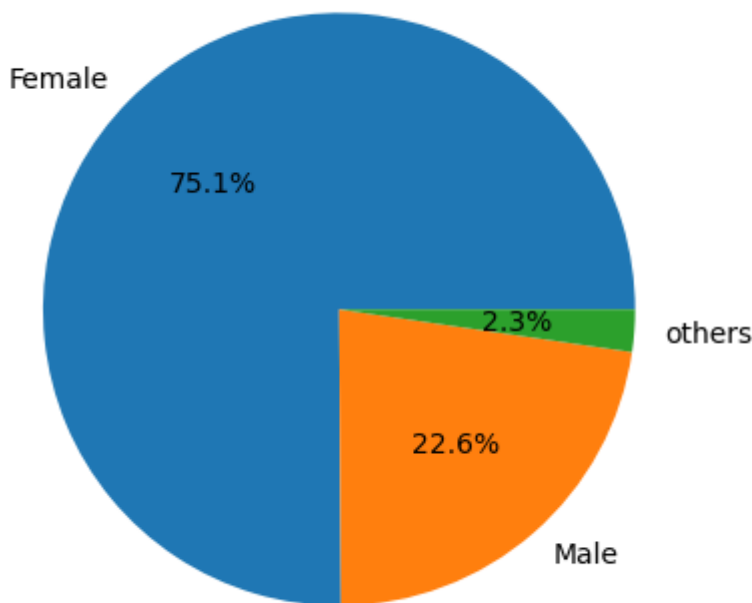
Here is the represntation for the above query:

In [54]:

```
1 d1=ds1['Gender'].value_counts()
2 plt.pie(d1,labels=['Female','Male','others'],autopct='%1.1f%%')
```

Out[54]:

```
(<matplotlib.patches.Wedge at 0x2867a4bb1c0>,
 <matplotlib.patches.Wedge at 0x2867a4bb0d0>,
 <matplotlib.patches.Wedge at 0x2867a4bbe50>],
 [Text(-0.7803692502900451, 0.7752572690415442, 'Female'),
 Text(0.7223314456896126, -0.8296006765715385, 'Male'),
 Text(1.0971265422611898, -0.07945659359679211, 'others')],
 [Text(-0.4256559547036609, 0.4228676012953877, '75.1%'),
 Text(0.39399897037615234, -0.4525094599481118, '22.6%'),
 Text(0.5984326594151944, -0.04333996014370478, '2.3%')])
```



2. which time slot is more preferable for the users?

In [55]:

```
1 ds['music_time_slot'].value_counts()
```

Out[55]:

```
Night          936
Afternoon      351
Morning        273
Name: music_time_slot, dtype: int64
```

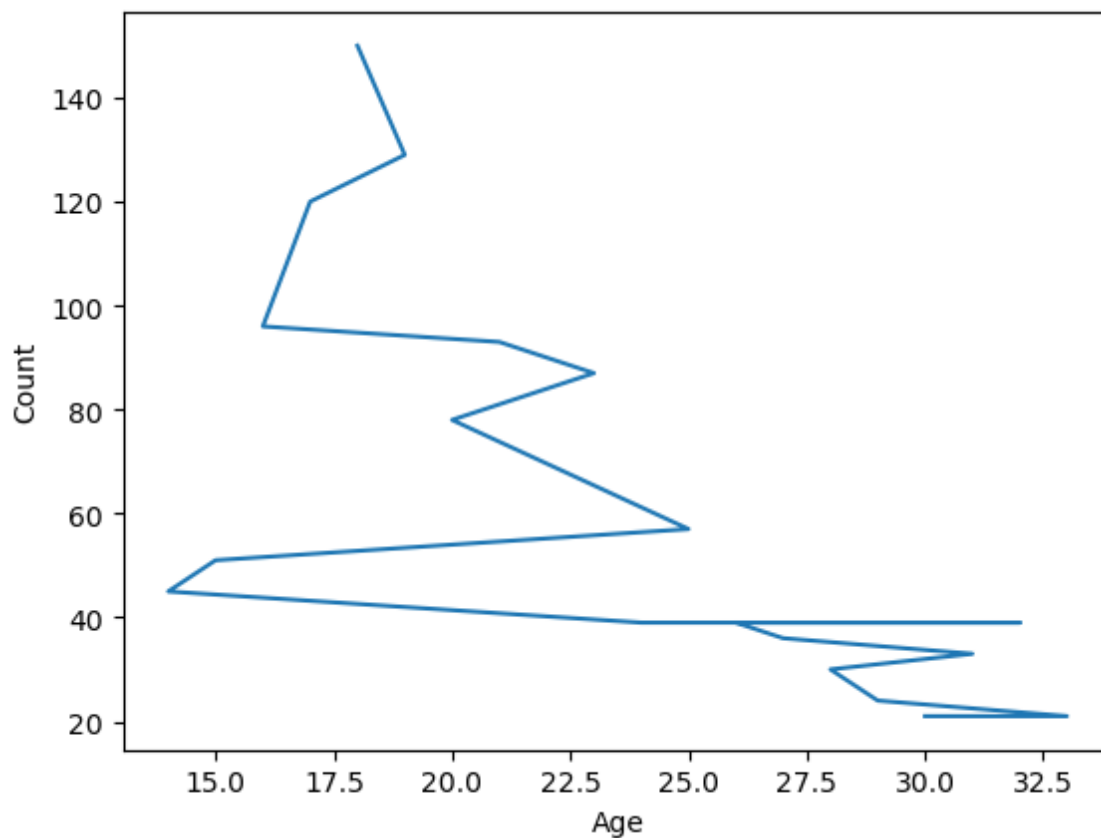
The users are preferred to listen the music in night times only

Here is the representation for the above query:

3. Age group distrubution in the users?

In [56]:

```
1 d2=ds1['Age'].value_counts().head(20).keys()
2 d3=ds1['Age'].value_counts().head(20)
3 y1=np.array(d2)
4 y2=np.array(d3)
5 plt.xlabel('Age')
6 plt.ylabel('Count')
7 plt.plot(y1,y2)
8 plt.show()
```



Most of the users in between the age groups 18-23 are there in the dataset

In [57]:

```
1 pip install wordcloud
```

Requirement already satisfied: wordcloud in c:\users\surya\anaconda3\lib\site-packages (1.9.2)Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: numpy>=1.6.1 in c:\users\surya\anaconda3\lib\site-packages (from wordcloud) (1.23.5)

Requirement already satisfied: pillow in c:\users\surya\anaconda3\lib\site-packages (from wordcloud) (9.4.0)

Requirement already satisfied: matplotlib in c:\users\surya\anaconda3\lib\site-packages (from wordcloud) (3.7.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)

Requirement already satisfied: cycler>=0.10 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)

Requirement already satisfied: packaging>=20.0 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (22.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.4)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.0.5)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\surya\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)

Requirement already satisfied: six>=1.5 in c:\users\surya\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)

4. In which device they are using the spotify?

In [58]:

```
1 ds1['spotify_listening_device'].value_counts()
```

Out[58]:

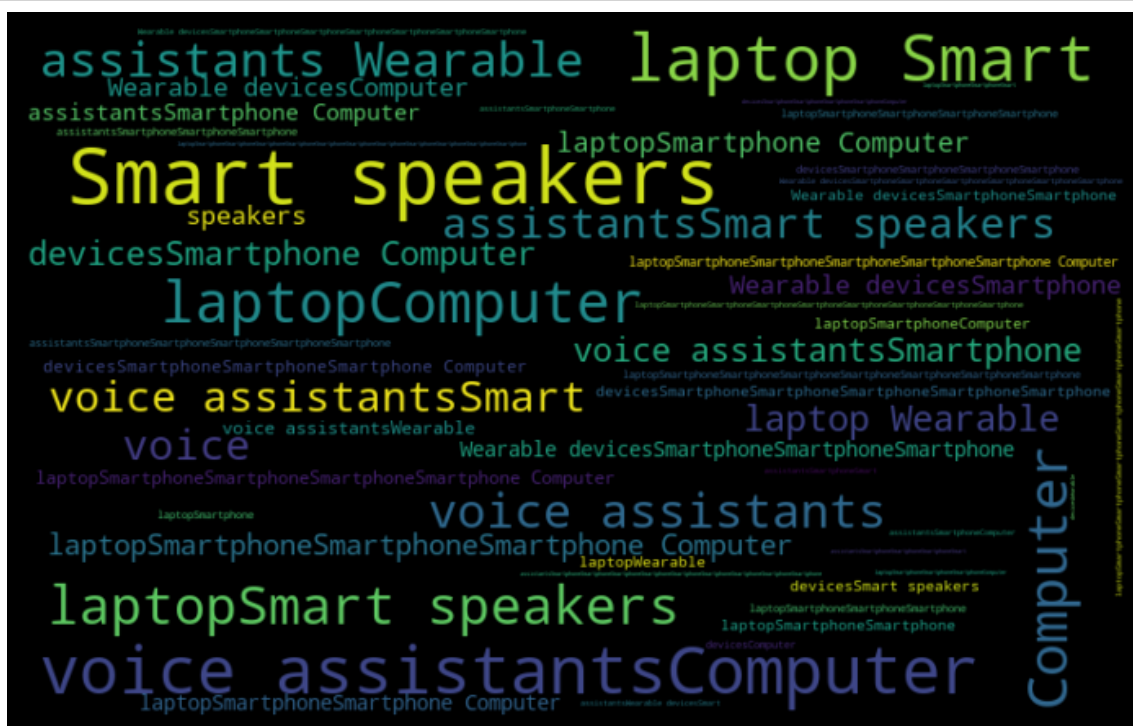
```
Smartphone
837
Smartphone, Computer or laptop
141
Computer or laptop
141
Smart speakers or voice assistants
111
Computer or laptop, Smart speakers or voice assistants
36
Smartphone, Smart speakers or voice assistants
27
Smartphone, Wearable devices
24
Smartphone, Computer or laptop, Smart speakers or voice assistants, Wearable devices
24
Smartphone, Computer or laptop, Smart speakers or voice assistants
18
Wearable devices
18
Smartphone, Computer or laptop, Wearable devices
15
Computer or laptop, Wearable devices
15
Smartphone, Smart speakers or voice assistants, Wearable devices
12
Computer or laptop, Smart speakers or voice assistants, Wearable devices
9
Smart speakers or voice assistants, Wearable devices
6
Name: spotify_listening_device, dtype: int64
```

In [59]:

```

1 from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
2 from PIL import Image
3
4 stopwords = set(STOPWORDS)
5 mask = np.array(Image.open('spotify.png'))
6
7 df1=pd.read_csv('Music_dataset.csv')
8
9 wordcloud = WordCloud(stopwords=stopwords, width=600, height=400,mask=mask, background_color='black')
10 plt.figure(figsize=(15,7), facecolor='k')
11 plt.imshow(wordcloud,interpolation = 'bilinear')
12 plt.axis('Off')
13 plt.tight_layout (pad=10)
14 plt.show()
15

```



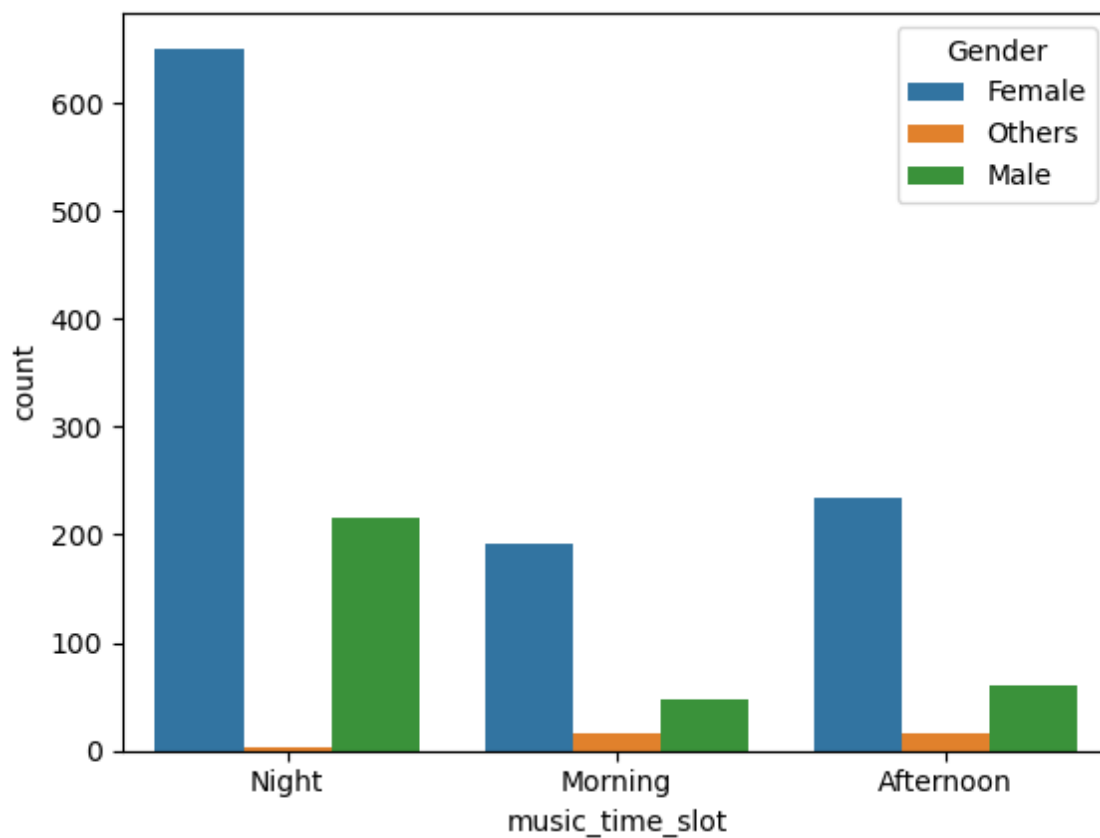
From this representation we can say that most of the users using computer and smart speakers to listen the music.

In [60]:

```
1 data=ds1
2 sns.countplot(data,x='music_time_slot',hue='Gender')
```

Out[60]:

<Axes: xlabel='music_time_slot', ylabel='count'>



In [61]:

```
1 ds1
```

Out[61]:

	Age	Gender	spotify_usage_period	Hours per day	spotify_listening_device	preferred_listening
0	18.0	Female	More than 2 years	3.0	Smart speakers or voice assistants	
2	18.0	Others	6 months to 1 year	4.0	Smart speakers or voice assistants	
4	18.0	Female	1 year to 2 years	4.0	Smartphone	
5	18.0	Male	More than 2 years	5.0	Smartphone	
6	18.0	Female	1 year to 2 years	3.0	Smartphone	
...
1551	43.0	Female	More than 2 years	1.0	Smartphone	
1555	22.0	Female	More than 2 years	2.0	Smartphone, Computer or laptop, Wearable devices	
1556	24.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	
1557	19.0	Female	More than 2 years	8.0	Smartphone	
1559	20.0	Female	More than 2 years	1.0	Smartphone, Computer or laptop	

1434 rows × 16 columns



5. Printing only required columns and performing head function.

In [62]:

```
1 ds1[["spotify_usage_period", "music_Influencial_mood", "frequently_lis_music"]].head
```

Out[62]:

	spotify_usage_period	music_Influencial_mood	frequently_lis_music
0	More than 2 years	Sadness or melancholy	leisure time
2	6 months to 1 year	Relaxation and stress relief	Study Hours, While Traveling
4	1 year to 2 years	Relaxation and stress relief	leisure time
5	More than 2 years	Uplifting and motivational	Workout session
6	1 year to 2 years	Relaxation and stress relief	Office hours
7	Less than 6 months	Social gatherings or parties	leisure time
8	Less than 6 months	Relaxation and stress relief, Uplifting and mo...	While Traveling, leisure time
9	More than 2 years	Relaxation and stress relief	Office hours, While Traveling, Workout session
10	More than 2 years	Relaxation and stress relief, Uplifting and mo...	While Traveling, leisure time
11	1 year to 2 years	Relaxation and stress relief, Uplifting and mo...	While Traveling, leisure time
13	More than 2 years	Relaxation and stress relief, Uplifting and mo...	Office hours, While Traveling
14	More than 2 years	Relaxation and stress relief	Office hours
15	More than 2 years	Relaxation and stress relief, Uplifting and mo...	Office hours, While Traveling, leisure time
16	Less than 6 months	Relaxation and stress relief	While Traveling, leisure time
17	Less than 6 months	Social gatherings or parties	Social gatherings

6. ratio blw podcast and music?

In [63]:

```
1 ds['preferred_listening_content'].value_counts()
```

Out[63]:

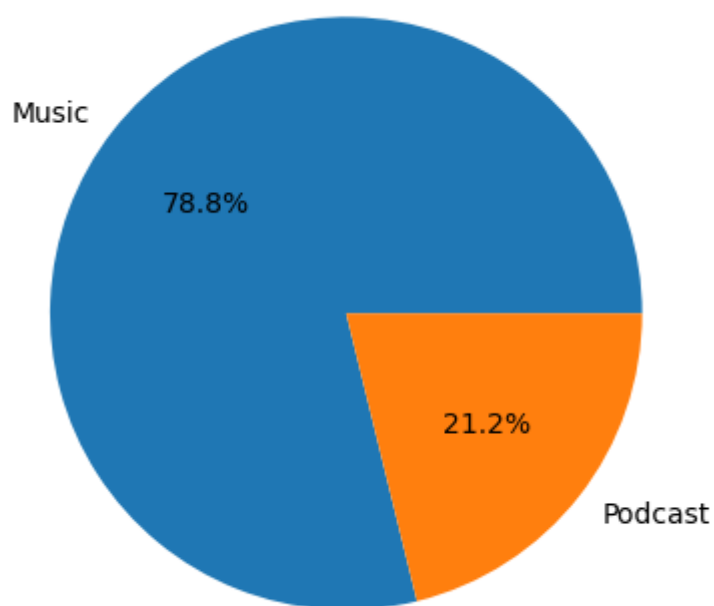
```
Music      1230
Podcast    330
Name: preferred_listening_content, dtype: int64
```

In [64]:

```
1 df2=ds['preferred_listening_content'].value_counts()
2 plt.pie(df2,labels=['Music','Podcast'],autopct='%1.1f%%')
```

Out[64]:

```
([<matplotlib.patches.Wedge at 0x2867acf9f90>,
 <matplotlib.patches.Wedge at 0x2867acf9ea0>],
 [Text(-0.8659018872996691, 0.6783906850560901, 'Music'),
 Text(0.8659019508151851, -0.6783906039845017, 'Podcast')],
 [Text(-0.472310120345274, 0.37003128275786734, '78.8%'),
 Text(0.47231015499010087, -0.3700312385370009, '21.2%')])
```



The pie chart shows that the ratio between music and podcasts is nearly 3:1

7. How many hours did the users use the spotify?

In [65]:

```
1 ds1['Hours per day'].value_counts()
```

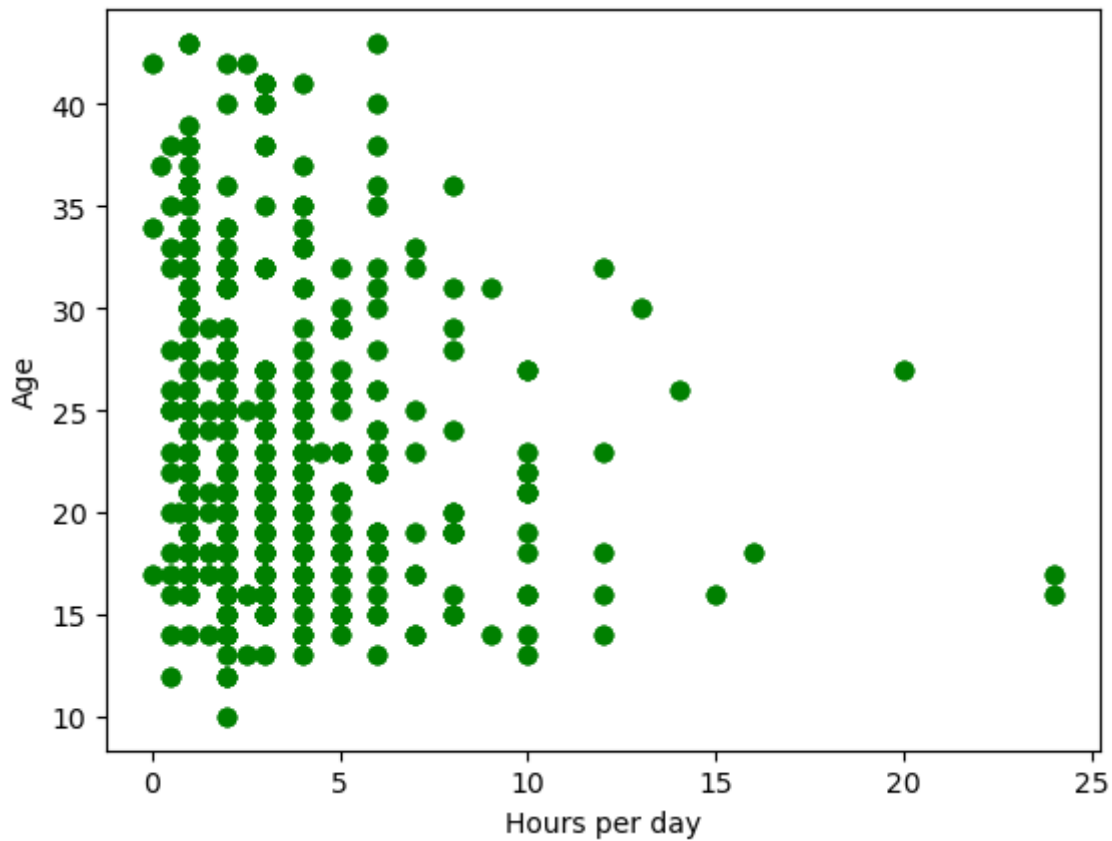
Out[65]:

2.00	348
3.00	243
1.00	204
4.00	159
5.00	114
6.00	96
0.50	48
8.00	45
10.00	42
1.50	33
7.00	27
2.50	15
12.00	15
0.00	9
24.00	6
9.00	6
0.25	3
0.70	3
4.50	3
20.00	3
16.00	3
15.00	3
14.00	3
13.00	3

Name: Hours per day, dtype: int64

In [66]:

```
1 x=ds1['Hours per day']
2 y=ds1['Age']
3 plt.ylabel('Age')
4 plt.xlabel('Hours per day')
5 plt.scatter(x,y,c='green')
6 plt.show()
```



8. pie chart for frequently_lis_music

In [67]:

```

1 l7 = ds1['frequently_lis_music'].value_counts().head(7).keys()
2 l8 = ds1['frequently_lis_music'].value_counts().head(7)
3 plt.pie(l8,labels=l7,autopct='%1.1f%%')

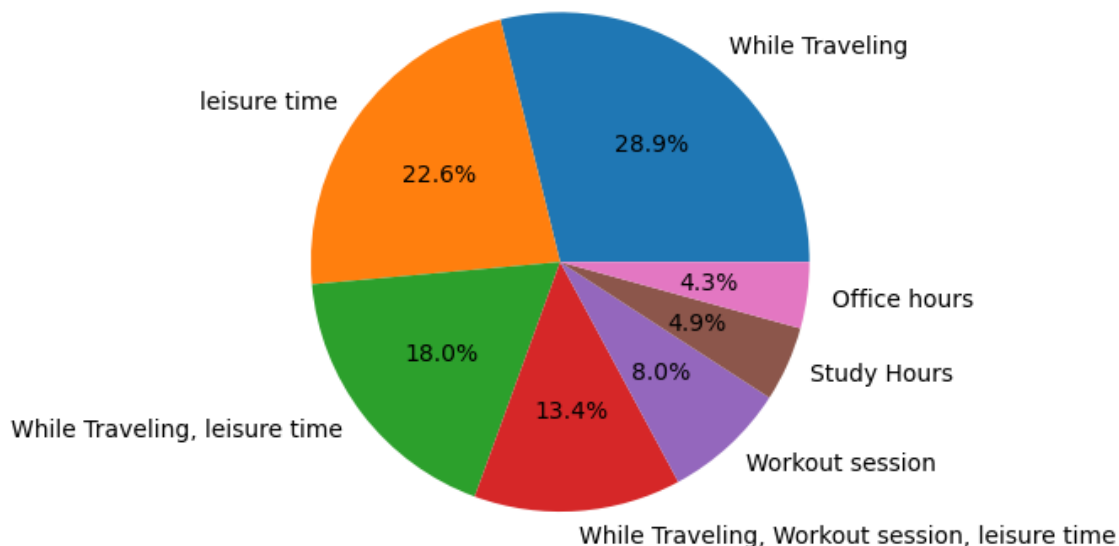
```

Out[67]:

```

([<matplotlib.patches.Wedge at 0x2867af0d120>,
 <matplotlib.patches.Wedge at 0x2867af0d030>,
 <matplotlib.patches.Wedge at 0x2867af0de10>,
 <matplotlib.patches.Wedge at 0x2867af0e4a0>,
 <matplotlib.patches.Wedge at 0x2867af0eb30>,
 <matplotlib.patches.Wedge at 0x2867af0f1c0>,
 <matplotlib.patches.Wedge at 0x2867af0f850>],
 [Text(0.6780918147992521, 0.8661359539369423, 'While Traveling'),
  Text(-0.8956862552973578, 0.638550023155114, 'leisure time'),
  Text(-0.8721875937808312, -0.6702900873911266, 'While Traveling, leisur
e time'),
  Text(0.07892070465195342, -1.0971652210935412, 'While Traveling, Workou
t session, leisure time'),
  Text(0.7457740232289247, -0.8085920518264716, 'Workout session'),
  Text(1.0035568428694164, -0.450414989903722, 'Study Hours'),
  Text(1.0900447339454662, -0.14765662192315615, 'Office hours')],
 [Text(0.3698682626177738, 0.47243779305651395, '28.9%'),
  Text(-0.4885561392531042, 0.3483000126300621, '22.6%'),
  Text(-0.4757386875168169, -0.36561277494061445, '18.0%'),
  Text(0.04304765708288368, -0.5984537569601134, '13.4%'),
  Text(0.40678583085214076, -0.44105021008716633, '8.0%'),
  Text(0.5473946415651361, -0.2456809035838483, '4.9%'),
  Text(0.5945698548793451, -0.08053997559444881, '4.3%')])

```



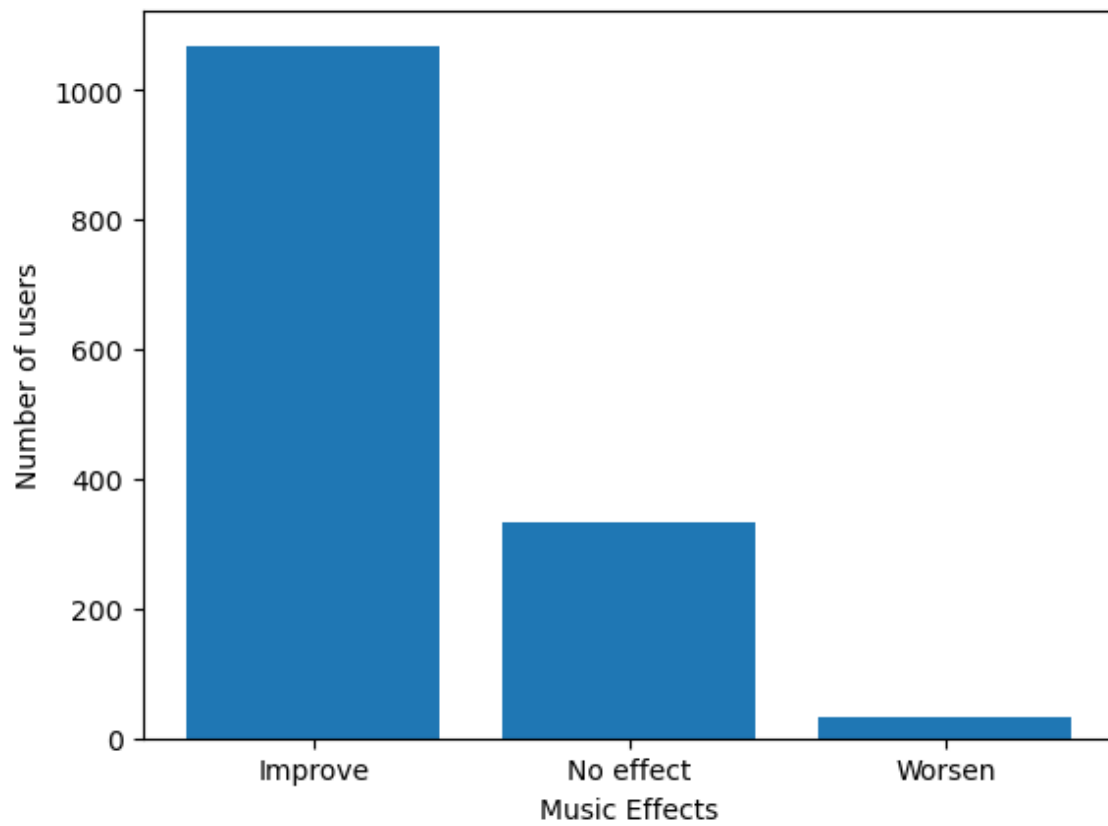
9. How many users asking for the improvement in music effects?

In [68]:

```
1 l9=ds1['Music effects'].value_counts().head().keys()
2 l10=ds1['Music effects'].value_counts().head()
3
4 plt.xlabel('Music Effects')
5 plt.ylabel('Number of users')
6 plt.bar(l9,l10)
```

Out[68]:

<BarContainer object of 3 artists>



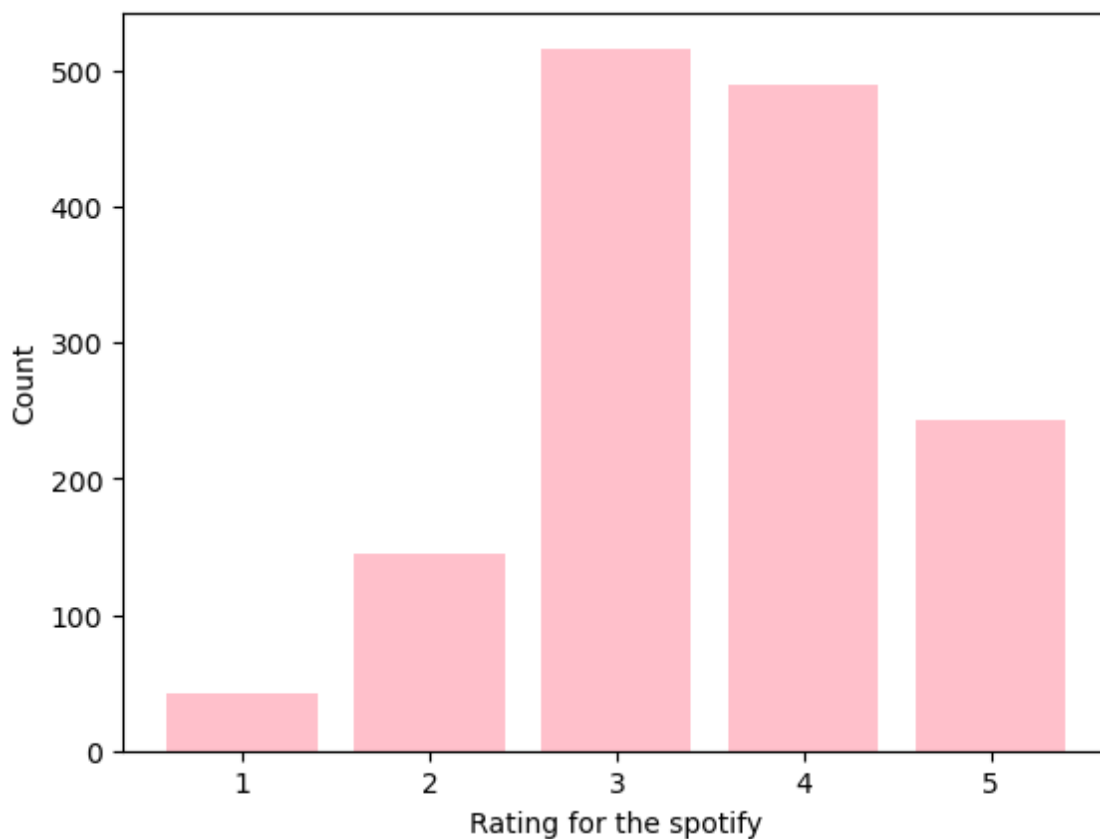
10. for rating of the spotify

In [69]:

```
1 l11=ds1['music_recc_rating'].value_counts().head().keys()
2 l12=ds1['music_recc_rating'].value_counts().head()
3
4 plt.xlabel('Rating for the spotify')
5 plt.ylabel('Count')
6 plt.bar(l11,l12,color='pink')
```

Out[69]:

<BarContainer object of 5 artists>



In []:

1