P9_2347107

September 20, 2023

1MCA-A PYTHON Programming(MCA171)

20-09-2023(Wednesday 09-11AM)

Regular lab Questions

P9: Demonstrate the use of "Matplotlib" modules to plot line and scatter plot

1. Import the domain dataset that you identified and use it as a data source for below mentioned questions.
2. Demonstrate the below-mentioned basic plots using 'matplotlib' library Histogram, Scator-plot, Line plot, Bargraph, Pie Chart, Area plot, Box plot, and pair plot.
3. Draw any two advanced graphs by referring to the mentioned link below. https://matplotlib.org/stable/gallery/index.html

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data=pd.read_csv("spotify.csv")
```

[ ]:
```
data.head(10)
```

[ ]:

	Unnamed: 0	acousticness	danceability	duration_ms	energy
0	0	0.01020	0.833	204600.0	0.434
1	1	0.19900	0.743	326933.0	0.359
2	2	0.03440	0.838	185707.0	0.412
3	3	0.60400	0.494	199413.0	0.338
4	4	0.18000	0.678	392893.0	0.561
5	5	0.00479	0.804	251333.0	0.560
6	6	0.01450	0.739	241400.0	0.472
7	7	0.02020	0.266	349667.0	0.348
8	8	0.04810	0.603	202853.0	0.944
9	9	0.00208	0.836	226840.0	0.603

	instrumentalness	key	liveness	loudness	mode	speechiness	tempo
0	0.021900	2	0.1650	-8.795	1	0.4310	150.062
1	0.006110	1	0.1370	-10.401	1	0.0794	160.083
2	0.000234	2	0.1590	-7.148	1	0.2890	75.044

```
3          0.510000    5    0.0922   -15.236     1       0.0261   86.468
4          0.512000    5    0.4390   -11.648     0       0.0694  174.004
5          0.000000    8    0.1640    -6.682     1       0.1850   85.023
6          0.000007    1    0.2070   -11.204     1       0.1560   80.030
7          0.664000   10    0.1600   -11.609     0       0.0371  144.154
8          0.000000   11    0.3420    -3.626     0       0.3470  130.035
9          0.000000    7    0.5710    -7.792     1       0.2370   99.994

   time_signature  valence  target         song_title            artist
0               4    0.286       1           Mask Off            Future
1               4    0.588       1            Redbone  Childish Gambino
2               4    0.173       1       Xanny Family            Future
3               4    0.230       1     Master Of None       Beach House
4               4    0.904       1     Parallel Lines       Junior Boys
5               4    0.264       1           Sneakin'             Drake
6               4    0.308       1        Childs Play             Drake
7               4    0.393       1    Gyöngyhajú lány             Omega
8               4    0.398       1  I've Seen Footage       Death Grips
9               4    0.386       1     Digital Animal       Honey Claws
```

[ ]:
```
data.isna().sum()
```

[ ]:
```
Unnamed: 0          0
acousticness        3
danceability        0
duration_ms         2
energy              1
instrumentalness    1
key                 0
liveness           13
loudness            0
mode                0
speechiness         1
tempo               0
time_signature      0
valence             2
target              0
song_title          0
artist              0
dtype: int64
```

[ ]:
```
data.describe()
```

[ ]:
```
         Unnamed: 0  acousticness  danceability   duration_ms       energy  \
count  2017.000000   2014.000000   2017.000000  2.015000e+03  2016.000000
mean   1008.000000      0.187695      0.618422  2.463637e+05     0.681857
std     582.402066      0.260169      0.161029  8.200214e+04     0.209949
```

```
min       0.000000      0.000003      0.122000  1.604200e+04     0.014800
25%     504.000000      0.009510      0.514000  2.000925e+05     0.563750
50%    1008.000000      0.062700      0.631000  2.293600e+05     0.715500
75%    1512.000000      0.265000      0.738000  2.703800e+05     0.846000
max    2016.000000      0.995000      0.984000  1.004627e+06     0.998000

       instrumentalness          key     liveness     loudness         mode  \
count       2016.000000  2017.000000  2004.000000  2017.000000  2017.000000
mean           0.133294     5.342588     0.191323    62.828327     0.612295
std            0.273230     3.648240     0.155843   201.495122     0.487347
min            0.000000     0.000000     0.018800   -33.097000     0.000000
25%            0.000000     2.000000     0.092175    -8.393000     0.000000
50%            0.000076     6.000000     0.127000    -6.248000     1.000000
75%            0.053925     9.000000     0.247000    -3.539000     1.000000
max            0.976000    11.000000     0.969000   996.000000     1.000000

       speechiness        tempo  time_signature      valence       target
count  2016.000000  2017.000000     2017.000000  2015.000000  2017.000000
mean      0.092652   121.567517        3.968270     0.497192     0.505702
std       0.089952    26.789328        0.255853     0.247028     0.500091
min       0.023100     8.000000        1.000000     0.034800     0.000000
25%       0.037500   100.189000        4.000000     0.296000     0.000000
50%       0.054900   121.427000        4.000000     0.493000     1.000000
75%       0.108000   137.849000        4.000000     0.691500     1.000000
max       0.816000   219.331000        5.000000     0.992000     1.000000
```

[ ]:
```
describe_df = data.describe().transpose()
describe_df
```

[ ]:
```
                   count           mean           std           min  \
Unnamed: 0        2017.0    1008.000000    582.402066      0.000000
acousticness      2014.0       0.187695      0.260169      0.000003
danceability      2017.0       0.618422      0.161029      0.122000
duration_ms       2015.0  246363.738958  82002.142160  16042.000000
energy            2016.0       0.681857      0.209949      0.014800
instrumentalness  2016.0       0.133294      0.273230      0.000000
key               2017.0       5.342588      3.648240      0.000000
liveness          2004.0       0.191323      0.155843      0.018800
loudness          2017.0      62.828327    201.495122    -33.097000
mode              2017.0       0.612295      0.487347      0.000000
speechiness       2016.0       0.092652      0.089952      0.023100
tempo             2017.0     121.567517     26.789328      8.000000
time_signature    2017.0       3.968270      0.255853      1.000000
valence           2015.0       0.497192      0.247028      0.034800
target            2017.0       0.505702      0.500091      0.000000

                        25%           50%           75%           max
```

```
Unnamed: 0           504.000000    1008.000000    1512.000000     2016.000
acousticness           0.009510       0.062700       0.265000        0.995
danceability           0.514000       0.631000       0.738000        0.984
duration_ms       200092.500000  229360.000000  270380.000000  1004627.000
energy                 0.563750       0.715500       0.846000        0.998
instrumentalness       0.000000       0.000076       0.053925        0.976
key                    2.000000       6.000000       9.000000       11.000
liveness               0.092175       0.127000       0.247000        0.969
loudness              -8.393000      -6.248000      -3.539000      996.000
mode                   0.000000       1.000000       1.000000        1.000
speechiness            0.037500       0.054900       0.108000        0.816
tempo                100.189000     121.427000     137.849000      219.331
time_signature         4.000000       4.000000       4.000000        5.000
valence                0.296000       0.493000       0.691500        0.992
target                 0.000000       1.000000       1.000000        1.000
```

[ ]:
```
# for column in describe_df.columns:
#     plt.plot(describe_df.index, describe_df[column], marker='o', label=column)

# plt.title('Summary Statistics Comparison')
# plt.xlabel('Statistics')
# plt.ylabel('Value')
# plt.xticks(rotation=45)
# plt.legend()

# plt.tight_layout()
# plt.grid(True)
# plt.show()
```

[ ]:
```
data.shape
```

[ ]:
```
(2017, 17)
```

[ ]:
```
data
```

[ ]:
```
      Unnamed: 0  acousticness  danceability  duration_ms  energy  \
0              0       0.01020         0.833     204600.0   0.434
1              1       0.19900         0.743     326933.0   0.359
2              2       0.03440         0.838     185707.0   0.412
3              3       0.60400         0.494     199413.0   0.338
4              4       0.18000         0.678     392893.0   0.561
…            …             …             …          …      …
2012        2012       0.00106         0.584     274404.0   0.932
2013        2013       0.08770         0.894     182182.0   0.892
2014        2014       0.00857         0.637     207200.0   0.935
2015        2015       0.00164         0.557     185600.0   0.992
2016        2016       0.00281         0.446     204520.0   0.915
```

```
      instrumentalness  key  liveness  loudness  mode  speechiness    tempo  \
0             0.021900    2    0.1650    -8.795     1       0.4310  150.062
1             0.006110    1    0.1370   -10.401     1       0.0794  160.083
2             0.000234    2    0.1590    -7.148     1       0.2890   75.044
3             0.510000    5    0.0922   -15.236     1       0.0261   86.468
4             0.512000    5    0.4390   -11.648     0       0.0694  174.004
…                  …  …       …        …          …       …
2012          0.002690    1    0.1290    -3.501     1       0.3330   74.976
2013          0.001670    1    0.0528    -2.663     1       0.1310  110.041
2014          0.003990    0    0.2140    -2.467     1       0.1070  150.082
2015          0.677000    1    0.0913    -2.735     1       0.1330  150.011
2016          0.000039    9    0.2180    -6.221     1       0.1410  190.013

      time_signature  valence  target                          song_title  \
0                  4    0.286       1                           Mask Off
1                  4    0.588       1                            Redbone
2                  4    0.173       1                       Xanny Family
3                  4    0.230       1                     Master Of None
4                  4    0.904       1                     Parallel Lines
…                  …       …      …                                  …
2012               4    0.211       0   Like A Bitch - Kill The Noise Remix
2013               4    0.867       0                              Candy
2014               4    0.470       0  Habit - Dack Janiels & Wenzday Remix
2015               4    0.623       0                     First Contact
2016               4    0.402       0              I Wanna Get Better

                artist
0               Future
1     Childish Gambino
2               Future
3          Beach House
4          Junior Boys
…                  …
2012    Kill The Noise
2013    Dillon Francis
2014         Rain Man
2015        Twin Moons
2016         Bleachers

[2017 rows x 17 columns]
```

[ ]:
```
acousticness=data['acousticness'].mean()
duration_ms=data['duration_ms'].mean()
energy=data['energy'].mean()
instrumentalness=data['instrumentalness'].mean()
liveness=data['liveness'].mean()
```

```
speechiness=data['speechiness'].mean()
valence=data['valence'].mean()
```

[ ]:
```
data['acousticness'].replace(np.nan,acousticness, inplace=True)
data['duration_ms'].replace(np.nan,duration_ms, inplace=True)
data['energy'].replace(np.nan,energy, inplace=True)
data['instrumentalness'].replace(np.nan,instrumentalness, inplace=True)
data['liveness'].replace(np.nan,liveness, inplace=True)
data['speechiness'].replace(np.nan,speechiness, inplace=True)
data['valence'].replace(np.nan,valence, inplace=True)

data.isna().sum()
```

[ ]:
```
Unnamed: 0          0
acousticness        0
danceability        0
duration_ms         0
energy              0
instrumentalness    0
key                 0
liveness            0
loudness            0
mode                0
speechiness         0
tempo               0
time_signature      0
valence             0
target              0
song_title          0
artist              0
dtype: int64
```

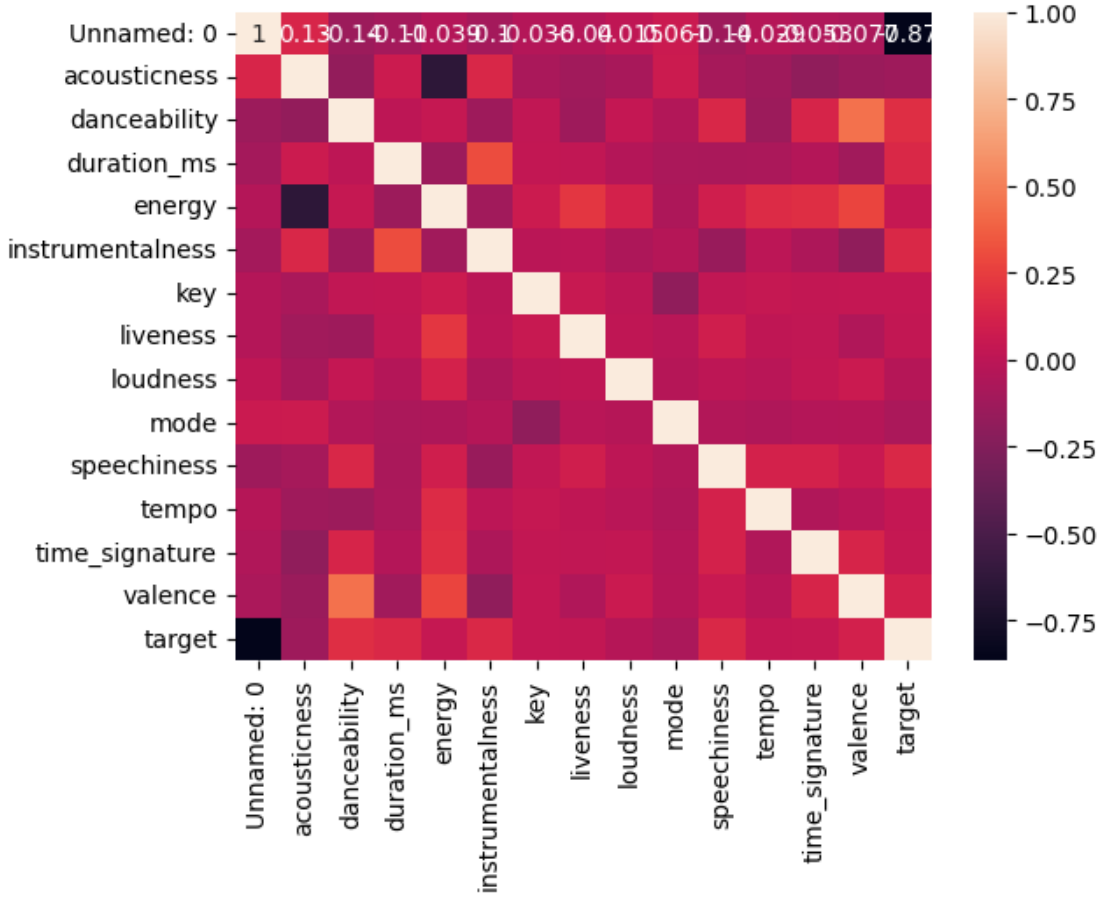
Null values removed using mean()

[ ]:
```
numeric_columns = data.select_dtypes(include=['number'])
correlation = numeric_columns.corr()
sns.heatmap(correlation, annot=True)
print(correlation)
```

```
                  Unnamed: 0  acousticness  danceability  duration_ms  \
Unnamed: 0          1.000000      0.133461     -0.139615    -0.105401
acousticness        0.133461      1.000000     -0.179780     0.071055
danceability       -0.139615     -0.179780      1.000000     0.004797
duration_ms        -0.105401      0.071055      0.004797     1.000000
energy             -0.038723     -0.644036      0.038216    -0.143670
instrumentalness   -0.104660      0.144202     -0.130578     0.305277
key                -0.035611     -0.081228      0.022325     0.023703
liveness           -0.040156     -0.117363     -0.133636     0.022506
```

```
loudness            0.014971     -0.088497      0.035409    -0.037555
mode                0.060639      0.068077     -0.047867    -0.074266
speechiness        -0.135215     -0.100090      0.142557    -0.079037
tempo              -0.028612     -0.126797     -0.141860    -0.077044
time_signature     -0.053235     -0.190313      0.128909    -0.040080
valence            -0.077349     -0.150535      0.441072    -0.121179
target             -0.865969     -0.130043      0.176706     0.147480

                     energy  instrumentalness       key  liveness  loudness  \
Unnamed: 0       -0.038723         -0.104660 -0.035611 -0.040156  0.014971
acousticness     -0.644036          0.144202 -0.081228 -0.117363 -0.088497
danceability      0.038216         -0.130578  0.022325 -0.133636  0.035409
duration_ms      -0.143670          0.305277  0.023703  0.022506 -0.037555
energy            1.000000         -0.119618  0.072891  0.216953  0.116680
instrumentalness -0.119618          1.000000 -0.009594  0.000115 -0.070483
key               0.072891         -0.009594  1.000000  0.049811  0.002329
liveness          0.216953          0.000115  0.049811  1.000000  0.011080
loudness          0.116680         -0.070483  0.002329  0.011080  1.000000
mode             -0.069327         -0.030136 -0.195040 -0.017607 -0.030716
speechiness       0.092557         -0.155291  0.019056  0.092132  0.002597
tempo             0.161915          0.001055  0.040271  0.010834 -0.014222
time_signature    0.183389         -0.070717  0.024937  0.017667  0.025140
valence           0.276094         -0.195200  0.030704 -0.050177  0.060755
target            0.038409          0.152623  0.033594  0.025636 -0.027781

                        mode  speechiness     tempo  time_signature   valence  \
Unnamed: 0          0.060639    -0.135215 -0.028612       -0.053235 -0.077349
acousticness        0.068077    -0.100090 -0.126797       -0.190313 -0.150535
danceability       -0.047867     0.142557 -0.141860        0.128909  0.441072
duration_ms        -0.074266    -0.079037 -0.077044       -0.040080 -0.121179
energy             -0.069327     0.092557  0.161915        0.183389  0.276094
instrumentalness   -0.030136    -0.155291  0.001055       -0.070717 -0.195200
key                -0.195040     0.019056  0.040271        0.024937  0.030704
liveness           -0.017607     0.092132  0.010834        0.017667 -0.050177
loudness           -0.030716     0.002597 -0.014222        0.025140  0.060755
mode                1.000000    -0.049506 -0.060825       -0.039038 -0.031703
speechiness        -0.049506     1.000000  0.111603        0.113514  0.053356
tempo              -0.060825     0.111603  1.000000       -0.055867 -0.013620
time_signature     -0.039038     0.113514 -0.055867        1.000000  0.132085
valence            -0.031703     0.053356 -0.013620        0.132085  1.000000
target             -0.072336     0.153876  0.035948        0.040182  0.106512

                     target
Unnamed: 0       -0.865969
acousticness     -0.130043
danceability      0.176706
duration_ms       0.147480
energy            0.038409
```

```
instrumentalness  0.152623
key               0.033594
liveness          0.025636
loudness         -0.027781
mode             -0.072336
speechiness       0.153876
tempo             0.035948
time_signature    0.040182
valence           0.106512
target            1.000000
```

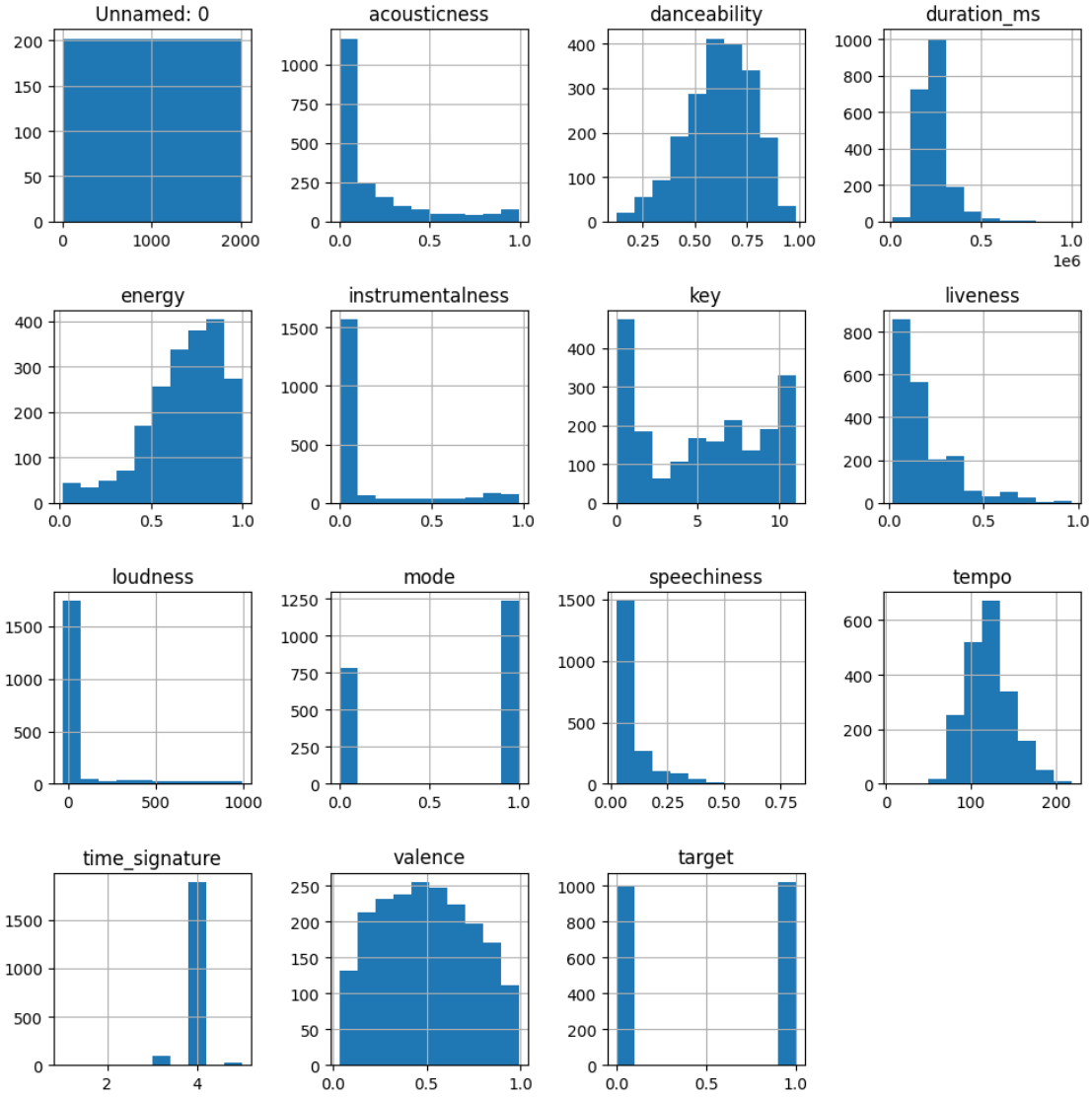


[ ]:
```
threshold = 0.7

high_correlation_columns = set()
for i in range(len(correlation.columns)):
    for j in range(i):
        if abs(correlation.iloc[i, j]) > threshold:
            col1 = correlation.columns[i]
            col2 = correlation.columns[j]
```

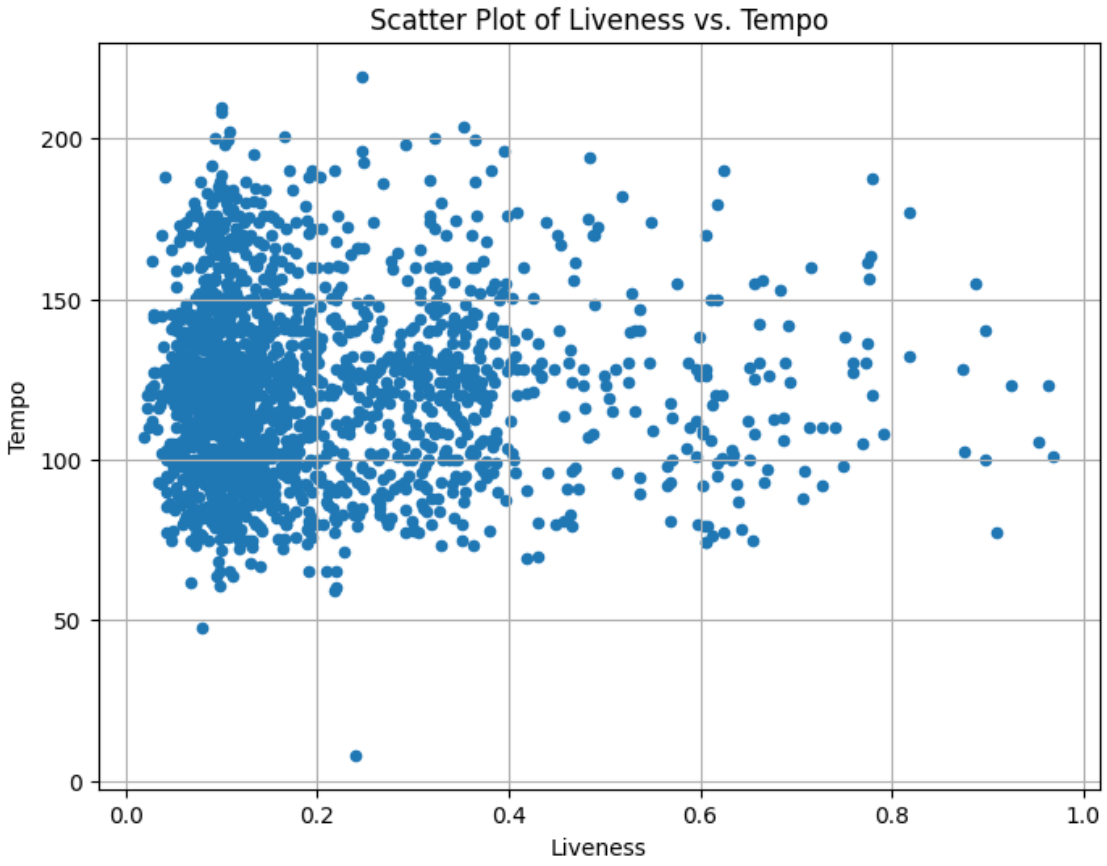
```
            high_correlation_columns.add(col1)
            high_correlation_columns.add(col2)

# Print the columns with high correlation
print("Columns with high correlation:")
print(high_correlation_columns)
```

```
Columns with high correlation:
{'target', 'Unnamed: 0'}
```

[ ]:
```
data.hist(figsize=(10, 10))
plt.tight_layout()
plt.show()
```



[ ]:
```
data.plot.scatter(x='liveness', y='tempo', figsize=(8, 6))
plt.title('Scatter Plot of Liveness vs. Tempo')
plt.xlabel('Liveness')
plt.ylabel('Tempo')
plt.grid(True)
plt.show()
```



[ ]:
```
print(data.groupby('mode').size())
print(data['mode'].unique())
mode0=data[data['mode']=='1'].copy()
mode1=data[data['mode']=='1'].copy()
```

```
mode
0     782
1    1235
dtype: int64
[1 0]
```

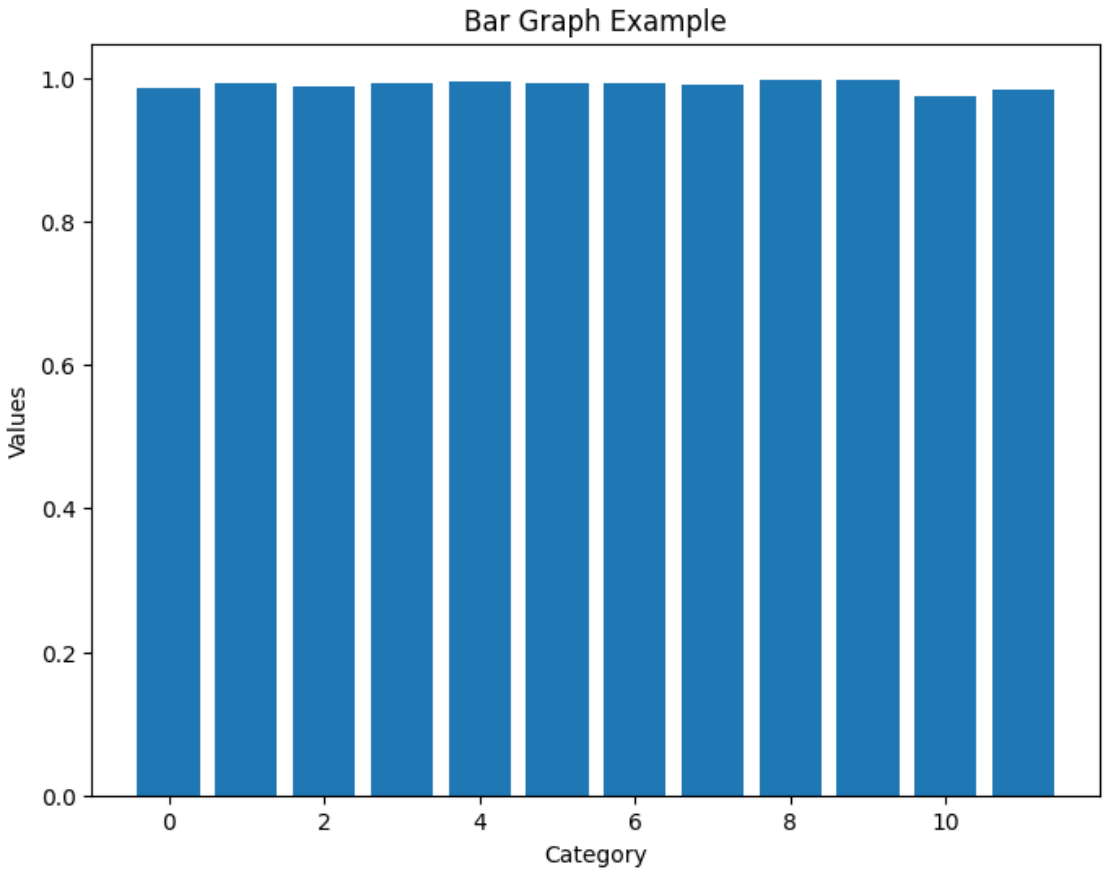
[ ]:
```
mode0
```

[ ]:
```
Empty DataFrame
Columns: [Unnamed: 0, acousticness, danceability, duration_ms, energy,
instrumentalness, key, liveness, loudness, mode, speechiness, tempo,
time_signature, valence, target, song_title, artist]
Index: []
```

[ ]:
```
mode1
```

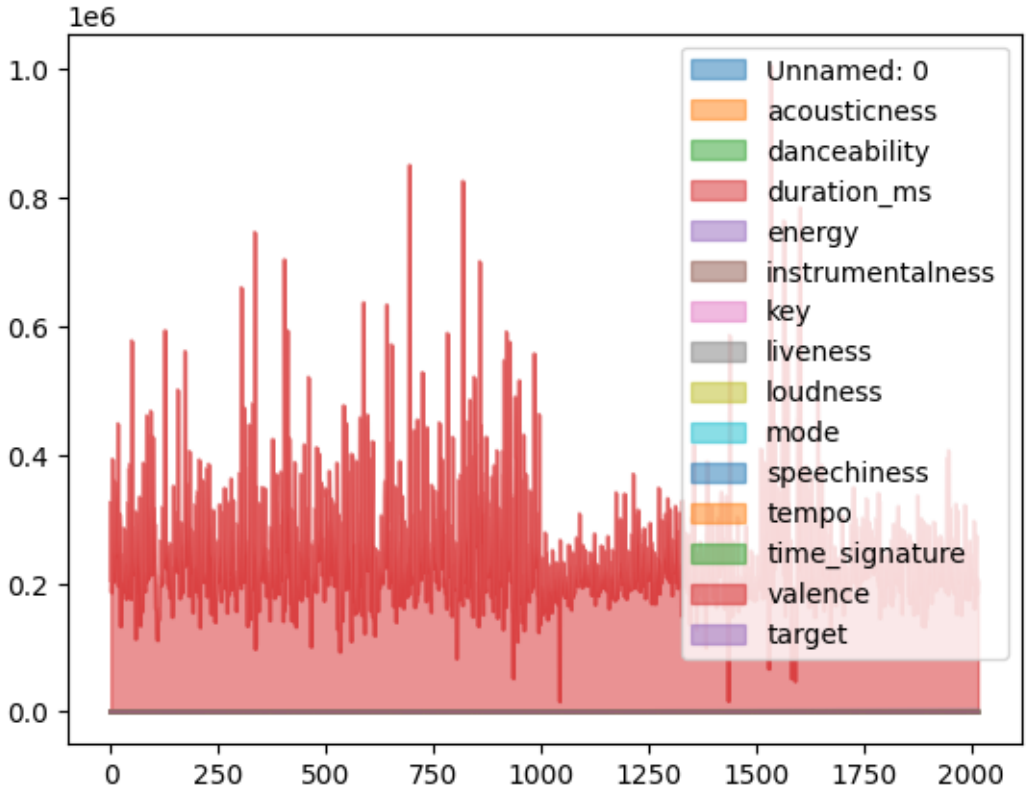
[ ]:
```
Empty DataFrame
Columns: [Unnamed: 0, acousticness, danceability, duration_ms, energy,
instrumentalness, key, liveness, loudness, mode, speechiness, tempo,
time_signature, valence, target, song_title, artist]
Index: []
```

[ ]:
```
plt.figure(figsize=(8, 6))
plt.bar(data['key'], data['energy'])
plt.xlabel('Category')
plt.ylabel('Values')
plt.title('Bar Graph Example')
plt.show()
```

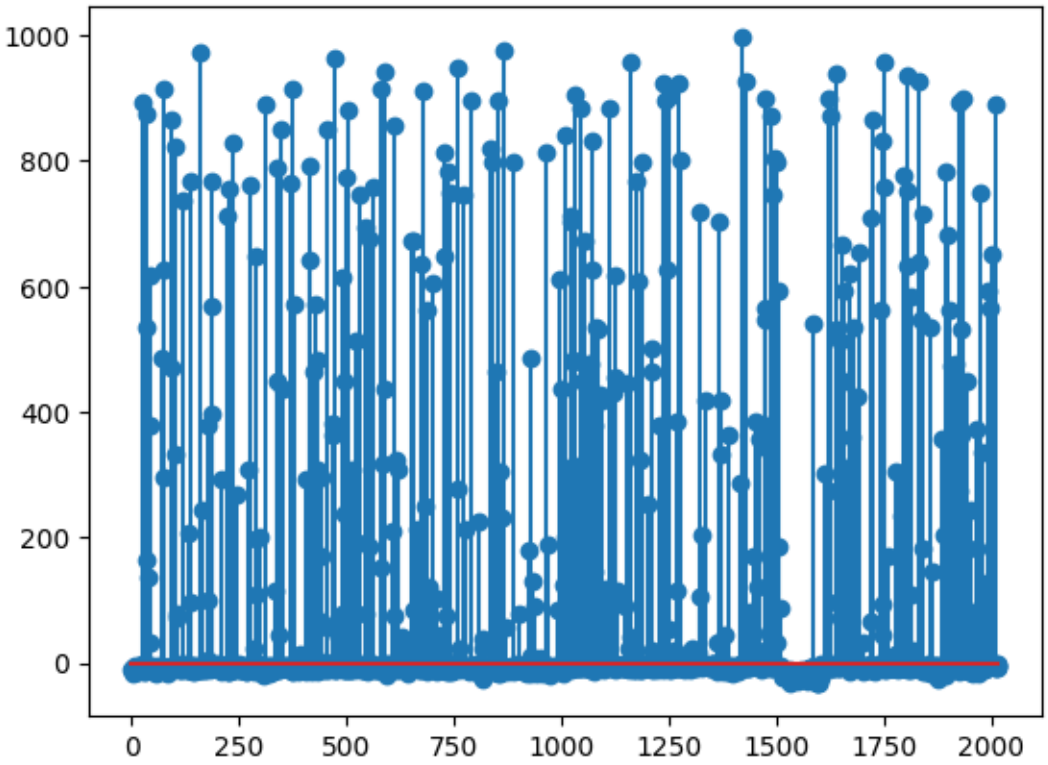


[ ]:
```
data.plot.area(stacked=False)
```

[ ]: <Axes: >



[ ]:
```
plt.stem(data['Unnamed: 0'], data['loudness'])
plt.show()
```



[ ]:
```
colors = ['lightcoral', 'lightblue', 'lightgreen', 'lightsalmon']
plt.figure(figsize=(8, 8))
plt.pie(data['key'], colors=colors, autopct='%1.1f%%', shadow=True,␣
  ↪startangle=140)
plt.title('Pie Plot Example')
plt.show()
```

Pie Plot Example