P8_2347107

September 15, 2023

For each output, give the interpretation with respect to the imported dataset.

2. Read the csv file and create and understand the data frame using describe(), shape, info().
3. Find if any missing values (null values) are in the data, handle all the rows with missing data in four different ways (delete, replace, fill, bfill), and print the data frame.
4. Filter based on any column using groupby().
5. Select 20 samples randomly and Create a data frame with Hirariclе Index

write a python code for it with a dataset on spotify songs

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

[ ]:
```
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.shape
```

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

[ ]:
```
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
```

[ ]:
```
data.columns[data.isna().any()].tolist()
```

[ ]:
```
['acousticness',
 'duration_ms',
 'energy',
 'instrumentalness',
 'liveness',
 'speechiness',
 'valence']
```

[ ]:
```
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
```

[ ]:
```
delData=data.dropna()

delData.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
```

[ ]:
```
RepData=data
RepData.head()
```

[ ]:

	Unnamed: 0	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence	target	song_title	artist
0	0	0.0102	0.833	204600.0	0.434	0.021900	2	0.1650	-8.795	1	0.4310	150.062	4	0.286	1	Mask Off	Future
1	1	0.1990	0.743	326933.0	0.359	0.006110	1	0.1370	-10.401	1	0.0794	160.083	4	0.588	1	Redbone	Childish Gambino
2	2	0.0344	0.838	185707.0	0.412	0.000234	2	0.1590	-7.148	1	0.2890	75.044	4	0.173	1	Xanny Family	Future
3	3	0.6040	0.494	199413.0	0.338	0.510000	5	0.0922	-15.236	1	0.0261	86.468	4	0.230	1	Master Of None	Beach House
4	4	0.1800	0.678	392893.0	0.561	0.512000	5	0.4390	-11.648	0	0.0694	174.004	4	0.904	1	Parallel Lines	Junior Boys

[ ]:
```
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()
```

[ ]:
```
RepData['acousticness'].replace(np.nan,acousticness, inplace=True)
RepData['duration_ms'].replace(np.nan,duration_ms, inplace=True)
RepData['energy'].replace(np.nan,energy, inplace=True)
RepData['instrumentalness'].replace(np.nan,instrumentalness, inplace=True)
```

```
RepData['liveness'].replace(np.nan,liveness, inplace=True)
RepData['speechiness'].replace(np.nan,speechiness, inplace=True)
RepData['valence'].replace(np.nan,valence, inplace=True)

RepData.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
```

[ ]:
```
FillData=data
```

[ ]:
```
FillData['acousticness'].fillna(acousticness, inplace=True)
FillData['duration_ms'].fillna(duration_ms, inplace=True)
FillData['energy'].fillna(energy, inplace=True)
FillData['instrumentalness'].fillna(instrumentalness, inplace=True)
FillData['liveness'].fillna(liveness, inplace=True)
FillData['speechiness'].fillna(speechiness, inplace=True)
FillData['valence'].fillna(valence, inplace=True)

FillData.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
```

[ ]:
```
Bfill=data
Bfill.bfill(axis ='columns')
```

[ ]:
```
      Unnamed: 0 acousticness danceability duration_ms energy instrumentalness  \
0              0       0.0102        0.833    204600.0  0.434           0.0219
1              1        0.199        0.743    326933.0  0.359          0.00611
2              2       0.0344        0.838    185707.0  0.412         0.000234
3              3        0.604        0.494    199413.0  0.338             0.51
4              4         0.18        0.678    392893.0  0.561            0.512
…            …            …            …          …      …                …
2012        2012      0.00106        0.584    274404.0  0.932          0.00269
2013        2013       0.0877        0.894    182182.0  0.892          0.00167
2014        2014      0.00857        0.637    207200.0  0.935          0.00399
2015        2015      0.00164        0.557    185600.0  0.992            0.677
2016        2016      0.00281        0.446    204520.0  0.915         0.000039

      key liveness loudness mode speechiness    tempo time_signature valence  \
0       2    0.165   -8.795    1       0.431  150.062              4   0.286
1       1    0.137  -10.401    1      0.0794  160.083              4   0.588
2       2    0.159   -7.148    1       0.289   75.044              4   0.173
3       5   0.0922  -15.236    1      0.0261   86.468              4    0.23
4       5    0.439  -11.648    0      0.0694  174.004              4   0.904
…    ..      …       …  …          …       …              …      …
2012    1    0.129   -3.501    1       0.333   74.976              4   0.211
2013    1   0.0528   -2.663    1       0.131  110.041              4   0.867
2014    0    0.214   -2.467    1       0.107  150.082              4    0.47
2015    1   0.0913   -2.735    1       0.133  150.011              4   0.623
2016    9    0.218   -6.221    1       0.141  190.013              4   0.402

      target                                 song_title            artist
0          1                                   Mask Off            Future
1          1                                    Redbone  Childish Gambino
2          1                                 Xanny Family            Future
3          1                             Master Of None       Beach House
4          1                             Parallel Lines       Junior Boys
…        …                                            …                 …
2012       0   Like A Bitch - Kill The Noise Remix    Kill The Noise
2013       0                                  Candy    Dillon Francis
```

```
2014      0  Habit - Dack Janiels & Wenzday Remix          Rain Man
2015      0                         First Contact        Twin Moons
2016      0                   I Wanna Get Better         Bleachers

[2017 rows x 17 columns]
```

[ ]:
```
Bfill.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
```

[ ]:
```
Group=data.groupby('key')
Group.first()
```

[ ]:
```
     Unnamed: 0  acousticness  danceability    duration_ms  energy  \
key
0            12       0.25300         0.603  356973.000000   0.434
1             1       0.19900         0.743  326933.000000   0.359
2             0       0.01020         0.833  204600.000000   0.434
3            49       0.18300         0.716  576888.000000   0.957
4            17       0.23300         0.789  447907.000000   0.659
5             3       0.60400         0.494  199413.000000   0.338
6            15       0.01900         0.637  246363.738958   0.832
7             9       0.00208         0.836  226840.000000   0.603
8             5       0.00479         0.804  251333.000000   0.560
9            14       0.44000         0.662  247288.000000   0.603
10            7       0.02020         0.266  349667.000000   0.348
11            8       0.04810         0.603  202853.000000   0.944

     instrumentalness  liveness  loudness  mode  speechiness    tempo  \
key
```

```
0             0.06190    0.1080   -11.062     1       0.0342  127.681
1             0.00611    0.1370   -10.401     1       0.0794  160.083
2             0.02190    0.1650    -8.795     1       0.4310  150.062
3             0.05800    0.1820    -4.814     1       0.0689  106.714
4             0.00049    0.1840   -12.654     0       0.0429  122.415
5             0.51000    0.0922   -15.236     1       0.0261   86.468
6             0.05630    0.3160    -6.637     1       0.1630   99.988
7             0.00000    0.5710    -7.792     1       0.2370   99.994
8             0.00000    0.1640    -6.682     1       0.1850   85.023
9             0.00000    0.0972    -8.317     0       0.0793  125.011
10            0.66400    0.1600   -11.609     0       0.0371  144.154
11            0.00000    0.3420    -3.626     0       0.3470  130.035

     time_signature  valence  target                song_title  \
key
0                 4    0.381       1                  Cemalim
1                 4    0.588       1                  Redbone
2                 4    0.286       1                Mask Off
3                 4    0.930       1    Odofo Nyi Akyiri Biara
4                 4    0.842       1  One Nation Under a Groove
5                 4    0.230       1            Master Of None
6                 4    0.317       1                      Char
7                 4    0.386       1            Digital Animal
8                 4    0.264       1                 Sneakin'
9                 4    0.351       1                  Oh lala
10                4    0.393       1           Gyöngyhajú lány
11                4    0.398       1         I've Seen Footage

                artist
key
0         Erkin Koray
1    Childish Gambino
2              Future
3          Ebo Taylor
4          Funkadelic
5         Beach House
6     Crystal Castles
7         Honey Claws
8               Drake
9                 PNL
10              Omega
11        Death Grips
```

[ ]:
```
Mode=data.groupby('key')
Mode.first()
```

[ ]:

key	Unnamed: 0	acousticness	danceability	duration_ms	energy	instrumentalness	liveness	loudness	mode	speechiness	tempo	time_signature	valence	target	song_title	artist
0	12	0.25300	0.603	356973.000000	0.434	0.06190	0.1080	-11.062	1	0.0342	127.681	4	0.381	1	Cemalim	
1	1	0.19900	0.743	326933.000000	0.359	0.00611	0.1370	-10.401	1	0.0794	160.083	4	0.588	1	Redbone	
2	0	0.01020	0.833	204600.000000	0.434	0.02190	0.1650	-8.795	1	0.4310	150.062	4	0.286	1	Mask Off	
3	49	0.18300	0.716	576888.000000	0.957	0.05800	0.1820	-4.814	1	0.0689	106.714	4	0.930	1	Odofo Nyi Akyiri Biara	
4	17	0.23300	0.789	447907.000000	0.659	0.00049	0.1840	-12.654	0	0.0429	122.415	4	0.842	1	One Nation Under a Groove	
5	3	0.60400	0.494	199413.000000	0.338	0.51000	0.0922	-15.236	1	0.0261	86.468	4	0.230	1	Master Of None	
6	15	0.01900	0.637	246363.738958	0.832	0.05630	0.3160	-6.637	1	0.1630	99.988	4	0.317	1	Char	
7	9	0.00208	0.836	226840.000000	0.603	0.00000	0.5710	-7.792	1	0.2370	99.994	4	0.386	1	Digital Animal	
8	5	0.00479	0.804	251333.000000	0.560	0.00000	0.1640	-6.682	1	0.1850	85.023	4	0.264	1	Sneakin’	
9	14	0.44000	0.662	247288.000000	0.603	0.00000	0.0972	-8.317	0	0.0793	125.011	4	0.351	1	Oh lala	
10	7	0.02020	0.266	349667.000000	0.348	0.66400	0.1600	-11.609	0	0.0371	144.154	4	0.393	1	Gyöngyhajú lány	
11	8	0.04810	0.603	202853.000000	0.944	0.00000	0.3420	-3.626	0	0.3470	130.035	4	0.398	1	I've Seen Footage	

```
0          Erkin Koray
1     Childish Gambino
2               Future
3           Ebo Taylor
4           Funkadelic
5          Beach House
6      Crystal Castles
7          Honey Claws
8                Drake
9                  PNL
10               Omega
11         Death Grips
```

[ ]:
```
random_samples = data.sample(n=20, random_state=42)

index = pd.MultiIndex.from_tuples([(f'Sample {i}',) for i in range(1, 21)],␣
↪names=['Sample'])

new_df = pd.DataFrame(random_samples.values, columns=random_samples.columns,␣
↪index=index)

# print(new_df)
regular_df = new_df.reset_index()
regular_df
```

[ ]:

	Sample	Unnamed: 0	acousticness	danceability	duration_ms	energy	\
0	Sample 1	1555	0.748	0.52	341667.0	0.0748	
1	Sample 2	526	0.0726	0.739	386907.0	0.526	
2	Sample 3	393	0.0021	0.646	219754.0	0.892	
3	Sample 4	1788	0.789	0.664	145707.0	0.32	
4	Sample 5	433	0.0198	0.517	245013.0	0.491	
5	Sample 6	1159	0.281	0.718	214867.0	0.609	
6	Sample 7	1090	0.27	0.639	307910.0	0.869	
7	Sample 8	429	0.0571	0.502	287827.0	0.632	
8	Sample 9	1801	0.0637	0.406	224848.0	0.638	
9	Sample 10	530	0.000078	0.351	141305.0	0.931	
10	Sample 11	1208	0.0762	0.809	195853.0	0.628	
11	Sample 12	1454	0.316	0.519	182720.0	0.829	
12	Sample 13	1910	0.0028	0.686	211000.0	0.915	
13	Sample 14	1619	0.00338	0.569	190155.0	0.896	
14	Sample 15	1022	0.013	0.631	187813.0	0.716	
15	Sample 16	678	0.0152	0.621	335156.0	0.813	
16	Sample 17	1477	0.0533	0.599	288877.0	0.667	
17	Sample 18	124	0.674	0.462	226333.0	0.449	
18	Sample 19	1900	0.35	0.926	199853.0	0.916	
19	Sample 20	462	0.185	0.731	207496.0	0.812	

	instrumentalness	key	liveness	loudness	mode	speechiness	tempo
0	0.633	8	0.102	-24.477	1	0.0497	107.327
1	0.0	4	0.215	-7.384	0	0.101	143.948
2	0.00626	10	0.191323	15.0	1	0.0331	119.998
3	0.0	4	0.152	-7.356	1	0.0322	141.916
4	0.000001	11	0.0786	-13.742	0	0.0341	104.996
5	0.000033	0	0.234	-4.699	0	0.0429	122.948
6	0.0	11	0.0802	-4.024	1	0.147	169.801
7	0.054	4	0.226	-9.971	0	0.271	82.738
8	0.0	9	0.107	-6.085	0	0.13	177.916
9	0.129	6	0.226	746.0	0	0.0539	126.394
10	0.000371	11	0.113	21.0	0	0.0561	130.008
11	0.0	5	0.0759	121.0	0	0.0901	123.659
12	0.000007	7	0.233	-4.447	1	0.0364	110.054
13	0.958	1	0.0633	-3.114	1	0.0489	128.083
14	0.0	10	0.0756	713.0	1	0.0432	88.541
15	0.0	6	0.691	-3.59	0	0.158	141.91
16	0.0	7	0.134	-4.267	1	0.0367	80.984
17	0.000632	0	0.0993	-9.271	0	0.0916	209.686
18	0.000006	10	0.0234	-2.221	0	0.0929	110.007
19	0.154	5	0.131	-6.691	1	0.0608	91.96

	time_signature	valence	target
0	4	0.134	0
1	4	0.374	1
2	4	0.935	1
3	3	0.71	0
4	4	0.331	1
5	4	0.47	0
6	4	0.766	0
7	4	0.207	1
8	4	0.404	0
9	4	0.604	1
10	4	0.797	0
11	4	0.629	0
12	4	0.784	0
13	4	0.113	0
14	4	0.637	0
15	4	0.524	1
16	4	0.811	0
17	4	0.748	1
18	4	0.903	0
19	4	0.861	1

	song_title	artist
0	The Man I Love	Marcus Roberts
1	Money Trees	Kendrick Lamar

2	Wait & See	Holy Ghost!
3	Perfect Harmony	Rags Cast
4	Midnight City	M83
5	Symphony	Clean Bandit
6	Sola (Remix) [feat. Daddy Yankee, Wisin, Farru…	Anuel Aa
7	Are you… Can you… Were you? (Felt)	Shabazz Palaces
8	Save My Soul	JoJo
9	Oh	FIDLAR
10	Sexy Bitch (feat. Akon) - Featuring Akon;explicit	David Guetta
11	Show You Love	Kato
12	Tearin' up My Heart - Radio Edit	*NSYNC
13	Raven	John Dahlbäck
14	Sleep Without You	Brett Young
15	Taylor Gang - Bonus Track	Wiz Khalifa
16	I'm the One	DJ Khaled
17	Danger and Dread	Brown Bird
18	Hollaback Girl	Gwen Stefani
19	Say Please	Teams vs. Star Slinger