Data Mining Assignment 1

Name: Harsh Arora
Roll NO:AE-1218
Subject: DSC Data Mining I
Course: BSc(Hons.)Computer Science
Semester:4th Semester
Date Of Submission: 17th February 2024
College: Acharya Narendra College

Given By: Sharanjit Kaur Ma'am

```python
import pandas as pd

#Importing The Dataset
dataset_anime=pd.read_csv(r"C:\Users\bharg\Downloads\anime.csv")
print(dataset_anime)
```

```
       anime_id                                              name  \
0         32281                                  Kimi no Na wa.
1          5114                  Fullmetal Alchemist: Brotherhood
2         28977                                          Gintama°
3          9253                                       Steins;Gate
4          9969                                     Gintama&#039;
...         ...                                              ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                         Under World
12291      5621                        Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                     Yasuji no Pornorama: Yacchimae!!

                                                  genre   type
episodes  \
0                    Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                      Sci-Fi, Thriller     TV
24
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                  ...    ...       .
..
12289                                         Hentai    OVA
1
12290                                         Hentai    OVA
1
12291                                         Hentai    OVA
4
12292                                         Hentai    OVA
1
12293                                         Hentai  Movie
1

       rating  members
0        9.37   200630
1        9.26   793665
2        9.25   114262
3        9.17   673572
4        9.16   151266
```

```
...       ...       ...
12289    4.15      211
12290    4.28      183
12291    4.88      219
12292    4.98      175
12293    5.46      142

[12294 rows x 7 columns]
```

```
#i.Find the rows and columns of the Dataframe
shape =dataset_anime.shape
print("Number of Rows and Columns of the Dataframe are:
",shape,"Respectively")
```

```
Number of Rows and Columns of the Dataframe are:  (12294, 7)
Respectively
```

```
#ii. Find the number of records in the Dataframe
print("Number of Records in the dataframe are :" ,len(dataset_anime))
```

```
Number of Records in the dataframe are : 12294
```

```
#iii. Find the Number of columns along with Column Names
number_of_columns = len(dataset_anime.columns)
column_names = list(dataset_anime.columns.values)
print("Number of columns are:", number_of_columns)
print("Names of all columns in the dataset are:", column_names)
```

```
Number of columns are: 7
Names of all columns in the dataset are: ['anime_id', 'name', 'genre',
'type', 'episodes', 'rating', 'members']
```

```
#iv. Display Data Types of all Columns
print("Date types of each Columns are are :",dataset_anime.dtypes)
```

```
Date types of each Columns are are : anime_id      int64
name          object
genre         object
type          object
episodes      object
rating       float64
members        int64
dtype: object
```

```
#v. Display Statistical Summary ofNumeric Columns Only
print("Statistical Summary of Numeric Columns
are:",dataset_anime.describe())
```

```
Statistical Summary of Numeric Columns are:             anime_id
rating       members
count  12294.000000  12064.000000  1.229400e+04
mean   14058.221653      6.473902  1.807134e+04
```

```
std    11455.294701      1.026746  5.482068e+04
min        1.000000      1.670000  5.000000e+00
25%     3484.250000      5.880000  2.250000e+02
50%    10260.500000      6.570000  1.550000e+03
75%    24794.500000      7.180000  9.437000e+03
max    34527.000000     10.000000  1.013917e+06
```

```
#vi. Display Statistical Summary of All Types of Columns
print("Statistical summary of all columns
is:",dataset_anime.describe(include='all'))
```

```
Statistical summary of all columns is:             anime_id
name   genre   type episodes  \
count   12294.000000                    12294   12232  12269    12294

unique           NaN                    12292    3264      6      187

top              NaN  Shi Wan Ge Leng Xiaohua  Hentai     TV        1

freq             NaN                        2     823   3787     5677

mean    14058.221653                      NaN     NaN    NaN      NaN

std     11455.294701                      NaN     NaN    NaN      NaN

min         1.000000                      NaN     NaN    NaN      NaN

25%      3484.250000                      NaN     NaN    NaN      NaN

50%     10260.500000                      NaN     NaN    NaN      NaN

75%     24794.500000                      NaN     NaN    NaN      NaN

max     34527.000000                      NaN     NaN    NaN      NaN


               rating       members
count   12064.000000  1.229400e+04
unique           NaN           NaN
top              NaN           NaN
freq             NaN           NaN
mean        6.473902  1.807134e+04
std         1.026746  5.482068e+04
min         1.670000  5.000000e+00
25%         5.880000  2.250000e+02
50%         6.570000  1.550000e+03
75%         7.180000  9.437000e+03
max        10.000000  1.013917e+06
```

```
#vii. Differentiate Between Shape and Size Property of the DataFrame
print("Shape of a Dataframe is used to get the dimensions of the
```

```
Pandas DataFrame. That is Rows And Columns")
print("Shape of the Given Dataframe is:", dataset_anime.shape)
print("Size of a Dataframe is equivalent to total number of elements.
That is rows x columns.")
print("Size of the Given Dataframe is:" ,dataset_anime.size)
```

```
Shape of a Dataframe is used to get the dimensions of the Pandas
DataFrame. That is Rows And Columns
Shape of the Given Dataframe is: (12294, 7)
Size of a Dataframe is equivalent to total number of elements. That is
rows x columns.
Size of the Given Dataframe is: 86058
```

```
#viii. Display a Particular Row and a Particular Column using iloc and
loc

#using iloc function(is used to get a paticular column or row by its
integer Value)
print("Accessing a Paticular Row using iloc :\n
",dataset_anime.iloc[[12]])
print("\n")

#using loc function(is used to get a Paticular Row or Column Using Its
Integer Value or Row/Column Name)
print("Accessing a Paticular Row using loc :\n",
dataset_anime.loc[[22]])
print("\n")
print("Accessing Rows and Columns Using loc:\
n",dataset_anime.loc[0,"name"])
```

```
Accessing a Paticular Row using iloc :
      anime_id     name
genre type  \
12       918  Gintama  Action, Comedy, Historical, Parody, Samurai,
S...   TV

   episodes  rating  members
12      201    9.04   336376


Accessing a Paticular Row using loc :
     anime_id          name
genre  \
22          1  Cowboy Bebop  Action, Adventure, Comedy, Drama, Sci-Fi,
Space

   type episodes  rating  members
22   TV       26    8.82   486824
```

```
Accessing Rows and Columns Using loc:
 Kimi no Na wa.
```

```
#ix. Display Row Number [1,2,5] of the DataFrame
print("Row Number[1] of The DataFrame are :\
n",dataset_anime.iloc[1],"\n")
print("Row Number[2] of The DataFrame are :\
n",dataset_anime.iloc[2],"\n")
print("Row Number[5] of The DataFrame are :\
n",dataset_anime.iloc[5],"\n")
```

```
Row Number[1] of The DataFrame are :
 anime_id                                                    5114
name                          Fullmetal Alchemist: Brotherhood
genre       Action, Adventure, Drama, Fantasy, Magic, Mili...
type                                                          TV
episodes                                                      64
rating                                                      9.26
members                                                   793665
Name: 1, dtype: object

Row Number[2] of The DataFrame are :
 anime_id                                                   28977
name                                                     Gintama°
genre       Action, Comedy, Historical, Parody, Samurai, S...
type                                                          TV
episodes                                                      51
rating                                                      9.25
members                                                   114262
Name: 2, dtype: object

Row Number[5] of The DataFrame are :
 anime_id                                                   32935
name        Haikyuu!!: Karasuno Koukou VS Shiratorizawa Ga...
genre                   Comedy, Drama, School, Shounen, Sports
type                                                          TV
episodes                                                      10
rating                                                      9.15
members                                                    93351
Name: 5, dtype: object
```

```
# x. Display a particular value in the data
print("Lets Get a Random Locataion Of The DataFrame by Using iloc
Function:\n", dataset_anime.iloc[3,1])
```

```
Lets Get a Random Locataion Of The DataFrame by Using iloc Function:
 Steins;Gate
```

```
# xi. Find All Rows Where a Numeric Field Value is More Than its
Average
```

```
avg_rating= dataset_anime["rating"].mean()
print("All Rows Where Rating is More Than its Average :",
[dataset_anime["rating"]>avg_rating])
```

```
All Rows Where Rating is More Than its Average : [0         True
1         True
2         True
3         True
4         True
         ...
12289    False
12290    False
12291    False
12292    False
12293    False
Name: rating, Length: 12294, dtype: bool]
```

```
#xii. Display Unique Values in Each Categorical Column
for col in dataset_anime.columns:
    print(col ,dataset_anime[col].unique(),"\n")
```

```
anime_id [32281  5114 28977 ...  5621  6133 26081]

name ['Kimi no Na wa.' 'Fullmetal Alchemist: Brotherhood'
'Gintama°' ...
 'Violence Gekiga David no Hoshi'
 'Violence Gekiga Shin David no Hoshi: Inma Densetsu'
 'Yasuji no Pornorama: Yacchimae!!']

genre ['Drama, Romance, School, Supernatural'
 'Action, Adventure, Drama, Fantasy, Magic, Military, Shounen'
 'Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen' ...
 'Hentai, Sports' 'Drama, Romance, School, Yuri' 'Hentai, Slice of
Life']

type ['Movie' 'TV' 'OVA' 'Special' 'Music' 'ONA' nan]

episodes ['1' '64' '51' '24' '10' '148' '110' '13' '201' '25' '22'
'75' '4' '26'
 '12' '27' '43' '74' '37' '2' '11' '99' 'Unknown' '39' '101' '47' '50'
 '62' '33' '112' '23' '3' '94' '6' '8' '14' '7' '40' '15' '203' '77'
'291'
 '120' '102' '96' '38' '79' '175' '103' '70' '153' '45' '5' '21' '63'
'52'
 '28' '145' '36' '69' '60' '178' '114' '35' '61' '34' '109' '20' '9'
'49'
 '366' '97' '48' '78' '358' '155' '104' '113' '54' '167' '161' '42'
'142'
 '31' '373' '220' '46' '195' '17' '1787' '73' '147' '127' '16' '19'
'98'
```

```
 '150' '76' '53' '124' '29' '115' '224' '44' '58' '93' '154' '92' '67'
 '172' '86' '30' '276' '59' '72' '330' '41' '105' '128' '137' '56'
'55'
 '65' '243' '193' '18' '191' '180' '91' '192' '66' '182' '32' '164'
'100'
 '296' '694' '95' '68' '117' '151' '130' '87' '170' '119' '84' '108'
'156'
 '140' '331' '305' '300' '510' '200' '88' '1471' '526' '143' '726'
'136'
 '1818' '237' '1428' '365' '163' '283' '71' '260' '199' '225' '312'
'240'
 '1306' '1565' '773' '1274' '90' '475' '263' '83' '85' '1006' '80'
'162'
 '132' '141' '125']

rating [ 9.37  9.26  9.25  9.17  9.16  9.15  9.13  9.11  9.1   9.06
9.05  9.04
  8.98  8.93  8.92  8.88  8.84  8.83  8.82  8.81  8.8   8.78  8.77
8.76
  8.75  8.74  8.73  8.72  8.71  8.69  8.68  8.67  8.66  8.65  8.64
8.62
  8.61  8.6   8.59  8.58  8.57  8.56  8.55  8.54  8.53  8.52  8.51
8.5
  8.49  8.48  8.47  8.46  8.45  8.44  8.43  8.42  8.41  8.4   8.39
8.38
  8.37  8.36  8.35  8.34  8.33  8.32  8.31  8.3   8.29  8.28  8.27
8.26
  8.25  8.24  8.23  8.22  8.21  8.2   8.19  8.18  8.17  8.16  8.15
8.14
  8.13  8.12  8.11  8.1   8.09  8.08  8.07  8.06  8.05  8.04  8.03
8.02
  8.01  8.    7.99  7.98  7.97  7.96  7.95  7.94  7.93  7.92  7.91
7.9
  7.89  7.88  7.87  7.86  7.85  7.84  7.83  7.82  7.81  7.8   7.79
7.78
  7.77  7.76  7.75  7.74  7.73  7.72  7.71  7.7   7.69  7.68  7.67
7.66
  7.65  7.64  7.63  7.62  7.61  7.6   7.59  7.58  7.57  7.56  7.55
7.54
  7.53  7.52  7.51  7.5   7.49  7.48  7.47  7.46  7.45  7.44  7.43
7.42
  7.41  7.4   7.39  7.38  7.37  7.36  7.35  7.34  7.33  7.32  7.31
7.3
  7.29  7.28  7.27  7.25  7.26  7.24  7.23  7.22  7.21  7.2   7.19
7.18
  7.17  7.16  7.14  7.15  7.13  7.12  7.11  7.1   7.09  7.08  7.07
7.06
  7.05  7.04  7.03  7.02  7.01  7.    6.99  6.98  6.97  6.96  6.95
6.94
```

```
  6.93  6.92  6.91  6.9   6.89  6.88  6.87  6.86  6.85  6.84  6.83
6.82
  6.81  6.8   6.79  6.78  6.75  6.77  6.76  6.74  6.73  6.72  6.71
6.7
  6.69  6.68  6.67  6.66  6.65  6.64  6.63  6.62  6.61  6.6   6.59
6.58
  6.57  6.56  6.55  6.54  6.53  6.52  6.51  6.47  6.5   6.49  6.48
6.46
  6.45  6.42  6.44  6.43  6.39  6.41  6.4   6.38  6.37  6.35  6.36
6.34
  6.33  6.32  6.31  6.3   6.29  6.28  6.27  6.26  6.25  6.22  6.24
6.23
  6.21  6.2   6.19  6.18  6.17  6.16  6.15  6.14  6.13  6.12  6.1
6.11
  6.09  6.08  6.06  6.07  6.05  6.04  6.03  6.01  6.02  6.    5.99
5.98
  5.97  5.96  5.95  5.94  5.93  5.92  5.91  5.89  5.9   5.88  5.87
5.86
  5.85  5.84  5.83  5.82  5.81  5.8   5.79  5.78  5.77  5.76  5.75
5.74
  5.73  5.72  5.7   5.71  5.69  5.68  5.67  5.66  5.65  5.64  5.63
5.62
  5.61  5.6   5.59  5.58  5.57  5.56  5.55  5.53  5.54  5.52  5.51
5.5
  5.49  5.48  5.46  5.47  5.45  5.44  5.43  5.42  5.41  5.4   5.39
5.38
  5.37  5.36  5.35  5.34  5.33  5.32  5.31  5.3   5.29  5.28  5.27
5.26
  5.24  5.25  5.23  5.22  5.21  5.2   5.19  5.14  5.18  5.17  5.16
5.15
  5.13  5.11  5.12  5.1   5.09  5.07  5.08  5.06  5.05  5.04  5.03
5.02
  5.01  5.    4.99  4.98  4.97  4.96  4.95  4.94  4.93  4.92  4.91
4.9
  4.89  4.88  4.84  4.87  4.86  4.85  4.83  4.82  4.81  4.8   4.79
4.78
  4.77  4.76  4.75  4.74  4.73  4.72  4.71  4.7   4.69  4.68  4.66
4.67
  4.65  4.64  4.63  4.62  4.6   4.59  4.58  4.57  4.56  4.55  4.54
4.53
  4.52  4.49  4.5   4.48  4.46  4.45  4.44  4.43  4.42  4.4   4.39
4.38
  4.36  4.35  4.34  4.32  4.31  4.3   4.28  4.27  4.26  4.25  4.24
4.23
  4.22  4.21  4.19  4.17  4.16  4.15  4.11  4.08  4.04  4.03  4.02  4.
  3.99  3.98  3.96  3.91  3.9   3.88  3.87  3.86  3.84  3.83  3.82
3.8
  3.78  3.76  3.75  3.74  3.73  3.71  3.7   3.68  3.65  3.63  3.62
3.6
```

```
  3.58  3.59  3.56  3.47  3.46  3.41  3.36  3.33  3.32  3.27  2.95
2.93
  2.78  2.67  2.37  2.14  2.    4.06  4.18  4.09  3.67  3.    4.33
3.89
  4.2   3.61  4.13  3.11  2.58  4.29  3.43  3.57  4.05  4.51  3.4
3.79
  3.81  3.92  3.48  3.38  3.95  4.61  3.34  3.02  2.69  3.5   3.97
3.85
   nan  3.77  3.53  9.33  4.14  4.12  3.2   3.93  3.52  2.55  9.
2.97
  4.1   3.39  3.17  9.5   3.25  2.8   2.91  2.75  4.41  3.72  3.94
1.67
  4.37  3.69  3.26  3.49  2.49  2.84  4.07  3.35  3.54  3.21  3.42
9.6
  3.28 10.    3.51  3.29  2.72  3.64  2.98  3.44  1.92  2.86  3.14]

members [200630 793665 114262 ...  27411  57355    652]


#xiii. Display Frequency of Each Distinct Value in Each Column
(#unique/value_counts)
for col in dataset_anime.columns:
    print(dataset_anime[col].value_counts())

anime_id
32281    1
30404    1
26013    1
26017    1
15787    1
        ..
12455    1
28789    1
373      1
2089     1
26081    1
Name: count, Length: 12294, dtype: int64
name
Shi Wan Ge Leng Xiaohua                              2
Saru Kani Gassen                                     2
Bakabon Osomatsu no Karee wo Tazunete Sansenri       1
Backkom Meogeujan Yeohaeng                           1
Backkom Mission Impossible                           1
                                                    ..
Yoroiden Samurai Troopers Kikoutei Densetsu          1
Yuu☆Yuu☆Hakusho: Mu Mu Hakusho                       1
3-gatsu no Lion meets Bump of Chicken                1
Bannou Bunka Neko-Musume                             1
Yasuji no Pornorama: Yacchimae!!                     1
Name: count, Length: 12292, dtype: int64
```

```
genre
Hentai                                                 823
Comedy                                                 523
Music                                                  301
Kids                                                   199
Comedy, Slice of Life                                  179
                                                      ...
Adventure, Drama, Fantasy, Game, Sci-Fi                  1
Adventure, Demons, Fantasy, Historical                   1
Action, Comedy, Drama, Mecha, Music, Sci-Fi, Shounen     1
Action, Comedy, Fantasy, Mecha, Sci-Fi, Shounen          1
Hentai, Slice of Life                                    1
Name: count, Length: 3264, dtype: int64
type
TV         3787
OVA        3311
Movie      2348
Special    1676
ONA         659
Music       488
Name: count, dtype: int64
episodes
1      5677
2      1076
12      816
13      572
26      514
       ...
358       1
366       1
201       1
172       1
125       1
Name: count, Length: 187, dtype: int64
rating
6.00    141
7.00     99
6.50     90
6.25     84
5.00     76
       ...
3.47      1
3.71      1
3.87      1
3.91      1
3.14      1
Name: count, Length: 598, dtype: int64
members
72       36
```

```
60       36
74       33
62       32
103      31
         ..
11941     1
5134      1
9447      1
2945      1
72174     1
Name: count, Length: 6706, dtype: int64
```

```
#xiv. count NULL values in The Data Frame
print("Number of Null Values in The DataFrame are:\n")
print(dataset_anime.isnull().sum().sum())
```

```
Number of Null Values in The DataFrame are:

317
```

```
#xv. Count NULL Values in Each Column in  the DataFrame
print(dataset_anime.isnull().sum())
```

```
anime_id      0
name          0
genre        62
type         25
episodes      0
rating      230
members       0
dtype: int64
```

```
#xvi. Display Rows with No Null Value
print(dataset_anime.dropna(axis=1))
```

```
       anime_id                                              name
episodes  \
0         32281                                    Kimi no Na wa.
1
1          5114                         Fullmetal Alchemist: Brotherhood
64
2         28977                                          Gintama°
51
3          9253                                     Steins;Gate
24
4          9969                                   Gintama&#039;
51
...         ...                                               ...
...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
1
```

```
12290      5543                                          Under World
1
12291      5621                       Violence Gekiga David no Hoshi
4
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
1
12293     26081                     Yasuji no Pornorama: Yacchimae!!
1

       members
0       200630
1       793665
2       114262
3       673572
4       151266
...        ...
12289      211
12290      183
12291      219
12292      175
12293      142

[12294 rows x 4 columns]
```

```
#xvii. Drop All Rows with Atleast One NA value
print(dataset_anime.dropna())
```

```
       anime_id                                              name  \
0         32281                                   Kimi no Na wa.
1          5114                  Fullmetal Alchemist: Brotherhood
2         28977                                          Gintama°
3          9253                                       Steins;Gate
4          9969                                     Gintama&#039;
...         ...                                               ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                          Under World
12291      5621                       Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                     Yasuji no Pornorama: Yacchimae!!

                                              genre   type
episodes  \
0                    Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                      Sci-Fi, Thriller     TV
24
```

```
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                   ...    ...      .
..
12289                                               Hentai    OVA
1
12290                                               Hentai    OVA
1
12291                                               Hentai    OVA
4
12292                                               Hentai    OVA
1
12293                                               Hentai  Movie
1

       rating  members
0        9.37   200630
1        9.26   793665
2        9.25   114262
3        9.17   673572
4        9.16   151266
...       ...      ...
12289    4.15      211
12290    4.28      183
12291    4.88      219
12292    4.98      175
12293    5.46      142

[12017 rows x 7 columns]
```

```
# xviii. Drop a Column on The Basis of all NA or any NA values, no in-
place change in data frame
print(dataset_anime.dropna(axis=1, how='any', inplace=False))
```

```
       anime_id                                              name
episodes  \
0         32281                                     Kimi no Na wa.
1
1          5114                      Fullmetal Alchemist: Brotherhood
64
2         28977                                         Gintama°
51
3          9253                                      Steins;Gate
24
4          9969                                    Gintama&#039;
51
...         ...                                               ...
...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
1
```

```
12290      5543                                          Under World
1
12291      5621                       Violence Gekiga David no Hoshi
4
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
1
12293     26081                     Yasuji no Pornorama: Yacchimae!!
1

       members
0       200630
1       793665
2       114262
3       673572
4       151266
...        ...
12289      211
12290      183
12291      219
12292      175
12293      142

[12294 rows x 4 columns]
```

```
#xix. Drop All Rows If They Have NA in all columns
print(dataset_anime.dropna(how='all', inplace=False))
```

```
       anime_id                                              name  \
0         32281                                    Kimi no Na wa.
1          5114                  Fullmetal Alchemist: Brotherhood
2         28977                                          Gintama°
3          9253                                       Steins;Gate
4          9969                                     Gintama&#039;
...         ...                                               ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                          Under World
12291      5621                       Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                     Yasuji no Pornorama: Yacchimae!!

                                              genre   type
episodes  \
0                    Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                      Sci-Fi, Thriller     TV
24
```

```
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                   ...    ...      .
..
12289                                               Hentai    OVA
1
12290                                               Hentai    OVA
1
12291                                               Hentai    OVA
4
12292                                               Hentai    OVA
1
12293                                               Hentai  Movie
1

       rating  members
0        9.37   200630
1        9.26   793665
2        9.25   114262
3        9.17   673572
4        9.16   151266
...       ...      ...
12289    4.15      211
12290    4.28      183
12291    4.88      219
12292    4.98      175
12293    5.46      142

[12294 rows x 7 columns]
```

```
#xx. Drop all Rows With Less Than 3 Correct values
value=3
print(dataset_anime.dropna(thresh=value,inplace=False))
```

```
       anime_id                                              name  \
0         32281                                  Kimi no Na wa.
1          5114                  Fullmetal Alchemist: Brotherhood
2         28977                                          Gintama°
3          9253                                       Steins;Gate
4          9969                                     Gintama&#039;
...         ...                                              ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                         Under World
12291      5621                        Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                    Yasuji no Pornorama: Yacchimae!!

                                                 genre   type
episodes  \
0                   Drama, Romance, School, Supernatural  Movie
```

```
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                       Sci-Fi, Thriller     TV
24
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                   ...    ...      .
..
12289                                             Hentai    OVA
1
12290                                             Hentai    OVA
1
12291                                             Hentai    OVA
4
12292                                             Hentai    OVA
1
12293                                             Hentai  Movie
1

       rating  members
0        9.37   200630
1        9.26   793665
2        9.25   114262
3        9.17   673572
4        9.16   151266
...       ...      ...
12289    4.15      211
12290    4.28      183
12291    4.88      219
12292    4.98      175
12293    5.46      142

[12294 rows x 7 columns]
```

```
#xxi. Fill NULL Values with Average of The Corresponding Numeric
Columns
numeric_columns =
dataset_anime.select_dtypes(include='number').columns
for col in numeric_columns:
    dataset_anime[col].fillna(dataset_anime[col].mean(), inplace=True)
print(dataset_anime)
```

```
       anime_id                                              name  \
0         32281                                    Kimi no Na wa.
1          5114                 Fullmetal Alchemist: Brotherhood
2         28977                                         Gintama°
3          9253                                      Steins;Gate
```

```
4          9969                                        Gintama&#039;
...         ...                                                ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                         Under World
12291      5621                   Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                  Yasuji no Pornorama: Yacchimae!!

                                                  genre   type
episodes  \
0                   Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                      Sci-Fi, Thriller     TV
24
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                  ...    ...       .
..
12289                                           Hentai    OVA
1
12290                                           Hentai    OVA
1
12291                                           Hentai    OVA
4
12292                                           Hentai    OVA
1
12293                                           Hentai  Movie
1

       rating  members
0        9.37   200630
1        9.26   793665
2        9.25   114262
3        9.17   673572
4        9.16   151266
...       ...      ...
12289    4.15      211
12290    4.28      183
12291    4.88      219
12292    4.98      175
12293    5.46      142

[12294 rows x 7 columns]
```

#xxii. Display min and max of Numeric Column and Mode of Categorical Column

```
numeric_stats = dataset_anime.describe().loc[['min', 'max']]
print("Minimum and Maximum of Numeric Columns are:          \
n",numeric_stats,"\n")

categorical_mode = dataset_anime.mode().iloc[0]
print("\nMode of Categorical Columns are :\n",categorical_mode)
```

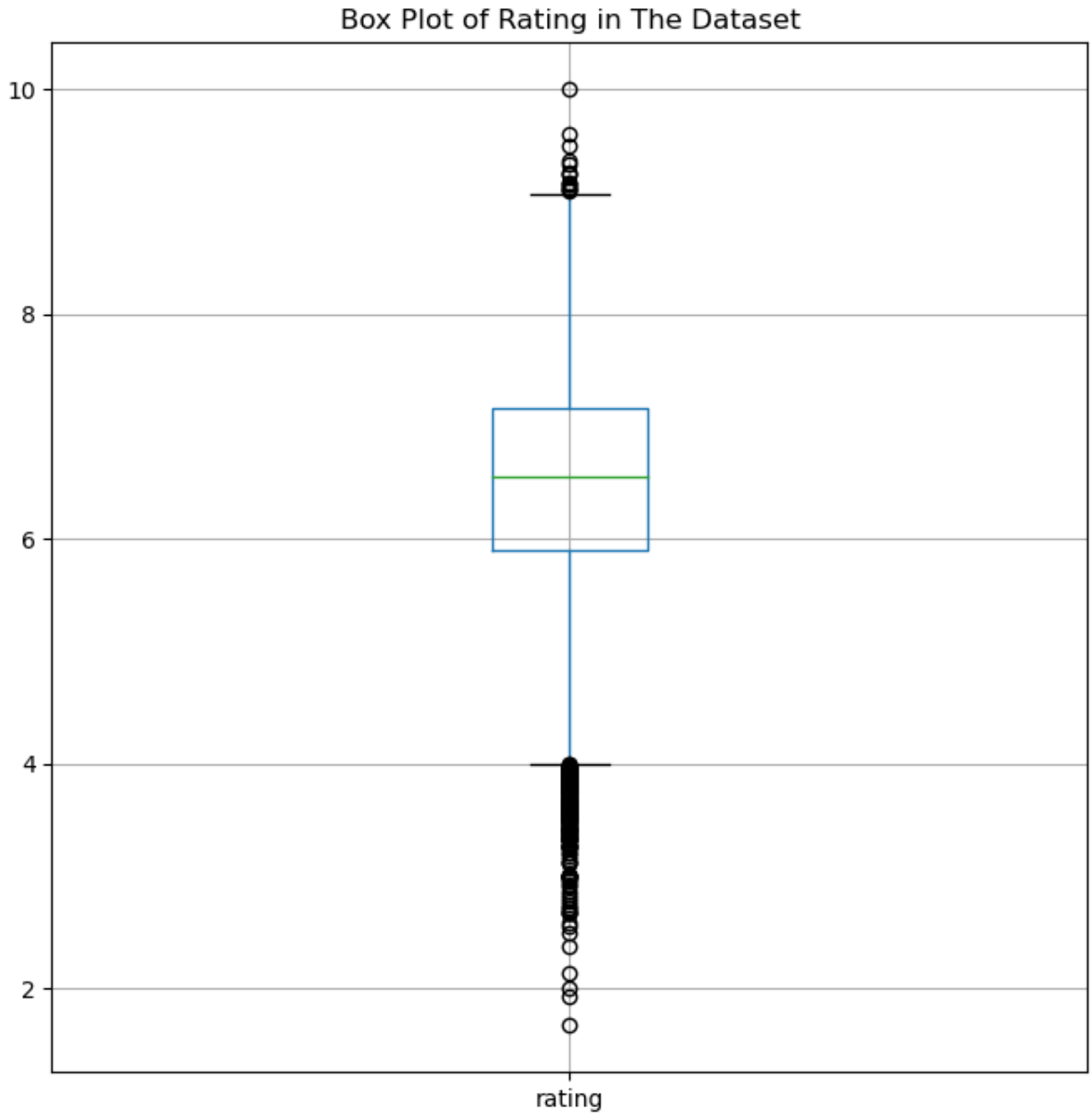
```
Minimum and Maximum of Numeric Columns are:
      anime_id  rating    members
min       1.0    1.67        5.0
max   34527.0   10.00  1013917.0


Mode of Categorical Columns are :
 anime_id                    1
name          Saru Kani Gassen
genre                  Hentai
type                       TV
episodes                    1
rating               6.473902
members                 60.0
Name: 0, dtype: object
```

```
#xxiii. Convert a Numeric Column in a Categorical Form (user-defined
values) (Discretize using cut/qcut)

bins_value = [4 ,5, 6, 7, 8, 9, float('inf')]
labels_value=['F','E', 'D', 'C', 'B', 'A']
dataset_anime['Categorical'] = pd.cut(dataset_anime['rating'],
bins=bins_value,labels=labels_value)
print(dataset_anime)
```

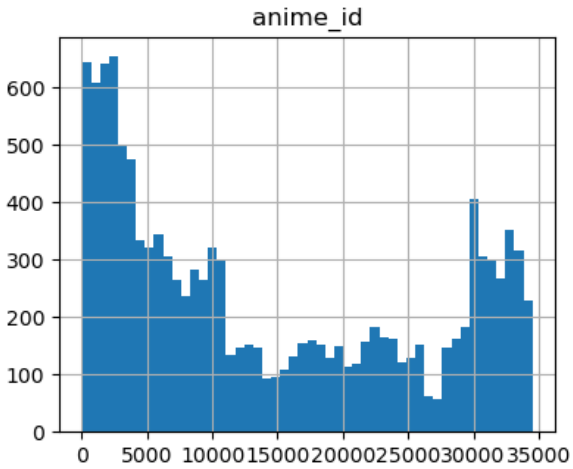
```
       anime_id                                                name  \
0         32281                                     Kimi no Na wa.
1          5114                  Fullmetal Alchemist: Brotherhood
2         28977                                          Gintama°
3          9253                                       Steins;Gate
4          9969                                     Gintama&#039;
...         ...                                               ...
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                         Under World
12291      5621                        Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                   Yasuji no Pornorama: Yacchimae!!

                                                  genre   type
episodes  \
0                   Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
```

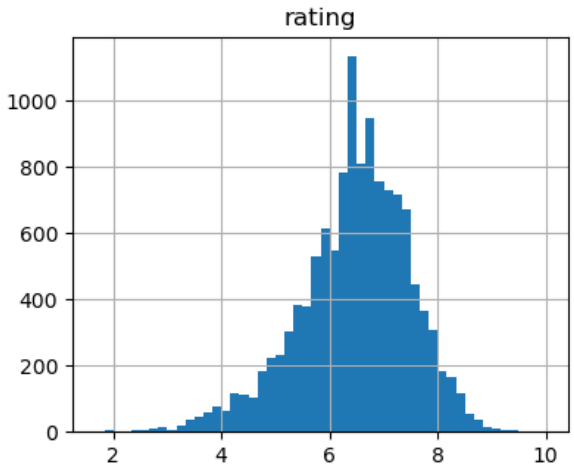
```
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                        Sci-Fi, Thriller     TV
24
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                   ...    ...       .
..
12289                                             Hentai    OVA
1
12290                                             Hentai    OVA
1
12291                                             Hentai    OVA
4
12292                                             Hentai    OVA
1
12293                                             Hentai  Movie
1

       rating  members Categorical
0        9.37   200630           A
1        9.26   793665           A
2        9.25   114262           A
3        9.17   673572           A
4        9.16   151266           A
...       ...      ...         ...
12289    4.15      211           F
12290    4.28      183           F
12291    4.88      219           F
12292    4.98      175           F
12293    5.46      142           E

[12294 rows x 8 columns]
```

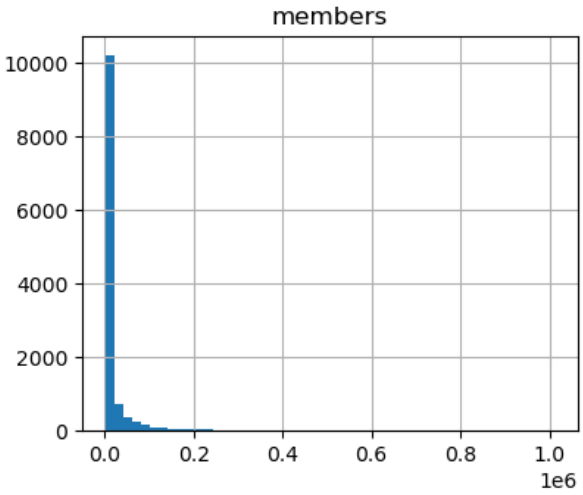
```
import matplotlib.pyplot as plt

#xxiv. Draw a Box Plot For Numeric Values to Display 5 Point
Statistical Summary Along With Outliers if Any.
plt.figure(figsize=(8, 8))
dataset_anime.boxplot(column='rating', sym='o')
plt.title('Box Plot of Rating in The Dataset')
plt.show()
```



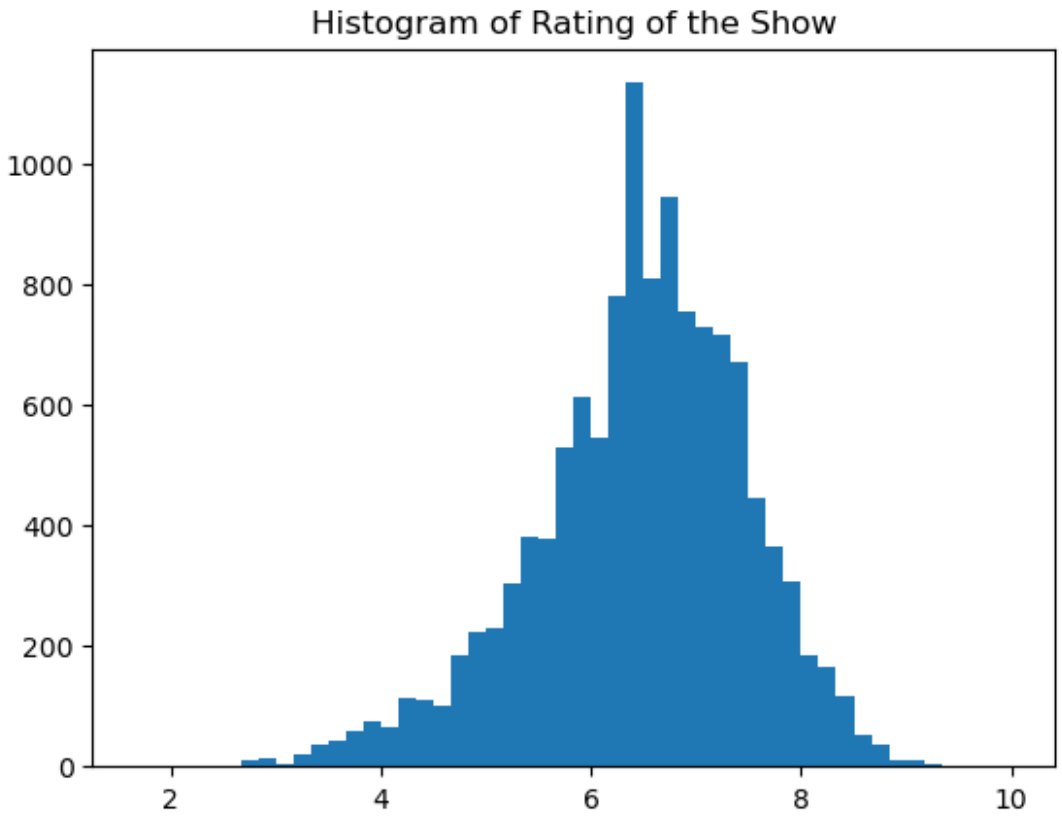

```
#xxv. Plot Distribution of Numeric Columns
dataset_anime.hist(figsize=(10, 8), bins=50)
plt.show()
```








```
#xxvi. Plot histogram
plt.figsize=(10, 8)
plt.hist(dataset_anime['rating'], bins=50,)
plt.title('Histogram of Rating of the Show')
plt.show()
```




```
#xxvii. Divide the numeric column using explicit values of quantiles
and label each category (#cats=pd.qcut(data, [0, 0.15, 0.3, 0.8, 1.])
and display number of values in each category
quantile_bins = [0, 0.15, 0.3, 0.8, 1.]
dataset_anime['Category'] = pd.qcut(dataset_anime['members'],
quantile_bins, labels=['A+', 'B+', 'C+', 'D+'])


category_counts = dataset_anime['Category'].value_counts()
print("Number of values in each category:",category_counts)
```

```
Number of values in each category: Category
C+    6144
D+    2459
A+    1871
B+    1820
Name: count, dtype: int64
```

```
#xxviii. Display All Values in First Numeric Column of the DataFrame
Which are More Than (mean+2std) of that Column
numeric_data = dataset_anime['members']
filtered_data = numeric_data[numeric_data > (numeric_data.mean() + 2 *
numeric_data.std())]
```

```
print("Values more than (mean + 2 * std):\n")
print(filtered_data)
```

```
Values more than (mean + 2 * std):

0        200630
1        793665
3        673572
4        151266
6        425855
         ...
4828     174337
5307     133971
5643     134349
6324     279183
10899    170054
Name: members, Length: 419, dtype: int64
```

```
#xxix. Explore use of crosstab.
cross_tab = pd.crosstab(dataset_anime['rating'],
dataset_anime['episodes'])

print("Cross-tabulation of Rating and Episodes are:\n")
print(cross_tab)
```

```
Cross-tabulation of Rating and Episodes are:

episodes  1  10  100  1006  101  102  103  104  105  108  ...  91  92
93  94  \
rating                                                    ...

1.67      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
1.92      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
2.00      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
2.14      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
2.37      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
...      ..  ..  ...   ...  ...  ...  ...  ...  ...  ...  ...  ..  ..
..  ..
9.33      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
9.37      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
9.50      1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0
9.60      0   0    0     0    0    0    0    0    0    0  ...   0   0
```

```
0   0
10.00     1   0    0     0    0    0    0    0    0    0  ...   0   0
0   0

episodes  95  96  97  98  99  Unknown
rating
1.67       0   0   0   0   0        0
1.92       0   0   0   0   0        0
2.00       0   0   0   0   0        0
2.14       0   0   0   0   0        0
2.37       0   0   0   0   0        0
...       ..  ..  ..  ..  ..      ...
9.33       0   0   0   0   0        0
9.37       0   0   0   0   0        0
9.50       0   0   0   0   0        0
9.60       0   0   0   0   0        1
10.00      0   0   0   0   0        0

[599 rows x 187 columns]
```

```
#Q2. Apply data cleaning techniques on any dataset. Techniques may
include handling missing values, outliers, inconsistent values. A set
of validation rules can be prepared based on the dataset and
validations can be performed.

print("Original dataset :")
print(dataset_anime.info())

#Missing Values
dataset_anime.isnull()
dataset_anime.fillna({'rating': dataset_anime['rating'].mean(),
'episodes': dataset_anime['episodes'].mode()[0]}, inplace=True)

# outliers
outlier_threshold = dataset_anime['rating'].mean() + 2 *
dataset_anime['rating'].std()
dataset_anime['rating'] = dataset_anime['rating'].apply(lambda x:
dataset_anime['rating'].mean() if x > outlier_threshold else x)

# For example converting all values to Lowercase
dataset_anime['episodes'] = dataset_anime['episodes'].str.lower()

# Display cleaned dataset information
print("\nCleaned dataset information:\n")
print(dataset_anime.info())
```

```
print("\nInvalid values in 'members':\n")
print(invalid_values)
```

```
Original dataset :
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12294 entries, 0 to 12293
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   anime_id     12294 non-null  int64
 1   name         12294 non-null  object
 2   genre        12232 non-null  object
 3   type         12269 non-null  object
 4   episodes     12294 non-null  object
 5   rating       12294 non-null  float64
 6   members      12294 non-null  int64
 7   Categorical  12028 non-null  category
 8   Category     12294 non-null  category
dtypes: category(2), float64(1), int64(2), object(4)
memory usage: 696.9+ KB
None

Cleaned dataset information:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12294 entries, 0 to 12293
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   anime_id     12294 non-null  int64
 1   name         12294 non-null  object
 2   genre        12232 non-null  object
 3   type         12269 non-null  object
 4   episodes     12294 non-null  object
 5   rating       12294 non-null  float64
 6   members      12294 non-null  int64
 7   Categorical  12028 non-null  category
 8   Category     12294 non-null  category
dtypes: category(2), float64(1), int64(2), object(4)
memory usage: 696.9+ KB
None

Invalid values in 'members':
      anime_id                                                name  \
0        32281                                     Kimi no Na wa.
1         5114                  Fullmetal Alchemist: Brotherhood
2        28977                                          Gintama°
3         9253                                       Steins;Gate
4         9969                                     Gintama&#039;
...        ...                                                ...
```

```
12289      9316       Toushindai My Lover: Minami tai Mecha-Minami
12290      5543                                        Under World
12291      5621                      Violence Gekiga David no Hoshi
12292      6133  Violence Gekiga Shin David no Hoshi: Inma Dens...
12293     26081                     Yasuji no Pornorama: Yacchimae!!

                                                  genre   type
episodes  \
0                    Drama, Romance, School, Supernatural  Movie
1
1      Action, Adventure, Drama, Fantasy, Magic, Mili...     TV
64
2      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
3                                       Sci-Fi, Thriller     TV
24
4      Action, Comedy, Historical, Parody, Samurai, S...     TV
51
...                                                  ...    ...       .
..
12289                                           Hentai    OVA
1
12290                                           Hentai    OVA
1
12291                                           Hentai    OVA
4
12292                                           Hentai    OVA
1
12293                                           Hentai  Movie
1

          rating  members Categorical Category
0      6.473902   200630           A        D
1      6.473902   793665           A        D
2      6.473902   114262           A        D
3      6.473902   673572           A        D
4      6.473902   151266           A        D
...         ...      ...         ...      ...
12289  4.150000      211           F        B
12290  4.280000      183           F        B
12291  4.880000      219           F        B
12292  4.980000      175           F        B
12293  5.460000      142           E        B

[12282 rows x 9 columns]
```