```
**DATA ANALYSIS OF GOOGLE APP'S RATING**
          Importing the necessary Modules
 In [58]: # Importing Necessary Libraries
           import numpy as np
           import pandas as pd
           import seaborn as sns
           import matplotlib.pyplot as plt
           %matplotlib inline
                                   ------ Reading Data ------
           Please ensure that your folder name **does not have a space or a forward slas(/)**
 In [39]: # import CSV of Google App store file.
           google_data = pd.read_csv("D:\Softwares\Data Analyst & Science Software\Data Science Projects\Datasets\Google App Ra
           ting Data Analysis Master\ipynb_checkpoints\googleplaystore.csv")
 In [40]: # Inspecting the first 10 rows
           google_data.head(10)
 Out[40]:
                                                                                                                    Last Current Android
                                                                                             Content
                                             Category Rating Reviews Size
                                                                           Installs Type Price
                                                                                              Rating
                                                                                                                 Updated
                                                                                                                                   Ver
                                                                                                                            Ver
               Photo Editor & Candy Camera &
                                                                                                                                  4.0.3
                                                                                          0 Everyone
                                                                                                       Art & Design 7-Jan-18
                                      ART_AND_DESIGN
                                                                           10,000+ Free
                                                                                                                           1.0.0
                        Grid & ScrapBook
                                                                                                                                 and up
                                                                                                            Art &
                                                                                                                  15-Jan-
                                                                                                                                  4.0.3
                      Coloring book moana ART AND DESIGN
                                                                          500,000+ Free
                                                                                                                           2.0.0
                                                                                          0 Everyone
                                                                                                     Design;Pretend
                                                                                                                                 and up
                                                                                                             Play
                 U Launcher Lite – FREE Live
                                                                                                                                  4.0.3
                                      ART_AND_DESIGN
                                                             87510 8.7M 5,000,000+ Free
                                                                                                                           1.2.4
           2
                                                                                          0 Everyone
                                                                                                       Art & Design
                      Cool Themes, Hide ...
                                                                                                                                 and up
                                                                                                                                4.2 and
                      Sketch - Draw & Paint ART_AND_DESIGN
           3
                                                       4.5 215644 25M 50,000,000+ Free
                                                                                               Teen
                                                                                                       Art & Design 8-Jun-18
                                                                                                                           with
                                                                                                                          device
              Pixel Draw - Number Art Coloring
                                                                                                            Art & 20-Jun-
                                                                                                                                4.4 and
                                      ART_AND_DESIGN
                                                                          100,000+ Free
                                                                                          0 Everyone
                                                                                                                            1.1
                                 Book
                                                                                                    Design;Creativity
                                                                                                                  26-Mar-
                                                                                                                                2.3 and
                   Paper flowers instructions ART_AND_DESIGN
                                                               167 5.6M
                                                                           50,000+ Free
                                                                                          0 Everyone
                                                                                                       Art & Design
                  Smoke Effect Photo Maker -
                                                                                                                                  4.0.3
                                                                                                                            1.1
                                      ART_AND_DESIGN
                                                               178 19M
                                                                           50,000+ Free
                                                                                          0 Everyone
                                                                                                       Art & Design
                           Smoke Editor
                                                                                                                                 and up
                                                                                                                                4.2 and
                                                                                                                  14-Jun-
                           Infinite Painter ART_AND_DESIGN
                                                             36815 29M
                                                                        1,000,000+ Free
                                                                                          0 Everyone
                                                                                                       Art & Design
                                                                                                                                3.0 and
                                                                                                                           2.9.2
                     Garden Coloring Book ART_AND_DESIGN
                                                                         1,000,000+ Free
                                                                                          0 Everyone
                                                                                                       Art & Design
                                                                                                                                  4.0.3
                                                               121 3.1M
                                                                                          0 Everyone
               Kids Paint Free - Drawing Fun ART_AND_DESIGN 4.7
                                                                           10,000+ Free
                                                                                                                  3-Jul-18
                                                                                                                            2.8
                                                                                                    Design;Creativity
                                                                                                                                 and up
 In [41]: # Insecting shape of the Data into rows and columns
           google_data.shape
Out[41]: (10841, 13)
 In [42]: # Summary Statistics of Data
           google_data.describe()
 Out[42]:
                     Rating
                 9367.000000
                    4.193338
                    0.537431
                   1.000000
            25%
                    4.000000
             50%
                    4.300000
                    4.500000
                   19.000000
 In [43]: # Inspecting the Rating using BoxPlot
           google_data.boxplot()
 Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x11992df0>
           17.5
           15.0
           12.5
            10.0
            7.5
            5.0
            2.5
 In [44]: # Inspecting the Data using histogram
           google_data.hist()
 Out[44]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x119926F0>]],
                 dtype=object)
                                  Rating
            7000
            6000
            5000
            4000
            3000
            2000
           1000
                    2.5 5.0
                             7.5
                                   10.0 12.5 15.0 17.5
 In [45]: google_data.info()
           <class 'pandas.core.frame.DataFrame'>
           RangeIndex: 10841 entries, 0 to 10840
           Data columns (total 13 columns):
                Column
                                 Non-Null Count Dtype
                                 -----
                App
                                 10841 non-null object
                                10841 non-null object
                Category
                                 9367 non-null float64
               Rating
                                10841 non-null object
           3
               Reviews
                                 10841 non-null object
               Installs
                                10841 non-null object
                                10840 non-null object
                Туре
                Price
                                10841 non-null object
                Content Rating 10840 non-null object
                Genres
                                10841 non-null object
               Last Updated
                                10841 non-null object
           10
            11 Current Ver
                                10833 non-null object
            12 Android Ver
                                10838 non-null object
           dtypes: float64(1), object(12)
           memory usage: 592.9+ KB
                                ------ Data Cleaning ------
           Count the number of missing values in the Dataframe
 In [46]: # Checking the number of null values in Dataframe
           google_data.isnull()
 Out[46]:
                  App Category Rating Reviews Size Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver
                                                                                                                False
               0 False
                                        False False
               1 False
                               False
                                        False False
                                                                                 False
                                                                                                      False
                                                                                                                False
                         False
                                                    False False False
                                                                          False
                                                                                            False
               2 False
                          False
                                        False False
                                                                           False
                                                                                             False
                                                                                                      False
                                                                                                                False
               3 False
                          False
                                False
                                        False False
                                                    False False False
                                                                          False
                                                                                 False
                                                                                             False
                                                                                                      False
                                                                                                                False
                          False
                                                                                 False
                                                                                             False
                                                                                                      False
                                                                                                                 False
           10836 False
                          False
                               False
                                        False False
                                                    False False False
                                                                           False
                                                                                 False
                                                                                             False
                                                                                                      False
                                                                                                                False
            10837 False
                                                                                 False
                                                                                                                False
                          False
                               False
                                        False False
                                                                           False
                                                                                             False
                                                                                                      False
            10838 False
                          False
                                                                          False
                                                                                 False
                                                                                             False
                                                                                                      False
                                                                                                                False
           10839 False
                                                    False False False
                                                                          False
                                                                                 False
                                                                                             False
                                                                                                      False
                                                                                                                False
                          False
                                        False False
           10840 False
                          False
                                        False False
                                                    False False False
                                                                           False
                                                                                 False
                                                                                             False
                                                                                                       False
                                                                                                                 False
           10841 rows × 13 columns
 In [47]: # Counting the number of missing values in each column
           google_data.isnull().sum()
 Out[47]: App
                                0
           Category
                              1474
           Rating
           Reviews
           Size
           Installs
           Type
           Price
           Content Rating
           Genres
           Last Updated
           Current Ver
           Android Ver
                                3
           dtype: int64
          Checking - how many ratings are more than 5 - outliners
 In [48]: # Checking the number of Ratings greater than 5
           google_data[google_data.Rating > 5]
 Out[48]:
                                                                                                                        Current
                                                                                                                                Android
                                        App Category Rating Reviews
                                                                    Size Installs Type
                                                                                                Rating
                                                                                                               Updated
                                                                                                                                   Ver
                   Life Made WI-Fi Touchscreen Photo
                                                                                                                        4.0 and
           10472
                                                                                                                 1.0.19
                                                      19.0
                                                                                                  NaN
                                                                                                                                   NaN
 In [49]: google_data.drop([10472] , inplace = True)
 In [50]: # Checking dataframe if the values is deleted or not
           google_data[10470:10475]
 Out[50]:
                                                                                                                                Android
                                                                                           Content
                                                                                                                  Last
                                                                                                                        Current
                                         Category Rating Reviews Size
                                                                       Installs Type Price
                                                                                            Rating
                                                                                                                Updated
                                                                                                                                   Ver
                                                                                                                                2.3 and
           10470
                                  COMMUNICATION
                                                                                                              10-Feb-17
                                                                                                                           0.1
                        Jazz Wi-Fi
                                                            49 4.0M
                                                                       10,000+
                                                                                                                               4.0.3 and
           10471
                   Xposed Wi-Fi-Pwd PERSONALIZATION
                                                          1042 404k
                                                                      100,000+ Free
                                                                                           Everyone Personalization 5-Aug-14
                                                                                                                          3.0.0
                   osmino Wi-Fi: free
                                                                                                                                4.4 and
           10473
                                          TOOLS
                                                        134203 4.1M 10,000,000+ Free
                                                                                                              7-Aug-18
                                                                                                                        6.06.14
                                                                                                                                2.2 and
           10474
                       Sat-Fi Voice
                                  COMMUNICATION
                                                                                           Everyone Communication 21-Nov-14
                                                            37 14M
                                                                        1,000+ Free
                                                                                                                                2.3 and
                                                                                                                          0.0.9
           10475
                     Wi-Fi Visualizer
                                          TOOLS
                                                          132 2.6M
                                                                       50,000+ Free
                                                                                                          Tools 17-May-17
                                                   3.9
 In [51]: # Again Inspecting using BoxPlot
           google_data.boxplot()
 Out[51]: <matplotlib.axes._subplots.AxesSubplot at 0x119afdd0>
            5.0
           4.5
           4.0
           3.5
           3.0
           2.5
           2.0
           1.5
           1.0
                                 Rating
 In [52]: # Again Inspecting using histogram
           google_data.hist()
 Out[52]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x11B0FEF0>]],
                 dtype=object)
                                  Rating
            4000
            3500
            3000
            2500
            2000
           1500
            1000
            500
                    1.5 2.0 2.5 3.0 3.5 4.0 4.5
           Removing Columns that are 90% empty
 In [53]: # Dropping the Columns which has 10%(0.1)
           threshold = len(google_data) * 0.1
           threshold
 Out[53]: 1084.0
In [54]: google_data.dropna(thresh = threshold, axis=1 , inplace=True)
 In [55]: print(google_data.isnull().sum())
                                0
           App
          Category
                                0
                             1474
           Rating
           Reviews
                                0
           Size
           Installs
           Type
           Price
           Content Rating
           Genres
           Last Updated
           Current Ver
           Android Ver
           dtype: int64
 In [57]: google_data.shape
 Out[57]: (10840, 13)
                    ------ Data Imputation and Manipulation ------ Data Imputation and Manipulation
           Fill the null values with appropriate values using aggregate functions such as mean, median or mode.
 In [59]: # Define a function impute_median
           def impute_median(series):
               return series.fillna(series.median())
 In [60]: google_data.Rating = google_data["Rating"].transform(impute_median)
 In [61]: # Count the number of null values in each colums
           google_data.isnull().sum()
 Out[61]: App
           Category
           Rating
           Reviews
           Size
           Installs
           Type
           Price
           Content Rating
           Genres
           Last Updated
           Current Ver
                             8
                             2
           Android Ver
           dtype: int64
 In [63]: # Modes of Categorical values
           print(google_data["Type"].mode())
           print(google_data["Current Ver"].mode())
           print(google_data["Android Ver"].mode())
           0 Free
           dtype: object
           O Varies with device
           dtype: object
           0 4.1 and up
           dtype: object
 In [64]: # Filling the missing Categorical values with the mode
           google_data["Type"].fillna(str(google_data["Type"].mode().values[0]), inplace=True)
           google_data["Current Ver"].fillna(str(google_data["Current Ver"].mode().values[0]), inplace=True)
           google_data["Android Ver"].fillna(str(google_data["Android Ver"].mode().values[0]), inplace=True)
 In [66]: # Again Counting the number of null values in each column
           google_data.isnull().sum()
 Out[66]: App
           Category
           Rating
           Reviews
           Size
           Installs
           Type
           Price
           Content Rating
           Genres
          Last Updated
           Current Ver
           Android Ver
                             0
           dtype: int64
In [71]: | ###### let's convert Price, Reviews and Ratings into Numerical Values
           google_data["Price"] = google_data["Price"].apply(lambda x: str(x).replace("$" , "") if "$" in str(x) else str(x))
           google_data["Price"] = google_data["Price"].apply(lambda x: float(x))
           google_data["Reviews"] = pd.to_numeric(google_data["Reviews"], errors="coerce")
 In [72]: google_data["Installs"] = google_data["Installs"].apply(lambda x: str(x).replace("+" , "") if "+" in str(x) else str
           google_data["Installs"] = google_data["Installs"].apply(lambda x: str(x).replace("," , "") if "," in str(x) else str
           google_data["Installs"] = google_data["Installs"].apply(lambda x: float(x))
 In [73]: google_data.head(10)
 Out[73]:
                                                                                                                    Last Current Android
                                                                                             Content
                                              Category Rating Reviews Size
                                                                           Installs Type Price
                                                                                                                 Updated
                                                                                                                            Ver
                                                                                                                                   Ver
               Photo Editor & Candy Camera &
                                                                                                                                  4.0.3
                                      ART_AND_DESIGN
                                                        4.1
                                                                159 19M
                                                                                                       Art & Design 7-Jan-18
                                                                                                                           1.0.0
                                                                                         0.0 Everyone
                         Grid & ScrapBook
                                                                                                            Art &
                                                                                                                                  4.0.3
                                                                                                                  15-Jan-
                       Coloring book moana ART_AND_DESIGN
                                                                                                                           2.0.0
                                                        3.9
                                                                967 14M
                                                                          500000.0 Free
                                                                                        0.0 Everyone
                                                                                                     Design;Pretend
                                                                                                                     18
                                                                                                                                 and up
                                                                                                             Play
                 U Launcher Lite – FREE Live ART_AND_DESIGN
                                                                                                                  1-Aug-
18
                                                                                                                                  4.0.3
           2
                                                       4.7
                                                             87510 8.7M 5000000.0 Free 0.0 Everyone
                                                                                                                           1.2.4
                                                                                                       Art & Design
                      Cool Themes, Hide ...
                                                                                                                                 and up
                                                                                                                                4.2 and
           3
                      Sketch - Draw & Paint ART_AND_DESIGN
                                                        4.5 215644 25M 50000000.0 Free
                                                                                                       Art & Design 8-Jun-18
                                                                                                                           with
                                                                                                                                    up
                                                                                                                          device
              Pixel Draw - Number Art Coloring ART_AND_DESIGN
                                                                                                            Art & 20-Jun-
                                                                                                                                4.4 and
                                                                          100000.0 Free 0.0 Everyone
                                                                                                    Design;Creativity
                                                                                                                                2.3 and
                                                                                         0.0 Everyone
                    Paper flowers instructions ART_AND_DESIGN
                                                                167 5.6M
                                                                           50000.0 Free
                                                                                                       Art & Design
                  Smoke Effect Photo Maker - ART_AND_DESIGN
                                                                                                                                  4.0.3
                                                                178 19M
                                                                           50000.0 Free
                                                                                       0.0 Everyone
                                                                                                       Art & Design
                                                                                                                            1.1
                                                                                                                                4.2 and
                                                                                                                  14-Jun-
                           Infinite Painter ART_AND_DESIGN
                                                              36815 29M
                                                                         1000000.0 Free
                                                                                        0.0 Everyone
                                                                                                       Art & Design
                                                                                                                                3.0 and
                                                       4.4 13791 33M 1000000.0 Free 0.0 Everyone
                      Garden Coloring Book ART_AND_DESIGN
                                                                                                       Art & Design
                                                                          10000.0 Free 0.0 Everyone Design; Creativity
                                                                                                                                  4.0.3
                Kids Paint Free - Drawing Fun ART_AND_DESIGN
                                                               121 3.1M
                                                                                                                            2.8
 In [74]: google_data.describe()
 Out[74]:
                                            Installs
                                                         Price
                      Rating
                                Reviews
           count 10840.000000 1.084000e+04 1.084000e+04 10840.000000
                    4.206476  4.441529e+05  1.546434e+07
                                                      1.027368
                                                      15.949703
             std
                    0.480342 2.927761e+06 8.502936e+07
                                                      0.000000
                    1.000000 0.000000e+00 0.000000e+00
            25%
                    4.100000 3.800000e+01 1.000000e+03
                                                      0.000000
                                                      0.000000
            50%
                    4.300000 2.094000e+03 1.000000e+05
                    4.500000 5.477550e+04 5.000000e+06
                                                      0.000000
            75%
                    5.000000 7.815831e+07 1.000000e+09
                                                     400.000000
                                          Data Visualization -----
 In [82]: grp = google_data.groupby('Category')
           x = grp['Rating'].agg(np.mean)
          y = grp['Price'].agg(np.sum)
          z = grp['Reviews'].agg(np.mean)
           print(x)
           print(y)
           print(z)
           Category
                                   4.355385
           ART_AND_DESIGN
           AUTO_AND_VEHICLES
                                   4.205882
           BEAUTY
                                   4.283019
           BOOKS_AND_REFERENCE
                                   4.335498
           BUSINESS
                                   4.182391
           COMICS
                                   4.160000
           COMMUNICATION
                                   4.180103
           DATING
                                   4.025641
           EDUCATION
                                   4.388462
                                   4.126174
           ENTERTAINMENT
           EVENTS
                                   4.395313
           FAMILY
                                   4.204564
           FINANCE
                                   4.151639
          FOOD_AND_DRINK
                                   4.185827
           GAME
                                   4.286888
           HEALTH_AND_FITNESS
                                   4.280059
           HOUSE_AND_HOME
                                   4.211364
           LIBRARIES_AND_DEMO
                                   4.207059
           LIFESTYLE
                                   4.131414
           MAPS_AND_NAVIGATION
                                   4.075182
           MEDICAL
                                   4.216199
           NEWS_AND_MAGAZINES
                                   4.161837
           PARENTING
                                   4.300000
           PERSONALIZATION
                                   4.328827
                                   4.197910
           PHOTOGRAPHY
                                   4.226651
           PRODUCTIVITY
           SHOPPING
                                   4.263077
           SOCIAL
                                   4.261017
           SP0RTS
                                   4.236458
                                   4.080071
           T00LS
           TRAVEL_AND_LOCAL
                                   4.132946
           VIDEO_PLAYERS
                                   4.084000
           WEATHER
                                   4.248780
           Name: Rating, dtype: float64
           Category
           ART_AND_DESIGN
                                      5.97
           AUTO AND VEHICLES
                                     13.47
                                      0.00
           BEAUTY
           BOOKS_AND_REFERENCE
                                    119.77
           BUSINESS
                                    185.27
           COMICS
                                      0.00
          COMMUNICATION
                                     83.14
           DATING
                                     31.43
                                     17.96
           EDUCATION
           ENTERTAINMENT
                                      7.98
           EVENTS
                                    109.99
           FAMILY
                                   2434.78
           FINANCE
                                   2900.83
           FOOD_AND_DRINK
                                      8.48
                                    287.30
           GAME
           HEALTH_AND_FITNESS
                                     67.34
           HOUSE_AND_HOME
                                      0.00
           LIBRARIES_AND_DEMO
                                      0.99
                                   2360.87
           LIFESTYLE
           MAPS_AND_NAVIGATION
                                     26.95
           MEDICAL
                                   1439.96
           NEWS_AND_MAGAZINES
                                      3.98
                                      9.58
           PARENTING
           PERSONALIZATION
                                    153.96
           PHOTOGRAPHY
                                    134.21
           PRODUCTIVITY
                                    250.93
           SHOPPING
                                      5.48
           SOCIAL
                                     15.97
           SP0RTS
                                    100.00
                                    267.25
           T00LS
           TRAVEL_AND_LOCAL
                                     49.95
           VIDEO_PLAYERS
                                     10.46
           WEATHER
                                     32.42
           Name: Price, dtype: float64
           Category
           ART_AND_DESIGN
                                   2.637600e+04
                                   1.369019e+04
           AUTO_AND_VEHICLES
           BEAUTY
                                   7.476226e+03
           BOOKS_AND_REFERENCE
                                   9.506090e+04
           BUSINESS
                                   3.033598e+04
           COMICS
                                   5.638793e+04
           COMMUNICATION
                                   2.107138e+06
           DATING
                                   3.115931e+04
           EDUCATION
                                   2.538191e+05
           ENTERTAINMENT
                                   3.971688e+05
           EVENTS
                                   2.515906e+03
           FAMILY
                                   2.080255e+05
                                   4.795281e+04
           FINANCE
           FOOD_AND_DRINK
                                   6.994748e+04
           GAME
                                   1.385859e+06
          HEALTH_AND_FITNESS
                                   1.111253e+05
           HOUSE_AND_HOME
                                   4.518619e+04
           LIBRARIES_AND_DEMO
                                   1.220139e+04
           LIFESTYLE
                                   3.372457e+04
           MAPS_AND_NAVIGATION
                                   2.237902e+05
           MEDICAL
                                   3.425432e+03
           NEWS_AND_MAGAZINES
                                   1.922292e+05
           PARENTING
                                   1.597218e+04
          PERSONALIZATION
                                   2.279238e+05
           PHOTOGRAPHY
                                   6.373631e+05
           PRODUCTIVITY
                                   2.691438e+05
           SHOPPING
                                   4.424662e+05
           SOCIAL
                                   2.105903e+06
           SP0RTS
                                   1.844536e+05
           T00LS
                                   3.240629e+05
           TRAVEL_AND_LOCAL
                                   2.427051e+05
           VIDEO_PLAYERS
                                   6.307439e+05
                                   1.781065e+05
           WEATHER
           Name: Reviews, dtype: float64
In [112]: plt.figure(figsize = (20,8))
         plt.plot(x , "ro")
           plt.xticks(rotation = 90)
           plt.show()
           4.40 -
           4.35
           4.30
           4.25
           4.20
           4.15
           4.10
           4.05
In [113]: plt.figure(figsize = (20,8))
          plt.plot(x , "ro", color="g")
           plt.xticks(rotation = 90)
           plt.title("Category Wise Rating")
           plt.xlabel("*** Category ***")
          plt.ylabel("*** Rating ***")
           plt.show()
                                                                    Category Wise Rating
             4.35
             4.25
           # 4.20
             4.15
In [118]: plt.figure(figsize = (20,8))
           plt.plot(y , "r--", color="b")
           plt.xticks(rotation = 90)
           plt.title("Category Wise Rating")
           plt.xlabel("*** Category ***")
          plt.ylabel("*** Rating ***")
           plt.show()
                                                                    Category Wise Rating
             2500
             2000
```

1000

In [117]: plt.figure(figsize = (20,8))

plt.show()

2000000

1500000

₹ 1000000

500000

In []:

plt.plot(z , "x", color="r")
plt.xticks(rotation = 90)

plt.title("Category Wise Rating")
plt.xlabel("*** Category ***")
plt.ylabel("*** Rating ***")

Category Wise Rating

In [123]: google_data.to_csv("D:\Softwares\Data Analyst & Science Software\Data Science Projects\Datasets\Google App Rating Da

ta Analysis Master\ipynb_checkpoints\googleplaystore_cleaned.csv" , index=False)