EDA - PLAYSTORE ANALYSIS

In [179…
```
import pandas as pd
import numpy as np
```

In [180…
```
df = pd.read_csv("F:\\PROJECTS\\DSC & DA\\EDA PLAYSTORE ANALYSIS\\googleplaystor
```

In [181…
```
pd.set_option("display.max_columns",None)
```

In [182…
```
df
```

Out[182...

	App	Category	Rating	Reviews	Size	Installs	Type
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free
...	...	...	...	...	...	...	...
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free

10841 rows × 13 columns

1. Find Total Number Apps in Google Play Store

In [183…
```
df.shape
print("Total Number Of Apps In Google Play Store = ", df.shape[0])
```

```
Total Number Of Apps In Google Play Store =  10841
```

2. Find the Total Number of Columns in Each app of Google Play Store

In [184…
```
df.shape
print("Total Number Of Columns = ", df.shape[1])
```

```
Total Number Of Columns =  13
```

---

1. Display Top 5 Rows of The Dataset

In [185…
```
print("Top 5 Rows Of Data-Set = ", df.head())
```

```
Top 5 Rows Of Data-Set =                                                 App
Category  Rating  \
0     Photo Editor & Candy Camera & Grid & ScrapBook  ART_AND_DESIGN     4.1
1                                Coloring book moana  ART_AND_DESIGN     3.9
2  U Launcher Lite – FREE Live Cool Themes, Hide ...  ART_AND_DESIGN     4.7
3                              Sketch - Draw & Paint  ART_AND_DESIGN     4.5
4              Pixel Draw - Number Art Coloring Book  ART_AND_DESIGN     4.3

  Reviews  Size     Installs  Type Price Content Rating  \
0     159   19M      10,000+  Free     0       Everyone
1     967   14M     500,000+  Free     0       Everyone
2   87510  8.7M   5,000,000+  Free     0       Everyone
3  215644   25M  50,000,000+  Free     0           Teen
4     967  2.8M     100,000+  Free     0       Everyone

                      Genres      Last Updated         Current Ver  \
0               Art & Design   January 7, 2018               1.0.0
1  Art & Design;Pretend Play  January 15, 2018               2.0.0
2               Art & Design    August 1, 2018               1.2.4
3               Art & Design      June 8, 2018  Varies with device
4    Art & Design;Creativity     June 20, 2018                 1.1

    Android Ver
0  4.0.3 and up
1  4.0.3 and up
2  4.0.3 and up
3    4.2 and up
4    4.4 and up
```

2. Check the Last 3 Rows of The Dataset

In [186…
```
print("Last 3 Rows Of Data-set = ", df.tail(3))
```

```
Last 3 Rows Of Data-set =                                                App
Category  \
10838                         Parkinson Exercices FR              MEDICAL
10839                  The SCP Foundation DB fr nn5n  BOOKS_AND_REFERENCE
10840  iHoroscope - 2018 Daily Horoscope & Astrology            LIFESTYLE

       Rating Reviews                Size     Installs  Type Price  \
10838     NaN       3                9.5M       1,000+  Free     0
10839     4.5     114  Varies with device       1,000+  Free     0
10840     4.5  398307                 19M  10,000,000+  Free     0

      Content Rating             Genres      Last Updated         Current Ver  \
10838       Everyone            Medical  January 20, 2017                 1.0
10839     Mature 17+  Books & Reference  January 19, 2015  Varies with device
10840       Everyone          Lifestyle     July 25, 2018  Varies with device

             Android Ver
10838         2.2 and up
10839  Varies with device
10840  Varies with device
```

3. Find Shape of Our Dataset (Number of Rows & Number of Columns)

In [187…
```
print("Shape Of Data-set Rows = {}, Cols = {}".format(df.shape[0], df.shape[1]))
```

```
Shape Of Data-set Rows = 10841, Cols = 13
```

4. Get Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement

In [188…
```
print("-----Information About Dataset-----")
print(df.info())
```

```
-----Information About Dataset-----
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10841 entries, 0 to 10840
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   App             10841 non-null  object
 1   Category        10841 non-null  object
 2   Rating          9367 non-null   float64
 3   Reviews         10841 non-null  object
 4   Size            10841 non-null  object
 5   Installs        10841 non-null  object
 6   Type            10840 non-null  object
 7   Price           10841 non-null  object
 8   Content Rating  10840 non-null  object
 9   Genres          10841 non-null  object
 10  Last Updated    10841 non-null  object
 11  Current Ver     10833 non-null  object
 12  Android Ver     10838 non-null  object
dtypes: float64(1), object(12)
memory usage: 1.1+ MB
None
```

5. Get Overall Statistics About The Dataframe

In [189…
```
print("----------Overall Stastics-----------")
print(df.describe(include='all'))
```

```
----------Overall Stastics-----------
           App Category       Rating Reviews                Size    Installs  \
count    10841    10841  9367.000000   10841               10841       10841
unique    9660       34          NaN    6002                 462          22
top     ROBLOX   FAMILY          NaN       0  Varies with device  1,000,000+
freq         9     1972          NaN     596                1695        1579
mean       NaN      NaN     4.193338     NaN                 NaN         NaN
std        NaN      NaN     0.537431     NaN                 NaN         NaN
min        NaN      NaN     1.000000     NaN                 NaN         NaN
25%        NaN      NaN     4.000000     NaN                 NaN         NaN
50%        NaN      NaN     4.300000     NaN                 NaN         NaN
75%        NaN      NaN     4.500000     NaN                 NaN         NaN
max        NaN      NaN    19.000000     NaN                 NaN         NaN

          Type  Price Content Rating Genres    Last Updated  \
count    10840  10841          10840  10841           10841
unique       3     93              6    120            1378
top       Free      0       Everyone  Tools  August 3, 2018
freq     10039  10040           8714    842             326
mean       NaN    NaN            NaN    NaN             NaN
std        NaN    NaN            NaN    NaN             NaN
min        NaN    NaN            NaN    NaN             NaN
25%        NaN    NaN            NaN    NaN             NaN
50%        NaN    NaN            NaN    NaN             NaN
75%        NaN    NaN            NaN    NaN             NaN
max        NaN    NaN            NaN    NaN             NaN

               Current Ver Android Ver
count                10833       10838
unique                2832          33
top     Varies with device  4.1 and up
freq                  1459        2451
mean                   NaN         NaN
std                    NaN         NaN
min                    NaN         NaN
25%                    NaN         NaN
50%                    NaN         NaN
75%                    NaN         NaN
max                    NaN         NaN
```

6. Total Number of App Titles Contain Astrology

In [190…
```
print("----------Apps Contains Astrology----------")
print(df[df['App'].str.contains('Astrology', case=False)])
no = len(df[df['App'].str.contains('Astrology', case=False)])
print("Total Number = {}".format(no))
```

```
----------Apps Contains Astrology----------
                                                    App   Category  Rating  \
1570   Horoscopes – Daily Zodiac Horoscope and Astrology  LIFESTYLE     4.6
1592                ဘဝ Astrology - Min Thein Kha BayDin  LIFESTYLE     4.7
10840      iHoroscope - 2018 Daily Horoscope & Astrology  LIFESTYLE     4.5

      Reviews Size     Installs  Type Price Content Rating     Genres  \
1570   161143  11M  10,000,000+  Free     0   Everyone 10+  Lifestyle
1592     2225  15M     100,000+  Free     0       Everyone  Lifestyle
10840  398307  19M  10,000,000+  Free     0       Everyone  Lifestyle

        Last Updated         Current Ver         Android Ver
1570   June 25, 2018          5.2.4(881)        4.0.3 and up
1592   July 26, 2018               4.2.1        4.0.3 and up
10840  July 25, 2018  Varies with device  Varies with device
Total Number = 3
```

7. Find Average App Rating

In [191…

```
print("Average Rating App = {}".format(round(df['Rating'].mean(),1)))
```

```
Average Rating App = 4.2
```

8. Find Total Number of Unique Category

In [192…

```
print("Total Number Of Unique Category = {}".format(df['Category'].nunique()))
```

```
Total Number Of Unique Category = 34
```

Displaying Unique 'Category Names'

In [193…

```
for name in df['Category'].unique():
    print(name)
```

```
ART_AND_DESIGN
AUTO_AND_VEHICLES
BEAUTY
BOOKS_AND_REFERENCE
BUSINESS
COMICS
COMMUNICATION
DATING
EDUCATION
ENTERTAINMENT
EVENTS
FINANCE
FOOD_AND_DRINK
HEALTH_AND_FITNESS
HOUSE_AND_HOME
LIBRARIES_AND_DEMO
LIFESTYLE
GAME
FAMILY
MEDICAL
SOCIAL
SHOPPING
PHOTOGRAPHY
SPORTS
TRAVEL_AND_LOCAL
TOOLS
PERSONALIZATION
PRODUCTIVITY
PARENTING
WEATHER
VIDEO_PLAYERS
NEWS_AND_MAGAZINES
MAPS_AND_NAVIGATION
1.9
```

9. Which Category Getting The Highest Average Rating?

In [194…
```
print("------Category Which Have Highest Rating------")
print(df.groupby("Category")['Rating'].mean().sort_values(ascending=False).head(
```

```
------Category Which Have Highest Rating------
Category
1.9    19.0
Name: Rating, dtype: float64
```

OR

In [195…
```
print("------Category Which Have Highest Rating------")
df.groupby("Category")['Rating'].mean().sort_values().tail(1)
```

```
------Category Which Have Highest Rating------
```

Out[195…
```
Category
1.9    19.0
Name: Rating, dtype: float64
```

10. Find Total Number of App having 5 Star Rating

In [196…
```
print("Number Of Apps Having 5 Star Ratings = {}".format((df.loc[:,['Rating']]=
```

```
Number Of Apps Having 5 Star Ratings = Rating    274
dtype: int64
```

OR

In [197…
```
print("Number Of Apps Having 5 Star Ratings = {}".format(len(df[df['Rating']==5]
```

```
Number Of Apps Having 5 Star Ratings = 274
```

11. Find Average Value of Reviews

In [198…
```
df['Reviews']=df['Reviews'].replace("3.0M", 3.0*1000000)
```

In [199…
```
df['Reviews']=df['Reviews'].astype("float")
```

In [200…
```
print("Average Values Of Reviews = {}".format(round(df['Reviews'].mean(), 2)))
```

```
Average Values Of Reviews = 444388.65
```

12. Find Total Number of Free and Paid Apps

In [201…
```
print("Total Number Of Free And Paid Apps {}".format(df['Type'].value_counts()))
```

```
Total Number Of Free And Paid Apps Type
Free    10039
Paid      800
0           1
Name: count, dtype: int64
```

In [202…
```
print("Total Number Of Free And Paid Apps {}".format(df.groupby("Type").agg('cou
```

```
Total Number Of Free And Paid Apps Type
0           1
Free    10039
Paid      800
Name: App, dtype: int64
```

13. Which App Has Maximum Reviews?

In [203…
```
mreview=df[df["Reviews"].max()==df["Reviews"]]["App"]
print("App Having Maximum Reviews = {}".format(mreview))
```

```
App Having Maximum Reviews = 2544    Facebook
Name: App, dtype: object
```

14. Display Top 5 Apps Having Highest Reviews

In [204…
```
index = df['Reviews'].sort_values(ascending=False).head().index
df.loc[index,['App']]
```

Out[204…

	App
2544	Facebook
3943	Facebook
381	WhatsApp Messenger
336	WhatsApp Messenger
3904	WhatsApp Messenger

15. Find Average Rating of Free and Paid Apps

In [205…
```
df.groupby("Type")["Rating"].mean()
```

Out[205…
```
Type
0       19.000000
Free     4.186203
Paid     4.266615
Name: Rating, dtype: float64
```

16. Display Top 5 Apps Having Maximum Installs

In [218…
```
df["Installss"]=df["Installs"].str.replace(",","")
```

In [219…
```
df['Installss']=df['Installss'].str.replace("+","")
```

In [220…
```
df['Installss']
```

Out[220…
```
0           10000
1          500000
2         5000000
3        50000000
4          100000
            ...
10836        5000
10837         100
10838        1000
10839        1000
10840    10000000
Name: Installss, Length: 10841, dtype: object
```

In [221…
```
df['Installss'].unique()
```

Out[221…
```
array(['10000', '500000', '5000000', '50000000', '100000', '50000',
       '1000000', '10000000', '5000', '100000000', '1000000000', '1000',
       '500000000', '50', '100', '500', '10', '1', '5', '0', 'Free'],
      dtype=object)
```

In [222…
```
df["Installss"]=df["Installss"].str.replace("Free","0")
```

In [223…
```
df['Installss'].unique()
```

Out[223…
```
array(['10000', '500000', '5000000', '50000000', '100000', '50000',
       '1000000', '10000000', '5000', '100000000', '1000000000', '1000',
       '500000000', '50', '100', '500', '10', '1', '5', '0'], dtype=object)
```

In [224…
```
df["Installss"]=df["Installss"].astype("int")
```

In [225…
```
df["Installss"].dtype
```

Out[225…
```
dtype('int64')
```

In [226…
```
indices=df["Installss"].sort_values(ascending=False).head().index
df.loc[indices]["App"]
```

Out[226…
```
5856    Google Play Games
5395        Google Photos
2853        Google Photos
2884        Google Photos
4170         Google Drive
Name: App, dtype: object
```

In [ ]: