Google_PlayStore_QUESTIONS/ANSWERED_BASED_EDA_Project_VIVEK_CHAUHAN

In [1]:
```
# import necessary libraries to work with dataset

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
# forward / is used for path.
data=pd.read_csv("C:/Users/VIVEK CHAUHAN/Desktop/eda-projects (1)/4-EDA Problem Statement (1)/4-googleplaystore (1).csv")
data
```

Out[2]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Curren Ve
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varie wit devic
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.
...	...	...	...	...	...	...	...	...	...	...	...	...
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0	Everyone	Education	July 25, 2017	1.4
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.
10838	Parkinson Exercices	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Curren Ve
	FR											
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varie wit devic
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varie wit devic

10841 rows × 13 columns

In [3]:
```
# display top 5 rows of the dataset

data.head(5)
```

Out[3]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

In [4]:
```
# check last 3 rows of the dataset

data.tail(3)
```

Out[4]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	An
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0	Everyone	Medical	January 20, 2017	1.0	2.
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0	Mature 17+	Books & Reference	January 19, 2015	Varies with device	
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	

In [5]:
```
# find the shapes of our datasets like numbers of rows and columns

data.shape
```

Out[5]: (10841, 13)

In [6]:
```
# print the number of rows

print("Number of rows",data.shape[0])
```

```
Number of rows 10841
```

In [7]:
```
# print the number of columns

print("Number of columns",data.shape[1])
```

```
Number of columns 13
```

In [8]:
```
# Get Information about our dataset like total number of rows,
# total number of columns, datatypes of each column & memory requirement
```

```
data.info()
```

```
<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
```

In [9]:
```
#  get overall statistics about the dataframe

data.describe(include="all")
```

Out[9]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
count	10841	10841	9367.000000	10841	10841	10841	10840	10841	10840	10841	10841	10833	10838
unique	9660	34	NaN	6002	462	22	3	93	6	120	1378	2832	33
top	ROBLOX	FAMILY	NaN	0	Varies with device	1,000,000+	Free	0	Everyone	Tools	August 3, 2018	Varies with device	4.1 and up
freq	9	1972	NaN	596	1695	1579	10039	10040	8714	842	326	1459	2451
mean	NaN	NaN	4.193338	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	0.537431	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	4.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	4.300000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	4.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	19.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [10]:
```
# total number of apps contains astrology case for upper & lower case both

data[data["App"].str.contains("Astrology",case=False)]
```

Out[10]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
1570	Horoscopes – Daily Zodiac Horoscope and Astrology	LIFESTYLE	4.6	161143	11M	10,000,000+	Free	0	Everyone 10+	Lifestyle	June 25, 2018	5.2.4(881)	4.0.3 and up
1592	ဘိုး Astrology - Min Thein Kha BayDin	LIFESTYLE	4.7	2225	15M	100,000+	Free	0	Everyone	Lifestyle	July 26, 2018	4.2.1	4.0.3 and up
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0	Everyone	Lifestyle	July 25, 2018	Varies with device	Varies with device

In [11]:
```
# total number of apps contains astrology

len(data[data["App"].str.contains("Astrology",case=False)])
```

Out[11]: 3

In [12]:
```
# find average app ratings

data["Rating"].mean()
```

Out[12]: 4.193338315362443

In [13]:
```
# find total number of unique category
# unique function will bring all the unique category as text value in the data
# nunique brings count of unique data in the dataframe

data["Category"].nunique()
```

Out[13]: 34

In [14]:
```
# Which Category Getting the highest Average Rating ?

data.groupby("Category")["Rating"].mean()
```

Out[14]:
```
Category
1.9                    19.000000
ART_AND_DESIGN          4.358065
AUTO_AND_VEHICLES       4.190411
BEAUTY                  4.278571
BOOKS_AND_REFERENCE     4.346067
BUSINESS                4.121452
COMICS                  4.155172
COMMUNICATION           4.158537
DATING                  3.970769
EDUCATION               4.389032
ENTERTAINMENT           4.126174
EVENTS                  4.435556
FAMILY                  4.192272
FINANCE                 4.131889
FOOD_AND_DRINK          4.166972
GAME                    4.286326
HEALTH_AND_FITNESS      4.277104
HOUSE_AND_HOME          4.197368
LIBRARIES_AND_DEMO      4.178462
LIFESTYLE               4.094904
MAPS_AND_NAVIGATION     4.051613
MEDICAL                 4.189143
NEWS_AND_MAGAZINES      4.132189
PARENTING               4.300000
PERSONALIZATION         4.335987
PHOTOGRAPHY             4.192114
PRODUCTIVITY            4.211396
SHOPPING                4.259664
SOCIAL                  4.255598
SPORTS                  4.223511
TOOLS                   4.047411
TRAVEL_AND_LOCAL        4.109292
VIDEO_PLAYERS           4.063750
WEATHER                 4.244000
Name: Rating, dtype: float64
```

In [15]:
```
# let's find which value has error 3.0m

data[data["Reviews"]=="3.0M"]
```

Out[15]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10472	Life Made WI-Fi Touchscreen Photo Frame	1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN	February 11, 2018	1.0.19	4.0 and up	NaN

In [16]:
```
# let's replace the value 3.0M to 3.0 for remove an error

data["Reviews"] = data["Reviews"].replace("3.0M",3.0)
```

In [17]:
```
# print all the column names which is present in the dataset

data.columns
```

Out[17]:
```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

In [18]:
```
# now cross check reviews column datatype

data['Reviews'].dtypes
```

Out[18]:
```
dtype('O')
```

In [19]:
```
# we need to change the object datatype to float datatype for reviews column

data['Reviews'] = data['Reviews'].astype("float")
```

In [20]:
```
# now again check the datatype for reviews column

data["Reviews"].dtypes
```

Out[20]:
```
dtype('float64')
```

In [21]:
```
# Find Average Value of Reviews

data["Reviews"].mean()
```

Out[21]:
```
444111.9265750392
```

In [22]:
```
# Find Total number of Apps having 5 Star Rating

data[data["Rating"]==5.0]
```

Out[22]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
329	Hojiboy Tojiboyev Life Hacks	COMICS	5.0	15.0	37M	1,000+	Free	0	Everyone	Comics	June 26, 2018	2.0	4.0.3 and up
612	American Girls Mobile Numbers	DATING	5.0	5.0	4.4M	1,000+	Free	0	Mature 17+	Dating	July 17, 2018	3.0	4.0.3 and up
615	Awake Dating	DATING	5.0	2.0	70M	100+	Free	0	Mature 17+	Dating	July 24, 2018	2.2.9	4.4 and up
633	Spine- The dating app	DATING	5.0	5.0	9.3M	500+	Free	0	Teen	Dating	July 14, 2018	4.0	4.0.3 and up
636	Girls Live Talk - Free Text and Video Chat	DATING	5.0	6.0	5.0M	100+	Free	0	Mature 17+	Dating	August 1, 2018	8.2	4.0.3 and up
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10721	Mad Dash Fo' Cash	GAME	5.0	14.0	16M	100+	Free	0	Everyone	Arcade	June 19, 2017	2.5a	4.1 and up
10742	GKPB FP Online Church	LIFESTYLE	5.0	32.0	7.9M	1,000+	Free	0	Everyone	Lifestyle	December 31, 2017	0.7.1	4.4 and up
10776	Monster Ride Pro	GAME	5.0	1.0	24M	10+	Free	0	Everyone	Racing	March 5, 2018	2.0	2.3 and up
10820	Fr. Daoud Lamei	FAMILY	5.0	22.0	8.6M	1,000+	Free	0	Teen	Education	June 27, 2018	3.8.0	4.1 and up
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4.0	3.6M	100+	Free	0	Everyone	Education	July 6, 2018	1.0	4.1 and up

274 rows × 13 columns

In [23]:
```
# Find Total number of Apps having 5 Star Rating

len(data[data["Rating"]==5.0])
```

Out[23]:
```
274
```

In [24]:
```
# other method to count the values of free and paid type of the apps

data["Type"].value_counts()
```

Out[24]:
```
Type
Free    10039
Paid      800
0           1
Name: count, dtype: int64
```

In [25]:
```
# Find Total Numbers of Free & Paid Apps

type1 = data["Type"]=="Free"
type1.value_counts()
```

Out[25]:
```
Type
True     10039
False      802
Name: count, dtype: int64
```

In [26]:
```
# other method to get which app has maximum reviews?
# we are comparing our max values with all the other data values in our dataset

data[data["Reviews"].max()==data["Reviews"]] # this query return the all the columns data
```

Out[26]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
2544	Facebook	SOCIAL	4.1	78158306.0	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device	Varies with device

In [27]:
```
# other method to get which app has maximum reviews?
# we are comparing our max values with all the other data values in our dataset
# we need only app like which app has max reviews

data[data["Reviews"].max()==data["Reviews"]]["App"]
```

Out[27]:
```
2544    Facebook
Name: App, dtype: object
```

In [28]:
```
# Which App has Maximum Reviews ?

a = data["Reviews"].max()
b = data["App"][data["Reviews"]==a]
b
```

Out[28]:
```
2544    Facebook
Name: App, dtype: object
```

In [29]:
```
# other method to find the top 5 apps having highest reviews

index = data["Reviews"].sort_values(ascending=False).head().index
print("this is the index number of all the highes reviews apps:",index)
data.iloc[index]["App"]
```

```
this is the index number of all the highes reviews apps: Index([2544, 3943, 381, 336, 3904], dtype='int64')
```

Out[29]:
```
2544              Facebook
3943              Facebook
381     WhatsApp Messenger
336     WhatsApp Messenger
3904    WhatsApp Messenger
Name: App, dtype: object
```

In [30]:
```
# this query is return all the columns data where the index is match

data.iloc[index]
```

Out[30]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver
2544	Facebook	SOCIAL	4.1	78158306.0	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device
3943	Facebook	SOCIAL	4.1	78128208.0	Varies with device	1,000,000,000+	Free	0	Teen	Social	August 3, 2018	Varies with device
381	WhatsApp Messenger	COMMUNICATION	4.4	69119316.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 3, 2018	Varies with device
336	WhatsApp Messenger	COMMUNICATION	4.4	69119316.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 3, 2018	Varies with device
3904	WhatsApp Messenger	COMMUNICATION	4.4	69109672.0	Varies with device	1,000,000,000+	Free	0	Everyone	Communication	August 3, 2018	Varies with device

In [31]:
```
# this query is return only index numbers

data.iloc[index].index
```

Out[31]:
```
Index([2544, 3943, 381, 336, 3904], dtype='int64')
```

In [32]:
```
# Display Top 5 Apps having Highest Reviews

index = data["Reviews"].sort_values(ascending=False).head().index
data.iloc[index]["App"]
```

Out[32]:
```
2544              Facebook
3943              Facebook
381     WhatsApp Messenger
336     WhatsApp Messenger
3904    WhatsApp Messenger
Name: App, dtype: object
```

In [33]:
```
# other method to find the average rating of free and paid apps

data.groupby("Type")["Rating"].mean()
```

Out[33]:
```
Type
0       19.000000
Free     4.186203
Paid     4.266615
Name: Rating, dtype: float64
```

In [34]:
```
# Find Average Rating of Free & Paid Apps

a = data["Type"]=="Free"
b = data["Rating"]
ans = b.where(a)
ans.mean()
```

Out[34]:
```
4.186202546163551
```

In [35]:
```
# Find Average Rating of Free & Paid Apps

a = data["Type"]=="Paid"
b = data["Rating"]
ans = b.where(a)
ans.mean()
```

Out[35]:
```
4.26661514683153
```