

# Google Play Store Analysis Using Python

The Google Play Store set has the information about the Apps, Rating & Downloads .

The Data set available from Flexible which is a Third Party Google Play Store which engine , and available on Kaggle dataset for free.

## Import Library

```
In [1]: import pandas as pd
```

```
In [2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\Syed Arif\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.25.1)
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}")
```

## Uploading Csv file

```
In [3]: df = pd.read_csv(r"C:\Users\Syed Arif\Downloads\Compressed\googleplaystore.csv")
```

## Data Preprocessing

### .head()

head is used show to the By default = 5 rows in the dataset

```
In [4]: df.head()
```

Out[4]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	D
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	De

**.tail()**

tail is used to show rows by Descending order

In [5]: `df.tail()`

Out[5]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price
10836	Sya9a Maroc - FR	FAMILY	4.5	38	53M	5,000+	Free	0
10837	Fr. Mike Schmitz Audio Teachings	FAMILY	5.0	4	3.6M	100+	Free	0
10838	Parkinson Exercices FR	MEDICAL	NaN	3	9.5M	1,000+	Free	0
10839	The SCP Foundation DB fr nn5n	BOOKS_AND_REFERENCE	4.5	114	Varies with device	1,000+	Free	0
10840	iHoroscope - 2018 Daily Horoscope & Astrology	LIFESTYLE	4.5	398307	19M	10,000,000+	Free	0

## .shape

It show the total no of rows & Column in the dataset

In [6]: `df.shape`

Out[6]: (10841, 13)

## .Columns

It show the no of each Column

In [7]: `df.columns`

Out[7]: Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',  
'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',  
'Android Ver'],  
dtype='object')

## .dtypes

This Attribute show the data type of each column

```
In [8]: df.dtypes
```

```
Out[8]: App                object
Category                object
Rating                 float64
Reviews                object
Size                   object
Installs                object
Type                   object
Price                  object
Content Rating          object
Genres                  object
Last Updated            object
Current Ver              object
Android Ver              object
dtype: object
```

## .unique()

In a column, It show the unique value of specific column.

```
In [9]: df["Category"].unique()
```

```
Out[9]: array(['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'], dtype=object)
```

## .nuique()

It will show the total no of unque value from whole data frame

```
In [10]: df.nunique()
```

```
Out[10]: App          9660  
Category      34  
Rating        40  
Reviews       6002  
Size          462  
Installs      22  
Type          3  
Price         93  
Content Rating 6  
Genres        120  
Last Updated  1378  
Current Ver   2832  
Android Ver   33  
dtype: int64
```

## .describe()

It show the Count, mean , median etc

```
In [11]: df.describe()
```

```
Out[11]:
```

	Rating
<b>count</b>	9367.000000
<b>mean</b>	4.193338
<b>std</b>	0.537431
<b>min</b>	1.000000
<b>25%</b>	4.000000
<b>50%</b>	4.300000
<b>75%</b>	4.500000
<b>max</b>	19.000000

## .value\_counts

It Shows all the unique values with their count

```
In [12]: df["Category"].value_counts()
```

```
Out[12]: FAMILY                1972
GAME                1144
TOOLS                843
MEDICAL             463
BUSINESS            460
PRODUCTIVITY        424
PERSONALIZATION     392
COMMUNICATION        387
SPORTS              384
LIFESTYLE           382
FINANCE             366
HEALTH_AND_FITNESS  341
PHOTOGRAPHY         335
SOCIAL              295
NEWS_AND_MAGAZINES  283
SHOPPING            260
TRAVEL_AND_LOCAL    258
DATING              234
BOOKS_AND_REFERENCE 231
VIDEO_PLAYERS       175
EDUCATION           156
ENTERTAINMENT       149
MAPS_AND_NAVIGATION 137
FOOD_AND_DRINK      127
HOUSE_AND_HOME       88
LIBRARIES_AND_DEMO   85
AUTO_AND_VEHICLES    85
WEATHER              82
ART_AND_DESIGN       65
EVENTS              64
PARENTING            60
COMICS              60
BEAUTY              53
1.9                  1
Name: Category, dtype: int64
```

## **.isnull()**

It shows the how many null values

```
In [13]: df.isnull()
```

```
Out[13]:
```

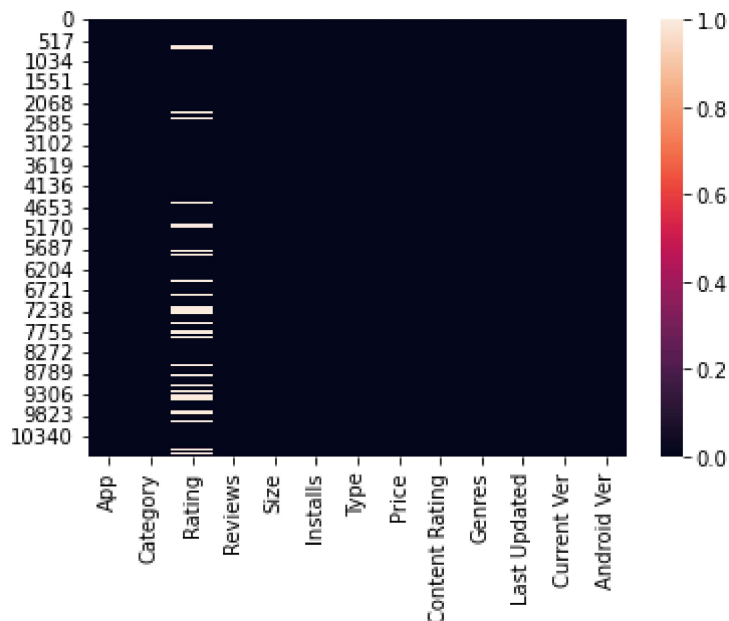
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...
10836	False	False	False	False	False	False	False	False	False	False	False
10837	False	False	False	False	False	False	False	False	False	False	False
10838	False	False	True	False	False	False	False	False	False	False	False
10839	False	False	False	False	False	False	False	False	False	False	False
10840	False	False	False	False	False	False	False	False	False	False	False

10841 rows × 13 columns



## How many Null value present show all the null values in Heatmap.

```
In [14]: sns.heatmap(df.isnull())
plt.show()
```



# How many Categories are in the dataset

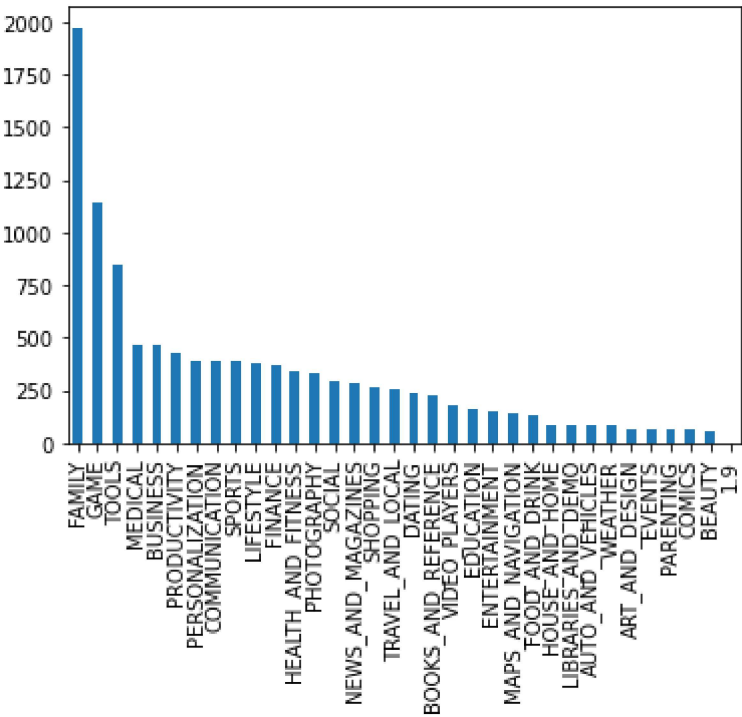
```
In [15]: df.Category.value_counts()
```

```
Out[15]: FAMILY                1972  
GAME                1144  
TOOLS               843  
MEDICAL             463  
BUSINESS            460  
PRODUCTIVITY        424  
PERSONALIZATION     392  
COMMUNICATION       387  
SPORTS              384  
LIFESTYLE            382  
FINANCE             366  
HEALTH_AND_FITNESS  341  
PHOTOGRAPHY         335  
SOCIAL              295  
NEWS_AND_MAGAZINES  283  
SHOPPING            260  
TRAVEL_AND_LOCAL    258  
DATING              234  
BOOKS_AND_REFERENCE 231  
VIDEO_PLAYERS       175  
EDUCATION           156  
ENTERTAINMENT       149  
MAPS_AND_NAVIGATION 137  
FOOD_AND_DRINK      127  
HOUSE_AND_HOME      88  
LIBRARIES_AND_DEMO  85  
AUTO_AND_VEHICLES   85  
WEATHER              82  
ART_AND_DESIGN       65  
EVENTS               64  
PARENTING            60  
COMICS               60  
BEAUTY               53  
1.9                  1  
Name: Category, dtype: int64
```



```
In [16]: df.Category.value_counts().plot(kind = "bar")
```

Out[16]: <AxesSubplot:>



## Show all records Where Rating is "5.0"

```
In [19]: df.sort_values(by = 'Rating', ascending = False).head()
```

Out[19]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
10472	Life Made WI-Fi Touchscreen Photo Frame	1.9	19.0	3.0M	1,000+	Free	0	Everyone	NaN
9511	Ek Bander Ne Kholi Dukan	FAMILY	5.0	10	3.0M	10,000+	Free	0	Everyone
10166	FA Player Essentials	SPORTS	5.0	7	68M	100+	Free	0	Everyone
7895	Dine In CT - Food Delivery	SHOPPING	5.0	4	1.6M	1,000+	Free	0	Everyone
5118	Eternal Light AG	SOCIAL	5.0	30	13M	100+	Free	0	Teen

# Show all record where rating less then or equal to 2

In [21]:

df[(df.Rating <= 2.0)]

Out[21]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
518	Just She - Top Lesbian Dating	DATING	1.9	953	19M	100,000+	Free	0	Mature 17+
554	Just She - Top Lesbian Dating	DATING	1.9	953	19M	100,000+	Free	0	Mature 17+
625	House party - live chat	DATING	1.0	1	9.2M	10+	Free	0	Mature 17+
663	Just She - Top Lesbian Dating	DATING	1.9	954	19M	100,000+	Free	0	Mature 17+
4006	C by GE	LIFESTYLE	1.7	686	6.6M	10,000+	Free	0	Everyone
...	...	...	...	...	...	...	...	...	...
10324	FE Mechanical Engineering Prep	FAMILY	1.0	2	21M	1,000+	Free	0	Everyone
10400	Familial Hypercholesterolaemia Handbook	MEDICAL	1.0	2	33M	100+	Free	0	Everyone
10562	FK Atlantas	SPORTS	1.5	2	26M	5+	Free	0	Everyone
10578	Check Lottery Tickets - Florida	FAMILY	2.0	27	31M	5,000+	Free	0	Teen
10591	Lottery Ticket Checker - Florida Results & Lotto	TOOLS	1.0	3	41M	500+	Free	0	Everyone

68 rows × 13 columns

# Show all the records where Category = "Dating"

```
In [25]: df[df['Category'] == "DATING"].head(3)
```

```
Out[25]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Up
476	Moco+ - Chat, Meet People	DATING	4.2	1545	Varies with device	10,000+	Paid	\$3.99	Mature 17+	Dating	Ju
477	Calculator	DATING	2.6	57	6.2M	1,000+	Paid	\$6.99	Everyone	Dating	O 25
478	Truth or Dare Pro	DATING	NaN	0	20M	50+	Paid	\$1.49	Teen	Dating	Sept 1

## check the Missing Values in our dataset

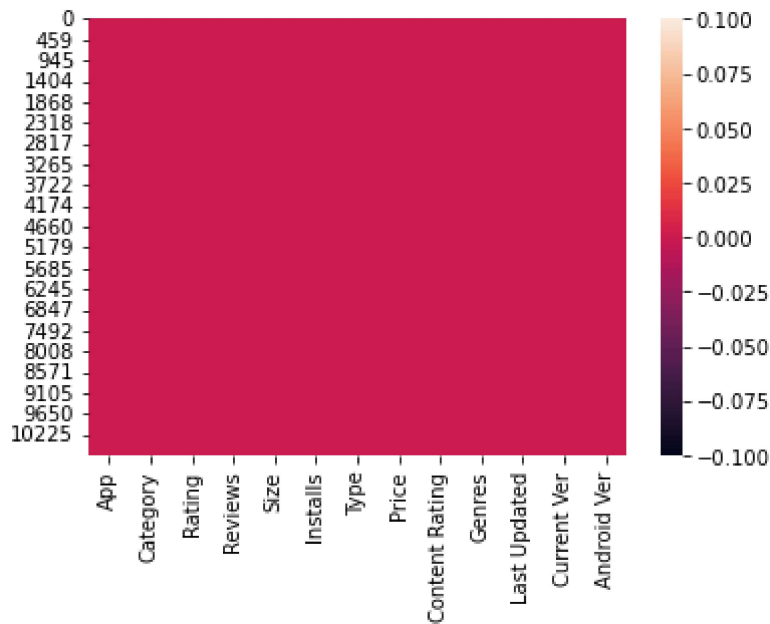
```
In [26]: df.isnull().sum()
```

```
Out[26]: App                0
Category                0
Rating                1474
Reviews                0
Size                  0
Installs              0
Type                  1
Price                0
Content Rating        1
Genres                0
Last Updated          0
Current Ver           8
Android Ver           3
dtype: int64
```

```
In [27]: df.dropna(axis = 0, inplace = True)
```

```
In [28]: sns.heatmap(df.isnull())
```

```
Out[28]: <AxesSubplot:>
```



## Show all record

```
In [37]: df[df['Category'] == "DATING"].max()
```

```
Out[37]: App                stranger chat - anonymous chat
Category                DATING
Rating                    5.0
Reviews                97699
Size                Varies with device
Installs                500,000+
Type                    Paid
Price                    0
Content Rating                Teen
Genres                    Dating
Last Updated                October 25, 2017
Current Ver                Varies with device
Android Ver                Varies with device
dtype: object
```

## Show all records of category by max review

```
In [42]: df.groupby('Category').Reviews.max()
```

```
Out[42]: Category
ART_AND_DESIGN          967
AUTO_AND_VEHICLES       994
BEAUTY                   964
BOOKS_AND_REFERENCE      98
BUSINESS                 984451
COMICS                   9952
COMMUNICATION            99559
DATING                   97699
EDUCATION                99020
ENTERTAINMENT            9941
EVENTS                   99
FAMILY                   997
FINANCE                  967
FOOD_AND_DRINK           91359
GAME                     9951
HEALTH_AND_FITNESS       979
HOUSE_AND_HOME           874
LIBRARIES_AND_DEMO       929
LIFESTYLE                987
MAPS_AND_NAVIGATION       96
MEDICAL                  9945
NEWS_AND_MAGAZINES       978
PARENTING                 99
PERSONALIZATION          98
PHOTOGRAPHY              98717
PRODUCTIVITY             9914
SHOPPING                 9975
SOCIAL                   9606
SPORTS                   9992
TOOLS                    999
TRAVEL_AND_LOCAL         9971
VIDEO_PLAYERS            9966
WEATHER                  981995
Name: Reviews, dtype: object
```

## Show all the Apps which are Paid

```
In [45]: df[(df["Type"] == "Paid")].head(3)
```

Out[45]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	
234	TurboScan: scan documents and receipts in PDF	BUSINESS	4.7	11442	6.8M	100,000+	Paid	\$4.99	Everyone	Business	
235	Tiny Scanner Pro: PDF Doc Scan	BUSINESS	4.8	10295	39M	100,000+	Paid	\$4.99	Everyone	Business	
290	TurboScan: scan documents and receipts in PDF	BUSINESS	4.7	11442	6.8M	100,000+	Paid	\$4.99	Everyone	Business	

## Show all the Apps which are Free

```
In [46]: df[(df["Type"] == "Free")].head(3)
```

Out[46]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	A
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Desi
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	A

## Show the Minimum Size

```
In [55]: df["Size"].min()
```

```
Out[55]: '1.0M'
```