

DATA ANALYSIS OF GOOGLE APP'S RATINGS

```
In [1]: import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib inline
import pandas as pd

In [2]: google_data = pd.read_csv('C:/Users/gh/Desktop/kaggle project/GoogleAppRating-DataAnalysis-master/googleplaystore.csv')

In [3]: google_data.head()

Out[3]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Glitch & 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	Design	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

```
In [4]: google_data.shape
Out[4]: (18841, 13)

In [5]: type(google_data)
Out[5]: pandas.core.frame.DataFrame

In [6]: google_data.describe()

Out[6]:
```

	Rating
count	9367.000000
mean	4.183338
std	0.537401
min	1.000000
25%	4.000000
50%	4.300000
75%	4.500000
max	19.000000

```
In [7]: google_data.boxplot()

Out[7]:
```

```
In [8]: google_data.hist()
Out[8]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x1516a220240>]],
      dtype=object)
```

Data Cleaning

```
In [10]: google_data.isnull().sum()

Out[10]:
```

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

check hoe many ratings are more than 5 -Outers

```
In [11]: google_data[google_data.Rating > 5]

Out[11]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version	Android Version
10472	Life Made With-Fi Touchscreen	1.9	18.0	3.0M	1,000+	Free	0	Everyone	NaN	February 23, 2014	1.0.19	4.0 and above	Na

```
In [12]: google_data.drop([10472],inplace=True)

In [13]: google_data[10470:10475]

Out[13]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
10470	Jazz Wi-Fi	COMMUNICATION	3.4	49	4.0M	10,000+	Free	0	Everyone	Communication	February 10, 2017	
10471	Xposed Wi-Fi-Pad	PERSONALIZATION	3.5	1042	404k	100,000+	Free	0	Everyone	Personalization	August 5, 2014	
10473	camino Wi-Fi free WiFi	TOOLS	4.2	134203	4.1M	10,000,000+	Free	0	Everyone	Tools	August 7, 2018	6
10474	Sa-Fi Voice	COMMUNICATION	3.4	37	14M	1,000+	Free	0	Everyone	Communication	November 23, 2014	
10475	Wi-Fi										May 17, 2017	

```
In [14]: google_data.boxplot()

Out[14]:
```

```
In [15]: google_data.hist()
Out[15]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000001516A5A0D08>]],
      dtype=object)
```

Remove columns that are 90% empty

```
In [16]: threshold = len(google_data)*0.1 #10% my rows
threshold

Out[16]: 1084.0

In [17]: google_data.dropna(thresh=threshold,axis=1,inplace=True)

In [18]: print(google_data.isnull().sum())

App              0
Category         0
Rating          1474
Reviews          0
Size             0
Installs         0
Type             1
Price            0
Content Rating   0
Genres           0
Last Updated     0
Current Ver      8
Android Ver      2
dtype: int64

Data Imputation and Manipulation

In [19]: #Fill the null values with appropriate values using aggregate functions such as mean, median

In [20]: google_data.fillna(google_data.median(), inplace=True) # replace missing with median

In [21]: google_data.isnull().sum()

Out[21]:
```

App	0
Category	0
Rating	1474
Reviews	0
Size	0
Installs	0
Type	1
Price	0
Content Rating	0
Genres	0
Last Updated	0
Current Ver	8
Android Ver	2
dtype:	int64

```
In [22]: google_data.head()

Out[22]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Glitch & 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	Design	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

```
In [23]: #modes of categorical values
print(google_data['Type'].mode())
print(google_data['Current Ver'].mode())
print(google_data['Android Ver'].mode())

0 Free
0 Varies with device
dtype: object
0 4.1 and up
dtype: object

In [24]: #Fill the missing categorical values with mode
google_data['Type'].fillna(google_data['Type'].mode,inplace=True)
google_data['Current Ver'].fillna(google_data['Current Ver'].mode,inplace=True)
google_data['Android Ver'].fillna(google_data['Android Ver'].mode,inplace=True)

In [25]: google_data.isnull().sum()

Out[25]:
```

App	0
Category	0
Rating	0
Reviews	0
Size	0
Installs	0
Type	0
Price	0
Content Rating	0
Genres	0
Last Updated	0
Current Ver	0
Android Ver	0
dtype:	int64

```
In [32]: #converting categorical value into numerical variable
google_data['Price']=google_data['Price'].apply(lambda x: str(x).replace('$','') if 'S' in s
str(x) else str(x))
google_data['Price']=google_data['Price'].apply(lambda x: float(x))

In [33]: google_data.head()

Out[33]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Glitch & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000.0	Free	0.0	Everyone	Design	January 15, 2018	2.0
2	U Launcher Lite - FREE Live Cool Themes, Hide	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000.0	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2

```
In [38]: google_data['Installs']=google_data['Installs'].apply(lambda x: str(x).replace(',','') if
google_data['Installs']=google_data['Installs'].apply(lambda x: str(x).replace(',','') if
', ' in str(x) else str(x))

In [39]: google_data.head()

Out[39]:
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Version
0	Photo Editor & Candy Camera & Glitch & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10000.0	Free	0.0	Everyone	Art & Design	January 7, 2018	1.0
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000.0	Free	0.0	Everyone	Design	January 15, 2018	2.0
2	U Launcher Lite - FREE Live Cool Themes, Hide	ART_AND_DESIGN	4.7	87510	8.7M	5000000.0	Free	0.0	Everyone	Art & Design	August 1, 2018	1.2

```
In [40]: google_data.describe()

Out[40]:
```

	Rating	Reviews	Installs	Price
count	10840.000000	1.084000e+04	1.084000e+04	10840.000000
mean	4.206476	4.441520e+05	1.546434e+07	1.027368
std	0.480342	2.927761e+06	8.502936e+07	15.949703
min	1.000000	0.000000e+00	0.000000e+00	0.000000
25%	4.100000	3.800000e+01	1.000000e+03	0.000000
50%	4.300000	2.094000e+03	1.000000e+05	0.000000
75%	4.500000	5.477500e+04	5.000000e+06	0.000000
max	5.000000	7.815831e+07	1.000000e+09	400.000000

data vizulization

```
In [43]: 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
ART_AND_DESIGN    4.35385
ART_AND_DESIGN    4.285882
AUTO_AND_VEHICLES  4.283919
BEAUTY             4.335498
BOOKS_AND_REFERENCE  4.182391
COMICS             4.168980
COMMUNICATION      4.189163
DATING             4.625641
EDUCATION           4.388462
ENTERTAINMENT      4.202174
EVENTS             4.395313
FAMILY             4.284564
FINANCE            4.155639
FOOD_AND_DRINK     4.185927
GAME              4.286888
HEALTH_AND_FITNESS  4.286959
HOUSE_AND_HOME     4.211394
LIBRARIES_AND_DEMO  4.297059
LIFESTYLE           4.133414
MAPS_AND_NAVIGATION  4.075182
MEDICAL            4.216199
NEWS_AND_MAGAZINES  4.161037
PARENTING          4.298909
PERSONALIZATION    4.328827
PHOTOGRAPHY        4.197910
PRODUCTIVITY       4.276651
SHOPPING           4.263077
SOCIAL             4.261017
SPORTS             4.236458
TOOLS             4.688071
TRAVEL_AND_LOCAL   4.132946
VIDEO_PLAYERS      4.184980
WEATHER            4.248780
Name: Rating, dtype: float64

Category
ART_AND_DESIGN    5.97
AUTO_AND_VEHICLES  13.47
BEAUTY             0.00
BOOKS_AND_REFERENCE  119.77
BUSINESS           185.27
COMICS             0.00
COMMUNICATION      83.14
DATING             31.43
EDUCATION           17.98
ENTERTAINMENT      7.98
EVENTS            109.99
FAMILY            2434.78
FINANCE           2990.83
FOOD_AND_DRINK     287.30
GAME              67.30
HEALTH_AND_FITNESS  0.00
HOUSE_AND_HOME     0.00
LIBRARIES_AND_DEMO  0.00
LIFESTYLE          2360.87
MAPS_AND_NAVIGATION  26.95
MEDICAL           1439.98
NEWS_AND_MAGAZINES  3.98
PARENTING          9.58
PERSONALIZATION    153.96
PHOTOGRAPHY        124.21
PRODUCTIVITY       259.93
SHOPPING           5.48
SOCIAL            15.97
SPORTS            100.00
TOOLS             267.25
TRAVEL_AND_LOCAL   49.95
VIDEO_PLAYERS      10.46
WEATHER           32.42
Name: Price, dtype: float64

Category
ART_AND_DESIGN    2.637609e+04
AUTO_AND_VEHICLES  1.369619e+04
BEAUTY             7.476226e+03
BOOKS_AND_REFERENCE  9.506090e+04
BUSINESS           3.333598e+04
COMICS             5.638793e+04
COMMUNICATION      2.107138e+06
DATING             3.159336e+04
EDUCATION           2.538191e+05
ENTERTAINMENT      3.971888e+05
EVENTS             2.151966e+03
FAMILY            2.989255e+05
FINANCE            4.795281e+04
FOOD_AND_DRINK     6.594748e+04
GAME              1.385559e+06
HEALTH_AND_FITNESS  1.112538e+05
HOUSE_AND_HOME     4.518619e+04
LIBRARIES_AND_DEMO  1.220135e+04
LIFESTYLE           3.372457e+04
MAPS_AND_NAVIGATION  2.379820e+05
MEDICAL            3.454520e+03
NEWS_AND_MAGAZINES  1.922292e+05
PARENTING          1.597218e+04
PERSONALIZATION    2.752596e+05
PHOTOGRAPHY        6.373631e+05
PRODUCTIVITY       2.691438e+05
SHOPPING           4.244620e+05
SOCIAL             2.105993e+06
SPORTS             1.844536e+05
TOOLS             3.248629e+05
TRAVEL_AND_LOCAL   6.278516e+05
VIDEO_PLAYERS      6.307439e+05
WEATHER            1.781065e+05
Name: Reviews, dtype: float64

In [44]: plt.figure(figsize=(12,5))
plt.plot(x,'r-',color='r')
plt.xticks(rotation=90)
plt.show()
```

```
In [48]: plt.figure(figsize=(16,5))
plt.plot(x,'r-',color='r')
plt.xticks(rotation=90)
plt.title('Category wise Pricing')
plt.xlabel('Categories-->')
plt.ylabel('Prices-->')
plt.show()
```

```
In [50]: plt.figure(figsize=(16,5))
plt.plot(x,'r-',color='r')
plt.xticks(rotation=90)
plt.title('Category wise Pricing')
plt.xlabel('Categories-->')
plt.ylabel('Prices-->')
plt.show()
```

```
In [51]: plt.figure(figsize=(16,5))
plt.plot(x,'r-',color='r')
plt.xticks(rotation=90)
plt.title('Category wise Reviews')
plt.xlabel('Categories-->')
plt.ylabel('Reviews-->')
plt.show()
```