

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
import sys
```

Reading Data

Please ensure that your folder name "does not have a space or a forward slash"/

```
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\sys_checks\points\googleplaystore.csv")
```

```
In [40]: # Inspecting the first 18 rows
google_data.head(18)
```

	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	7-Jan-18	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Design/Puzzle/Toy	15-Jan-18	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	50,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	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	8-Jun-18	Varies with device	4.2 and up
4	Pixol Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design/Creativity	20-Jun-18	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.3	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
7	Infinite Painter	ART_AND_DESIGN	4.1	36915	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.6.11	4.2 and up
8	Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
9	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design/Creativity	3-Jul-18	2.8	4.0.3 and up

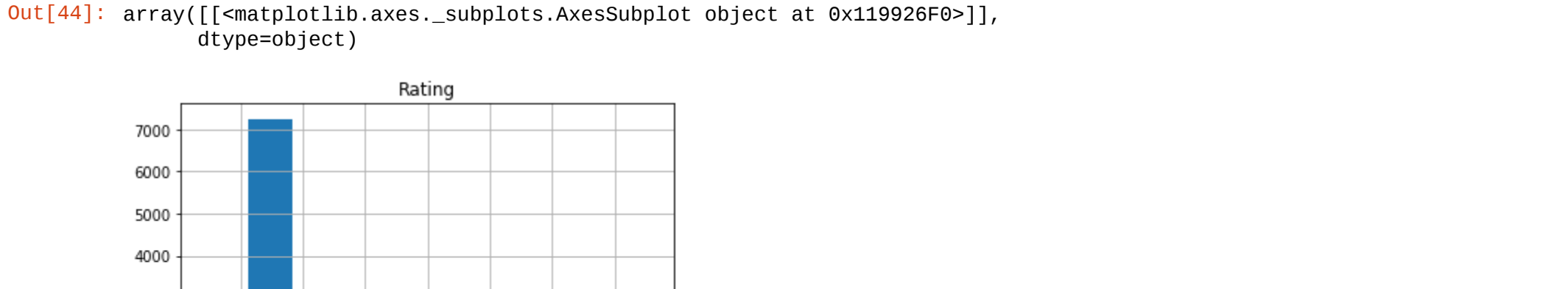
```
In [41]: # Inspecting shape of the Data into rows and columns
google_data.shape
```

```
Out[41]: (18841, 13)
```

```
In [42]: # Summary Statistics of Data
google_data.describe()
```



```
In [44]: # Inspecting the Data using histogram
google_data.hist()
```



```
In [45]: google_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18841 entries, 0 to 18840
Data columns (total 13 columns):
# Column                Non-Null Count  Dtype
---  ---                ---
0 App                    18841 non-null object
1 Category               18841 non-null object
2 Rating                 9367 non-null float64
3 Reviews                18841 non-null object
4 Size                   18841 non-null object
5 Installs               18840 non-null object
6 Price                  18841 non-null object
7 Content Rating         18840 non-null object
8 Genres                 18841 non-null object
9 Last Updated           18841 non-null object
10 Current Ver           18833 non-null object
11 Android Ver           18838 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()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	False	False	False	False	False	False	False	False	False	False	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	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
...
18836	False	False	False	False	False	False	False	False	False	False	False	False	False
18837	False	False	False	False	False	False	False	False	False	False	False	False	False
18838	False	False	True	False	False	False	False	False	False	False	False	False	False
18839	False	False	False	False	False	False	False	False	False	False	False	False	False
18840	False	False	False	False	False	False	False	False	False	False	False	False	False

18841 rows x 13 columns

```
In [47]: # Counting the number of missing values in each column
google_data.isnull().sum()
```

Out[47]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
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													

Checking - how many ratings are more than 5 - outliers

```
In [48]: # Checking the number of Ratings greater than 5
google_data[google_data.Rating > 5]
```

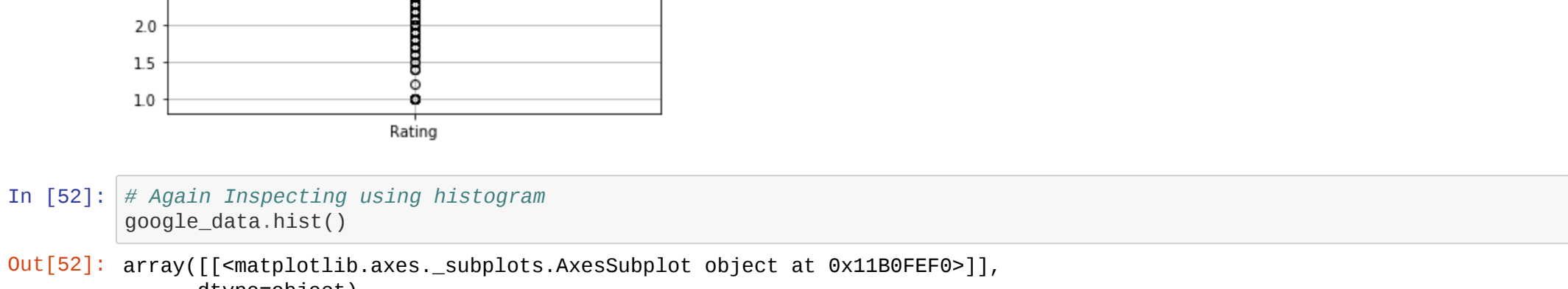
	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10472	Life Made Wi-Fi Touchscreen Photo Frame	ART_AND_DESIGN	1.9	18.0	3.0M	1,000+	Free	0	Everyone	N/A	11-Feb-18	1.0.19	4.0 and up

```
In [49]: google_data.drop([10472], inplace = True)
```

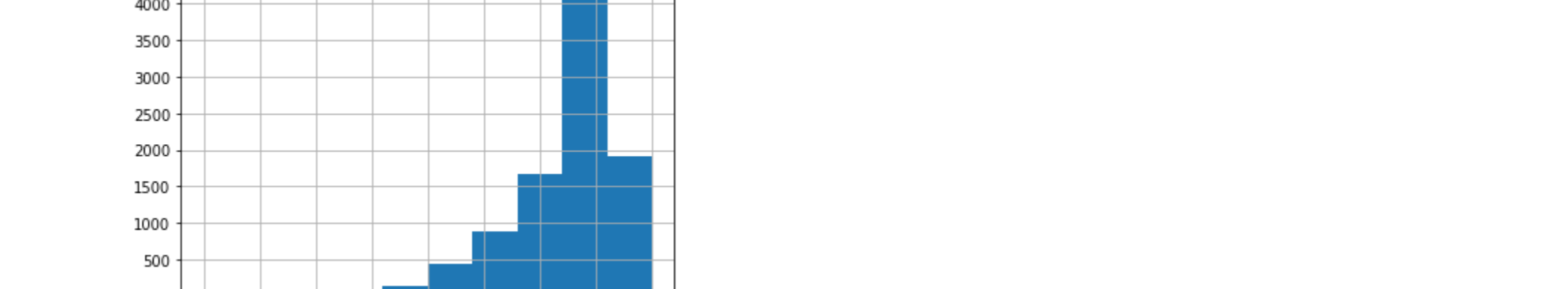
```
In [50]: # Checking dataframe if the values is deleted or not
google_data[10470:10475]
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
10470	Jazz Wi-Fi	COMMUNICATION	3.4	49	4.0M	10,000+	Free	0	Everyone	Communication	10-Feb-17	0.1	2.3 and up
10471	Xposed Wi-Fi-Pad	PERSONALIZATION	3.5	1042	404k	100,000+	Free	0	Everyone	Personalization	5-Aug-14	3.0.0	4.0.3 and up
10473	remmy Wi-Fi free Wi-Fi	TOOLS	4.2	134203	4.1M	10,000,000+	Free	0	Everyone	Tools	7-Aug-18	6.06.14	4.4 and up
10474	Set-Fi Voice	COMMUNICATION	3.4	37	14M	1,000+	Free	0	Everyone	Communication	21-Nov-14	2.2.1.5	2.2 and up
10475	Wi-Fi Visualizer	TOOLS	3.9	122	2.6M	50,000+	Free	0	Everyone	Tools	17-May-17	0.0.9	2.3 and up

```
In [51]: # Again Inspecting using BoxPlot
google_data.boxplot()
```



```
In [52]: # Again Inspecting using histogram
google_data.hist()
```



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]: 1884.0
```

```
In [54]: google_data.dropna(thresh= threshold, axis=1, inplace=True)
```

```
In [55]: print(google_data.isnull().sum())
```

Out[55]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
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 [57]: google_data.shape
```

```
Out[57]: (18840, 13)
```

Data Imputation and Manipulation

Fill the null values with appropriate values using aggregate functions such as mean, median or mode.

```
In [58]: # Define a function impute_median
def impute_median(series):
    return series.fillna(series.median())
```

```
In [59]: google_data.Rating = google_data["Rating"].transform(impute_median)
```

```
In [61]: # Count the number of null values in each column
google_data.isnull().sum()
```

Out[61]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
App	0												
Category	0												
Rating	0												
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 [63]: # Modes of Categorical values
print(google_data["Type"].mode())
print(google_data["Current Ver"].mode())
print(google_data["Android Ver"].mode())
```

Out[63]:

	Free	Varies with device	4.1 and up
dtype: object			

```
In [64]: # Filling the missing Categorical values with the mode
google_data["Type"] = google_data["Type"].fillna(google_data["Type"].mode().values[0], inplace=True)
google_data["Current Ver"] = google_data["Current Ver"].fillna(google_data["Current Ver"].mode().values[0], inplace=True)
google_data["Android Ver"] = google_data["Android Ver"].fillna(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
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 [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(x)
google_data["Installs"] = google_data["Installs"].apply(lambda x: str(x).replace(",", "")) if "," in str(x) else str(x)
google_data["Installs"] = google_data["Installs"].apply(lambda x: float(x))
```

```
In [73]: google_data.head(18)
```

```
Out[73]:
```

	Price	0
	Content Rating	0
	Genres	0
	Last Updated	0
	Current Ver	0
	Android Ver	0
	dtype: int64	

```
In [71]: #assigne last'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(x))
google_data["Installs"] = google_data["Installs"].apply(lambda x: str(x).replace(",", "").if "" in str(x) else str(x))
google_data["Installs"] = google_data["Installs"].apply(lambda x: float(x))

In [73]: google_data.head(16)

Out[73]:
```

	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	100000	Free	0.0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500000	Free	0.0	Everyone	Design/Puzzle/Toy	15-Jan-18	2.0.0	4.0.3 and up
2	U Launcher Lite - FREE Live Cool Themes, Hide Icons	ART_AND_DESIGN	4.7	87510	8.7M	50000000	Free	0.0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50000000	Free	0.0	Teen	Art & Design	8-Jun-18	Varies with device	4.2 and up
4	Pixol Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100000	Free	0.0	Everyone	Art & Design	20-Jun-18	1.1	4.4 and up
5	Paper flowers instructions	ART_AND_DESIGN	4.3	167	5.6M	50000	Free	0.0	Everyone	Design/Creativity	26-Mar-17	1	2.3 and up
6	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50000	Free	0.0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
7	Infinite Painter	ART_AND_DESIGN	4.1	36915	29M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	6.1.6.11	4.2 and up
8	Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1000000	Free	0.0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
9	Kids Paint Free - Drawing Fun	ART	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up
10	Coloring book for kids	ART_AND_DESIGN	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up
11	Coloring book for kids	ART_AND_DESIGN	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up
12	Coloring book for kids	ART_AND_DESIGN	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up
13	Coloring book for kids	ART_AND_DESIGN	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up
14	Coloring book for kids	ART_AND_DESIGN	4.2	10000	15M	1000000	Free	0.0	Everyone	Art & Design	14-Jun-18	1.0.0	4.0.3 and up