Play Store Review Data Analysis : Nirnoy Ghosh

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

Importing the data set.

In [2]:
```
playdt = pd.read_csv(r"C:\Users\ASUS\OneDrive\Desktop\Nirnoy\Placement\NITS\Project\Google Data\googleplaystore_user_reviews.csv")
playdt
```

Out[2]:

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.00	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.25	0.288462
2	10 Best Foods for You	NaN	NaN	NaN	NaN
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.40	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.00	0.300000
...	...	...	...	...	...
64290	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64291	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64292	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64293	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN
64294	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN

64295 rows × 5 columns

Showing information about the data.

In [3]:
```
playdt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 64295 entries, 0 to 64294
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   App                     64295 non-null  object
 1   Translated_Review       37427 non-null  object
 2   Sentiment               37432 non-null  object
 3   Sentiment_Polarity      37432 non-null  float64
 4   Sentiment_Subjectivity  37432 non-null  float64
dtypes: float64(2), object(3)
memory usage: 2.5+ MB
```

Data Types of the features.

In [4]:
```
playdt.dtypes
```

Out[4]:
```
App                        object
Translated_Review          object
Sentiment                  object
Sentiment_Polarity        float64
Sentiment_Subjectivity    float64
dtype: object
```

Dealing with Null values

Here we are dropping all the Nan values containg rows, because these rows contains no informations.

In [5]:
```
playdt=playdt.dropna()
playdt
```

Out[5]:

	App	Translated_Review	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity
0	10 Best Foods for You	I like eat delicious food. That's I'm cooking ...	Positive	1.000000	0.533333
1	10 Best Foods for You	This help eating healthy exercise regular basis	Positive	0.250000	0.288462
3	10 Best Foods for You	Works great especially going grocery store	Positive	0.400000	0.875000
4	10 Best Foods for You	Best idea us	Positive	1.000000	0.300000
5	10 Best Foods for You	Best way	Positive	1.000000	0.300000
...	...	...	...	...	...
64222	Housing-Real Estate & Property	Most ads older many agents ..not much owner po...	Positive	0.173333	0.486667
64223	Housing-Real Estate & Property	If photos posted portal load, fit purpose. I'm...	Positive	0.225000	0.447222
64226	Housing-Real Estate & Property	Dumb app, I wanted post property rent give opt...	Negative	-0.287500	0.250000
64227	Housing-Real Estate & Property	I property business got link SMS happy perform...	Positive	0.800000	1.000000
64230	Housing-Real Estate & Property	Useless app, I searched flats kondapur, Hydera...	Negative	-0.316667	0.400000

37427 rows × 5 columns

In [6]:
```
playdt.isna().sum()
```

Out[6]:
```
App                       0
Translated_Review         0
Sentiment                 0
Sentiment_Polarity        0
Sentiment_Subjectivity    0
dtype: int64
```

Counts the no of apps

In [7]:
```
playdt["App"].value_counts()
```

Out[7]:
```
Bowmasters                                 312
Angry Birds Classic                        273
Helix Jump                                 273
Calorie Counter - MyFitnessPal             254
Candy Crush Saga                           240
                                          ...
Fruit Block - Puzzle Legend                  1
Free Live Talk-Video Call                    1
Calendar+ Schedule Planner App               1
Caf - Mon Compte                             1
Bed Time Fan - White Noise Sleep Sounds      1
Name: App, Length: 865, dtype: int64
```

How many unique apps in the dataset presents.

In [8]:
```
playdt["App"].unique()
len(playdt["App"].unique())
```

Out[8]:
```
865
```

Statistical Analysis the Sentiment of the apps

In [9]:
```
playdt["Sentiment"]
```

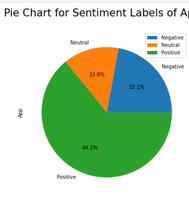
Out[9]:
```
0        Positive
1        Positive
3        Positive
4        Positive
5        Positive
           ...
64222    Positive
64223    Positive
64226    Negative
64227    Positive
64230    Negative
Name: Sentiment, Length: 37427, dtype: object
```

In [10]:
```
levels=playdt.groupby('Sentiment')[['App']].count()
levels
```

Out[10]:

	App
Sentiment	
Negative	8271
Neutral	5158
Positive	23998

In [11]:
```
levels.plot.pie(subplots=True,figsize=(12, 6),autopct='%1.1f%%');
plt.title('Pie Chart for Sentiment Labels of Apps \n',fontsize=21)
plt.show()
```

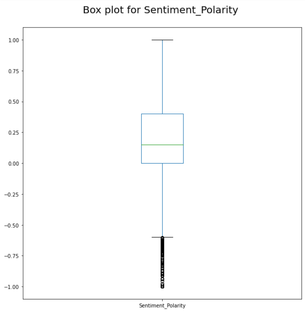


Statistical Analysis for Sentiment_Polarity

In [12]:
```
playdt["Sentiment_Polarity"].describe()
```

Out[12]:
```
count    37427.000000
mean         0.182171
std          0.351318
min         -1.000000
25%          0.000000
50%          0.150000
75%          0.400000
max          1.000000
Name: Sentiment_Polarity, dtype: float64
```

In [13]:
```
plt.figure(figsize=(10, 10))
playdt["Sentiment_Polarity"].plot.box();
plt.title("Box plot for Sentiment_Polarity \n",fontsize=20);
```



From the above plot we can see that many points are behaves like outliers.

Statistical Analysis for Sentiment_Subjectivity

In [14]:
```
playdt["Sentiment_Subjectivity"].describe()
```

Out[14]:
```
count    37427.000000
mean         0.492770
std          0.259984
min          0.000000
25%          0.357143
50%          0.514286
75%          0.650000
max          1.000000
Name: Sentiment_Subjectivity, dtype: float64
```

In [15]:
```
plt.figure(figsize=(10, 10))
playdt["Sentiment_Subjectivity"].plot.box();
plt.title("Box plot for Sentiment_Subjectivity \n",fontsize=20);
```