IS 603: Project Deliverable 3

Flipkart sales Analysis

This Dataset contains information on Products Name, Prices, Reviews, Rates, and Summary for Sentiment Analysis Purposes. The dataset can be used for a variety of applications such as price prediction, Sentiment Analysis, Auto Ai generated Reviews, and market Research.

After the Data cleaning our relevant data is 189874 rows and 5 columns from 194276 rows and 5 columns.

This dataset contains a total of 104 types of different products on flipkart.com. This dataset contains 189874 rows and 5 columns. Using Customer Review or Summary you can use it for sentiment Analysis purposes which gives us an idea about whether a product should be purchased or not based on Positive, Negative Reviews. Train your model using Natural Language Processing and you can make Ai applications. The dataset contains a .csv file format.

Descriptive Statistical Analysis:

- Arithmetic mean
- Median
- Mode
- Dispersion
- Range Variance
- Standard Deviation

In [14]:
```
df.head()
```

Out[14]:

	ProductName	Price	Rating	Review	Summary
0	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Super!	Great cooler.. excellent air flow and for this...
1	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Awesome	Best budget 2 fit cooler. Nice cooling
2	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	The quality is good but the power of air is de...
3	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	1.0	Useless product	Very bad product it's a only a fan
4	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	Ok ok product

In [36]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 189874 entries, 0 to 189873
Data columns (total 5 columns):
 #   Column       Non-Null Count   Dtype  
---  ------       --------------   -----  
 0   ProductName  189870 non-null  object 
 1   Price        189869 non-null  object 
 2   Rating       189869 non-null  float64
 3   Review       189866 non-null  object 
 4   Summary      189856 non-null  object 
dtypes: float64(1), object(4)
memory usage: 7.2+ MB
```

In [29]:
```
df['Rating'].mean() #mean of the rating column
```

Out[29]:
```
4.114415728739289
```

In [30]:
```
df['Rating'].median() #median of rating column
```

Out[30]:
```
5.0
```

In [31]:
```
df['Rating'].mode() #mode value of rating column
```

Out[31]:
```
0    5.0
Name: Rating, dtype: float64
```

In [34]:
```
df['Rating'].var() #variance of rating column
```

Out[34]:
```
1.7027009960004422
```

In [35]:
```
df['Rating'].std() #standard deviation of rating column
```

Out[35]:
```
1.3048758546315593
```

Sentiment Analysis Result :

In [50]:
```
sentiments = SentimentIntensityAnalyzer()
df["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in df["Summary"]]
df["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in df["Summary"]]
df["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in df["Summary"]]
df["Compound"] = [sentiments.polarity_scores(i)["compound"] for i in df["Summary"]]
# data.head()
score = df["Compound"].values
sentiment = []
for i in score:
    if i >= 0.05 :
        sentiment.append('Positive')
    elif i <= -0.05 :
        sentiment.append('Negative')
    else:
        sentiment.append('Neutral')
df["Sentiment"] = sentiment
```

In [52]:
```
df.head()
```

Out[52]:

	ProductName	Price	Rating	Review	Summary	Sentiment	Positive	Negative	Neutral	Compound
0	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Super!	Great cooler.. excellent air flow and for this...	Positive	0.565	0.000	0.435	0.9576
1	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Awesome	Best budget 2 fit cooler. Nice cooling	Positive	0.760	0.000	0.240	0.8591
2	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	The quality is good but the power of air is de...	Positive	0.163	0.000	0.837	0.2382
3	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	1.0	Useless product	Very bad product it's a only a fan	Negative	0.228	0.376	0.396	-0.3597
4	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	Ok ok product	Positive	0.815	0.000	0.185	0.5267