

Flipkart Products Review Dataset

IS603 – Decision Making Support Systems



By (Group-2)

Vineeth Petnakota (RX82617)

Sravani Linga (OO40177)

Swati Roy (PR57930)

Project Description

- Our project aims to develop a DSS application for Flipkart
- The objective is to optimize supply chain management, inventory cost, and determine future scope using predictive analysis
- Product reviews will be analyzed to improve product quality, customer service, marketing, advertising, and pricing strategy
- Dashboard visualization will assist sales managers in making informed decisions about purchases, discounts, and future sales
- The dataset includes consumer comments and helps maintain inventory levels for profitable goods
- Inventory data, such as number of purchases and product categories, will be analyzed
- Sentiment analysis will be used to provide insights into customer sentiment, increase customer satisfaction, and drive sales for Flipkart

Dataset Description

- Dataset-related information:
- Link:<https://www.kaggle.com/datasets/mansithummar67/flipkart-product-review-dataset>
- This dataset contains a total of 104 types of different products on flipkart.com.
- There are total number of 194276 rows and 5 columns.

Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	ProductName	189870 non-null	object
1	Price(Rs)	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

Data Preprocessing

The preprocessed dataset has 189874 rows and 5 columns, which is 4402 rows less than the original dataset.

Data has undergone pre-processing and analysis. Some preparation has been done, such as removing unwanted symbols(?"?) in the ProductName column so the readability and meaningful product name is obtained. And also unwanted (É??)(,?π) symbols in the price column has been eliminated so only numeric values in the prices have been updated in the preprocessing phase.

Apart from the above pre-processing in the Rate column had string data which must have only numbers, the unwanted string data (Bajaj DX, Nova favourite etc) strings have been eliminated and only numeric data is obtained in the Rate column which is now easy to analyse.

- **Data Before Preprocessing**

C	D	E
Rate	Review	Summary
5	Rate	Now and for this price. It cooling power of air is decent fan
5	Sort	and provides good air
1	A Z ↓ Ascending	Z A ↓ Descending
3	By color: None	trial and perfectly work
5	Filter	/
4	By color: None	zeze producer
5	Choose One	
5	Search	
5	3	
5	4	
5	5	
5	Bajaj DX 2 L/W Dry Iron	able price
5	Nova Plus Amaze NI 10 1100 W I	beautiful designs afford
5	Pigeon Favourite Electric Kettle?	experience.
5	S	
5	(Blanks)	
5	<input checked="" type="checkbox"/> Auto Apply	
	<input type="button" value="Apply Filter"/>	<input type="button" value="Clear Filter"/>
5	Simply awesome	Very good performance and nice look
4	Wonderful	Product is good having strong thrust of air flow.. M
5	Terrific	Very good ???ç
2	Expected a better product	Bad quality
5	Excellent	Very nice ???ç product
5	Terrific purchase	Very soundless product and low price other compa

19	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
20	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
21	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
22	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
23	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
24	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
25	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black,	,Cn655	
26	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, L...),Cn655	,Cn655	1 Fair
27	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	1 Don't waste your money
28	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	5 Excellent
29	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	4 Wonderful
30	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	2 Could be way better
31	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	5 Awesome
32	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	1 Worst experience ever!
33	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	1 Don't waste your money
34	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	1 Not recommended at all
35	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	3 Good
36	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	4 Good choice
37	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	4 Good quality product
38	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	4 Nice product
39	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	3 Good
40	S . K Store Carbon Steel Collapsible Wardrobe-+~+(Finish Color - Black, DIY(Do-It-Yourself)	,Cn655	1 Utterly Disappointed

Data After Preprocessing

The screenshot shows a Jupyter Notebook interface with several code cells and their outputs.

In [44]: `import pandas as pd`
`df=pd.read_csv("/Users/vineethpetnakota/desktop/flipkart_product_final.csv",encoding="unicode_escape")`

In [45]: `df.head()`

Out[45]:

	ProductName	Price	Rate	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 [43]: `df.dtypes`

Out[43]:

ProductName	object
Price	object
Rate	object
Review	object
Summary	object
dtype:	object

In [32]: `df.shape`

Out[32]: (189874, 5)

On the right side, there is a preview of a CSV file with columns A, B, C, and D. The data includes various product reviews and ratings.

A	B	C	D
1 ProductName		Price	Rate
2 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	5 Super!	
3 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	5 Awesome	
4 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	3 Fair	
5 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	1 Useless product	
6 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	3 Fair	
7 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	5 Awesome	
8 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	5 Highly recommended	
9 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	3 Nice	
0 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	1 Unsatisfactory	
1 Candes 12 L Room/Personal Air Cooler(White, Black, Elegant High Speed-Honey Comb Cooling Pad & Ice Chamber, Blower)	3,999	4 Worth the money	
2 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Great product	
3 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Mind-blowing purchase	
4 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Highly recommended	
5 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Brilliant	
6 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Classy product	
7 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Must buy!	
8 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Fabulous!	
9 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Worth every penny	
0 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Super!	
1 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Great product	
2 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Awesome	
3 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Worth every penny	
4 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Just wow!	
5 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Terrific	
6 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Awesome	
7 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Excellent	
8 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Worth every penny	
9 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Terrific	
0 Candes 60 L Room/Personal Air Cooler(White, Black, CRETA)	8,999	5 Awesome	

Sentiment Analysis



Sentiment Analysis using Textblob python library from nltk to analyse textual data.



We looked at positive, negative, and neutral reviews. So that we may more precisely identify product reviews.



Our project's sentiment analysis focuses on reviews.



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.

Code snippets for sentiment analysis using Textblob

```
In [51]: def sentiment_analysis(polarity):
    if polarity > 0:
        return 'Positive'
    elif polarity < 0:
        return 'Negative'
    else:
        return 'Neutral'
df['Sentiment'] = df['Polarity'].apply(sentiment_analysis)
```

```
In [50]: subjectivity_list = []
polarity_list = []
for review in df['Summary']:
    tb = TextBlob(review)
    subjectivity_list.append(tb.sentiment.subjectivity)
    polarity_list.append(tb.sentiment.polarity)

#Add the result to the DataFrame
df['Subjectivity'] = subjectivity_list
df['Polarity'] = polarity_list
df.head()
```

Out[50]:

	ProductName	Price(Rs)	Rating	Review	Summary	Subjectivity	Polarity
0	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Super!	Great cooler.. excellent air flow and for this...	0.812500	0.725000
1	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Awesome	Best budget 2 fit cooler. Nice cooling	0.566667	0.666667
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...	0.633333	0.433333
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	0.933333	-0.455000
4	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	Ok ok product	0.500000	0.500000

Code output with sentiment analysis

In [84]:	df.head(20)							
Out [84]:								
Product Data:								
The following table displays the first 20 rows of the product dataset, including columns for Product Name, Price (Rs), Rating, Review, Summary, Subjectivity, Polarity, and Sentiment.								
	ProductName	Price(Rs)	Rating	Review	Summary	Subjectivity	Polarity	Sentiment
0	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Super!	Great cooler.. excellent air flow and for this...	0.812500	0.725000	Positive
1	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Awesome	Best budget 2 fit cooler. Nice cooling	0.566667	0.666667	Positive
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...	0.633333	0.433333	Positive
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	0.933333	-0.455000	Negative
4	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Fair	Ok ok product	0.500000	0.500000	Positive
5	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Awesome	The cooler is really fantastic and provides go...	0.680000	0.420000	Positive
6	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	5.0	Highly recommended	Very good product	0.780000	0.910000	Positive
7	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	3.0	Nice	Very nice	1.000000	0.780000	Positive
8	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	1.0	Unsatisfactory	Very bad cooler	0.866667	-0.910000	Negative
9	Candes 12 L Room/Personal Air Cooler(White, Bl...	3,999	4.0	Worth the money	Very good	0.780000	0.910000	Positive
10	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Great product	Beautiful product good material and perfectly ...	0.866667	0.850000	Positive
11	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Mind-blowing purchase	Awesome	1.000000	1.000000	Positive
12	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Highly recommended	Good	0.600000	0.700000	Positive
13	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Brilliant	Wonderful product, Must buy	1.000000	1.000000	Positive
14	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Classy product	Nice air cooler, smart cool breeze producer	0.764286	0.388095	Positive
15	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Must buy!	Awsm	0.000000	0.000000	Neutral
16	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Fabulous!	Nice product ???	1.000000	0.600000	Positive
17	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Worth every penny	Great cooler..	0.750000	0.800000	Positive
18	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Super!	Nice product	1.000000	0.600000	Positive
19	Candes 60 L Room/Personal Air Cooler(White, Bl...	8,999	5.0	Great product	Good ???	0.600000	0.700000	Positive

Statistical Analysis

- We did statistical analysis for this data set in terms of
 - **Arithmetic mean**
 - **Median**
 - **Mode**
 - **Dispersion**
 - **Range Variance**
 - **Standard Deviation**

```
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
```

Visualization

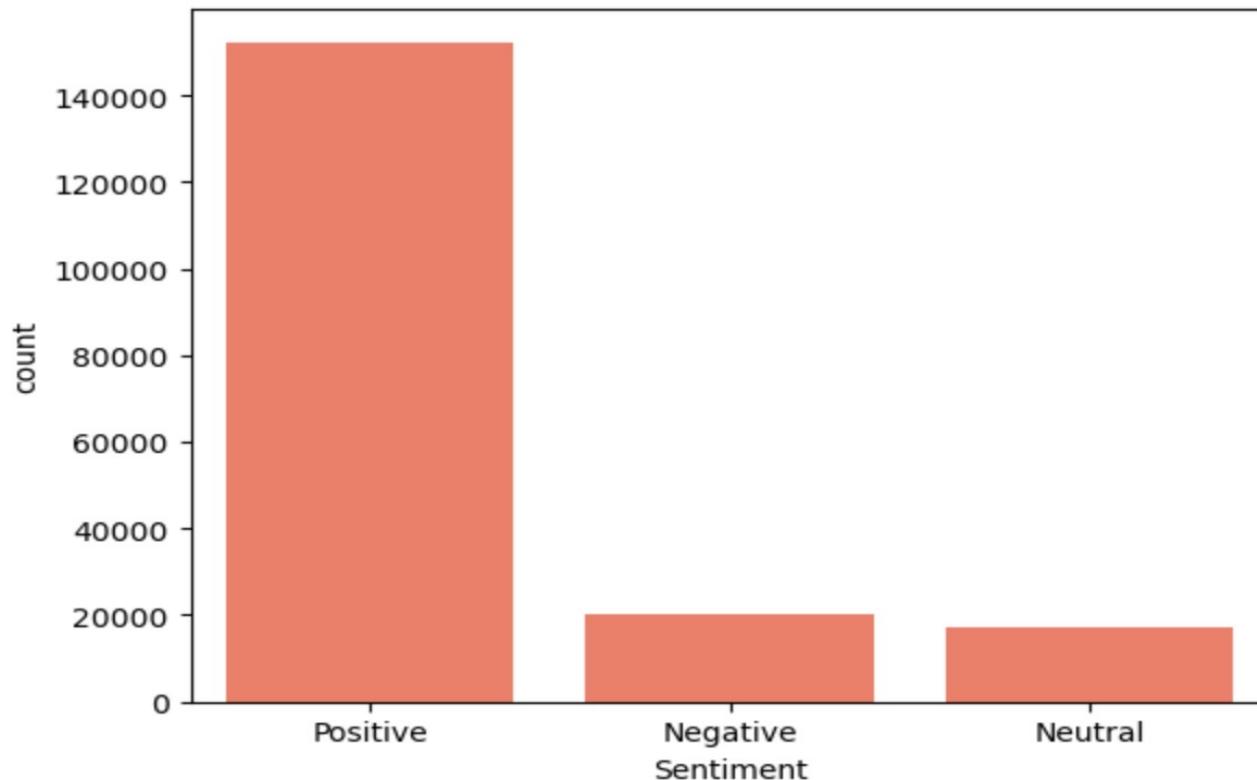
- Quantitative Data Visualization was used in our project. Based on the review and summary, we used the comparative technique for the various product names.
- The product was compared based on review comments (excellent to worst reviews).
- Our data was best shown using Bar chart.
- This will assist to decide if the product with the highest rating and the most reviews can be handled more effectively to improve sales, while the product with the lowest rating will focus on marketing strategy to raise sales.



Positive,Negative and Neutral Review Graph

```
In [54]: sns.countplot(x=df["Sentiment"], color = '#FF7256')
```

```
Out[54]: <AxesSubplot:xlabel='Sentiment', ylabel='count'>
```



Word Cloud

```
import pandas as pd
import matplotlib.pyplot as plt
from wordcloud import WordCloud

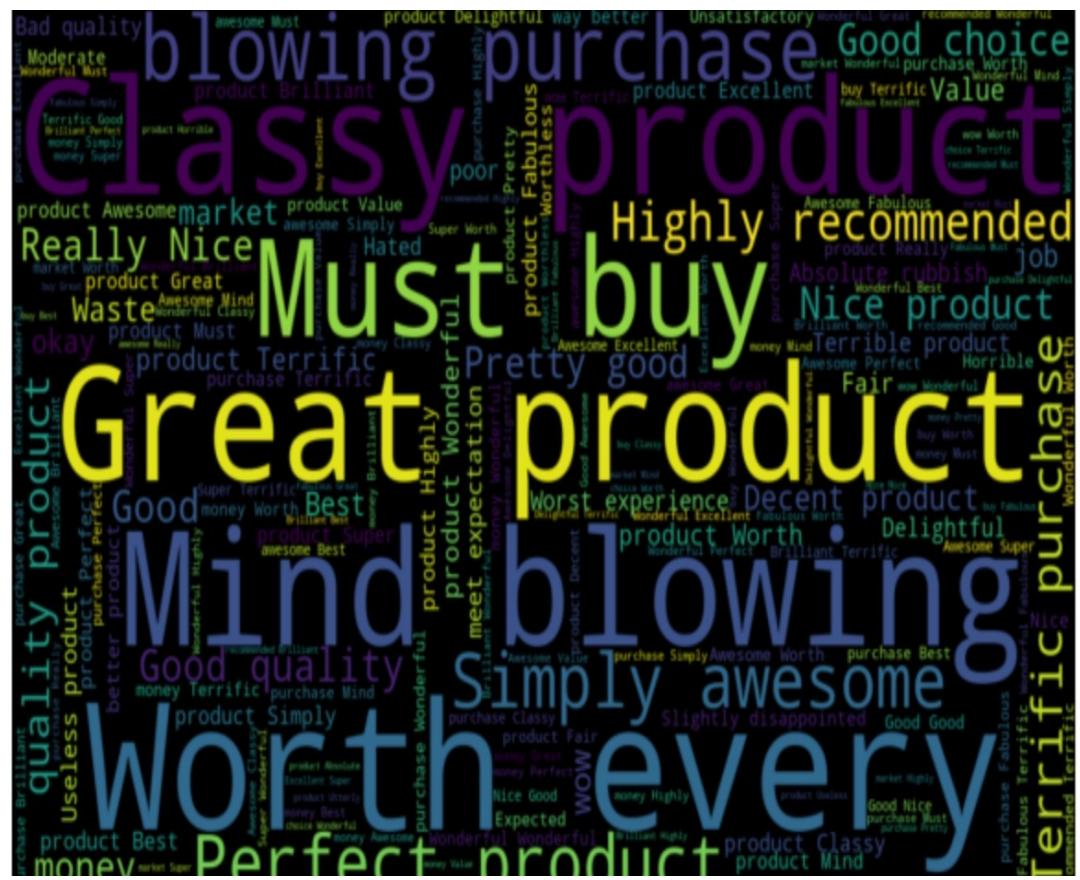
w = WordCloud()
stop_words = list(w.stopwords)
custom_stop_words = ['Nan', 'penny']
stop_words = set(stop_words + custom_stop_words)

text_column = df['Review']

# Concatenate the text from the selected column
text = ' '.join(text_column)

# Create a WordCloud object
wordcloud = WordCloud(width=1200, height=800, background_color='black', stopwords=st

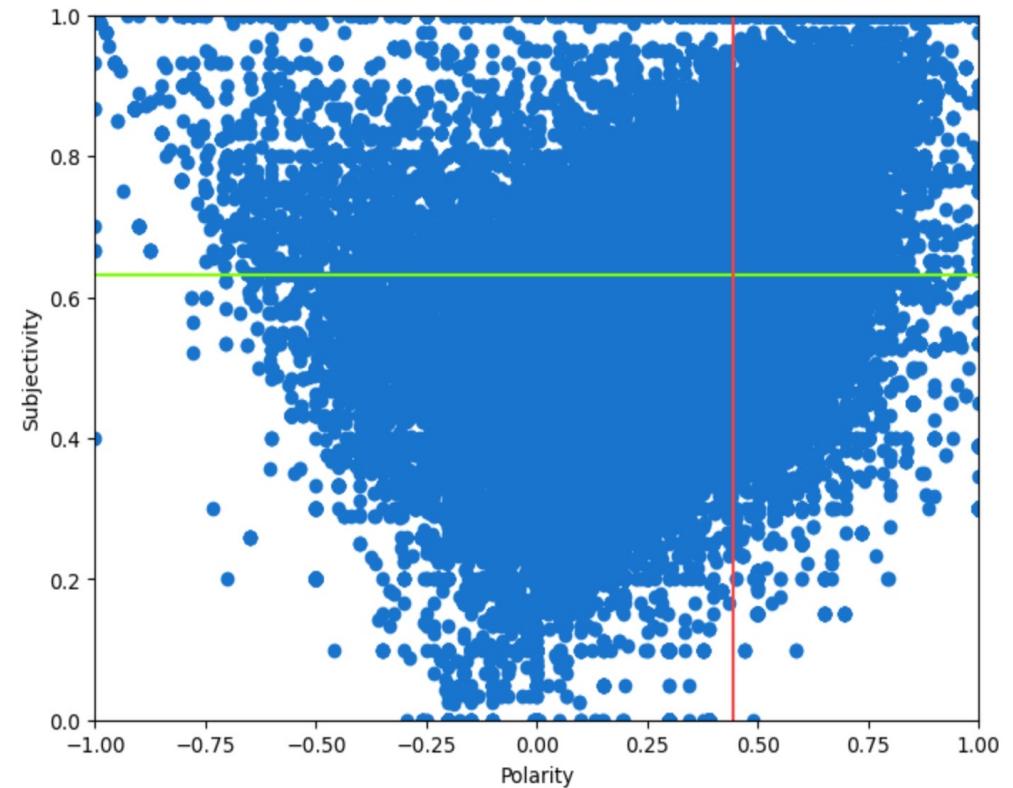
# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

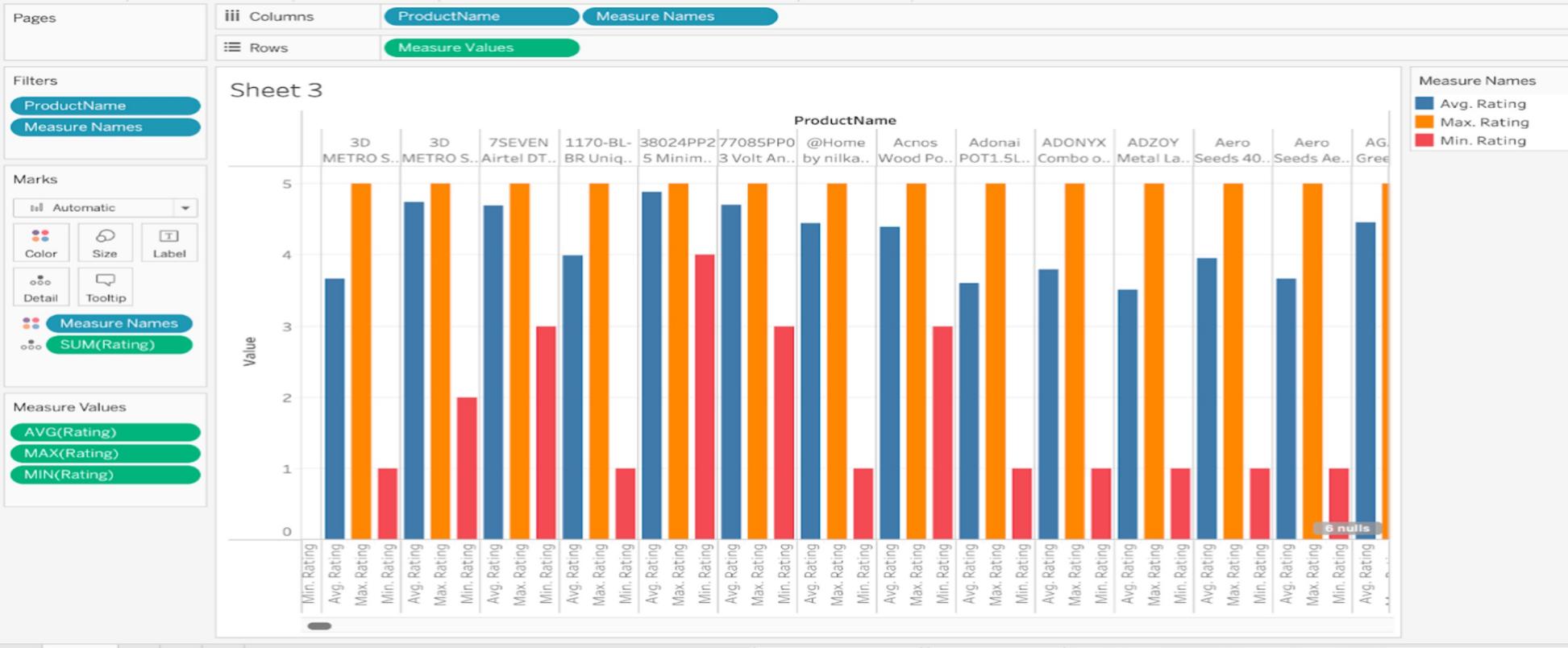


Sentiment Analysis plot with subjectivity vs Polarity

#Visualization of the sentiment analysis Subjectivity vs Polarity

```
plt.figure(figsize=(8,6))
plt.scatter(df.Polarity, df.Subjectivity, color = '#1874CD')
plt.axhline(y = np.mean(df.Subjectivity), color = '#7FFF00')
plt.axvline(x = np.mean(df.Polarity), color = '#FF4040')
plt.xlabel("Polarity")
plt.xlim((-1, 1))
plt.ylabel("Subjectivity")
plt.ylim(0, 1)
plt.show()
```

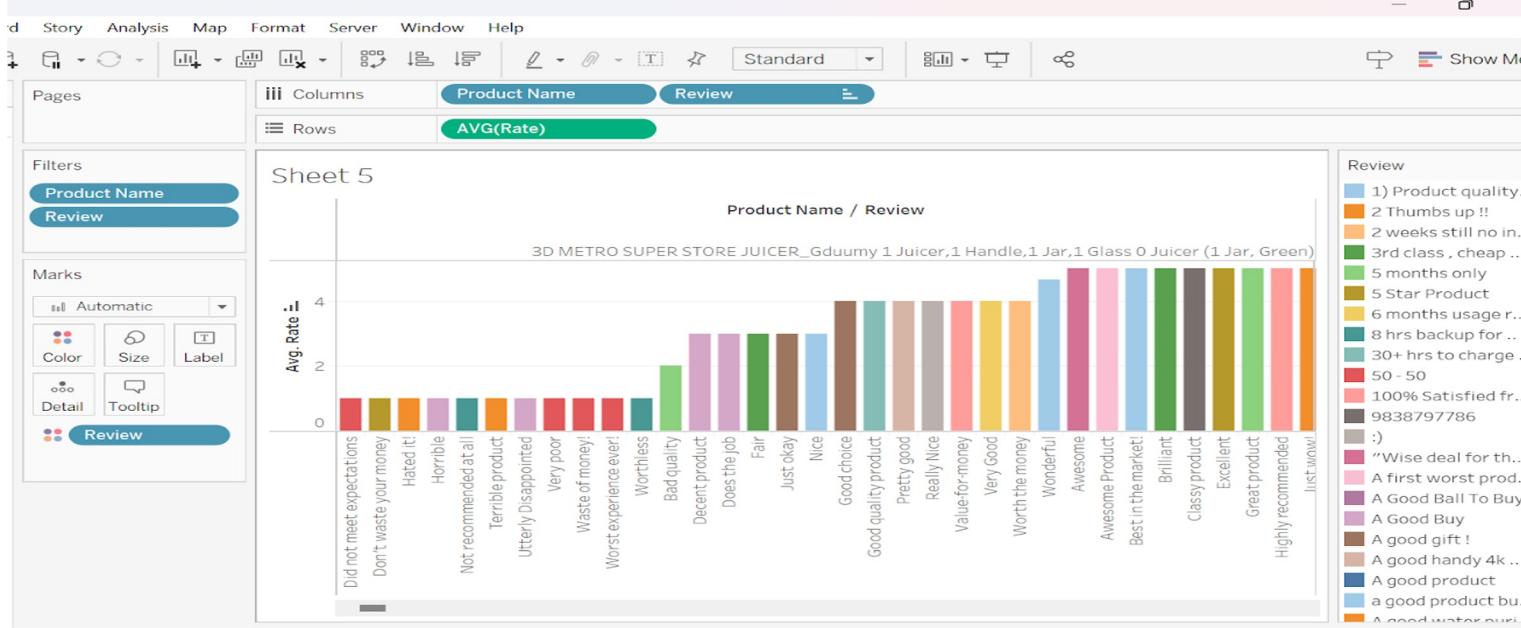




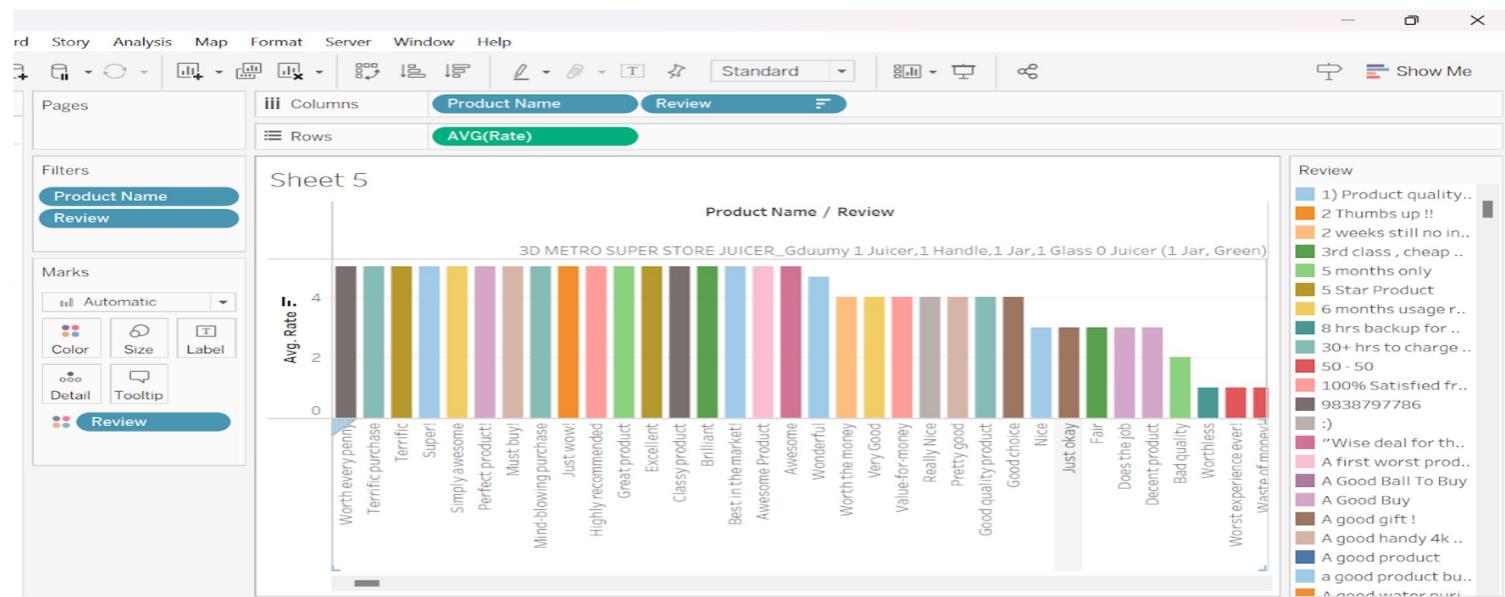
Each Product based on min , average and max rating



Standard deviation and variance of the products



**Product
based on
maximum
review**





- **Limitations**

- **Dataset limitations:**

- Limited scope: Due to portion of the platform's products are available for a limited time.
- Our model have shown sentimental analysis for the over all data but our analysis was not able to show the sentimental analysis for each product wise.
- Bias :The data set may be biased towards certain products or categories, as well as towards certain types of reviews or ratings.
- Data Privacy: Flipkart may impose limits on the use of their data, and it is critical that any research or predictions based on the data conform with data privacy laws and regulations.

Thank You

