M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 & ISO 14001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

18AIE010T – WEB & SOCIAL MEDIA MINING

TEAM MEMBERS:
- **GURUMEETA SR (927621BAD012)**
- **PRANISHKA N (927621BAD037)**
- **YUVASHREE S (927621BAD063)**

STAFF:
Dr. T SARAVANAN
AP/AI

SUBMIT DATE : MARCH 04, 2024

FLIPKART DATA MINING

INTRODUCTION:

Brief overview of the project objective: Implementing mining techniques on a dataset from Flipkart to extract valuable insights into consumer behaviour, product trends, and market dynamics.

DATASET DESCRIPTION:

Description of the Flipkart dataset used, including the types of data available (e.g., product sales, customer reviews, user interactions).

Mention of any preprocessing steps performed to clean and prepare the data for analysis.

MINING TECHNIQUES IMPLEMENTED:

Explanation of the mining techniques applied, including:

- Data preprocessing: Cleaning and transforming the raw dataset to make it suitable for analysis.
- Exploratory data analysis (EDA): Visualizing and summarizing key aspects of the data to understand its characteristics and distributions.
- Sentiment analysis on reviews: Analyzing customer reviews to extract sentiment polarity and identify trends in product feedback.
- Recommendation systems: Building models to suggest relevant products to customers based on their past preferences and behavior.
- Social network analysis: Analyzing relationships between products and customers to identify influential products and customer segments.

IMPLEMENTATION:

Python, with its rich ecosystem of libraries, is an ideal choice for implementing mining techniques on the Flipkart dataset. Learners can utilize libraries such as NetworkX for network analysis, Pandas for data manipulation, and Matplotlib for visualization. By writing Python code, learners can preprocess the dataset, apply mining algorithms, and visualize results in an interactive and intuitive manner.

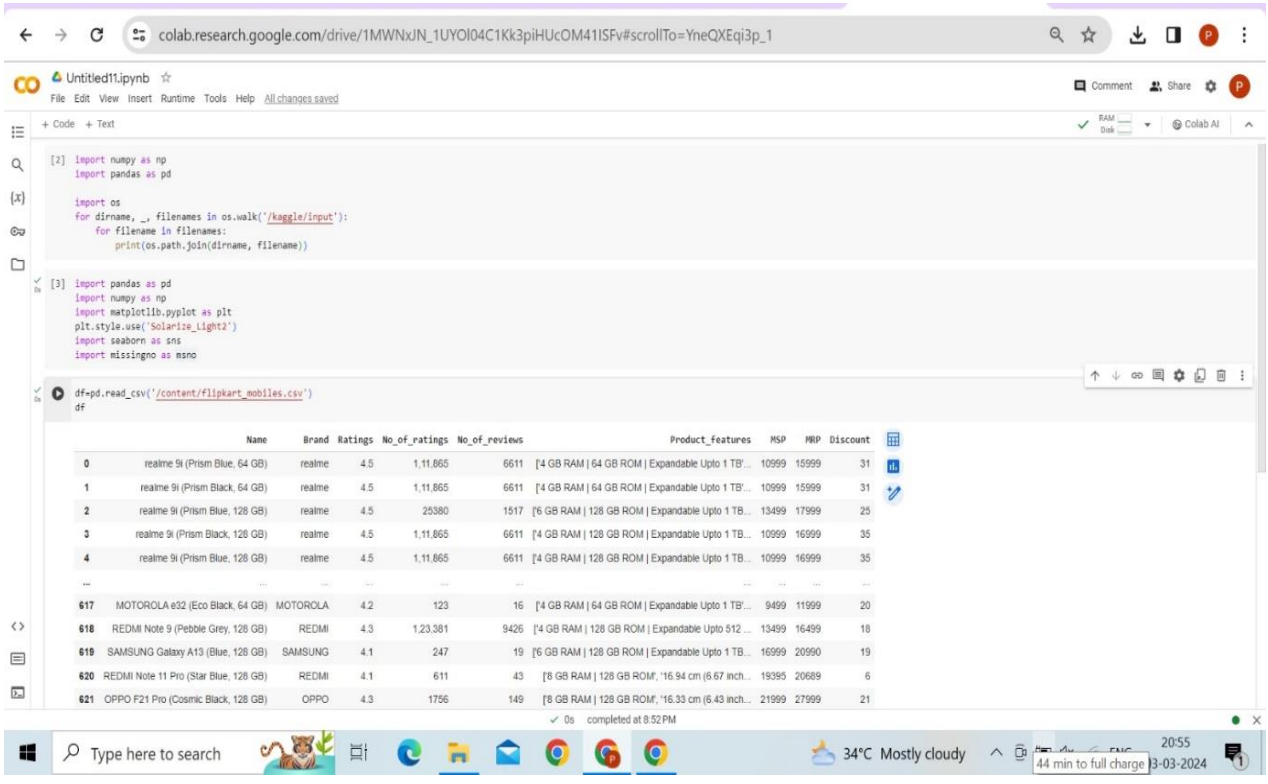
Coding exercises form the core of implementing mining techniques in Python. Learners can engage in hands-on activities, including:

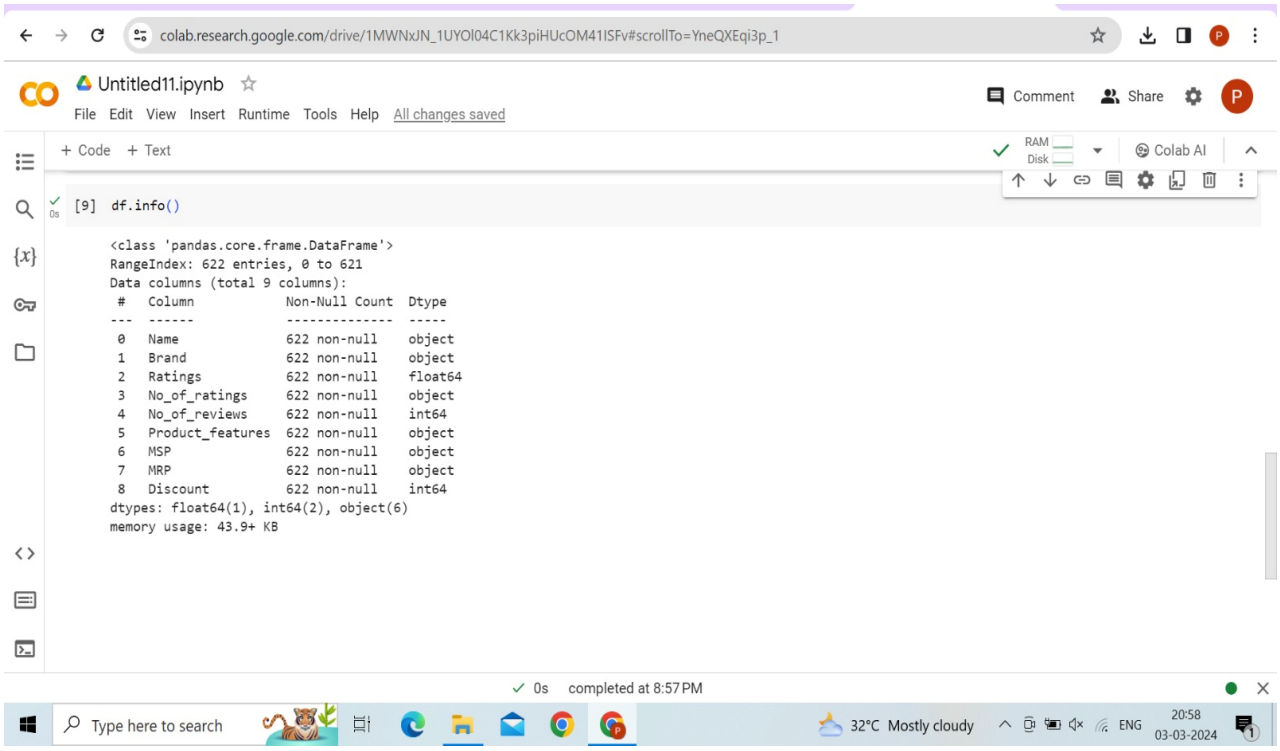
1)Loading and pre-processing the Flipkart dataset using Pandas.

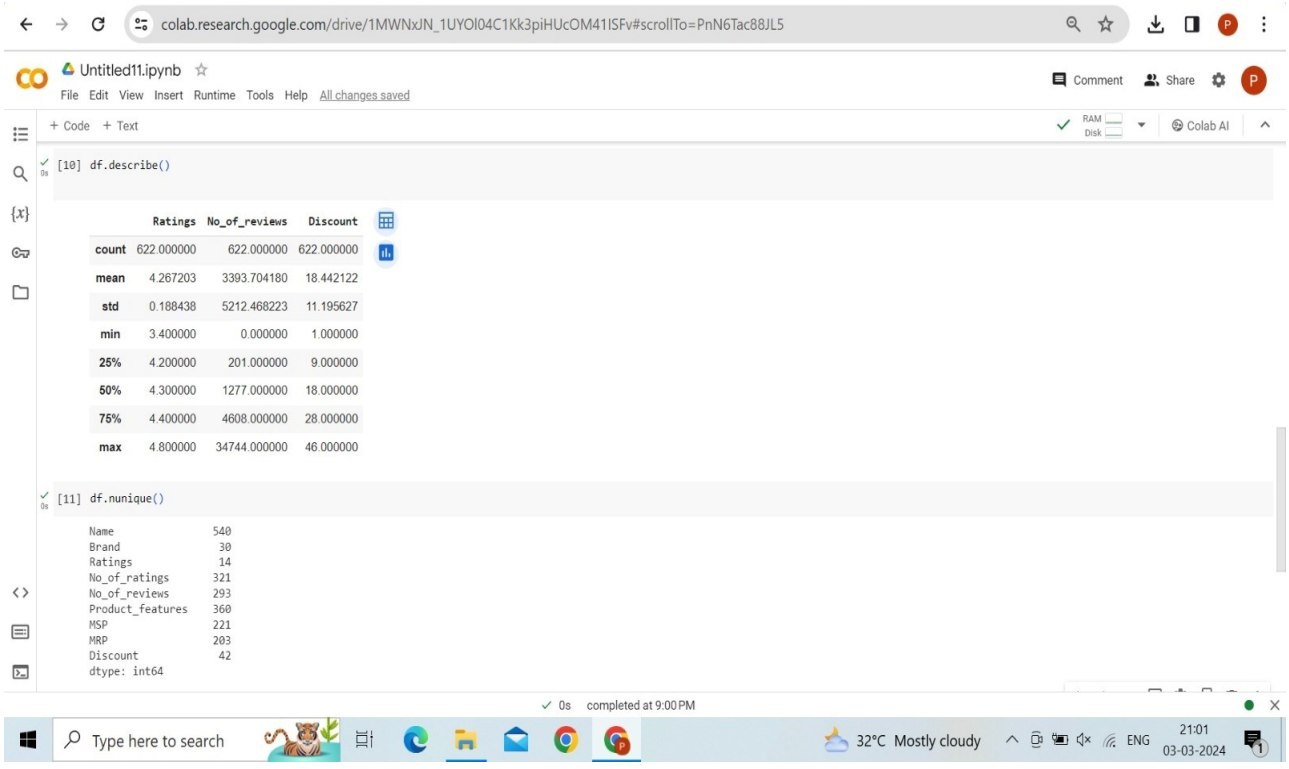
2)Constructing a social network graph using NetworkX.

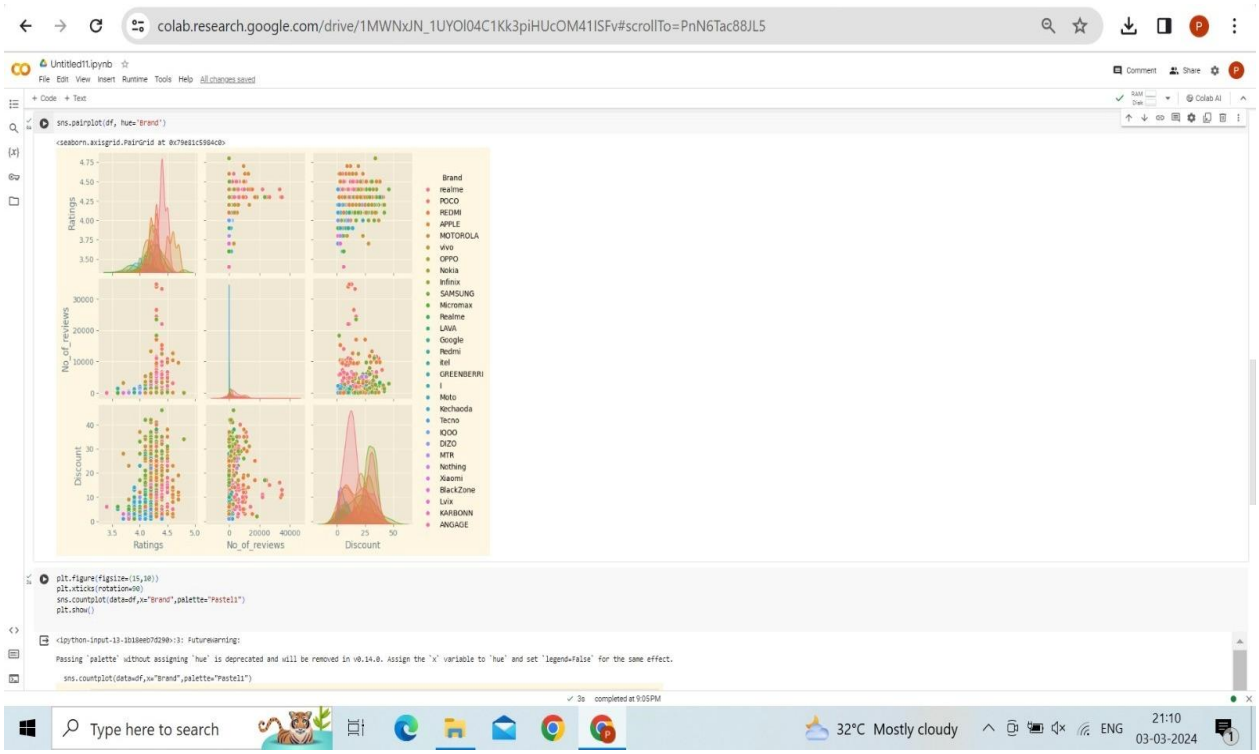
3)Applying community detection algorithms to identify clusters within the network.

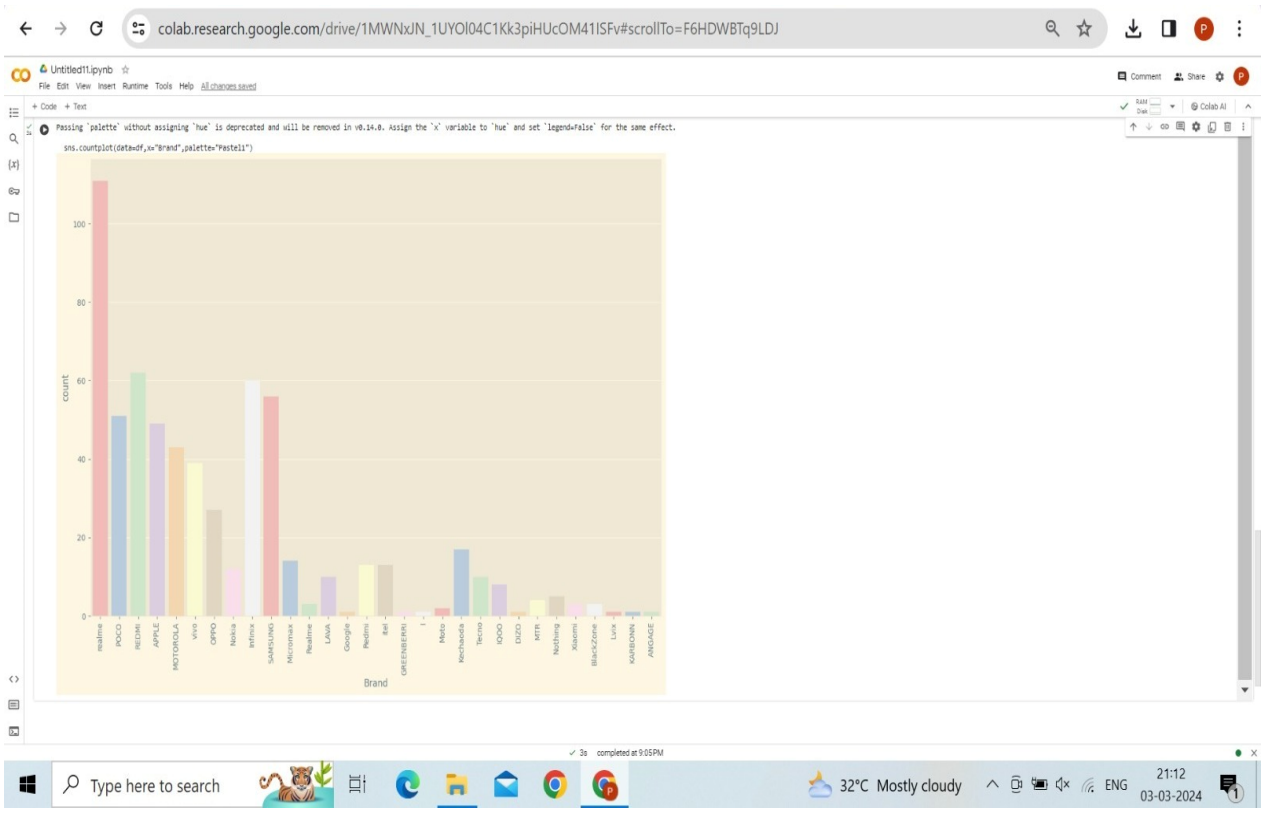
4)Calculating centrality measures to assess node importance and influence.

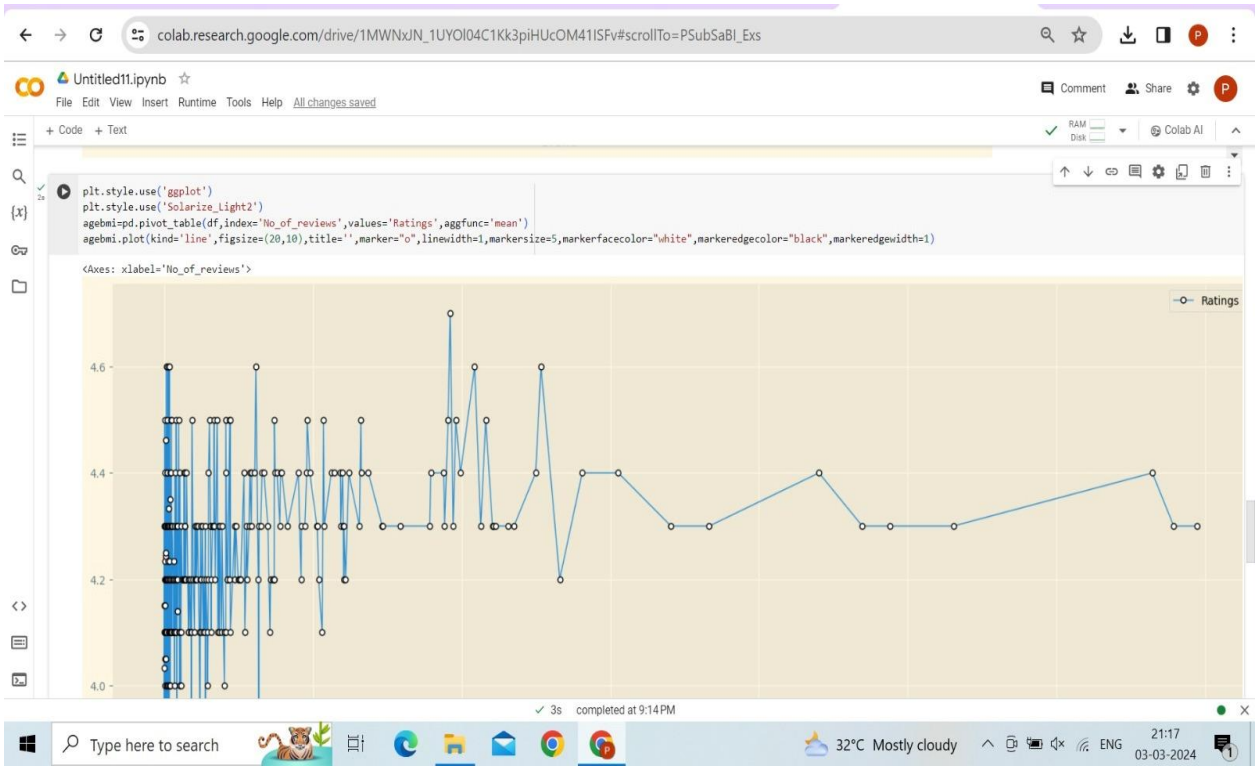


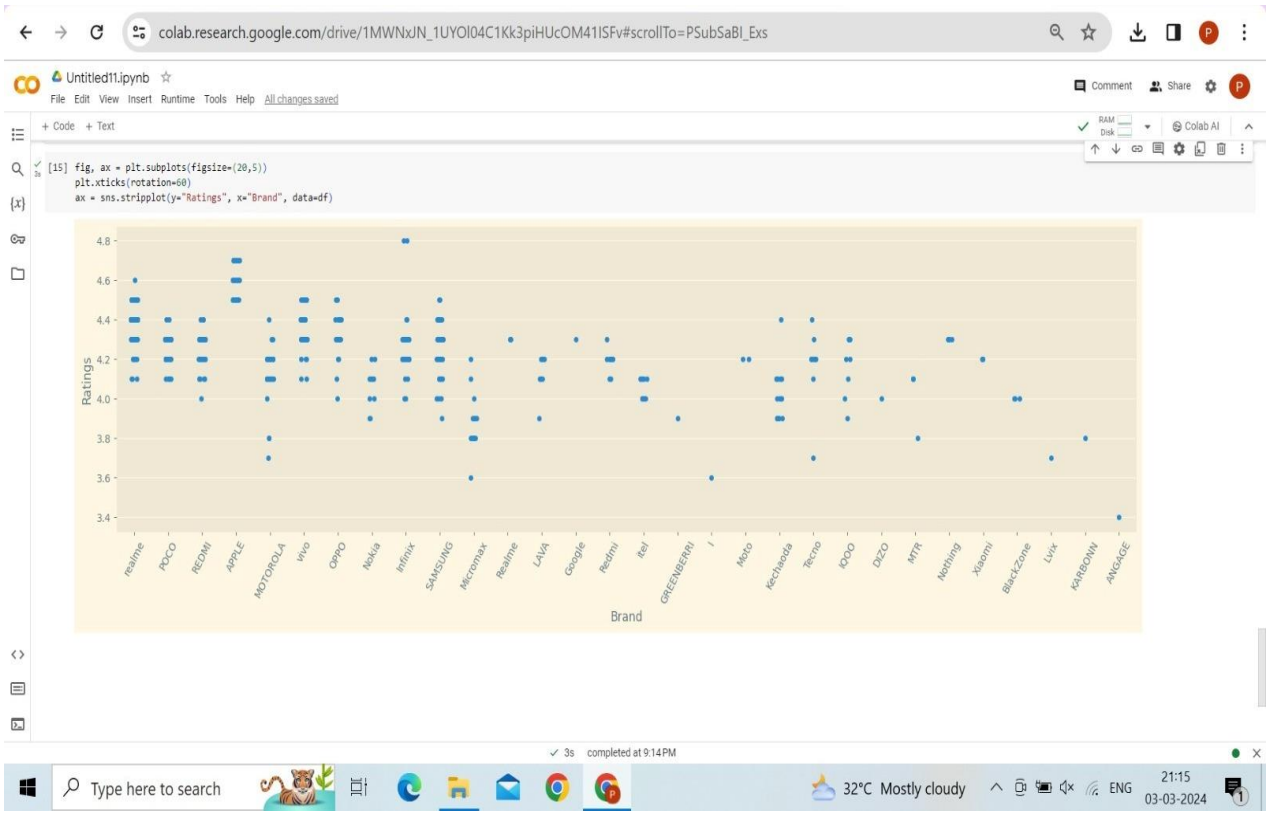

colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=YneQXEqi3p_1
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[9] df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 622 entries, 0 to 621
Data columns (total 9 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 Name 622 non-null object
1 Brand 622 non-null object
2 Ratings 622 non-null float64
3 No_of_ratings 622 non-null object
4 No_of_reviews 622 non-null int64
5 Product_features 622 non-null object
6 MSP 622 non-null object
7 MRP 622 non-null object
8 Discount 622 non-null int64
dtypes: float64(1), int64(2), object(6)
memory usage: 43.9+ KB
0s completed at 8:57 PM
Type here to search
32°C Mostly cloudy
ENG
20:58
03-03-2024

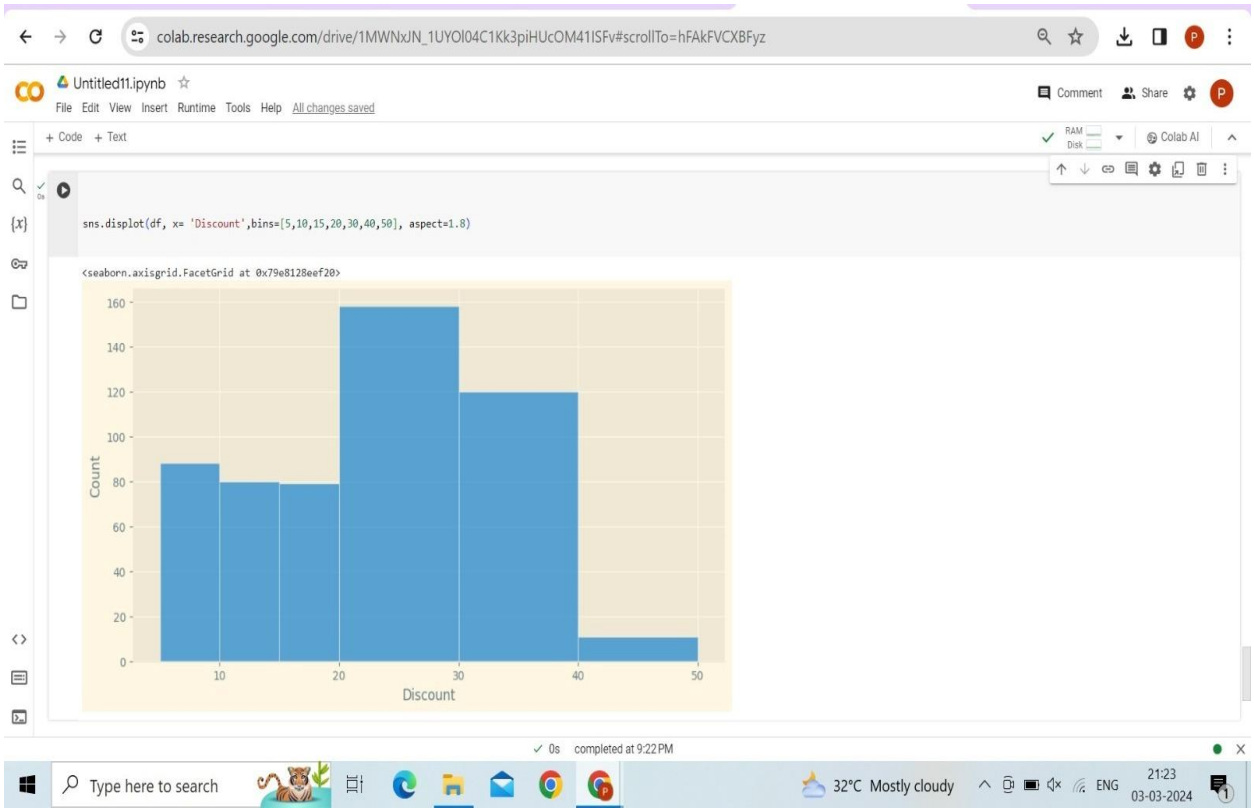

colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=PnN6Tac88JL5
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Colab AI
[10] df.describe()
Ratings No_of_reviews Discount
count 622.000000 622.000000 622.000000
mean 4.267203 3393.704180 18.442122
std 0.188438 5212.468223 11.195627
min 3.400000 0.000000 1.000000
25% 4.200000 201.000000 9.000000
50% 4.300000 1277.000000 18.000000
75% 4.400000 4608.000000 28.000000
max 4.800000 34744.000000 46.000000
[11] df.nunique()
Name 540
Brand 30
Ratings 14
No_of_ratings 321
No_of_reviews 293
Product_features 360
MSP 221
MRP 203
Discount 42
dtype: int64
0s completed at 9:00 PM
Type here to search
32°C Mostly cloudy
ENG
21:01
03-03-2024

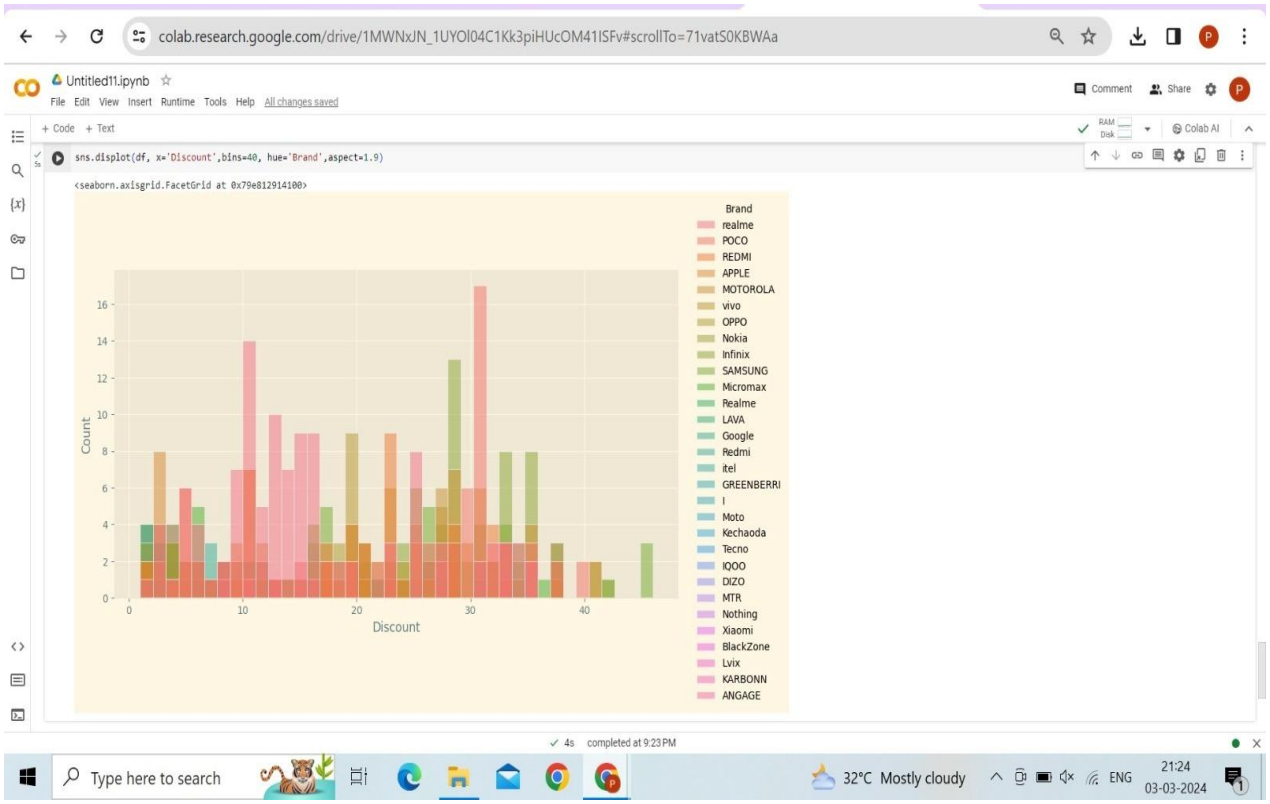

colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=PnN6Tac88JL5
Untitled11.ipynb
sns.pairplot(df, hue='Brand')
Ratings
No_of_reviews
Discount
Brand
realme
POCO
REDMI
APPLE
MOTOROLA
vivo
OPPO
Nokia
Infinix
SAMSUNG
Micromax
Realme
LAVA
Google
Redmi
itel
GREENBERRI
I
Moto
Kechaoda
Tecno
IQOO
DIZO
MTR
Nothing
Xiaomi
BlackZone
Lvix
KARBONN
ANGAGE
plt.figure(figsize=(15,10))
plt.xticks(rotation=90)
sns.countplot(data=df,x="Brand",palette="Pastel1")
plt.show()
FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.
completed at 9:05PM
Type here to search
32°C Mostly cloudy
ENG
21:10
03-03-2024

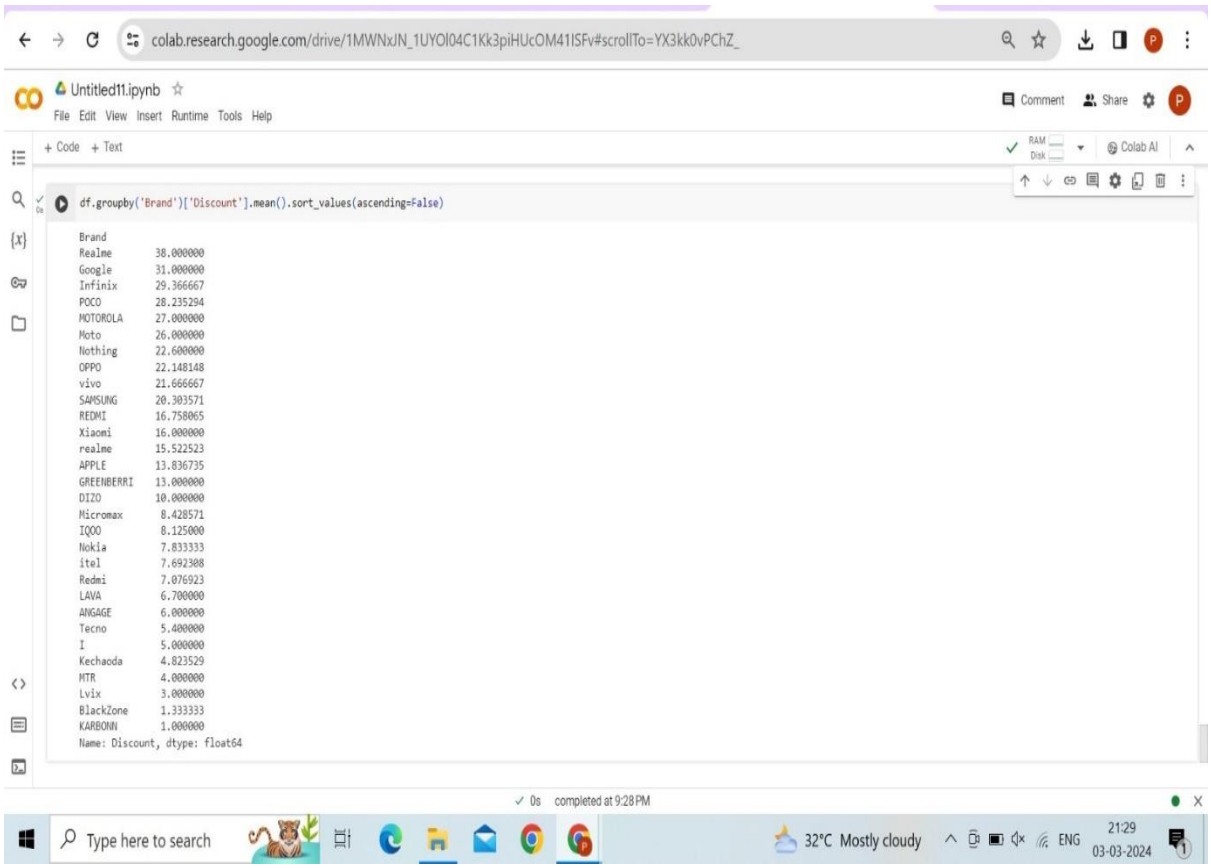

colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=F6HDWBTq9LDJ
Untitled11.ipynb
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.
sns.countplot(data=df,x="Brand",palette="Pastel1")
count
Brand
completed at 9:05PM
Type here to search
32°C Mostly cloudy
ENG
21:12
03-03-2024


colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=PSubSaBI_Exs
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
Colab AI
plt.style.use('ggplot')
plt.style.use('Solarize_Light2')
agebmi=pd.pivot_table(df,index='No_of_reviews',values='Ratings',aggfunc='mean')
agebmi.plot(kind='line',figsize=(20,10),title='',marker="o",linewidth=1,markersize=5,markerfacecolor="white",markeredgecolor="black",markeredgewidth=1)
<Axes: xlabel='No_of_reviews'>
Ratings
4.6
4.4
4.2
4.0
3s completed at 9:14 PM
Type here to search
32°C Mostly cloudy
ENG
21:17
03-03-2024


colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=PSubSaBI_Exs
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
Colab AI
[15] fig, ax = plt.subplots(figsize=(20,5))
plt.xticks(rotation=60)
ax = sns.stripplot(y="Ratings", x="Brand", data=df)
Ratings
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
realme
POCO
REDMI
APPLE
MOTOROLA
vivo
OPPO
Nokia
Infinix
SAMSUNG
Micromax
Realme
LAVA
Google
Redmi
itel
GREENBERRI
I
Moto
Kechaoda
Tecno
IQOO
DIZO
MTR
Nothing
Xiaomi
BlackZone
Lvix
KARBONN
ANGAGE
Brand
3s completed at 9:14 PM
Type here to search
32°C Mostly cloudy
ENG
21:15
03-03-2024


colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=hFAkFVCXBFyz
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Colab AI
sns.displot(df, x= 'Discount',bins=[5,10,15,20,30,40,50], aspect=1.8)
<seaborn.axisgrid.FacetGrid at 0x79e8128eef20>
Count
Discount
0s completed at 9:22 PM
Type here to search
32°C Mostly cloudy
ENG
21:23
03-03-2024


colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=71vatS0KBWAa
Untitled11.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
Colab AI
sns.displot(df, x='Discount',bins=40, hue='Brand',aspect=1.9)
<seaborn.axisgrid.FacetGrid at 0x79e812914100>
Brand
realme
POCO
REDMI
APPLE
MOTOROLA
vivo
OPPO
Nokia
Infinix
SAMSUNG
Micromax
Realme
LAVA
Google
Redmi
itel
GREENBERRI
I
Moto
Kechaoda
Tecno
IQOO
DIZO
MTR
Nothing
Xiaomi
BlackZone
Lvix
KARBONN
ANGAGE
Count
Discount
4s completed at 9:23 PM
Type here to search
32°C Mostly cloudy
ENG
21:24
03-03-2024


colab.research.google.com/drive/1MWNxJN_1UYOl04C1Kk3piHUcOM41ISFv#scrollTo=YX3kk0vPChZ_
Untitled11.ipynb
File Edit View Insert Runtime Tools Help
Comment
Share
+ Code + Text
Colab AI
df.groupby('Brand')['Discount'].mean().sort_values(ascending=False)
Brand
Realme 38.000000
Google 31.000000
Infinix 29.366667
POCO 28.235294
MOTOROLA 27.000000
Moto 26.000000
Nothing 22.600000
OPPO 22.148148
vivo 21.666667
SAMSUNG 20.303571
REDMI 16.758065
Xiaomi 16.000000
realme 15.522523
APPLE 13.836735
GREENBERRI 13.000000
DIZO 10.000000
Micromax 8.428571
IQOO 8.125000
Nokia 7.833333
itel 7.692308
Redmi 7.076923
LAVA 6.700000
ANGAGE 6.000000
Tecno 5.400000
I 5.000000
Kechaoda 4.823529
MTR 4.000000
Lvix 3.000000
BlackZone 1.333333
KARBONN 1.000000
Name: Discount, dtype: float64
0s completed at 9:28 PM
Type here to search
32°C Mostly cloudy
ENG
21:29
03-03-2024

CONCLUSION:

Encouraging learners to implement mining techniques on the Flipkart dataset using Python facilitates hands-on learning and skill development. By leveraging Python's versatility and libraries like NetworkX, learners can explore the intricacies of social network analysis and gain insights into online interactions. Through coding exercises, learners not only deepen their understanding of mining techniques but also acquire valuable skills applicable to various data analysis tasks. Ultimately, harnessing social network analysis with Python empowers learners to navigate complex datasets and extract meaningful insights from online social networks.