Theory Assingment 1

Name:- Nupur S. Batki

PRN no.:- 202401050035

Roll no. :- CC-44

Q. Formulate 20 problem statements for a sales dataset using Numpy and Pandas and Apply Numpy and pandas methods to find the solution for the formulated problem statements.

Ans :-

Dataset taken from :-

https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset?resource=download

Screen shot of codes and questions :-

Files

Analyze your files with code written by Gemini — Upload

..
sample_data
retail_sales_dataset.csv

Disk 70.74 GB available

```
[1] import pandas as pd
    df=pd.read_csv('/content/retail_sales_dataset.csv')
    print(df.head())
```

```
   Transaction ID        Date Customer ID  Gender  Age Product Category  \
0               1  2023-11-24     CUST001    Male   34           Beauty
1               2  2023-02-27     CUST002  Female   26         Clothing
2               3  2023-01-13     CUST003    Male   50      Electronics
3               4  2023-05-21     CUST004    Male   37         Clothing
4               5  2023-05-06     CUST005    Male   30           Beauty

   Quantity  Price per Unit  Total Amount
0         3              50           150
1         2             500          1000
2         1              30            30
3         1             500           500
4         2              50           100
```

```
[32] import matplotlib.pyplot as plt
```

```
[2] print(df.shape)
```

```
(1000, 9)
```

```
[3] print(df.columns)
```

```
Index(['Transaction ID', 'Date', 'Customer ID', 'Gender', 'Age',
       'Product Category', 'Quantity', 'Price per Unit', 'Total Amount'],
      dtype='object')
```

theory assingment.ipynb - Cola

colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnlIpiQE?authuser=0#scrollTo=PHzFf9oUDJyD

theory assingment.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text

RAM Disk

Files

Analyze your files with code written by Gemini — Upload

..
sample_data
retail_sales_dataset.csv

Disk 70.74 GB available

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Transaction ID    1000 non-null   int64
 1   Date              1000 non-null   object
 2   Customer ID       1000 non-null   object
 3   Gender            1000 non-null   object
 4   Age               1000 non-null   int64
 5   Product Category  1000 non-null   object
 6   Quantity          1000 non-null   int64
 7   Price per Unit    1000 non-null   int64
 8   Total Amount      1000 non-null   int64
dtypes: int64(5), object(4)
memory usage: 70.4+ KB
None
```

1. Total no. of Orders

```
[6] total_orders = df['Transaction ID'].nunique()
    print("Total Orders:", total_orders)
```

```
Total Orders: 1000
```

2. Total Sales Amount

0s completed at 9:46 PM

31°C Clear — Search — ENG IN — 21:51 28-04-2025

theory assingment.ipynb - Cola

colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnlIpiQE?authuser=0#scrollTo=PHzFf9oUDJyD

theory assingment.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text

RAM Disk

Files

Analyze your files with code written by Gemini — Upload

..
sample_data
retail_sales_dataset.csv

Disk 70.74 GB available

2. Total Sales Amount

```
[8] total_sales = df['Quantity'].sum()
    print("Total Sales:", total_sales)
```

```
Total Sales: 2514
```

3. Average Sales Per Order

```
[9] average_sales = df['Quantity'].mean()
    print("Average Sales per Order:", average_sales)
```

```
Average Sales per Order: 2.514
```

4.Maximum Quantity Sold in a Single Order

```
[10] max_quantity = df['Quantity'].max()
     print("Maximum Quantity Sold:", max_quantity)
```

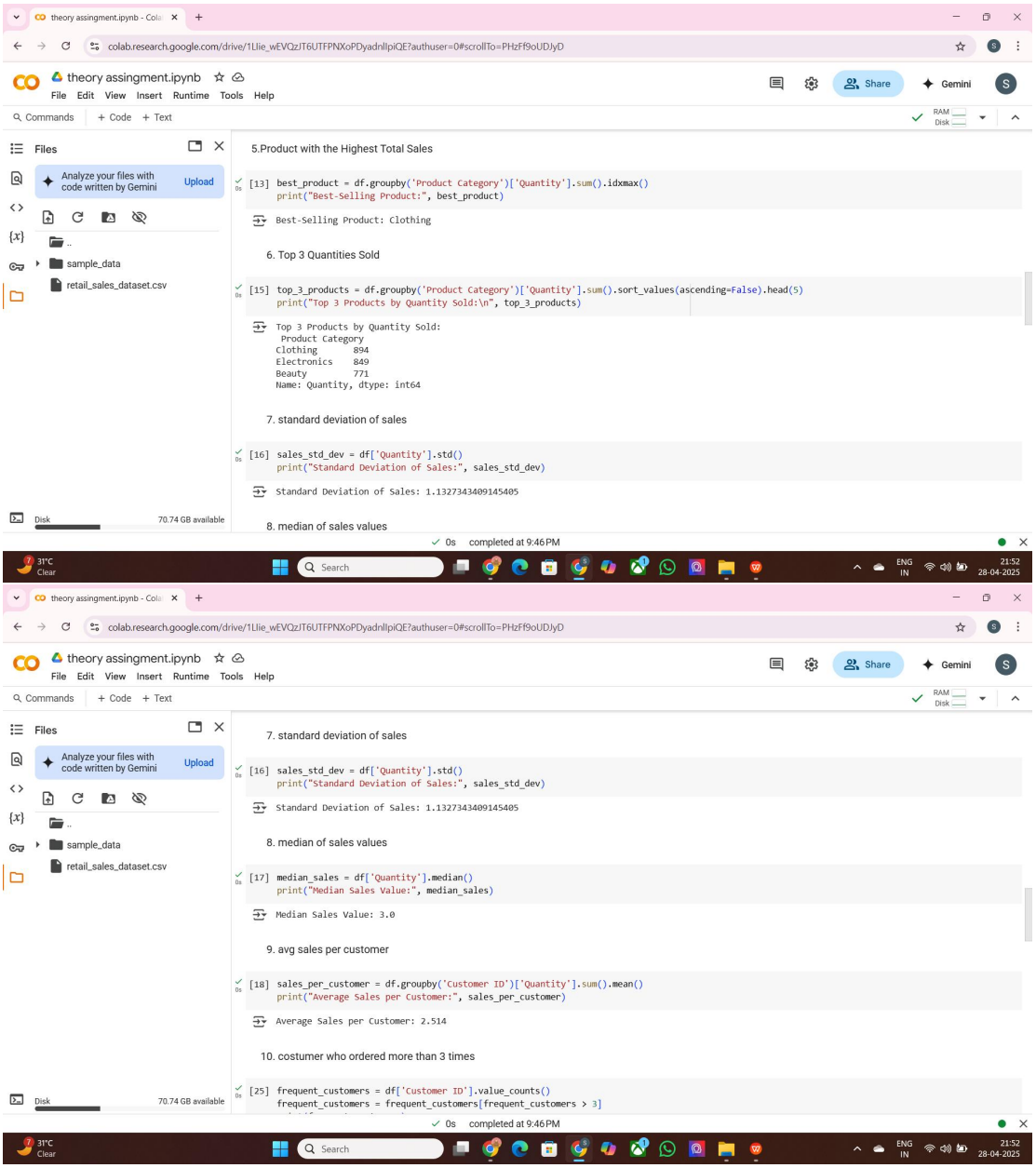
```
Maximum Quantity Sold: 4
```

5.Product with the Highest Total Sales

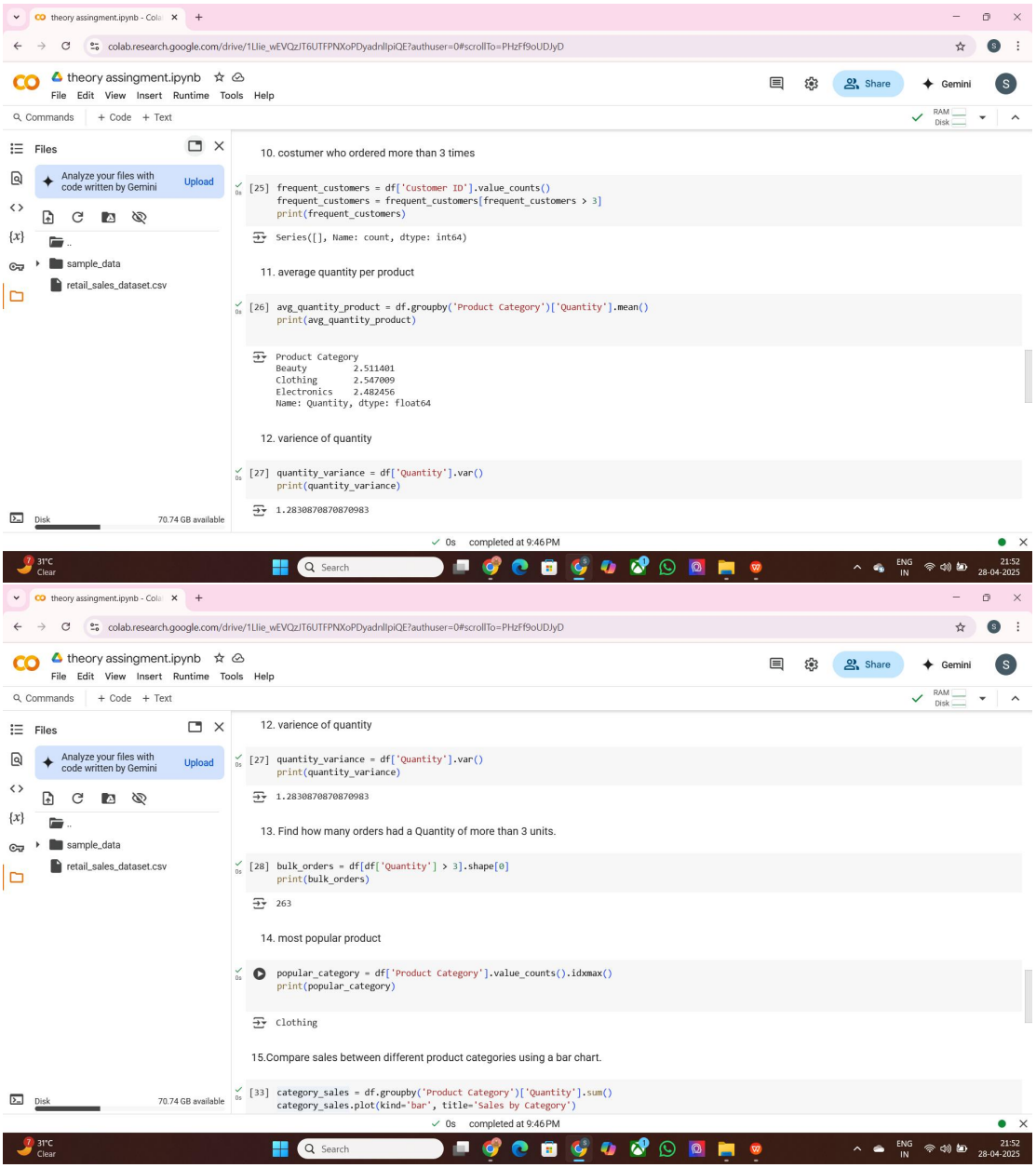
```
[13] best_product = df.groupby('Product Category')['Quantity'].sum().idxmax()
     print("Best-Selling Product:", best_product)
```

0s completed at 9:46 PM

31°C Clear — Search — ENG IN — 21:51 28-04-2025


theory assingment.ipynb - Cola
colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnIlpiQE?authuser=0#scrollTo=PHzFf9oUDJyD
theory assingment.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
retail_sales_dataset.csv
Disk
70.74 GB available
5.Product with the Highest Total Sales
[13] best_product = df.groupby('Product Category')['Quantity'].sum().idxmax()
print("Best-Selling Product:", best_product)
Best-Selling Product: Clothing
6. Top 3 Quantities Sold
[15] top_3_products = df.groupby('Product Category')['Quantity'].sum().sort_values(ascending=False).head(5)
print("Top 3 Products by Quantity Sold:\n", top_3_products)
Top 3 Products by Quantity Sold:
Product Category
Clothing 894
Electronics 849
Beauty 771
Name: Quantity, dtype: int64
7. standard deviation of sales
[16] sales_std_dev = df['Quantity'].std()
print("Standard Deviation of Sales:", sales_std_dev)
Standard Deviation of Sales: 1.1327343409145405
8. median of sales values
0s completed at 9:46 PM
31°C Clear
Search
ENG IN
21:52 28-04-2025

theory assingment.ipynb - Cola
colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnIlpiQE?authuser=0#scrollTo=PHzFf9oUDJyD
theory assingment.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
retail_sales_dataset.csv
Disk
70.74 GB available
7. standard deviation of sales
[16] sales_std_dev = df['Quantity'].std()
print("Standard Deviation of Sales:", sales_std_dev)
Standard Deviation of Sales: 1.1327343409145405
8. median of sales values
[17] median_sales = df['Quantity'].median()
print("Median Sales Value:", median_sales)
Median Sales Value: 3.0
9. avg sales per customer
[18] sales_per_customer = df.groupby('Customer ID')['Quantity'].sum().mean()
print("Average Sales per Customer:", sales_per_customer)
Average Sales per Customer: 2.514
10. costumer who ordered more than 3 times
[25] frequent_customers = df['Customer ID'].value_counts()
frequent_customers = frequent_customers[frequent_customers > 3]
0s completed at 9:46 PM
31°C Clear
Search
ENG IN
21:52 28-04-2025


theory assingment.ipynb - Cola
colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnIlpiQE?authuser=0#scrollTo=PHzFf9oUDJyD
theory assingment.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
RAM
Disk
Files
Analyze your files with code written by Gemini
Upload
sample_data
retail_sales_dataset.csv
Disk
70.74 GB available
10. costumer who ordered more than 3 times
[25] frequent_customers = df['Customer ID'].value_counts()
frequent_customers = frequent_customers[frequent_customers > 3]
print(frequent_customers)
Series([], Name: count, dtype: int64)
11. average quantity per product
[26] avg_quantity_product = df.groupby('Product Category')['Quantity'].mean()
print(avg_quantity_product)
Product Category
Beauty 2.511401
Clothing 2.547009
Electronics 2.482456
Name: Quantity, dtype: float64
12. varience of quantity
[27] quantity_variance = df['Quantity'].var()
print(quantity_variance)
1.2830870870870983
0s completed at 9:46 PM

theory assingment.ipynb - Cola
colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnIlpiQE?authuser=0#scrollTo=PHzFf9oUDJyD
theory assingment.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Files
Analyze your files with code written by Gemini
Upload
sample_data
retail_sales_dataset.csv
Disk
70.74 GB available
12. varience of quantity
[27] quantity_variance = df['Quantity'].var()
print(quantity_variance)
1.2830870870870983
13. Find how many orders had a Quantity of more than 3 units.
[28] bulk_orders = df[df['Quantity'] > 3].shape[0]
print(bulk_orders)
263
14. most popular product
popular_category = df['Product Category'].value_counts().idxmax()
print(popular_category)
Clothing
15.Compare sales between different product categories using a bar chart.
[33] category_sales = df.groupby('Product Category')['Quantity'].sum()
category_sales.plot(kind='bar', title='Sales by Category')
0s completed at 9:46 PM

theory assingment.ipynb - Cola

colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnllpiQE?authuser=0#scrollTo=PHzFf9oUDJyD

theory assingment.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text

RAM Disk

Files

Analyze your files with code written by Gemini Upload

..

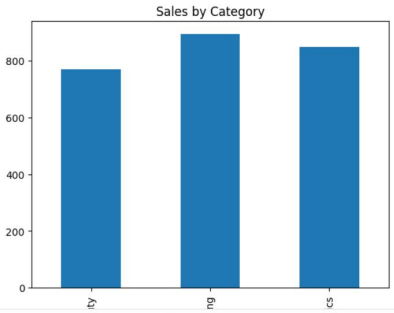
sample_data

retail_sales_dataset.csv

Disk 70.74 GB available

15.Compare sales between different product categories using a bar chart.

```
category_sales = df.groupby('Product Category')['Quantity'].sum()
category_sales.plot(kind='bar', title='Sales by Category')
plt.show()
```



0s completed at 9:46 PM

31°C Clear

Search

ENG IN 21:53 28-04-2025

theory assingment.ipynb - Cola

colab.research.google.com/drive/1Llie_wEVQzJT6UTFPNXoPDyadnllpiQE?authuser=0#scrollTo=PHzFf9oUDJyD

theory assingment.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text

RAM Disk

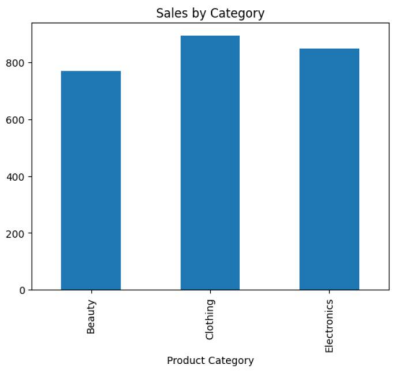
Files

Analyze your files with code written by Gemini Upload

..

sample_data

retail_sales_dataset.csv



Disk 70.74 GB available

0s completed at 9:46 PM

31°C Clear

Search

ENG IN 21:53 28-04-2025