

Task-1 : Exploratory Data Analysis (EDA) and Business Insights

1.Introduction:

1. Customers.csv

Contains customer profile details like unique IDs, names, regions, and signup dates. It provides demographic data used for analyzing customer behavior and segmentation.

2. Products.csv

Contains details about each product, including product ID, name, category, and price. It helps analyze customer purchase preferences and spending patterns.

3. Transactions.csv

Records individual transactions with customer ID, product ID, date, quantity, and total value. It is the key dataset for understanding purchase behavior and calculating customer metrics like total spend and frequency.

2.Dataset Relationships

1. Customers.csv:

- Acts as the primary dataset for customer-related features.
- Links to Transactions.csv using CustomerID.

2. Products.csv:

- Provides product-related features for Transactions.csv.
- Links to Transactions.csv using ProductID.

3. Transactions.csv:

- Acts as the key dataset for **behavioral analytics**.
- Combines customer and product information to calculate spending patterns and preferences.

3. Business Insights:

1. Regional Customer Distribution and Revenue Analysis

- **Insight:** Customers from specific regions (e.g., North America, Europe) contribute a higher percentage of total transactions and revenue. This insight can help prioritize regions for marketing campaigns.

2. Product Category Insights

- **Insight:** Certain product categories (e.g., Electronics, Apparel) dominate sales in terms of revenue and quantity sold. This indicates which products to stock more frequently.

3. Customer Behaviour Analysis

- **Insight:** Customers from specific regions (e.g., North America) have higher average spending compared to others. Identifying high-value customers can aid in personalized marketing.

4. Sales Trends Over Time

- **Insight:** Sales peak during certain months, indicating seasonal demand. Companies can time promotional campaigns for these periods.

5. High-Demand Products

- **Insight:** Certain products (e.g., top 5 bestsellers) account for a significant share of sales. Stocking these products is critical for maintaining high revenue.

4. Recommendations

- Suggestions for marketing strategies.
- Inventory management tips based on product demand.
- Region-focused sales campaigns.

5. Exploratory Data Analysis(EDA):

memory usage: 54.8+ KB

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity
0	T00001	C0199	P007	2024-08-25 12:38:23	1
1	T00112	C0146	P067	2024-05-27 22:23:54	1
2	T00166	C0127	P067	2024-04-25 07:38:55	1
3	T00272	C0087	P067	2024-03-26 22:55:37	2
4	T00363	C0070	P067	2024-03-21 15:10:10	3

	TotalValue	Price
0	300.68	300.68
1	300.68	300.68
2	300.68	300.68
3	601.36	300.68
4	902.04	300.68

Missing Values in Customers:

CustomerID	CustomerName	Region	SignupDate
0	0	0	0

dtype: int64

Missing Values in Products:

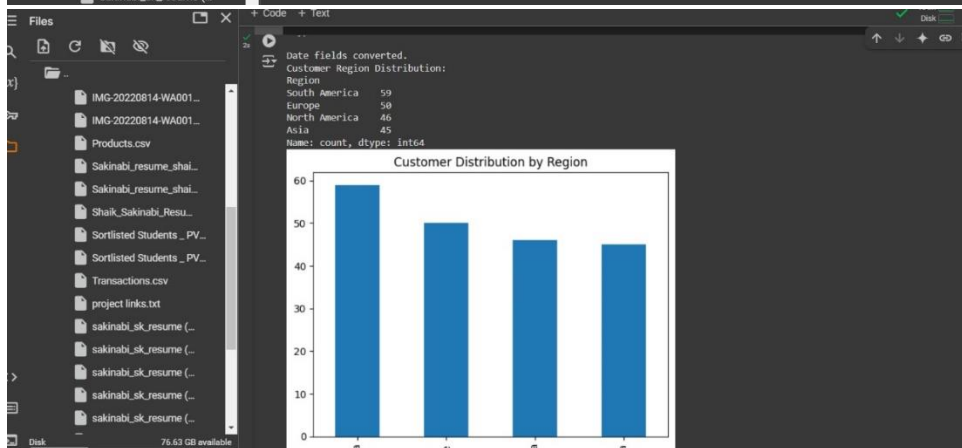
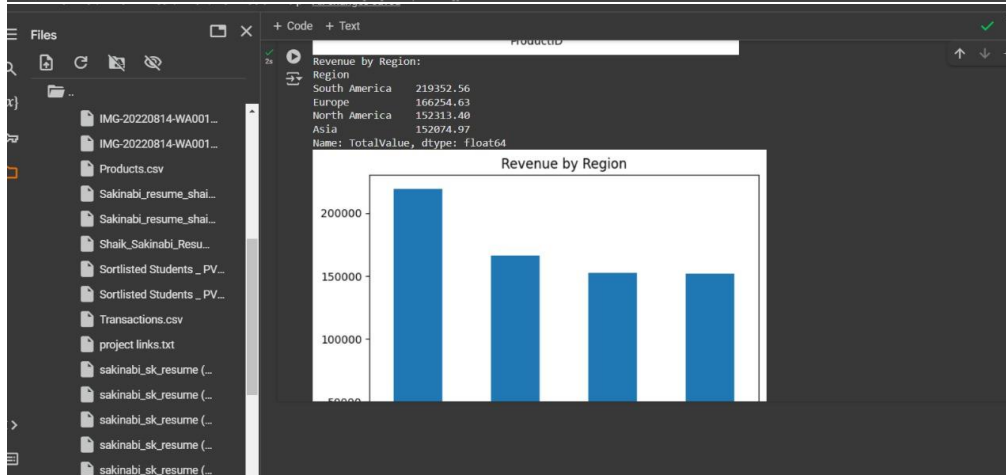
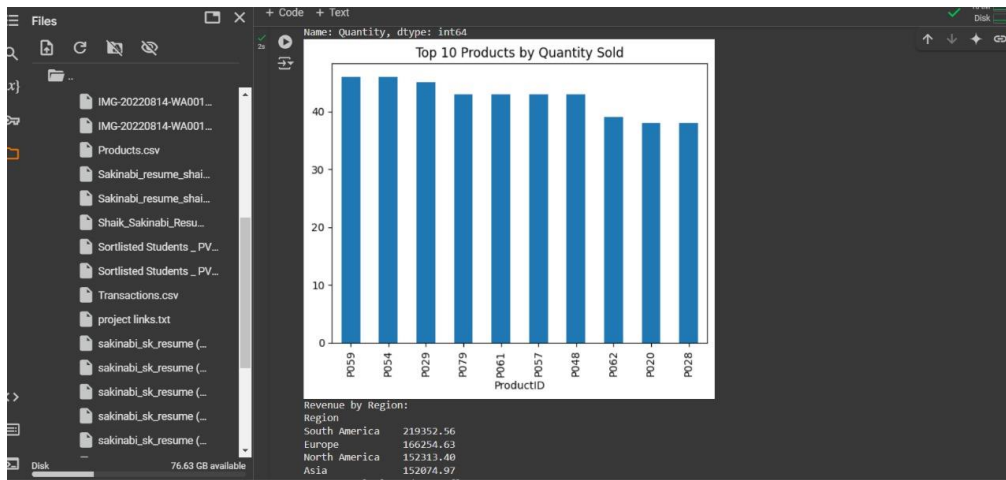
ProductID	ProductName	Category	Price
0	0	0	0

dtype: int64

Missing Values in Transactions:

TransactionID	CustomerID	ProductID	TransactionDate
0	0	0	0

dtype: int64





Sakinabi Shaik_EDA.ipynb

File Edit View Insert Runtime Tools Help All changes saved

combined_analysis()

```
if __name__ == "__main__":
    main()
```

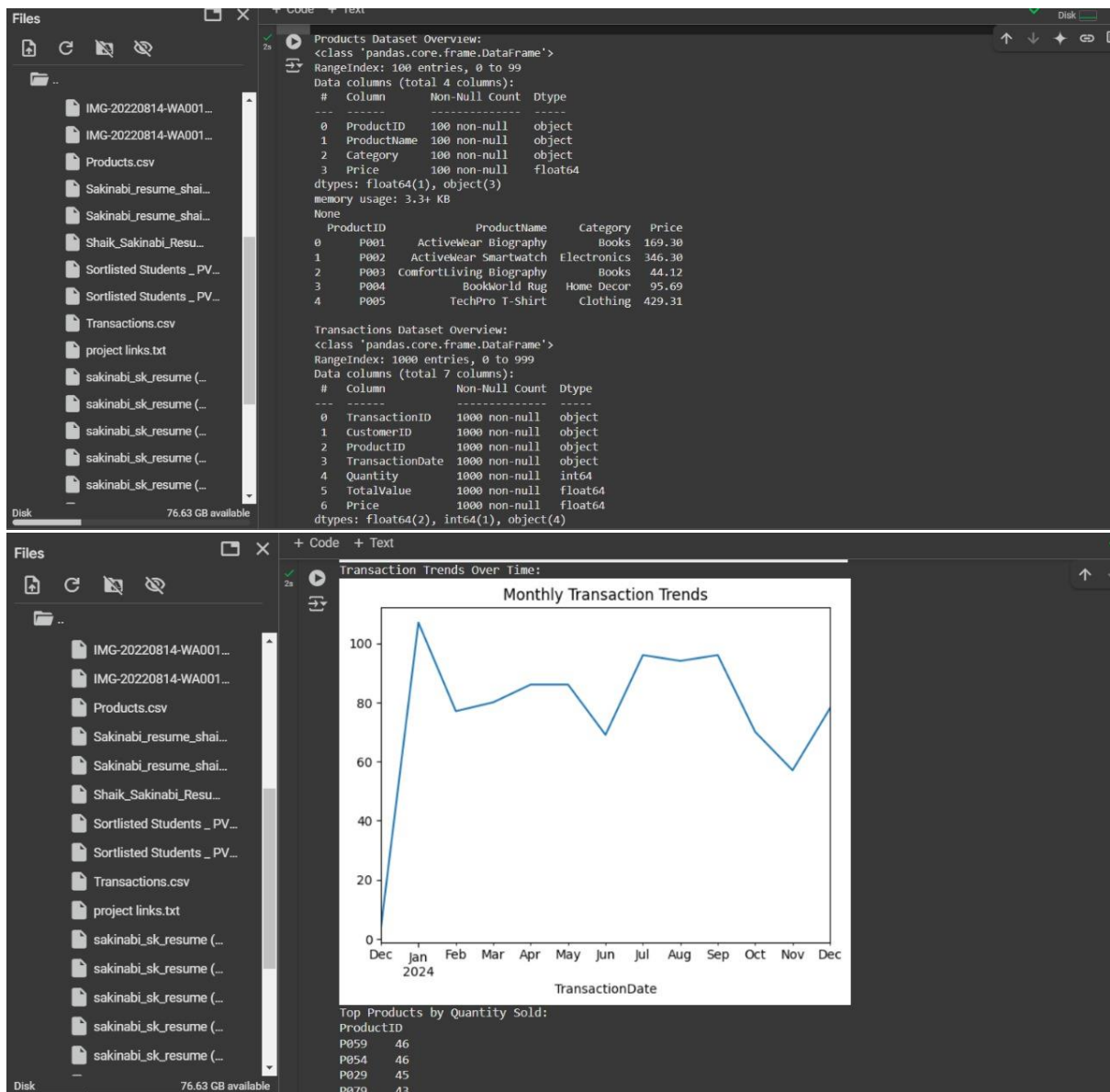
Customers Dataset Overview:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   column      Non-Null Count  Dtype
---  --
 0   CustomerID  200 non-null     object
 1   CustomerName 200 non-null     object
 2   Region      200 non-null     object
 3   SignedDate  200 non-null     object
dtypes: object(4)
memory usage: 14.4+ KB
```

	CustomerID	CustomerName	Region	SignedDate
0	CO001	Lawrence Carroll	South America	2022-07-10
1	CO002	Elizabeth Letz	Asia	2022-02-11
2	CO003	Michael Rivera	South America	2024-01-07
3	CO006	Kathleen Rodriguez	South America	2022-10-09
4	CO005	Laura Weber	Asia	2022-06-15

Products Dataset Overview:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 4 columns):
 #   column      Non-Null Count  Dtype
---  --
 0   ProductID   100 non-null     object
 1   ProductName 100 non-null     object
```



Recommendations:

The EDA provided valuable insights into customer behavior, product performance, and transaction trends:

1. **Customer Insights:** Key regions driving transactions were identified, revealing the geographic distribution of the customer base. High-value customers were concentrated in specific regions.
2. **Product Insights:** Certain product categories, like Electronics and Apparel, showed consistent demand, indicating popular segments for sales growth.
3. **Transaction Trends:** The analysis highlighted peak transaction periods, enabling businesses to optimize marketing campaigns during these high-activity times.

4. Revenue Drivers: High-value customers contributed disproportionately to revenue, emphasizing the importance of loyalty programs and personalized strategies for retention.
5. Opportunities for Growth: Low-activity regions and underperforming product categories can be targeted for strategic expansion or promotional activities.