Exploratory Data Analysis(EDA) and Business insights

Overview

The provided Python script is a comprehensive analysis workflow for exploring and visualizing customer, product, and transaction data from three CSV files: Customers.csv, Products.csv, and Transactions.csv. The analysis involves data exploration, summary statistics, and visualization to gain insights into customer distribution by region and product purchasing patterns.

Dependencies

To execute the script, the following Python libraries are required:

pandas: For data manipulation and reading CSV files. numpy:

For numerical operations. matplotlib: For data visualization.

seaborn: For aesthetically pleasing statistical plots.

Steps in the Script

Install and Import Necessary Libraries pip install pandas

scikit-learn numpy tensorflow matplotlib import numpy as

np import pandas as pd import matplotlib.pyplot as plt

import seaborn as sns

Libraries are installed and imported to handle data manipulation, computation, and visualization.

Load Data The script reads data from three CSV files:

Customers.csv

Products.csv

Transactions.csv

```
Cust_data = pd.read_csv('Customers.csv')
```

```
Prod_data = pd.read_csv('Products.csv')
```

```
Trans_data = pd.read_csv('Transactions.csv')
```

Explore and Preview Data The .head() function displays the first few rows of each dataset for an initial overview:

```
Cust_data.head()
Prod_data.head()

Trans_data.head()
```

Extract Key Data Insights

Customer Data: Count unique values in the CustomerID, CustomerName, Region, and SignupDate columns.

Product Data: Count unique values in the ProductID, ProductName, Category, and Price columns.

Transaction Data: Count unique values in the TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price columns.

```
Cust_data_info = {
    'CustomerID': len(Cust_data['CustomerID'].unique()),
    'CustomerName': len(Cust_data['CustomerName'].unique()),
    'Region': len(Cust_data['Region'].unique()),
    'SignupDate': len(Cust_data['SignupDate'].unique())
}
Prod_data_info = {
    'ProductID': len(Prod_data['ProductID'].unique()),
    'ProductName': len(Prod_data['ProductName'].unique()),
    'Category': len(Prod_data['Category'].unique()),
    'Price': len(Prod_data['Price'].unique())
}
Trans_data_info = {
    'TransactionID': len(Trans_data['TransactionID'].unique()),
    'CustomerID': len(Trans_data['CustomerID'].unique()),
    'ProductID': len(Trans_data['ProductID'].unique()),
    'TransactionDate': len(Trans_data['TransactionDate'].unique())
```

'Quantity': len(Trans_data['Quantity'].unique()),

'TotalValue': len(Trans_data['TotalValue'].unique()),

'Price': len(Trans_data['Price'].unique())

}

This provides summary statistics about the datasets.

Dataset Metadata The .info() method is used to display the structure of the datasets, including column names, data types, and null values:

Cust_data.info()

Prod_data.info()

Trans_data.info()

Analyze Regions and Product Popularity

Unique Regions: Count the number of unique regions in the customer data.

Region Distribution: Count the number of customers per region. Top

Products: Identify the top 5 most frequently purchased products.

unique_regions = Cust_data['Region'].nunique() region_counts

= Cust_data['Region'].value_counts() top_products =

Trans_data['ProductID'].value_counts().head(5)

Visualize Insights

Customer Distribution by Region: The number of customers per region is visualized using a bar chart.

plt.figure(figsize=(8, 5)) sns.barplot(x=region_counts.index,

y=region_counts.values, palette="pastel") plt.title("Customer Distribution by

Region", fontsize=14) plt.ylabel("Number of Customers") plt.xlabel("Region")

plt.show()

Top 5 Products by Purchases: The top 5 most purchased products are visualized using another bar chart.

```
plt.figure(figsize=(8, 5)) sns.barplot(x=top_products.index,
y=top_products.values, palette="muted") plt.title("Top 5 Products by
Number of Purchases", fontsize=14) plt.ylabel("Number of Purchases")
plt.xlabel("ProductID") plt.show()
```

Outputs

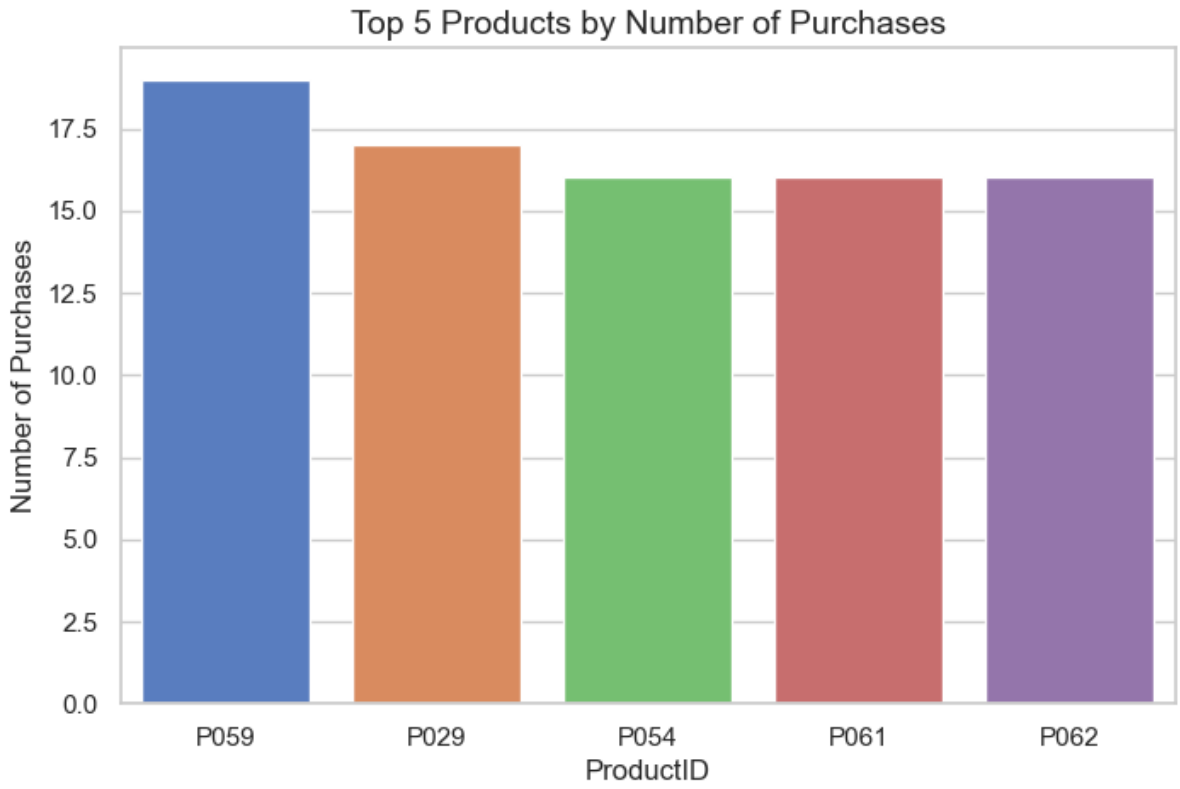
Summary Statistics:

- The number of unique customers, regions, products, and transactions.
- Insights into the diversity of product categories and pricing.

Visualizations:

- A bar chart showing customer distribution across regions.

A bar chart showcasing the top 5 products by purchase count.



Key Findings:

The most populous regions in terms of customers.

The most popular products based on purchase frequency.