Name: Kalhar Patel

Mail: kalharpatel9@gmail.com

Documentation for Task 1: Exploratory Data Analysis (EDA):

1. Objective:

This task focuses on performing Exploratory Data Analysis (EDA) to uncover patterns, insights, and trends from the eCommerce Transactions dataset. The analysis covers customer, product, and transaction data to identify actionable business insights.

2. Dataset Overview:

The analysis was conducted on the following datasets:

1. Customers.csv: Contains details such as CustomerID, CustomerName, Region, and SignupDate.
2. Products.csv: Includes ProductID, ProductName, Category, and Price.
3. Transactions.csv: Consists of TransactionID, CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

3. Methodology:

A. Data Loading and Preprocessing

The datasets were loaded using pandas for analysis. Basic data exploration was performed to understand the structure of the data. Date columns in the Customers and Transactions datasets were converted to datetime format.

B. Customer Analysis

Key statistics were derived, such as total customers, customer distribution by region, and signup date range.

C. Product Analysis

The analysis included calculating the total number of products, distribution by category, and statistical summaries of product prices.

D. Transaction Analysis

Key metrics such as total transactions, total revenue, average transaction value, and daily revenue trends were calculated.

Name: Kalhar Patel

Mail: kalharpatel9@gmail.com

E. Customer Purchase Behavior

Metrics like purchase frequency and total spending per customer were derived. Product performance was analyzed based on total sales, revenue, and units sold by category.

4. *Code Highlights*:

Loading Datasets

```
customers_df = pd.read_csv('Customers.csv', encoding='utf8')
products_df = pd.read_csv('Products.csv', encoding='utf8')
transactions_df = pd.read_csv('Transactions.csv', encoding='utf8')
```

Customer Statistics

```
customer_stats = {
  'total_customers': len(customers_df),
  'customers_by_region': customers_df['Region'].value_counts(),
  'signup_date_range': (customers_df['SignupDate'].min(),
customers_df['SignupDate'].max())
}
```

Product Performance

```
product_performance = transactions_df.merge(products_df, on='ProductID')\
  .groupby('Category')\
  .agg({
    'TransactionID': 'count',
    'TotalValue': 'sum',
    'Quantity': 'sum'
  })\
  .rename(columns={
    'TransactionID': 'total_sales',
    'TotalValue': 'total_revenue',
    'Quantity': 'units_sold'
  })
```

5. *Key Business Insights:*

Based on the analysis, the following key business insights were identified:
1. Total Revenue: $689,995.56.
2. Average Transaction Value: $690.00.
3. Most Active Customer Region: South America.
4. Best Performing Category: Books.
5. Customer Purchase Frequency: 5.03 transactions per customer.