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

Derive Business Insights:

Example Insights:

1. **Customer Distribution**: The highest number of customers come from North America, contributing 45% of total transactions. Marketing efforts should focus on this region.

2. **Product Categories**: Electronics and Fashion categories contribute 70% of total revenue, suggesting these are high-priority product lines.

3. **Seasonal Demand**: Sales spike during November and December, indicating strong holiday demand. Stock and promotional strategies should align with this trend.

4. **Customer Engagement**: 40% of customers make repeat purchases within six months, showing opportunities for loyalty programs.

5. **Top Products**: Premium electronic gadgets dominate sales, emphasizing the importance of maintaining competitive pricing and availability.

Step 1: Understand the Dataset

- **Files Overview**:

    1. Customers.csv: Contains customer details like ID, name, region, and signup date.

    2. Products.csv: Contains product details like ID, name, category, and price.

    3. Transactions.csv: Contains transaction details, including customer-product relations, transaction values, and dates.

Step 2: Data Preparation

1. **Import Required Libraries**:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

Load the Datasets: Download the datasets from the provided links and load them into dataframes:

customers = pd.read_csv("Customers.csv")

products = pd.read_csv("Products.csv")

transactions = pd.read_csv("Transactions.csv")

Inspect the Data: Use these commands to understand the structure of the datasets:

print(customers.head())

print(products.head())

print(transactions.head())

Check for Missing Values and Duplicates: Identify and handle missing or duplicate data:

print(customers.isnull().sum())

print(products.isnull().sum())

print(transactions.isnull().sum())

customers.drop_duplicates(inplace=True)

products.drop_duplicates(inplace=True)

transactions.drop_duplicates(inplace=True)

Merge the Datasets: Combine the datasets for a unified analysis:

merged_data = transactions.merge(customers, on='CustomerID').merge(products, on='ProductID')

Step 3: Perform Exploratory Data Analysis (EDA)

General Data Summary:

- **Explore the number of unique customers, products, and transactions**

print("Number of customers:", customers['CustomerID'].nunique())

print("Number of products:", products['ProductID'].nunique())

print("Number of transactions:", transactions['TransactionID'].nunique())

Customer Demographics:

- **Analyze customer distribution by region**

```
sns.countplot(data=customers, x='Region', palette='viridis')

plt.title("Customer Distribution by Region")

plt.show()

sns.countplot(data=customers, x='Region', palette='viridis')

plt.title("Customer Distribution by Region")

plt.show()
```

Transaction Trends:

- **Analyze transaction trends over time**

```
merged_data['TransactionDate'] =
pd.to_datetime(merged_data['TransactionDate'])

merged_data['Month'] = merged_data['TransactionDate'].dt.to_period('M')

monthly_sales = merged_data.groupby('Month')['TotalValue'].sum()

monthly_sales.plot(kind='line', marker='o', figsize=(12,6))

plt.title("Monthly Sales Trends")

plt.xlabel("Month")

plt.ylabel("Total Sales (USD)")

plt.show()
```