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

Below is the implementation and insights generation process for Task 1.

EDA Code

```
# Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load datasets
customers = pd.read_csv("Customers.csv")
products = pd.read_csv("Products.csv")
transactions = pd.read_csv("Transactions.csv")

# Convert date columns to datetime format
customers["SignupDate"] = pd.to_datetime(customers["SignupDate"])
transactions["TransactionDate"] =
pd.to_datetime(transactions["TransactionDate"])

# Merge datasets
merged_data = transactions.merge(customers, on="CustomerID", how="left")
merged_data = merged_data.merge(products, on="ProductID", how="left")

# 1. Distribution of Customers by Region
plt.figure(figsize=(8, 5))
sns.countplot(data=customers, x="Region", palette="viridis",
order=customers["Region"].value_counts().index)
plt.title("Customer Distribution by Region", fontsize=16)
plt.ylabel("Number of Customers")
plt.xlabel("Region")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# 2. Signup Trends Over the Years
customers["SignupYear"] = customers["SignupDate"].dt.year
signup_trends = customers.groupby("SignupYear").size()
plt.figure(figsize=(10, 6))
signup_trends.plot(kind="bar", color="teal", alpha=0.8)
plt.title("Signup Trends Over the Years", fontsize=16)
plt.xlabel("Year")
plt.ylabel("Number of Signups")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# 3. Top-Selling Product Categories by Revenue
category_sales = merged_data.groupby("Category")
["TotalValue"].sum().sort_values(ascending=False)
plt.figure(figsize=(10, 6))
category_sales.plot(kind="bar", color="orange", alpha=0.8)
plt.title("Top-Selling Product Categories by Revenue", fontsize=16)
plt.ylabel("Total Revenue (USD)")
plt.xlabel("Product Category")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# 4. Transactions Over Time
transactions_over_time =
merged_data.groupby(merged_data["TransactionDate"].dt.date)["TotalValue"].sum()
plt.figure(figsize=(12, 6))
```

```
transactions_over_time.plot(color="blue", alpha=0.8)
plt.title("Transactions Over Time", fontsize=16)
plt.ylabel("Total Revenue (USD)")
plt.xlabel("Transaction Date")
plt.tight_layout()
plt.show()

# 5. High-Value Customers
high_value_customers = merged_data.groupby("CustomerID")
["TotalValue"].sum().sort_values(ascending=False).head(10)
print("Top 10 High-Value Customers:\n", high_value_customers)
```

Business Insights

Here are 5 insights derived from the EDA:

1. Regional Customer Distribution:

The majority of customers are concentrated in specific regions, indicating the need for targeted marketing campaigns. Regions with fewer customers present opportunities for expansion.

2. Signup Trends:

A significant increase in customer signups was observed in recent years, suggesting effective customer acquisition strategies. Analyzing factors that drove these signups can help replicate success.

3. Top Revenue-Generating Categories:

Certain product categories dominate revenue generation. These categories should be prioritized for promotions and inventory stocking to maximize profits.

4. Seasonal Transaction Patterns:

Revenue trends show peaks during specific time periods, possibly indicating seasonal demand. This insight can help optimize inventory and marketing strategies during high-demand months.

5. High-Value Customers:

A small percentage of customers contribute to the majority of revenue. Loyalty programs or personalized offers can be designed to retain these customers and increase lifetime value.