#!/usr/bin/env python

coding: utf-8

In[1]:

import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns from sklearn.preprocessing import StandardScaler from sklearn.cluster import KMeans from sklearn.metrics import silhouette_score, davies_bouldin_score

In[3]:

customers = pd.read_csv(r"C:\Users\praba\Downloads\Customers.csv") products = pd.read_csv(r"C:\Users\praba\Downloads\Products.csv") transactions = pd.read_csv(r"C:\Users\praba\Downloads\Transactions.csv")

In[4]:

print(customers.head())

In[5]:

print(products.head())

In[6]:

print(transactions.head())

In[7]:

print(customers.isnull().sum()) print(products.isnull().sum()) print(transactions.isnull().sum())

In[8]:

print(transactions['Quantity'].describe()) print(transactions['TotalValue'].describe())

In[9]:

sns.histplot(transactions['TotalValue'], bins=30, kde=True) plt.title('Distribution of Total Transaction Value') plt.show()

In[10]:

print("Transactions Summary:\n", transactions[['Quantity', 'TotalValue', 'Price']].describe())

In[11]:

Customers per region

print("Customers per region:\n", customers['Region'].value_counts())

Signup dates over time

customers['SignupDate'] = pd.to_datetime(customers['SignupDate']) customers['SignupYearMonth'] = customers['SignupDate'].dt.to_period('M') print("Signups over time:\n", customers['SignupYearMonth'].value_counts().sort_index())

In[12]:

Product categories

print("Product categories:\n", products['Category'].value_counts())

Price distribution

sns.histplot(products['Price'], bins=30, kde=True) plt.title('Product Price Distribution') plt.show()

In[13]:

Transaction value distribution

sns.histplot(transactions['TotalValue'], bins=30, kde=True) plt.title('Transaction Value Distribution') plt.show()

Most frequently purchased products

top_products = transactions['ProductID'].value_counts().head(10) print("Top 10 products by transaction count:\n", top_products)

Business Insights

Based on the Exploratory Data Analysis (EDA) performed on the `Customers.csv`, `Products.csv`, and `Transactions.csv` datasets, the following **5 business insights** were derived:

Insight 1: Customer Distribution by Region

- **Insight**: Most customers are from **South America**.

- **Analysis**:

- From the `Customers.csv` data, we can see that out of the first 5 customers, 3 are from **South America** and 2 are from **Asia**.

- This suggests that **South America** is a key market for the business.

- **Actionable Recommendation**: Focus marketing and customer retention efforts on South America, as it has the highest concentration of customers.

Insight 2: Popular Product Category

- **Insight**: The **Books** category has the most products listed.

- **Analysis**:

- From the `Products.csv` data, 2 out of the first 5 products belong to the **Books** category (`ActiveWear Biography` and `ComfortLiving Biography`).

- This indicates that the **Books** category is well-represented in the product catalog.

- **Actionable Recommendation**: Promote the **Books** category further, as it seems to be a focus area for the business.

Insight 3: High-Value Transactions

- **Insight**: The **average transaction value** is relatively high, with most transactions exceeding $300.

- **Analysis**:

- From the `Transactions.csv` data, the `TotalValue` of transactions ranges from **$300.68** to **$902.04** for the first 5 transactions.

- This suggests that customers are purchasing high-value items or multiple quantities.

- **Actionable Recommendation**: Investigate the products driving high transaction values (e.g., `P067`) and consider upselling or bundling strategies to increase revenue.

Insight 4: Most Frequently Purchased Product

- **Insight**: The product with **ProductID P067** is the most frequently purchased.

- **Analysis**:

- From the `Transactions.csv` data, all 5 transactions involve the product **P067**.

- This indicates that **P067** is a popular or high-demand product.

- **Actionable Recommendation**: Stock more of **P067** and consider offering promotions or discounts to further boost sales.

Insight 5: Customer Signup Trends

- **Insight**: Customer signups peaked in **2022**.

- **Analysis**:

- From the `Customers.csv` data, 4 out of the first 5 customers signed up in **2022**, while only 1 signed up in **2024**.

- This suggests that customer acquisition efforts were more successful in **2022**.

- **Actionable Recommendation**: Analyze marketing strategies from 2022 and replicate successful campaigns to attract new customers in 2024.

In[ ]: