Experiment No: 9

Basic Data Analysis using pandas and matplotlib

Objective:

To analyze customer transaction data and segment customers based on their shopping behavior using data analysis and clustering techniques in Python.

Task Description:

You are working for a large e-commerce platform, and your task is to perform customer segmentation based on their shopping behavior. You have access to a dataset containing information about customer transactions. The dataset is in a CSV format and contains the following columns:

1. Customer ID: Unique identifier for each customer.
2. Total Amount Spent: The total amount spent by each customer on the platform.
3. Total Items Purchased: The total number of items purchased by each customer.
4. Last Purchase Date: The date of the customer's most recent purchase.
5. Average Purchase Value: The average value of each customer's purchases.

Using NumPy and Pandas, your goal is to perform the following tasks:

1. **Data Loading:** Load the dataset into a Pandas DataFrame for analysis.
2. **Data Cleaning:** Check for missing values, duplicates, or any inconsistencies in the data. If found, clean the data appropriately.
3. **Descriptive Statistics:** Calculate basic statistics such as mean, median, and standard deviation of TotalAmountSpent and TotalItemsPurchased.
4. **Customer Segmentation:** Divide the customers into segments based on their shopping behavior. You can use techniques like K-means clustering or any other method you prefer. For example, you might create segments like "High Spenders," "Frequent Shoppers," and "Inactive Customers."
5. **Visualization:** Create visualizations (e.g., scatter plots, bar charts) to represent the different customer segments you've identified.
6. **Customer Insights:** Provide insights into each customer segment. What distinguishes one segment from another? How can the e-commerce platform tailor its marketing strategies for each segment?
7. **Customer Engagement Recommendations**: Based on your analysis, provide recommendations for the ecommerce platform on how to engage with each customer segment more effectively. For example, should they offer discounts, provide personalized product recommendations, or run targeted marketing campaigns?

This problem requires you to use Pandas for data manipulation, NumPy for numerical operations, and potentially machine learning libraries for customer segmentation. It showcases the power of data analysis and segmentation for making data-driven decisions in e-commerce.

Dataset for the above analysis- https://www.kaggle.com/datasets/puneetbhaya/online-retail/

Dataset contains following information: Description, Quantity, Invoice Date, Unit Price, Customer ID, and Country.