Certified Data Analysts

Assignment 5

Customer Purchase Analysis using R

Scenario
You are working as a data analyst at an e-commerce company. The company wants to understand customer purchasing behavior to improve its marketing strategies. A sample dataset of customer transactions over the past year has been provided. The dataset includes the following columns:

CustomerID: Unique identifier for each customer
Gender: Gender of the customer (Male/Female)
Age: Age of the customer
AnnualIncome: Annual income in USD
SpendingScore: Score assigned by the marketing team based on customer spending habits (0–100)
PurchaseCategory: Category of the product purchased (Electronics, Clothing, Groceries, etc.)
TransactionDate: Date of the transaction

Use R to perform the following tasks:

1. **Data Import and Cleaning**
   - Load the dataset (CSV format) into R.
   - Handle any missing values appropriately.
   - Convert appropriate columns to categorical types (e.g., Gender, PurchaseCategory).

2. **Exploratory Data Analysis**
   - Display summary statistics for numerical columns.
   - Create a histogram of SpendingScore and boxplot of AnnualIncome by Gender.
   - Find the top 3 most purchased categories.

3. **Customer Segmentation (Clustering)**
   - Using SpendingScore and AnnualIncome, perform K-means clustering to segment customers into 3 groups.
   - Visualize the clusters using a scatter plot.

4. **Time Series Insights**
   - Convert TransactionDate into Date format.
   - Plot the number of transactions per month.
   - Identify any noticeable trends or seasonality.

5. **Statistical Analysis**
   - Perform a t-test to check if there is a significant difference in the average SpendingScore between Male and Female customers.

The submission requirements:

1. **Report:**
   - **Platform:** Google Drive
   - **Format:** word and PDF
   - **Page Limit:** Maximum 3 pages
   - **Content:** Include your analysis, findings, recommendations, and any other relevant information. (R Script file .R)