SYNOPSIS on

TITLE – CUSTOMER SHOPPING DATA ANALYSIS

TARUN S-AF0378031

ABOUT THE TOPIC (DATASET):

Data Structure

The dataset should be organized in a tabular format, where each row represents a customer record on a specific date. The columns might include:

1. invoice_no: This column contains unique identifiers for each transaction or invoice. Each value in this column represents a distinct sale or transaction.
2. customer_id: This column holds unique identifiers for customers. Each value represents a different customer who has made a purchase.
3. gender: This column indicates the gender of the customers. It typically contains categorical values such as 'Male', 'Female', or other gender identities.
4. age: This column shows the age of the customers. The values are numerical and represent the age in years.
5. category: This column categorizes the items purchased. It contains categorical values representing different types of products or services.
6. quantity: This column specifies the number of units of each item purchased in a transaction. The values are numerical.
7. price: This column indicates the price of a single unit of the item purchased. The values are numerical and typically represent the price in a specific currency.

8. payment_method: This column shows the method used by customers to make the payment. It contains cateorical values such as 'Credit Card', 'Cash', 'Debit Card', 'Online Payment', etc.
9. invoice_date: This column records the date and possibly the time when the transaction occurred. The values are typically in a date-time format.
10. shopping_mall: This column indicates the name or identifier of the shopping mall where the transaction took place. It contains categorical values representing different shopping malls.

These columns collectively provide detailed information about each transaction, including who made the purchase, what was purchased, how much was purchased, the cost, how it was paid for, and where and when the transaction occurred. This dataset can be very useful for analyzing customer behavior, sales trends, inventory management, and financial performance of the shopping malls

Data Preprocessing

1. Data Cleaning: Handle missing values and ensure data consistency.
2. Normalization: Normalize numeric features if required.
3. Encoding Categorical Variables: Encode categorical variables using one-hot encoding or label encoding.

Implementation Steps

1. Data Ingestion: Load the dataset into a data analysis environment (e.g., Python, R).
2. Preprocessing: Clean and prepare the data for analysis.
3. Exploratory Data Analysis (EDA): Conduct EDA to understand data distribution and initial patterns.

4. Modeling: Develop and validate models for classification, time series analysis, and clustering.
5. Visualization: Create visualizations to communicate insights effectively (e.g., correlation heatmaps, survival curves, cluster visualizations).
6. Reporting: Summarize findings in reports or dashboards for stakeholders.

Data set: customer_shopping_data

Technologies: pandas, Microsoft Excel, Microsoft PowerBi, seaborn, matplotlib.

Software Requirements:

Operating System – Windows, Linux and mac

IDLE – Jupyter Notebook

Hardware Requirements: RAM – Minimum 4GB Processor – Minimum intel i3