***Problem Statement: Customer Segmentation Using Machine Learning***

***Team number - 5***

Madhuri Sagiraju – 2111CS020256

Pilli Akshitha - 2111CS020040

Darapureddy Kiran – 2111CS020222

***Reading Data:***

* The code starts by importing necessary libraries like pandas, numpy, and matplotlib for data manipulation and visualization.
* It reads a CSV file named 'data.csv' using pandas' read\_csv() function and stores it in a DataFrame named data.

***Data Exploration:***

* data.head(): Displays the first few rows of the dataset.
* data.shape: Gives the shape of the dataset (number of rows, number of columns).
* data.columns: Returns the column names in the dataset.
* Data ['CustomerID'].isna().sum(): Counts the number of missing values in the 'CustomerID' column.
* Data ['Description'].isna().sum(): Counts the number of missing values in the 'Description' column.
* Data ['Country'].unique(): Returns unique values in the 'Country' column, indicating the countries present in the dataset.
* Data ['Description'].unique(): Returns unique values in the 'Description' column, showing the types of purchases.
* data.isna().sum(): Checks for missing values in the entire dataset.
* Data [pd.notnull (data['CustomerID'])]: Filters out rows where 'CustomerID' is not null.

***Data Cleaning and Filtering:***

* Missing values in the 'CustomerID' column are handled by filtering out rows where 'CustomerID' is not null.
* Further filtering is done to focus on specific subsets of data, such as customers from the United Kingdom and purchases with positive quantities.

***RFM Analysis:***

* Computes Recency, Frequency, and Monetary (RFM) analysis specifically for customers in the United Kingdom.
* Groups data by 'CustomerID' and aggregates:
* Recency: Number of days since the last purchase.
* Frequency: Total number of purchases made.
* Monetary: Total amount spent.
* Divides the RFM values into quartiles and assigns quartile ranks to each customer.
* Calculates an RFM score by concatenating quartile ranks.
* Identifies the best customers based on the RFM score.

***Exploratory Data Analysis (EDA):***

* Conducts various analyses on the original dataset such as:
* Total quantity sold by country and customer.
* Maximum quantity sold by country, stock code, and day of the week.

***Year-wise Analysis:***

* Splits the dataset into two years: 2010 and 2011.
* Analyzes the distribution of purchases across different months for each year.

***Algorithm Used : (RFM)***

Here RFM is used for the analysis and customer segmentation.

***Recency (R):***

**Definition:** Recency refers to how recently a customer has made a purchase or engaged with the business.

**Calculation:** It is usually calculated as the difference between the current date and the date of the customer's last purchase. The shorter the time since the last purchase, the higher the recency score.

***Frequency (F):***

**Definition:** Frequency represents how often a customer makes a purchase or interacts with the business.

**Calculation:** It is determined by counting the number of transactions a customer has made over a specific period. A customer with a higher frequency score is one who makes more frequent purchases.

***Monetary (M):***

**Definition:** Monetary value refers to the total amount of money a customer has spent on purchases.

**Calculation:** It is calculated by summing up the monetary value of all transactions made by the customer. Customers with a higher monetary value score are those who contribute more to the business's revenue.