**Customer Segmentation Analysis**

**Task Overview:**

The task revolves around customer segmentation analysis, aiming to understand and categorize customers based on their behaviour, demographics, and preferences. This analysis will enable the company to develop targeted marketing strategies, enhancing customer engagement and satisfaction.

**Packages and Datasets:**

Several essential Python packages are imported, including pandas, NumPy, and scikit-learn, to facilitate data analysis, visualization, and machine learning model development. The dataset is obtained from an internal source, containing information such as customer demographics, transaction history, and behavioural patterns.

The dataset I chose is from Kaggle : <https://www.kaggle.com/datasets/vetrirah/customer>

**Context**

An automobile company has plans to enter new markets with their existing products (P1, P2, P3, P4 and P5). After intensive market research, they’ve deduced that the behaviour of new market is similar to their existing market.

**Content**

In their existing market, the sales team has classified all customers into 4 segments (A, B, C, D ). Then, they performed segmented outreach and communication for different segment of customers. This strategy has work exceptionally well for them. They plan to use the same strategy on new markets and have identified 2627 new potential customers.

You are required to help the manager to predict the right group of the new customers.

**Exploratory Data Analysis (EDA):**

**Understanding the Dataset:**

The dataset is thoroughly examined, providing insights into the various features and their relevance to customer segmentation. The types of data and the presence of missing values are analyzed to ensure the quality of the dataset. Furthermore, the exploration includes statistical summaries and visualizations, shedding light on customer trends and behaviors.

**Data Cleaning and Preprocessing:**

**Handling Data Anomalies:**

Procedures to identify and handle outliers and inconsistencies within the dataset are implemented, ensuring that the data is reliable for subsequent analyses.

**Imputing Missing Values:**

Advanced techniques, such as imputing missing values based on customer segments or using predictive models, are applied to maintain the integrity of the dataset while preserving essential information.

**Feature Engineering:**

Existing features are carefully selected, and new features are engineered to enrich the dataset, providing additional insights for the segmentation process. The data is appropriately split into training and validation datasets, separating the predictors (independent variables) from the target variable.

**Model Development and Evaluation:**

**Model Selection:**

Various machine learning algorithms, including clustering techniques, are explored and implemented to identify the most suitable approach for customer segmentation. The selection is based on the model's performance metrics and its ability to capture the underlying patterns in the data effectively.

**Performance Evaluation:**

The model's performance is evaluated using appropriate metrics such as silhouette score, Davies-Bouldin index, and visual representations like cluster plots and silhouette plots. This evaluation allows for the validation of the model's effectiveness in customer segmentation.

**Insights and Recommendations:**

Based on the segmented customer groups, key insights and recommendations are provided to the management, highlighting specific strategies to improve customer engagement, loyalty, and overall business growth.

**Conclusion:**

A comprehensive summary of the project's findings, including the successful segmentation of customers and the impact of this segmentation on business strategies, is presented. Additionally, future steps and potential areas for further analysis are discussed to continue enhancing customer satisfaction and overall business success.

Link for my code : <https://colab.research.google.com/drive/1CN8ZTdZu0v9a-mKOcQ4psUW67VIq-sHU?usp=sharing>