**Customer Lifetime Value**

**Steps**

* Project Understanding
* Data Collection
* Exploratory Data Analysis (EDA)
* Data Cleaning
* Data Visualization
* Feature Engineering
* CLV Modeling
* Model Evaluation
* Predictions and Insights
* Reporting and Presentation
* Implementation and Monitoring
* **Project Understanding**

This project focuses on predicting Customer Lifetime Value (CLV) based on a range of independent features. The goal is to provide the company with insights into future profitability and customer behavior. By accurately estimating CLV, the organization can identify high-value customers and implement targeted strategies to retain them. Additionally, the project aims to highlight customers at risk of churn, enabling proactive measures to engage and retain these individuals. Ultimately, this initiative seeks to enhance customer satisfaction and loyalty, thereby improving the company’s overall reputation in the marketplace.

* **Data Collection**

The dataset for this project was obtained from the UCI Machine Learning Repository. It is important to note that the dataset does not include Customer Lifetime Value (CLV) as a predefined variable. Therefore, one of the key tasks in this project will be to calculate CLV using the available independent features. This will allow for a deeper analysis of customer behavior and enable the development of effective retention strategies.

[Data link](https://archive.ics.uci.edu/dataset/502/online+retail+ii)

* **EDA**

The EDA phase involves loading the dataset and examining its structure, including the number of records, feature types, and summary statistics. Key tasks include identifying missing values and outliers through visualizations like box plots and histograms. Relationships between variables are explored using scatter plots and correlation matrices to uncover potential patterns. Segmenting the data by customer demographics provides insights into different behaviors that influence Customer Lifetime Value (CLV). Overall, EDA lays the groundwork for data cleaning, feature engineering, and accurate CLV calculations.

* **Data Cleaning**

In the Data Cleaning phase, we address missing values and outliers identified during the EDA. Missing data is handled through appropriate strategies such as imputation or removal, depending on the extent and significance of the gaps. Outliers are assessed to determine whether they should be retained or removed based on their potential impact on analysis. Additionally, data types are standardized to ensure consistency, and new features may be engineered to enhance the dataset's predictive power for Customer Lifetime Value (CLV). This process ensures that the dataset is accurate, complete, and ready for modeling.

* **Data Visualization**

In the Data Visualization phase, various tools such as Power BI, Tableau, and Python are employed to create insightful visual representations of the data.

**Note:** The three phases of Exploratory Data Analysis (EDA), data cleaning, and visualization are interrelated, and I will conduct them concurrently.

* **Feature Engineering**

In the Feature Engineering phase, new features like recency, frequency, and monetary value are created to enhance the model's predictive power for Customer Lifetime Value (CLV). The data is then split into training and testing sets for evaluation, and feature scaling (using methods like Standard Scaler) is applied to normalize variables. This ensures that all features contribute equally to the model, improving its accuracy and generalization.

* **CLV Modeling**