DATASET OVERVIEW:

Auto insurance and customer value analysis. This dataset comes from an auto insurance company. C LV is the total revenue, the client will derive from their entire relationship with a customer which is a co ntinuous target variable. The dataset contains 24 features and 9134 observations. The features and th eir description are as follows:

METADATA:

Feature	Explanation	Data types
Customer	Unique customer ID	Object
State	US province to where the customer belongs to	Categorical
Response	Refers to whether customers have responded to marketing calls or not	Categorical
Coverage	Nature of insurance coverage	Categorical
Education	Education level of customer	Categorical
Effective To Date	Expiry date of policy	object
Gender	Gender of the customer	Categorical
Employment Status	Current Employment status of the customer	Categorical
Income	Customer annual income in USD	Continuous
Location Code	Type of location where customer lives	Categorical
Marital Status	Marital status of the customer	Categorical
Vehicle Size	Size of vehicle	Categorical
Vehicle Class	Type of vehicle	Categorical

Sales Channel	Channel of sales	Categorical
Renew Offer Type	Offer given during renewal	Categorical
Total Claim Amount	Amount claimed till date	Continuous
Monthly Premium Auto	Monthly premium for auto insurance	Continuous
Months Since Last Claim	No. of months before which the last claim was made	Continuous
Months Since Policy Inception	No. of months before which the policy commenced	Continuous
Number of Open Complaints	No. of unresolved complaints from the customer	Continuous
Number of Policies	No. of policies with the current customer	Continuous
Policy Type	Type of policy	Categorical
Policy	Policy sub category	Categorical
Customer Life Time Value	CLV of the customer for the auto insurance company	Continuous

OBJECTIVE:

To Analyse the data and develop focused customer retention programs by grouping the customers for auto Insurance marketing and creating a model to Predict future customers Grouping based as on Given Features. A secondary Goal can Also be predicting CLV for a particular customer based on the data provided.

There is Three parts to this goal:

- First is to Identify the groups of Customers.
- Second is to Build a Classification model that may be able to Predict such customers to identified groups in future given similar data on set Features.
- Third is to build a regression model to predict CLV for a given Customer

APPROACH:

We start with data cleaning and pre-processing.

1.In data cleaning and pre-processing :

- Check for missing values
- Check for outliers
- If outliers is present impute the missing values using KNN ,decision tree, mode or median
- If outliers are not present impute with mean or KNN
- Check data for any unusual values (flags like 9999, nan,?,- etc)
- Check if we have date- time and convert it to type date.

2.EDA:

Using various libraries like matplotlib, seaborn, tableau, etc we do complete Descriptive and Inferential analysis of the dataset.

- We do univariate analysis of numerical data.
- Univariate value count analysis of Categorical data.
- Do Bivariate analysis of required features with respect to Target variable (CLV).
- Check Correlation amongst features using pairplots and heatmaps.
- Check Outliers using Boxplot or QQ-PLOT.
- We do outlier treatment if required.
- We encode for converting Categorical into Numerical columns (which might increase the dimensions of the Dataset).
- We do use various feature selection techniques to see which is favourable feature for dimensionality reduction.
- We validate metadata inferences and add inferences of our own.

3.MODEL BUILDING:

CLUSTERING:

- Using Clustering we find out the best possible clusters to get the Customers belonging to similar Clusters (K=Means or Agglomerative Techniques used.)
- We check our clusters using various clustering metrics
- We divide cluster into labels and create a new column.
- Do group by based on this cluster.
- Perform Bivariate analysis on new groups using all highly relevant features like income, class types etc.

CLASSIFICATION:

- After analysing various clusters, we segment Customers accordingly and classify data and create a Classification Model based on those clusters.
- We are going to identify the best Classification Models.
- Do different validation and testing to justify the best selected Model.

REGRESSION:

- We finally create a regression model to predict CLV.
- We Ensure that our regression model is best by checking R2 and Adj R2 and other different parameters and tweak them to achieve the best model if required.

CONCLUSION:

- We provide business inputs based on our final models. Providing inputs as to which features to be focused on which customer.
- We also provide business suggestions on based on our Models to make their approach and marketing easier.