**Intelligent Customer Retention: Using Machine Learning for Enhanced Prediction of Telecom Customer Churn**

**Define Analysis**

* Exploratory Data Analysis
* Descriptive Statistical
* Visual Analysis
* Univariate Analysis
* Multivariate Analysis
* Splitting data into train and test
* Scaling the Data

**Exploratory Data Analysis**

* In this milestone, we will see the exploratory data analysis

**Descriptive Statistical**

* Descriptive analysis is to study the basic features of data with the statistical process.
* Here a panda has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features.
* And we can find mean, STD, min, max. And percentile values of continuous features.

**Visual Analysis**

* Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data.
* It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions.

**Univariate Analysis**

* In simple words, univariate analysis is understanding the data with a single feature Here we have displayed two different graphs such as despot and count plot.
* The Seaborne package provides a wonderful function despot. With the help of despot, we can find the distribution of the feature. To make multiple graphs in a single plot, we use subplot.
* In our dataset we have some categorical features. With the count plot function, we are going to count the unique category in those features.
* We have created a dummy data frame with categorical features. With for loop and subplot we have plotted this below graph.
* From the plot we came to know, Applicants income is skewed towards left side, where as credit history is categorical

**Multivariate Analysis**

* In simple words, multivariate analysis is to find the relation between multiple features.
* Here we have used a swarm plot from the seaborne package.

**Splitting data into train and test**

* Now let's split the Dataset into train and test sets
* Changes: first split the dataset into x and y and then split the data set
* Here x and y variables are created. On x variable, do is passed with dropping the target variable And on y variable is passed
* For splitting training and testing data we are using the train, test, split () function from sclera as parameters, we are passing x, y, test size, random state.

**Scaling the Data**

* Scaling is one the important process, we have to perform on the dataset, because of data measures in different ranges can leads to mislead in prediction
* Models such as KNN, Logistic regression need scaled data, as they follow distance based method and Gradient Descent concept.