1. Importing all the needed libraries.
2. Importing the data set. Training data into train variable and test data into test variable.
3. Train and test data basic info.
4. Data types.
5. Number of missing values in our dataset.
6. Mean, min, max, std. of the Continuous data
7. Handling Missing data.
8. Holding policy type. – I filled the missing values with mode of the that column.
9. Holding\_Policy\_Duration- Firstly I switched the “14+” with a new str 20 which convertible to int. I used the ***“Random Sample Imputation”***  for this columns.

I took the random values from the column dropping NaN and merged it to the same column where there were NaN values.

1. Health Indicator – I used the ***“Random Sample Imputation”***  for this columns.

I took the random values from the column dropping NaN and merged it to the same column where there were NaN values.

Again used dataset.insa().sum()- to check the if still there are any missing values

1. Visualized representation of the relation between “Response” and other Features.
2. Data preprocessing for using it in our model
3. Dropping “ID” and “Region\_Code”.
4. Converting “Holding\_Policy\_Duration” from object type to float64 type.
5. Separating dependent and independent variable.

X=independent

Y=dependent

1. Converting categorical data into numerical form using onehotencoding technique.
2. Spliting of Training data into train and test dataset
   1. Xtrain : independent train
   2. Xvalid : dependent train
   3. Ytrain : independent test
   4. Yvalid : dependent test
3. Applying {' Logistic Regression ' , ' K Neighbors Regressor ' , ' Decision Tree Regressor '} methods to our datasets.
4. Calculating (roc\_auc\_score) to each model and selecting one with highest score.
5. Final submission.