**In class problems (Regression).**

1. Import the insurance dataset and get to know the dataset.
2. Identify the presence of null values.
3. Perform data preprocessing –
4. Encode the string column values to integer using the df.replace() function.

Hint: This can also be easily done using LabelEncoder() function from scikit-learn if you don’t know all the different values in the column.

How to use:

Step 1: from sklearn.preprocessing import LabelEncoder

Step 2: encode = LabelEncoder()

Step 3: dataset["region"] = encode.fit\_transform(dataset["region"])

Separate the data into input and output data variables X and y.

1. Data Visualization –
2. Perform data correlation.
3. Plot the columns with high correlation.
4. Machine Learning --
5. Split the variables X and y into training and testing data with testing data size being 35%.
6. Use one of the regression algorithms of your choice to train the data.
7. Test on the training data and report the accuracy.