

Problem Statement or Requirement: A client's requirement is, he wants to predict the insurance charges based on several parameters. The Client has provided the dataset of the same. As a data scientist, you must develop a model which will predict the insurance charges.

1.) Identify your problem statement :

**The problem statement is to be predict the Insurance charge based on the client's personal information**

2.) Tell basic info about the dataset (Total number of rows, columns):

**The dataset has 5 independent parameters (age,sex,bmi,children,Smoker) and one dependent parameter (Charges). It has 1338 rows and 6 columns**

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data) :

**Sex and the Smoker field values are preprocessed and converted to numeric values prior to the data model training.**

4.) Develop a good model with r2\_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

**Developed the Simple Linear Regression, Multiple Linear Regression, Decision Tree and Random Forest models.**

5.) All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

Sl.No.	Model Name	R2 Score
1	Simple Linear Regression	0.7894790349867009
2	Multiple Linear Regression	0.7681647919672667
3	Decision Tree	0.7094671863165635
4	Random Forest models	0.8498329315421834

6.) Mention your final model, justify why u have chosen the same.

**Based on the R2 Score Random Forest has the highest value of R2 and that is concluded as the best model for this input dataset**