1.) Identify your problem statement

- **Develop a model which will predict the insurance charges.**
- **As the output value is number, choosing machine learning.**
- **In ML, choosing supervised learning method as the input and output are very clear.**
- **In supervised learning, selecting regression algorithm as the output is numeric value.**

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

- **In the provided dataset (insurance_pre.csv) there are 6 columns and 1338 rows exists.**
- **In that data set, 3 columns contain numerical values and 2 columns contain categorical values.**
- **In that data set, except "charges" column remaining are input columns.**

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

- **As there are 2 categorical columns (sex column and smoker column) exist in the provided data set, we need to convert those categorical values into numerical values using "One Hot Encoding".**

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.

Need to develop multiple models using MLR (Multiple Linear Regression), SVM (Support Vector Machine), DT (Decision Tree), and RF(Random Forest) regression algorithm by modifying the playable parameters and come up with the good model which gives the good r2_score.

5.) All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

Consolidated_R2_Score_Values_For_Insu

I will attach the PDF file as a separate file.

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

The final model is Random Forest algorithm based model which uses below parameters

criterion='absolute_error', max_features = 'log2', C-100