**Predicting Health insurance Charges**

Objective:

The objective of this data science problem is to develop a predictive model that can accurately estimate health insurance charges for individuals based on their demographic and health-related information.

Your task is to use the data below to achieve the above objective :

<https://drive.google.com/file/d/1ooRL_3LehZ2IurpzL67zEwOyWsoqPFBG/view?usp=sharing>

1. Define an approach you would take to solve the problem and document it
2. Get the data and determine what type of machine learning problem it is
3. Outline possible algorithms you would use to create the model.

Hint :

So far you know about Linear regression. There are a host of other regression algorithms : Decision Tree regression, Random Forest regression, SVM regression etc. Do your research.

1. Conduct exploratory analysis to understand the distribution of variables, identify any correlations, and gain insights into the dataset.
2. Handle missing values, encode categorical variables, and scale numerical features if necessary.

Hint: Encoding is converting from categorical to numerical using libraries like Label Encoder and OHE (One Hot Encoder). Scaling features is ensuring there’s no disparity in the variations for the features so that no feature is given preference. There are libraries like Standard Scaling and Min-Max scaling for that

1. Extract additional features if needed, such as interaction terms or polynomial features.
2. Evaluate different regression algorithms (e.g., linear regression, decision tree regression, random forest regression, etc.) and select the one with the best performance based on evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared.
3. Train the selected model on the training dataset.
4. Evaluate the trained model's performance on the test dataset using appropriate evaluation metrics.

Have Fun !!