Problem Statement

Given a dataset with features like age,bmi,male/female etc, predict the insurance

Solution

- **Stage 1 -** Identifying the domain. The dataset for the problem statement contains only numerical data . Hence the domain that we would choose is Machine Learning
- **Stage 2 -** We have clear requirements. The dataset has both input and output columns. Hence we will go with Supervised Learning
- **Stage 3 -** Output column is numeric , so will use Regression.
- **Deciding Regression Algorithm**
    - We have multiple features to be considered, Therefore Simple Linear Regression is ruled-out
    - We have 4 remaining algorithms to evaluate - Multiple Linear Regression,Support Vector Regression,Decision Tree and Random Forest
    - Each of these algorithms has many hyper parameters that can be tuned to get the best results. Wrote a python program **best_model_decider.ipynb** that evaluates each algorithm with different combinations of hyper parameters and saves the best model. Here are the results from the python program
        - MLR **r_score=0.7894790349867009**
        - SVR C=3000, kernel=rbf, **r_score=0.8663393963090398**
        - DT criterion=absolute_error, splitter=best, max_features=log2,**r_score=0.7655436098574626**
        - RF n_estimators=200, max_depth=10, min_samples_split=8,max_features=sqrt,random_state=0, **r_score=0.8844928536280982**
- Saved the best model which is RandomForest and then wrote a program **insurance_predictor.py** which will load the saved model, accept user inputs and predict insurance. Sample inputs and outputs are given below

```
Enter no. of inputs:2
(Input #1)
----------------
Enter age:25
Enter Bmi:29.99
Enter no. of children:2
Is Male(y/n)?:y
Is smoker(y/n)?:y
```

```
(Input #2)
----------------
Enter age:35
Enter Bmi:32.346
Enter no. of children:3
Is Male(y/n)?:n
Is smoker(y/n)?:n

---------Insurance Predictions------------
        age=25,
        bmi=29.99,
        children=2,
        gender=male,
        smoker=yes,
        predictedInsurance=26835.07

        age=35,
        bmi=32.346,
        children=3,
        gender=female,
        smoker=no,
        predictedInsurance=8434.64
```