Predict the Insurance premium/charges for a customer
Project Proposal

Shruthi Subbaiah Machimada
001880621

Objective:
To create a Linear regression model to predict the amount of insurance premium a customer has to pay, based on factors such as Age, gender, bmi.
This will help insurance companies and users save money by accurately predict the personalized health care costs for a user, based on his profile.

Data:
The data can be found at : https://www.kaggle.com/mirichoi0218/insurance
Input variables:
Age
Sex
BMI
Children
Smoker
Region

Outcome variable:
charges

Steps:

1. Loading and cleaning the data
2. Exploratory data analysis to understand the distribution of our data, the variables and their correlation
3. Convert Categorical values into factors
4. Create a Linear regression model, perform ANOVA to filter out the input variables that might not be needed.
5. Divide the dataset into Training and Test set.
6. Using the Training set, create a Linear regression model again, with the new set of input variables, repeat this process until we are satisfied with the model performance (R^2 and R^2_{adj})
7. Use the model on the test set, measure its performance.
8. Use the model to score new data.