# Regression-Medical-Insurance-Cost

**Abstract:**

Health insurance is a type of insurance that covers medical expenses that arise due to an illness. These expenses could be related to hospitalization costs, cost of medicines or doctor consultation fees.
To give you some background, insurance companies should collect higher premium than the amount paid to the insured person. Due to this, insurance companies invest a lot of time, effort, and money in creating models that accurately predicts health care costs.
**The goal of this project** is to use the data provided in the Medical Cost Dataset exploring which factors are important to predicting medical costs, and then we will perform a regression analysis to predict the price for every individual patient.

**Design:**

No one plans to fall ill or get hurt, but a serious illness can strike anyone at any time. The cost of treating the illness can cause severe financial strain on the savings you have accumulated over time. This means that you might have to compromise on providing your child the best quality education or defaulting on your home loan payments. Today, the cost of medical treatment is continuously rising.

So, we will see which model is good to predict the cost.

**Data:**

This dataset was web scraping on GitHub by html processes and were exploring by jupyter notebook to predict and select the best model to predict the a proximal cost for every customer.

Data size: (1338 Rows x 6 Columns).

Rows: 1338

Columns: 6

which are: (Age – Sex – Bmi – Children – Smoker – Region)

**Algorithms:**

The dataset was clean when we web scraping it.
Calculate the factor how have the height correlation by heatmap.
Change the categorical value by using dummy function.
Calculate the r^2 for the models and see what is the good one.
Calculate the mean squared error.
Calculate the mean absolute error.
Calculate the root mean squared error.

**Tools:**

The technologist that I used are Python, Jupyter Notebook.

The libraries are Pandas, NumPy, Matplotlib, seaborn, Scikit Learn, statsmodels, patsy, BeautifulSoup, requests.

**Communication:**

Smoking has the highest impact on medical costs, even though the costs are growing with age, bmi and children.

Also, people who have children generally smoke less.

We use some models to find the best R^2 and the Polynomial Regression turned out to be the best model.