**Tutorial**

This dataset uses different factors that can contribute to predicting the possible charges related to a person’s health care. This prediction is particularly useful for insurance companies. The insurance dataset includes:

* Age: age of the primary beneficiary.
* Sex: insurance contractor gender, female, male.
* BMI: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg/m^2) using the ratio of height to weight, ideally 18.5 to 24.9.
* Children: Number of children covered by health insurance/Number of dependents.
* Smoker: Is the person a smoker or not.
* Region: the beneficiary’s residential area in the US, northeast, southeast, southwest, northwest.
* Charges: Individual medical costs billed by health insurance.

Based on this dataset create a multiple linear model that predict the charges based on the other variable:

1. Explain what the coefficients in your model means.
2. Use and interpret the Multiple R squared value to give an indication of the strength of your model.

Check if the assumptions of multiple linear regression are met:

1. Visualize the relationship between the two variables you have chosen in a Scatter Plot. Your plot should show a trend line and be fully labelled
2. Check if There is multicollinearity in your data
3. Check if the The values of the residuals are independent
4. Check if the variance of the residuals is constant.
5. Check if the values of the residuals are normally distributed.
6. Check if There are influential cases biasing your model