

Health Insurance claim

Problem Statement (Situation):

“Finding out the health parameters that affect health insurance claims”

An insurance company in the US is reviewing its insurance claims/charges and is trying to do a cause and effect analysis for future business decisions. It has collected data for its customers who have made claims till recent time. The data-points collected are age, gender, bmi, number of children/dependents, smoking habit, region they belong to, charges/bills claimed under the insurance. This analysis would have a bearing on what premium should the company charge a customer availing an insurance policy.

The insurance company has collected a dataset of 1338 customers-claims. Please refer to the data dictionary below:

Data Dictionary:

Attribute	Description
Age	Age of the customer/claimant who has claimed insurance for medical treatment charges
Sex	Gender of the customer/claimant
bmi	Health parameter: person's weight in kilograms divided by the square of height in meters
Children	No. of children the claimant has
Smoker	Whether the claimant smokes or not
Region	Region to which the claimant belongs
Charges	The exact medical charges for which the claimant has claimed insurance

Learning Outcome (Result):

Understand implementation of Exploratory Data Analysis to understand the nature of different data-attributes, through pivot tables and different types of visualizations

Understand how to use various statistical/analytical tools in MS Excel like Summary statistics, Histogram, correlation table, Pivot tables, Regression analysis (using Data analysis tool pack