**Problem Statement**

A key challenge for the insurance industry is to charge each customer an appropriate premium for the risk they represent. The ability to predict a correct claim amount has a significant impact on insurers' management decisions and financial statements. Predicting the cost of claims in an insurance company is a real-life problem that needs to be solved in a more accurate and automated way. Several factors determine the cost of claims based on health factors like BMI, age, smoking, health conditions and others. Insurance companies apply numerous techniques for analysing and predicting health insurance costs

**Data Definition**

* age: Age of the policyholder (Numeric)
* sex: Gender of the policyholder (Categoric)
* weight: Weight of the policyholder (Numeric)
* 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 (Numeric)
* no\_of\_dependents: Number of dependent persons on the policyholder (Numeric)
* smoker: Indicates policyholder is a smoker or a non-smoker (non-smoker=0;smoker=1) (Categoric)
* claim: The amount claimed by the policyholder (Numeric) in dollars
* blood pressure: Blood pressure reading of policyholder (Numeric)
* diabetes: Indicates policyholder suffers from diabetes or not (non-diabetic=0; diabetic=1) (Categoric)
* regular\_ex: A policyholder regularly exercises or not (no-exercise=0; exercise=1) (Categoric)
* job\_title: Job profile of the policyholder (Categoric)
* city: The city in which the policyholder resides (Categoric)
* hereditary\_diseases: A policyholder suffering from hereditary diseases or not (Categoric)