

# Linear Regression - Insurance

---

## Data:

The insurance.csv dataset contains 1338 observations and 7 attributes. The data contains medical costs of people characterized by certain attributes. Let's see if we can dive deep into this data to find some valuable insights.

## Attributes:

Age: age of 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 ( $\text{kg} / \text{m}^2$ ) 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: Smoking

Region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.

Charges: Individual medical costs billed by health insurance.

## Key asks:

- Do charges of people who smoke differ significantly from the people who don't?
- Does BMI of males differ significantly from that of females?
- Is the proportion of smokers significantly different in different genders?
- Is the distribution of BMI across women with no children, one child and two children the same?