

Problem Statement

The cost of health insurance varies for individuals based on their personal and lifestyle characteristics. Insurance companies need to accurately estimate the expected medical charges of a customer to set fair premiums.

In this project, we use the Insurance dataset containing fields such as:

- Age – age of the insured person
- Sex – gender
- BMI – body mass index
- Children – number of dependents
- Smoker – whether the person is a smoker
- Region – geographic region
- Charges – individual medical costs billed by health insurance

Using IBM SPSS Modeler, we aim to:

1. Explore the dataset to find patterns and relationships between variables (e.g., smoker vs charges, BMI vs charges).
2. Derive new fields such as:
 - BMI_Category (Underweight, Normal, Overweight, Obese)
 - Age_Group (Young Adult, Middle Aged, Senior Citizen)
3. Reclassify categorical variables (e.g., merging regions into “North” and “South” groups).
4. Interpret which factors most significantly affect medical costs.