

Predicting Health Insurance Premiums in the US

University : The George Washington university.

Team Members : Keerthana Aravindhan, Amit Sopan Shendge, Vamsidhar Boddu.

Project Description:

People are always confused about their medical insurance and don't know the cost of insurance at different ages and conditions.

This project aims to develop a predictive model that estimates health insurance premiums for individuals based on a range of demographic and medical history factors. This helps insurance providers and policyholders understand how different factors influence insurance premiums and to predict future premium changes.

SMART questions for Insurance dataset :

- Specific:

1. To what extent is "region" a useful variable for estimating insurance costs? Does the data show any regional trends that affect premiums? Else what specific factors affect premiums?

- Measurable:

2. In comparison to non-smokers, how much does being a "smoker" add to the rise in insurance costs?
3. How much does age impact insurance premiums, and is this impact consistent across different regions?

- Action oriented:

4. How can insurance companies use the data on smoking habits and exercise frequency to devise strategies for premium adjustments?

- Relevant:

5. How relevant is gender in determining insurance premiums, and is there a gender-based disparity in premiums?

- Time bound:

6. How can we provide individuals with real-time estimates of their health insurance premiums based on their unique characteristics beforehand?

Modeling Methods:

Regression models like Linear Regression, Random Forest Regression can be used for prediction.

Data Source:

This dataset was created using a script that generated a million records of randomly sampled data points, ensuring that the data represented the population of insured individuals in the US. This dataset includes information on **12 variables**, like age, gender, BMI, children, smoking status, region etc. **It has 100,000+ observations.**

Link :

<https://www.kaggle.com/datasets/sridharstreaks/insurance-data-for-machine-learning/data>

GIT Link : <https://github.com/Keerthana0620/DATS6103-Data-Mining-Project>