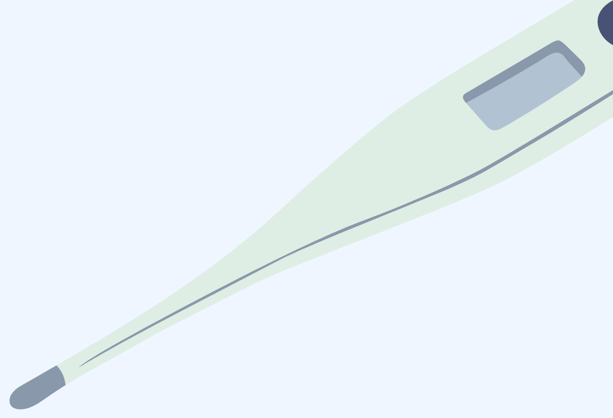


NUS SDS Mini-Hackathon AY25/26

NUS Statistics and
Data Science Society

Timeline

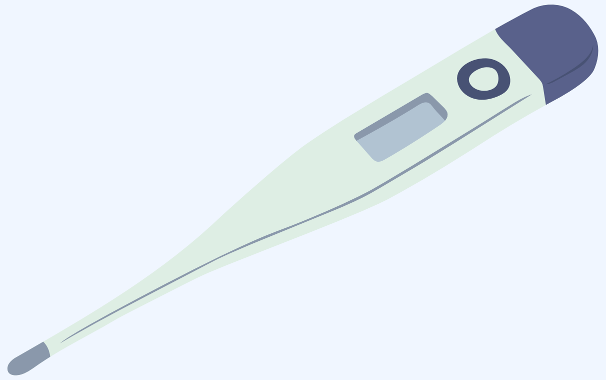
1. **Problem Statement Release:**
20 Oct (Mon)

2. **Submission Deadline:**
30 Oct(Thu) 2359

3. **Finalist Announcement:**
3 Nov (Mon)

4. **Final Round (In-person):**
5 Nov (Wed)

NUS Statistics and
Data Science Society



Problem statement

Predicting Insurance Costs and Analyzing Key Factors

Challenge: Can you predict medical insurance charges based on demographic and lifestyle data, and identify which factors are most important?

Prediction

Build regression models (e.g., linear regression, decision trees etc.) to predict insurance charges using features like age, BMI, smoking habits, and region

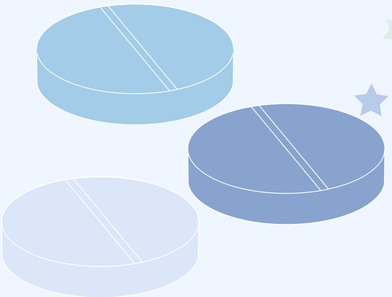
Feature Analysis

Identify the top factors driving insurance costs using simple techniques like feature importance or visualizations.

Fairness check

Quickly assess if the charges differ across groups (e.g., gender, region) and discuss any fairness concerns.

NUS Statistics and
Data Science Society



Link & Intro to the dataset

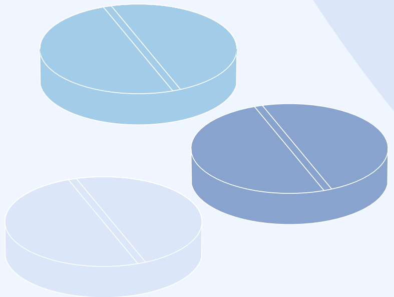
Medical Insurance Cost

Link to Dataset

This dataset contains the medical insurance cost information for 1338 individuals.

Variable names/ columns:

age: Age of primary beneficiary (int)
sex: Gender of beneficiary (male, female)
bmi: Body Mass Index, a measure of body fat based on height and weight (float)
children: Number of children covered by health insurance (int)
smoker: Smoking status of the beneficiary (yes, no)
region: Residential region in the US (northeast, northwest, southeast, southwest)
charges: Medical insurance cost billed to the beneficiary (float)



NUS Statistics and
Data Science Society

Rubrics

Creativity & Novelty (20%)

- Unique angles, innovative methods, or original perspectives in predicting insurance cost & analysing key factors

Methodological Soundness (20%)

- Correct use of ML techniques, robustness of preprocessing, justification of models
- Clear code readability and structure

Interpretability of Analysis (30%)

Results are clearly explained; accessible to both technical and non-technical audiences

Presentation & Communication (20%)

Clear visuals, engaging storytelling and effective slides

Feasibility & Relevance (10%)

Solutions are practical, dataset limitations acknowledged, insights applicable to real-world use

NUS Statistics and Data Science Society

Submission requirements

Teams must submit the following by 30 Oct 2025, 23:59 SGT via Google forms:
Google Form Submission Link

Slide Deck (≤10 slides, PDF or PPTX)

1. Problem framing & objective
2. Exploratory Data Analysis (EDA)
3. Regression & modeling approach (baseline + advanced)
4. Key findings & visualizations
5. Feature impact & fairness analysis
6. Practical recommendations
7. Difficulties faced & methods used to overcome
8. Appendix (please include links to your relevant code files here) [*Appendix not counted in page limit*]

NUS Statistics and
Data Science Society