

MFR Project 2025

Prof. Eva Cantoni

Description

The dataset represents **healthcare costs for patients undergoing a new therapy for chronic disease management**. The response variable, **Healthcare_Cost**, measures the cost of treatment in thousands of dollars. The healthcare system aims to understand how different patient characteristics influence treatment costs.

Covariates

- **Age**: Patient's age in years.
- **BMI**: Body Mass Index (kg/m^2), a measure of obesity.
- **Severity_Score**: A composite score (1-10) measuring disease severity.
- **Exercise_Hours**: Average weekly hours spent exercising.
- **Height_m**: The patient's height in meters.
- **Weight_kg**: The patient's weight in kilograms.
- **Health_Insurance_Premium**: The monthly insurance premium in dollars.
- **Sedentary_Hours**: The average hours spent sitting per day.

Research question

Which factors should be considered to impact costs and therefore should be taken into account for insurance premium calculation? In addition, do a proposition for an insurance premium calculation such that the insurance company is guaranteed to have a profit for a population of a million individuals from which the sample has been taken.

MFR Project 2025

Prof. Eva Cantoni

Description

This dataset represents a **study of hospital visit frequency among chronic patients**. The **response variable** is the number of hospital visits in a year (`Hospital_Visits`), influenced by patient lifestyle and demographics.

Covariates

- **Age:** Patient's age in years.
- **BMI:** Body Mass Index (kg/m^2), a measure of obesity.
- **Diet_Type:** Patient's diet type: `Healthy`, `Moderate`, `Unhealthy`.
- **Smoking_Status:** Smoking status: `Non-Smoker`, `Smoker`.
- **Exercise_Hours:** Average weekly hours spent exercising.
- **Hospital_Visits:** Number of hospital visits per year.
- **Height_m:** The patient's height in meters.
- **Weight_kg:** The patient's weight in kilograms.
- **Healthcare_Cost:** The accumulated healthcare costs per patient in a year, in dollars.

Research question

Your company aims to design a preventative healthcare campaign targeting high-risk individuals who frequently visit hospitals. You have to choose one important risk factor to focus your campaign on. Using statistical modeling, determine the most important risk factor influencing hospital visits and make a campaign proposition. Justify your decision with data and include data in your proposed campaign.

MFR Project 2025

Prof. Eva Cantoni

Description

The dataset comes from a study examining service request times in customer support. The goal is to analyze how customer features (experience, issue complexity, time of day, etc.) affect service time.

Covariates

- **Group:** Customer support collaborator ID handling the request.
- **Loyalty_Years:** Number of years the customer has used the service.
- **Issue_Complexity:** Complexity of the issue: **Low**, **Medium**, **High**.
- **Time_of_Day:** Time the request was submitted (0-24 hours).
- **Urgency_Level:** 0 = **Normal**, 1 = **Urgent** request.
- **Service_Time:** Time taken to resolve the service request (minutes).
- **Total_Hours:** Represents the approximate total number of hours a customer has interacted with the service.
- **Request_Length:** A numeric encoding of request complexity.

Research question

The company is accused to discriminate customers based on loyalty, privileging the oldest customers. Does the data support this claim?