IST 687

Introduction to Data Science

Submitted by:

Yuktha Lakshmi Arcot BadarinathPrakyath Davanam PraveenGughapriyaa ElangoAmisha GangwarChaitanya Krishna Mupparaju

Health Care Management Organization

A close-up of a calculator and a calculator

Description automatically generated with medium confidence

Abstract

In the field of healthcare, costs are a significant factor, and some patients may present a greater financial strain than others. There are a lot of factors that go into making one customer more expensive than another, and if a company can figure out why some customers are more expensive than others, they will see a significant increase in their bottom line. The HMO is able to set the premium plans up correctly, which allows for more effective management of the firm. This document presents our research into leveraging data science (exploratory data analysis as well as prediction models) to anticipate how and why some people have higher bills and deliver actionable insights to Health Management Organizations to help them make better decisions. The research was carried out using exploratory data analysis and prediction models.

**Dataset & Variables**



The data set we are using consists of 7582 rows and 14 columns, details about the people were given in each of these columns such as age, location, gender, etc.

Text, letter

Description automatically generated

**Process**

We began by retrieving the information from the specified URL, and then we put it into the variable denoted by the letter 'DF.' This is the dataset in its raw form. The first thing that we did was look through the database to see if there was any mention of a NA for it anyplace. In order to fill in the blanks left by the N/A values in the BMI, Hypertension, and Cost columns, we made use of the interpolation function, which is included in the imputeTS package. In order to finish our exploratory data analysis, we made use of all of the variables (EDA).

A picture containing text

Description automatically generated

As we are trying to find why some people are more spending more than others in the medical costs the cost variable becomes absolutely important and that is the reason we have examined the cost attribute. NEW JERSEY is the state with the lowest cost where CONNECTICUT is the state with the highest cost.

