

Insurance Data Analysis

1.0 Problem Statement:

An insurance agency, ABC Insurance, has a large dataset containing information about their policyholders and claims. They want to perform exploratory data analysis (EDA) on this dataset to gain insights that can help them make better business decisions and improve their operations.

The agency wants to analyze the different body types and the environment that affect the premium. The disease's effect or the cost of treatment differs depending on the circumstances. For example, a smoker's medical insurance premium may be higher than that of a healthy person, because smokers are more likely to develop chronic diseases. The agency wants to analyze the data to research healthcare premium costs.

2.0 Objective: To analyze the dataset that will help to create a model that will predict the cost of medical insurance based on various input features

3.0 Domain: Healthcare

4.0 Dataset: insurance dataset (insurance.csv)

5.0 Dataset Description:

Table 1.1

age	Age of the person
sex	Female or Male
BMI	BMI value to estimate an individual's health and fitness condition
children	number of children (1,2,3,4, or 5)
smoker	The person is a smoker or not
region	Specifies the region (northeast, northwest, southeast, southwest)
charges	the amount of insurance

6.0 Steps to Be Followed:

1. Import libraries such as Pandas, matplotlib, NumPy, and seaborn and load the insurance dataset
2. Check the shape of the data along with the data types of the column
3. Check missing values in the dataset and find the appropriate measures to fill in the missing values
4. Explore the relationship between the feature and target column using a count plot of categorical columns and a scatter plot of numerical columns
5. Perform data visualization using plots of feature vs feature
6. Check if the number of premium charges for smokers or non-smokers is increasing as they are aging
7. After each step, specify the observations