TLADI LT INSURANCE EDA PROJECT

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.

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

I did the steps below to accomplish this project:

1. Imported libraries such as Pandas, matplotlib, NumPy, and seaborn and loaded the insurance dataset

2. Checked the shape of the data along with the data types of the column

3. Checked missing values in the dataset and find there were no missing values , therefore no action needed.

4. Explored the relationship between the feature and target column using a count plot of categorical columns and a scatter plot of numerical columns

5. Performed data visualization using plots of feature vs feature

6. Checked if the number of premium charges for smokers or non-smokers is increasing as they are aging

7. After each step, observed the results