

## Problem Statement:

# Patient Charges - Clustering and Regression

## The Impact of Medicine to your Wallet



### Brief Introduction

In this project, the main aim is to show ways to go deep into the data story-telling even though the dataset is small. Also, you will work on a model that could give us an approximation as to what will be the charges of the patients. Nevertheless, you must go deeply into what factors influenced the charge of a specific patient. In order to do this you must look for patterns in our data analysis and gain extensive insight of what the data is telling us. Lastly, you will go step by step to understand the story behind the patients in this dataset only through this way we could have a better understanding of what features will help our model have a closer accuracy to the true patient charge.

### Table of Contents:

#### I. Importing Libraries

#### II. Distribution of Medical Charges

#### III. Age Analysis

- a) Turning Age into Categorical Variables
- b) Is there a relationship between BMI and Age?
- c) Comparing Independent Variables with ANOVA
- d) Who got charged more on Average by Age

#### IV. Weight Status

- a) Turning BMI into Categorical Variables
- b) Weight Status vs Charges
- c) Obesity and the Impact of Smoking to the Wallet
- d) Distribution of Charges (Obese Smoker vs Obese non-Smoker)
- e) Separation in Charges between Obese Smokers vs Non-Obese Smokers

#### V. Regional Analysis:

- a) Building a Contingency Table
- b) Average Patient Charge by Region
- c) Average charge by region depending on weight condition

#### VI. Unsupervised Learning Algorithms:

- a) Performing Clustering Manually
- b) K-Means Clustering
- c) Hierarchical Clustering