U.S. Health Insurance Linear regression model.

The focus of this analysis is to present a descriptive analysis and build a linear model to predict future health insurance charges. The data set was taken from a Kaggle base entry on a U.S census bureau of healthcare costs.

About the data set

The data set is Healthcare charges of a sample size of 1,338 individuals. The charges are healthcare insurance costs based on a yearly rate. The Charges are based on whether the individual is a smoker or not, age, sex, region, family size, and BMI (Body Mass Index).

Insights:

Was able to determine that

- individuals who smoke have a higher average cost than those who don't smoke.
- Also, BMI and your age are moderately correlated to charges an individual faces.
- There so a descending trend in charges as individuals have a larger family size.

Prediction Model

I simple linear regression shows that a prediction of charges based on the if the customer smokes, bmi, and age. Also, a coefficient correlation chart to shows which columns are making a major impact on insurance cost.

Results:

Model outputs of 74% accuracy
Smokers have a 0.79 a coefficient as it relates to charges
Age has a 0.3 coefficient as it relates to charges
BMI has a 0.2 coefficient as it relates to charges