

# Insurance Scoring

Based on [consumerfinance.gov](https://consumerfinance.gov), insurance is **a way to manage your risk**.

When you buy insurance, you purchase protection against unexpected financial losses. The insurance company pays you or someone you choose if something bad happens to you.

**BUT**, how much you should pay to get that protection?  
and, **WHY?**

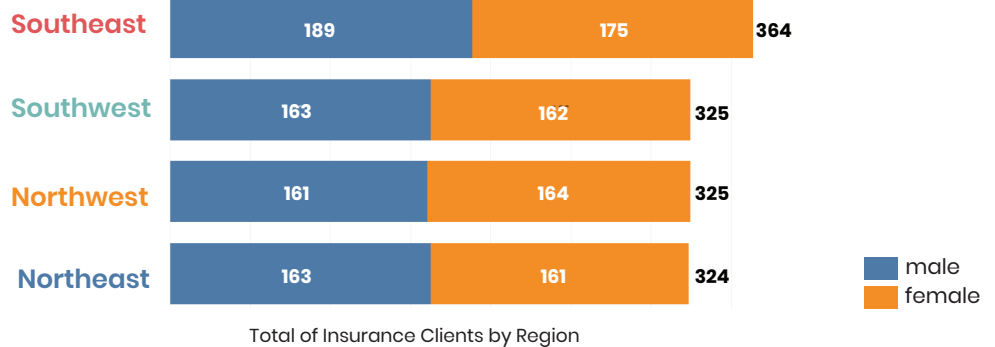
With the data from 4 regions we have, we will try to elaborate the insurance scoring.

Total of insurance clients is  
**1,338 persons.**

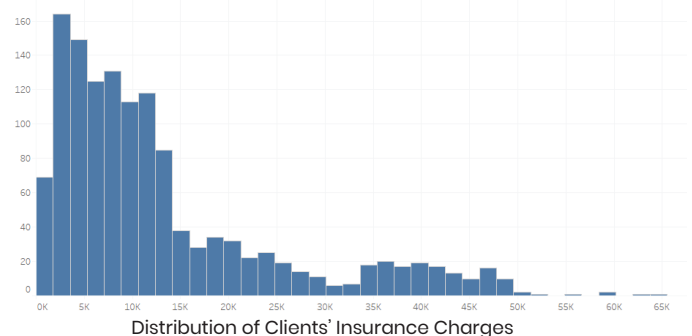
Total of male clients is  
**676 persons.**

Total of female clients is  
**662 persons.**

The client is mostly from Southeast region, it is 3% higher than other regions.



While the age is varied from 15-65 years old, average age is 39. The highest number is 20 years old with the total of 145 clients.



## Age and Charges of Clients

### BMI

The BMI range score is based on [cdc.gov](https://www.cdc.gov/obesity/basics/adult-defining.html), divided into 4 categories: obese, overweight, healthy weight, and underweight. (<https://www.cdc.gov/obesity/basics/adult-defining.html>)

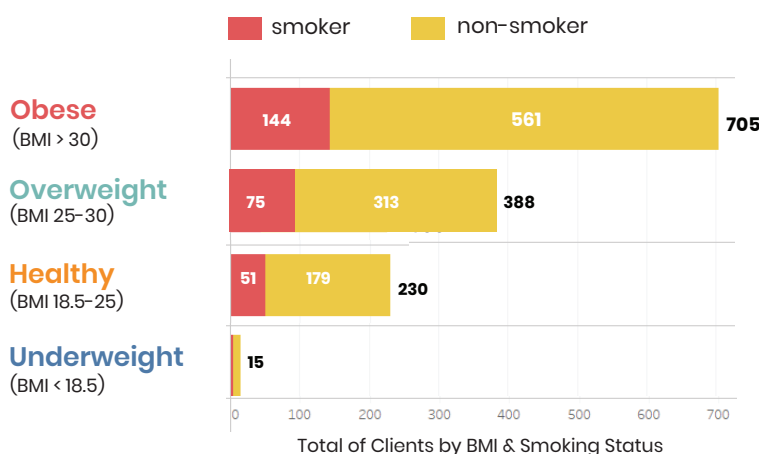
The distribution of clients' insurance charges is **skewed to the right**. The median charges is 9,382 while the average charges is 13,270. It means **price of the insurance is considered lower for most range of age**.

Given a smoker, there are around:

**5/10 male** clients who are smokers

**4/10 female** clients who are smokers

**Smoking Status**



Based on gender, the average BMI is:

**male** 30.94

**female** 30.37

Based on smoking status, the average BMI is:

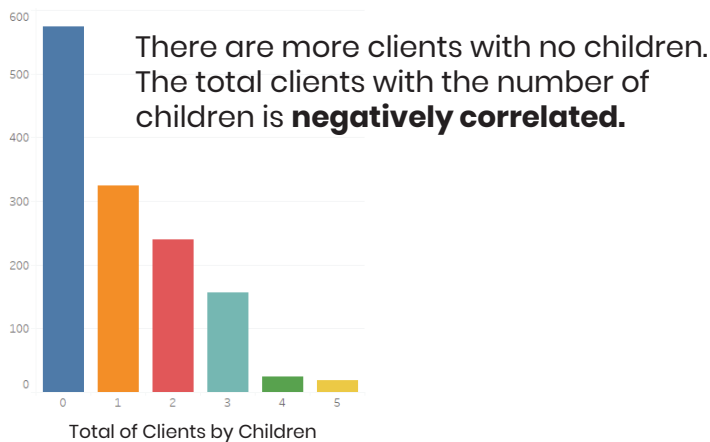
**smoker** 30.7

**non-smoker** 30.6

As seen on the graph, **50%** of the clients are obese. The distribution is most likely **skewed to the left**.

The charges for those overweight and obese given a smoker is **four times higher** than the non-smoker.

## Children



## Smoker + Overweight and Obese

avg. charges **35,116**

## Non-smoker + Overweight and Obese

avg. charges **8,929**

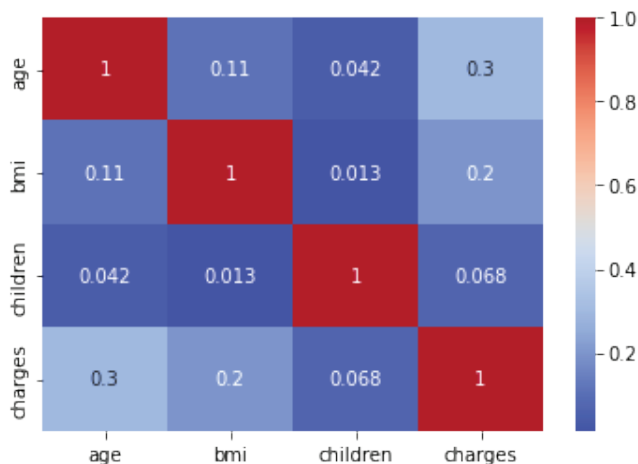
With the variance of charges:

Smoker **11,541**

Non-smoker **5,993**

In spite of the gender and BMI, **smokers have higher chance to be charged more** than non-smokers.

## What features correlated the **most** with charges?



As per diagram, **number of children is the most correlated** with the charge of insurance.

It means when **the total of children increases, so does the charges**.

But, there is *lack of categorical data*, such as: gender and smoking status, from the dataset.

### • Do smokers get higher charges than non-smokers?

\*(Test using t-test and alpha = 0.05)

H0: Smokers get lower charges than non-smokers.

H1: Smokers get higher charges than non-smokers.

With p-value = 8.271435842179102e-283.

We accept H1 that **smokers do get higher charges than non-smokers**.

### • Do male clients get higher charges than female clients?

\*(Test using t-test and alpha = 0.05)

H0: Male clients get lower charges than female clients.

H1: Male clients get higher charges than female clients.

With p-value = 0.03613272100592976.

We accept H1 that **male clients do get higher charges than female clients**.

### • Do older clients get higher charges than younger clients?

\*(Test using t-test and alpha = 0.05)

H0: Older clients get lower charges than younger clients.

H1: Older clients get higher charges than younger clients.

First, the data is divided into 2 categories, those are:

= clients from 18-41 years old;

= clients > 41 years old.

With p-value = 4.926339120398916e-25.

We accept H1 that **older clients do get higher charges than younger clients**.

## Conclusion

If you are **a male, older than 41 years old, a smoker, have children, and have BMI > 25**, it means **your insurance has higher probability to be charged than** people that are younger, not smoking, have no children, and have lower BMI.

Sources:

- Dataset (.csv)

- Python file (.ipynb)