Problem: Analysis of health insurance csv from Kaggle.com.

This graph clearly shows that people who smoke must pay high for their health insurance cost, as there is a high probability for that person to become ill, and therefore have problems related to heart, lungs etc.

Chart, box and whisker chart

Description automatically generatede

I also tried putting the number of children and individual has and tried to guess the average price paid by that individual, but at the end something which came as a conclusion is that the number of children does not make a huge difference in the health insurance cost, but if they smoke, their insurance prices were high.

Chart, bar chart

Description automatically generated

We see that most people have their health insurance in the range of 1000$ to 10000$, as that is around the average people must pay, where people who smoke make some kind of an exception, as their prices end up being particularly high.

Chart, histogram

Description automatically generated

The area where the people resided should have made some difference in the health insurance prices, but when I see northwest, northeast, southwest, and southeast, there is no much difference in the prices.

Northwest – 12417

Southwest – 12346

Northeast – 13406

Southeast – 14735

Smoker Southwest – 32269

Smoker Southwest – 8019

So, this is an example of smoker vs nonsmoker in a particular area.

Prediction of linear regression: This ends up being the per unit increase in the age, smoker, and children. This is kind of decided by the average of all the coordinates on a graph.

Table

Description automatically generated with medium confidence