HW1

Maedeh Karkhaneh Yousefi

February 17, 2022

The data consists of 7 columns, with titles being: 'age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'. It consists of 1338 rows of entries. Types include float64, int64, object. I filtered the charges between 10000 range of numbers, so that it would be easier to represent in *countplots*. Therefore, the specific values of charges in the table are replaced with:$<= 10k$, $> 10k$ & $<= 20k$, $> 20k$ & $<= 30k$, $> 30k$ & $<= 40k$ and $> 40k$.

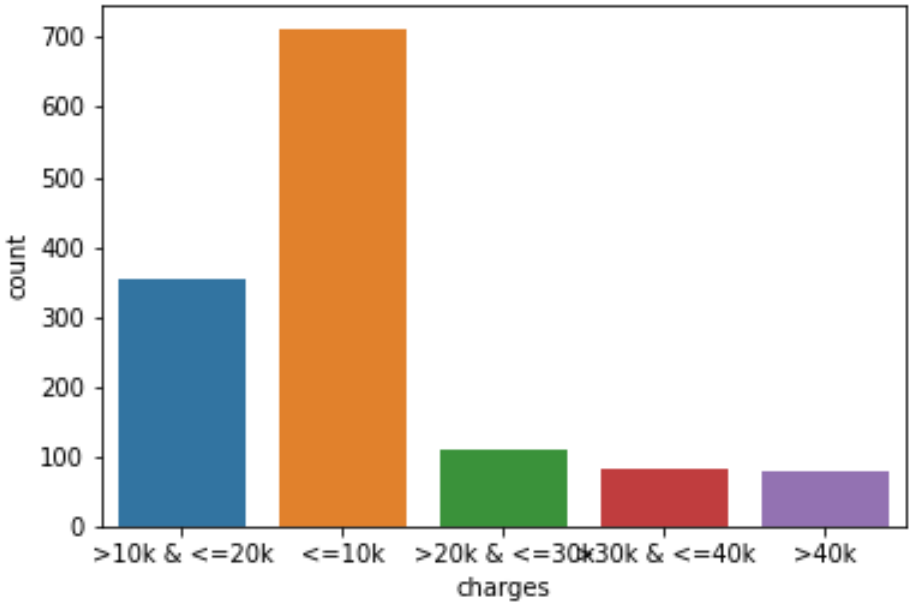

Figure 1

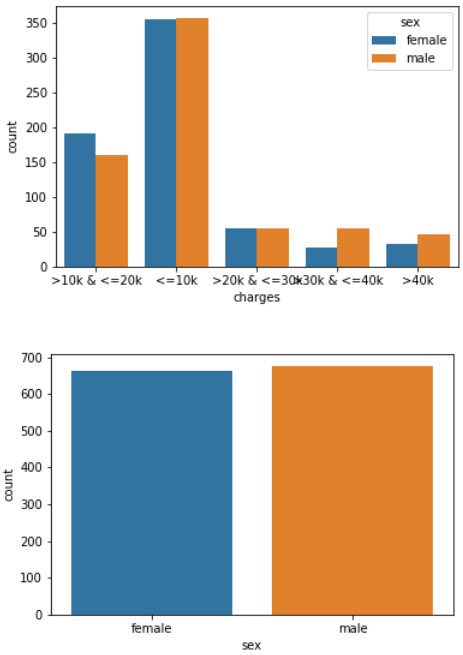

Figure 2: The number of females and males is almost equal. From the first plot it can be seen that when charges go above 30000, males have more contribution.

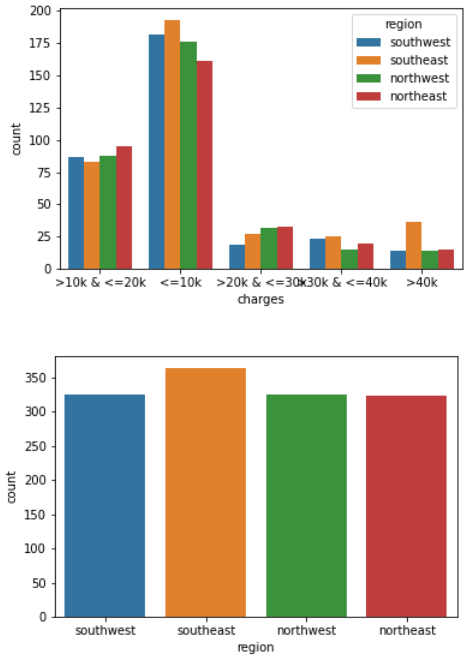

Figure 3: Regions are equally distributed except for southeast, which is a little bit more than others. This could be why it has more contribution to some ranges of prices.

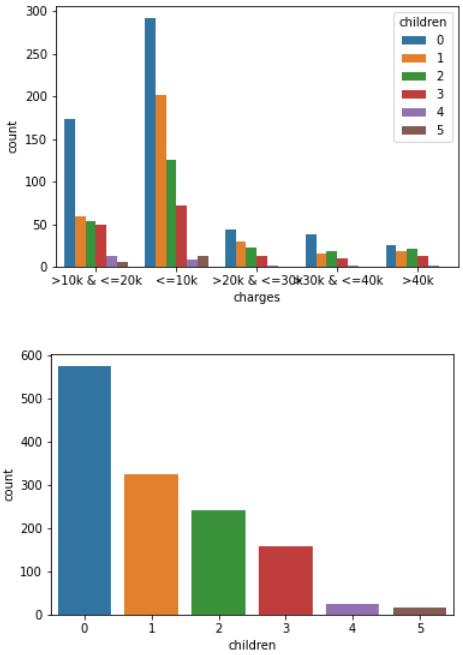

Figure 4: It seems like the number of children doesn't have much effect on charges.

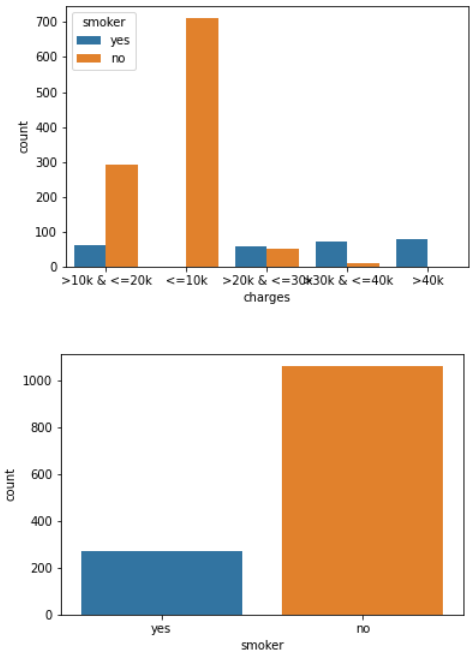

Figure 5: From plots we can conclude that being an smoker results in higher charges.