

Problem Statement

- ▶ Variables that contributes to the profitability of insurance premium

Data Cleaning

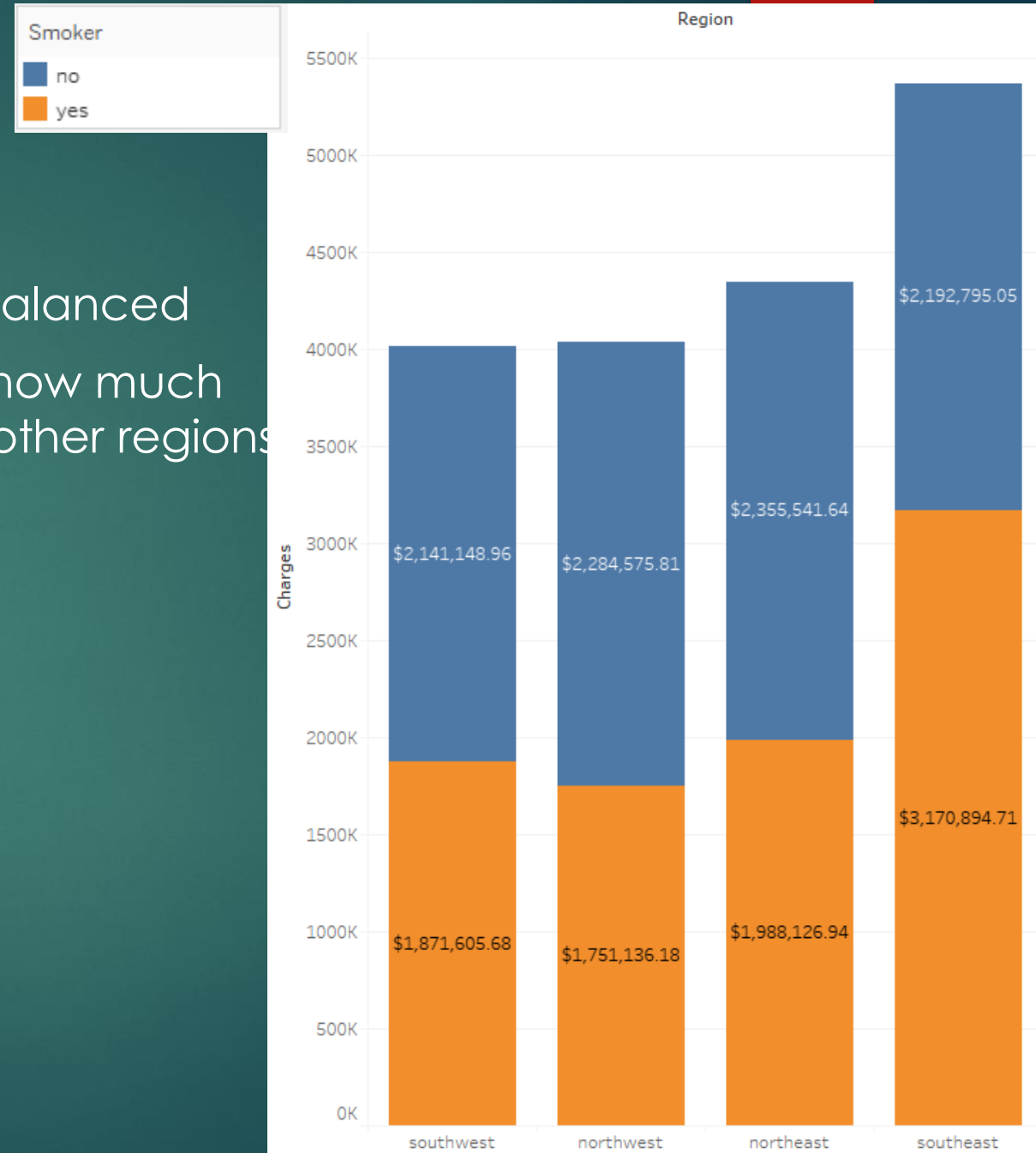
	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age         1338 non-null   int64
1   sex         1338 non-null   object
2   bmi         1338 non-null   float64
3   children    1338 non-null   int64
4   smoker      1338 non-null   object
5   region      1338 non-null   object
6   charges     1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

No missing values from any columns while age, BMI, children and charges are in either int or float. Data can be used straight away.

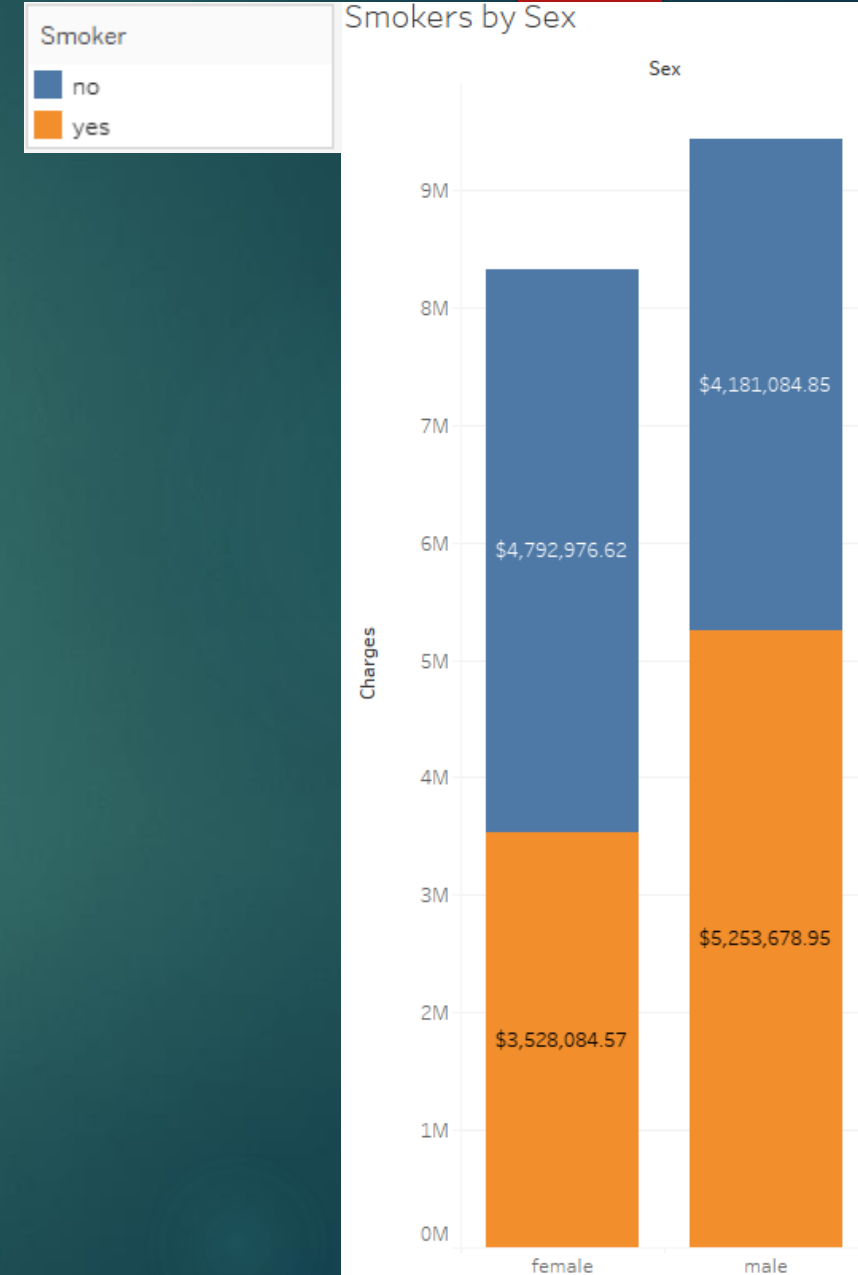
Insights - Region

- ▶ Southeast region pays the most and rest of the regions are comparable and well-balanced
- ▶ The available variables that could impact on how much the premium is paid for southeast against the other regions are BMI or children to find the profitability



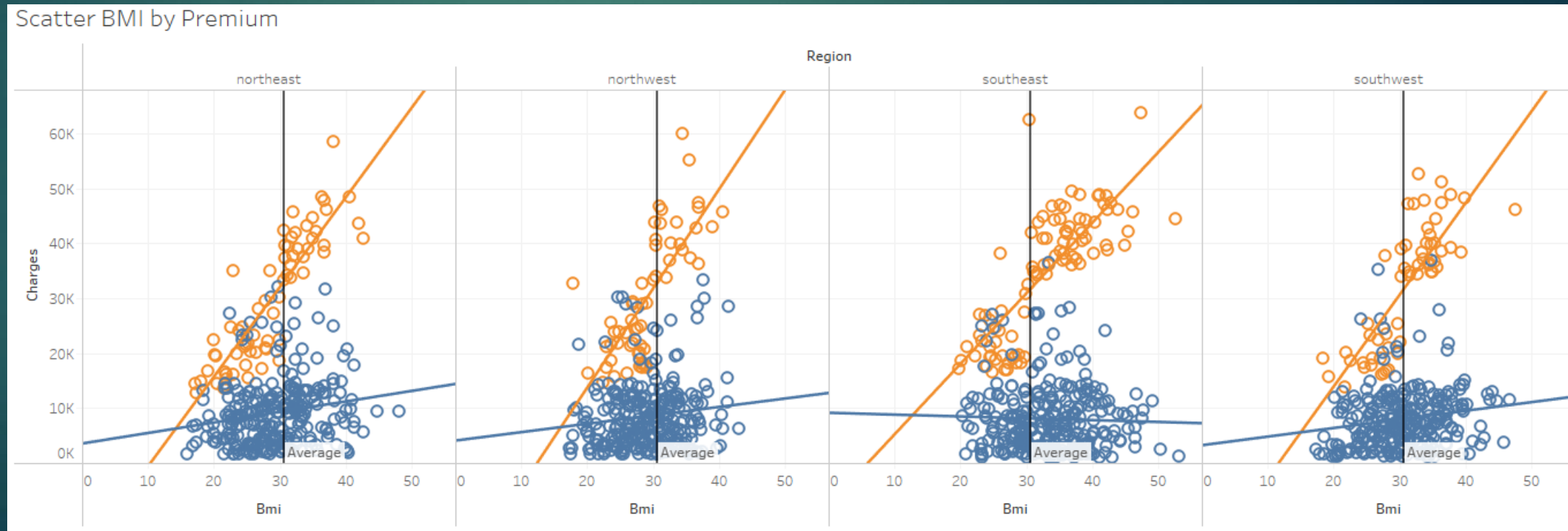
Insights – Sex factor

- Sexuality does not matter whether there is money to be made, although male smokers do pay substantially more



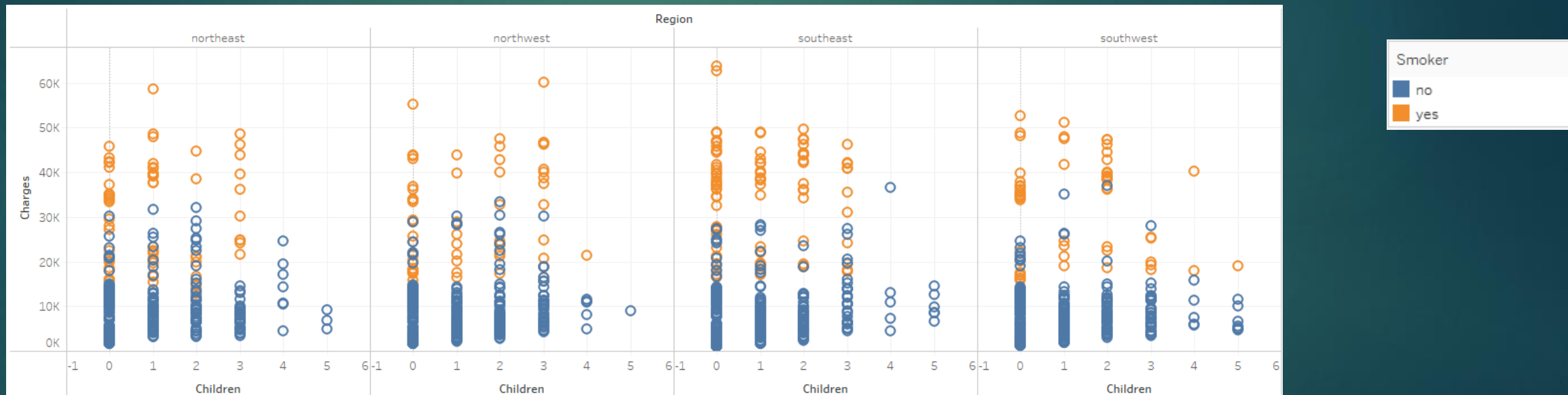
Insights – Region by BMI

- Mean line by BMI is shown, no correlation between the relations of premium paid against BMI
- However, a positive correlation of smokers with higher BMI seems to suggest that they do pay at higher premium



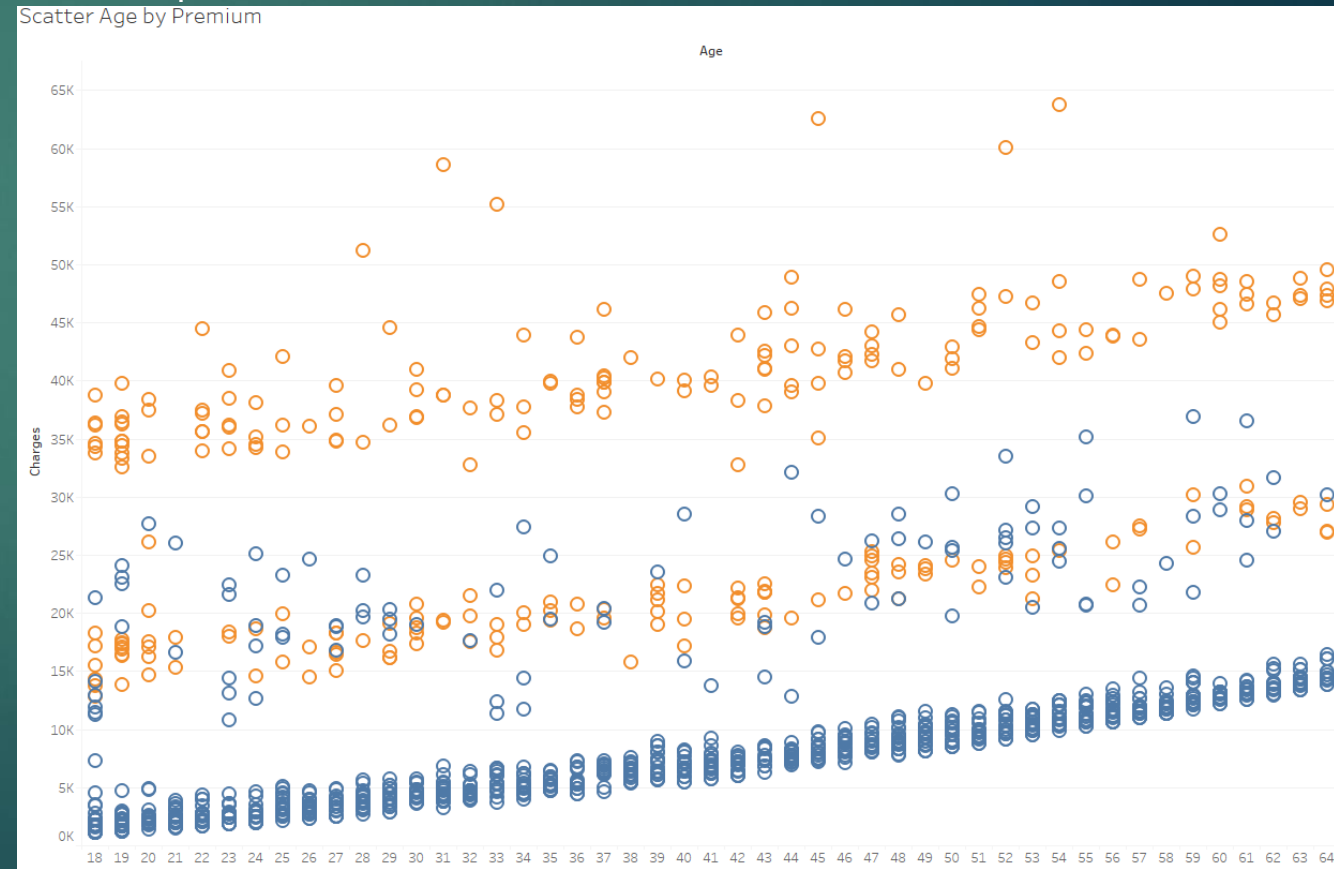
Insights – Region by Children

- Overall, parents with less than 4 children are willing to buy insurance but smokers will pay much more
- Possibly due to financial means of a family, but cannot be confirmed due to lack of data.



Insights – Age factor

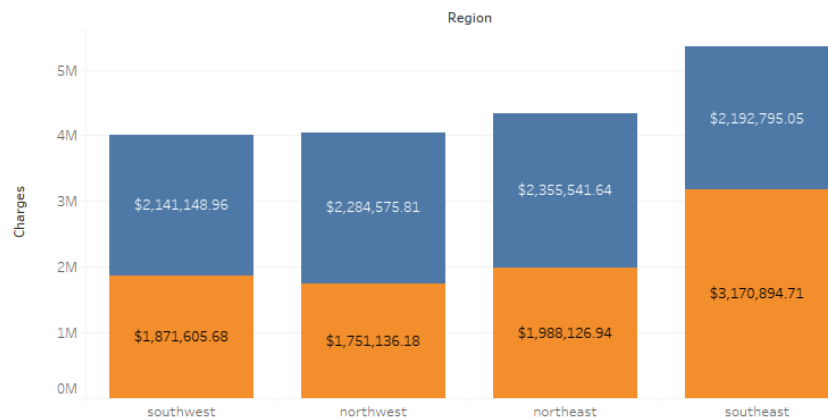
- ▶ Positive correlation showing an uptrend irrespective of regions, the skew of older age pays significantly more compared to youngsters especially if they are smokers
- ▶ Obvious outliers spotted that could be impacted from different variables that were not provided, more information is needed to ascertain the problem



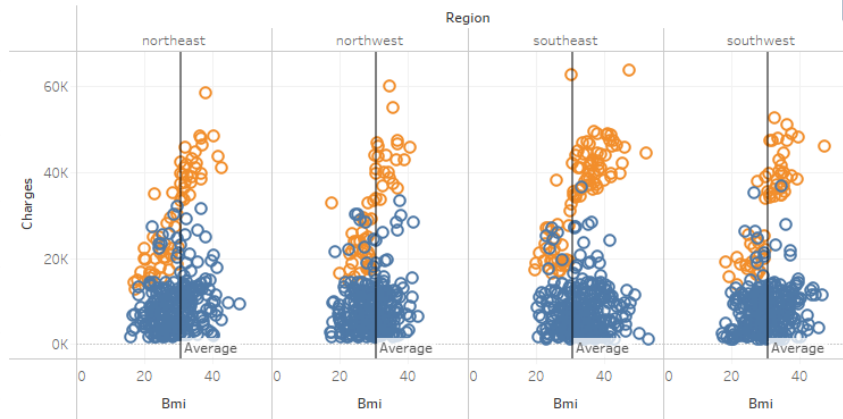
Dashboard

Insurance Premium Study

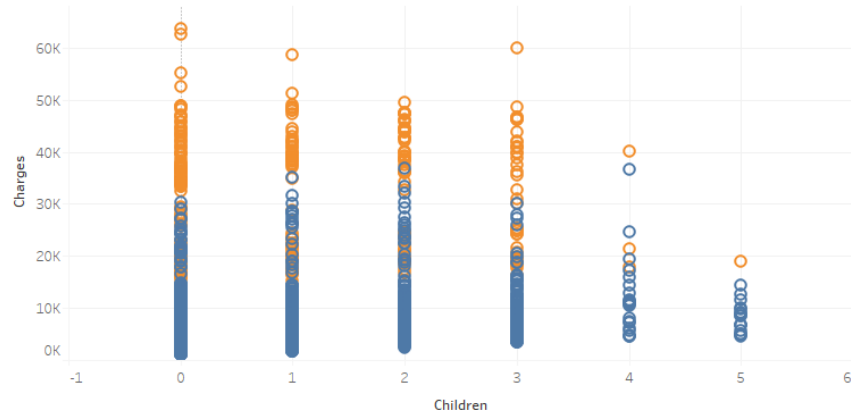
Region



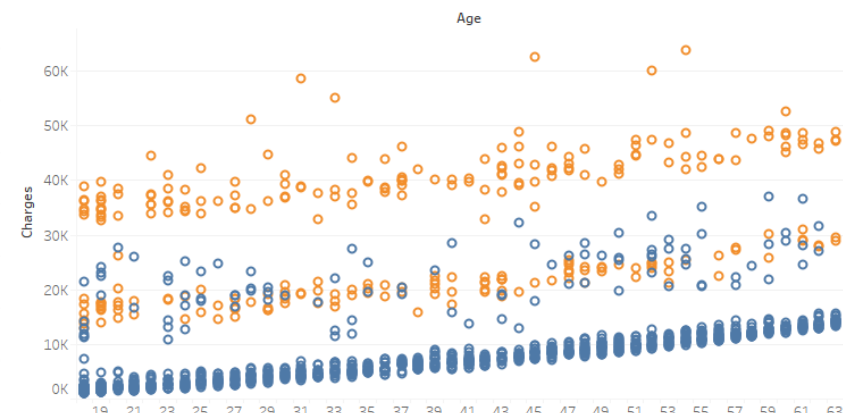
Scatter BMI by Premium



Scatter Children by Premium



Scatter Age by Premium



<https://public.tableau.com/app/profile/chua.ding.zhang/viz/InsurancePremiumPaid/Dashboard1?publish=yes>

Conclusion

- ▶ There is a positive correlation between smokers with higher BMI, or older age and it will yield higher premium paid
- ▶ Customers with fewer children, preferentially, less than 4 would pay for insurance. And more than 3 suggest that they are less likely to smoke
- ▶ Therefore, smokers generally pay at much higher premium due to direct or indirect complications, such as health or child