

CS6: Data Visualization Final Project



Question:

What factors do health insurance companies consider most when setting charges?





Objectives

- Showcase Data Visualization techniques learned in the course on an uncomplicated set of data using [colab](#).
- Present the key findings using a range of graphs.
- Draw basic statistical conclusions from the data analysis and visualization.
- **Answer the main question:** What factors do health insurance companies consider most when setting charges?





Data Set

Medical Cost Personal Dataset

(Public Domain Data)

→ Entries

1338 (adults aged between
18 and 64 years)

→ Variables

Seven

→ Source

[Kaggle](#)



seaborn

Find a Dataset

Import
Libraries

Understand
Data

Clean and
Prepare

Vizualise

Understand Data: Variables

Age	Sex	BMI	Children	Smoker	Region	Charges
Primary beneficiary	Insurance contractor (M/F)	Body Mass Index	Number of dependents	Tobacco smoker (Y/N)	Residential area in the US (NE, SE, SW, NW)	Individual medical costs billed by health insurance

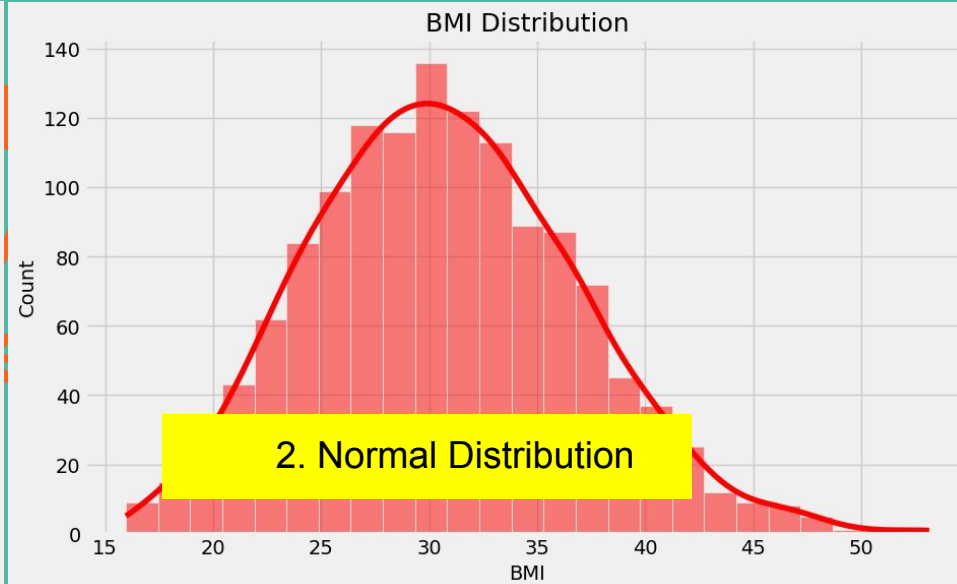
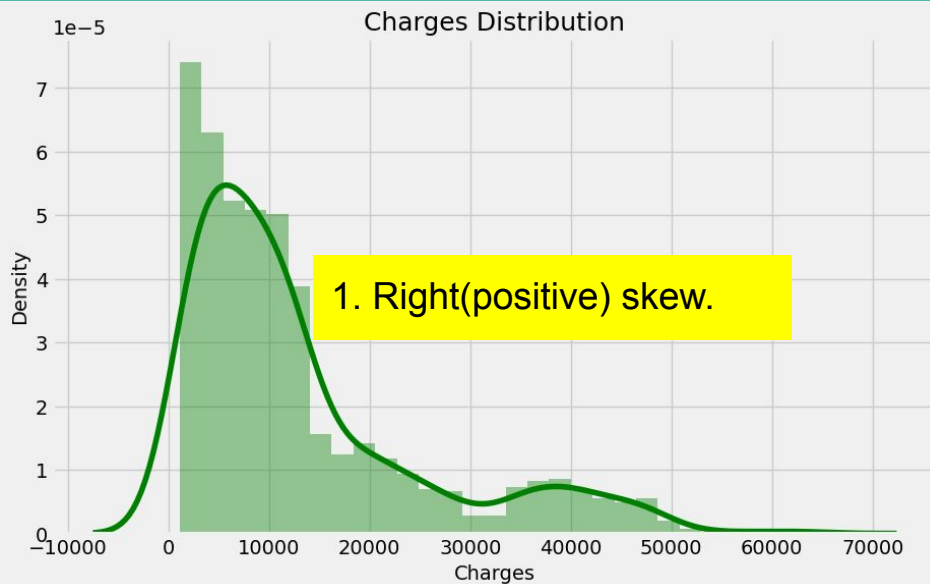
Understanding Data: Descriptive Stats.

	Age (Years)	BMI	Children	Charges (\$)
mean	39	31	1	13270
std	14	6	1	12110
min	18	16	0	1122
25%	27	26	0	4740
50%	39	30	1	9382
75%	51	35	2	16640
max	64	53	5	63770

Understand Data: Distribution of values for a single variable

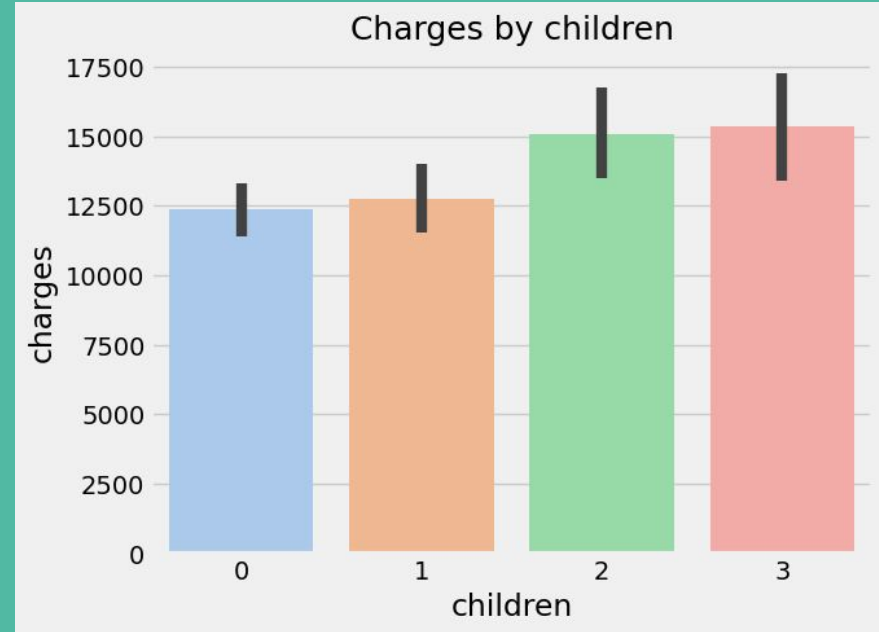
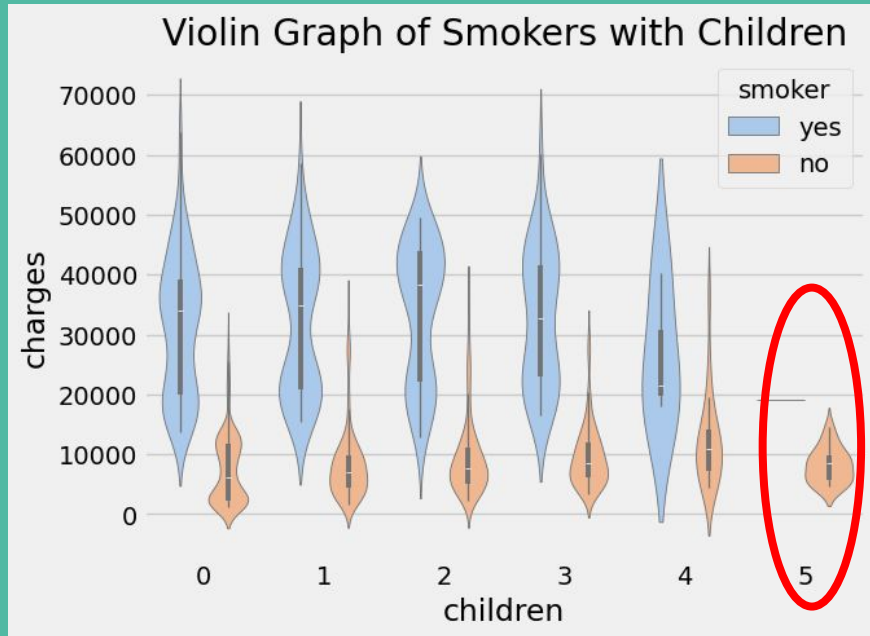
1. Most of the charges are between 0 – 10000\$

2. Most people have BMI between 27 - 34



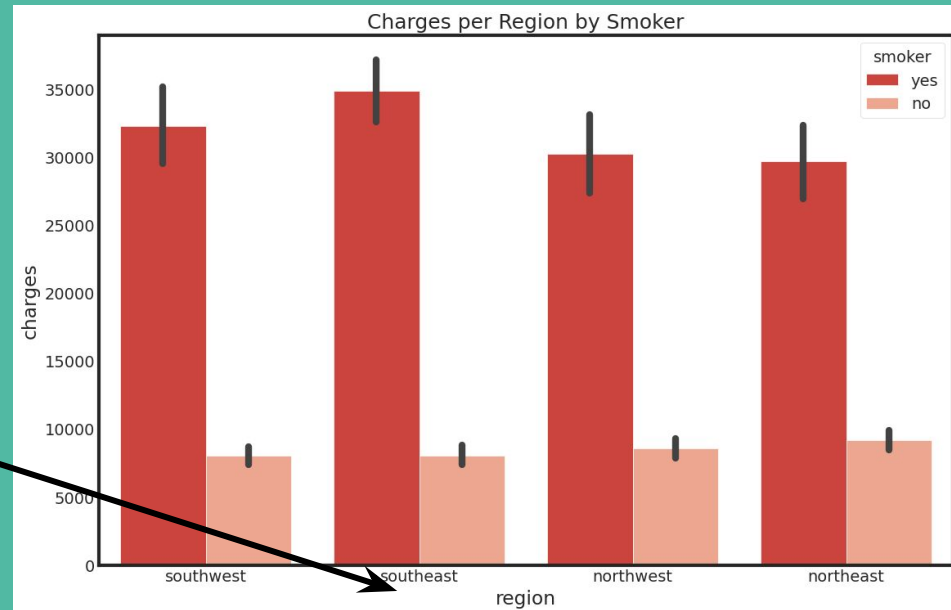
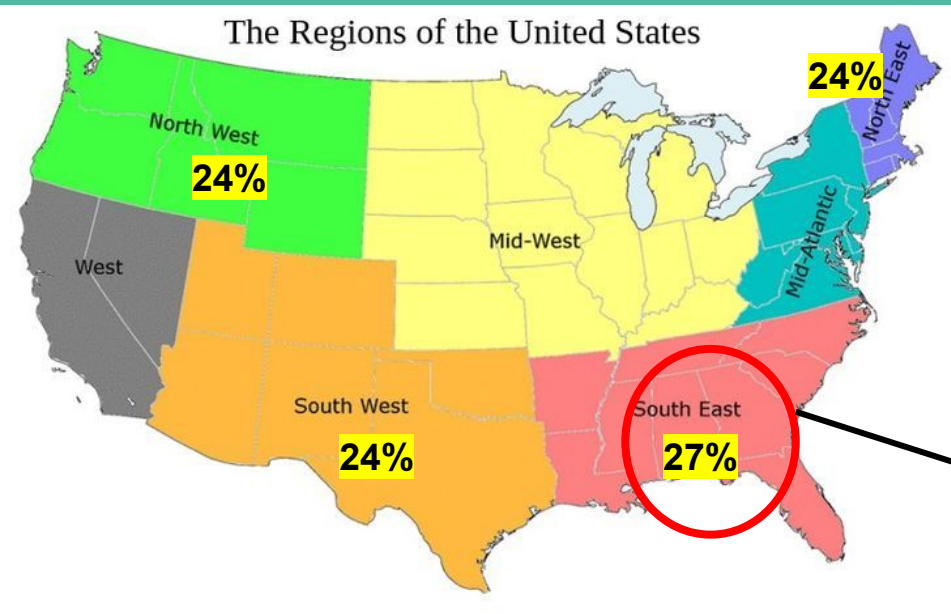
Key Findings from Data Visualisation:

People who have children smoke less and are charged less than smokers who tend to have fewer or no children (Violin). Individuals with 2-3 children have higher average insurance premiums compared to those with 0-1 children (Bar).



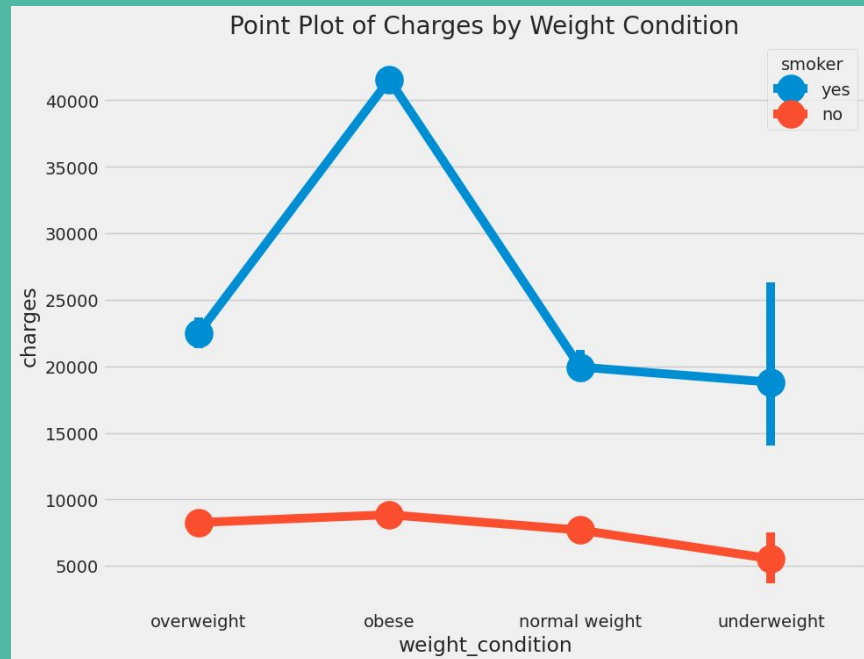
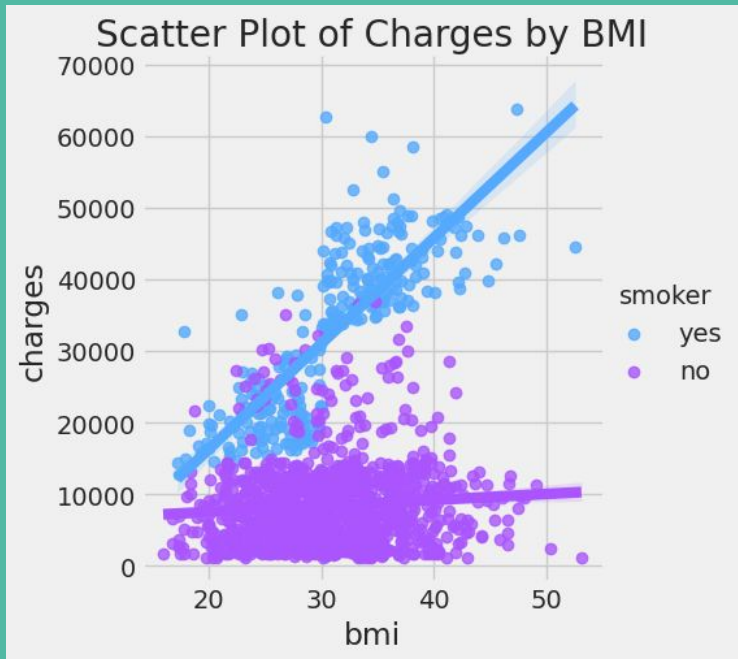
Key Findings from Data Visualisation:

Region value counts indicate there is almost an equal distribution between the four regions with Southeast having slightly more people and the highest number of smokers. This region also pays the highest charges.



Key Findings from Data Visualisation:

Charges and BMI seem positively correlated. Dramatically steeper line for smokers in comparison to non-smokers. Obese smokers pay significantly higher charges.



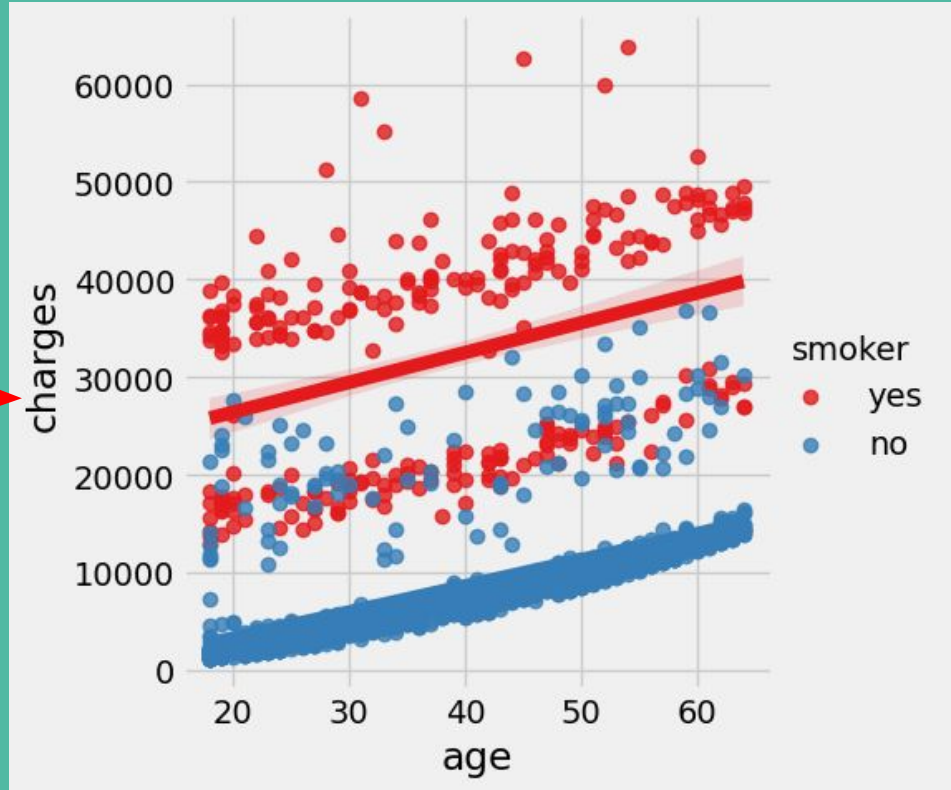
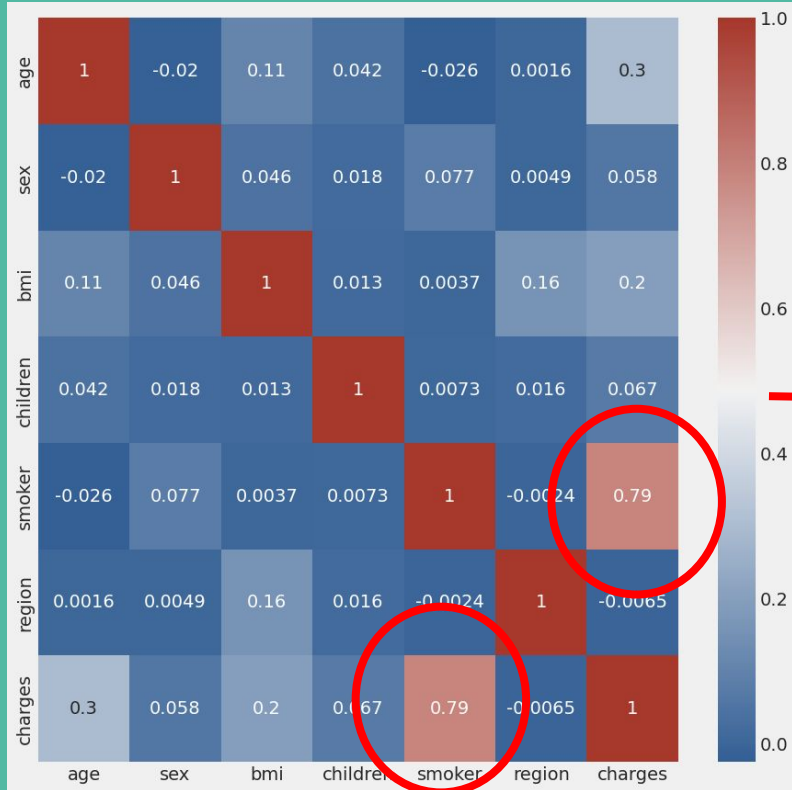
Key Findings from Data Vizualisation:

The average charges for smokers are almost 5 times higher than those for non-smokers.



Key Findings from Data Visualisation:

Older people are charged more, but the highest charges apply to smokers, even if young!



Question:

What factors do health insurance companies consider most when setting charges?

Answer:

Smoking has the highest impact on medical costs, even though the costs are growing with age, BMI and the number of children.

