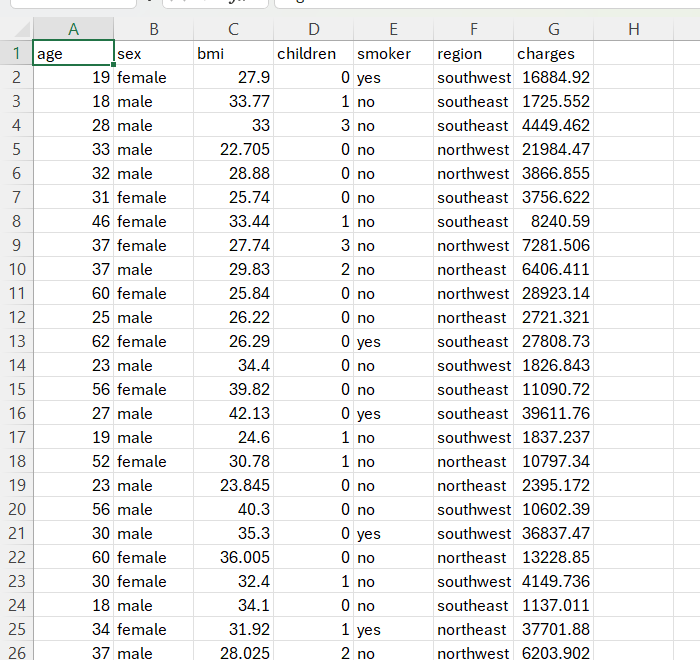
->No missing values

**One hot encoding**

-used to represent categorical data as numerical data

-recommended because it does not assume ordering among the categories.

Data before One hot encoding



Data after One hot encoding

**A screenshot of a computer

Description automatically generated**A black screen with white text

Description automatically generated

**Individual box plots of each column that are originally numerical**

->One hot encoded columns are not included because it does not make sense to box plot binary values.

A screenshot of a graph

Description automatically generated

**Box Plot For Age**

* Distribution has a wide range (20-65 years)
* The median age is around 39-40
* Relatively balanced distribution with no visible outliers, indicating that all ages are within a common range.

**Box Plot For BMI**

* The median BMI is approximately 30(overweight)
* Several outliers are visible beyond 40,Indicating that some individuals have higher BMI
* The plot shows a significant spread, indicating that Bmi values vary widely among inidividuals.

**Box Plot For Children**

* The median is 1, indicating most individuals have one child.
* Majority of individuals have 0 to 2 children
* No visible outliers, indicating that the number of children per individual falls within a typical range.

**Box Plot For Charges**

* The distribution is highly skewed, with several high-value outliers above 30,000
* Outliers indicates a small number of individuals with higher medical expenses.
* Median charges are relatively low compared to the maximum values, indicating a small number of high medical costs.
* Charges may not follow a normal distribution, instead influenced by high cost cases.