

APPLIED DATA SCIENCE ASSIGNMENT 1

REPORT

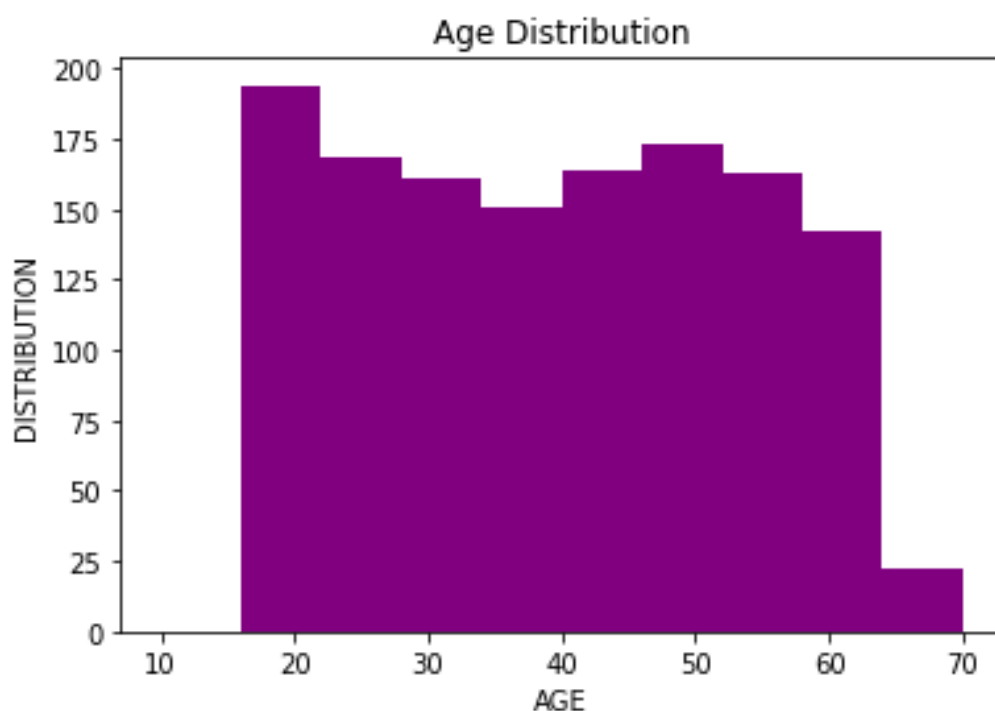
DEVANSHI PANKAJ WADKAR
STUDENT ID: 22065576

GIT HUB REPO LINK: <https://github.com/Devanshi-45/APPLIED-DATA-SCIENCE-1>

DATASET FROM KAGGLE:

<https://www.kaggle.com/datasets/willianoliveiragibin/healthcare-insurance>

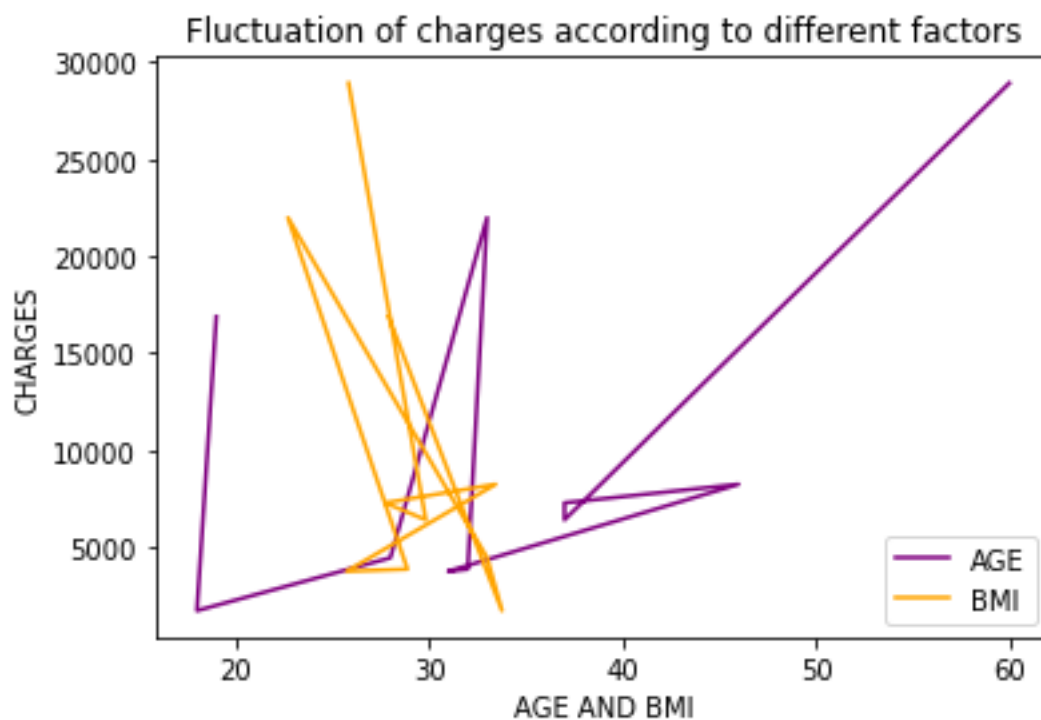
Visualization 1: Histogram - Distribution of ages of customers



Histograms are a good way to show the distribution of the data which is usually continuous and in a specific interval.

This visualization is used to show the distribution of the customers purchasing the healthcare insurance. We get to see that there are major customers in the age of 18-22.

Visualization 2: Line plot – Changes in price according to Age and BMI



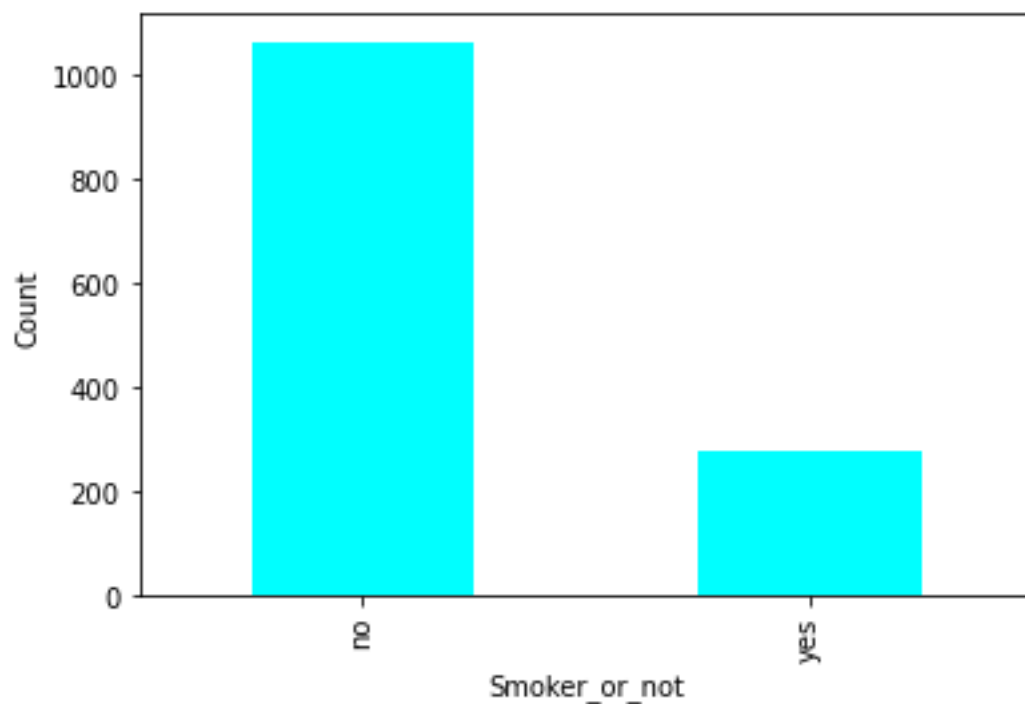
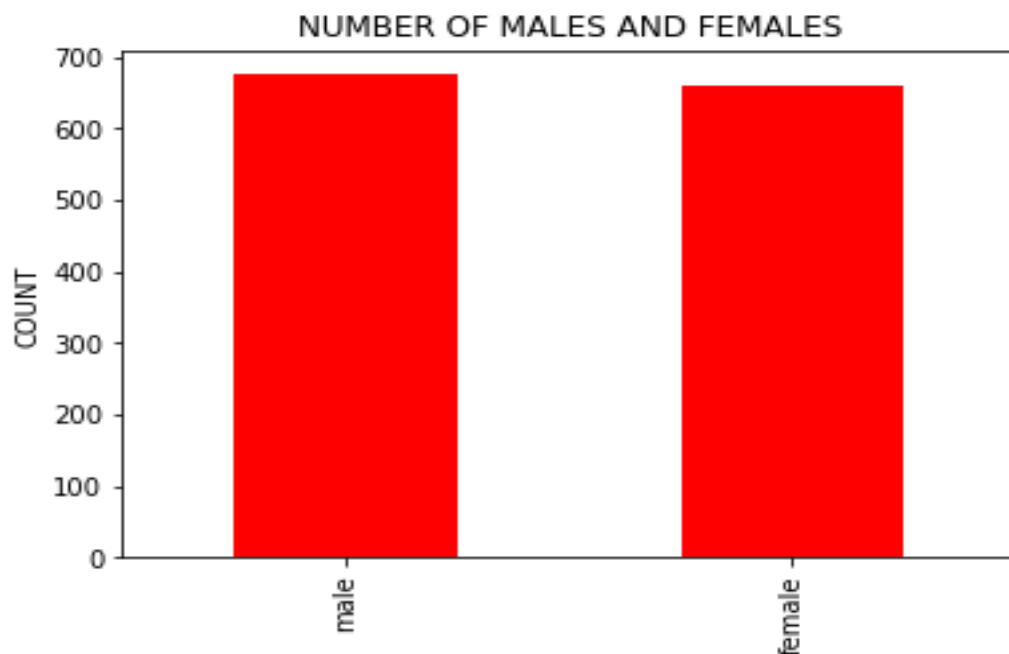
These are line plots which are usually used whenever we want to observe changes in one variable with respect to the other i.e. changes in some quantity w.r.t to other quantity.

Here there are two different line plots which describe the change in the charges of the insurance based on two different factors they are Age and BMI

Here it is quite hard to predict the change because of the number of entries in the dataset but we can see as the age increases the charges tend to increase. But the same cannot be said for the other.

The legend is specified to recognize the line for that particular factor.

Visualization 3: Bar plots- Count of sex and smokers

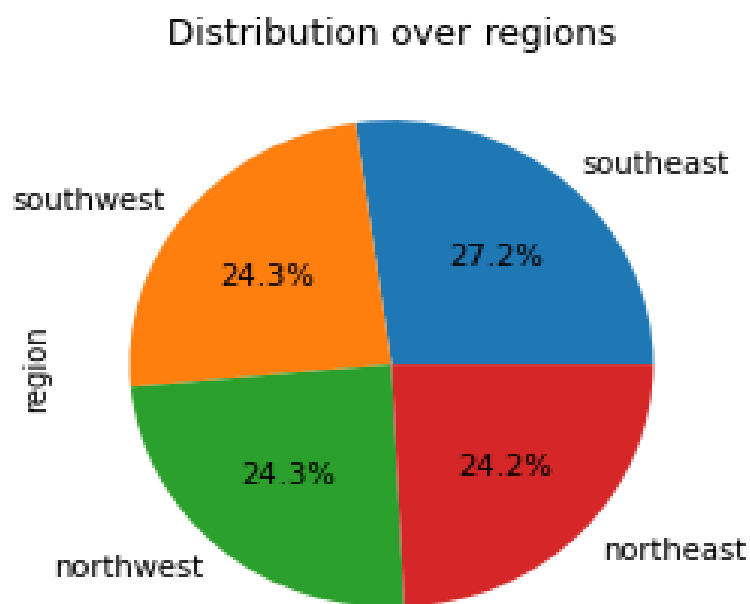


Bar charts or plots are commonly used for categorical type of data.

Here we have used the bar plots to plot two different things, the count of males and females in the customers it shows that there is not much difference between the number of male and female.

While the second bar plot shows the count of number of smokers and non-smokers, it depicts there are more non-smokers amongst the customers with a difference of almost 800.

Visualization 4: Pie Charts – Distribution of customers amongst different regions.



The pie chart is used frequently to show the percentage distribution of a particular thing.

Here the pie chart is used for showing the distribution of customers who stay in different regions of the country and the highest number of customers stay in the Southeast with 27.2% and the remaining are almost equally located in the Southwest, Northeast and Northwest simultaneously.