Heart Attack Analysis & Prediction

Data
I propose to work on the heart attack prediction dataset from Kaggle. This dataset and the relative problem is of health and the likelihood of patients to suffer heart attack. The context of the data is that given certain variables, is one able to build a model that will correctly predict if a patient will suffer a heart attack.

Description
This project object is to detect whether patients have heart disease or not by given several features from patients. The motivation of this project is to save human resources in medical centers and improve accuracy of diagnosis. In our project we use different methods to detect heart disease such as Logistic Regression, SVM, Naïve Bayes, Random Forest and Artificial neural network. And among all these algorithms Random Forest gives us the best accuracy of 91.8%

Tools
✓ NumPy
✓ Pandas
✓Matplotlib
✓ Seaborn

Rustle :

I called my dataset using pandas, numpy, matplotlib.pyplot, seaborn
Then I used the Data Clean

Data Cleaning ¶

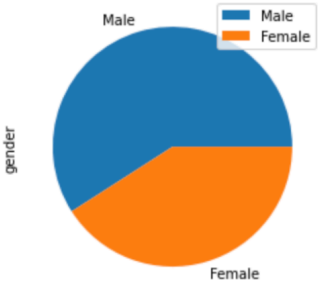
```
data.rename(columns={'smoking_status':'smoking','ever_married':'married'} , inplace=True)
data.head()
```

	id	gender	age	hypertension	heart_disease	married	avg_glucose_level	bmi	smoking	stroke
0	9046	Male	67.0	0	1	Yes	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	174.12	24.0	never smoked	1

```
data.duplicated().sum()
```

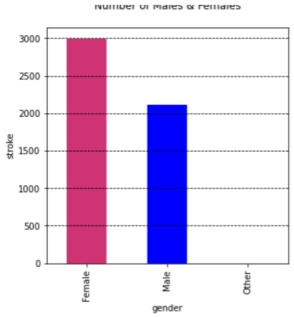
```
0
```

```
data.drop(columns=['bmi'] , inplace=True)
```

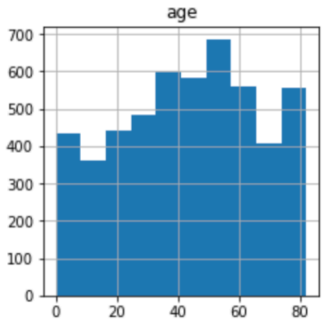


Pie chart

Result :Almost 59% of patients are males

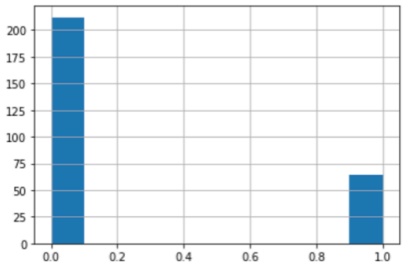


Bar chart

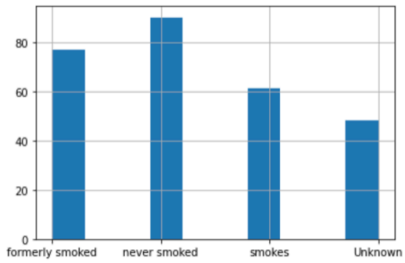


Histograms

Result : Most of patients almost 94.5% are over 50 years old



Ruslt: Almost hypertension 64



Ruslt :Almost 50% of the patients were either smoking or formerly smoked