Cardiovascular Disease dataset

Introduction:

Heart and blood vessel disease (also called heart disease) includes numerous problems, many of which are related to a process called atherosclerosis.

Atherosclerosis is a condition that develops when a substance called plaque builds up in the walls of the arteries. This buildup narrows the arteries, making it harder for blood to flow through. If a blood clot forms, it can block the blood flow. This can cause a heart attack or stroke.

the cardiovascular disease dataset is an open-source dataset found on **Kaggle**

About The Dataset:

The data was obtained from the kaggle website: **https://www.kaggle.com/sulianova/cardiovascular-disease-dataset**

The dataset Contains 70000 Rows and 11 Columns:

- id: ID number
- age: in days
- gender: 1 - women, 2 - men
- height: cm
- wight: kg
- ap_hi: Systolic blood pressure
- ap_lo: Diastolic blood pressure
- cholesterol: 1: normal, 2: above normal, 3: well above normal
- gluc: 1: normal, 2: above normal, 3: well above normal

- smoke: whether patient smokes or not

The dataset is available as the `.csv` file. a sample of data is shown in the following table:

```
df.head()
```

Out[5]:

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

Tools:

- There are tools that will be used to achieve the goal of this project, such as Jupyter notebook,Numpy,Matplotlib,pandas.

Questions This Project Will Answered:

- **Is the one who smokes the most prone to heart disease?**
- More susceptible to heart disease, women or men?
- Who's more likely to have heart disease?
- How much is cholesterol in heart disease?

TO DO:

- Explore the data and come up with EDA phases then use a model to fit the data.
- **NOTE:** the used features may be increased or changed and the model as well.