**PROGRESS REPORT**

**ON**

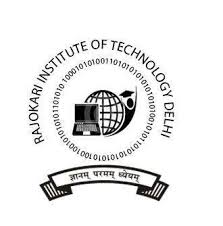
**HEART DISEASE PREDCTION SYSTEM**

**USING MACHINE LEARNING**

SUMITTED FOR PARTIAL FULFILMENT OF DIPLOMA IN IT ENABLE SERVICES & MANAGEMENT

BY

**RAHUL KUMAR**

**1813111034**

**Submitted to – Mrs. Sunita chaurasia**

**In-charge of ITES&M**

[RAJOKARI INSTITUTE OF TECHNOLOGY -(GOVT), RAJOKRI VILLAGE, RAJOKRI NEW DELHI, DELHI 110038](https://www.bing.com/local?lid=YN4070x4833719631170764844&id=YN4070x4833719631170764844&q=Rajokari+Farms&name=Rajokari+Farms&cp=28.52436637878418%7e77.11663818359375&ppois=28.52436637878418_77.11663818359375_Rajokari+Farms),INDIA

**HEART DISEASE PREDCTION SYSTEM**

**USING MACHINE LEARNING**

Introduction

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analysing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

**Scenario**

You have just been hired as a **Data Scientist** at a Hospital with an alarming number of patients coming in reporting various cardiac symptoms. A cardiologist measures vitals & hands you this data to **perform Data Analysis** and **predict** whether certain patients have Heart Disease. We would like to make a **Machine Learning algorithm** where we can train our AI to learn & improve from experience. Thus, we would want to classify patients as either positive or negative for Heart Disease.

**Goal**

* **Predict** whether a patient should be diagnosed with Heart Disease. This is a **binary** outcome.

**Positive** (+) = 1, patient diagnosed with Heart Disease

**Negative** (-) = 0, patient not diagnosed with Heart Disease

* Experiment with various **Classification Models** & see which yields greatest **accuracy.**
* **Examine trends & correlation** within our data
* Determine which **feature** are **most important** to Positive/Negative Heart Disease diagnosis

**Features & Predictor:**

Our**Predictor** (Y, Positive or Negative diagnosis of Heart Disease) is determined by 13 **features** (X):

1. **Age** (#)  
2. **Sex** : 1= Male, 0= Female (*Binary*)  
3. (**cp**)chest pain type (4 values -*Ordinal*):Value 1: typical angina ,Value 2: atypical angina, Value 3: non-anginal pain , Value 4: asymptomatic  
4. (**trestbps**) resting blood pressure (#)  
5. (**chol**) serum cholesterol in mg/dl (#)  
6. (**fbs**)fasting blood sugar > 120 mg/dl(*Binary*)(1 = true; 0 = false)  
7. (**restecg**) resting electrocardiography results(values 0,1,2)  
8. (**thalach**) maximum heart rate achieved (#)  
9. (**exang**) exercise induced angina (*binary*) (1 = yes; 0 = no)  
10. (**oldpeak**) = ST depression induced by exercise relative to rest (#)  
11. (**slope**) of the peak exercise ST segment (*Ordinal*) (Value 1: up sloping , Value 2: flat , Value 3: down sloping )  
12. (**ca**) number of major vessels (0–3, *Ordinal*) colored by fluoroscopy  
13. (**thal**) maximum heart rate achieved — (*Ordinal*): 3 = normal; 6 = fixed defect; 7 = reversible defect

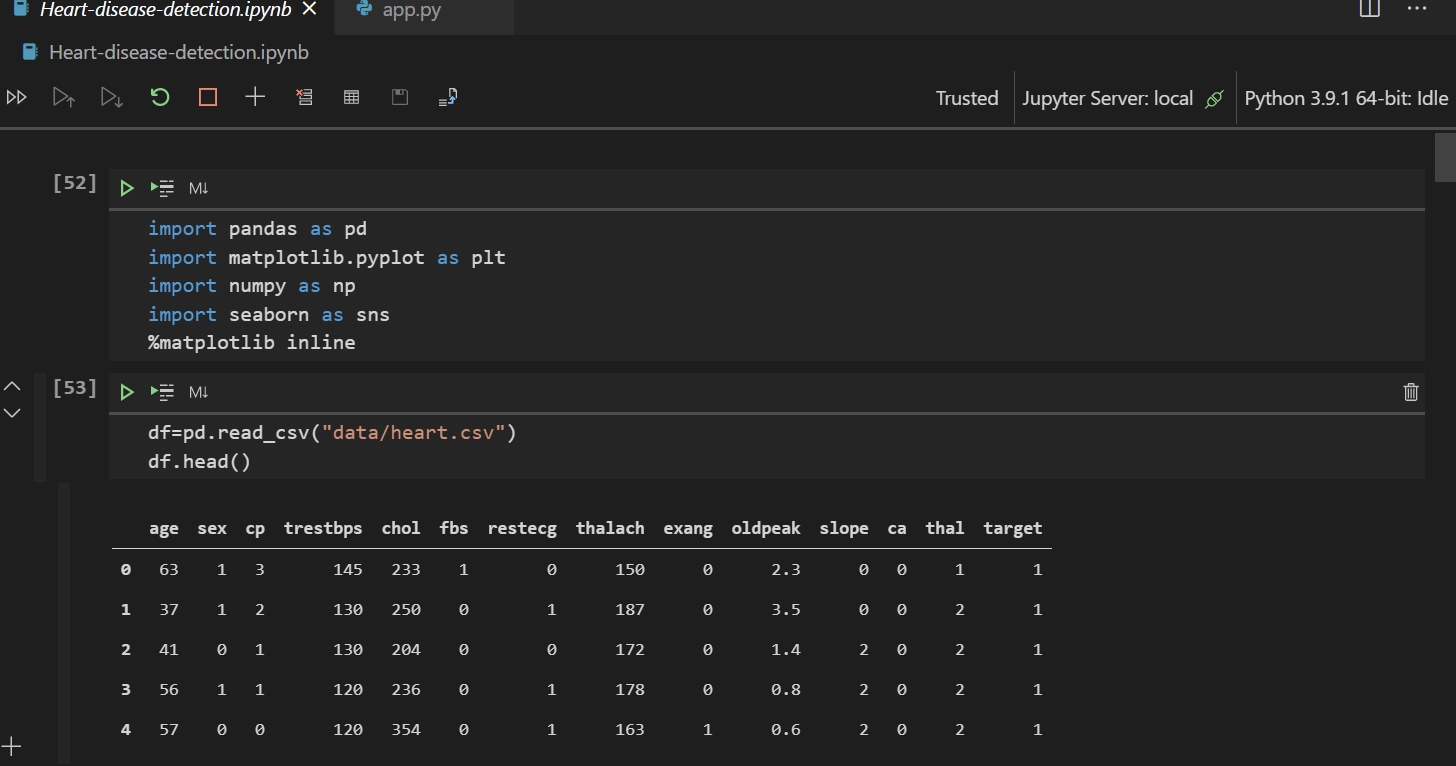
Note: Our data has 3 types of data:

**Continuous (#)**: which is quantitative data that can be measured

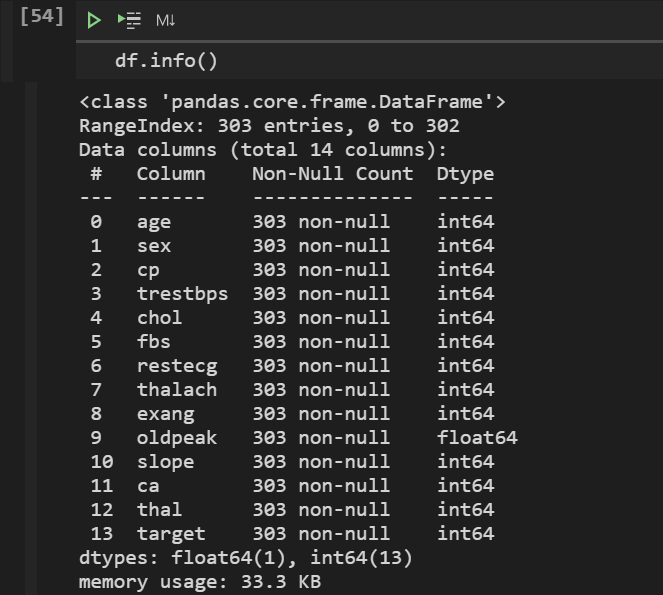
**Binary Data**: data whose unit can take on only two possible states ( 0 &1 )

**Ordinal Data**: Categorical data that has a order to it (0,1,2,3, etc)

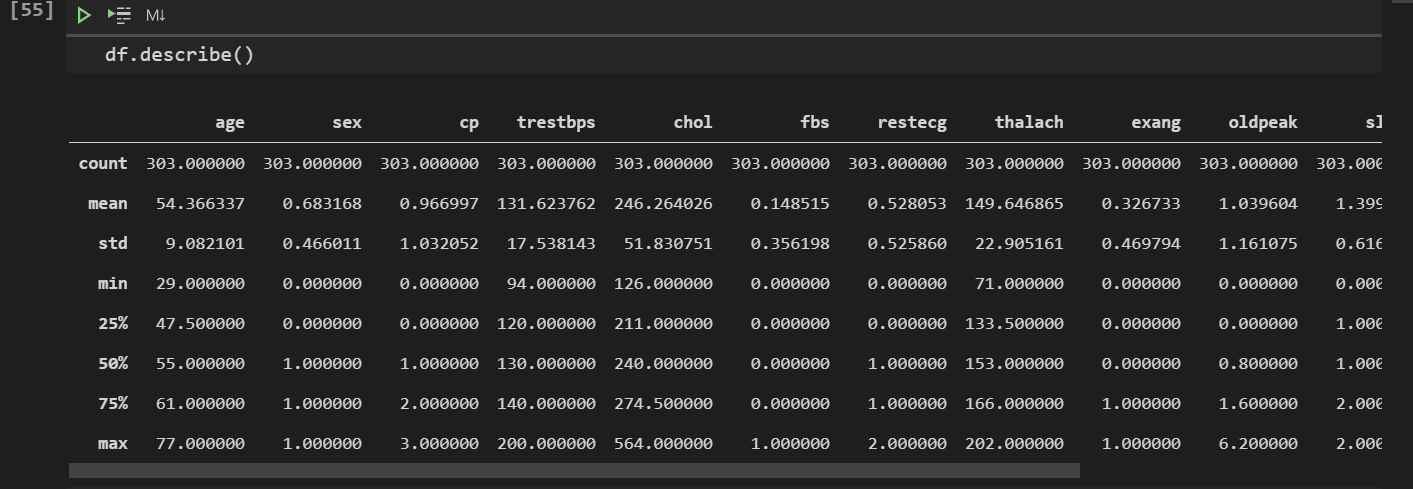
**Data Wrangling**



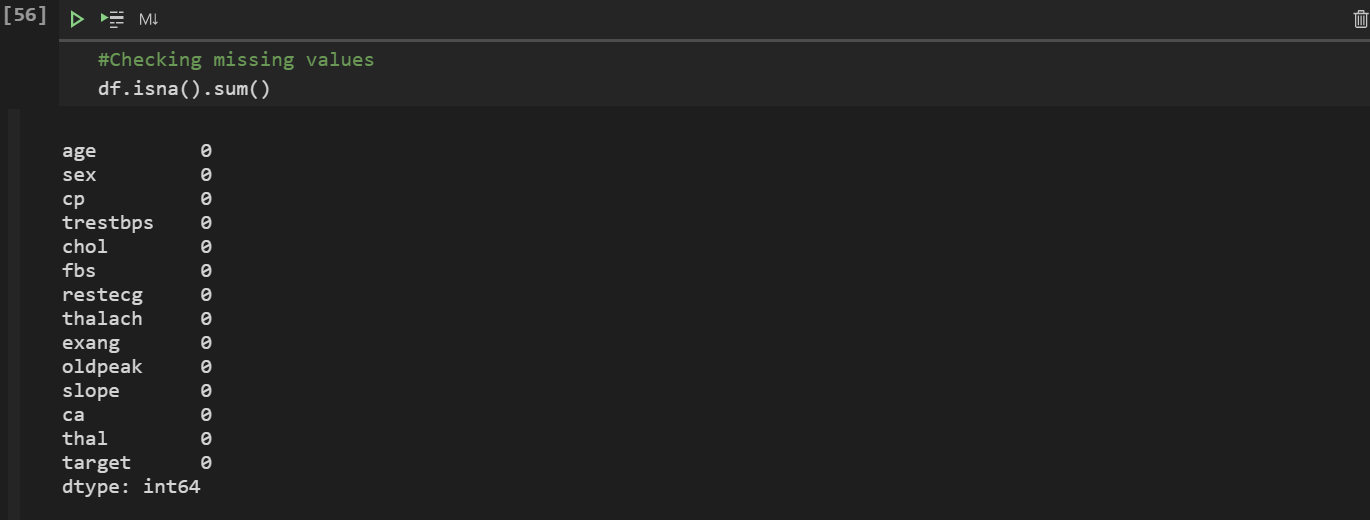
Helps us get an idea of what data we are working with.



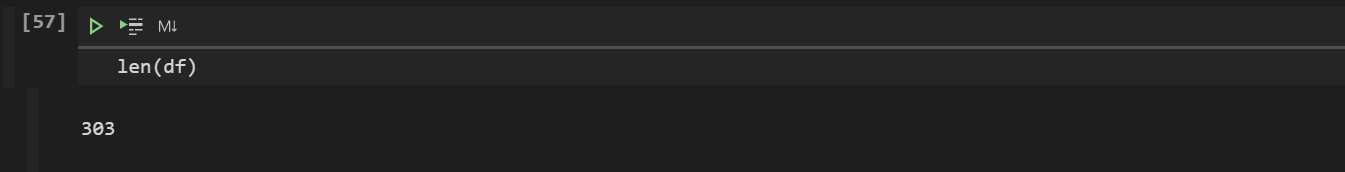
Display the Datatype ,DataFrame, RangeIndex, columns ,memory usage



**summarizes the count, mean, standard deviation, min, and max for numeric variables.**

****

**Display the Number of Missing Values for each column. We luckily have none.**

****

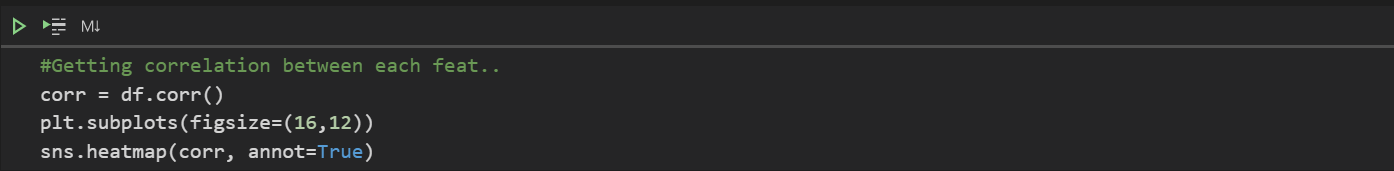
**Display the length of DataFrame**

**Exploratory Data Analysis**

# Correlations

**Correlation Matrix**- let’s you see **correlations** between all variables.

Within seconds, you can see whether something is positively or negatively correlated with our **predictor (target)**.

**



We can see there is a **positive correlation** between chest pain (cp) & target (our predictor). This makes sense since, the greater amount of chest pain results in a greater chance of having heart disease. Cp (chest pain), is a ordinal feature with 4 values:

Value 1: typical angina ,

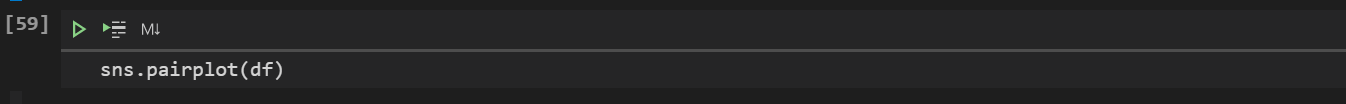
Value 2: atypical angina,

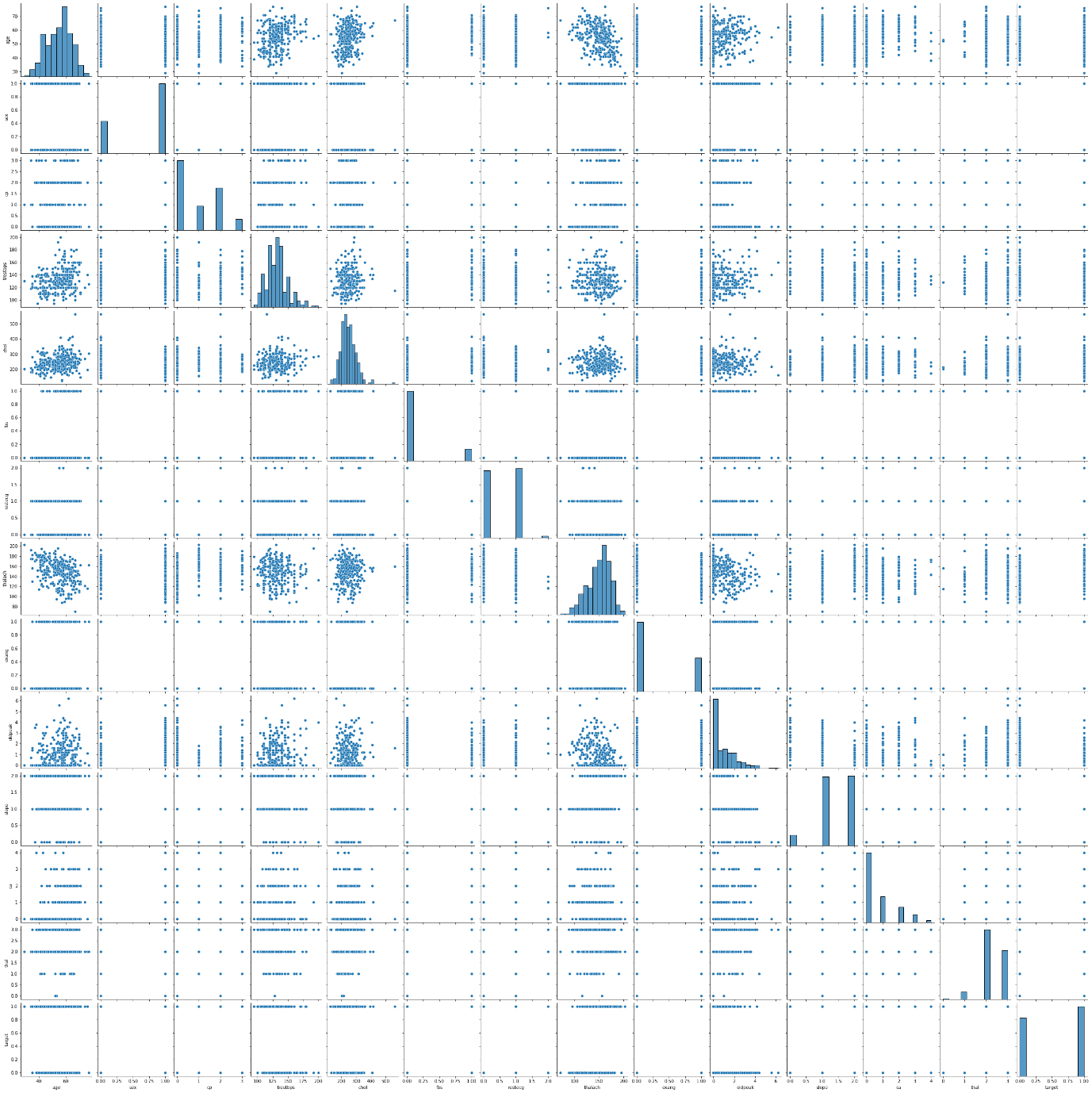
Value 3: non-anginal pain ,

Value 4: asymptomatic.

In addition, we see a **negative correlation** between exercise induced angina (exang) & our predictor. This makes sense because when you excercise, your heart requires more blood*,* but narrowed arteries slow down blood flow*.*

Pairplots are also a great way to immediately see the correlations between all variables. But you will see me make it with only continuous columns from our data, because with so many features, it can be difficult to see each one. So instead I will make a pairplot with only our continuous features.





Chose to make a smaller pairplot with **only** the **continuous** variables, to dive deeper into the relationships. Also a great way to see if theirs a positive or negative **correlation**!



**It is display the 0 and 1**