Detailed Project Report (DPR)
Heart Disease Diagnostic Analysis

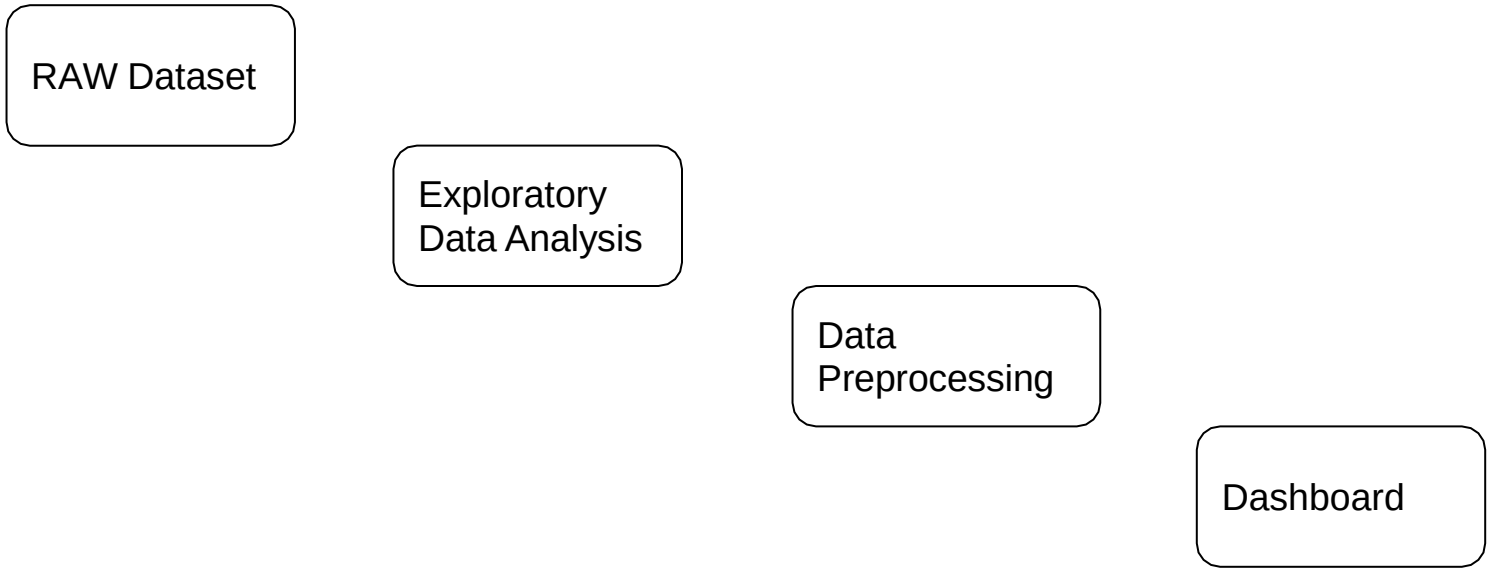
Document Version Control

Date Issue	Version	Description	Author
10/04/2023	1	Initial DPR - V 1.0	Prateek Kumar
30/05/2023	2	Initial DPR - V 2.0	Prateek Kumar

Objective

The objective of the project is to build an dashboard that will help to understand the heart disease data and its pattern.

Architecture



About the Project

Health is real wealth in the pandemic time we all realized the brute effects of covid-19 on all irrespective of any status. You are required to analyze this health and medical data for better future preparation.

This dataset contains details of the age, BP, cholesterol and about having heart disease or not.Using the attributes in the dataset we can predict heart disease risk for an individual and identify the risk factors.

Dataset Information

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

Attribute Information

- age: The person's age in years
- sex: The person's sex (1 = male, 0 = female)
- cp: The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
- trestbps: The person's resting blood pressure (mm Hg on admission to the hospital)
- chol: The person's cholesterol measurement in mg/dl
- fbs: The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)

Attribute Information

- restecg: Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
- thalach: The person's maximum heart rate achieved
- exang: Exercise induced angina (1 = yes; 0 = no)
- oldpeak: ST depression induced by exercise relative to rest
- slope: the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
- ca: The number of major vessels (0-3)
- thal: A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
- num: Heart disease (0 = no, 1 = yes)

Dataset Information

Age : Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. It is estimated that 82 percent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.

Sex : Men are at greater risk of heart disease than premenopausal women. Once past menopause, it has been argued that a woman's risk is similar to a man's although more recent data from the WHO and UN disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.

Resting Blood Pressure : Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.

Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of heart attack.

Dataset Information

Cholesterol: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of heart attack.

Resting ECG: For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercise ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.

Max heart rate achieved: The increase in the cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.

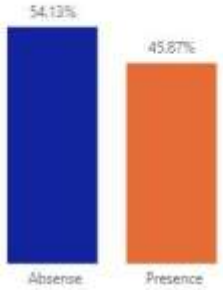
ST Depression: In unstable coronary artery disease, ST-segment depression is associated with a 100% increase in the occurrence of three-vessel/left main disease and to an increased risk of subsequent cardiac events. In these patients an early invasive strategy substantially decreases death/myocardial infarction.

Absense Presence

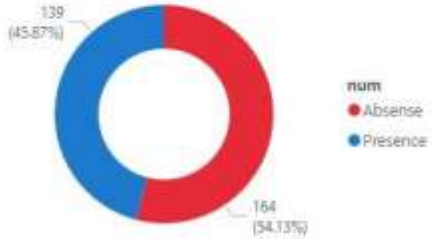
HEART DISEASE DIAGNOSTIC DASHBOARD

Female Male

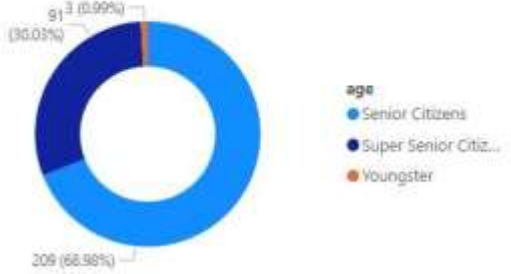
Heart Disease Gender Count



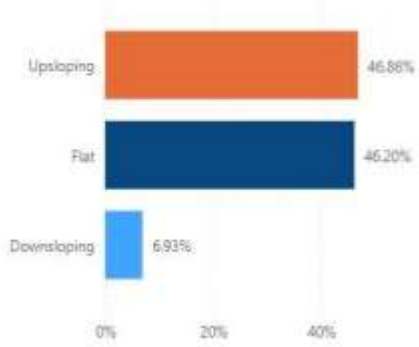
Exercise Induced Angina



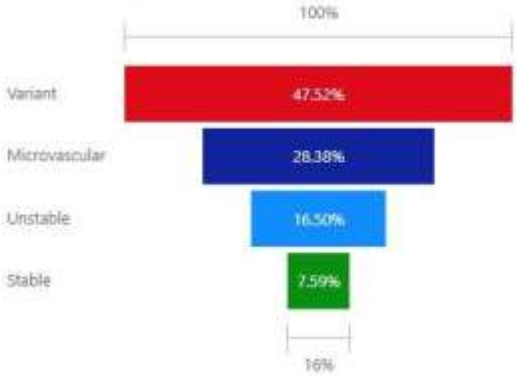
Chest Pain vs Age



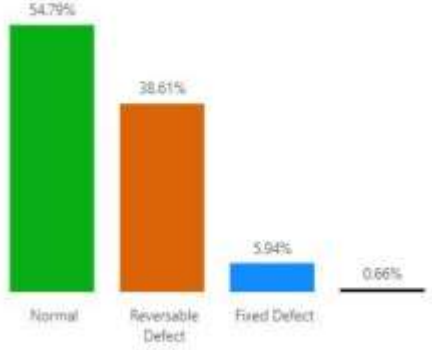
SLOPE : Normal trace during exercise



Chest Pain vs Gender



A Blood Disorder called Thalassemia



Q & A

Q1) What is the size of your data ?

The size of the data in terms of KB is 38.0, with with 303 rows and 14 columns names are encrypted for security reasons

Q2) What are the data type ?

The columns consisted of both integer and float values.

Q & A

Q3) What were the libraries that you used Python ?

I used Pandas, Matplotlib, Seaborn, Plotly and Streamlit.

Q4) What's the complete flow you followed in this Project ?

We had a full-fledged data pipeline which would extract the data from the S3 bucket, which had come from Big Data pipeline created by the client side.

Q & A

Q5) Where did you get the data ?

The data was provided by iNeuron.

https://drive.google.com/drive/folders/165Pjmfb9W9PGy0rZjHEA22LW0Lt3Y-Q8

THANK YOU