HEART DISEASE DIAGNOSTIC-ANALYSIS

DETAILED PROJECT REPORT

BY- TANUSHREE POOJARY

PROJECT DETAIL

Project Title	Heart Disease Diagnostic – Analysis
Technology	Business Intelligence
Domain	Healthcare
Project Difficulty level	Intermediate
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel, Tableau

OBJECTIVE

- The goal of this project is to analyse the heart disease occurrence, based on a combination of features that describes the heart disease.
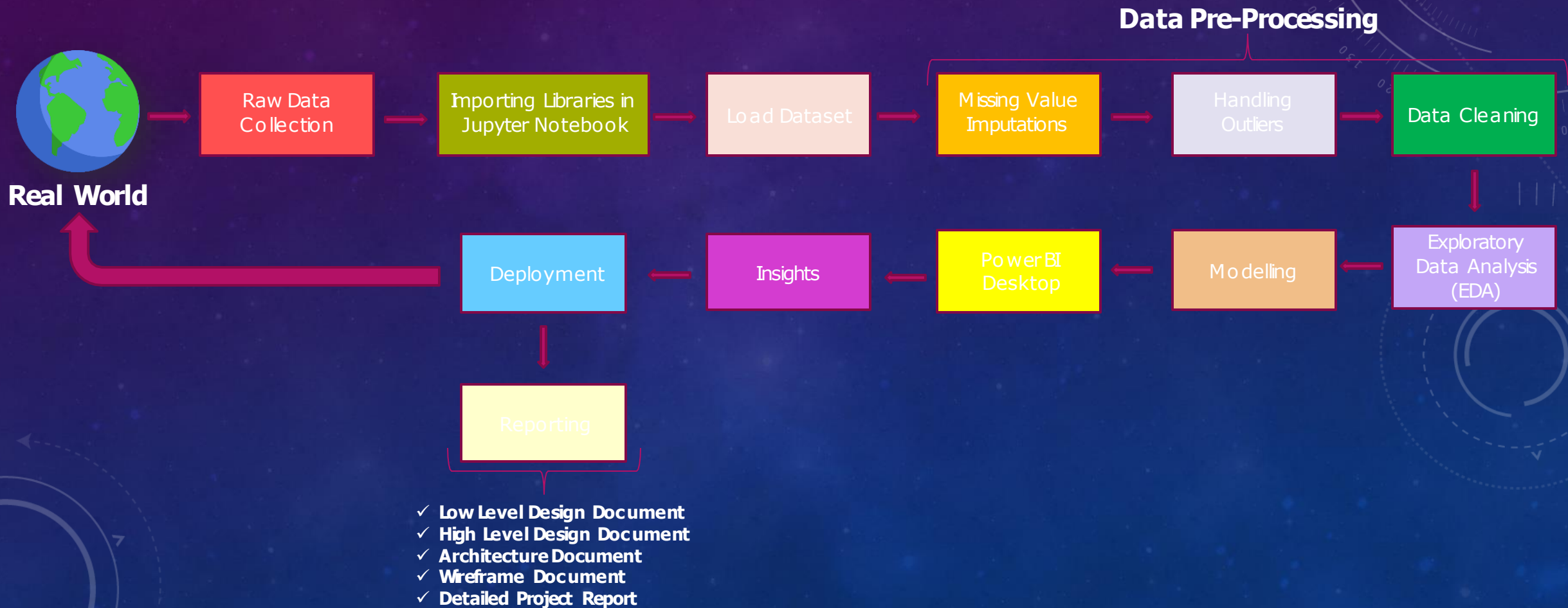
PROBLEM STATEMENT

► 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 analyse this health and medical data for better future preparation.

► A dataset is formed by taking into consideration some of the information of 303 individuals.

DATASET 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)

- 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: down sloping)

- 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)

ARCHITECTURE

Data Pre-Processing

Real World → Raw Data Collection → Importing Libraries in Jupyter Notebook → Load Dataset → Missing Value Imputations → Handling Outliers → Data Cleaning

Data Cleaning → Exploratory Data Analysis (EDA) → Modelling → Power BI Desktop → Insights → Deployment → **Real World**

Deployment → Reporting

- ✓ **Low Level Design Document**
- ✓ **High Level Design Document**
- ✓ **Architecture Document**
- ✓ **Wireframe Document**
- ✓ **Detailed Project Report**

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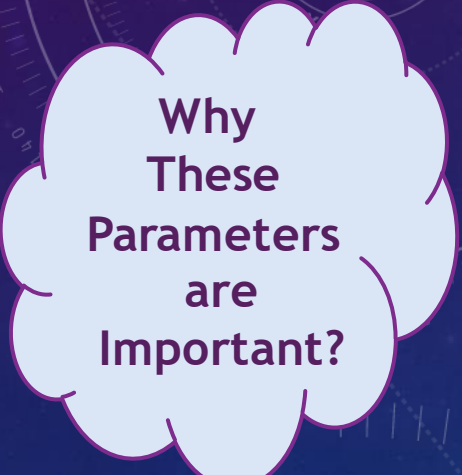
Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of heart attack.

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.

Why These Parameters are Important?

STEPS FOLLOWED

- **Data Extraction**: This step involves extracting the data from different sources relevant to the problem statement or obtaining data from the client.

- **Data Pre-processing**: Once the raw data is obtained, we need to ensure that the data is free from errors. We perform Exploratory Data Analysis followed by Data Cleaning which involves imputing missing values, removing duplicates, finding anomalies or outliers, and treating them.

- **Data Exporting**: The preprocessed data is exported to a .csv file to be used for

- analysis.

- **Data Loading and Modification**: The preprocessed data in the .csv file is loaded into the Tableau Desktop for analysis purposes and modified for simplicity purposes.

- **Data Analysis**: Once the data is loaded, we perform the data analysis using Tableau features and store the visualizations in Tableau worksheets.

- **Deployment**: The prepared visualizations are deployed on the Tableau Online Software where they will be available publicly

1. DATA EXTRACTION

- The dataset used for analysis is the Heart Disease dataset . It actually contains 76 attributes out of which only 14 are used. We will be using the Cleveland dataset.

- Dataset source: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

- The dataset is available is a .csv file - 'heart_disease_dataset.csv

2. DATA PREPROCESSING

- After Exploratory Data Analysis carried out on the dataset we have certain observations with the dataset.

- There is no column in the dataset with missing values.There are a few columns which actually contain categorical values but have been incorrectly labeled as numeric. As a part of data preprocessing we will convert them to categorical values and there are a few columns which have unusual values / outliers. We will impute these values with the median / mode value obtained from the remaining values of the columns.

3. DATA EXPORTING

- Once the data has been cleaned in the data preprocessing stage, we will export the cleaned dataset into a new file with .csv format.

- The new dataset file has name - 'preprocessed_heart_disease_dataset.csv

4. DATASET LOADING AND MODIFICATION

- The exported .csv dataset file – 'preprocessed_heart_disease_dataset.csv' will be imported into Tableau Public Desktop. Since this a .csv file, we will choose the 'Microsoft Excel' file option when prompted to import dataset into Tableau.

- Since the dataset contains many categorical columns which store the categories in the form of integers we will convert these numbers into meaningful phrases which will be understandable to the viewer and also easy to understand the terms used in the visualizations

5. DATA ANALYSIS

- Once the data has been loaded into the Tableau Desktop software, we perform the analysis for the various medical parameters provided in the dataset and study the relationship between them.

- Based on these patterns, we try to draw approximate inferences about the data

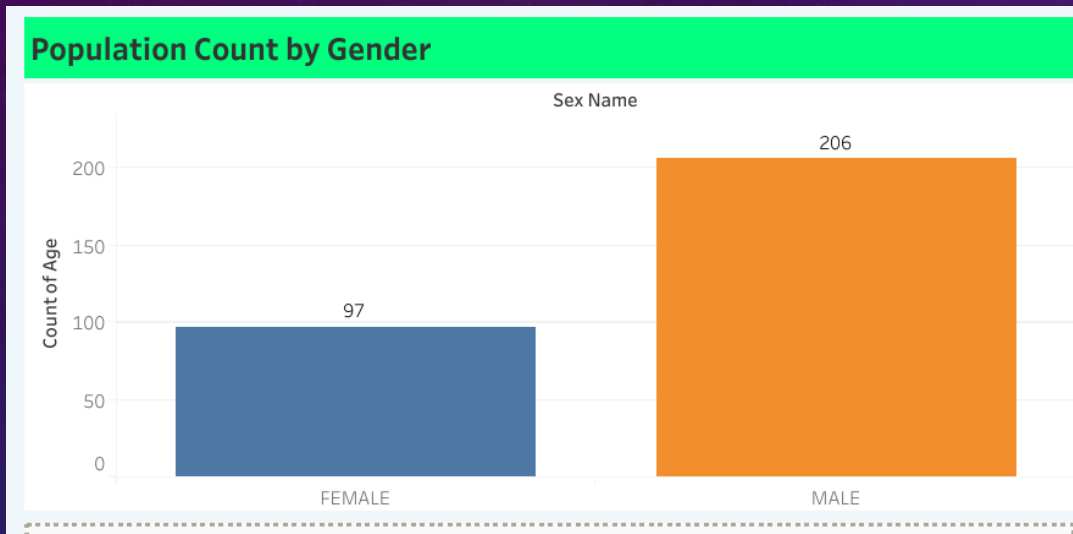- provided on the basis of visualizations created

6. DEPLOYMENT

- All the different worksheets that have been created are compiled together into a tableau workbook. Each worksheet is named based on the type of visualization performed in the chart.

- When we save all the worksheets on tableau software, it connects to the tableau public software via personal email id and credentials. The dashboard is uploaded onto the tableau public software on personal profile and this is visible to public.

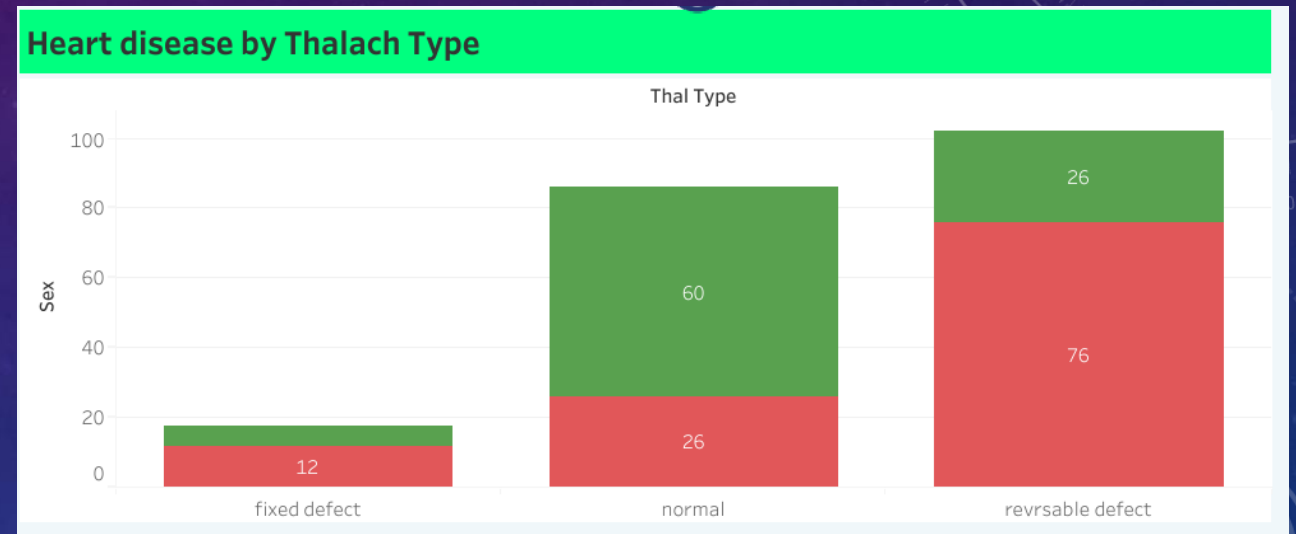- The link for the worksheets is at :

https://public.tableau.com/app/profile/tanushreepoojary/viz/HeartDiseaseDiagnosticAnalysisproject/HeartDiseaseDiagnosticAnalysis?publish=yes

INSIGHT

Population Count by Gender
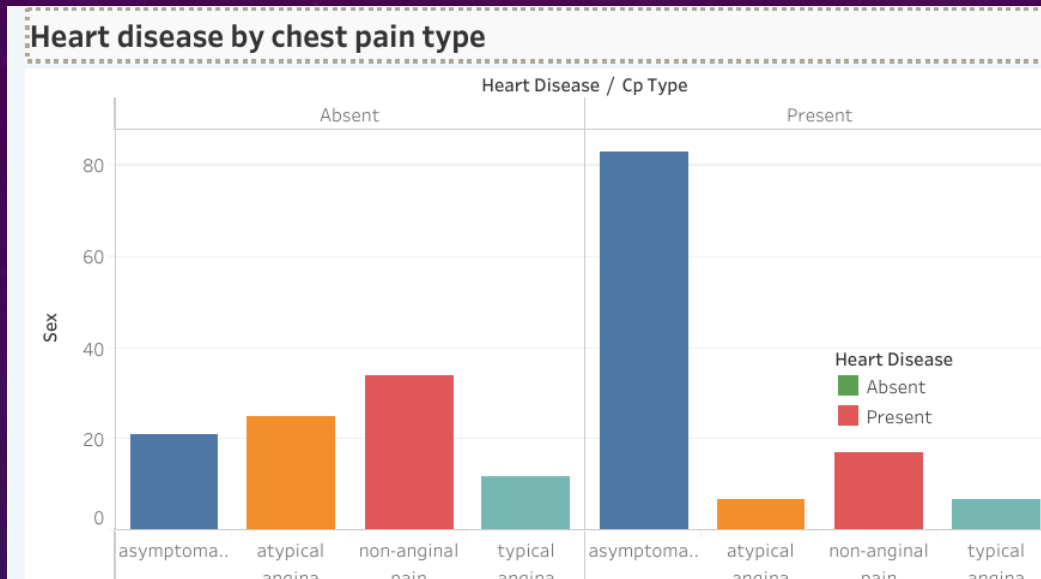

Heart disease by Thalach Type

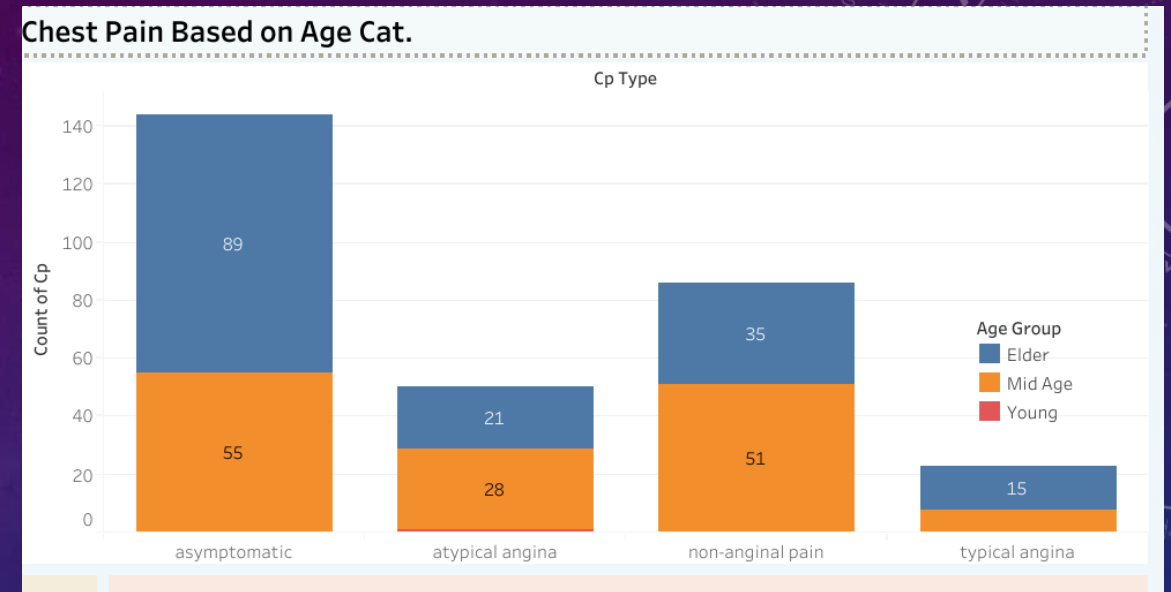- Majority of the population contributes to Male over number of Women

- We observe that the Normal type of Thalassemia is common in both Male and Female and is approximately equal chances to contract the disorder.

- As opposed to the Normal type, the Fixed and Reversible types of Thalassemia are more likely to be found in the Male sex. The proportion of the Males having these disorders is much greater than the Females.
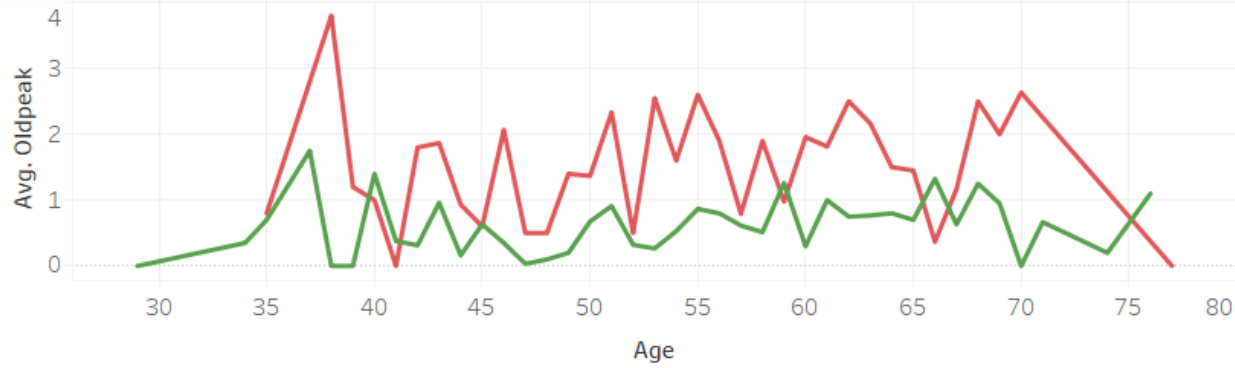
Heart disease by chest pain type

There is very high number of Asymptomatic Pain in Elderly age Category

Chest Pain Based on Age Cat.

It seems people having asymptomatic chest pain have a higher chance of heart disease.

Asymptomatic Chest pain means neither causing nor exhibiting symptoms of heart disease.

ST depression vs Age

we can observe from here that ST depression mostly increases between the age group of 30-40.

ST depression refers to a finding on an electrocardiogram, wherein the trace in the ST segment is abnormally low below the baseline.

Cholestrol vs Sex name

Sex Name

FEMALE — 261.75

MALE — 239.60

Here, we can observe that Cholesterol and maximum heart rate In creasing in the age group of 50-60.

KEY PERFORMANCE INDICATOR (KPI)

1. Percentage of People Having Heart Disease

2. Age Distribution including Gender

3. Gender Distribution Based on Heart Disease

4. Chest Pain Experienced by People Suffering from Heart Disease

5. Blood Pressure, Cholesterol Level and Maximum Heart Rate of People According to their Age and Heart Disease Patients.

6. ST Depression Experienced by People According to their age and heart disease.

THANKYOU