WIREFRAME DOCUMENT

Heart Disease Diagnostic Analysis

BY
VIKASH KUMAR MAHAPATRA

Revision Number & Date :- 1.3 02/09/2022

Document Version Control

Date Issued	Version	Description	Author
05 AUGUST 2022	1.0	Created Document And Add Document Version Control	Vikash
25 AUGUST 2022	1.1	Created Content Page ,Introduction Page & Dataset Information Page	Vikash
01 SEPT 2022	1.2	Created Exploratory Analysis, Power-Bi dashboard, Add Page Number in Content Section	Vikash
02 SEPT 2022	1.3	Final Revision	Vikash

CONTENTS

ABSTRACT

CVD(Cardiovascular Disease) are group of disorders of heart and blood vessels. Heart disease is a general term that means the heart isn't working normally. Babies can be born with heart disease. This is called congenital heart disease. If people get heart disease later, it is called acquired heart disease and most of them are acquired. The most common types of acquired heart disease are CAD(Coronary Artery Disease), CHF(Congestive Heart Failure), Bad Heart Rhythms. India has one of the highest burdens of CVD worldwide. According to a report the annual numbers of deaths from CVD in India rise from 2.26million(1990) to 4.77million(2020).Coronary heart disease prevalence rate in India have been estimated over the past several decades & have ranged from 1.6% to 7.4% in rural populations and from 1% to 13.2% in urban population. The prevalence rate increased with age from 22% in 45-54 to 38% in age 70 and above as per report.

INTRODUCTION

WHY THIS WIREFRAME DOCUMENT

This document represent the complex tasks involved in a project in an easy-to-understand visual format. By these we can also communicate any changes to all stakeholders quickly and efficiently. This also assist in keeping project on track and help to reduce misunderstanding between stakeholders.

PROBLEM STATEMENT

From earlier times we all heard health is wealth, in the pandemic scenario everybody got a better of understanding that, we also see harsh effects during this Covid pandemic scenario whether you are healthy or not everyone got affected even children also. We are required to analyze the medical data for better future preparation to avoid risk. Find key metrics and factors and show the meaningful relationships attributes.

DATASET INFORMATION

The file name heart_disease_dataset.csv contains different medical parameters data of the selective given peoples. Sample data below.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
48	1	2	110	229	0	0	168	0	1	3	0	7	1
54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0
58	1	2	120	284	0	2	160	0	1.8	2	0	3	1

- Age: Person's age in years.
- Sex: Gender of a person (1->Male 0->Female)
- Cp: Donates Chest pain Experienced which are categorized as(1-> Typical angina , 2->Atypical angina , 3-> Non-angina pain , 4-> Asymptomatic)
- Trestbps : Resting blood pressure (mm hg on admission to hospital)
- Chol: Cholesterol level of a person measure in mg/dl.
- Fbs: Fasting Blood sugar level of a person (>120 mg/dl then denotes as 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 : Maximum heart rate achieved by a person
- Exang: Exercised induced angina (1->Yes ,0-> No)

- Oldpeak: ST depression induced by exercise relative to rest
- Slope: the slope of the peak exercise ST segment(1-> unsloping , 2-> flat , 3-> downsloping)
- Ca : The number of major vessels(0-3)
- Thal : A blood disorder known as thalassemia categorized as(3->normal ,6-> fixed defect ,7-> reversible defect)
- Num : Heart disease(0->NotHave-Disease 1->Have-Disease)

EXPLORATORY ANALYSIS

```
#Extracting CSV Data From System using Pandas Library
Heart_Df = pd.read_csv('heart_disease_dataset.csv')
```

```
# Shape of dataset i.e. no of rows and columns
Heart_Df.shape
```

```
(303, 14)
```

In the above figure we have load dataset read the data with help of pandas. We created a data frame name as Heart_Df.Also got shape of the dataset i.e. 303 * 14.

```
#Statistics about the numerical columns in a dataset
Heart_Df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
age	303.0	54.438944	9.038662	29.0	48.0	56.0	61.0	77.0
sex	303.0	0.679868	0.467299	0.0	0.0	1.0	1.0	1.0
cp	303.0	3.158416	0.960126	1.0	3.0	3.0	4.0	4.0
trestbps	303.0	131.689769	17.599748	94.0	120.0	130.0	140.0	200.0
chol	303.0	246.693069	51.776918	126.0	211.0	241.0	275.0	564.0
fbs	303.0	0.148515	0.356198	0.0	0.0	0.0	0.0	1.0
restecg	303.0	0.990099	0.994971	0.0	0.0	1.0	2.0	2.0
thalach	303.0	149.607261	22.875003	71.0	133.5	153.0	166.0	202.0
exang	303.0	0.326733	0.469794	0.0	0.0	0.0	1.0	1.0
oldpeak	303.0	1.039604	1.161075	0.0	0.0	0.8	1.6	6.2
slope	303.0	1.600660	0.616226	1.0	1.0	2.0	2.0	3.0
ca	303.0	-1319.468647	11432.566205	-100000.0	0.0	0.0	1.0	3.0
thal	303.0	-655.363036	8111.366638	-100000.0	3.0	3.0	7.0	7.0
num	303.0	0.458746	0.499120	0.0	0.0	0.0	1.0	1.0

```
#Data Types of Columns
Heart_Df.dtypes
```

```
age           int64
sex           int64
cp            int64
trestbps      int64
chol          int64
fbs           int64
restecg       int64
thalach       int64
exang         int64
oldpeak     float64
slope         int64
ca            int64
thal          int64
num           int64
dtype: object
```

In the above figure we read some statistical value of the dataset as well as datatypes present in the dataset. We have use of Heart_Df.describe() for statistical value and Heart_Df.dtypes for datatypes.

```
# Finding out the null variables
Heart_Df.isna().sum()
```

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
num         0
dtype: int64
```

In the above figure we find out if there any null values present in the dataset or not. We have use of Heart_Df.isna.sum() check null value.

```
#Making copy of original data
Heart_Anlys_Df = Heart_Df.copy()
```

In the above figure we have copied the dataset to new dataframe that is Heart_Analys_Df to make some necessary changes without messing around with original dataset.

```python
#Function for Disease have or not
def modify_heartDis(col_name):
    if col_name == 1:
        return "Have-Disease"
    elif col_name == 0:
        return "NotHave-Disease"
```

```python
#Function for Sex Male and Female
def modify_genderCode(sc_name):
    if sc_name == 0:
        return "Female"
    elif sc_name ==1:
        return "Male"
```

```python
#Function for AgeGroup Young Adult upto(18to35),Middle-Aged(36to54) and Senior(55 above)
def create_agesection(data_select):
    if data_select>=18 and data_select<=35:
        return "Young-Adult"
    elif data_select>=36 and data_select<=54:
        return "Middle-Aged"
    elif data_select>=55:
        return "Senior"
```

From previous page image we have created some functions like modify_heartDis to add have-disease in place of 1 and not-have disease in place of 0 into the dataset likewise modify_genderCode to give Female in place of 0 and Male in place of 1 another function create_agesection for segeration of age group in the dataset.

```python
#Applying converted data into our dataset with new column - Heart Disease
Heart_Anlys_Df['Heart-Disease']=Heart_Anlys_Df['num'].apply(modify_heartDis)
```

```python
#Applying converted data into our dataset with new column - Gender
Heart_Anlys_Df['Gender']=Heart_Anlys_Df['sex'].apply(modify_genderCode)
```

```python
#Applying converted data into our dataset with new column - Age-Group
Heart_Anlys_Df['Age-Group']=Heart_Anlys_Df['age'].apply(create_agesection)
```

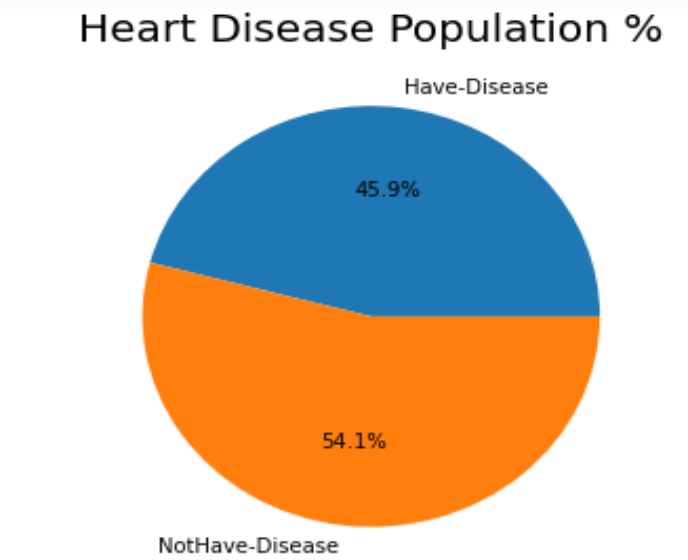
In the above figure we apply the function values to new dataset with new columns name such as Heart-Disease, Gender and Age-Group.

```
In [16]:
1 #Check new made columns
2 Heart_Anlys_Df.head(10)
```

Out[16]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num	Heart-Disease	Gender	Age-Group
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0	NotHave-Disease	Male	Senior
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1	Have-Disease	Male	Senior
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1	Have-Disease	Male	Senior
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0	NotHave-Disease	Male	Middle-Aged
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0	NotHave-Disease	Female	Middle-Aged
5	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0	NotHave-Disease	Male	Senior
6	62	0	4	140	268	0	2	160	0	3.6	3	2	3	1	Have-Disease	Female	Senior
7	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0	NotHave-Disease	Female	Senior
8	63	1	4	130	254	0	2	147	0	1.4	2	1	7	1	Have-Disease	Male	Senior
9	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1	Have-Disease	Male	Middle-Aged

In the above figure we read the top ten records after modifying it with new dataframe that is Heart_Anlys_Df.



From above figure we can see that more half i.e. around 54.13 %(53.1% approx.) From given People data from csv don’t have heart disease while around 45.87 %(45.9% aprrox.) People have heart disease.

```
# No of people have Disease
No_of_People=Heart_Anlys_Df.groupby('Heart-Disease')['num'].count()
No_of_People
```

```
Heart-Disease
Have-Disease       139
NotHave-Disease    164
Name: num, dtype: int64
```

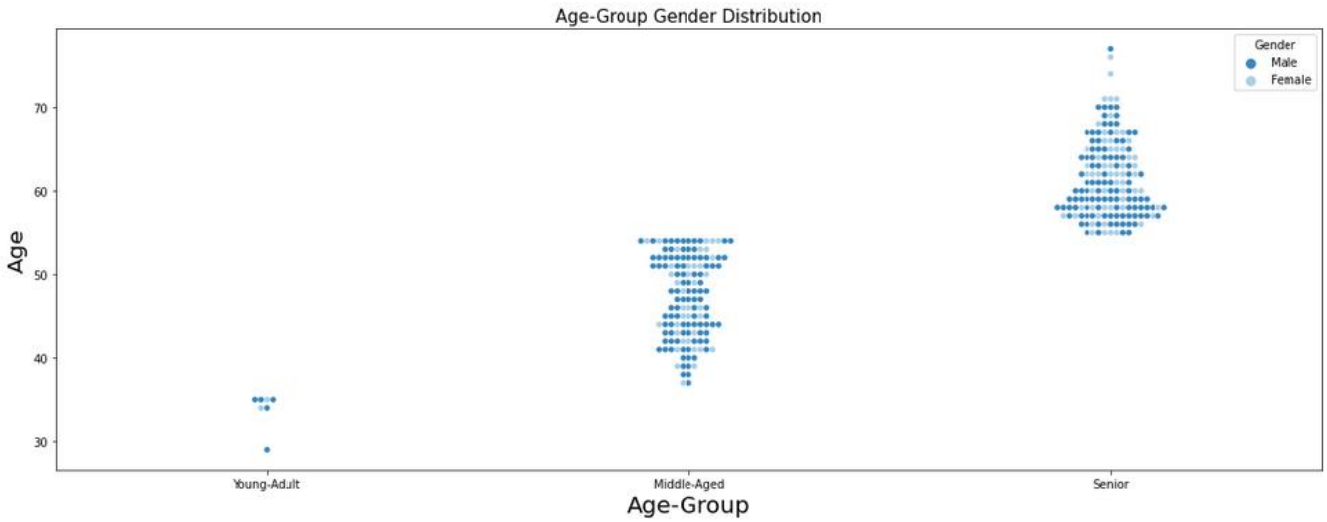
```
# Gender Distribution in Dataset
Gender_Dis=Heart_Anlys_Df.groupby('Gender')['sex'].count()
Gender_Dis
```

```
Gender
Female     97
Male      206
Name: sex, dtype: int64
```

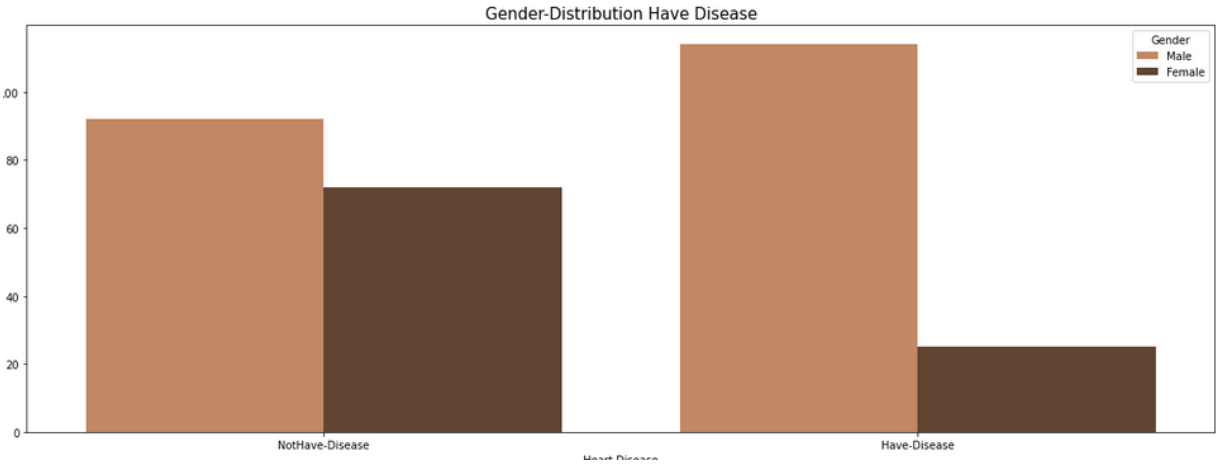
```
# Age Distribution in Dataset
Age_Sec=Heart_Anlys_Df.groupby('Age-Group')['age'].count()
Age_Sec
```

```
Age-Group
Middle-Aged    136
Senior         160
Young-Adult      7
Name: age, dtype: int64
```

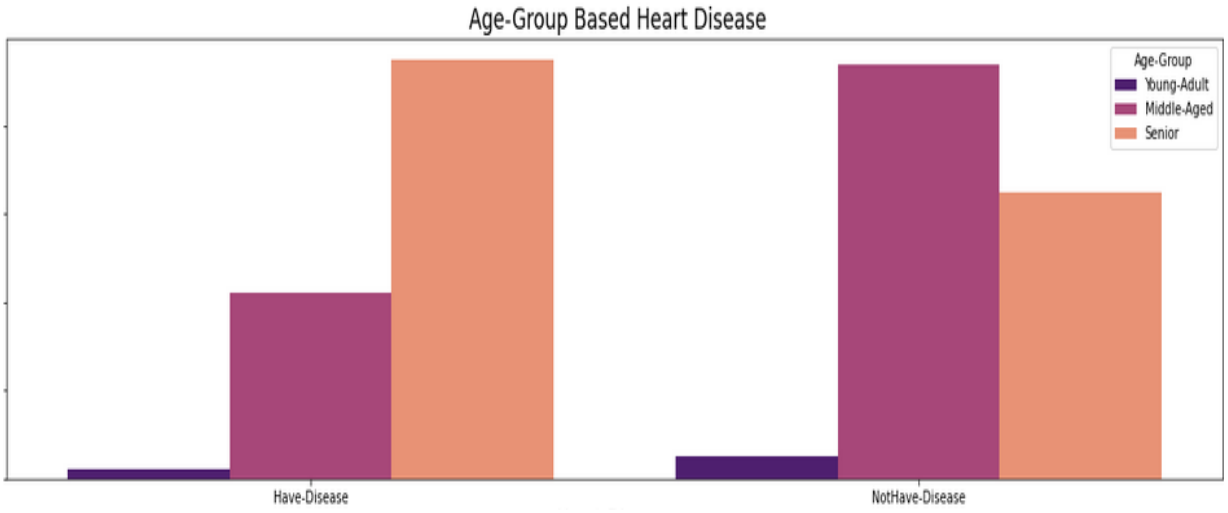
In above figure shows complete compilation of certain data like out of 303 records 139 people have disease and 164 people doesn't have , 206 people are male while 97 people are female. And in age group Young-Adult count is 7, Middle-Aged count is 136 and most people are in Senior where count is 160.



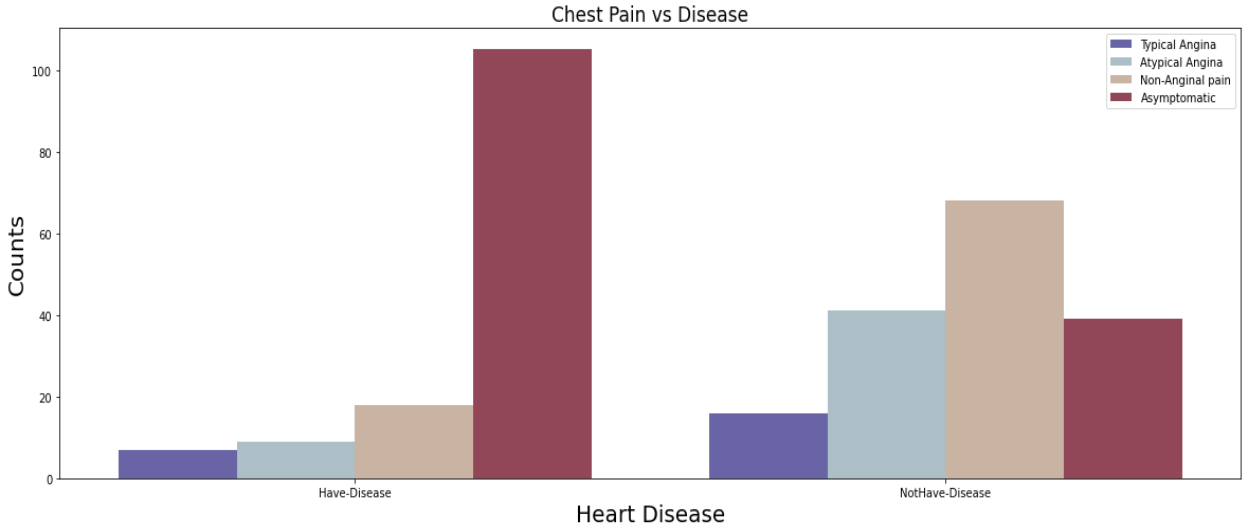
From the above figure people are mostly from senior age group constitutes the most then comes the middle age group and there are less no young people in dataset. Also ratio of male is higher in every age-group over female.



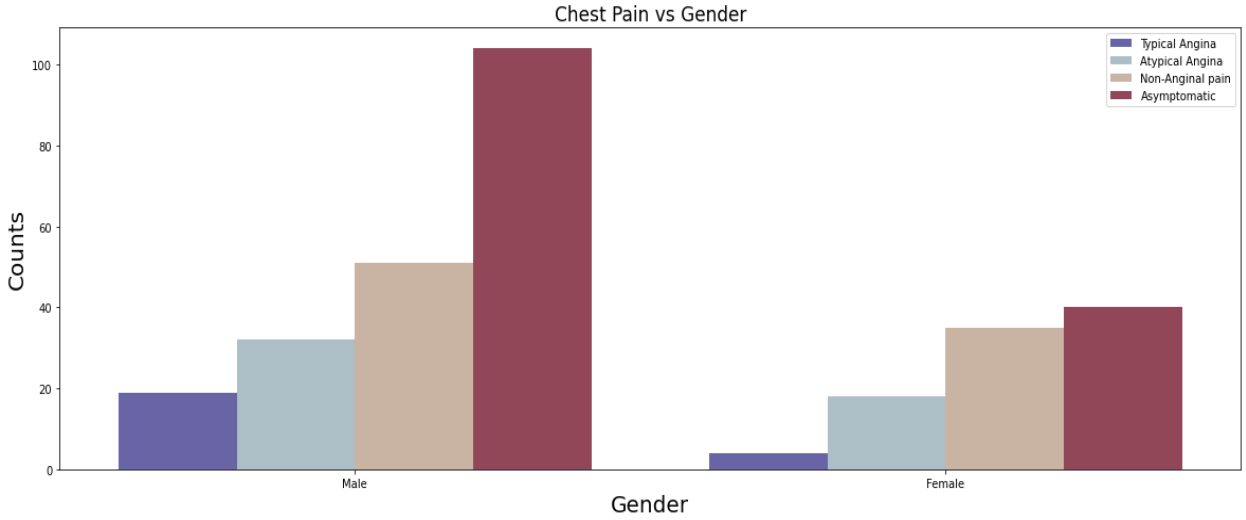
In above figure showing max number of male people are having heart disease in compare to female people. Also number male is max in not having disease in compare to female



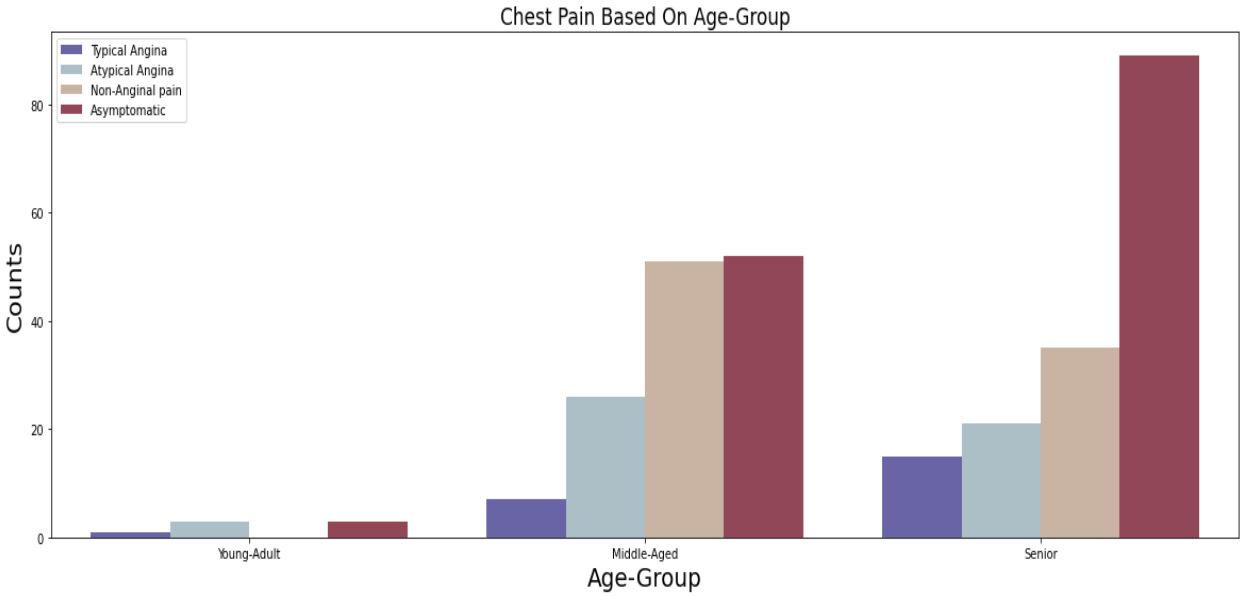
From above figure we can see that people from age group senior are mostly have heart disease then comes the middle age and least is found in the young-adult. Whereas in not having disease case max are middle age people then comes senior age and least in young age.



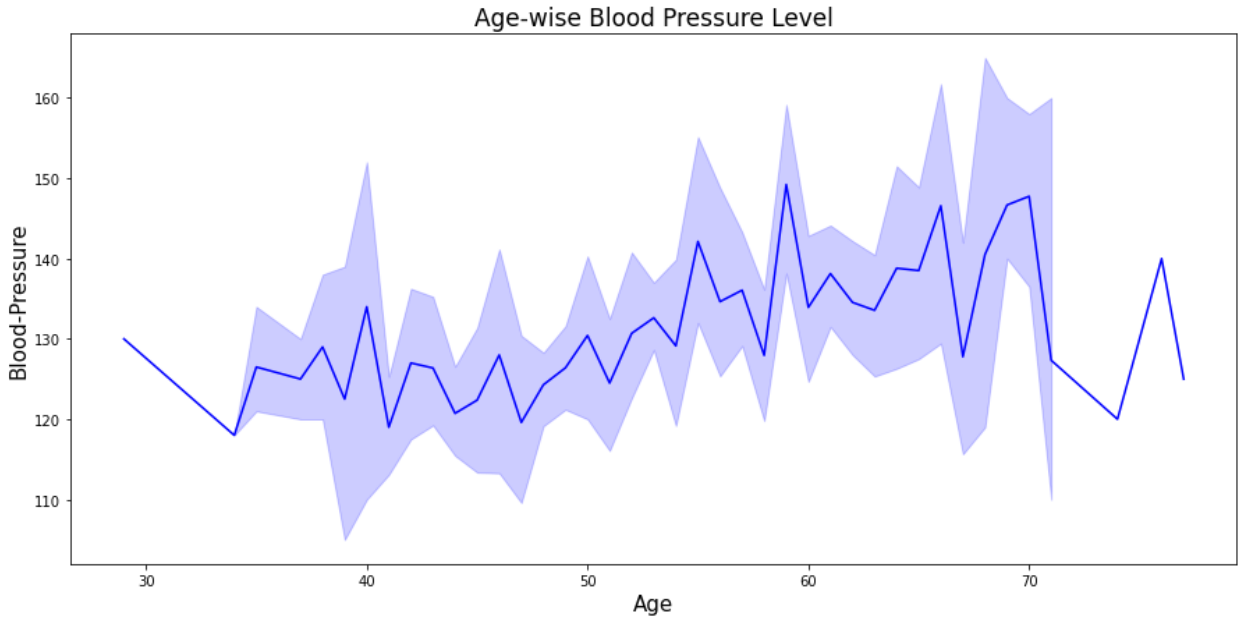
From above figure we can see that out of four categorized levels of chest pain experienced by people the asymptomatic (silent heart attack) chest pain is most deadly in heart disease in comparisons to other chest pain.



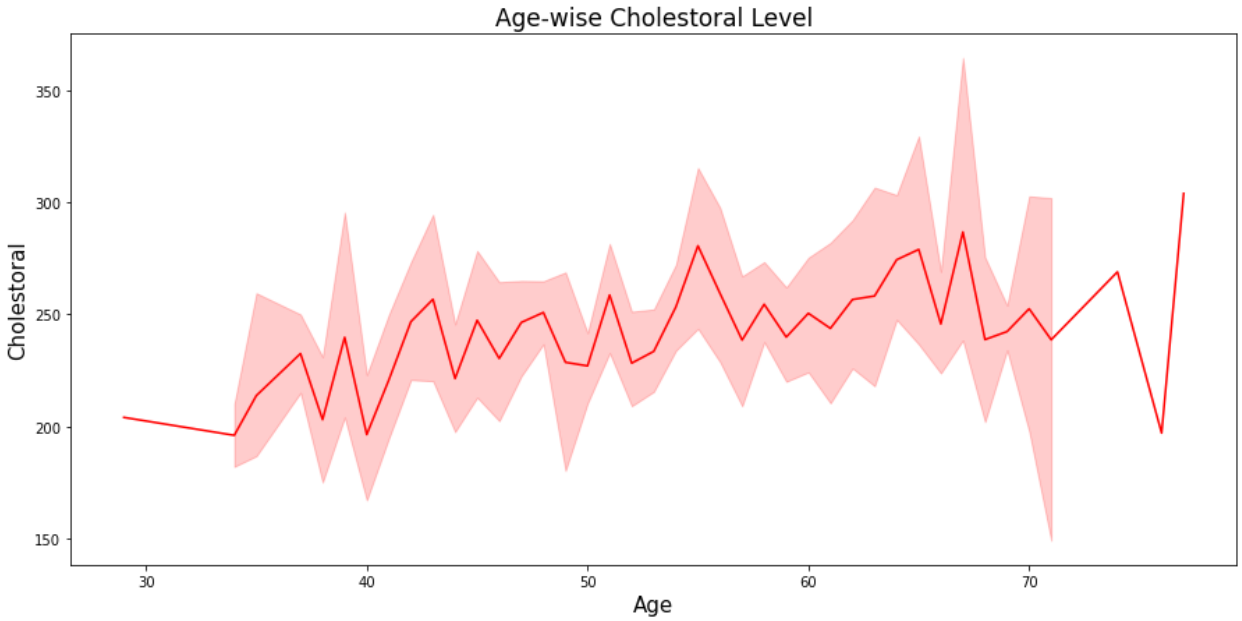
From above figure we can see that majorly male peoples face the asymptomatic (silent heart attack) chest pain which is most deadly in heart disease in comparisons to female peoples.



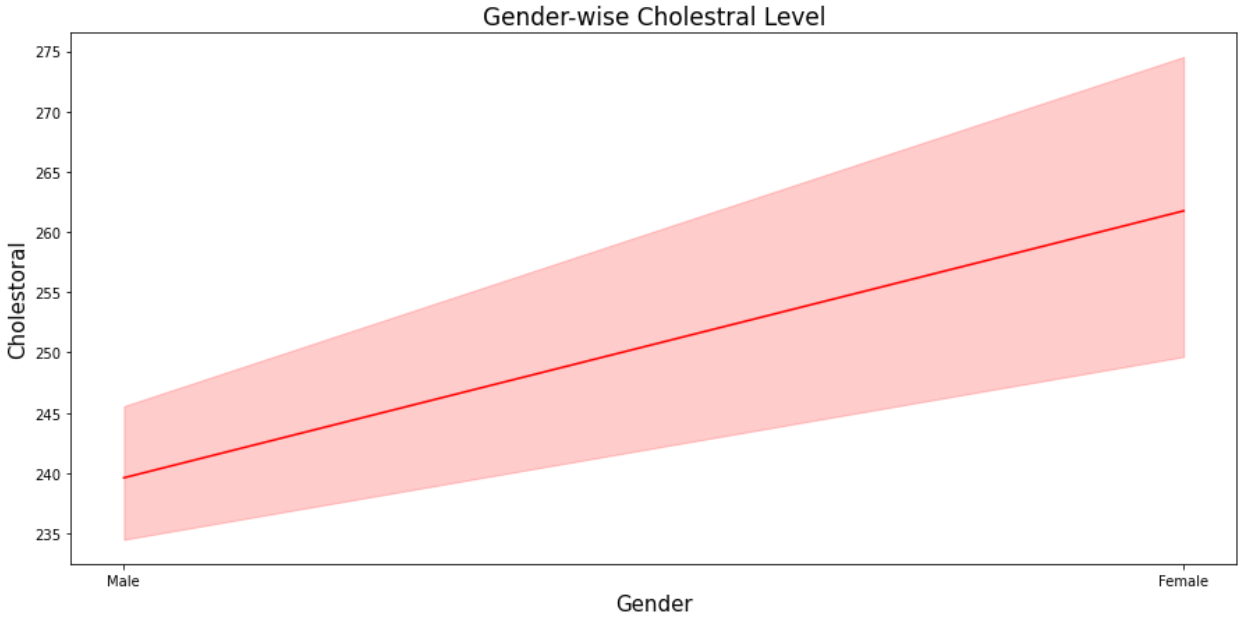
From above figure we can see that majorly people from senior age group experience more Asymptomatic (silent heart attack) than other age group individual.



From above figure we can see that person from age 50-60 there is increase in blood-pressure level i.e. 130 to 150 then there was deep after age of 70.



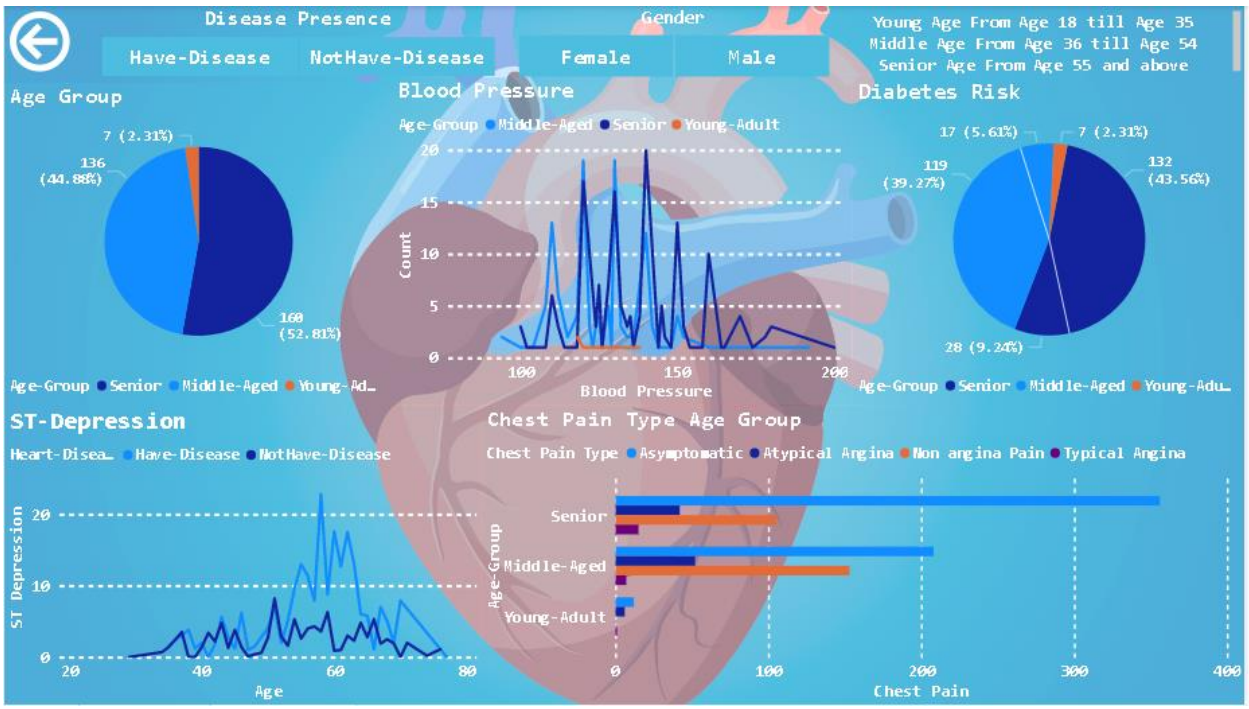
From above figure we can see that there was increase in Cholesterol from age 50 till 70.



From above figure we can see that Female have higher cholesterol in comparison to Male.

POWER-BI DASHBOARD

In the above figure in dashboard contains two criteria for selection of data on basis of gender and disease presence. The left side pie chart denotes out of total sample data how many have disease and many don’t in percentage form, whereas the right side pie chart denotes diabetes risk percentage on the basis of gender. The center graph denotes the age wise disease presence or absence. In bottom left side bar graph display people who have cholesterol and blood pressure have heart problem or not. In bottom center position on the basis of gender people having cholesterol and blood pressure comparison. In bottom right side bar graph denotes people who have angina when they do heavy exercise or long duration exercise (certain complication arises sudden chest pain or breathing problem or heart attack) out which how many male as well female have heart disease or not.



In above figure criteria selection on the basis of disease presence and gender. The first left side pie chart display age group segregation on the percentage form taking from sample data. The second right side pie chart display age group segregation on basis of diabetes risk. The center graph shows the blood pressure on the age group basis. In the bottom left side graph show st-depression on the basis of disease presence. In bottom right side bar graph display chest pain on the age group basis.