BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Course – Advanced Data Visualization (ADV)

Name	Tejal Subhash Komb
UID	2021600037
Batch	A
Lab no	3

Aim :- Design Interactive Dashboards and Storytelling using Power BI Python to be performed on the dataset - Disease spread / Healthcare.

Objectives :- 1) To Create complex visualizations using Python and Power BI
2) To Develop interactive dashboards for dynamic data exploration
3) To Perform EDA to uncover patterns and insights
4) To Combine and clean data using Python and Power BI.

Dataset :- Heart Disease Diagnostic Data

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
42	1	0	140	226	0	1	178	0	0	2	0	2	1
61	1	2	150	243	1	1	137	1	1	1	0	2	1
40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
71	0	1	160	302	0	1	162	0	0.4	2	2	2	1

Dataset Description:
This dataset contains medical diagnostic information related to heart disease. Each row represents an individual patient's data, and the columns represent various medical attributes. Below is a brief description of each column.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

1) **age:** Age of the patient.
2) **sex:** Gender of the patient (1 = male, 0 = female).

3) cp (chest pain type):

i. 0: Typical angina
ii. 1: Atypical angina
iii. 2: Non-anginal pain
iv. 3: Asymptomatic

4) **trestbps:** Resting blood pressure (in mm Hg).
5) **chol:** Serum cholesterol in mg/dl.
6) **fbs:** Fasting blood sugar > 120 mg/dl (1 = true, 0 = false).

7) **restecg:** Resting electrocardiographic results.

i. 0: Normal
ii. 1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
iii. 2: Showing probable or definite left ventricular hypertrophy by Estes' criteria

8) **thalach**: Maximum heart rate achieved.
9) **exang:** Exercise-induced angina (1 = yes, 0 = no).
10) **oldpeak:** ST depression induced by exercise relative to rest.

11) **slope**: The slope of the peak exercise ST segment.

a. 0: Upsloping
b. 1: Flat
c. 2: Downsloping

12) **ca:** Number of major vessels (0-3) colored by fluoroscopy.

13) **thal:** Thalassemia.

i. 1: Normal
ii. 2: Fixed defect
iii. 3: Reversible defect

14) t**arget**: Diagnosis of heart disease (1 = presence of heart disease, 0 = absence of heart disease).

This dataset is useful for analyzing the risk factors associated with heart disease and can be used for predictive modeling and statistical analysis to identify patterns and trends in cardiovascular health.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Implementation :- 1] Implementation Using Python -

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go

df = pd.read_csv('heart.csv')
df.head()

def clean_data(df):
    df = df.dropna()
    return df

missing_values = df.isnull().sum()
print("Missing values in each column:\n", missing_values)

plt.figure(figsize=(10, 6))
sns.countplot(x='sex', data=df)
plt.title('Count of Heart Disease by Sex')
plt.show()

plt.figure(figsize=(8, 8))
df['target'].value_counts().plot.pie(autopct='%1.1f%%', colors=['skyblue',
'orange'], startangle=90, explode=[0.1, 0])
plt.title('Heart Disease Distribution')
plt.ylabel('')
plt.show()
```

```
plt.figure(figsize=(10, 6))
sns.histplot(df['age'], kde=True, bins=20)
plt.title('Age Distribution')
plt.show()

plt.figure(figsize=(10, 6))
sns.lineplot(x=df.index, y='chol', data=df, marker='o')
plt.title('Cholesterol Levels Over Time')
plt.show()

plt.figure(figsize=(10, 6))
sns.scatterplot(x='age', y='trestbps', hue='target', data=df)
plt.title('Scatter Plot of Age vs Resting Blood Pressure')
plt.show()

plt.figure(figsize=(10, 6))
sns.scatterplot(x='age', y='chol', size='trestbps', hue='target', data=df,
sizes=(20, 200))
plt.title('Bubble Plot of Age vs Cholesterol with Resting Blood Pressure')
plt.show()

from wordcloud import WordCloud
cp_labels = {
    0: 'Typical Angina',
    1: 'Atypical Angina',
    2: 'Non-Anginal Pain',
    3: 'Asymptomatic'
}
df['cp_label'] = df['cp'].map(cp_labels)
```

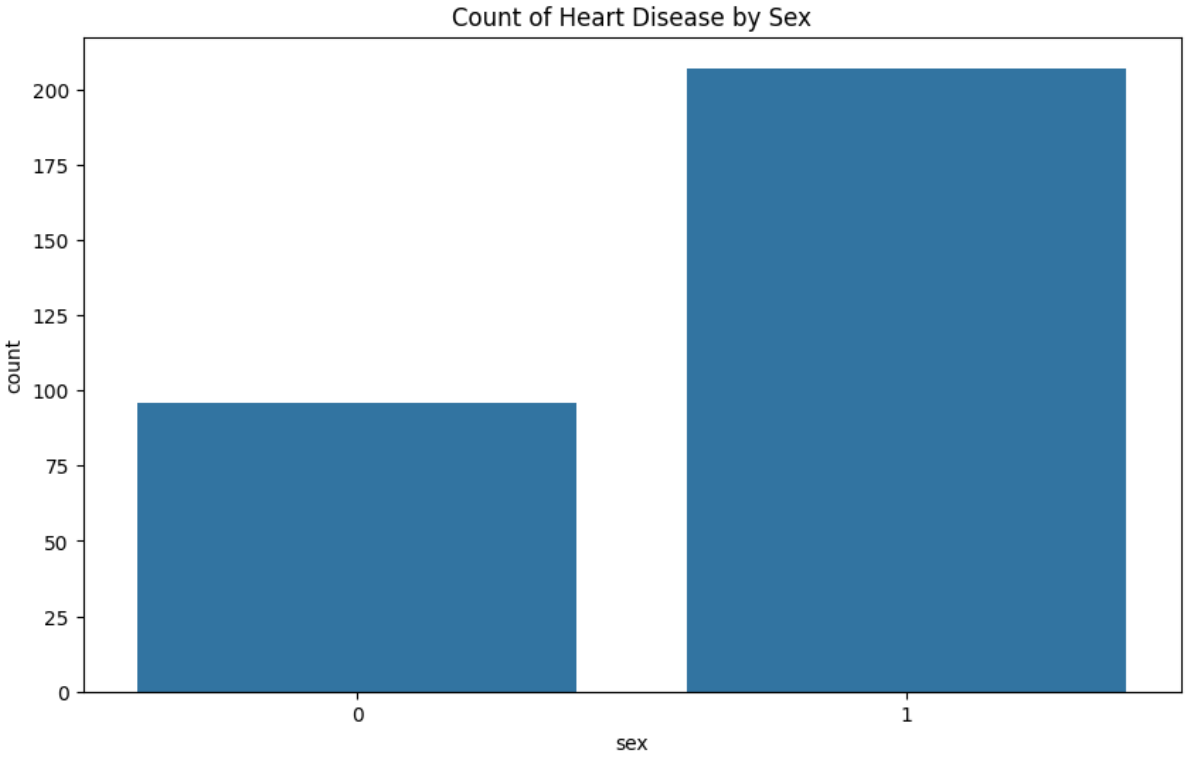
```
# Generate Word Cloud from the cp_label column
plt.figure(figsize=(10, 6))
wordcloud = WordCloud(background_color='white', width=800, height=400).generate('
'.join(df['cp_label']))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of Chest Pain Types')
plt.show()

plt.figure(figsize=(10, 6))
sns.boxplot(x='sex', y='chol', data=df)
plt.title('Boxplot of Cholesterol Levels by Sex')
plt.show()

import squarify

plt.figure(figsize=(10, 6))
sizes = df['cp'].value_counts()
squarify.plot(sizes=sizes, label=sizes.index, alpha=.8)
plt.axis('off')
plt.title('Treemap of Chest Pain Types')
plt.show()
```

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Output :-

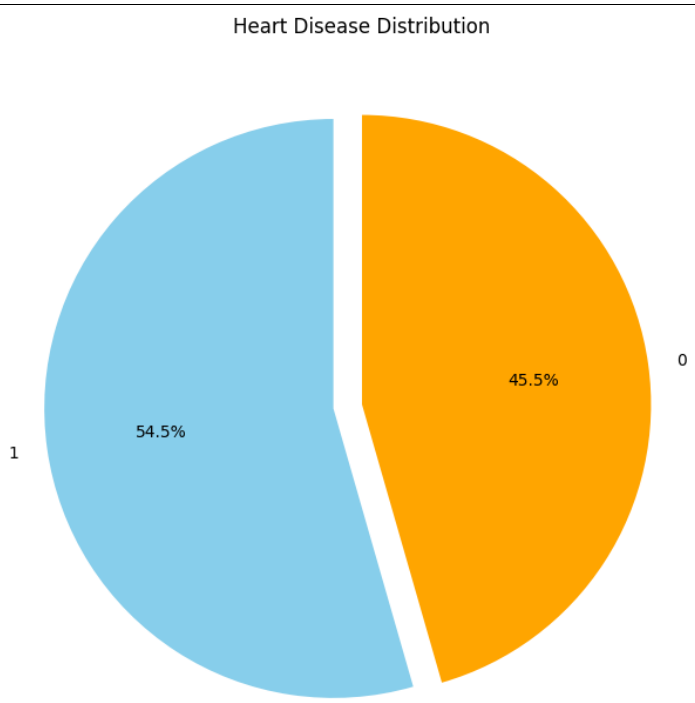
1] Bar Chart :- Count of Heart Disease by Sex



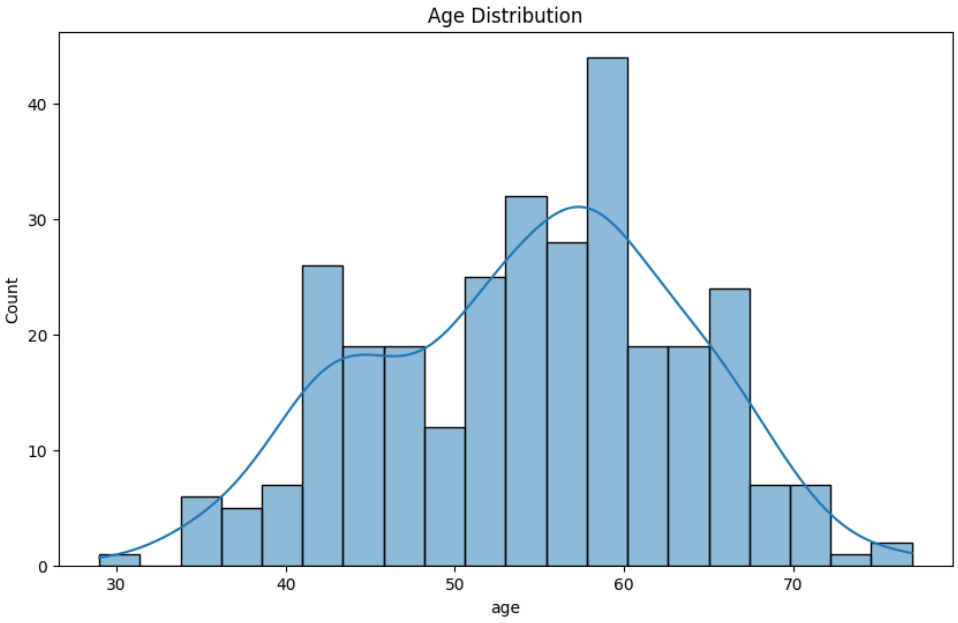
- **Observation:** This could indicate that males are more susceptible to heart disease or that the dataset might have more male patients. Understanding this relationship can help in tailoring preventive measures for different genders.

2. Pie Chart: Heart Disease Distribution

- **Observation:** The pie chart indicates that a majority of the patients in the dataset have been diagnosed with heart disease. This helps in understanding the overall prevalence of heart disease in the population represented by the dataset. The proportion is crucial for healthcare planning and resource allocation.



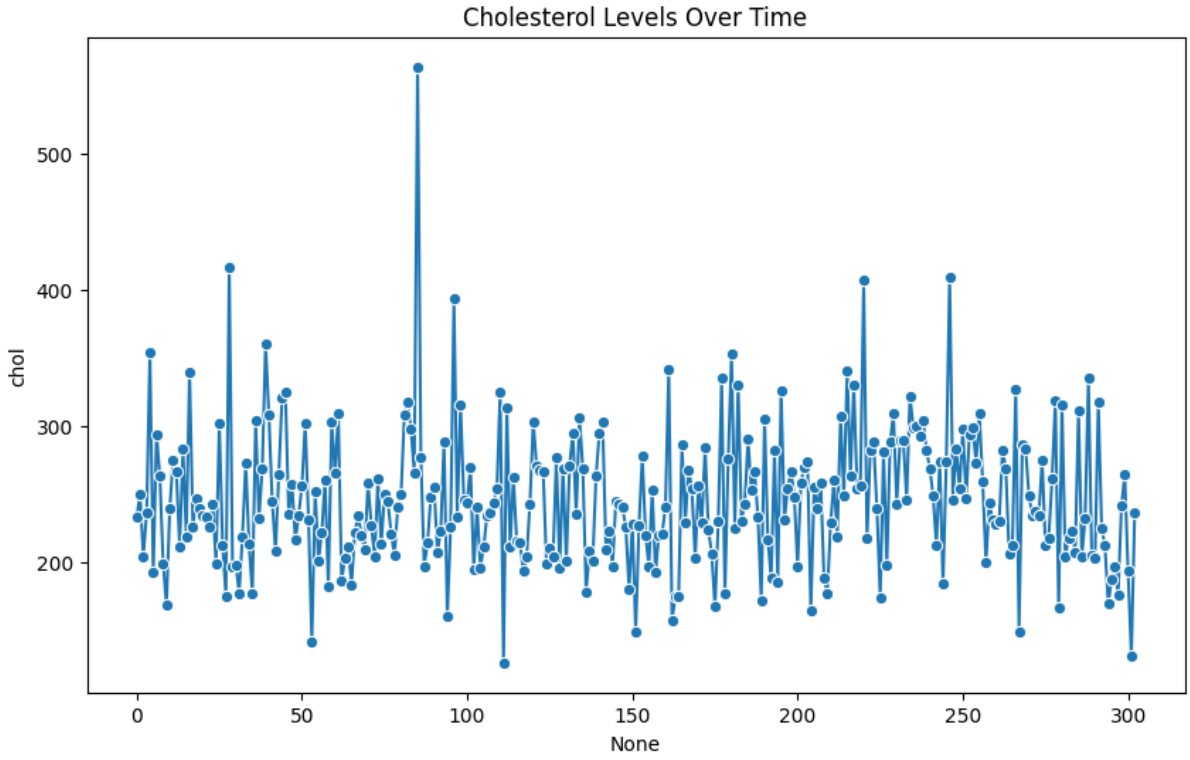
3. Histogram: Age Distribution



- **Observation:** The histogram reveals that most patients in the dataset are aged between 50 and 60 years. This age group seems to be the most affected by heart disease, suggesting that age is a critical risk factor. The distribution also shows a gradual decline in the number of cases in younger and older age groups.

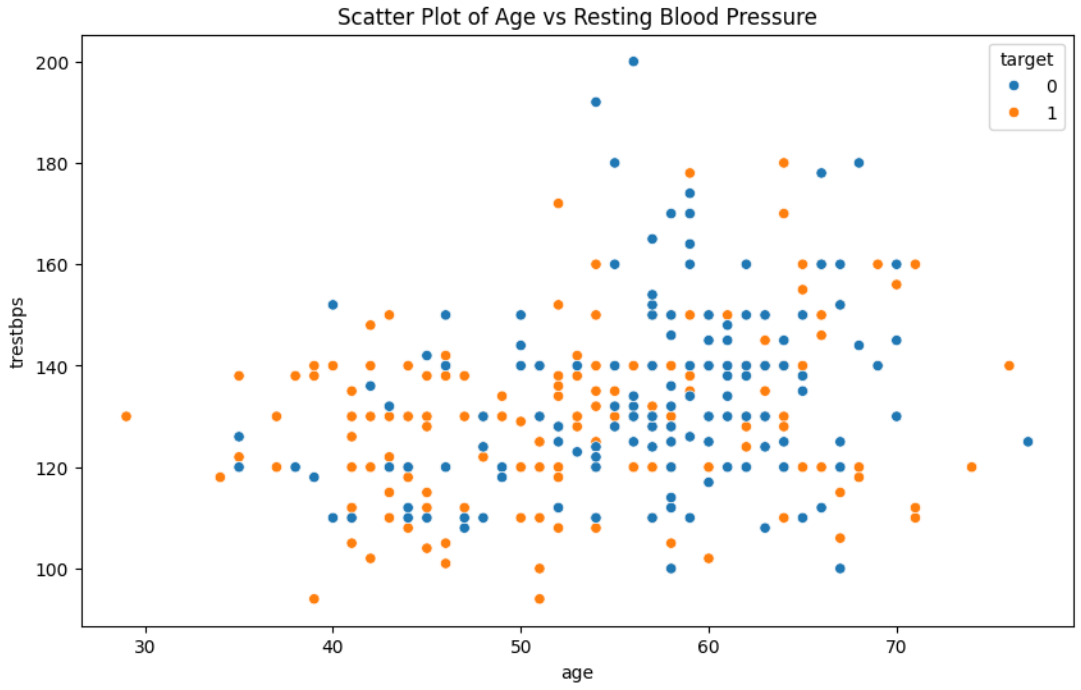
4. Timeline Chart (Scatter plot with line): Cholesterol Levels Over Time

- **Observation:** The timeline chart, which plots cholesterol levels against the order of data points (not time per se), shows fluctuations in cholesterol levels among patients. There isn't a consistent pattern over time, indicating that cholesterol levels vary widely among individuals. This variability could be due to different lifestyles, genetics, or treatment regimens.



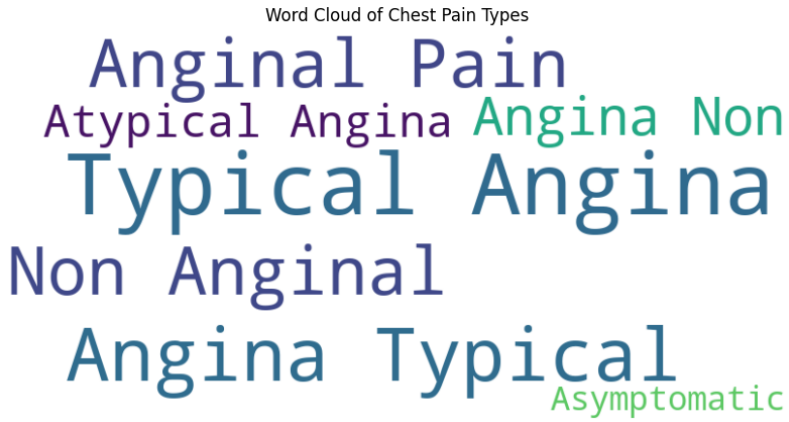
5. Scatter Plot: Age vs Resting Blood Pressure

- **Observation:** The scatter plot shows the relationship between age and resting blood pressure. There is no clear linear relationship, but older patients generally tend to have higher resting blood pressure. This is important as high blood pressure is a known risk factor for heart disease, and its management is crucial, especially in older populations.



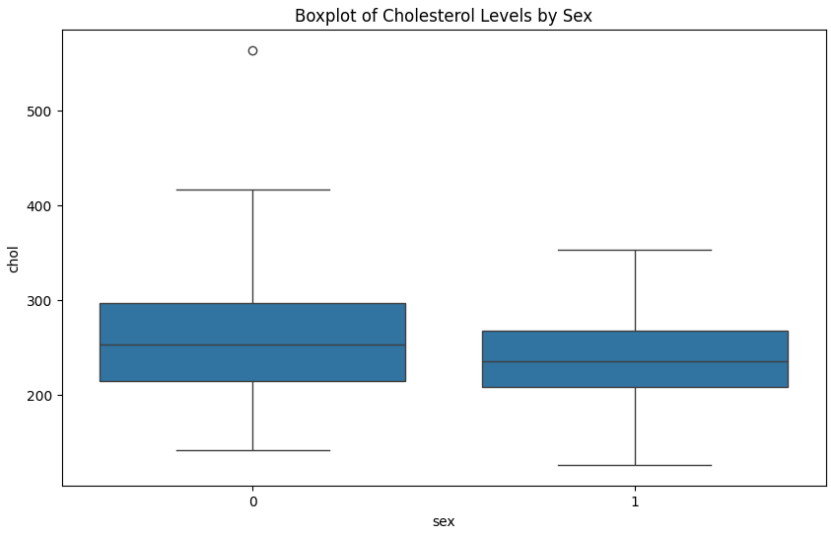
6. Word Cloud: Chest Pain Types

- **Observation:** The word cloud visualizes the frequency of different chest pain types. Chest pain type 0 (typical angina) and type 1 (atypical angina) appear most frequently, suggesting they are common symptoms among patients with heart disease. Understanding the prevalence of different chest pain types can guide diagnostic and treatment approaches.



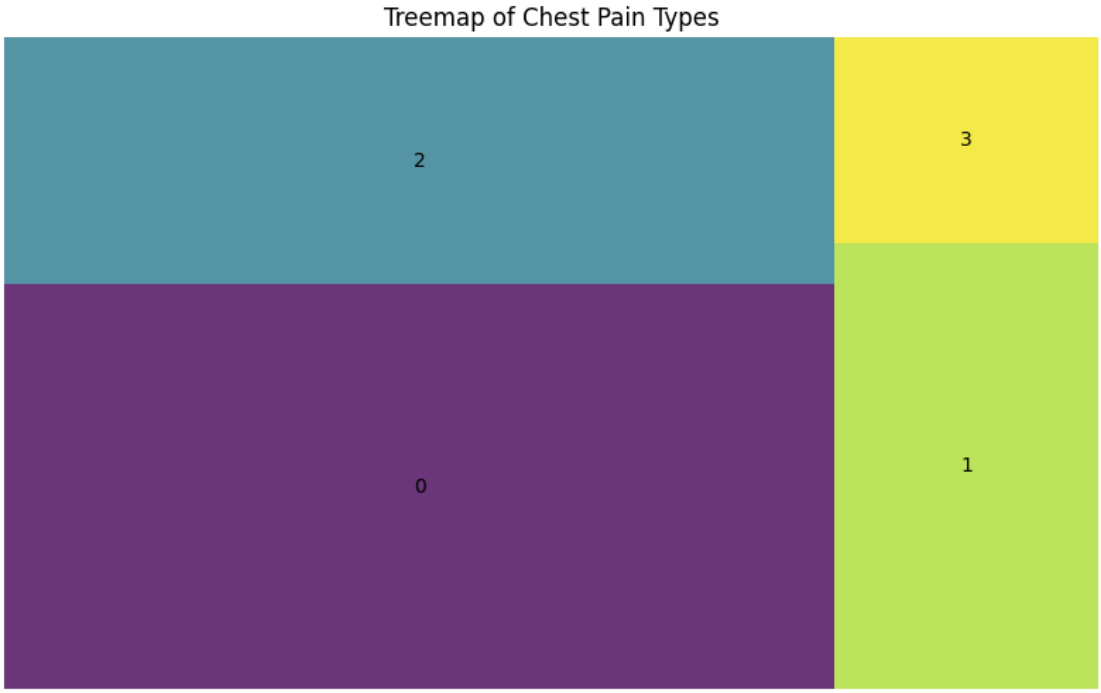
7. Box and Whisker Plot: Cholesterol Levels by Sex

- **Observation:** The box plot shows that cholesterol levels in males are slightly higher than in females, with a wider range of cholesterol levels observed in males. This suggests that while both sexes are at risk, males might have a higher variation in cholesterol levels, which could be a risk factor for heart disease.



8. Treemap: Chest Pain Types

- **Observation:** The treemap shows the distribution of chest pain types among patients, with chest pain type 0 being the most common. This visualization helps in quickly understanding the proportion of different symptoms within the dataset, which can guide further analysis or treatment strategies.



Conclusion :- In this experiment, we aimed to design graph plot relation using Python to analyze heart disease diagnostic data. The key objectives were to create complex visualizations, develop interactive dashboards for exploring data dynamically, perform exploratory data analysis (EDA) to uncover patterns, and combine and clean the data for accurate analysis.