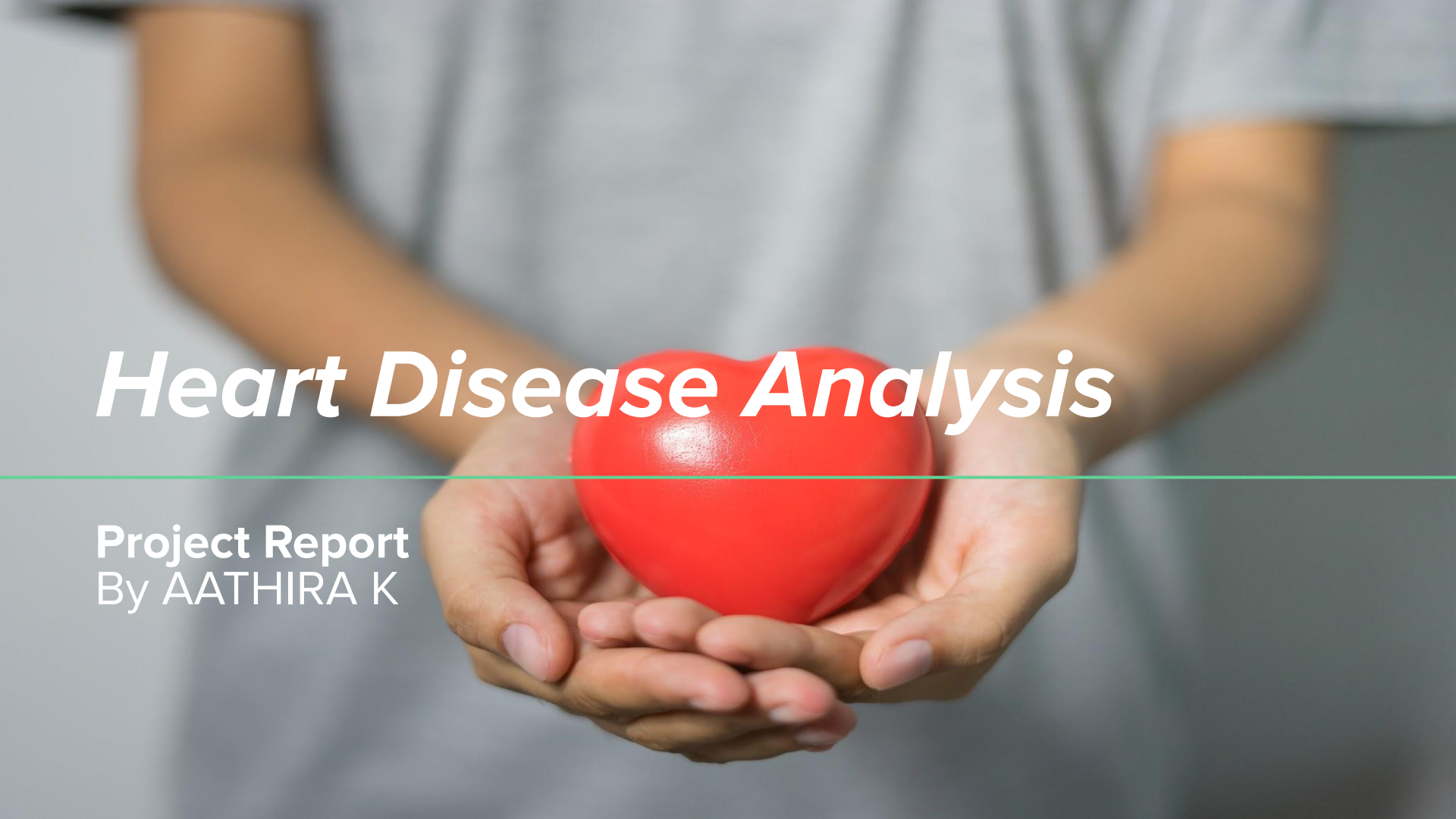


Heart Disease Analysis

Project Report

By AATHIRA K

Problem Statement

Heart disease remains one of the leading causes of death worldwide, necessitating a comprehensive analysis of health data to better understand its risk factors, progression, and outcomes. The objective of this project is to analyze demographic, clinical, and lifestyle data to identify patterns and correlations that can predict the likelihood of heart disease. By leveraging data analytics techniques, this project aims to develop predictive models and actionable insights that can aid healthcare providers in early diagnosis, personalized treatment plans, and preventive measures, ultimately reducing the incidence and severity of heart disease.

Table of Contents

- **Problem Statement**
- **Sample Dataset**
- **Graphical representation based on Several parameters**
- **Dashboards**
- **Cruical Findings**
- **Conclusion**

Sample Dataset

```
# We are reading our data
df = pd.read_csv("Heart Disease data.csv")
```

```
[ ] # First 5 rows of our data
    df.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Dataset Details

Data contains;

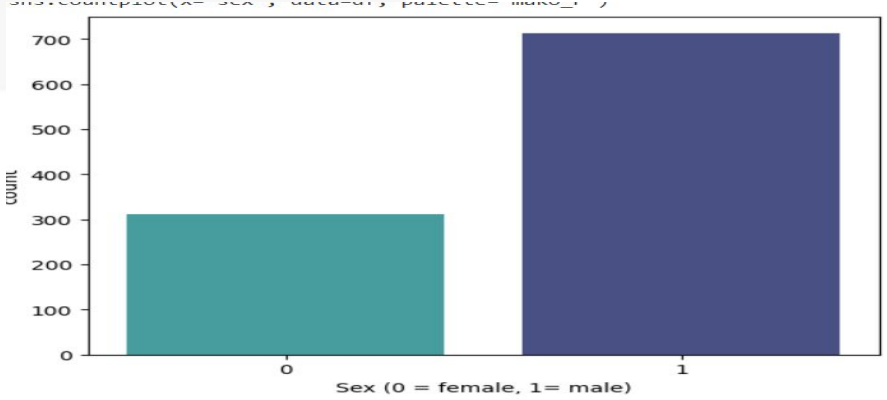
- age - age in years
- sex - (1 = male; 0 = female)
- cp - chest pain type
- trestbps - resting blood pressure (in mm Hg on admission to the hospital)
- chol - serum cholestoral in mg/dl
- fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- restecg - resting electrocardiographic results
- thalach - 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
- ca - number of major vessels (0-3) colored by flourosopy
- thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
- target - have disease or not (1=yes, 0=no)

Total people having disease and does’nt have disease based on Dataset

```
countNoDisease = len(df[df.target == 0])
countHaveDisease = len(df[df.target == 1])
print("Percentage of Patients Haven't Heart Disease: {:.2f}%".format((countNoDisease / (len(df.target))*100)))
print("Percentage of Patients Have Heart Disease: {:.2f}%".format((countHaveDisease / (len(df.target))*100)))
```

```
Percentage of Patients Haven't Heart Disease: 48.68%
Percentage of Patients Have Heart Disease: 51.32%
```

```
sns.countplot(x='sex', data=df, palette="mako_r")
plt.xlabel("Sex (0 = female, 1= male)")
plt.show()
```



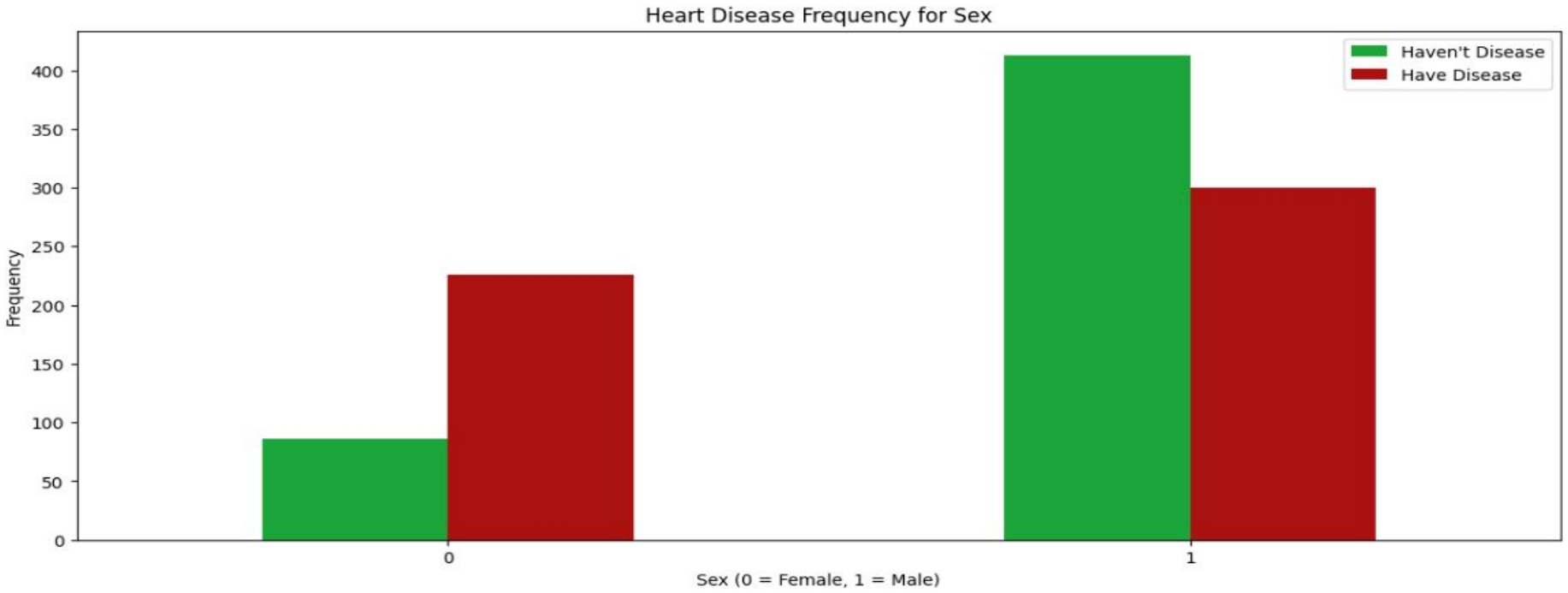
Disease Analysis based on Sex

```
countFemale = len(df[df.sex == 0])
countMale = len(df[df.sex == 1])
print("Percentage of Female Patients: {:.2f}%".format((countFemale / (len(df.sex))*100)))
print("Percentage of Male Patients: {:.2f}%".format((countMale / (len(df.sex))*100)))
```

```
Percentage of Female Patients: 30.44%
Percentage of Male Patients: 69.56%
```

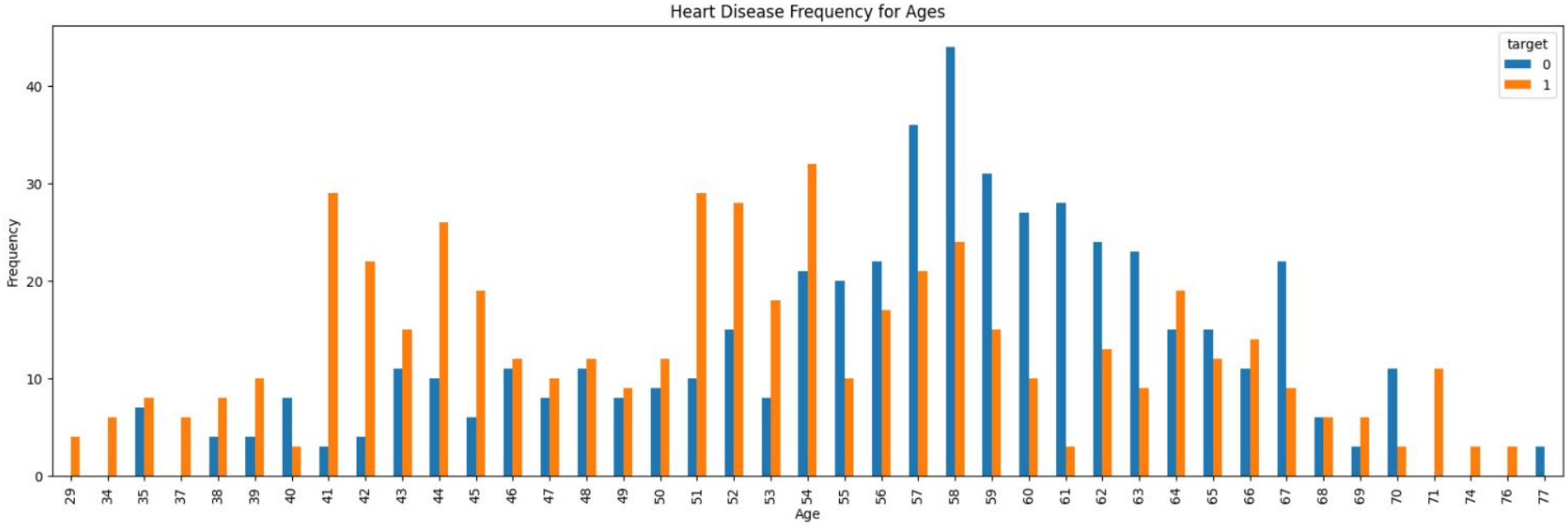
Graphical Representation

```
pd.crosstab(df.sex,df.target).plot(kind="bar",figsize=(15,6),color=['#1CA53B','#AA1111' ])
plt.title('Heart Disease Frequency for Sex')
plt.xlabel('Sex (0 = Female, 1 = Male)')
plt.xticks(rotation=0)
plt.legend(["Haven't Disease", "Have Disease"])
plt.ylabel('Frequency')
plt.show()
```



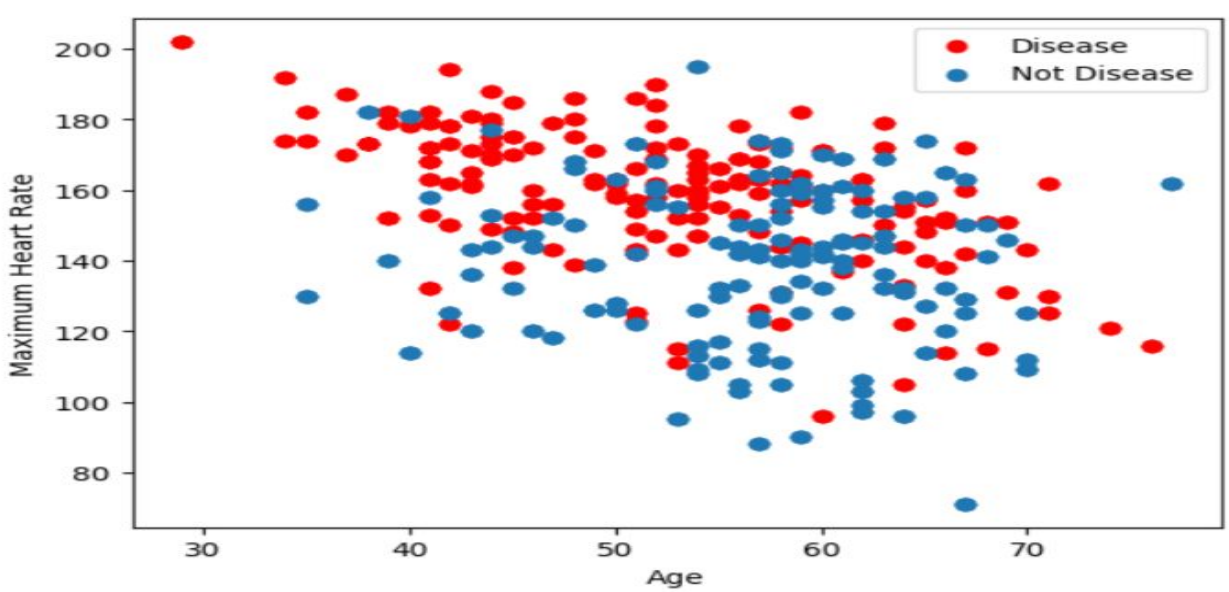
Analysis Based on Age

```
pd.crosstab(df.age,df.target).plot(kind="bar",figsize=(20,6))
plt.title('Heart Disease Frequency for Ages')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.savefig('heartDiseaseAndAges.png')
plt.show()
```



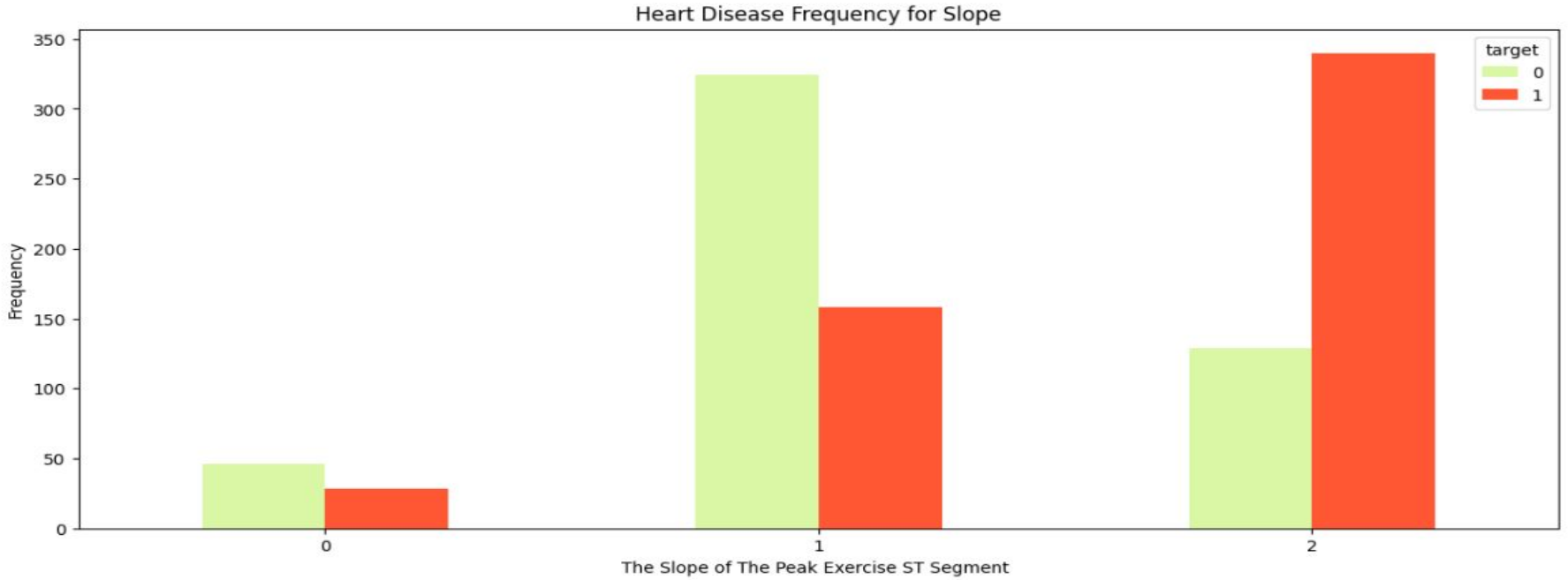
Analysis based on Maximum Heart Rate

```
[ ]  plt.scatter(x=df.age[df.target==1], y=df.thalach[(df.target==1)], c="red")
     plt.scatter(x=df.age[df.target==0], y=df.thalach[(df.target==0)])
     plt.legend(["Disease", "Not Disease"])
     plt.xlabel("Age")
     plt.ylabel("Maximum Heart Rate")
     plt.show()
```



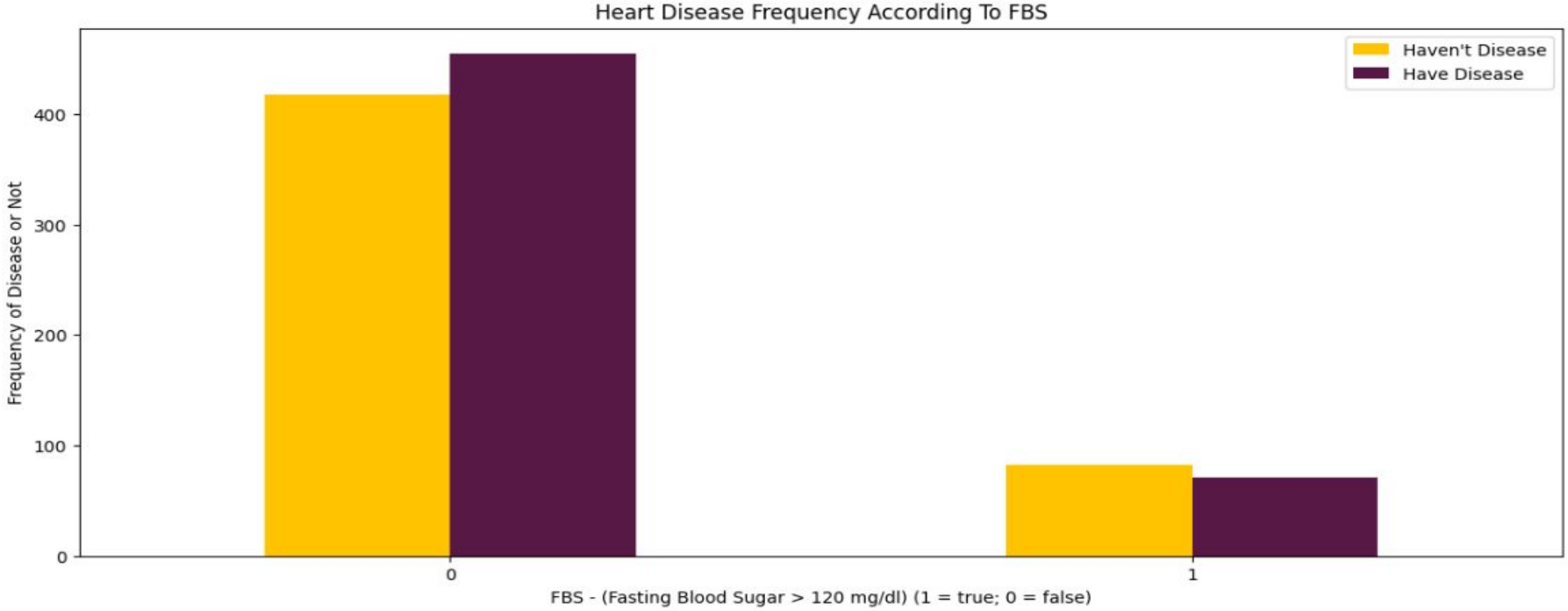
Analysis based on Slope of ST segment

```
pd.crosstab(df.slope,df.target).plot(kind="bar",figsize=(15,6),color=['#DAF7A6','#FF5733' ])
plt.title('Heart Disease Frequency for Slope')
plt.xlabel('The Slope of The Peak Exercise ST Segment ')
plt.xticks(rotation = 0)
plt.ylabel('Frequency')
plt.show()
```



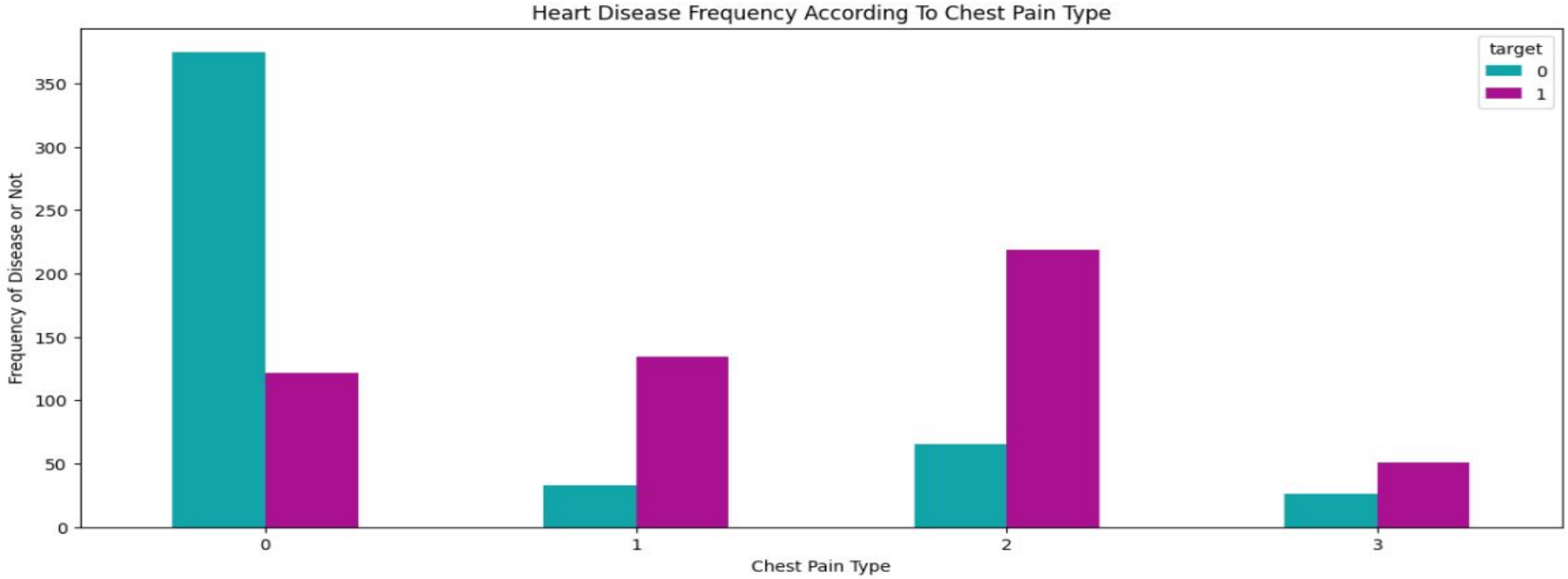
Based on Fasting Blood Pressure

```
[ ] pd.crosstab(df.fbs,df.target).plot(kind="bar",figsize=(15,6),color=['#FFC300','#581845' ])
    plt.title('Heart Disease Frequency According To FBS')
    plt.xlabel('FBS - (Fasting Blood Sugar > 120 mg/dl) (1 = true; 0 = false)')
    plt.xticks(rotation = 0)
    plt.legend(["Haven't Disease", "Have Disease"])
    plt.ylabel('Frequency of Disease or Not')
    plt.show()
```

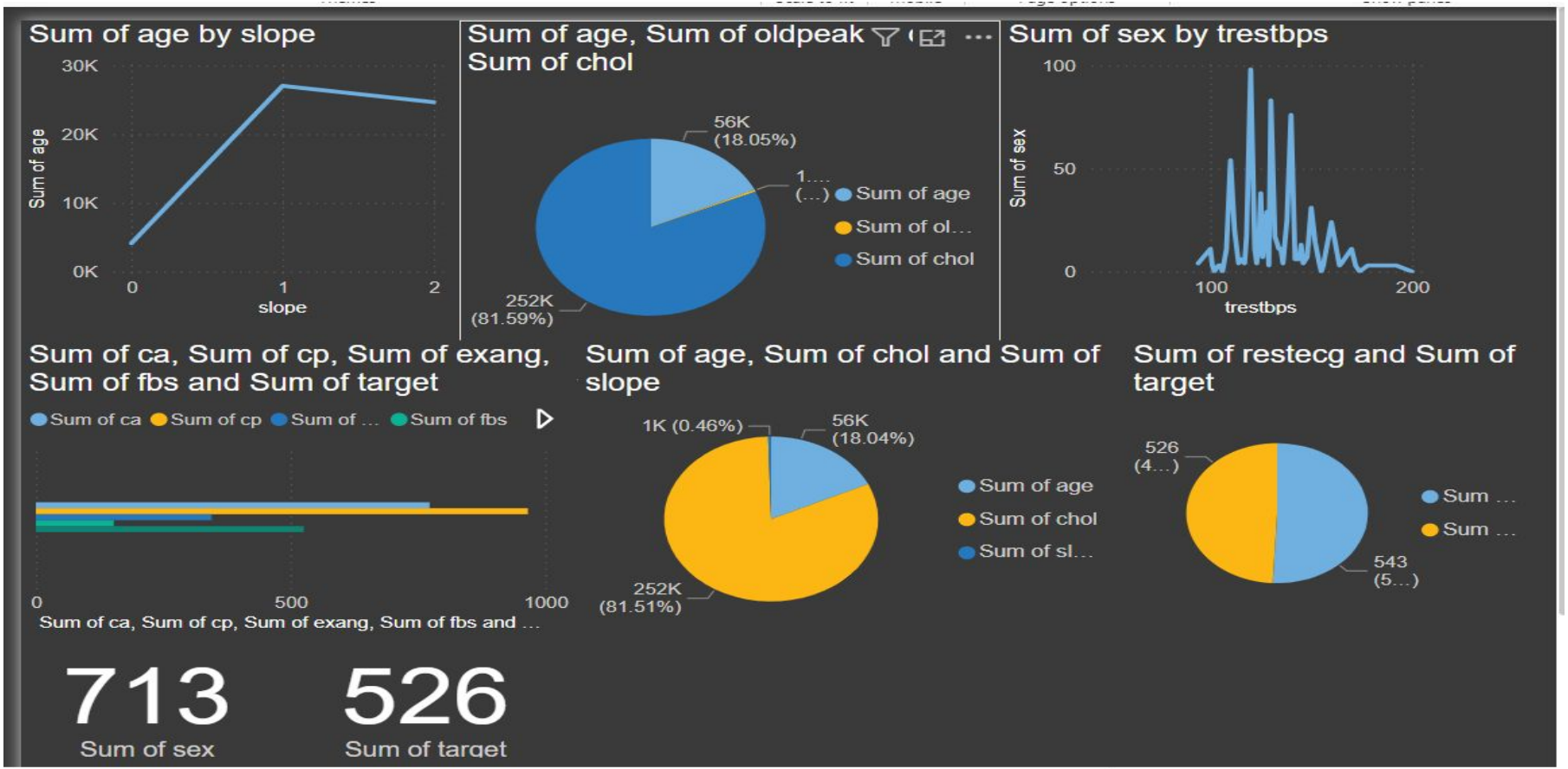


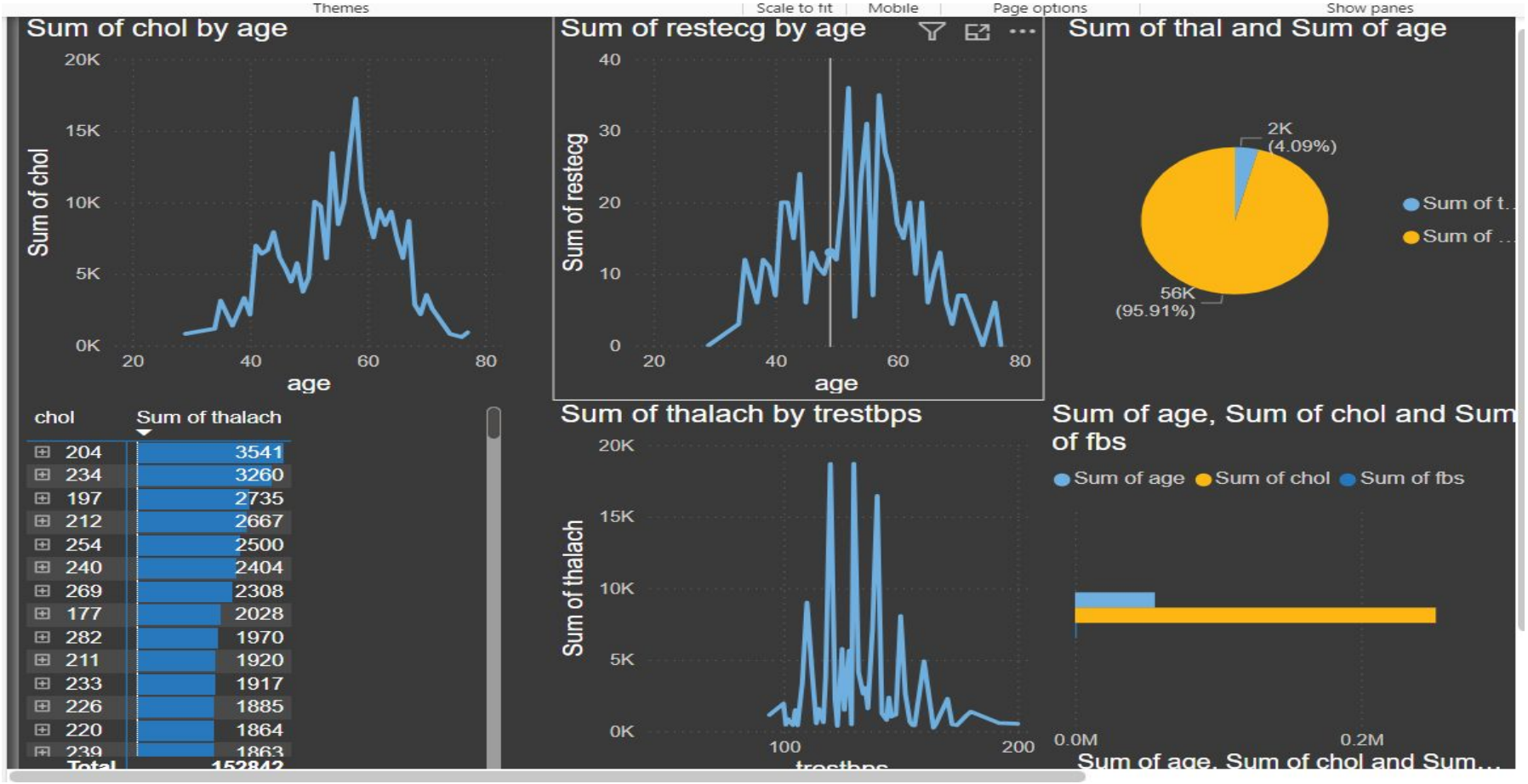
Based on Chest Pain Type

```
pd.crosstab(df.cp,df.target).plot(kind="bar",figsize=(15,6),color=['#11A5AA','#AA1190' ])
plt.title('Heart Disease Frequency According To Chest Pain Type')
plt.xlabel('Chest Pain Type')
plt.xticks(rotation = 0)
plt.ylabel('Frequency of Disease or Not')
plt.show()
```



Dashboards


Sum of age by slope
30K
20K
10K
0K
Sum of age
0
1
2
slope
Sum of age, Sum of oldpeak
Sum of chol
56K
(18.05%)
1....
(...)
252K
(81.59%)
Sum of age
Sum of ol...
Sum of chol
Sum of sex by trestbps
100
50
0
Sum of sex
100
200
trestbps
Sum of ca, Sum of cp, Sum of exang,
Sum of fbs and Sum of target
Sum of ca
Sum of cp
Sum of ...
Sum of fbs
0
500
1000
Sum of ca, Sum of cp, Sum of exang, Sum of fbs and ...
713
Sum of sex
526
Sum of target
Sum of age, Sum of chol and Sum of
slope
1K (0.46%)
56K
(18.04%)
252K
(81.51%)
Sum of age
Sum of chol
Sum of sl...
Sum of restecg and Sum of
target
526
(4...)
543
(5...)
Sum ...
Sum ...


Themes
Scale to fit
Mobile
Page options
Show panes
Sum of chol by age
20K
15K
10K
5K
0K
Sum of chol
20
40
60
80
age
Sum of restecg by age
40
30
20
10
0
Sum of restecg
20
40
60
80
age
Sum of thal and Sum of age
2K
(4.09%)
56K
(95.91%)
Sum of t...
Sum of ...
chol
Sum of thalach
204 3541
234 3260
197 2735
212 2667
254 2500
240 2404
269 2308
177 2028
282 1970
211 1920
233 1917
226 1885
220 1864
239 1863
Total 152842
Sum of thalach by trestbps
20K
15K
10K
5K
0K
Sum of thalach
100
200
trestbps
Sum of age, Sum of chol and Sum of fbs
Sum of age
Sum of chol
Sum of fbs
0.0M
0.2M
Sum of age, Sum of chol and Sum...

Crucial Findings

When data is trained using Decision Tree Algorithm:

```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier()
dtc.fit(x_train.T, y_train.T)

acc = dtc.score(x_test.T, y_test.T)*100
accuracies['Decision Tree'] = acc
print("Decision Tree Test Accuracy {:.2f}%".format(acc))
```

```
Decision Tree Test Accuracy 100.00%
```

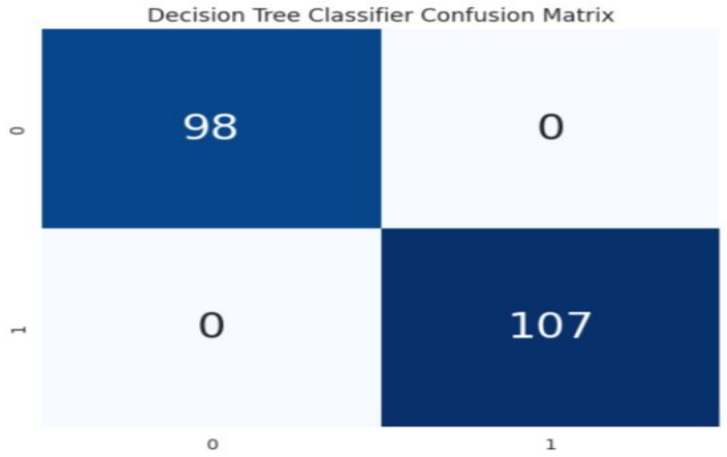
Confusion matrix

```
# Predicted values
y_head_lr = lr.predict(x_test.T)
knn3 = KNeighborsClassifier(n_neighbors = 3)
knn3.fit(x_train.T, y_train.T)
y_head_knn = knn3.predict(x_test.T)
y_head_svm = svm.predict(x_test.T)
y_head_dtc = dtc.predict(x_test.T)
y_head_rf = rf.predict(x_test.T)
```

```
[ ] from sklearn.metrics import confusion_matrix

cm_lr = confusion_matrix(y_test,y_head_lr)
cm_knn = confusion_matrix(y_test,y_head_knn)
cm_svm = confusion_matrix(y_test,y_head_svm)
cm_dtc = confusion_matrix(y_test,y_head_dtc)
cm_rf = confusion_matrix(y_test,y_head_rf)
```

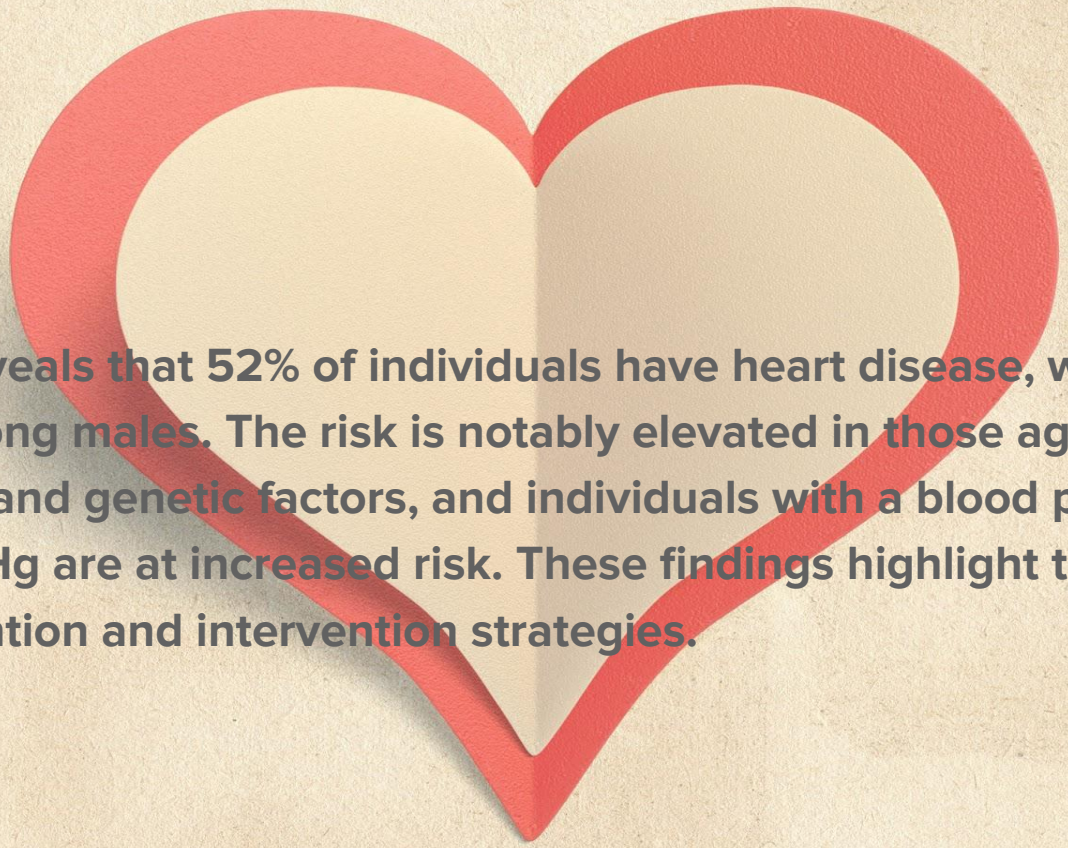
```
plt.figure(figsize=(24,12))

plt.suptitle("Confusion Matrixes",fontsize=24)
plt.subplots_adjust(wspace = 0.4, hspace= 0.4)
```

```
plt.subplot(2,3,5)
plt.title("Decision Tree Classifier Confusion Matrix")
sns.heatmap(cm_dtc,annot=True,cmap="Blues",fmt="d",cbar=False, annot_kws={"size": 24})
```



Conclusions

- **Based on the given dataset:**
- **Total people having disease is 52% and doesnt having disease is 48.62%.**
- **Based on Sex: Males have more chances of having heart disease as compared to that of Females.**
- **Based on Age:People belonging to an age group of 42 -52 have more chances of having Heart Disease , mostly because of Lifestyle,Stress Management,Generic Issues etc.**
- **Based on blood pressure people having a BP range of 120 -130 have most chances of heart diseases.**



The analysis reveals that 52% of individuals have heart disease, with a higher prevalence among males. The risk is notably elevated in those aged 42-52, likely due to lifestyle and genetic factors, and individuals with a blood pressure range of 120-130 mmHg are at increased risk. These findings highlight the need for targeted prevention and intervention strategies.


THANKYOU
Thank
You