AJEENKYA
D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY

A

MINI PROJECT REPORT ON

“Heart Disease Detection and Analysis ”

FOR

Term Work Examination

Bachelor of Computer Application in Artificial Intelligence and Machine Learning (BCA - AIML)

Year 2024-2025

Ajeenkya DY Patil University, Pune

-Submitted By-

Mr : Sachin Swami

Under the guidance of

Prof. Vivek More

Ajeenkya DY Patil University

D Y Patil Knowledge City,
Charholi Bk. Via Lohegaon,
Pune - 412105
Maharashtra (India)

Date: 11/04/ 2025

CERTIFICATE

This is to certified that Sachin Swami
A student's of **BCA(AIML) SEM-IV** URN No 2023-B-19112005
has Successfully Completed the Dashboard Report On

"Heart Disease Detection and Analysis"

As per the requirement of
Ajeenkya DY Patil University, Pune was carried out under my supervision.
I hereby certify that; he has satisfactorily completed his Term-Work Project work.

Place: - Pune

Examiner

INDEX

INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide. Early detection and prediction of heart-related issues can significantly improve patient outcomes. In this project, we aim to build a machine learning model to predict the presence of heart disease based on various health-related attributes such as age, blood pressure, cholesterol levels, and more. We utilized a public dataset (`heart.csv`) and applied data analysis, preprocessing, visualization, and modeling techniques to develop an efficient and accurate prediction system.

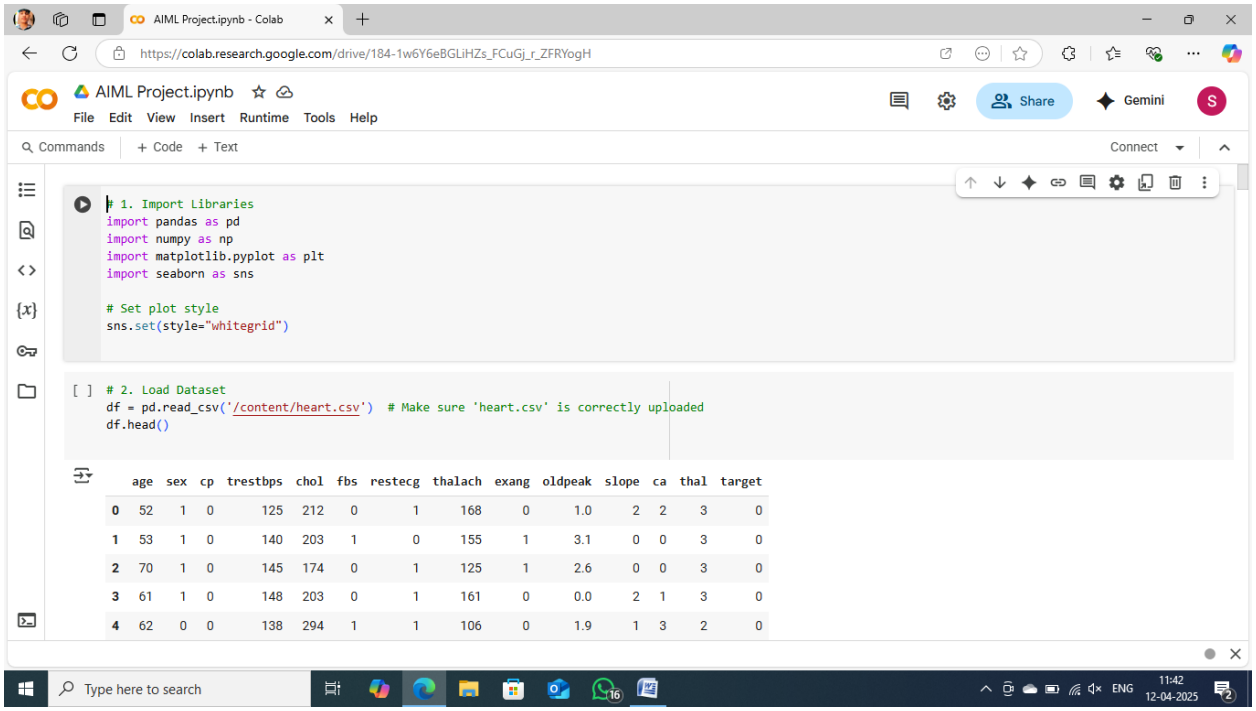
METHODOLOGY & APPROACH

1. **Data Collection:**
   The dataset used for this project was sourced from a publicly available heart disease dataset containing 303 records with 14 attributes including age, sex, chest pain type,

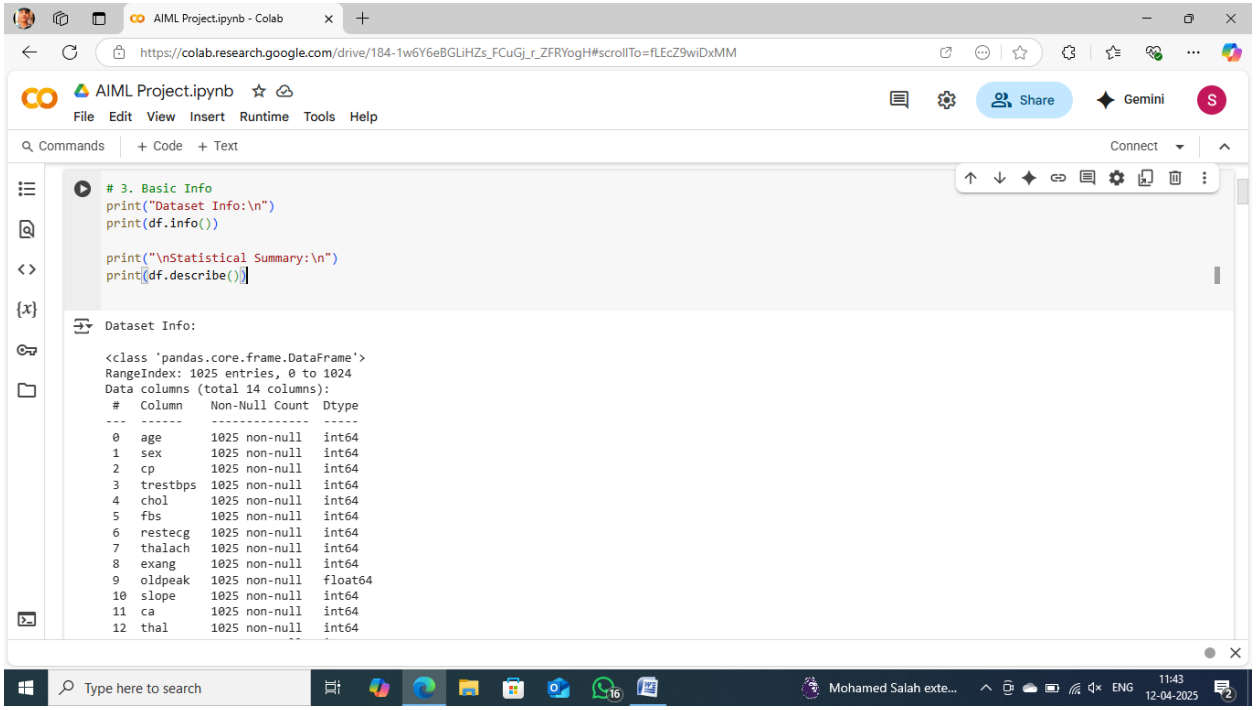
resting blood pressure, cholesterol, fasting blood sugar, resting ECG results, maximum heart rate achieved, and others.

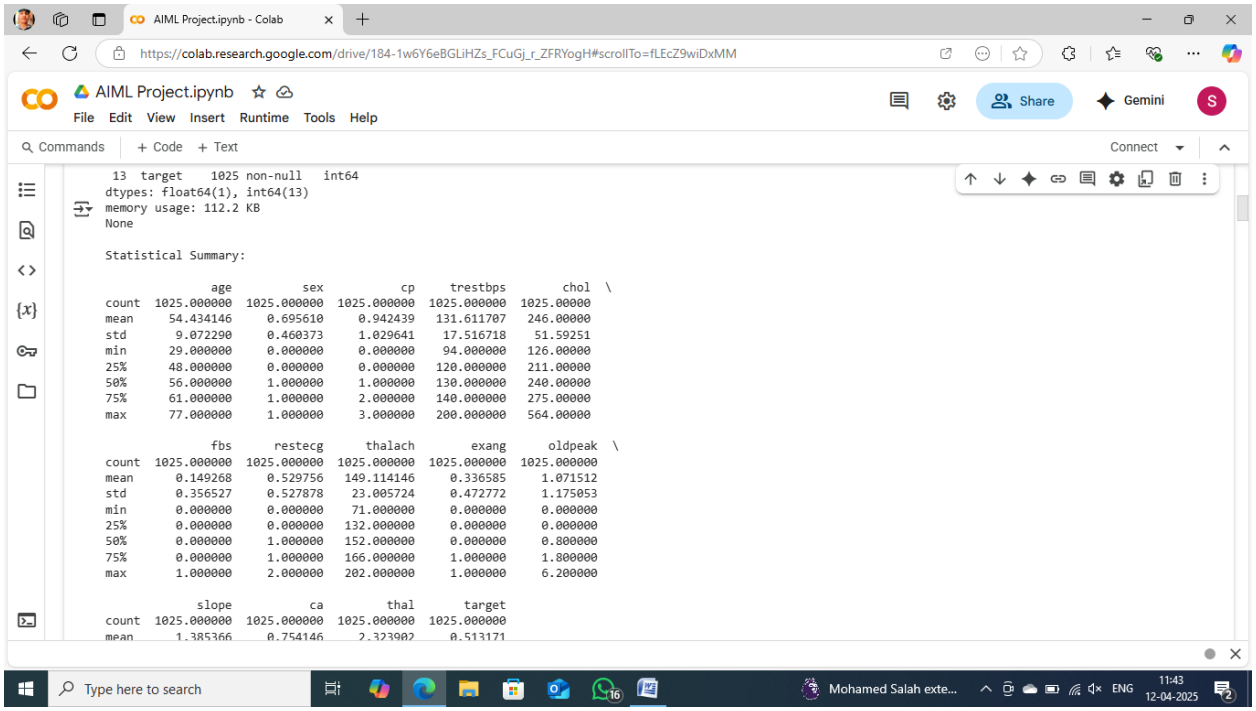
2. **Data Preprocessing:**
    - Checked for missing values and handled them appropriately (though the dataset had minimal to no missing values).
    - Removed duplicate entries to maintain data quality.
    - Identified outliers through boxplots to ensure the robustness of the model.
3. **Exploratory Data Analysis (EDA):**
    - Visualized distributions of features like age and cholesterol.
    - Explored relationships between features and the target variable (heart disease presence).
    - Created a correlation heatmap to understand interdependencies among features.
4. **Model Building:**
    - Applied various classification algorithms (example: Logistic Regression, Decision Trees, Random Forests, etc.) for prediction.
    - Split the data into training and testing sets.
    - Evaluated model performance using metrics such as accuracy, precision, recall, and F1-score.
5. **Model Evaluation:**
    - Selected the best-performing model based on evaluation metrics.
    - Plotted confusion matrix and ROC curve for visual analysis of model performance.

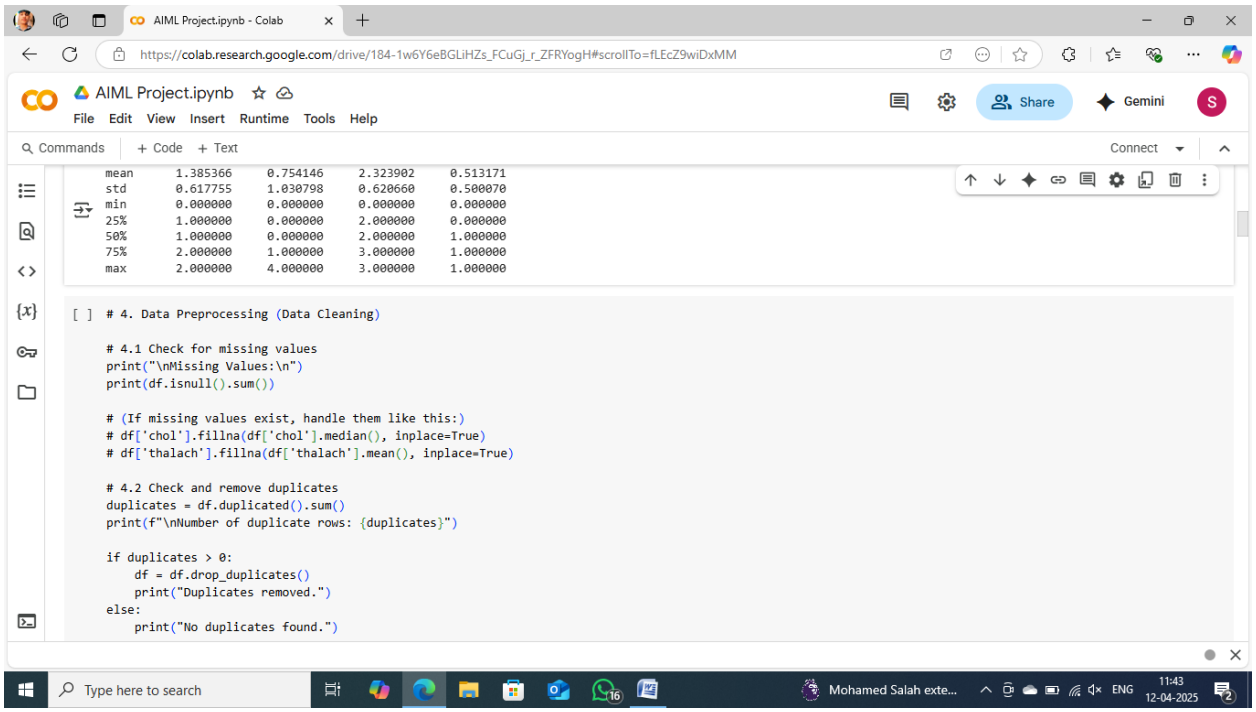
Implementation and Code :

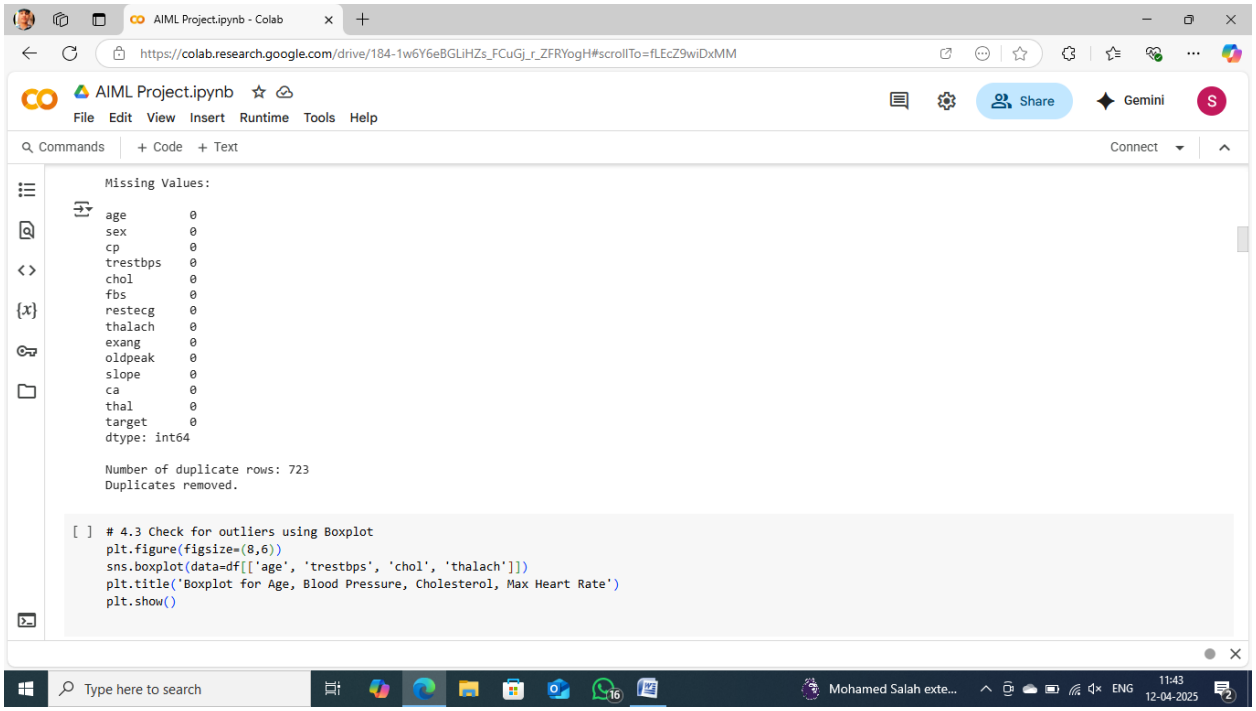


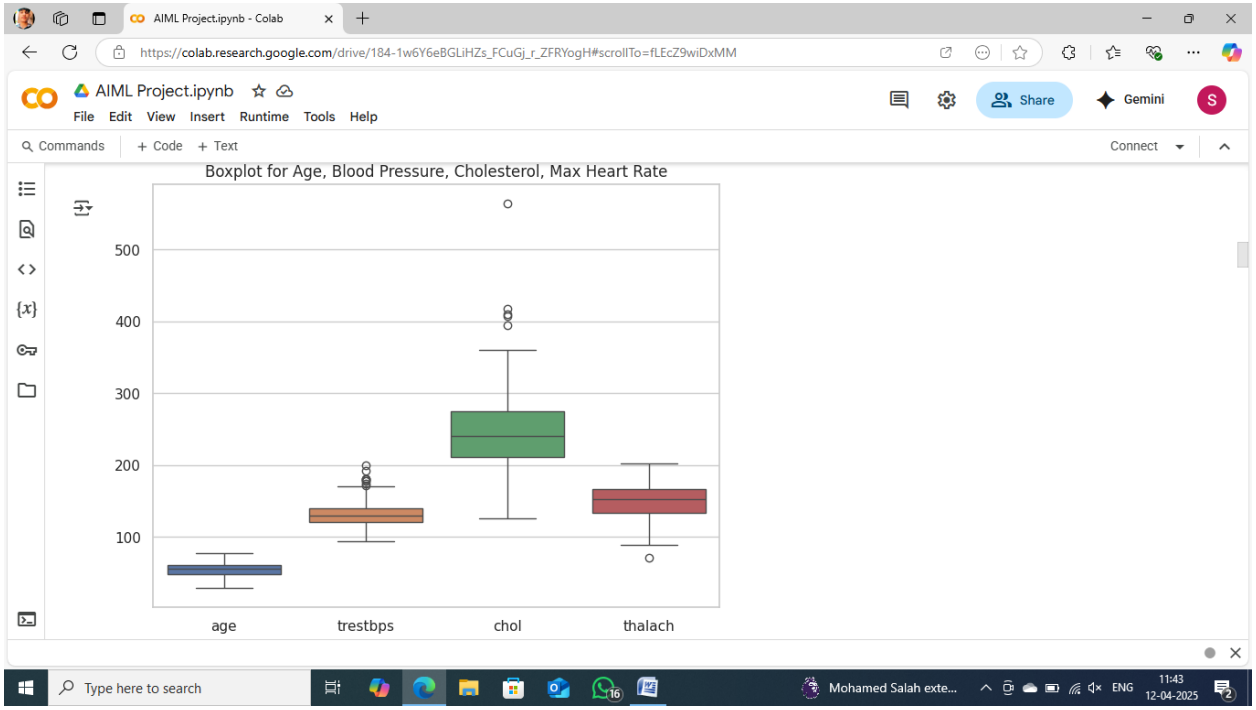
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0




AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=fLEcZ9wiDxMM
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
13 target 1025 non-null int64
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
None
Statistical Summary:
age sex cp trestbps chol \
count 1025.000000 1025.000000 1025.000000 1025.000000 1025.00000
mean 54.434146 0.695610 0.942439 131.611707 246.00000
std 9.072290 0.460373 1.029641 17.516718 51.59251
min 29.000000 0.000000 0.000000 94.000000 126.00000
25% 48.000000 0.000000 0.000000 120.000000 211.00000
50% 56.000000 1.000000 1.000000 130.000000 240.00000
75% 61.000000 1.000000 2.000000 140.000000 275.00000
max 77.000000 1.000000 3.000000 200.000000 564.00000
fbs restecg thalach exang oldpeak \
count 1025.000000 1025.000000 1025.000000 1025.000000 1025.000000
mean 0.149268 0.529756 149.114146 0.336585 1.071512
std 0.356527 0.527878 23.005724 0.472772 1.175053
min 0.000000 0.000000 71.000000 0.000000 0.000000
25% 0.000000 0.000000 132.000000 0.000000 0.000000
50% 0.000000 1.000000 152.000000 0.000000 0.800000
75% 0.000000 1.000000 166.000000 1.000000 1.800000
max 1.000000 2.000000 202.000000 1.000000 6.200000
slope ca thal target
count 1025.000000 1025.000000 1025.000000 1025.000000
mean 1.385366 0.754146 2.323902 0.513171
Type here to search
Mohamed Salah exte...
ENG
11:43
12-04-2025


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=fLEcZ9wiDxMM
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
mean 1.385366 0.754146 2.323902 0.513171
std 0.617755 1.030798 0.620660 0.500070
min 0.000000 0.000000 0.000000 0.000000
25% 1.000000 0.000000 2.000000 0.000000
50% 1.000000 0.000000 2.000000 1.000000
75% 2.000000 1.000000 3.000000 1.000000
max 2.000000 4.000000 3.000000 1.000000
[ ] # 4. Data Preprocessing (Data Cleaning)
# 4.1 Check for missing values
print("\nMissing Values:\n")
print(df.isnull().sum())
# (If missing values exist, handle them like this:)
# df['chol'].fillna(df['chol'].median(), inplace=True)
# df['thalach'].fillna(df['thalach'].mean(), inplace=True)
# 4.2 Check and remove duplicates
duplicates = df.duplicated().sum()
print(f"\nNumber of duplicate rows: {duplicates}")
if duplicates > 0:
df = df.drop_duplicates()
print("Duplicates removed.")
else:
print("No duplicates found.")
Type here to search
Mohamed Salah exte...
ENG
11:43
12-04-2025


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=fLEcZ9wiDxMM
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
Missing Values:
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
target 0
dtype: int64
Number of duplicate rows: 723
Duplicates removed.
[ ] # 4.3 Check for outliers using Boxplot
plt.figure(figsize=(8,6))
sns.boxplot(data=df[['age', 'trestbps', 'chol', 'thalach']])
plt.title('Boxplot for Age, Blood Pressure, Cholesterol, Max Heart Rate')
plt.show()
Type here to search
Mohamed Salah exte...
ENG
11:43
12-04-2025


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=fLEcZ9wiDxMM
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
Boxplot for Age, Blood Pressure, Cholesterol, Max Heart Rate
500
400
300
200
100
age
trestbps
chol
thalach
Type here to search
Mohamed Salah exte...
ENG
11:43
12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=qB9YLkb3D-bq

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

```
# 5.1 Age Distribution
plt.figure(figsize=(8,6))
sns.histplot(df['age'], bins=30, kde=True)
plt.title('Age Distribution of Patients')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

AIML Project.ipynb - Colab

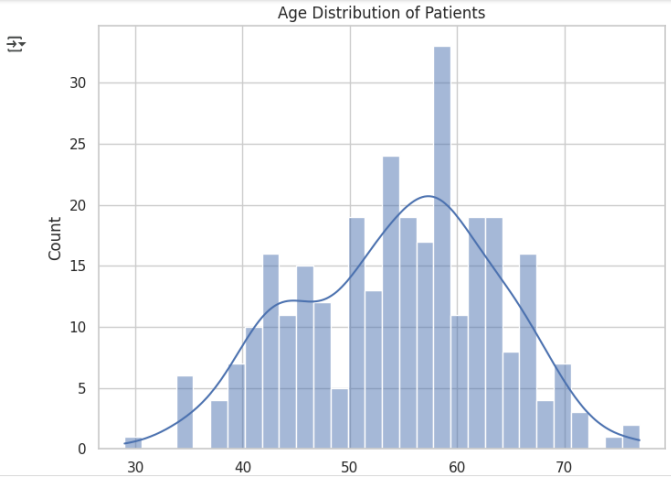
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=qB9YLkb3D-bq

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g

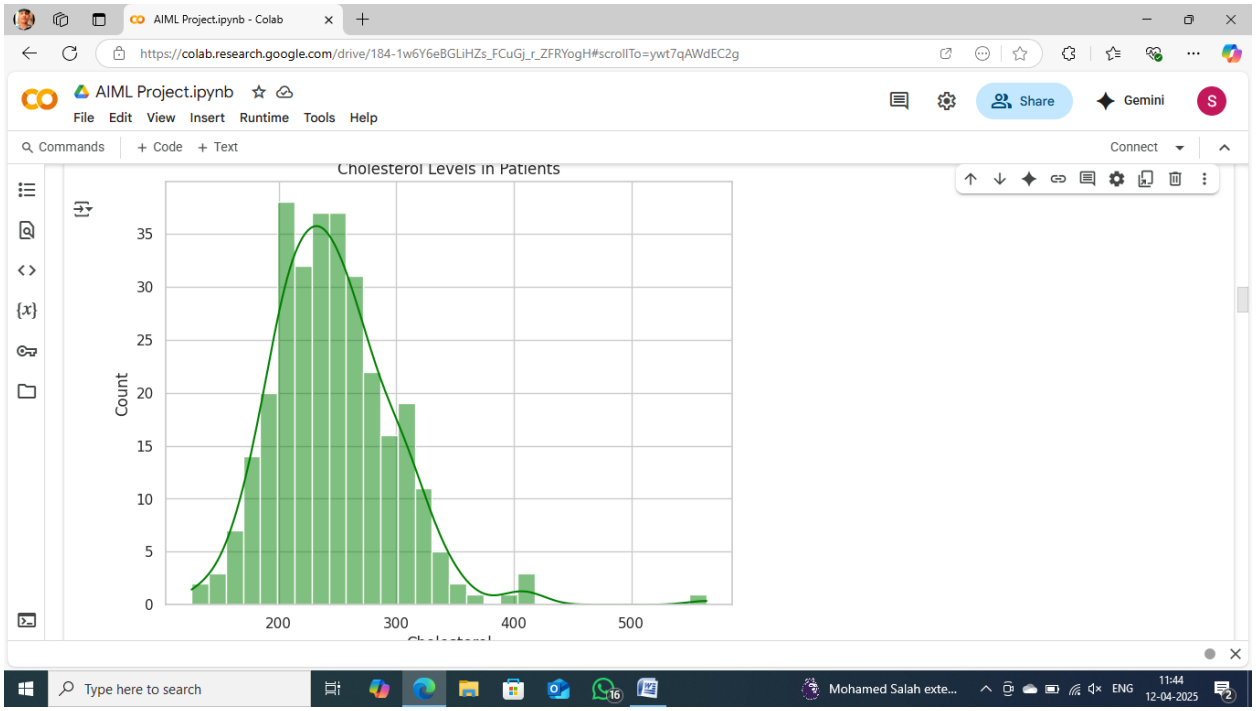
AIML Project.ipynb

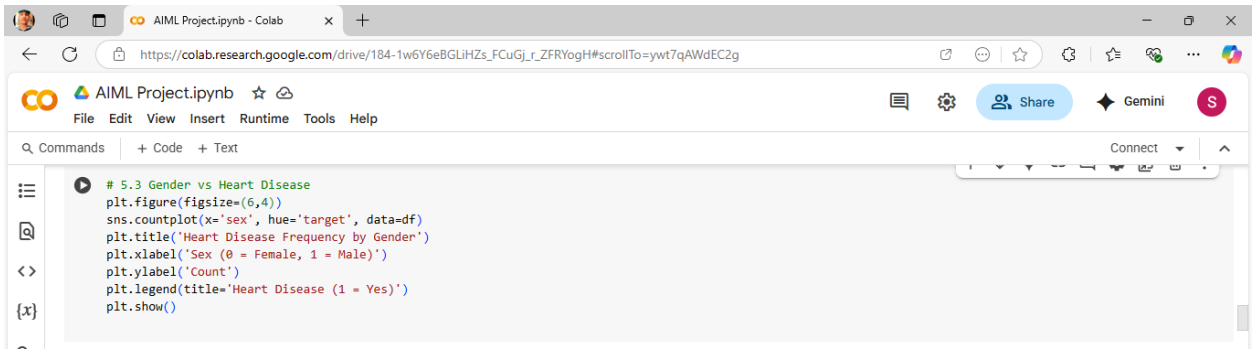
File Edit View Insert Runtime Tools Help

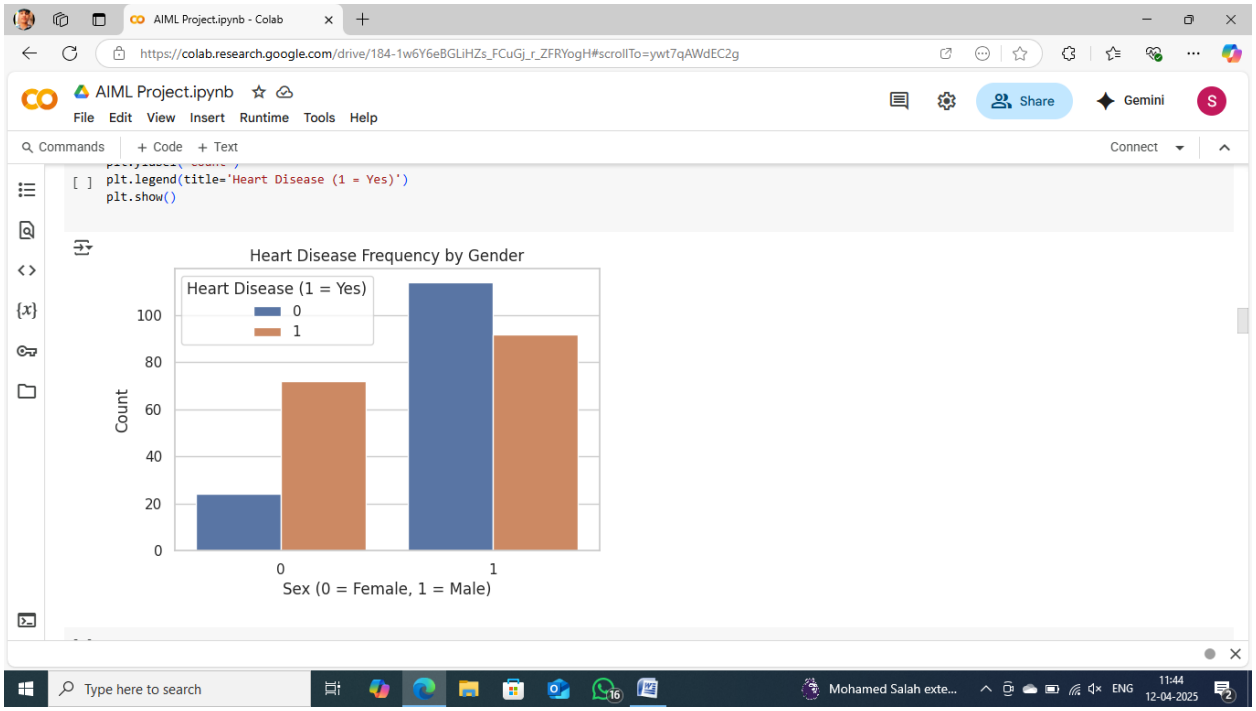
Share Gemini

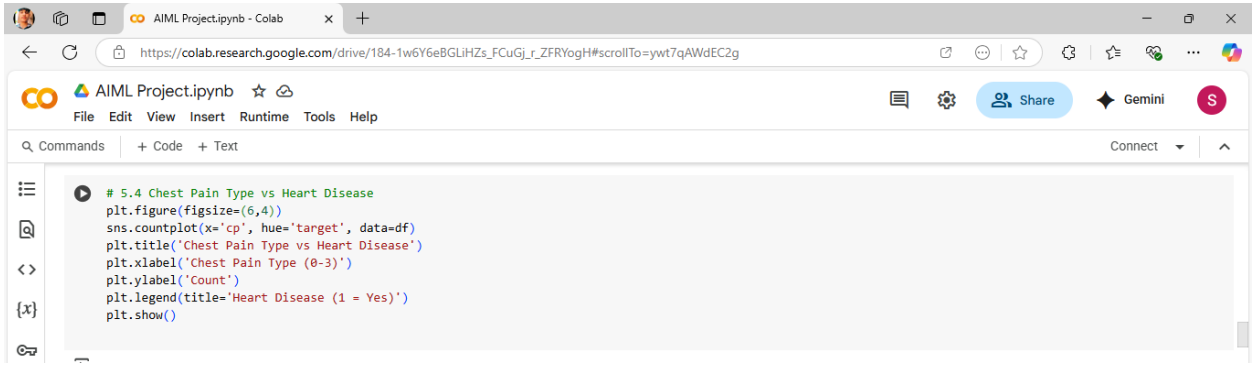
Commands + Code + Text Connect

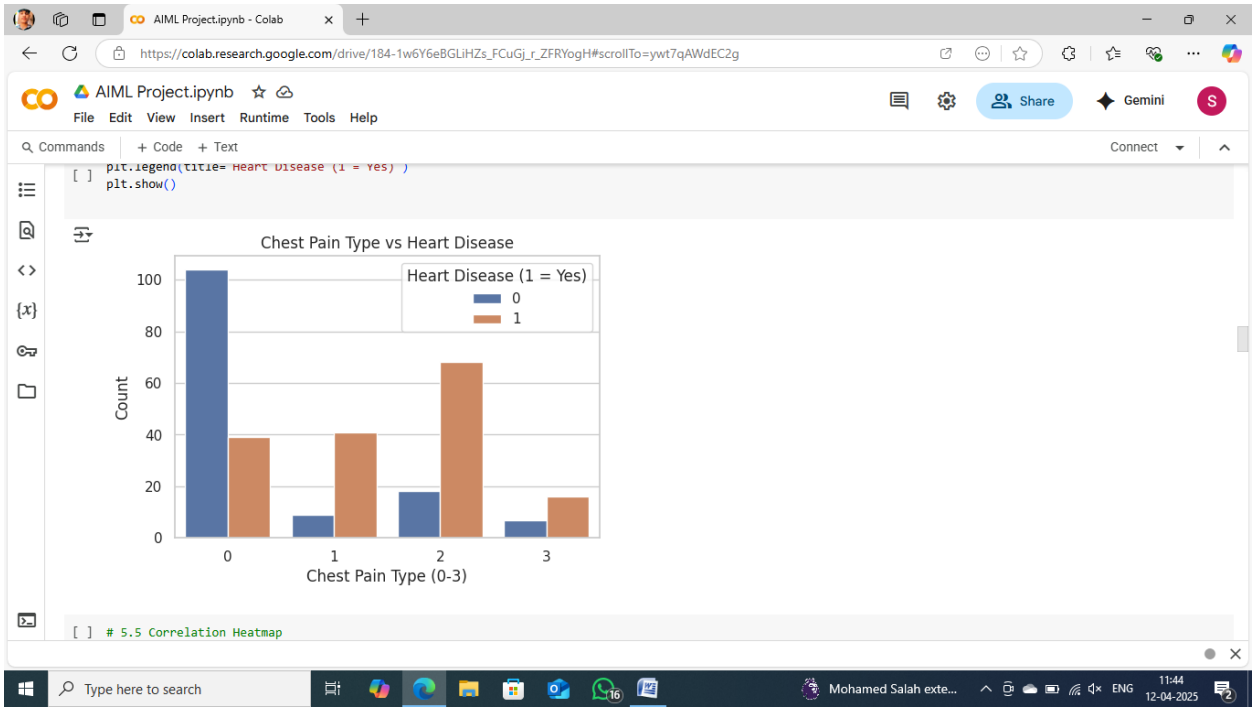
```
# 5.2 Cholesterol Distribution
plt.figure(figsize=(8,6))
sns.histplot(df['chol'], bins=30, kde=True, color='green')
plt.title('Cholesterol Levels in Patients')
plt.xlabel('Cholesterol')
plt.ylabel('Count')
plt.show()
```

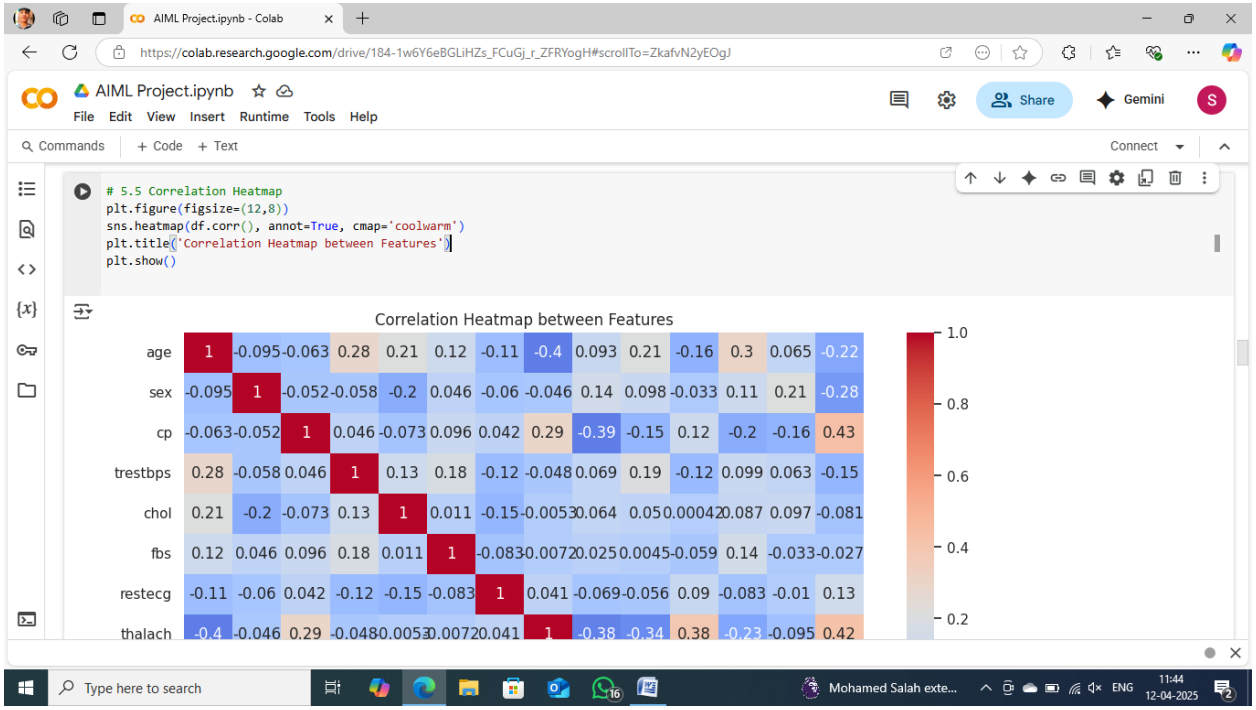

AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
Cholesterol Levels in Patients
Count
0
5
10
15
20
25
30
35
200
300
400
500
Type here to search
Mohamed Salah exte...
ENG
11:44
12-04-2025

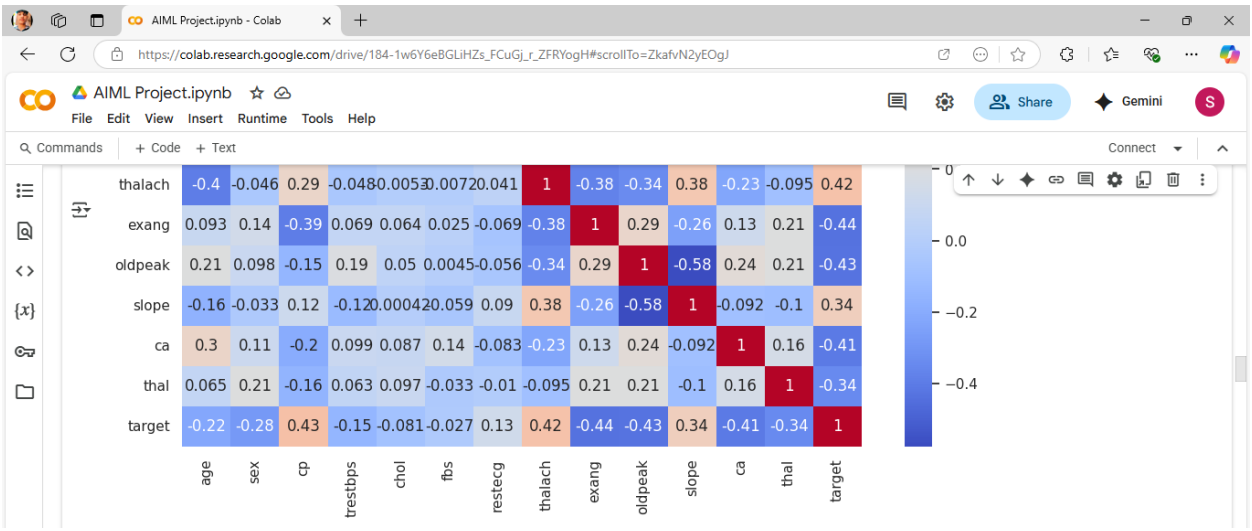

AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
# 5.3 Gender vs Heart Disease
plt.figure(figsize=(6,4))
sns.countplot(x='sex', hue='target', data=df)
plt.title('Heart Disease Frequency by Gender')
plt.xlabel('Sex (0 = Female, 1 = Male)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
[ ] plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
Heart Disease Frequency by Gender
Heart Disease (1 = Yes)
0
1
Count
100
80
60
40
20
0
0
1
Sex (0 = Female, 1 = Male)
Type here to search
Mohamed Salah exte...
ENG
11:44
12-04-2025

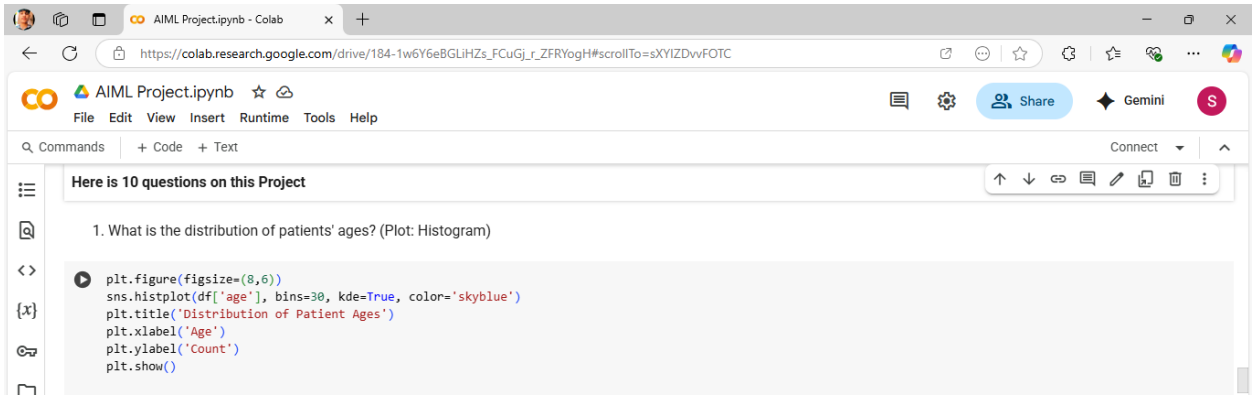

AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
# 5.4 Chest Pain Type vs Heart Disease
plt.figure(figsize=(6,4))
sns.countplot(x='cp', hue='target', data=df)
plt.title('Chest Pain Type vs Heart Disease')
plt.xlabel('Chest Pain Type (0-3)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()

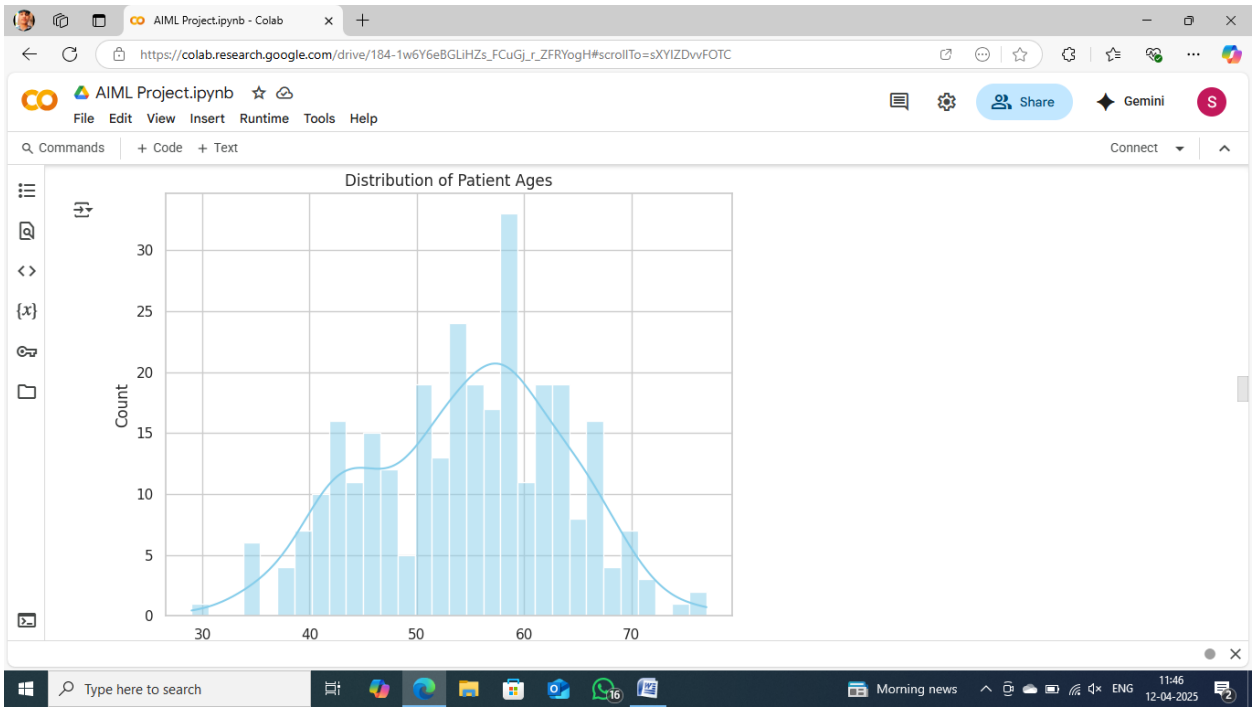

AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ywt7qAWdEC2g
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
Chest Pain Type vs Heart Disease
Heart Disease (1 = Yes)
0
1
Count
Chest Pain Type (0-3)
5.5 Correlation Heatmap
Type here to search
Mohamed Salah exte...
ENG
11:44
12-04-2025


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ZkafvN2yEOgJ
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands + Code + Text
Connect
# 5.5 Correlation Heatmap
plt.figure(figsize=(12,8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap between Features')
plt.show()
Correlation Heatmap between Features
age
sex
cp
trestbps
chol
fbs
restecg
thalach
Type here to search
Mohamed Salah exte...
ENG
11:44
12-04-2025


AIML Project.ipynb - Colab
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=ZkafvN2yEOgJ
AIML Project.ipynb
File Edit View Insert Runtime Tools Help
Share
Gemini
Commands
+ Code
+ Text
Connect
thalach -0.4 -0.046 0.29 -0.048 0.0053 0.0072 0.041 1 -0.38 -0.34 0.38 -0.23 -0.095 0.42
exang 0.093 0.14 -0.39 0.069 0.064 0.025 -0.069 -0.38 1 0.29 -0.26 0.13 0.21 -0.44
oldpeak 0.21 0.098 -0.15 0.19 0.05 0.0045 -0.056 -0.34 0.29 1 -0.58 0.24 0.21 -0.43
slope -0.16 -0.033 0.12 -0.12 0.00042 0.059 0.09 0.38 -0.26 -0.58 1 -0.092 -0.1 0.34
ca 0.3 0.11 -0.2 0.099 0.087 0.14 -0.083 -0.23 0.13 0.24 -0.092 1 0.16 -0.41
thal 0.065 0.21 -0.16 0.063 0.097 -0.033 -0.01 -0.095 0.21 0.21 -0.1 0.16 1 -0.34
target -0.22 -0.28 0.43 -0.15 -0.081 -0.027 0.13 0.42 -0.44 -0.43 0.34 -0.41 -0.34 1
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
0.0
−0.2
−0.4

10 questions based on the project :





AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

2. How does gender (sex) relate to the presence of heart disease? (Plot: Countplot)

```
plt.figure(figsize=(6,4))
sns.countplot(x='sex', hue='target', data=df, palette='Set1')
plt.title('Heart Disease Frequency by Gender')
plt.xlabel('Sex (0 = Female, 1 = Male)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
```

AIML Project.ipynb - Colab

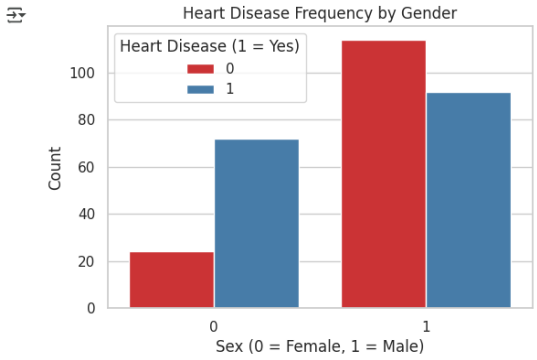
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



3. What type of chest pain is most common among patients with heart disease? (Plot: Countplot)

Type here to search

Morning news

ENG

11:46

12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

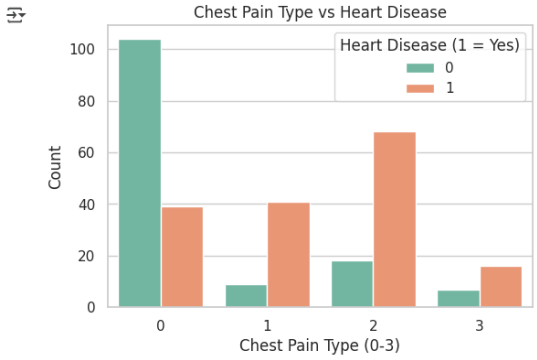
File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

3. What type of chest pain is most common among patients with heart disease? (Plot: Countplot)

```
plt.figure(figsize=(6,4))
sns.countplot(x='cp', hue='target', data=df, palette='Set2')
plt.title('Chest Pain Type vs Heart Disease')
plt.xlabel('Chest Pain Type (0-3)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
```



4. What is the distribution of cholesterol levels among patients? (Plot: Histogram)

Type here to search

Morning news

ENG

11:46

12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

4. What is the distribution of cholesterol levels among patients? (Plot: Histogram)

```
plt.figure(figsize=(8,6))
sns.histplot(df['chol'], bins=30, kde=True, color='green')
plt.title('Cholesterol Levels of Patients')
plt.xlabel('Cholesterol')
plt.ylabel('Count')
plt.show()
```

AIML Project.ipynb - Colab

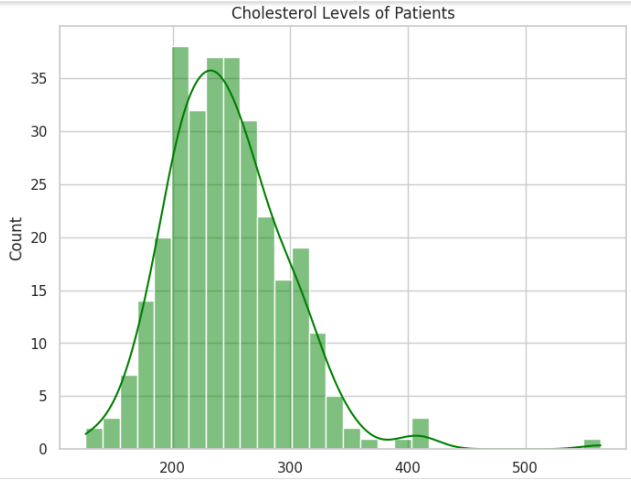
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



Type here to search

Morning news

ENG 11:46 12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

5. How does fasting blood sugar (fbs) affect heart disease? (Plot: Countplot)

```
plt.figure(figsize=(6,4))
sns.countplot(x='fbs', hue='target', data=df, palette='pastel')
plt.title('Fasting Blood Sugar > 120 mg/dl vs Heart Disease')
plt.xlabel('Fasting Blood Sugar (1 = True)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
```

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

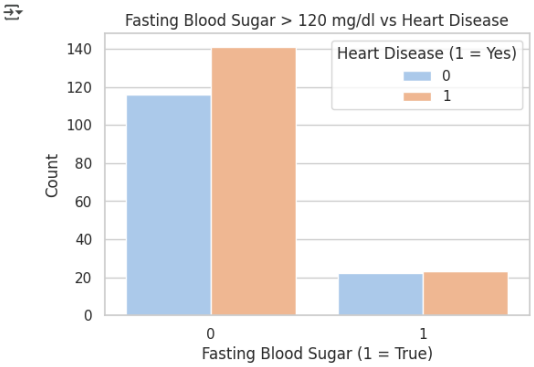
AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

plt.show()



6. What is the relationship between maximum heart rate achieved (thalach) and heart disease? (Plot: Boxplot)

Type here to search

Morning news

ENG

11:46

12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

6. What is the relationship between maximum heart rate achieved (thalach) and heart disease? (Plot: Boxplot)

```
plt.figure(figsize=(8,6))
sns.boxplot(x='target', y='thalach', data=df, palette='coolwarm')
plt.title('Max Heart Rate vs Heart Disease')
plt.xlabel('Heart Disease (1 = Yes)')
plt.ylabel('Maximum Heart Rate Achieved')
plt.show()
```

```
<ipython-input-17-fa5544c59e8c>:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same eff

  sns.boxplot(x='target', y='thalach', data=df, palette='coolwarm')
```

Max Heart Rate vs Heart Disease

AIML Project.ipynb - Colab

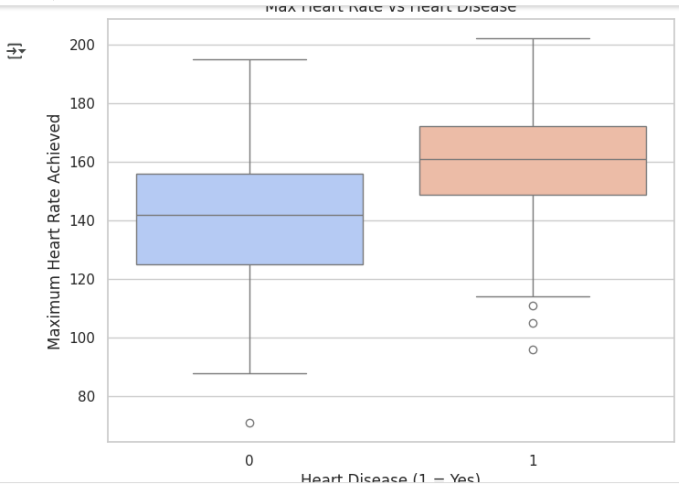
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



Type here to search

Morning news

ENG 11:46 12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

7. How is resting blood pressure (trestbps) distributed among patients? (Plot: Histogram)

```
plt.figure(figsize=(8,6))
sns.histplot(df['trestbps'], bins=30, kde=True, color='orange')
plt.title('Resting Blood Pressure of Patients')
plt.xlabel('Resting Blood Pressure (trestbps)')
plt.ylabel('Count')
plt.show()
```

AIML Project.ipynb - Colab

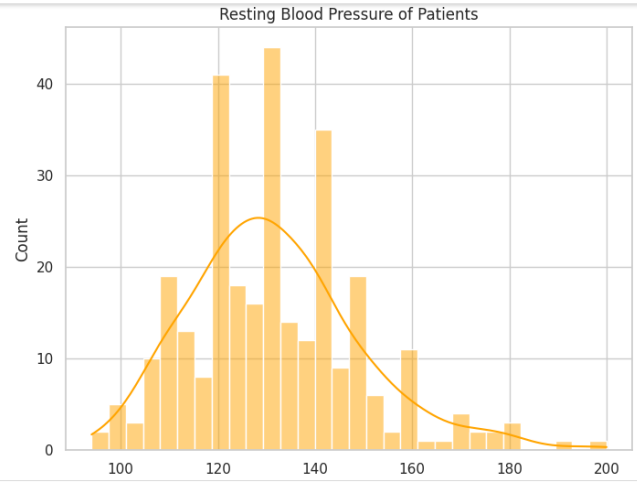
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



Type here to search

ENG 11:47 12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

8. What is the distribution of the 'slope' of the peak exercise ST segment? (Plot: Countplot)

```
plt.figure(figsize=(6,4))
sns.countplot(x='slope', hue='target', data=df, palette='viridis')
plt.title('Slope of Peak Exercise ST Segment vs Heart Disease')
plt.xlabel('Slope (0 = Upsloping, 1 = Flat, 2 = Downsloping)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
```

AIML Project.ipynb - Colab

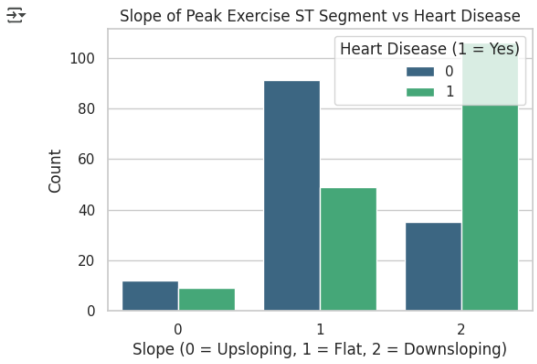
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



9. How does the number of major vessels (ca) relate to heart disease? (Plot: Countplot)

Type here to search

Virat Kohli scripts hist...

ENG 11:47 12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

9. How does the number of major vessels (ca) relate to heart disease? (Plot: Countplot)

```
plt.figure(figsize=(6,4))
sns.countplot(x='ca', hue='target', data=df, palette='magma')
plt.title('Major Vessels Colored by Fluoroscopy vs Heart Disease')
plt.xlabel('Number of Major Vessels (ca)')
plt.ylabel('Count')
plt.legend(title='Heart Disease (1 = Yes)')
plt.show()
```

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

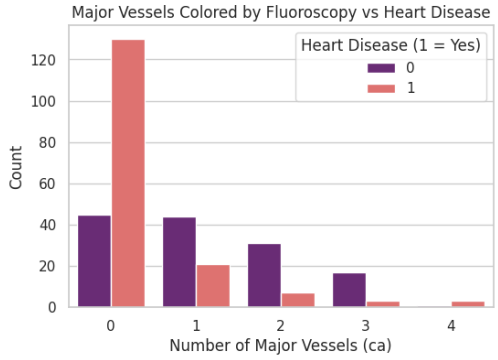
AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect

plt.show()



10. What are the strongest feature correlations with heart disease? (Plot: Heatmap)

Type here to search

Virat Kohli scripts hist...

ENG 11:47 12-04-2025

AIML Project.ipynb - Colab

https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

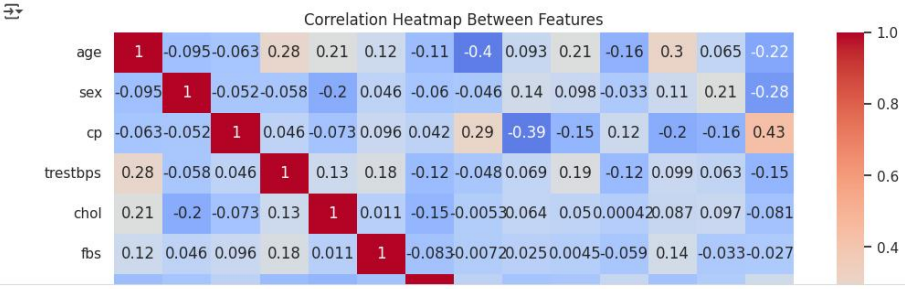
Share Gemini

Commands + Code + Text Connect

10. What are the strongest feature correlations with heart disease? (Plot: Heatmap)

python Copy Edit

```
plt.figure(figsize=(12,8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap Between Features')
plt.show()
```



Type here to search Virat Kohli scripts hist... ENG 11:47 12-04-2025

AIML Project.ipynb - Colab

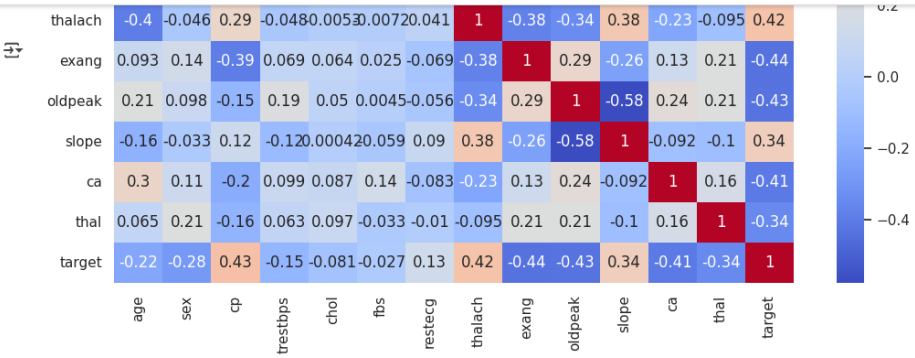
https://colab.research.google.com/drive/184-1w6Y6eBGLiHZs_FCuGj_r_ZFRYogH#scrollTo=sXYIZDvvFOTC

AIML Project.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text Connect



Dataset Information and Github Repository Link :

- **Dataset Type:** It's a **heart disease dataset** — typically used for predicting the presence or absence of heart disease in a patient based on medical attributes.
- **Attributes/Columns include:** (from the common `heart.csv` dataset structure)
  - `age` – Age of the person
  - `sex` – Gender (1 = male, 0 = female)
  - `cp` – Chest pain type (4 values)
  - `trestbps` – Resting blood pressure
  - `chol` – Serum cholesterol in mg/dl
  - `fbs` – Fasting blood sugar > 120 mg/dl
  - `restecg` – Resting electrocardiographic results (0, 1, 2)
  - `thalach` – Maximum heart rate achieved
  - `exang` – Exercise induced angina
  - `oldpeak` – ST depression induced by exercise
  - `slope` – Slope of the peak exercise ST segment
  - `ca` – Number of major vessels colored by fluoroscopy
  - `thal` – Thalassemia
  - `target` – Target variable (1 = heart disease present, 0 = not present)
- **Initial Steps Done in Notebook:**
  - Imported libraries (`pandas`, `numpy`, `matplotlib`, `seaborn`).
  - Loaded the dataset.
  - Displayed basic info and statistical summary.
  - Checked for missing values and duplicates.
  - Visualized data with boxplots.

Github Project Link :

https://github.com/sachinswami00/Heart-Disease

CONCLUSION & FUTURE SCOPE

Through this comprehensive project, we explored and visualized key factors associated with heart disease using a publicly available dataset.

Key Insights:

- **Age Distribution:** Heart disease is more common in middle-aged and older individuals, particularly those above 50 years.

- **Gender Impact:** Males were found to have a slightly higher risk of heart disease compared to females in this dataset.
- **Chest Pain Type:** The type of chest pain (especially typical angina and asymptomatic pain) is strongly linked with heart disease.
- **Cholesterol Levels:** High cholesterol levels were common among patients, although there were some natural outliers.
- **Fasting Blood Sugar:** Fasting blood sugar over 120 mg/dl had some association but was not the strongest indicator alone.
- **Heart Rate:** Patients with heart disease generally exhibited lower maximum heart rates during exercise.
- **Resting Blood Pressure:** Although variations exist, extremely high blood pressure was less common, suggesting blood pressure alone may not be a strong predictor.
- **Slope and Major Vessels:** A flat slope in the ST segment and more major vessels colored by fluoroscopy were correlated with a higher likelihood of heart disease.
- **Feature Correlations:** Strongest correlations with heart disease were found in features like chest pain type (cp), maximum heart rate achieved (thalach), number of major vessels (ca), and exercise-induced angina .

Project Summary: We used different visualization techniques (histograms, countplots, boxplots, heatmaps) to uncover hidden patterns in the data. This exploratory data analysis helps in understanding which features are important for predicting heart disease. It also lays the foundation for building predictive machine learning models in the future.

Future Recommendations:

- Apply machine learning models like Logistic Regression, Random Forest, or XGBoost to predict heart disease.
- Use feature engineering to improve model performance.
- Explore more recent and diverse datasets for better generalization across different populations.

In conclusion, data-driven insights like these are critical in early diagnosis and prevention strategies for heart disease, ultimately helping save lives.

Future Scope:

- **Model Improvement:** Experiment with advanced models like XGBoost, LightGBM, or deep learning techniques for potentially higher accuracy.
- **Feature Engineering:** Incorporate additional features such as lifestyle habits, genetic data, and medication history to enhance model performance.
- **Deployment:** Develop a user-friendly web or mobile application to make predictions accessible to users and healthcare providers.
- **Real-time Data:** Integrate real-time patient monitoring data for continuous health assessment.