

```
ufprlooma dgeodo g=l ef.icloelsa.bu pilmopaodr(t) files
uploaded = files.upload()
```

Choose Files heart.xlsx

heart.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 57094 bytes, last modified: 3/6/2026 - 100% done
Saving heart.xlsx to heart (2).xlsx

```
# استيراد المكتبات
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# قراءة البيانات
df = pd.read_excel("heart.xlsx")

# عرض أول 5 صفوف
print(df.head())

# معلومات عن البيانات
print(df.info())

# -----

# Missing Values التحقق من
print("Missing Values:")
print(df.isnull().sum())

# لو فيه قيم مفقودة نعوضها بالمتوسط
df = df.fillna(df.mean())

# -----

# IQR باستخدام Outliers اكتشاف
```

```
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1

# إزالة القيم الشاذة
df_clean = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]

print("Shape after removing outliers:", df_clean.shape)

# -----

# Data Visualization

# Plot 1
plt.figure()
sns.histplot(df_clean['age'], kde=True)
plt.title("Age Distribution")
plt.show()

# Plot 2
plt.figure()
sns.boxplot(x=df_clean['chol'])
plt.title("Cholesterol Outliers")
plt.show()

# Plot 3
plt.figure()
sns.countplot(x=df_clean['target'])
plt.title("Heart Disease Count")
plt.show()

# Plot 4 (زيادة للتأكيد)
plt.figure()
sns.scatterplot(x='age', y='chol', data=df_clean)
plt.title("Age vs Cholesterol")
plt.show()
```

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

	ca	thal	target
0	2	3	0
1	0	3	0
2	0	3	0
3	1	3	0
4	3	2	0

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):

#	Column	Non-Null Count	Dtype
-----	-----	-----	-----
0	age	1025 non-null	int64
1	sex	1025 non-null	int64
2	cp	1025 non-null	int64
3	trestbps	1025 non-null	int64
4	chol	1025 non-null	int64
5	fbs	1025 non-null	int64
6	restecg	1025 non-null	int64
7	thalach	1025 non-null	int64
8	exang	1025 non-null	int64
9	oldpeak	1025 non-null	float64
10	slope	1025 non-null	int64
11	ca	1025 non-null	int64
12	thal	1025 non-null	int64
13	target	1025 non-null	int64

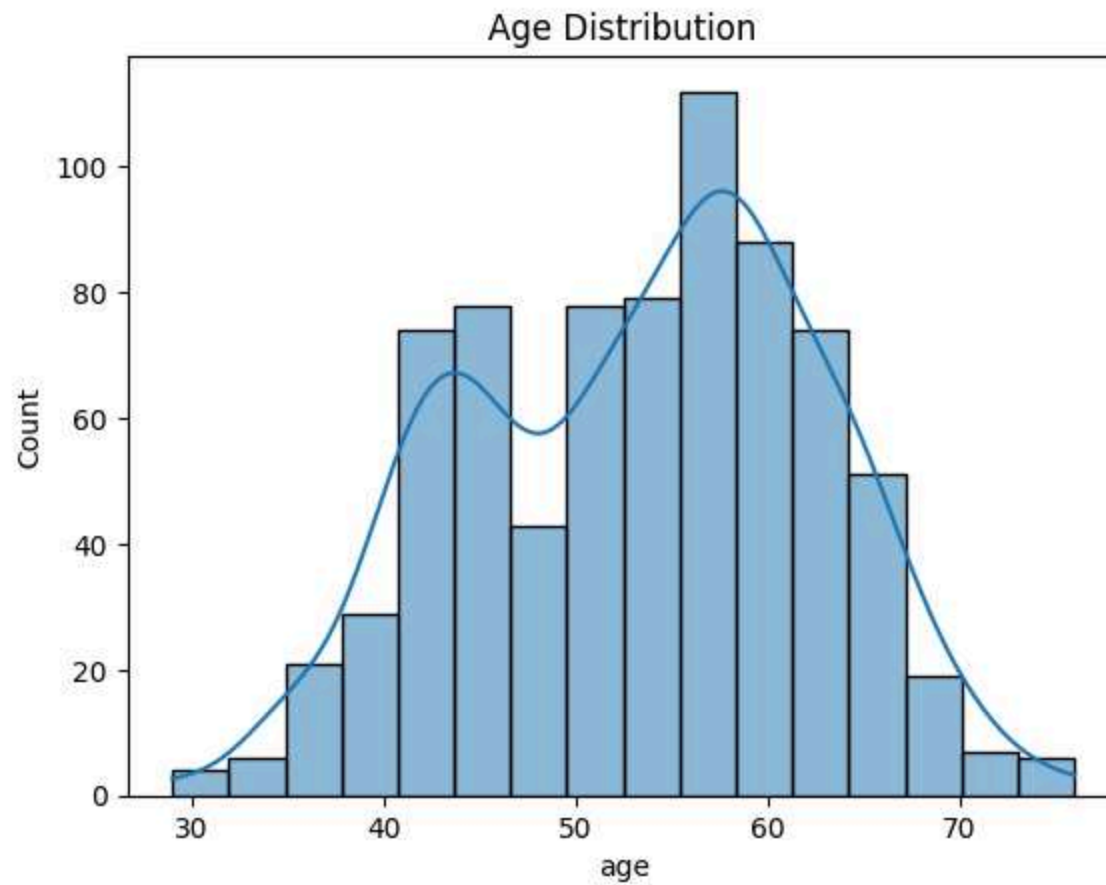
dtypes: float64(1), int64(13)
memory usage: 112.2 KB
None

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
```

```
Shape after removing outliers: (769, 14)
```



Cholesterol Outliers

