

## ✓ Importing pandas library

```
import pandas as pd
```

## ✓ Reading dataset

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

```
df
# 303 rows x 14 columns
```

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows x 14 columns

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

## ✓ Cleaning dataset

```
df = df.drop_duplicates()
```

```
df
```

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

302 rows x 14 columns

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

## ✓ Duplicates removed

```
df.isna().sum()
# No null values, it's clean
```

```
↵
```

	0
age	0
sex	0
cp	0
trtbps	0
chol	0
fbs	0
restecg	0
thalachh	0
exng	0
oldpeak	0
slp	0
caa	0
thall	0
output	0

```
df.dtypes
```

## Visualization plot

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df.columns
```

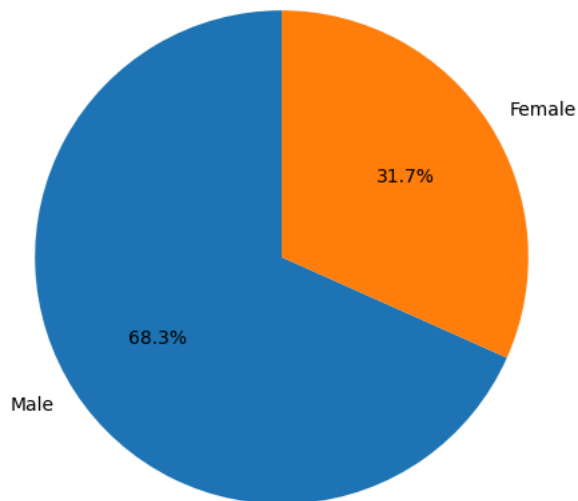
```
↵ Index(['age', 'sex', 'cp', 'trtbps', 'chol', 'fbs', 'restecg', 'thalachh',
        'exng', 'oldpeak', 'slp', 'caa', 'thall', 'output'],
        dtype='object')
```

## Pie graph

```
sex_counts = df['sex'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(sex_counts, labels=['Male', 'Female'], autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Sex')
plt.show()
```

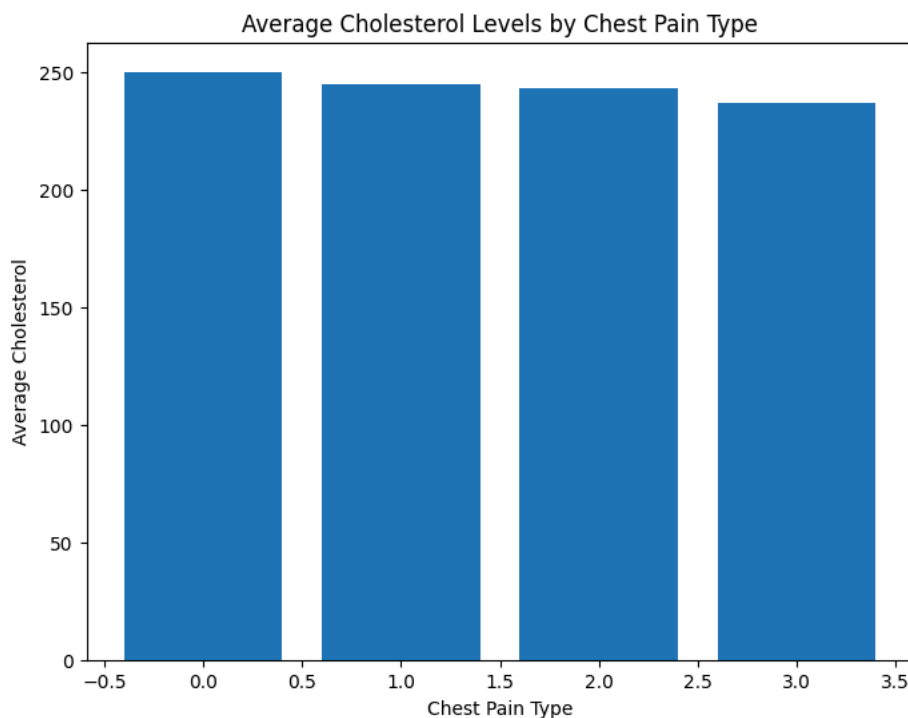


Distribution of Sex



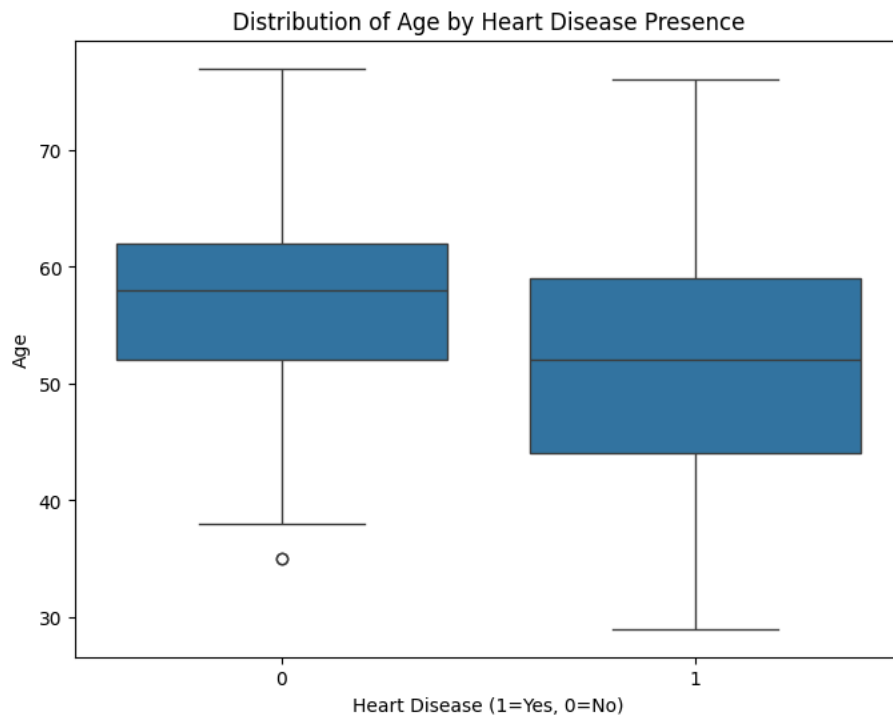
## Bar plot

```
avg_chol_by_cp = df.groupby('cp')['chol'].mean()
plt.figure(figsize=(8, 6))
plt.bar(avg_chol_by_cp.index, avg_chol_by_cp.values)
plt.xlabel('Chest Pain Type')
plt.ylabel('Average Cholesterol')
plt.title('Average Cholesterol Levels by Chest Pain Type')
plt.show()
```



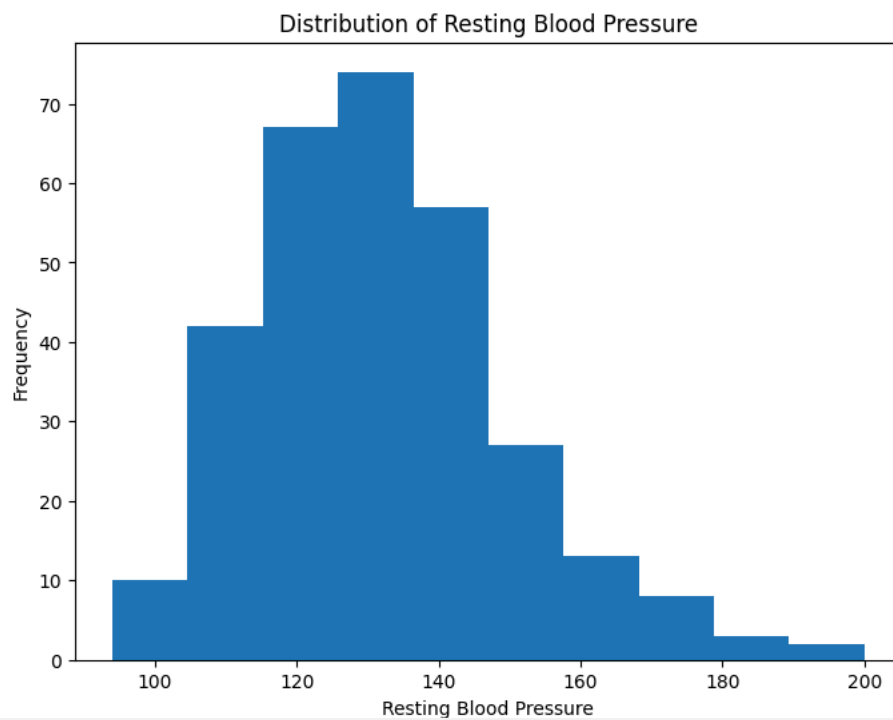
## Box plot

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='output', y='age', data=df)
plt.xlabel('Heart Disease (1=Yes, 0=No)')
plt.ylabel('Age')
plt.title('Distribution of Age by Heart Disease Presence')
plt.show()
```



## ✓ Histogram

```
plt.figure(figsize=(8, 6))
plt.hist(df['trtbps'], bins=10)
plt.xlabel('Resting Blood Pressure')
plt.ylabel('Frequency')
plt.title('Distribution of Resting Blood Pressure')
plt.show()
```



## ✓ Scatter plot

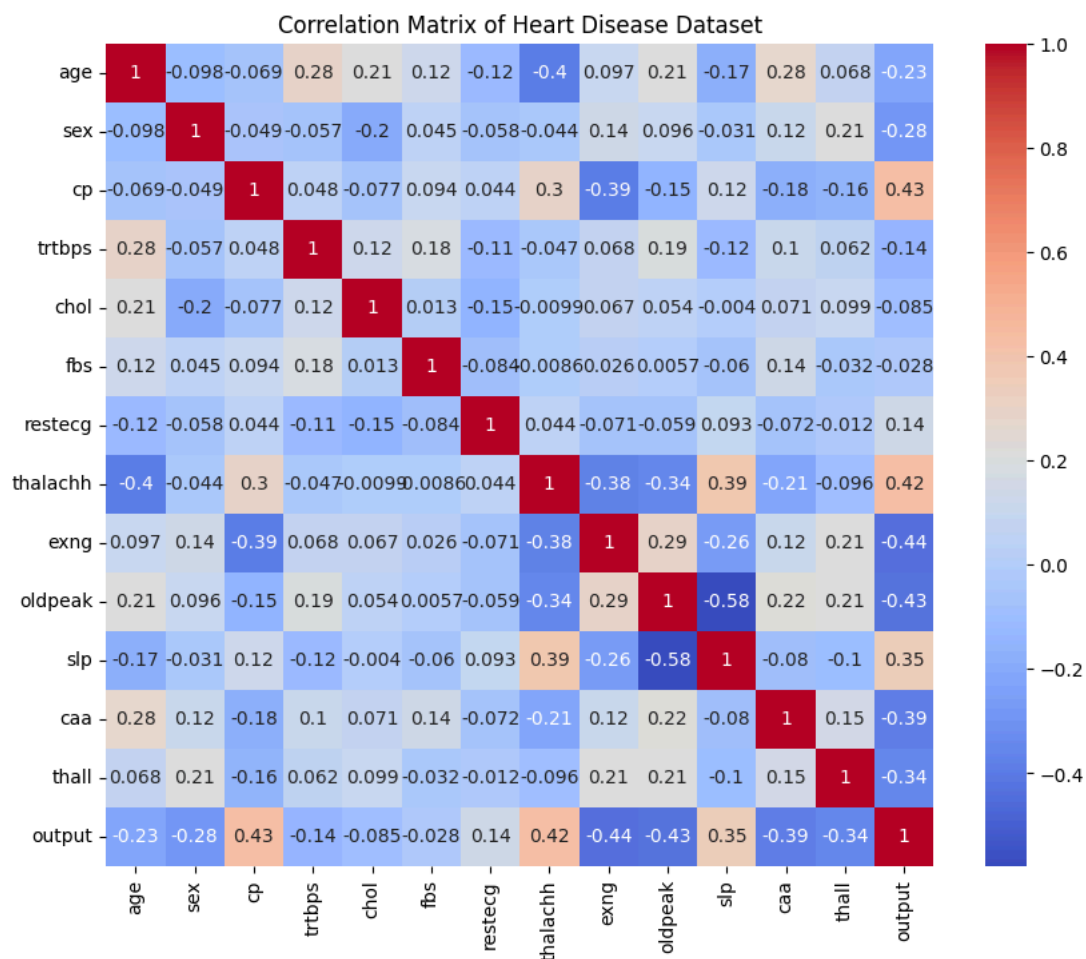
```
plt.figure(figsize=(8, 6))
plt.scatter(df['age'], df['thalachh'])
plt.xlabel('Age')
plt.ylabel('Maximum Heart Rate')
```

```
plt.title('Relationship between Age and Maximum Heart Rate')
plt.show()
```



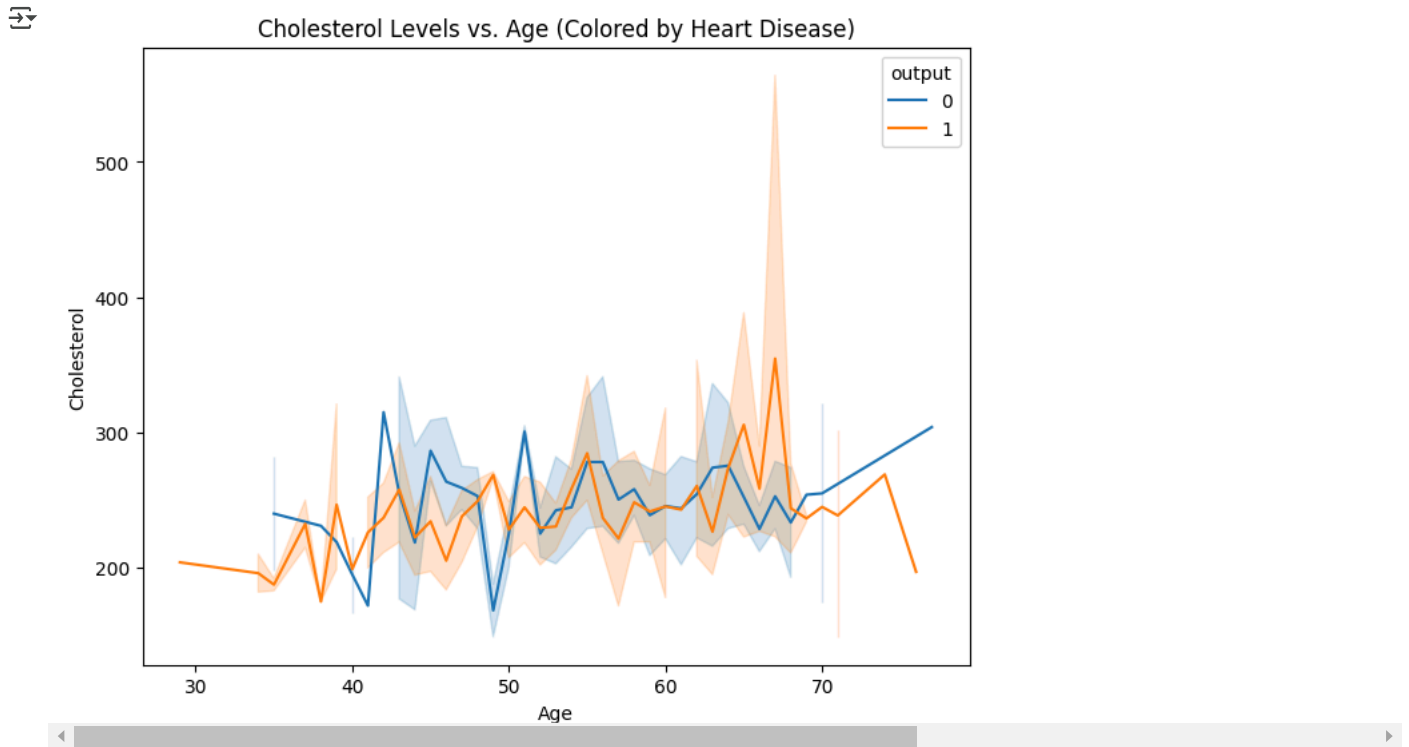
## Heat plot

```
plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix of Heart Disease Dataset')
plt.show()
```



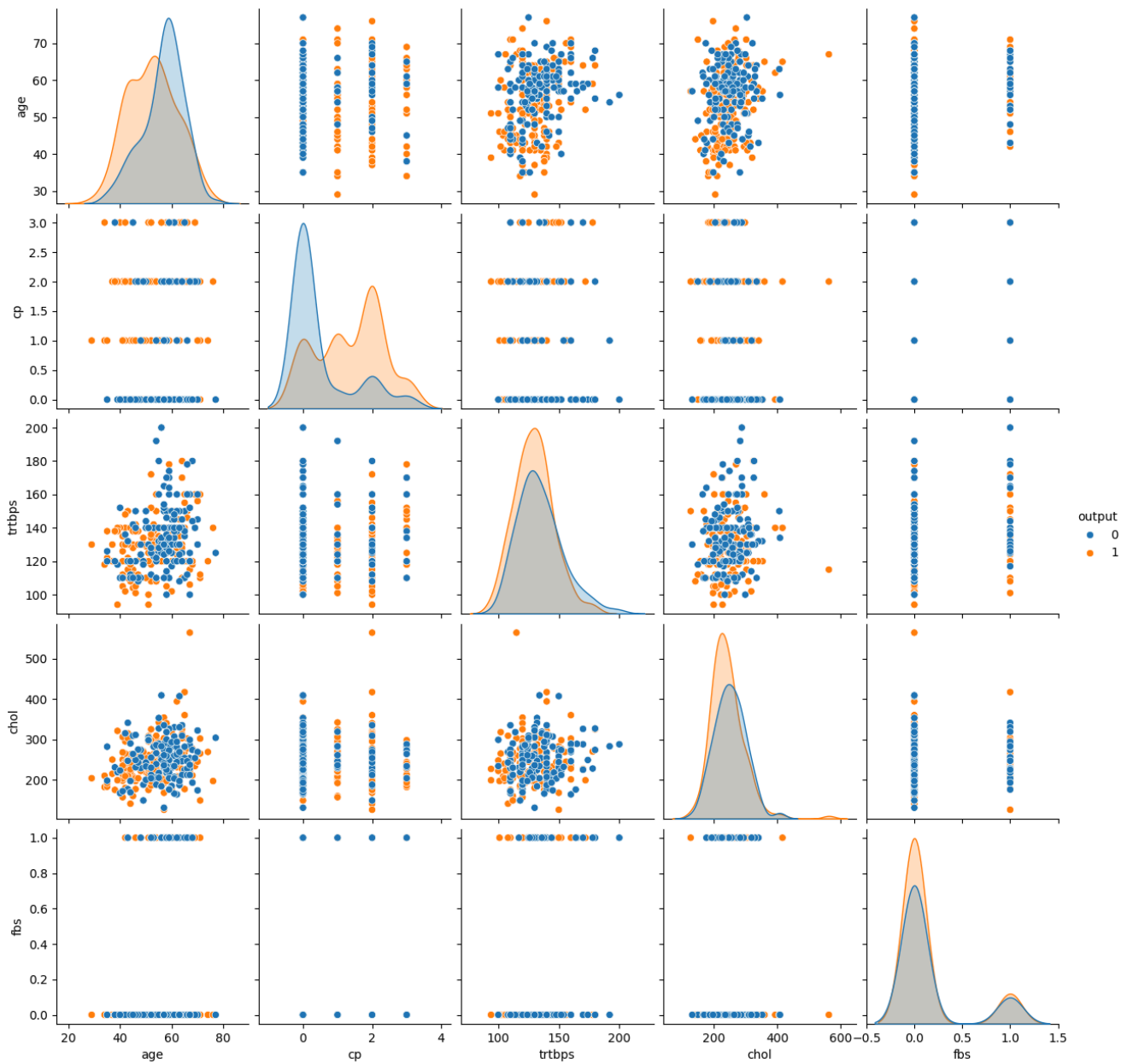
## Line plot

```
plt.figure(figsize=(8, 6))
sns.lineplot(data=df, x='age', y='chol', hue='output')
plt.xlabel('Age')
plt.ylabel('Cholesterol')
plt.title('Cholesterol Levels vs. Age (Colored by Heart Disease)')
plt.show()
```



## Pair plot

```
temp_df = df[['age', 'cp', 'trtbps', 'chol', 'fbs', 'output']]
sns.pairplot(temp_df, hue='output')
plt.show()
```



## Word Cloud

Generated code may be subject to a license | Km103/nlp\_project

```
!pip install wordcloud
```

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

```
with open('samplewords.txt', 'r') as f:
    text = f.read()
```