

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
data = pd.read_csv("/content/heart.csv")
```

```
data.tail()
```

```
↗
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

```
data.columns.values
```

```
↗ array(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg',
        'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
        dtype=object)
```

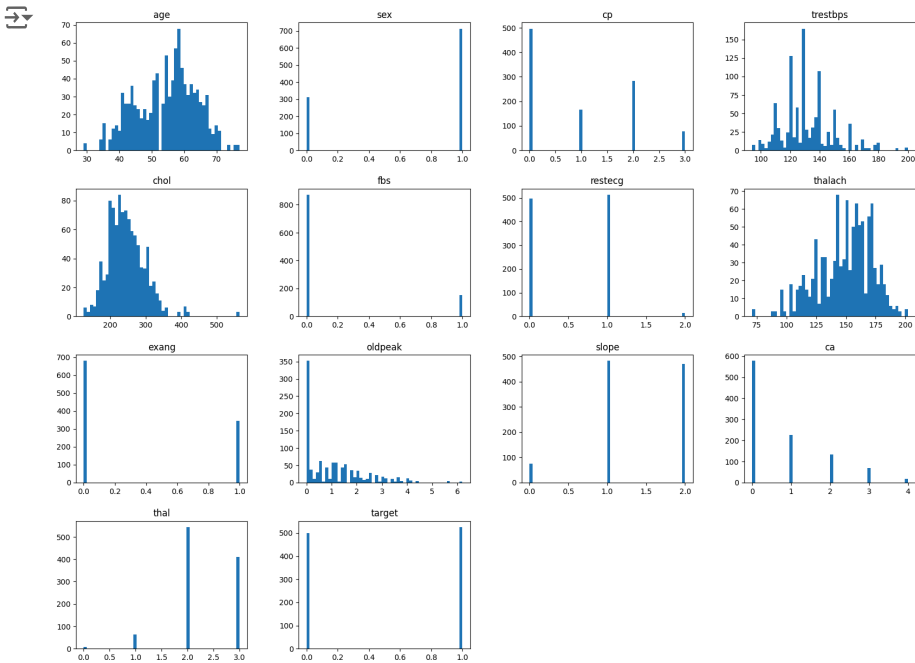
```
data.isna().sum()
```

```
↗ 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
```

```
data.info()
```

```
↗ <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
```

```
data.hist(bins = 50, grid = False, figsize=(20,15));
```



```
data.describe()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000	1025.000000
mean	54.434146	0.695610	0.942439	131.611707	246.000000	0.149268	0.356527	130.000000	0.000000	0.460373	1.000000	0.000000	150.000000	0.000000
std	9.072290	0.460373	1.029641	17.516718	51.59251	0.356527	0.460373	27.516718	0.000000	0.460373	1.000000	0.000000	27.516718	0.000000
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	75.000000	0.000000	0.000000	0.000000	0.000000	75.000000	0.000000
25%	48.000000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	120.000000	0.000000	0.000000	0.000000	0.000000	120.000000	0.000000
50%	56.000000	1.000000	1.000000	130.000000	240.000000	0.000000	0.000000	130.000000	0.000000	0.000000	1.000000	0.000000	130.000000	0.000000
75%	61.000000	1.000000	2.000000	140.000000	275.000000	0.000000	0.000000	140.000000	0.000000	0.000000	2.000000	0.000000	140.000000	0.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.000000	2.000000	4.000000	202.000000	1.000000

```

questions =["1. How many have heart disease and how many people doesn't have heart disesease? ",
            "2. People of which sex has most heart disease?",
            "3. People of which sex has which type of chest pain most?",
            "4. People with chest pain are most pron to have heart disease?",
            "5. How does fasting blood sugar (fbs) relate to heart disease occurrence?",
            "6. How do the number of major vessels colored by fluoroscopy (ca) compare for people with and wi",
            "7. What is the age distribution of people with and without heart disease?"
            ]
questions

```

```

["1. How many have heart disease and how many people doesn't have heart disesease? ",
 '2. People of which sex has most heart disease?',
 '3. People of which sex has which type of chest pain most?',
 '4. People with chest pain are most pron to have heart disease?',
 '5. How does fasting blood sugar (fbs) relate to heart disease occurrence?',
 '6. How do the number of major vessels colored by fluoroscopy (ca) compare for people with and without heart disease?',
 '7. What is the age distribution of people with and without heart disease?']

```

```

# 1. How many have heart disease and how many people doesn't have heart disesease?
data.target.value_counts()

```

```

target
1    526
0    499
Name: count, dtype: int64

```

```

data.target.value_counts().plot(kind = "bar", color=["orchid","salmon"])
plt.title("Heart disease values")
plt.xlabel("1 = Heart disease, 0= No Heart disease")
plt.ylabel("Amount");

```



```

#2. People of which sex has most heart disease?
pd.crosstab(data.target,data.sex)

```

```

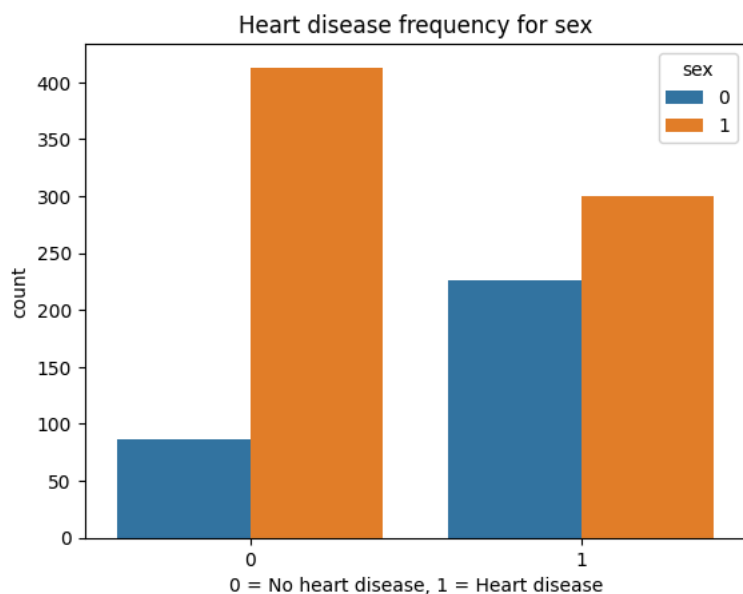
sex    0    1
target
0      86  413
1     226  300

```

```

sns.countplot(x= "target", data=data, hue= "sex")
plt.title("Heart disease frequency for sex")
plt.xlabel("0 = No heart disease, 1 = Heart disease");

```



#3. People of which sex has which type of chest pain most?

`data.cp.value_counts()`

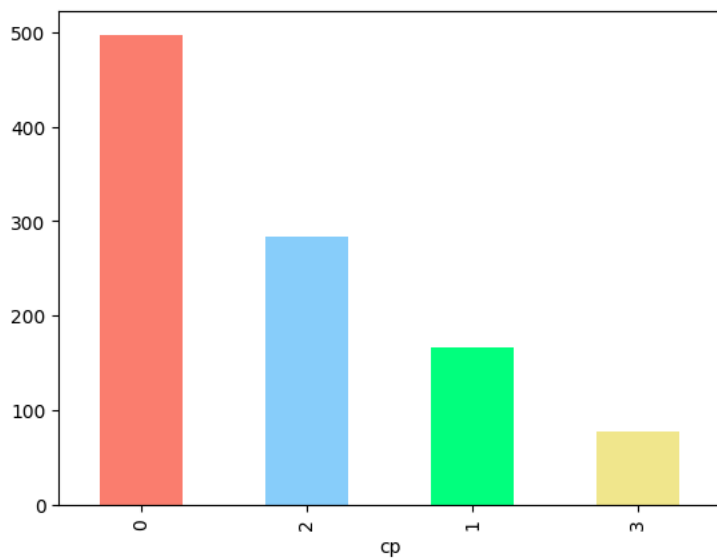


```
cp
0    497
2    284
1    167
3     77
Name: count, dtype: int64
```

`data.cp.value_counts().plot(kind = "bar",color = ["salmon", "lightskyblue", "springgreen","khaki"])`



<Axes: xlabel='cp'>

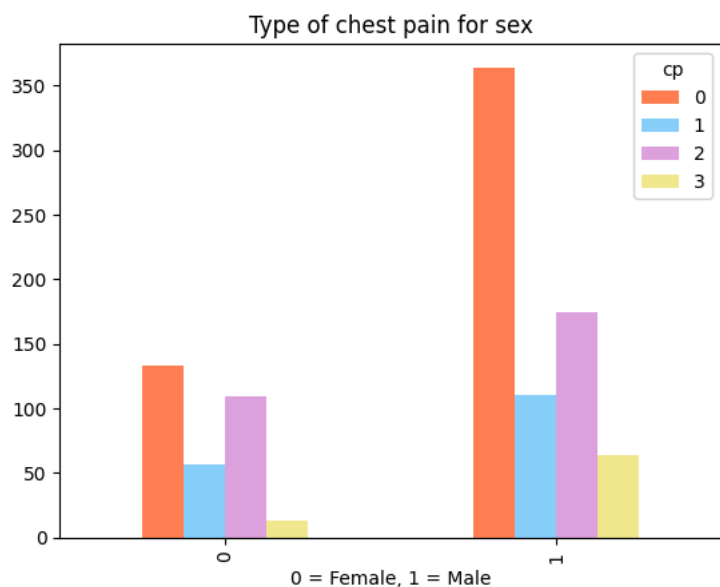


`pd.crosstab(data.sex,data.cp)`



cp	0	1	2	3
sex				
0	133	57	109	13
1	364	110	175	64

```
pd.crosstab(data.sex,data.cp).plot(kind= "bar", color = ["coral","lightskyblue","plum","khaki"])
plt.title("Type of chest pain for sex")
plt.xlabel("0 = Female, 1 = Male");
```



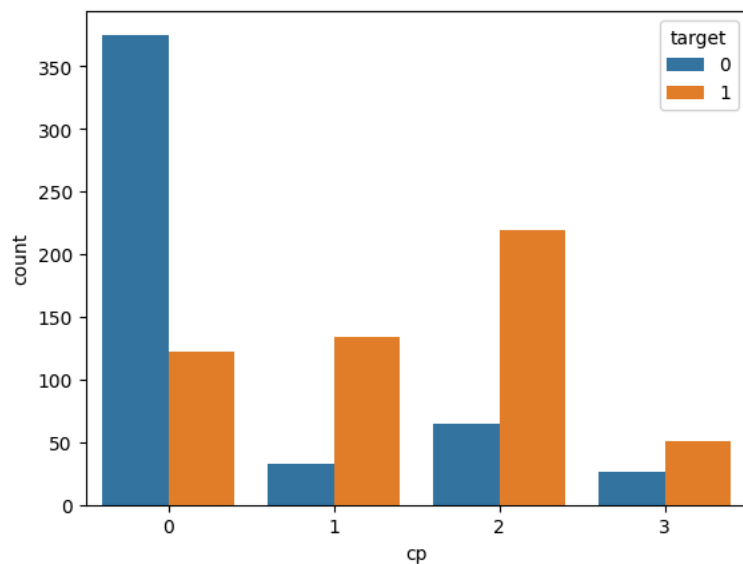
#4. People with chest pain are most pron to have heart disease?

```
pd.crosstab(data.cp,data.target)
```



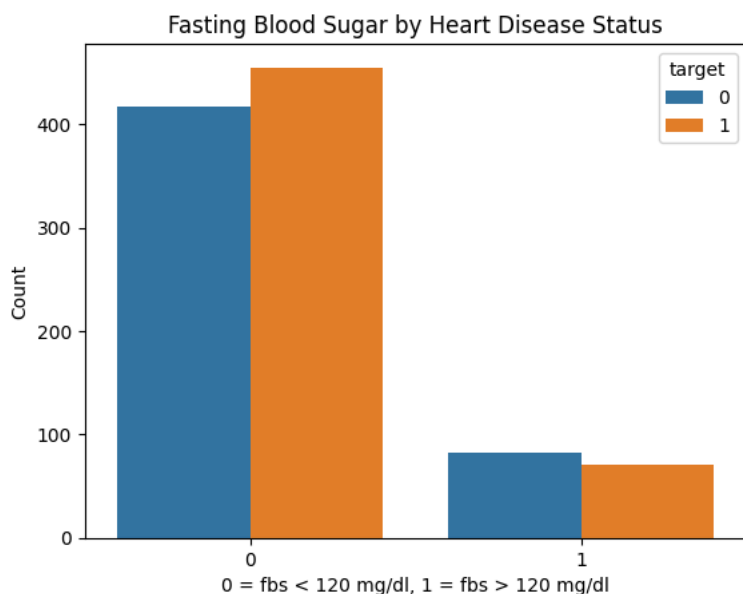
target	0	1
cp		
0	375	122
1	33	134
2	65	219
3	26	51

```
sns.countplot(x="cp", data = data, hue= "target");
```



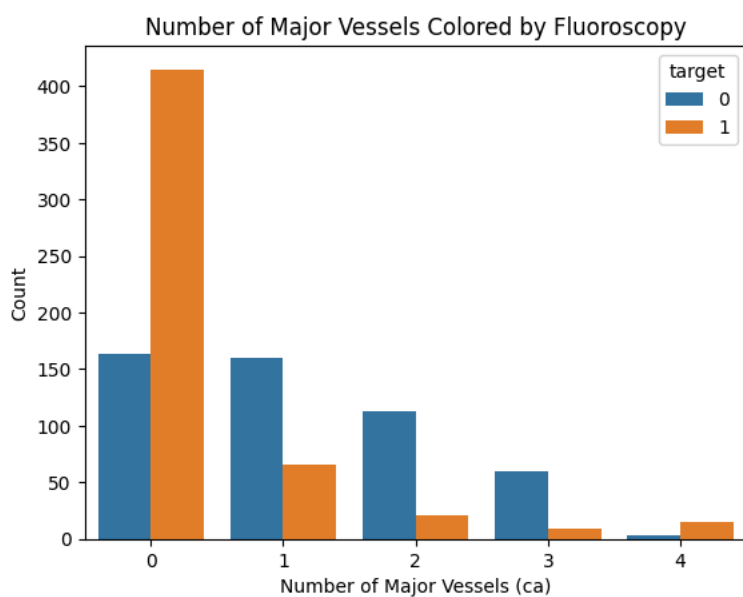
#5.How does fasting blood sugar (fbs) relate to heart disease occurrence?

```
sns.countplot(x="fbs", data=data, hue="target")
plt.title("Fasting Blood Sugar by Heart Disease Status")
plt.xlabel("0 = fbs < 120 mg/dl, 1 = fbs > 120 mg/dl")
plt.ylabel("Count")
plt.show()
```



#6.How do the number of major vessels colored by fluoroscopy (ca) compare for people with and without heart disease?

```
sns.countplot(x="ca", data=data, hue="target")  
plt.title("Number of Major Vessels Colored by Fluoroscopy")  
plt.xlabel("Number of Major Vessels (ca)")  
plt.ylabel("Count")  
plt.show()
```

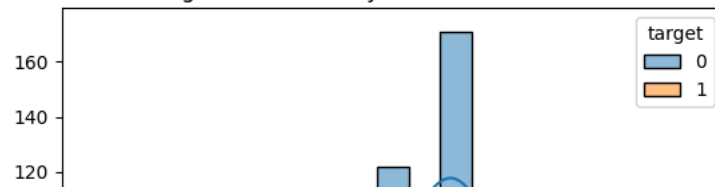


#7.What is the age distribution of people with and without heart disease?

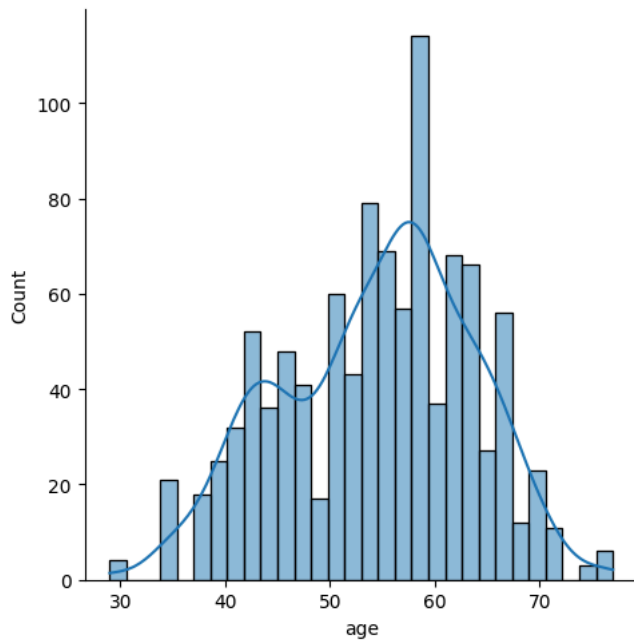
```
sns.histplot(data=data, x="age", hue="target", multiple="stack", kde=True)  
plt.title("Age Distribution by Heart Disease Status")  
plt.xlabel("Age")  
plt.ylabel("Count")  
plt.show()
```



Age Distribution by Heart Disease Status



```
sns.displot(x="age", data = data, bins = 30, kde= True);
```



```
sns.displot(x="thalach", data = data, bins = 30, kde = True, color = "chocolate");
```

