

```
In [60]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import os
```

```
In [61]: df=pd.read_csv(r"C:\Users\Prakash Enerprener\OneDrive\Desktop\data science\Heart Disease Prediction with Logist
```

```
In [62]: df
```

```
Out[62]:
```

	age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	2
2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0
...
298	45	1	1	110	264	0	0	132	0	1.2	2	0.0	7.0	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2.0	7.0	2
300	57	1	4	130	131	0	0	115	1	1.2	2	1.0	7.0	3
301	57	0	2	130	236	0	2	174	0	0.0	2	1.0	3.0	1
302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0

303 rows × 14 columns

```
In [63]: import pandas as pd

# Assuming df is your DataFrame containing the heart dataset
# Convert heart_diagnosis column to binary format
df['heart_diagnosis'] = df['heart_diagnosis'].apply(lambda x: 1 if x > 0 else 0)

# Alternatively, you can use a map function:
# df['heart_diagnosis'] = df['heart_diagnosis'].map(lambda x: 1 if x > 0 else 0)

# Print the first few rows to verify the conversion
print(df.head())
```

	age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	\
0	63	1	1	145	233	1	2	150	0	2.3	
1	67	1	4	160	286	0	2	108	1	1.5	
2	67	1	4	120	229	0	2	129	1	2.6	
3	37	1	3	130	250	0	0	187	0	3.5	
4	41	0	2	130	204	0	2	172	0	1.4	

	slope	ca	thal	heart_diagnosis
0	3	0.0	6.0	0
1	2	3.0	3.0	1
2	2	2.0	7.0	1
3	3	0.0	3.0	0
4	1	0.0	3.0	0

```
In [64]: df=pd.DataFrame(df)
```

```
In [65]: df.head()
```

```
Out[65]:
```

	age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0

```
In [66]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   age                    303 non-null   int64
1   gender                 303 non-null   int64
2   cp                     303 non-null   int64
3   trestbps               303 non-null   int64
4   chol                  303 non-null   int64
5   fbs                   303 non-null   int64
6   restecg               303 non-null   int64
7   thalach               303 non-null   int64
8   exang                 303 non-null   int64
9   oldpeak               303 non-null   float64
10  slope                 303 non-null   int64
11  ca                    299 non-null   float64
12  thal                  301 non-null   float64
13  heart_diagnosis       303 non-null   int64
dtypes: float64(3), int64(11)
memory usage: 33.3 KB
```

```
In [67]: len(df)
```

```
Out[67]: 303
```

```
In [68]: df.columns
```

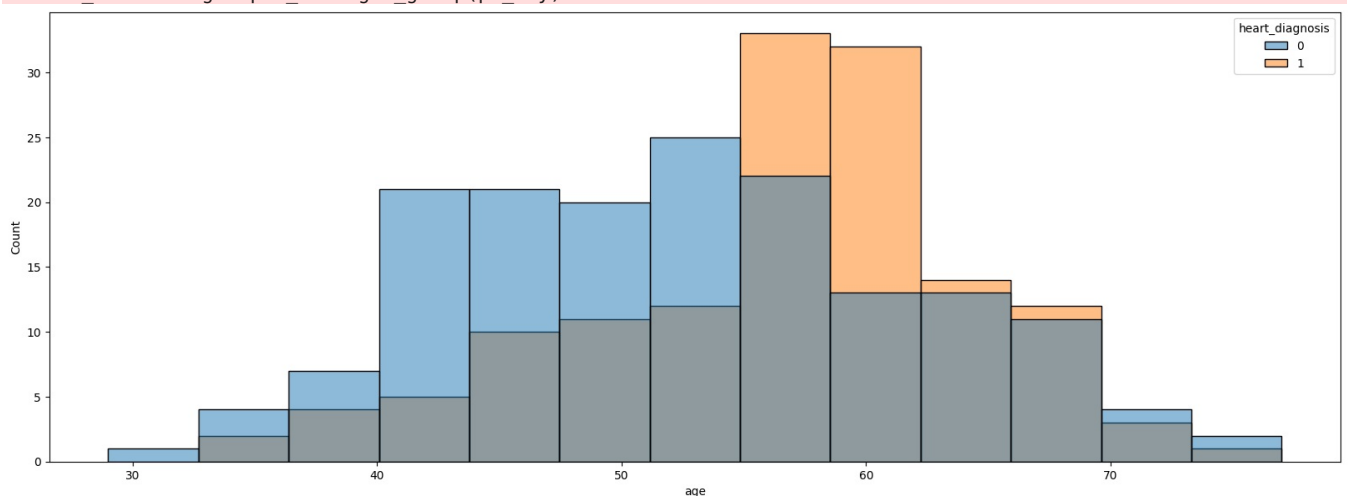
```
Out[68]: Index(['age', 'gender', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach',
              'exang', 'oldpeak', 'slope', 'ca', 'thal', 'heart_diagnosis'],
              dtype='object')
```

Explotery data analysis

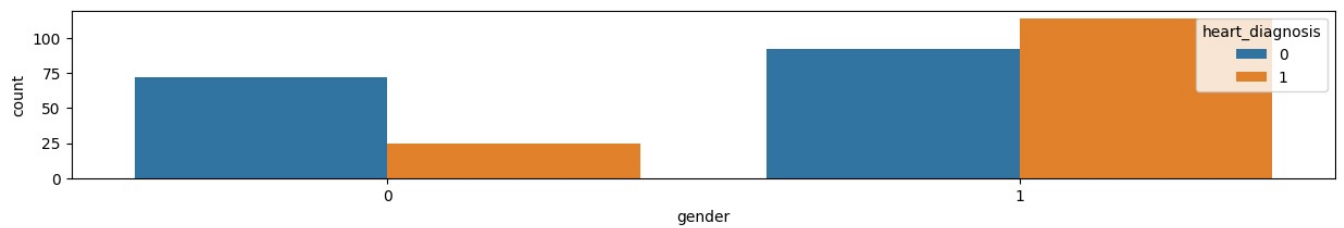
```
In [69]: plt.figure(figsize=(20,7))
sns.histplot(data=df,x='age',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping
with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. P
ass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping
with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. P
ass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping
with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. P
ass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```



```
In [70]: plt.figure(figsize=(15,2))
sns.countplot(data=df,x='gender',hue='heart_diagnosis')
plt.show()
```



```
In [71]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='cp',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

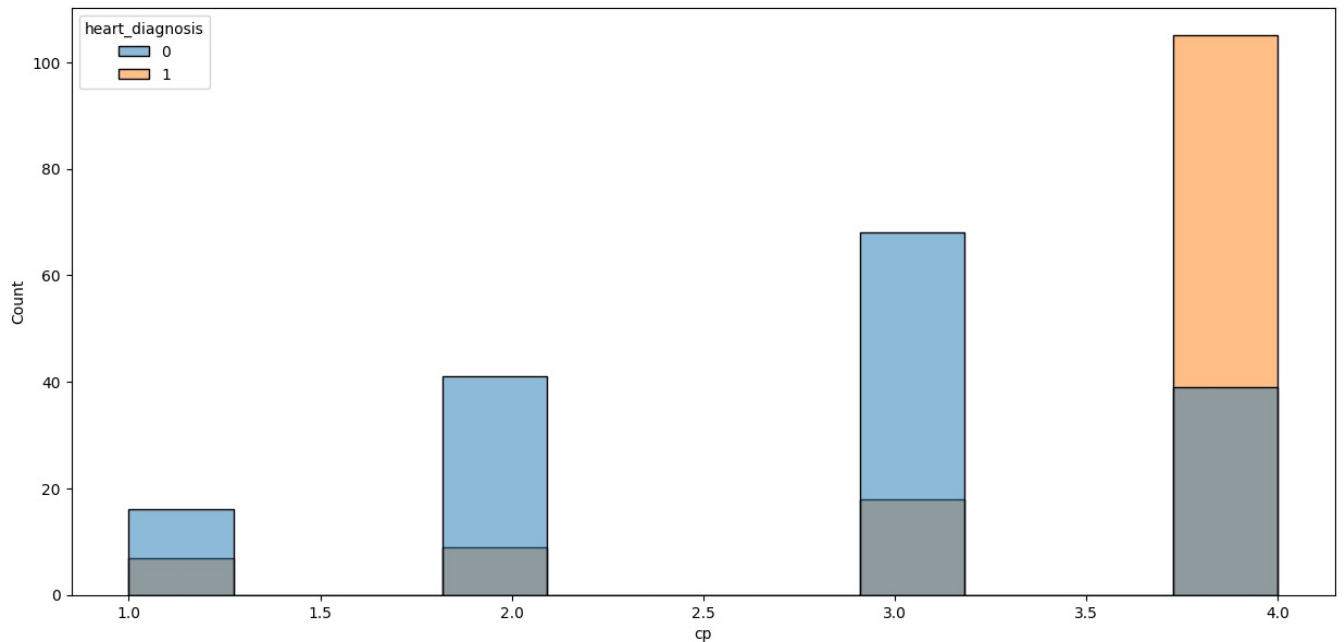
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



```
In [72]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='trestbps',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

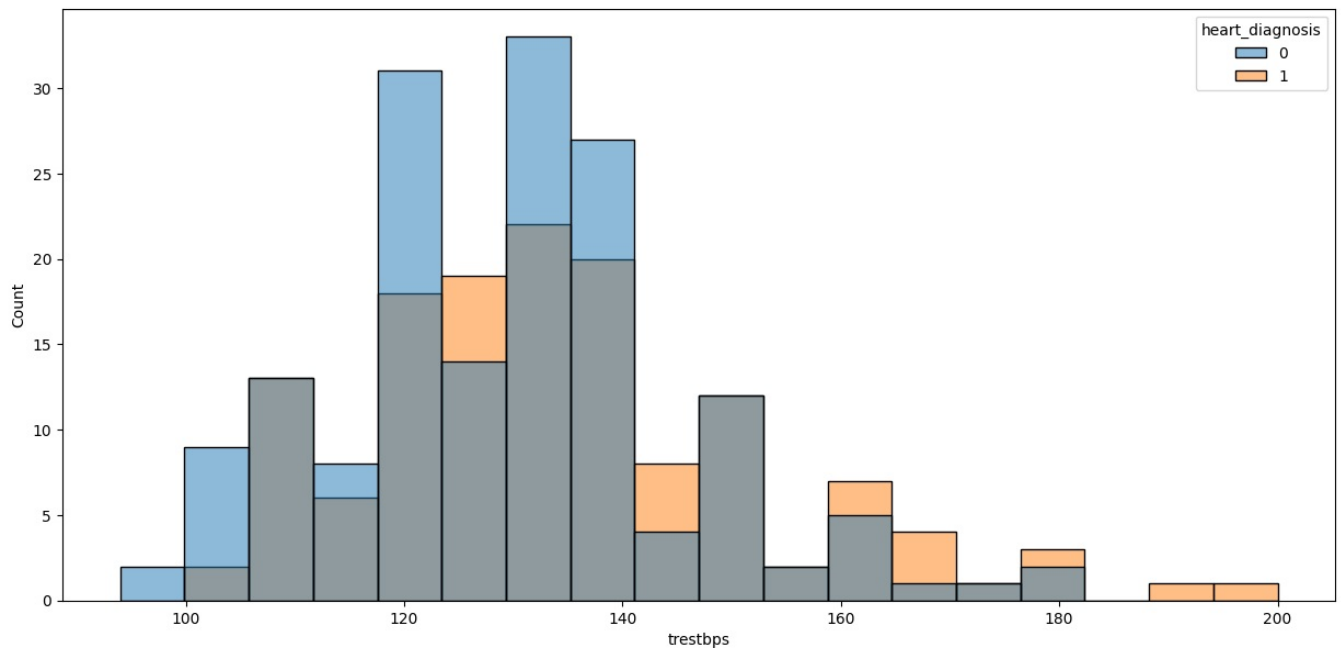
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of name to silence this warning.

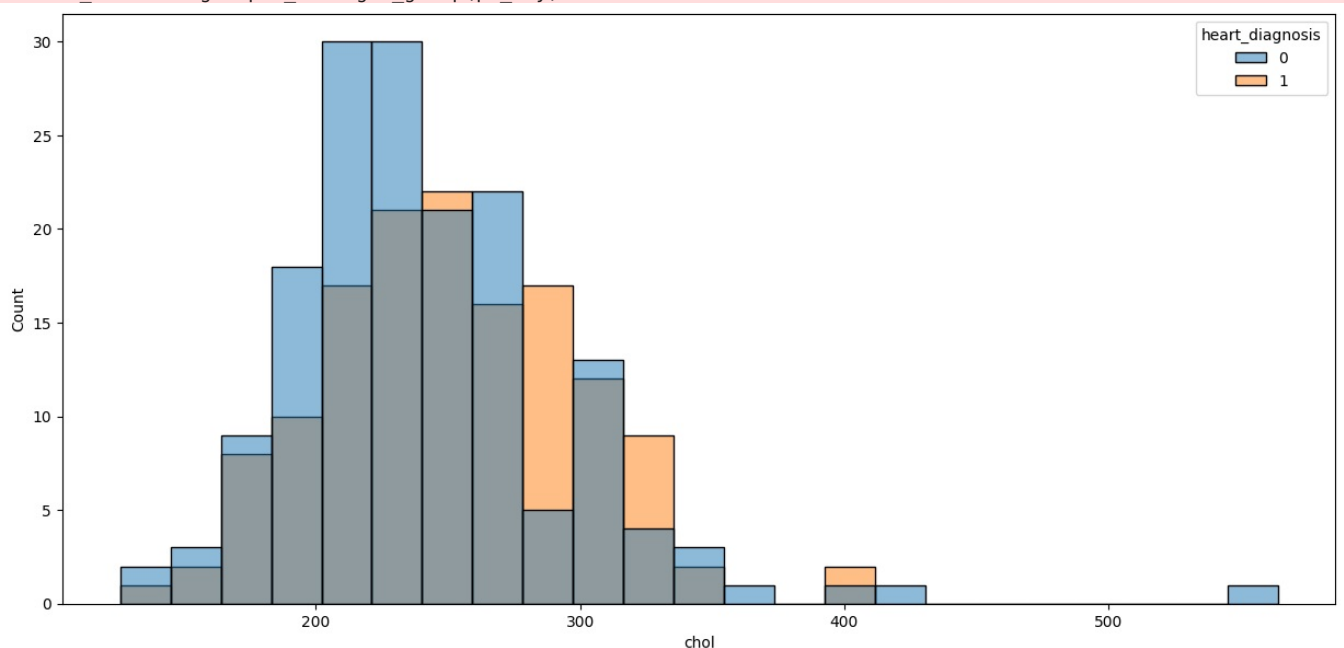
```
data_subset = grouped_data.get_group(pd_key)
```



```
In [73]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='chol',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass (name,) instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```



```
In [74]: plt.figure(figsize=(8,4))
sns.histplot(data=df,x='fbs',hue='heart_diagnosis')
```

```
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

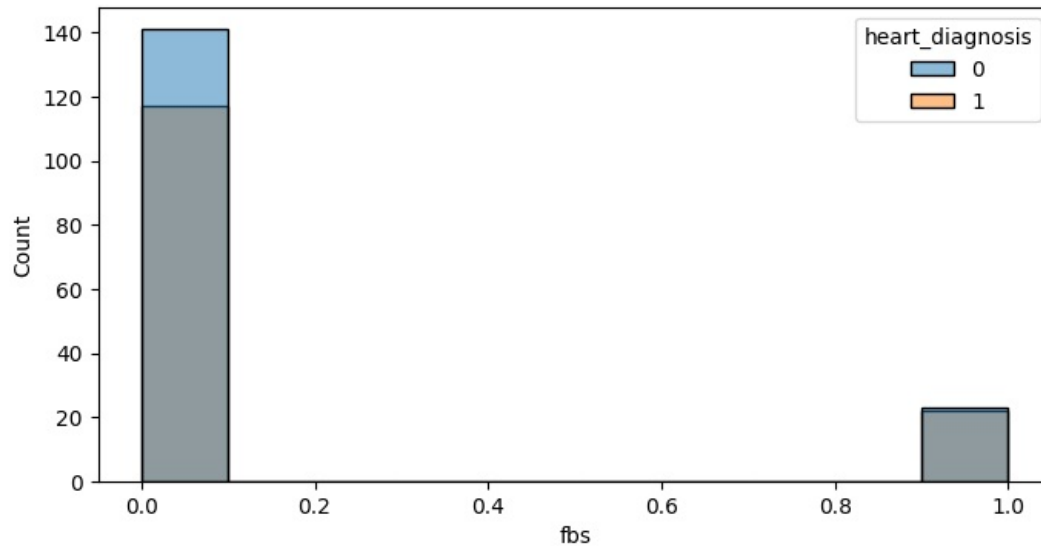
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



```
In [75]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='restecg',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

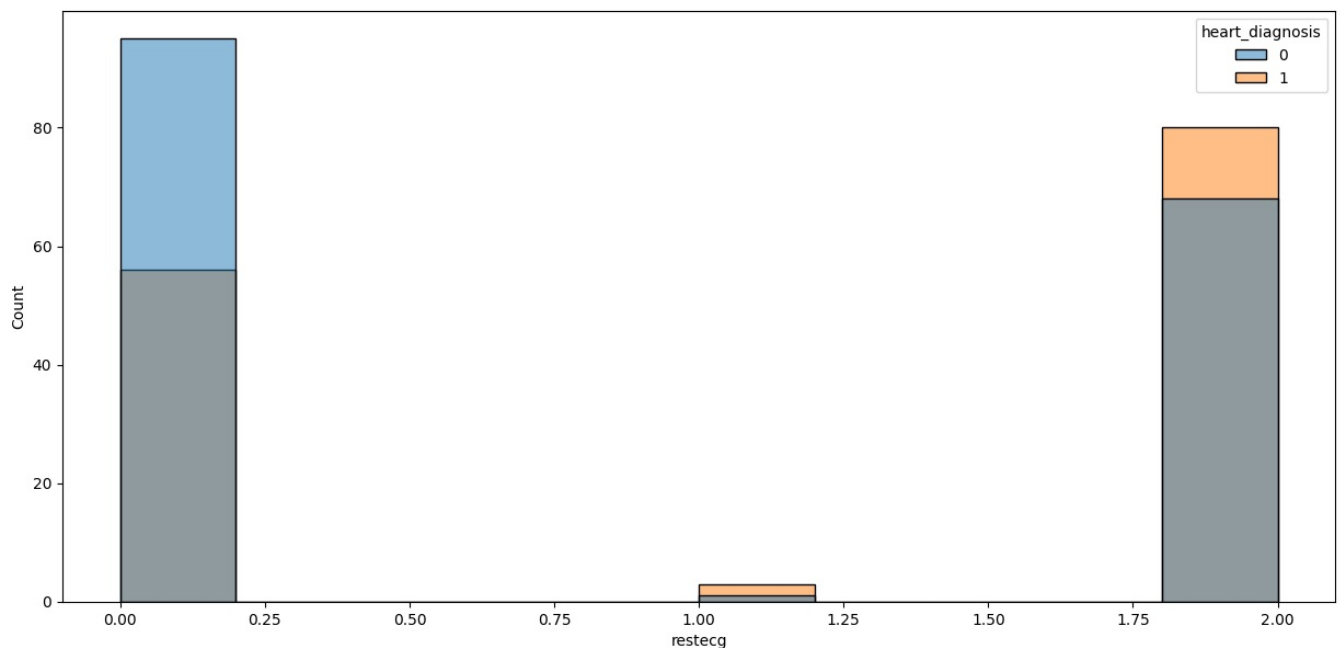
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



```
In [76]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='thalach',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

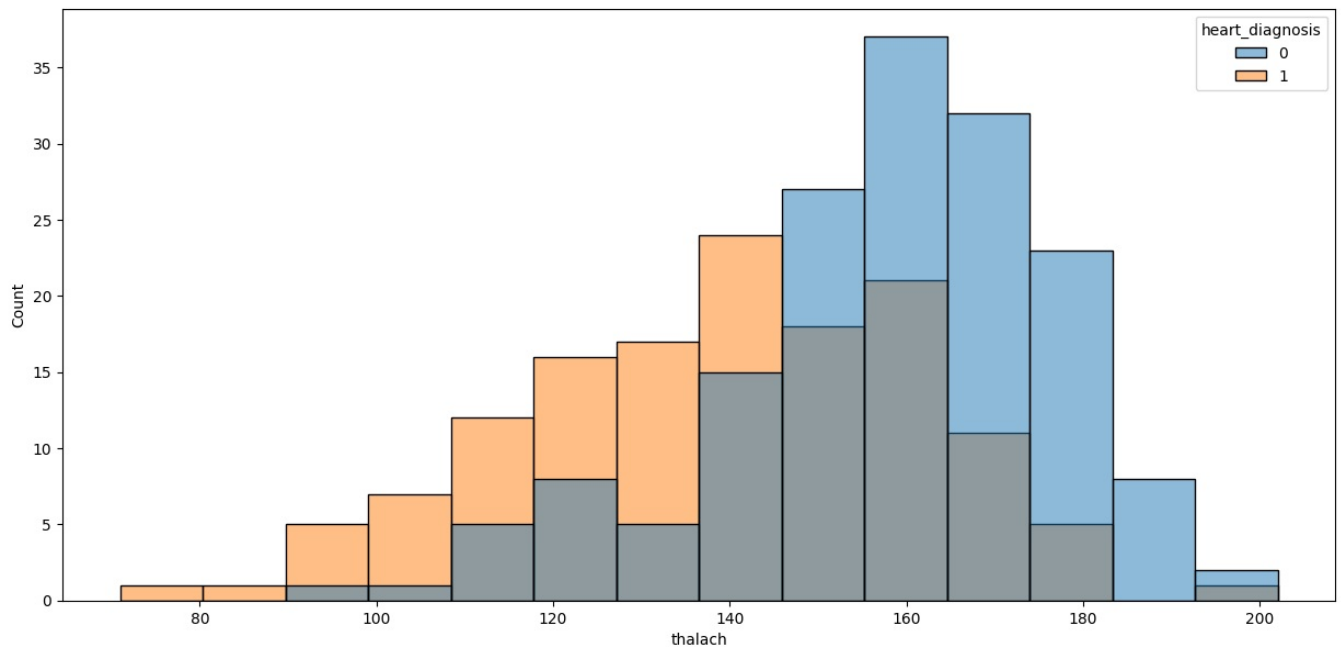
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



```
In [77]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='exang',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

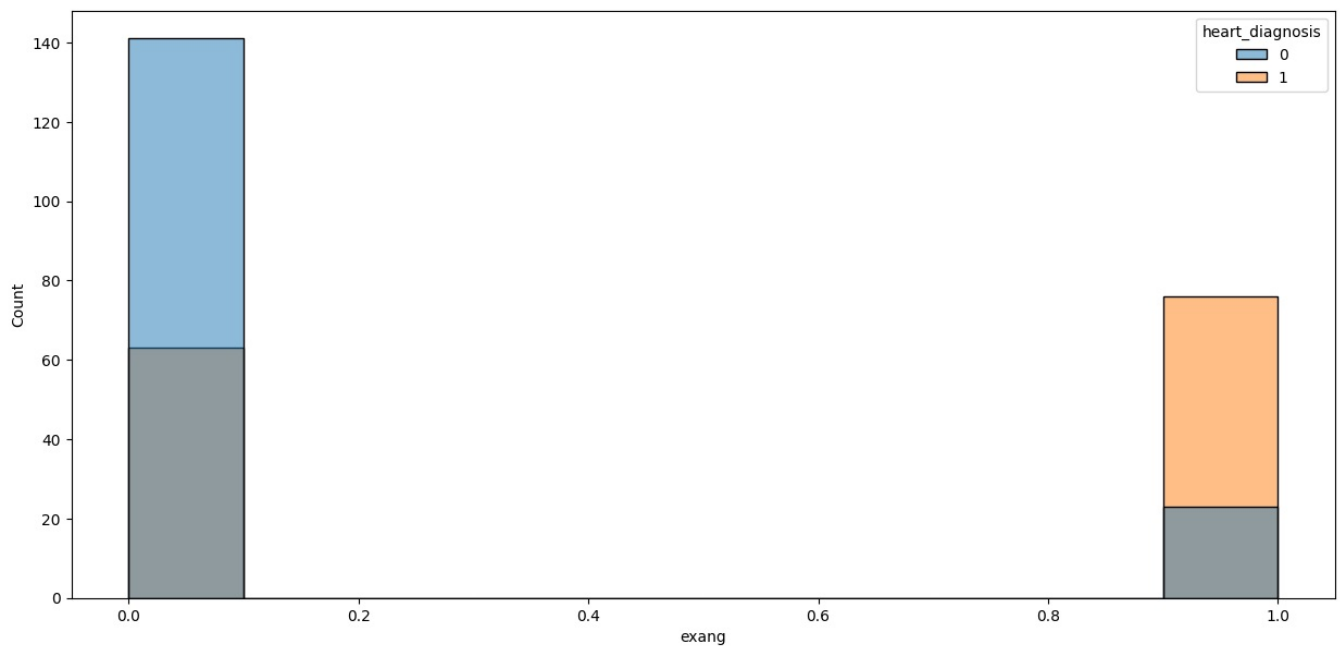
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

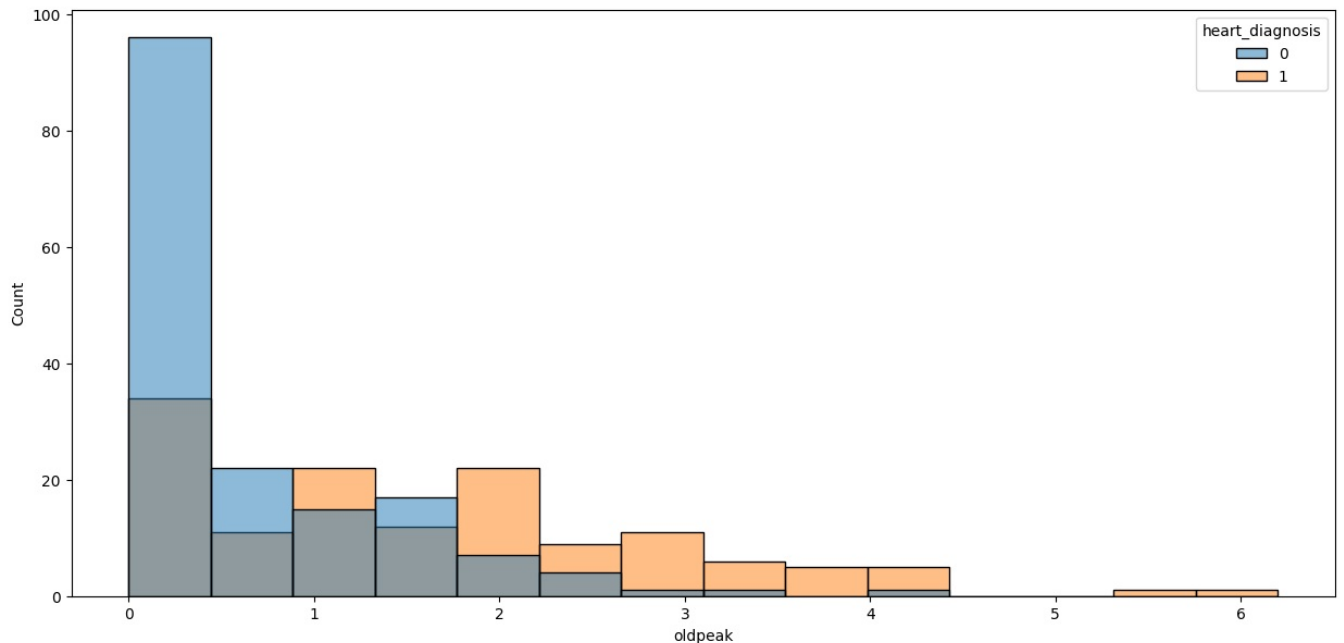
```
data_subset = grouped_data.get_group(pd_key)
```



```
In [78]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='oldpeak',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping
with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. P
ass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn\_oldcore.py:1075: FutureWarning: When grouping
with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. P
ass `(name,)` instead of `name` to silence this warning.
    data_subset = grouped_data.get_group(pd_key)
```

```
In [79]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='slope',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

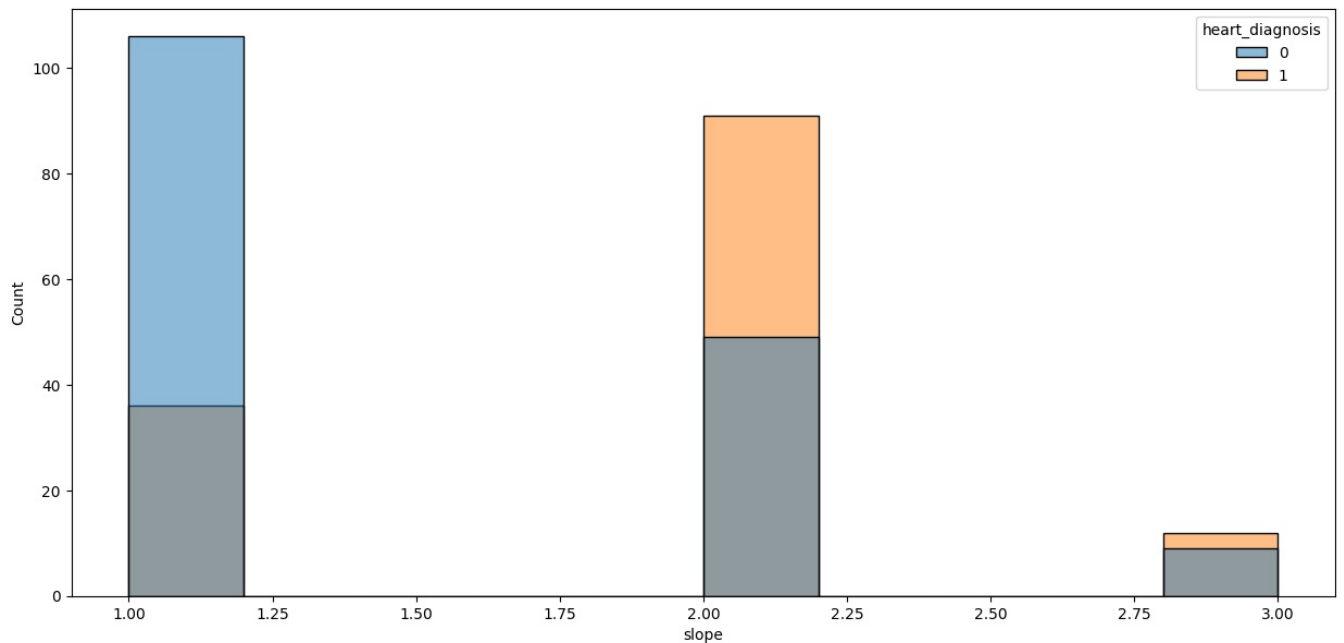
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



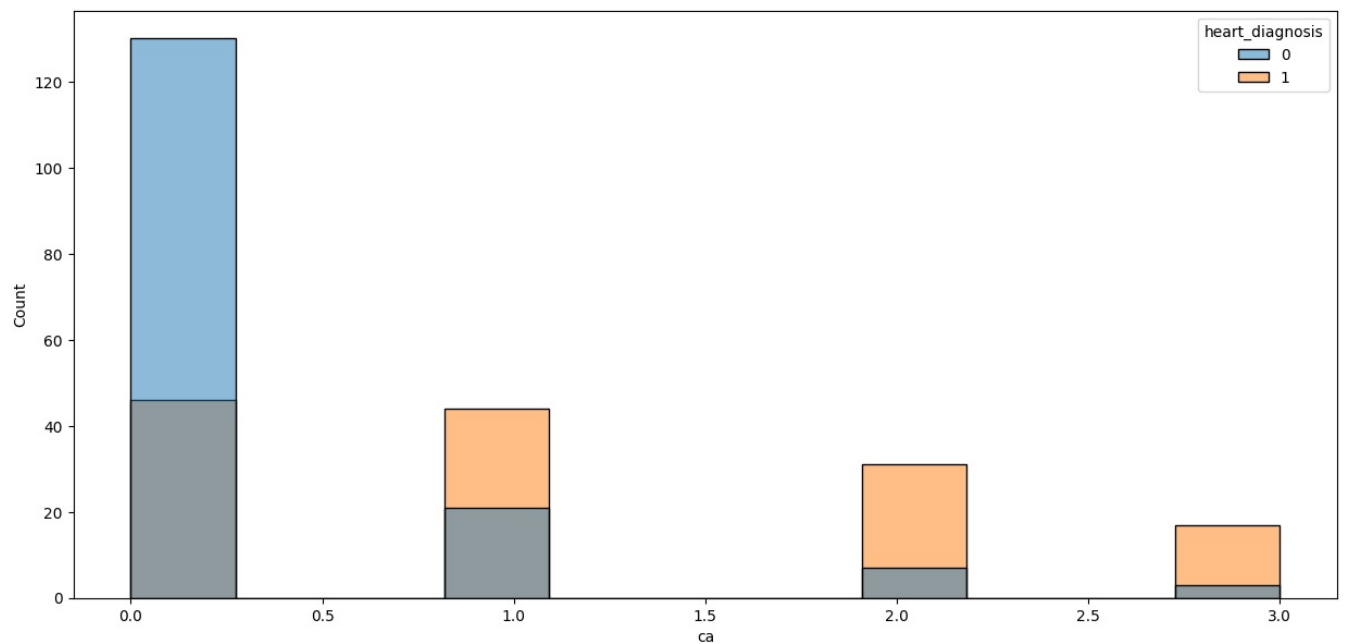
```
In [80]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='ca',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

data_subset = grouped_data.get_group(pd_key)



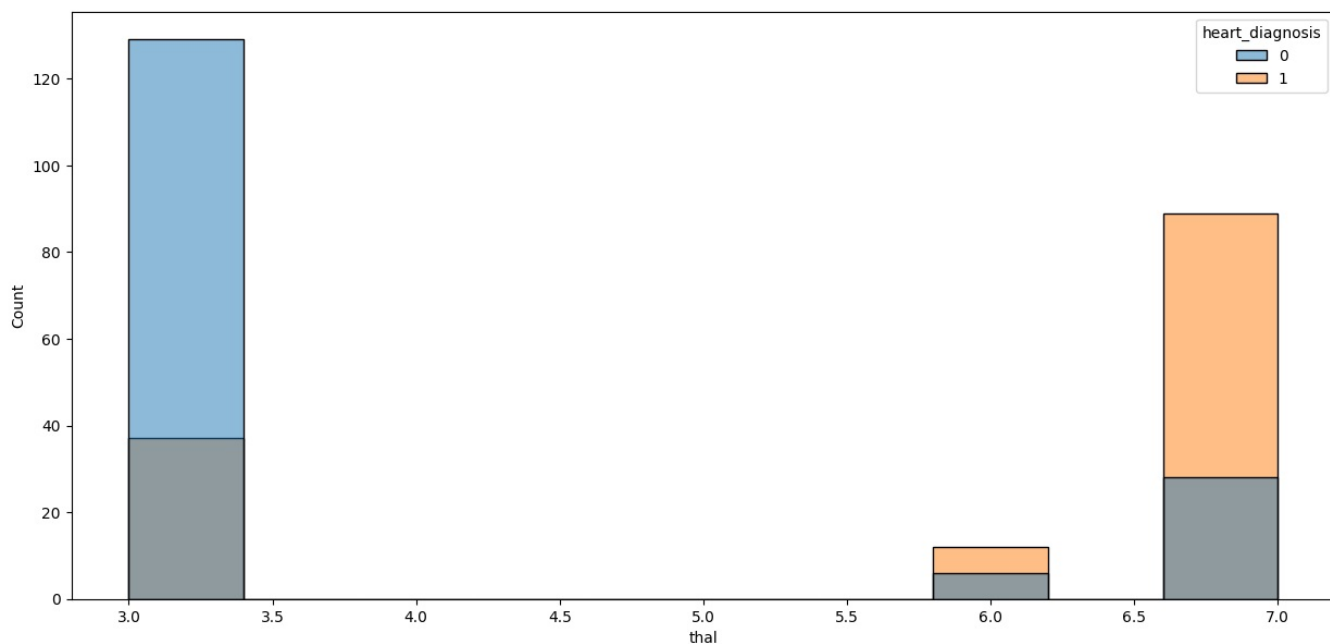
```
In [81]: plt.figure(figsize=(15,7))
sns.histplot(data=df,x='thal',hue='heart_diagnosis')
plt.show()
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

data_subset = grouped_data.get_group(pd_key)
C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\seaborn_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

data_subset = grouped_data.get_group(pd_key)



2.1 Missing value Treatment

```
In [82]: df.isnull().sum()/len(df)*100
```

```
Out[82]: age                0.000000
gender              0.000000
cp                  0.000000
trestbps            0.000000
chol                0.000000
fbs                 0.000000
restecg             0.000000
thalach             0.000000
exang               0.000000
oldpeak             0.000000
slope               0.000000
ca                  1.320132
thal                0.660066
heart_diagnosis     0.000000
dtype: float64
```

```
In [83]: df.head()
```

```
Out[83]:
```

	age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0

```
In [84]: df_numerical_values=df.select_dtypes(include='number')
```

```
In [85]: df_categorical_values=df.select_dtypes(include="object")
```

```
In [86]: df_binary_values=df.select_dtypes(include="bool")
```

```
In [87]: df_numerical_values
```

```
Out[87]:
```

	age	gender	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0
...
298	45	1	1	110	264	0	0	132	0	1.2	2	0.0	7.0	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2.0	7.0	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1.0	7.0	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1.0	3.0	1
302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0

303 rows × 14 columns

function to impute missing values

```
In [88]: heart_data=df.fillna(df.mean())
```

```
In [89]: heart_data
```

```
Out[89]:
```

	age	gender	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.000000	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.000000	3.0	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2.000000	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.000000	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.000000	3.0	0
...
298	45	1	1	110	264	0	0	132	0	1.2	2	0.000000	7.0	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2.000000	7.0	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1.000000	7.0	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1.000000	3.0	1
302	38	1	3	138	175	0	0	173	0	0.0	1	0.672241	3.0	0

303 rows × 14 columns

```
In [90]: # Transform categorical data using one-hot encoding
heart_data_encoded = pd.get_dummies(heart_data, drop_first=True)
```

```
In [91]: heart_data_encoded
```

```
Out[91]:
```

	age	gender	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	heart_diagnosis
0	63	1	1	145	233	1	2	150	0	2.3	3	0.000000	6.0	0
1	67	1	4	160	286	0	2	108	1	1.5	2	3.000000	3.0	1
2	67	1	4	120	229	0	2	129	1	2.6	2	2.000000	7.0	1
3	37	1	3	130	250	0	0	187	0	3.5	3	0.000000	3.0	0
4	41	0	2	130	204	0	2	172	0	1.4	1	0.000000	3.0	0
...
298	45	1	1	110	264	0	0	132	0	1.2	2	0.000000	7.0	1
299	68	1	4	144	193	1	0	141	0	3.4	2	2.000000	7.0	1
300	57	1	4	130	131	0	0	115	1	1.2	2	1.000000	7.0	1
301	57	0	2	130	236	0	2	174	0	0.0	2	1.000000	3.0	1
302	38	1	3	138	175	0	0	173	0	0.0	1	0.672241	3.0	0

303 rows × 14 columns

```
In [92]: from sklearn.model_selection import train_test_split

# Separate features and target variable
X = heart_data_encoded.drop("heart_diagnosis", axis=1)
y = heart_data_encoded["heart_diagnosis"]

# Split the data into train and test sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
In [93]: from imblearn.under_sampling import RandomUnderSampler
from imblearn.over_sampling import RandomOverSampler

# Undersample the data
undersampler = RandomUnderSampler(random_state=42)
X_train_undersampled, y_train_undersampled = undersampler.fit_resample(X_train, y_train)

# Oversample the data
oversampler = RandomOverSampler(random_state=42)
X_train_oversampled, y_train_oversampled = oversampler.fit_resample(X_train, y_train)

# Print the class distribution after undersampling and oversampling
print("\nClass distribution after undersampling:")
print(pd.Series(y_train_undersampled).value_counts())

print("\nClass distribution after oversampling:")
print(pd.Series(y_train_oversampled).value_counts())

Class distribution after undersampling:
heart_diagnosis
0      107
1      107
Name: count, dtype: int64

Class distribution after oversampling:
heart_diagnosis
0      135
1      135
Name: count, dtype: int64
```

```
In [94]: from sklearn.linear_model import LogisticRegression
from sklearn.metrics import roc_auc_score, accuracy_score, classification_report

# Apply Logistic Regression model on the skewed data
logreg_skewed = LogisticRegression(random_state=42)
logreg_skewed.fit(X_train, y_train)

# Apply Logistic Regression model on the undersampled data
logreg_undersampled = LogisticRegression(random_state=42)
logreg_undersampled.fit(X_train_undersampled, y_train_undersampled)

# Apply Logistic Regression model on the oversampled data
logreg_oversampled = LogisticRegression(random_state=42)
logreg_oversampled.fit(X_train_oversampled, y_train_oversampled)

# Print model results
def print_model_results(model, X_test, y_test, model_type):
    y_pred = model.predict(X_test)
    roc_auc = roc_auc_score(y_test, y_pred, multi_class='ovr')
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Model Type: {model_type}")
    print("ROC AUC Score:", roc_auc)
    print("Accuracy Score:", accuracy)
    print("Classification Report:")
    print(classification_report(y_test, y_pred))

print("Model Results:")
print()
print_model_results(logreg_skewed, X_test, y_test, "Skewed Data")
print()
print_model_results(logreg_undersampled, X_test, y_test, "Undersampled Data")
print()
print_model_results(logreg_oversampled, X_test, y_test, "Oversampled Data")
```

Model Results:

Model Type: Skewed Data

ROC AUC Score: 0.8685344827586206

Accuracy Score: 0.8688524590163934

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	29
1	0.88	0.88	0.88	32
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

Model Type: Undersampled Data

ROC AUC Score: 0.9014008620689655

Accuracy Score: 0.9016393442622951

Classification Report:

	precision	recall	f1-score	support
0	0.90	0.90	0.90	29
1	0.91	0.91	0.91	32
accuracy			0.90	61
macro avg	0.90	0.90	0.90	61
weighted avg	0.90	0.90	0.90	61

Model Type: Oversampled Data

ROC AUC Score: 0.8685344827586206

Accuracy Score: 0.8688524590163934

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.86	0.86	29
1	0.88	0.88	0.88	32
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\sklearn\linear_model_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\sklearn\linear_model_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

C:\Users\Prakash Enerprener\anaconda4\Lib\site-packages\sklearn\linear_model_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

```
In [95]: # Get feature importance
feature_importance_skewed = pd.DataFrame({'Feature': X.columns, 'Importance': logreg_skewed.coef_[0]})
feature_importance_undersampled = pd.DataFrame({'Feature': X.columns, 'Importance': logreg_undersampled.coef_[0]})
feature_importance_oversampled = pd.DataFrame({'Feature': X.columns, 'Importance': logreg_oversampled.coef_[0]})

# Compare the models
print("\nFeature Importance - Skewed Data:")
print(feature_importance_skewed.sort_values(by='Importance', ascending=False))

print("\nFeature Importance - Undersampled Data:")
print(feature_importance_undersampled.sort_values(by='Importance', ascending=False))

print("\nFeature Importance - Oversampled Data:")
print(feature_importance_oversampled.sort_values(by='Importance', ascending=False))
```

Feature Importance - Skewed Data:

	Feature	Importance
11	ca	1.367356
1	gender	0.571258
8	exang	0.504076
9	oldpeak	0.415343
12	thal	0.315353
2	cp	0.307734
6	restecg	0.203676
10	slope	0.119330
3	trestbps	0.011205
4	chol	0.005604
7	thalach	-0.034924
0	age	-0.041848
5	fbs	-0.112033

Feature Importance - Undersampled Data:

	Feature	Importance
11	ca	1.238453
1	gender	1.035724
8	exang	0.681077
10	slope	0.489897
2	cp	0.353480
12	thal	0.241606
9	oldpeak	0.239721
6	restecg	0.106071
3	trestbps	0.005967
4	chol	0.005146
0	age	-0.027341
7	thalach	-0.034562
5	fbs	-0.141779

Feature Importance - Oversampled Data:

	Feature	Importance
11	ca	1.178348
1	gender	1.127028
8	exang	0.794667
10	slope	0.338263
9	oldpeak	0.325869
2	cp	0.255005
12	thal	0.237955
6	restecg	0.132076
3	trestbps	0.012872
4	chol	0.006706
0	age	-0.036006
7	thalach	-0.037409
5	fbs	-0.385303

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js