CM4

January 29, 2021

1 Histogram for Heart Diseases dataset

1.1 Required libraries

[1]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from sklearn.model_selection import train_test_split
import seaborn as sns
sns.set()
```

1.1.1 Grouping the features by their variable type

[2]:
```
binary = ['sex', 'fbs', 'exang']
catergorical = ['cp', 'restecg', 'slope', 'thal']
ordinal = ['ca']
numrical = ['age', 'oldpeak', 'trestbps', 'chol', 'thalach']
```

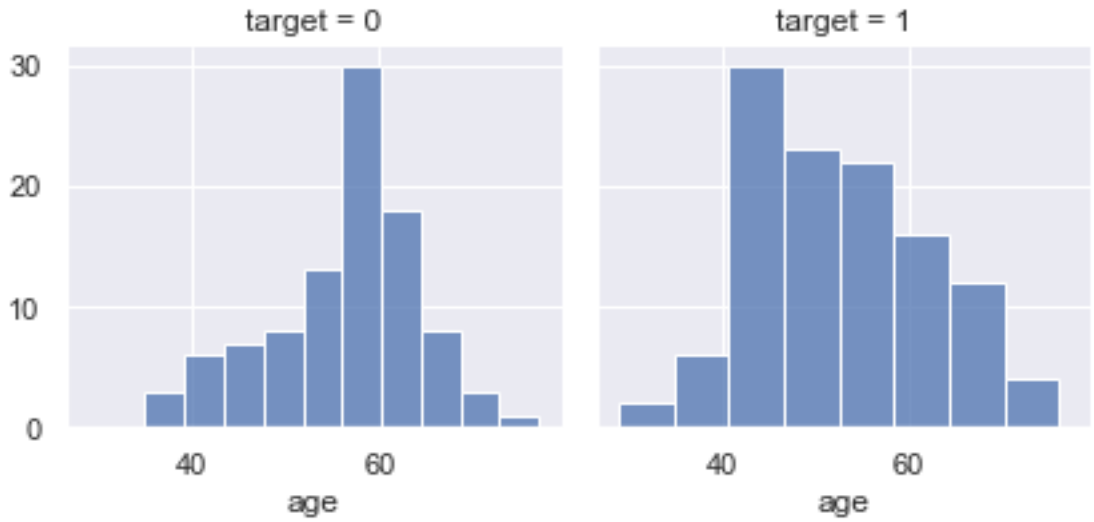
1.2 Plotting histogram for features of heart disease datset.

[3]:
```
df_heart= pd.read_csv("heart_disease_missing.csv")
```

The features were selected to show the present and absent heart diseases.

[4]:
```
heart= sns.FacetGrid(df_heart, col="target")
heart.map(sns.histplot, "age")
```

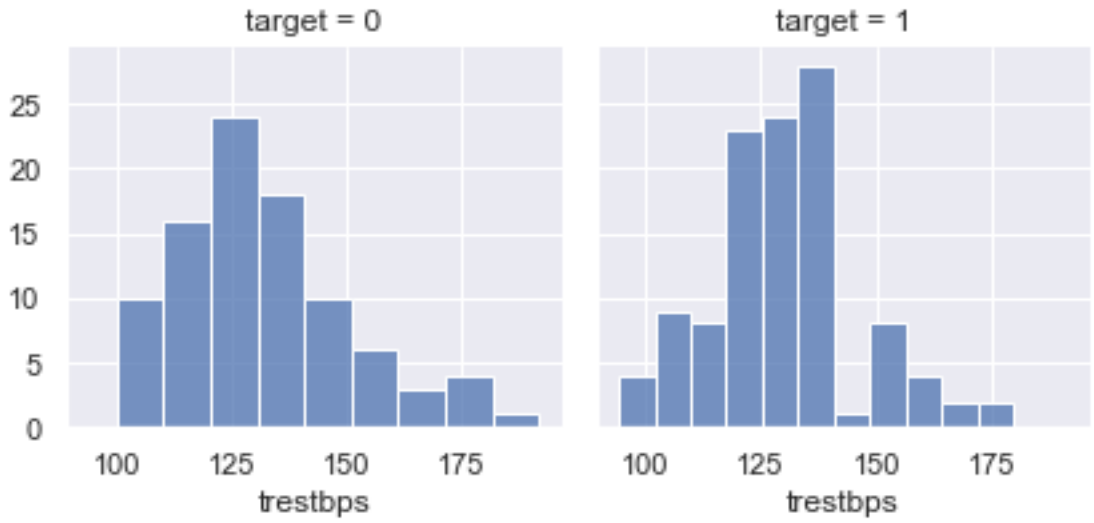
[4]:
```
<seaborn.axisgrid.FacetGrid at 0x20be8e6cc70>
```



You can conclude from the above histogram that young people are more likely to have heart diseases than older people, this relation can be used to further investigate the reason for young poeple suffer from heart diseases.

[5]:
```
heart= sns.FacetGrid(df_heart, col="target")
heart.map(sns.histplot, "trestbps")
```

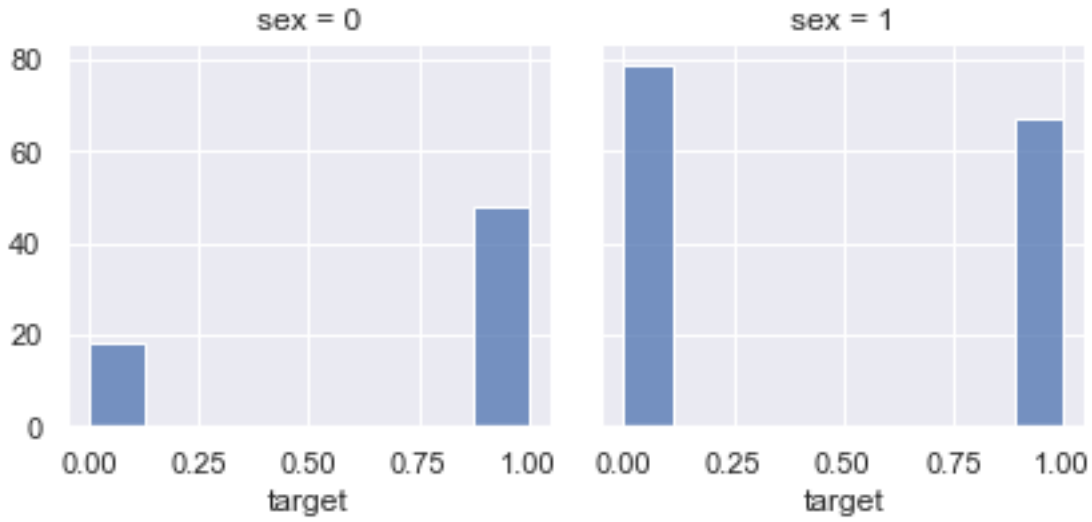
[5]: <seaborn.axisgrid.FacetGrid at 0x20bee5db550>



It is easily observed from above plot that high resting blood pressure can lead to heart diseases and visa vera is also true.

[6]:
```
heart= sns.FacetGrid(df_heart, col="sex")
heart.map(sns.histplot, "target")
```

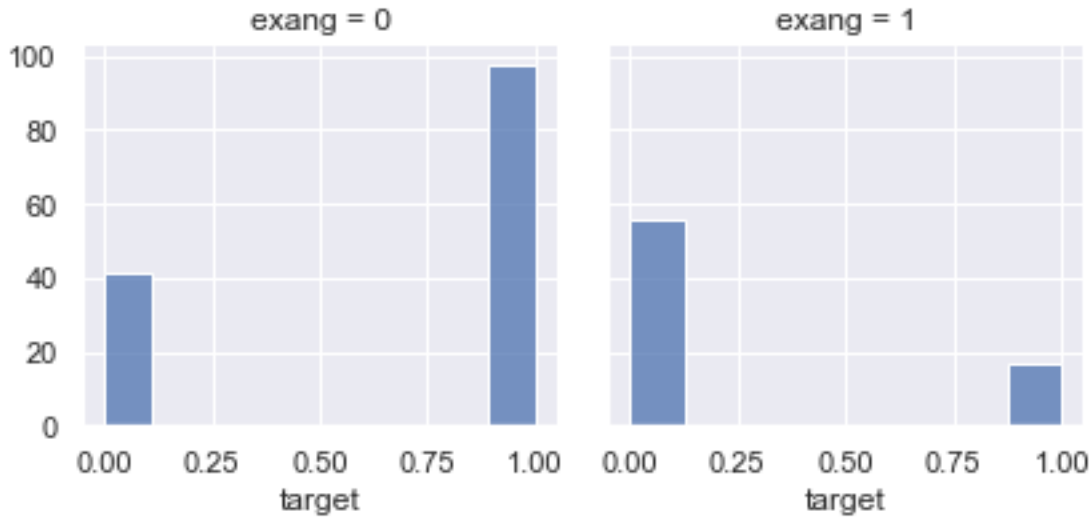
[6]: <seaborn.axisgrid.FacetGrid at 0x20bee543a90>



The graph shows that females are more prone to heart diseases compared to male.

[7]:
```
heart= sns.FacetGrid(df_heart, col="exang")
heart.map(sns.histplot, "target")
```

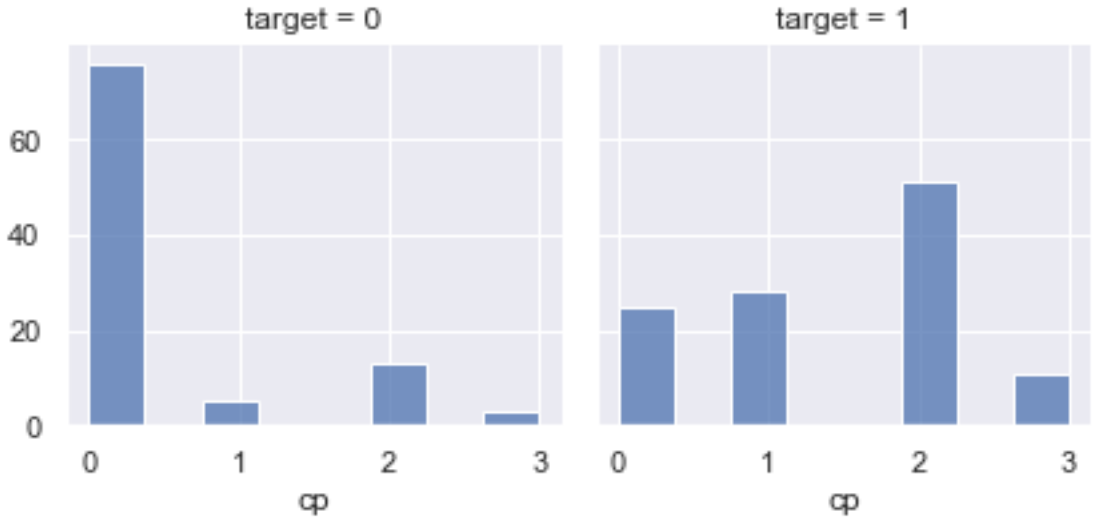
[7]: <seaborn.axisgrid.FacetGrid at 0x20bee552340>



If exercise induced angina is absent then person is more likely to suffer from heart diseases.

[8]:
```
heart= sns.FacetGrid(df_heart, col="target")
heart.map(sns.histplot, "cp")
```

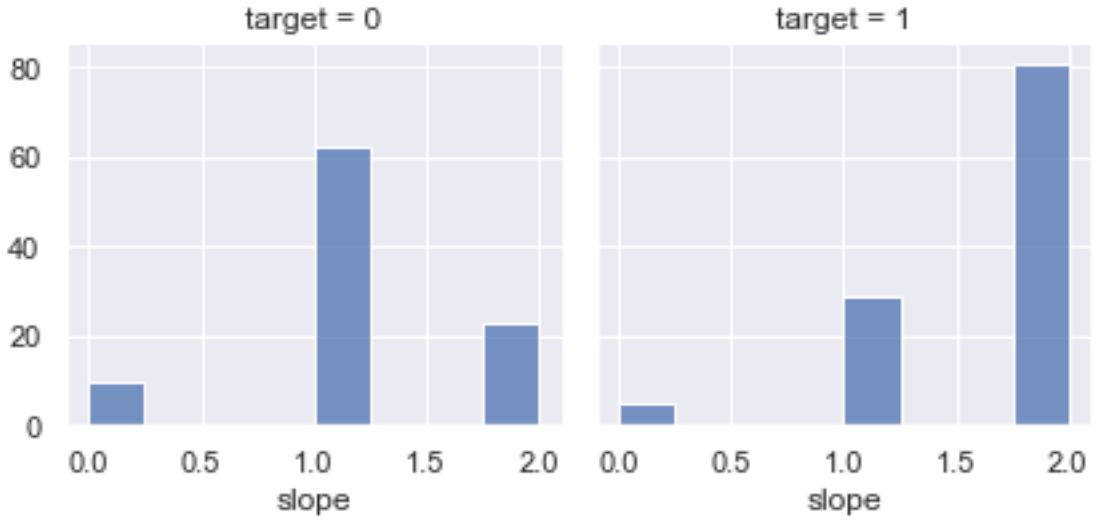
[8]: <seaborn.axisgrid.FacetGrid at 0x20bee8496a0>



Person with asymptomatic angina has low probability of suffering from heart diseases, whereas non-angina case has more chance of heart diseases.

[9]:
```
heart= sns.FacetGrid(df_heart, col="target")
heart.map(sns.histplot, "slope")
```

[9]: <seaborn.axisgrid.FacetGrid at 0x20bee8d7a60>



If the slope of the peak exercise ST segment is flat then there are more chances of getting a heart diseases.

1.3 References

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html
https://seaborn.pydata.org/generated/seaborn.FacetGrid.html
https://seaborn.pydata.org/generated/seaborn.FacetGrid.map.html#seaborn.FacetGrid.map