

Assignment 7

Details

1. Author : Syed Farhan Naqvi
2. Roll Number : 33346
3. Batch : M11
4. Class : TE11

Problem Statement

Visualize the data using Python libraries matplotlib, seaborn by plotting the graphs for assignment no. 2 and 3

Implementation details

1. Dataset URLs
 1. Facebook metrics : <https://archive.ics.uci.edu/ml/datasets/Facebook+metrics>
(visualization done in different notebook)
 2. Heart Disease : <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
2. Python version : 3.7.4

Dataset details

1. Facebook Metrics :
 1. Given dataset is a representative of some of the Facebook metrics which are associated with the posts on social media.
 2. These metrics are indicative of the engagement of the users with the corresponding post.
 3. It includes various types of posts and their details
2. Heart Disease Dataset :
 1. This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date.
 2. The "goal" field refers to the presence of heart disease in the patient.

3. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).
4. The names and social security numbers of the patients were recently removed from the database, replaced with dummy values

▼ Importing required libraries

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

B) Heart Disease dataset

▼ 1) Loading the dataset

```
heart_dataset = pd.read_csv("./processed.cleveland.csv", header=None)
heart_dataset.head()
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0	0
1	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
2	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
3	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
4	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0

▼ 2) Renaming columns

```
heart_dataset.columns = [
    "age",
    "sex",
    "chest_pain",
```

```

    "trestbps",
    "cholesterol",
    "fbs",
    "restecg",
    "thalach",
    "exang",
    "oldpeak",
    "slope",
    "ca",
    "thal",
    "num"
]

```

```
heart_dataset.head()
```

	age	sex	chest_pain	trestbps	cholesterol	fbs	restecg	thalach	exang	oldpeak	slope
0	63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	0
1	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	0
2	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	0
3	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	0
4	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	0

3) Quartile spread of thalach feature

```

# Generating bar graph
fig = plt.figure(figsize=(8, 8))

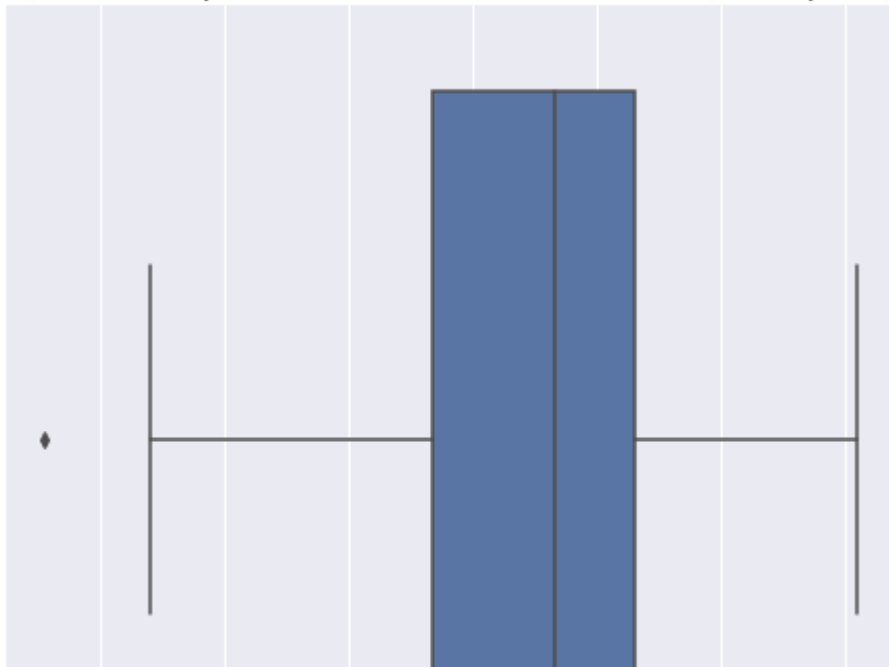
# Adds subplot on position 1
ax = fig.add_subplot(111)

sns.boxplot(x=heart_dataset.thalach)
plt.title("Quartile spread of thalach feature (Box plot)", fontsize=20)

plt.show()

```

Quartile spread of thalach feature (Box plot)



▼ 4) Distribution of age in entire dataset

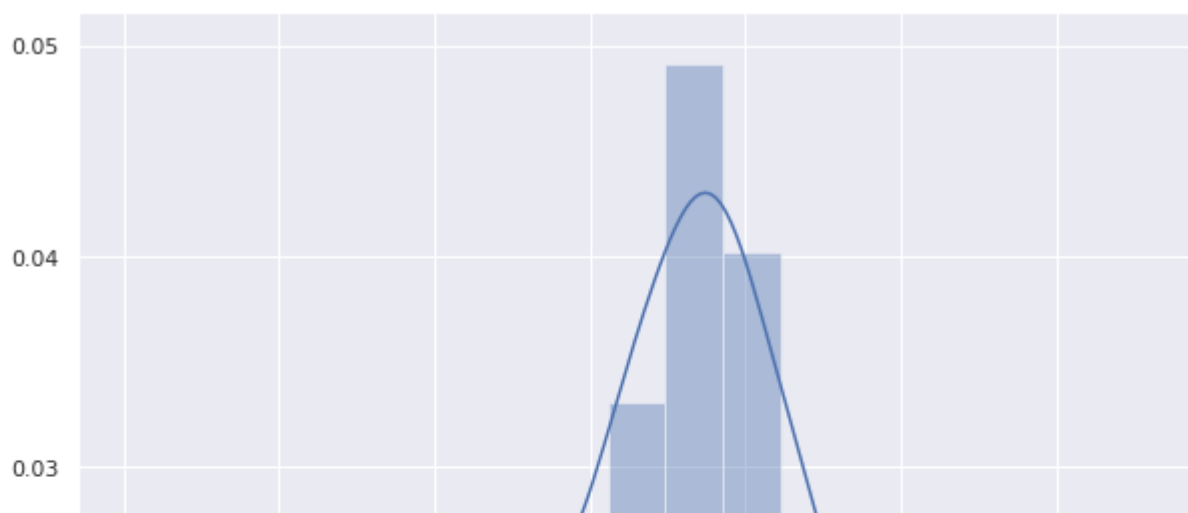


```
# Generating bar graph
fig = plt.figure(figsize=(10, 10))

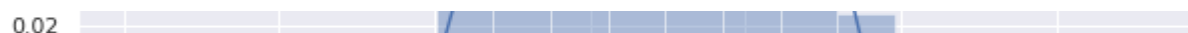
# Adds subplot on position 1
ax = fig.add_subplot(111)

sns.distplot(x=heart_dataset.age)
plt.show()
```

```
/home/varadmash/anaconda3/envs/python3.7_TF2.0/lib/python3.7/site-packages/seaborn/distr
warnings.warn(msg, FutureWarning)
```



▼ 5) Checking correlation using heatmap



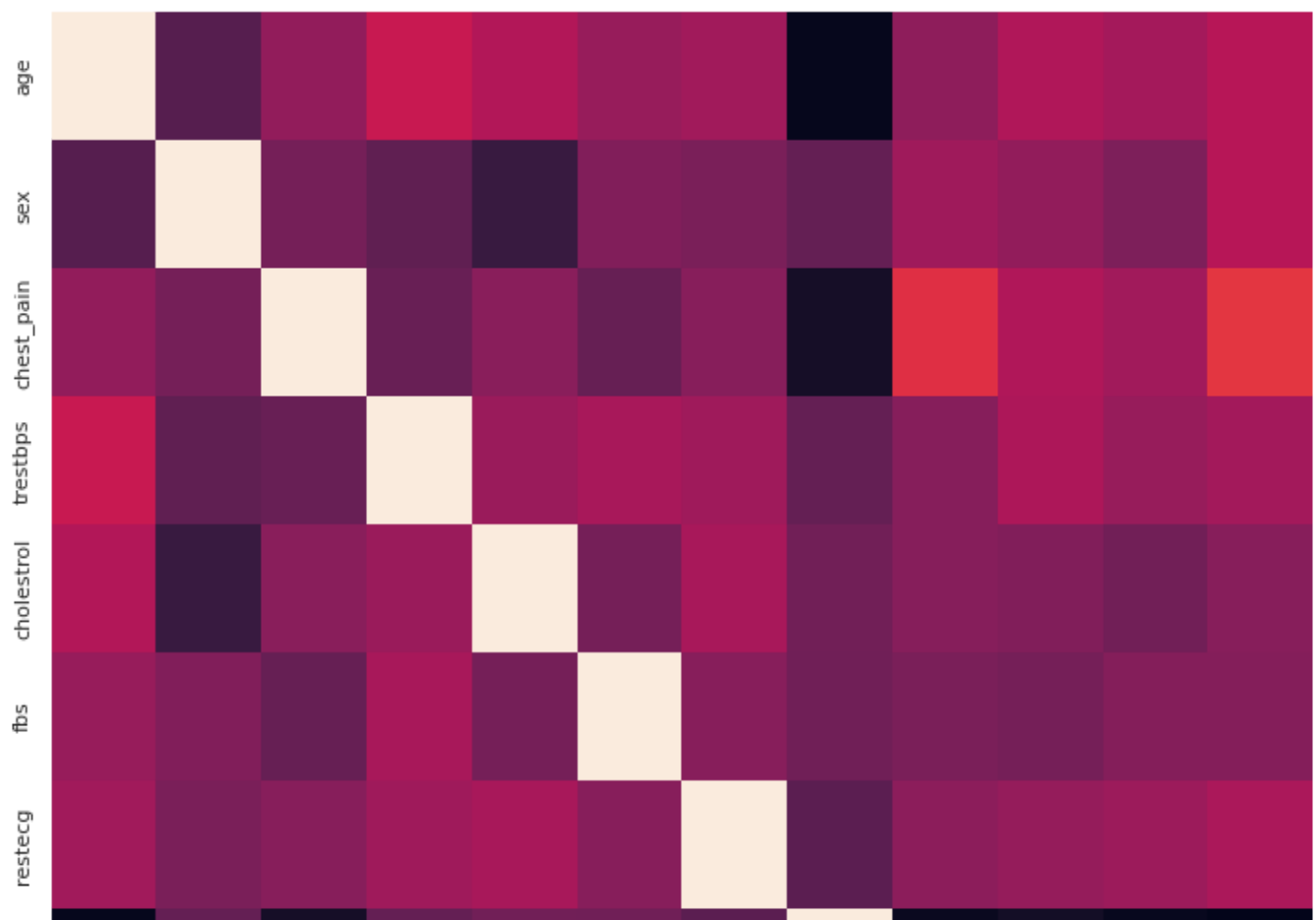
```
# Generating bar graph
fig = plt.figure(figsize=(15, 15))

# Adds subplot on position 1
ax = fig.add_subplot(111)

sns.heatmap(heart_dataset.corr())

plt.plot()
```

[]



▼ Conclusion

1. Implemented following visualization methods :

1. Pie chart
2. Bar chart
3. Count plot
4. Box plot
5. Distribution plot (Histogram)
6. Heatmap



End of Notebook

