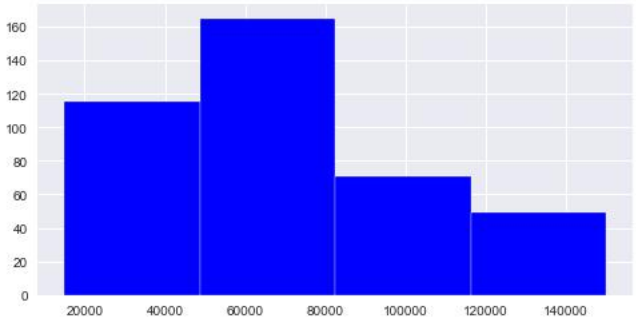
Lab Assignment 3
Chaudhary Hamdan
1905387
Date: 02-02-2022

Question:

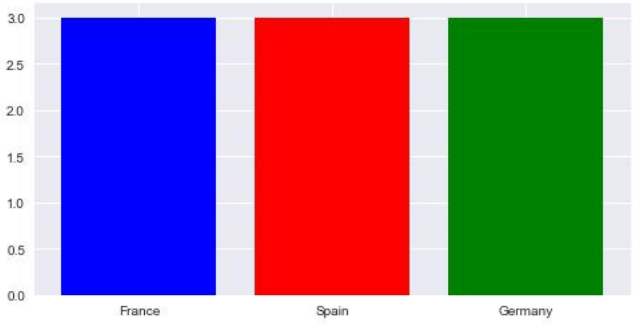
1. Plot a histogram with blue color bars of size 4, and edges should be distinguished from each other, for the dataset social_nework for the feature estimated salary.

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Social_Network_Ads.csv')
plt.hist(
      df['EstimatedSalary'],
      color='blue',
      edgecolor='white',
      bins=4
      )
plt.show()
```



2. On the dataset ‘data’, draw barplot to show the count of categorical feature ‘Country’

```
"""
Created on Wed Feb  2 12:09:12 2022

@author: Chaudhary Hamdan
"""


import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Data.csv')
df.dropna(axis=0, inplace=True)

index = np.arange(len(df['Country'].unique()))
counts = [3,3,3]
plt.bar(index, counts, color = ['blue', 'red', 'green'])
plt.xticks(index, df.Country.unique())


plt.show()
```



3. Remove missing values from the dataframe ceated from dataset ‘data’ and display the dimension of dataframe in both cases.

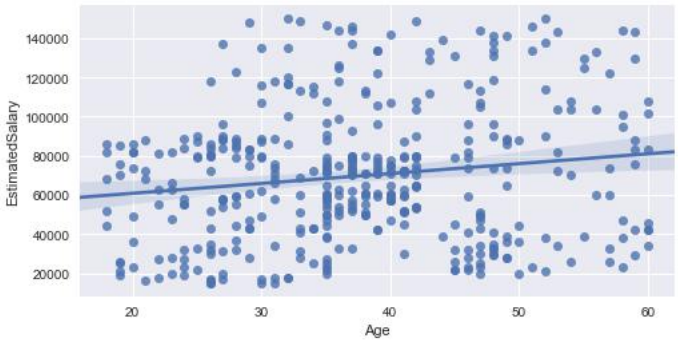
```
"""
Created on Wed Feb  2 12:09:12 2022

@author: Chaudhary Hamdan
"""

import pandas as pd

df = pd.read_csv('Data.csv')

print('Before removing NA Values:', df.shape)

df.dropna(axis=0, inplace=True)

print('After removing NA Values:', df.shape)
```

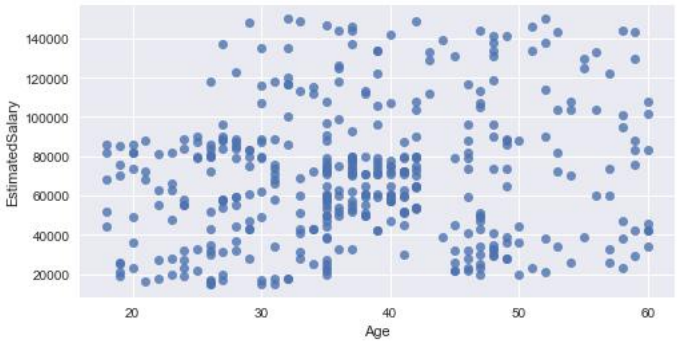
```
In [84]: runfile('C:/Users/KIIT/Desktop/TnT Lab/Lab 3/q3.py', wdir='C:/
Users/KIIT/Desktop/TnT Lab/Lab 3')
Before removing NA Values: (10, 4)
After removing NA Values: (7, 4)
```

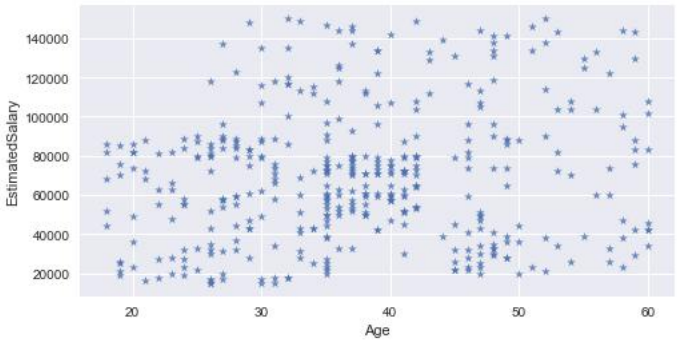
4. Scatter polt age vs estimated salary on gridview
A) show regression fit line
B) Regression fit line should not be visible
C) Use * symbol to show data points without the regression fit line

```
"""
Created on Wed Feb  2 12:25:39 2022

@author: KIIT
"""

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Social_Network_Ads.csv')
#print(df)

sns.set(style='darkgrid')

print('A')
sns.regplot(
      x=df.Age,
      y=df.EstimatedSalary,
      )
plt.show()

print('B')
sns.regplot(
      x=df.Age,
      y=df.EstimatedSalary,
      fit_reg=False
      )
plt.show()

print('C')
sns.regplot(
```

```
x=df.Age,
y=df.EstimatedSalary,
fit_reg=False,
marker='*'
)
```





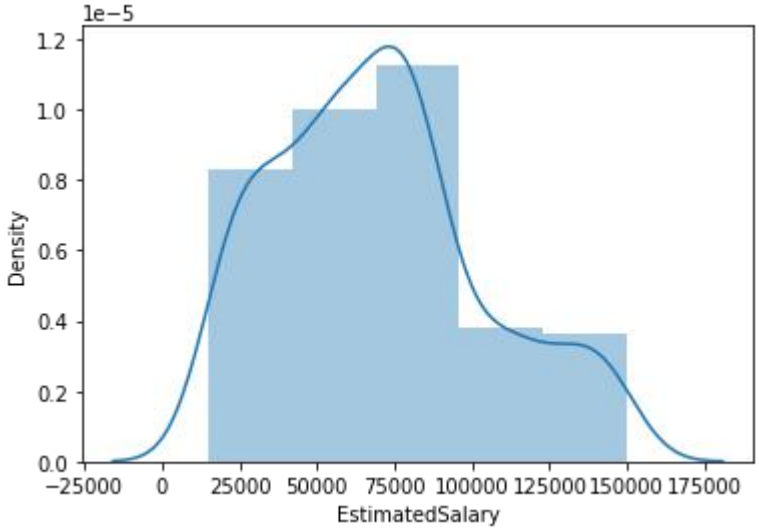


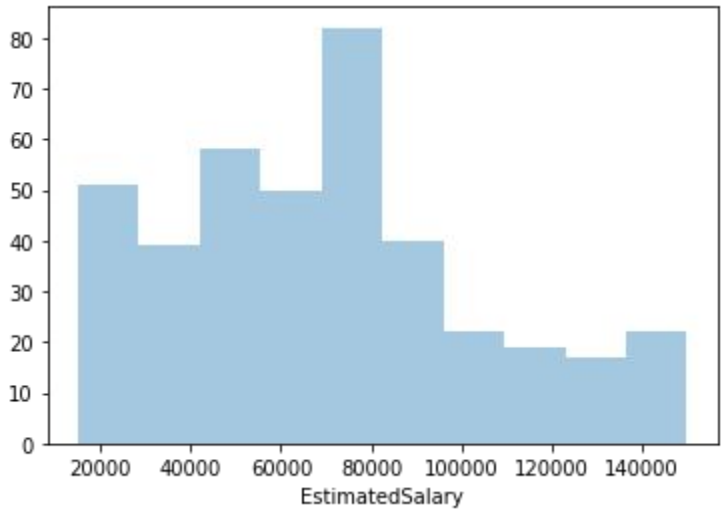
5. Scatter plot age vs estimated salary vs purchased on dataset ‘social network.csv’

```
"""
Created on Wed Feb  2 12:25:39 2022

@author: KIIT
"""


import pandas as pd
import seaborn as sns

df = pd.read_csv('Social_Network_Ads.csv')
#print(df)

sns.set(style='darkgrid')

sns.lmplot(
        x='Age',
        y='EstimatedSalary',
        data=df,
        hue='Purchased',
        fit_reg=False,
        legend=True,
        )
```

6. Plot Histogram for estimated salary attribute on dataset ‘social network.csv’
A) with default kernel density estimate
B) Without kernel density estimate

```
"""
Created on Wed Feb  2 12:25:39 2022

@author: KIIT
"""

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('https://raw.githubusercontent.com/hamdan-codes/tnt-lab-6th-sem/main/Lab%202/Social_Network_Ads.csv')

print('A')
ax = sns.distplot(
      df.EstimatedSalary,
      bins=5
      )
plt.show()

print('B')
ax = sns.distplot(
      df.EstimatedSalary,
      bins=10,
      kde=False
      )
plt.show()
```


1e−5
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Density
−25000
0
25000
50000
75000
100000
125000
150000
175000
EstimatedSalary


80
70
60
50
40
30
20
10
0
20000
40000
60000
80000
100000
120000
140000
EstimatedSalary

7. a)show Bar plot frequency distribution of country attribute on dataset ‘data.csv’
b)show Grouped bar plot of country and purchased
c)show Box and whiskers plot for age vs country

```
"""
Created on Wed Feb  2 12:25:39 2022

@author: KIIT
"""

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Data.csv')

print('A')
sns.countplot(
      x='Country',
      data=df
      )
plt.show()

print('B')
sns.countplot(
      x='Country',
      data=df,
      hue='Purchased'
      )
plt.show()

print('C')
sns.boxplot(
      x='Age',
      y='Country',
```

```
        hue='Purchased',
        data=df
        )
plt.show()
```