Pooja Agarwal
1905330

Lab 3

Q1: 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.

CODE:

```
1   import pandas as pd
2   import matplotlib.pyplot as plt
3
4   df = pd.read_csv('Social_Network_Ads.csv')
5   plt.hist(
6       df['EstimatedSalary'],
7       color='blue',
8       edgecolor='white',
9       bins=4
10      )
11  plt.show()
```

GRAPH:

Pooja Agarwal
1905330
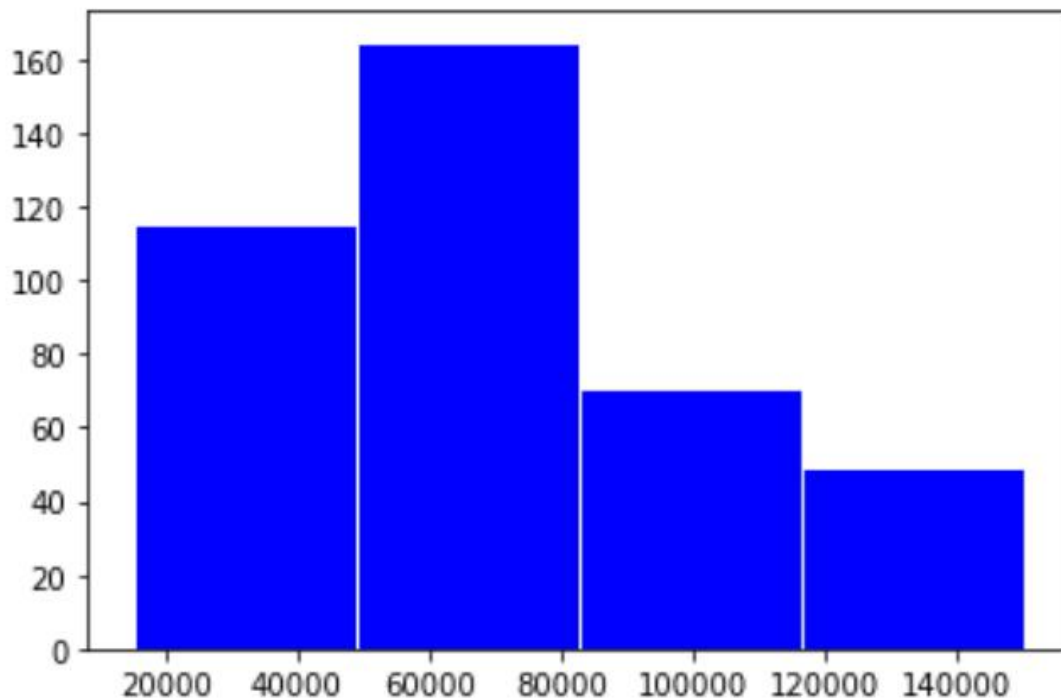
Q2: On the dataset 'data', draw barplot to show the count of categorical feature 'Country'.

CODE:

```
1   import numpy as np
2   import pandas as pd
3   import matplotlib.pyplot as plt
4
5   df = pd.read_csv('Data.csv')
6   df.dropna(axis=0, inplace=True)
7
8   index = np.arange(len(df['Country'].unique()))
9   counts = [3,3,3]
10  plt.bar(index, counts, color = ['blue', 'red', 'green'])
11  plt.xticks(index, df.Country.unique())
12
13
14  plt.show()
```

GRAPH:

Pooja Agarwal
1905330
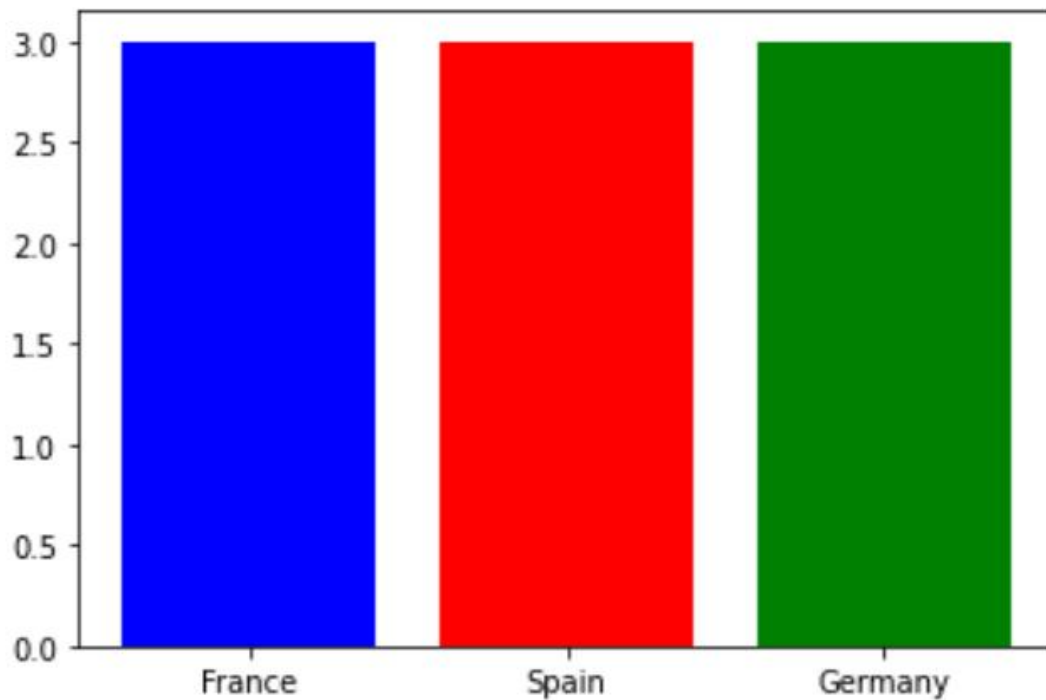
Q3: Remove missing values from the dataframe ceated from dataset 'data'
and display the dimension of dataframe in both cases.

CODE:

```
1    import pandas as pd
2
3    df = pd.read_csv('Data.csv')
4
5    print('Before removing NA Values:', df.shape)
6
7    df.dropna(axis=0, inplace=True)
8
9    print('After removing NA Values:', df.shape)
```

OUTPUT:

```
In [3]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab3/q3.py', wdir='C:/Users/KIIT/
Desktop/Assignments/TNT/Lab3')
Before removing NA Values: (10, 4)
After removing NA Values: (7, 4)
```

Pooja Agarwal
1905330

Q4: 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
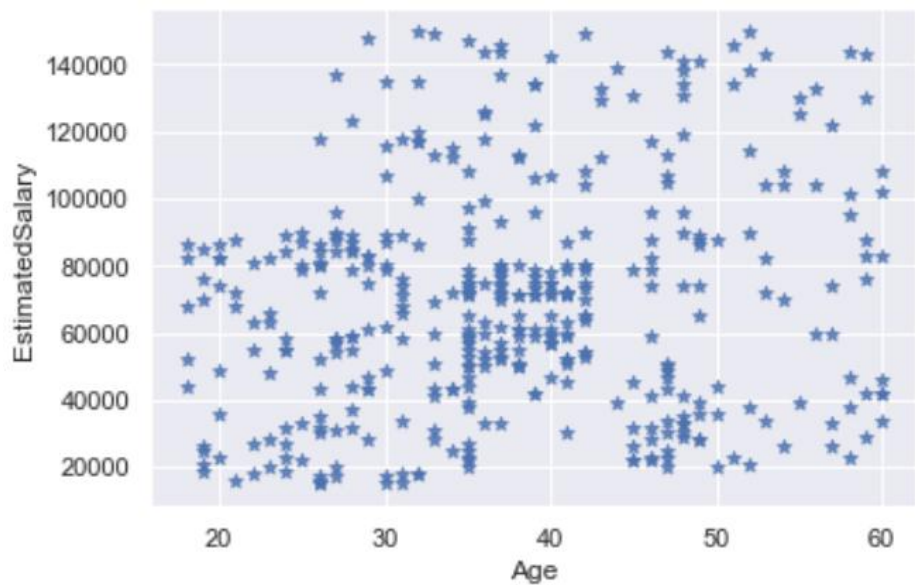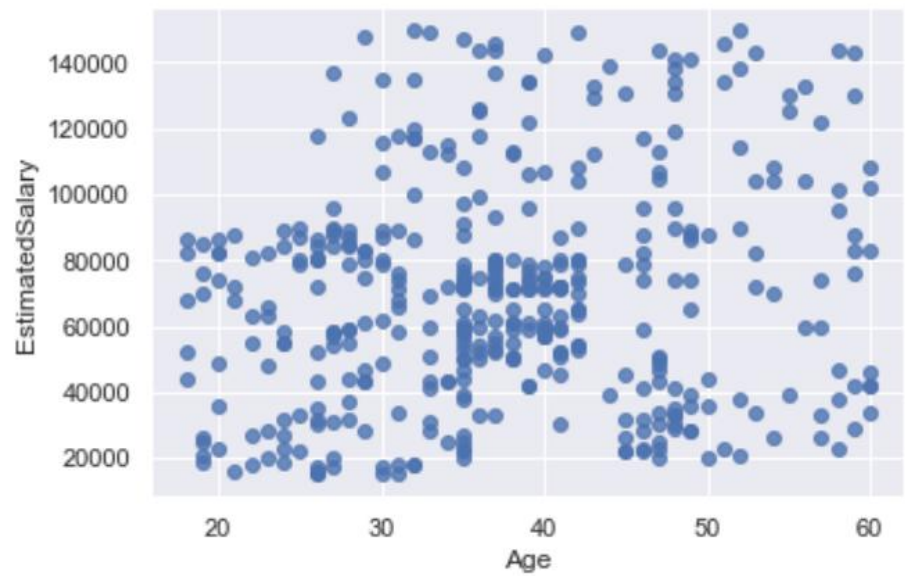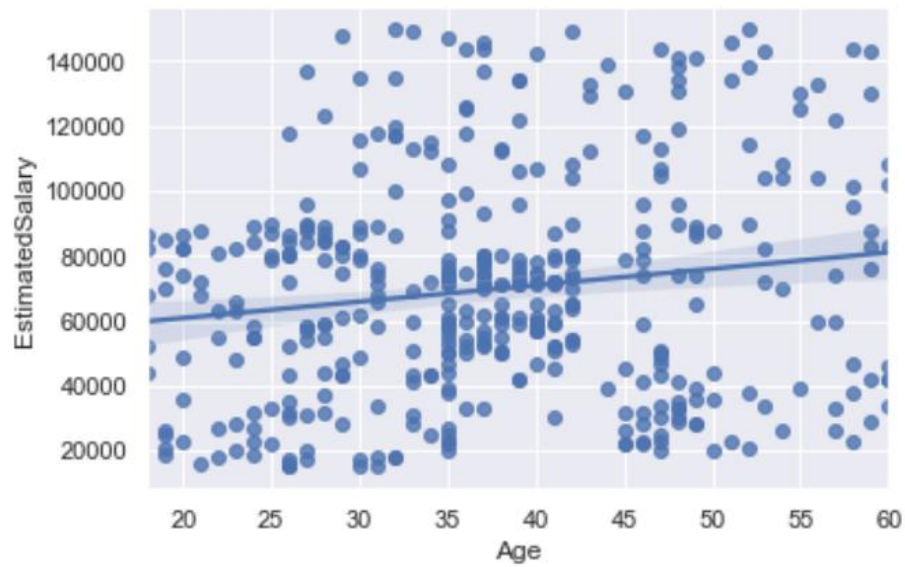
CODE:

```
1    import pandas as pd
2    import seaborn as sns
3    import matplotlib.pyplot as plt
4
5    df = pd.read_csv('Social_Network_Ads.csv')
6    #print(df)
7
8    sns.set(style='darkgrid')
9
10   print('A')
11   sns.regplot(
12       x=df.Age,
13       y=df.EstimatedSalary,
14       )
15   plt.show()
16
17   print('B')
18   sns.regplot(
19       x=df.Age,
20       y=df.EstimatedSalary,
21       fit_reg=False
22       )
23   plt.show()
24
25   print('C')
26   sns.regplot(
27       x=df.Age,
28       y=df.EstimatedSalary,
29       fit_reg=False,
30       marker='*'
31       )
```

OUTPUT:

```
In [4]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab3/q4.py', wdir='C:/Users/KIIT/
Desktop/Assignments/TNT/Lab3')
A
B
C
```

Pooja Agarwal
1905330

GRAPH:

Q5: Scatter plot age vs estimated salary vs purchased on dataset 'social network.csv'.

CODE:

```
1   import pandas as pd
2   import seaborn as sns
3
4   df = pd.read_csv('Social_Network_Ads.csv')
5   #print(df)
6
7   sns.set(style='darkgrid')
8
9   sns.lmplot(
10          x='Age',
11          y='EstimatedSalary',
12          data=df,
13          hue='Purchased',
14          fit_reg=False,
15          legend=True,
16          )
```

GRAPH:

Pooja Agarwal
1905330
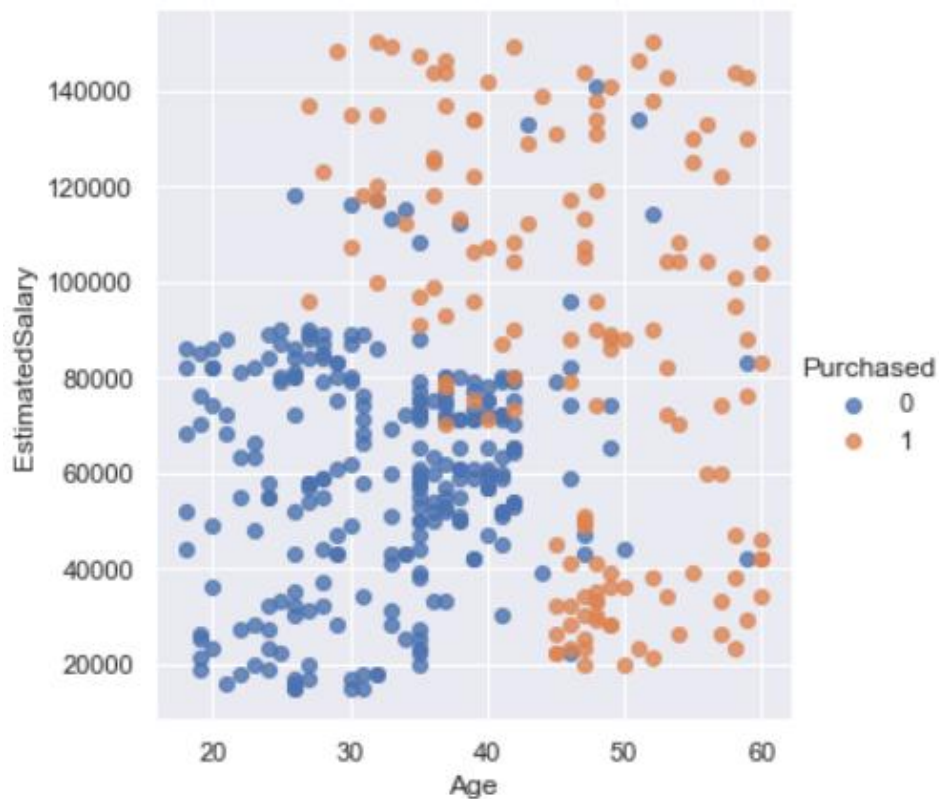
Q6: Plot Histogram for estimated salary attribute on dataset 'social network.csv'
A) with default kernel density estimate
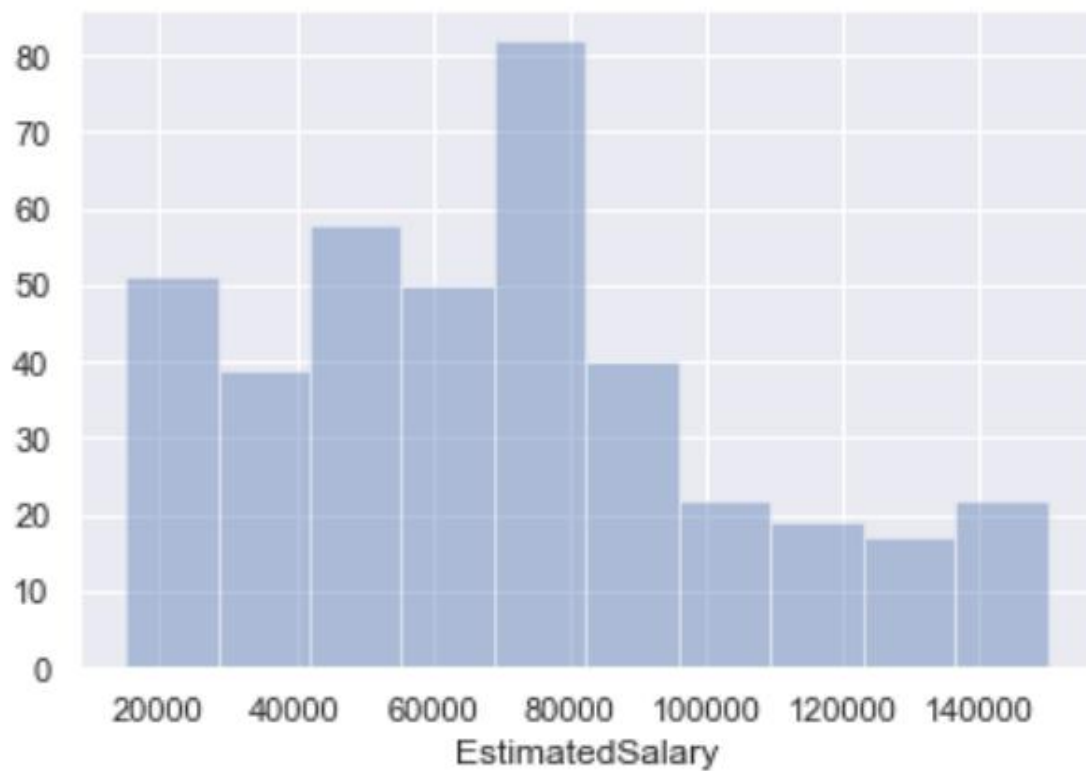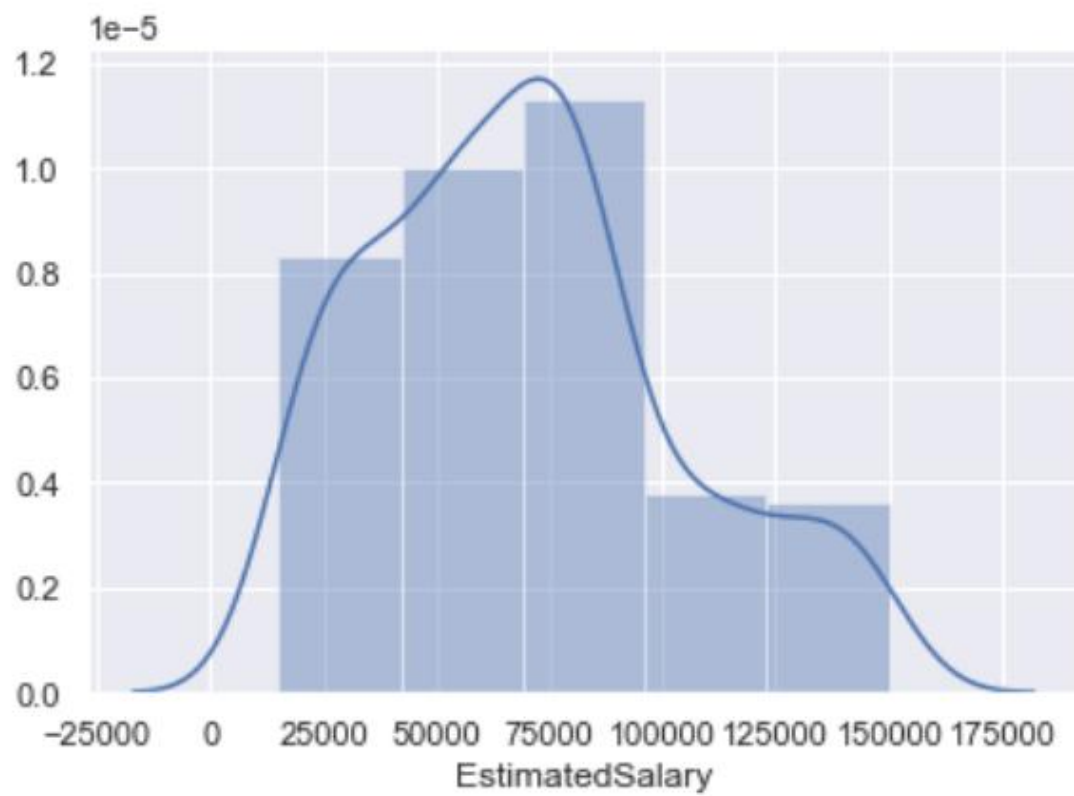B) Without kernel density estimate

CODE:

```
1   import pandas as pd
2   import seaborn as sns
3   import matplotlib.pyplot as plt
4
5   df = pd.read_csv('Social_Network_Ads.csv')
6
7   print('A')
8   ax = sns.distplot(
9           df.EstimatedSalary,
10          bins=5
11          )
12  plt.show()
13
14  print('B')
15  ax = sns.distplot(
16          df.EstimatedSalary,
17          bins=10,
18          kde=False
19          )
20  plt.show()
21
```

OUTPUT:

```
In [6]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab3/q6.py', wdir='C:/Users/KIIT/
Desktop/Assignments/TNT/Lab3')
A
B
```

Pooja Agarwal
1905330

GRAPH:

Q7: 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

CODE:

```python
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()
```

OUTPUT:

```
In [7]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab3/q7.py', wdir='C:/Users/KIIT/
Desktop/Assignments/TNT/Lab3')
A
B
C
```

Pooja Agarwal
1905330

GRAPH: