```
import seaborn as sns
import matplotlib as mp
import math
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
df=pd.read_csv("C:\\New Folder\\CDR_Purchase.csv")
df
```

```
   User ID  Gender  Age  EstimatedSalary  Purchased
0 15624510    Male   19            19000          0
1 15628060    Male   35            20000          0
2 15643093  Female   26            43000          0
3 15708651  Female   27            57000          0
4 15708076    Male   19            76000          0
5 15724883    Male   27            58000          0
6 15728773    Male   27            84000          0
7 15733883  Female   48            65000          0
8 15746811  Female   19            48000          0
9 15748078    Male   53            23000          1
10 15749160  Female   19            76000          0
11 15753944    Male   30            29000          0
12 15755062  Female   19            65000          0
```

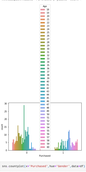
400 rows × 5 columns

```
sns.countplot(df['Purchased',df]
```

```
<Axes:xlabel='Purchased', ylabel='count'>
```

```
sns.countplot(df['Purchased',hue='age',df])
```

```
<Axes:xlabel='Purchased', ylabel='count'>
```




```
sns.countplot(df['Purchased',hue='sex',df])
```

```
<Axes:xlabel='Purchased', ylabel='count'>
```

```
df['Purchased',plot.hist()
```

```
<Axes:xlabel='Purchased'>
```

```
df['Purchased',sort_values()
```

```
<Axes:xlabel='Purchased'>
```

```
df.shape
```

```
(400, 5)
```

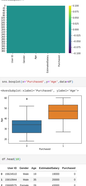
wrangling data

```
df.isna().sum()
```

```
User ID          0
Gender           0
Age              0
EstimatedSalary  0
Purchased        0
dtype: int64
```

```
sns.heatmap(df.isna())<matplotlib.axes._subplots.AxesSubplot:>
```

```
<Axes:>
```




```
sns.boxplot('Purchased',y='age',df)
```

```
<Axes:xlabel='Purchased', ylabel='Age'>
```

```
df.head(4)
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15628060	Male	35	20000	0
2	15643093	Female	26	43000	0
3	15708651	Female	27	57000	0

```
df[df.get_dummies('Purchased',drop_first=True),
get_head()
```

```
Male
0    1
1    1
2    0
3    0
dtype: int64
```

```
df[df.concat([df,gen],axis=True)
df.head()
```

	User ID	Gender	Age	EstimatedSalary	Purchased	Male
0	15624510	Male	19	19000	0	1
1	15628060	Male	35	20000	0	1
2	15643093	Female	26	43000	0	0
3	15708651	Female	27	57000	0	0
4	15708076	Male	19	76000	0	1

```
df[df.drop('User ID','Gender',axis=1)
df.head()
```

	Age	EstimatedSalary	Purchased	Male
0	19	19000	0	1
1	35	20000	0	1
2	26	43000	0	0
3	27	57000	0	0
4	19	76000	0	1

```
y.head()
```

```
0    0
1    0
2    0
3    0
4    0
Name: Purchased, dtype: int64
```

training the data

```
x1,y1=df[df['Purchased',axis=1]
y1.df['Purchased']
x1.head()
```

	Age	EstimatedSalary	Male
0	19	19000	1
1	35	20000	1
2	26	43000	0
3	27	57000	0
4	19	76000	1

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=1)
```

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
X_train=scaler.fit_transform(X_train)
X_test=scaler.transform(X_test)
```

```
array([[-0.95258966, -1.05117195,
        1.79667907,  1.60020993,
        1.60020993,  1.60020993,
        1.60020993,  1.60020993,
        1.60020993
```