1 . Importing Required Package

```
import pandas as pd
import numpy as np
import seaborn as sbn
import matplotlib.pyplot as plt
```

2 . Loading the Dataset

```
db = pd.read_csv('/content/Mall_Customers.csv')
db
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

3. Visualizations

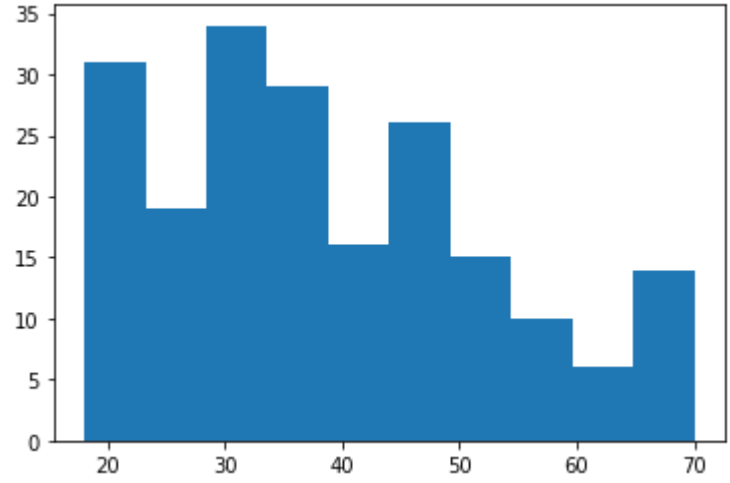
3.1 UniVariate Analysis

```
plt.hist(db['Annual Income (k$)'])
```

```
(array([24., 22., 28., 38., 30., 36.,  8.,  6.,  4.,  4.]),
 array([ 15. ,  27.2,  39.4,  51.6,  63.8,  76. ,  88.2, 100.4, 112.6,
        124.8, 137. ]),
 <a list of 10 Patch objects>)
```

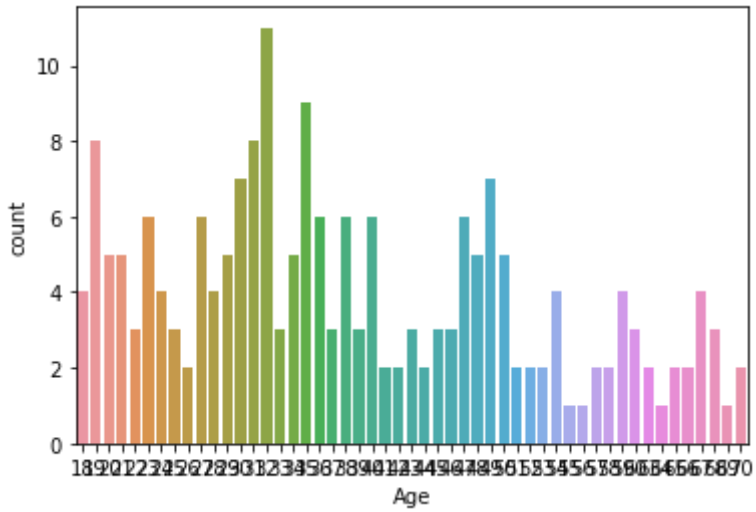

```
plt.hist(db['Age'])
```

```
(array([31., 19., 34., 29., 16., 26., 15., 10.,  6., 14.]),
 array([18. , 23.2, 28.4, 33.6, 38.8, 44. , 49.2, 54.4, 59.6, 64.8, 70. ]),
 <a list of 10 Patch objects>)
```



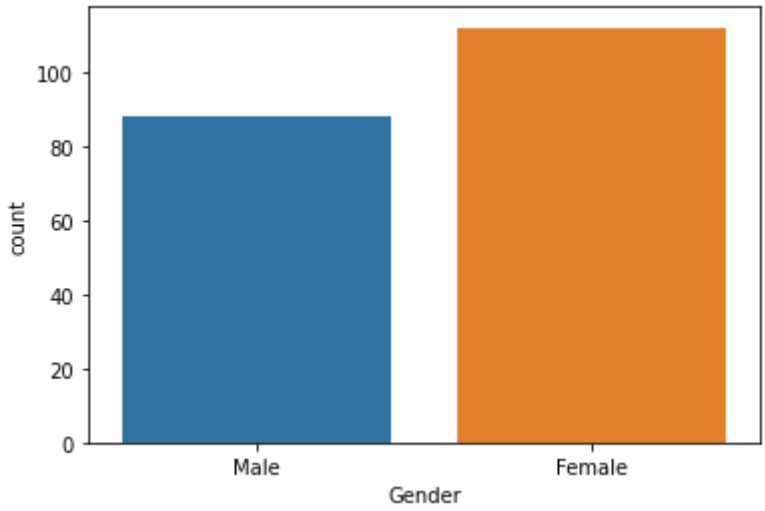

```
sbn.countplot(db['Age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: P
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fbe87154710>
```



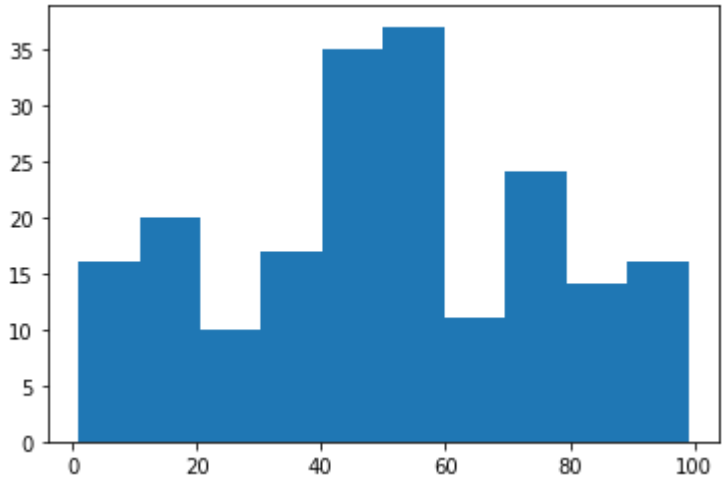
```
sbn.countplot(db['Gender'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: P
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fbe86fd98d0>
```




```
plt.hist(db['Spending Score (1-100)'])
```

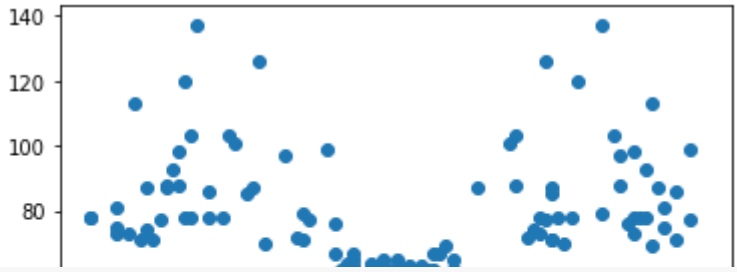
```
(array([16., 20., 10., 17., 35., 37., 11., 24., 14., 16.]),
 array([ 1. , 10.8, 20.6, 30.4, 40.2, 50. , 59.8, 69.6, 79.4, 89.2, 99. ]),
 <a list of 10 Patch objects>)
```



3.2 Bi-Variate Analysis

```
plt.scatter(db['Spending Score (1-100)'],db['Annual Income (k$)'])
```

```
<matplotlib.collections.PathCollection at 0x7fbe86e29d10>
```

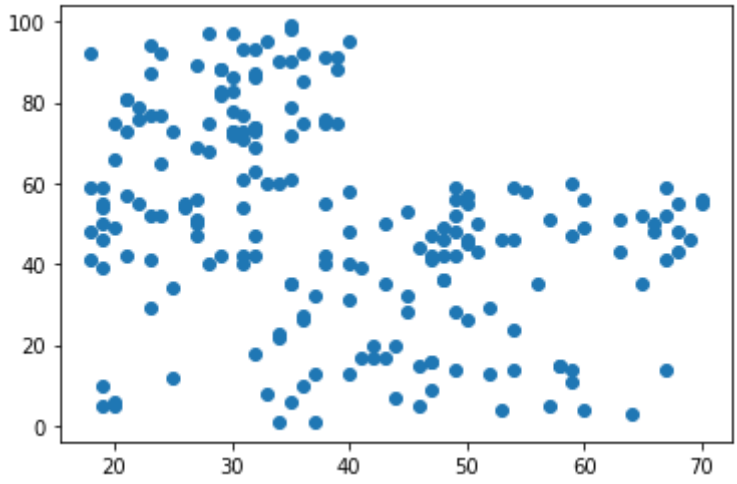



```
plt.scatter(db['Gender'],db['Annual Income (k$)'])
```

```
<matplotlib.collections.PathCollection at 0x7fbe86e17e90>
```


```
plt.scatter(db['Age'],db['Spending Score (1-100)'])
```

```
<matplotlib.collections.PathCollection at 0x7fbe86d798d0>
```




```
plt.scatter(db['Age'],db['Annual Income (k$)'])
```

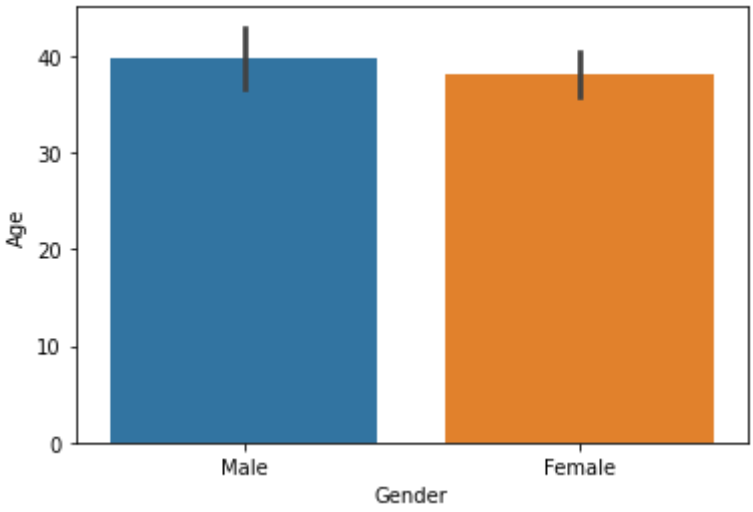
```
<matplotlib.collections.PathCollection at 0x7fbe86ced210>
```


```
sbn.heatmap(db.corr(), annot = True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fbe86ca3590>
```


```
sbn.barplot(db['Gender'], db['Age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fbe86c53990>
```



3.3 Multi-Variate Analysis

[ ] ↳ *2 cells hidden*

4. Perform descriptive statistics on the dataset

[ ] ↳ *6 cells hidden*

5. Check for Missing values and deal with them

[ ] ↳ *3 cells hidden*

6 . Find the outliers and replace them outliers

[ ] ↳ *13 cells hidden*

7. Check for Categorical columns and perform encoding

[ ] ↳ *4 cells hidden*

8 . Scaling the data

[ ] ↳ *1 cell hidden*

9 . Perform any of the clustering algorithms

[ ] ↳ *5 cells hidden*

10 . Add the cluster data with the primary dataset

[ ] ↳ *2 cells hidden*

11 . Split the data into dependent and independent variables

[ ] ↳ *2 cells hidden*

12 . Split the data into training and testing

[ ] ↳ *4 cells hidden*

13 . Build the Model

[ ] ↳ *1 cell hidden*

14. Train the Model

[ ] ↳ *1 cell hidden*

15 . Test the Model

[ ] ↳ *1 cell hidden*

16. Measure the performance using Evaluation Metrics

[ ] ↳ *1 cell hidden*