Customer Segmentation Analysis

In [ ]:

```
#import required libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
```

Load Dataset

In [1]:

```
df=pd.read_csv('Mall_Customers.csv')
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
 in
----> 1 df=pd.read_csv('Mall_Customers.csv')

NameError: name 'pd' is not defined
```

In [ ]:

```
dff=df.copy() #Make a copy of dataset
```

In [ ]:

```
dff.head()
```

Out[ ]:

	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

In [ ]:

```
dff.tail()
```

Out[ ]:

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
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

In [ ]:

```
dff.shape
```

Out[ ]:

```
(200, 5)
```

In [ ]:

```
dff.info()
```

```
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Gender                  200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

Handle Missing Values

In [ ]:

```
dff.isnull().any() #In this Dataset,there is no missing value
```

Out[ ]:

```
CustomerID                False
Gender                    False
Age                       False
Annual Income (k$)        False
Spending Score (1-100)    False
dtype: bool
```

In [ ]:

```
dff.Gender.value_counts()
```

Out[ ]:

```
Female    112
Male       88
Name: Gender, dtype: int64
```

In [ ]:

```
dff=dff.drop(columns=['CustomerID'],axis=1) #Drop unused column
```

In [ ]:

```
dff.head()
```

Out[ ]:

	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40

Data Visualization

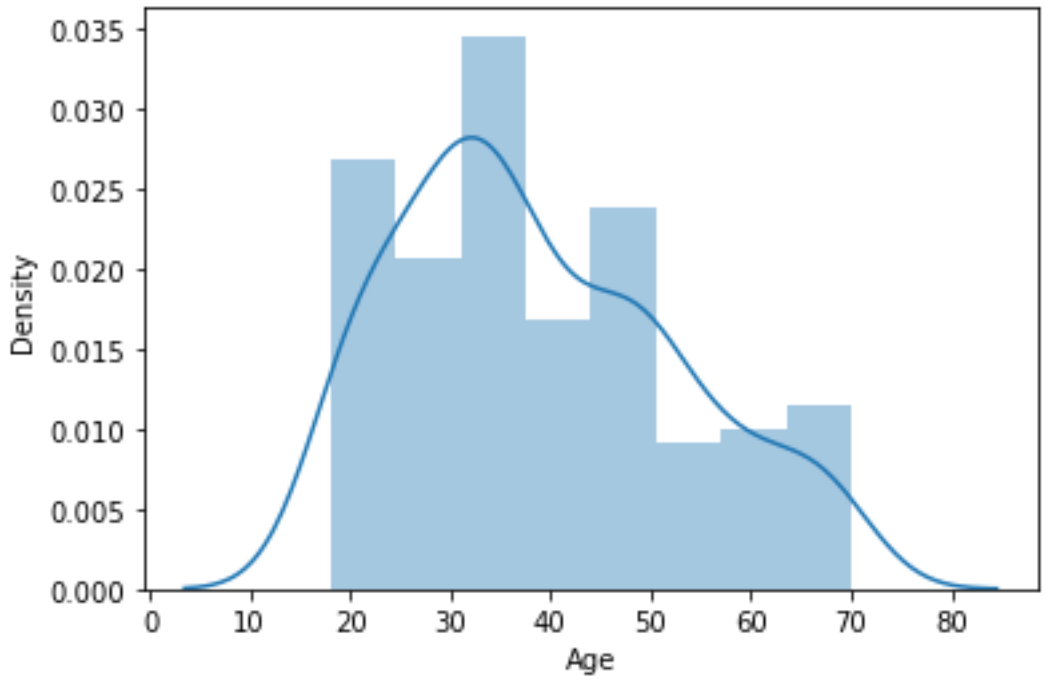
i. Uni-variate Analysis

In [ ]:

```
sns.distplot(dff.Age)
```

```
C:\Users\God\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Futur
eWarning: `distplot` is a deprecated function and will be removed in a future
 version. Please adapt your code to use either `displot` (a figure-level func
tion with similar flexibility) or `histplot` (an axes-level function for hist
ograms).
  warnings.warn(msg, FutureWarning)
```

Out[ ]:



The Age values lies between 18 and 70