Implement K-Mean cluster algorithm on Mall_Customers dataset by finding number of clusters using Elbow Method. ¶

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [3]:
```
data=pd.read_csv('C:/satish (coding)/csv files/Mall_Customers.csv')
```

In [4]:
```
data
```

Out[4]:

	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
...	...	...	...	...	...
245	246	Male	30	297	69
246	247	Female	56	311	14
247	248	Male	29	313	90
248	249	Female	19	316	32
249	250	Female	31	325	86

250 rows × 5 columns

In [5]:
```
x=data.iloc[:,[2,3]]
```

In [6]:
```
x
```

Out[6]:

	Age	Annual Income (k$)
0	19	15
1	21	15
2	20	16
3	23	16
4	31	17
...	...	...
245	30	297
246	56	311
247	29	313
248	19	316
249	31	325

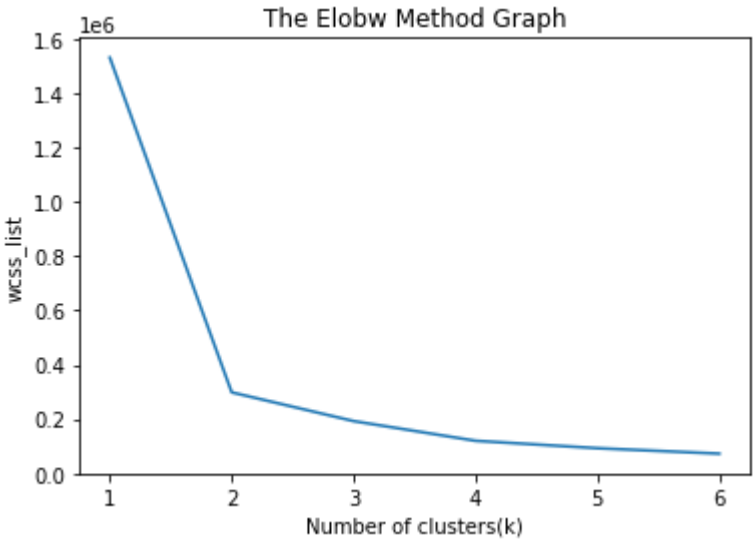
250 rows × 2 columns

In [7]:
```
from sklearn.cluster import KMeans
```

In [17]:
```
wcss_list= []
for i in range(1, 7):
    kmeans = KMeans(n_clusters=i,init='k-means++', random_state= 42)
    kmeans.fit(x)
    wcss_list.append(kmeans.inertia_)
plt.plot(range(1, 7), wcss_list)
plt.title('The Elobw Method Graph')
plt.xlabel('Number of clusters(k)')
plt.ylabel('wcss_list')
plt.show()
```

```
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
```

```
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:870: FutureWarning: The default value of `n_init` will change fro
m 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the w
arning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster
\_kmeans.py:1382: UserWarning: KMeans is known to have a memory leak on Windo
ws with MKL, when there are less chunks than available threads. You can avoid
it by setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
```



In [18]:
```
kmeans=KMeans(n_clusters=2,init='k-means++',random_state=43)
```

In [19]:
```
y_predict=kmeans.fit_predict(x)
```

```
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster\_k
means.py:870: FutureWarning: The default value of `n_init` will change from 10
to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\91733\AppData\Roaming\Python\Python39\site-packages\sklearn\cluster\_k
means.py:1382: UserWarning: KMeans is known to have a memory leak on Windows wi
th MKL, when there are less chunks than available threads. You can avoid it by
setting the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
```

In [20]:
```
y_predict
```

Out[20]:
```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1])
```

In [21]:
```
x['pre']=y_predict
x
```

```
C:\Users\91733\AppData\Local\Temp\ipykernel_23044\475348706.py:1: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  x['pre']=y_predict
```

Out[21]:

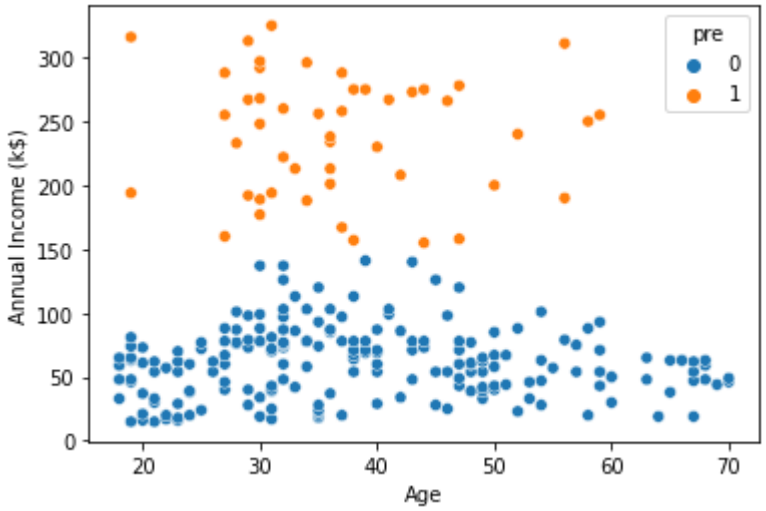
	Age	Annual Income (k$)	pre
0	19	15	0
1	21	15	0
2	20	16	0
3	23	16	0
4	31	17	0
...	...	...	...
245	30	297	1
246	56	311	1
247	29	313	1
248	19	316	1
249	31	325	1

250 rows × 3 columns

In [28]:
```
import seaborn as sns

sns.scatterplot(x='Age',y='Annual Income (k$)',hue='pre',data=x)
```

Out[28]: <AxesSubplot:xlabel='Age', ylabel='Annual Income (k$)'>



In [29]:
```
kmeans.
```

Out[29]:
```
array([[ 38.87128713,  61.35148515,   2.52475248],
       [ 36.89583333, 239.6875    ,   1.375     ]])
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: