

```
In [1]: import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
import matplotlib.pyplot as mtp
from sklearn.cluster import KMeans
```

```
In [2]:
```

Out[2]:

	ID	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	100000001	0	0	67	2	124670	1	2
1	100000002	1	1	22	1	150773	1	2
2	100000003	0	0	49	1	89210	0	0
3	100000004	0	0	45	1	171565	1	1
4	100000005	0	0	53	1	149031	1	1

```
In [6]:
```

Out[6]:

	ID	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
1995	False	False	False	False	False	False	False	False
1996	False	False	False	False	False	False	False	False
1997	False	False	False	False	False	False	False	False
1998	False	False	False	False	False	False	False	False
1999	False	False	False	False	False	False	False	False

2000 rows × 8 columns

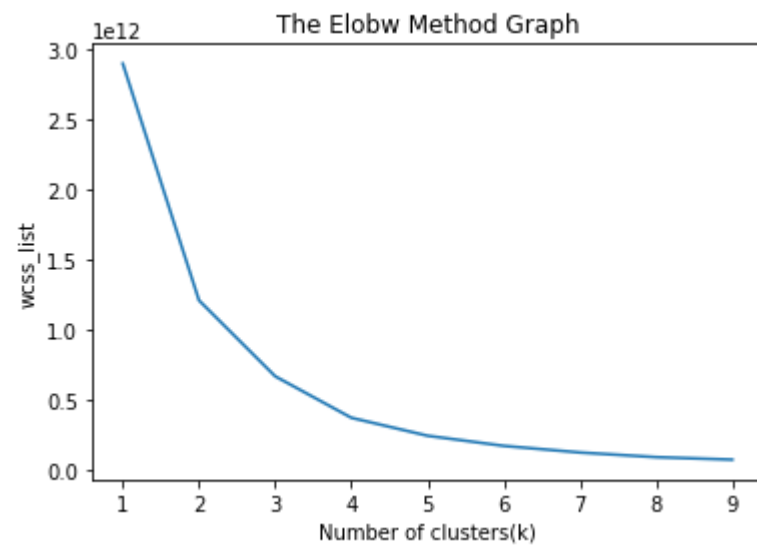
```
In [12]: x=df.iloc[:,1:]
```

Out[12]:

	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
0	0	0	67	2	124670	1	2
1	1	1	22	1	150773	1	2
2	0	0	49	1	89210	0	0
3	0	0	45	1	171565	1	1
4	0	0	53	1	149031	1	1
...
1995	1	0	47	1	123525	0	0
1996	1	1	27	1	117744	1	0
1997	0	0	31	0	86400	0	0
1998	1	1	24	1	97968	0	0
1999	0	0	25	0	68416	0	0

2000 rows × 7 columns

```
In [20]: from sklearn.cluster import KMeans
import matplotlib.pyplot as mtp
wcss=[]
for i in range(1,10):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state= 42)
    kmeans.fit(x)
    wcss.append(kmeans.inertia_)
mtp.plot(range(1, 10), wcss)
mtp.title('The Elobw Method Graph')
mtp.xlabel('Number of clusters(k)')
mtp.ylabel('wcss_list')
mtp.show()
```



```
In [37]: #training the K-means model on a dataset
kmeans = KMeans(n_clusters=2, init='k-means++', random_state= 42)
```

```
In [38]: ...
```

```
Out[38]: array([0, 1, 0, ..., 0, 0, 0])
```

```
In [39]: from sklearn.metrics import silhouette_score
Silhouette Score(n=2): 0.5834469001696239
```

```
In [ ]:
```