

K-Means Clustering

In []:

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

In [3]:

```
data_set = pd.read_csv("https://raw.githubusercontent.com/AP-State-Skill-Development-Co
rporation/Datasets/master/Classification/Social_Network_Ads.csv")
data_set.head()
```

Out[3]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

In [4]:

```
data_set.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 5 columns):
User ID          400 non-null int64
Gender           400 non-null object
Age              400 non-null int64
EstimatedSalary  400 non-null int64
Purchased        400 non-null int64
dtypes: int64(4), object(1)
memory usage: 15.7+ KB
```

In [5]:

```
data_set.columns
```

Out[5]:

```
Index(['User ID', 'Gender', 'Age', 'EstimatedSalary', 'Purchased'], dtype
='object')
```

In [16]:

```
X = data_set.iloc[:,[3,4]].values  
print(X)
```

```
[[ 19000    0]
 [ 20000    0]
 [ 43000    0]
 [ 57000    0]
 [ 76000    0]
 [ 58000    0]
 [ 84000    0]
 [150000    1]
 [ 33000    0]
 [ 65000    0]
 [ 80000    0]
 [ 52000    0]
 [ 86000    0]
 [ 18000    0]
 [ 82000    0]
 [ 80000    0]
 [ 25000    1]
 [ 26000    1]
 [ 28000    1]
 [ 29000    1]
 [ 22000    1]
 [ 49000    1]
 [ 41000    1]
 [ 22000    1]
 [ 23000    1]
 [ 20000    1]
 [ 28000    1]
 [ 30000    1]
 [ 43000    0]
 [ 18000    0]
 [ 74000    0]
 [137000    1]
 [ 16000    0]
 [ 44000    0]
 [ 90000    0]
 [ 27000    0]
 [ 28000    0]
 [ 49000    0]
 [ 72000    0]
 [ 31000    0]
 [ 17000    0]
 [ 51000    0]
 [108000    0]
 [ 15000    0]
 [ 84000    0]
 [ 20000    0]
 [ 79000    0]
 [ 54000    0]
 [135000    1]
 [ 89000    0]
 [ 32000    0]
 [ 44000    0]
 [ 83000    0]
 [ 23000    0]
 [ 58000    0]
 [ 55000    0]
 [ 48000    0]
 [ 79000    0]
 [ 18000    0]
 [117000    0]
 [ 20000    0]
```

```
[ 87000    0]
[ 66000    0]
[120000    1]
[ 83000    0]
[ 58000    0]
[ 19000    0]
[ 82000    0]
[ 63000    0]
[ 68000    0]
[ 80000    0]
[ 27000    0]
[ 23000    0]
[113000    0]
[ 18000    0]
[112000    1]
[ 52000    0]
[ 27000    0]
[ 87000    0]
[ 17000    0]
[ 80000    0]
[ 42000    0]
[ 49000    0]
[ 88000    0]
[ 62000    0]
[118000    1]
[ 55000    0]
[ 85000    0]
[ 81000    0]
[ 50000    0]
[ 81000    0]
[116000    0]
[ 15000    0]
[ 28000    0]
[ 83000    0]
[ 44000    0]
[ 25000    0]
[123000    1]
[ 73000    0]
[ 37000    0]
[ 88000    0]
[ 59000    0]
[ 86000    0]
[149000    1]
[ 21000    0]
[ 72000    0]
[ 35000    0]
[ 89000    0]
[ 86000    0]
[ 80000    0]
[ 71000    0]
[ 71000    0]
[ 61000    0]
[ 55000    0]
[ 80000    0]
[ 57000    0]
[ 75000    0]
[ 52000    0]
[ 59000    0]
[ 59000    0]
[ 75000    0]
[ 72000    0]
```

[75000	0]
[53000	0]
[51000	0]
[61000	0]
[65000	0]
[32000	0]
[17000	0]
[84000	0]
[58000	0]
[31000	0]
[87000	0]
[68000	0]
[55000	0]
[63000	0]
[82000	0]
[107000	1]
[59000	0]
[25000	0]
[85000	0]
[68000	0]
[59000	0]
[89000	0]
[25000	0]
[89000	0]
[96000	1]
[30000	0]
[61000	0]
[74000	0]
[15000	0]
[45000	0]
[76000	0]
[50000	0]
[47000	0]
[15000	0]
[59000	0]
[75000	0]
[30000	0]
[135000	1]
[100000	1]
[90000	0]
[33000	0]
[38000	0]
[69000	0]
[86000	0]
[55000	0]
[71000	0]
[148000	1]
[47000	0]
[88000	0]
[115000	0]
[118000	0]
[43000	0]
[72000	0]
[28000	0]
[47000	0]
[22000	0]
[23000	0]
[34000	0]
[16000	0]
[71000	0]
[117000	1]

```
[ 43000    0]
[ 60000    0]
[ 66000    0]
[ 82000    0]
[ 41000    0]
[ 72000    0]
[ 32000    0]
[ 84000    0]
[ 26000    0]
[ 43000    0]
[ 70000    0]
[ 89000    0]
[ 43000    0]
[ 79000    0]
[ 36000    0]
[ 80000    0]
[ 22000    0]
[ 39000    0]
[ 74000    0]
[134000    1]
[ 71000    0]
[101000    1]
[ 47000    0]
[130000    1]
[114000    0]
[142000    1]
[ 22000    0]
[ 96000    1]
[150000    1]
[ 42000    0]
[ 58000    0]
[ 43000    0]
[108000    1]
[ 65000    0]
[ 78000    0]
[ 96000    0]
[143000    1]
[ 80000    0]
[ 91000    1]
[144000    1]
[102000    1]
[ 60000    0]
[ 53000    0]
[126000    1]
[133000    1]
[ 72000    0]
[ 80000    1]
[147000    1]
[ 42000    0]
[107000    1]
[ 86000    1]
[112000    0]
[ 79000    1]
[ 57000    0]
[ 80000    0]
[ 82000    0]
[143000    1]
[149000    1]
[ 59000    0]
[ 88000    1]
[104000    1]
```

```
[ 72000    0]
[146000    1]
[ 50000    0]
[122000    1]
[ 52000    0]
[ 97000    1]
[ 39000    0]
[ 52000    0]
[134000    1]
[146000    1]
[ 44000    0]
[ 90000    1]
[ 72000    0]
[ 57000    0]
[ 95000    1]
[131000    1]
[ 77000    0]
[144000    1]
[125000    1]
[ 72000    0]
[ 90000    1]
[108000    1]
[ 75000    0]
[ 74000    0]
[144000    1]
[ 61000    0]
[133000    0]
[ 76000    1]
[ 42000    1]
[106000    1]
[ 26000    1]
[ 74000    1]
[ 71000    0]
[ 88000    1]
[ 38000    1]
[ 36000    1]
[ 88000    1]
[ 61000    0]
[ 70000    1]
[ 21000    1]
[141000    0]
[ 93000    1]
[ 62000    0]
[138000    1]
[ 79000    0]
[ 78000    1]
[134000    1]
[ 89000    1]
[ 39000    1]
[ 77000    0]
[ 57000    0]
[ 63000    0]
[ 73000    1]
[112000    1]
[ 79000    0]
[117000    1]
[ 38000    1]
[ 74000    1]
[137000    1]
[ 79000    1]
[ 60000    0]
```

[54000	0]
[134000	0]
[113000	1]
[125000	1]
[50000	0]
[70000	0]
[96000	1]
[50000	0]
[141000	1]
[79000	0]
[75000	1]
[104000	1]
[55000	0]
[32000	1]
[60000	0]
[138000	1]
[82000	1]
[52000	0]
[30000	1]
[131000	1]
[60000	0]
[72000	0]
[75000	0]
[118000	1]
[107000	1]
[51000	0]
[119000	1]
[65000	0]
[65000	0]
[60000	1]
[54000	0]
[144000	1]
[79000	0]
[55000	0]
[122000	1]
[104000	1]
[75000	0]
[65000	0]
[51000	1]
[105000	1]
[63000	0]
[72000	1]
[108000	1]
[77000	0]
[61000	0]
[113000	1]
[75000	0]
[90000	1]
[57000	0]
[99000	1]
[34000	1]
[70000	1]
[72000	0]
[71000	1]
[54000	0]
[129000	1]
[34000	1]
[50000	1]
[79000	0]
[104000	1]
[29000	1]


```
[ 47000      1]
[ 88000      1]
[ 71000      0]
[ 26000      1]
[ 46000      1]
[ 83000      1]
[ 73000      0]
[130000      1]
[ 80000      0]
[ 32000      1]
[ 74000      0]
[ 53000      0]
[ 87000      1]
[ 23000      1]
[ 64000      0]
[ 33000      1]
[139000      1]
[ 28000      1]
[ 33000      1]
[ 60000      1]
[ 39000      1]
[ 71000      0]
[ 34000      1]
[ 35000      1]
[ 33000      1]
[ 23000      1]
[ 45000      1]
[ 42000      1]
[ 59000      0]
[ 41000      1]
[ 23000      1]
[ 20000      1]
[ 33000      0]
[ 36000      1]]
```

In [7]:

```
from sklearn.cluster import KMeans
```

In [8]:

```
elbow = []
for i in range(1,11):
    kmeans = KMeans(n_clusters = i,init = "k-means++",random_state = 42)
    kmeans.fit(X)
    elbow.append(kmeans.inertia_)
```

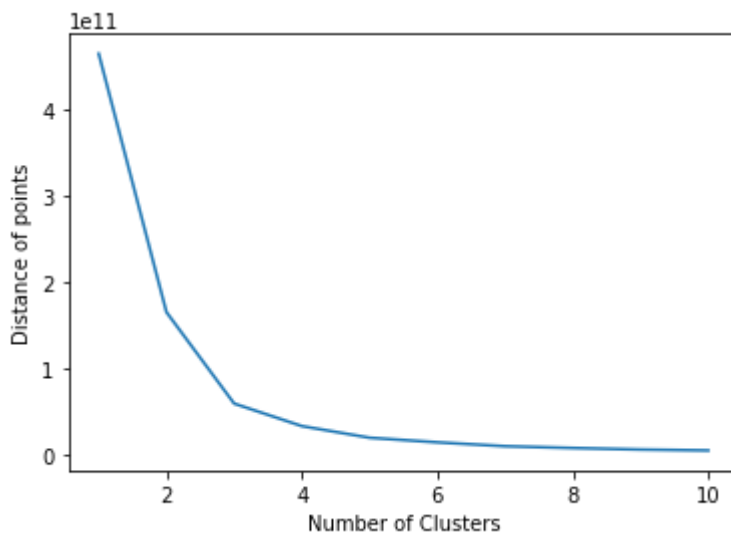
In [9]:

```
print(elbow)
```

```
[463878477591.8775, 165190293074.85718, 59489922495.063545, 33459707476.55
077, 19983360116.390648, 14690873439.203066, 10202678741.900618, 810746288
1.326336, 6348472196.173229, 5283655668.908146]
```

In [10]:

```
plt.plot(range(1,11),elbow)
plt.xlabel("Number of Clusters")
plt.ylabel("Distance of points")
plt.show()
```



In [11]:

```
kmeans = KMeans(n_clusters = 5,init = "k-means++",random_state = 42)
y_means = kmeans.fit_predict(X)
```

In [12]:

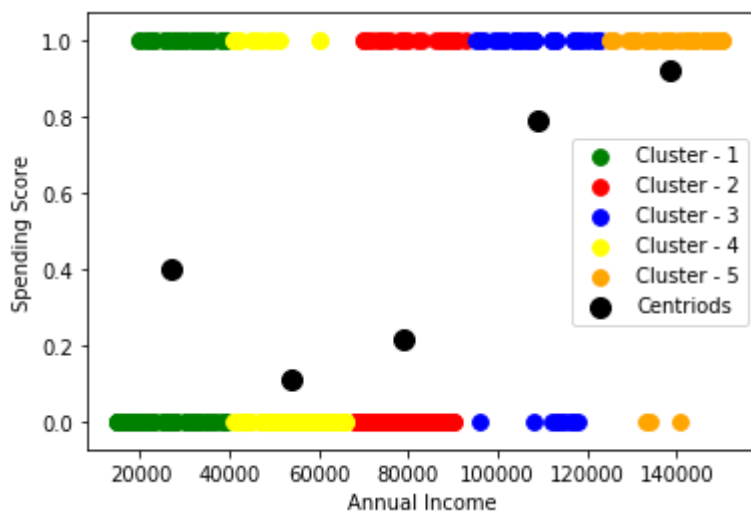
y_means

Out[12]:

```
array([0, 0, 3, 3, 1, 3, 1, 4, 0, 3, 1, 3, 1, 0, 1, 1, 0, 0, 0, 0, 0, 3,
       3, 0, 0, 0, 0, 0, 3, 0, 1, 4, 0, 3, 1, 0, 0, 3, 1, 0, 0, 3, 2, 0,
       1, 0, 1, 3, 4, 1, 0, 3, 1, 0, 3, 3, 3, 1, 0, 2, 0, 1, 3, 2, 1, 3,
       0, 1, 3, 1, 1, 0, 0, 2, 0, 2, 3, 0, 1, 0, 1, 3, 3, 1, 3, 2, 3, 1,
       1, 3, 1, 2, 0, 0, 1, 3, 0, 2, 1, 0, 1, 3, 1, 4, 0, 1, 0, 1, 1, 1,
       1, 1, 3, 3, 1, 3, 1, 3, 3, 3, 1, 1, 1, 3, 3, 3, 3, 0, 0, 1, 3, 0,
       1, 1, 3, 3, 1, 2, 3, 0, 1, 1, 3, 1, 0, 1, 2, 0, 3, 1, 0, 3, 1, 3,
       3, 0, 3, 1, 0, 4, 2, 1, 0, 0, 1, 1, 3, 1, 4, 3, 1, 2, 2, 3, 1, 0,
       3, 0, 0, 0, 0, 1, 2, 3, 3, 3, 1, 3, 1, 0, 1, 0, 3, 1, 1, 3, 1, 0,
       1, 0, 0, 1, 4, 1, 2, 3, 4, 2, 4, 0, 2, 4, 3, 3, 3, 2, 3, 1, 2, 4,
       1, 1, 4, 2, 3, 3, 4, 4, 1, 1, 4, 3, 2, 1, 2, 1, 3, 1, 1, 4, 4, 3,
       1, 2, 1, 4, 3, 2, 3, 2, 0, 3, 4, 4, 3, 1, 1, 3, 2, 4, 1, 4, 4, 1,
       1, 2, 1, 1, 4, 3, 4, 1, 3, 2, 0, 1, 1, 1, 0, 0, 1, 3, 1, 0, 4, 1,
       3, 4, 1, 1, 4, 1, 0, 1, 3, 3, 1, 2, 1, 2, 0, 1, 4, 1, 3, 3, 4, 2,
       4, 3, 1, 2, 3, 4, 1, 1, 2, 3, 0, 3, 4, 1, 3, 0, 4, 3, 1, 1, 2, 2,
       3, 2, 3, 3, 3, 3, 4, 1, 3, 2, 2, 1, 3, 3, 2, 3, 1, 2, 1, 3, 2, 1,
       1, 3, 2, 0, 1, 1, 1, 3, 4, 0, 3, 1, 2, 0, 3, 1, 1, 0, 3, 1, 1, 4,
       1, 0, 1, 3, 1, 0, 3, 0, 4, 0, 0, 3, 0, 1, 0, 0, 0, 0, 3, 3, 3, 3,
       0, 0, 0, 0])
```

In [14]:

```
plt.scatter(X[y_means == 0,0],X[y_means == 0,1],s = 60,c = "green",label = "Cluster - 1")
plt.scatter(X[y_means == 1,0],X[y_means == 1,1],s = 60,c = "red",label = "Cluster - 2")
plt.scatter(X[y_means == 2,0],X[y_means == 2,1],s = 60,c = "blue",label = "Cluster - 3")
plt.scatter(X[y_means == 3,0],X[y_means == 3,1],s = 60,c = "yellow",label = "Cluster - 4")
plt.scatter(X[y_means == 4,0],X[y_means == 4,1],s = 60, c = "orange",label = "Cluster - 5")
plt.scatter(kmeans.cluster_centers[:,0],kmeans.cluster_centers[:,1],s = 100,c = "black",label = "Centriods")
plt.xlabel("Annual Income")
plt.ylabel("Spending Score")
plt.legend(loc = "best")
plt.show()
```



In [15]:

```
print(X)
```

```
[[ 19000    0]
 [ 20000    0]
 [ 43000    0]
 [ 57000    0]
 [ 76000    0]
 [ 58000    0]
 [ 84000    0]
 [150000    1]
 [ 33000    0]
 [ 65000    0]
 [ 80000    0]
 [ 52000    0]
 [ 86000    0]
 [ 18000    0]
 [ 82000    0]
 [ 80000    0]
 [ 25000    1]
 [ 26000    1]
 [ 28000    1]
 [ 29000    1]
 [ 22000    1]
 [ 49000    1]
 [ 41000    1]
 [ 22000    1]
 [ 23000    1]
 [ 20000    1]
 [ 28000    1]
 [ 30000    1]
 [ 43000    0]
 [ 18000    0]
 [ 74000    0]
 [137000    1]
 [ 16000    0]
 [ 44000    0]
 [ 90000    0]
 [ 27000    0]
 [ 28000    0]
 [ 49000    0]
 [ 72000    0]
 [ 31000    0]
 [ 17000    0]
 [ 51000    0]
 [108000    0]
 [ 15000    0]
 [ 84000    0]
 [ 20000    0]
 [ 79000    0]
 [ 54000    0]
 [135000    1]
 [ 89000    0]
 [ 32000    0]
 [ 44000    0]
 [ 83000    0]
 [ 23000    0]
 [ 58000    0]
 [ 55000    0]
 [ 48000    0]
 [ 79000    0]
 [ 18000    0]
 [117000    0]
 [ 20000    0]
```

```
[ 87000    0]
[ 66000    0]
[120000    1]
[ 83000    0]
[ 58000    0]
[ 19000    0]
[ 82000    0]
[ 63000    0]
[ 68000    0]
[ 80000    0]
[ 27000    0]
[ 23000    0]
[113000    0]
[ 18000    0]
[112000    1]
[ 52000    0]
[ 27000    0]
[ 87000    0]
[ 17000    0]
[ 80000    0]
[ 42000    0]
[ 49000    0]
[ 88000    0]
[ 62000    0]
[118000    1]
[ 55000    0]
[ 85000    0]
[ 81000    0]
[ 50000    0]
[ 81000    0]
[116000    0]
[ 15000    0]
[ 28000    0]
[ 83000    0]
[ 44000    0]
[ 25000    0]
[123000    1]
[ 73000    0]
[ 37000    0]
[ 88000    0]
[ 59000    0]
[ 86000    0]
[149000    1]
[ 21000    0]
[ 72000    0]
[ 35000    0]
[ 89000    0]
[ 86000    0]
[ 80000    0]
[ 71000    0]
[ 71000    0]
[ 61000    0]
[ 55000    0]
[ 80000    0]
[ 57000    0]
[ 75000    0]
[ 52000    0]
[ 59000    0]
[ 59000    0]
[ 75000    0]
[ 72000    0]
```

[75000	0]
[53000	0]
[51000	0]
[61000	0]
[65000	0]
[32000	0]
[17000	0]
[84000	0]
[58000	0]
[31000	0]
[87000	0]
[68000	0]
[55000	0]
[63000	0]
[82000	0]
[107000	1]
[59000	0]
[25000	0]
[85000	0]
[68000	0]
[59000	0]
[89000	0]
[25000	0]
[89000	0]
[96000	1]
[30000	0]
[61000	0]
[74000	0]
[15000	0]
[45000	0]
[76000	0]
[50000	0]
[47000	0]
[15000	0]
[59000	0]
[75000	0]
[30000	0]
[135000	1]
[100000	1]
[90000	0]
[33000	0]
[38000	0]
[69000	0]
[86000	0]
[55000	0]
[71000	0]
[148000	1]
[47000	0]
[88000	0]
[115000	0]
[118000	0]
[43000	0]
[72000	0]
[28000	0]
[47000	0]
[22000	0]
[23000	0]
[34000	0]
[16000	0]
[71000	0]
[117000	1]

[43000	0]
[60000	0]
[66000	0]
[82000	0]
[41000	0]
[72000	0]
[32000	0]
[84000	0]
[26000	0]
[43000	0]
[70000	0]
[89000	0]
[43000	0]
[79000	0]
[36000	0]
[80000	0]
[22000	0]
[39000	0]
[74000	0]
[134000	1]
[71000	0]
[101000	1]
[47000	0]
[130000	1]
[114000	0]
[142000	1]
[22000	0]
[96000	1]
[150000	1]
[42000	0]
[58000	0]
[43000	0]
[108000	1]
[65000	0]
[78000	0]
[96000	0]
[143000	1]
[80000	0]
[91000	1]
[144000	1]
[102000	1]
[60000	0]
[53000	0]
[126000	1]
[133000	1]
[72000	0]
[80000	1]
[147000	1]
[42000	0]
[107000	1]
[86000	1]
[112000	0]
[79000	1]
[57000	0]
[80000	0]
[82000	0]
[143000	1]
[149000	1]
[59000	0]
[88000	1]
[104000	1]


```
[ 72000    0]
[146000    1]
[ 50000    0]
[122000    1]
[ 52000    0]
[ 97000    1]
[ 39000    0]
[ 52000    0]
[134000    1]
[146000    1]
[ 44000    0]
[ 90000    1]
[ 72000    0]
[ 57000    0]
[ 95000    1]
[131000    1]
[ 77000    0]
[144000    1]
[125000    1]
[ 72000    0]
[ 90000    1]
[108000    1]
[ 75000    0]
[ 74000    0]
[144000    1]
[ 61000    0]
[133000    0]
[ 76000    1]
[ 42000    1]
[106000    1]
[ 26000    1]
[ 74000    1]
[ 71000    0]
[ 88000    1]
[ 38000    1]
[ 36000    1]
[ 88000    1]
[ 61000    0]
[ 70000    1]
[ 21000    1]
[141000    0]
[ 93000    1]
[ 62000    0]
[138000    1]
[ 79000    0]
[ 78000    1]
[134000    1]
[ 89000    1]
[ 39000    1]
[ 77000    0]
[ 57000    0]
[ 63000    0]
[ 73000    1]
[112000    1]
[ 79000    0]
[117000    1]
[ 38000    1]
[ 74000    1]
[137000    1]
[ 79000    1]
[ 60000    0]
```

[54000	0]
[134000	0]
[113000	1]
[125000	1]
[50000	0]
[70000	0]
[96000	1]
[50000	0]
[141000	1]
[79000	0]
[75000	1]
[104000	1]
[55000	0]
[32000	1]
[60000	0]
[138000	1]
[82000	1]
[52000	0]
[30000	1]
[131000	1]
[60000	0]
[72000	0]
[75000	0]
[118000	1]
[107000	1]
[51000	0]
[119000	1]
[65000	0]
[65000	0]
[60000	1]
[54000	0]
[144000	1]
[79000	0]
[55000	0]
[122000	1]
[104000	1]
[75000	0]
[65000	0]
[51000	1]
[105000	1]
[63000	0]
[72000	1]
[108000	1]
[77000	0]
[61000	0]
[113000	1]
[75000	0]
[90000	1]
[57000	0]
[99000	1]
[34000	1]
[70000	1]
[72000	0]
[71000	1]
[54000	0]
[129000	1]
[34000	1]
[50000	1]
[79000	0]
[104000	1]
[29000	1]

```
[ 47000    1]
[ 88000    1]
[ 71000    0]
[ 26000    1]
[ 46000    1]
[ 83000    1]
[ 73000    0]
[130000    1]
[ 80000    0]
[ 32000    1]
[ 74000    0]
[ 53000    0]
[ 87000    1]
[ 23000    1]
[ 64000    0]
[ 33000    1]
[139000    1]
[ 28000    1]
[ 33000    1]
[ 60000    1]
[ 39000    1]
[ 71000    0]
[ 34000    1]
[ 35000    1]
[ 33000    1]
[ 23000    1]
[ 45000    1]
[ 42000    1]
[ 59000    0]
[ 41000    1]
[ 23000    1]
[ 20000    1]
[ 33000    0]
[ 36000    1]]
```

In []: