K-means clustering

In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
```

In [3]:

```
df = pd.read_csv("dataset.csv")
df.columns = ["CustomerID", "Gender", "Age", "Annual_Income", "Spending_Score"]
df.head(5)
```

Out[3]:

	CustomerID	Gender	Age	Annual_Income	Spending_Score
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 [4]:

```
df.shape
```

Out[4]:

```
(200, 5)
```

In [5]:

```
df.isnull().sum()
```

Out[5]:

```
CustomerID         0
Gender             0
Age                0
Annual_Income      0
Spending_Score     0
dtype: int64
```

In [6]:

```
X = df.iloc[:, [3,4]].values
X
```

Out[6]:

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
```

```
[ 20,  15],
[ 20,  77],
[ 20,  13],
[ 20,  79],
[ 21,  35],
[ 21,  66],
[ 23,  29],
[ 23,  98],
[ 24,  35],
[ 24,  73],
[ 25,   5],
[ 25,  73],
[ 28,  14],
[ 28,  82],
[ 28,  32],
[ 28,  61],
[ 29,  31],
[ 29,  87],
[ 30,   4],
[ 30,  73],
[ 33,   4],
[ 33,  92],
[ 33,  14],
[ 33,  81],
[ 34,  17],
[ 34,  73],
[ 37,  26],
[ 37,  75],
[ 38,  35],
[ 38,  92],
[ 39,  36],
[ 39,  61],
[ 39,  28],
[ 39,  65],
[ 40,  55],
[ 40,  47],
[ 40,  42],
[ 40,  42],
[ 42,  52],
[ 42,  60],
[ 43,  54],
[ 43,  60],
[ 43,  45],
[ 43,  41],
[ 44,  50],
[ 44,  46],
[ 46,  51],
[ 46,  46],
[ 46,  56],
[ 46,  55],
[ 47,  52],
[ 47,  59],
[ 48,  51],
[ 48,  59],
[ 48,  50],
[ 48,  48],
[ 48,  59],
[ 48,  47],
[ 49,  55],
[ 49,  42],
[ 50,  49],
[ 50,  56],
[ 54,  47],
[ 54,  54],
[ 54,  53],
[ 54,  48],
[ 54,  52],
[ 54,  42],
[ 54,  51],
[ 54,  55],
[ 54,  41],
[ 54,  44],
[ 54,  57],
```

```
[ 54,  57],
[ 54,  46],
[ 57,  58],
[ 57,  55],
[ 58,  60],
[ 58,  46],
[ 59,  55],
[ 59,  41],
[ 60,  49],
[ 60,  40],
[ 60,  42],
[ 60,  52],
[ 60,  47],
[ 60,  50],
[ 61,  42],
[ 61,  49],
[ 62,  41],
[ 62,  48],
[ 62,  59],
[ 62,  55],
[ 62,  56],
[ 62,  42],
[ 63,  50],
[ 63,  46],
[ 63,  43],
[ 63,  48],
[ 63,  52],
[ 63,  54],
[ 64,  42],
[ 64,  46],
[ 65,  48],
[ 65,  50],
[ 65,  43],
[ 65,  59],
[ 67,  43],
[ 67,  57],
[ 67,  56],
[ 67,  40],
[ 69,  58],
[ 69,  91],
[ 70,  29],
[ 70,  77],
[ 71,  35],
[ 71,  95],
[ 71,  11],
[ 71,  75],
[ 71,   9],
[ 71,  75],
[ 72,  34],
[ 72,  71],
[ 73,   5],
[ 73,  88],
[ 73,   7],
[ 73,  73],
[ 74,  10],
[ 74,  72],
[ 75,   5],
[ 75,  93],
[ 76,  40],
[ 76,  87],
[ 77,  12],
[ 77,  97],
[ 77,  36],
[ 77,  74],
[ 78,  22],
[ 78,  90],
[ 78,  17],
[ 78,  88],
[ 78,  20],
[ 78,  76],
[ 78,  16],
[ 78,  89],
```

```
       [ 78,   1],
       [ 78,  78],
       [ 78,   1],
       [ 78,  73],
       [ 79,  35],
       [ 79,  83],
       [ 81,   5],
       [ 81,  93],
       [ 85,  26],
       [ 85,  75],
       [ 86,  20],
       [ 86,  95],
       [ 87,  27],
       [ 87,  63],
       [ 87,  13],
       [ 87,  75],
       [ 87,  10],
       [ 87,  92],
       [ 88,  13],
       [ 88,  86],
       [ 88,  15],
       [ 88,  69],
       [ 93,  14],
       [ 93,  90],
       [ 97,  32],
       [ 97,  86],
       [ 98,  15],
       [ 98,  88],
       [ 99,  39],
       [ 99,  97],
       [101,  24],
       [101,  68],
       [103,  17],
       [103,  85],
       [103,  23],
       [103,  69],
       [113,   8],
       [113,  91],
       [120,  16],
       [120,  79],
       [126,  28],
       [126,  74],
       [137,  18],
       [137,  83]], dtype=int64)
```

In [7]:

```
scaler = StandardScaler()
X = scaler.fit_transform(X)
X
```

Out[7]:

```
array([[-1.73899919, -0.43480148],
       [-1.73899919,  1.19570407],
       [-1.70082976, -1.71591298],
       [-1.70082976,  1.04041783],
       [-1.66266033, -0.39597992],
       [-1.66266033,  1.00159627],
       [-1.62449091, -1.71591298],
       [-1.62449091,  1.70038436],
       [-1.58632148, -1.83237767],
       [-1.58632148,  0.84631002],
       [-1.58632148, -1.4053405 ],
       [-1.58632148,  1.89449216],
       [-1.54815205, -1.36651894],
       [-1.54815205,  1.04041783],
       [-1.54815205, -1.44416206],
       [-1.54815205,  1.11806095],
       [-1.50998262, -0.59008772],
       [-1.50998262,  0.61338066],
       [-1.43364376, -0.82301709],
       [-1.43364376,  1.8556706 ],
```

```
[-1.39547433, -0.59008772],
[-1.39547433,  0.88513158],
[-1.3573049 , -1.75473454],
[-1.3573049 ,  0.88513158],
[-1.24279661, -1.4053405 ],
[-1.24279661,  1.23452563],
[-1.24279661, -0.7065524 ],
[-1.24279661,  0.41927286],
[-1.20462718, -0.74537397],
[-1.20462718,  1.42863343],
[-1.16645776, -1.7935561 ],
[-1.16645776,  0.88513158],
[-1.05194947, -1.7935561 ],
[-1.05194947,  1.62274124],
[-1.05194947, -1.4053405 ],
[-1.05194947,  1.19570407],
[-1.01378004, -1.28887582],
[-1.01378004,  0.88513158],
[-0.89927175, -0.93948177],
[-0.89927175,  0.96277471],
[-0.86110232, -0.59008772],
[-0.86110232,  1.62274124],
[-0.82293289, -0.55126616],
[-0.82293289,  0.41927286],
[-0.82293289, -0.86183865],
[-0.82293289,  0.5745591 ],
[-0.78476346,  0.18634349],
[-0.78476346, -0.12422899],
[-0.78476346, -0.3183368 ],
[-0.78476346, -0.3183368 ],
[-0.70842461,  0.06987881],
[-0.70842461,  0.38045129],
[-0.67025518,  0.14752193],
[-0.67025518,  0.38045129],
[-0.67025518, -0.20187212],
[-0.67025518, -0.35715836],
[-0.63208575, -0.00776431],
[-0.63208575, -0.16305055],
[-0.55574689,  0.03105725],
[-0.55574689, -0.16305055],
[-0.55574689,  0.22516505],
[-0.55574689,  0.18634349],
[-0.51757746,  0.06987881],
[-0.51757746,  0.34162973],
[-0.47940803,  0.03105725],
[-0.47940803,  0.34162973],
[-0.47940803, -0.00776431],
[-0.47940803, -0.08540743],
[-0.47940803,  0.34162973],
[-0.47940803, -0.12422899],
[-0.4412386 ,  0.18634349],
[-0.4412386 , -0.3183368 ],
[-0.40306917, -0.04658587],
[-0.40306917,  0.22516505],
[-0.25039146, -0.12422899],
[-0.25039146,  0.14752193],
[-0.25039146,  0.10870037],
[-0.25039146, -0.08540743],
[-0.25039146,  0.06987881],
[-0.25039146, -0.3183368 ],
[-0.25039146,  0.03105725],
[-0.25039146,  0.18634349],
[-0.25039146, -0.35715836],
[-0.25039146, -0.24069368],
[-0.25039146,  0.26398661],
[-0.25039146, -0.16305055],
[-0.13588317,  0.30280817],
[-0.13588317,  0.18634349],
[-0.09771374,  0.38045129],
[-0.09771374, -0.16305055],
[-0.05954431,  0.18634349],
[-0.05954431, -0.35715836],
```

```
[-0.02137488, -0.04658587],
[-0.02137488, -0.39597992],
[-0.02137488, -0.3183368 ],
[-0.02137488,  0.06987881],
[-0.02137488, -0.12422899],
[-0.02137488, -0.00776431],
[ 0.01679455, -0.3183368 ],
[ 0.01679455, -0.04658587],
[ 0.05496398, -0.35715836],
[ 0.05496398, -0.08540743],
[ 0.05496398,  0.34162973],
[ 0.05496398,  0.18634349],
[ 0.05496398,  0.22516505],
[ 0.05496398, -0.3183368 ],
[ 0.09313341, -0.00776431],
[ 0.09313341, -0.16305055],
[ 0.09313341, -0.27951524],
[ 0.09313341, -0.08540743],
[ 0.09313341,  0.06987881],
[ 0.09313341,  0.14752193],
[ 0.13130284, -0.3183368 ],
[ 0.13130284, -0.16305055],
[ 0.16947227, -0.08540743],
[ 0.16947227, -0.00776431],
[ 0.16947227, -0.27951524],
[ 0.16947227,  0.34162973],
[ 0.24581112, -0.27951524],
[ 0.24581112,  0.26398661],
[ 0.24581112,  0.22516505],
[ 0.24581112, -0.39597992],
[ 0.32214998,  0.30280817],
[ 0.32214998,  1.58391968],
[ 0.36031941, -0.82301709],
[ 0.36031941,  1.04041783],
[ 0.39848884, -0.59008772],
[ 0.39848884,  1.73920592],
[ 0.39848884, -1.52180518],
[ 0.39848884,  0.96277471],
[ 0.39848884, -1.5994483 ],
[ 0.39848884,  0.96277471],
[ 0.43665827, -0.62890928],
[ 0.43665827,  0.80748846],
[ 0.4748277 , -1.75473454],
[ 0.4748277 ,  1.46745499],
[ 0.4748277 , -1.67709142],
[ 0.4748277 ,  0.88513158],
[ 0.51299713, -1.56062674],
[ 0.51299713,  0.84631002],
[ 0.55116656, -1.75473454],
[ 0.55116656,  1.6615628 ],
[ 0.58933599, -0.39597992],
[ 0.58933599,  1.42863343],
[ 0.62750542, -1.48298362],
[ 0.62750542,  1.81684904],
[ 0.62750542, -0.55126616],
[ 0.62750542,  0.92395314],
[ 0.66567484, -1.09476801],
[ 0.66567484,  1.54509812],
[ 0.66567484, -1.28887582],
[ 0.66567484,  1.46745499],
[ 0.66567484, -1.17241113],
[ 0.66567484,  1.00159627],
[ 0.66567484, -1.32769738],
[ 0.66567484,  1.50627656],
[ 0.66567484, -1.91002079],
[ 0.66567484,  1.07923939],
[ 0.66567484, -1.91002079],
[ 0.66567484,  0.88513158],
[ 0.70384427, -0.59008772],
[ 0.70384427,  1.27334719],
[ 0.78018313, -1.75473454],
[ 0.78018313,  1.6615628 ],
```

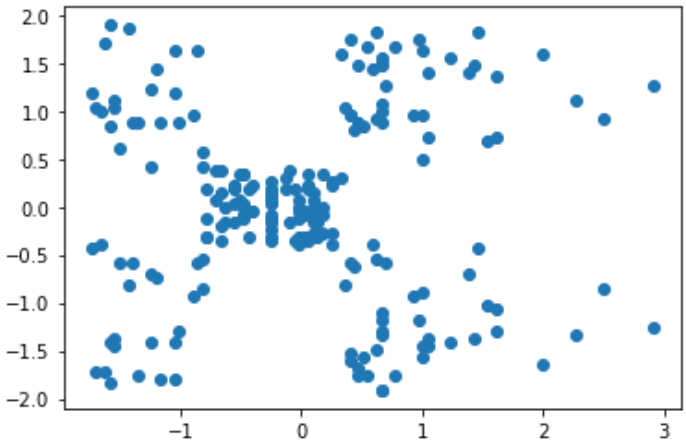
```
[ 0.93286085, -0.93948177],
[ 0.93286085,  0.96277471],
[ 0.97103028, -1.17241113],
[ 0.97103028,  1.73920592],
[ 1.00919971, -0.90066021],
[ 1.00919971,  0.49691598],
[ 1.00919971, -1.44416206],
[ 1.00919971,  0.96277471],
[ 1.00919971, -1.56062674],
[ 1.00919971,  1.62274124],
[ 1.04736914, -1.44416206],
[ 1.04736914,  1.38981187],
[ 1.04736914, -1.36651894],
[ 1.04736914,  0.72984534],
[ 1.23821628, -1.4053405 ],
[ 1.23821628,  1.54509812],
[ 1.390894  , -0.7065524 ],
[ 1.390894  ,  1.38981187],
[ 1.42906343, -1.36651894],
[ 1.42906343,  1.46745499],
[ 1.46723286, -0.43480148],
[ 1.46723286,  1.81684904],
[ 1.54357172, -1.01712489],
[ 1.54357172,  0.69102378],
[ 1.61991057, -1.28887582],
[ 1.61991057,  1.35099031],
[ 1.61991057, -1.05594645],
[ 1.61991057,  0.72984534],
[ 2.00160487, -1.63826986],
[ 2.00160487,  1.58391968],
[ 2.26879087, -1.32769738],
[ 2.26879087,  1.11806095],
[ 2.49780745, -0.86183865],
[ 2.49780745,  0.92395314],
[ 2.91767117, -1.25005425],
[ 2.91767117,  1.27334719]])
```

In [8]:

```
df.iloc[:, [3,4]] = X
```

In [9]:

```
plt.scatter(X[:, 0], X[:, 1])
plt.show()
```

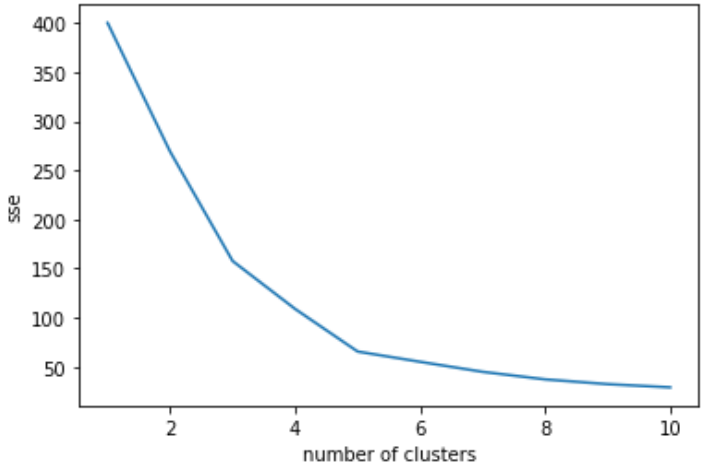


In [11]:

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

plt.plot(range(1,11), sse)
```

```
plt.xlabel("number of clusters")
plt.ylabel("sse")
plt.show()
```



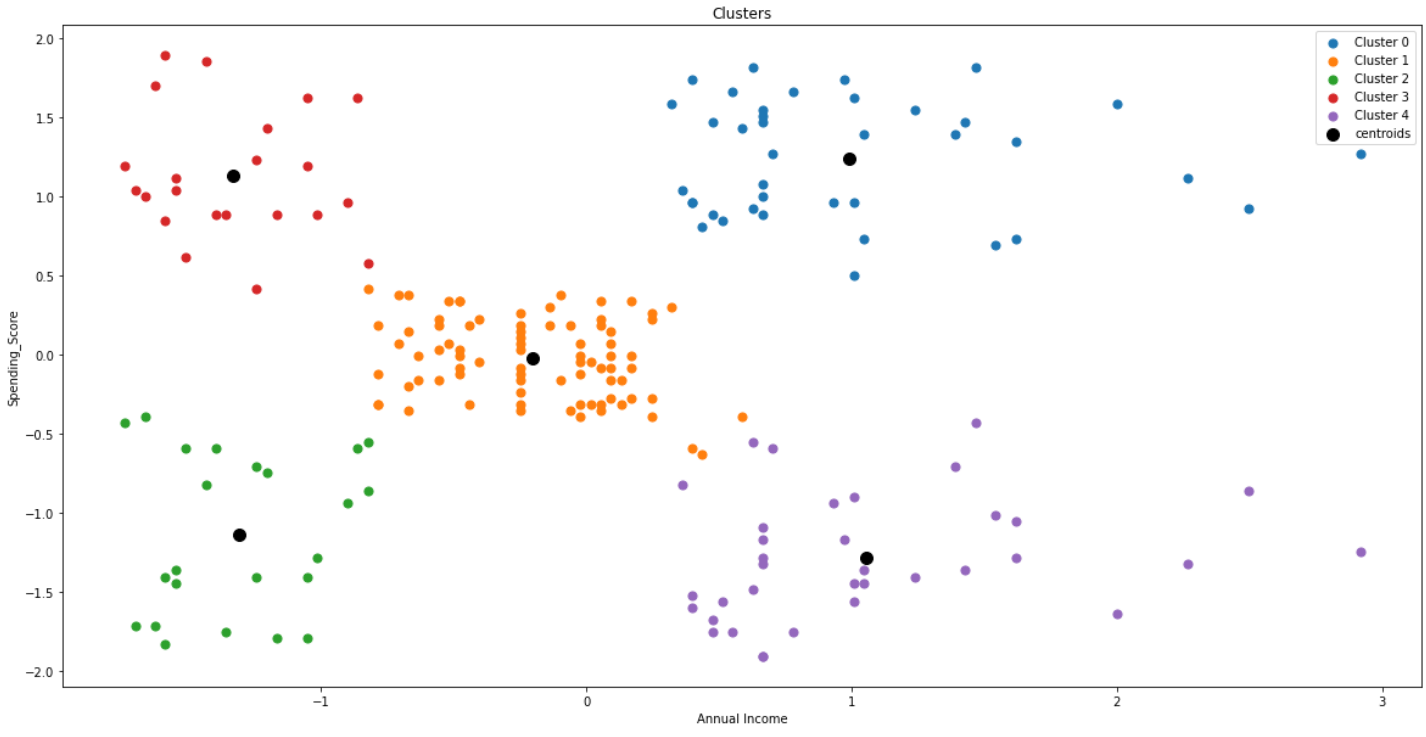
In [12]:

```
n_clusters = 5
model = KMeans(n_clusters=n_clusters, init="k-means++")
pred = model.fit_predict(X)
```

In [13]:

```
plt.figure(figsize=(20,10))
for i in range(0, n_clusters):
    plt.scatter(X[pred == i, 0], X[pred == i, 1], s=50, label="Cluster %d" % i)

plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:,1], s = 100, c = 'blac
k', label='centroids')
plt.title("Clusters")
plt.xlabel("Annual Income")
plt.ylabel("Spending_Score")
plt.legend()
plt.show()
```



In [ ]: