

```
In [2]: import numpy as np  
import pandas as pd
```

```
In [3]: dataset=pd.read_csv("employee_dataset.csv")
```

In [4]: dataset

Out[4]:

	Employee ID	Gender	Age	Annual Income (\$)	Expenses (\$)
0	E001	Female	49	88650	10
1	E002	Male	36	87324	66
2	E003	Female	52	84159	94
3	E004	Male	26	46557	22
4	E005	Female	42	69495	96
5	E006	Male	41	35110	37
6	E007	Non-binary	51	101626	63
7	E008	Non-binary	21	83278	47
8	E009	Non-binary	41	97782	42
9	E010	Non-binary	45	116230	14
10	E011	Non-binary	23	59884	45
11	E012	Male	37	116857	16
12	E013	Female	42	104406	91
13	E014	Male	25	49506	93
14	E015	Male	48	76283	32
15	E016	Non-binary	42	40903	100
16	E017	Non-binary	43	65536	11
17	E018	Male	23	94683	97
18	E019	Non-binary	26	107737	83
19	E020	Female	51	45029	17
20	E021	Male	25	36281	21
21	E022	Female	45	85407	42
22	E023	Female	36	34853	82
23	E024	Non-binary	35	115280	41
24	E025	Female	51	107610	24
25	E026	Female	60	86687	56
26	E027	Non-binary	35	34038	92
27	E028	Non-binary	56	107736	53
28	E029	Male	60	52479	57
29	E030	Female	41	33111	16

In [8]: x=dataset.iloc[:,[3,4]].values

In [9]: x

```
Out[9]: array([[ 88650,    10],
               [ 87324,    66],
               [ 84159,    94],
               [ 46557,    22],
               [ 69495,    96],
               [ 35110,    37],
               [101626,    63],
               [ 83278,    47],
               [ 97782,    42],
               [116230,    14],
               [ 59884,    45],
               [116857,    16],
               [104406,    91],
               [ 49506,    93],
               [ 76283,    32],
               [ 40903,   100],
               [ 65536,    11],
               [ 94683,    97],
               [107737,    83],
               [ 45029,    17],
               [ 36281,    21],
               [ 85407,    42],
               [ 34853,    82],
               [115280,    41],
               [107610,    24],
               [ 86687,    56],
               [ 34038,    92],
               [107736,    53],
               [ 52479,    57],
               [ 33111,    16]], dtype=int64)
```

In [10]: `from sklearn.cluster import KMeans`

In [11]: `wcss=[]`

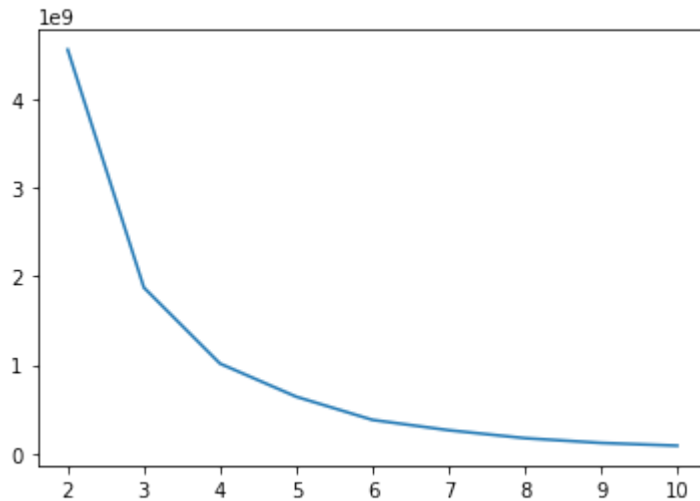
```
In [17]: for i in range(2,11):
          kmeans=KMeans(n_clusters = i,random_state = 42)
          kmeans.fit(x)
          wcss.append(kmeans.inertia_)
```

In [18]: `wcss`

```
Out[18]: [4553267720.515837,
          1871145227.2676768,
          1013939900.5749998,
          640850876.375,
          379667171.7380953,
          263333958.0714286,
          173804544.96666667,
          120145670.73333336,
          88972473.73333335]
```

```
In [19]: import matplotlib.pyplot as plt
plt.plot(range(2,11),wcss)
```

Out[19]: [



```
In [22]: kmeans=KMeans(n_clusters = 5,init = "k-means++",random_state = 42)
```

```
In [24]: y_means = kmeans.fit_predict(x)
```

```
In [25]: y_means
```

Out[25]: array([0, 0, 0, 1, 3, 4, 2, 0, 0, 2, 3, 2, 2, 1, 3, 1, 3, 0, 2, 1, 4,  
0,  
4, 2, 2, 0, 4, 2, 1, 4])

```
In [33]: plt.scatter(x[y_means==0,0], x[y_means==0,1],s=100, c='red',label='clus
plt.scatter(x[y_means==1,0], x[y_means==1,1],s=100, c='blue',label='clus
plt.scatter(x[y_means==2,0], x[y_means==2,1],s=100, c='orange',label='c
plt.scatter(x[y_means==3,0], x[y_means==3,1],s=100, c='green',label='cl
plt.scatter(x[y_means==4,0], x[y_means==4,1],s=100, c='black',label='cl
```

Out[33]: <matplotlib.collections.PathCollection at 0x20d4e338f10>

