```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
```

```
In [2]: dataset = pd.read_csv('Mall_Customers.csv')
        dataset.head()
```

Out[2]:

	CustomerID	Genre	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

```
In [3]: dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   CustomerID              200 non-null    int64
 1   Genre                   200 non-null    object
 2   Age                     200 non-null    int64
 3   Annual Income (k$)      200 non-null    int64
 4   Spending Score (1-100)  200 non-null    int64
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
```

```
In [4]: dataset.isnull().sum()
```

```
Out[4]: CustomerID              0
        Genre                   0
        Age                     0
        Annual Income (k$)      0
        Spending Score (1-100)  0
        dtype: int64
```

```
In [5]: X = dataset.iloc[:, [3, 4]].values
```

Modelling

```
In [6]: from sklearn.cluster import DBSCAN
```

```
In [7]: dbs = DBSCAN(eps=5, min_samples=5)
```

```
In [8]: y_dbs = dbs.fit_predict(X)
```

```
In [9]: y_dbs
```

```
Out[9]: array([-1,  0, -1,  0, -1,  0, -1, -1, -1,  0, -1, -1, -1,  0, -1,  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,  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,  1,  1,  1,  1,  1,  1,  1,
                1,  1,  1,  1, -1, -1,  2, -1,  4,  3,  2,  3,  2, -1,  2,  3,  4,
                3,  2,  3,  2,  3,  4, -1,  4,  3,  4, -1,  2, -1,  4, -1,  4, -1,
                2, -1,  4,  3,  2,  3,  2, -1,  4, -1,  4, -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], dtype=int64)
```

```
In [10]: np.unique(y_dbs)
```

```
Out[10]: array([-1,  0,  1,  2,  3,  4], dtype=int64)
```

```
In [11]: # Visualising the clusters
         plt.scatter(X[y_dbs == -1, 0], X[y_dbs == -1, 1], s = 100, c = 'red', label =
         'Cluster -1')
         plt.scatter(X[y_dbs == 0, 0], X[y_dbs == 0, 1], s = 100, c = 'magenta', label
         = 'Cluster 0')
         plt.scatter(X[y_dbs == 1, 0], X[y_dbs == 1, 1], s = 100, c = 'blue', label =
         'Cluster 1')
         plt.scatter(X[y_dbs == 2, 0], X[y_dbs == 2, 1], s = 100, c = 'green', label =
         'Cluster 2')
         plt.scatter(X[y_dbs == 3, 0], X[y_dbs == 3, 1], s = 100, c = 'cyan', label =
         'Cluster 3')
         plt.scatter(X[y_dbs == 4, 0], X[y_dbs == 4, 1], s = 100, c = 'yellow', label =
         'Cluster 4')

         plt.title('Clusters of customers')
         plt.xlabel('Annual Income (k$)')
         plt.ylabel('Spending Score (1-100)')
         plt.legend()
         plt.show()
```