

Importing Libraries

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
```

About Data

Customers are asked to rate the store out of 10 i.e. **"satisfaction" point** given by the customers to the store. Using an algorithm, store gives **"Loyalty points"** to customers based on the no. of items purchased by them in the last year + the amount of money spent by them in the store.

Loading Data

```
In [2]: data = pd.read_csv(r'C:\Users\vamsi\Desktop\ML\14.MeanShift Clustering\market_data.csv')
```

Data Exploration

```
In [3]: data.head()
```

Out[3]:

	Satisfaction	Loyalty
0	4	-1.33
1	6	-0.28
2	5	-0.99
3	7	-0.29
4	4	1.06

```
In [4]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Satisfaction    30 non-null    int64
1   Loyalty         30 non-null    float64
dtypes: float64(1), int64(1)
memory usage: 608.0 bytes
```

```
In [5]: data.describe()
```

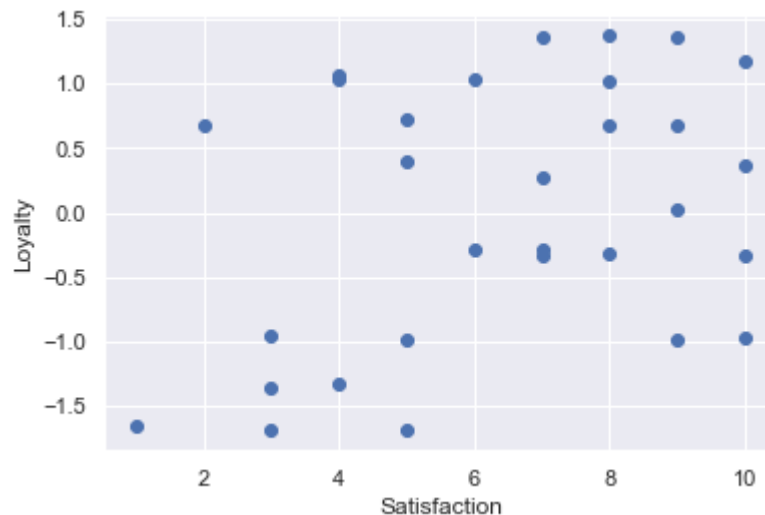
```
Out[5]:
```

	Satisfaction	Loyalty
count	30.000000	30.000000
mean	6.400000	0.001000
std	2.620871	1.016476
min	1.000000	-1.690000
25%	4.250000	-0.967500
50%	7.000000	0.150000
75%	8.750000	0.947500
max	10.000000	1.380000

Let's plot

```
In [6]: plt.scatter(data['Satisfaction'], data['Loyalty'])  
plt.xlabel('Satisfaction')  
plt.ylabel('Loyalty')
```

```
Out[6]: Text(0, 0.5, 'Loyalty')
```



Preparing input data

```
In [7]: from sklearn import preprocessing
data_scaled = data.copy()
data_scaled = preprocessing.scale(data)
data_scaled[:5]
```

```
Out[7]: array([[ -0.93138063, -1.3318111 ],
               [-0.15523011, -0.28117124],
               [-0.54330537, -0.99160391],
               [ 0.23284516, -0.29117733],
               [-0.93138063,  1.05964534]])
```

Mean-Shift Clustering

Selecting Band-width

```
In [8]: from sklearn.cluster import estimate_bandwidth
bandwidth = estimate_bandwidth(data_scaled, quantile=0.3)
```

```
In [9]: bandwidth
```

```
Out[9]: 1.2909951858425464
```

Training Model

```
In [10]: from sklearn.cluster import MeanShift
clustering = MeanShift(bandwidth)
clustering.fit(data_scaled)
```

C:\Users\vamsi\anaconda3\lib\site-packages\sklearn\utils\validation.py:67: FutureWarning: Pass bandwidth=1.2909951858425464 as keyword args. From version 0.25 passing these as positional arguments will result in an error
warnings.warn("Pass {} as keyword args. From version 0.25 "

```
Out[10]: MeanShift(bandwidth=1.2909951858425464)
```

Clustering Labels

```
In [11]: clustering.labels_
```

```
Out[11]: array([1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1,
                0, 1, 0, 0, 0, 0, 1, 0], dtype=int64)
```

```
In [12]: cluster_data= data.copy()
cluster_data['cluster_pred'] = clustering.fit_predict(data_scaled)
```

In [13]: cluster_data

Out[13]:

	Satisfaction	Loyalty	cluster_pred
0	4	-1.33	1
1	6	-0.28	1
2	5	-0.99	1
3	7	-0.29	0
4	4	1.06	0
5	1	-1.66	1
6	10	-0.97	0
7	8	-0.32	0
8	8	1.02	0
9	8	0.68	0
10	10	-0.34	0
11	5	0.39	0
12	5	-1.69	1
13	2	0.67	1
14	7	0.27	0
15	9	1.36	0
16	8	1.38	0
17	7	1.36	0
18	7	-0.34	0
19	9	0.67	0
20	10	1.18	0
21	3	-1.69	1
22	4	1.04	0
23	3	-0.96	1
24	6	1.03	0
25	9	-0.99	0
26	10	0.37	0
27	9	0.03	0
28	3	-1.36	1
29	5	0.73	0

Visualisation

```
In [14]: plt.scatter(data['Satisfaction'], data['Loyalty'], c = cluster_data['cluster_pred'])  
plt.xlabel('Satisfaction')  
plt.ylabel('Loyalty')
```

Out[14]: Text(0, 0.5, 'Loyalty')

