

Customer Spending Score Prediction

The use of machine learning can be seen almost everywhere around us, be it Facebook recognizing you or your friends, or YouTube recommending you a video or two based on your history —

Machine Learning is everywhere!

However, the ‘magic’ of machine learning is not just limited to only these areas.

The Problem

Malls or shopping complexes are often indulged in the race to increase their customers and hence making huge profits. To achieve this task machine learning is being applied by many stores already. It is amazing to realize the fact that how machine learning can aid in such ambitions. The shopping complexes make use of their customers’ data and develop ML models to target the right ones. This not only increases sales but also makes the complexes efficient.

1	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
2	1	Male	19	15	39
3	2	Male	21	15	81
4	3	Female	20	16	6
5	4	Female	23	16	77
6	5	Female	31	17	40
7	6	Female	22	17	76
8	7	Female	35	18	6
9	8	Female	23	18	94
10	9	Male	64	19	3

Fig. 1 Sample Dataset

Here we have the following features :

1. CustomerID: It is the unique ID given to a customer

2. Gender: Gender of the customer
3. Age: The age of the customer
4. Annual Income(k\$): It is the annual income of the customer
5. Spending Score: It is the score(out of 100) given to a customer by the mall authorities, based on the money spent and the behavior of the customer.

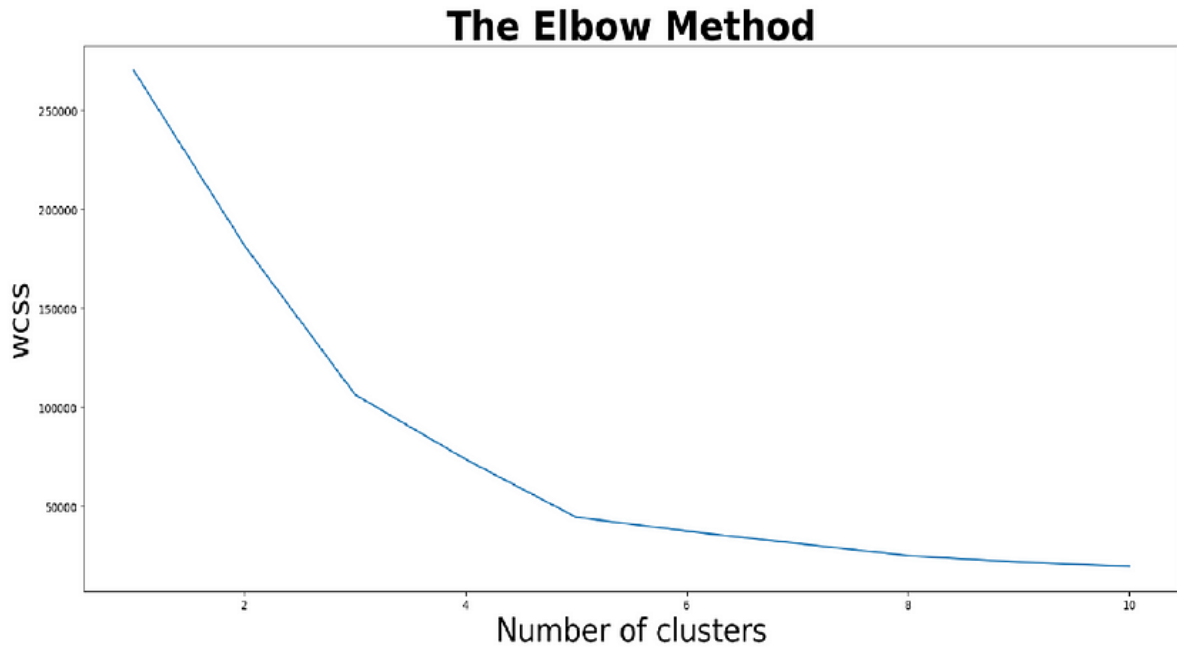
Steps:

1. Data Preprocessing
2. Exploratory Data Analysis
3. Feature Transformation
4. Model Selection
5. Evaluation

I will use the K-Means Clustering algorithm to cluster the data.
To implement K-Means clustering, we need to look at the **Elbow Method**.

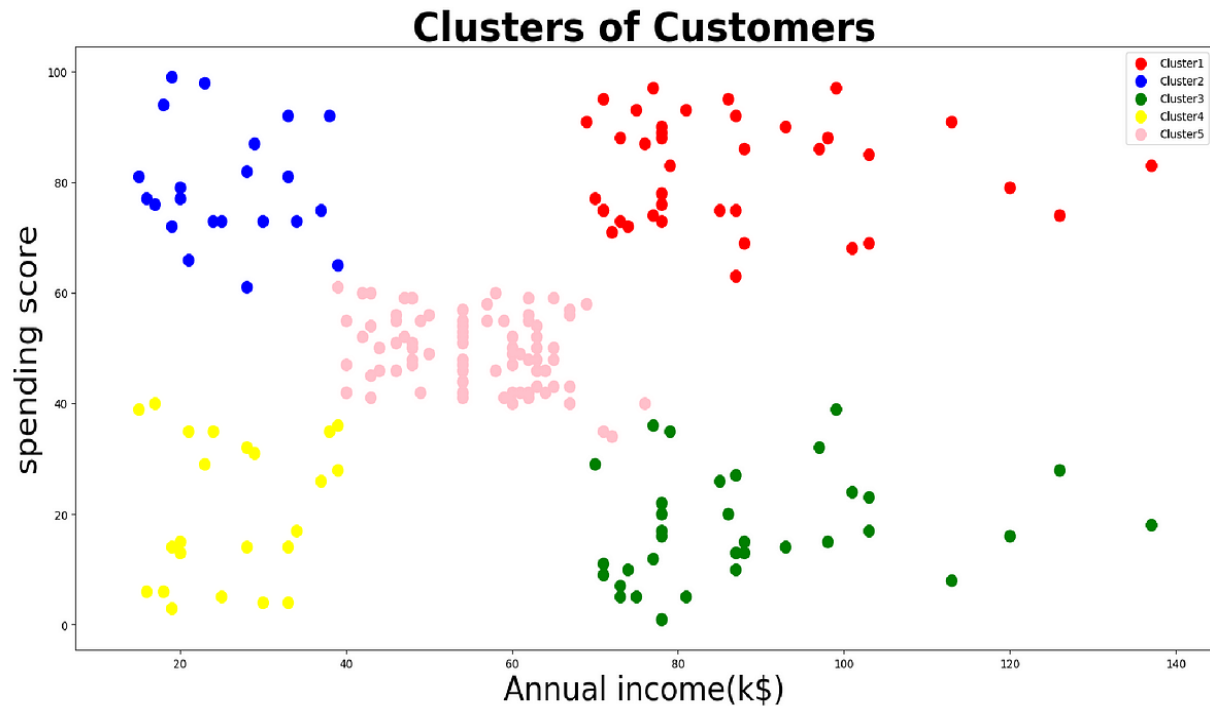
The Elbow method is a method of interpretation and validation of consistency within-cluster analysis designed to help to find the appropriate number of clusters in a dataset.

The following figure demonstrates the elbow method:



It is clear from the figure that we should take the number of clusters equal to 5, as the slope of the curve is not steep enough after it.

Finally, let us plot the clusters:



The data(clusters) are plotted on a spending score Vs annual income curve.

Let us now analyze the results of the model.

Analyzing the Results

We can see that the mall customers can be broadly grouped into 5 groups based on their purchases made in the mall.

Cluster 4(yellow colored): We can see people have low annual income and low spending scores, this is quite reasonable as people having low salaries prefer to buy less, in fact, these are the wise people who know how to spend and save money. The shops/mall will be least interested in people belonging to this cluster.

Cluster 2(blue colored): we can see that people have low income but higher spending scores, these are those people who for some reason love to buy products more often even though they have a low income. Maybe it's because these people are more than satisfied with the mall services. The shops/malls might not target these people that effectively but still will not lose them.

Cluster 5(pink colored): we see that people have average income and an average spending score, these people again will not be the prime targets of the shops or mall, but again they will be considered and other data analysis techniques may be used to increase their spending score.

Cluster 1(red-colored): we see that people have high income and high spending scores, this is the ideal case for the mall or shops as these people are the prime sources of profit. These people might be the regular customers of the mall and are convinced by the mall's facilities.

Cluster 3(green colored): we see that people have high income but low spending scores, this is interesting. Maybe these are the people who are unsatisfied or unhappy by the mall's services. These can be the prime targets of the mall, as they have the potential to spend money. So, the mall authorities will try to add new facilities so that they can attract these people and can meet their needs.

Finally, based on our machine learning technique we may deduce that to increase the profits of the mall, the mall authorities should target people belonging to cluster 3 and cluster 5 and should also maintain its standards to keep the people belonging to cluster 1 and cluster 2 happy and satisfied.

To conclude, I would like to say that it is amazing to see how machine learning can be used in businesses to enhance profit.