Customer Segmentation

Data Science Course Project

What is Customer Segmentation?

Customers Data

The first step of segmentation is gathering customer data

Applying ML Models

We use unsupervised learning techniques to segment them into clusters

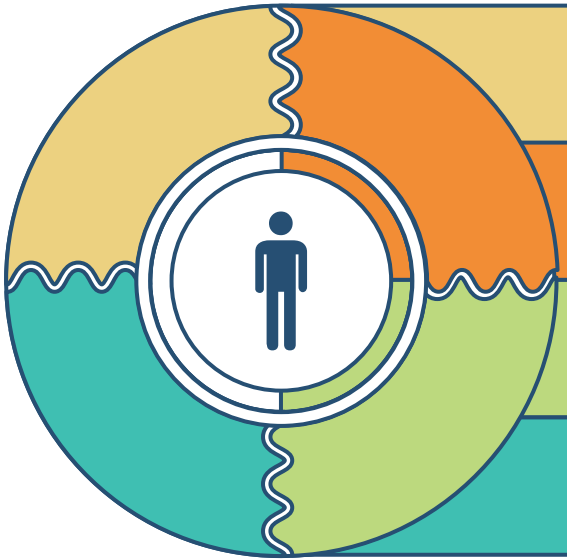
Drawing Conclusions

After we plot the segments of customers' data we can draw several conclusions

Acting on the output

After drawing the conclusions and understanding the clusters we can optimize our customer approach

Features in the Dataset Used



Gender

Customers are classified on the basis of their genders i.e Male(44%) and Female(56%)

Age

Age plays and important role is customer segmentation hence the ages are mentioned in front of each customer

Annual Income

Annual Income is ought to be mentioned in customer segmentation projects. The Annual income of customers in mentioned in thousands

Spending Score

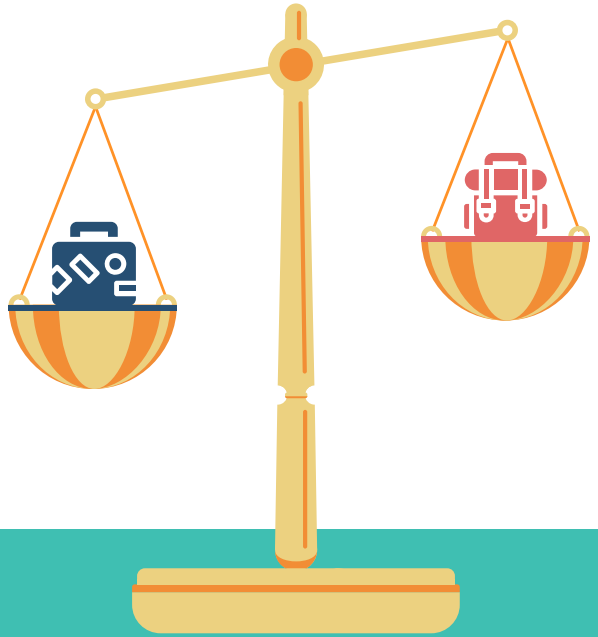
This is a feature in which a special spending score is assigned to each customer based on his/her buying behaviour and net spend

Gender

44%

Male

The given dataset contains 44% Male population that is 88 out of 200 rows of data.



56%

Female

The Majority of the dataset contains female customers. This was analyzed with the help of a pie chart

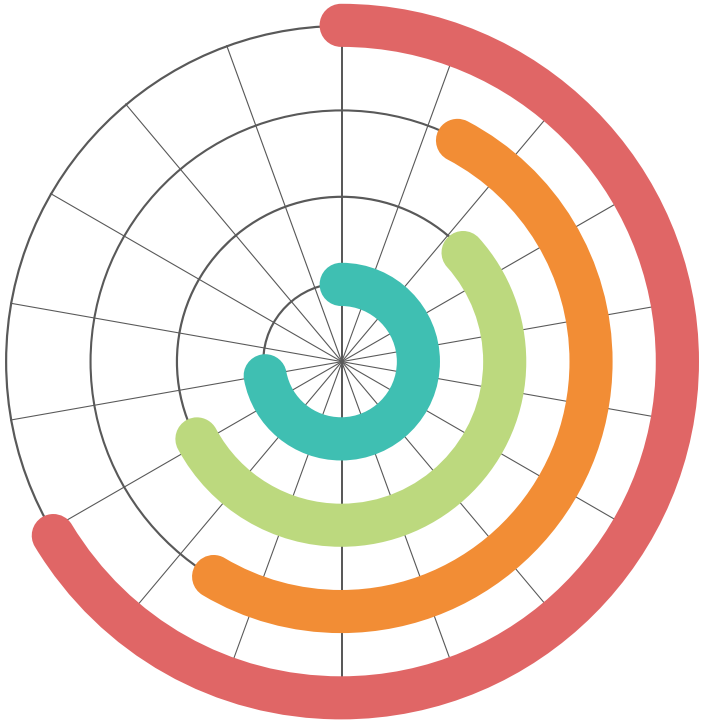
Age

21 Customers who have an age between 20-25

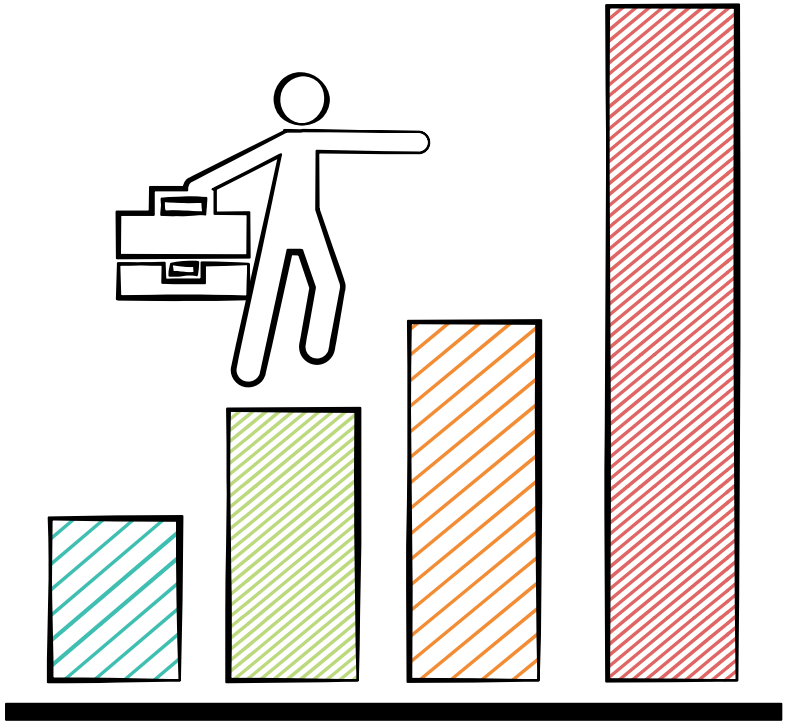
24 Customers who have an age between 25-30 and 35-40

26 Customers who have an age between 45-50

36 Customers who have an age between 30-35



Annual Income



70-80k

36 customers have an annual income between this range

60-70k

28 customers have an annual income between this range

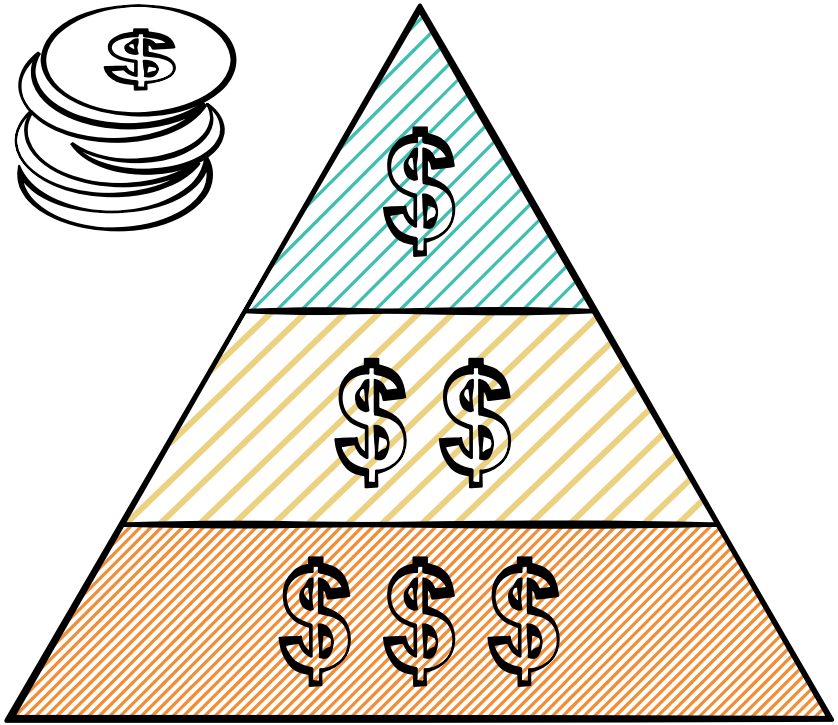
50-60k

24 customers have an annual income between this range

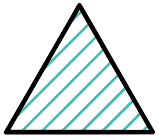
30-40k

18 customers have an annual income between this range

Spending Score

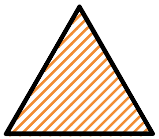


Around 53 customers out of 200 have a spending score less than 40



83 Customers have a spending score between 40 and 70

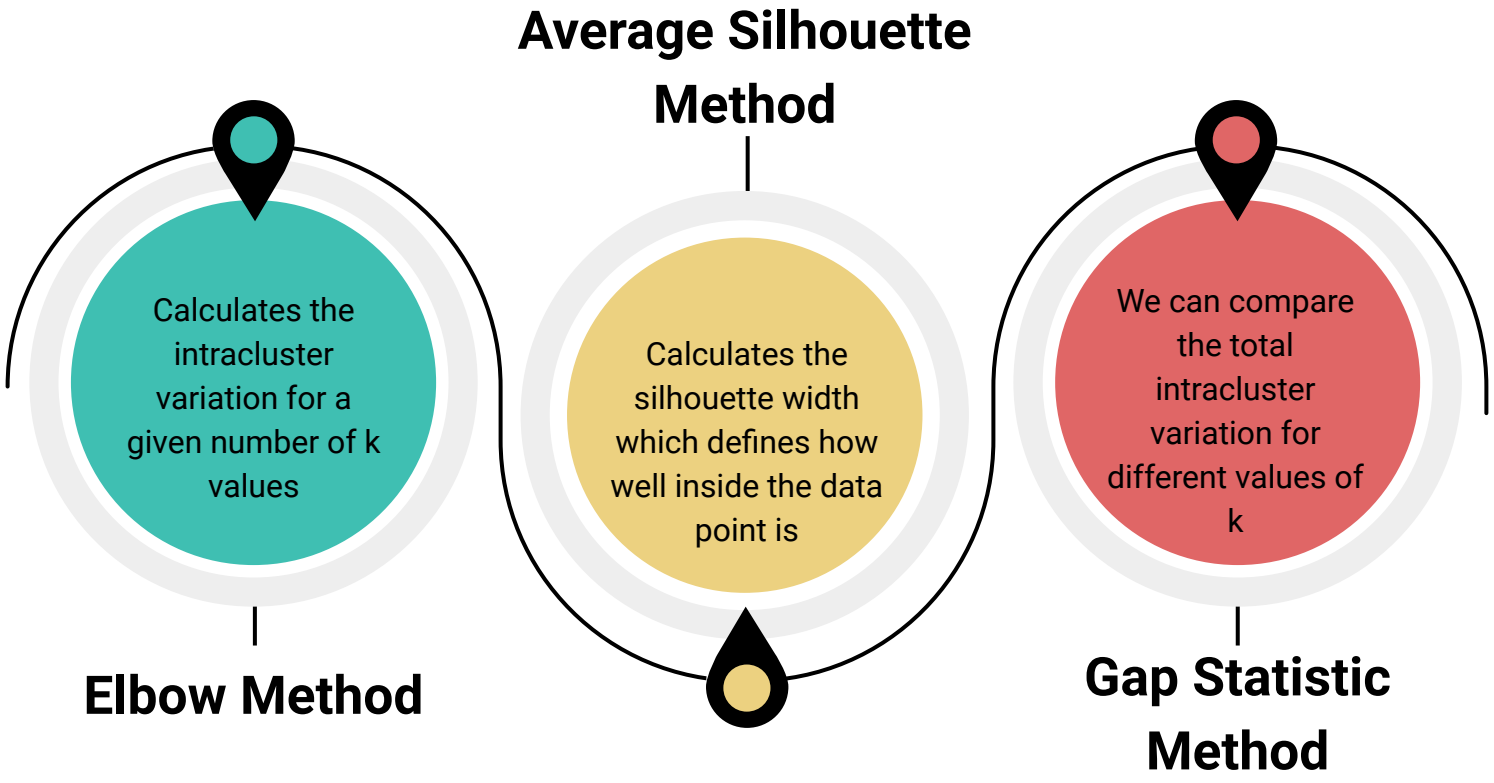
While 54 customers have spending score ranging from 70 to 100

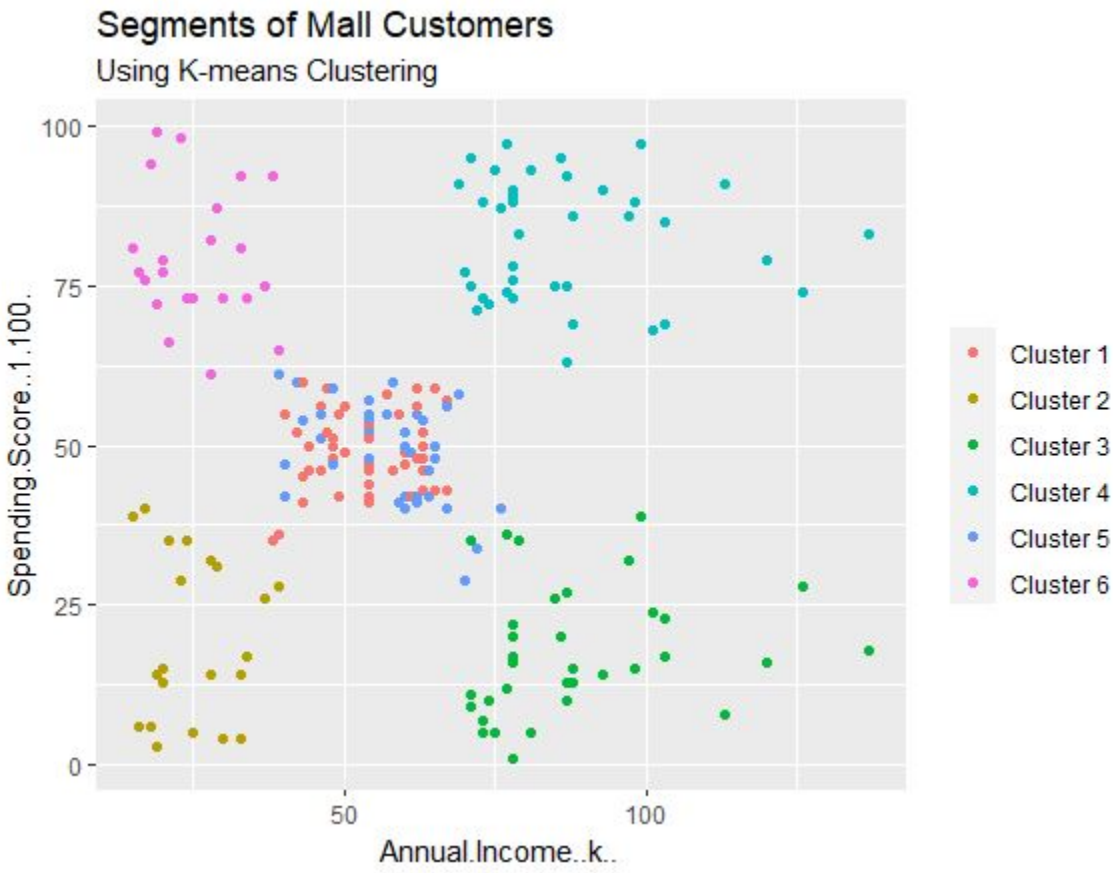


K-means Clustering

1. Deciding the number of clusters we want to create(K)
2. The algorithm selects random k data points
3. The selected points serve as the centroids for each clusters
4. Data points are allocated to each clusters based on euclidean distance
5. The iterative algorithm keeps on assigning data points till there is no change
6. The cluster centroids get updated based on the calculated mean
7. The total sum of squares of intracluster variation is calculated
8. This is done iteratively until lowest intracluster variation is found

Different Methods used to find optimal clusters




Segments of Mall Customers
Using K-means Clustering
Spending.Score..1.100.
100
75
50
25
0
50
100
Annual.Income..k..
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6

Cluster Analysis post segmentation

Cluster 6

Has low income but high spending score

Cluster 5

Has medium income and medium spending score

Cluster 4

Has high income and high spending score

Cluster 1

Has medium income and medium spending score

Cluster 2

Has low income and low spending score

Cluster 3

Has high income but low spending score

Let’s see the demonstration now!

Thank You :)