

CUSTOMER SEGMENTATION

ABSTRACT

In marketing, *customer segmentation* is the process of grouping customers by common traits. Discerning buying habits by customer type helps to market appropriately. For instance, it reveals the sizes of the various segments, how much we make from them, etc. This can help decide how to apportion the marketing budget.

In data science, *clustering* is the process of grouping objects by some common traits.

See the connection? Clustering — a data science method — is a good fit for customer segmentation — a use case. That said, as often happens, there is more to it when we start digging. We call this modeling.

CUSTOMER SEGMENTATION

ACKNOWLEDGMENT

A great deal of time and lot of effort has gone into completing this mini project report and documenting it. The number of hours spent in getting through various books and other materials related to this topic chosen by me have reaffirmed its power and utility in doing this lab work.

Several special people have contributed significantly to this effort. First of all, I we are grateful to our organisation Exposys Data Labs,Bangalore which provides us an opportunity in fulfilling our most cherished desire of reaching the goal.

We acknowledge and express our sincere thanks to our beloved Director and Principal

Our sincere and hearty thanks to our beloved parents, friends and others for their all-time support and co-operation.

Project Associate : Chetan M

CUSTOMER SEGMENTATION

TABLE OF CONTENTS

1. INTRODUCTION.....	4
2.METHEDOLOGY	5
3. APPLICATIONS.....	6
4. FEATURES.....	8
5. PROPOSED METHOD.....	9 -10
6. IMPLEMENTATION.....	11-15
7. CONCLUSION.....	16

INTRODUCTION

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.

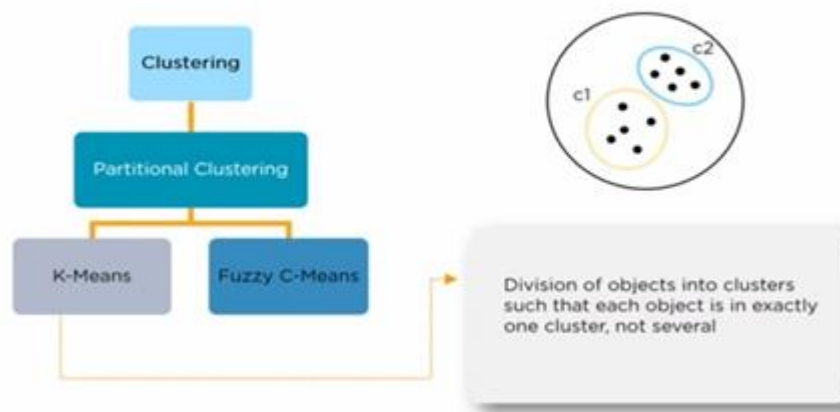
[Andrey Bu](#), who has more than 5 years of machine learning experience and currently teaches people his skills, says that “the objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.”

A cluster refers to a collection of data points aggregated together because of certain similarities. You’ll define a target number k , which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.

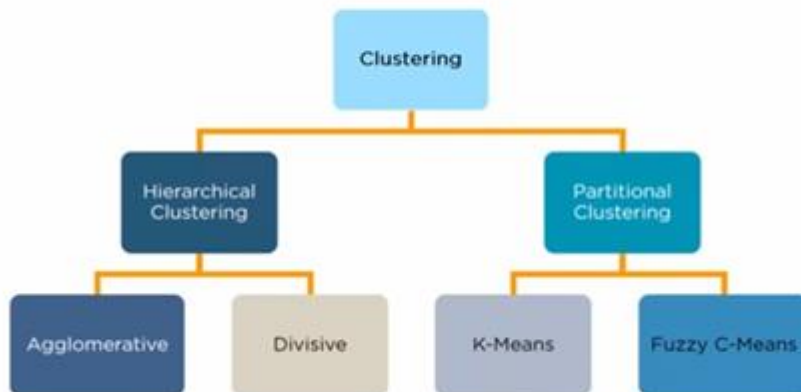
In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

To scale efficiently and effectively, expansion stage companies need to focus their efforts not on a broad universe of potential customers, but rather on a specific subset of customers who are most similar to their best current customers. The key to doing so is through customer segmentation.

Types of Clustering

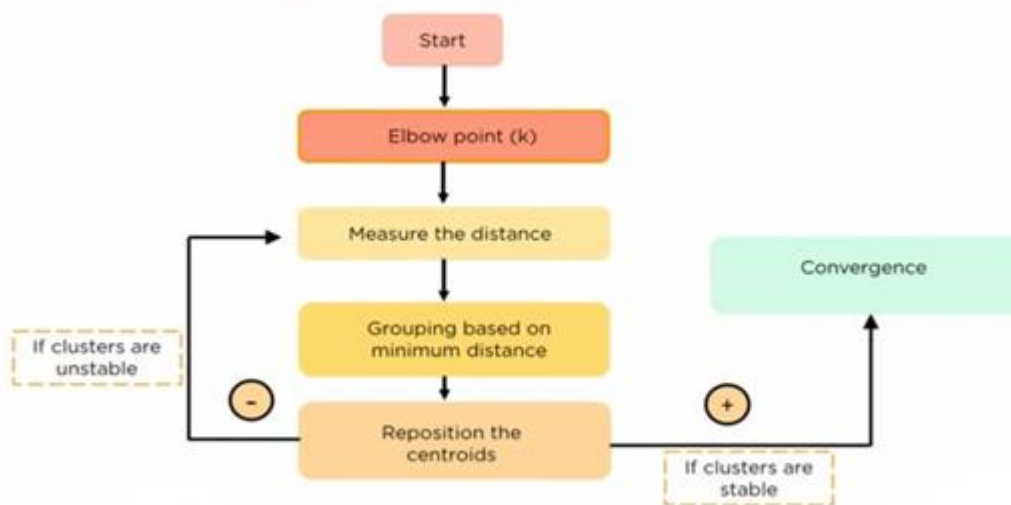


Types of Clustering



METHODOLOGY

How does K-Means clustering work?



Applications of Customer Segmentation

At the expansion stage, executing a marketing strategy without any knowledge of how your target market is segmented is akin to firing shots at a target 100 feet away — while blindfolded. The likelihood of hitting the target is a matter of luck more than anything else.

Without a deep understanding of how a company's best current customers are segmented, a business often lacks the market focus needed to allocate and spend its precious human and capital resources efficiently. Furthermore, a lack of best current customer segment focus can cause diffused go-to-market and product development strategies that hamper a company's ability to fully engage with its target segments. Together, all of those factors can ultimately impede a company's growth.

If best current customer segmentation is done right, however, the business benefits are numerous. For example, a best current customer segmentation exercise can tangibly impact your operating results by:

1. **Improving your whole product:** Having a clear idea of who wants to buy your product and what they need it for will help you differentiate your company as the best solution for their individual needs. The result will be increased satisfaction and better performance against competitors. The benefits also extend beyond your core product offering, since any insights into your best customers will allow your organization to offer better customer support, professional services, and any other offerings that make up their whole product experience.
2. **Focusing your marketing message:** In parallel with improvements to the product, conducting a customer segmentation project can help you develop more focused marketing messages that are customized to each of your best segments, resulting in higher quality inbound interest in your product.
3. **Allowing your sales organization to pursue higher percentage opportunities:** By spending less time on less lucrative opportunities and more on your most successful segments, your sales team will be able to increase its win rate, cover more ground, and ultimately increase revenues.
4. **Getting higher quality revenues:** Not all revenue dollars are created equal. Sales into the wrong segment can be more expensive to sell and maintain, and may have a higher churn rate or lower upsell potential after the initial purchase has been made. Staying away from these types of customers and focusing on better ones will increase your margins and promote the stability of your customer base.

Conducting best current customer segmentation research can have numerous other ancillary benefits, of course, but this guide will focus primarily on how it can impact the four cited above. The bottom line is that if you are able to sell more of your product to your most profitable customers, then you will be able to scale the business more efficiently and ensure that everything you do — from lead generation to new product development — revolves around the right things.



Advantages of Customer Segmentation:

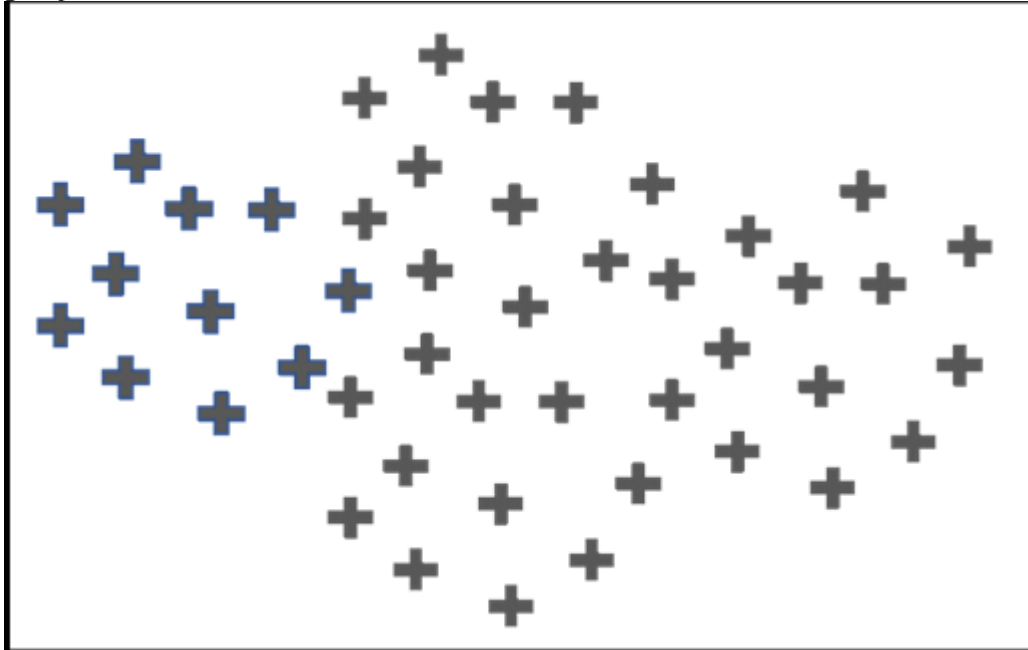
1. Determine appropriate product pricing.
2. Develop customized marketing campaigns.
3. Design an optimal distribution strategy.
4. Choose specific product features for deployment.
5. Prioritize new product development efforts.

Grouping Records into Clusters:

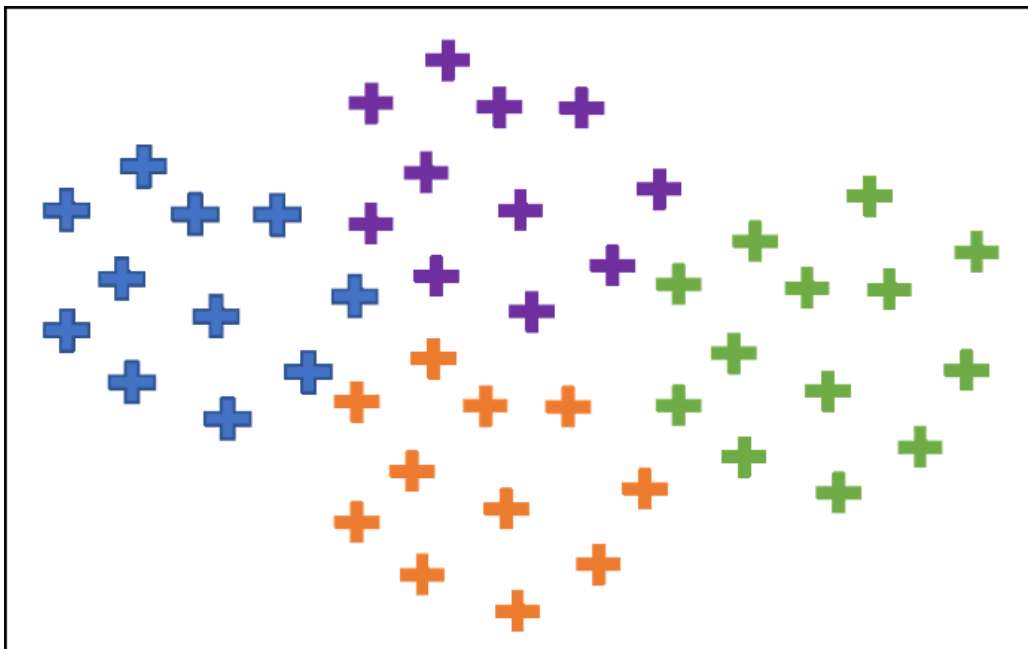
Next , we need to group the records onto clusters using this similarity measure. Pairs of records in the same cluster should be similar to each other, pairs in different clusters not. This is easier said than done. Its not clear that we have even fully thought through our problem formulation .The word ‘group’ seems to imply that a record should go into exactly one cluster.

Features:

Clustering fundamentals: Clustering is an unsupervised machine learning technique, where there are no defined dependent and independent variables. The patterns in the data are used to identify / group similar observations.



Original Dataset



After clustering

The objective of any clustering algorithm is to ensure that the distance between datapoints in a cluster is very low compared to the distance between 2 clusters. In other words, members of a group are very similar, and members of different groups are extremely dissimilar. We will use k-means clustering for creating customer segments based on their income and spend data.

K Means Clustering Algorithm

1. Specify number of clusters K .
2. Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.



K Means Clustering where $K=3$

The Elbow Method

Calculate the Within Cluster Sum of Squared Errors (WSS) for different values of k , and choose the k for which WSS first starts to diminish. In the plot of WSS-versus k , this is visible as an elbow.

CUSTOMER SEGMENTATION

The steps can be summarized in the below steps:

1. Compute K-Means clustering for different values of K by varying K from 1 to 10 clusters.
2. For each K, calculate the total within-cluster sum of square (WCSS).
3. Plot the curve of WCSS vs the number of clusters K.
4. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

IMPLEMENTATION

```
# Mall Customers - Unsupervised Learning

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# Importing the dataset

dataset = pd.read_csv('Mall_Customers.csv')

df = dataset.copy()

# Visualizing data

df.head()

# Getting the insides of the data

df.isnull().sum()

df.describe()

df.info()

# Making the independent variables matrix

X = df.iloc[:, [3, 4]].values # 3 & 4 column for plotting

print(X)

# One Hot Encoding the categorical data - Gender

df = pd.get_dummies(df, columns = ['Gender'], prefix = ['Gender'])
```

CUSTOMER SEGMENTATION

```
print(df)

#Using KMeans for clustering

from sklearn.cluster import KMeans

wcss = []

for i in range(1, 11):

    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10)

    kmeans . fit(X)

    wcss.append(kmeans.inertia_)

print(wcss)  #variance values are stored in empty list wcss

font_title = {'family' : 'normal',

              'weight' : 'bold',

              'size' : 35}

font_axes = {'family' : 'normal',

             'weight' : 'normal',

             'size' : 28}

#-----Elbow_method-----#

#Plotting Number of Clusters Vs wcss - The Elbow Method

plt.plot(range(1, 11), wcss)

plt.title('The Elbow Method', **font_title)
```

CUSTOMER SEGMENTATION

```
plt.xlabel('Number of clusters', **font_axes)
```

```
plt.ylabel('wcss', **font_axes)
```

```
plt.show()
```

```
#-----Clustering-----#
```

```
#Taking number of clusters = 5
```

```
kmeans = KMeans(n_clusters = 5, init = 'k-means++', max_iter = 300, n_init = 10)
```

```
print(kmeans)
```

```
y_kmeans = kmeans.fit_predict(X)
```

```
print(y_kmeans)
```

```
# PLOtting the clusters
```

```
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster1')
```

```
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster2')
```

```
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster3')
```

```
plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'yellow', label = 'Cluster4')
```

```
plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'pink', label = 'Cluster5')
```

```
plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s=300, c='cyan', label='centroid')
```

CUSTOMER SEGMENTATION

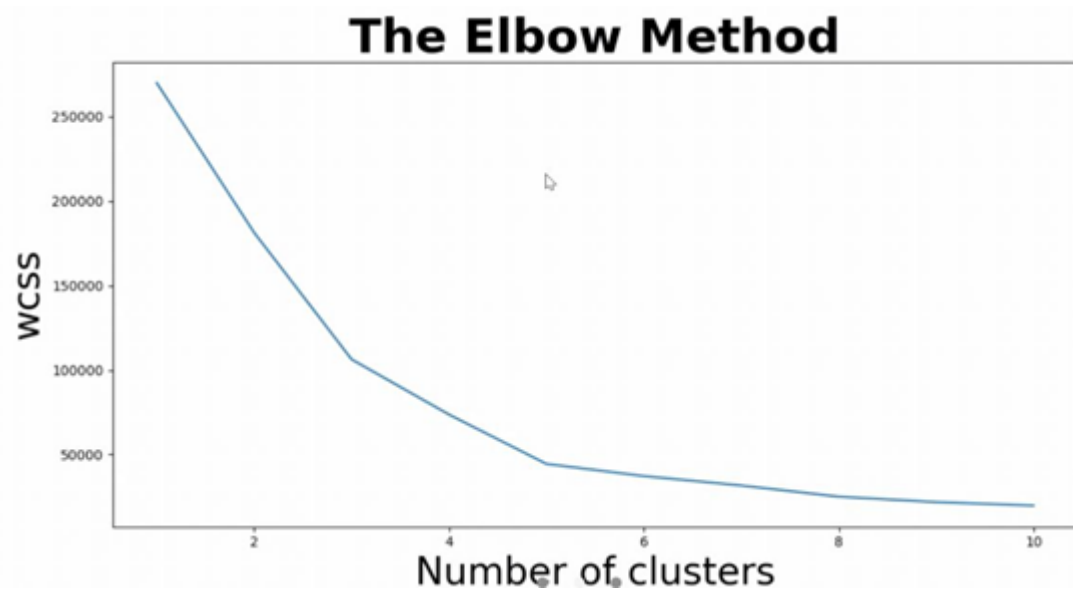
```
plt.title('Clusters of Customers', **font_title)
```

```
plt.xlabel('Annual income(k$)', **font_axes)
```

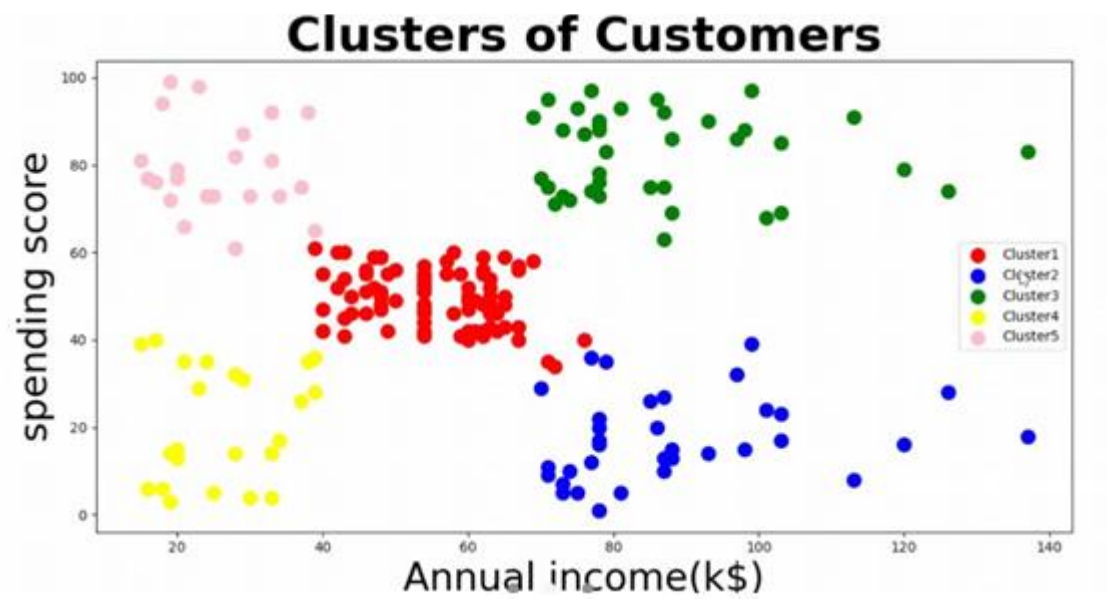
```
plt.ylabel('spending score', **font_axes)
```

```
plt.legend()
```

```
plt.show()
```



CUSTOMER SEGMENTATION



CUSTOMER SEGMENTATION

CONCLUSION

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major application of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.

In this post, we have seen how clustering as a data science method is a good fit for market segmentation as a use case. We have examined issues that arise when a factor has clean values; when a factor has a long tail of unstructured values; and when a factor has a small set of ordinal values.

To be effective, you must prepare and plan for the various challenges and hurdles that each step may present, and always make sure to adapt your process to any new information or feedback that might change its output.

Additionally, you cannot force feed this process on your business. If the key stakeholders that will be impacted by the best current customers segmentation process do not fully buy-in, then the outputs produced from it will be relatively meaningless.

If you properly manage the best current customer segmentation process, however, the impact it can have on every part of your organization — sales, marketing, product development, customer service, etc. — is immense. Your business will possess stronger customer focus and market clarity, allowing it to scale in a far more predictable and efficient manner.

Ultimately, that means no longer needing to take on every customer that is willing to pay for your product or service, which will allow you to instead hone in on a specific subset of customers that present the most profitable opportunities and efficient use of resources. That is critical for every business, of course, but at the expansion stage, it can often be the difference between incredible success and certain failure.