

THE CUSTOMER SEGMENTATION MODEL

WHAT IS CUSTOMER SEGMENTATION?

Customer segmentation simply means grouping customers according to various character.

It's a way for organizations to understand their customers. Knowing the differences between customer groups, it's easier to make strategic decisions regarding product growth and marketing.

The opportunities to segment are endless and depend mainly on how much customer data you have at your use. Starting from the basic criteria, like gender, hobby, or age, it goes all the way to things like “time spent of website X” or “time since user opened our app”.

There are different methodologies for customer segmentation, and they depend on four types of parameters:

- geographic,
- demographic,
- behavioral,
- psychological.

Geographic customer segmentation is very simple, it's all about the user's location. This can be implemented in various ways. You can group by country, state, city, or zip code.

Demographic segmentation is related to the structure, size, and movements of customers over space and time. Many companies use gender differences to create and market products. Parental status is another important feature. You can obtain data like this from customer surveys.

Behavioral customer segmentation is based on past observed behaviors of customers that can be used to predict future actions. For example, brands that customers purchase, or moments when they buy the most. The behavioral aspect of customer segmentation not only tries to understand reasons for purchase but also how those reasons change throughout the year.

Psychological segmentation of customers generally deals with things like personality traits, attitudes, or beliefs. This data is obtained using customer surveys, and it can be used to gauge customer sentiment.

ADVANTAGES OF CUSTOMER SEGMENTATION:

- budgeting,
- product design,
- promotion,
- marketing,
- customer satisfaction.

MACHINE LEARNING FOR CUSTOMER SEGMENTATION:

Machine learning methodologies are a great tool for analyzing customer data and finding insights and patterns. Artificially intelligent models are powerful tools for decision-makers. They can precisely identify customer segments, which is much harder to do manually or with conventional analytical methods.

There are many machine learning algorithms, each suitable for a specific type of problem. One very common **machine learning algorithm that's suitable for customer segmentation problems is the k-means clustering algorithm.** There are other clustering algorithms as well such as DBSCAN, Agglomerative Clustering, and BIRCH, etc.

EXPLORING CUSTOMER DATASET AND ITS FEATURES:

Let's analyze a customer dataset. Our dataset has 24,000 data points and four features. The features are:

- Customer ID – This is the id of a customer for a particular business.
- Products Purchased – This feature represents the number of products purchased by a customer in a year.
- Complaints – This column value indicates the number of complaints made by the customer in the last year
- Money Spent – This column value indicates the amount of money paid by the customer over the last year.

```
customersdata.head()
```

	customer_id	products_purchased	complains	money_spent
0	649	1	0.0	260.0
1	1902	1	0.0	79.2
2	2155	3	0.0	234.2
3	2375	1	0.0	89.0
4	2407	2	0.0	103.0

PRE-PROCESSING THE DATA SET:

Before feeding the data to the k-means clustering algorithm, we need to pre-process the dataset. Let's implement the necessary pre-processing for the customer dataset.

```
: customersdata.shape
```

```
: (24574, 4)
```

```
: customersdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24574 entries, 0 to 24573
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customer_id           24574 non-null  int64
1   products_purchased    24574 non-null  int64
2   complains             24574 non-null  float64
3   money_spent           24574 non-null  float64
dtypes: float64(2), int64(2)
memory usage: 768.1 KB
```

```
customersdata.describe()
```

	customer_id	products_purchased	complains	money_spent
count	2.457400e+04	24574.000000	24574.000000	24574.000000
mean	4.509005e+06	1.742085	0.001051	191.503347
std	2.592493e+06	1.088471	0.027208	171.373344
min	6.490000e+02	1.000000	0.000000	0.000000
25%	2.275220e+06	1.000000	0.000000	89.000000
50%	4.518730e+06	1.000000	0.000000	142.400000
75%	6.768568e+06	2.000000	0.000000	237.000000
max	8.999186e+06	13.000000	1.000000	3131.700000

CLUSTERING ALGORITHM

A clustering machine learning algorithm is an unsupervised machine learning algorithm. It's used for discovering natural groupings or patterns in the dataset. It's worth noting that clustering algorithms just interpret the input data and find natural clusters in it

```
# Import required libraries

import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
import plotly.express as px
import plotly.graph_objects as go
import matplotlib.pyplot as plt
#Load customers data
customersdata = pd.read_csv("customers-data.csv")
```

```
customersdata.head()
```

	customer_id	products_purchased	complains	money_spent
0	649	1	0.0	260.0
1	1902	1	0.0	79.2
2	2155	3	0.0	234.2
3	2375	1	0.0	89.0
4	2407	2	0.0	103.0

VISUALIZATION OF CUSTOMER SEGMENTS:

In this section, we'll be implementing some code using plotly express. This way we'll visualize the clusters in three dimensions, formed by our k-means algorithm. Plotly express is a library based on plotly that works on several types of datasets and generates highly-styled plots. First,

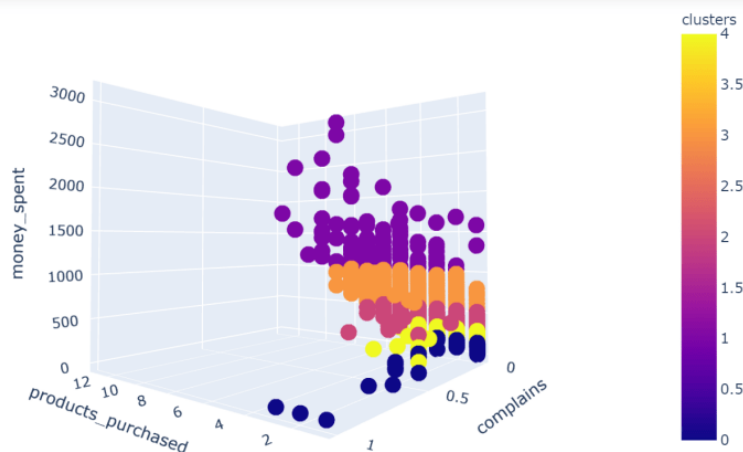
let's add a new column named 'clusters' to the existing customer data dataset. This column will be able to tell which customer belongs to what cluster.

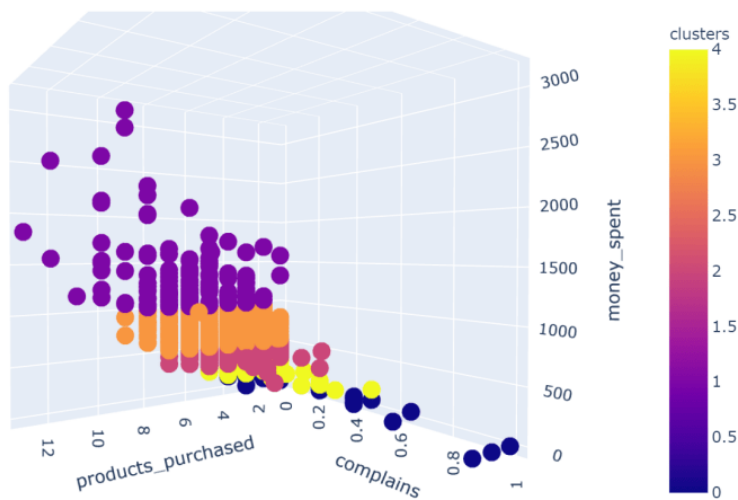
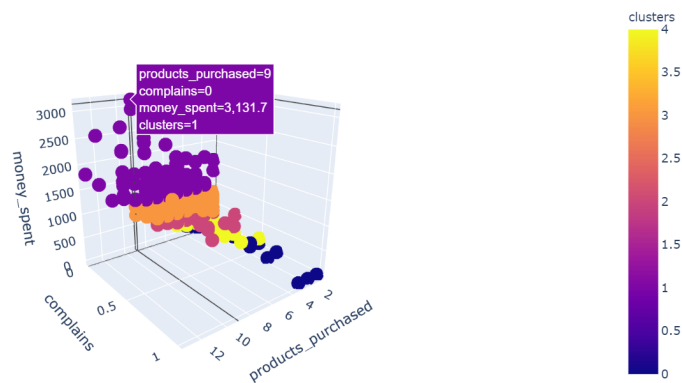
```
# Create data arrays
cluster_centers = kmeans_model_new.cluster_centers_
data = np.exp1(cluster_centers)
points = np.append(data, cluster_centers, axis=1)
points
# Add "clusters" to customers data
points = np.append(points, [[0], [1], [2], [3], [4]], axis=1)
customersdata["clusters"] = kmeans_model_new.labels_

customersdata.head()
```

	customer_id	products_purchased	complains	money_spent	clusters
0	649	1	0.0	260.0	4
1	1902	1	0.0	79.2	0
2	2155	3	0.0	234.2	4
3	2375	1	0.0	89.0	0
4	2407	2	0.0	103.0	0

```
# visualize clusters
figure=px.scatter_3d(customersdata,color='clusters',x="products_purchased",y=
"complains",z="money_spent",category_orders = {"clusters": ["0", "1", "2",
"3","4"]})
figure.update_layout()
figure.show()
```





CONCLUSION:

It's not wise to serve all customers with the same product model, email, text message campaign, or ad. Customers have different needs. A one-size-for-all approach to business will generally result in less engagement, lower-click through rates, and ultimately fewer sales. Customer segmentation is the cure for this problem.

Finding an optimal number of unique customer groups will help you understand how your customers differ, and help you give them exactly what they want. Customer segmentation improves customer experience and boosts company revenue. That's why segmentation is a must if you want to surpass your competitors and get more customers. Doing it with machine learning is definitely the right way to go.