Customer Segmentation with the Aid of Machine Learning

One of the fundamental principles of any successful business is recognizing the pivotal role that customers play in its success. Without satisfied customers, no business can thrive or survive in the long run. Meeting the needs and expectations of customers is essential for maintaining a positive reputation, building brand loyalty, and generating revenue.

Therefore, businesses must prioritize providing exceptional customer service, understanding their customers' preferences and behaviors, and continuously adapting their offerings to meet changing market demands. A strategy for achieving customer satisfaction is to utilize customer segmentation.

Source: Unsplash

What is Customer segmentation?

Customer segmentation is the process of dividing a customer base into smaller subgroups of customers who share similar characteristics, behaviors, or needs. It can be based on a variety of criteria, including demographic information such as age, gender, income, and location, as well as psychographic factors such as personality, values, and lifestyle. By segmenting customers based on various criteria, businesses can tailor their products and services to cater to the unique needs of each customer group. It is a key strategy for businesses looking to improve their marketing efforts and overall customer experience. Here are some of the advantages of customer segmentation:

1. Better understanding of customers
2. Improved customer experience
3. Increased profitability
4. Better resource allocation
5. Competitive advantage
6. Enhanced marketing effectiveness

In summary, customer segmentation offers a range of benefits for businesses.

How is it done?

Before the emergence of machine learning algorithms, businesses relied on various traditional methods for customer segmentation. These methods included conducting surveys, organizing focus groups, analyzing customer databases, examining sales data, and evaluating geographic data. While these traditional methods can still be effective in small business, they are often less accurate and less efficient than modern data-driven approaches.

The Machine Learning way

One way to achieve successful segmentation is by utilizing clustering algorithms. Clustering algorithms are a type of unsupervised machine learning technique that groups together similar data points based on their characteristics. In customer segmentation, clustering algorithms can be used to group customers together based on similarities in their behavior, preferences, demographics, and other relevant factors. There are several types of clustering algorithms, including:

1. K-means clustering
2. Hierarchical clustering
3. DBSCAN: Density-based spatial clustering of applications with noise (DBSCAN)
4. Gaussian mixture models (GMM)

This article will focus on using k-means clustering to perform customer segmentation using a bank dataset as a case study.

The dataset can be found on Kaggle at:
https://www.kaggle.com/datasets/ahmedmohameddawoud/bank-customers-segmentation.
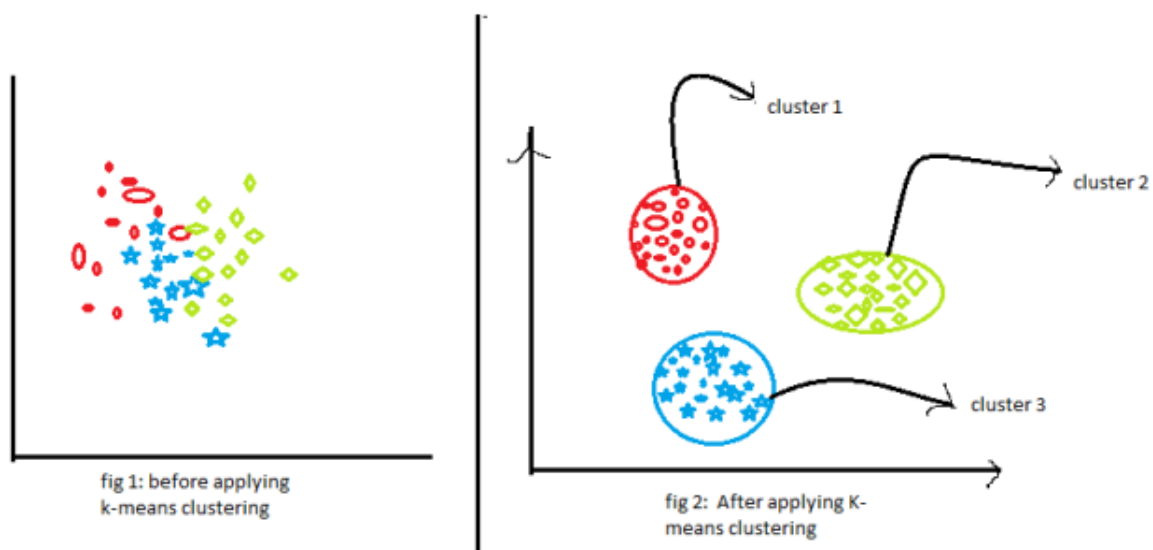The notebook can be found at: https://github.com/Anene-Chukwunenye/k-means_bank_segmentaion. The dataset includes various features such as age, occupation, marital status, education, as well as campaign-related factors and the results of those campaign efforts. These details are elaborated on in the provided notebook.

The main focus of this case study is on a bank that aims to enhance the number of term deposit subscriptions among its customers. The bank aims to achieve this goal through the implementation of marketing campaigns.

Our objective is to assist the bank in identifying different customer segments based on their subscription patterns, using k-means clustering. This will enable the bank to focus their marketing efforts in the appropriate direction for each segment to yield better outcomes.

K- means clustering

K-means clustering is a machine learning algorithm that is commonly used for unsupervised clustering of data. In this algorithm, data points are grouped into clusters based on their similarities. The algorithm starts by randomly selecting K cluster centers from the dataset, where K is the number of desired clusters. Customer Segmentation 4 Then, for each data point in the dataset, the algorithm calculates the distance between that point and each of the K cluster centers. The data point is then assigned to the cluster with the closest cluster center. This process continues iteratively until the optimal cluster centers are found. The outcome of k-means clustering is the formation of K clusters, with each cluster containing data points that are similar to each other based on their features. K-means clustering is widely used in customer segmentation, image processing, and other applications where unsupervised learning is required.

fig 1: before applying k-means clustering

fig 2: After applying K-means clustering

Source: https://shashwatgaur001.medium.com/k-means-clustering-and-its-use-cases-7128e7aa4226

Implementation of K-means Clustering Model

The end objective of this project is to determine in a practical manner how Customer Segmentation using machine learning can positively benefit the Financial Sector.
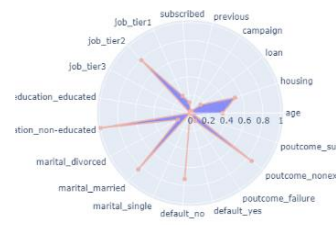
Steps taken to achieve the objective:

- Data collection
  - Loading the dataset
- Data cleaning
  - Removing missing values
- Data analysis- exploratory data analysis
  - Feature engineering
- K-means model implementation
  - Perform several iteration/tests to determine the best number of clusters to group the bank customers
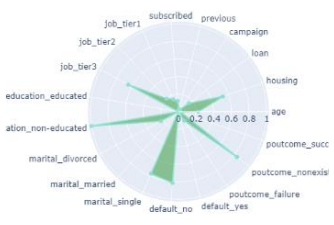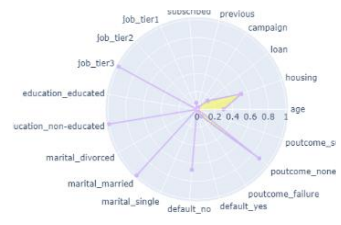
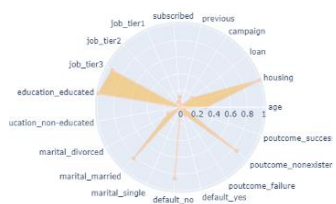Graphical visualization of the clusters:
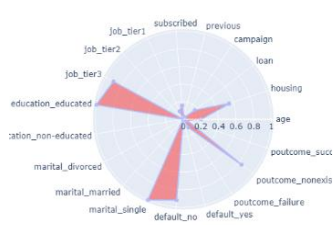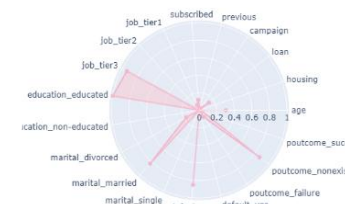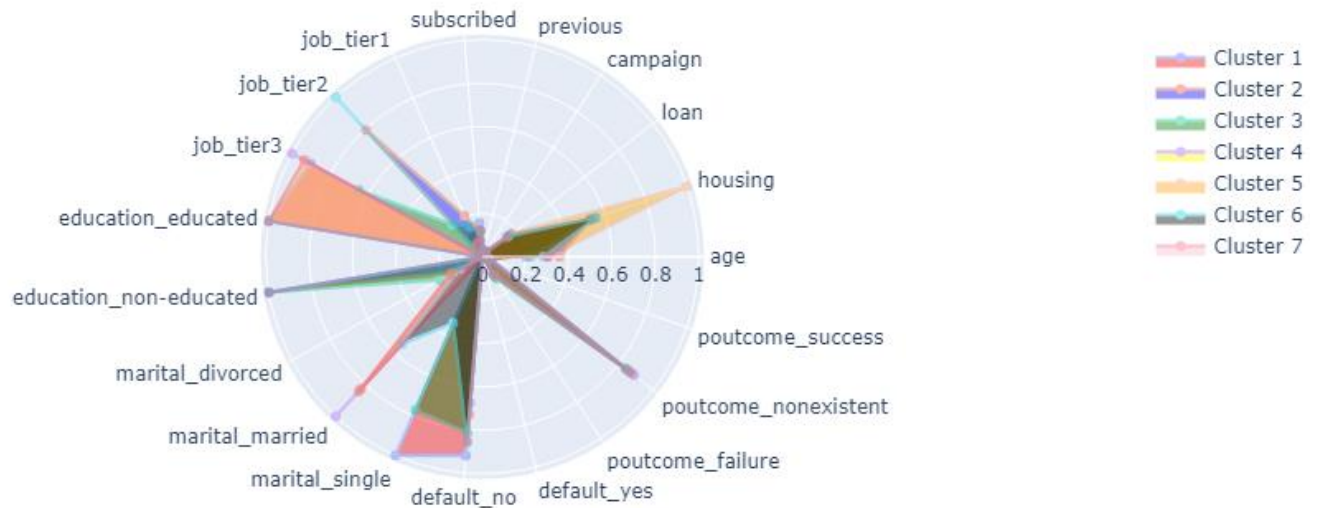

cluster1 cluster2 cluster3 cluster4

cluster5 cluster6 cluster7

combined visualization of all the clusters

Clusters Interpretation:

➢ Cluster1: shows a large category of young, educated, but mostly single individuals having tier3 jobs. A large number have not had any loan defaults, few have housing loans and the outreach campaign is relatively nonexistent.

➢ Cluster2: details more of married, non-educated individuals with tier2 jobs with a few having house loans. Cluster2 has a higher number of older customers.

➢ Cluster3: a cluster of mostly single non-educated customers. Most have not defaulted from their loans

➢ Cluster4: reveals an odd category; married non-educated customers with tier3 jobs and housing loans.

➢ Cluster5: details a very large number of married educated customers with tier3 jobs. Most are average in age and have housing loans.

➢ Cluster6: indicates a number of educated customers, some married and some single. They have tier2 jobs and a number have housing loans

➢ Cluster7: a little bit similar to cluster 5 except there is a lesser number of customers with loans.

Real-Case Uses of Customer Segmentation

Companies have started to discover the importance of using ML in clustering their customers in increasing profits. The list below shows just a few examples of companies using machine learning in customer segmentation:

- First Bank: First Bank is one of the oldest and largest banks in Nigeria, and the bank uses customer segmentation to better understand and serve its customers. First Bank has identified different customer segments, including personal banking customers, SME banking customers, and corporate banking customers.
- GTBank: Guaranty Trust Bank is a leading Nigerian bank that uses customer segmentation to create tailored products and services for different customer segments. The bank has identified four customer segments, including retail banking customers, SME banking customers, corporate banking customers, and e-banking customers.
- MTN: This telecommunication company has made clusters of their customers based on certain habits. What services are used and how much they spend on those services are the important information by which they create the clusters.
- North Face: A reputable fashion brand has partitioned their customers based on different lifestyle segments like, "explorer", "traveler", "hiker", "extreme athlete", there are still a lot more segments. With the information about the different segments relevant products are then push to each segment.
- Bolt: One of the world's largest ride-hailing companies, they have learnt to classify their customers based on riding history. With the clusters they have determined individuals who frequently request rides and those that do not. They are able to easily target promos and discounts to their customers

Conclusion

Through the use of machine learning we have demonstrated how effectively machine learning can greatly benefit the financial sector. Given the bank customer's data, our K-means model was able to partition the customers into different groups of individuals with similar characteristics. This can help the bank to understand the customer's behavior, preferences, and needs, and tailor their marketing strategies, loans and other product offerings to specific customer groups.