Customer Segmentation WebApp

I have done my Project on data science. Customer Segmentation is a popular application of unsupervised learning. Using clustering, identify segments of customers to target the potential user base. They divide customers into groups according to common characteristics like gender, age, interests, and spending habits so they can market to each group effectively. I used K-means clustering to make the clusters in the dataset. The main aim was to use and implement K-means clustering and visualize the distributions of data. Then analyze the new data onto clusters.

I used Jupiter notebook to run the test Python code and then ran it on vs code. This project has two folders static and templates, static has the generated clusters images, and templates have the HTML files for the website.

On the website it will ask to upload the CSV file after uploading the file it will make the clusters according to the dataset and generate the images of the clusters after clicking the submit button.
The app.py file has the main code for the backend of the website where we used K-means to create the clusters and have done some data cleaning to get the actual data, other than that the file has the k-means model downloaded which I've downloaded writing the code in Jupiter notebook.

Language: Python

Table of Contents

chapter:1

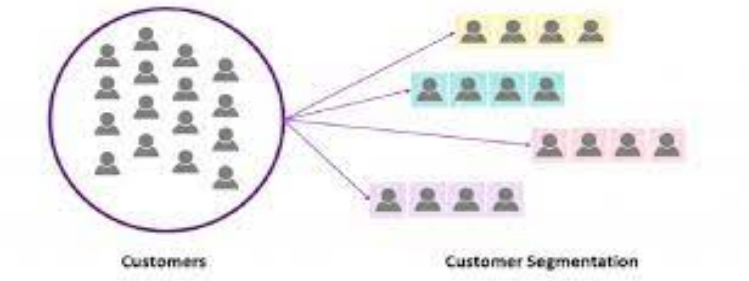
Introduction

Data Science is about data gathering, analysis, and decision-making.Data Science is about finding patterns in data, through analysis, and making future predictions.

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience.

Clustering algorithms: These algorithms divide customers into groups based on their characteristics and behavior. For example, k-means Clustering can be used to find the k number of clusters in a dataset.

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on



chapter:2

Existing Method

Algorithm used: K-means

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

It allows us to cluster the data into different groups and is a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training.

It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

chapter:3

Proposed method with Architecture

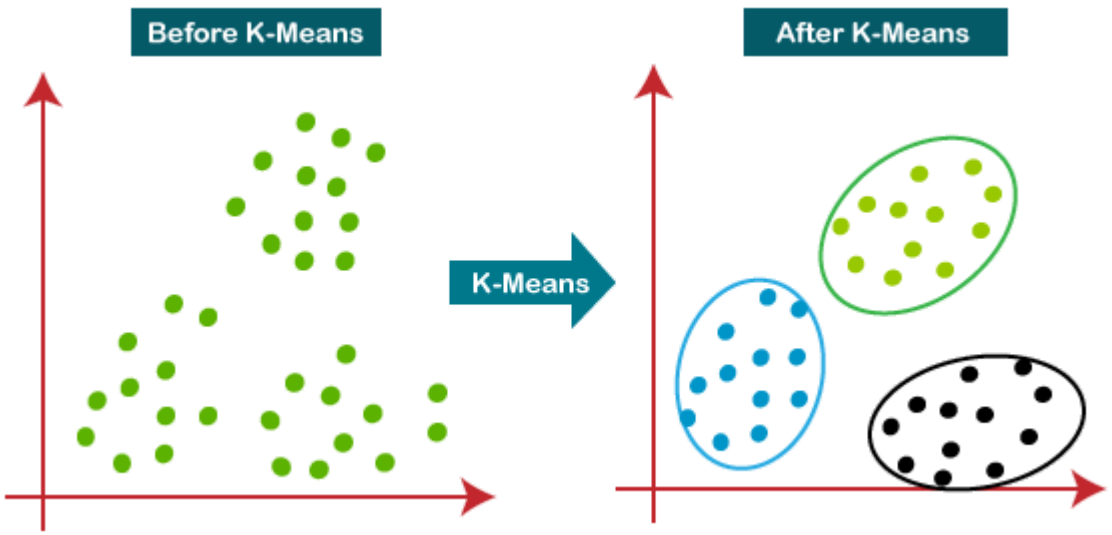
Customer Segmentation was made using the language Python and K-means algorithm. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest K-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has data points with some commonalities, and it is away from other clusters.

The below diagram explains the working of the K-means Clustering Algorithm:



Methodology

There are several different methods for using machine learning to perform customer segmentation, including:-

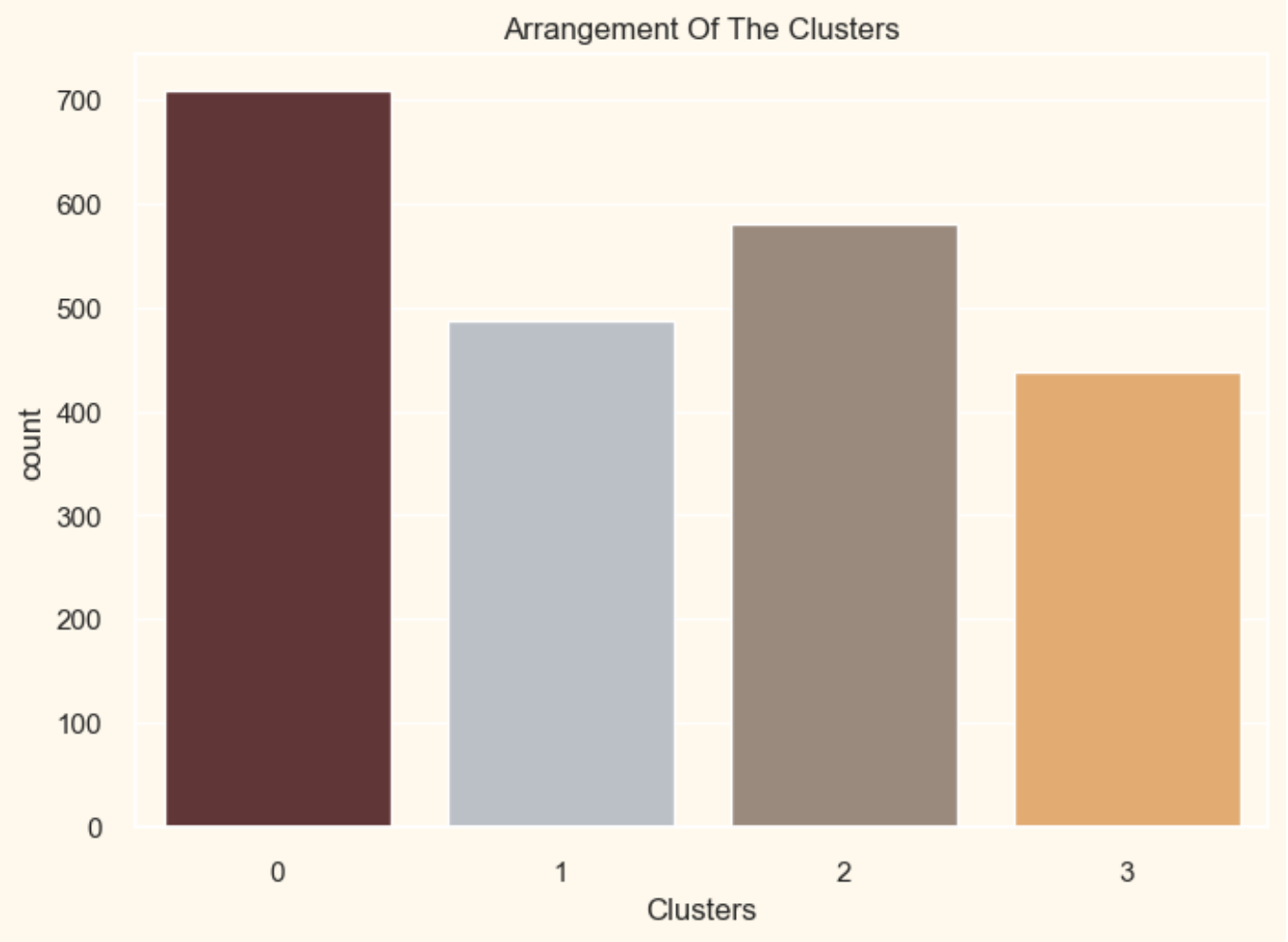
- **Clustering algorithms:** These algorithms divide customers into groups based on their characteristics and behavior. For example, **k-means Clustering** can be used to find the k number of clusters in a dataset.
- **Decision trees:** These algorithms use a tree-like model to identify the most important variables that influence customer behavior. By using decision trees, companies can determine which customers are most likely to respond to certain marketing campaigns or products.
- **Neural networks:** These algorithms can be used to model complex relationships between customers and their behavior. Neural networks can identify patterns in customer data that are not easily recognizable through traditional methods.
- **Association rule learning:** This method finds the relationships between customer attributes and behaviors, such as buying habits and product preferences. Association rule learning can help companies understand which products are frequently purchased together and target customers accordingly.

Implementation

Steps for implementation:

1. **Importing Libraries**
2. **Data Cleaning & Processing**
3. **model training**
4. **performing the Elbow Method (to get the number of clusters)**
5. **Training of K-Means model and Saving it**

For example in an image:



Conclusion

Unsupervised Clustering was done. Dimensionality reduction and agglomerative Clustering were both used. We developed four clusters and utilized them to profile clients in clusters based on their family configurations, income levels, and spending habits. This may be applied to creating better marketing plans.

In conclusion, customer segmentation is a critical aspect of marketing strategy, and machine learning has become an increasingly popular tool for automating the process. By using machine learning algorithms to process vast amounts of customer data, companies can quickly identify new trends and patterns, target specific customer segments with tailored promotions, and make more informed marketing decisions. With its ability to process data in real-time, eliminate the need for manual analysis, and continuously improve over time, machine learning is a powerful tool for customer segmentation.