**Online Retail customer segmentation**



**Project Overview**

In marketing, market segmentation is the process of dividing a broad consumer or business market, normally consisting of existing and potential customers, into sub-groups of consumers based on some type of shared characteristics.

Customer segmentation has a lot of potential benefits. It helps a company to develop an effective strategy for targeting its customers. This has a direct impact on the entire product development cycle, the budget management practices, and the plan for delivering targeted promotional content to customers. For example, a company can make a high-end product, a budget product, or a cheap alternative product, depending upon whether the product is intended for its most high yield customers, frequent purchasers or for the low-value customer segment. It may also fine-tune the features of the product for fulfilling the specific needs of its customers.

# Objectives of Online retail customer segmentation

# The objectives of online retail customer segmentation marketing are focused on optimizing marketing efforts, enhancing customer experiences, and driving business growth. Here are some common objectives:

# Targeted Marketing Campaigns:

# The primary objective of online retail customer segmentation is to create targeted marketing campaigns. By dividing customers into distinct segments based on their demographics, behaviors, preferences, or purchase history, businesses can tailor their marketing messages and offers to specific segments. This approach increases the relevance and effectiveness of marketing campaigns, leading to higher conversion rates and improved ROI.

1. **Personalized Customer Experience:**

# By understanding the unique needs, preferences, and behaviors of different customer segments, businesses can deliver customized product recommendations, personalized emails, and targeted promotions. This personalization enhances the customer experience, builds customer loyalty, and encourages repeat purchases.

1. **Customer Retention and Loyalty:**

# Effective customer segmentation helps identify the most valuable customer segments and develop strategies to retain and increase their loyalty. By understanding the characteristics and behaviors of loyal customers, businesses can implement retention tactics such as personalized offers, loyalty programs, and exclusive rewards to foster long-term relationships and encourage repeat business.

# New Customer Acquisition:

# By analyzing existing customer data, businesses can identify potential segments that align with their target audience and develop marketing strategies to attract and convert prospects within those segments. This approach maximizes the chances of acquiring customers who are more likely to become loyal and valuable to the business.

1. **Product Development and Merchandising:**

By analyzing segment-specific data, businesses can make informed decisions regarding product assortment, pricing strategies, and merchandising efforts. This helps optimize product offerings, ensure inventory alignment, and improve overall customer satisfaction.

1. **Competitive Advantage:**

Effective customer segmentation provides a competitive advantage in the online retail space. By understanding customers' needs and preferences better than competitors, businesses can differentiate themselves by delivering targeted and personalized experiences. This helps in building a strong brand reputation, attracting and retaining customers, and staying ahead of the competition.

# Importance of online retail customer segmentation marketing

# Here are several key reasons why it is crucial for businesses operating in the online retail industry:

# Customer segmentation allows businesses to better understand their customers and what they want. By analyzing data such as purchasing history, demographics, and behavioral patterns, businesses can gain insights into what drives each segment's buying decisions.

# It also helps businesses to allocate their marketing budget more effectively, by focusing their efforts on the segments that are most likely to generate revenue.

# Additionally, customer segmentation can help businesses to identify new opportunities for growth. By identifying gaps in the market or unmet customer needs within each segment, businesses can create new products or services that are specifically designed to meet those needs.

# 

# Problem Statement/ Description:

Cluster analysis is a popular technique used to identify customer segments.

However, businesses often face challenges when attempting to implement customer segmentation by cluster analysis. The problem statement revolves around these challenges and can be summarized as follows:

1. **Lack of a clear understanding of customer segments:**

Businesses struggle to define meaningful and actionable customer segments due to a lack of insights into customer behaviors, preferences, and demographic information. This makes it difficult to develop targeted marketing strategies and effectively allocate resources.

1. **Complex data analysis and interpretation:**

Customer data is often vast and diverse, consisting of various variables such as purchase history, demographic information, online interactions, and more. Analyzing and interpreting this data to identify distinct clusters can be complex and time-consuming, requiring expertise in statistical techniques and data mining.

1. **Selection of appropriate clustering algorithms:**

There are several clustering algorithms available, such as k-means, hierarchical clustering, and DBSCAN. Choosing the most suitable algorithm for a specific business and dataset can be challenging, as different algorithms have varying strengths and weaknesses.

In this project we will learn how to build a prototype of a Customer Segmentation model.

**Project Flow-**



## Requirements Objectives:

The project will involve the following requirements objectives:

1. **Data Cleaning and Preprocessing:** Dataset contains missing values, inconsistent formats, and other issues that need to be addressed before analysis can begin.
2. **Exploratory Data Analysis (EDA**): The objective is to determine the extent of missing data and implement techniques such as imputation or deletion to fill in or handle missing values effectively, enabling a comprehensive analysis.
3. **Feature Engineering:** Creating new features or transforming existing ones to extract meaningful information from the data.
4. **Customer Segmentation:** Using unsupervised learning techniques such as clustering to group customers with similar characteristics together.
5. **Evaluation and Interpretation**: Assessing the performance of the segmentation model and interpreting the results to extract actionable insights.

## Dataset:

Dataset contains transactional data based on spending score of a person by different profession.

Dataset explains how the person spend his money based on his profession, based on work experience and based on score, family size and qualification.

The dataset has 11 variables:

1. **Customer ID:** Unique identifier for each customer
2. **Gender:** Male or Female
3. **Ever Married:** Married or Unmarried customer
4. **Age:** Describe age of customer
5. **Graduated:** Describe qualification of customer whether Graduate or Undergraduate
6. **Profession:** IdentifyProfession of customer
7. **Work\_Experience:** Describe work experience of customer
8. **Spending\_Score:** Describe customer’s spending score
9. **Family\_Size:** Describe no. of persons living in family
10. **Var\_1:** Identify unique category of customer
11. **Segmentation:** Describer segmentation category of customer

## The dataset can be downloaded from the following link on Kaggle:

https://[www.kaggle.com/vijayuv/onlineretail](http://www.kaggle.com/vijayuv/onlineretail)

## Deliverables:

**The deliverables for this project include:**

1. A cleaned and preprocessed dataset ready for analysis.
2. EDA report with visualizations and insights into the dataset.
3. Feature engineering report with a description of the new features created or transformed.
4. Customer segmentation model report with a description of the approach taken, the number of clusters identified, and the characteristics of each segment.
5. Evaluation and interpretation report with a description of the performance of the segmentation model and actionable insights extracted from the results.

# Project Framework for Online Retail Customer Segmentation:

## This project will be structured around the following framework:

1. **Data Cleaning and Preprocessing**

**Remove duplicates:** Remove any duplicate transactions from the dataset.

**Handle missing values:** Identify missing values in the dataset and determine the best approach for handling them (e.g., imputation, deletion).

**Check for inconsistencies:** Identify any inconsistencies in the dataset, such as different formats for the same variable or values that do not make sense.

## Exploratory Data Analysis

**Descriptive statistics:** Calculate summary statistics for each variable, such as mean, median, standard deviation, and range.

**Visualizations**: Create visualizations, such as heatmap, histograms, scatterplots, and boxplots, bargraph,to identify patterns and trends in the data.

**Insights**: Use the results of the EDA to gain insights into the dataset and identify potential customer segments.

## Feature Engineering

**Create new variables:** Create new variables that capture additional information about the customers or their transactions (e.g., Customer segmentation by dummy variables).

**Transform variables**: Transform existing variables to make them more useful for segmentation.

**Select variables:** Identify the most relevant variables for customer segmentation based on their importance and correlation with other variables.

## Customer Segmentation

**Choose a clustering algorithm:** Doing analysis by Supervise and unsupervised learning algorithm for customer segmentation (e.g., One Hot encoding, Decision Tree, Random Forest, k-means clustering).

**Determine the number of clusters:** Determine the optimal number of clusters based on the spending score score or other performance metrics.

**Cluster analysis:** Perform cluster analysis to identify the characteristics of each segment and interpret the results. (Cluster analysis using Age and Family Size)

## Evaluation and Interpretation

**Evaluation metrics:** Evaluate the performance of the segmentation model using Age and Family Size.

**Interpretation:** Interpret the results of the segmentation model to extract actionable insights about the different customer segments.

**Business recommendations**: Make business recommendations based on the insights gained from the customer segmentation analysis.

# Code Explanation :

Here is the simple explanation for the code which is provided in the code.py file.

**Section 1**: Importing the Required Libraries

In this section, we import the required libraries that will be used throughout the project. We will be using Pandas for data manipulation, Numpy for numerical operations, Matplotlib and Seaborn for data visualization, and Scikit-learn for preprocessing and machine learning.

**Section 2**: Loading and Cleaning the Dataset

In this section, we load the Online Retail dataset into a Pandas DataFrame and clean the data by removing missing values. We also engineered a new feature with dummy variable.

**Section 3:** Feature Engineering and Scaling

In this section, we engineer dummy variables on category var1. We then scale the features using StandardScaler from scikit-learn to ensure that each feature is on the same scale.

**Section 4:** Model Training and Selection

In this section, we use the KMeans clustering algorithm from scikit-learn to train a customer segmentation model using the scaled features. We use the elbow method to determine the optimal number of clusters, and then fit the KMeans model with that number of clusters. We save the selected dataset with the cluster labels as a new CSV file.

**Section 5:** Evaluation and Interpretation

In this section, we evaluate the performance of the segmentation model, which measures the similarity of data points within a cluster compared to other clusters. We then interpret the results of the segmentation model by calculating summary statistics for each customer segment, such as mean, median and Standard Deviation. Finally, we make business recommendations based on the insights gained from the segmentation analysis.

## Implementation Guide:

**To implement these future work steps, we can follow these steps:**

1. First, we can expand the feature engineering step to create additional features that might be useful for segmentation. We can also explore external data sources to see if they can provide additional insights.
2. Next, we can perform hyperparameter tuning for the KMeans model using cross- validation techniques to find the optimal number of clusters for our dataset.
3. We can also explore alternative clustering algorithms and compare their performance to that of the KMeans algorithm.
4. Finally, we can create visualizations and reports to communicate the results of the segmentation analysis to stakeholders.

By implementing these future work steps, we can improve the segmentation analysis and provide more meaningful insights to stakeholders.

**Conclusion-**

In summary, online retail customer segmentation marketing is essential for businesses to achieve targeted marketing, deliver personalized experiences, optimize resources, enhance customer retention and loyalty, improve customer acquisition efforts, optimize product development and merchandising, and gain a competitive advantage. By leveraging customer segmentation effectively, businesses can drive growth and success in the dynamic online retail environment.

Solving the above challenges and implementing customer segmentation by cluster analysis effectively can provide businesses with valuable insights into their customer base, enabling them to personalize marketing efforts, improve customer satisfaction, increase customer retention, and drive revenue growth.