

Title: Customer segmentation

First name	Last Name	IIT Email
Srujan	Shetty	sshetty16@hawk.iit.edu
Aliance	Tedonfouet	atedonfouet@hawk.iit.edu
Karthik	Gande	Kgande@hawk.iit.edu
Shilpitha Reddy	Peruvala	speruvala@hawk.iit.edu
Sri Bhavya	Chaganti	schaganti@hawk.iit.edu

Important Notes:

- Each group must submit ONLY one copy by a single team member!
- Do not worry! Since you will list all team members in the table above.

1. Introduction

Introduce the background of your application and give me the motivations why you want to do that.

The application revolves around customer segmentation within the context of an online retail business. Customer segmentation is a crucial aspect of marketing strategy, aiming to divide a customer base into groups that share similar characteristics such as purchasing behavior, preferences, and demographics. By understanding the distinct needs and behaviors of different customer segments, businesses can tailor their marketing efforts more effectively, improve customer satisfaction, and ultimately drive higher revenues.

The motivation behind this application is to enhance the marketing strategies and overall business performance of the online retail company. By segmenting customers based on their purchasing behavior and patterns, the company can gain insights into the preferences of different customer groups. This allows for personalized marketing campaigns, product recommendations, and pricing strategies, leading to increased customer satisfaction and loyalty. Additionally, customer segmentation can help identify high-value customers for targeted retention efforts and optimize inventory management by understanding demand patterns.

2. Data Sets

Briefly introduce your data sets, such as which application or domain the data belongs to, where did you collect it, how large it is, how many features there are, what is your target variable, and so forth

Tell where the data is, such as giving Kaggle URL

Tell the size of the data

Tell the variables

Tell the target variables, if you are going to perform predictive tasks

The data set used for this application is the "Online Retail" data set, which contains transaction records of a UK-based non-store online retail company. The data set was collected between 01/12/2010 and 09/12/2011. The company specializes in selling unique all-occasion gifts etc.

Data Location:

The data set can be found on Kaggle at the following URL:
<https://www.kaggle.com/datasets/puneetbhaya/online-retail/data>

Size of the Data:

The size of the data set is relatively large (541909 columns and 8 rows), containing transactional records spanning almost a year. The exact size can be determined by checking the file size, depending on the volume of transactions during the specified time period.

Variables:

The variables included in the data set are as follows:

InvoiceNo: A unique identifier for each invoice. Multiple transactions with the same invoice number indicate that they were part of the same purchase.

StockCode: Identifier for the items contained in each invoice.

Description: Textual description of each stock item.

Quantity: The quantity of each item purchased.

InvoiceDate: Date and time of each purchase.

UnitPrice: The price of each item.

CustomerID: Identifier for the customer making the purchase.

Country: Country of the customer.

Target Variable: In this, the target variable may vary depending on the specific task being performed. For example, if the goal is to segment customers based on their purchasing behavior, the target variable would be the customer segments themselves. However, if the objective is to predict future purchases or customer churn, the target variable would be relevant metrics such as purchase frequency or customer retention status.

3. Research Problems

List your research problems, that is, what kinds of problems you want to solve.

You cannot simply say I want to explore the data and find the patterns

If you decide to work on a classification task, you must identify labels.

If your project is involved with multiple data mining tasks, you should clearly mention each problem and

why you want to do that.

You should provide finer-grained research problems that can be solved by data analysis/mining techniques. If it is an implementation project, you should introduce the challenges in implementing or development and how you will evaluate them

Research Problem:

Customer segmentation involves dividing a customer base into distinct groups based on similarities in their purchasing behavior, preferences, demographics, or other relevant characteristics. The primary goal of customer segmentation is to identify homogeneous groups of customers with similar needs and behaviors so that tailored marketing strategies and promotions can be developed for each segment.

In the context of the online retail business dataset, customer segmentation can offer several benefits. In summary, customer segmentation is a critical research problem for online retail businesses seeking to optimize their marketing strategies, enhance customer satisfaction, and drive revenue growth. By dividing the customer base into meaningful segments and tailoring marketing efforts accordingly, businesses can unlock opportunities for increased engagement, loyalty, and profitability.

4. Potential Solutions

For each problem you list above, figure out feasible solutions, and introduce your plan to perform experiments

Customer Segmentation:

Feasible Solution:

Utilize clustering algorithms such as K-means, hierarchical clustering, or Gaussian mixture models to segment customers based on features such as purchase frequency, recency, monetary value, and product preferences.

Alternatively, employ dimensionality reduction techniques like PCA (Principal Component Analysis) or t-SNE (t-distributed Stochastic Neighbor Embedding) to visualize and identify natural clusters in the data. This alternative is for reference sharing.

Experiment Plan:

Data Preprocessing: Clean the dataset, handle missing values, and transform categorical variables into numerical representations if necessary.

Feature Engineering: Extract relevant features such as purchase frequency, recency, monetary value, and product categories.

Model Selection: Experiment with different clustering algorithms (e.g., K-means, hierarchical clustering) and dimensionality reduction techniques (e.g., PCA, t-SNE) to identify the optimal approach for customer segmentation.

Evaluation Metrics: Use silhouette score, Davies–Bouldin index, or visual inspection of cluster separation to evaluate the quality of segmentation.

Interpretation and Validation: Analyze the resulting customer segments to understand their characteristics and validate the segmentation by assessing the distinctiveness and actionable insights provided by each segment.

5. Evaluations

There could be multiple solutions for a same problem, You must figure out how to evaluate them and the details about your evaluations, for example, hold-out or N-folds evaluation?, which metrics you will use for evaluations.

For customer segmentation, metrics such as silhouette score or Davies–Bouldin index can be used to evaluate the effectiveness of clustering algorithms.

6. Expected Outcomes

Introduce your expected outcomes for your project

Clear Customer Groups: We'll define distinct groups of customers based on their buying habits and characteristics.

Better Marketing Strategies: We'll learn what each customer group prefers, so we can tailor marketing messages and offers to them.

Happier Customers: By giving customers what they want, we'll increase their satisfaction and loyalty.

More Profitable Customers: By understanding our customers better, we can encourage them to spend more overtime.

Continuous Improvement: We'll keep refining our approach based on feedback and data, making our strategies even more effective over time.