

CUSTOMER SEGMENTATION REPORT



TABLE OF CONTENTS

- 01** Project Overview
- 02** Project Methodology
- 03** Challenges
- 04** Conclusion

PROJECT OVERVIEW

Project Overview:

The "Customer Segmentation and Classification" project aims to leverage unsupervised machine learning techniques for customer segmentation and subsequently explore the potential for supervised model training to enable targeted marketing. The project focuses on understanding shopping patterns and providing personalized marketing recommendations based on customer behavior.

Problem Statement:

The challenge lies in the limitations of traditional, one-size-fits-all marketing strategies. Recognizing and addressing the unique needs and preferences of individual customers is essential for building stronger customer relationships and driving sustained business growth. The project addresses the following problem by Segmenting customers based on latent customer behavioral, psychological and financial characteristics

Objectives:

- Implement unsupervised learning (K-Means clustering) for customer segmentation.
- Explore the feasibility of supervised model training for targeted marketing.
- Develop a real-time API for predicting customer segments.

The Customer Segmentation and Classification project emerged from the need to enhance marketing strategies by understanding customer behavior. This report provides a comprehensive overview of the project, outlining the key objectives, methodology, and outcomes.



PROJECT METHODOLOGY

2.1 Data Collection:

For data collection, we sourced transaction and customer data from various channels, including point-of-sale systems and customer databases. The primary data repository is an AWS S3 bucket, ensuring secure storage and easy access to the datasets. AWS S3 provides a scalable and reliable solution, aligning with the project's data-driven approach.

2.2 Data Processing:

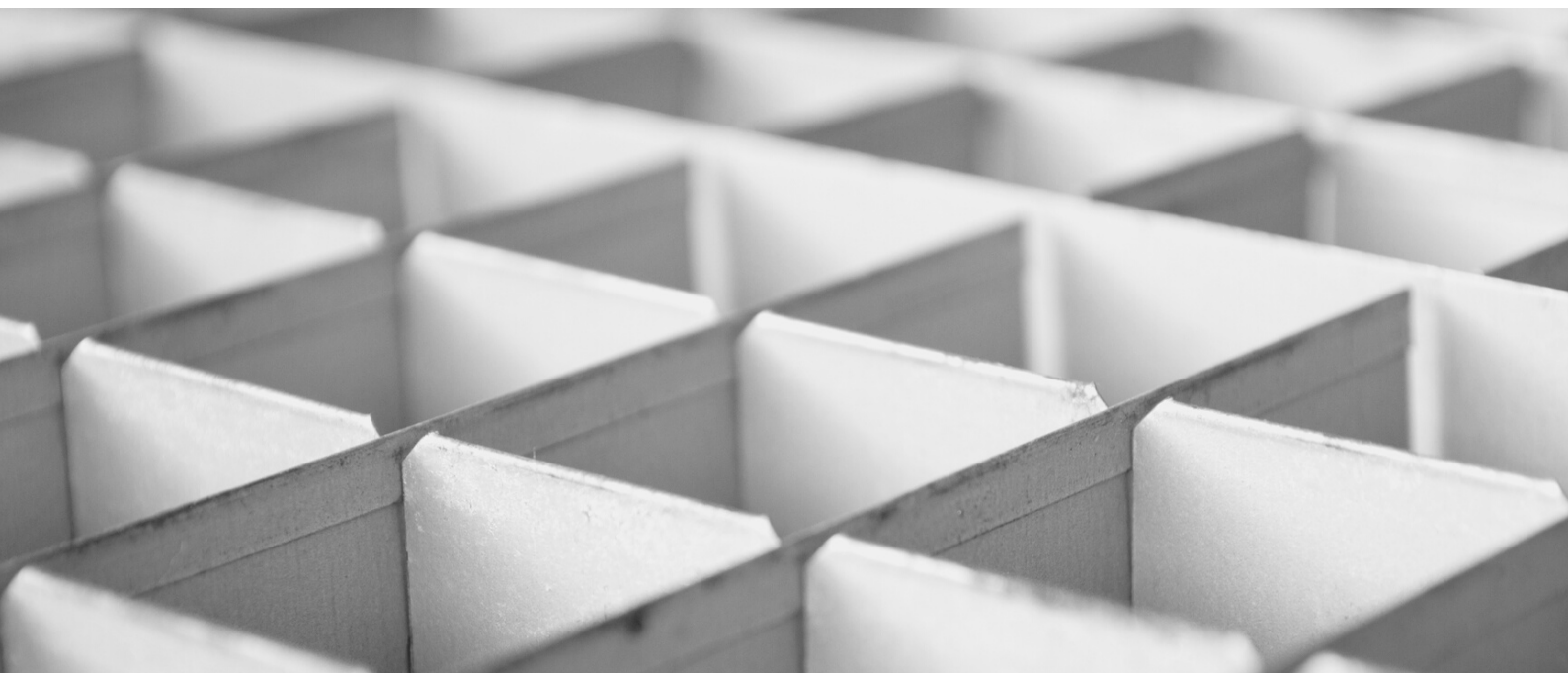
Data processing involves several critical steps to ensure quality and consistency. The Data Validation Module employs custom checks to validate the data structure, confirming alignment with predefined features. The Data Transformation Module is responsible for handling missing values, outliers, and encoding categorical variables. The entire process is version-controlled using Data Versioning Control (DVC) to track changes in the data over time.

2.3 Customer Segmentation:

Customer segmentation relies on the application of the K-Means clustering algorithm. K-Means is utilized to group customers based on their shopping behavior, considering features such as purchase frequency, average transaction value, and product category preferences. The resulting clusters provide actionable insights into distinct customer segments.

2.4 API Development:

The API Module is developed using Flask to facilitate real-time predictions. The API endpoints are designed to accept customer data as input, enabling dynamic predictions. The API seamlessly integrates with the clustering results or the trained supervised model, ensuring that recommendations align with the segmented customer groups. The endpoints include /predict for real-time predictions and /train for retraining the system if necessary.



CHALLENGES

4.1 Data Quality Challenges:

Challenge:

- Inconsistent data quality from various sources posed a challenge in maintaining uniformity for clustering.
- Lack of Proper System Resources

Solution:

- Implemented rigorous data validation checks to identify and rectify inconsistencies.
- Established communication channels with data providers for continuous quality improvement.
- A prototype has been only developed and further improvements can be made on top of this prototype

4.2 Model Performance Challenges:

Challenge:

- Initial iterations of the K-Means clustering model exhibited suboptimal performance due to sensitivity to outliers.

Solution:

- Applied robust preprocessing techniques to handle outliers and normalize data.
- Iteratively adjusted hyperparameters and conducted feature engineering to enhance model stability.

4.3 System Integration Challenges:

Challenge:

- Integrating the API module with existing systems presented compatibility and versioning challenges.

Solution:

- Adopted a modular design approach, allowing seamless integration with existing systems.
- Implemented version control through Git for API development, ensuring consistent updates.

CONCLUSION

In conclusion, the project has successfully addressed the challenges associated with customer segmentation and targeted marketing. The implementation of K-Means clustering has provided valuable insights into customer behavior, enabling the delivery of personalized marketing strategies. Key achievements include:

- **Accurate Segmentation:** The K-Means clustering algorithm effectively categorized customers into distinct segments based on their shopping patterns.
- **Real-time Predictions:** The API module facilitates real-time predictions, aligning with clustering outcomes

. Future Work:

6.1 UX Design Enhancement:

Recommendation:

- Implement a user-friendly dashboard to visually represent customer segments and marketing insights.

6.2 Supervised Learning Exploration:

Recommendation:

- Explore the integration of supervised learning models to further refine personalized marketing recommendations.

6.3 Automated Retraining:

Recommendation:

- Introduce automated retraining mechanisms to ensure the model stays relevant to evolving customer behavior.

6.4 Additional Features:

Recommendation:

- Incorporate customer feedback loops for continuous improvement.
- Explore the integration of more advanced machine learning models for richer insights.