Customer Segmentation

Beginner Project Report submitted in partial fulfillment.

of the requirement for the internship of

Company-Zidio Software Development

Submitted By
DS Group 4(2 months intern)
Darshan Jain
Lavya Thaper
Hritik karan
Puri Pranav

Under the Guidance of

Mr. Chandan Mishra
Mr. Pitabas Pradhan

ACKNOWLEGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. We express our profound gratitude we give to our **Mr. Chandan Mishra Sir**, our respectable project guide, for his gigantic support and guidance. Without his counseling our project would not have seen the light of the day.

We extend our sincere thanks to **Mr. Pitabas Pradhan Sir**, Internship co-ordinator for offering valuable advice at every stage of this undertaking.
We would like to thank all the staff members who willingly helped us. We are grateful to ZIDIO SOFTWARE DEVELOPMENT 2 MONTH INTERNSHIP for giving us this
opportunity.

The days we have spent in the internship will always be remembered and also be reckoned as guiding in our career.

1. **Darshan Jain**
2. **Lavya Thaper**
3. **Hritik Karan**
4. **Puri Pranav**

Table of Contents

Introduction

Embarking on a Data Science Journey at Zidio Development! 🚀

In the dynamic world of business analytics, customer segmentation is crucial for personalized marketing and enhanced customer engagement. At Zidio Development, we recognize the importance of understanding our customers at a granular level to drive strategic decisions and foster lasting relationships. This project explores customer segmentation using Python's powerful libraries—NumPy, Pandas, Scikit-Learn, and Seaborn.

The journey began with meticulous data preparation, addressing missing values and extracting essential metrics through descriptive statistics. Advanced data visualization techniques, such as heatmaps, revealed intricate patterns and correlations, forming the basis for effective segmentation strategies.

Normalization of the data using MinMaxScaler preceded the application of K-Means Clustering to identify distinct customer segments. The Elbow Method and Silhouette Method determined the optimal number of clusters, ensuring meaningful and actionable segments.

This project aimed to derive actionable insights that inform targeted marketing campaigns and personalized customer engagement strategies. The ultimate goal was to enhance customer experiences and drive business growth through informed, data-driven decisions.

Objectives

- **Data Preparation**: Ensure data integrity by addressing missing values, handling outliers, and standardizing data formats to provide a clean dataset for analysis.
- **Descriptive Statistics and Initial Analysis**: Utilize descriptive statistics to understand the basic metrics of the dataset and uncover initial insights that guide the segmentation process.
- **Data Visualization**: Employ advanced visualization techniques, including heatmaps, scatter plots, and pair plots, to identify patterns, correlations, and trends critical for segmentation.
- **Feature Engineering**: Create new features that capture nuanced customer behaviors and preferences, enhancing the segmentation process.
- **Data Normalization**: Apply MinMaxScaler to normalize the data, ensuring that all features contribute equally to the clustering algorithm.
- **Clustering Analysis**: Implement K-Means Clustering to identify distinct customer segments based on their behaviors and characteristics.
- **Cluster Validation**: Use the Elbow Method and Silhouette Method to determine the optimal number of clusters and evaluate their effectiveness in creating meaningful segments.
- **Segment Profiling**: Analyze each segment to understand its unique characteristics, behaviors, and preferences, providing a detailed profile of each customer group.
- **Actionable Insights**: Derive insights from the segmented data to inform targeted marketing campaigns, personalized customer engagement strategies, and business decision-making.
- **Strategic Recommendations**: Develop practical recommendations based on the insights gained, aimed at optimizing marketing efforts, improving customer satisfaction, and driving business growth.

Problem Definition

In today's competitive business environment, understanding and catering to diverse customer needs is paramount. Companies often struggle with generic marketing strategies that fail to resonate with individual customers, leading to suboptimal engagement and conversion rates. The challenge lies in effectively segmenting a heterogeneous customer base to identify distinct groups with similar characteristics and preferences.

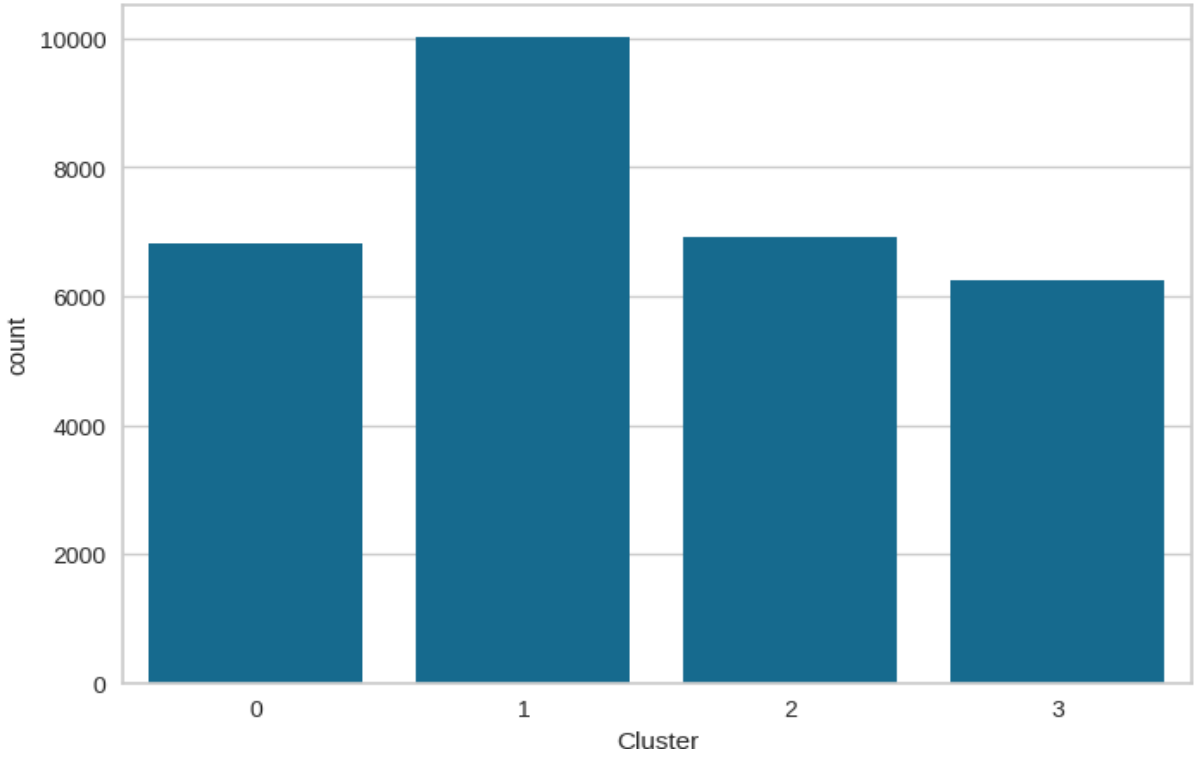
Specific Challenges Addressed:

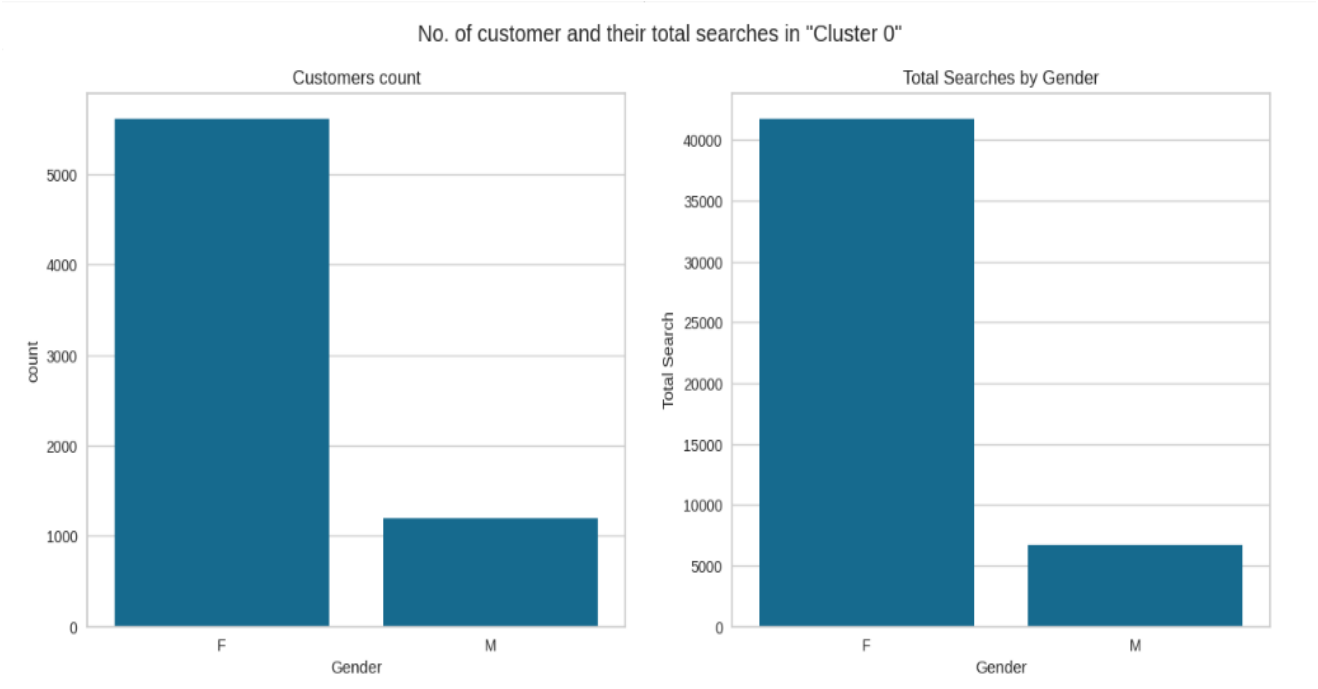
1. **Data Integrity and Quality**: The initial dataset contained missing values, inconsistencies, and noise that needed to be addressed to ensure reliable analysis.
2. **Complex Customer Behavior**: Customers exhibit varied behaviors and preferences, making it difficult to pinpoint commonalities and define clear segments without advanced analytical techniques.
3. **Optimal Segmentation**: Determining the optimal number of segments is crucial. Too few segments might overlook unique customer traits, while too many could result in fragmented and less actionable insights.
4. **Actionable Insights**: Beyond identifying segments, the goal was to derive insights that could translate into practical strategies for targeted marketing and personalized customer engagement.
5. **Strategic Implementation**: Implementing the insights in a way that aligns with business goals and enhances customer satisfaction requires a deep understanding of the derived segments and their implications.

Software Requirements and Tools used

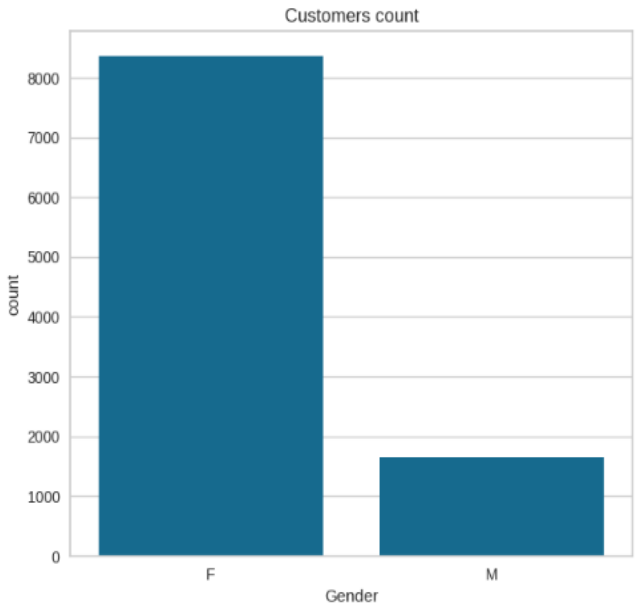
- **Python 3.8**: The programming language used for all data manipulation and analysis.
- **NumPy 1.20.3**: For numerical computations and handling arrays.
- **Pandas 1.2.4**: For data manipulation and preprocessing.
- **Scikit-Learn 0.24.2**: For implementing machine learning algorithms, including K-Means Clustering.
- **Seaborn 0.11.1**: For advanced data visualization and creating heatmaps.
- **Matplotlib 3.4.2**: Used alongside Seaborn for creating detailed plots and visualizations.
- **Google Colab**: Used for running the notebook, leveraging its cloud-based CPU runtime for computations.
- **Operating System**: The project was developed and tested on Windows 10 and Windows 11 environments.
- **Elbow Method and Silhouette Method**: Specific methods from Scikit-Learn used for evaluating the optimal number of clusters.
- **MinMaxScaler**: A feature in Scikit-Learn used for data normalization.

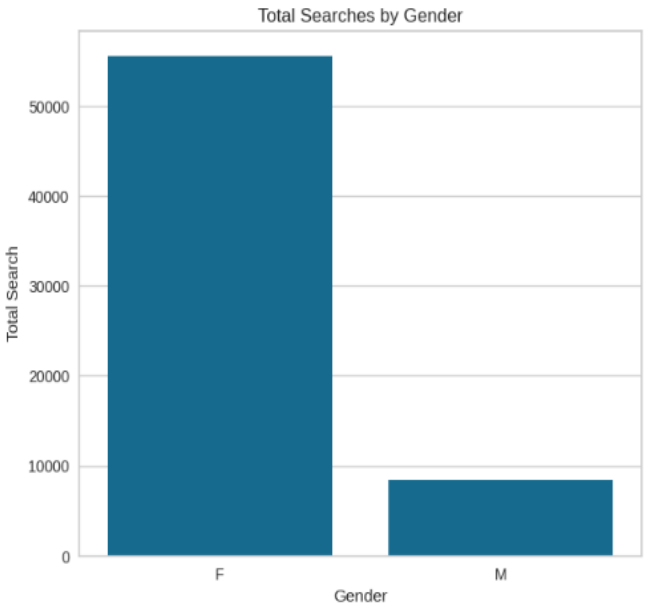
Implementation Screenshot Of Customer Segmentation



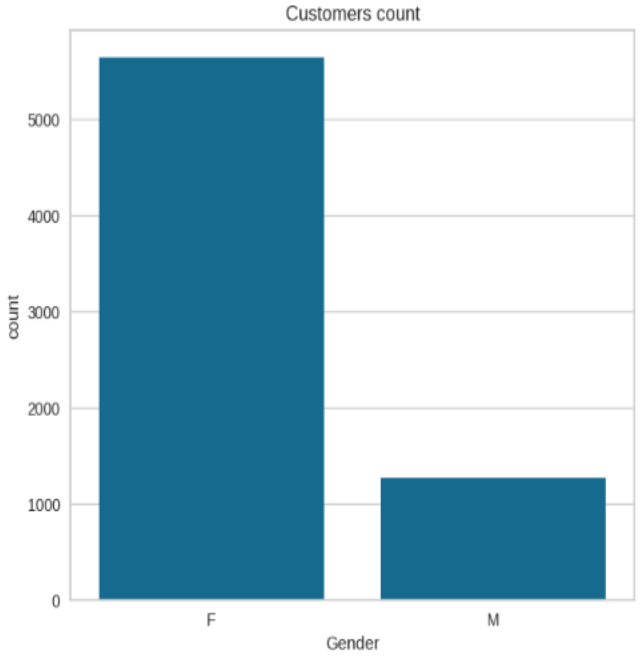


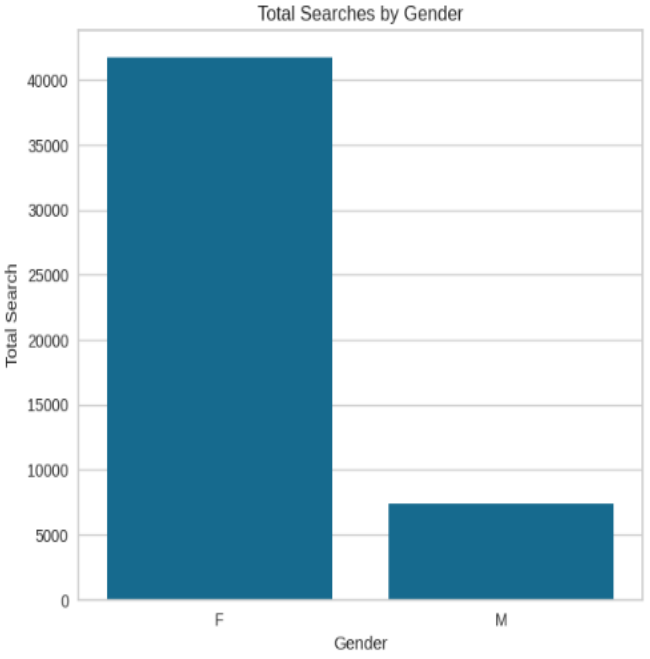
No. of customer and their total searches in "Cluster 1"

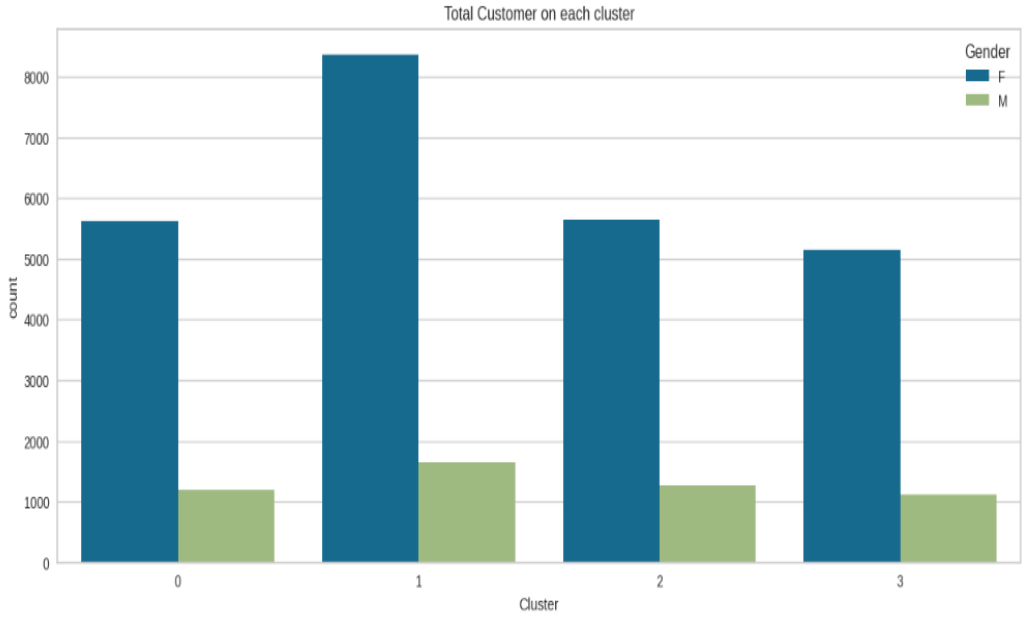


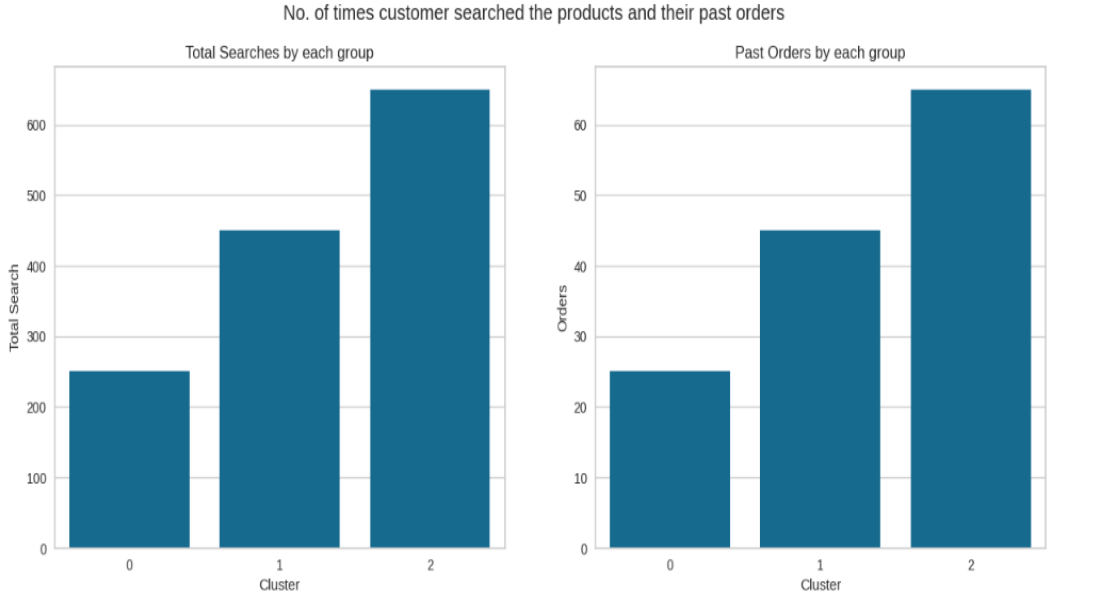


No. of customer and their total searches in "Cluster 2"






Total Customer on each cluster
Gender
F
M
count
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
Cluster


No. of times customer searched the products and their past orders
Total Searches by each group
Total Search
600
500
400
300
200
100
0
0
1
2
Cluster
Past Orders by each group
Orders
60
50
40
30
20
10
0
0
1
2
Cluster

Conclusion and Future Scope

The customer segmentation project at Zidio Development was a comprehensive exploration into the world of data-driven marketing strategies. Utilizing Python's powerful libraries—NumPy, Pandas, Scikit-Learn, and Seaborn—we successfully navigated through complex datasets to uncover valuable insights that drive informed business decisions and enhance customer experiences.

By meticulously preparing the data, employing advanced visualization techniques, and implementing robust clustering algorithms, we identified distinct customer segments that reflect varied behaviors and preferences. The application of the Elbow Method and Silhouette Method ensured the optimal number of clusters, providing meaningful and actionable segments.

The insights derived from this project enabled the development of targeted marketing campaigns and personalized customer engagement strategies. These insights underscore the pivotal role of data-driven segmentation in optimizing business outcomes and fostering deeper customer relationships.

Key takeaways from this project include the significance of robust data preprocessing, the power of advanced visualization in uncovering critical patterns, and the practical application of machine learning algorithms to derive actionable insights. The project also highlighted the importance of using the right tools and methodologies to address specific business challenges.

Looking forward, the skills and knowledge gained from this project will be instrumental in exploring new avenues in data science, continuing to harness its transformative potential in delivering strategic solutions. The experience at Zidio Development has not only enhanced my technical abilities but also provided a deeper understanding of the strategic importance of customer segmentation in driving business excellence and customer-centric innovation.

Reference

Github Link - https://github.com/Darshanj229/Customer-Segmentation

THANK YOU