

ONLINE RETAIL

Machine Learning (Unsupervised)

By: Team 4

Our Team



Brian Maxwell Kateren



Muhamad Ilyas Haikal



Rahmat Faisal M



Riska Amylatul Askiyah

AGENDA

01

THE OBJECTIV

02

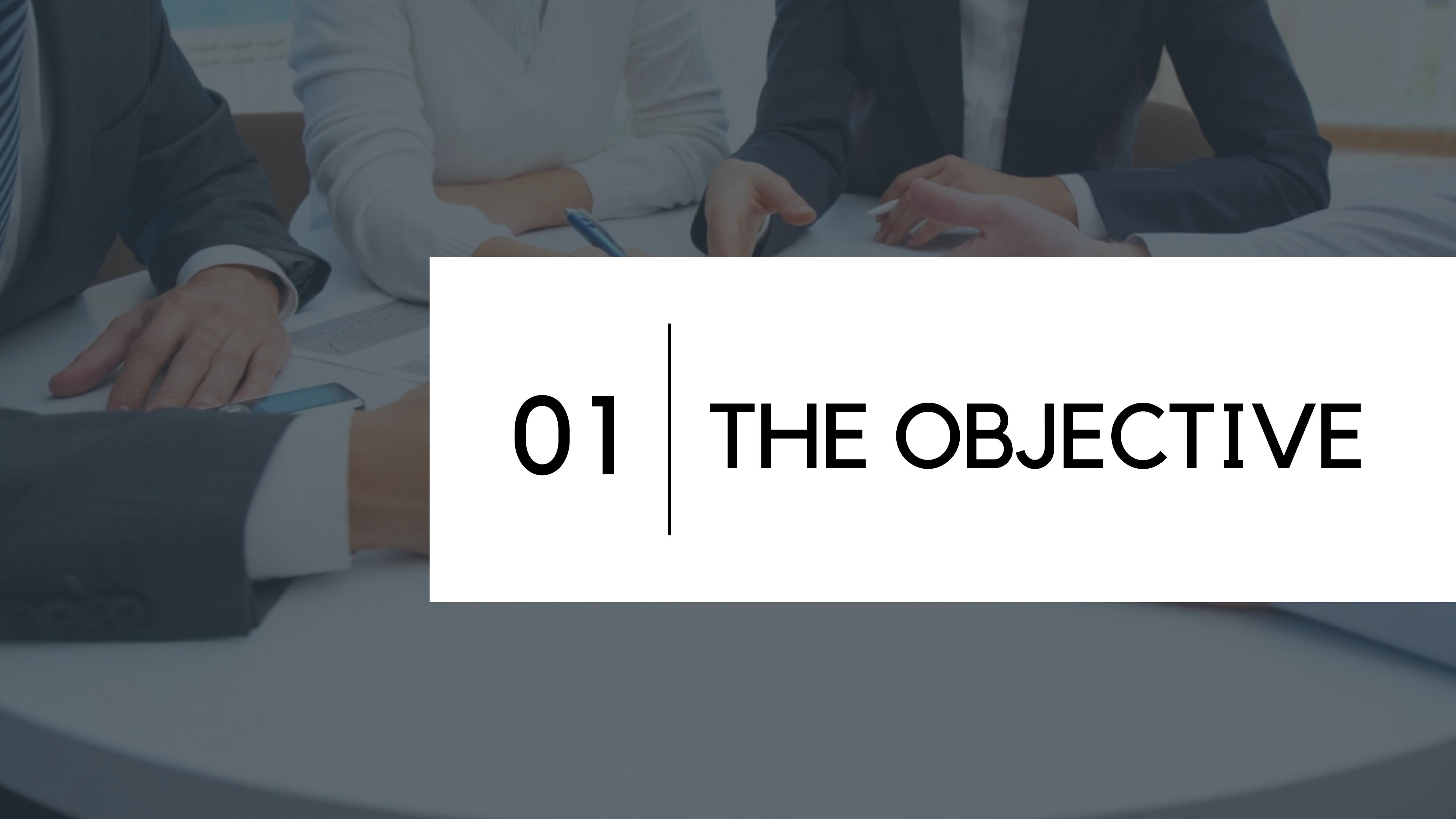
DATA PREPARATION

03

ANALYSIS & FINDING

04

SUMMARY & CONCLUTION

A photograph showing several people in a professional setting, likely a business meeting. They are seated around a table covered with papers, looking down at them. One person's hands are visible, holding a pen and pointing at a document. The scene is well-lit, suggesting a bright office environment.

01 | THE OBJECTIVE

BACKGROUND

E-commerce has become a new channel to support businesses development. Through e-commerce, businesses can get access and establish a wider market presence by providing cheaper and more efficient distribution channels for their products or services. E-commerce has also changed the way people shop and consume products and services. Many people are turning to their computers or smart devices to order goods, which can easily be delivered to their homes.

GOAL

- How was the sales trend over the months?
- What are the most frequent purchased products?
- How many products does the customer purchase in each transaction?
- How does each cluster/group contribute to the company's revenue?
- What are the interval date between transactions?
- Based on your findings, what strategy could you recommend to the business to gain more profit?

Challenges & Methodology

CHALLENGES

- Huge amount of data
- There is a missing value
- Ambiguous meaning of columns of features

Methodology

- Data Preparation
- Data Preprocessing
- RFM Model Building
- Clustering
- Predictive Model Building
- Exploration Data Analysis

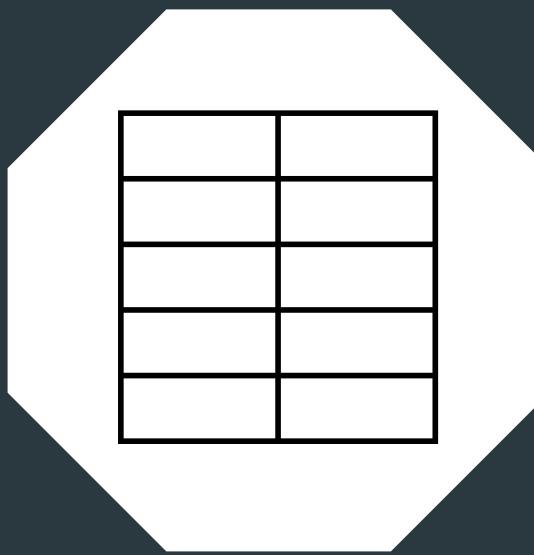


02 | DATA PREPARATION

Initial Data Quality Report



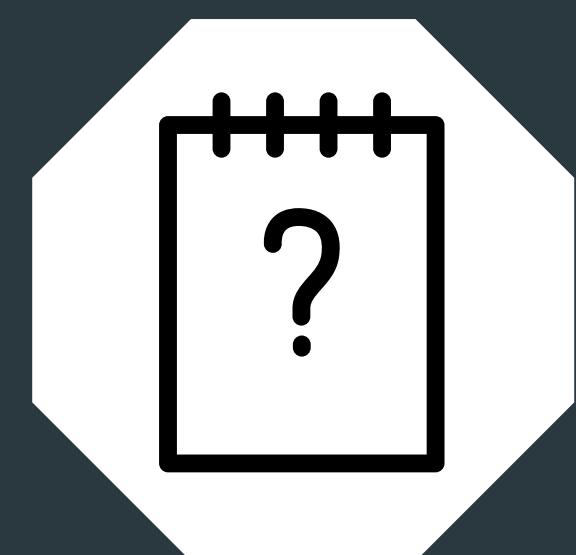
Total of 541,909
Data



Total of 8
Column



2 Columns with
Missing Value



136,534 Missing
Value

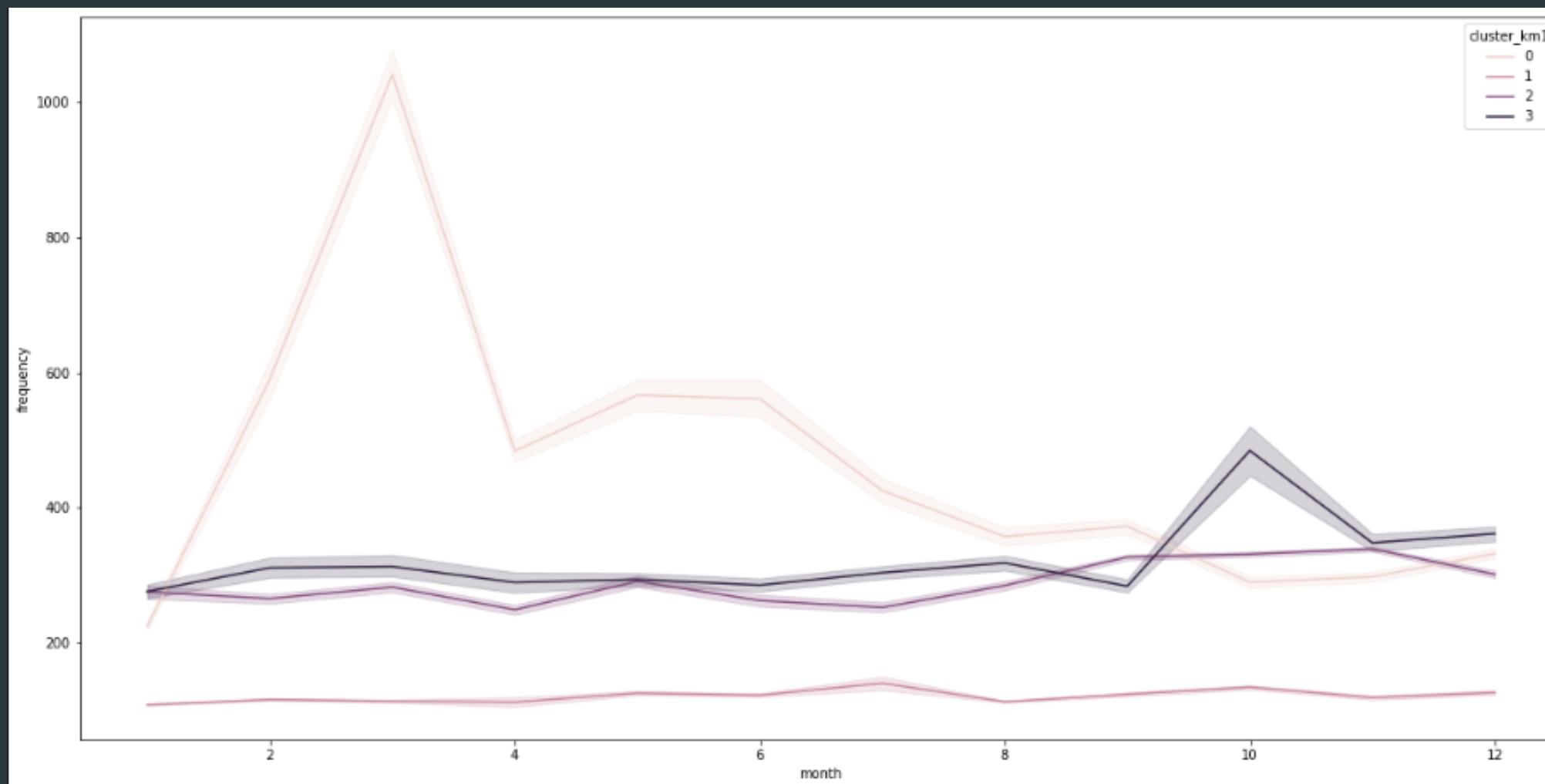
Preparation Step

- 1 ————— 2 ————— 3 —————
- Data Quality Check
- Preprocessing
- K-Means
- 4 ————— 5 ————— 6 —————
- Logistic Regression
- Optimization and Validation
- Imbalanced Data



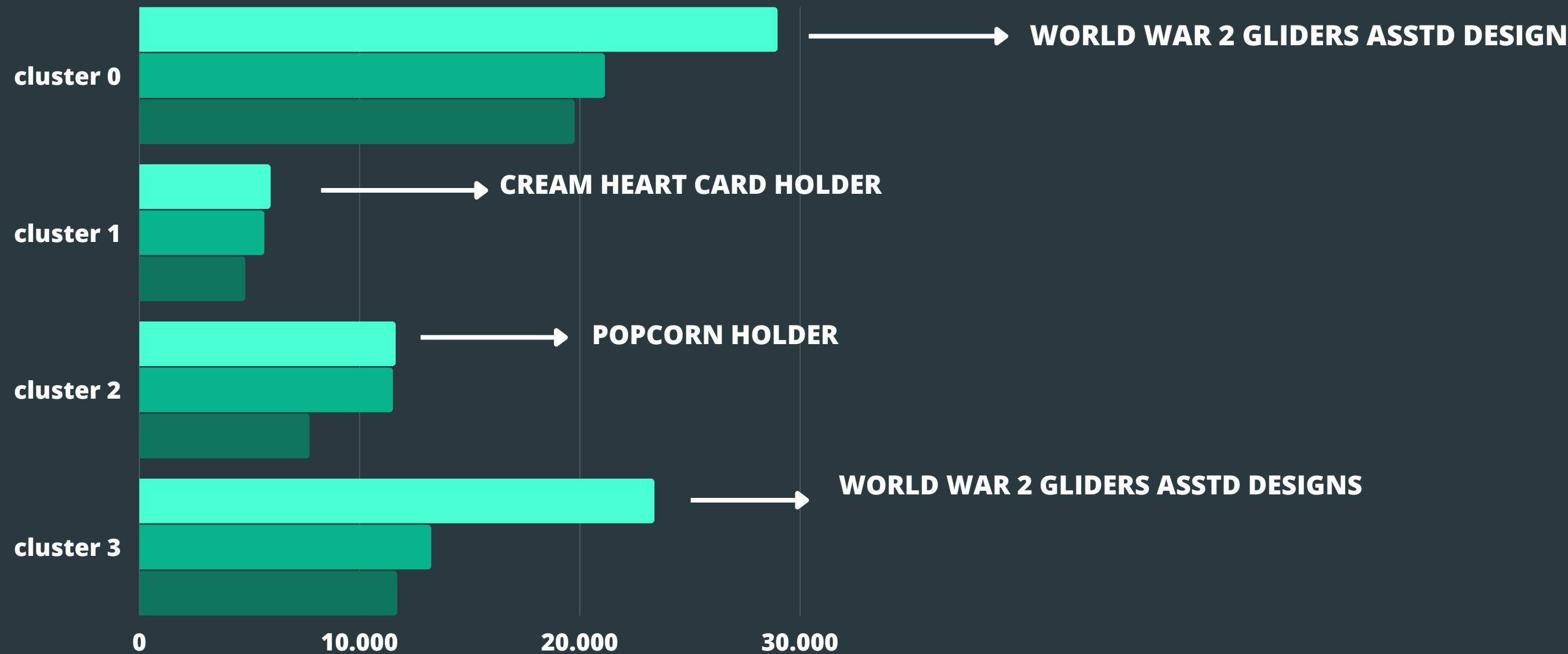
03 | ANALYSIS & FINDING

Sales Over the Months



the trend of sales frequency in cluster 0 can be seen as relatively stable and tends to increase every month even though there was a decline at the end of the year, in each cluster, there was a decrease in sales at the end of the year

Most Frequent Purchased Product



products does the customer purchase in each transaction



Then it can be concluded that:

For cluster 0

- The number of products that were purchased the most in one transaction was 277 purchased

For cluster 2

- The number of products that were purchased the most in one transaction was 542 purchased

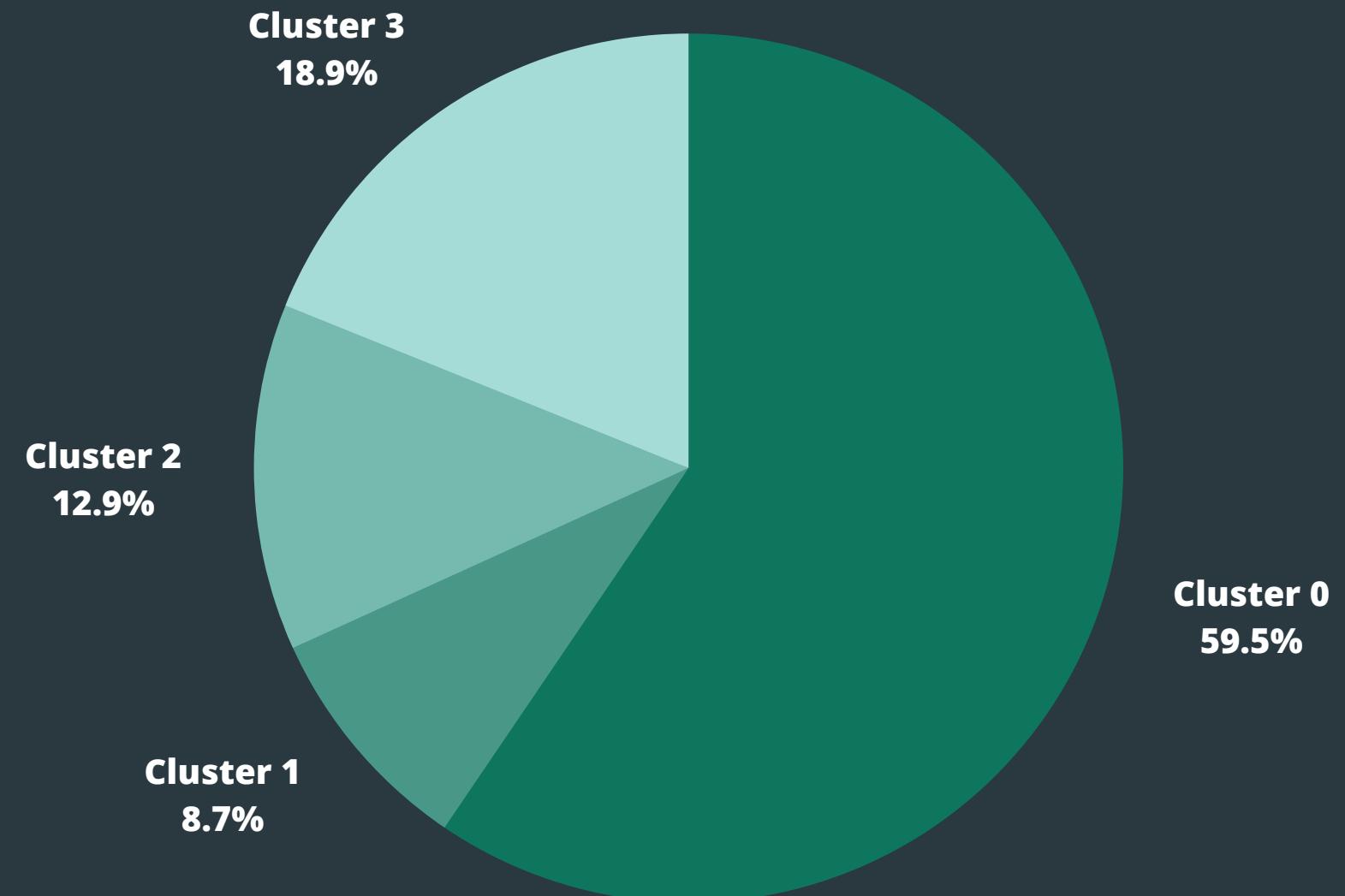
For cluster 3

- The number of products that were purchased the most in one transaction was 399 purchased

For cluster 1

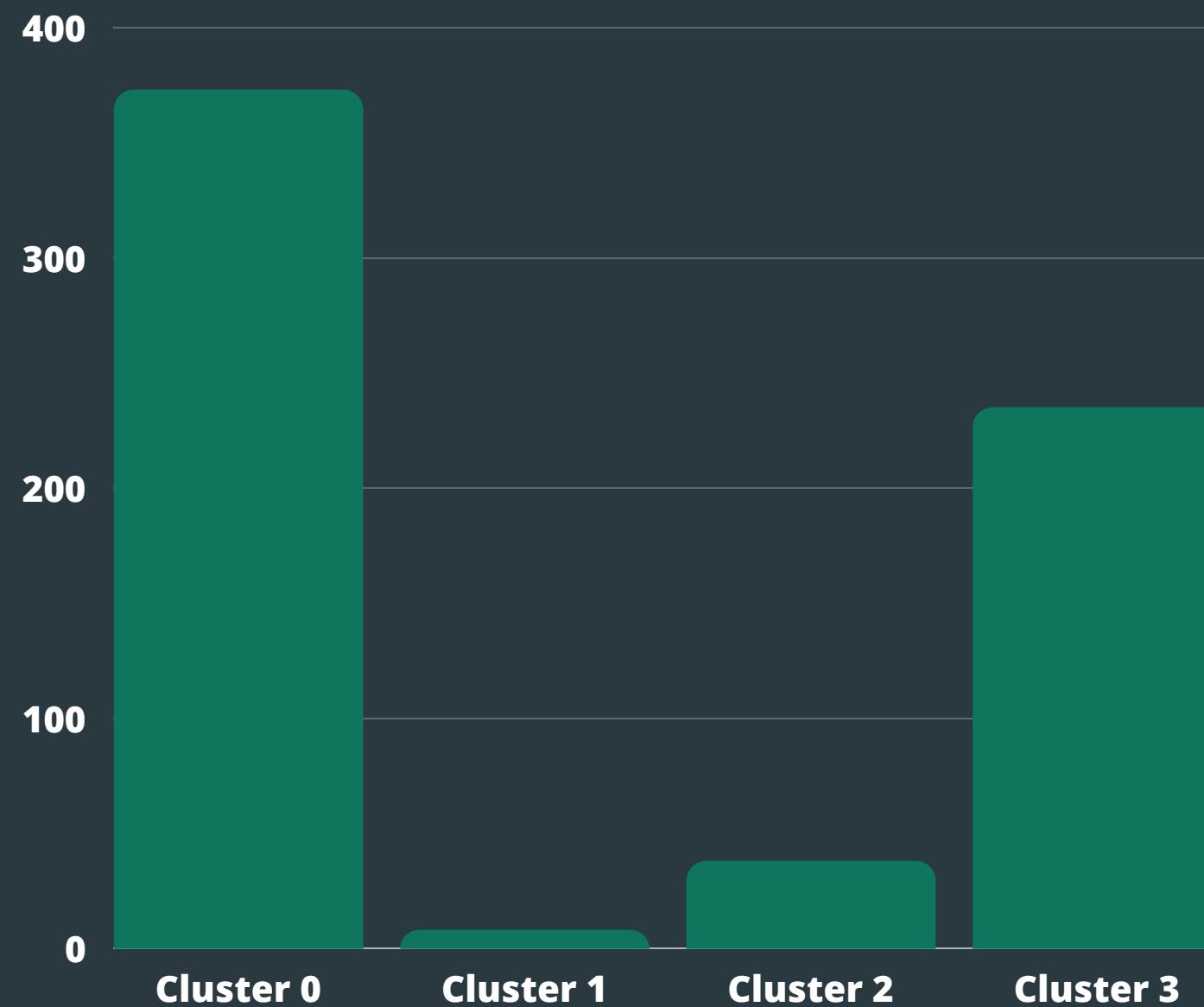
- The number of products that were purchased the most in one transaction was 125 purchased

clusters that contribute to the company's revenue

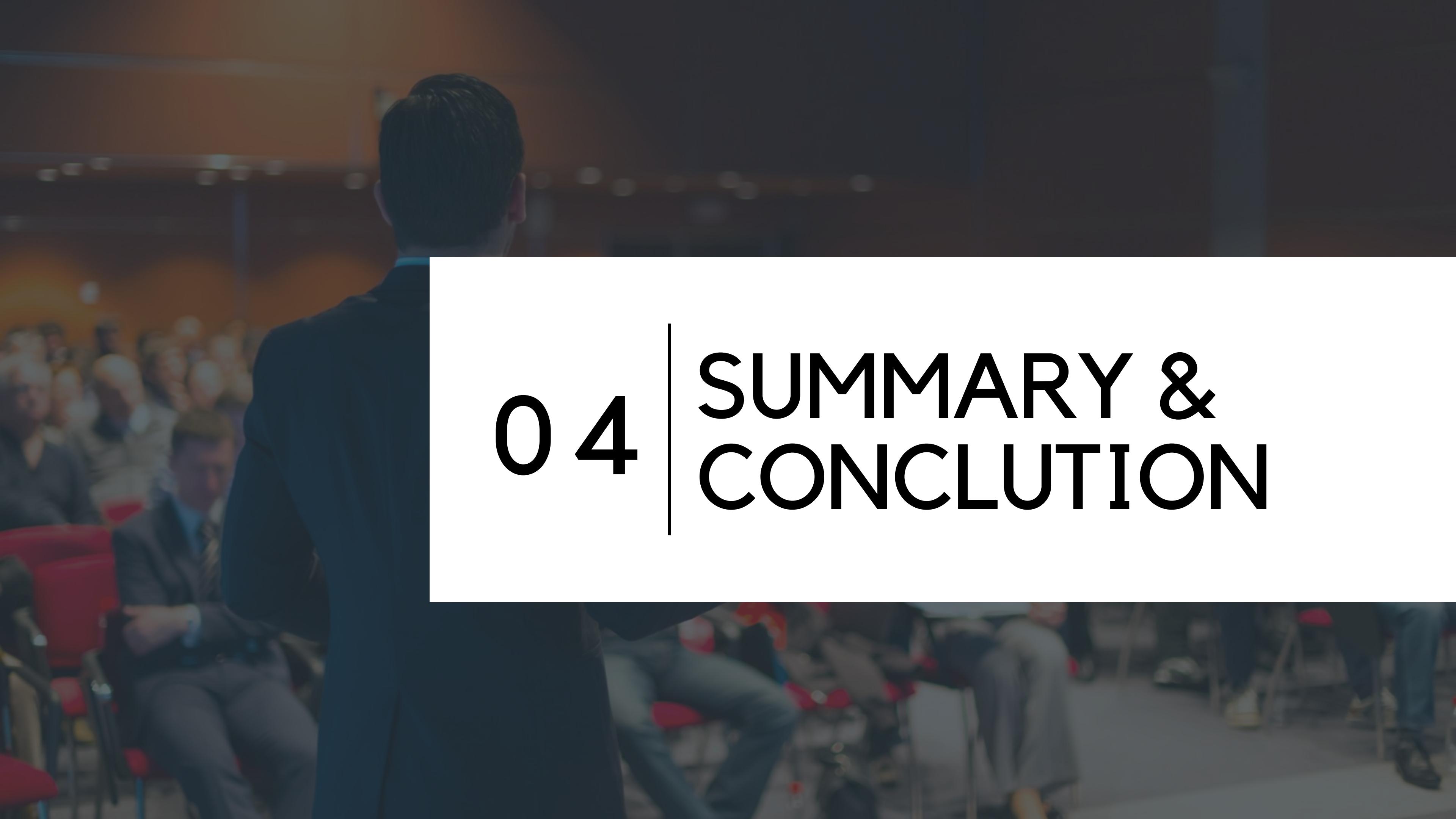


According to the graph above, cluster 0 contributes a sizable amount of revenue to the company, amounting to about 59,5%, followed by cluster 3 at about 18,9%, cluster 2 at about 12,9%, and cluster 1 at about 8,72%.

Interval Date Between Transaction



from the results of these calculations, we can conclude that cluster 0 has the largest interval value and cluster 1 has the smallest interval value

A photograph of a man in a dark suit and tie standing at a podium, facing an audience. He is gesturing with his right hand. The audience is visible in the foreground, seated in rows. The background is a blurred indoor setting.

04

SUMMARY & CONCLUSION

what strategy could you recommend to the business to gain more profit?

Cluster 0

We could offer a discount on the top 3 most popular items, such as the T-Heart Holder, Regency cakestand 3 tier, and big bag retrospot, for cluster 0 customers who have a high cancellation rate and make up the majority of the population.

Cluster 2

The top 3 things for cluster 2 (the high-spending and a few of the best customers) that we could discount are the regency cakestand with 3 tiers, the gigantic bag in red retrospot, and the carriage.

Cluster 1

For cluster 1 (The Best Customers) with high quantity purchases, high total costs, and high recency, we could create bundle promotions to motivate them and loyalty points for their regular purchases and large spends; the bundling/discount includes the top 3 products, which are the spaceboy lunch box, round snack box set of 4 boxes, and postage.

Cluster 3

Regency Cakestand 3 Tier, Jumbo Bag Red Retrospot, and Lunch Bag Red Retrospot are the top 3 products we could discount for Cluster 3 (The Big Spenders and Some of the Best Customers).

CONCLUSION

372

Cluster 0 has the largest date interval value

6634

Cluster 0 is the cluster that cancels the most

236

Based on the average recency per segment, the lost cheap customer segment is the segment with the highest average recency

Top 3

The most purchased product are White Hanging Heart T-Light Holder, Recency 3 Tier, Jumbo Bag Retrosport

A group of people are gathered around a table in a room with large windows. In the foreground, two women are looking towards the right. In the background, a man is smiling. A white rectangular overlay contains the text.

THANK YOU

Mentor : Cahya Putera