DSW | Powered by **Telkomsel**
DATA SCIENCE WEEKEND

LEVERAGING CUSTOMER SEGMENTATION TO ELEVATE STRATEGIC APPROACH IN PREVENTING CHURN

Yang Ini

Our Team

Cornelius Justin
Satryo Hadi

Mathematics
Universitas Indonesia

Cindy Fara
Nabila
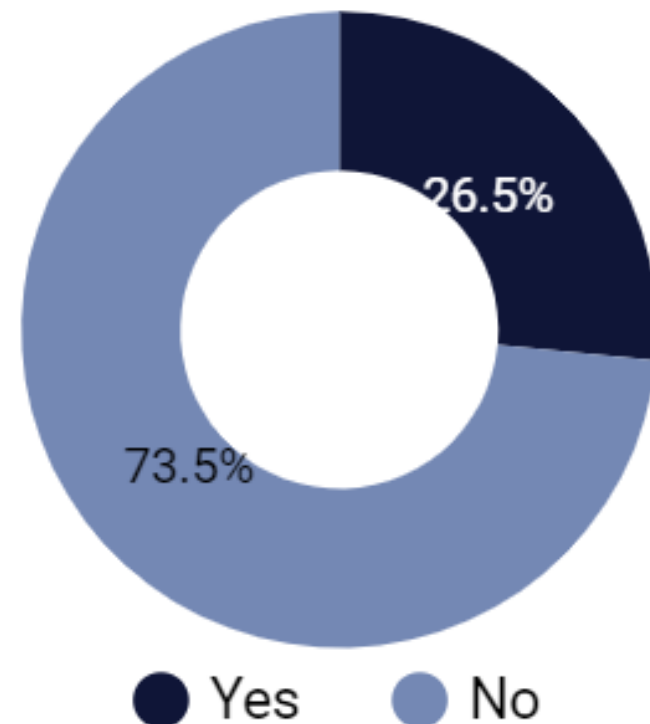
Statistics
Universitas Indonesia

Safira Raissa
Rahmi

Mathematics
Universitas Indonesia

Churn is experienced by 26.5% of telco customer base

BUSINESS CHALLENGE

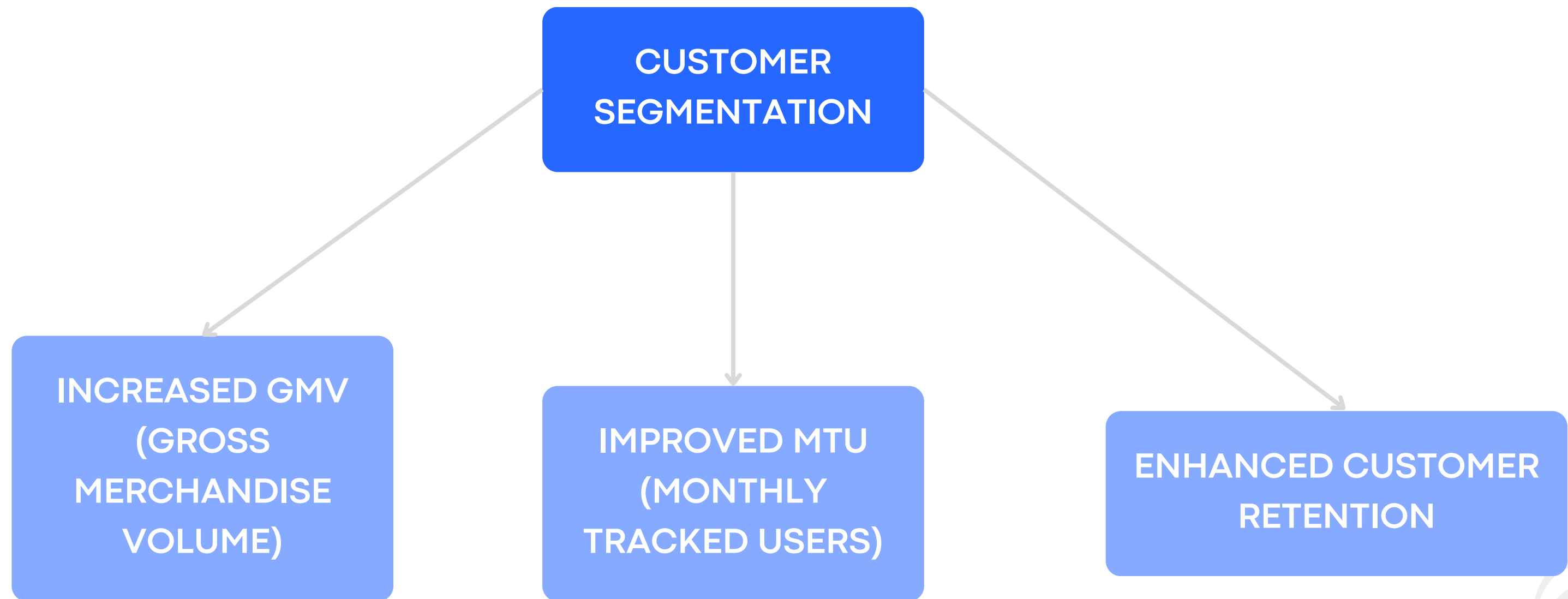The churn rate is still relatively high.

26.5%

73.5%

● Yes ● No

APPROACH

Analyzing customer segments to identify the appropriate strategy for preventing customer churn high.

PROPOSED SOLUTIONS

Segmenting the non-churned customers and executing campaigns based on the analysis results
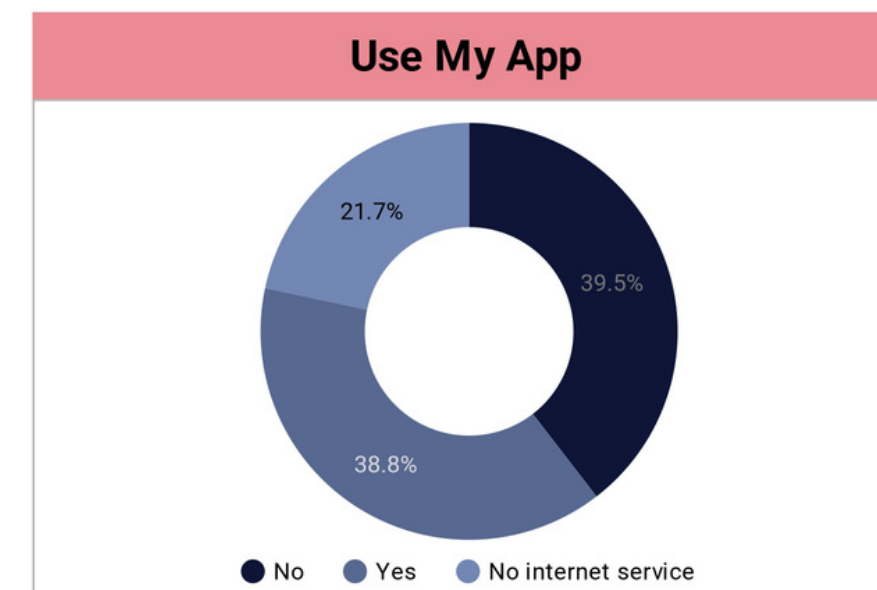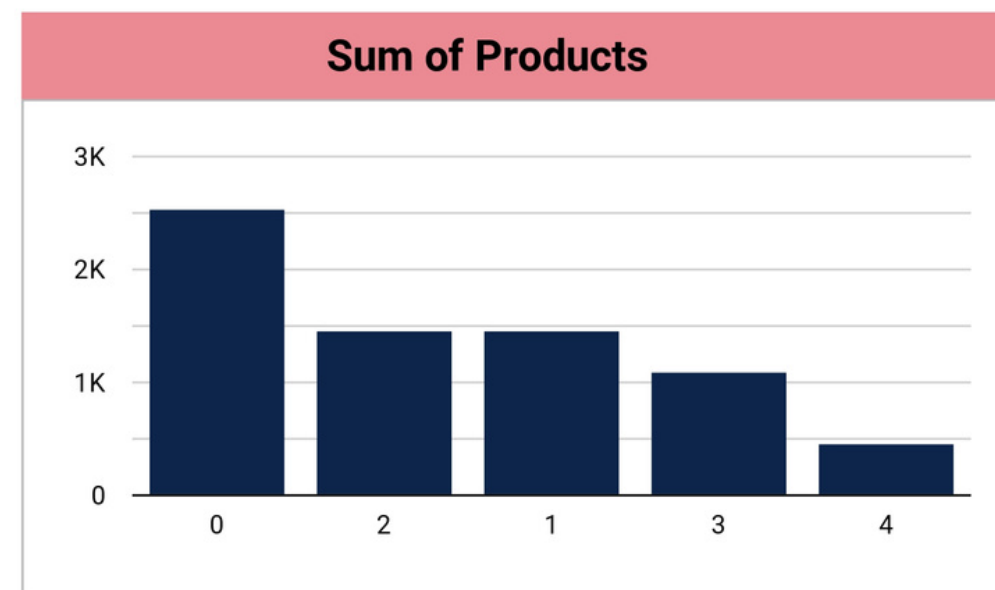
Expected Impact

CUSTOMER SEGMENTATION

INCREASED GMV (GROSS MERCHANDISE VOLUME)

IMPROVED MTU (MONTHLY TRACKED USERS)

ENHANCED CUSTOMER RETENTION
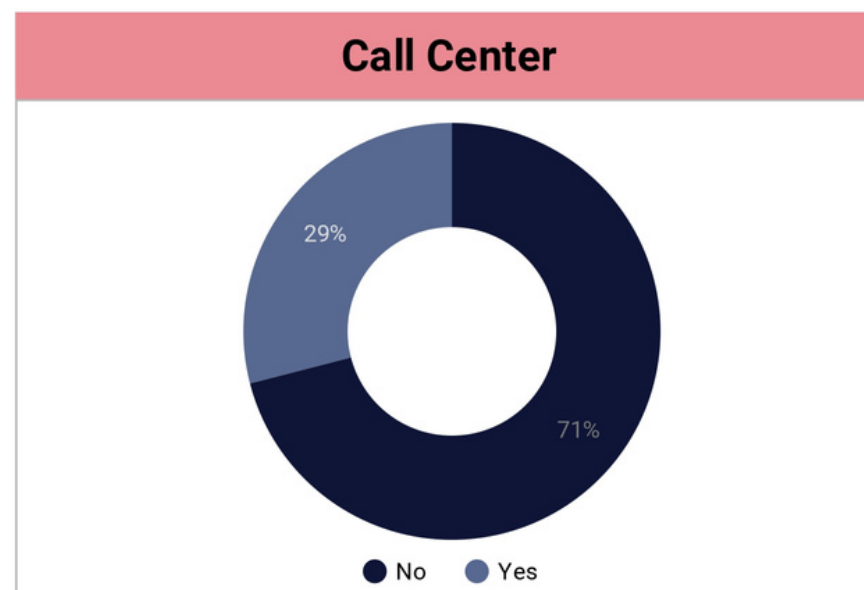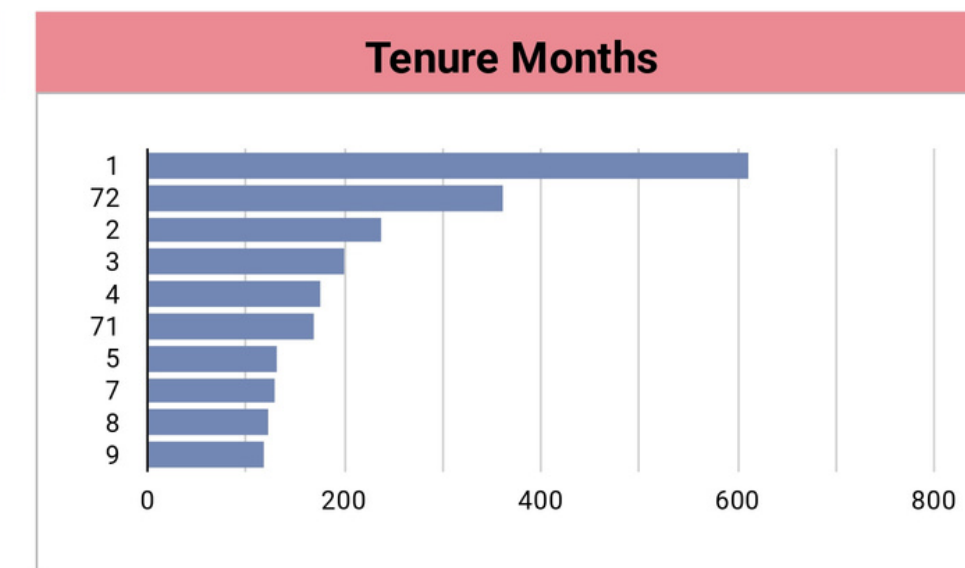
DATA AND METHODOLOGY

Current Condition

Telco Customer Churn
Yang Ini

Location
Record Count 2,012 ▢ 5,031

Churn Label
26.5%
73.5%
● Yes ● No

Tenure Months

Call Center
29%
71%
● No ● Yes

Sum of Products

Use My App
21.7%
39.5%
38.8%
● No ● Yes ● No internet service
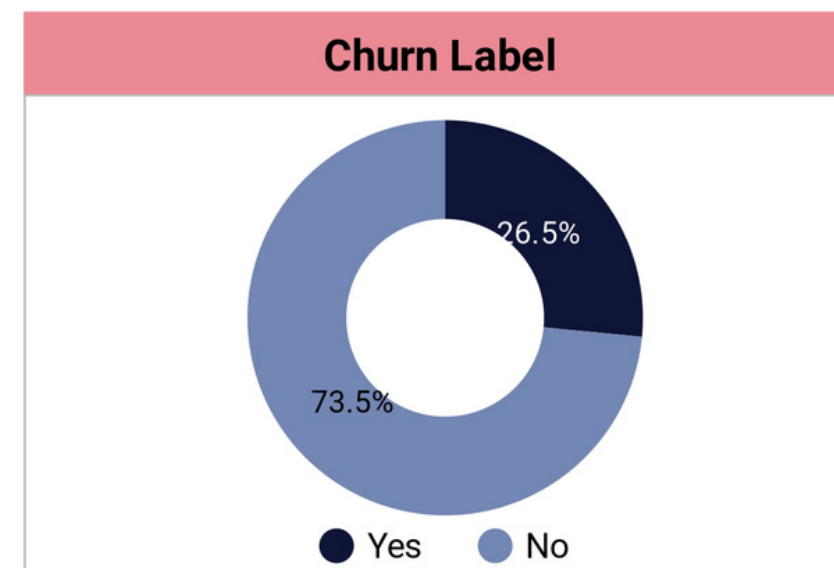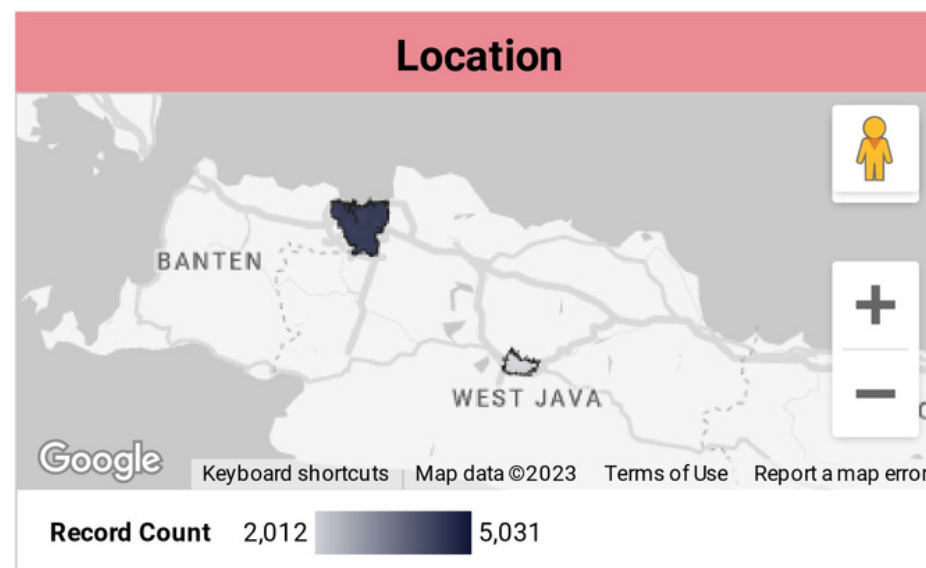
For an interactive dashboard

RFM APPROACH

We use the RFM approach to segment our customers. Based on that, we get 3 features for our modelling and analysis.

Features

R	Recency	Tenure Months
F	Frequency	Calculate product usage (Games Product, Music Product, Education Product, Video Product)
M	Monetary	Monthly Purchase
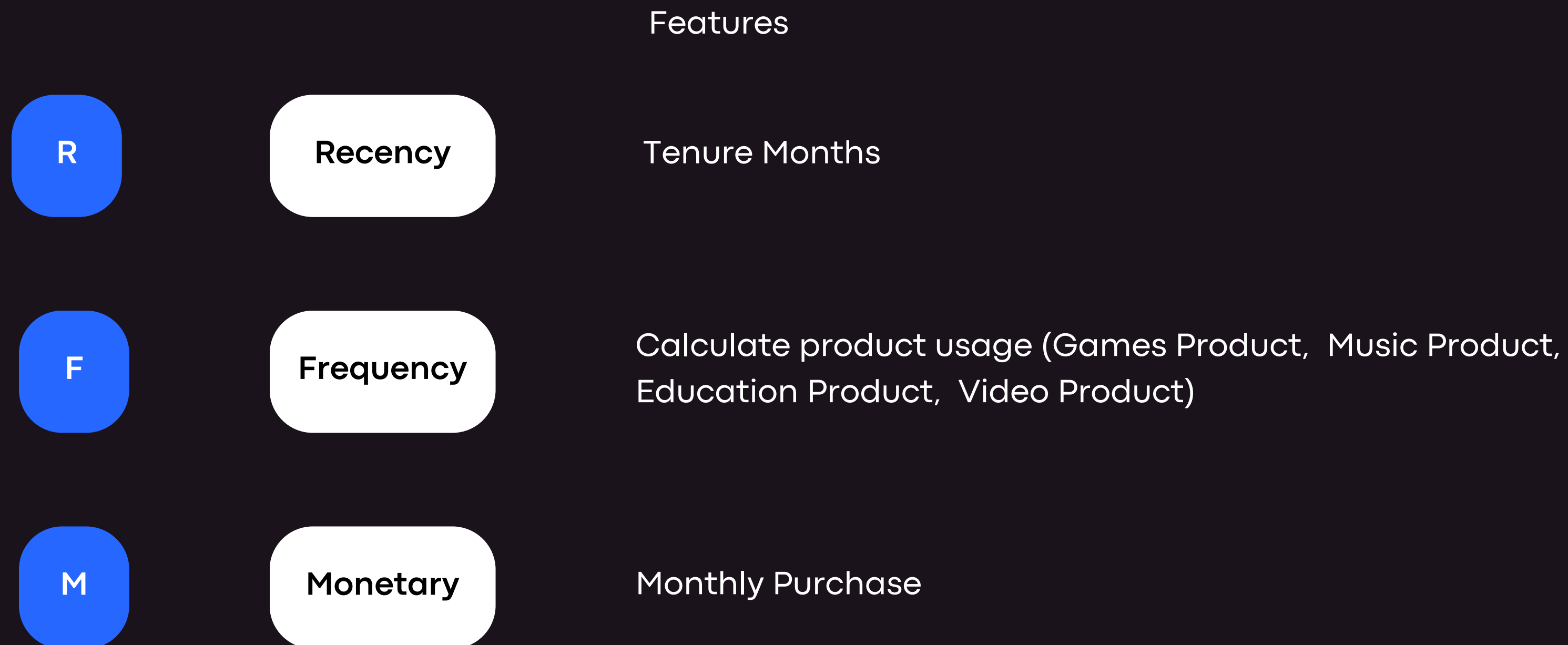
METHODOLOGY

MODEL

- Unsupervised model approach, clustering model.
- Using hard clustering, K-Means model.
- Drawbacks K-Means:
  - Computationally efficient and straightforward to implement.
  - Scales well with an increasing number of data points.
  - Suitable for both numerical and categorical data.
  - Clusters generated by K-Means are often easy to interpret.

EVALUATION

- Using Silhouette Score.

K-MEANS
GENERAL PROCEDURES

K-means solve the following minimization problems

$$\arg min_c \sum_{j=1}^{k} \sum_{x \in c_j} d(x, \mu_j) = \arg min_c \sum_{j=1}^{k} \sum_{x \in c_j} ||x - \mu_j||_2^2$$

1. Choose the number of clusters (k) and randomly initialize cluster centroids.
2. Assign each data point to the nearest centroid, forming k clusters.
3. Recalculate the centroids as the mean of data points within each cluster.
4. Repeat the assignment and centroid update steps until convergence or a specified number of iterations.
5. Define stopping criteria, such as minimal centroid movement or a maximum number of iterations, to determine when the algorithm converges.

FLOWCHART METHODOLOGY

Start

Getting Data → Data *cleaning & preprocessing* → Hyperparameter Tuning (number of centroids) → K–Means Models for Customers Segmentation

Stop
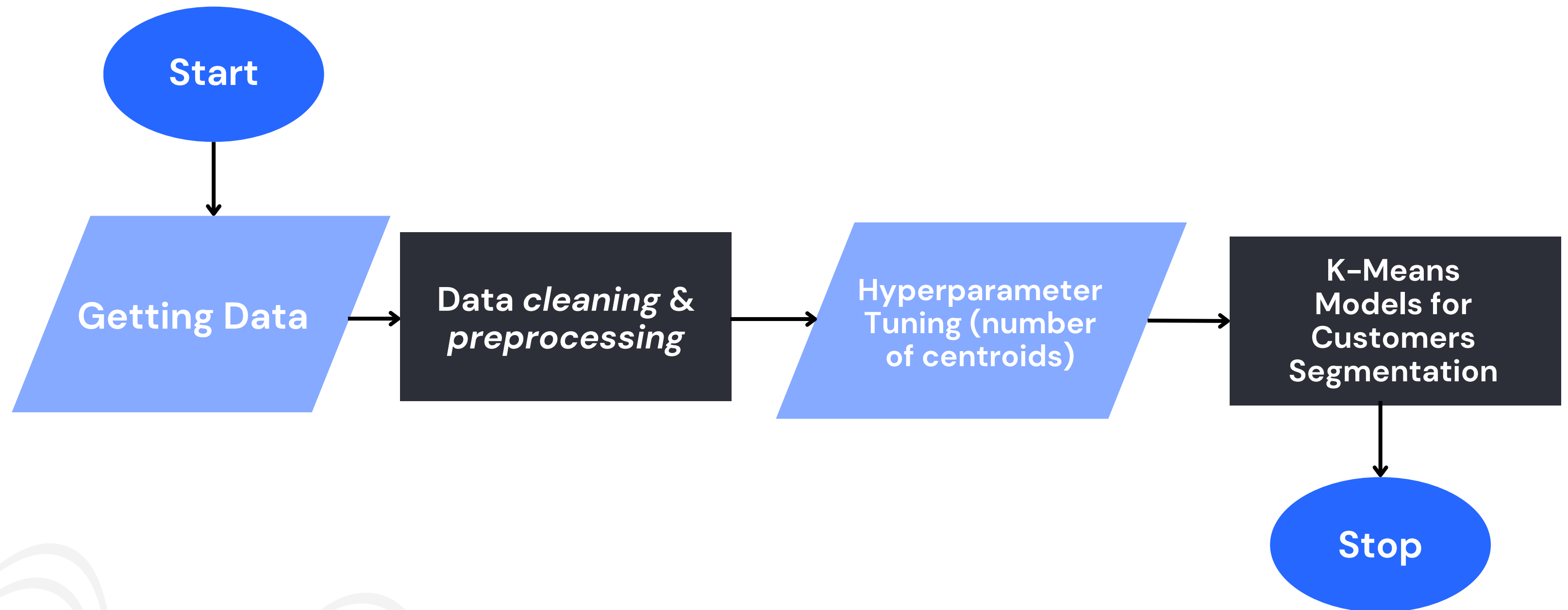
For python (google colaboratory)notebook

EVALUATION
SILHOUETTE SCORES

Values meaning for silhouette scoring:

- 0.71 - 1.0: A strong structure has been found.
- 0.51 - 0.70: A reasonable structure has been found.
- 0.26 - 0.50: The structure is weak and could be artificial. Consider trying different clustering methods or features.
- < 0.25: No substantial structure has been found.

Evaluating clustering models using formulas:

$$s(i) = \frac{b(i) - a(i)}{\text{larger of } b(i) \text{ and } a(i)}$$

a(i) = average distance inside cluster
b(i) = average distance nearest other cluster

Our model get **silhouette score of 0.6366122757251996** for **2 clusters** created, it falls into the "reasonable structure has been found" category.
It suggests that the **clusters are well-defined** and the data points within each cluster are more similar to each other than to points in other clusters.

RESULT AND ANALYSIS

Analyzing two created clusters for non-churn customers.

46.3%

53.7%

1

0
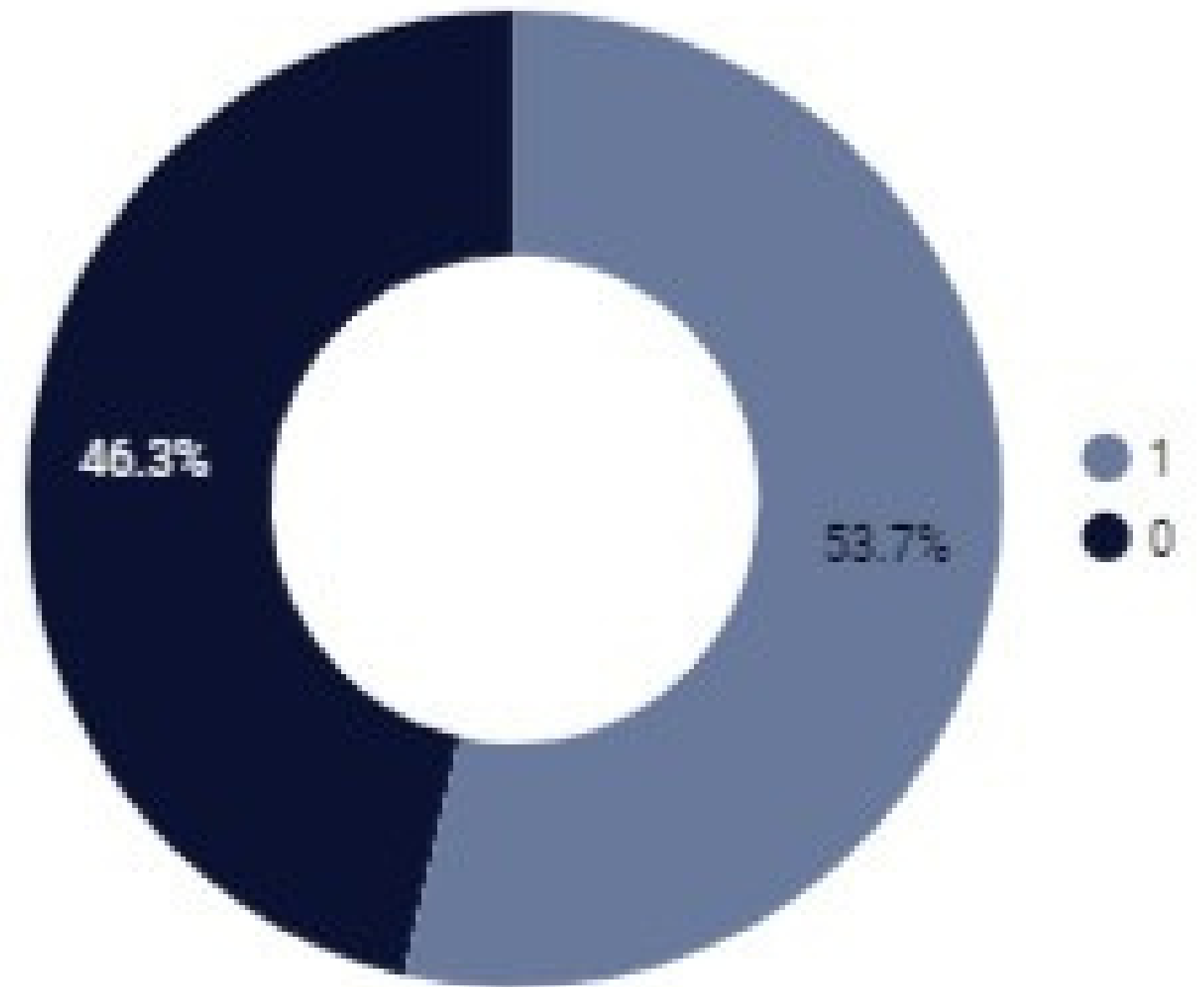
RESULT

CREEATED CLUSTER

- Cluster 0 : Covered 2398 (46,3%) of Total Non-Churn Customers
- Cluster 1 : Covered 2776 (53,7%) of Total Non-Churn Customers

EXPECTED INSIGHT

- Learn differences between two clusters of Non-Churn Customers
- Determine scope priorities treatment for each cluster of Non-Churn Customers
- Determine scope priorities treatment for both clusters of Non-Churn Customers

Analyzing Clusters Differences

MEAN TABLE

Cluster_Labels	Count Used Product	Tenure Months	Monthly Purchase (Thou. IDR)
0	3.868224	52.052544	112.693736
1	0.638329	25.059438	51.095784

STANDARD DEVIATION TABLE

Cluster_Labels	Count Used Product	Tenure Months	Monthly Purchase (Thou. IDR)
0	1.250175	18.135601	23.334988
1	0.949325	21.494966	28.567924

- Cluster 0 exhibits higher mean values for all RFM features compared to Cluster 1.
- Standard deviation analysis strengthens the analytics argument presented earlier.
- Cluster 0 has lower standard deviation in Tenure Months and Monthly Purchase, suggesting more consistency.
- Cluster 1 displays lower standard deviation in Count Used Product, implying greater uniformity in this feature.
- The conclusion is **cluster 0 is higher value costumers than cluster 1**.

Characteristic Each Clusters Non-Churn Customers

COUNT USED ALL PRODUCT TABLE

Cluster_Labels	mean_Count Used Product	std_Count Used Product
0	3.868224	1.250175
1	0.638329	0.949325

MEAN COUNT USED EACH PRODUCT TABLE

Cluster_Labels	Games Product	Music Product	Education Product	Call Center	Video Product	Use MyApp
0	0.560467	0.654295	0.672644	0.588824	0.692661	0.699333
1	0.136888	0.121398	0.095101	0.115994	0.083573	0.085375

STANDARD DEVIATION COUNT USED EACH PRODUCT TABLE

Cluster_Labels	Games Product	Music Product	Education Product	Call Center	Video Product	Use MyApp
0	0.496434	0.475697	0.469346	0.492150	0.461488	0.458644
1	0.343791	0.326648	0.293407	0.320276	0.276797	0.279489

- Cluster 0 has higher mean and standard deviation in total product count.
- **Top 2 products** for Cluster 0: Use My App and Video Product.
- **Top 2 products** for Cluster 1: Games Product and Music Product.
- Customers in cluster 0 is more non-uniform in product spending based on standard deviation.

Characteristic Each Clusters Non-Churn Customers

MEAN AND STANDARD DEVIATION TENURE MONTHS TABLE

Cluster_Labels	mean_Tenure Months	std_Tenure Months
0	52.052544	18.135601
1	25.059438	21.494966

- Cluster 0 has higher mean in Tenure Months.
- Customers in cluster 0 is more uniform in Tenure Months spending based on standard deviation.
- Customers in cluster 0 spend more time than cluster 1 for using products.

Characteristic Each Clusters Non-Churn Customers

<u>MEAN AND STANDARD DEVIATION
MONTHLY PURCHASE (THOU. IDR) TABLE</u>

Cluster_Labels	mean_Monthly Purchase (Thou. IDR)	std_Monthly Purchase (Thou. IDR)
0	112.693736	23.334988
1	51.095784	28.567924

- Cluster 0 has higher mean in Monthly Purchase.
- Customers in cluster 0 is more uniform in Monthly Purchase spending based on standard deviation.
- Customers in cluster 0 spend more purchasing than cluster 1 for using products.
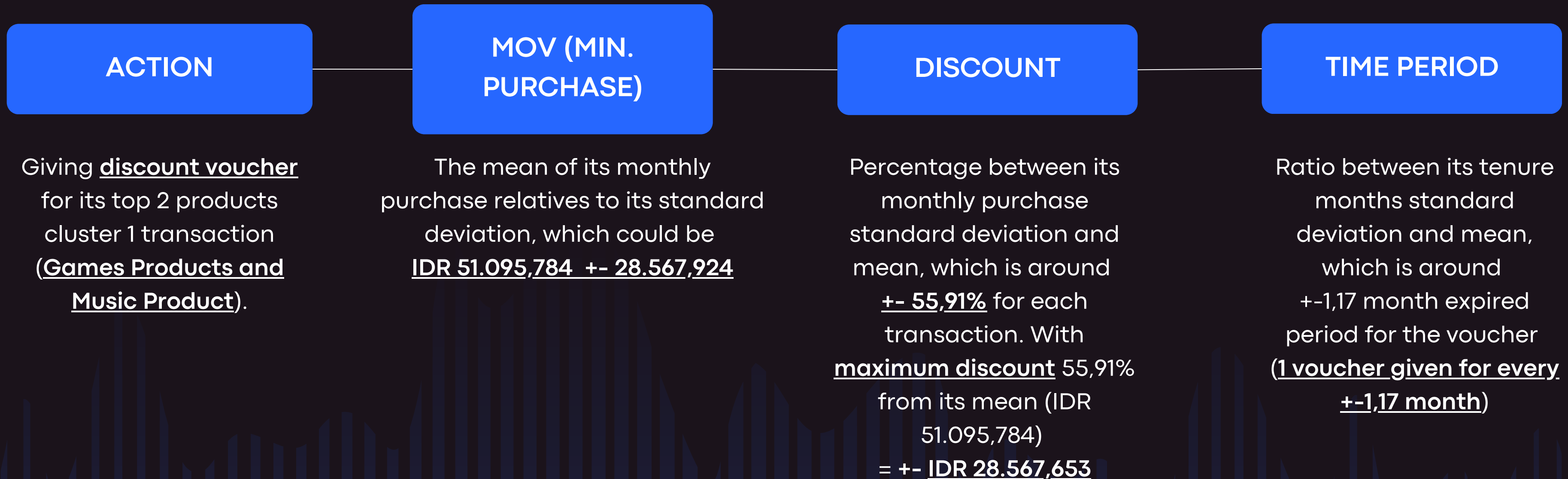
Priorities Treatment Action
Each Cluster

FOR CLUSTER 0

ACTION

MOV (MIN. PURCHASE)

DISCOUNT

TIME PERIOD

Giving **discount voucher** for its top 2 products cluster 0 transaction (**Use My App and Video Product**).

The mean of its monthly purchase relatives to its standard deviation, which could be **IDR 112.693,736 +- 23.334,988**

Percentage between its monthly purchase standard deviation and mean, which is around +- **20,71%** for each transaction. With **maximum discount** 20,71% from its mean (IDR 112.693,736) = +- **IDR 23.338,873**

Ratio between its tenure months standard deviation and mean, which is around +- 2,87 month expired period for the voucher (**1 voucher given for every +-2,87 month**)

Priorities Treatment Action
Each Cluster

FOR CLUSTER 1

ACTION	MOV (MIN. PURCHASE)	DISCOUNT	TIME PERIOD
Giving **discount voucher** for its top 2 products cluster 1 transaction (**Games Products and Music Product**).	The mean of its monthly purchase relatives to its standard deviation, which could be **IDR 51.095,784 +- 28.567,924**	Percentage between its monthly purchase standard deviation and mean, which is around **+- 55,91%** for each transaction. With **maximum discount** 55,91% from its mean (IDR 51.095,784) = +- **IDR 28.567,653**	Ratio between its tenure months standard deviation and mean, which is around +-1,17 month expired period for the voucher (**1 voucher given for every +-1,17 month**)

PRIORITIES TREATMENT ACTION FOR BOTH CLUSTERS

ACTION	MOV (MIN. PURCHASE)	COINS CASHBACK	TIME PERIOD
Giving **coins cashback** (that can be used for discount) to products that not in top 2 products for cluster 0 and cluster 1 (**Education Product** only, because coins **can't be applied in call center**).	Mean of cluster 0 and cluster 1's mean relative to mean of their standard deviations in monthly purchase, which **IDR 81.894,76 +- 25.951,456**	Percentage between mean of cluster 0 and cluster 1's standard deviation in monthly purchase and mean of theirs mean , which is around **+- 31,67%** for each transaction. With **maximum coins cashback** 31,67% from 81.894,76 = **+- IDR 25.936,070**	Ratio between mean of cluster 0 and cluster 1's standard deviation in tenure months and mean of theirs mean, which is around +-0,514 month expired period for the voucher coins cashback **(1 voucher given for every +-0,514 month)**

ADD ON FUTURE ACTION STEP

VOUCHER AND COIN CASHBACK

- Doing **A/B testing** with confident range (mean+-standard deviation) for each calculation parts in voucher and coin cashback schemes.
- **Improving** top products and **evaluate** top least quantity products.
- Gathering **partnership to balance** budget for voucher and coin cashback schemes.

OTHERS

- Improving call center elements.
- Getting more data to get more insightful analysis.
- Doing another model's approach to get new impactful analysis.

CONCLUSION

In utilizing the K-Means clustering model for customer segmentation, our approach revealed clusters that are computationally efficient, scalable, and interpretable. The silhouette score of 0.6366 indicates a well-defined structure, with Cluster 0 covering 46.3% and Cluster 1 covering 53.7% of non-churn customers. Which with analysis, we could conclude that cluster 0 is higher value costumers than cluster 1.

Distinguishing treatment priorities for each cluster, we propose tailored discount vouchers for top products and exclusive coins cashback for specific items. A/B testing, product optimization, budget balancing, and strategic partnerships are key to refining our approach. Future steps involve enhancing call center elements, acquiring more data for deeper insights, and exploring alternative models for impactful analyses.

Thank You

- Yang Ini -