



KALBE
Nutritionals



Optimization Strategy for Sales and Customer Loyalty at Kalbe Nutritionals: A Data-Driven Approach Using RFM, Cohort Analysis, and CLV

Hijir Della Wirasti

Final Project

Business Intelligence

Batch 13

Saturday, 08 March 2025



GitHub:
<https://github.com/hijirdella/Kalbe-Nutritionals-BI-Final-Project-Sales-RFM-Cohort-Churn-CLV-Optimization>

Outline

A

Introduction

Projects Overview

B

C. Main Project

1

Project Background

Business Problem

2

3

Data Analysis and Insights

**Dashboards and
Visualizations**

4

5

**Recommendations and
Actionable Insights**

Appendix

6

Education



Institut Teknologi Bandung
Bachelor's in Ocean Engineering
GPA: 3.21 (2011-2016)

Relevant coursework: Information of Technology B, Statistic and Probability Analysis, Numerical Analysis, Engineering Economics, Field Data Acquisition and Analysis, and Modelling in Coastal Engineering.



Universitas Pendidikan Indonesia
Bachelor's in Music Education
GPA: 3.57 (2013-2019)



Telkom University
Master's in Information Systems
GPA: 3.71 (2022-Present)

Relevant coursework: Data Governance and Information Technology, Business Analysis and Company Data, Integration and Architecture of Company Applications, and Information Assurance and Security.

Working



PT Dinamaritama Konsultan Rekayasa

Ocean Engineer Consultant
July 2016 – March 2022

Analyzed millions of wave, current, and geotechnical data points for designing docks, coastal protection structures, and other maritime infrastructure across 20 government projects.



Sony Music Entertainment Indonesia

Artist & Repertoire Executive
October 2022 – October 2023

Released 136 songs annually for 29 Sony artists, analyzed music trends via dashboards, and guided curation, collaborations, and artist development.



Believe

Artist & Repertoire Manager
December 2023 – May 2024

Successfully acquired Javanese Pop and hyperlocal artists, utilizing sales and marketing analytics with dashboard data for revenue projections, contract management, and A&R strategies.

A. Introduction

Self-Overview

I am a data-driven professional passionate about leveraging analytics, visualization, and predictive modeling to drive business decisions. With a diverse background in **engineering, music, and information systems**, I bring a unique blend of **technical expertise and creativity** to solve complex problems.

My expertise lies in **Business Intelligence**, with a strong focus on **data analysis, A/B testing, customer segmentation, uplift modeling, and churn prediction** using **SQL, Python, Tableau, and Power BI**. I thrive on transforming raw data into actionable insights that optimize strategies and enhance decision-making.

As I continue to grow in the **data industry**, my goal is to **bridge the gap between data and business impact**, ensuring organizations harness the full power of their data assets. I am eager to collaborate, innovate, and contribute to **data-driven success stories**.



South Jakarta



hijirdw@gmail.com



<https://www.linkedin.com/in/hijirdella/>



<https://github.com/hijirdella>

B. Projects Overview

Data Warehouse Design



Design a Data Warehouse (DWH) for Employee & Sales, including an ERD diagram, Star Schema, Data Mart tables, SQL scripts, and a presentation on schema, relationships, and insights. [View Project](#)

PostgreSQL ETL Procedure for Data Warehouse



This project contains dwh.generate_sales(), a stored procedure for ETL in PostgreSQL, moving transaction data to staging and loading it into the data warehouse. Includes documentation and a step-by-step presentation. [View Project](#)

House Price Analysis (EDA) and Correlation-Insights



Analyzing housing data using EDA, preprocessing, and statistical techniques like Pearson Correlation, Chi-Square Test, Linear Regression, T-Test, and ANOVA to identify trends and relationships. [View Project](#)

Bank Marketing Campaign Impact Analysis on Term Deposits



Analyze marketing campaign impact on term deposits using EDA, correlation, chi-square tests, and linear regression. [View Project](#)

Automated ETL Pipeline with Airflow for School Data



Showcasing an Airflow ETL pipeline to extract school data from an API, transform it, and load it into PostgreSQL. Includes Python scripts, documentation, and examples. [View Project](#)

Netflix Recommendation System Using Machine Learning



A Netflix Recommendation System using machine learning models (KNN, Decision Tree, Random Forest, Logistic Regression, Naive Bayes, K-Means) evaluated for accuracy and clustering, with insights into content trends and user preferences. [View Project](#)

During this bootcamp, I have worked on various projects focused on data analysis, business modeling, and data visualization.



Customer Analysis and RFM Segmentation for Superstore

Analyzing Superstore customer data with RFM segmentation, EDA, and visualizations, identifying key insights, and presenting findings with a report and Tableau dashboard. [View Project](#)



Funnel & Cohort Analysis Dashboard

This project features a Business Intelligence case study using Funnel and Cohort Analysis in Tableau, including data preprocessing, visualization, and trend analysis. It provides interactive dashboards and a presentation with key insights and recommendations. [View Project](#)



Bank Customer Churn Prediction and CLV Optimization

Analyzing bank customer churn and optimizing Customer Lifetime Value (CLV) using 11 machine learning models (e.g., Logistic Regression, Random Forest, XGBoost). Includes EDA, churn prediction, CLV analysis, and model evaluation with metrics like Precision, Recall, F1-score, and Confusion Matrix. [View Project](#)



A/B Testing for Spotify Playlist Recommendation

This project evaluates the impact of a new playlist recommendation algorithm on engagement, retention, and conversion rates using t-tests and chi-square tests. It includes randomized test design, feature engineering, statistical analysis, and visualizations to optimize recommendations. [View Project](#)



Uplift Modeling for Marketing Promotions

This project analyzes the impact of Discount vs. BOGO offers compared to No Offer (Control) using S-Learner & Uplift Random Forest. It includes EDA, AUUC, Gain Chart, and model evaluation to optimize marketing conversions. [View Project](#)



Olist E-Commerce Executive & Operational Dashboard

An interactive Tableau dashboard providing both executive and operational insights into Olist's sales trends, customer satisfaction, order fulfillment, and strategic performance for data-driven decision-making. [View Project](#)

C. Main Project



<https://github.com/hijirdella/Kalbe-Nutritionals-BI-Final-Project-Sales-RFM-Cohort-Churn-CLV-Optimization>



<https://public.tableau.com/app/profile/hijir.della.wirasti5486/viz/KalbeNutritionalsBusinessPerformanceDashboardUnderstandingSalesDynamicsCustomerRetentionRFMCohortandCLVStrategies/Sales-Dashboard>



01. [Combine Kalbe Data.ipynb](#)
02. [Preprocessing - EDA - RFM - Kalbe Farma.ipynb](#)
03. [Churn Analysis - Kalbe Farma.ipynb](#)



PostgreSQL

04. [Cohort Query.sql](#)



1. Project Background

- **General Description:** This project aims to enhance sales and customer loyalty at Kalbe Nutritionals using a data-driven approach. By implementing RFM analysis, cohort analysis, and CLV estimation, the project seeks to understand customer behavior and identify effective strategies for improving retention and customer value.
- **Primary Objectives:** Improve marketing strategy effectiveness, reduce customer churn, and increase Customer Lifetime Value (CLV).
- **Expected Outcomes:** Increased customer retention, higher sales, and more targeted marketing strategies.
- **Project Importance:** In the competitive nutrition industry, understanding and meeting customer needs is crucial to maintaining market share and increasing profitability.
- **Beneficiaries:** Management teams, marketing departments, and Kalbe Nutritionals customers.



2. Business Problem

- **Business Challenge:** Kalbe Nutritionals faces challenges in retaining customers and increasing sales in a highly competitive market. High churn rates and a lack of deep understanding of customer segmentation hinder marketing effectiveness.
- **Problem Importance:** A high churn rate can reduce revenue and increase customer acquisition costs. Without proper segmentation, marketing campaigns may be ineffective, leading to wasted resources.
- **Impact of Solving the Problem:** By identifying valuable customer segments and understanding their behavior, Kalbe Nutritionals can design effective retention strategies, reduce churn, and increase profitability.



Literature Review

The implementation of data-driven strategies such as **RFM Analysis (Recency, Frequency, Monetary), Cohort Analysis, and Customer Lifetime Value (CLV)** can significantly impact **sales and customer loyalty** at **Kalbe Nutritionals**. Below are some data and figures supporting the effectiveness of this approach:

Increase in Customer Loyalty and Revenue

- **Eastwood**: Achieved a **21% increase in email marketing profits** after implementing RFM analysis.
- **L'Occitane**: Experienced a **25-fold increase in revenue per email** through customer segmentation using the RFM model.
- **Frederick's of Hollywood**: Recorded a **6-9% conversion rate** in their campaigns after adopting RFM analysis. ([Desai et al. 2021](#))

Reduction in Churn Rate

- Companies utilizing churn prediction can **reduce churn rates by up to 30%** with more targeted strategies.
- By **identifying at-risk customers early**, businesses can implement retention strategies such as **loyalty programs and exclusive discounts** ([Yulianti, 2018](#)).

Marketing Cost Efficiency

- With **better customer segmentation**, businesses can allocate marketing budgets more efficiently, focusing on high-value customers.
- Using **RFM and CLV analysis** can **reduce Customer Acquisition Cost (CAC) by up to 25%**, as companies can concentrate on **more profitable customers** ([Feaseo, 2022](#)).

Better Customer Behavior Insights

- Cohort analysis helps businesses track customer behavior over time, improving retention strategies. Companies using this approach saw a 15% retention increase in six months ([DQLab, 2023](#)).

Optimized Customer Lifetime Value (CLV)

- RFM-based CLV calculations and clustering techniques like K-Means can improve marketing efficiency by 30% and long-term profitability by 20% ([Laksono et al. 2023](#)).

3. Data Analysis and Insights



DATASET OVERVIEW



This **dataset** consists of four CSV files: Customer, Store, Product, and Transaction. It is a dummy dataset created for an FMCG case study covering a one-year period, collected through a membership program.

This case study utilizes four key datasets:

- **Customer Data:** Contains demographic details such as age, gender, marital status, and income levels.
- **Store Data:** Provides store-related information including store name, group affiliation, trade type (Modern or General Trade), and geographical coordinates.
- **Product Data:** Lists product names, unique IDs, and pricing details.
- **Transaction Data:** Includes transactional records such as transaction date, product details, quantity purchased, and total purchase amount.

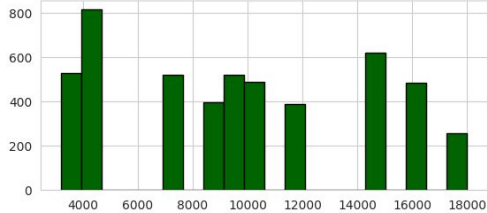
Univariate Analysis

Key Insights & Recommendations:

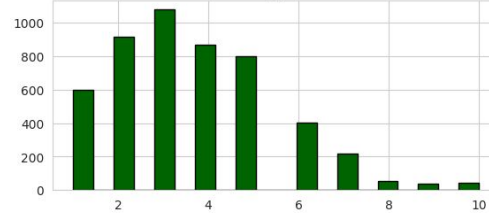
- **Popular price points: 4000-5000 & 14000-16000** → **Focus promotions here**
- **Small purchase trend (1-5 items)** → **Use upselling & bundling**
- **Most transactions: 10,000 - 30,000 IDR** → **Encourage higher spend with discounts**
- **Main audience: 25-55 years old** → **Target working professionals**
- **Low to mid-income customers** → **Offer paylater & budget-friendly deals**
- **High-income segment** → **Introduce premium products**

Distribution of Numerical Features

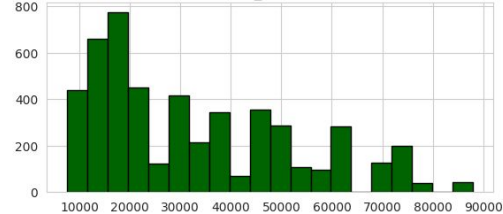
price



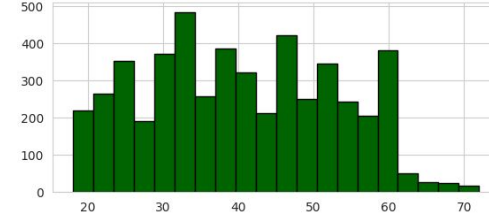
qty



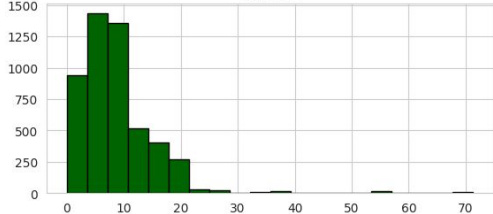
total_amount



age



income

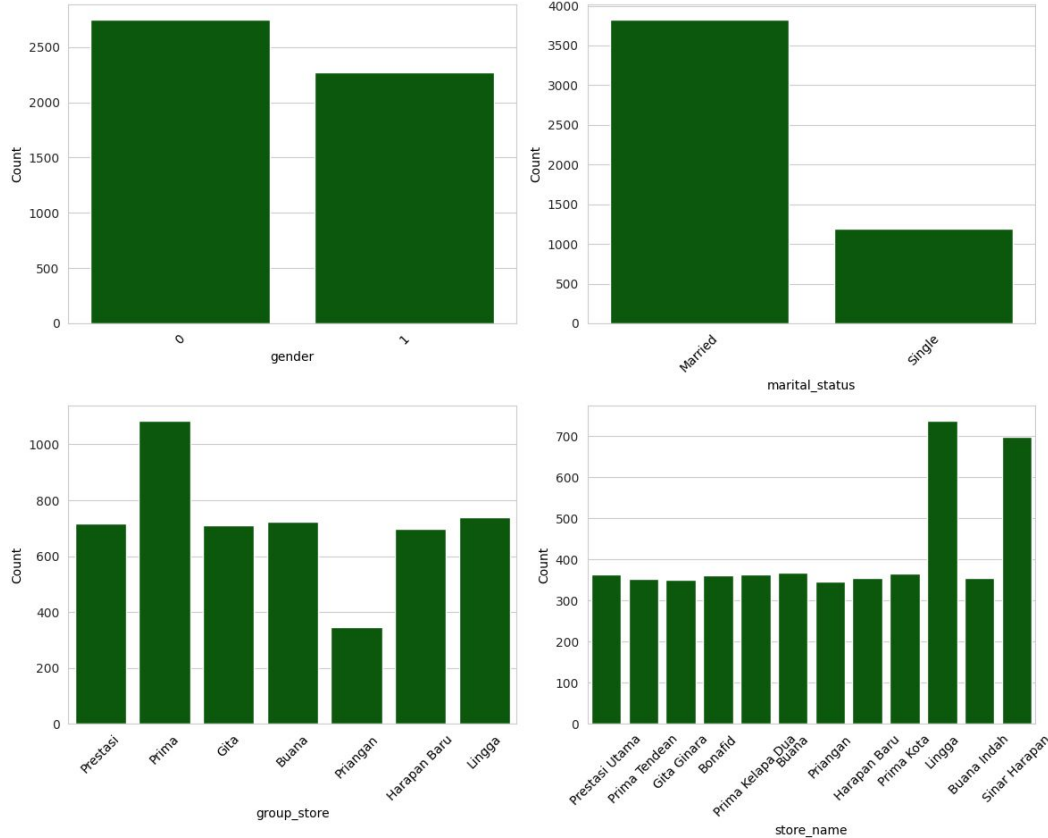


Uniavariate Analysis

Key Insights & Recommendations:

- Gender Distribution:**
 - More female (0) than male (1) customers.
 - Target female-oriented promotions & product selections.**
- Marital Status:**
 - Majority of customers are **married**.
 - Family/bundle deals & household product promotions** may perform well.
- Group Store Distribution:**
 - Prima dominates sales**, followed by other groups like Gita and Lingga.
 - Strengthen partnerships with top-performing stores** for better sales.
- Store Name Distribution:**
 - Lingga & Sinar Harapan lead in store transactions.**
 - Optimize inventory & exclusive promotions in these stores.**

Distribution of Categorical Features





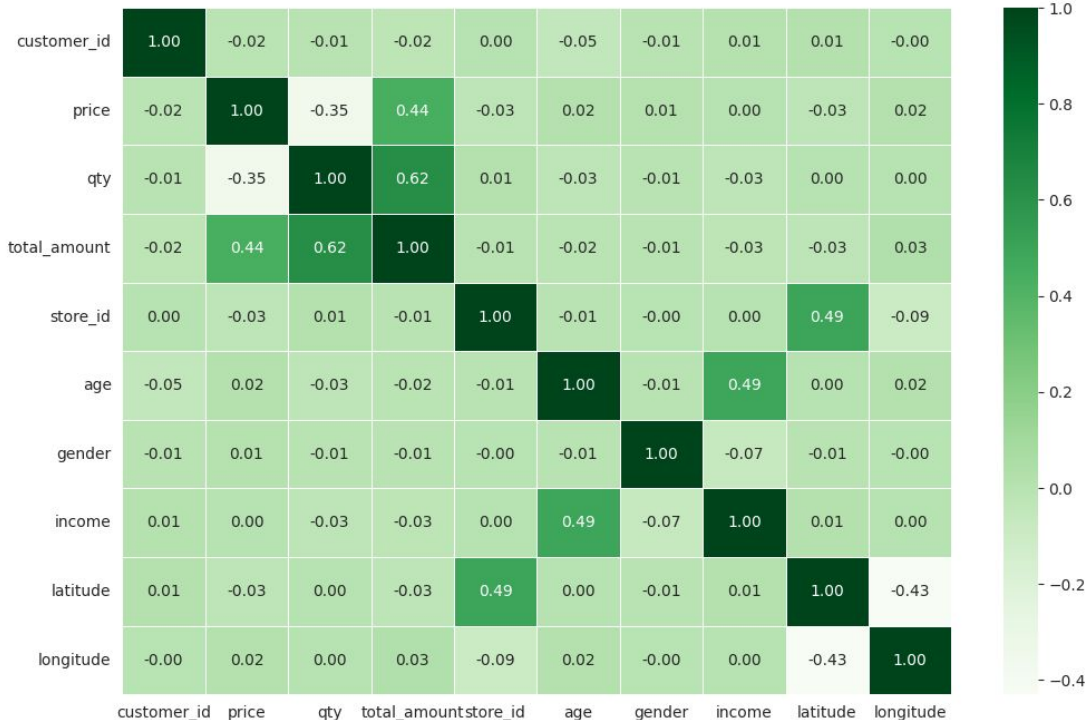
Multivariate Analysis

Key Insights from Pairplot Analysis:

- ✓ **Price vs. Quantity:** Lower-priced products are bought in higher quantities.
- ✓ **Total Amount vs. Quantity:** Higher quantity leads to higher spending, but not always linearly.
- ✓ **Income vs. Total Amount:** Higher-income customers **don't** necessarily spend more.
- ✓ **Age vs. Income:** Younger customers have more varied income levels.

Multivariate Analysis

Correlation Matrix of All Numerical Features



Key Insights from Correlation Matrix:

- ✓ **Price & Total Amount (0.44):** Higher-priced items contribute more to total spending.
- ✓ **Quantity & Total Amount (0.62):** More items purchased = higher total spending.
- ✓ **Income & Age (0.49):** Older customers tend to have higher income.
- ✓ **Income & Total Amount (~0):** No strong correlation → high-income customers don't always spend more.

4. DASHBOARD

[Tableau Link](#)



Sales Dashboard

Month of Transaction Date

January 2022

December 2022

Segment

{All}

Store Id

{All}

Store Name

{All}

Group Store

{All}

Product Name

{All}

SALES

162,043,000

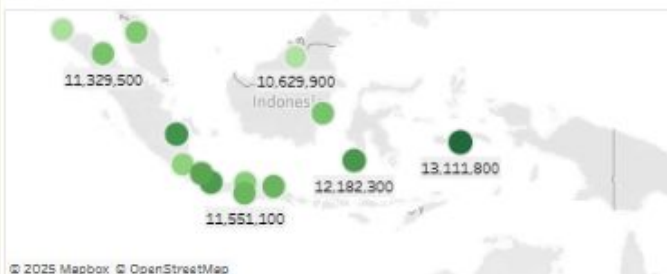
ORDERS

18,296

CUSTOMERS

447

Map



© 2025 Mapbox © OpenStreetMap

Sales by Demography



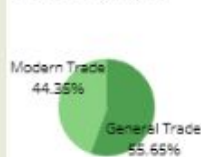
Income vs Spent



Top Customer



Sales by Type



Sales by Store Id



Sales by Store Name



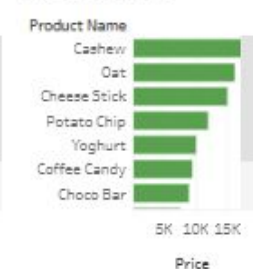
Sales by Group Store



Top Product



Product Price



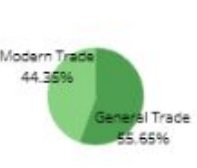
SALES ANALYSIS



Overall Performance: Sales, orders, and customer trends appear stable over time.



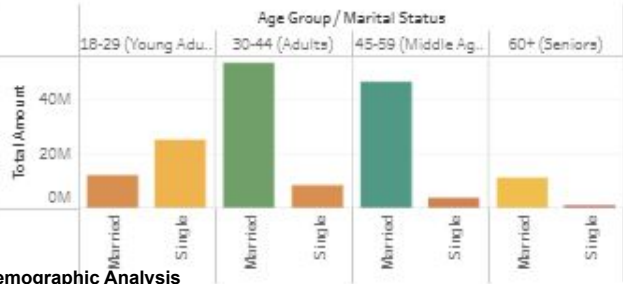
Sales by Type



Sales by Type

Indicating traditional retail channels still play a crucial role.

Sales by Demography



Demographic Analysis

- **Married adults (30-44 years old)** contribute the highest sales.
- **Younger (18-29) and senior (60+) segments** have lower sales, suggesting potential for targeted marketing.

Geographic Insights

- Eastern Indonesia & Sulawesi lead in sales, indicating strong market presence.
- Jakarta & Java contribute significantly but are behind emerging regions.
- Potential for growth in Sumatra and Kalimantan, where sales are slightly lower.
- Expansion opportunities in Bali/Nusa Tenggara, which have strong demand.

Sales by Store Id



Sales by Store Name



Sales by Group Store



Store & Group Store Performance

- **Top-performing stores:** Lingga, Sinar Harapan, and Prestasi Utama.
- **Best-performing group stores:** Prima and Lingga.

Top Product



Product Price



Top Products & Pricing

- **Best-selling products:** Cheese Stick, Choco Bar, and Coffee Candy.
- **Higher-priced products like Cashew and Oat** might need premium marketing strategies.

Recency, Frequency, Monetary Dashboard

Month of Transaction Date

January 2022

December 2022

Segment

(All)

Store Id

(All)

Product Name

(All)

Type

(All)

Metrics

Sales

Age Group

(All)

Marital Status

(All)

RFM Score

(All)

RECENTY

812

FREQUENCY

12

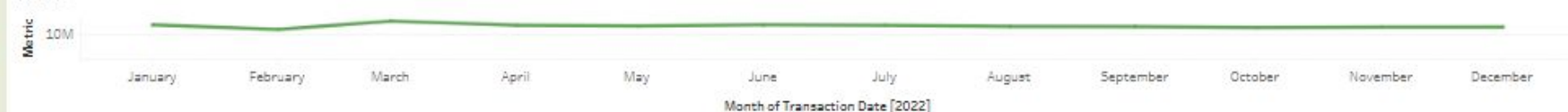
MONETARY

384,600

AOV

31,867

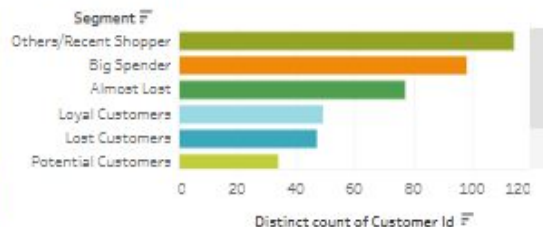
Trend



RFM Segmentation



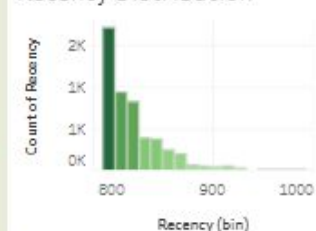
Top RFM Segment



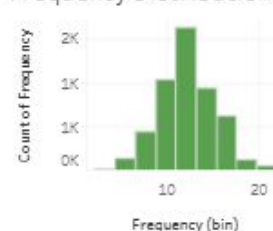
RFM Score Distribution



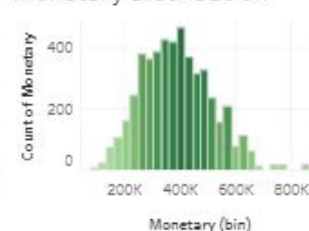
Recency Distribution



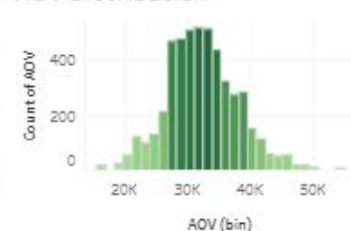
Frequency Distribution



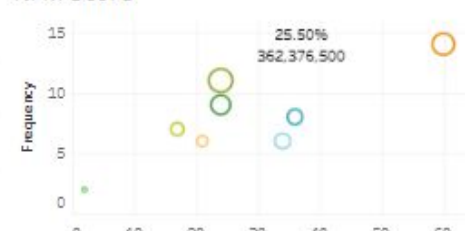
Monetary Distribution



AOV Distribution



RFM Score



RFM ANALYSIS

REGENCY

FREQUENCY

MONETARY

AOV

812

12

384,600

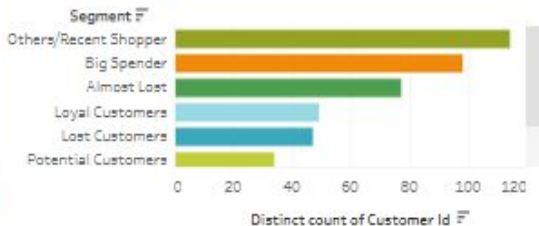
31,867

Key Metrics: Recency (812) indicates declining engagement, Frequency (12) shows moderate repeat purchases, Monetary (384,600) reflects good spending, and AOV (31,867) highlights high-value transactions per order.

RFM Segmentation



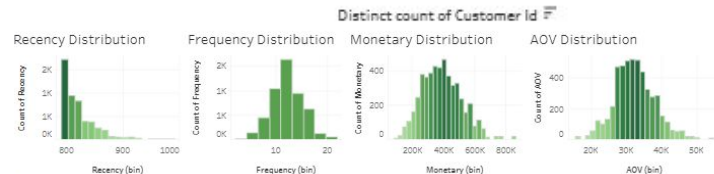
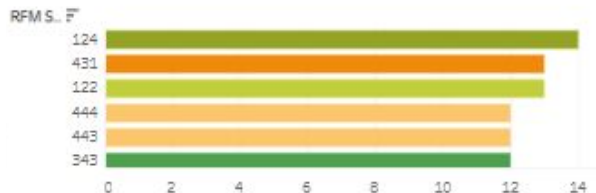
Top RFM Segment



RFM Analysis

- **Largest Segments: Recent Shoppers (25.50%) and Big Spenders (21.92%)** – Key focus for retention and upselling.
- **At-Risk: Almost Lost (17.23%) and Lost Customers (10.51%)** – Need re-engagement strategies.
- **Loyal Customers (10.69%)** – Maintain with personalized offers.
- **Smaller Segments: Potential Customers (7.61%), Lost Cheap (5.37%), and Best Customers (0.89%)** – Growth and retention opportunities.

RFM Score Distribution



Score Analysis:

Most customers have mid-to-high RFM scores, meaning opportunities exist for upselling and loyalty programs.

RFM Distribution Insights

1. **Recency (Left-Skewed)** – Most customers haven't purchased in a long time. **Reactivation needed.**
2. **Frequency (Normal)** – Majority buy 10-15 times. **Encourage more repeat purchases.**
3. **Monetary (Normal)** – Most spend 300K-500K. **Upsell strategies can boost spending.**
4. **AOV (Slight Right-Skewed)** – with most transactions between 25K-40K, peaking at 30K-35K, suggesting opportunities for **bundling, volume discounts, and premium upselling.**

Cohort Dashboard

Retention Rate

56% 100%

Year of Cohort Month

2022

RETENTION RATE

66.4%

Month Offset

(All)

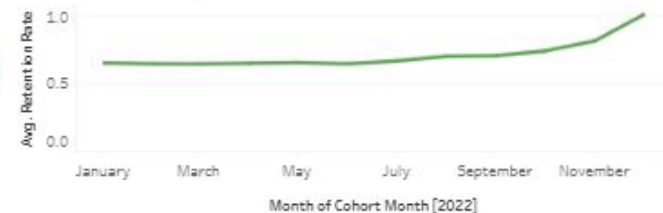
Month of Cohort Month

(All)

Cohort Analysis

Month of Cohort ..	0	1	2	3	4	5	6	7	8	9	10	11
January	100%	59%	61%	64%	65%	60%	59%	62%	63%	63%	59%	61%
February	277	164	170	178	180	167	164	171	174	174	164	169
March	248	153	148	155	154	152	140	149	154	153	145	
April	100%	60%	62%	61%	61%	56%	58%	61%	60%	61%		
May	272	179	165	162	170	163	162	151	156			
June	100%	60%	64%	59%	59%	60%	58%	60%				
July	281	170	180	165	165	168	164	169				
August	100%	56%	59%	58%	57%	57%	61%					
September	278	156	164	162	159	159	170					
October	100%	58%	59%	62%	60%	59%						
November	273	159	162	168	163	160						
December	100%	63%	61%	60%	64%							
	271	171	166	163	173							
	100%	59%	58%	64%								
	266	158	153	169								
	100%	61%	60%									
	266	162	159									
	100%	62%										
	262	162										
	100%											
	273											

Retention Rate per Cohort



User Count per Cohort Over Time



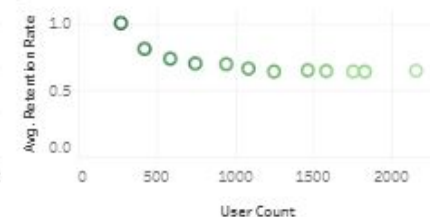
Retention Rate Over Time



Cumulative Retention Rate



Retention Rate vs. Initial User Count per Cohort



COHORT ANALYSIS

Cohort Analysis

Month of Cohort ...	0	1	2	3	4	5	6	7	8	9	10	11
January	100%	59%	61%	64%	65%	60%	59%	62%	63%	63%	59%	61%
February	277	184	170	178	180	167	164	171	174	174	184	189
March	100%	62%	60%	63%	62%	61%	58%	60%	62%	62%	58%	
April	246	153	148	155	154	152	140	149	154	153	145	
May	100%	60%	62%	61%	61%	56%	56%	61%	60%	61%		
June	285	170	178	174	174	161	166	173	171	174		
July	100%	66%	61%	60%	63%	60%	60%	56%	57%			
August	272	179	165	162	170	163	162	151	156			
September	100%	60%	64%	59%	59%	60%	56%	60%				
October	281	170	180	165	165	168	164	161				
November	100%	56%	59%	58%	57%	57%						
December	278	156	164	162	159	159	170					
	100%	58%	59%	62%	60%	59%						
	273	159	162	168	163	160						
	100%	63%	61%	60%	64%							
	271	171	166	163	173							
	100%	59%	56%	64%								
	266	158	153	169								
	100%	61%	60%									
	266	162	159									
	100%	62%										
	262	162										
	100%											
	273											

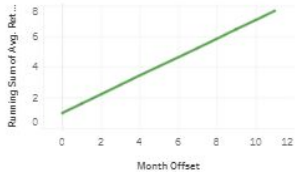
Key Insights:

- ✓ **Strong Early Retention (~60%+)**, but declines by month 6 → Enhance post-acquisition engagement.
- ✓ **Some cohorts (March, May, August) retain better** → Analyze and replicate success factors.
- ✓ **Later cohorts (Q3 & Q4) show improved retention** → Recent strategies are working; sustain them.
- ✓ **Long-term retention drops below 20% by month 9-12** → Strengthen loyalty and re-engagement efforts.

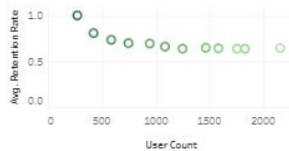
Retention Rate Over Time



Cumulative Retention Rate



Retention Rate vs. Initial User Count per Cohort



Retention Rate per Cohort



Retention Rate per Cohort

- Retention is **relatively stable but slightly increasing towards the end**.
- **Action:** Identify recent strategies boosting retention and optimize them.

User Count per Cohort Over Time



User Count per Cohort Over Time

- **Most users drop off quickly after the first month.**
- **Action:** Improve onboarding and early engagement incentives.

- ✓ **Early Drop-Off** – Retention falls sharply in the first month; strengthen onboarding and engagement.
- ✓ **Loyal Core Users** – Cumulative retention grows over time; leverage loyalty programs.
- ✓ **Acquisition Quality** – Larger cohorts show lower retention; improve targeting strategies.

CHURN & CLV Dashboard

Month of Transaction Date

January 2022

December 2022

Store Id

(All)

Store Name

(All)

Product Name

(All)

Type

(All)

Metrics

Sales

Marital Status

(All)

Age Group

(All)

Gender Label

(All)

RETENTION RATE

66.4%

CHURN RATE

6.7%

USERS

447

ACTIVE CUSTOMER

417

CHURN CUSTOMER

30

CLV

3,613,391

REVENUE IMPACT
OF CHURN

8,291,300

AVERAGE
PURCHASE VALUE

32,279

AVG CUSTOMER
LIFESPAN (M)

9.667

AVG CUSTOMER
LIFESPAN (D)

299.0

PURCHASE
FREQUENCY

11.23

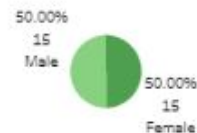
Churn Rate Over Time



Churn Demography



Churn by Gender



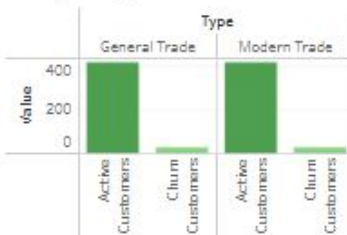
Churn by Age



Churn Percentage (Map)



Churn by Store Type



Store-Wise Churn Rate



Churn by Product Category

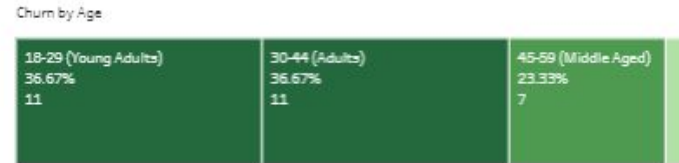
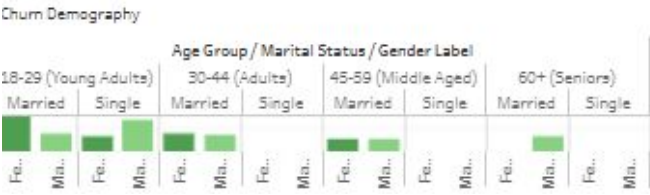


CHURN & CLV ANALYSIS

RETENTION RATE	CHURN RATE	USERS	ACTIVE CUSTOMER	CHURN CUSTOMER	CLV	REVENUE IMPACT OF CHURN	AVERAGE PURCHASE VALUE	AVG CUSTOMER LIFESPAN (M)	AVG CUSTOMER LIFESPAN (D)	PURCHASE FREQUENCY
66.4%	6.7%	447	417	30	3,613,391	8,291,300	32,279	9.667	299.0	11.23



Customers show **strong retention (66.4%)** and **high CLV (3.61M)**, indicating long-term value, but **churn (6.7%) is rising mid-year, especially among younger segments (18-44) and in key store (Harapan Baru dan Priangan)**. **Action:** Strengthen early engagement, target high-churn locations, and optimize product offerings to sustain loyalty.



5. Business Recommendation



Business Recommendation

Recomm.	Insights	Actionable Items
Expand High-Sales Regions	Eastern Indonesia & Sulawesi lead sales, while Jakarta & Java have potential for growth	<ol style="list-style-type: none">1. Strengthen distributor partnerships in top regions.2. Invest in marketing to boost Jakarta & Java performance.3. Explore expansion in Sumatra & Kalimantan.
Boost Retention & Reduce Churn	Retention is 66.4% , but churn (6.7%) is rising, especially among 18-44-year-olds and in Harapan Baru & Lingga .	<ol style="list-style-type: none">1. Implement personalized retention campaigns.2. Re-engage high-churn locations with exclusive offers.3. Strengthen onboarding & loyalty programs.
Enhance RFM-Based Customer Engagement	Big Spenders (21.92%) & Loyal Customers (10.96%) drive sales, while Almost Lost (17.23%) are at risk.	<ol style="list-style-type: none">1. Offer VIP perks for top customers.2. Run win-back campaigns for Almost Lost.3. Convert Recent Shoppers (25.5%) into repeat buyers..
Optimize Product & Pricing Strategy	High AOV (31,867) but churn on certain products (e.g., Coffee Candy & Choco Bar).	<ol style="list-style-type: none">1. Upsell premium products like Cashew & Oat.2. Reduce churn on low-retention products with bundles/discounts.3. Implement repeat purchase incentives (e.g., buy 3x, get 1 free).
Leverage Best-Selling & High-Priced Products	Yoghurt, Oat, & Crackers lead in sales, while Potato Chips, Coffee Candy, & Choco Bar are high-priced.	<ol style="list-style-type: none">1. Promote best-sellers with targeted ads & bundling.2. Introduce premium packaging & branding for high-priced products.3. Offer trial packs or discounts to drive demand for premium items.
Increase Sales of Low-Performing Products	These products (Thai Tea, Cashew, Ginger Candy) have lower sales and need better positioning.	<ol style="list-style-type: none">1. Bundle Thai Tea with best-sellers (e.g., Coffee Candy).2. Rebrand Cashew as a premium snack with gift-worthy packaging.3. Position Ginger Candy for health-conscious consumers with immunity-boosting messaging.
Encourage Repeat Purchases	Cheese Stick is the top-selling product , but can be optimized further.	<ol style="list-style-type: none">1. Exclusive Cheese Stick multipacks (e.g., family-size or snack packs) to drive bulk purchases.2. Limited-edition seasonal flavors to create excitement & urgency.3. Loyalty points for frequent buyers (e.g., buy 5, get 1 free).

Thank You !

Any QUESTIONS ?

hijirdw@gmail.com

<https://www.linkedin.com/in/hijirdella/>

<https://github.com/hijirdella/Kalbe-Nutritionals-BI-Final-Project-Sales-RFM-Cohort-Churn-CLV-Optimization>



6. Appendix

DATASET OVERVIEW



Case Study - Customer



File Edit View Insert Format Data Tools Extensions Help

Q Menus 100% \$ % .0 .00 123 Def

G30 fx

	A	B	C	D	E
1	CustomerID	Age	Gender	Marital Status	Income
2	1	55	1	Married	5,12
3	2	60	1	Married	6,23
4	3	32	1	Married	9,17
5	4	31	1	Married	4,87



Case Study - Product



File Edit View Insert Format Data

Q Menus 100%

N12 fx

	A	B	C
1	ProductID	Product Name	Price
2	P1	Choco Bar	8800
3	P2	Ginger Candy	3200
4	P3	Crackers	7500
5	P4	Potato Chip	12000



Case Study - Store



File Edit View Insert Format Data Tools Extensions Help

Q Menus 100% \$ % .0 .00 123 Default... - {

M20 fx

	A	B	C	D	E	F
1	StoreID	StoreName	GroupStore	Type	Latitude	Longitude
2	1	Prima Tendean	Prima	Modern Trade	-6,2	106,816,666
3	2	Prima Kelapa Du	Prima	Modern Trade	-6,914,864	107,608,238
4	3	Prima Kota	Prima	Modern Trade	-7,797,068	110,370,529
5	4	Gita Ginara	Gita	General Trade	-6,966,667	110,416,664



Case Study - Transaction



File Edit View Insert Format Data Tools Extensions Help

Q Menus 100% \$ % .0 .00 123 Default... - 10 + B I A

J20 fx

	A	B	C	D	E	F	G	H
1	TransactionID	CustomerID	Date	ProductID	Price	Qty	TotalAmount	StoreID
2	TR11369	328	01/01/2022	P3	7500	4	30000	12
3	TR16356	165	01/01/2022	P9	10000	7	70000	1
4	TR1984	183	01/01/2022	P1	8800	4	35200	4
5	TR35256	160	01/01/2022	P1	8800	7	61600	4

DATASET PROFILE – CUSTOMER

	CustomerID	Age	Gender	Marital Status	Income
0	1	55	1	Married	5,12
1	2	60	1	Married	6,23
2	3	32	1	Married	9,17
3	4	31	1	Married	4,87
4	5	58	1	Married	3,57

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 447 entries, 0 to 446
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CustomerID      447 non-null   int64
1   Age             447 non-null   int64
2   Gender          447 non-null   int64
3   Marital Status  444 non-null   object
4   Income          447 non-null   object
dtypes: int64(3), object(2)
memory usage: 17.6+ KB
```

	CustomerID	Age	Gender
count	447.000000	447.000000	447.000000
mean	224.000000	39.782998	0.458613
std	129.182042	12.848719	0.498842
min	1.000000	0.000000	0.000000
25%	112.500000	30.000000	0.000000
50%	224.000000	39.000000	0.000000
75%	335.500000	50.500000	1.000000
max	447.000000	72.000000	1.000000

	Marital Status	Income
count	444	447
unique	2	369
top	Married	0
freq	340	16

Recommended Data Improvements:

- **Fix the income format** (replace commas with dots if necessary to ensure numerical accuracy).
- **Handle missing marital status values**, potentially using mode imputation.
- **Investigate age 0 entries** to check for data entry errors.

DATASET PROFILE – STORE

	StoreID	StoreName	GroupStore	Type	Latitude	Longitude
0	1	Prima Tendean	Prima	Modern Trade	-6,2	106,816,666
1	2	Prima Kelapa Dua	Prima	Modern Trade	-6,914,864	107,608,238
2	3	Prima Kota	Prima	Modern Trade	-7,797,068	110,370,529
3	4	Gita Ginara	Gita	General Trade	-6,966,667	110,416,664
4	5	Bonafid	Gita	General Trade	-7,250,445	112,768,845
5	6	Lingga	Lingga	Modern Trade	-5,135,399	11,942,379
6	7	Buana Indah	Buana	General Trade	3,316,694	114,590,111
7	8	Sinar Harapan	Harapan Baru	General Trade	554,829	95,323,753
8	9	Lingga	Lingga	Modern Trade	-3,654,703	128,190,643
9	10	Harapan Baru	Harapan Baru	General Trade	3,597,031	98,678,513
10	11	Sinar Harapan	Prestasi	General Trade	533,505	101,447,403
11	12	Prestasi Utama	Prestasi	General Trade	-2,990,934	104,756,554
12	13	Buana	Buana	General Trade	-126,916	116,825,264
13	14	Priangan	Priangan	Modern Trade	-5,45	10,526,667

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   StoreID      14 non-null     int64
1   StoreName    14 non-null     object
2   GroupStore   14 non-null     object
3   Type         14 non-null     object
4   Latitude     14 non-null     object
5   Longitude    14 non-null     object
dtypes: int64(1), object(5)
memory usage: 804.0+ bytes
```

StoreID	
count	14.0000
mean	7.5000
std	4.1833
min	1.0000
25%	4.2500
50%	7.5000
75%	10.7500
max	14.0000

	StoreName	GroupStore	Type	Latitude	Longitude
count	14	14	14	14	14
unique	12	7	2	14	14
top	Lingga	Prima	General Trade	-6,2	106,816,666
freq	2	3	8	1	1

Recommended Data Improvements:

- **Fix latitude and longitude formatting** (convert commas to dots for numerical consistency).
- **Check for duplicate stores** in `StoreName` and `GroupStore`, as "Lingga" and "Sinar Harapan" appear multiple times.
- **Analyze store type distribution** to see if "Modern Trade" and "General Trade" differ significantly in geographical placement.

DATASET PROFILE – PRODUCT

	ProductID	Product Name	Price
0	P1	Choco Bar	8800
1	P2	Ginger Candy	3200
2	P3	Crackers	7500
3	P4	Potato Chip	12000
4	P5	Thai Tea	4200
5	P6	Cashew	18000
6	P7	Coffee Candy	9400
7	P8	Oat	16000
8	P9	Yoghurt	10000
9	P10	Cheese Stick	15000

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ProductID    10 non-null     object
1   Product Name  10 non-null     object
2   Price        10 non-null     int64
dtypes: int64(1), object(2)
memory usage: 372.0+ bytes
```

	Price
count	10.000000
mean	10410.000000
std	4890.455557
min	3200.000000
25%	7825.000000
50%	9700.000000
75%	14250.000000
max	18000.000000

	ProductID	Product Name
count	10	10
unique	10	10
top	P1	Choco Bar
freq	1	1

Recommended Data Improvements:

- **Further categorize products** into types (e.g., snacks, beverages, dairy) for better segmentation.
- **Analyze price segmentation** to identify pricing trends and potential product bundling opportunities.

DATASET PROFILE – TRANSACTION

	TransactionID	CustomerID	Date	ProductID	Price	Qty	TotalAmount	StoreID
0	TR11369	328	01/01/2022	P3	7500	4	30000	12
1	TR16356	165	01/01/2022	P9	10000	7	70000	1
2	TR1984	183	01/01/2022	P1	8800	4	35200	4
3	TR35256	160	01/01/2022	P1	8800	7	61600	4
4	TR41231	386	01/01/2022	P9	10000	1	10000	4

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5020 entries, 0 to 5019
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   TransactionID    5020 non-null   object
1   CustomerID       5020 non-null   int64
2   Date             5020 non-null   object
3   ProductID        5020 non-null   object
4   Price            5020 non-null   int64
5   Qty              5020 non-null   int64
6   TotalAmount      5020 non-null   int64
7   StoreID          5020 non-null   int64
dtypes: int64(5), object(3)
memory usage: 313.9+ KB
```

	CustomerID	Price	Qty	TotalAmount	StoreID
count	5020.000000	5020.000000	5020.000000	5020.000000	5020.000000
mean	221.263745	9684.800797	3.644622	32279.482072	7.489841
std	129.672955	4600.708780	1.855295	19675.462455	4.028502
min	1.000000	3200.000000	1.000000	7500.000000	1.000000
25%	108.000000	4200.000000	2.000000	16000.000000	4.000000
50%	221.000000	9400.000000	3.000000	28200.000000	7.000000
75%	332.000000	15000.000000	5.000000	47000.000000	11.000000
max	447.000000	18000.000000	10.000000	88000.000000	14.000000

	TransactionID	Date	ProductID
count	5020	5020	5020
unique	4908	365	10
top	TR71313	02/03/2022	P5
freq	3	31	814

Recommended Data Improvements:

- **Check for duplicate transactions** (since there are 4,908 unique TransactionIDs but 5,020 rows).
- **Analyze product demand trends** to see which items drive the most sales.
- **Investigate high-value transactions** to understand customer purchasing behavior.
- **Segment transactions by store** to identify top-performing locations.

Customer Segmentation

1. **Best Customer** (R=1, F=1, M=1)
Highly engaged, frequent buyers with high spending. Prioritize loyalty programs and exclusive rewards.
2. **Potential Customers** (R=1, F=1, M=2) or (R=1, F=2, M=2) or (R=2, F=1, M=1) or (R=2, F=2, M=2)
Emerging valuable customers. Encourage repeat purchases with personalized offers.
3. **Lost Cheap** (R=4, F=4, M=4 or 3)
Inactive, low-value buyers. Low priority for retention, but could be re-engaged with discounts.
4. **Big Spender** (M=1)
High-value customers regardless of frequency. Offer premium deals and upsell opportunities.
5. **Loyal Customers** (F=1)
Frequent buyers with consistent purchases. Strengthen relationships through VIP programs.
6. **Others/Recent Shopper** (R=1 or R=2)
New or occasional buyers. Encourage further engagement through follow-ups and promotions.
7. **Almost Lost** (R=3)
Customers showing decreased engagement. Use retention strategies like personalized reactivation campaigns.
8. **Lost Customers** (R=4)
Inactive buyers. Re-engage with strong incentives like exclusive discounts or special deals.

REFERENCES

- Desai, S. K., & Kalyandurgmath, K. (2021). Strategy to increase lifetime value of a customer using RFM. International Journal of Creative Research Thoughts (IJCRT), 9(7), ISSN: 2320-2882.
- Yulianti, Yulianti. (2018). Metode Data mining Untuk prediksi Churn Pelanggan. Jurnal ICT Akademi Telkom Jakarta. 9. 46-52.
- Laksono, B. C., & Wulansari, I. Y. (n.d.).(2023). Pemodelan dan penerapan metode RFM pada estimasi nilai konsumen (Customer Lifetime Value) menggunakan K-Means Clustering Machine Learning. Politeknik Statistika STIS.