

PORTFOLIO

RIFKI MUHAMMAD ARBIAN

2024

TABLE OF CONTENTS

- 1 About me
- 2 Education, Skills & Certification
- 3 Working Experience
- 4 My Project
- 5 Summary
- 6 Conclusion / Business Insight
- 7 Link

Hello,

I'M RIFKI MUHAMMAD ARBIAN

1

2

3

4

5

6

7

8

9

DATA ENTHUSIAST

A Bachelor of Law graduate who passionate about business is seeking to transition into a data analyst and business intelligence career.

I have completed the Business Intelligence Bootcamp organized by dibimbing.id. Aiming to apply knowledge of data analysis and business acumen to contribute effectively to organizational objectives.

Reach me



rifkimad10@gmail.com



+62 859 4722 9254



linkedin.com/in/rifkimuharbian99/





Education

2015 - 2020

Faculty Of Law - Islam Indonesia University

GPA : 3.2 / 4.0

1

2

3

4

5

6

7

8

9

Bootcamp

2024

Business Intelligence Bootcamp - dibimbing.id

Batch 9, Januari - June 2024

Final Grade = 92.82

Skills & Tools



*SQL, Python, Tableau, DBeaver, Powerpoint,
Excel, PowerBI, FB Ads, and Ms Word.*



CERTIFICATE

OF COMPLETION AND PASSING THE EXAM

Student ID : 38202

Certificate no: 201029BI09063010

THIS CERTIFICATE IS PROUDLY PRESENTED TO

RIFKI MUHAMMAD ARBIAN

FOR SUCCESSFULLY COMPLETING

Business Intelligence Bootcamp

HELD ON: 13 Januari 2024 – 23 Juni 2024

ZAKY MUHAMMAD SYAH
CEO DIBIMBING.ID



VEGI FATURRAHMAN
HEAD OF BOOTCAMP



Verify at:
<https://dibimbng.id/certicate-validation?cn=201029BI09063010>



SERTIFIKAT KOMPETENSI LULUSAN

2022/11/2777/831149

Dengan bangga diberikan kepada :

Rifki Muhammad Arbian

Telah Menyelesaikan Pelatihan :

Belajar Manajemen Pengembangan Produk Digital dan Penerapan Cara Kerja SCRUM Untuk Menjadi Calon Manajer Produk

Jakarta, 4 November 2022

Elnofian
Expert of Faculty Member Pijar Mahir



Verify this certificate at:
<https://pijarmahir.id/sertifikat/FUBlxnH1J>



CERTIFICATE OF COMPLETION

Rifki Muhammad Arbian

has successfully completed the online course:
Introduction to Data Analytics

This professional has demonstrated initiative and a commitment to deepening their skills and advancing their career. Well done!

09th Oct 2024

Certificate code : 7440471

Krishna Kumar
CEO, Simplilearn



CERTIFICATE OF COMPLETION

Rifki Muhammad Arbian

has successfully completed the online course:
Power BI for Beginners

This professional has demonstrated initiative and a commitment to deepening their skills and advancing their career. Well done!

15th Oct 2024

Certificate code : 7457400

Krishna Kumar
CEO, Simplilearn



Working Experience

1

2

3

4

5

6

7

8

9

- **Staff Notary & PPAT**

Notary&PPAT Lies Setyorini SH, MH, MPD (Oktober 2021 - Januari 2023)

Drafted contracts, sales or credit agreements, establishing limited liability companies (PT), fiduciary guarantees, and then entered them into the online system to be recorded in the national archives.

- **FB Ads and Advertising**

Freelance (Januari 2023 - April 2023)

Finded winning products by checking purchase data on Facebook Ads and then created ads tailored to the appropriate target market demographics.

- **Owner & Storekeeper**

Toko Setyo (April 2023 - Present)

Finding winning products that match the target market, then becoming the person who sells these products to increase company sales, conducting stock-taking, and creating a system to transform the store from a conventional setup to one where everything is recorded digitally.

My Previous Project

1

2

3

4

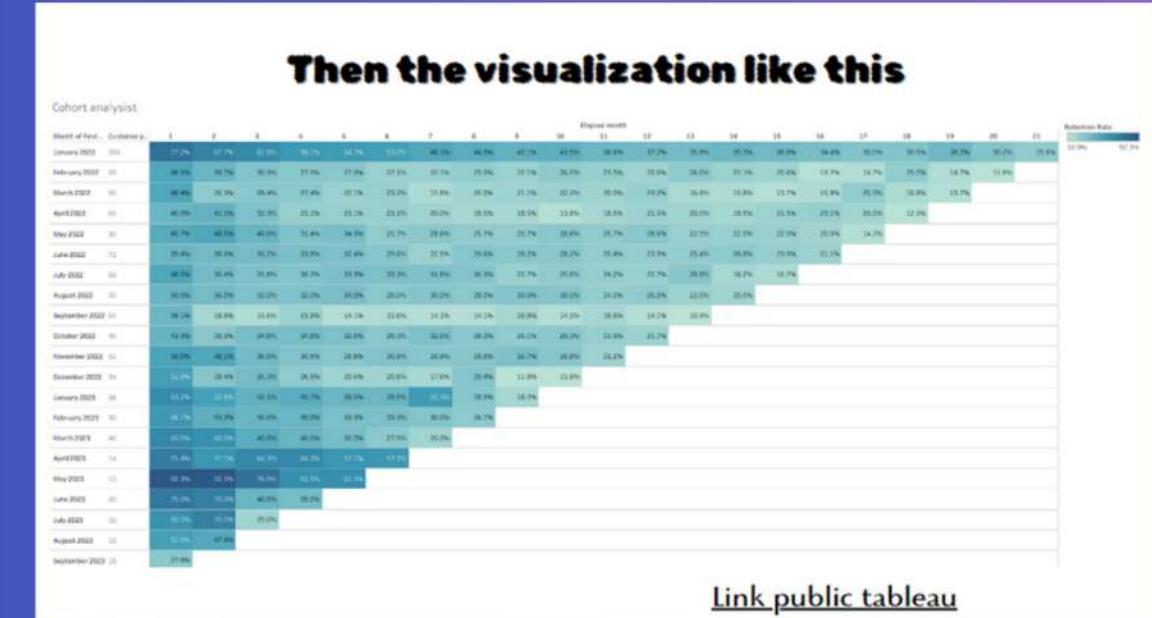
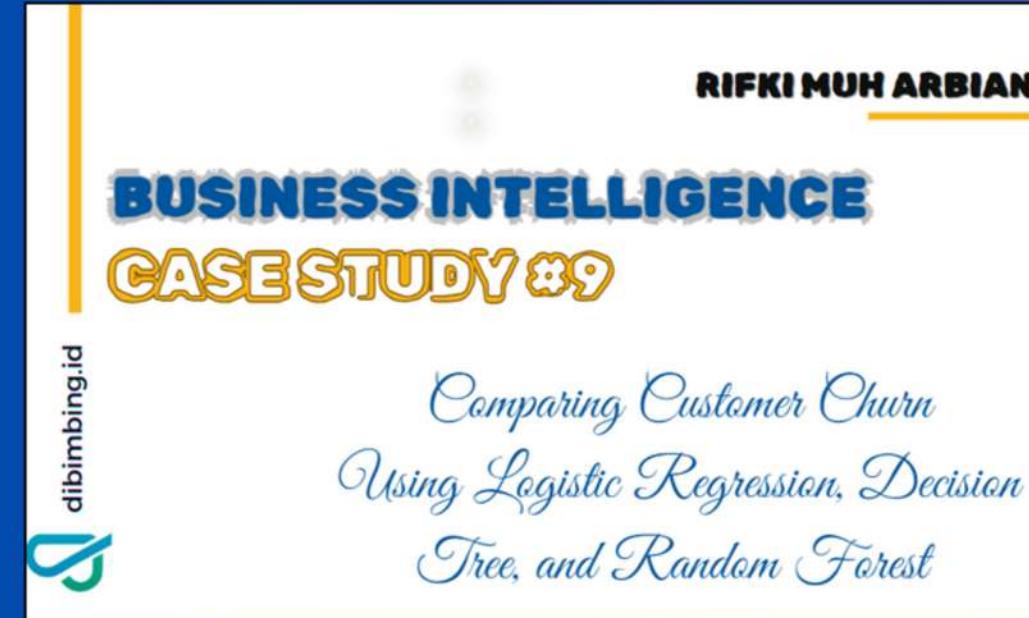
5

6

7

8

9



This project focuses on customer churn using Python. It involves applying machine learning with three different models to determine which model is the best for predicting customer churn.

This project is about funnel and cohort analysis. Using Tableau, it analyzes the visualizations of funnel analysis (customer behavior) and cohort analysis (retention rate) for a product or company.

The screenshot shows a DBeaver interface with several tabs. One tab displays a complex SQL query involving CTEs to filter transactions from BRI with a successful status, ranking them by total revenue. Another tab shows the resulting table of transactions, which includes columns for bank_code, transaction_status, product_category, payment_method, total_revenue, and rank.

rank	bank_code	transaction_status	product_category	payment_method	total_revenue	rank
1	BRI	SUCCESS	fashion	MANUAL_TRANSFER	241,699,390	1
2	BRI	SUCCESS	gadget	MANUAL_TRANSFER	166,144,903	2
3	BRI	SUCCESS	home & living	MANUAL_TRANSFER	108,332,990	3
4	BRI	SUCCESS	fashion	VIRTUAL_ACCOUNT	81,874,668	4
5	BRI	SUCCESS	furniture	MANUAL_TRANSFER	83,638,266	5
6	BRI	SUCCESS	food	MANUAL_TRANSFER	82,857,152	6
7	BRI	SUCCESS	electronics	VIRTUAL_ACCOUNT	58,426,253	7
8	BRI	SUCCESS	toys	MANUAL_TRANSFER	56,727,536	8
9	BRI	SUCCESS	electronics	MANUAL_TRANSFER	54,719,266	9
10	BRI	SUCCESS	home & living	VIRTUAL_ACCOUNT	51,128,634	10
11	BRI	SUCCESS	furniture	VIRTUAL_ACCOUNT	27,453,755	11
12	BRI	SUCCESS	sports	MANUAL_TRANSFER	27,264,436	12
13	BRI	SUCCESS	sports	VIRTUAL_ACCOUNT	26,371,084	13
14	BRI	SUCCESS	food	VIRTUAL_ACCOUNT	26,228,892	14

This project involves SQL, where I am using the DBeaver tool. My objective is to create CTEs for filtering data from 2 or even 3 datasets that contain transactions made using BRI, with a transaction status of 'successful.' Then, I will rank each transaction according to its product category.

My Project

1

2

3

4

5

6

7

8

9

- Title** : Customer Segmentation and Customer Lifetime Value using All u need mart dataset
- Objective** : 1. Segmentation customer
2. Predictive Modeling
- Analyst** : 1. Conducting data cleaning and then analyzing the company's sales over the specified period, in this context from January 2019 to March 2019.
2. Creating customer segmentation into 8 segments by grouping based on RFM metrics.
3. Performing predictive modeling using machine learning with BetaGeoFitter and GammaGammaFitter.
- Tools** : Python and Tableau

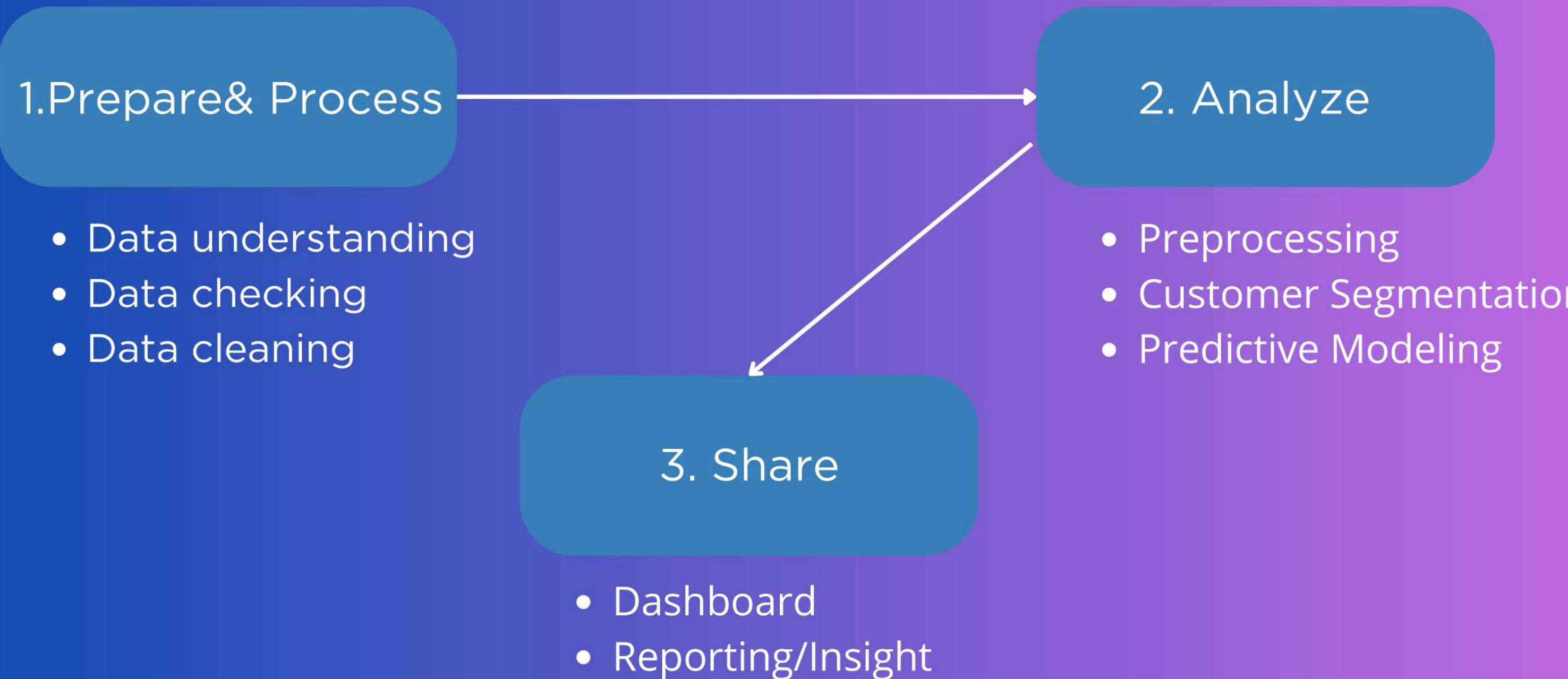
Business Problem

It was found that we have a dataset of transaction records from All U Need Mart, spread across several cities in Myanmar, occurring **between January 2019 and March 2019**. All U Need Mart sells various product lines, including Sports and Travel, Fashion Accessories, Food and Beverages, Electronic Accessories, Home and Lifestyle, and Health and Beauty.

From this dataset, we are expected to

- create **insights** to effectively increase the company's revenue.
- create **customer segmentation** to better categorize customers, which will help us find the most suitable approach for each segment.
- Build a predictive model using machine learning to project the next three months, estimating the **frequency** or **monetary** value each customer would spend at our company.

Stages of work



So this is the dataset. It contain

invoice_id	The transaction invoice ID
customer_id	The customer ID
branch	The branch where the transaction took place (e.g., A, B, C)
city	The city where the branch is located (e.g., Yangon, Mandalay, Naypyitaw)
customer_type	The type of customer (Normal or Member).
gender	The customer's gender (Male or Female).
product_line	The category of the product purchased (e.g., Sports and Travel, Home and Lifestyle, Electronic Accessories).
unit_price	The unit price of the product (in decimal format, e.g., 72.61).
quantity	The quantity of the product purchased.
tax	The tax from the transaction.
total	The total cost of the transaction, including tax.
date	The date of the transaction (format: dd/mm/yyyy).
time	The time of the transaction (format: HH).
payment	The payment method (e.g., Credit card, Cash).
cogs	Cost of Goods Sold (the cost of goods sold).
gross_margin_ratio	The gross margin ratio.
gross_income	The gross income from the transaction.
rating	The rating of the transaction (in decimal format, e.g., 6.9).

Data understanding & Data Checking

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1826 entries, 0 to 1825
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   invoice_id      1826 non-null    object  
 1   customer_id     1826 non-null    int64  
 2   branch          1826 non-null    object  
 3   city             1826 non-null    object  
 4   customer_type   1826 non-null    object  
 5   gender           1826 non-null    object  
 6   product_line    1826 non-null    object  
 7   unit_price      1826 non-null    float64 
 8   quantity         1826 non-null    int64  
 9   tax              1826 non-null    float64 
 10  total            1826 non-null    float64 
 11  date             1826 non-null    datetime64[ns]
 12  time             1826 non-null    object  
 13  payment          1826 non-null    object  
 14  cogs             1826 non-null    float64 
 15  gross_margin_ratio 1826 non-null    float64 
 16  gross_income    1826 non-null    float64 
 17  rating           1826 non-null    float64 
dtypes: datetime64[ns](1), float64(7), int64(2), object(8)
memory usage: 256.9+ KB
```

```
df.unique()
```

	Count
invoice_id	1000
customer_id	240
branch	3
city	3
customer_type	2
gender	2
product_line	6
unit_price	943
quantity	10
tax	990
total	990
date	89
time	506
payment	3
cogs	990
gross_margin_ratio	1
gross_income	990
rating	61
dtype: int64	

The dataset consists of 1,826 rows with 18 columns. Several unique values that need to more attention with the following details:

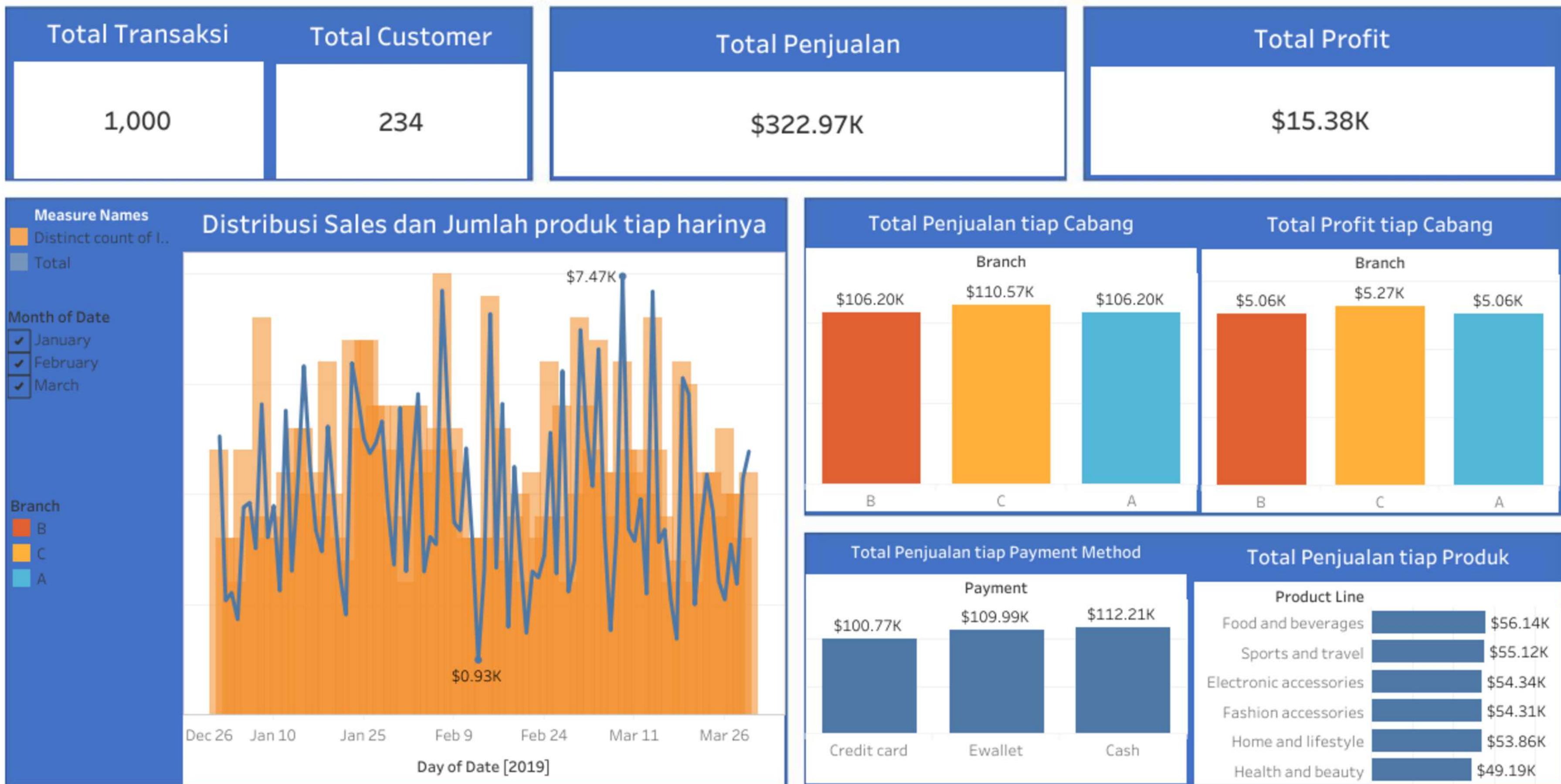
- Invoice_id : 1,000
- branch : 3
- city : 3
- product_line : 6
- payment : 3

We conducted a check for unique or non-duplicated values, as shown here. From this, we noticed an anomaly: for each invoice_id or transaction data, there should be only one unique invoice_id per transaction. However, out of the total 1,826 rows, there are only 1,000 unique invoice_id. I then conducted further investigation to find out the reason for this discrepancy.

Sure enough, there are a total of 826 rows with duplicate invoice_ids. From here, we need to perform cleansing the data.

Now, the total data ready for use contains 1,000 rows and 18 columns. After this data cleansing, we proceed to the next steps...

Report Dashboard



Preprocessing for RFM

In this Preprocessing, we used RFM Metrics. the objectives we need to achieve are identifying customer segmentation and calculating customer lifetime value.

So first of it, we need to identifying rfm metrics. RFM metrics is a method used to analyze and segment customers based on three key factors:

1. Recency (R): How **recently** a customer made a purchase. The more recent the transaction, the higher the recency score.
2. Frequency (F): How **often** a customer makes a purchase within a certain period. Customers who make purchases more frequently have a higher frequency score.
3. Monetary (M): The **total amount of money** a customer has spent. Customers who spend more money receive a higher monetary score.

from the dataset, column that we used for preprocessing RFM are

- **invoice_id**
- **customer_id**
- **date**
- **total**

```
Monetary Quartiles:  
0.25    1530.787125  
0.50    2327.734500  
0.75    3089.837625  
Name: monetary, dtype: float64  
Frequency Quartiles:  
0.25    6.0  
0.50    7.0  
0.75    9.0  
Name: frequency, dtype: float64  
Recency Quartiles:  
0.75    17.0  
0.50    9.0  
0.25    4.0  
Name: recency, dtype: float64
```

So, here I used the quantile method to identify the

- **Recency**,
- **Frequency**, and
- **Monetary** metrics

dividing them into four categories with thresholds as shown here.

For example, for Frequency, values of 9 and above are assigned a score of 1, while values of 5 and below are assigned a score of 4.

RFM Metrics	Segmentation
'111'	Best Customer
'[1-4][1-2][1-4]'	Loyal Customer
'[1-4][1-4]1'	Big Spenders
'444'	Lost Customer
'[3-4][1-4][1-3]'	Retarget Customer
'[1-4]4[3-4]'	One-time Customer
'14[1-4]'	New Customer
'[1-4][3-4][2-4]'	Occational Customer

So here, I created naming conventions for each customer segment based on the descriptions of the quantile thresholds we discussed earlier. I derived these names from a case study on Kaggle.com involving the MegaStores Sales Dataset.

Report

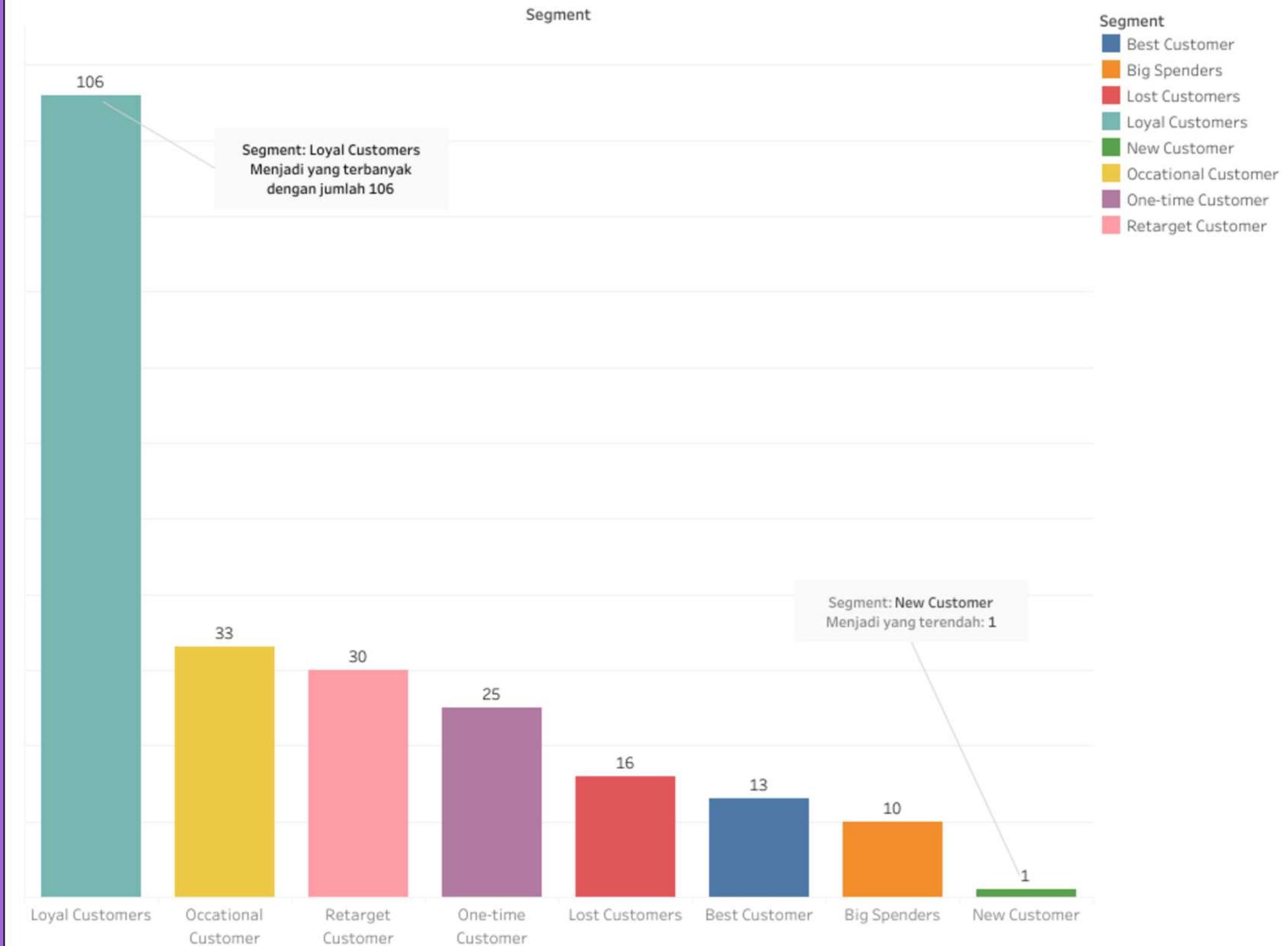
Customer Segmentation

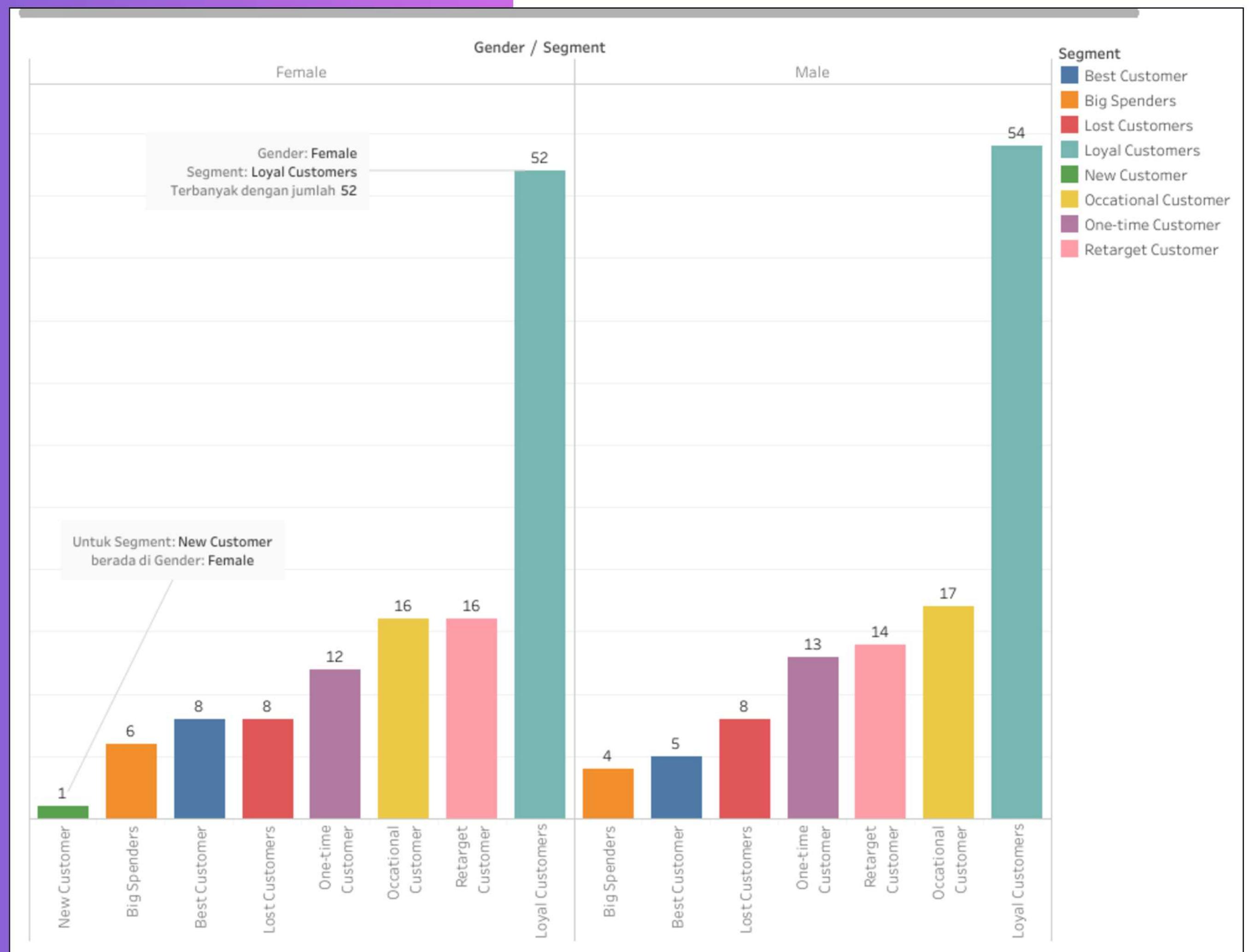
Berikut untuk Distribusi Customer, setelah dilakukan pengelompokan menggunakan RFM Metrics. Untuk Segment terbanyak ada di Loyal Customer

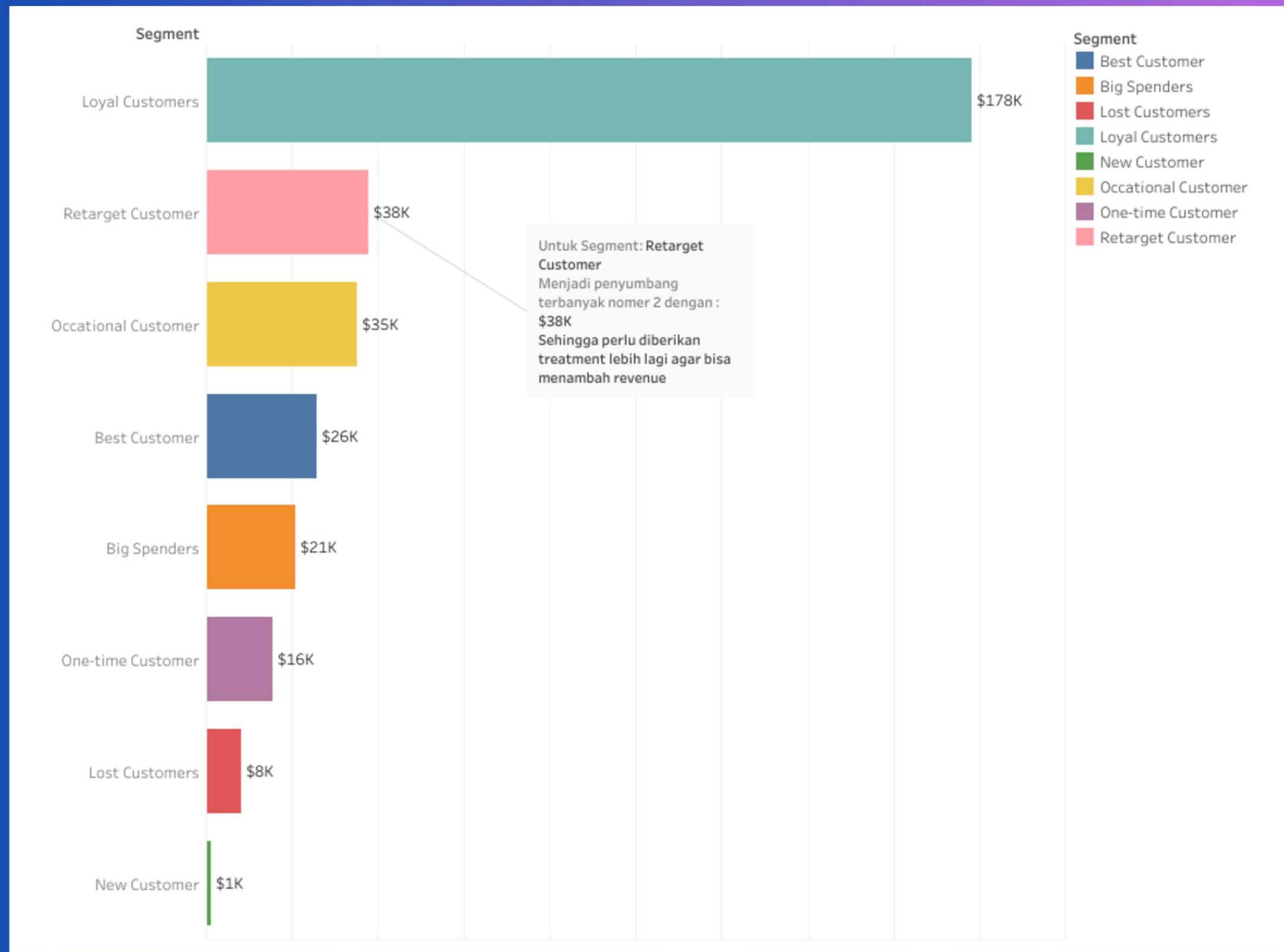
Kemudian mengidentifikasi menggunakan variabel gender. Didapatkanlah distribusi untuk tiap segment customer berdasarkan gendernya. Untuk gender female lebih banyak dengan total 119 dan male 115.

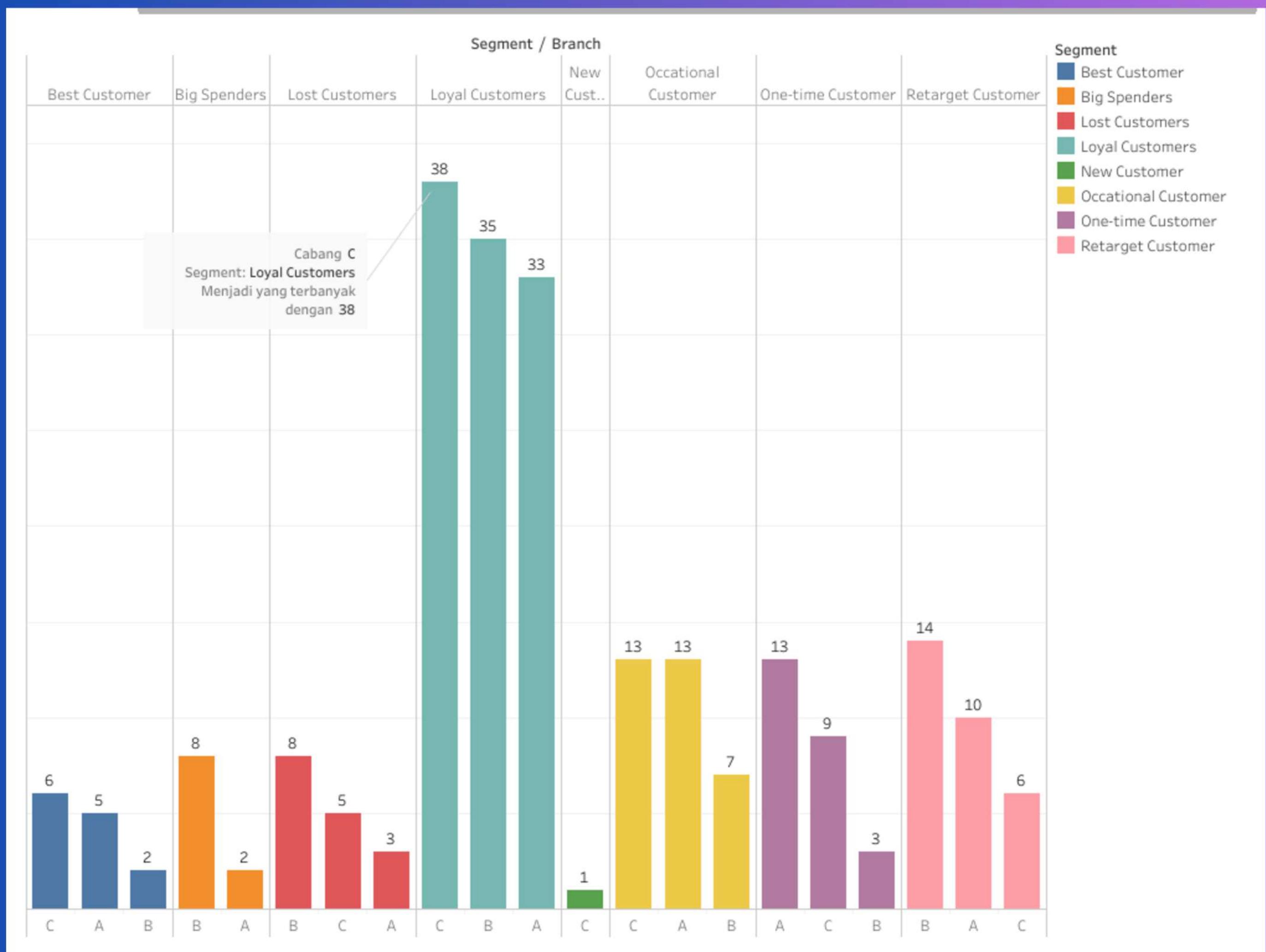
Melakukan identifikasi untuk mencari total penjualan terbanyak. Berikut untuk hasilnya didapatkan Loyal Customer sebagai penyumbang penjualan terbanyak dengan \$178K.

Kemudian menambahkan variabel untuk di tiap cabang, didapatkan persebarannya sebagai berikut. Dari hasil persebaran tersebut, seg..



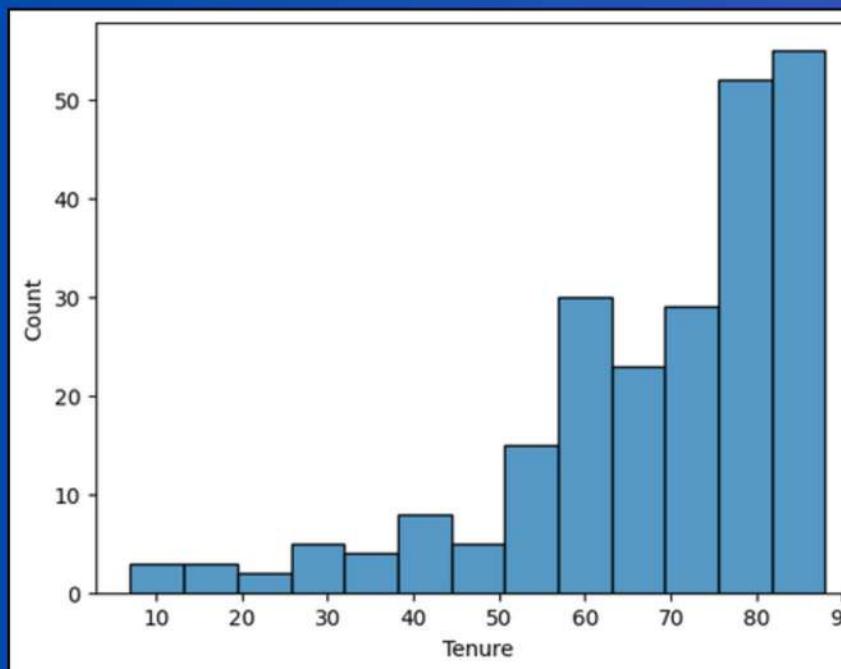




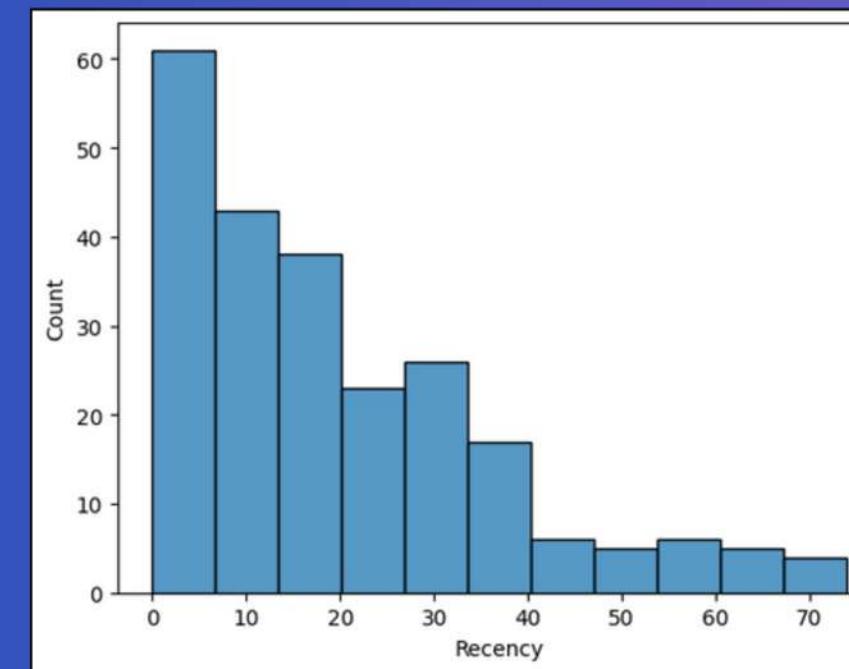


Preprocessing for Customer Lifetime Value

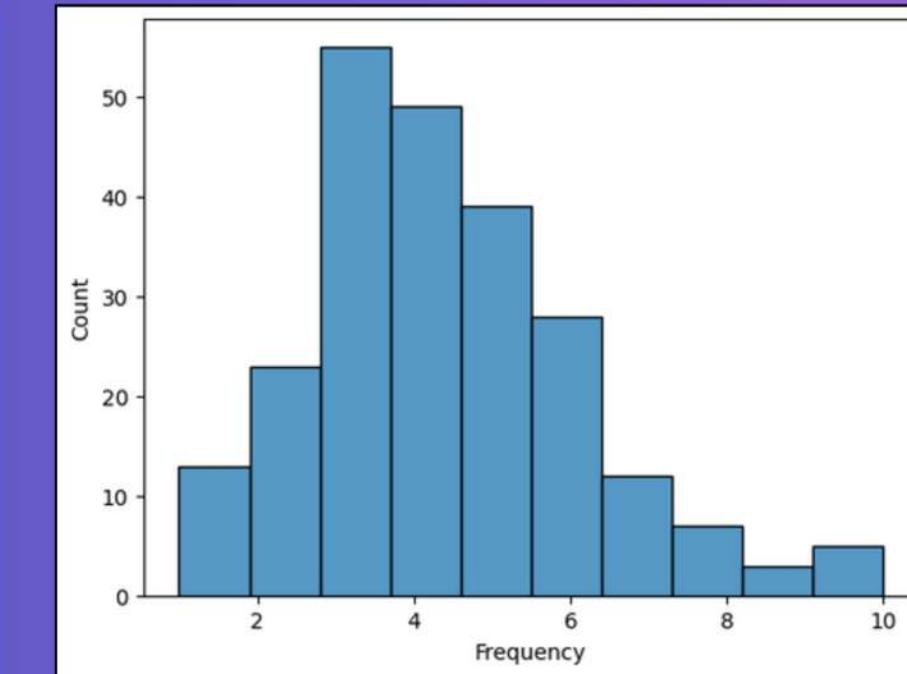
Tenure



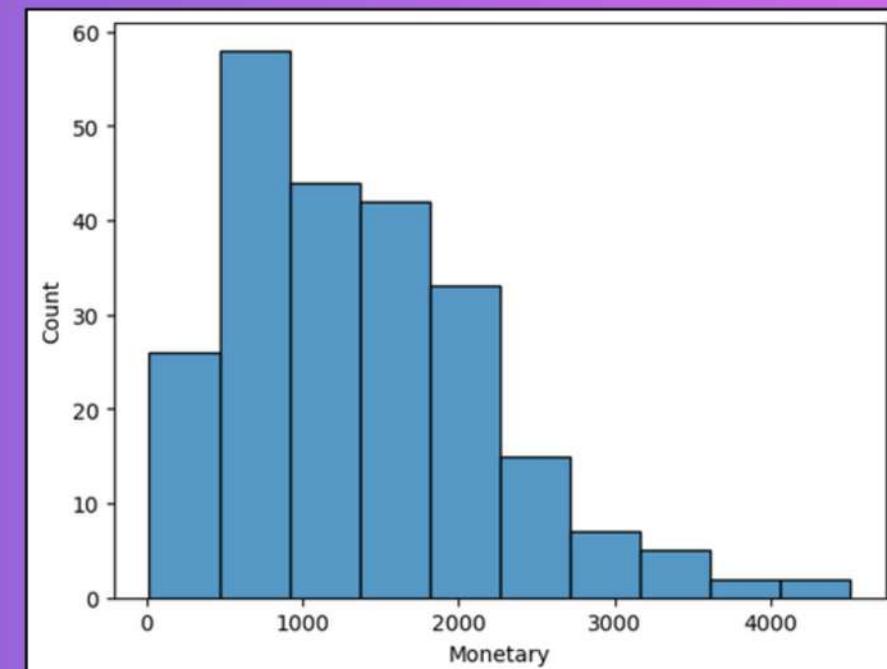
Recency



Frequency



Monetary



"This tenure is the **first transaction date**. Therefore, to find this tenure, subtract the last recorded transaction in the data from the first transaction of each customer."

Recency is the day when the customer made a transaction calculated from the **last recorded date**, in this context, it is calculated by subtracting from March 30, 2019.

Frequency is **how often** the customer makes transactions.

From here, we will proceed to calculate using machine learning.

Monetary is the **total amount spent** by each customer.

	customer_id	Tenure	Recency	Frequency	Monetary	CLTV_BG	CLTV_Gamma
0	1	68.0	21.0	25	972.9195	4.105305e-02	41.258602
1	2	83.0	36.0	9	418.1625	5.475896e-01	245.146759
2	3	72.0	39.0	25	916.3665	7.059486e-01	668.211628
3	4	78.0	8.0	16	1853.6280	7.441169e-04	1.442312
4	5	78.0	60.0	4	880.7925	1.450534e+00	1429.325508
...
229	236	72.0	35.0	16	1911.7350	6.202445e-01	1240.018003
230	237	88.0	25.0	16	2035.9815	5.841037e-02	124.329742
231	238	81.0	12.0	25	1975.7955	6.982429e-04	1.424240
232	239	81.0	62.0	4	1199.6775	1.404421e+00	1883.932533
233	240	82.0	11.0	81	1879.7205	3.010731e-07	0.000571

234 rows × 7 columns



Customer Lifetime Value here I used **BetaGeoFitter** to find the frequency and **GammaGammaFitter** to find the monetary value by projecting **3 months** ahead. The result is over there.

Below that, I also included the minimum and maximum thresholds to provide a clearer picture. For example, let's consider **customer_id 2**, which has a **BG score** of **0.5475**, placing it in the **75%** threshold. This indicates that the customer is likely to shop at the store again and is expected to spend around **\$245**, as seen in the **50%** threshold in the gamma column.

	customer_id	Tenure	Recency	Frequency	Monetary	CLTV_BG	CLTV_Gamma
count	234.000000	234.000000	234.000000	234.000000	234.000000	2.340000e+02	2.340000e+02
mean	120.683761	68.482906	19.837607	21.991453	1380.199782	4.519982e-01	3.662557e+02
std	69.459303	17.468357	17.067127	19.875361	852.800213	7.829888e-01	6.433120e+02
min	1.000000	7.000000	0.000000	1.000000	16.275000	4.081556e-15	5.983867e-12
25%	60.250000	61.000000	6.000000	9.000000	726.894000	5.958070e-05	1.355070e-01
50%	121.500000	74.000000	16.500000	16.000000	1274.395500	4.172366e-02	4.742106e+01
75%	180.750000	81.000000	28.000000	25.000000	1878.835875	6.578523e-01	5.097885e+02
max	240.000000	88.000000	74.000000	100.000000	4510.002000	4.181491e+00	4.317704e+03



Summary



- All u need mart has 3 branches spread across Myanmar with total sales of \$322.9K and a total profit of \$15.38K. The branch with the highest sales is branch C, and the highest product sales are in Food and Beverage.
- Divided into 8 segments, the most numerous segment is 'Loyal Customer' with 106 customers out of a total of 234, and the least numerous is 'New Customer' with 1.
- Looking at the visual segments, those with a positive connotation total 130, while those with a negative connotation total 104.

1

2

3

4

5

6

7

8

9



Conclusion



- There are a total of 104 customers or 45% of customers who need to be identified for the problem that makes them minimally contribute to this company.
- It could be due to several variables, so need to find out the specific problem.
- Can provide appropriate treatment to attract both existing and new customers to generating profit.

1

2

3

4

5

6

7

8

9

1

2

3

4

5

6

7

8

9

Link Google Collab



or [Click here](#)

For your appreciation to my post, i would say



Thank You



Reach me



rifkimad10@gmail.com



+62 859 4722 9254



linkedin.com/in/rifkimuharbian99/