FINAL PROJECT DATA ANALYST

▶ # CUSTOMER BEHAVIOR AND SEGMENTATION ANALYSIS FOR IPHONE TRANSACTIONS IN INDONESIA

ROY FIRMAN SIHOMBING

TABLE OF CONTENTS

Roy Firman Sihombing

I am a Mathematics graduate with a strong interest in data science. Currently, I am enrolled in the Data Science Bootcamp at Dibimbing.id to sharpen my skills in Python, SQL, data analysis, and visualization. I have hands-on experience in warehouse data management, sales tracking, and spreadsheet automation. Known for strong analytical thinking, problem-solving skills, and clear communication, I thrive both in team collaborations and independent work.

Based in Tangerang, Indonesia.

✉️ You can contact me at royfirmans2d3j@gmail.com

📬LinkedIn: linkedin.com/in/roy-firman-sihombing

💻 Portfolio Website: roy-firman-sihombing.free.nf

EDUCATION

✓ Dibimbing.id

May 2025 - Present

Data Science Bootcamp
- Learning Python basics, data structures, statistics, and hypothesis testing.
- Data cleaning, analysis, and visualization using Pandas, Matplotlib, Seaborn.
- Dashboard creation using Power BI and Tableau.

✓ Sriwijaya University

2020 - 2024

B.Sc. in Mathematics (GPA 3.82/4.00)
- Relevant coursework: SQL, PHP, HTML (Database), Python (AI & Neural Networks), C++, Excel (Cryptography & Programming).
- Final Project: Inventory Model for Deteriorating Pharmaceutical Items with Logarithmic Demand Rate.
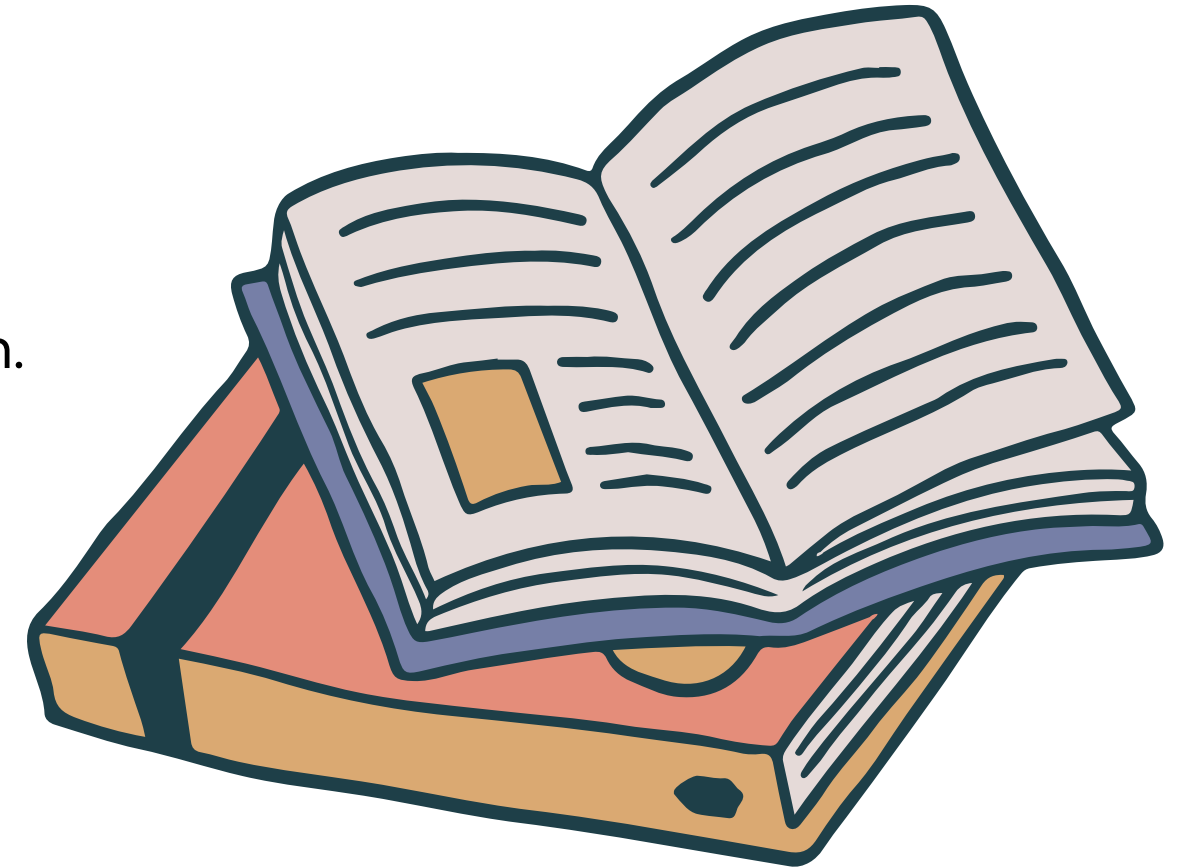
WORKING EXPERIENCE

PT Sumber Alam Putra Lestari

Sept 2024 - Present

Administration Staff
- Digitized stock records from manual logs to spreadsheets for better accuracy and accessibility.
- Managed sales requests, delivery notes, and stock transfers.
- Improved stock tracking efficiency with structured and regularly updated data.

CoastM Education

Aug 2021 - Aug 2023

Chief Executive Officer
- Led business strategy, managed operations, and developed partnerships.
- Secured a total of 26 million in business funding from three rounds.

PROJECT OVERVIEW

🏪 Market Channel Analysis

Using Superstore data (2014–2017), three dashboards were created: Geographic Profit, Sales, and Customer Segmentation.

📍 NYC leads in profit ($70K)

💲 $2.3M in sales dominated by Office Supplies & Standard Class

👥 Most customers fall into the Need Attention/At Risk segment

Previous Project Data Analyst at Dibimbing.id

👤 Customer Segmentation Analysis

Using Superstore data (2014–2017), Customer Segmentation Analysis shows:

📍 New York City leads in customers (355), followed by Los Angeles (304)

💲 $2.30M in sales and $286.4K profit with a 12.47% average margin

👥 Largest customer group: At Risk (38% sales, 36% profit)

MAIN PROJECT

Dataset From Kaggle bit.ly/DatasetProjectDA

Project Background

This project analyzes iPhone sales transaction data in the Indonesian market from 2022 to 2024.
The analysis uses Python (data cleaning, feature engineering, exploratory data analysis) and Power BI (interactive dashboards) to provide a comprehensive overview of customer behavior, sales trends, and optimal pricing strategies.

Project Objectives
- Identify high-value customers.
- Understand purchasing trends based on time, location, and demographics.
- Optimize promotional strategies and customer retention efforts.

Beneficiaries

📌 Marketing Management Team

📌 Sales Team

📌 Strategic Management

Business Problem

⚠ The company faces several challenges

❌ No clear mapping of customer segments and purchasing behavior.

❌ Difficulty in identifying customers at risk of churning or already lost.

❌ Lack of insights into the impact of discounts on sales and profit margins.

❌ Price differences between cities have not been leveraged for pricing strategy.

✅ Impact of addressing these issue

⬆ Improved retention of high-value customers.

⬆ More efficient promotions based on customer segment behavior.

⬆ Optimized pricing strategies across different sales locations.

Data Understanding

Data Sources:
- transactions.csv – Main transaction data.
- transaction_details.csv – Product details for each transaction.
- products.csv – Product specifications (name, color, storage, price).
- customers.csv – Customer data (name, age, city).

Data Structure & Size:
- Total: ~200,000 transactions initially, reduced to ~125,000 after cleaning, and ~37,000 unique customers.
- Period: 2022–2024.
- Key columns: transaction date, customer ID, product ID, quantity, price, discount, total, city, age group

Sample Dataset

transaction_id	transaction_date	customer_id	name	age	city	product_id	product_name	storage	color
2	71024	2024-01-28	36119	Kamila Irawan	45	Jakarta	1	iPhone 15	128GB	Black
3	45980	2024-05-13	21642	Wirda Kusumo	23	Makassar	6	iPhone 14 Plus	256GB	Green
4	35845	2024-09-19	28348	Tgk. Yuni Suryono	19	Yogyakarta	4	iPhone 15 Pro Max	1TB	Gold
6	49723	2024-04-15	36190	Digdaya Gunarto	27	Bandung	8	iPhone 14 Pro Max	128GB	Yellow
8	55065	2023-01-04	36596	Tgk. Zulaikha Setiawan	56	Bandung	3	iPhone 15 Pro	512GB	Silver

shipping_method	payment_method	quantity	unit_price	price	coupon_code	discount	total	selisih_harga	keterangan_harga	usia_group
POS Indonesia	GoPay	3	11000000	12000000	DISCOUNT10	15	28050000.0	-1000000	Normal Di Bawah	36-45
GoSend	Cash on Delivery	4	12000000	12000000	DISCOUNT10	15	40800000.0	0	Normal	18-25
POS Indonesia	GoPay	3	17000000	18000000	NONKUPON	15	43350000.0	-1000000	Normal Di Bawah	18-25
JNE	Cash on Delivery	4	13000000	17000000	GRATISONGKIR	5	49400000.0	-4000000	Normal Di Bawah	26-35
POS Indonesia	OVO	4	15000000	16000000	SALE20	10	54000000.0	-1000000	Normal Di Bawah	56-65

DATA PREPARATION

01   **Merge Data**

transactions.csv
transaction_details.csv   →   iphone_transaction.csv
products.csv
customers.csv

03   Feature Engineering

- Total Price After Discount
- Age Group   **Add New Column**
- Price Category
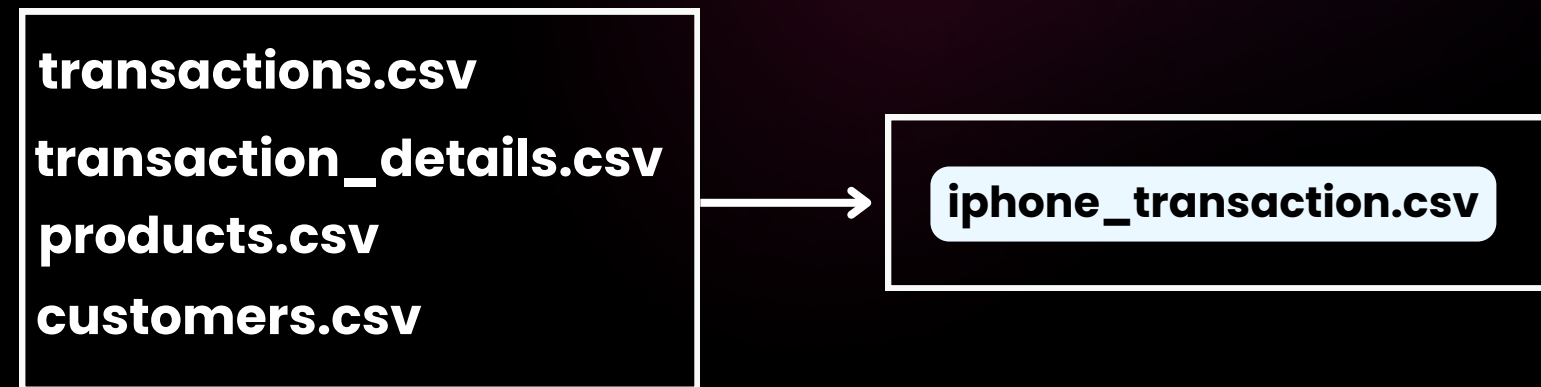
02   **Data Cleaning**

- Handling Outliers (Abnormal Price)
- Handling Missing Values
- Remove Duplicates

Data Analysis and Insights

Number of Transactions Each Month

Last Transaction Date
12/23/2024

January February March April May June July August Septe... October Nove... Decem...

*Data Bulan Desember 2024 Belum Lengkap

During the period 2022 - 2024 the largest increase in transactions occurred in January 2022 increasing by 1,961

Sales	**3.79 Triliun**
Quantity	**313,089**
Total Order	**71,169**
Avg Sales Order	**53.22 Juta**
Products	**12**

The sales trend fluctuates over time. The largest increase was in January 2022, with 1,961 more transactions than the previous month. The dip in February is normal due to fewer days, while December 2024 shows a drop because data only covers up to December 23, 2024.

Sales Top 5 Product

Product	Revenue
iPhone 14 Pro	0.45T
iPhone 13 Pro...	0.44T
iPhone 15 Pro	0.41T
iPhone 15 Plus	0.38T
iPhone 13 Pro	0.36T

Sales Volume by Storage Capacity

512GB 128GB 256GB 1TB 64GB
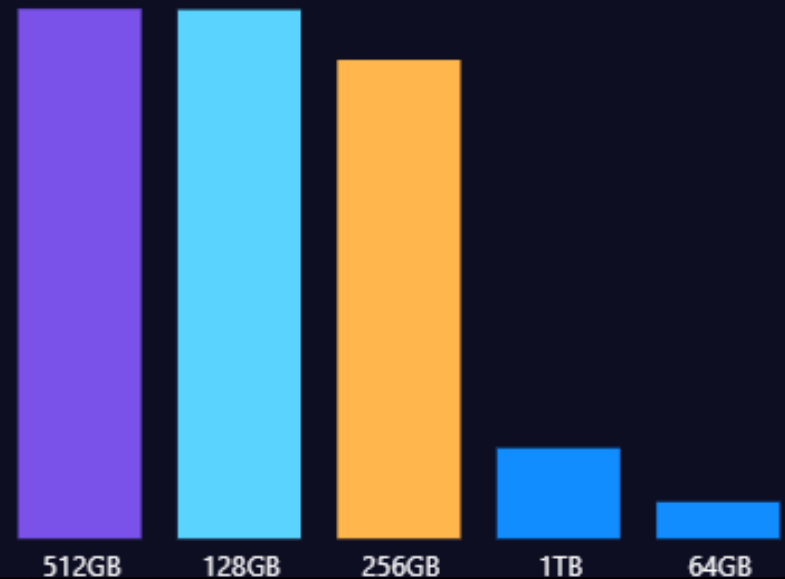
The iPhone 14 Pro leads in revenue, followed by the iPhone 13 Pro and iPhone 15 Pro. For storage, 512GB is most popular for long-term use, 128GB is common as today's minimum standard, and 256GB is a balanced mid-tier choice.

Product Price Levels in Top 5 Cities

● Normal ● Normal Di Atas ● Normal Di Bawah

	Jakarta	Bandung	Yogyakarta	Medan	Denpasar
Normal Di Bawah	39.54%	39.54%	39.45%	39.54%	39.24%
Normal Di Atas	35.76%	35.91%	35.49%	36.06%	35.90%
Normal	24.69%	24.55%	25.06%	24.40%	24.86%
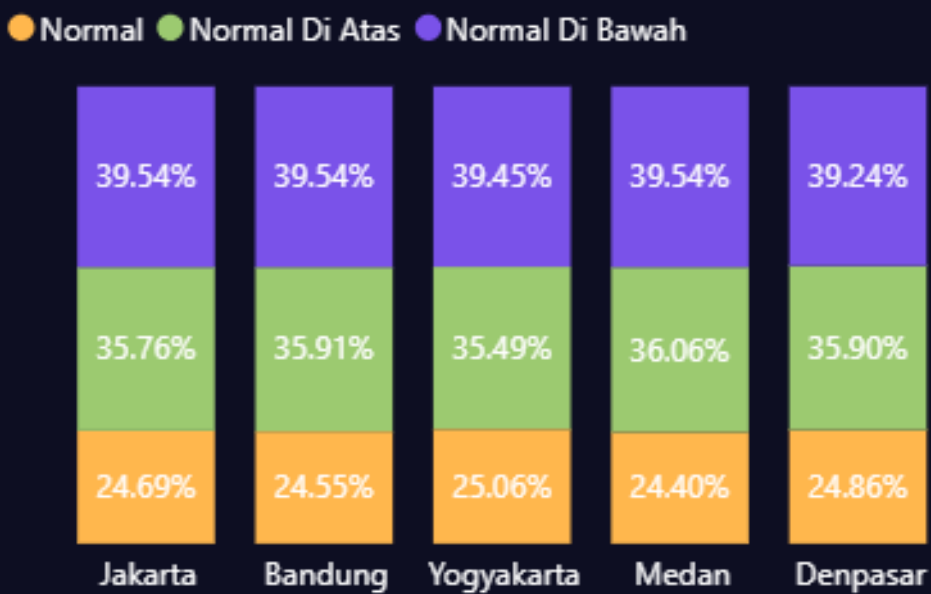
Most purchases occur when the product price is below the standard level, but many customers still buy even when prices are slightly above normal. This suggests demand is relatively strong, with buyers valuing the product enough to purchase outside discount periods.

Data Analysis and Insights

Customer Segmentation

Sales by City

- Bandung ● Denpasar ● Jakarta ● Makassar ● Medan ● Surabaya ● Yogyakarta

Medan
MALAYSIA
SINGAPORE
INDONESIA
Makassar
Jakarta
Yogyakarta
Surabaya
Bandung
Denpasar

Microsoft Bing
© 2025 Microsoft Corporation Terms

Top 5 City by Total Customers

32,399
Total Customer

- ● Jakarta
- ● Bandung
- ● Yogyakarta
- ● Medan
- ● Surabaya

Jakarta
Top City by Sales

Most customers are in Jakarta, which also records the highest sales. This shows a strong market concentration in the capital, while Bandung and Yogyakarta follow as secondary markets.

Customer Segment	Segment Description	Total Customers	Total Sales
Champions	Membeli baru-baru ini, sering membeli, dan membelanjakan paling banyak.	5K	0.89T
Loyal Customers	Sering membeli dan membelanjakan cukup besar, responsif terhadap promosi.	5K	0.54T
Potential Loyalist	Pelanggan baru yang sudah membeli lebih dari satu kali dan jumlah pembeliannya cukup tinggi.	8K	0.99T
At Risk	Dulu pernah aktif berbelanja dan menghabiskan banyak, tapi sudah lama tidak membeli lagi.	12K	1.02T
Lost	Skor recency, frequency, dan monetary sangat rendah. Sudah lama tidak aktif.	7K	0.36T

Champions are the biggest spenders, Loyal Customers buy often and respond to promos, Potential Loyalists have strong buying history but need nurturing, At Risk customers have reduced activity, and Lost customers are inactive.

Data Analysis and Insights

Customer Segmentation

Customer Segment Age Distribution

● 26-35 ● 36-45 ● 46-55 ● 56-65

Segment	26-35	36-45	46-55	56-65
Champions	38.61%	20.68%	21.10%	19.62%
Loyal Custo...	41.65%	20.08%	19.32%	18.95%
Potential Loy...	39.34%	20.26%	19.55%	20.85%
At Risk	39.43%	20.57%	20.33%	19.67%
Lost	41.95%	17.79%	18.18%	22.08%
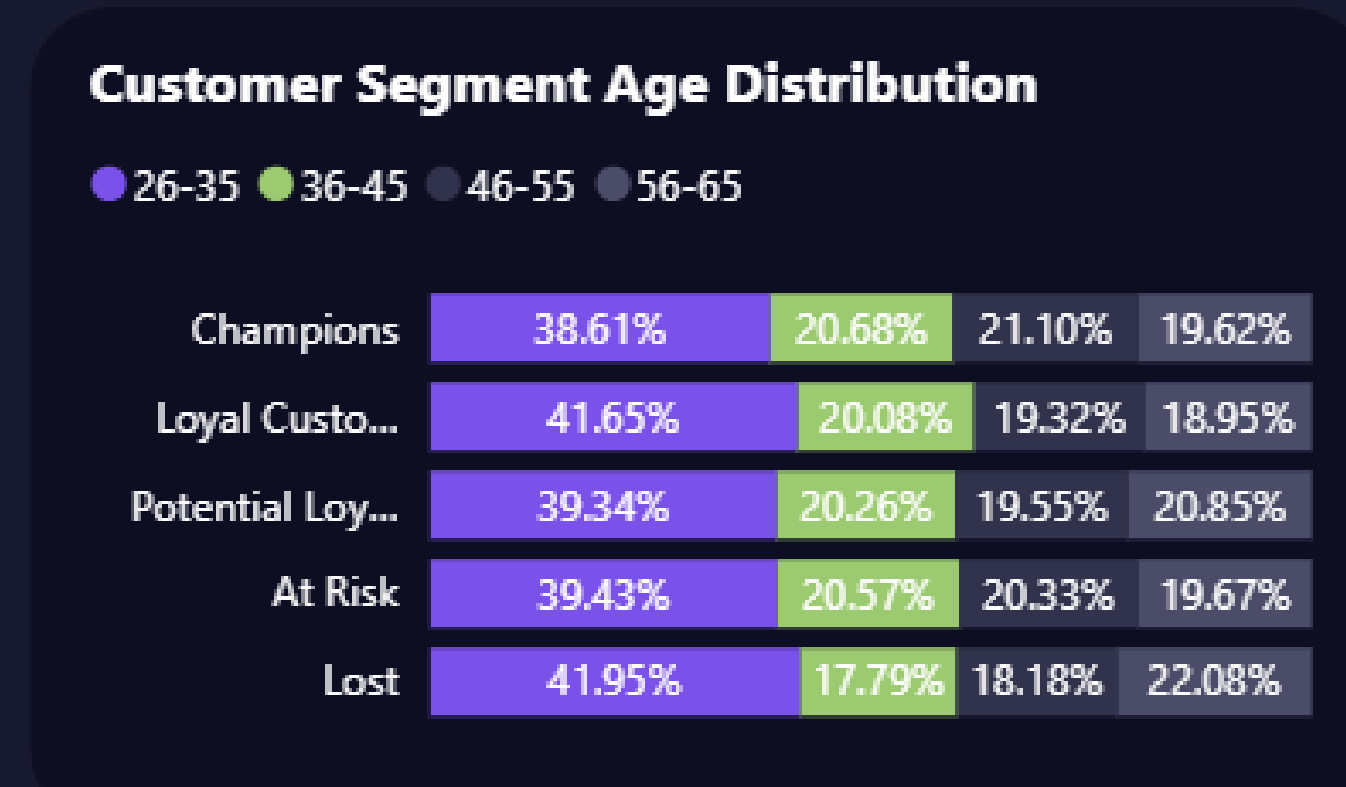
Across all segments, the 26–35 age group dominates, followed by 36–45. This shows that the main customer base is in the young-to-middle adult range, with relatively consistent proportions across segments.

Total Sales by Discount Usage

● Use Discount
● No Discount

Distribution of Discount Usage

● No Discount ● Use Discount

	Champions	Potential Loyalist	Loyal Customers	At Risk	Lost
Use Discount	54.48%	53.20%	62.56%	50.78%	54.92%
No Discount	45.52%	46.80%	37.44%	49.22%	45.08%

All segments consistently make more purchases using discounts, with Loyal Customers having the highest proportion of discount usage. Meanwhile, the At Risk segment records the lowest discount usage among all groups. This shows that discounts are a strong driver across the board, but their impact varies between segments.

Dashboard Power BI

Sales Performance & Product Insights

Iphone Transactions (Indonesia Market) 2022 - 2024

Year	Customer Segment
All	All

Number of Transactions Each Month

Last Transaction Date 12/23/2024

January February March April May June July August Septe... October Nove... Decem...

*December 2024 Data is Not Yet Complete

During the period 2022 - 2024 the largest increase in transactions occurred in January 2022 increasing by 1,961

Sales	**3.79 Triliun**
Quantity	**313,089**
Total Order	**71,169**
Avg Sales Order	**53.22 Juta**
Products	**12**

Monthly Sales and Growth Trends

MonthYear	Total Sales	Sales Last Month	Sales Growth	
Jan 2022	104.37 Miliar			
Feb 2022	98.56 Miliar	104.4 Miliar	▼	-0.06
Mar 2022	111.67 Miliar	98.6 Miliar	▲	0.13
Apr 2022	102.59 Miliar	111.7 Miliar	▼	-0.08
May 2022	107.11 Miliar	102.6 Miliar	▲	0.04
Jun 2022	108.45 Miliar	107.1 Miliar	▲	0.01
Jul 2022	107.28 Miliar	108.4 Miliar	▼	-0.01
Aug 2022	109.88 Miliar	107.3 Miliar	▲	0.02
Sep 2022	108.87 Miliar	109.9 Miliar	▼	-0.01
Oct 2022	107.21 Miliar	108.9 Miliar	▼	-0.02
Nov 2022	107.39 Miliar	107.2 Miliar	▲	0.00
Dec 2022	107.11 Miliar	107.4 Miliar	▼	0.00

Sales Top 5 Product

Product	Sales
iPhone 14 Pro	0.45T
iPhone 13 Pro...	0.44T
iPhone 15 Pro	0.41T
iPhone 15 Plus	0.38T
iPhone 13 Pro	0.36T

Product Price Levels in Top 5 Cities

● Normal ● Normal Di Atas ● Normal Di Bawah

	Jakarta	Bandung	Yogyakarta	Medan	Denpasar
Normal Di Bawah	39.54%	39.54%	39.45%	39.54%	39.24%
Normal Di Atas	35.76%	35.91%	35.49%	36.06%	35.90%
Normal	24.69%	24.55%	25.06%	24.40%	24.86%

Sales Volume by Storage Capacity

512GB 128GB 256GB 1TB 64GB

Dashboard Power BI

Customer Segmentation
Iphone Transactions (Indonesia Market) 2022 - 2024

Product Name: All
Year: All
Year: All

Sales by City
● Bandung ● Denpasar ● Jakarta ● Makassar ● Medan ● Surabaya ● Yogyakarta

Top 5 City by Total Customers

32,399 Total Customer

● Jakarta
● Bandung
● Yogyakarta
● Medan
● Surabaya

Customer Segment Age Distribution
● 26-35 ● 36-45 ● 46-55 ● 56-65

	26-35	36-45	46-55	56-65
Champions	38.61%	20.68%	21.10%	19.62%
Loyal Custo...	41.65%	20.08%	19.32%	18.95%
Potential Loy...	39.34%	20.26%	19.55%	20.85%
At Risk	39.43%	20.57%	20.33%	19.67%
Lost	41.95%	17.79%	18.18%	22.08%

Total Sales by Discount Usage
● Use Discount
● No Discount

Jakarta
Top City by Sales

Customer Segment	Segment Description
Champions	Membeli baru-baru ini, sering membeli, dan membelanjakan paling banyak.
Loyal Customers	Sering membeli dan membelanjakan cukup besar, responsif terhadap promosi.
Potential Loyalist	Pelanggan baru yang sudah membeli lebih dari satu kali dan jumlah pembeliannya cukup tinggi.
At Risk	Dulu pernah aktif berbelanja dan menghabiskan banyak, tapi sudah lama tidak membeli lagi.
Lost	Skor recency, frequency, dan monetary sangat rendah. Sudah lama tidak aktif.

Total Customers / Total Sales

Segment	Total Customers	Total Sales
Champions	5K	0.89T
Loyal Customers	5K	0.54T
Potential Loyalist	8K	0.99T
At Risk	12K	1.02T
Lost	7K	0.36T

Distribution of Discount Usage
● No Discount ● Use Discount

	Champions	Potential Loyalist	Loyal Customers	At Risk	Lost
No Discount	54.48%	53.20%	62.56%	50.78%	54.92%
Use Discount	45.52%	46.80%	37.44%	49.22%	45.08%

Recommendations

01   Reactivate the "At Risk" segment aged 26–35 and 36–45 with promotions tailored to their product preferences.

02   Offer exclusive loyalty programs for Champions and Loyal Customers in Jakarta to maintain the highest sales contribution.

03   Target Potential Loyalists in Bandung and Yogyakarta with bundles of iPhone 14 Pro or 512GB and 256GB variants to boost purchase frequency.

04   Adjust discount schemes for each segment to effectively drive purchases while maintaining profit margins.

Let's Connect

I'm always open to networking, collaboration, or just a quick chat.

linkedin.com/in/roy-firman-sihombing

github.com/Roysihombing

roy-firman-sihombing.free.nf/