



Python Assignment

User Segmentation for New Mutual Fund Service Campaign

RevoU - FSDA Batch 5

Google Colab: [Click here](#)

01

**Business
Overview**

02

**Data
Cleaning**

03

**Exploratory
Data Analysis**

04

**User
Segmentation**

05

**Regression and
Benefit-Cost
Analysis**

06

Recommendation

01

Business Overview

You can enter a subtitle
here if you need it



Business Background



Indonesian investment startup that builds an app that can buy and sell mutual funds, which consist of stocks, bonds, money market, and mixed- investments.

Our startup wants to create thematic campaigns based on user preference. For this reason, user segmentation is needed to provide recommendations for each campaign that will be made.

Dataset: [Click here](#)

Dataset

Transaction and invested
amount data

Daily User Transaction

1

Users

User profile and summary of
their investment

2

3

Churn

User that churn or not

```
response = requests.get(url)

# checking response.status_code (if you get 502, try rerunning the code)
if response.status_code != 200:
    print(f"Status: {response.status_code} - Try rerunning the code!")
else:
    print(f"Status: {response.status_code}\n")

# using BeautifulSoup to parse the response object
soup = BeautifulSoup(response.content, "html.parser")

# finding Post images in the soup
images = soup.find_all("img", attrs={"alt": "Post Image"})

# downloading images
for i, image in enumerate(images):
    url = image.get("src")
    if url:
        response = requests.get(url)
        with open(f"image_{i}.jpg", "wb") as f:
            f.write(response.content)
```

02

Data Cleaning

Cleaning Process

Data Types

1

- Convert 'user_id' to string
- Convert date to datetime
- Convert numeric to float

Irrelevant/typo values

2

- No irrelevant or typo values

Missing Value

3

- In column 'referral_code_used', fill null with 'no referral'
- In column 'transaction_amount', for each buy/sell replace null with zero
- In column 'invested_amount', for each mutual funds replace null with zero

Duplicates

4

- There's no duplicates

Outliers

5

- Based on the boxplot made, there are indications of outliers. But because there is no limit of the amount of investment made by the user. So, we keep all the existing numeric variables.

Cleaning Process

03

Exploratory Data Analysis



Total User

8.277

Total User



Average Transactions Amount



Buy Amount

Rp2.001.688,76

Sell Amount

Rp450.616,89

**Total Invested
Amount**

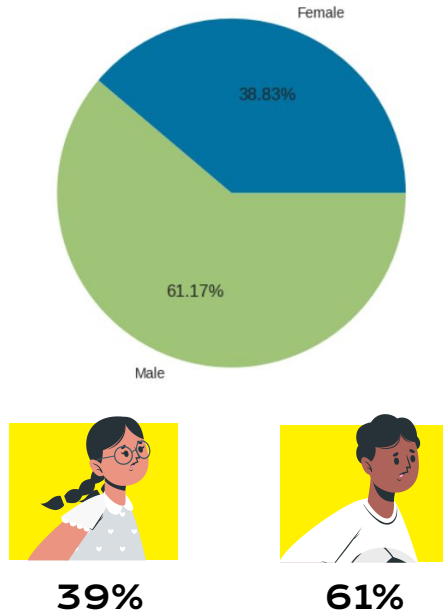
Rp2.442.450,32

**Monthly Invested
Amount**

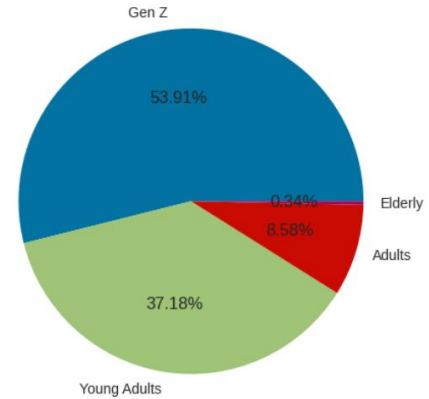
Rp3.007.509,97

User Demographic

Gender



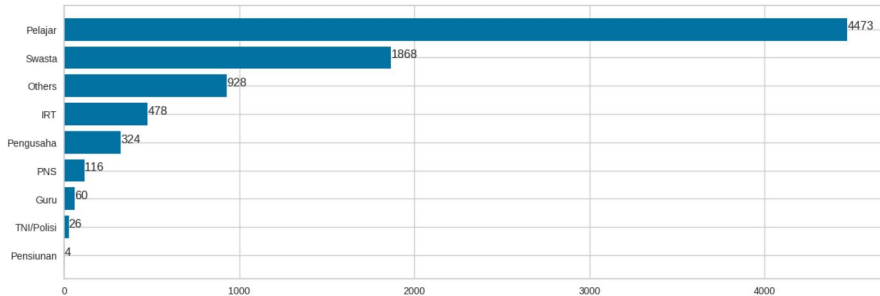
Age Group



4462 Gen-Z Users
(15 -25 Years Old)

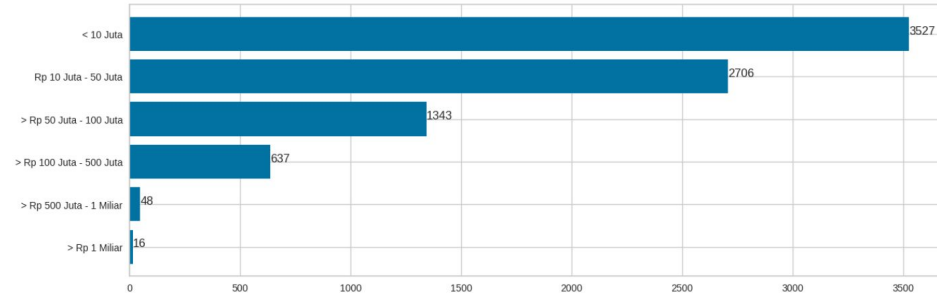
User Demographic

Occupation



The majority of users come from the student, with **55,96%**

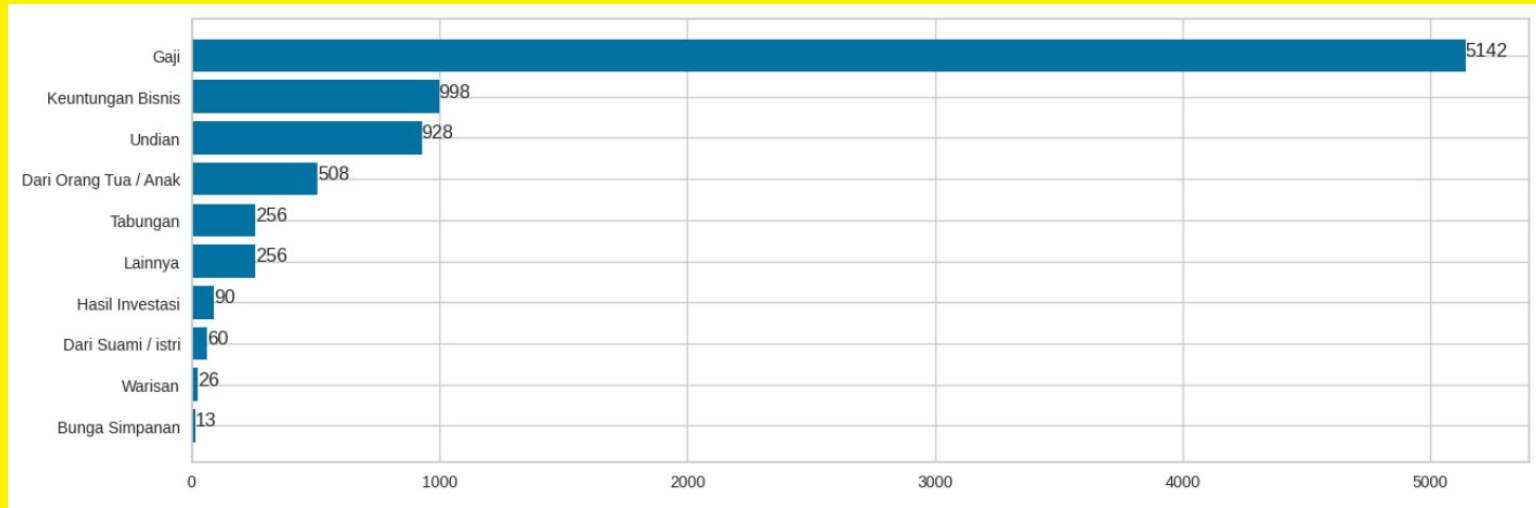
Income Range



The majority of user income is <IDR 10 million, with **43,13%**

User Demographic

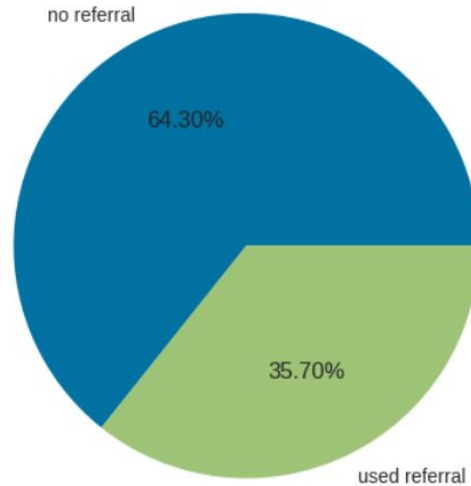
Income Source



The majority of users incomes comes from salary, with **62,72%**

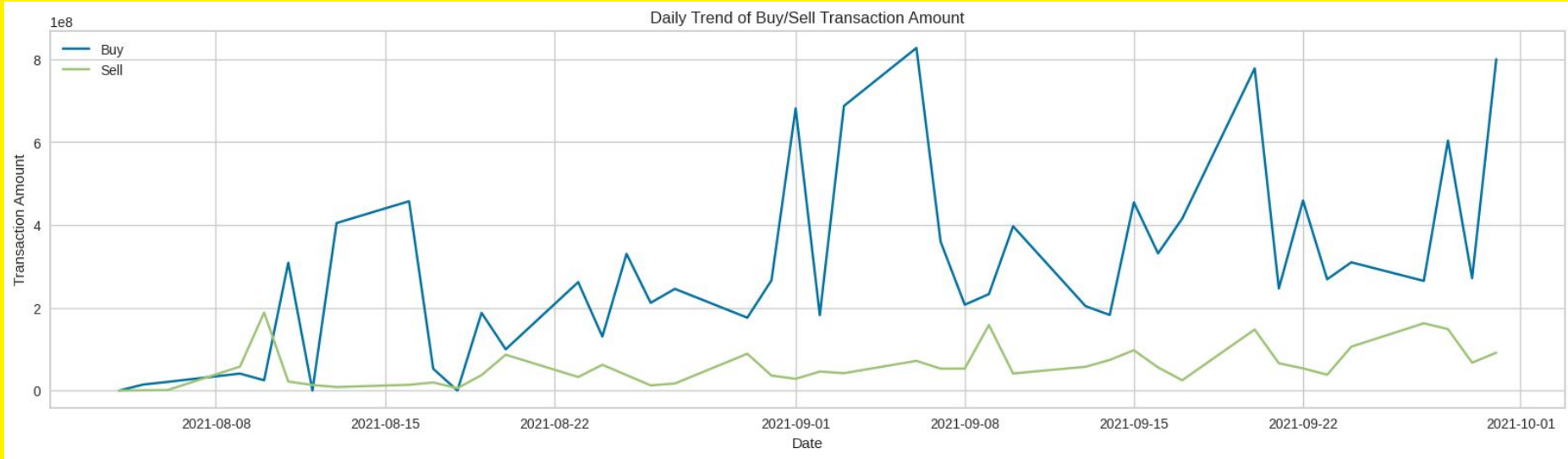
User Demographic

Referral Code Used



The majority of users did not use the referral code, with **64,30%**

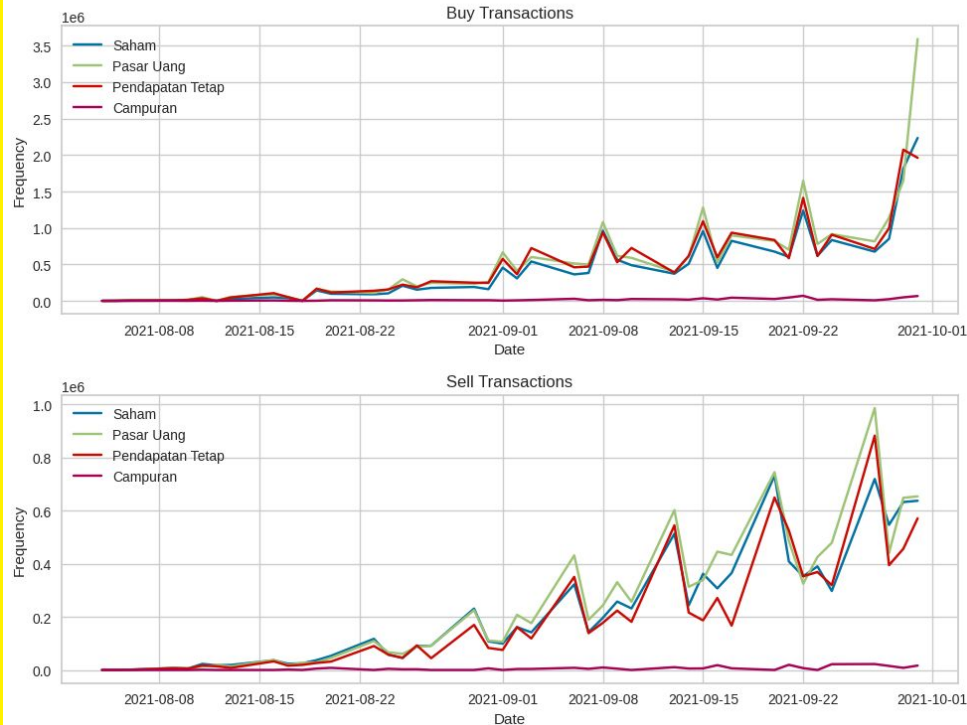
Daily Transaction Trend



There is a fluctuating trend for daily buying and selling transactions from August 2021 to September 2021. However in general, the selling value is lower than the buying value.

Daily Transaction Frequency

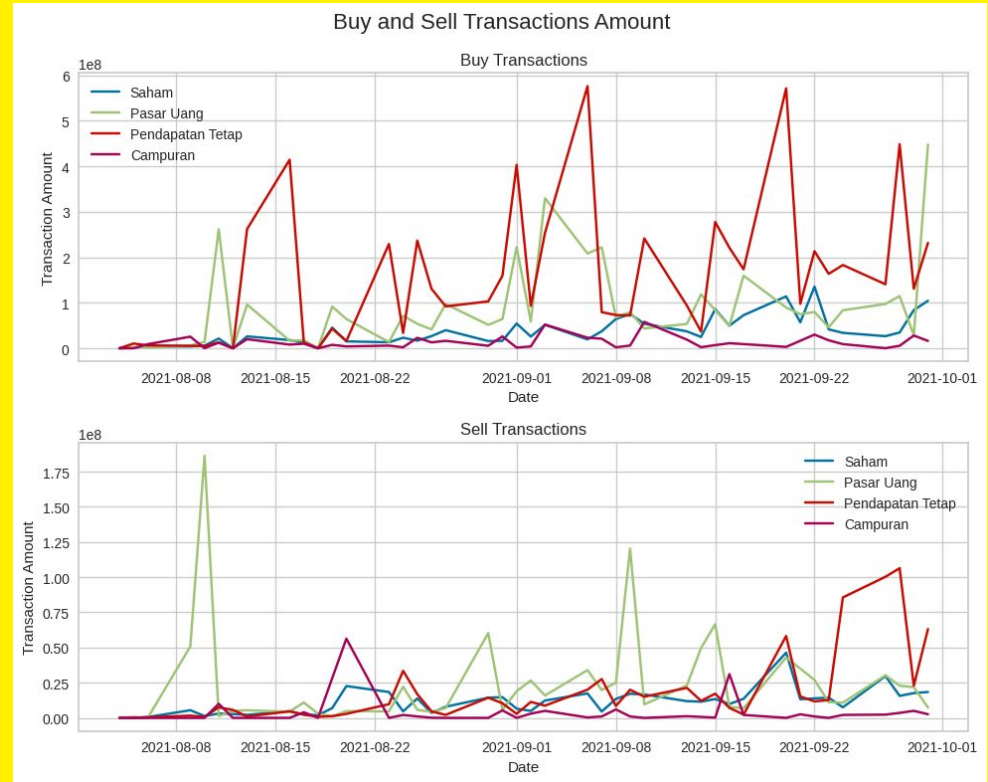
Buy and Sell Transactions Frequency



There is a fluctuating trend towards the daily transaction frequency for each mutual funds from August 2021 to September 2021. However, mixed mutual funds generally have the lowest trend compared to other types.

Daily Transaction Amount

There is a fluctuating trend towards the daily transaction amount for each mutual funds from August 2021 to September 2021. However, mixed mutual funds generally have the lowest trend compared to other types.





04

User Segmentation

K-means clustering

Variables used

- User Age
- Last Amount Total
- Transaction Frequency

The number of clusters is 4, determined from the elbow method and silhouette analysis.

Cluster Characteristics

Newcomers

New to investing, and may not have much experience with investing and have started investing with very little value.

Beginners

Still new to investing, but starting to understand a little about investing and have the courage to invest even a little.

Frequent User

Understanding of investment, having the highest frequency of transactions with quite a lot of investment value.

Rich User

Expert in investment, has quite a lot of transaction frequency with a very large investment value.



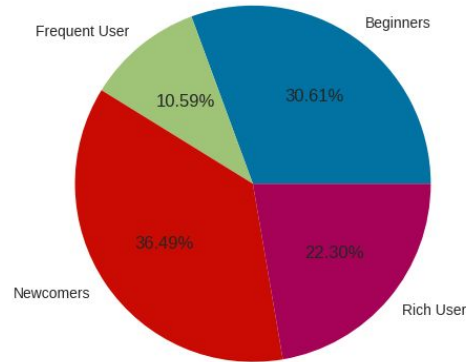
Cluster Characteristics

Newcomers

About 2798 user, spend on average Rp309.756 in one transaction.

Frequent User

About 812 user, spend on average Rp1.536.123 in one transaction.



Beginners

About 2347 user, spend on average Rp570.830 in one transaction.

Rich User

About 1710 user, spend on average Rp2.438.934 in one transaction.

Business Recommendation

Newcomers

Offer lower fees, and provide benefits on their first few investments to encourage them to invest.

Beginners

Offer benefits if they can invest more. Provide personal recommendations for the types of investments they have made.

Frequent User

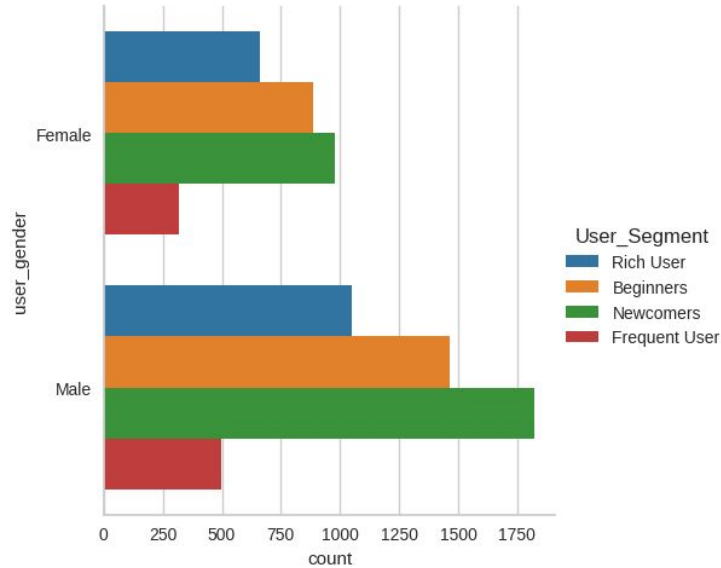
Offer membership programs and provide benefits from the program. Provide personal recommendations for the types of investments they have made.

Rich User

Offer exclusive access to premium services that will greatly facilitate investment that other users will not get.

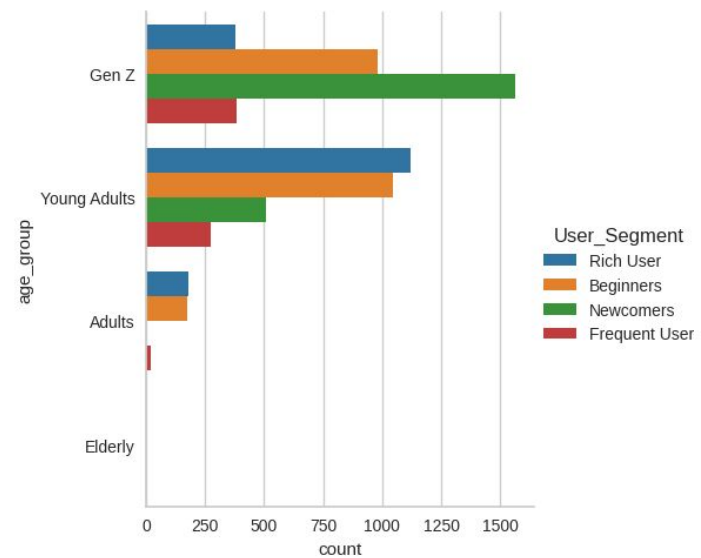
Cluster Demographic

Gender



Most user form all cluster are **male**

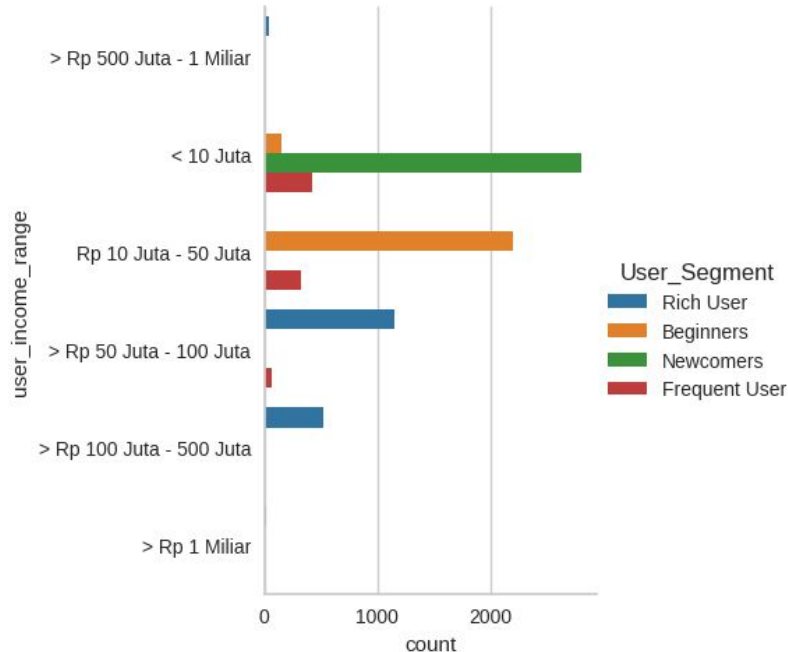
Age Group



Gen z and young adults dominate in every cluster.

Cluster Demographic

Income Range

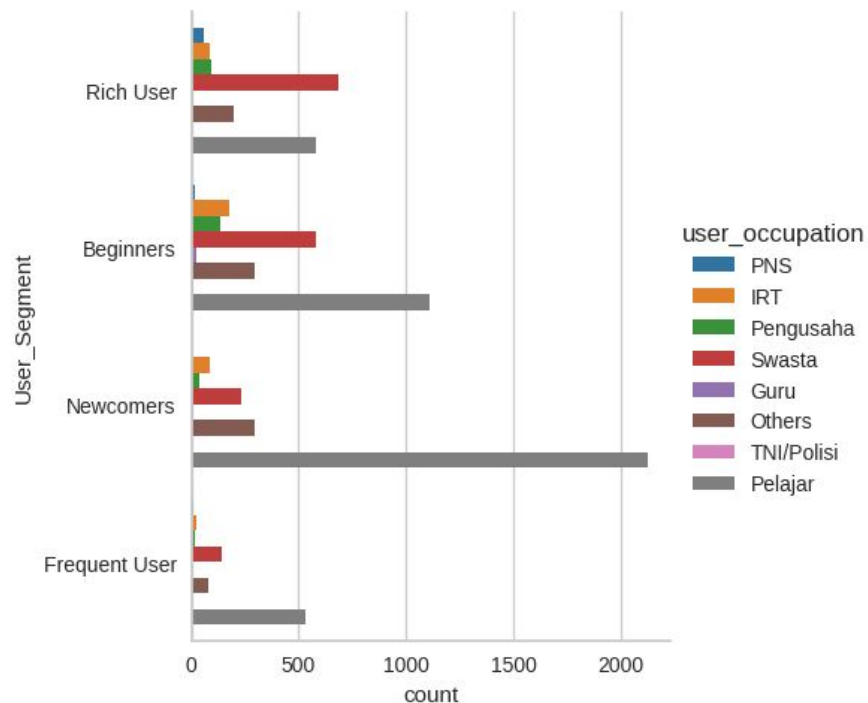


Income Range for each cluster:

- Newcomers: < 10 Juta (Lowest Income)
- Beginners: < 10 Juta - 50 Juta
- Frequent User: < 10 Juta - 100 Juta
- Rich User: > 50 Juta - > 1 Miliar (Highest Income)

Cluster Demographic

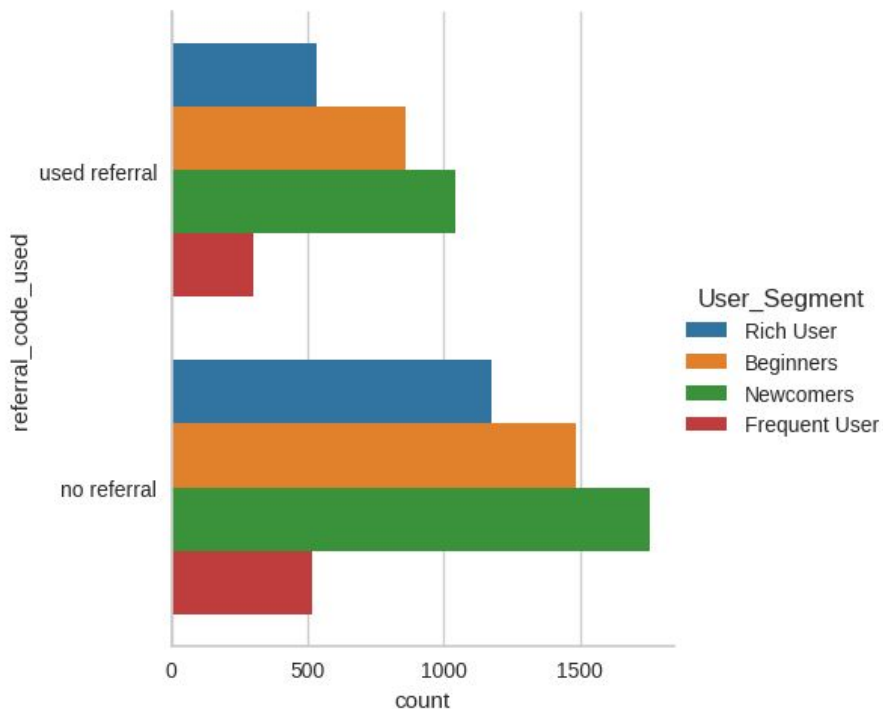
Occupation



In general **student** dominates in each cluster.

Cluster Demographic

Referral Used



In general, each cluster is dominated by users who do **not use referral codes.**

Average Invested Amount

Amount	Newcomers	Beginners	Frequent User	Rich User
Buy amount	Rp91.966	Rp165.483	Rp1.388.062	Rp1.009.055
Sell amount	Rp57.288	Rp74.163	Rp557.559	Rp270.266
Total invested amount	Rp309.756	Rp570.830	Rp1.536.123	Rp2.438.934

05

Regression and Benefit-Cost Analysis



New Problem

Marketing team only have a budget for campaign 30% of total users on each segment. So they seek help to choose who will be targeted with the budget limitation and the projection of how profitable the campaign. The cost per campaign is Rp1.000 and transaction fee (buy and sell) is 0.15%

Solution

1

Doing logistic regression

Targeted users are top 30% of users who have highest churn probability from each campaign.

2

Calculate the assumed transaction value

Calculate the average of total buy amount of targeted users for each campaign.

3

Perform benefit cost analysis

Top 30% Targeted Users Who Have The Highest Churn Probability in Each Campaign

Users	Campaign 1	Campaign 2	Campaign 3	Campaign 4
All	2798	2347	812	1710
30%	839	704	244	513

Churn?	Campaign 1	Campaign 2	Campaign 3	Campaign 4
Churn	58.88%	65.21%	61.54%	67.35%
Not Churn	41.12%	34.79%	38.46%	32.65%

- Targeted users for campaign 1 are 839 users (58.88% churn)
- Targeted users for campaign 2 are 641 users (65.21% churn)
- Targeted users for campaign 3 are 234 users (61.54% churn)
- Targeted users for campaign 4 are 441 users (67.35% churn)

Calculate The Assumed Transaction Value

	Campaign 1	Campaign 2	Campaign 3	Campaign 4
Transaction Value	Rp1.205.283	Rp1.471.829	Rp972.325	Rp1.707.937

After we have calculated the assumed transaction value, we can calculate the expected return using this formula:

Expected Return = (Potential Profit) x (Targeted User - User Churn) - (Cost Losing) x (User Churn) - Total Campaign Cost

Benefit Cost Analysis: [Click here](#)

Campaign for Newcomers

User Base		Notes
Total User	2,798	total user in the cluster
% Targeted User	30%	percentage of user that can be included based on budget
Targeted User	839	absolute number of user that can be included based on budget
% User Churn from Targeted User	58.88%	percentage of user that are churned
User Churn from Targeted User	494	absolute number of user that are churned

Cost		
Campaign Cost per User	1,000	cost per campaign
Total Campaign Cost	839,400	total cost of campaign
Assumption Cost Losing User	1,808	cost if user churn
Transaction Fee	0.15%	

Revenue		
Assumption Transaction Value	1,205,284	transaction value when user transact because marketing campaign
Assumption % Potential Profit	150%	percentage of potential profit if campaign is success
Assumption Potential Profit	2,712	absolute number of potential profit if campaign is success

Expected Return	-795,829	expected revenue from campaign
-----------------	----------	--------------------------------

Campaign for Beginners

User Base		Notes
Total User	2,347	total user in the cluster
% Targeted User	30%	percentage of user that can be included based on budget
Targeted User	704	absolute number of user that can be included based on budget
% User Churn from Targeted User	65.21%	percentage of user that are churned
User Churn from Targeted User	459	absolute number of user that are churned
Cost		
Campaign Cost per User	1,000	cost per campaign
Total Campaign Cost	704,100	total cost of campaign
Assumption Cost Losing User	2,208	cost if user churn
Transaction Fee	0.15%	
Revenue		
Assumption Transaction Value	1,471,829	transaction value when user transact because marketing campaign
Assumption % Potential Profit	150%	percentage of potential profit if campaign is success
Assumption Potential Profit	3,312	absolute number of potential profit if campaign is success
Expected Return	-905,777	expected revenue from campaign

Campaign for Frequent User

User Base		Notes
Total User	812	total user in the cluster
% Targeted User	30%	percentage of user that can be included based on budget
Targeted User	244	absolute number of user that can be included based on budget
% User Churn from Targeted User	61.54%	percentage of user that are churned
User Churn from Targeted User	150	absolute number of user that are churned
Cost		
Campaign Cost per User	1,000	cost per campaign
Total Campaign Cost	243,600	total cost of campaign
Assumption Cost Losing User	1,458	cost if user churn
Transaction Fee	0.15%	
Revenue		
Assumption Transaction Value	972,326	transaction value when user transact because marketing campaign
Assumption % Potential Profit	150%	percentage of potential profit if campaign is success
Assumption Potential Profit	2,188	absolute number of potential profit if campaign is success
Expected Return	-257,601	expected revenue from campaign

Campaign for Rich User

User Base		Notes
Total User	1,710	total user in the cluster
% Targeted User	30%	percentage of user that can be included based on budget
Targeted User	513	absolute number of user that can be included based on budget
% User Churn from Targeted User	67.35%	percentage of user that are churned
User Churn from Targeted User	346	absolute number of user that are churned

Cost		
Campaign Cost per User	1,000	cost per campaign
Total Campaign Cost	513,000	total cost of campaign
Assumption Cost Losing User	2,562	cost if user churn
Transaction Fee	0.15%	

Revenue		
Assumption Transaction Value	1,707,937	transaction value when user transact because marketing campaign
Assumption % Potential Profit	150%	percentage of potential profit if campaign is success
Assumption Potential Profit	3,843	absolute number of potential profit if campaign is success

Expected Return	-757,662	expected revenue from campaign
-----------------	----------	--------------------------------

Recommendation

If the marketing team applies this strategy which only focuses on this target user, then none of the campaigns will be profitable, because all expected return values are negative. For this reason, we need to do campaign analysis by reviewing whether the campaign is attractive enough for the target. We can also adjust existing incentives and modify campaign targets. We also need to do a test of our campaign, by comparing the original campaign with the campaign that will be made.

Thanks!

Contact me:

rafiqnaufal97@gmail.com

+62 812 809 05778

[linkedin.com/in/rafiqnaufal](https://www.linkedin.com/in/rafiqnaufal)

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Storyset**

Please keep this slide for attribution

