

USER SEGMENTATION

By : Intania – Data Scientist Automation Dana



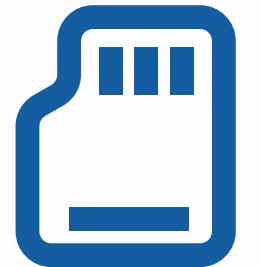
Overview

- About Dataset
- Profile Overview of the Users
- Popularity by category
- Segmentation by Peak Season
- Segmentation by CLV
- Conclusion
- Recommendation



About the dataset

User transaction data was collected from March 1st to May 31st, 2024, covering transactions from three Javanese provinces, spanning from west to east.



278.931
Total Transaction

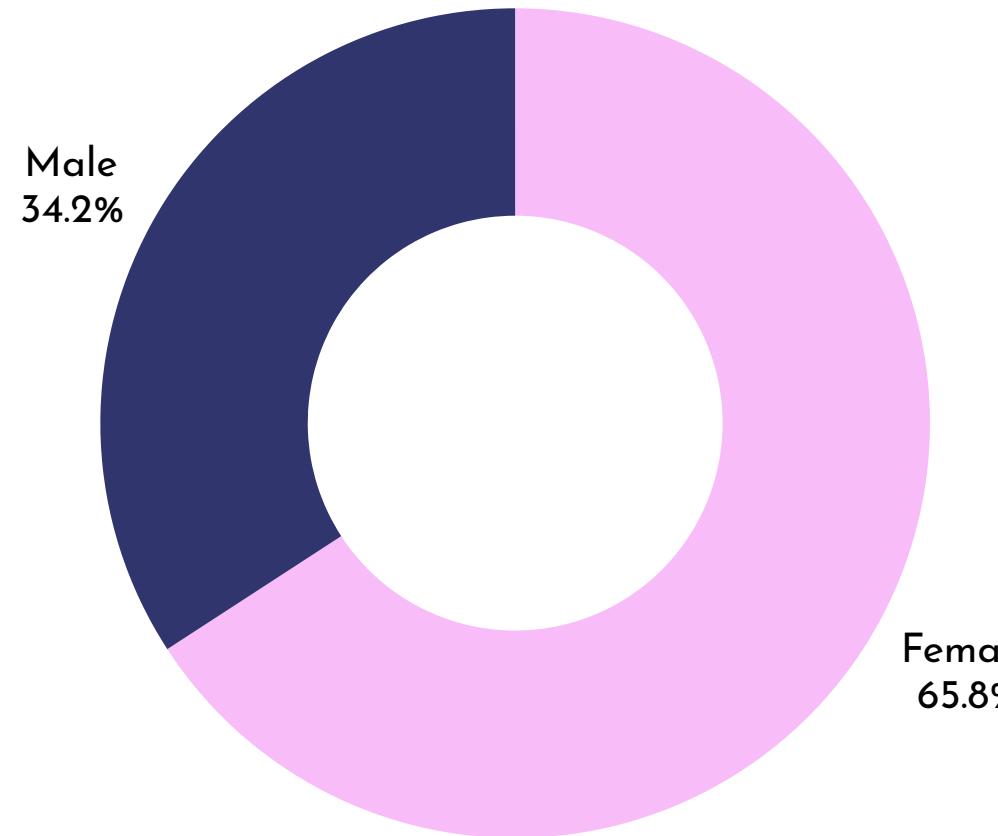


24.321
Total User



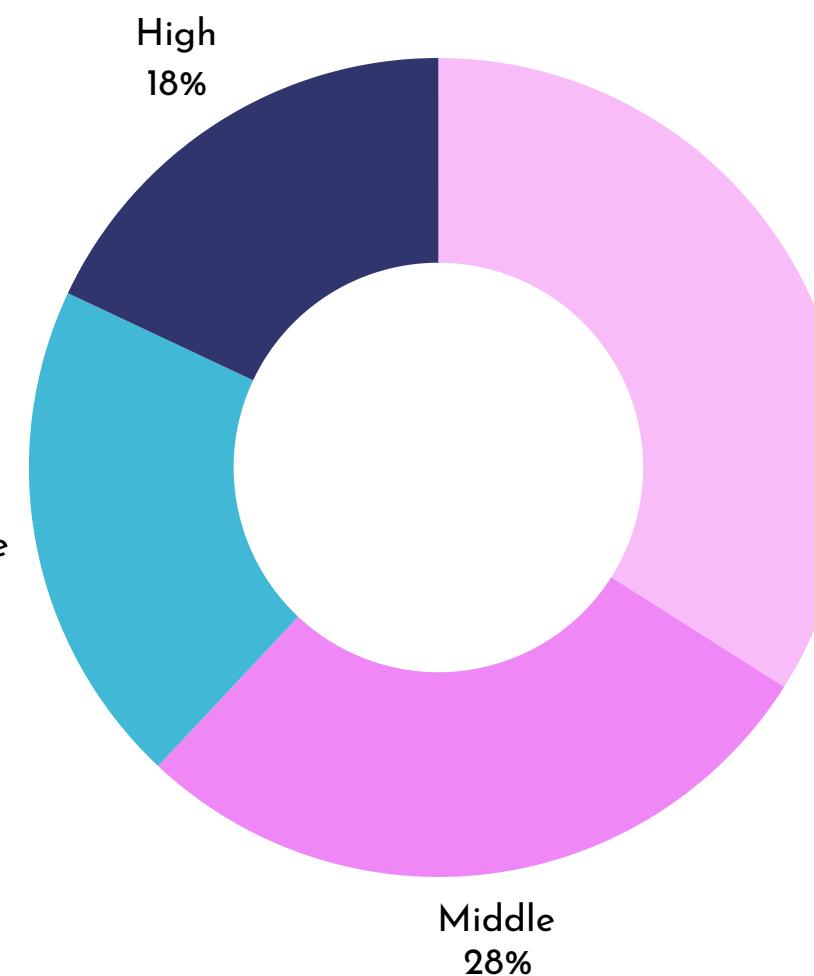
5.820
Total Merchant

PROFILE OF THE USERS IN DATASET (1/2)



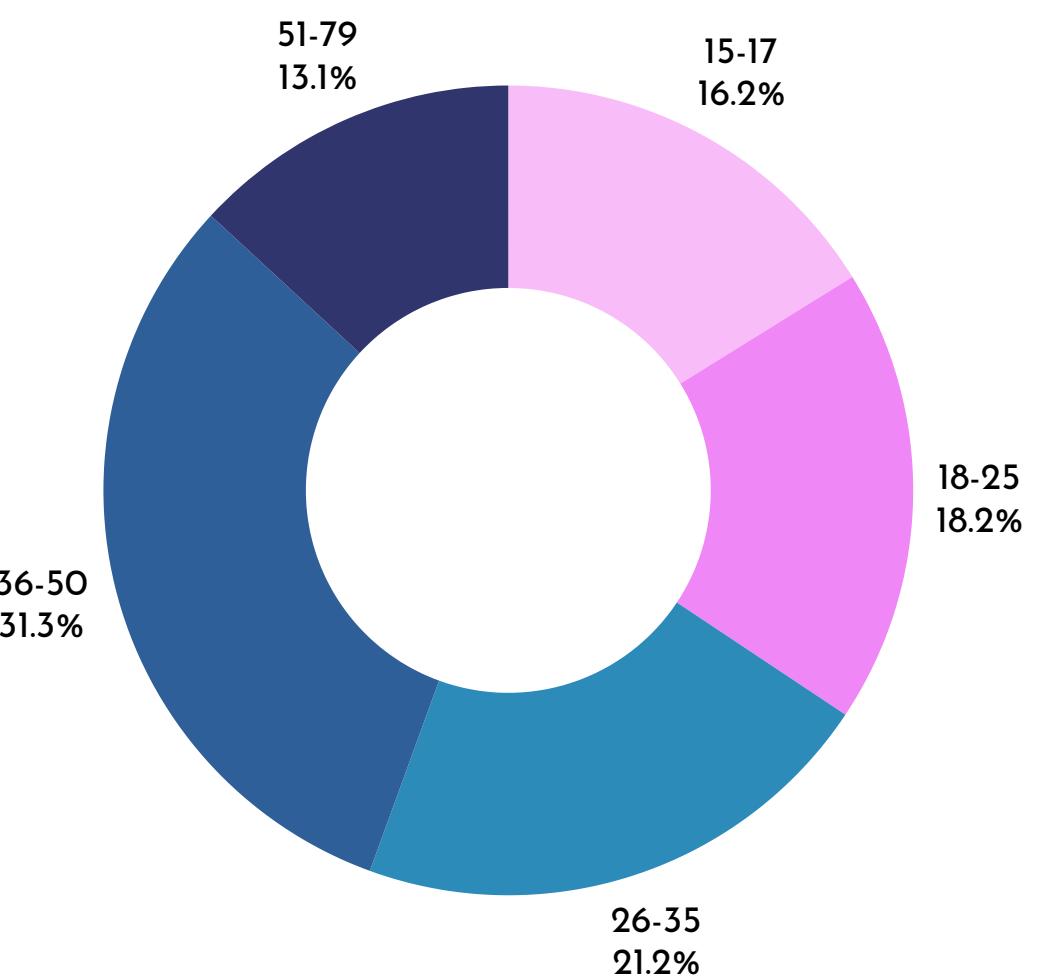
Gender

More than 65% of the transactions are dominated by female customers.



Income

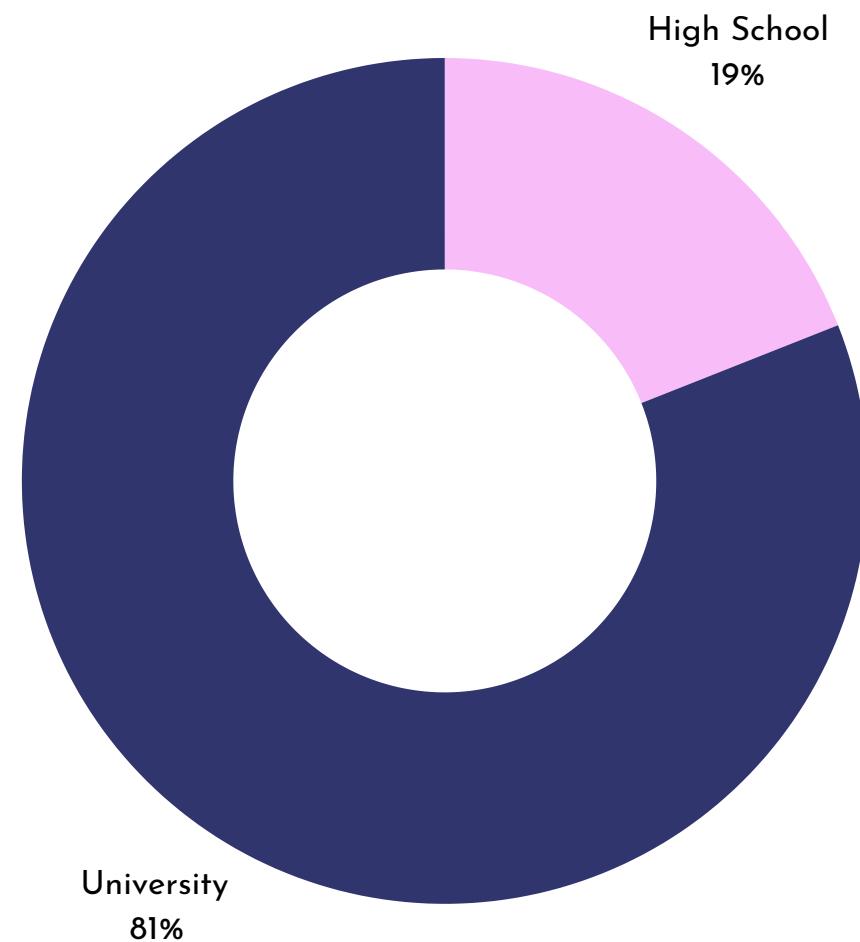
There are four income levels, and more than a quarter of the users belong to the lower-income class and middle-income class.



Age

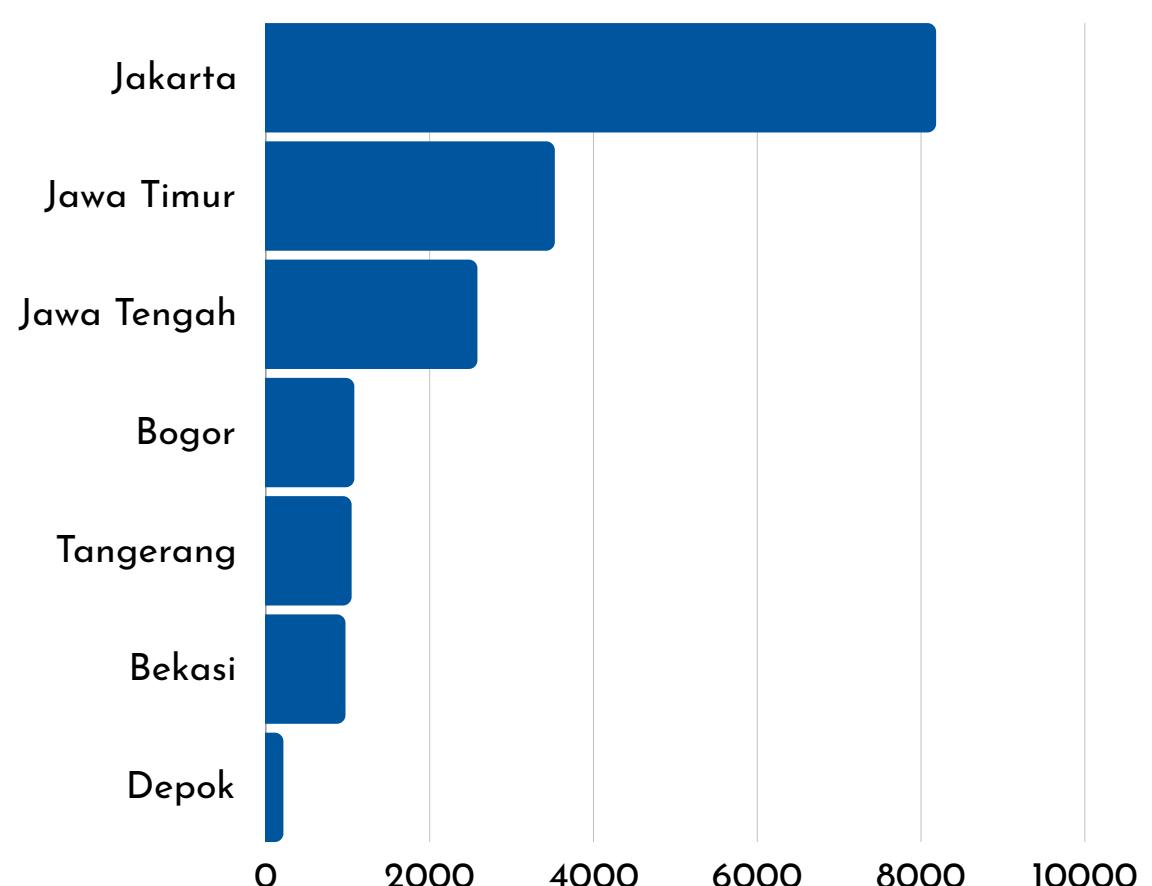
The largest percentage of users falls within the 36-50 age group, while the smallest percentage consists of those aged over 51 years.

PROFILE OF THE USERS IN DATASET (2/2)



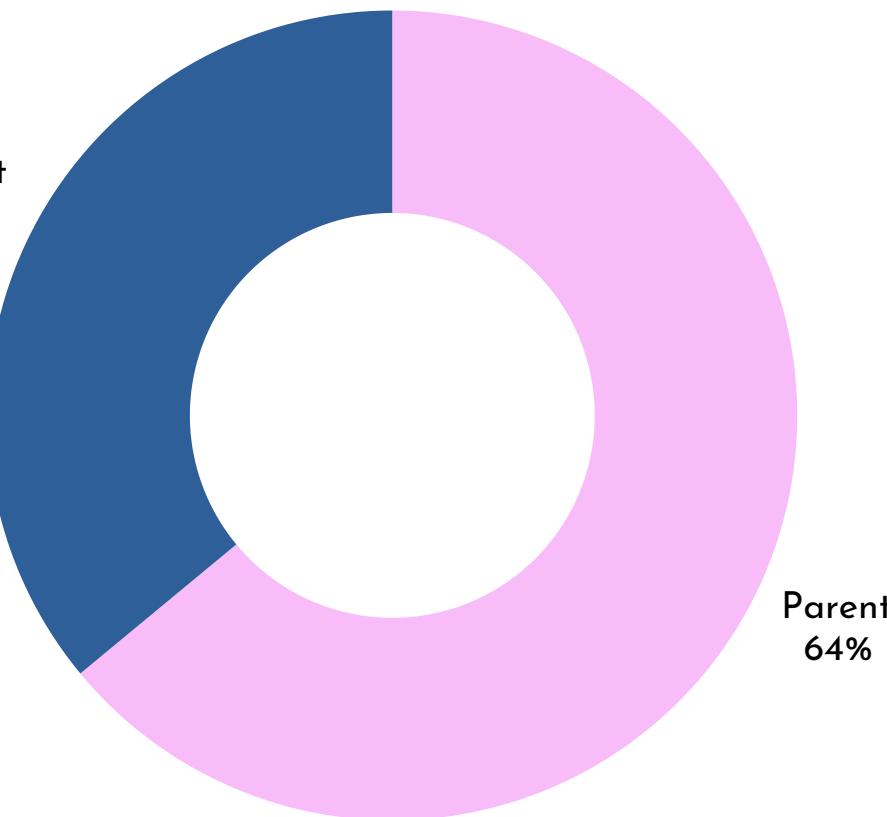
Education

Based on education, nearly all users have at least a bachelor's degree.



Geography

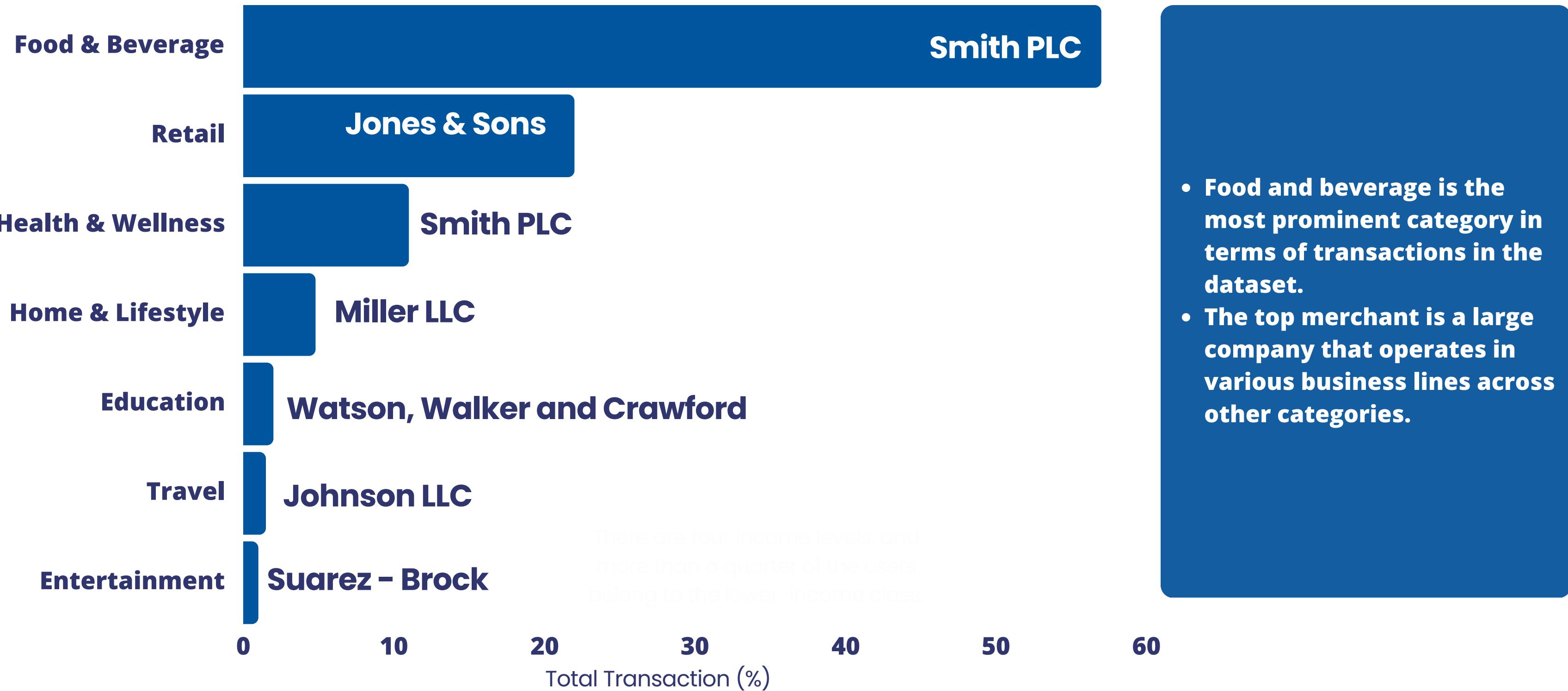
The dataset shows there are 24,321 users that are dispersed from Jabodetabek, East and Central Java.



Parental Status

More than 60% of the users are a parent

Popularity by Category Usage & Top Merchant



SEGMENTATION

-PEAK SEASON-

ANALYSIS PROCESS USER SEGMENTATION BASED ON PEAK SEASON (RAMADHAN)

Description & Objectives

During the exploratory data analysis (EDA), I identified certain periods with significantly higher transaction volumes. To understand this trend better, I analyzed the characteristics of individuals with high transaction amounts during these peak times.

Objectives:

To identify the key characteristics that differentiate each group, enabling better business decisions and improvements for the next peak season

Steps to do

- Split the transaction data into two sets: one that excludes all data from the peak season and another that includes only the transactions during the peak season.
- Combine each dataset with the user_features dataset, which has been engineered to capture user features.
- Find the central tendency of transaction amounts for each user.
- Determine the difference between the datasets for peak season and non-peak season transactions.
- Group the users based on their quartile ranges.

Results

The steps allow me to group users into four distinct categories:

- Group 0: Users with transaction amounts that are not significantly different from usual activities
- Group 1: Users who are 0.5 times more likely to spend above the average.
- Group 2: Users with a transaction amount that is 1.7 times higher than average.
- Group 3: Users with the highest transaction amounts during the peak season.

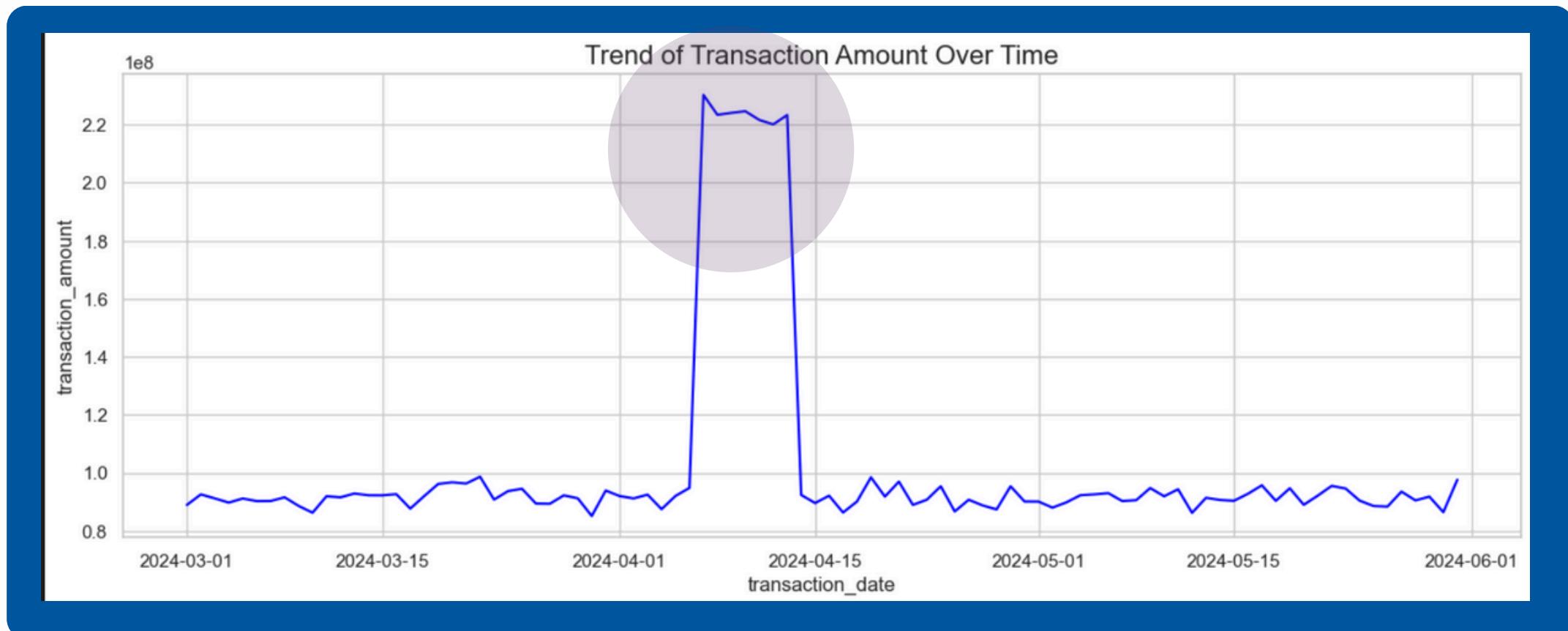
Key Insights

Most groups, based on characteristics such as age, gender, and parental status, share similar traits. The key distinguishing factors are income level and neuroticism characteristics.

In Group 3, nearly half of the members have a lower income level. This may suggest that during the peak season, various promotions enable them to make more transactions.

Additionally, individuals in Group 3 tend to score higher in neuroticism, indicating that they are more emotional in their purchasing decisions.

HIGH TRANSACTION AMOUNT 07 - 13 APRIL 2024



>50%

Users tend to spend at least 10% more than usual.

+17%

increase of average spending

>50%

Individuals with neurotic traits tend to have higher transaction amounts.

>100%

Increase in transaction amount

During the Ramadan festive period, it is likely to happen because Indonesia's society is predominantly Muslim, and they have access to UTHR (Tunjangan Hari Raya), which can increase their buying power."



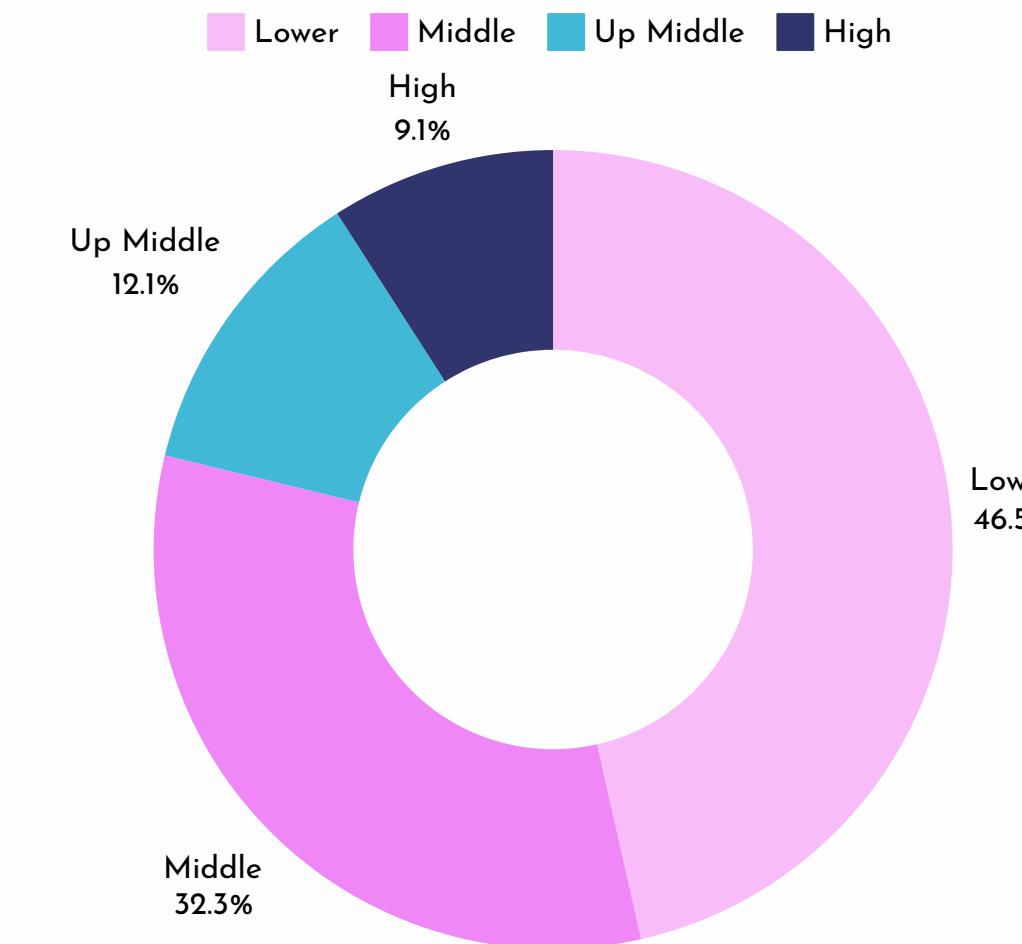
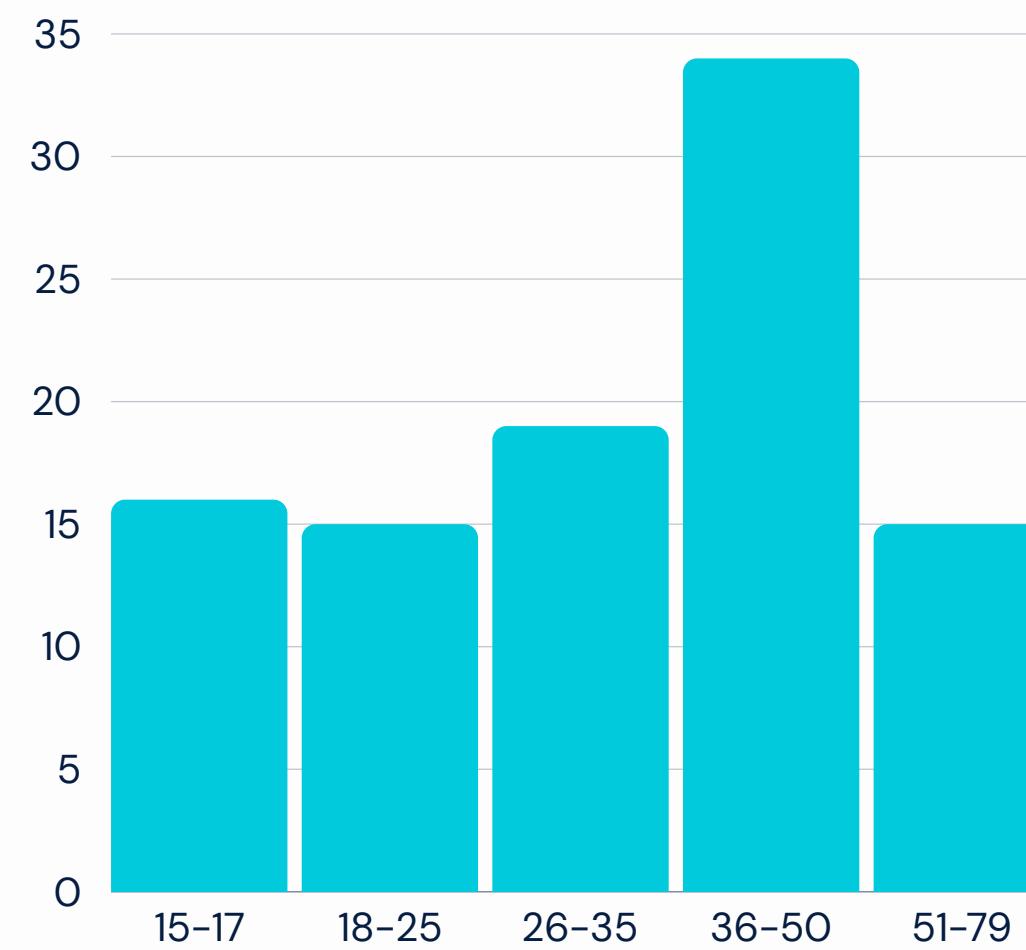
PEAK - High

This group consists of 1,310 users who, during the peak season, can have transaction amounts more than four times higher than usual. One key distinguishing trait among them is their high level of neuroticism.

1310
users

4.7X
difference of
transactions during
the peak season

73%
has trait
neuroticism



Distribution by Age

Distribution by Income Level

PEAK - Up Middle

The group consists of a total of 3,880 users who can have transaction amounts that are 1 to 3 times higher than the average. This group is predominantly made up of individuals aged 36 to 50, with a majority having a middle-income level.

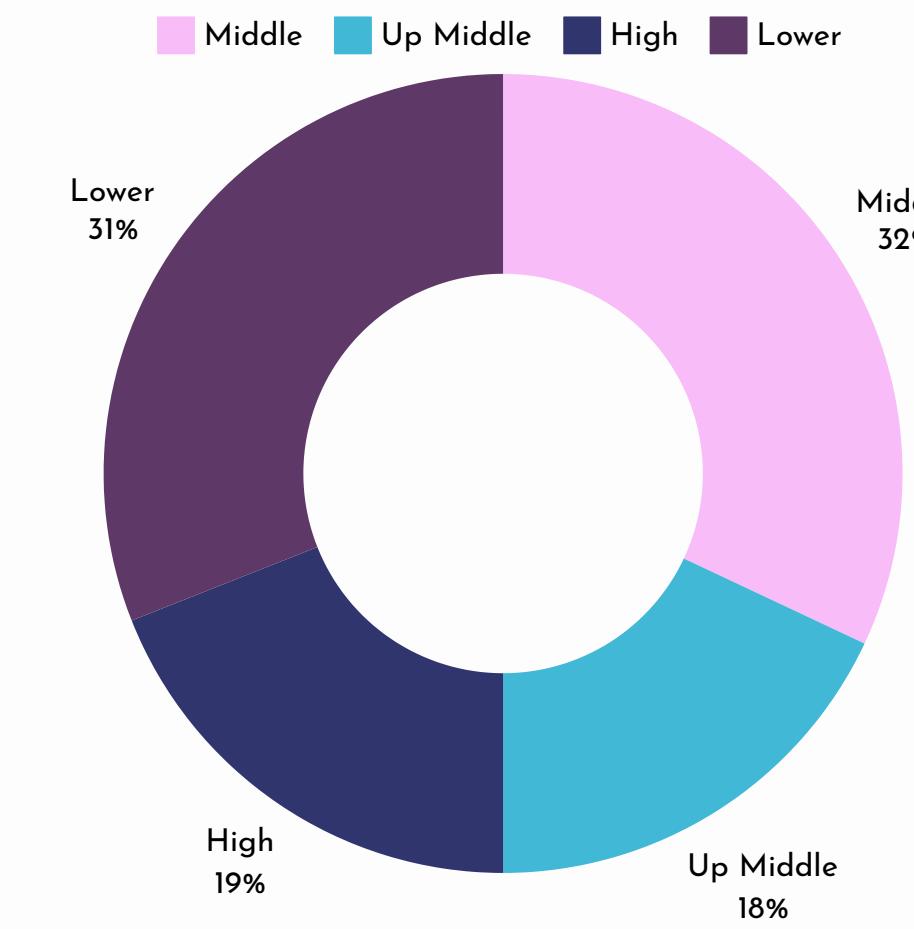
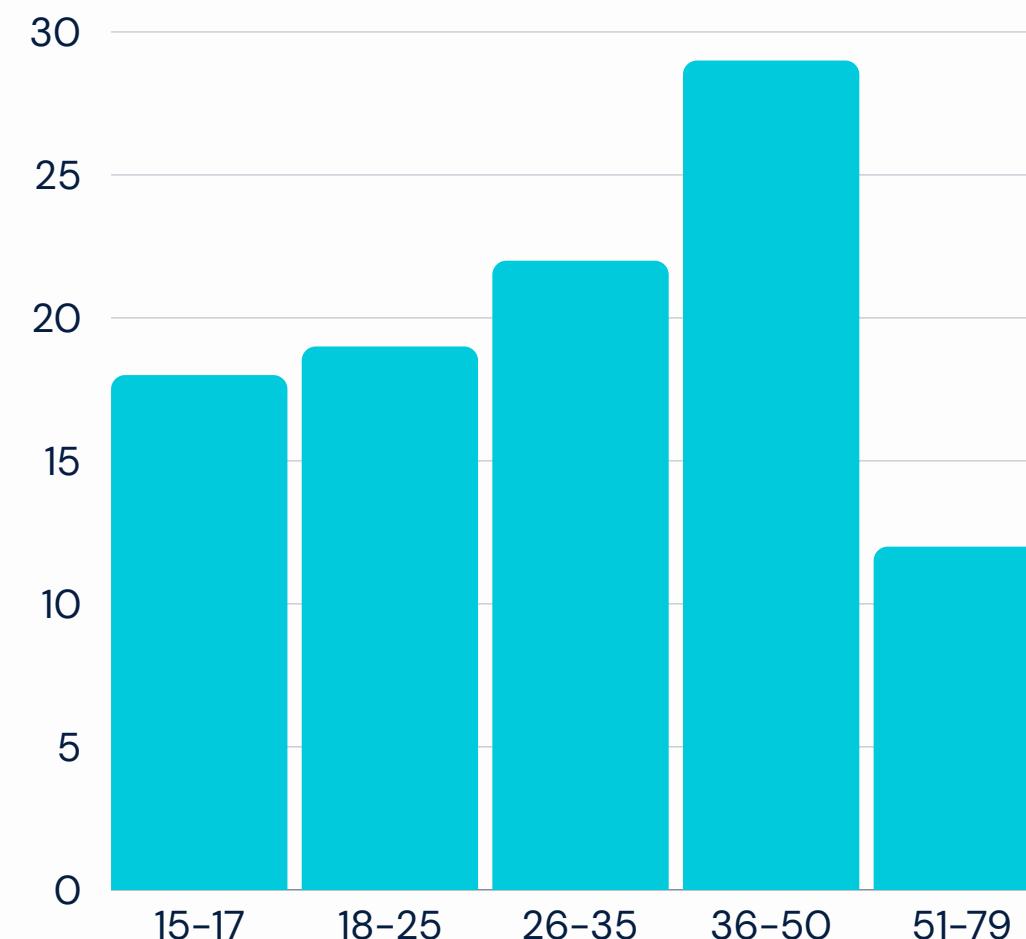
3880
users

1.7X

difference of
transactions during
the peak season

53%

has trait
neuroticism



Distribution by Age

Distribution by Income Level

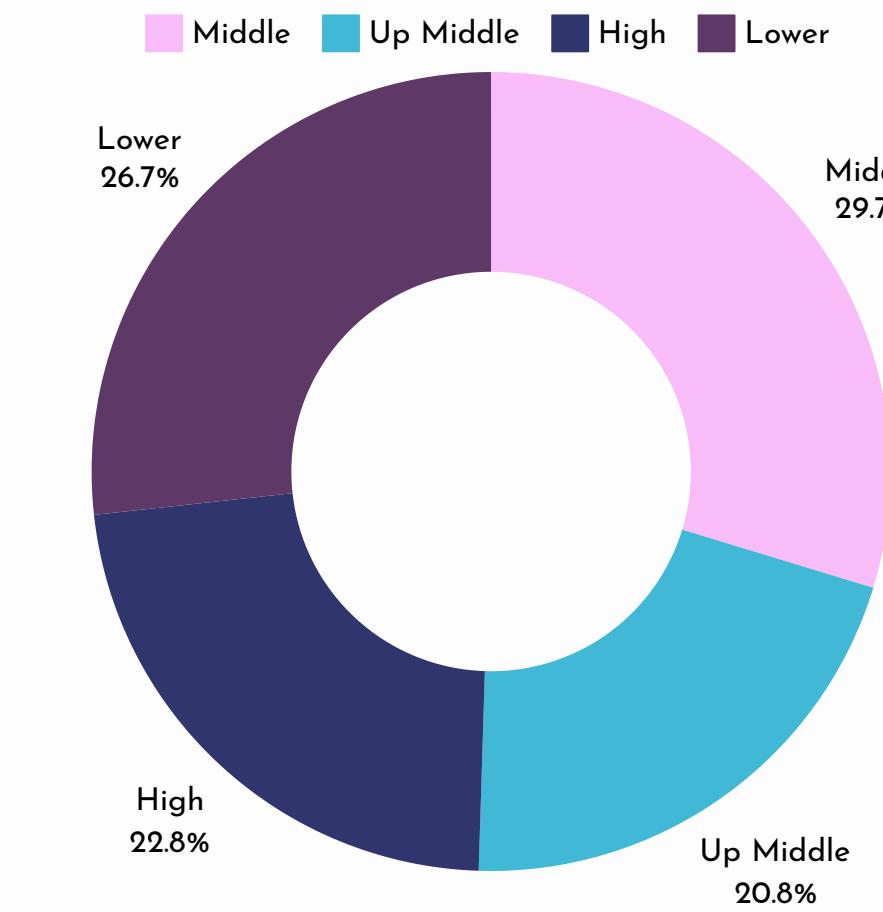
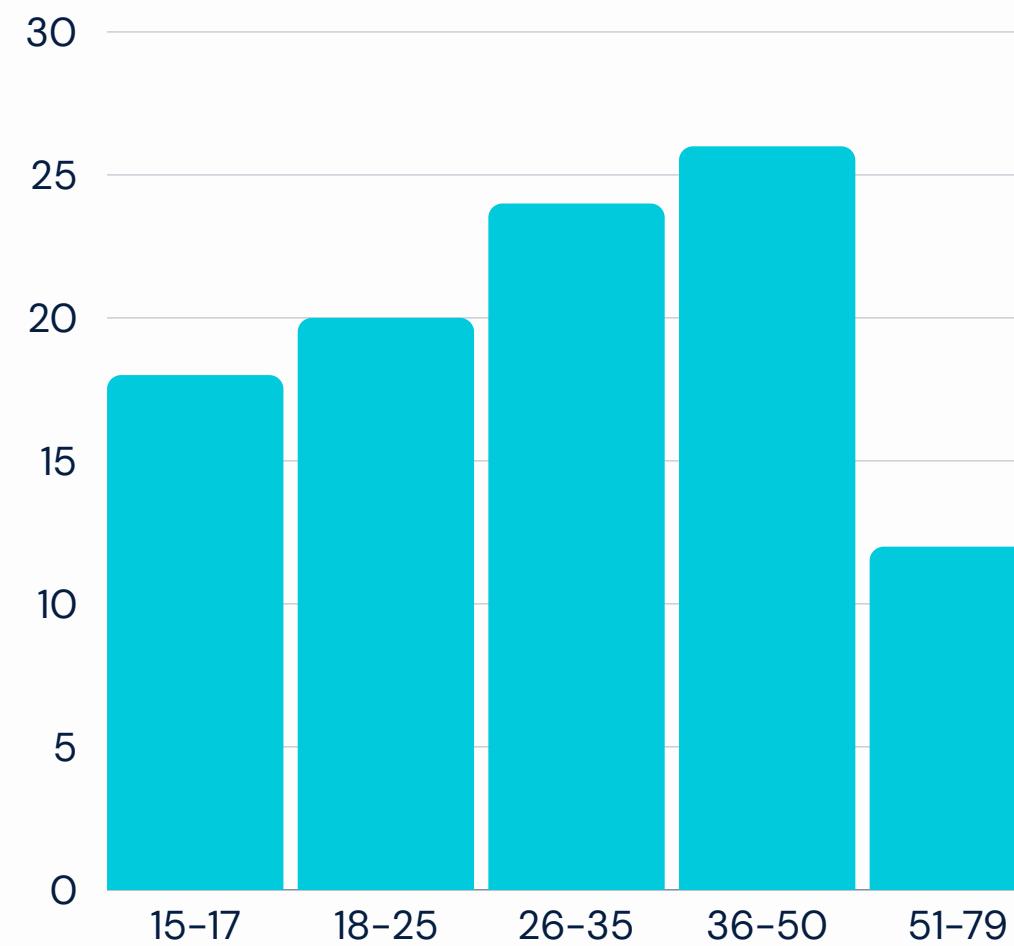
PEAK - Middle

Users in this group have a more relative distribution of ages, ranging from 36 to 50 and down to 15 to 17. This group is primarily led by individuals with middle-income levels, followed by users from other income brackets.

5190
users

0.5X
difference of
transactions during
the peak season

45%
has trait
neuroticism



Distribution by Age

Distribution by Income Level



Summary (Peak Season)

After segmentation, we found that most groups are predominantly made up of individuals aged 36 to 50, who tend to be in the lower income class.

Another notable trait is that users in the high CLV group exhibit a high proportion of neuroticism, indicating emotional instability.

Considering that this analysis occurred during Ramadan and assuming that Dana offered numerous promotions while users received their THR (holiday allowance), it is highly probable that these factors created an opportunity for larger transactions, allowing users to take advantage of various promotions.



SEGMENTATION

-CLV-

ANALYSIS PROCESS USER SEGMENTATION BASED ON CUSTOMER LIFETIME VALUE

Description & Objectives

Assuming this dataset is from the Dana App, one important metric to consider is Customer Lifetime Value (CLV). CLV is a crucial metric that estimates the total revenue a business can expect from a single customer throughout their relationship.

By understanding CLV, companies can identify valuable customers, tailor their marketing strategies, and optimize resource allocation.

Steps to do

- For each user, I calculated their average transaction amount, number of transactions, and customer lifespan to determine their Customer Lifetime Value (CLV).
- I also performed additional data cleaning to remove outliers, ensuring the data focuses more on the big picture.
- Afterward, I binned the CLV using the same interval strategy to analyze the proportion of users based on their CLV ranges, followed by segmentation.

Results

The steps allow me to group users into four distinct categories:

- Group 1: Users with a maximum CLV value of Rp 474,225
- Group 2: Users with a CLV range from Rp 474,226 to Rp 947.350.
- Group 3: Users with a CLV range from Rp 947.351 to Rp 1,420,475
- Group 4: Users with a CLV range from Rp 1,420,476 to Rp 1,893,600

Key Insights

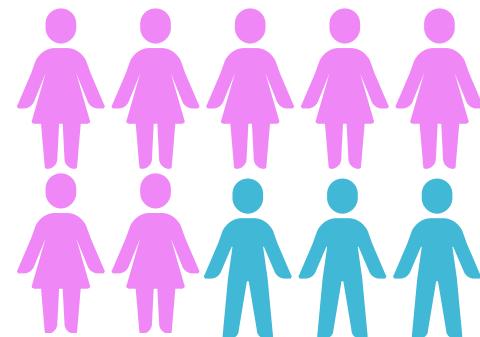
Most groups based on CLV share similar characteristics in education, gender, parental status, vehicle ownership, and device type.

The key distinguishing factors are age and income level, along with two personality traits: primarily extraversion, followed by neuroticism.

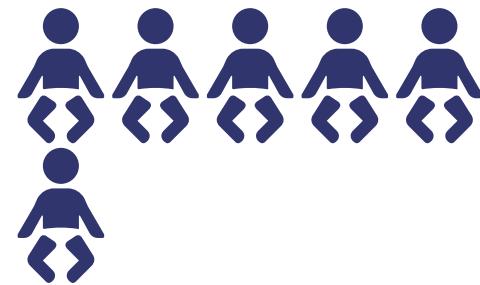
Individuals with the highest CLV are mainly aged 26 to 35 with high income, while those with low CLV are typically aged 36 to 50 with low income. High CLV users also tend to exhibit strong extraversion, indicating openness to social activities.

Over 41% of users fall into the middle CLV (Customer Lifetime Value) category.

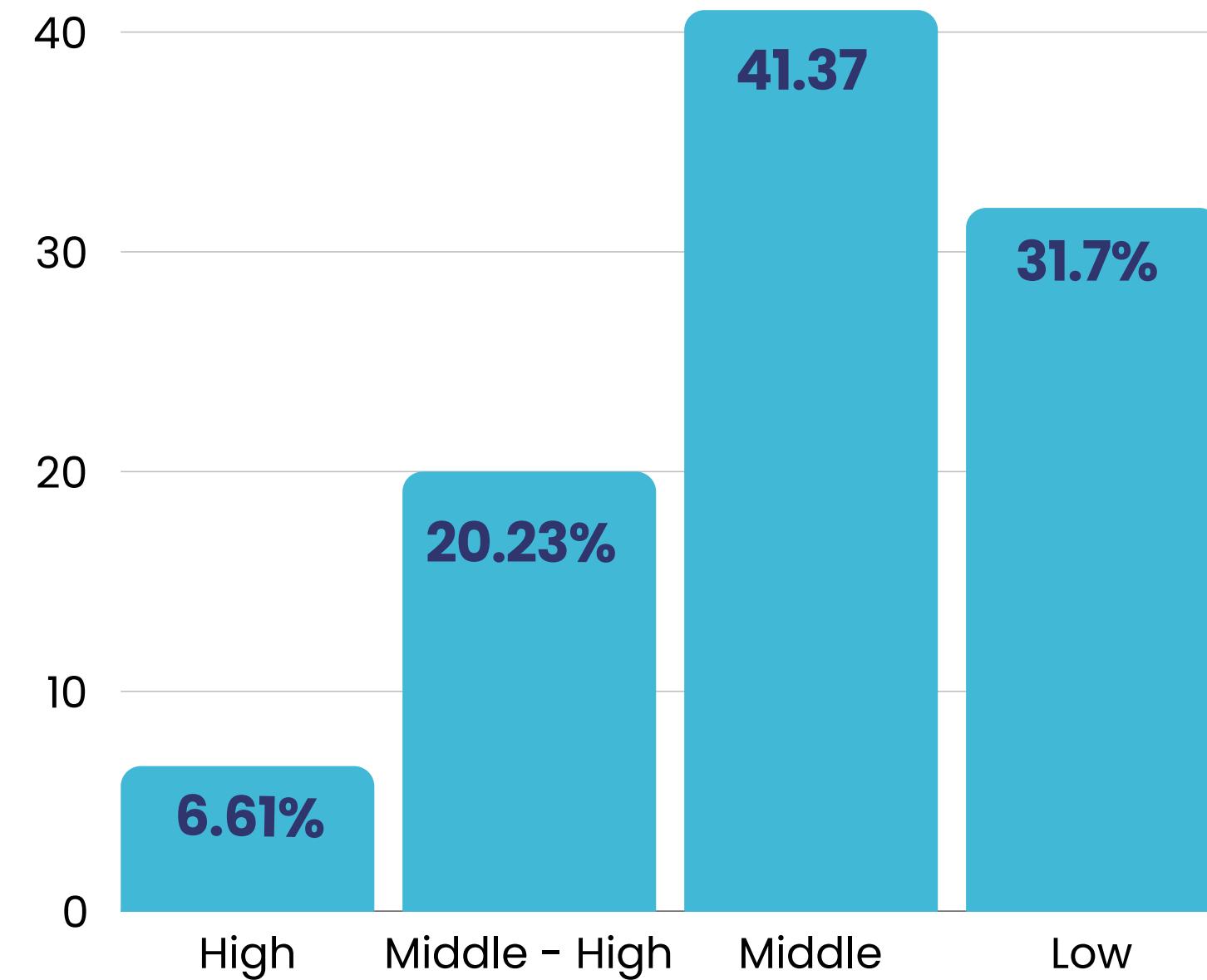
>62%
Transaction dominated
by Female



>60%
User is a parent



- Low : Transaction range Rp 1.100 – Rp 474.225
- Middle : Transaction range Rp 474.226 – Rp 947.350
- Middle – High : Transaction range Rp 947.351 – Rp 1.420.475
- High : Transaction range Rp 1.420.476 – Rp 1.893.600



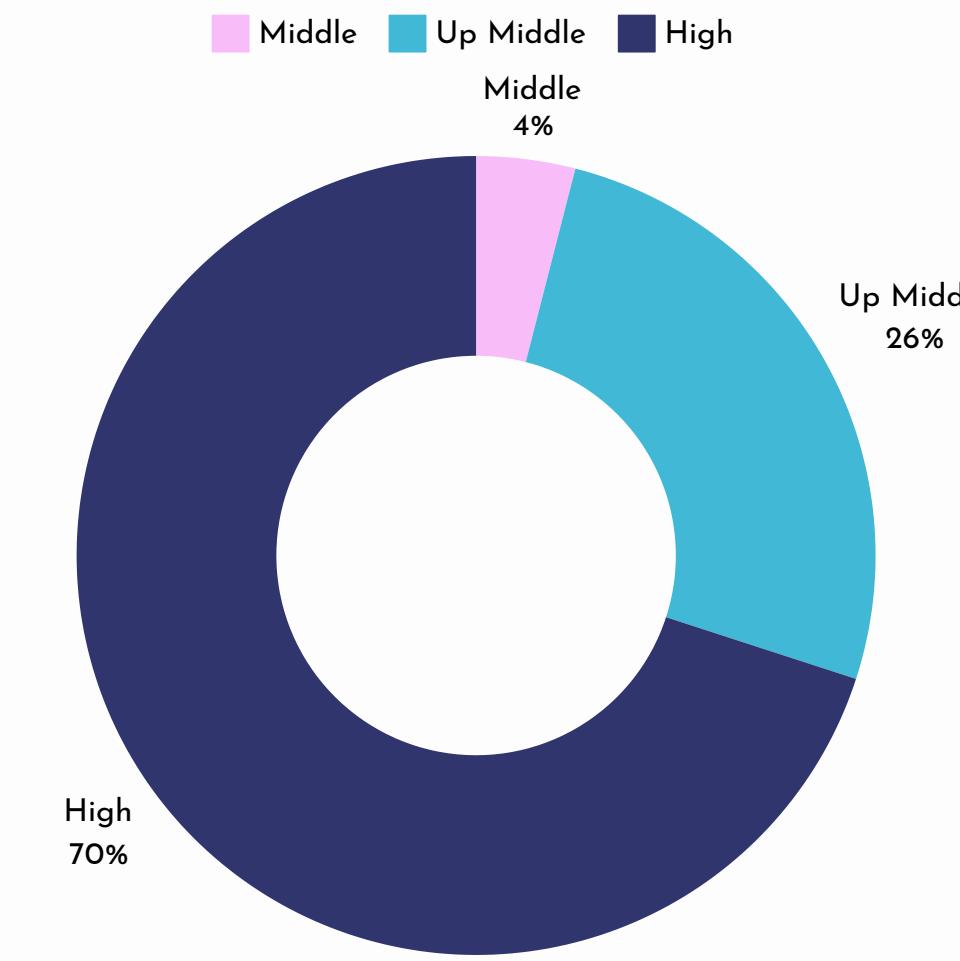
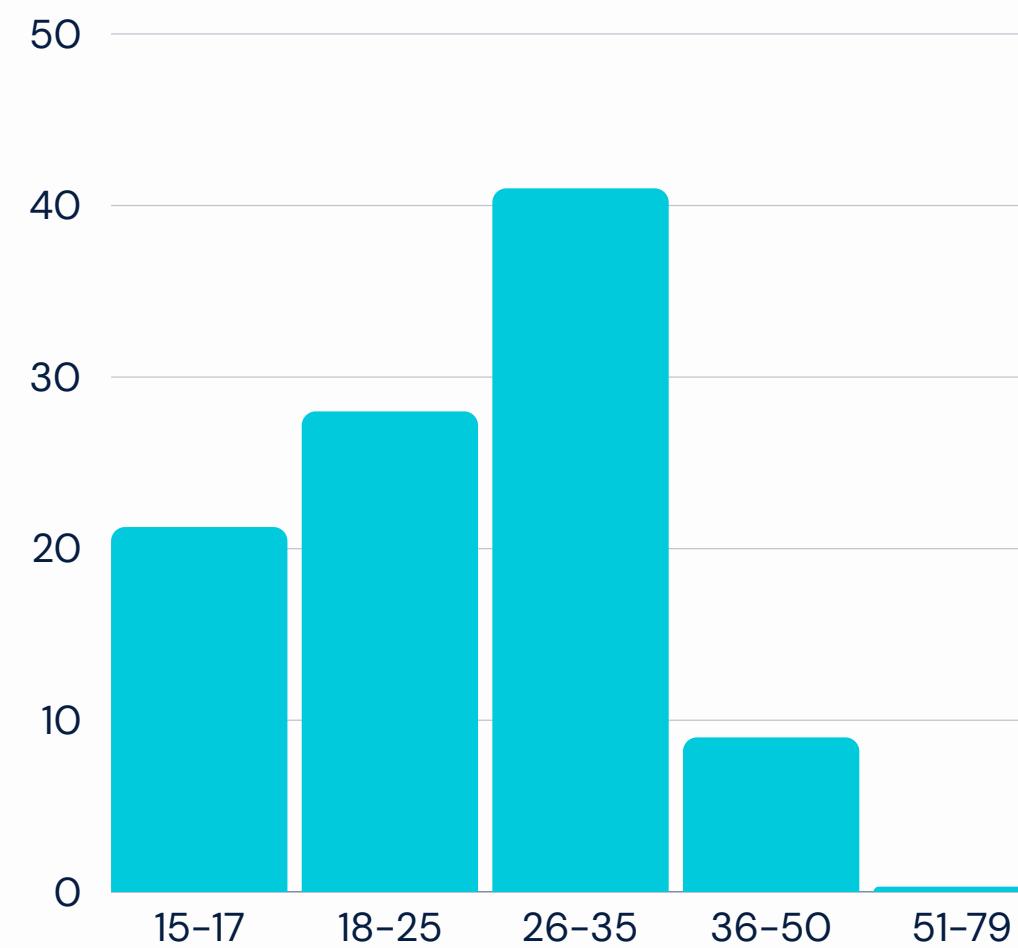
CLV - HIGH

This group consists of 1,575 users, and the most notable characteristic is that none of them fall into the low-income category, with all members owning homes.

1575
users

49.9 K
in rupiah
Avg. Transaction Value

2.7
month
User LifeSpan



48%
has trait
neuroticism

83%
has trait
extraversion

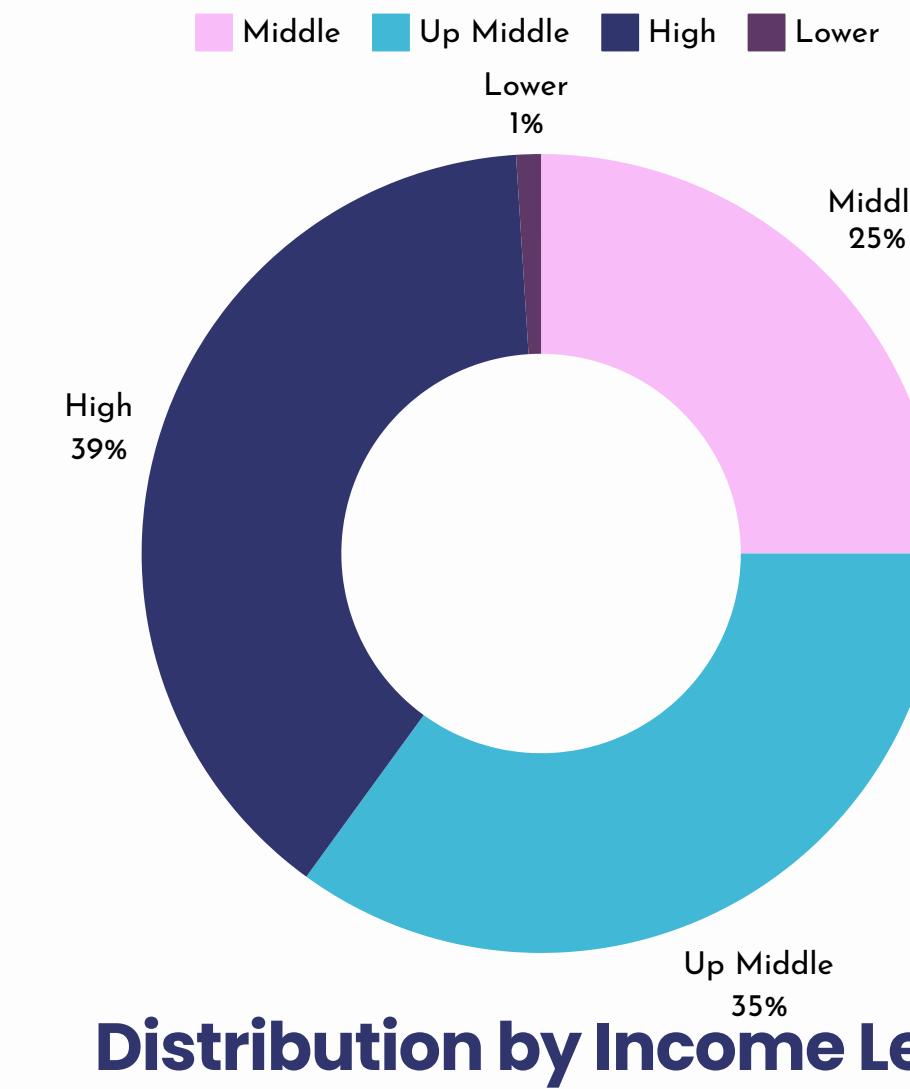
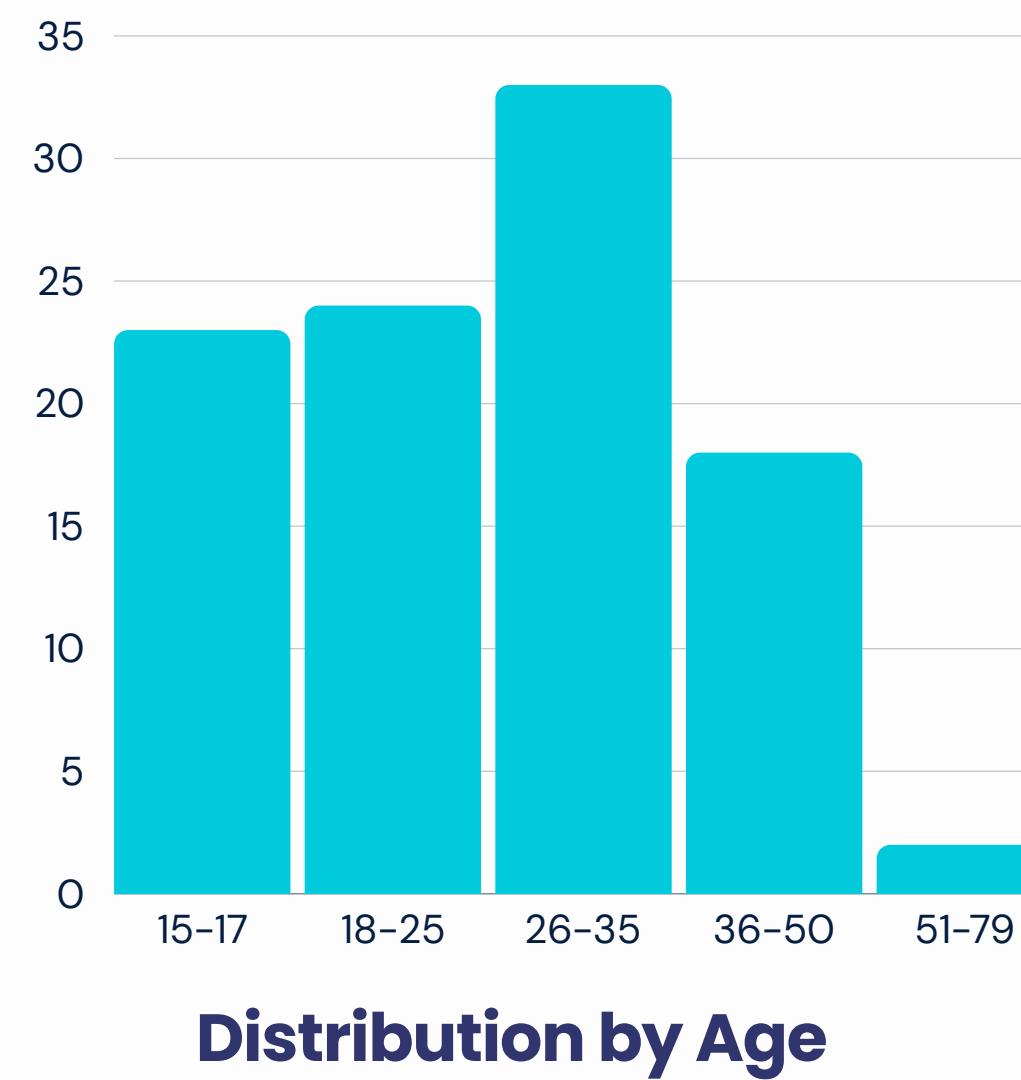
CLV - Up Middle

A total of 4,817 users show statistical similarities with those who have a high CLV. Notably, this group is characterized by higher tendencies towards extraversion and neuroticism.

4817
users

41.5 K
in rupiah
Avg. Transaction Value

2.6
month
User LifeSpan



49%
has trait
neuroticism

76%
has trait
extraversion

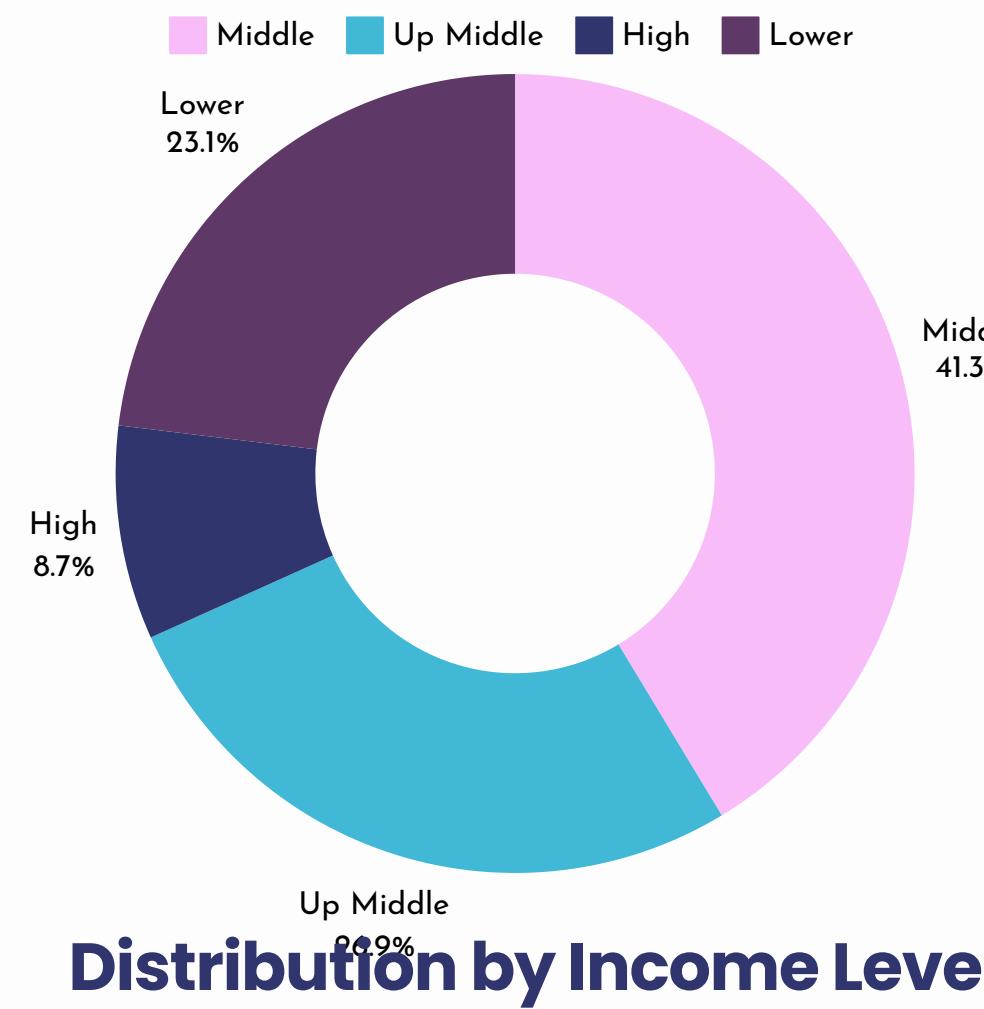
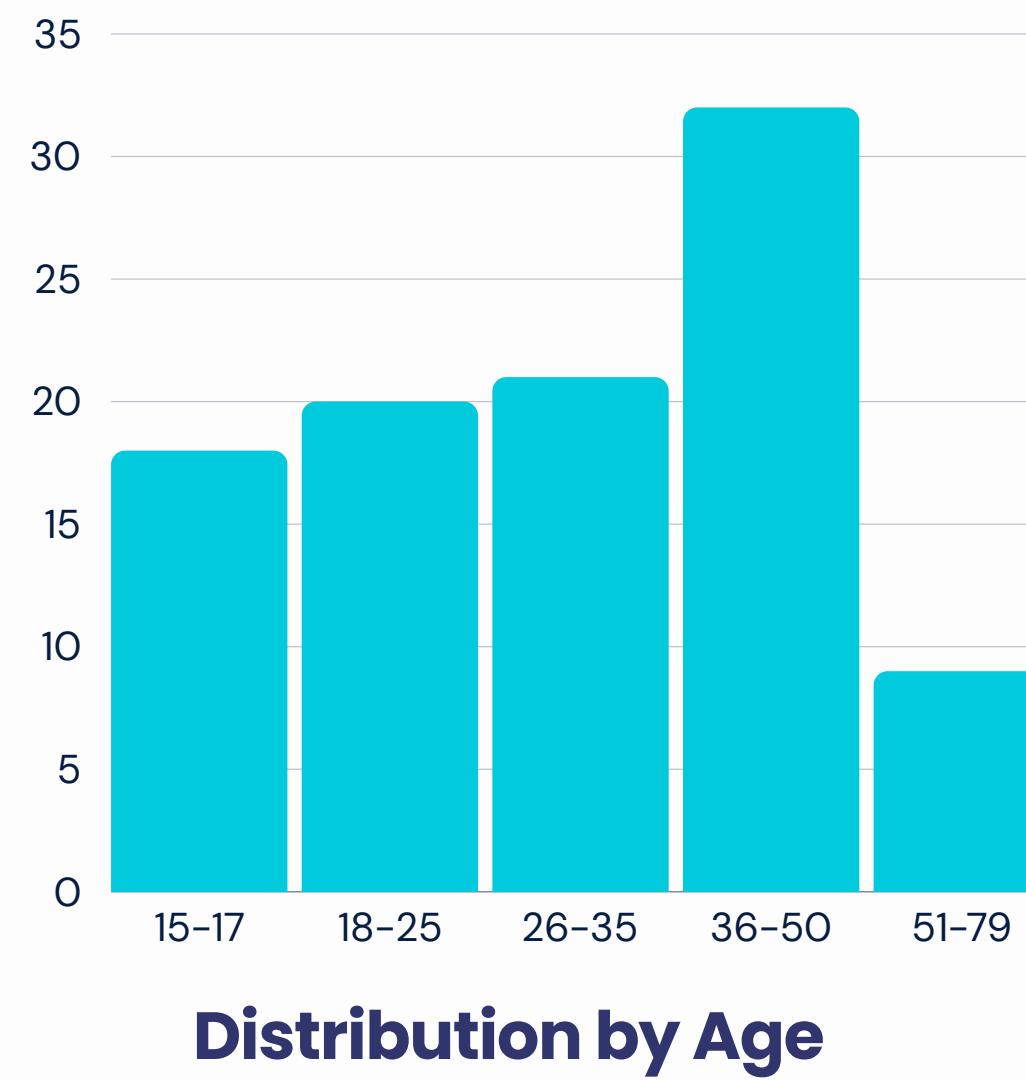
CLV - Middle

This are the highest user compare to other CLV'group. 60% of the group are extraversion

9848
users

31.7K
in rupiah
Avg. Transaction Value

2.4
month
User LifeSpan



51%
has trait
neuroticism

34%
has trait
extraversion

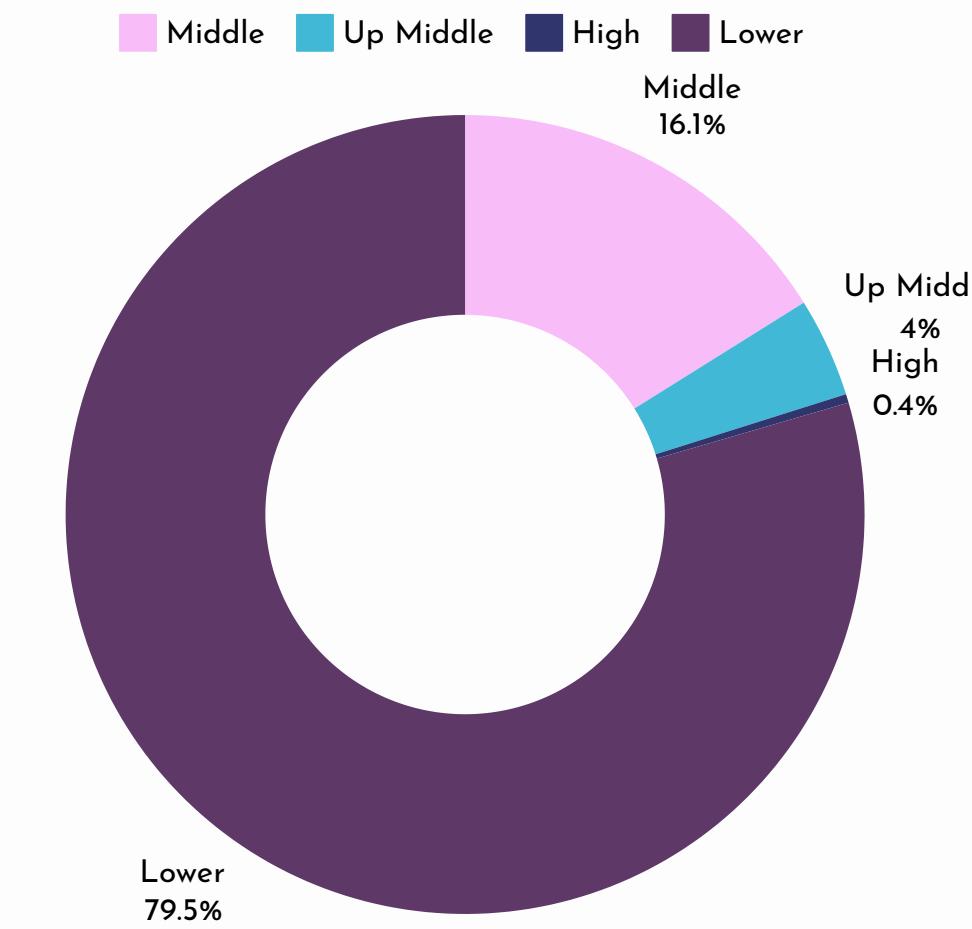
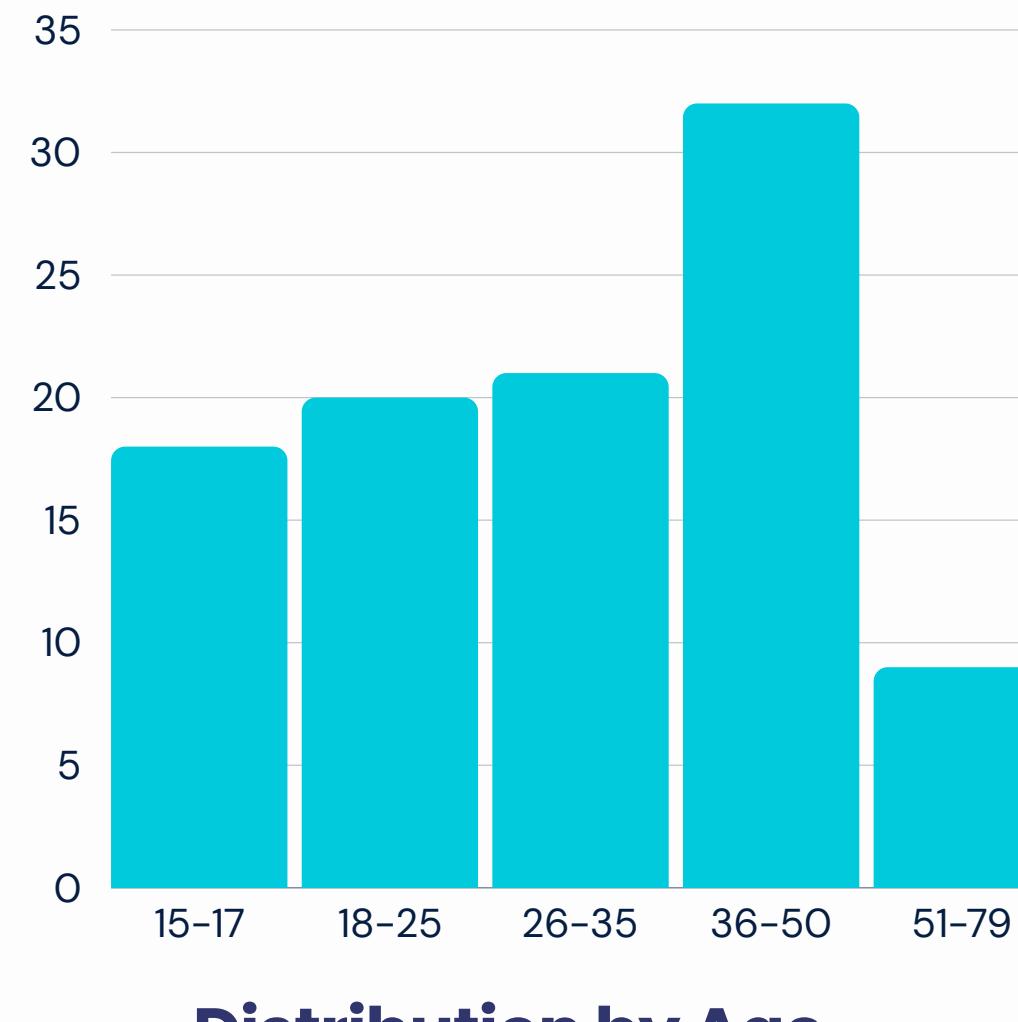
CLV - Low

A total of 7,560 users are included in this group, with a high percentage coming from the lower-income class. Psychologically, one key trait of this group is their lower levels of extraversion.

7560
users

21.9 K
in rupiah
Avg. Transaction Value

2
month
User LifeSpan



Distribution by Income Level

52%

has trait
neuroticism

34%

has trait
extraversion



Summary (CLV)

Segmentation using CLV yielded interesting results. Although individuals aged 36 to 50 make up the majority of users, the age group contributing the highest CLV is 26 to 35, particularly those with higher income levels, which enables them to make more transactions. Notably, this group also exhibits a higher proportion of extraversion, indicating they engage in many social activities.

In contrast, the middle and lower CLV groups are predominantly composed of individuals aged 36 to 50 with middle to lower income levels. This group does not show higher values in extraversion but demonstrates a greater tendency toward neuroticism, indicating emotional instability that may lead to impulsive buying.

CONCLUSION & RECOMMENDATION

Conclusion

Income level, age, neuroticism, and extraversion are the main features that differentiate one group from another.

01

The largest percentage of users is in the age group of 36 to 50, followed by those aged 25 to 36, 18 to 25, and 15 to 17. The smallest percentage is from the older generation. Moreover, over 50% of transactions are made by female users.

02

Based on the engineered features, almost all characteristics—such as gender, parental status, education level, gadget adoption, employment, and the three main personality traits of openness, conscientiousness, and agreeableness—show similar patterns.

03

Based on segmentation during the peak season, individuals aged 36 to 50 with lower incomes are most likely to have higher transaction amounts. They also show a significant proportion of neuroticism characteristics. Combined with the context of the Ramadan festive season, it is likely that users are using this opportunity to indulge their shopping desires

04

Meanwhile, based on CLV segmentation, it is interesting to note that the group with higher CLV consists of individuals aged 26 to 35 with higher income levels. People in this group also exhibit higher traits of extraversion, indicating greater engagement in social activities.

In contrast, the low CLV group comprises individuals aged 26 to 35 with lower incomes and a higher value in the trait of neuroticism.

Recommendation

Implementing tailored strategies for attracting and retaining customers can lead to more cost-efficient and effective results.

- 01**

Collaborate with companies that operate across various industries to enhance exposure and encourage users to increase their transactions.
- 02**

Focus on users with characteristics of extraversion and neuroticism, as these two traits can have a greater impact on the purchasing behavior of users.
- 03**

Customers with higher CLV should be engaged with effective retention strategies. Dive deeper into their needs and preferences, identifying any activity surges—particularly as many are parents or in the early stages of family life. Additionally, given their extraversion trait, provide information about various activities that align with their interests.
- 04**

Customers with middle and lower CLV should be targeted with strategies to attract them, such as increasing brand collaborations and offering tailored deals, as this group has a higher tendency toward impulse buying.

THANK YOU!



Intania
Data Enthusiast
 +62818-0702-0973



in

