

FINAL PROJECT PURWADHIKA DATA SCIENCE AND MACHINE LEARNING

RFM Customer Segmentation and Product Recommendation



CREATED BY :

BAYU FATHURRAHMAN
GIOVANO ADITYA GRAHA
NILA WILDANUL HUSNA

MENTORED BY :

MUHAMMAD AULIA, ARIES FITRIAWAN

About us



BAYU FATHURRAHMAN

- **Bachelor of Engineering,**
Petroleum Engineering
- Previously freelanced online as a
socmed/ search engine evaluator
- Aspiring Data Scientist/ Engineer



<https://github.com/Bysslockhart>



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



GIOVANO ADITYA GRAHA

- **Bachelor of Science,**
Mathematics Major
- Previously worked in the
insurance industry
- Data Scientist soon to be



<https://github.com/GiovanoAG>



www.linkedin.com/in/giovano-aditya-graha



NILA WILDANUL HUSNA

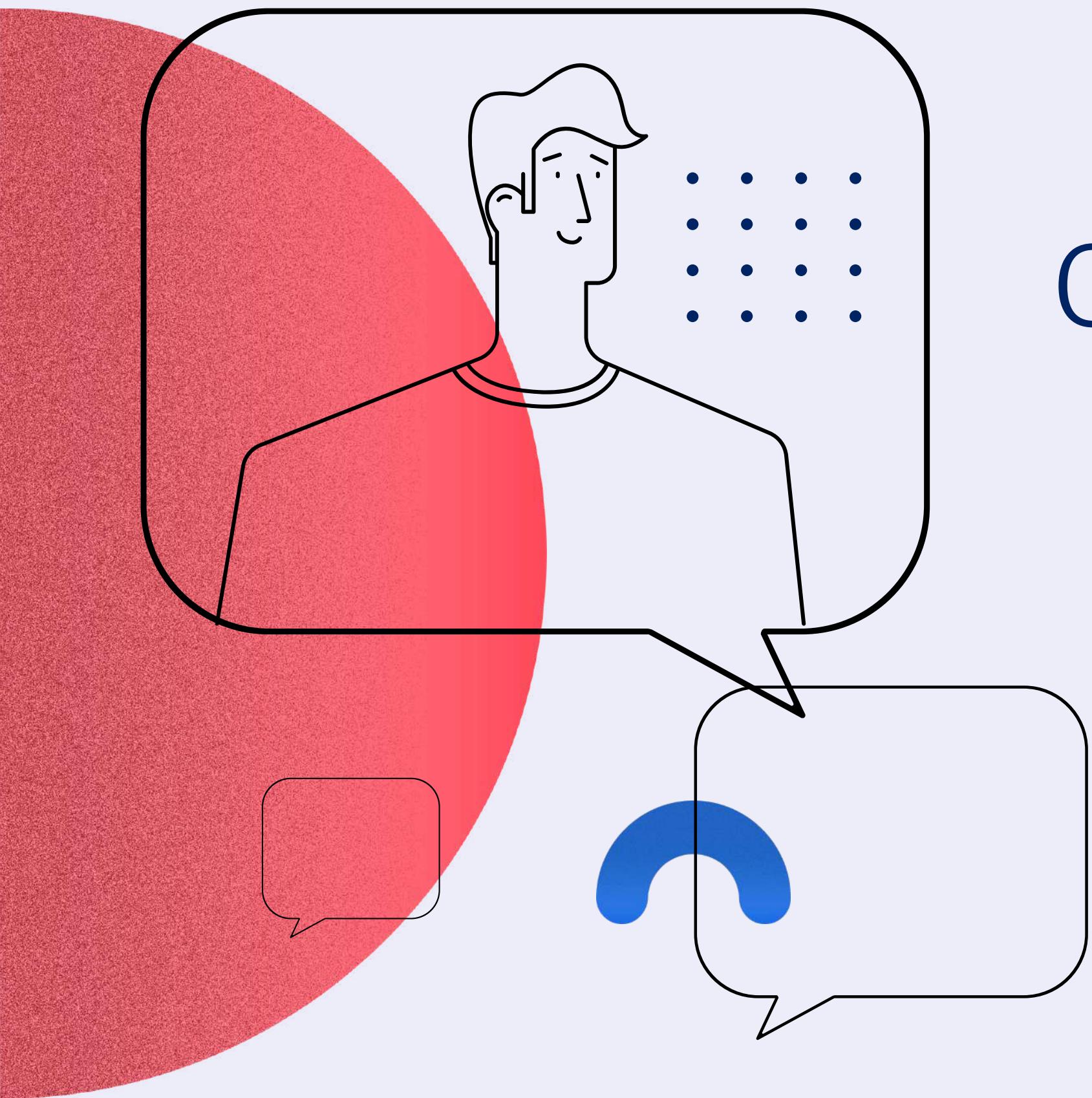
- **Bachelor of Mathematics,**
Mathematics Major
- Previously worked as Front end
Web Developer Freelance
- Data Scientist soon to be



<https://github.com/uildanila>



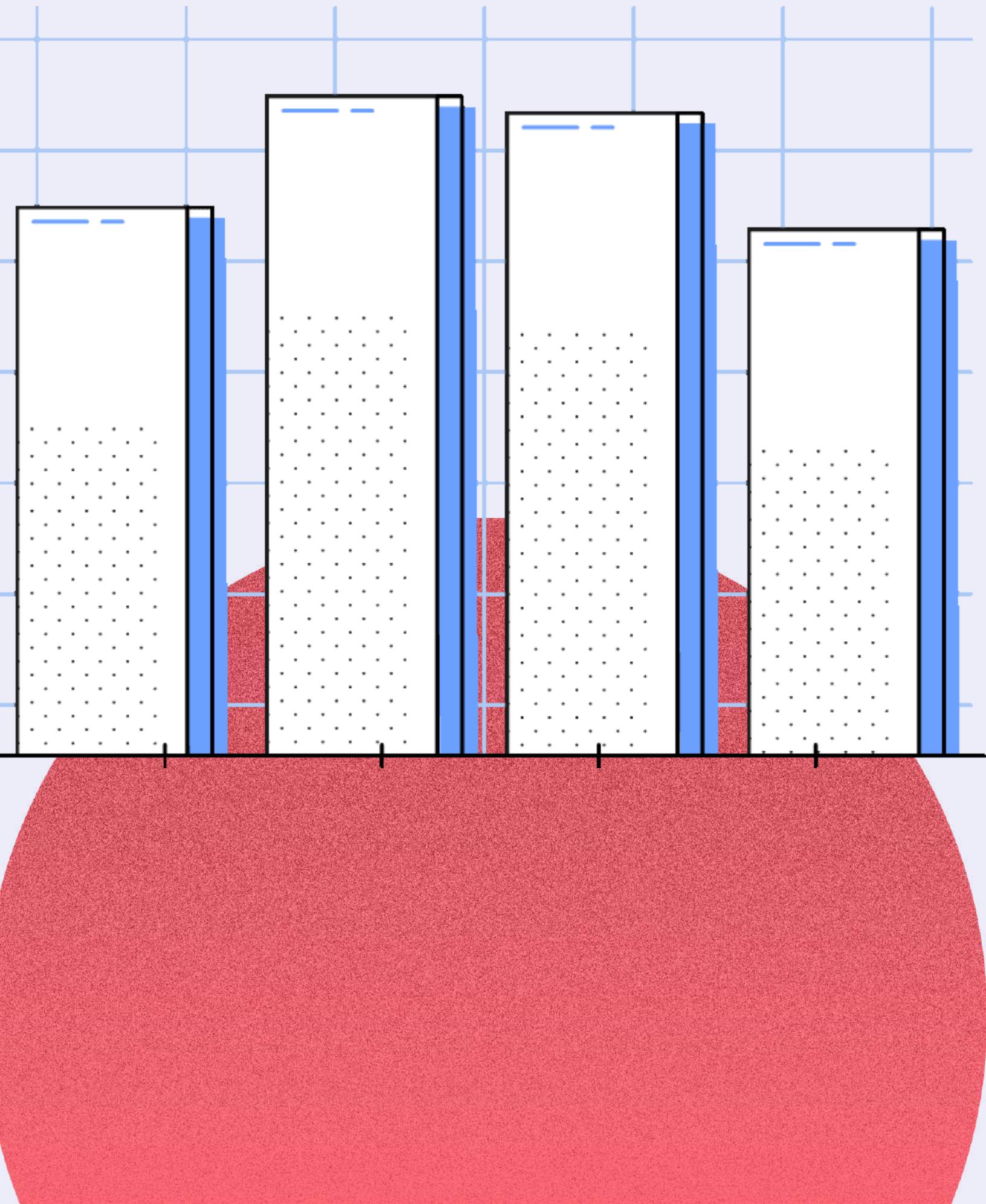
www.linkedin.com/in/nila-wildanul-husna



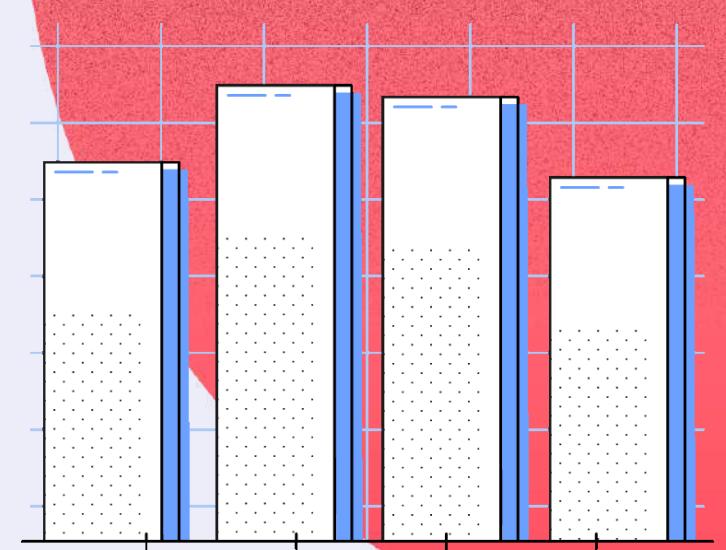
OUTLINES:

- Business Problem Understanding
- Data Understanding
- Exploratory Data Analysis
- Model Development
- Conclusion and Future Development

BUSINESS PROBLEM UNDERSTANDING



BUSINESS UNDERSTANDING



INTRODUCTION

- **Brazilian E-Commerce Public Dataset by Olist**
- Contains details of more than 100.000 orders from 2016 to 2018
- Growth and empowering commerce since they were developed

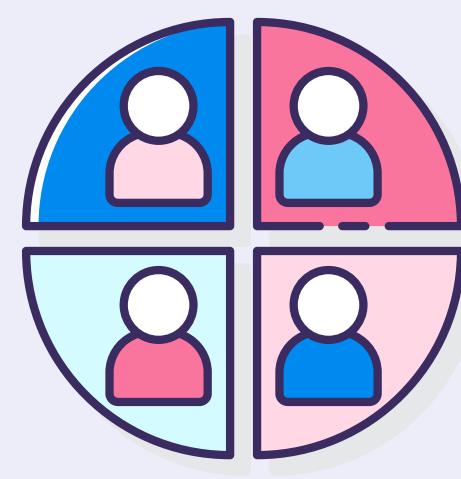
Limitations	Problem Statement
<ul style="list-style-type: none">• Based on Olist's Application• Presently Dataset	<ul style="list-style-type: none">• As our customer base increase, so does our customer variety• Categorize them into several segmentation for a more precise and personalized approach.

Analytic Approach	Goal
RFM Clustering ➔ Recommendation System	Increase Frequency and Monetary value.

BUSINESS UNDERSTANDING

OBJECTIVE

Making RFM segmentation model based on Customer transaction.



Customer Transaction Data

Learn the **Recency, Frequency, and Monetary** value of each customer

Classify Customers into several classes/segmentations and provided the basis for suggesting the right action (RecSys - Product Recommendation)

BENEFIT :



Know more about our customer

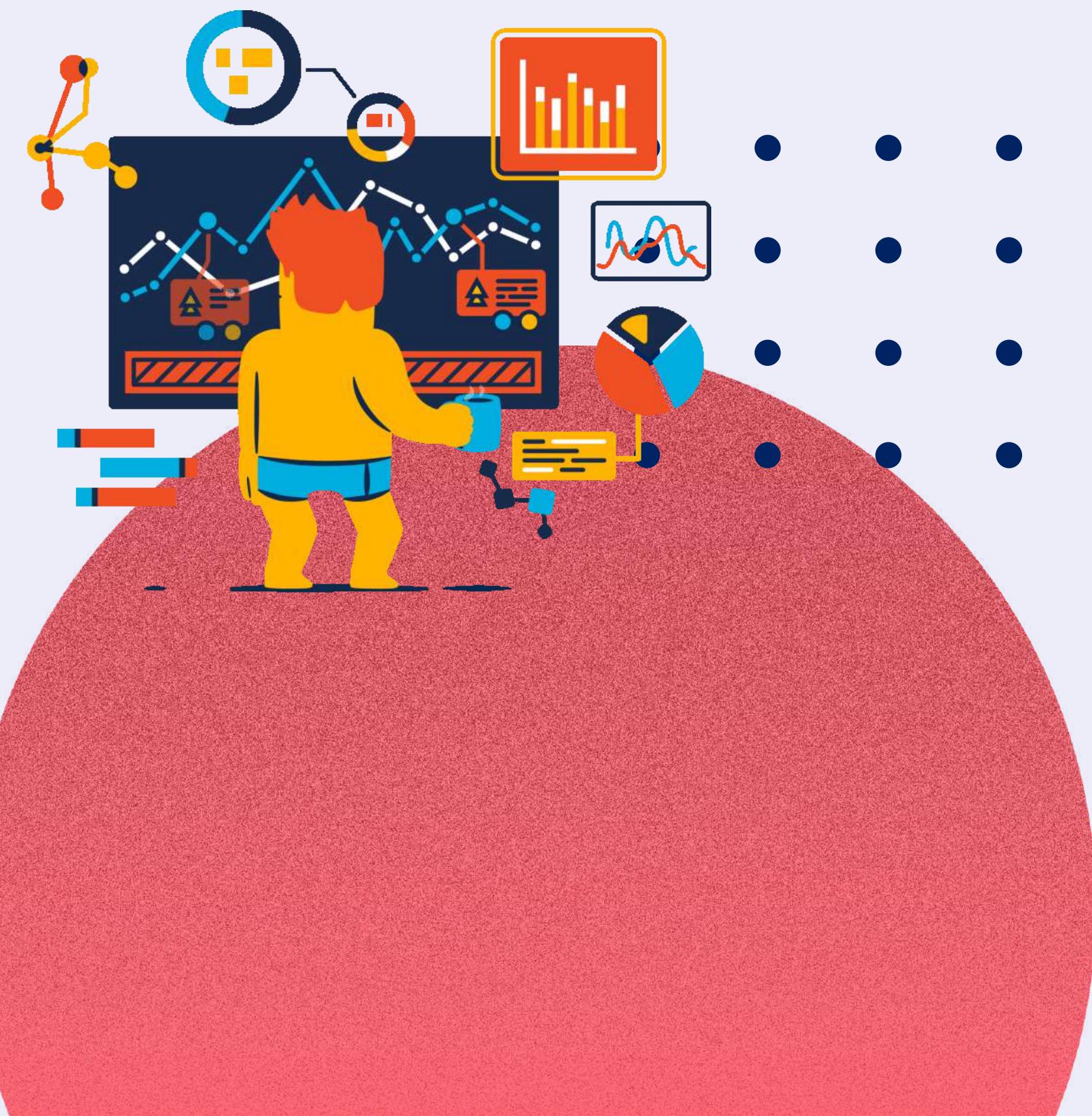


Provide a more personalized action

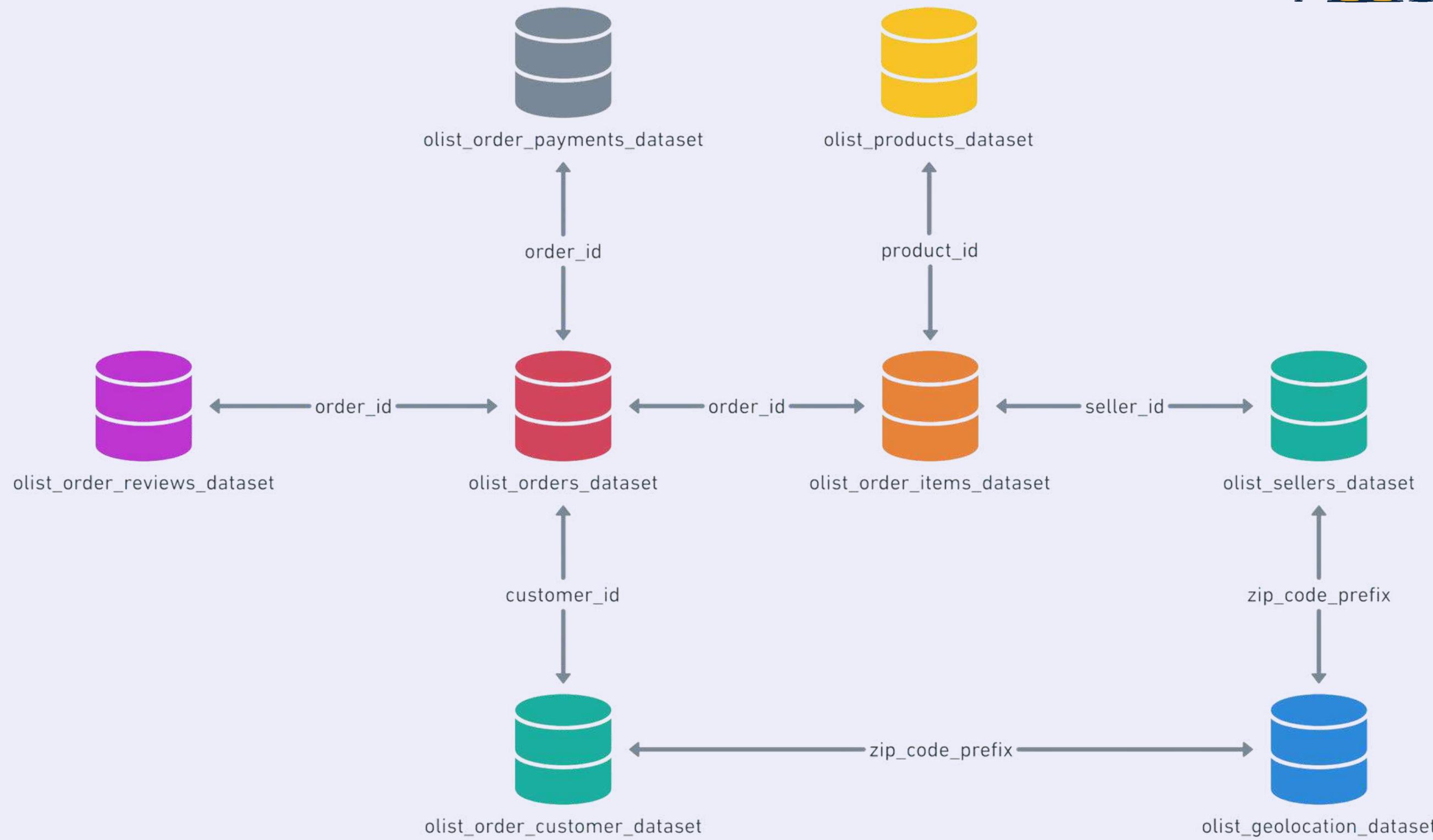


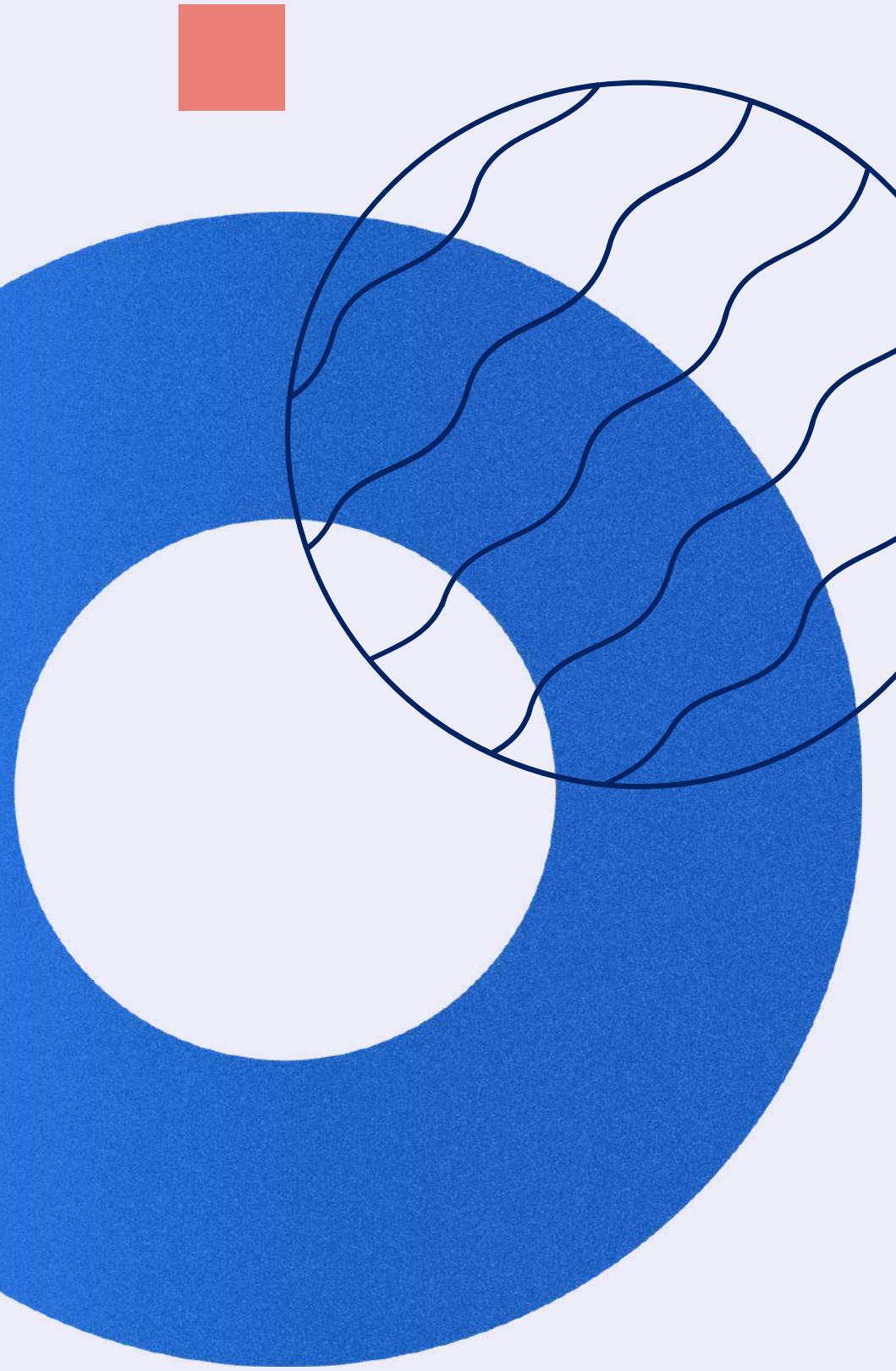
Can be used as a tool to maintain our customer

DATA UNDERSTANDING



DATA UNDERSTANDING





EXPLORATORY DATA ANALYSIS (EDA)

Our Insight

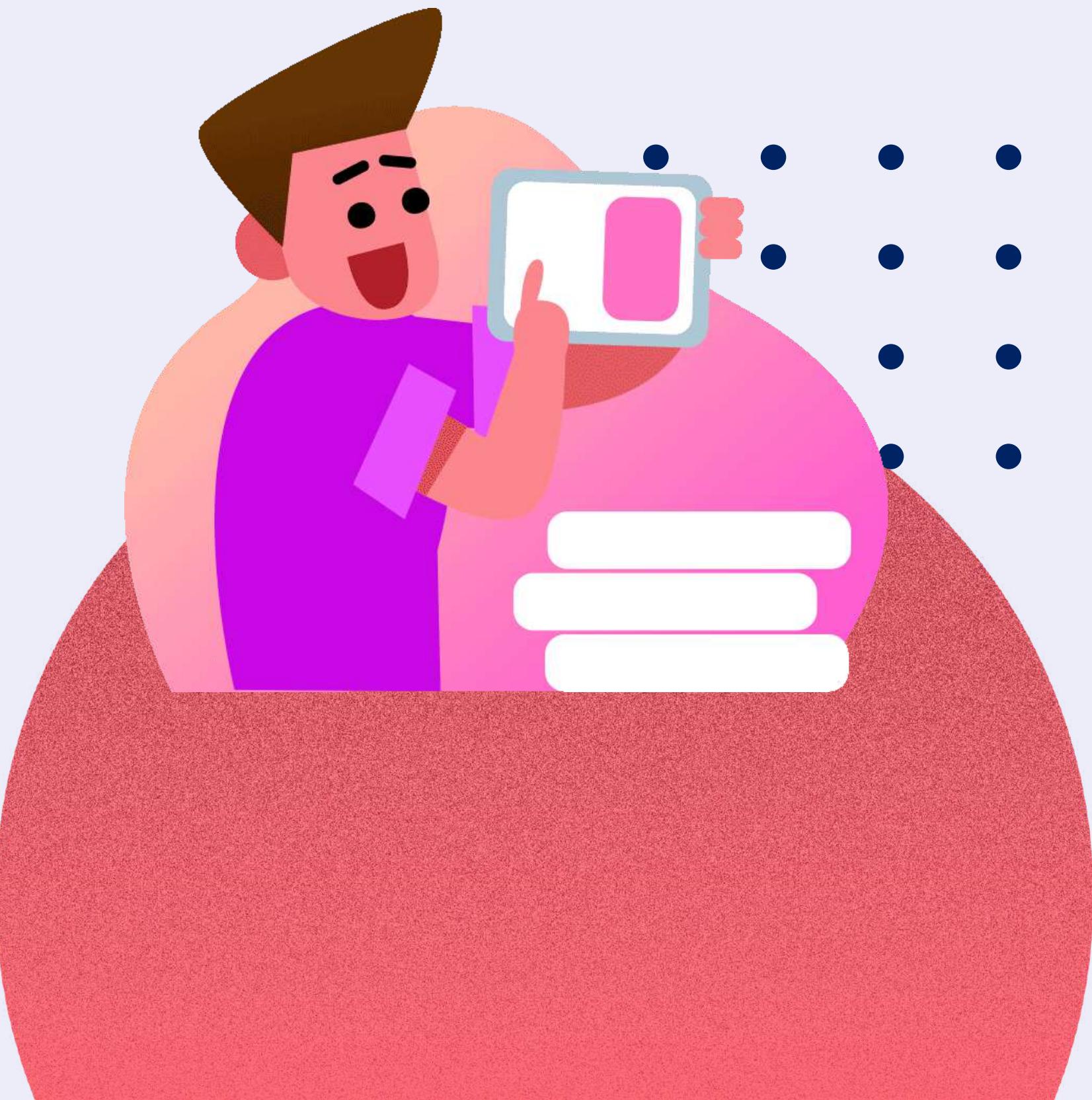


Customer Reach & Growth



Seller and Product Sold Track

METHODOLOGY





STEP #1 - SEGMENTATION

RFM Segmentation Metrics

R

Recency

F

Frequency

M

Monetary

The combination for each score of those dimensions will create different groups, and then the groups can be clustered into several segments

We will use K-Means **Clustering** to get the segmentation



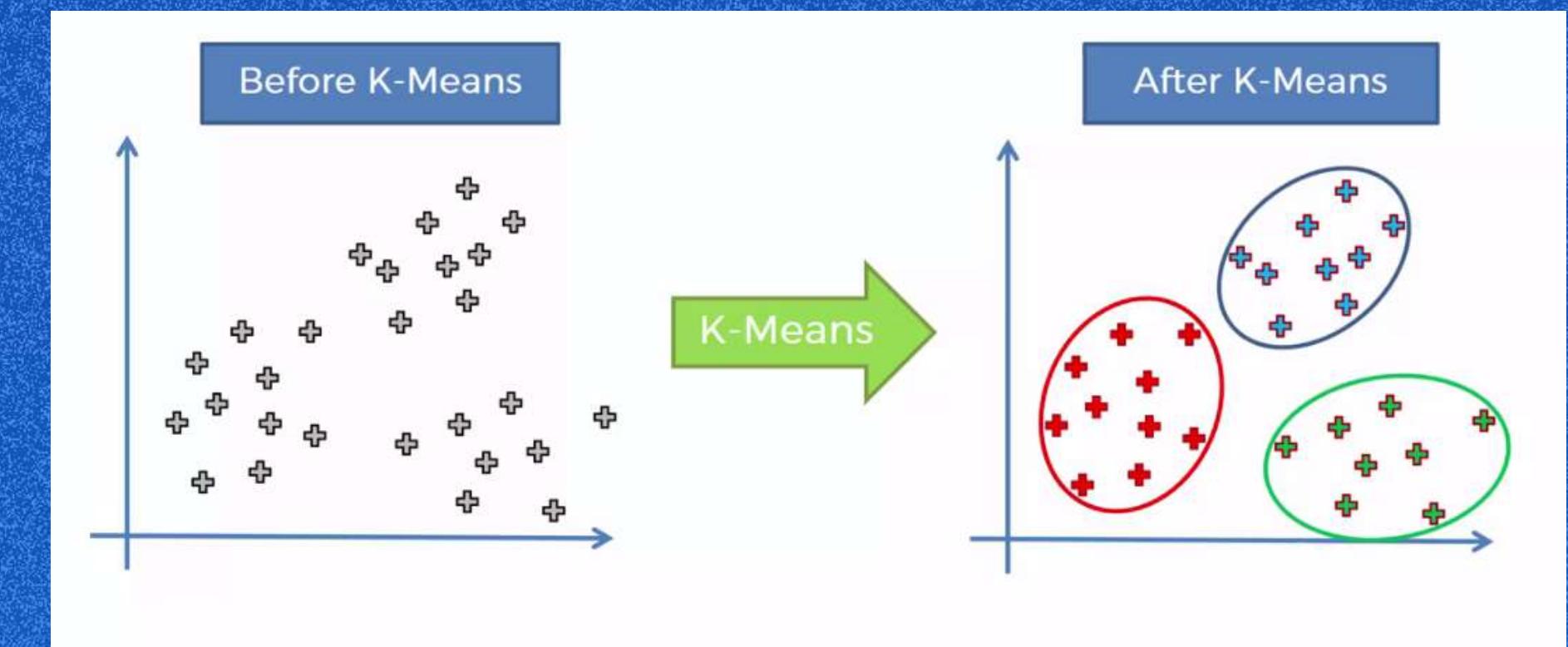
WHY RFM ?

- Who are our best customers ?
- Which customer is likely to churn?
- ***Who has the potential to become valuable customers ?***



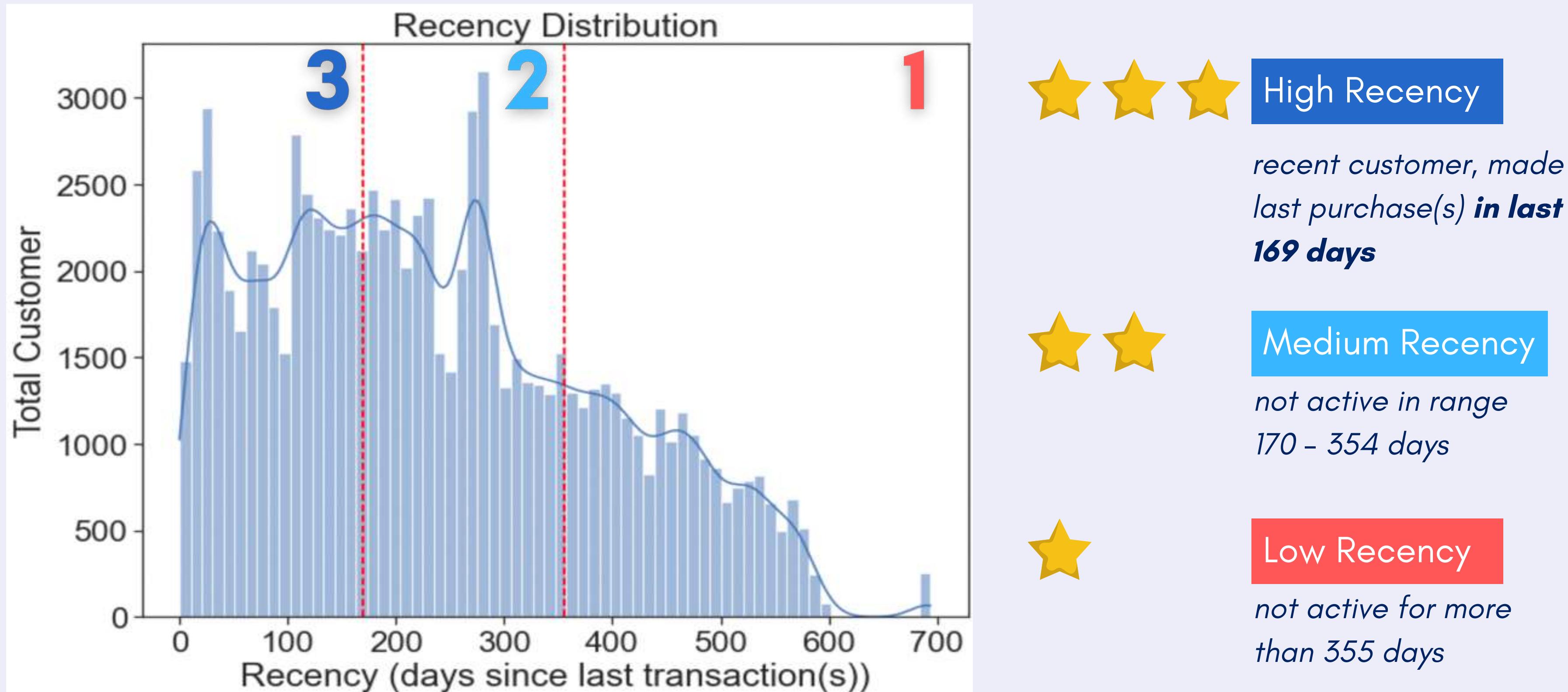
"CLUSTERING" IS THE PROCESS OF GROUPING SIMILAR ENTITIES TOGETHER

NON-HIERARCHICAL CLUSTERING INVOLVES FORMATION OF NEW CLUSTERS BY MERGING OR SPLITTING THE CLUSTERS



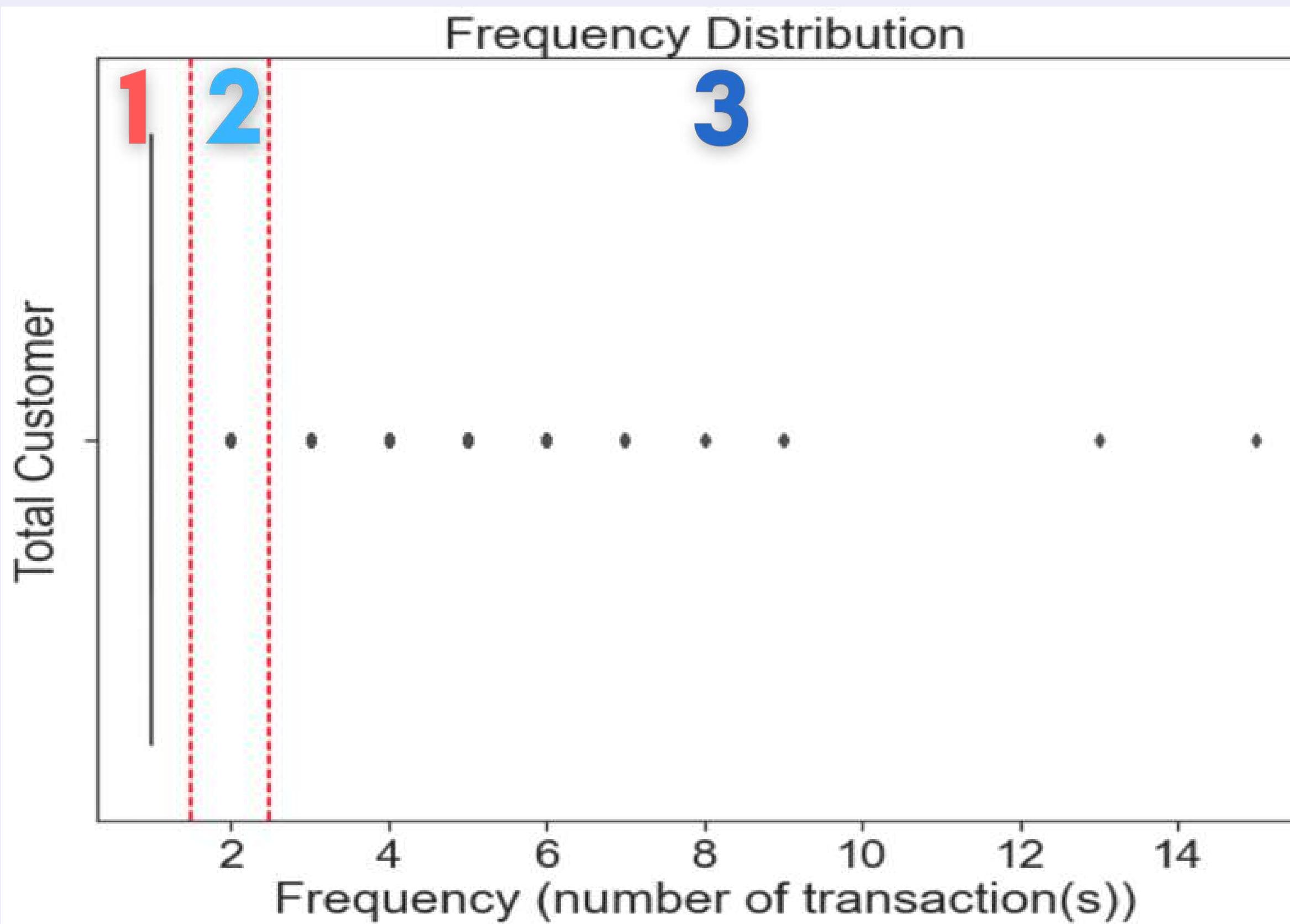
CLUSTERING : RECENCY

Days since last order



CLUSTERING : FREQUENCY

Number of Total Order



High Frequency



Medium Frequency



Low Frequency

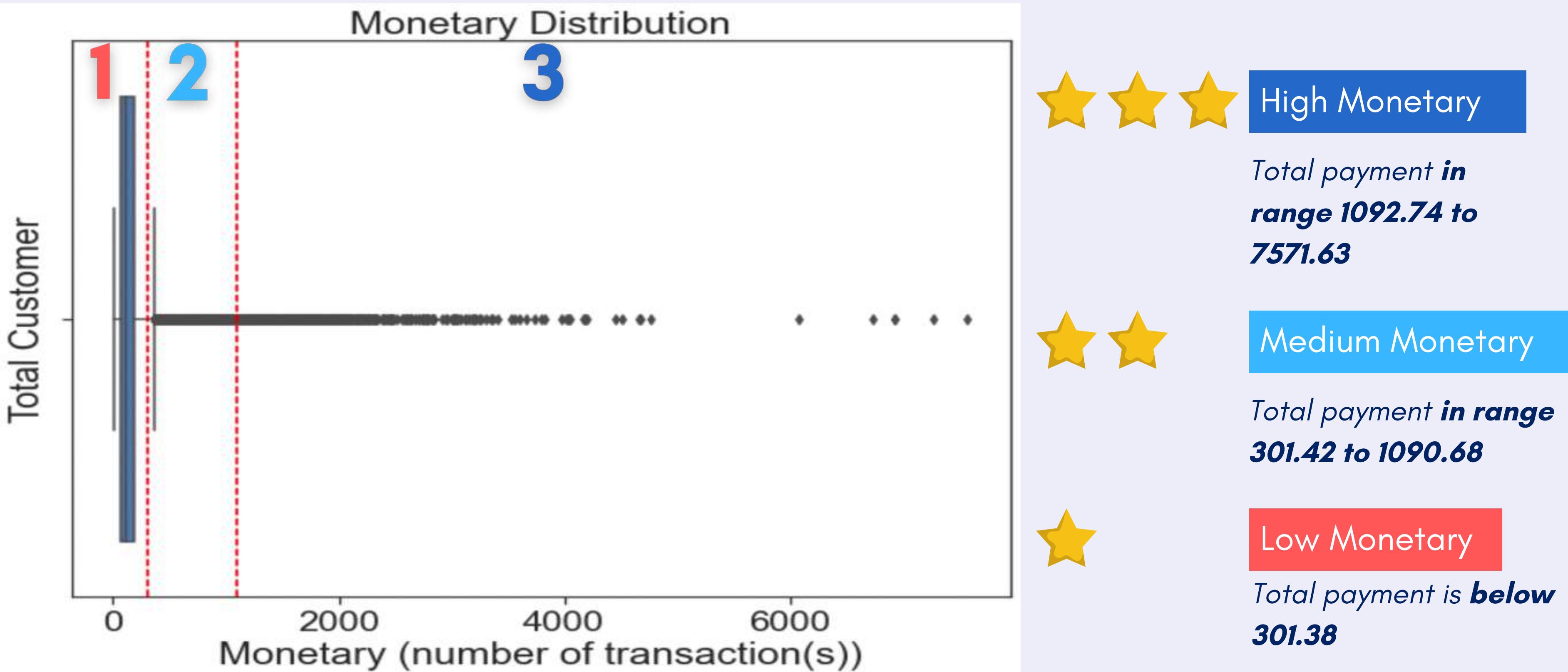
Multi order
(transactions above or equals 3 times)

Double order
(made 2 transactions)

Single order
(only made 1 transaction)

CLUSTERING : MONETARY

Amount of Total Payment Value





RFM SEGMENTATION

Customer Segment	Segment Description
Loyal Customer (Segment A)	Do transaction recently, often and spend the most money.
Casual Customer (Segment B)	Bought most recently, but not often. In medium recency, frequency and monetary values.
Need Attention (Segment C)	Our relatively new customer, has the highest recency score but still low on the frequency and monetary value.
About To Sleep - Lost Customer (Segment D)	Lowest recency, frequency and monetary scores. This customer are lost and/ or has never made any transaction in a very long time.

RFM SEGMENTATION SCORING RESULT

Recency Score	Frequency Score	Monetary Score	Segment
3	3	3	Loyal Customer
		2	Loyal Customer
		1	Need Attention
	2	3	Loyal Customer
		2	Need Attention
		1	Need Attention
	1	3	Loyal Customer
		2	Loyal Customer
		1	Need Attention
2	3	3	Casual Customer
		2	Casual Customer
		1	Casual Customer
	2	3	Casual Customer
		2	Casual Customer
		1	About To Sleep - Lost Customer
	1	3	Casual Customer
		2	About To Sleep - Lost Customer
		1	About To Sleep - Lost Customer
1	3	3	About To Sleep - Lost Customer
		2	About To Sleep - Lost Customer
		1	About To Sleep - Lost Customer
	2	3	About To Sleep - Lost Customer
		2	About To Sleep - Lost Customer
		1	About To Sleep - Lost Customer
	1	3	About To Sleep - Lost Customer
		2	About To Sleep - Lost Customer
		1	About To Sleep - Lost Customer

RFM SCORING DETAILS



EDA of Segmentation Result

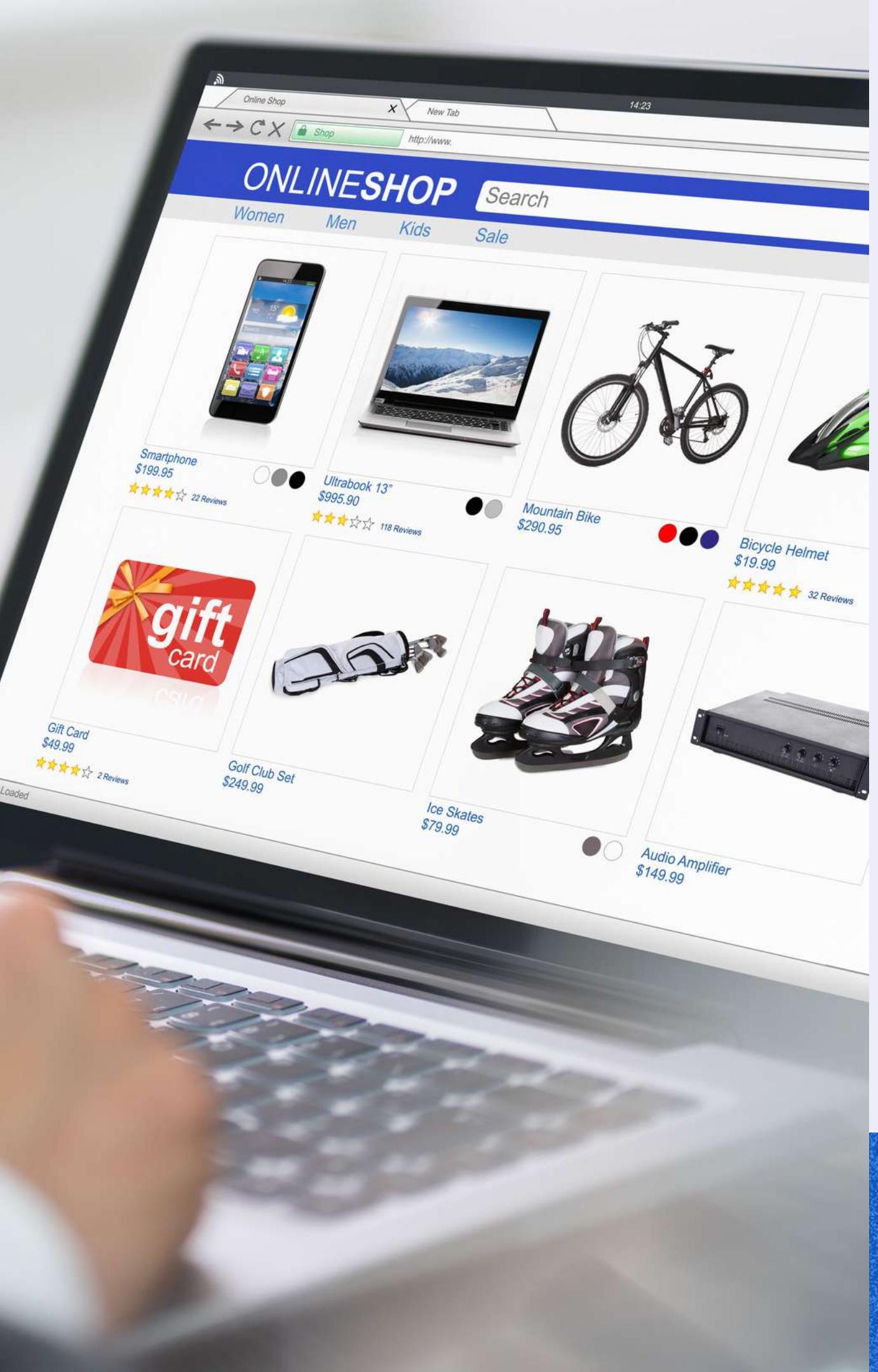


Inactive Days (R)

Number of Orders (F)

Total Payment (M)

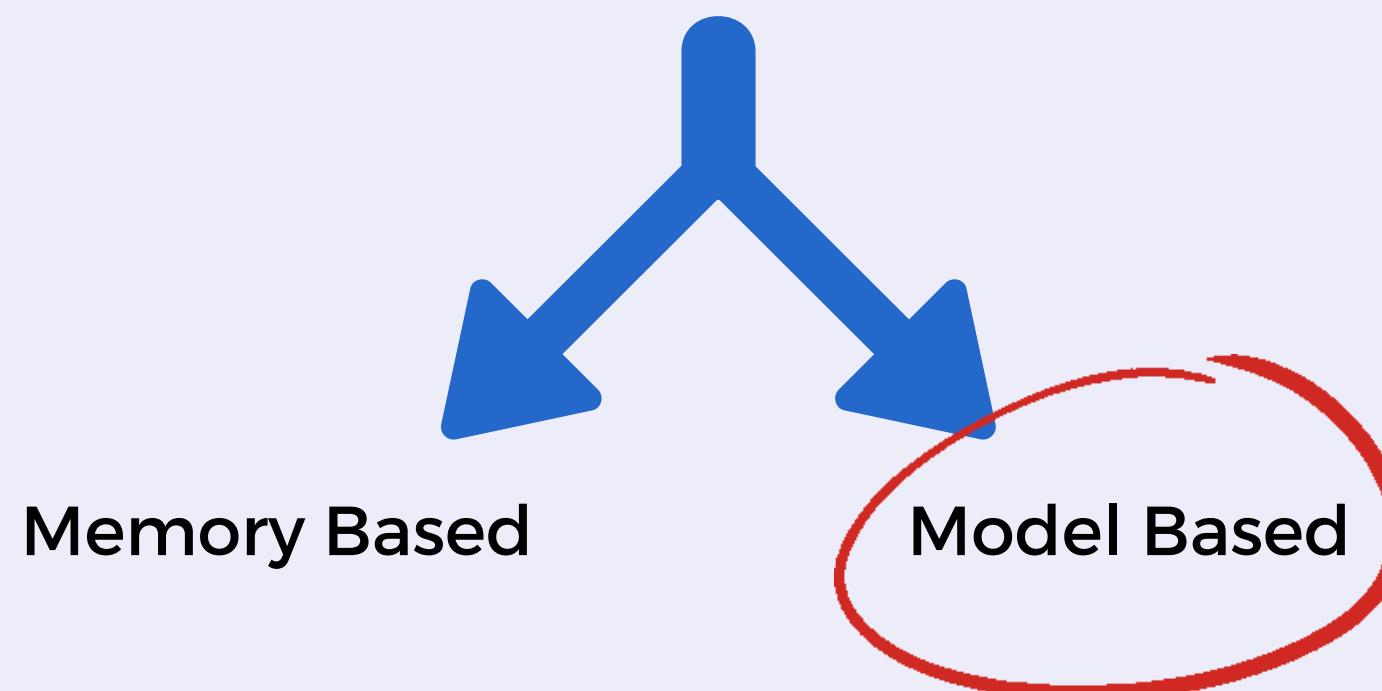
CUSTOMER SEGMENTATION	COUNT	MIN	MAX	MIN	MAX	MIN	MAX
Loyal Customer (Segment A)	3,552	0.0	169.0	1	15	301.42	7274.88
Casual Customer (Segment B)	972	170.0	354.0	1	13	83.04	6081.54
Need Attention (Segment C)	31,941	0.0	169.0	1	7	9.59	1075.23
About To Sleep - Lost Customer (Segment D)	55,063	170.0	694.0	1	7	10.07	7571.63



STEP #2 - RECOMMENDATION SYSTEM

Collaborative Filtering

Collaborative filtering is a technique that can filter out items that a user might like based on the interest of similar users.



SVD (Singular Value Decomposition)



EVALUATION

RECOMMENDATION SYSTEM - #EVALUATION PER CUSTOMER

#CASE 1 - CUSTOMER 'C74'

ACTUAL



	product	product_category_name_english	seller	price	review_score
1	P33	perfumery	S16	59.9	5.0
851	P21326	consoles_games	S127	98.0	5.0

RECOMMENDATION AND PREDICTED REVIEW RESULT



	product	product_category_name_english	seller	price	avg_review_score	Predicted_Review
0	P10381	toys	S33	319.99	5.0	5.000000
1	P2345	fashion_bags_accessories	S432	49.00	5.0	5.000000
2	P26200	housewares	S2370	89.90	5.0	4.978884

EVALUATION



	customer	product	review_score	Predicted_Review
0	C74	P33	5.0	4.632600
1	C74	P21326	5.0	4.852404

RECOMMENDATION SYSTEM

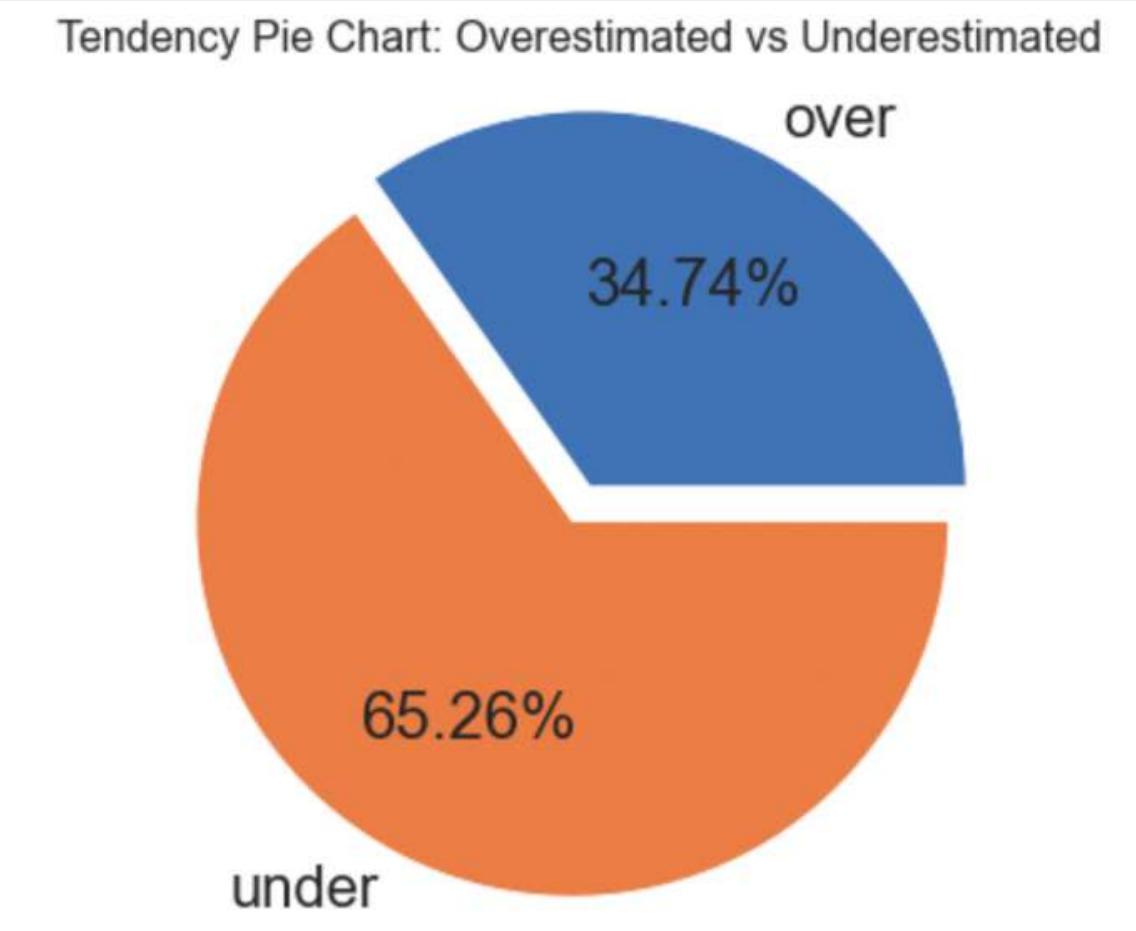
#EVALUATION CUSTOMER SEGMENT C (NEED ATTENTION)

COMPARISON RESULT (ACTUAL VS PREDICTED REVIEW SCORE)

	customer	product	Predicted_Review	review_score	discrepancy	tendency
0	C1	P1	4.162262	4.0	0.162262	over
1	C1	P20012	4.800593	5.0	-0.199407	under
2	C1	P3424	4.767873	5.0	-0.232127	under
3	C10181	P13323	4.572971	5.0	-0.427029	under
4	C10181	P2113	4.014990	4.0	0.014990	over
...
33992	C9645	P1208	4.537367	5.0	-0.462633	under
33993	C9645	P22292	4.873845	5.0	-0.126155	under
33994	C9645	P5220	4.883238	5.0	-0.116762	under
33995	C9682	P17427	4.801570	5.0	-0.198430	under
33996	C9682	P5240	4.790880	5.0	-0.209120	under

MAE SCORE : 0.9875

TENDENCY PERCENTAGE



CONCLUSION & FUTURE DEVELOPMENT



CONCLUSION

- Customer base is still rapidly growing, **attention to this type of customers to increase their frequency and monetary value.**
- Using our RecSys model, customer will **get top product recommendation** based on their purchase history.
- **Customer class 'C' have potential to be our loyal customer** if we increase their frequency and monetary value.
- With our RecSys Model, we've got MAE score of 0.98. **However, as our focus is not to predict customer's review score but to recommend products that might be interested in**, this result doesn't necessarily mean that the products we recommended would miss

• • • •
• • • •
• • • •

RECOMMENDATION - FOR BUSINESS

- For Class 'A' (Loyal Customer) and 'B' (Casual Customers), **we could implement loyalty program**
- **Customer class 'C' have potential to be loyal customer**, recommend products from our model.
- For class 'D', we could try to email blast them with our product recommendation and/or other promotions such as **Comeback Customer Sale**.
- We could **consider building warehouse(s)** in Sao Paulo in order **to cut 'delivery time' and maybe even 'freight value'**

• • • •
• • • •
• • • •

RECOMMENDATION - FOR DEVELOPMENT

- Try to develop current model to get a better result, **using different algorithm such as SlopeOne, CoClustering, SVDpp, etc.**
- **Tuning using different hyperparameter values**, to get a better metrics (MAE)
- **Use different approach other than RFM**, such as Customer Lifetime Value, etc.

• • •
• • •
• • •



DEPLOYMENT



[WebApp Deployment Model](#)



**THANK
YOU**

The text "THANK YOU" is centered in the upper half of the image. The word "THANK" is in dark blue, bold, sans-serif capital letters, and "YOU" is in red, bold, sans-serif capital letters. Both words have a thin pink outline. Above the text, several hands of different skin tones (white, brown, orange) reach upwards from the bottom right, center, and left side, all pointing towards a single blue downward-pointing triangle at the top center.