# OLIST BRAZILIAN E-COMMERCE

## Recommendation System

*Satria Fajar Bagaskara*
*Dibimbing Data Scientist Batch 18*

linkedin.com/in/satriabagaskara/

github.com/dataaga/Olist_CollaborativeFiltering

# TABLE OF CONTENTS


01 PROBLEM BACKGROUND

02 THE DATA THAT WE HAVE

03 OBJECTIVE

04 EXPLORATORY DATA ANALYSIS

05 RECOMMENDATION SYSTEM

06 CONCLUSION & SUGGESTION

# 01 PROBLEM BACKGROUND

# TOO MANY PRODUCTS MAKES PEOPLE CONFUSES

IN OTHER HAND, THE SELLER HAVE OBLIGATORY TO PROVIDE WHOLE PRODUCTS FOR ALL DAILY NEEDS

Photo by Atoms on Unsplash

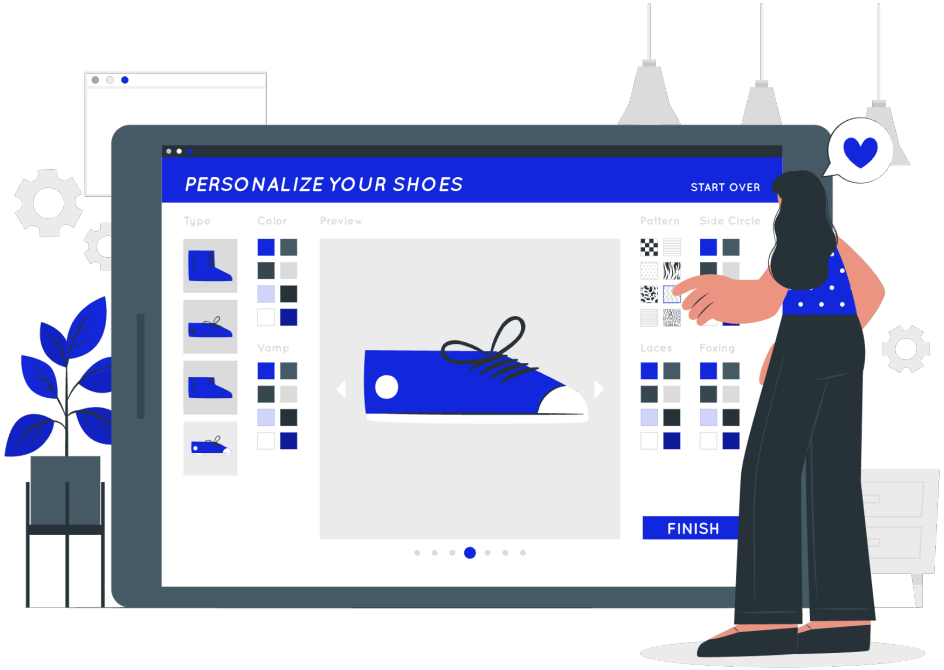
# People is tending to following the MAJORITY OF CHOICE*

which end up with popular products

*Bandwagon effect: https://www.wallstreetmojo.com/bandwagon-effect/

Photo by chuttersnap on Unsplash

Photo by JESHOOTS.COM on Unsplash

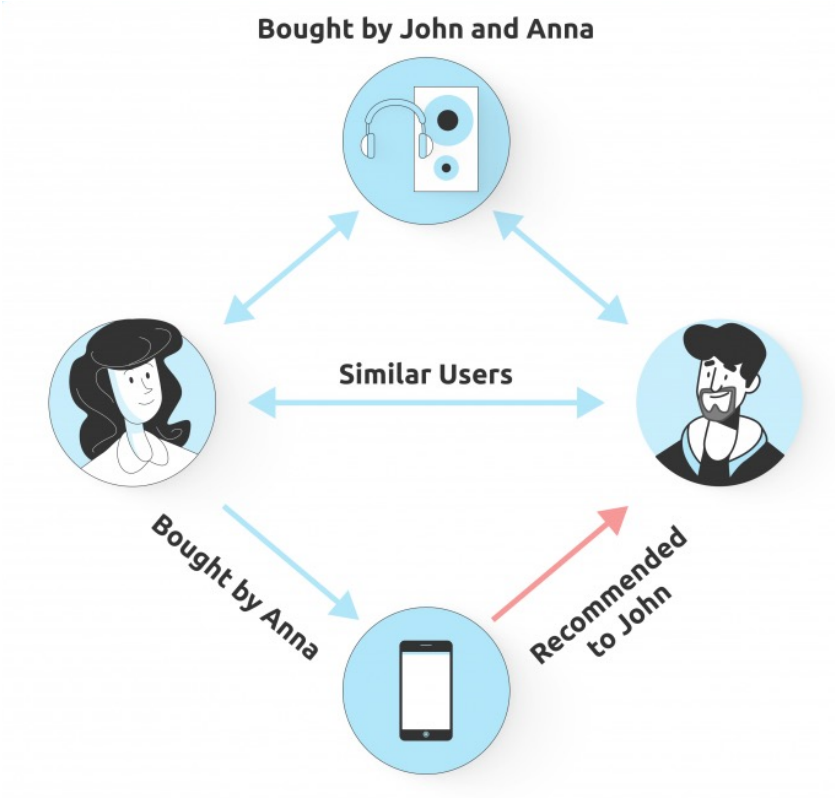



Because of that, we need to make

# Personalized Buying Experience

## by recommending their own preferences

which the tools called recommendation system




**One of method that can be used is**

# Collaborative Filtering

by using preferences of *similar users* in *the past* to *get your* preference in *the future*.

Source: https://influencermarketinghub.com/ecommerce-recommendation-system/

# 02 THE DATA THAT WE HAVE

# DATA SOURCE

**100k**

**ORDERS**

**2016 until 2018**

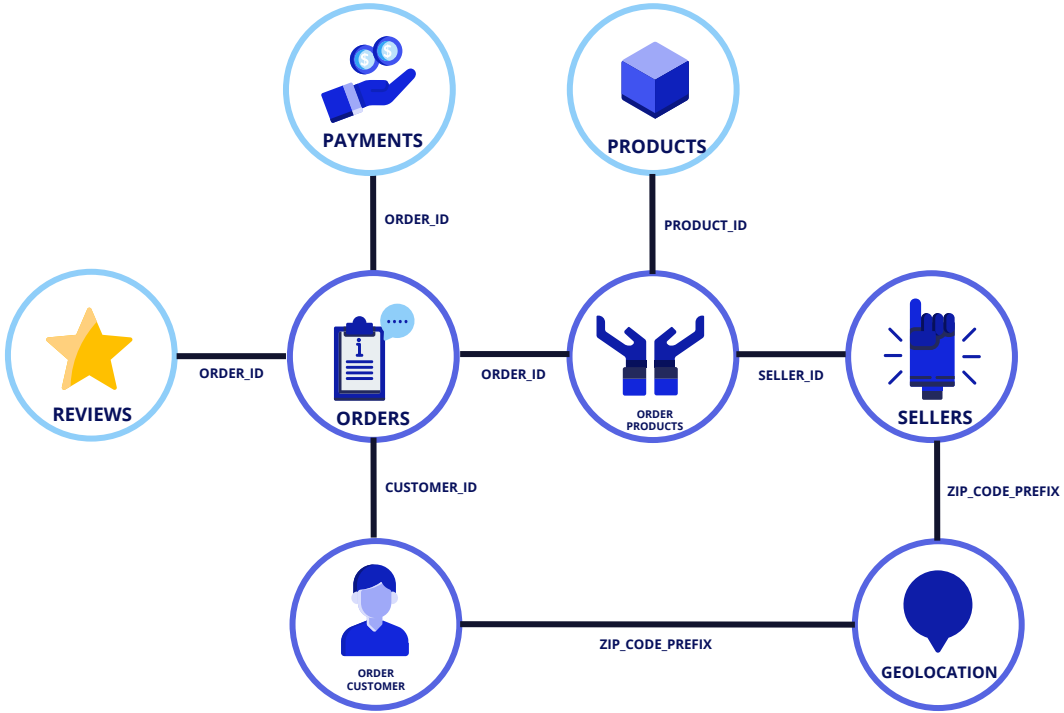
**PERIOD**

**9**

**DATASETS**

**Olist** is the largest department store in Brazilian e-commerce. Olist **connects small businesses** from all over Brazil to channels without hassle and with a single contract. Those merchants are able to **sell their products** through the Olist Store and **ship them directly to the customers** using Olist logistics partners.

# OLIST DATA SCHEME



# 03 OBJECTIVE

# GOALS

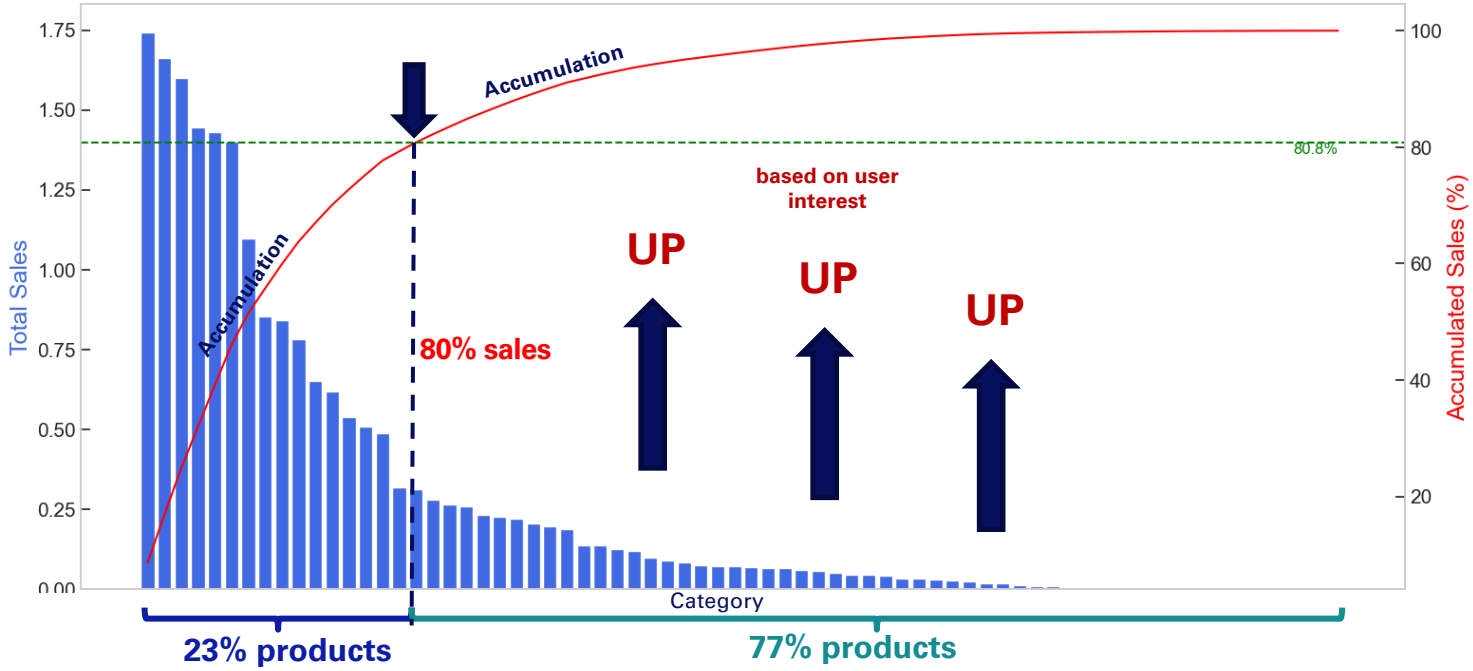
**PRODUCT** **CUSTOMER** **RATING**

Olist needs **collaborative filtering** to *increase sales* by give nearest suggestion **products** based on **customer interest and preferences**

# 04 EXPLORATORY DATA ANALYSIS

# 1. PRODUCT

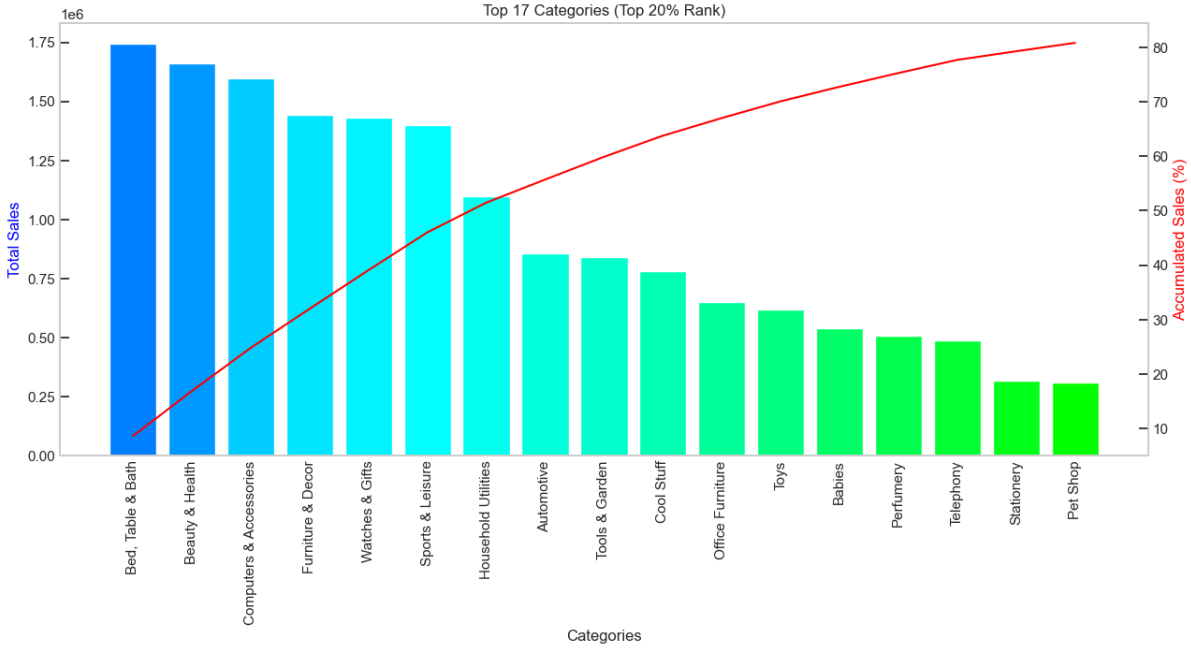
# TOP 17 CATEGORIES FROM 72 CATEGORIES




**23% products** contribute **80% sales** which align with the pareto & long-tail
Hence there're still a lot of opportunities to recommend other products

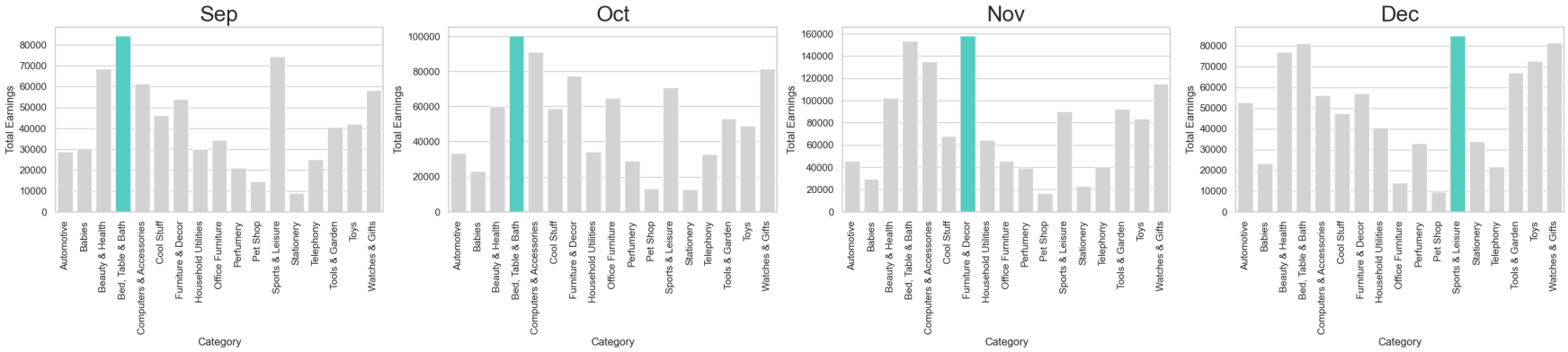
https://www.nngroup.com/articles/long-tail/

# TOP 17 CATEGORIES (TOP 20% RANK)



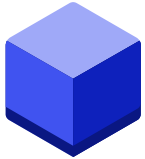

**Top 3** are **Bed, Table & Bath, Beauty & Health,** and **Computer & Accessories**
**Next top 3** are **Furniture & Décor, Watches & Gifts,** and **Sports & leisure**

# MONTHLY BEST SELLING CATEGORIES



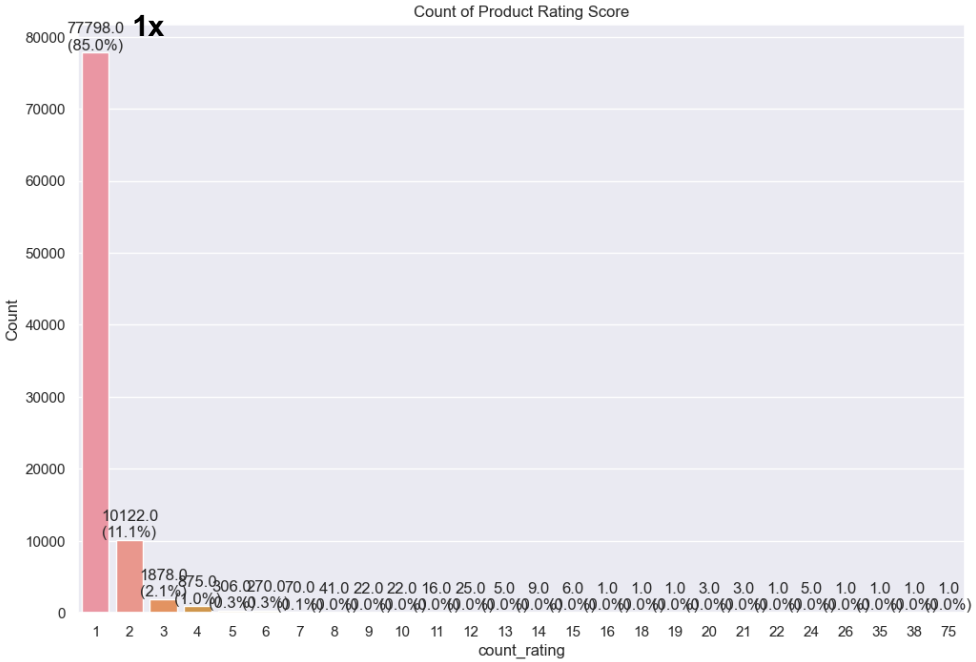

**3x consecutively** on **Computer & Accessories**
**4x times** on **Bed Table & Bath**
**3x times** on **Beauty & Health**

Hence monthly top category product could use for first-time customer since they still don't have their purchase data.



# 2. RATING

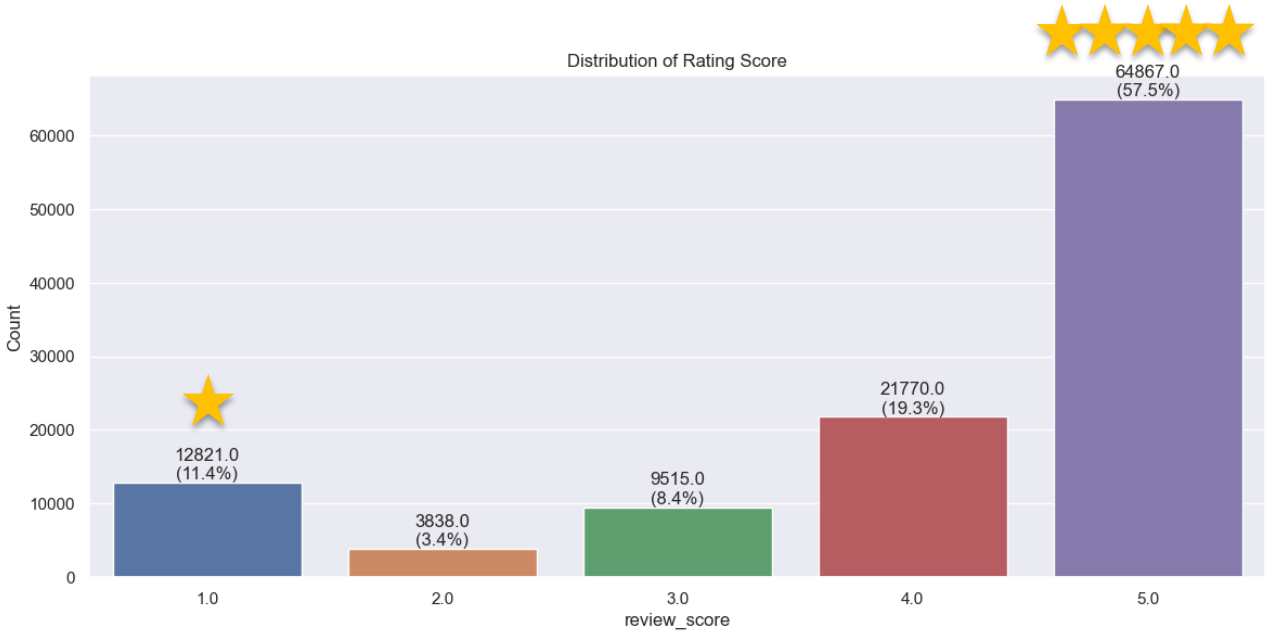
# AMOUNT OF PRODUCT RATING SCORE




Mostly Customer **Rated** only **1 Times** with almost **85%**

*It seems likely rate by First-Time Customer, and could be Repeat Customer don't want to give rating*

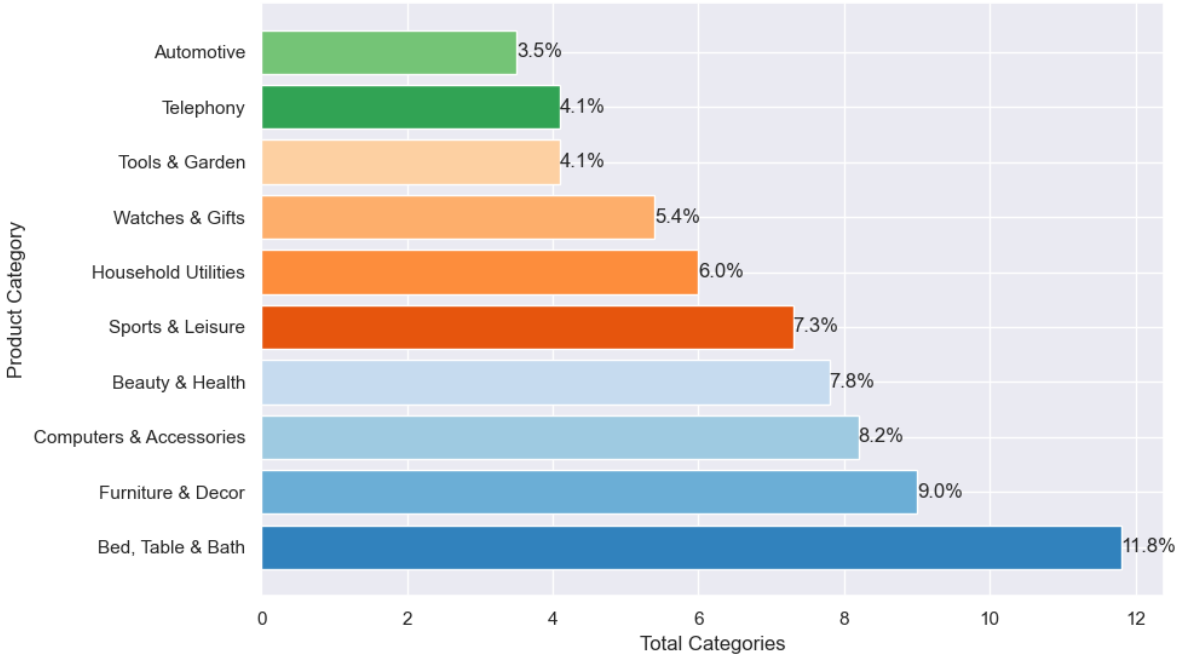
# RATING SCORE OF PRODUCT




**5 Rating** Score with mostly **57%**

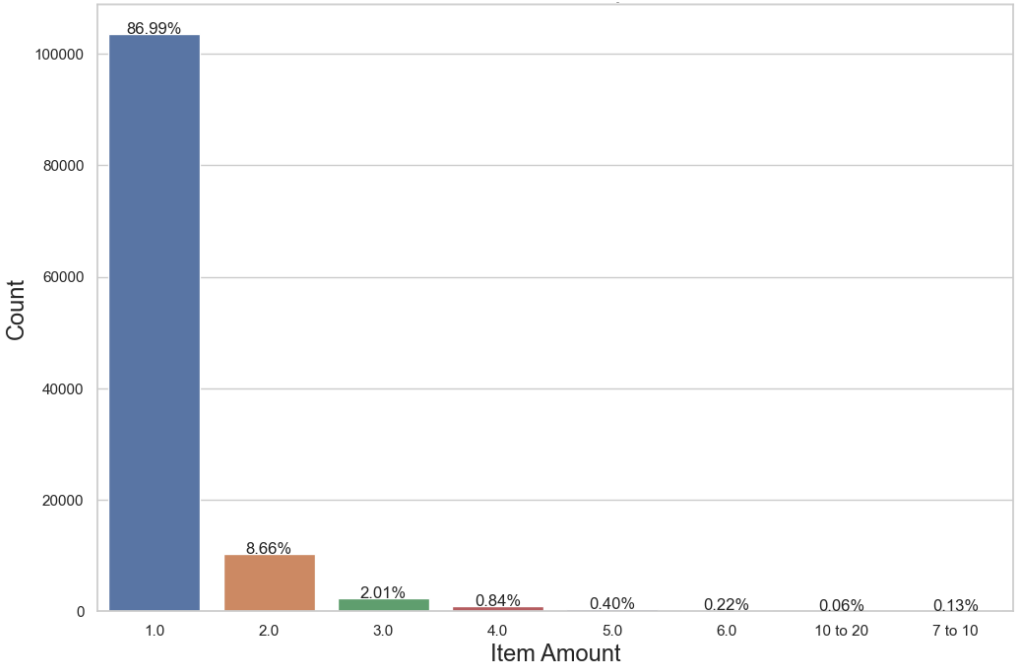
***except*** *for* ***1 star score*** *had high score, furthermore I would inspect on the further step.*

# TOTAL PRODUCT CATEGORY WITH 1 SCORE




Unfortunately, some of **Top Category Product** has **Rated** with **1 Score**
Hence there's need service improvement

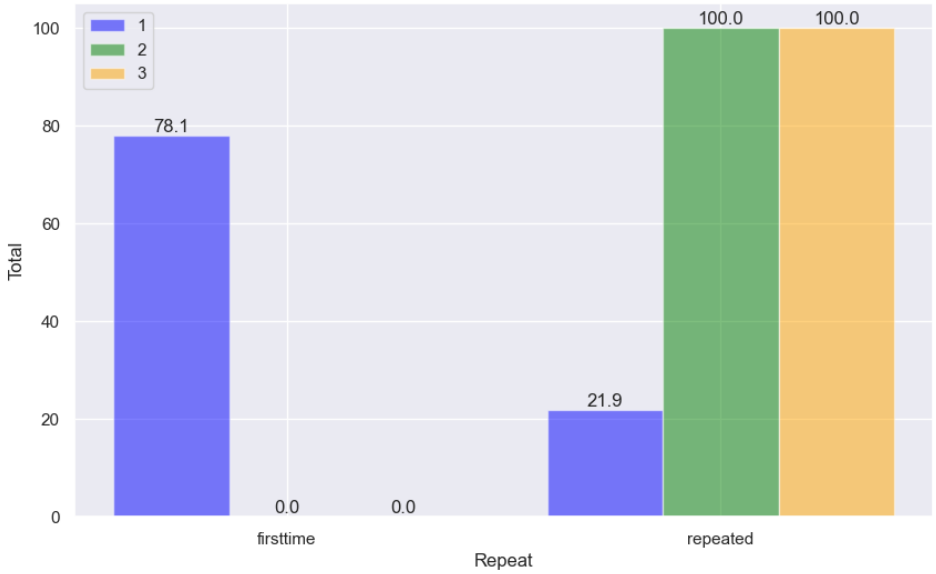
# DISTRIBUTION OF PRODUCT PER ORDER




Almost **87% order** is only bought **1 product**
Hence it align with the given product ratings count that still 1x times

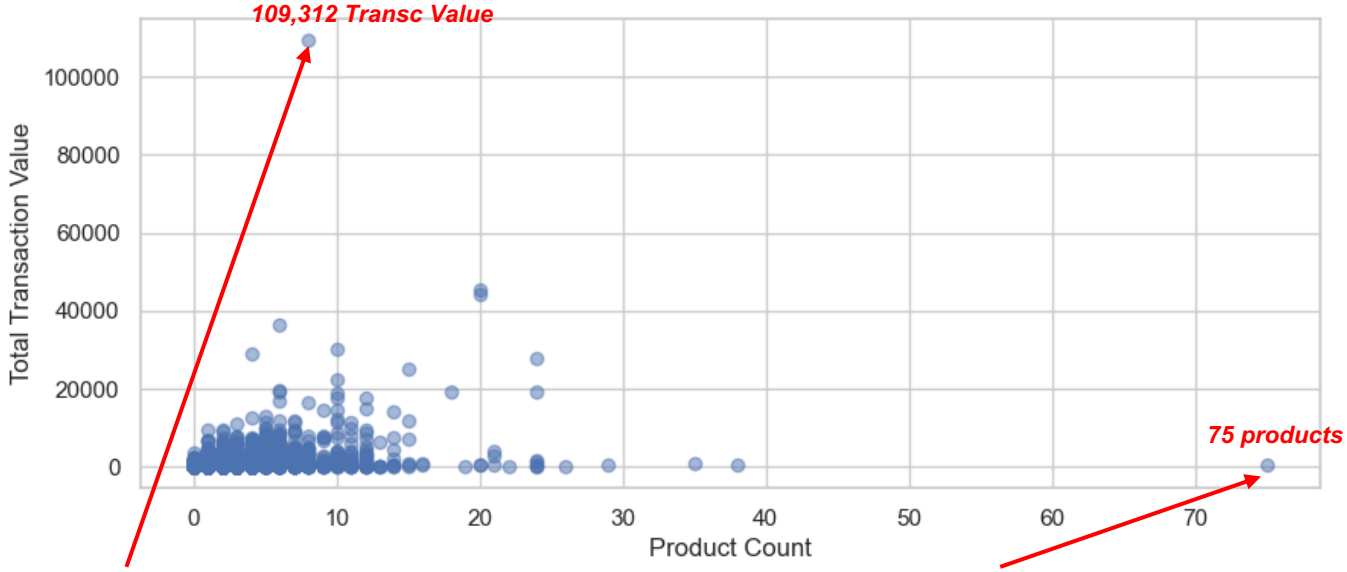
# 3. CUSTOMER

# First-Time vs Repeater:
# HOW MANY PRODUCT PER ORDER?




78% First-Time Customer only **buy 1 product**
vs
100% Repeat Customer buy **2-3 products**

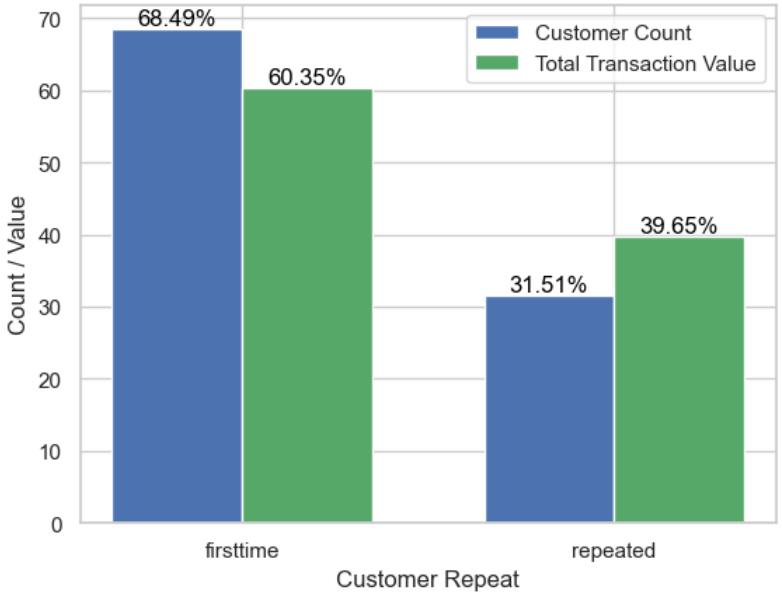
# Customer per Order:
# HIGHEST PRODUCT AMOUNT
vs
# HIGHEST TRANSCATION VALUE




Surprisely, the customer bought 8x **landline telephone** using credit card
Need further analysis to validate this.

**Most of them included to 17 Top Products** such as furniture & décor; bed, table & bath; and household utilities using voucher

# Transaction Value:
# REPEAT vs FIRST-TIME CUSTOMER?



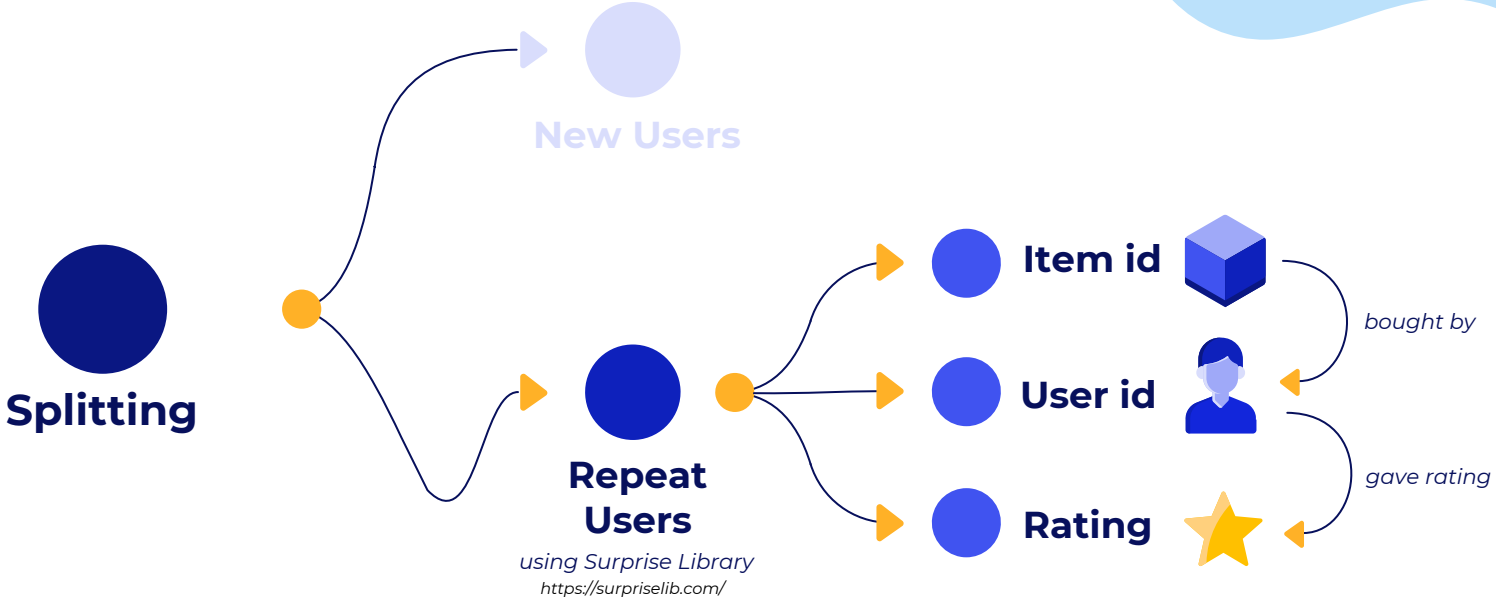

**Repeat customers made repeat purchases of the product by seeing total transaction value is higher than their amount.**

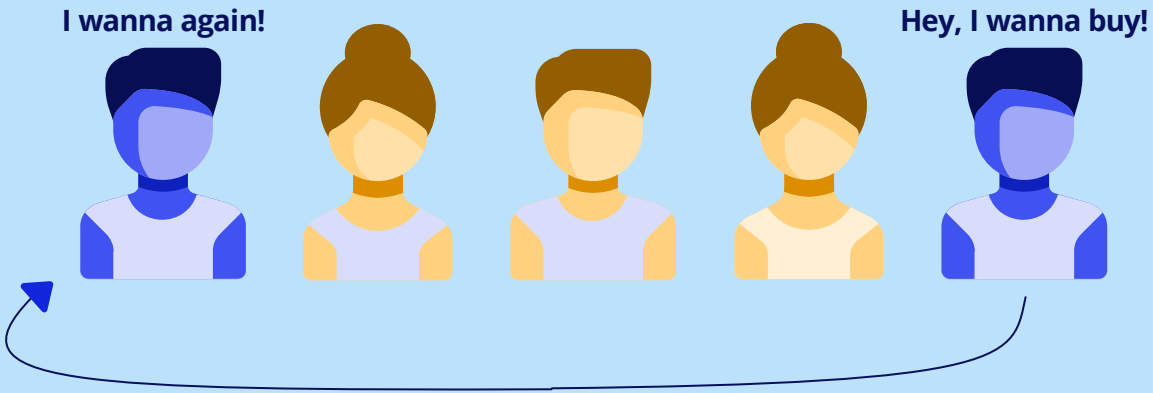
Hence there're still lot of opportunities to gain more selling on repeat customer and first-time customer

05

# RECOMMENDATION SYSTEM

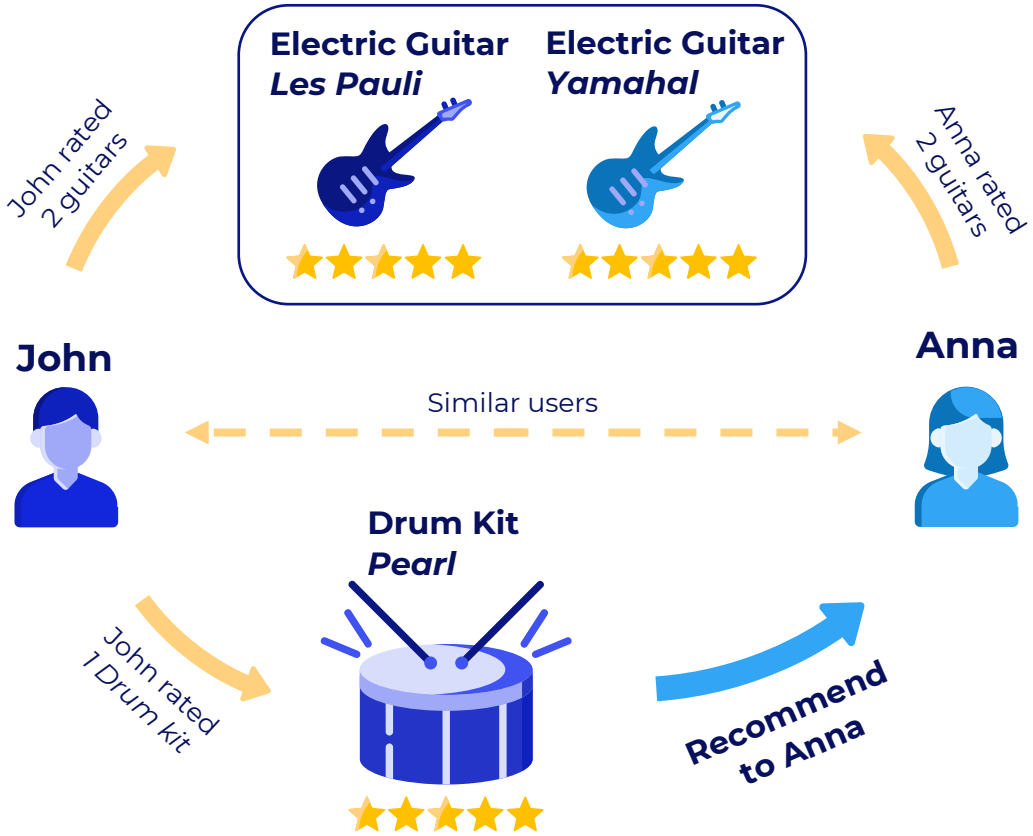
# RECOMMENDER
## FOR REPEAT USERS






# Recommender for Repeat Customer

# RECOMMENDATION SYSTEM




Collaborative Filtering models use the **collaborative power** of ratings provided by community ratings and user ratings

Source: https://www.stat.ipb.ac.id/main/wp-content/uploads/2021/01/Introduction-to-RecSys_BERT4Rec-IPB.pdf

# RECOMMENDATION SYSTEM

## Collaborative Filtering - *User Based*

| | | | |
|---|---|---|---|
| Jona | ★★ | ★★★ | |
| John | ★★★★ | ★★★★★ | ★★ |
| Anna | ★★★★ | ? | ★★ |

Source: https://www.stat.ipb.ac.id/main/wp-content/uploads/2021/01/Introduction-to-RecSys_BERT4Rec-IPB.pdf

# 7 Algorithm Models
## on Surprise Library

Normalpred

SVD

SVD++

NMF

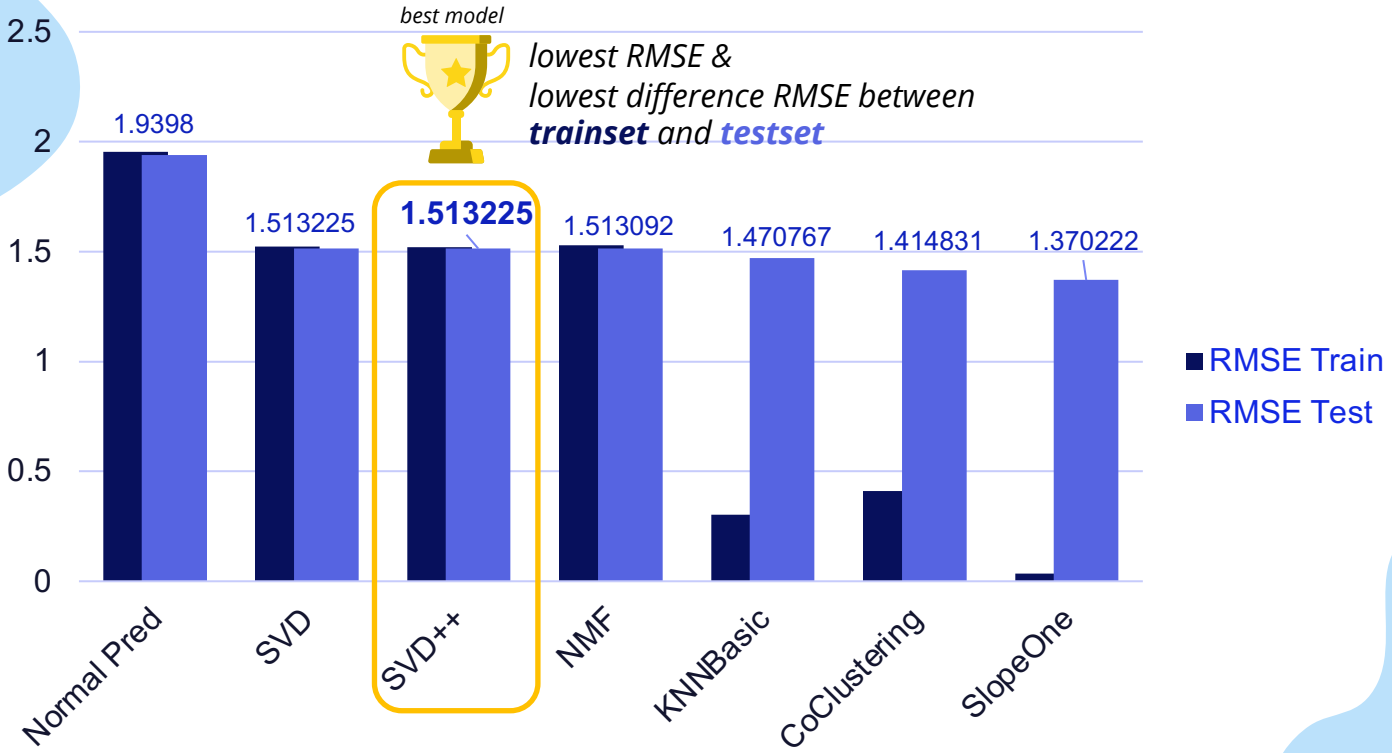
KNNBasic

Coclustering

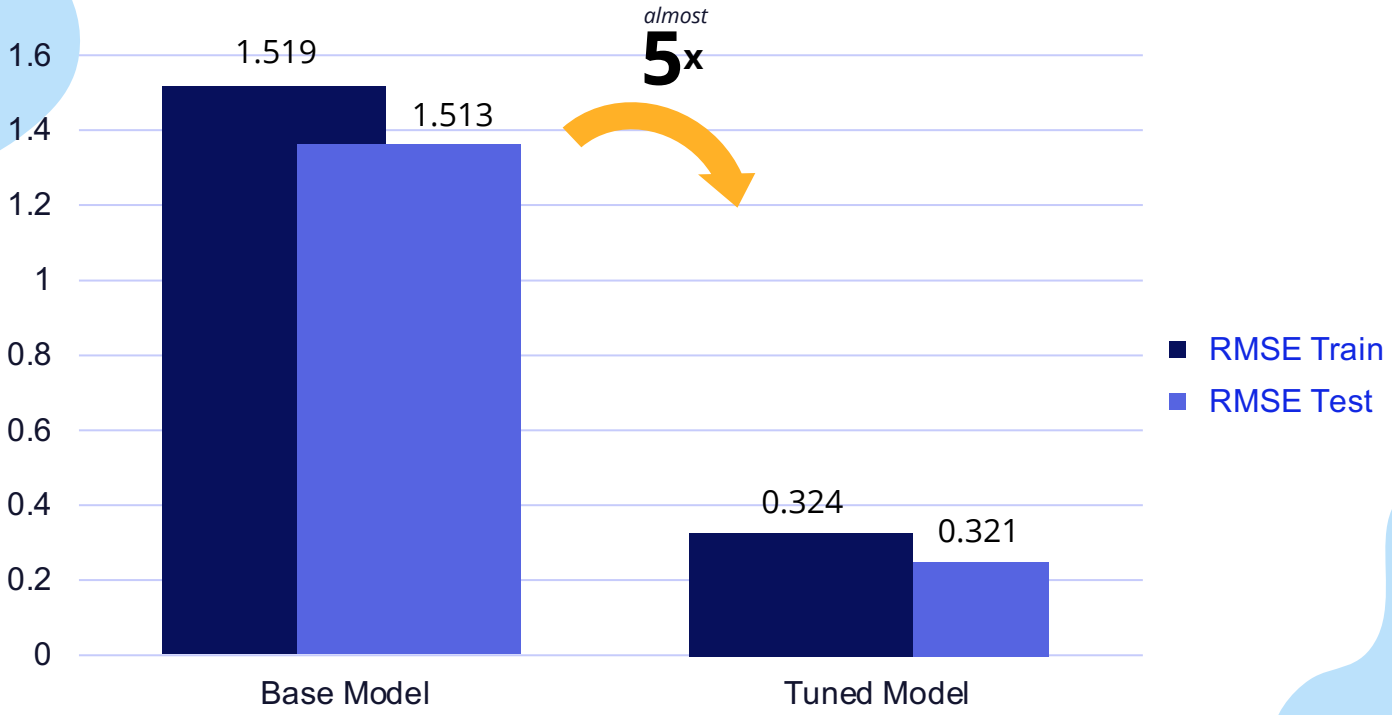
SlopeOne

**FINDING BEST MODEL**

# Modeling Result



# Improved Model after using GridSearchCV



# Testing The Model

## Best Prediction

User | Item | Actual | Predicted | Error

| | uid | iid | rui | est | details | Iu | Ui | err |
|---|---|---|---|---|---|---|---|---|
| 3188 | 4f82360423c19895d2a8183e6d8ec701 | bf66f5d110af02b50af37038afe90bd9 | 1.0 | 1.0 | {'was_impossible': False} | 2 | 2 | 0.0 |
| 100 | a133f658a4d462264b8d4d8cbb282393 | 7c6fbb3a5346dfd607386155c6f628f8 | 5.0 | 5.0 | {'was_impossible': False} | 2 | 3 | 0.0 |
| 2123 | b0d60f871dec79cae101d9e74c816407 | 65223c26538a2226610efc437e488b77 | 1.0 | 1.0 | {'was_impossible': False} | 2 | 2 | 0.0 |

Joni bought Lexung Car with **Actual Rating 1** Score otherwise the **Predicted Rating 1** Score too, Hence the **difference/error is 0**. Meanwhile this Car was rated by 2 person.

## Worst Prediction

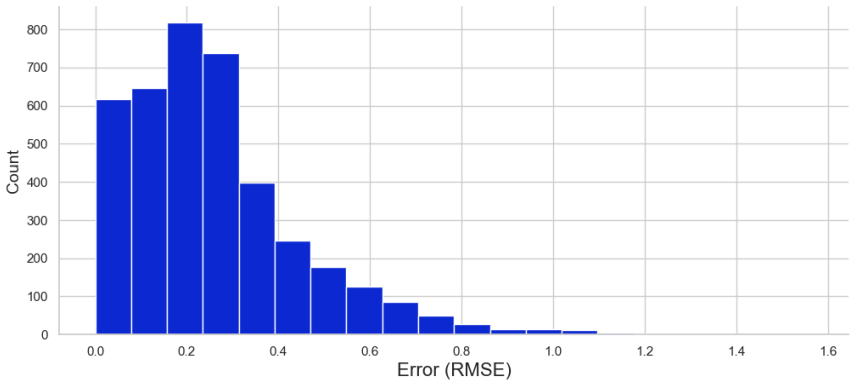
User | Item | Actual | Predicted | Error

| | uid | iid | rui | est | details | Iu | Ui | err |
|---|---|---|---|---|---|---|---|---|
| 2568 | c49760f3652f57abae1908ae0a452350 | 87d780fa7d2cf3710aa02dc4ca8db985 | 1.0 | 2.076481 | {'was_impossible': False} | 1 | 11 | 1.076481 |
| 618 | 50ad7151ad494370e8dc57a57351d17e | f71973c922ccaab05514a36a8bc741b8 | 1.0 | 2.081859 | {'was_impossible': False} | 1 | 9 | 1.081859 |
| 260 | 4f88e6285f805c9a823ba23291a65ca3 | f77dd338d9f75229a09cbb9a18fd0c9a | 1.0 | 2.091208 | {'was_impossible': False} | 2 | 8 | 1.091208 |

Kamila bought Cupid Doll with **Actual Rating 1** Score otherwise the **Predicted Rating 2.1** Score, Hence the **difference/error is 1.1**. Meanwhile this Doll was rated by 11 person.

# Evaluating The Model

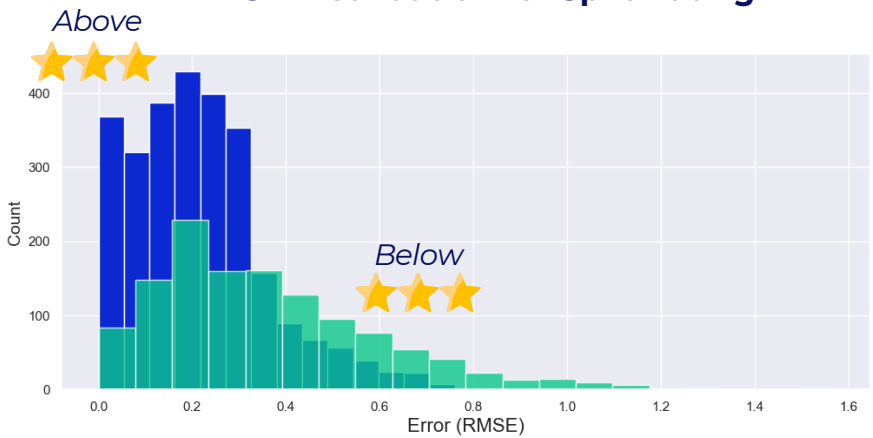
## RMSE Distribution for All Rating




The model had the **good performance** so far for **all rating score**

For all rating, the model had RMSE with skewed right distribution

## RMSE Distribution for Split Rating



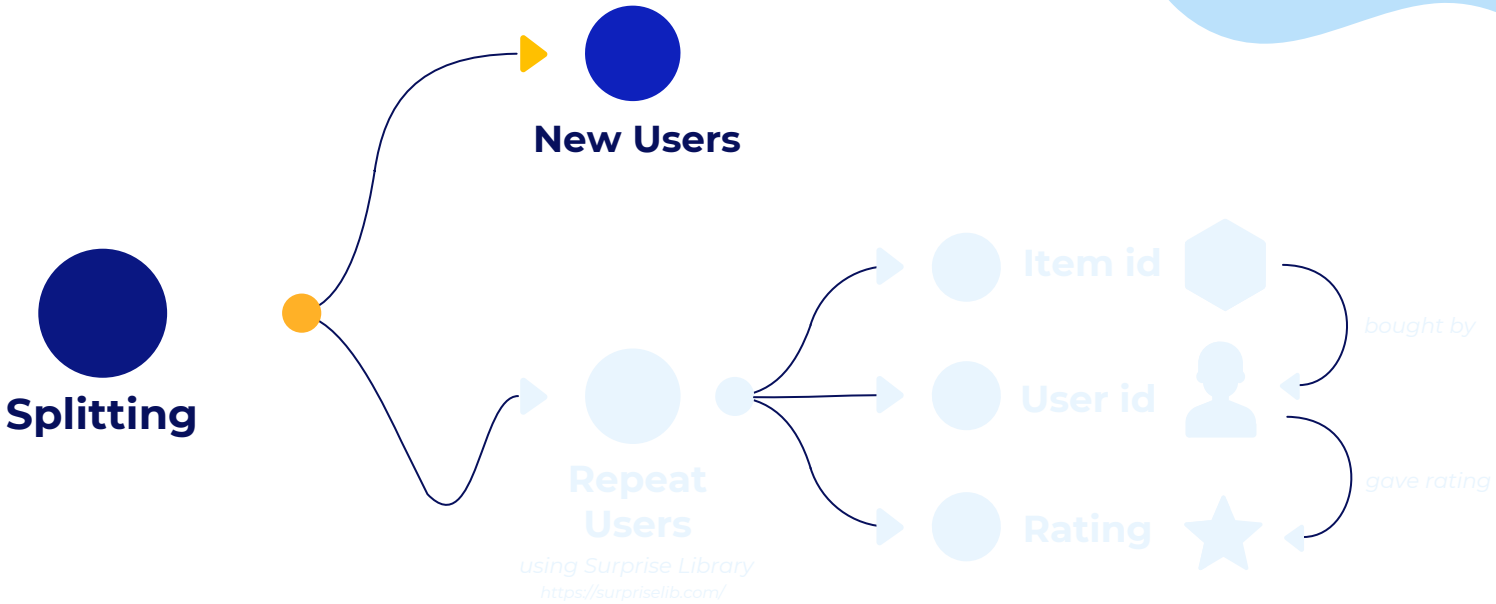

The model had the **good performance** so far for **above and below 3 rating score**

For above 3 rating score, the model had RMSE with skewed right distribution, meanwhile for below 3 rating score the model had same distribution but more long tail and slightly lower head

Hey, I wanna buy!

# Recommender for First-Time Customer

# RECOMMENDER
## FOR NEW USERS

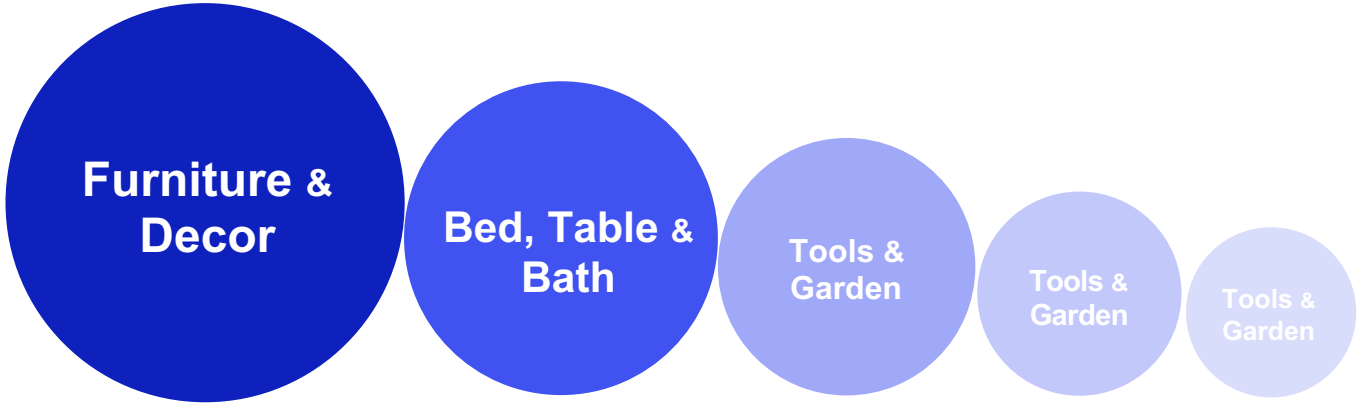


# Cold-start Problem
## from First-time Customer

| | | | |
|---|---|---|---|
| Jona | ★★ | ★★★ | ? |
| John | ? | ★★★★★ | ★★ |
| New User | ? | ? | ? |

https://en.wikipedia.org/wiki/Cold_start_(recommender_systems)

# Top 5 Popular Product Category
## for First Time-Customer




for all customers

# Top 5 Popular Product Category
## for First Time-Customer

example for customer:11e379501f73ec65c77def13ec8ae678

06

# CONCLUSION & SUGGESTION

# Conclusion & Suggestions

**Conclusion:**

The analysis reveals three key problems for the Olist. Firstly, a significant **80% of transaction value** is generated from **only 17 from 72 categories of products (23%),** indicating there's still **potential products** in the remaining **55 categories**. Secondly, there is a **urgent issue** with **bad ratings** that mostly product on **top 17 categories**, which may lead to **customer dissatisfaction** and impact **customer retention**. Thirdly, there're still **78% first-time customer** that **bought 1 product** instead of repeat customer that bought **2-3 products.**

**Suggestions:**

1. Implement a recommendation system that leverages **collaborative filtering**, specifically the SVD++ algorithm. This approach will provide **personalized product recommendations** based on user-to-product data, **improving** the overall **customer experience** and driving sales.
2. Try to **solve the issue of bad ratings** by identifying the underlying causes. Causes might be from product quality, delivery, customer support, or other factors affecting customer satisfaction. Resolving these issues would help **retain loyal customers**.
3. Since there're still a lot first-time customer than repeat customer**, overcome** the **cold-start problem** for **first-time customers** with two strategies. Firstly, implement a **Top 10 Popular Product Category Recommendation** to guide customers towards popular products. Secondly, implement a **Top 10 Product Category in State Recommendation** to offer products that are **specifically popular** in the **customer's state**. Additionally could be added from **Top Product Category Monthly Recommendation** These approaches would increase the **likelihood of repeat purchases** and contributing data collection from the customer.

**“The Purpose of Computation is Insight, Not Numbers.”**

—HAMMING, 1962

# REFERENCES


1. https://surpriselib.com/
2. https://www.moyak.com/papers/collaborative-filtering.html
3. https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada
4. https://developers.google.com/machine-learning/recommendation/collaborative/basics
5. https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0
6. https://towardsdatascience.com/recsys-series-part-4-the-7-variants-of-matrix-factorization-for-collaborative-filtering-368754e4fab5
7. https://towardsdatascience.com/collaborative-filtering-and-embeddings-part-1-63b00b9739ce
8. https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0
9. https://www.ethanrosenthal.com/2015/11/02/intro-to-collaborative-filtering/
10. https://www.stat.ipb.ac.id/main/wp-content/uploads/2021/01/Introduction-to-RecSys_BERT4Rec-IPB.pdf
11. https://medium.datadriveninvestor.com/how-to-build-a-recommendation-system-for-purchase-data-step-by-step-d6d7a78800b6
12. https://towardsdatascience.com/building-and-testing-recommender-systems-with-surprise
13. https://towardsdatascience.com/a-simple-approach-to-building-a-recommendation-system-d0f4de1a1f50e-step-by-step-d4ba702ef80b
14. https://celikkam.medium.com/surprise-recommender-library-few-lines-for-everything-d2de3ceac3d4
15. https://towardsdatascience.com/recommendation-system-series-part-1-an-executive-guide-to-building-recommendation-system-608f83e2630a

# THANK YOU!

Let’s connect!

linkedin.com/in/satriabagaskara/

github.com/dataaga/Olist_CollaborativeFiltering

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