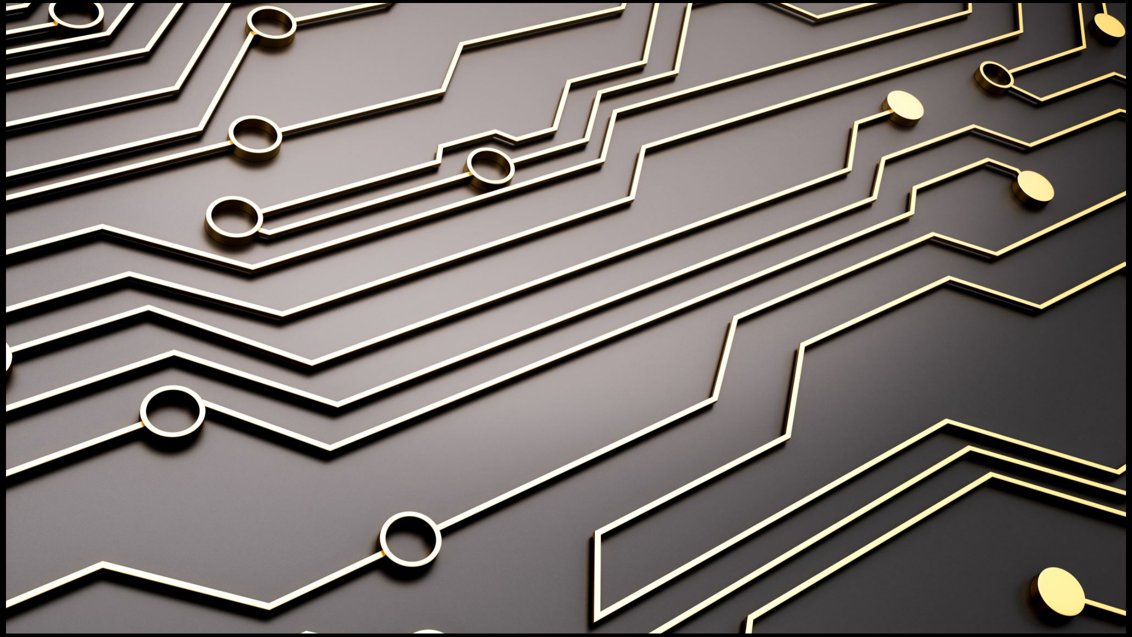
milkbasket



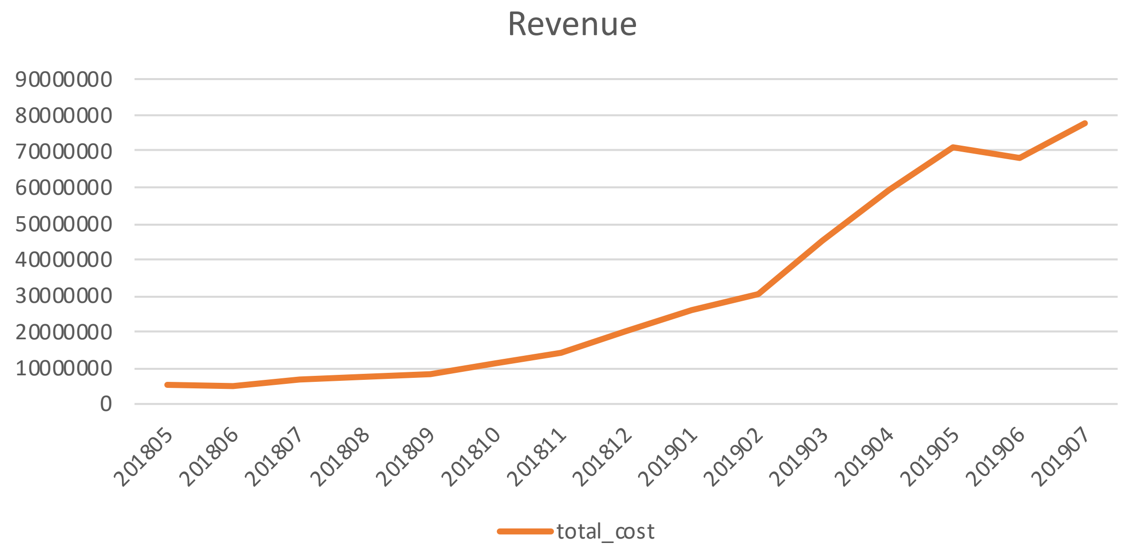
# milkbasket: Trend analysis and recommendations to improve business metrics

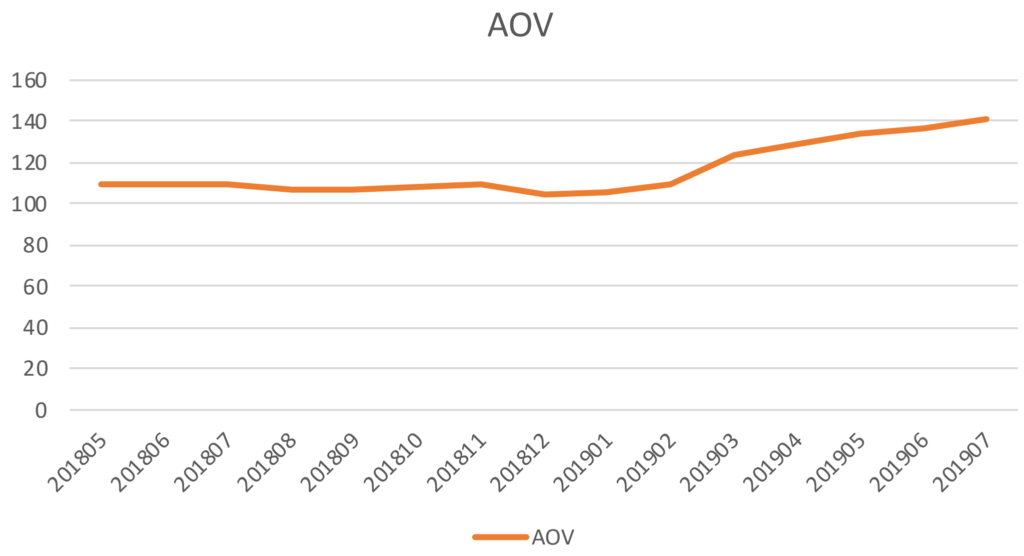
**Submitted by: Team T_WIN**

# Agenda

- Observations and Insights: Exploratory Analysis of order data to gain observations and insights
- Product Recommendation Engine: Build a model to generate recommendations based on multiple parameters
- Churn Analysis: Identify potential risk of churning out for a customer

# Observations and Insights

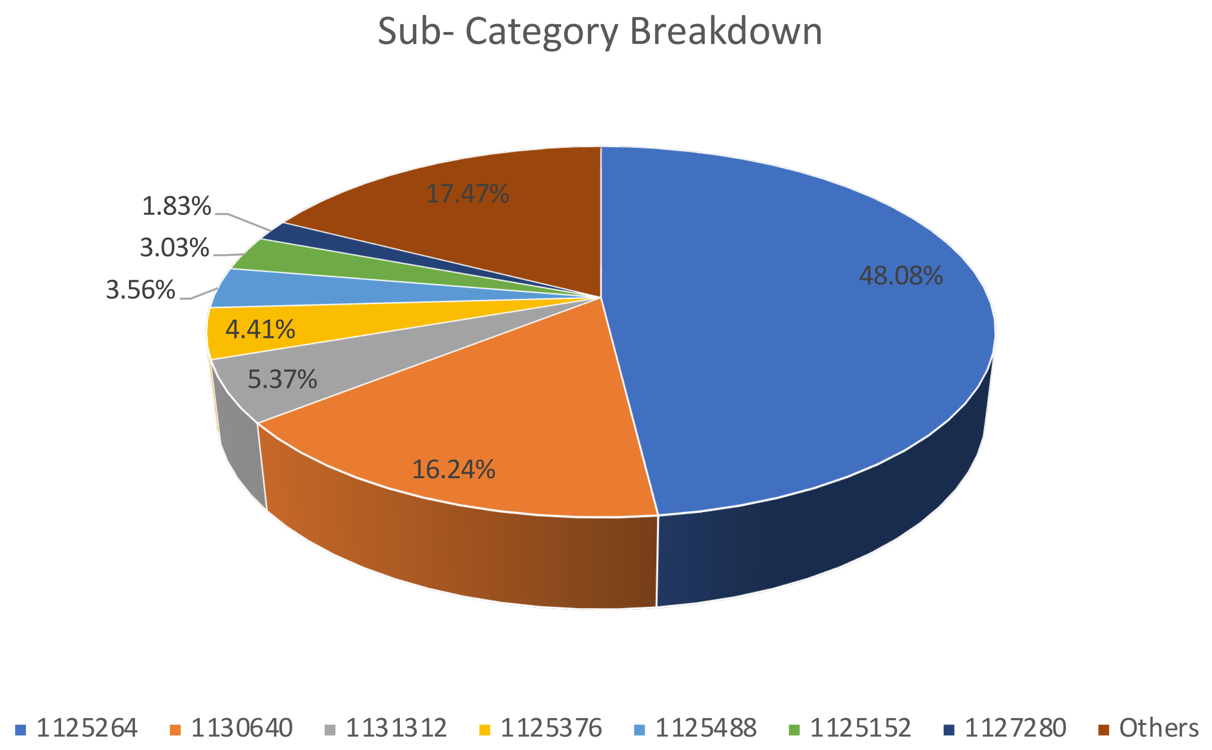


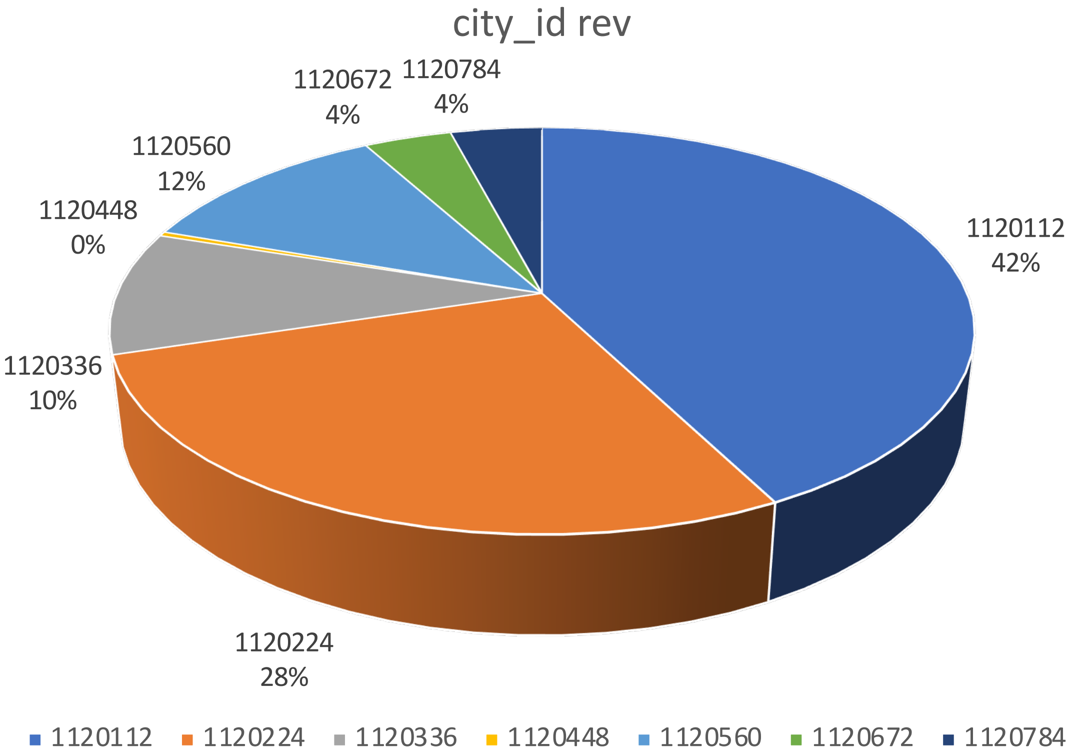


Milk basket had a continuous growth but on Mar'19 there was a growth of 50% MoM which also increased the AOV from (104-109) to 126 and then after went to 146 in 4 months

# Revenue Breakdown by sub category



# Analysis by cities





1. Out of new 5 cities, which started its operation after December, 2018, two cities i.e. 1120336 and 1120560 with average ~ INR 50,00,000 per month and has huge contribution to jump in growth rate for overall revenue, especially in March 2019
2. As time is progressing, new cities are catching up and %contribution is significant ~30%

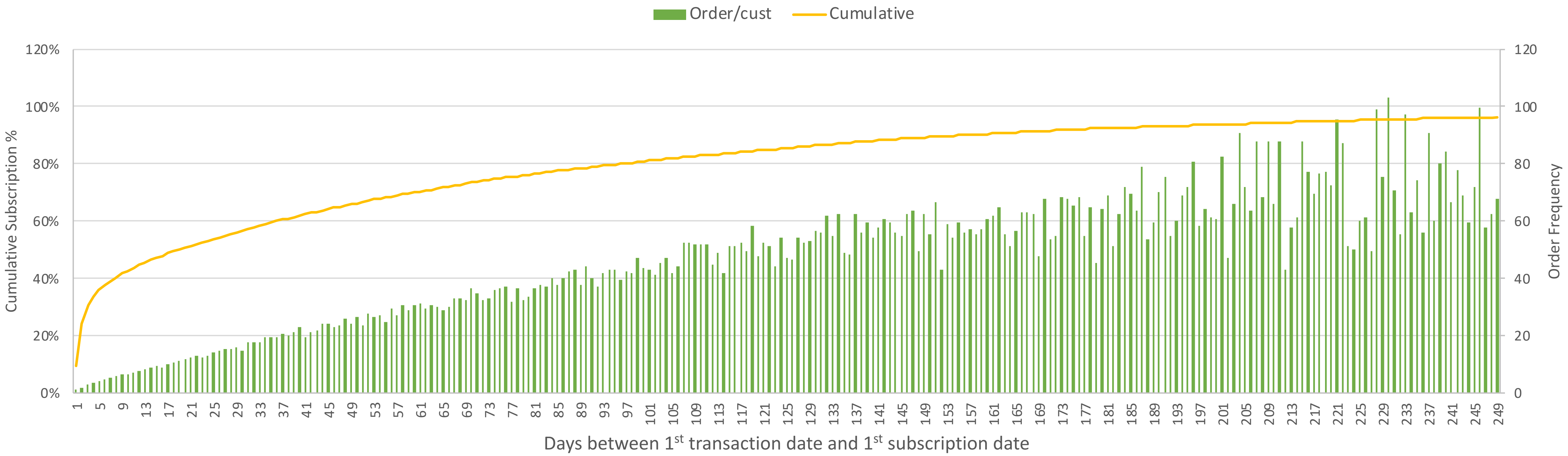
# Retention Rate

| % Retention after # of orders | New City (#  Orders) | Old City (# Orders) |
|---|---|---|
| 99% | 7 | 2 |
| 98% | 13 | 3 |
| 97% | 17 | 4 |
| 96% | 20 | 5 |
| 95% | 22 | 6 |

The repeat purchase rate has  been very high for Milk Basket. For the new cities (opened after Nov'18), this rate has further increased to 1% cumulative customer drop out after nearly 7 orders. While for the older cities, it has been reached just after the 1st order

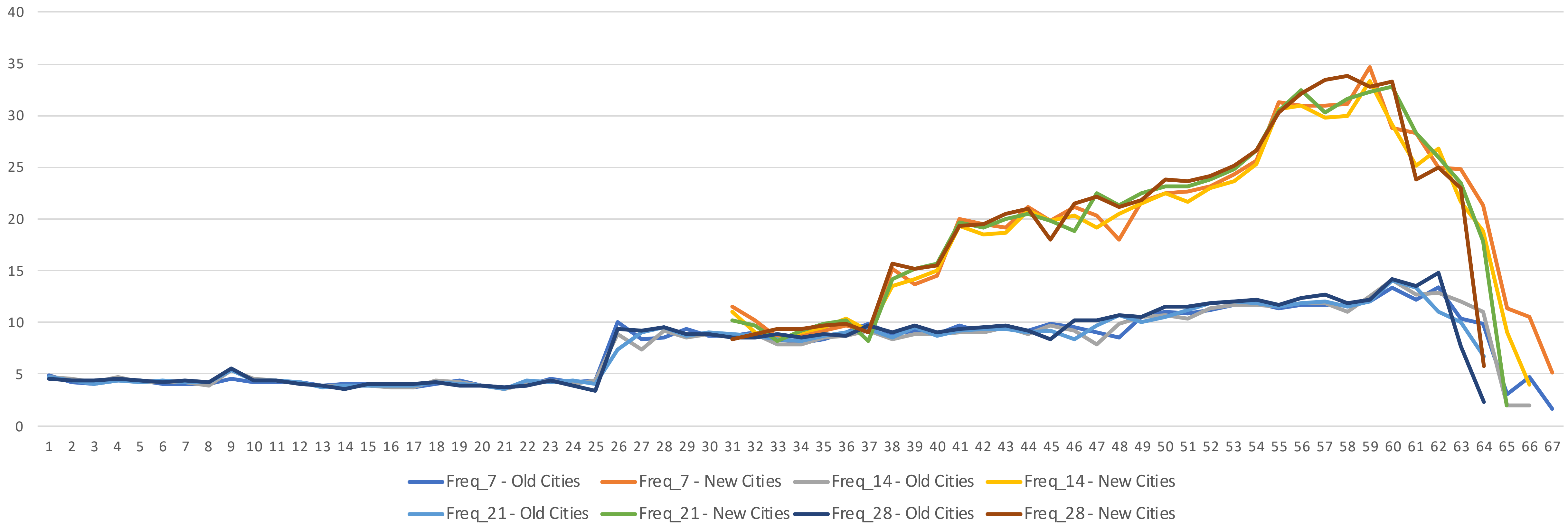
* Subscriptions are not included

# Subscriptions




Off all the customers (19K)who subscribed at least one-time, the first 25% customers subscribe within first 3 orders while the next 25% customers take around 10 orders

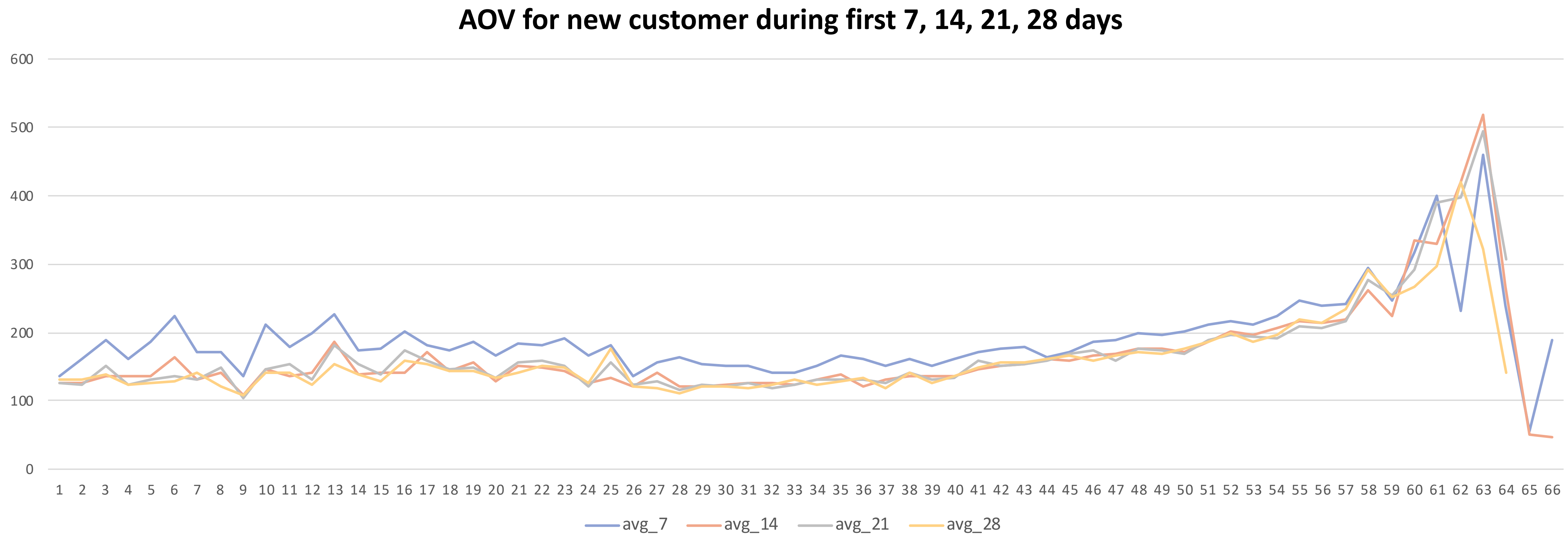
# Repeat frequency rate




Freq_7- old cities is a cohort of customer from his first transaction date to his next 7 days from first transaction

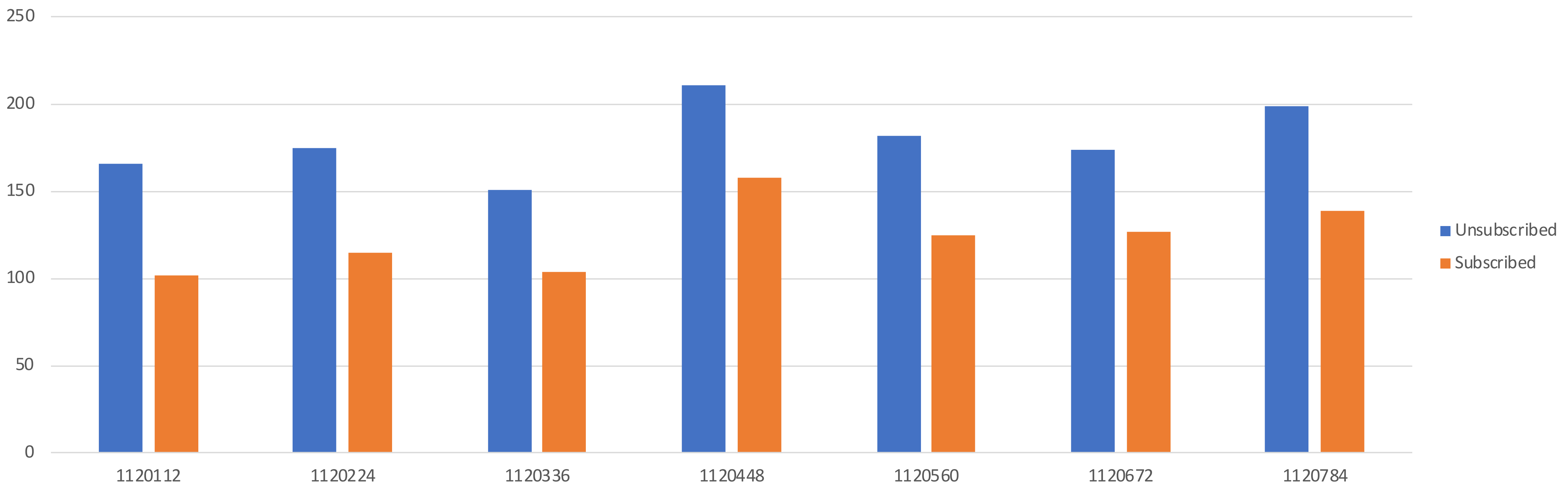
We wanted to understand how many repeat customer are there for batch of every 7 days from his first transaction date for both old cities (1120112,1120224) and new cities (remaining 5)

# AOV for new customers




The AOV for new customers has been increasing for the first 7 days, 14 days, 21 days, 28 days

# AOV by subscription

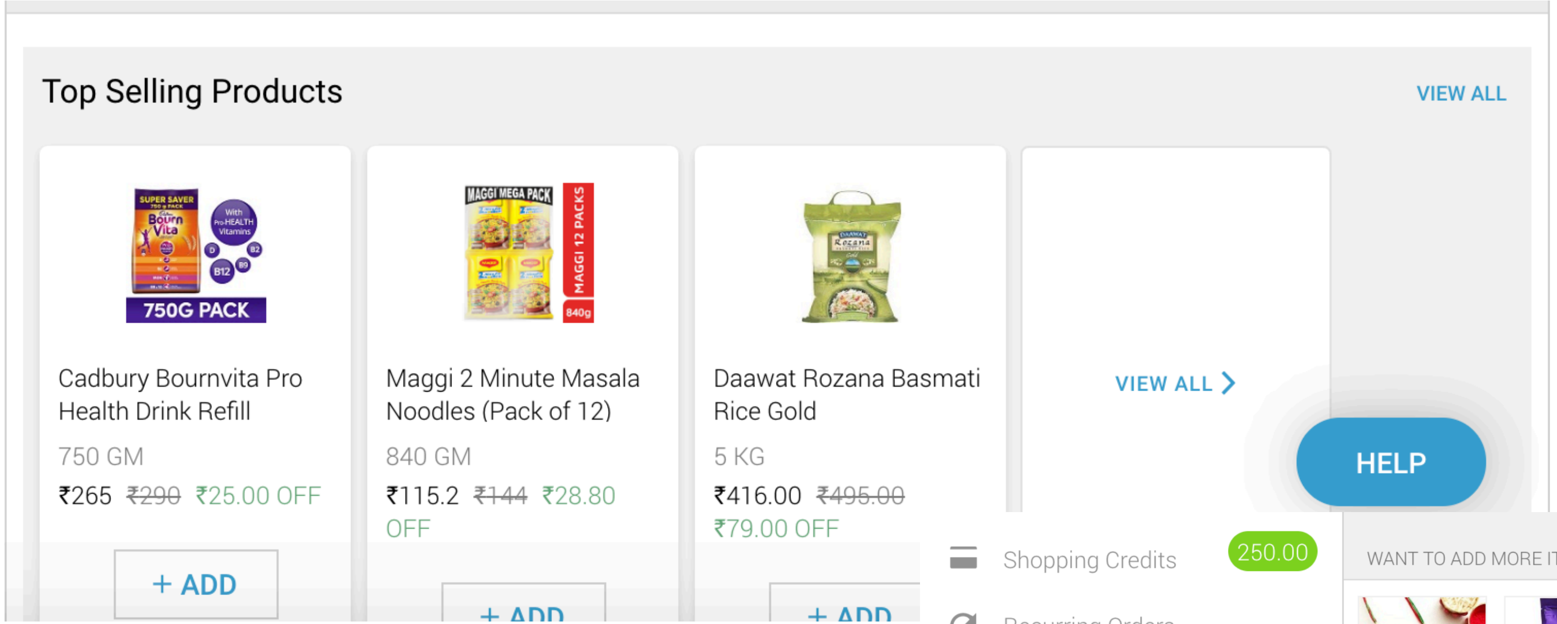



Total unsubscribed AOV is ~ 50% higher than the subscribed AOV

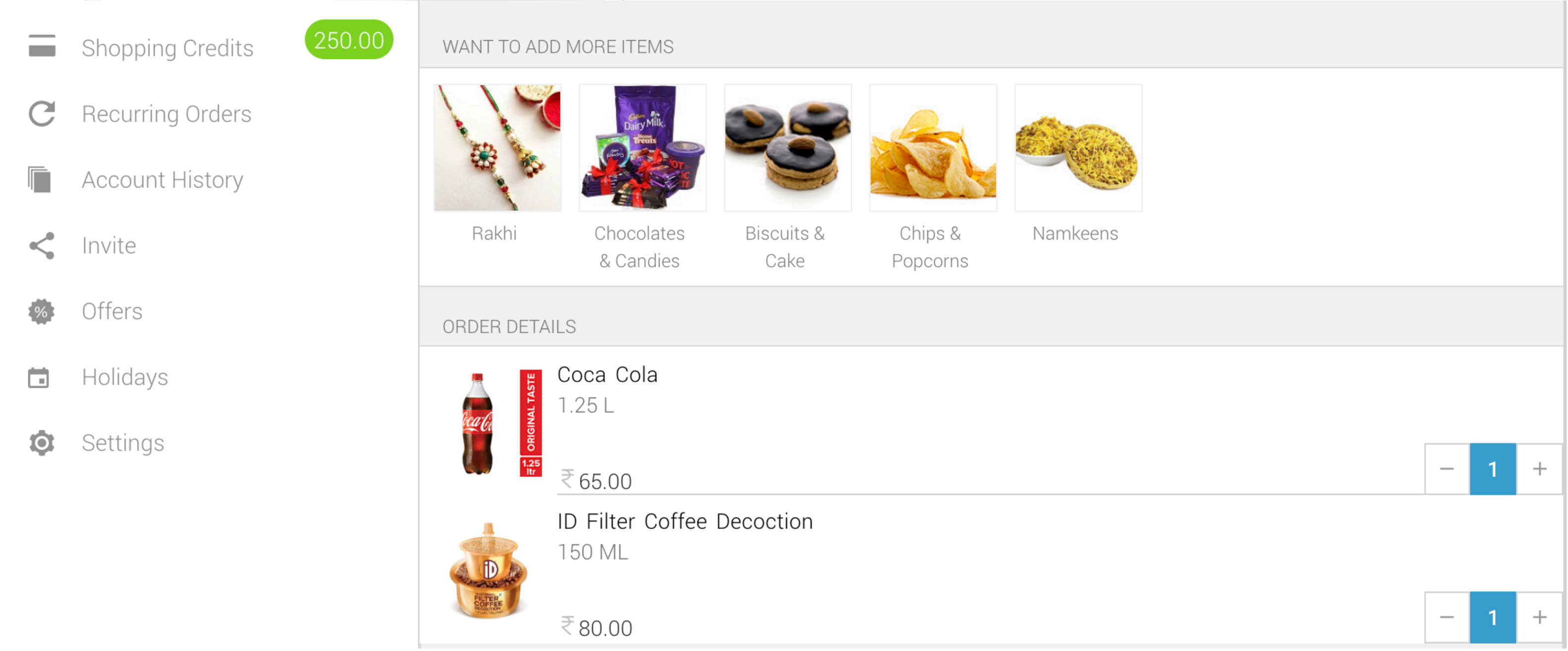
# Product Recommendation Landscape

# pRODUCT rECOMMENDATION

# Current product recommendations on Milkbasket



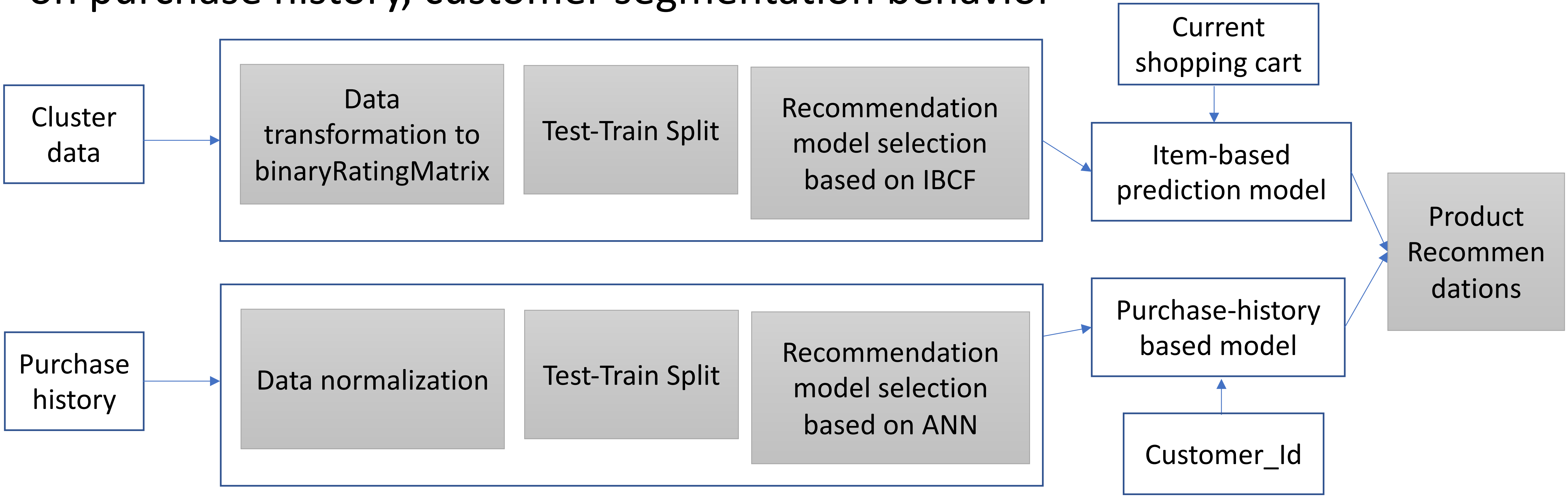

Product recommendations on home-page based on top selling products




Product category recommendations on CART page just before checkout

# Recommendation engine architecture

**Objective:** To generate relevant product recommendations for customer based on purchase history, customer segmentation behavior




Model Information

**Item-based prediction model**

- Algorithm: Item-based collaborating filter
- Stack used: Recommenderlab

**Purchase history-based model**

- Algorithm: ANN
- Stack used: numpy pandas, Keras

# Confusion Matrix for Item-based prediction model

| Number of products recommended | Precision | Recall | True Positive Rate | False Positive Rate |
|---|---|---|---|---|
| 5 | 0.11394562 | 0.5622100 | 0.5622100 | 0.025856762 |
| 10 | 0.06841389 | 0.6529378 | 0.6529378 | 0.051973459 |
| 15 | 0.05250823 | 0.6702758 | 0.6702758 | 0.072132157 |
| 20 | 0.04868339 | 0.6713285 | 0.6713285 | 0.080969971 |

# Churn Propensity Model

**Objective:** We are trying to predict the customers who are going to get churned in the next 30 days and give them relevant offers to reduce the overall churn rate

The approach here is to understand that what all variables are defining a customer to become dormant in the next 30 days and predict the propensity values for that customer.

**Insights:** There are about 5,190 customers out of 16,711 customers who has actually became dormant for 30 days from the previous data order

From this data we can clearly say that building a churn model is important to predict the propensity values

For this problem we have taken many variables as follows.

**Input Variables:** Customer_id, City_code,

Dormancy days from previous order

How is the customer behaviour in terms of

(AOV, No of subscriptions, No of products bought, No of Orders, No of Ordered Days) *at two levels of duration*

1. How's his behaviour during 1 month before dormant
2. How's his behaviour during 3$^{rd}$ month to 2$^{nd}$ Month before dormant

**Output Variable:** dormant (1/0)

Accuracy: 0.8, AUC: 0.72947

Thank you!