greatlearning
Power Ahead

Customer Churn Analysis for E-Commerce Company

Akash Kamble
DSBA - June'21 Batch

Agenda

Business Problem Understanding

3 - 5

Insights from Analysis

6 - 10

Modelling Approach

11 - 14

Recommendations & Offers

15 - 18

Agenda

Business Problem Understanding

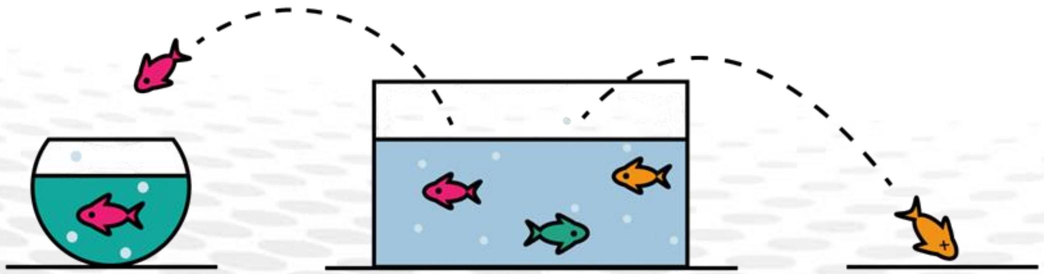
Insights from Analysis

Modelling Approach

Recommendations & Offers

Business Problem Understanding

- An **E-Commerce company** is facing a lot of competition in the current market and it has become a challenge to **retain the existing customers** in the current situation.
- Hence, the company wants to develop a model through which they can do **churn prediction** of the accounts and provide **segmented offers** to the potential churners.
- So, we are going to develop a churn prediction model using various **regression, ML and ensemble techniques**. In addition, we will be providing the **insights, business recommendations** and **segmented offers** that helps the company to arrest the **churn rate**.



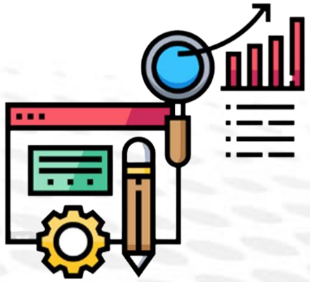
Business Problem Understanding

Assumptions/ Constraints

- ‘rev_per_month’ have unit digit values such as 9, 7, 6 etc. and values in ‘cashback’ are three digit i.e., 160, 121 etc. which cannot happen in reality. So, we **converted** the ‘rev_per_month’ by multiplying the same by 100.

Scope

- Deliver a **optimized model** which is able to predict the class of interest
- **Business recommendations** on the basis of insights
- Suggest **segmented offers** to prevent customer churn



Objective

- Built various models such as **linear, non-linear and ensemble models** and choose the most optimum model with **good performance parameters**.

Agenda

Business Problem Understanding

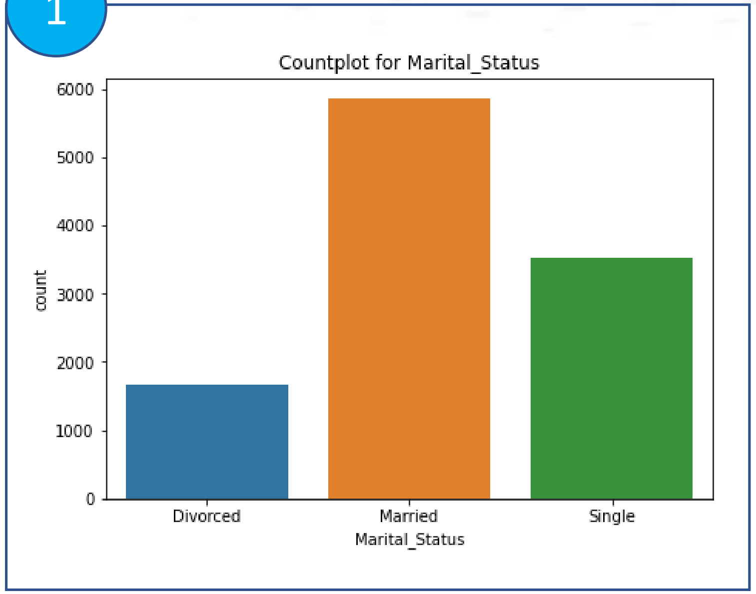
Insights from Analysis

Modelling Approach

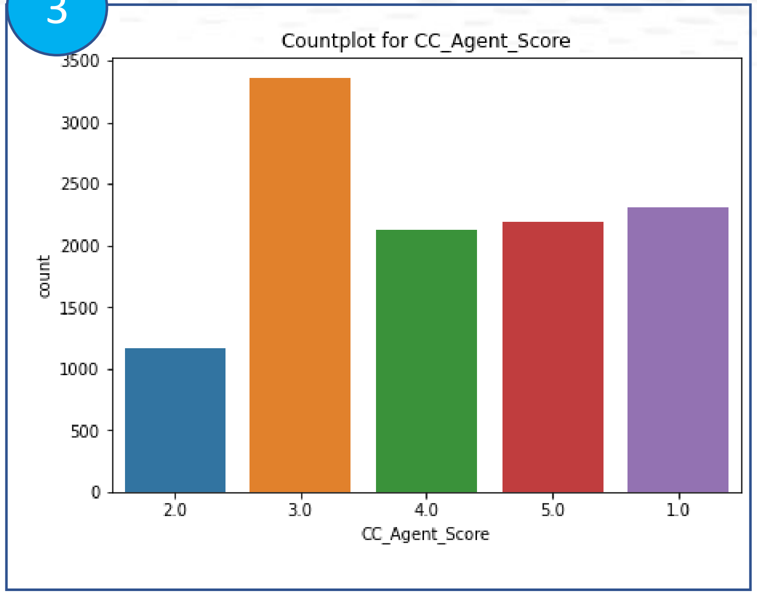
Recommendations & Offers

Insights from Analysis

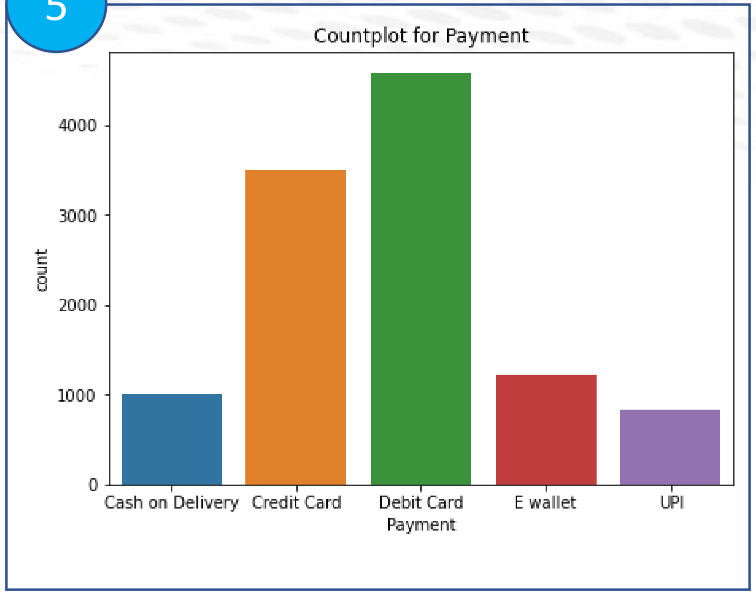
1

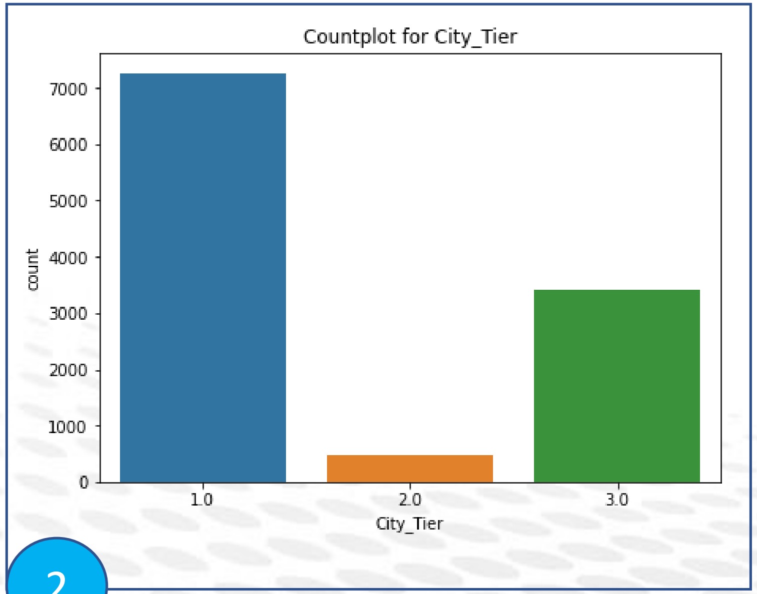


3

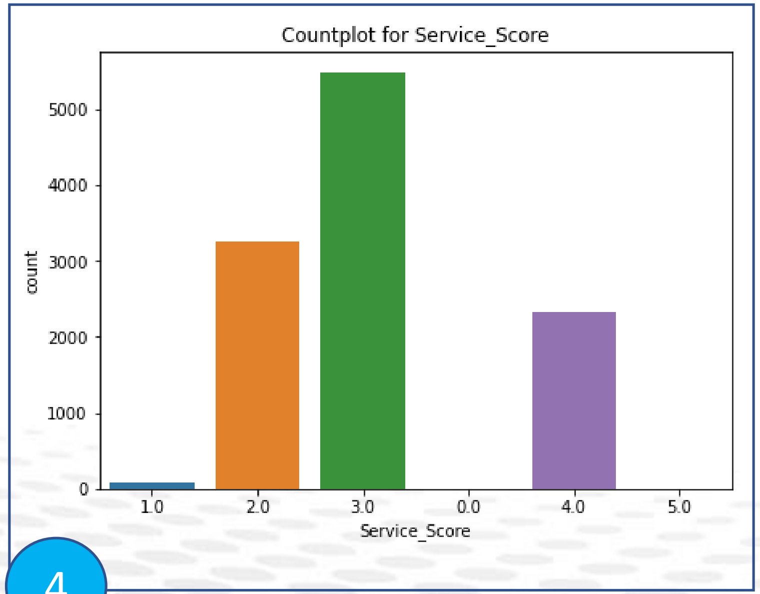


5

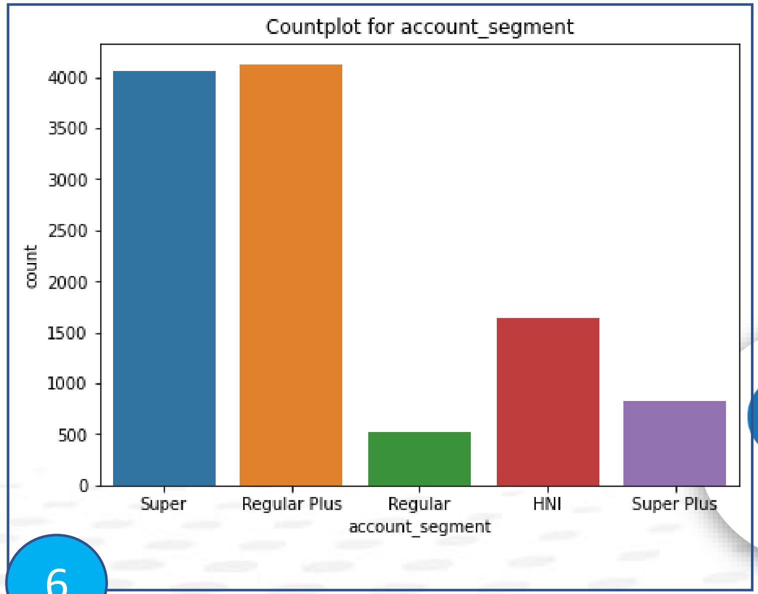




2



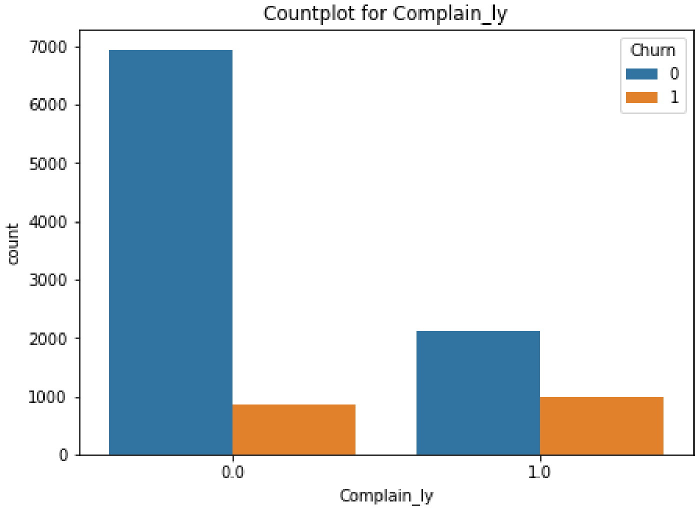
4



6

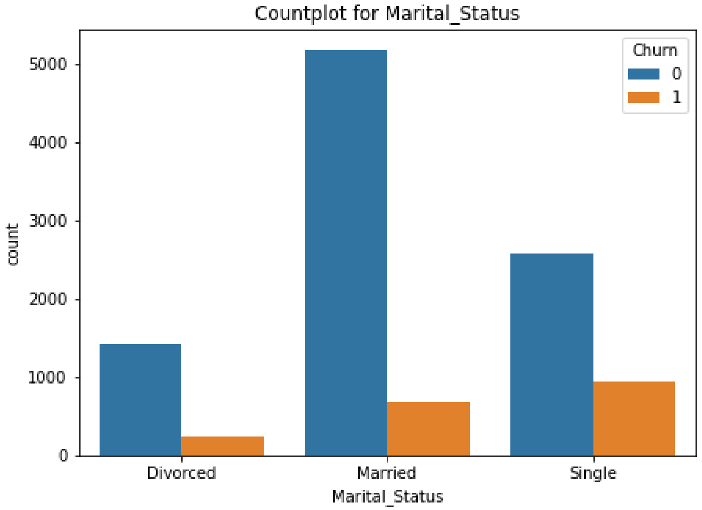
Insights from Analysis

Complaints Logged by Users



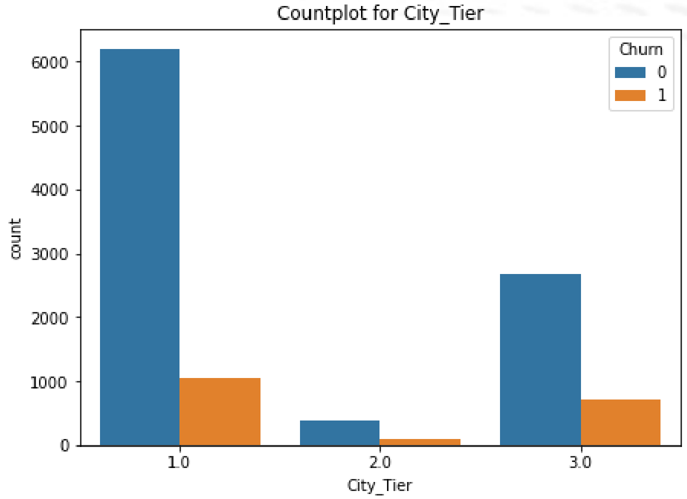
1. Out of total customers who logged a complaint, ~**1/3rd** resulted **in churn**.
2. This particular variable is a **good predictor** of potential future churners.
3. There are also customers who didn't log a complain but still resulted in churn which is ~**1/8th of non-complainers**

Marital Status



1. Customers having marital status as **'Single'** are prone to be **potential churners**. **~25%** of the total customers having 'Single' status are churners.
2. There are **relatively low** churners in **'Married'** & **'Divorced'** status.

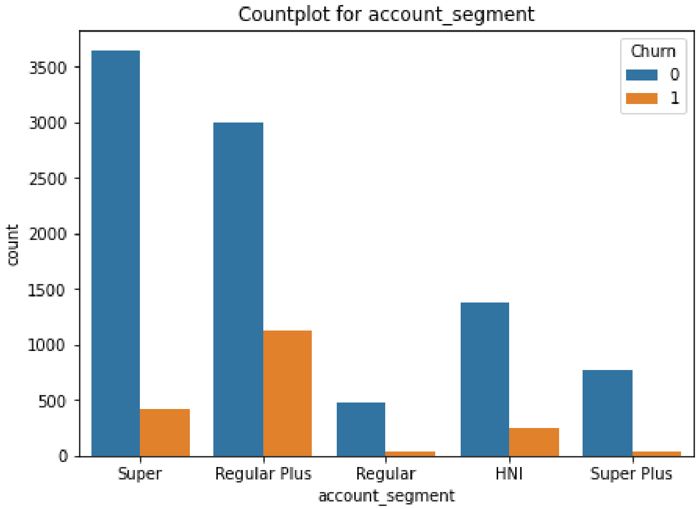
City Tier



1. There are **relatively higher churners** in Tier-3 cities.
2. There are churners present in Tier-1 cities which constitute **1/7th of total customers present in Tier-1 cities**.

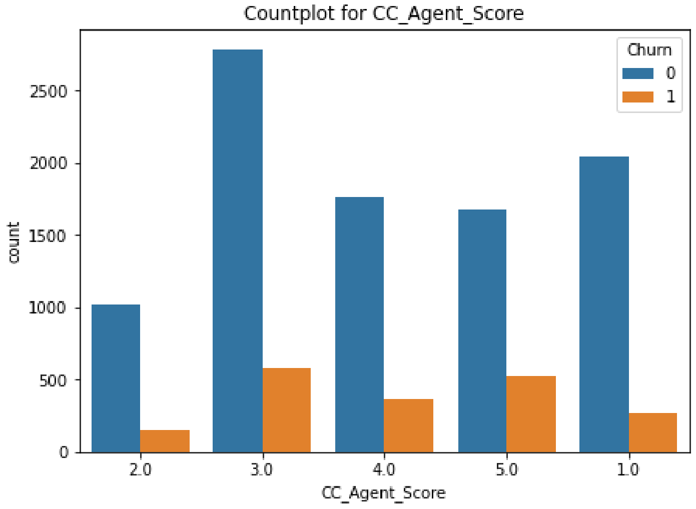
Insights from Analysis

Account Segments



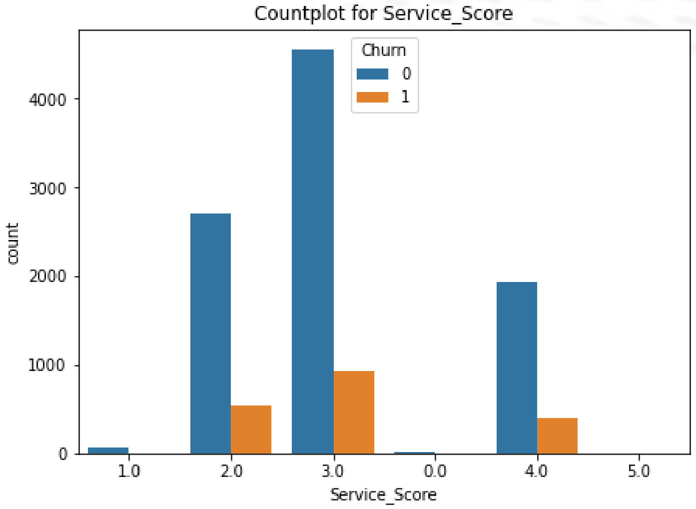
1. There are **proportionately more churners** present in the **'Regular Plus'** account segment. Followed by 'Super' account segment.

Customer Care Service Score



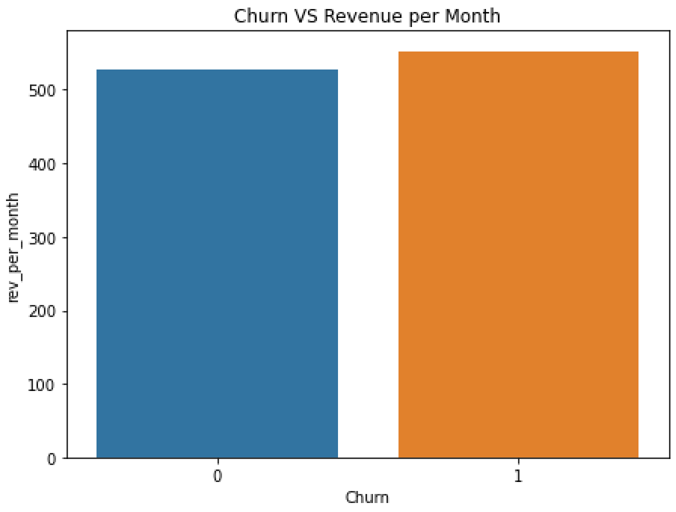
1. Most of the customers had given a **score of 3** for customer care service.
2. Majority of the **churners** are present in **customer group of score 3, 4 & 5**.

Company Service Score

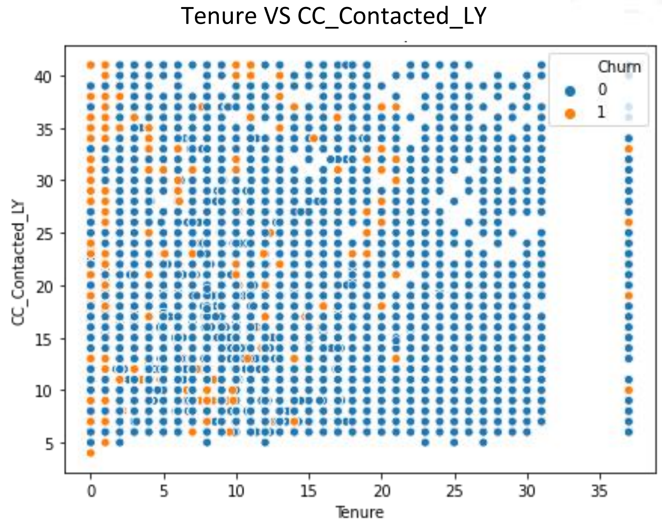


1. Most of the customers had given a **score of 3** for services provided by the company.
2. Majority of the **churners** are present in **customer group of score 2 & 3**.

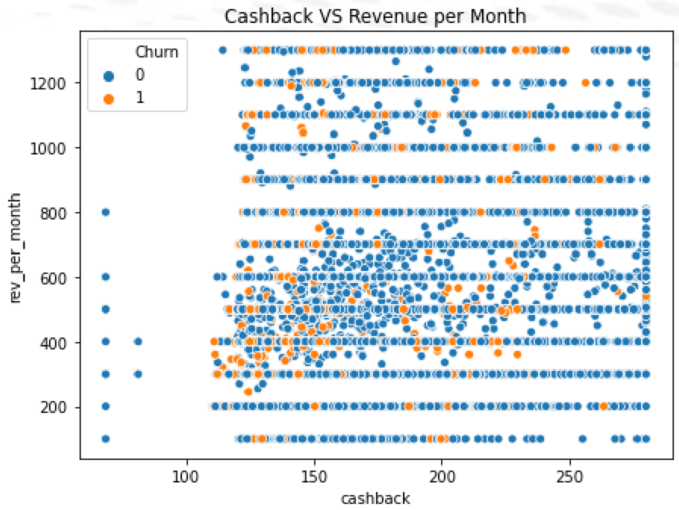
Insights from Analysis



1. Churners generate comparatively **higher average revenue per month** than non-churners.
2. Therefore, when a customer churns it has **higher negative impac**t on **revenue and profits**.



1. There seems to be more churners in zone **(tenure<5 & no. of times contacted >25)**.
2. Customer tends to churn more when tenure is less than 5.
3. Also, customer tends to churn when he/she is frequently connecting with customer care service.



1. For **cashback** in range of **INR 100-150** and **avg. revenue/month** in range of **INR 200-600**, there seems to be more churners present.
2. This could be due to lower cashbacks given by the company.

Agenda

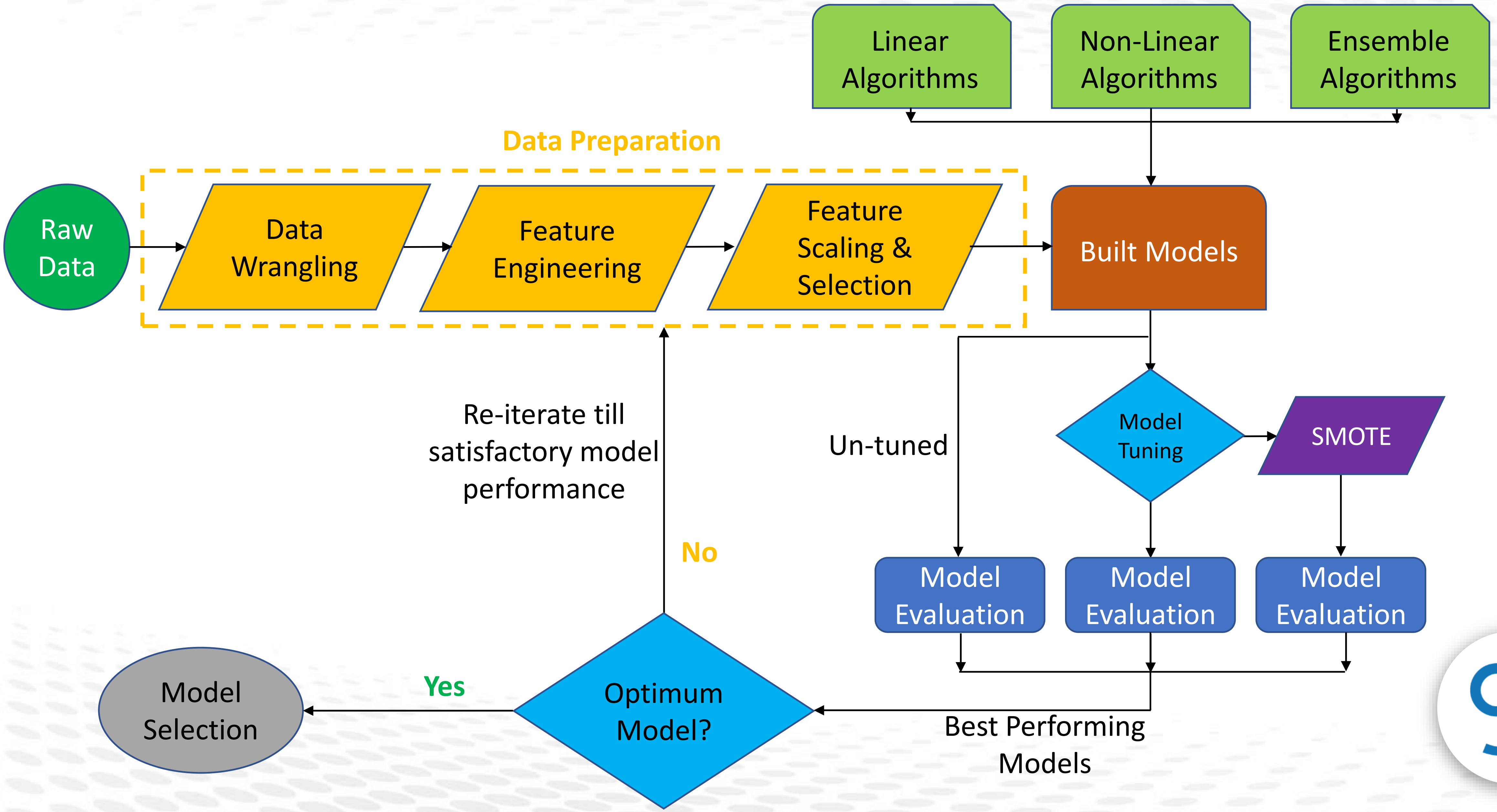
Business Problem Understanding

Insights from Analysis

Modelling Approach

Recommendations & Offers

Modelling Approach



Modelling Approach

Data Wrangling	**Treating anomalies, outliers & null values**
Feature Engineering	**Feature created ('cashback/coupon'), EDA & Logarithmic Transformation of few variables**
Feature Scaling & Selection	**Feature Selection using VIF values, creating dummy variables and scaling the numeric data**
Built Models (Un-tuned, Tuned & SMOTE)	**Built models using different algorithms such as Logistic regression, LDA, Random Forest, Artificial Neural Networks, Ada Boost and XG Boost.**
Model Evaluation	**Model evaluation using accuracy, precision, recall, F-1 score and AUC score. XG Boost with 0.88 recall score**
Optimum Model Selection	**XG Boost selected as optimum model over ANN as ANN has black box nature & unable to explain the importance of variables**

Modelling Approach

Model Performance Comparison for Class '1':

Models			Train Dataset					Test Dataset				
			Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score	AUC
Level 1: Un-tuned Models	Linear Models	Logistic Regression	0.90	0.76	0.58	0.66	0.90	0.90	0.78	0.57	0.66	0.89
		LDA	0.90	0.72	0.62	0.66	0.90	0.90	0.73	0.60	0.66	0.88
	Non-linear Models	Random Forest	0.96	0.99	0.78	0.87	0.99	0.93	0.95	0.64	0.77	0.97
		ANN	0.99	0.98	0.97	0.97	1.00	0.96	0.90	0.84	0.87	0.98
	Ensemble Models	Ada Boost	0.90	0.76	0.59	0.67	0.92	0.90	0.76	0.59	0.67	0.91
		XG Boost	1.00	1.00	1.00	1.00	1.00	0.97	0.96	0.85	0.90	0.99
Level 2: Hyperparameters Tuned Models	Linear Models	Logistic Regression	0.90	0.76	0.56	0.65	0.90	0.90	0.76	0.56	0.65	0.88
		LDA	0.89	0.72	0.60	0.66	0.90	0.89	0.73	0.58	0.65	0.88
	Non-linear Models	Random Forest	1.00	1.00	1.00	1.00	0.99	0.97	0.98	0.84	0.90	0.99
		ANN	1.00	1.00	1.00	1.00	1.00	0.97	0.93	0.88	0.91	0.99
	Ensemble Models	Ada Boost	0.90	0.76	0.60	0.67	0.92	0.90	0.76	0.59	0.67	0.91
		XG Boost	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.88	0.92	0.99
Level 3: SMOTE with Tuned Models	Linear Models	Logistic Regression	0.85	0.86	0.85	0.85	0.93	0.84	0.51	0.71	0.60	0.87
		LDA	0.85	0.86	0.85	0.85	0.92	0.84	0.52	0.74	0.61	0.87
	Non-linear Models	Random Forest	1.00	1.00	1.00	1.00	1.00	0.96	0.89	0.87	0.88	0.99
		ANN	1.00	1.00	1.00	1.00	1.00	0.96	0.89	0.88	0.88	0.98
	Ensemble Models	Ada Boost	0.89	0.89	0.89	0.89	0.96	0.87	0.61	0.72	0.66	0.90
		XG Boost	1.00	1.00	1.00	1.00	1.00	0.97	0.93	0.88	0.90	0.90

Performance Parameters

- We can observe in the table above that hyper-parameter tuned XG Boost model had performed best considering performance parameters such as recall, precision, F-1 score etc.

Agenda

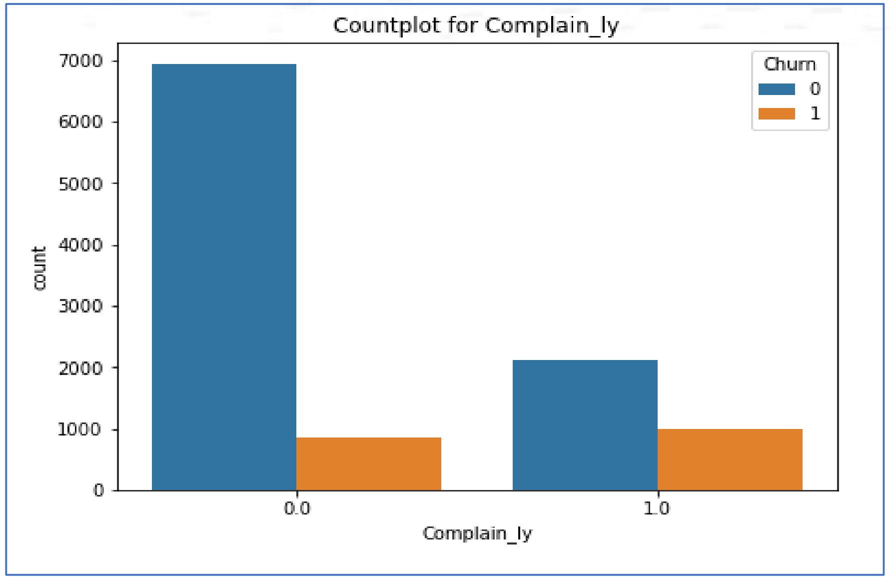
Business Problem Understanding

Insights from Analysis

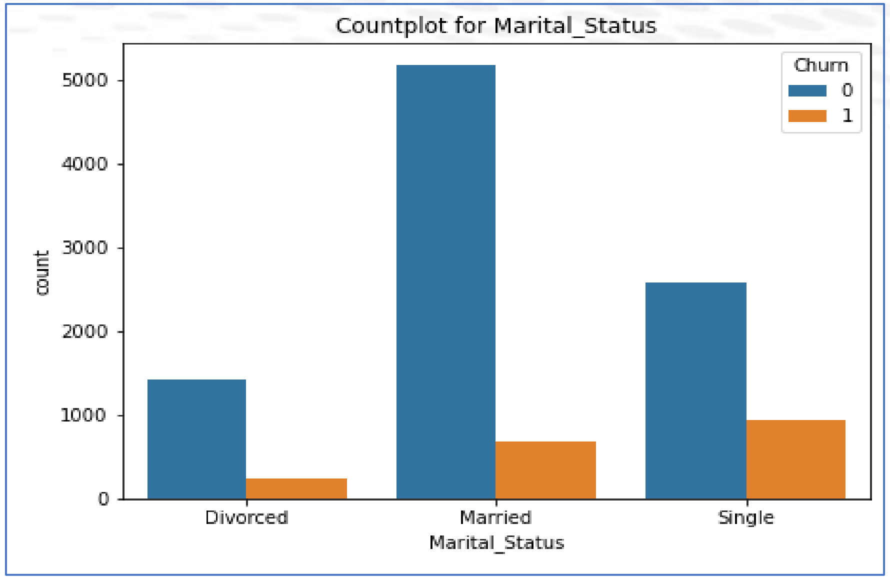
Modelling Approach

Recommendations & Offers

Recommendations & Offers

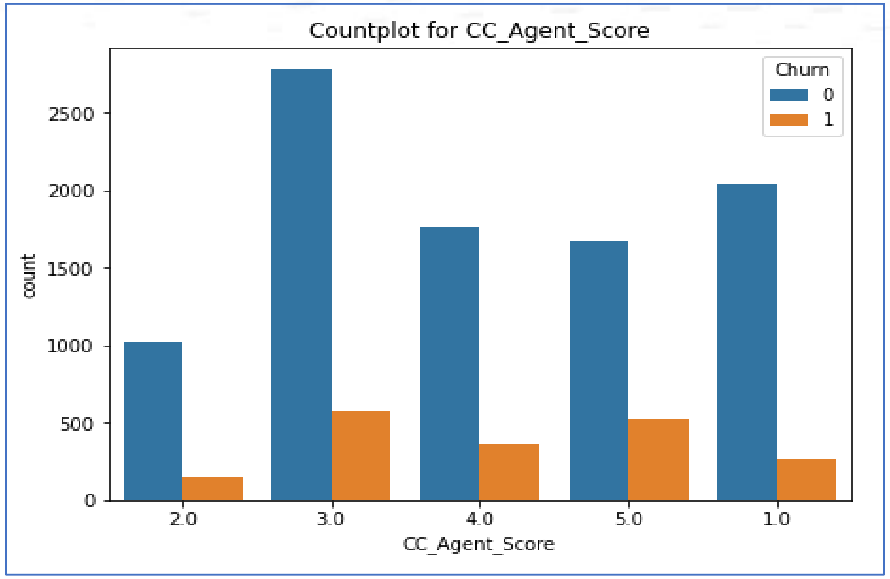


1. Refer the graph above, out of total customers those who had filed a complained in the past, **1/3rd customers had churned**.
2. In such scenarios, customer service department should be given a free hand to **compensate** the painful customer in **monetary terms** depending upon the situation.
3. The monetary compensation could be in various forms such as immediate **replacement** of damaged product, coupons for delay in delivery of the product, exclusive festival offers for such customers, follow-ups on past & current issues etc.

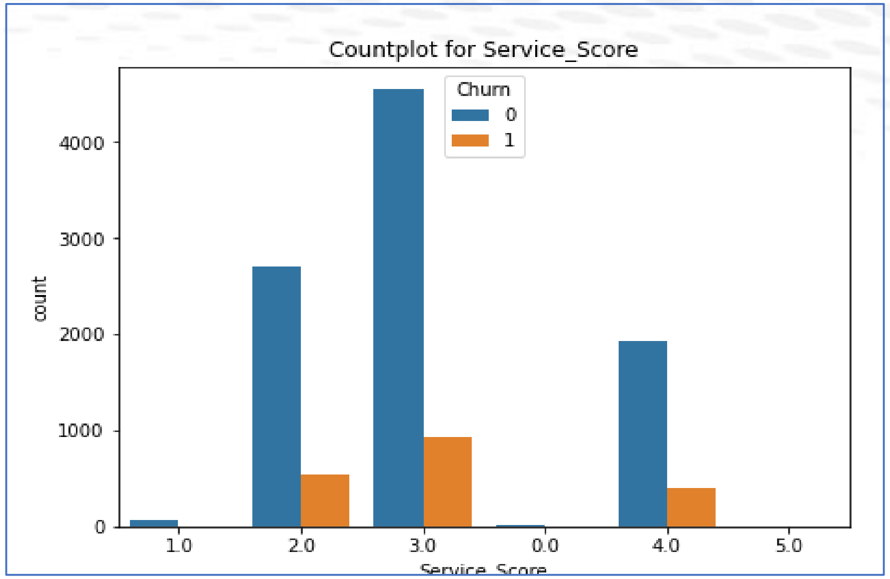


1. It can be said that customers with marital status as **'Single' have proportionately more churn population**.
2. E-commerce Company should tie up with **online dating apps** such as 'Tinder', 'Happen' etc and should bring offers in collaboration with dating apps such as **free 1 month of subscription** on 'xyz' dating app on cart value of above INR 5000, given that the customer age is above 18 years. Age, gender etc such information can readily availed through account information.

Recommendations & Offers

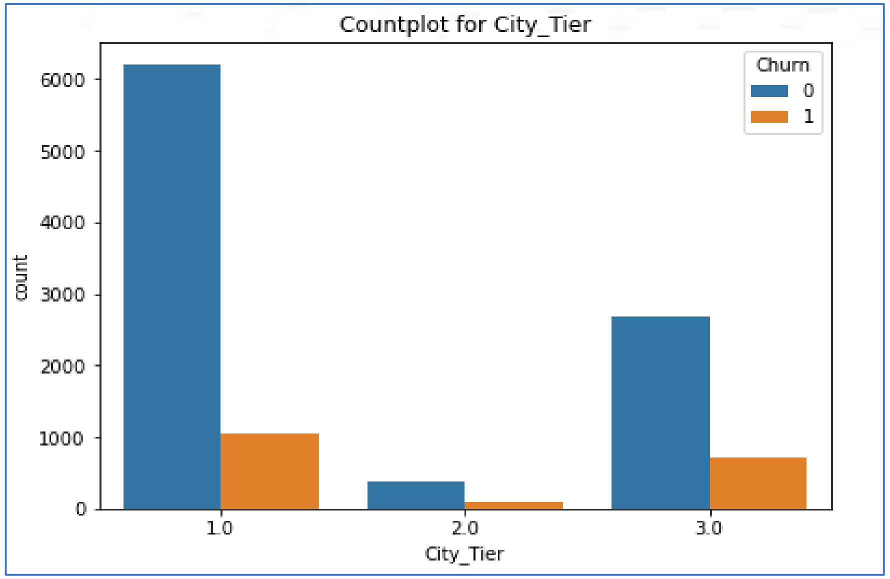


1. We can see there are more churners from group of customers giving **3 customer care service score**.
2. To arrest the churn in above situations, E-commerce Company should train its customer care service department employees with **best practices**, etiquettes and job training. Also**, immediate feedback on customer care executive/employee** should be taken through online mode such as survey form, calls etc.
3. Executives/employees with poor scores should be provided with specific trainings, or such executive should be moved to backend departments.
4. When employees are aware that they are been monitored, their performance will naturally improve.

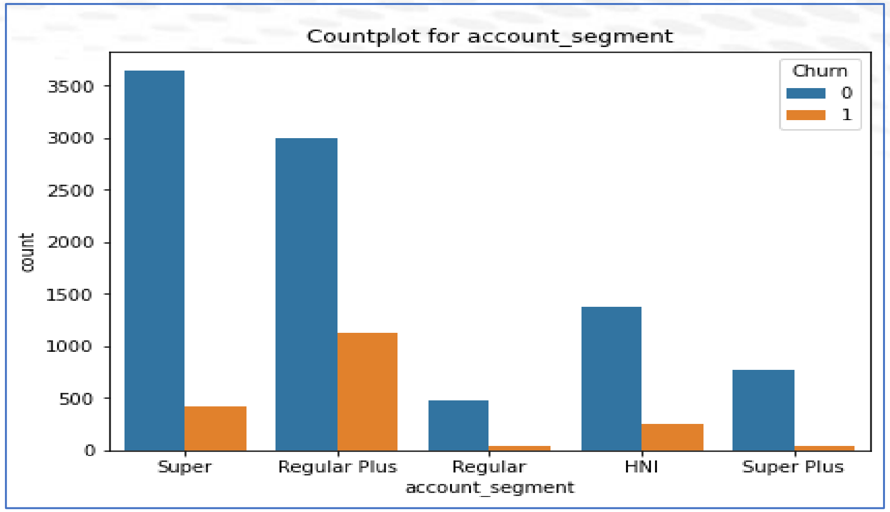


1. Customers giving 3 score for company's services are probable churners also more number of customer belongs to 3 score category, so it is an important area to look at and improve the retention plan.
2. Customer under this category should be closely monitored and feedbacks should be collected on regular basis for any issues faced. **The feedbacks should be recorded and reviewed internally in the team.**
3. The issues faced should be proactively closed and executive should **acknowledge the customer** about same.

Recommendations & Offers



1. There are proportionately more churners in Tier-3 cities as compared to others tier of cities.
2. These churners could be present due to **non-delivery of certain items/services** to those Tier-3 cities by the E-commerce company.
3. Company should improve its **supply chain network** in Tier-3 cities. Company should combine warehouse facility for 3-5 Tier-3 cities for better connectivity, faster shipments, and limit damage of goods due poor handling/logistic service.
4. Average tier-3 city churner have revenue more than non-churners. Therefore, it is not a variable to ignore.



1. It can be inferred from above graph that **proportionately more churners** belong to **'Regular Plus'** account segment.
2. This segment can be considered as **upper-medium income segment** where customer is ready to spend/buy more given that there are offers or some monetary advantage on the purchase.
3. This segment's customer may result in churn when not given offers and other benefits.
4. So, in order to retain such segment of customer we can execute following strategies:
   a. Flash sale – Buy top selling groceries at just INR 1
   b. Top selling groceries could be 'dal', 'biscuits', 'salt' etc.
   c. Such flash sales will attract more and more 'Regular Plus' segment customer and retain the existing ones.
   d. Create Items Combos using market basket analysis and provide combos at discounts.
   e. Buy Combo of Scotch-Brite scrub pad, Vim dish wash bar and Pitanbari (a powder used to clean copper utensils) at just INR 199/- only

Thank You