RECOMMENDATION SYSTEM FOR RETAIL STORES

Submitted by:

Team WeLookUp

Ashish Kumar Wani, Sai Krishnan Mohan, Srinivas Vattikutti & Swati Gupta

IMPROVE RETENTION

Continuously catering to users' preferences makes them more likely to remain loyal subscribers of the service

ACCELERATE WORK

Analysts can save up to 80% time when served tailored suggestions for materials necessary for their further research

INCREASE SALES

Various research show an increase in upselling revenue ranging from 10-50% caused by accurate "You Might Also Like" product recommendations

FORM HABITS

Serving accurate content can trigger cues, building strong habits and influencing usage patterns in customers

Project Summary

Key Objectives

Primary goal of the project has been to:

1. Provide customer segmentation and persona profiles based on buying behaviours (RFM analysis)
2. Recommend highly personalised Top 10 products to each customer as s/he embarks on a purchase journey

We have also worked on the following as the secondary project objectives (over and above the primary goals as stated above) –

1. Recommending pre-configured and personalized baskets to each customer (pro-active & personalized marketing campaigns)
2. Estimating Weekly/Monthly Product demand at store level

Our team has developed a highly personalized **Product Recommendation** solution for a physical store leveraging the immense amount of customer, product and transaction level from a big retailer in India. While the key focus has been on devising an e-commerce like item recommendation engine, we have also developed customer profiles that could be used by the retailer as well as the customer himself or herself, to enrich their shopping experience.

Why Recommendation Systems

Shoppers - online and offline - often suffer from the paradox of choice: a range of options so vast that it can feel overwhelming (e.g., the "250 varieties of cookies, 75 iced teas, 230 soups, 175 salad dressings, 275 cereals and 40 toothpastes" that Barry Schwartz mentions in his TED presentation on the topic) and result in low customer satisfaction, regardless of which option is selected. Recommender systems help people navigate broad ranges of options in the digital world, but without the assistance of a salesperson. In the physical stores, however, there has been little help for shoppers. Because of expectations created by web shopping, consumers increasingly expect offline stores to have the goods they want and make them easy to find. So, the ante is raised in the physical world. Global cues are anything but promising for the physical retail industry. Sears and Toys-R-Us were the two highest profile failures of 2018. But there are plenty of other traditional retailers struggling too. Against that backdrop is the growth of ecommerce and Amazon specifically. As traditional retailers adapt to an increasingly complex retail environment, with the looming possibility of recession in 2019, they will need to use their stores, technology and data in smart ways. They'll also need to potentially rethink the store experience entirely.

The idea of any Recommendation System is to recommend items to customers even before they know that they want it. Modern-day recommender systems contribute to millions of dollars in revenue in retail industry alone. The beauty of it is that its all data driven as Recommendation System find relationships between users and between items, just bases on actions, that are usually difficult to detect using simple analysis.

Our Approach (Data & Methodology)

The recommender system has been built based on the purchase patterns of customers in India leveraging the datasets provided by one of India's big retailers. The data comes from 7 stores across Amritsar, Hubli, Indore, Jamshedpur, Ludhiana, and Madurai. Approximately 80 lakh rows of product data and 1 lakh+ rows of transactional data have been provided.

We began with an extensive descriptive analysis of the data provided to gain insights into the magnitude and variation of purchase and overall sales pattern by various categorical attributes we had in data. This helped us create a deeper understanding of the retailer's business and customers' behaviour in general. Detailed report covering the hindsight analysis is provided in the later sections of this document.

The exploratory data analysis showed that all the 7 stores differed hugely in the customer purchase patterns and hence, for the purpose of detailed analysis, we picked up one store (Indore Malhar Mall). Recommender System development has also been pursued only for this store. We believe a similar approach would be applicable to all other stores as well. Each store would have its own unique customer strategy that would have to be developed after studying their specific portfolio, local nuances and customer behaviour in a similar detail.

The team has worked on both non-personalised recommenders and highly persoanlised "Top N recommender" techniques for Malhar Mall Store. In both the scenarios we have suggested best strategy available for deployment by the store after a rigorous model validation exercise.

KEY FINDINGS & RECOMMENDATIONS

Malhar Mall sees 5500+ unique customers visit its store every month. It is one of the key stores of the retailer and it seems to be focusing heavily on new customer acquisitions which is evident in its increasing customer growth month on month.

We find that ~78% of unique customers of the store are not as engaged. They don't visit the store regularly. They are either old customers who no more shop with the store, or they are one and done kind of customers, possibly not staying near the store. We believe such customers need not be focused upon for product recommendations as they need a different win-back effort from strategy perspective.

For the remaining 22%, deploying a product recommendation system would complement existing revenue enhancement strategies of the store. Given the immense amount of customer purchase data available, we have been able to develop a robust **Top 10 product recommender** using *User based collaborative filtering algorithm*, that can be integrated with existing customer smartphone apps of the retailer to provide real time product recommendations to each customer given the purchases they make with the store. To begin with, the algorithm just uses the POS data (as that is what was available to us during the project), however, over time customer's interaction with the app can help improve the recommendations by means of leveraging their digital footprint / clicks / ratings / feedback input into the model.

The detailed actionable insights and recommendations are provided in the concluding sections of the report.

DEPLOYMENT PLAN

There are two key phases to recommend products to any customer– during a current purchase journey and/or during the inactive stage, where customer is not interacting with the retailer. While in the former phase, the customer is on an explicit purchase event, the challenge in the latter phase is that the retailer won't know if the customer is looking to buy anything or not. An ideal product recommender would provide relevant recommendations and drive the customer to begin a purchase journey. Typically, such recommenders are built into a smartphone app that the customer can login into (Amazon web-app for example). This app serves highly personalised product recommendations at each of these phases as soon as the customer logs into it. In absence of this app,

In summary, the deployment typically would depend on the specific need and the technological maturity of the store. The team recommends that our solution gets integrated within an existing customer mobile app of the store. In case there is no such app existing, we propose developing a new one that puts transactions at the centre of the consideration and brings the entire POS on the smartphone.

Hence the idea is to first replace the physical, plastic made loyalty card by a screen in the app showing the same barcode. Second, in place of the paper receipt we upload a digital receipt on the customer's smartphone. The process is implemented in real time, transferring the receipt to the smartphone within a few seconds. Finally, the smartphone app stores the personal receipts and offers the customer relevant information about the purchases made.

This also ensures that the existing loyalty programmes are integrated into the app and over time the app provides immense amount of data from which the product recommender could learn from. Using the same clickstream analysis techniques as that of successful online retailers we track every movement within the app eventually and use the valuable information about consumers' intentions, preferences, and feedback.

Alternately, if the retailer is technologically not as mature, the current solution, can then be integrated with the existing POS system with a severe limitation of being used at the end of each purchase journey only. This would mean missing a powerful sales opportunity with a customer, who has come into the shop with an open wallet and even more open mind to spend! The recommendations at this stage would mostly be for the next purchase event of the customer and hence may not be as effective. However, this would still be an excellent "starting point" for a retailer who is currently not deploying any personalised product recommendation system whatsoever. The recommendations from the proposed system could be served on any existing loyalty platform that the customer has access to or in terms of simple message notifications.

Contents

1. Introduction and Problem Definition

Compared to online-retailers, bricks-and-mortar stores have only limited possibilities to understand consumer preferences, their intentions, and their feedback. The former is able to evaluate clickstream data collected on their webpages alongside the actual purchase data to put together a comprehensive view on individual customers. Bricks-and mortar stores on the other hand, have to rely solely on the evaluation of scanner data collected at the point of sale (POS). we attempt to solve this problem for offline retailers by devising a recommender system that would work solely with POS data to begin with and generate relevant personalized recommendations for valued customers at the store.

2. Exploratory Data Analysis (EDA)

Detailed below are the key highlights from the EDA we did across Customers, Stores, and Products respectively on the complete data available to us.

a. Data Sources

We had been provided with 2 sets of data - Product Sales and Transactional. Listed below are the key attributes available in these two datasets -

Table 1: Source Data Attributes

Product Sales Data	Transactional Data
Customer ID	Customer ID
Date of Birth	Date of Birth
Gender	Gender
State	State
Pin Code	Pin Code
Transaction Date	Transaction Date
Store Code	Store Code
Store Description	Store Description
Till No	Till No
Transaction Number by Till	Transaction Number by Till
Promotion Code	Tender Type
Promotion Description	Payment Amount by Tender
Product Code	Payment Used
Product Description	
Sale Price After Promotion	
Discount Used	

Here is the high-level summary of the source data: -

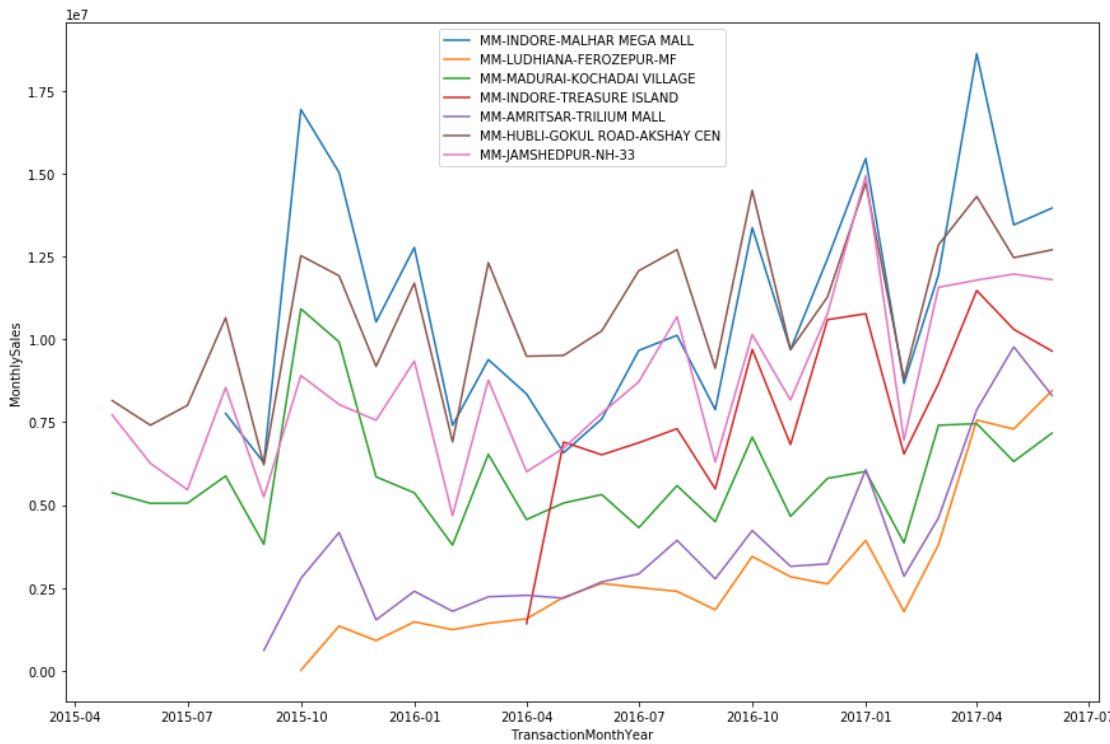
Table 2: Source data summary

Product Data	Value
Number of Records	7,981,262
Number of unique stores	7
Start Date	01/05/15
End Date	30/06/17
Number of days data available	791

b. Store Data

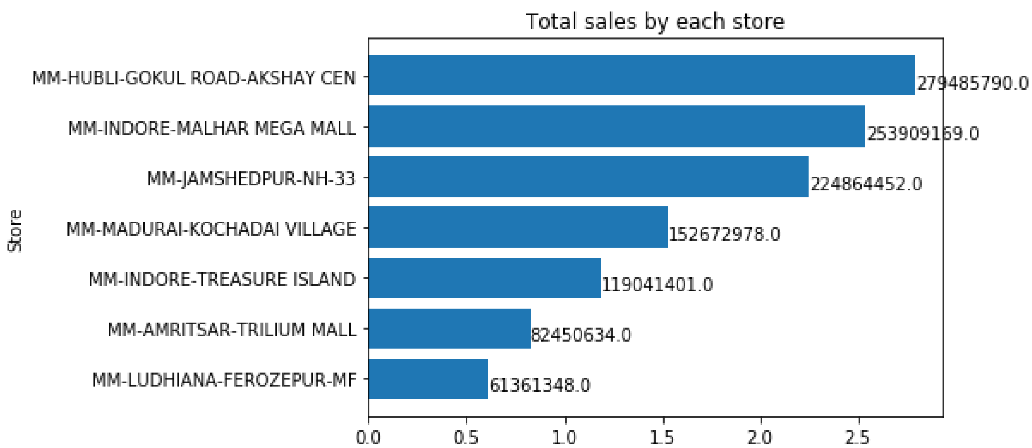
There are totally 7 stores that started operations over different time periods . Madurai, Jamshedpur, and Hubli are the oldest stores, whereas Indore-Treasure Island store is the newest one.

Figure 1: Store revenue contribution over time



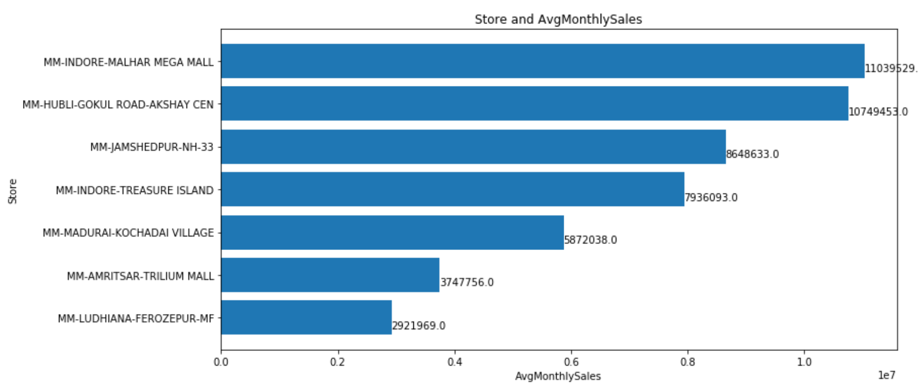
Hubli store is generating the maximum revenue overall, while Indore Malhar and Jamshedpur stores have the maximum revenue per till.

Figure 2: Overall store revenue



Indore Malhar is the store with maximum average monthly revenue (as explained by the following graph)

Figure 3: Average monthly store revenue



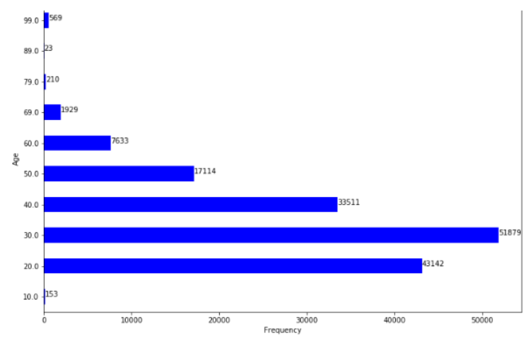
c. Customer Data

There are 175,978 Unique Customer across all the stores. Indore Malhar Mall has the maximum unique customer base, while Ludhiana store has the lowest. Most of the customers are from age range 20-40 years. Transactions are mostly done by the male customers. Also, we observed that the number of transactions with cash is more but the value of transactions with credit/debit cards are higher than the cash.

Table 3: Frequency of Payment

Mode of payment	Percentage of Transactions
Cash	42%
Cards	34%
BB Profit Club	11%
Payback	3%
Others	10%

Figure 4: Age distribution of customers



d. Product Data

There are 204,115 unique products. The table below shows the top 10 products across stores by the number of units sold

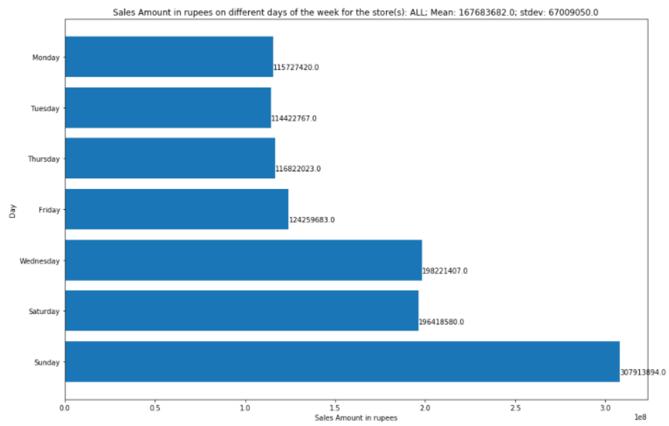
Table 4: Top 10 Products by Volume Sold

Product	Units sold
BB-CB-27X30X208SWG NEW	60,994
BB-CB-20X25X208SWG NEW	58,437
SUGAR MEDIUM LOOSE	51,338
BB-CB-20X25X168SWG-Suitable for ROI New	41,500
TOMATO LOOSE	35,307
BB-CB-27X30X168SWG-Suitable for ROI New	34,794
Fiber bag 45 GSM 20x25	31,149
ONION LOOSE	27,963
TATA SALT PP 1Kg	27,727
POTATO LOOSE	25,360

e. Sales

Our analysis says that Saturdays, Sundays, and Wednesdays are the days with the higher sales volumes. Also, the highest sale is during first 4 days of the month.

Figure 5: Revenue by Weekdays



f. EDA Summary

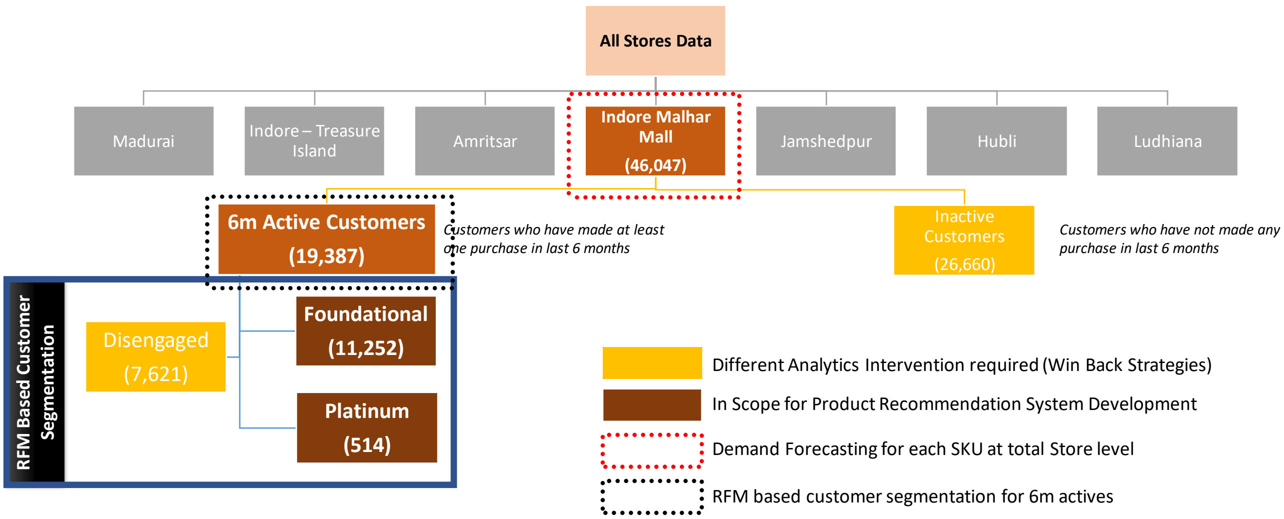
Following are the key insights from our exploratory data analysis.

- Only 25% customers visited the stores more than 5 times. Around 50% visited twice or less
- Ludhiana, Amritsar and Madurai stores are smaller in terms of no. of tills, customers, no. of unique products sold
- Madurai store has stagnating sales/customers when compared to others which may need some interventions
- Although Indore Treasure Island store has larger customer base, revenue per till is less
- Average Revenue Per Customer (ARPC) is not growing across the stores
- More credit notes are issued in Indore Malhar and Jamshedpur stores. That means there are more items getting returned. Talk to the business to understand more
- Indore Treasure Island store has highest parking fees reimbursement
- Debit/Credit Cards are used to pay 49% of the bill amount (cash 36%) though the stores are in tier 2 cities. Counter intuitive
- Challenges/Limitation in Datasets
    - Not a live customer and cannot talk to the business to get further details
    - Number of units sold for each product is not known. Only price after discount is given
    - Category of product is not given
    - Total revenue varies from store to store (INR 6 to 27 crores). We do not have size/scale of store and staffing size to evaluate whether it is justified
    - Large data that needs larger machine than a laptop for operations like pivot tables, apriori etc
- 412,729 unique transactions (transaction_number_by_till)

3. Methodology

We have selected Indore Malhar Mall to demonstrate how customer analytics helps in developing personalised marketing strategies that generate significant top-line impact for any offline retailer. The scope of the project is restricted to **customer segmentation** (for stronger customer understanding and lay the foundation to subsequent niche & targeted marketing efforts), **product recommendation system** and **understanding product demand** for the store. Highlighted below is the schematic of the overall methodology –

Figure 6: Overall Methodology



Final waterfall of data and the respective models we developed have been highlighted in the table below –

Table 5: Data Waterfall

Data	Unique Customers	# of transactions	Analytics Models Developed	Objectives
Total Data Available	1,65,055	6,86,419	None	
Only Indore Store (Store number 2655)	46,047	1,52,960	Demand Forecasting by SKUs, Association Rules / Market Basket	Accurate Procurement Planning, Revenue Leakage Reduction
6 months active	19,387	92,479	RFM Based Customer Segmentation	Targeted Customer Marketing, Higher Retention
At least 5 products purchased	16,548	89,097	Content Based, Neighborhood based, Matrix factorisation Based, Deep Learning on top of RFM based Customer Segments	Personalized Customer Marketing, Higher Revenue Per Customer
At least 10 products purchased	13,910	84,851		

a. Customer Segmentation

Recency-Frequency-Monetary Value (RFM Value) based customer segmentation has been done on 6 months active customers of Malhar Mall. Customers with at least one transaction in last 6 months are defined as **active customers**. Typically, revenue enhancement strategies are deployed on active customers as they have the highest propensity to respond to marketing efforts, given they have been recently engaged and buying with the store. Inactive customers need a totally different intervention from win-back perspective and hence have not been considered in scope for our project.

Elbow method has been used to define 3 as the optimal number of clusters on the given customer data. Subsequently, **K-means clustering** has been done using the RFM attributes of the customer to identify the 3 customer segments. Following is the RFM definition used –

Recency – how recent is the last purchase of the customer (calculated in weeks)
Frequency – average monthly number of purchases made by the customer at the store in last 2 years
Monetary – total spend by the customer in 2 years

b. Non-Personalized Recommendation Systems

We have implemented Association Rule mining as the first non-personalized recommender algorithm. Two versions have been run. In the first version, entire data is considered for the rules build and in the second one, very frequently purchased items like tomatoes, onions etc have been filtered out to unearth some non-obvious associations.

While we have taken product sales data for Indore Malhar mall for modelling, the same program functions can be used for other stores by passing the store name like the following:
'MM-INDORE-MALHAR MEGA MALL', 'MM-LUDHIANA-FEROZEPUR-MF',
'MM-MADURAI-KOCHADAI VILLAGE', 'MM-INDORE-TREASURE ISLAND',
'MM-AMRITSAR-TRILIUM MALL', 'MM-HUBLI-GOKUL ROAD-AKSHAY CEN',
'MM-JAMSHEDPUR-NH-33'

Association Rule Mining

Association Rule Mining (ARM) is used to identify patterns in data. It is helpful for retail data, web click analysis data etc. ARM can find features which occur together and features which are "correlated". For example, people who buy tea are likely to buy sugar. In other words, If (a customer buys tea powder), then (she buys sugar also). This does not necessarily mean that if people buy sugar, they buy tea powder. In General, if condition A tends to B it does not necessarily mean that B tends to A. Directionality matters here.

The measures of effectiveness of the rule are Support, Confidence and Lift.

In case of say If A -> Then B association rule, Support can be calculated as the fraction of rows containing both A and B or joint probability of A and B. Among rows containing A, Confidence is the fraction of rows containing B or conditional probability of B given A. Lift is the ratio Confidence to Support. If the lift is < 1 then A and B are negatively correlated else positively correlated and if it is equal to 1 it is not correlated.

Apriori algorithms and FP Growth algorithms are generally used for association rule mining.

Table 6: Apriori vs FP-Growth

	Apriori	FP-Growth
Technique	Generate singletons, pairs, triplets, etc.	Insert sorted items by frequency into a pattern tree
Runtime	Candidate generation is extremely slow. Runtime increases exponentially depending on the number of different items.	Runtime increases linearly, depending on the number of transactions and items
Memory Usage	Saves singletons, pairs, triplets, etc.	Stores a compact version of the database
Parallelizability	Candidate generation is very parallelizable	Data is inte-dependent; each node needs the root

We have used arules package/apriori algorithm for Association Rule Mining.

Apriori algorithm uses an iterative approach known as level-wise search. k itemsets are used to explore (k+1) itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted as L1. Next, L1 is used to find L2, the set of frequent 2-itemsets, which is used to find L3, and so on, until no more frequent k itemsets can be found. The finding of each Lk requires one full scan of the database.

At final iteration we will end up with many k itemsets which is basically called association rules. To select interesting rules from the set of all possible rules various constraint measures such as lift, support and confidence are applied.

For implementation details please refer to the Python workbook provided along with this report.

c. Personalized Recommendation Systems

We have used the following algorithms for Model Development & Validation for “Loyalists” segment of Malhar store –

User Based Collaborative Filtering

This relies on similar users given their past purchase behaviour and recommending items that closest users have bought in the past. For example, if Anny bought chocolates, tooth-brush and chewing gums in the past and Sam bought only chocolates and tooth brush, we might want to recommend chewing gums to Sam given that he looks to be similar buyer like Anny!

User-item matrix was created to begin with having user as rows and items as columns. Three types of actual ratings have been tried for each model – actual purchase count of the item, scaled purchase count and just a purchase flag (which indicates whether the user has purchased a particular item in the user-item matrix). Once the ratings matrix is formulated, we find user similarities using Cosine, MSD and Pearson similarity measures. For each user, a set of top 10 closest neighbours is then identified, and the items purchased by these closest neighbours becomes the starting item recommendation candidates. We have then scored these unique items by multiplying their scaled rankings with user similarity score of the respective user. Finally ordering these items on this score helps to get to the top 10 recommendations for each user!

Leave one out CV has been used to validate the top 10 recommender solution built as above. We have removed one item from each user purchase list and evaluated the recommender’s ability to get the removed item featuring in the top 10 recommendations. Hit rate is thus the key evaluation metric used for all collaborative filtering models.

Item Based Collaborative Filtering

The method is pretty similar to User Based Collaborative Filtering. We replace user similarities with item similarities. We look at items users have liked in the past and then recommend similar items. Item-User matrix is created to begin with having items as rows and users as columns. Item Similarity is calculated using Cosine, MSD and Pearson similarity measures. The matrix again has 3 kind of ratings as in the case of User Based collaborative filtering - actual purchase count of the item, scaled purchase count and just a purchase flag.
Rest of the process remains same in terms of generating the candidate items that are similar to items a user has already bought and then scoring / ranking them using the actual rating and similarity score.

We use LOOCV (Leave-one-out cross validation) here as well for cross validation and take Hit Rate as the key evaluation metric.

KNN Recommenders

Since it is difficult to evaluate collaborative filtering methods without running costly real-world experiments (as they are not based on making rating predictions), we can’t measure their accuracy offline. This is where KNN recommenders are useful as they use the same concept as collaborative filtering but are able to make rating predictions.

We have tried both user and item based KNN recommenders. The methodology used is given below –

User based KNN –

Here we are trying to predict the rating of each item for each user, we restrict our nearest neighbours to k users who have bought the item in question. This helps get k ratings on each item and we can calculate mean similarity score weighted by these k ratings to arrive at the final rating prediction for this item.

Item based KNN –

This follows a similar approach as above; We begin with items similar to the one in question and select the top K neighbours with highest similarity score and also purchased by the given user. We then compute the mean similarity score for each item weighted by its ratings by the user. This is the final rating prediction of this item for the selected user.

Since now we have rating predictions, we can use classic train-test validation methodology. We divided the data into 75% training and 25% test to administer model validation.

Model Based Methods – SVD

While Collaborative filtering gives good results, we have also explored model-based methods, matrix factorisation specifically to see if that can help improve on rating prediction accuracy and relevance. This method also is seen to perform better with sparse data as compared to collaborative filtering method.

Singular Value Decomposition is one such Matrix Factorisation Method. This algorithm tries to find latent features of users and items in its own using underlying techniques like PCA and SGD / ALS. The idea is to describe users and items as combinations of different amount of each latent feature.

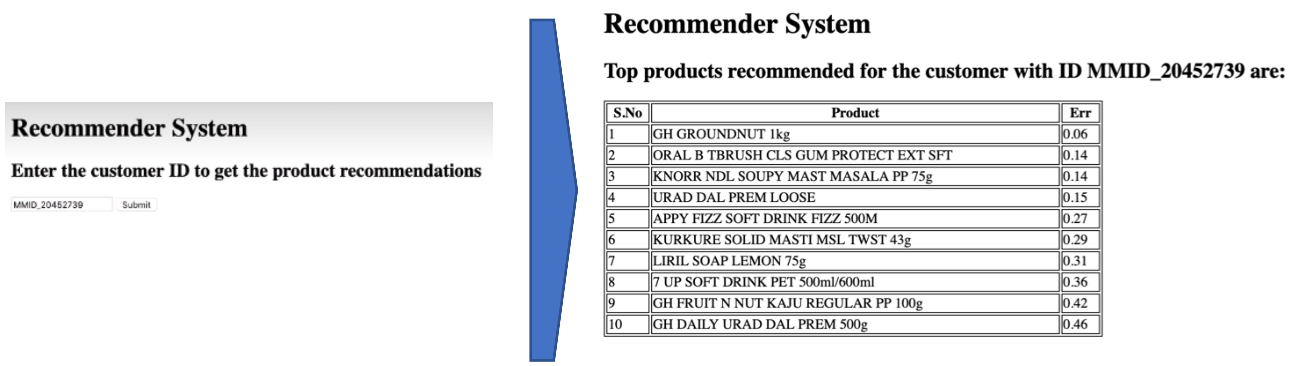
Since this is also a rating prediction algorithm, we can use classic train-test validation approach. As with KNN recommenders, we divide the data here into 75% training and 25% test for model validation.

Model Based Methods – SVD++

SVD++ is a variant on SVD algorithm, that uses non negative matrix factorisation. It is the algorithm that won the Netflix Prize and is quite popular in general. The difference is in the loss function that is run while running the stochastic gradient descent algorithm internally. In SVD++, this loss function takes into account the idea that merely rating an item at all, is some sort of implicit interest in the item, irrespective of the rating itself.

Please refer section 4g for results by technique. Our team has developed a sample user interface for the store to derive personalized recommendations by customer. Illustrations from screens for user input and system output are as follows: -

Figure 7: User Interface for Personalized Recommender System



S.No	Product	Err
1	GH GROUNDNUT 1kg	0.06
2	ORAL B TBRUSH CLS GUM PROTECT EXT SFT	0.14
3	KNORR NDL SOUPY MAST MASALA PP 75g	0.14
4	URAD DAL PREM LOOSE	0.15
5	APPY FIZZ SOFT DRINK FIZZ 500M	0.27
6	KURKURE SOLID MASTI MSL TWST 43g	0.29
7	LIRIL SOAP LEMON 75g	0.31
8	7 UP SOFT DRINK PET 500ml/600ml	0.36
9	GH FRUIT N NUT KAJU REGULAR PP 100g	0.42
10	GH DAILY URAD DAL PREM 500g	0.46

d. Product Demand by SKUs

We have forecasted weekly demand for Maggi Noodles and Watermelons, which have been two high demand products in the Malhar Mega Mall in 2017. We have used 2017 data for Malhar Mega Mall which has been filtered to exclude details of commonly sold products such as potatoes, tomatoes and onions. We have considered time-series (TS) forecasting approaches for the selected products versus driver-based machine learning driven forecasting approach based on the data available to us and it has been our endeavour to apply simple to complex TS approaches and assess the final results basis achieved accuracy.

We have tried the following four approaches and compared their results: -

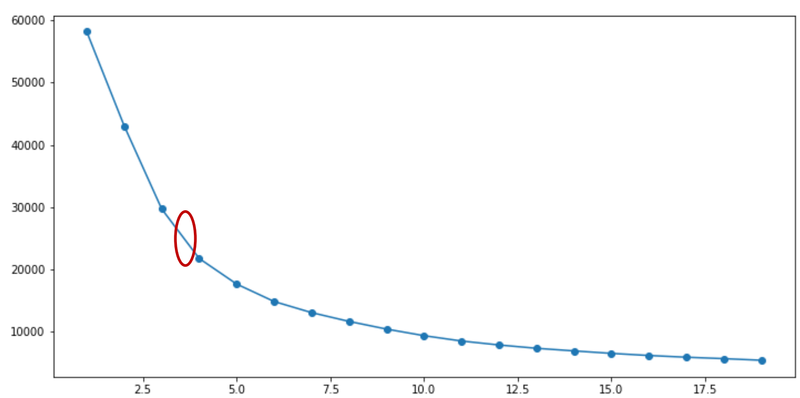
1. ARIMA
2. Holt's Method
3. Holt Winter's Method
4. Simple Moving Average

4. Main Result

e. Customer Segmentation and Profiles

There are 19,387 unique active customers shopping with Malhar Mall currently. We defined the RFM attributes for each of these customers and then using elbow method find out the optimal number of clusters existing inherently in the data. Following is the output from the elbow method run from k=1 to k=20

Figure 8: Elbow Chart



We see that 3 or 4 clusters are optimal for this data. Actual running of these clusters showed that 3 is the optimal number as with k=4 we get an extremely small cluster with unstable size.

Following are the final insights from customer segmentation and segment profiles that help us better understand the customer base with the retailer –

"Loyalists" are the high value customers who are typically old timers. Only 27% of them are new with first transaction in last 90 days. They are very frequent shoppers too, with 1.5 transactions per month on average. This is a very small cluster with only 2.5% of the customer base, but given the very high loyalty index, these customers need to be looked at separately with very personalized targeting and regular marketing interventions in order to generate maximum revenue opportunities and also ensure that they continue to stay engaged with the store

"Foundational Customers" form the core of the customer base with 58% of the portfolio. They are mostly new customers with 42% of them making their first purchase in last 90 days. This group needs to be retained and focused upon as they contribute more than 50% to the revenues of the store.

"Disengaged Customers" are the remaining 40% of the portfolio. These are mostly old customers who have preferred not to come back and shop with the retailer. They are mostly attrited customers, who need to be won back using very targeted marketing campaigns. These remain out of scope for Product recommendations.

Clusters	# of Customers	Avg Recency (Weeks)	Avg Frequency (Per Month)	Avg Monetary Value	Number of Unique Prds Bought
Disengaged Customers	7,621	17.74	0.13	₹ 5,041	33
Foundational Customers	11,252	5.0	0.19	₹ 7,131	45
Loyalists	514	4.6	1.51	₹ 62,244	221
Grand Total	**19,387**	**10.0**	**0.20**	**₹ 7,770**	**45**

Following figure shows Vintage (in days) distribution for each customer segment. Vintage indicates how old the customer (or customer group) is.

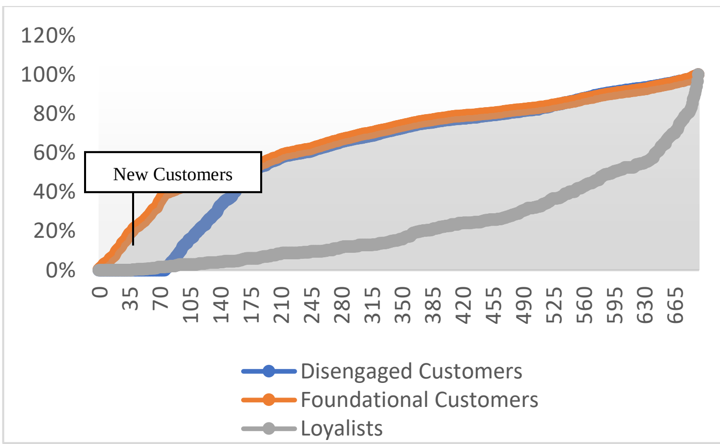


Figure 9: Customer Vintage in Days by Customer Segments

f. Non-Personalized Recommendation System Results

We have considered Indore Malhar mall data and implemented two versions of association rule mining algorithms. In the first version, the whole data is considered. In the second one, most commonly purchased commodity items like tomatoes, onions etc are filtered out. Below is the sample output for frequent item sets/market baskets:
Sample outputs for Indore Malhar Mall

Market basket

Before filtering the data

```
     items                                                          support    count
[1]  {ONION LOOSE,POTATO LOOSE,TOMATO LOOSE}                        0.08804244 3236
[2]  {CORIANDER,POTATO LOOSE,TOMATO LOOSE}                          0.06575976 2417
[3]  {CUCUMBER GREEN LOOSE,POTATO LOOSE,TOMATO LOOSE}               0.06439940 2367
[4]  {CUCUMBER GREEN LOOSE,ONION LOOSE,TOMATO LOOSE}                0.06279418 2308
[5]  {CORIANDER,ONION LOOSE,TOMATO LOOSE}                           0.06271256 2305
[6]  {CORIANDER,CUCUMBER GREEN LOOSE,TOMATO LOOSE}                  0.05748878 2113
[7]  {CORIANDER,ONION LOOSE,POTATO LOOSE}                           0.05542103 2037
[8]  {LADYFINGER LOOSE,POTATO LOOSE,TOMATO LOOSE}                   0.05479527 2014
[9]  {BOTTLE GOURD LONG,POTATO LOOSE,TOMATO LOOSE}                  0.05457761 2006
[10] {CUCUMBER GREEN LOOSE,ONION LOOSE,POTATO LOOSE}                0.05389743 1981
[11] {Carrot English Loose,POTATO LOOSE,TOMATO LOOSE}               0.05076860 1866
[12] {BB-CB-27X30X168SWG-Suitable for ROI New,POTATO LOOSE,TOMATO LOOSE} 0.05057815 1859
[13] {BOTTLE GOURD LONG,ONION LOOSE,TOMATO LOOSE}                   0.05038770 1852
[14] {BB-CB-27X30X168SWG-Suitable for ROI New,ONION LOOSE,TOMATO LOOSE}  0.05003401 1839
[15] {LADYFINGER LOOSE,ONION LOOSE,TOMATO LOOSE}                    0.04997959 1837
```

After filtering the data

```
     items                                                                              support     count
[1]  {DOVE SOAP MOISTURE CREAM 3*100g,PEARS SOAP PURE GENTLE 3*125g,RED LABEL CTC TEA PP 1kg}  0.004338213 155
[2]  {GROUNDNUT LOOSE,POHA THICK LOOSE,SABUDANA BIG LOOSE}                              0.003918386 140
[3]  {BITTER GOURD LOOSE,RADISH RED,SPINACH}                                            0.003414593 122
[4]  {GROUNDNUT LOOSE,MOONG DAL SPLIT PREM LOOSE,POHA THICK LOOSE}                      0.003386605 121
[5]  {MINT LEAVES,RADISH RED,SPINACH}                                                   0.003330628 119
[6]  {DOVE SOAP MOISTURE CREAM 3*100g,KISSAN KETCHUP DOY PACK 950g,RED LABEL CTC TEA PP 1kg}  0.003246662 116
[7]  {KISSAN KETCHUP DOY PACK 950g,RED LABEL CTC TEA PP 1kg,RIN DET POWDER ADVANCE 6Kg}  0.003218674 115
[8]  {BEANS FRENCH FLAT,MINT LEAVES,SPINACH}                                            0.003134709 112
[9]  {DOVE SOAP MOISTURE CREAM 3*100g,KISSAN KETCHUP DOY PACK 950g,PEARS SOAP PURE GENTLE 3*125g} 0.003050743 109
[10] {GROUNDNUT LOOSE,POHA THICK LOOSE,TUR DAL ECONOMY LOOSE}                           0.002994766 107
[11] {BEANS FRENCH FLAT,BITTER GOURD LOOSE,SPINACH}                                     0.002994766 107
[12] {DOVE SOAP MOISTURE CREAM 3*100g,RED LABEL CTC TEA PP 1kg,RIN DET POWDER ADVANCE 6Kg}  0.002910801 104
[13] {BEANS FRENCH FLAT,RADISH RED,SPINACH}                                             0.002910801 104
[14] {BITTER GOURD LOOSE,MINT LEAVES,SPINACH}                                           0.002854824 102
[15] {ASH GOURD,BEANS FRENCH FLAT,BITTER GOURD LOOSE}                                   0.002826835 101
```

Association rules

Before filtering the data

```
     lhs                                                      rhs                     support     confidence lift     count
[1]  {BB-CB-27X30X208SWG NEW,Kiddie Club Card 2016}        => {Kiddie Currency Rs 500} 0.001931710 0.6513761  161.7657 71
[2]  {BB-CB-20X25X208SWG NEW,Kiddie Club Card 2016}        => {Kiddie Currency Rs 100} 0.001169909 0.5810811  140.5108 43
[3]  {Kiddie Currency Rs 100,POTATO LOOSE}                 => {Kiddie Club Card 2016}  0.001006666 0.9487179  139.4805 37
[4]  {Kiddie Club Card 2016,POTATO LOOSE}                  => {Kiddie Currency Rs 100} 0.001006666 0.5692308  137.6452 37
[5]  {BB-CB-20X25X208SWG NEW,Kiddie Currency Rs 100}       => {Kiddie Club Card 2016}  0.001169909 0.9347826  137.4317 43
[6]  {BB-CB-20X25X208SWG NEW,Kiddie Currency Rs 500}       => {Kiddie Club Card 2016}  0.001033873 0.9268293  136.2624 38
[7]  {Kiddie Club Card 2016,TOMATO LOOSE}                  => {Kiddie Currency Rs 100} 0.001033873 0.5588235  135.1287 38
[8]  {Kiddie Currency Rs 100,TOMATO LOOSE}                 => {Kiddie Club Card 2016}  0.001033873 0.8837209  129.9247 38
[9]  {BB-CB-20X25X208SWG NEW,Kiddie Club Card 2016}        => {Kiddie Currency Rs 500} 0.001033873 0.5135135  127.5283 38
[10] {BB-CB-27X30X208SWG NEW,Kiddie Currency Rs 500}       => {Kiddie Club Card 2016}  0.001931710 0.8658537  127.2978 71
[11] {BB-CB-27X30X208SWG NEW,Kiddie Currency Rs 100}       => {Kiddie Club Card 2016}  0.001360359 0.8474576  124.5932 50
[12] {BB-CB-27X30X208SWG NEW,T24 STARTER KIT MP}           => {T24 Paid Recharge}      0.001061080 0.9512195  122.6739 39
[13] {BB-CB-16X20X168SWG-Suitable for ROI New,T24 STARTER KIT MP} => {T24 Paid Recharge} 0.001088287 0.9302326 119.9674 40
[14] {PB Card FVRL,T24 STARTER KIT MP}                     => {T24 Paid Recharge}      0.001061080 0.9285714  119.7531 39
[15] {PB Card FVRL,T24 Paid Recharge}                      => {T24 STARTER KIT MP}     0.001061080 1.0000000  119.7231 39
```

After filtering the data

```
     lhs                                       rhs                                  support confidence      lift count
[1]  {PB Card FVRL,
      T24 Paid Recharge}                    => {T24 STARTER KIT MP}               0.001091550  1.0000000 118.70100    39
[2]  {BB-CB-16X20X168SWG-Suitable for ROI New,
      T24 Paid Recharge}                    => {T24 STARTER KIT MP}               0.001119539  1.0000000 118.70100    40
[3]  {BB-CB-16X20X168SWG-Suitable for ROI New,
      T24 STARTER KIT MP}                   => {T24 Paid Recharge}                0.001119539  0.9302326 118.70100    40
[4]  {PB Card FVRL,
      T24 STARTER KIT MP}                   => {T24 Paid Recharge}                0.001091550  0.9285714 118.48903    39
[5]  {DOVE SOAP MOISTURE CREAM 3*100g,
      VIM DROP DW ACTV GEL YLW LEMON 250m}  => {VIM DROP DW ACTIVE GEL YLW LEMON 750ml} 0.001203504  0.6515152  54.26104    43
[6]  {DUKES WFR BISC ORANGE WAFFY 75g,
      DUKES WFR BISC STRAWBERRY WAFFY 75g}  => {DUKES WFR BISC PINEAPPLE WAFFY 75g}     0.001063562  0.5135135  52.57113    38
[7]  {RIN DET POWDER ADVANCE 6Kg,
      VIM DROP DW ACTV GEL YLW LEMON 250m}  => {VIM DROP DW ACTIVE GEL YLW LEMON 750ml} 0.001007585  0.6206897  51.69375    36
[8]  {DOVE SOAP MOISTURE CREAM 3*100g,
      LUX SOAP VELVET TOUCH 3*150g}         => {LUX SOAP SOFT&TOUCH 3*150g}       0.001007585  0.6000000  49.85442    36
[9]  {DUKES WFR BISC ORANGE WAFFY 75g,
      DUKES WFR BISC STRAWBERRY WAFFY 75g}  => {DUKES WFR BISC VANILLA WAFFY 75g} 0.001175516  0.5675676  22.13823    42
[10] {DAWAT BASMATI RICE DEVAYA PP 5kg,
      FORTUNE PLUS SOYABEAN OIL JR 5L}      => {MADHUR CRYSTAL SUGAR PP 5Kg}      0.001315458  1.0000000  22.02774    47
[11] {DUKES WFR BISC ORANGE WAFFY 75g,
      DUKES WFR BISC PINEAPPLE WAFFY 75g}   => {DUKES WFR BISC VANILLA WAFFY 75g} 0.001539366  0.5392157  21.03236    55
[12] {DNU,
      DUKES WFR BISC ORANGE WAFFY 75g}      => {DUKES WFR BISC VANILLA WAFFY 75g} 0.001175516  0.5185185  20.22505    42
[13] {DAWAT BASMATI RICE DEVAYA PP 5kg,
      MAHAKOSH SOYABEAN OIL JR 5L}          => {MADHUR CRYSTAL SUGAR PP 5Kg}      0.001119539  0.8510638  18.74702    40
[14] {AMARANTHUS RED,
      BANANA CHAKKARAKELI}                  => {CHILLI LIGHT GREEN}               0.001287470  0.5111111  17.72960    46
[15] {FORTUNE PLUS SOYABEAN OIL JR 5L,
      GH RICE BASMATI DIL KHUSH 5Kg}        => {MADHUR CRYSTAL SUGAR PP 5Kg}      0.001371435  0.8032787  17.69442    49
```

g. Personalized Recommendation System Results

Following table shows the results of running various personalized recommendation algorithms on Malhar Mall "Loyalists" segment using the purchase data for last 2 years-

Algorithm	Rating Attribute	Similarity Measure	Validation Approach	Accuracy	HR	ARHR
User Based Collaborative Filtering	Purchase Frequency	Cosine	LOOCV		13%	4.6%
		Pearson			18%	6.6%
	Purchase Flag	Cosine			17%	4.4%
		Pearson			8%	0.3%
Item Based Collaborative Filtering	Purchase Frequency	Cosine			2%	0.05%
		Pearson			0%	0%
	Purchase Flag	Cosine			3%	2.5%
		Pearson			7%	0.2%
User Based KNN	Purchase Frequency	Cosine	Train 75% / Test 25%	3.19		
		Pearson		3.2		
	Purchase Flag	Cosine		0	0	0
		Pearson		0	0	0
Item Based KNN	Purchase Frequency	Cosine		3.57		
		Pearson		3.35		
	Purchase Flag	Cosine		0	0	0
		Pearson		0	0	0
SVD	Purchase Frequency			3.14	5%	3.12%
	Purchase Flag			0.035	0	0
SVD++	Purchase Frequency			3.16	7%	4.48%
	Purchase Flag			0.013	0	0
SVD Tuned	Purchase Frequency			3.14	4%	4%
	Purchase Flag			0.012	0	0

We get the best Hit Rate and Average Reciprocal Hit Ranking (ARHR) using the User Based Collaborative Filtering approach with Pearson Similarity (highlighted in yellow above). Hit rate reflects how often we can recommend a product that the customer actually bought and is typically a better metric to evaluate recommender system than accuracy. Hence, we select **User Based Collaborative Filtering approach** with Pearson Similarity as the final model for the "Loyalists" segment at Malhar Mall.

In order to understand the type of recommendations thrown by the model, we test it for a customer - MMID_20412951. This is a female customer of the store, who has made 59 transactions in the 2 years of data we have. She transacted 14 months of the 24 months, bought 130 unique products overall and spent Rs 1.5 Lakhs roughly on the purchases in 2 years!! A super loyal customer shopping 2.5 times a month on average and spending ~Rs 2600 per transaction. Following table highlights the Top 10 recommender output for this customer from various algorithms: -

Table 7: Output of Top 10 Recommender Algorithms for a customer

Collaborative Filtering Methods				
Puchase Frequency Rating	Simple User CF Cosine	Simple Item CF Cosine	User KNNBasic cosine	Item KNNBasic cosine
	AMUL MILK GOLD MILKY MILK PP 500ml	FERRERO RO CHOCOLATE T16 200G 2	BOTTLE GOURD LONG	FRESH FOR YOU SOAN PAPDI BX 1Kg
	SUGAR MEDIUM LOOSE	KINDER JOY CHOCOLATE LEI BOYS 20g 75	LEMON LOOSE	DJTR-0150-BT-B-SF-SAT,34,JET BLACK
	SANCHI MILK GOLD PP 500ml	BRITANNIA GD CHOCONUT CHOCHIP 2*120g OP	METHI	RATH VANASPATI PP 1L
	BB-CB-16X20X208SWG NEW	FIAMA DW SOAP EXOTIC DREAM BX 115g	GINGER	AFSH-4166-PK76-B-FS-RF-P, S, DEEP PETROL
	PARLE SLT BIS KRACKJACK 80g	DERMICOOL TALC LAVENDER 150G	AMUL MILK GOLD MILKY MILK PP 500ml	AFSH-0160-FC-C-HS-SF-SL, S, CORAL
	MAHAKOSH SOYABEAN OIL PP 1L	PALMOLIVE SHW GEL AROMA MORNG TONIC 250M	PROMO JAC D CURTAIN ALPS 1.37x2.1m Multi	FRESH&PURE SALT REFINED PP 1kg
	TATA SALT PP 1Kg	SAFFOLA OATS MASALA CLASSIC PP 39g	AQUAFINA WATER 1L	AFSH-4166-PK72-C-FS-RF-P, S, OXFORD NAVY
	BB-CB-20X25X168SWG-Suitable for ROI New	LAKME COMPLEXION CARE CREAM BRONZE 30g	BAGRRY MUSELI CRUNCHY BX 400g	Cannon SD Jacquard D1 B, 75CMX1.5M, ROSE
	PARLE SLT BIS MONCO CLASSIC REG 80g	BAJAJ HAIROIL BRAHMI AMLA 300ml	SOFY SAN PAD BODYFIT OVERNIGHT XXXL 3P	Set Of 6 Bowls Mapple
	SOO FRESH BREAD WHITE 400g	PRESTIGE GTM03L BLACK GT GAS STOVE	24 LM TUR DAL PP 500g	LPY-0010-PDP-207-SH, M, FUSHIA
	SOO FRESH MUFFIN VANILLA 6P	BOURNVITA CKS CRNCH CHOCLTY 4+1 120g OP		
	Simple User CF Pearson	Simple Item CF Pearson	User KNNBasic Pearson	Item KNNBasic Pearson
	AMUL MILK GOLD MILKY MILK PP 500ml	WAGHBAKRI CTC TEA LEAF PP 1K	LETTUCE ARUGULA	EKTAA UJJAINI POHA PP 500g
	SUGAR MEDIUM LOOSE	BRITANNIA BISC GDAY CASHEW 100g	HALDIRAM NI NAMK CHANA CHOOR 150G	Fiber bag 45 GSM 20x25
	BB-CB-20X25X168SWG-Suitable for ROI New	MAMY POKO PANTS BABY DIAPER LRG 62/64P	SANCHI MILK GOLD PP 500ml	SUNFEAST NDL YIPPEE MAGIC MSALA 240/280g
	SANCHI MILK GOLD PP 500ml	DABUR HAIROIL AMLA 450ml	AMUL MILK GOLD MILKY MILK PP 500ml	CORNITOS NACHO CHEESE 60g
	LUX SOAP VELVET TOUCH 3*150g	MOTHERS REC PAPAD PUNJABI 200g	LIPTON TEA YELLOW LABEL BX 250g	GH DAILY URAD DAL PREM 1kg
	BB-CB-16X20X208SWG NEW	MOTHERS R PICKLE MANGO 200G	FEM HAIR REMOV ANTI DARKENING ROSE 40g	GH CHANA MAUSAMI 1kg
	PARLE SLT BIS KRACKJACK 80g	COLOUR NAT HRCLR BROWN NO4 100m	MOTHERS REC PAPAD URAD 200g	ALL OUT INSECT KILL MULTI CN 600ml
	SOO FRESH MUFFIN VANILLA 6P	MOUNTAIN DEW SOFT DRINK 1.25L	LAKME PERFECT LIQ FOUND MARBLE 27ML	BOURNVITA HEALTH DRINK PP 500g
	MAHAKOSH SOYABEAN OIL PP 1L	DABUR CHYAWANPRASH 1KG	LAKME COMPLEXION CARE CREAM BEIGE 30g	HARI DARSHAN AGARBATHI TATHASTU
	AMUL BUTTER 100G	THUMS UP SOFT DRINK BT 1.75L	LAK N/E 31 TRUE WEAR COLOUR CRUSH 9ml	AMUL CHEESE SLICE 100G
	TATA SALT PP 1Kg	CRMT CLING FILM 30MTR		
Puchase Flag Based Rating	Simple User CF Cosine	Simple Item CF Cosine	User KNNBasic cosine	Item KNNBasic cosine
	FB SIS NAMKEENS	BB-CB-20X25X168SWG-Suitable for ROI New	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE
	BOTTLE GOURD LONG	BB-CB-16X20X168SWG-Suitable for ROI New	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g
	BB-CB-20X25X168SWG-Suitable for ROI New	BRITANNIA NUTRICHOE DIGEST 4+1 100g OP	AAKASH NAMKEEN MUMBAI MIX 170G	AAKASH NAMKEEN MUMBAI MIX 170G
	CAPSICUM GREEN	STAYFREE SC COT WG 2*20P+SC DRY XL 7P OP	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml
	LEMON LOOSE	LUX SOAP VELVET TOUCH 3*150g	GALAXY CHOC MILK BAR 40g 31	GALAXY CHOC MILK BAR 40g 31
	CAULIFLOWER	NIVEA TALC MUSK GENTLE CARE 400g	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg
	GINGER	SENSODYNE TPASTE FRESH MINT BX 130g	TAJMAHAL CTC TEA BX 250g	TAJMAHAL CTC TEA BX 250g
	CABBAGE	NIVEA BODY LTN MILK ALMOND OIL BT 200ml	PRATHA AGARBATTI ROSE BX 15P	PRATHA AGARBATTI ROSE BX 15P
	BRINJAL BHARTA PURPLE	TIDE DET POW PLUS JASMIN ROSE 4Kg	POMEGRANATE PREMIUM PKD	POMEGRANATE PREMIUM PKD
	ORANGE NAGPUR LOOSE	COLGATE TPASTE MAXFRESH RED 3+1 150g OP	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g
	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg	BOTTLE GOURD LONG		
	Simple User CF Pearson	Simple Item CF Pearson	User KNNBasic Pearson	Item KNNBasic Pearson
	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE	AFL-KNIT-CLASSIC-SHORTS, XL, NAVY BLUE
	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g	AIRWICK RM EVERFRESH LEMON GARDEN BX 50g
	AAKASH NAMKEEN MUMBAI MIX 170G	AAKASH NAMKEEN MUMBAI MIX 170G	AAKASH NAMKEEN MUMBAI MIX 170G	AAKASH NAMKEEN MUMBAI MIX 170G
	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml	LIFEBUOY HW KITCHEN FRESH REFIL PP 185ml
	GALAXY CHOC MILK BAR 40g 31	GALAXY CHOC MILK BAR 40g 31	GALAXY CHOC MILK BAR 40g 31	GALAXY CHOC MILK BAR 40g 31
	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg	AASHIRVAAD SELECT SHARBATI ATTA PP 5Kg
	TAJMAHAL CTC TEA BX 250g	TAJMAHAL CTC TEA BX 250g	TAJMAHAL CTC TEA BX 250g	TAJMAHAL CTC TEA BX 250g
	PRATHA AGARBATTI ROSE BX 15P	PRATHA AGARBATTI ROSE BX 15P	PRATHA AGARBATTI ROSE BX 15P	PRATHA AGARBATTI ROSE BX 15P
	POMEGRANATE PREMIUM PKD	POMEGRANATE PREMIUM PKD	POMEGRANATE PREMIUM PKD	POMEGRANATE PREMIUM PKD
	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g	AIRWICK RM EVERFRESH LAVENDER DEW BX 50g
	AAKASH NAMKEEN LAJAWAB MX 200G	AAKASH NAMKEEN LAJAWAB MX 200G		

	Matrix Factorisation Methods		
	SVD	SVD++	SVD Tuned
Purchase Frequency Rating	WATERMELON KIRAN	LETTUCE ARUGULA	SOO FRESH BREAD WHITE 400g
	GH TUR DAL PREMIUM 5Kg	LEMON LOOSE	GH TUR DAL PREMIUM 5Kg
	AGRIPURE ATTA SAMPOORNA 10kg	MUSHROOM BUTTON	BEANS CLUSTER
	PAPAYA ROUND SMALL	AMUL MILK GOLD MILKY MILK PP 500ml	WATERMELON KIRAN
	LUX SOAP VELVET TOUCH 3*150g	SOO FRESH FOCACCIA OLIVE 200g	WHEAT SHARBATI TOPAZ LOOSE
	SOO FRESH BREAD WHITE 400g	CAPSICUM GREEN	HORLICKS HEALTHDRINK CHOCOLATE BIB 1K
	AMARANTHUS RED	SOO FRESH BREAD WHITE 400g	SOFRSH VD BREAD WHITE 400g
	RADISH RED	MADHUR CRYSTAL SUGAR PP 5Kg	VIM UTENSL BAR 3*200g
	SOO FRESH BREAD WHOLE WHEAT MLTGRN 400g	GH TUR DAL PREMIUM 5Kg	Colocasia Leaves
	FORTUNE SUNFLOWER OIL JR 5L	FB SIS NAMKEENS	MAGGI ATTA NOODLES MASALA PP 300g
Purchase Flag Based Rating	TUR DAL PREM LOOSE	SABW-4161-SAJAM1652-B-JM-KN, 3-4, NAVY	GNG NP SUPRA CHAIR 66x58x83, Beige, NA
	FACES GLAM ON PERFECT MASCARA BLACK 8ml	Uncle Tell Me A Story	DAVE POW SPICE GARAM MASALA 250g
	MAMY POKO PANTS BABY DIAPER SMALL 80P	MTR SNACKMIX RAVA IDLI PP 500g	PRATHA GHEE WATI JR 20P
	PBBCD-4161-KTF-15-HS-C-KTF, 3-6, RED	BUSH-3161-BC20-B-FS-CX, L-RF, BLUE	GH SOYABEAN 500g
	ACT II POPCORN SPICY PUDINA PP 35g	SS16-WNBELS-L1290-H&S, 39, GOLDEN	PARLE CANDY MELODY CHOCOLATY 391g 43
	BAKER STREET BREAD STICKS HERBED BX 100g	GH WHO SPICE RED CHILLI 200g	MAPRO CHOCOLATE FALCHOOS 15P
	KELLOGGS OATS MASST MASALA PP 39g	Recron 16 x 24 / 40x61cm	SOO FRESH DIPS MANGO JALAPENO 200g
	CRMT HANDWASH SANDAL PP 900ml	CTTHM-3151-THRML PY, 11-12, GREY MELANGE	DETTOL HW LQ SKIN CARE PP 175ml
	HAMAM SOAP 150G	BUTR-3151-MAUL-A-SF-TW-SL, 30, GINGER	DEL MONTE MAYONNAISE THICK&CREAMY 900g
	TT-0010-PO3-PSTLPRT, XXL, NA	NFE ASSORTED PENCIL POUCH C	BAJAJ HAIROIL ALMOND DROP 100ml

Considering the customer mostly purchases grocery on a regular basis and a usual basket contains items of daily use like tea, oil, ketchup, oats, soaps, lotions toothpastes etc, we believe that the output from collaborative filtering is most relevant and gives her options to diversify her bucket by including complementary items like milk, bread, salt and sugar, that she has surprisingly not purchased with the store yet. The store may consider recommendations like these and run campaigns to lure the customer into increasing the basket size.

h. Product Demand Estimation Results

We focused on the Malhar Mega Mall for this exercise with the objective to predict demand for two high selling products in the store. We shortlisted a couple of high demand products (Maggie noodles and watermelon) and estimated the weekly demand for them. The comparison of results (MAPE of estimated weekly demand versus actuals) is as follows:

Table 8: Weekly Demand Estimation Results by Time-series Modelling Technique

SKU	ARIMA	Holt's Method	Holt Winter's Method	Simple Moving Average
Maggi Noodles Masala (425 gm)	0.517	0.429	50.575	0.052 (last week prediction)
Watermelon (Kiran)	218.167	41.209	1928.57	1.835 (last week prediction)

We see that simple moving average and Holt's method give us the best results across the four approaches.

The simple moving average model can be used to estimate the demand for these products and in a similar way, models can be trained and tested for other high demand items and other stores.

5. Actionable Items

It is evident that any marketing strategy for the retailer must be driven at a store level, which is why we decided to focus on one store and develop a sample product recommendation strategy. The figure below summarises our proposed actionable strategy for Indore Malhar Mall –

Table 9: Proposed Marketing Strategy for Malhar Mall

Customer Segments	Products Purchased	# of Customers	Proposed Action
Inactive Customers		26,660	"Churned" Customers with no transaction in last 6 months; No need to invest in Product Recommendations; Reactivation Campaigns to get them back into the store
Disengaged Customers		7,621	Low Value Customers; No need to invest in Product Recommendations; Win Back Campaigns to lure them into more buying
Foundational Customers	< 5 Unique Products	1,494	Association Rules Based Notifications; Pick on the products customer bought and recommend associated products and complete the basket
	>= 5 Unique Products	9,758	Top 10 Recommender (User Based Collaborative Filtering Using Pearson Similarity)
Loyalists		514	
Grand Total		**46,047**	

As an outcome of this project we have developed the Top 10 Product Recommender for Loyalist segment of customers. This can be tested in controlled environment by the store to see the impact on Revenue over 6-8 months. As suggested earlier, if the store has an existing customer app, the recommendation system can be integrated into it to provide real time recommendations before and during the purchase journey. In absence of such a smartphone customer app, we can deploy the recommendation system offline to provide the recommendations over any channel the store uses for customer outreach (text notifications or emails). We have developed a user interface for store counter operative or customer connect manager representative to key in customer ID and see top products recommended for the customer. This can be deployed in the POS checkout counter and for planning / promotional offers.

We have also developed a non-personalized product recommender system for generalized product promotions in the store based on products being purchased together. This can be used to design promotion offers and incentivize customers to purchase on Mondays, Tuesdays Thursdays and Fridays (days with relatively less purchases).

Apart from the above, we have also developed time-series based weekly demand estimation models for Maggi noodles (Masala) and Watermelons to aid demand planning and inventory replenishment decisions.

6. Conclusions & Recommendations

In this project, we worked on POS data with a particular store and delivered the following –

1. RFM based customer segmentation

2. Top 10 product recommender that can be deployed for the high value customers of the store with the objective of revenue enhancement
3. A non-personalised recommender algorithm to create standard promotional offers
4. Time-series based SKU specific demand forecast for the store to guide inventory replenishment

The biggest challenge in using solely POS data to come up with a recommender system is that we don't have explicit user ratings / preferences available to understand the customer. Working with implicit ratings like purchase frequency or purchase flag is one of the available options we have worked with. There are other ways one may want to approach this given time and requirements – possibly using scaled purchase frequencies or weighted purchase frequencies.

We also didn't have any item categorisation available with us that constrained us from using content-based recommendation algorithms. We would highly recommend working with richer data from product and customer interaction perspective to improve the model.

On the demand forecasting front, the provided approach can be evolved further to make it 'driver based' and machine learning driven. Inclusion of more internal and exogenous data (market, macro-economic) would also help build a stronger demand forecast.

7. Glossary

Term	Description
ALS	Alternating Least Squares – Regression based predictive modelling technique
ARHR	Average Reciprocal Hit Rank This metric considers where on the Top 10 recommender the hit appears. More weightage is given to recommending relevant items on top of the list
ARIMA	Auto Regressive Integrated Moving Average – Time series predictive modelling technique
HIT RATE	Key metric used to evaluate effectiveness of Top N recommender Defined as # of users in test set for whom the recommender worked / Total # users in test set
KNN	K-nearest neighbours classification technique
LOOCV	Leave-one-out cross validation technique
PCA	Principal Component Analysis – dimensionality reduction technique
POS	Point of Sale
SGD	Stochastic Gradient Descent – Popular Optimization Technique
SVD	Singular Value Decomposition – Matrix Factorization Technique
SVD++	A variant of SVD algorithm

8. References

1. Business Analytics - The Science of Data-Driven Decision Making, *U.Dinesh Kumar (2018)*
2. Xia Ning and George Karypis , *SLIM: Sparse Linear Methods for Top-N Recommender Systems*
3. Sanjeevan Sivapalan, Alireza Sadeghian, Hossein Rahanam & Asad M. Madni, *Recommender Systems in E-Commerce*
4. G. Krishna Kishore, D. Suresh Babu , *Recommender System based on Customer Behaviour for Retail Stores*
5. Thorben Keller & Matthias Raffelsieper, Cosibon: *An e-Commerce like platform enabling bricks-and-mortar stores to use sophisticated product recommender systems*
6. Recommender Systems and Deep Learning in Python – Lazy Programmer Inc
7. Building Recommendation System with Machine Learning and AI- Sundog Education
8. http://wp.vcu.edu/univ200choice/wp-content/uploads/sites/5337/2015/01/The-Paradox-of-Choice-Barry-Schwartz.pdf
9. https://stats.stackexchange.com/questions/16456/apriori-algorithm-in-plain-english,
10. https://en.wikipedia.org/wiki/Matrix_factorization_(recommender_systems)