Final Project
Instacart Recommendation Systems

Jaini Bhansali

Tushar Goel

12/12/2017

Contents

Business Case

Recommendation Systems

Recommendation systems have impacted and redefined our lives in many ways. The internet has opened us to wide a range of possibilities making it important to provide relevant information in order to alleviate the problem of information overload. Hence, a recommender system solves this problem by searching through a large volume of dynamically generated content to provide personalized content and services.

This approach has percolated to various industries and markets. **Instacart i**s one such company that is leading the way for online grocery shopping. Instacart allows users to purchase products online from a variety of departments and deliver these at a user's door step. Following this business model requires the need of recommender systems to make a user's experience better and in turn encourage them to return buying with Instacart. This also has an additional advantage as it also lures new customers to return due to the fruitful user experience.

Instacart Recommendation Systems

Our Goal

Our aim was to create a recommendation system for Instacart. The recommendation system would help Instacart as follows:

1. Instacart has a faithful following of users. Hence, Recommender system has the ability to predict whether a particular user would prefer an item or not, based on the user's profile. This helps the seller and the consumer.
2. Instacart benefits as recommender systems reduce the transaction cost of finding and selecting items in an online shopping environment
3. Instacart Users benefit as this will improve the decision making process and quality
4. Recommender systems enhance revenue as it effectively enables selling products and hence, benefits by increasing revenues.
5. This also allows users to beyond catalog searches.

Our Approach

We have primarily used the following recommendation Systems:

1. Collaborative Filtering Technique – Model Based Filtration technique
2. Association Rule Mining – Apriori Algorithm

Stakeholders Benefited

1. **Instacart** will benefit by enhanced revenues and retain its faithful customers luring them to come back each time
2. **Instacart Customers** will be the beneficiaries of this system as they will experience a better user experience, and also allow them to narrow their search based on the wide variety of products that Instacart provides. Thus, improving the decision making experience as well.

In order to achieve our goal we used the Instacart Dataset that has been made public. The details of the dataset are present below.

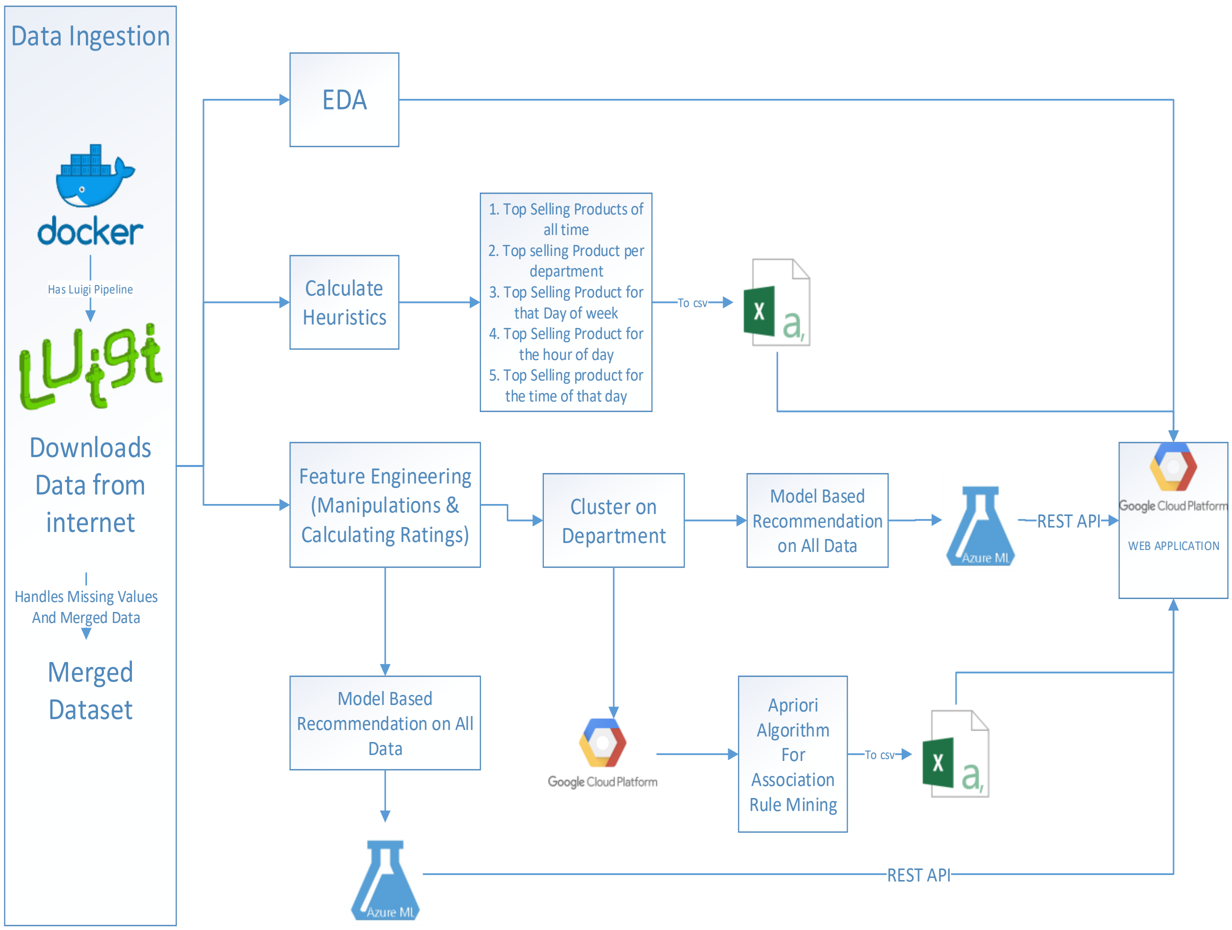
Instacart Dataset

The Instacart Dataset provided us with 3 million orders. The dataset provided information on the **product**, the **department** it belonged to and the **aisle** it belonged with. Additional characteristics also included the **time of day of purchase**, the **day of the week** of the purchase and if the product had been **reordered** by the customer before.

As our initial step we tried to understand the data and the nature of the dataset. Exploratory Data Analysis would help understand trends and patterns

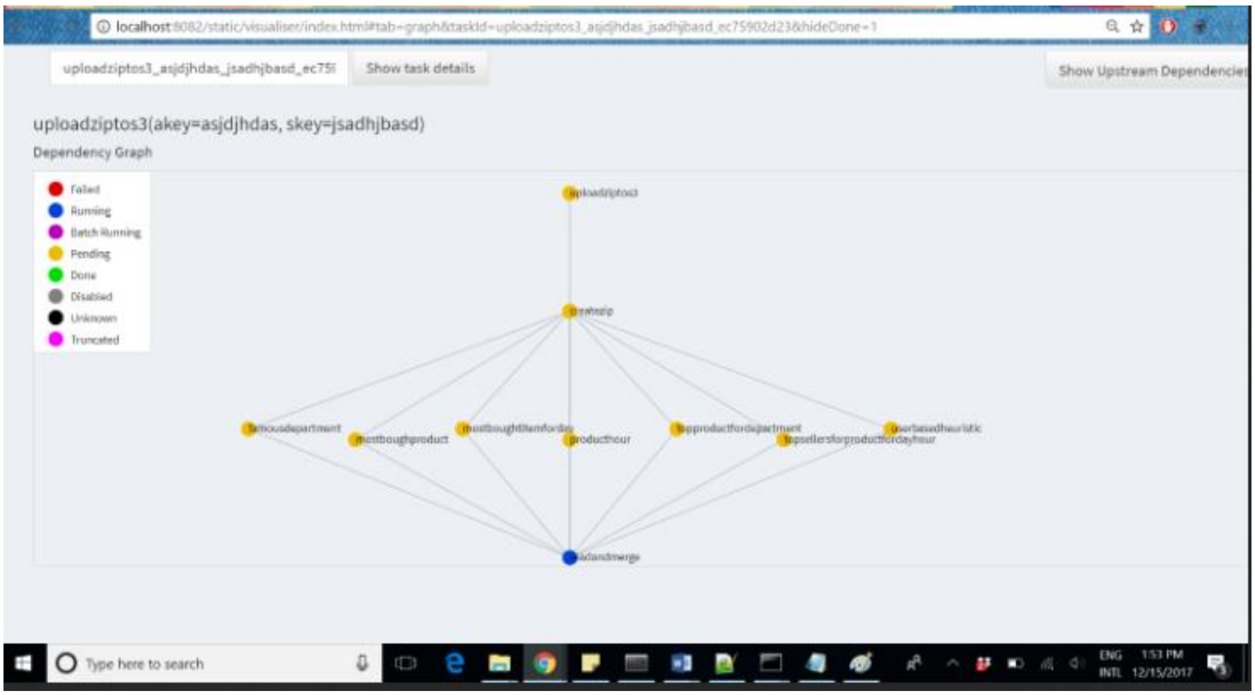
Project Work Flow

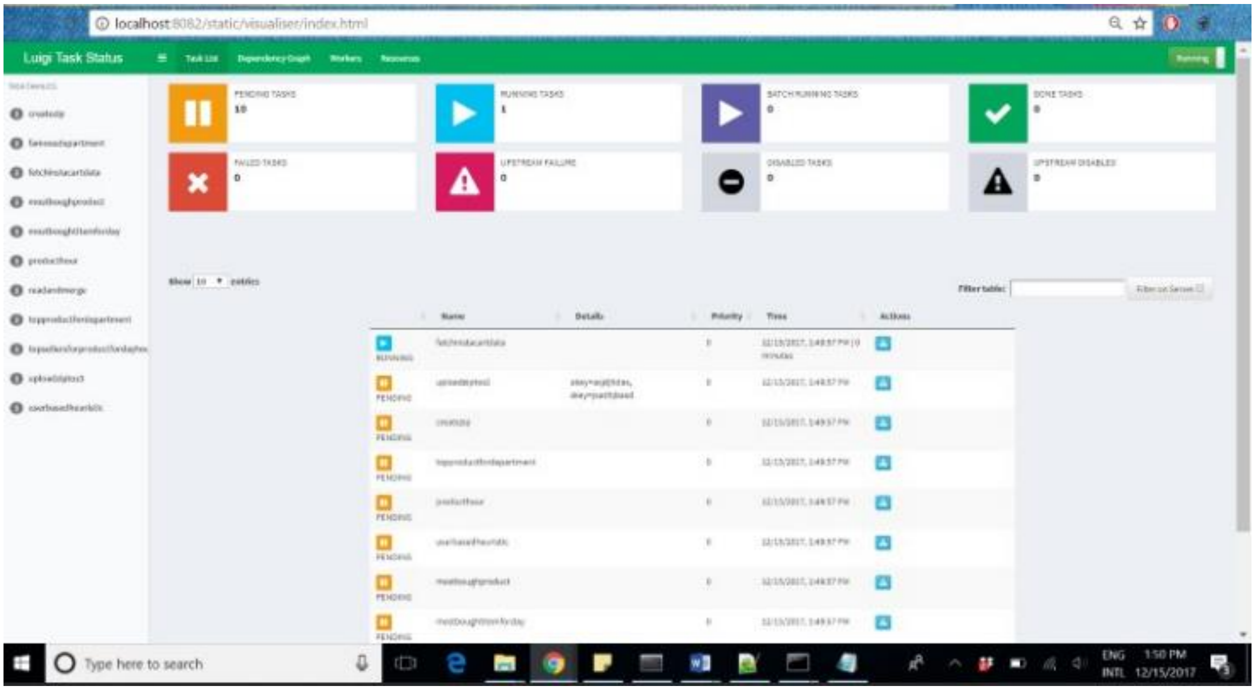
Data Ingestion was done followed by dockerizing and creating the Luigi Data Pipeline. The data is used for Exploratory Data Analysis, Heuristics Calculation, Feature Engineering and Creation of Recommendation Systems . The result set of the models is placed in the CSV's as due to volume of data and then used on the WebApp hosted on the Google Cloud Platform



Data Ingestion

We merged the Instacart dataset and handled missing values in the Prior Orders csv's. Next, we dockerized the image and created a Luigi Pipiline for the same.




localhost:8082/static/visualiser/index.html
Luigi Task Status
Type here to search

Exploratory Data Analysis

In order to understand the data we first tried to understand the inventory, the users and the patterns, if any.

Instacart Inventory Details

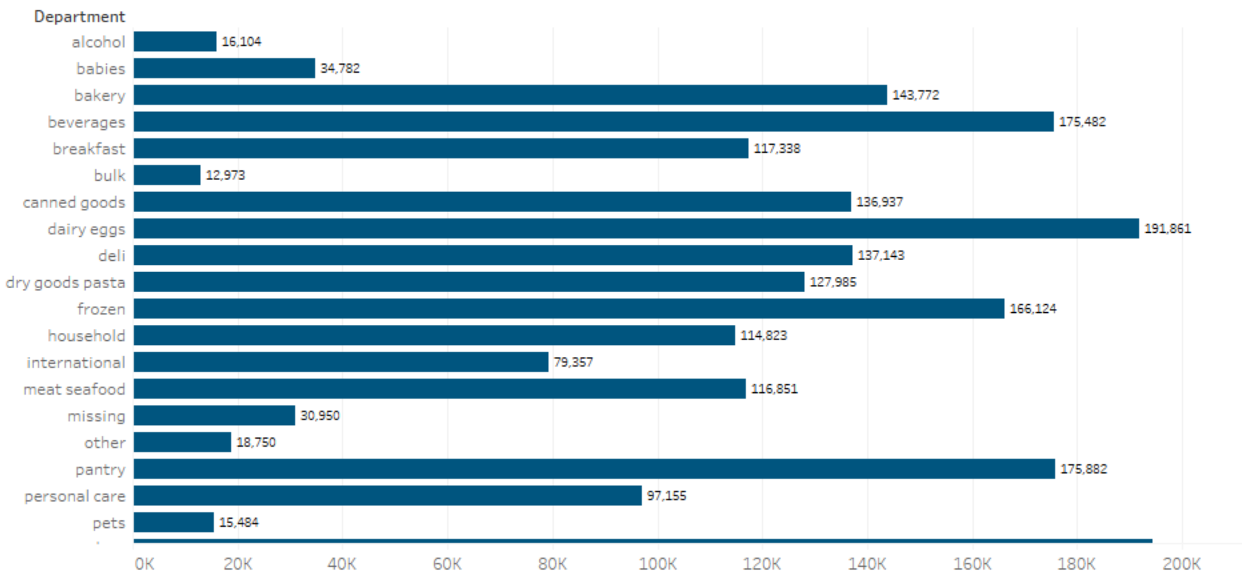
Instacart has a total of 49689 products, 21 Departments and 134 Aisles.

Product Count by Department and Aisle

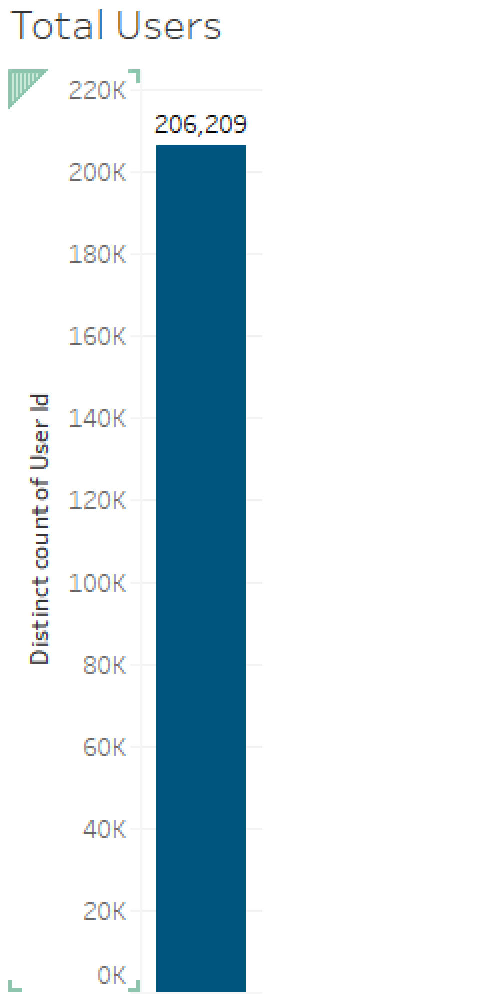
Department	Aisle	Product Count
pets	cat food care	499
	dog food care	473
	Total	972
produce	packaged vegetables fruits	615
	fresh vegetables	569
	fresh fruits	382
	fresh herbs	86
	packaged produce	32
	Total	1,684
snacks	candy chocolate	1,246
	chips pretzels	989
	cookies cakes	874
	energy granola bars	832
	crackers	747
	nuts seeds dried fruit	582
	fruit vegetable snacks	356
	popcorn jerky	316
	mint gum	168
	ice cream toppings	85
	trail mix snack mix	69
	Total	6,264
Grand Total		49,689

Total Instacart Users

Count of Users by Department



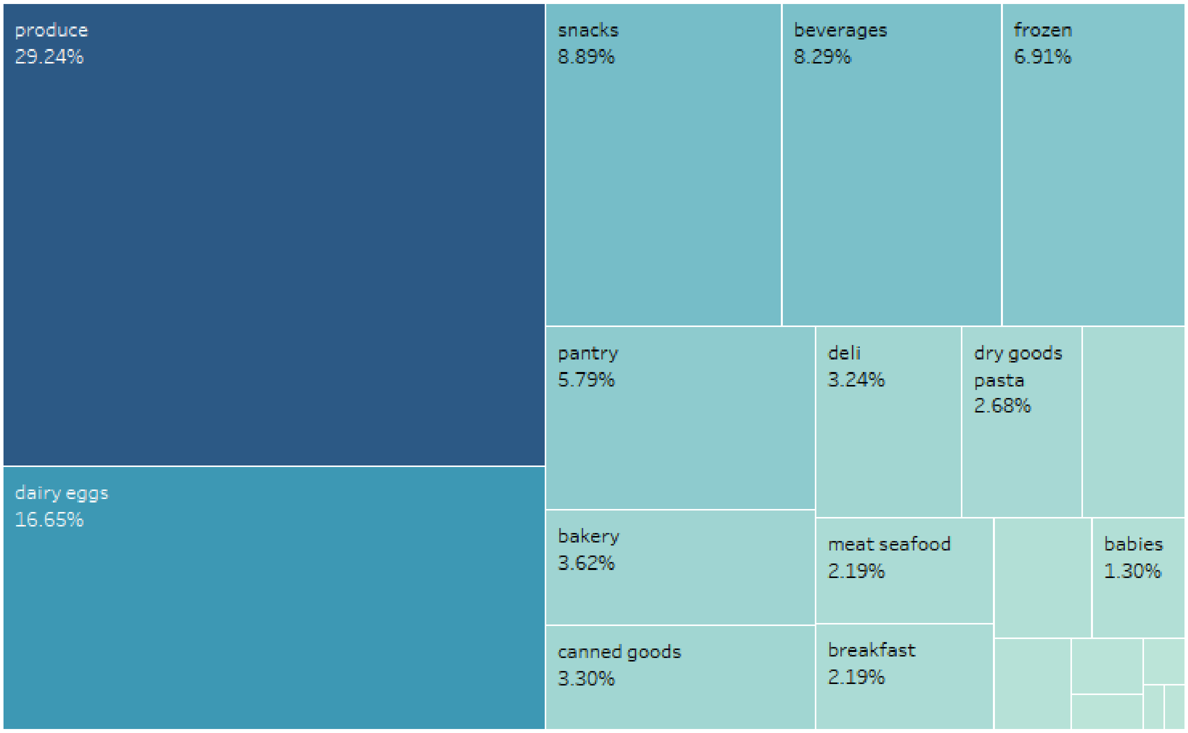
Total Count of Instacart Customers



Sales Distribution by Department

Next, observe the sales distribution by Department

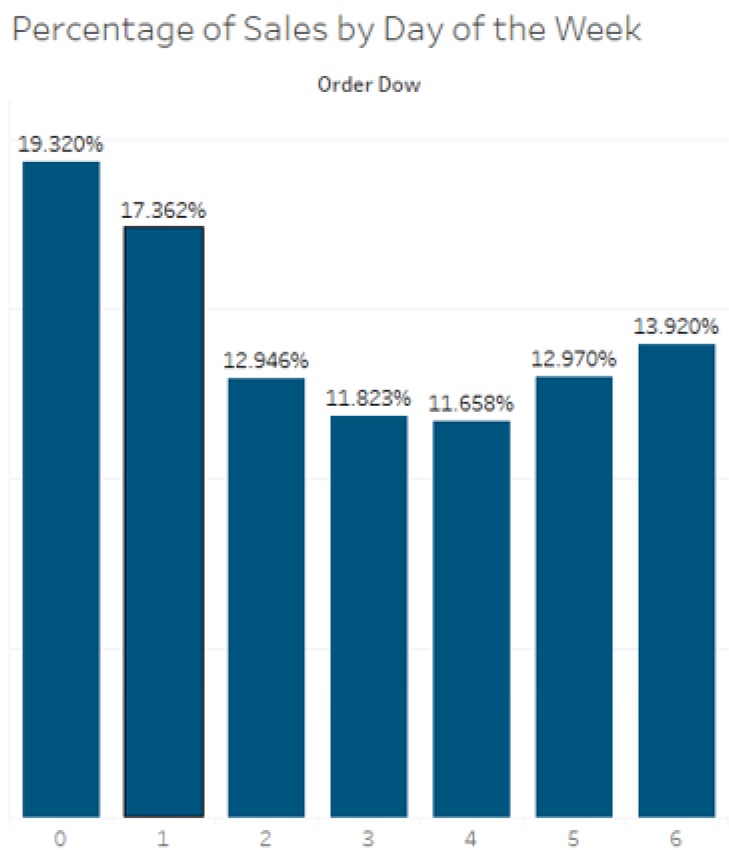
Departments Contributing to Most Sales



Analysis: We observe that Produce, Dairy Eggs and Snacks are the maximum Departments that contribute to sales

Sales by Day of the Week

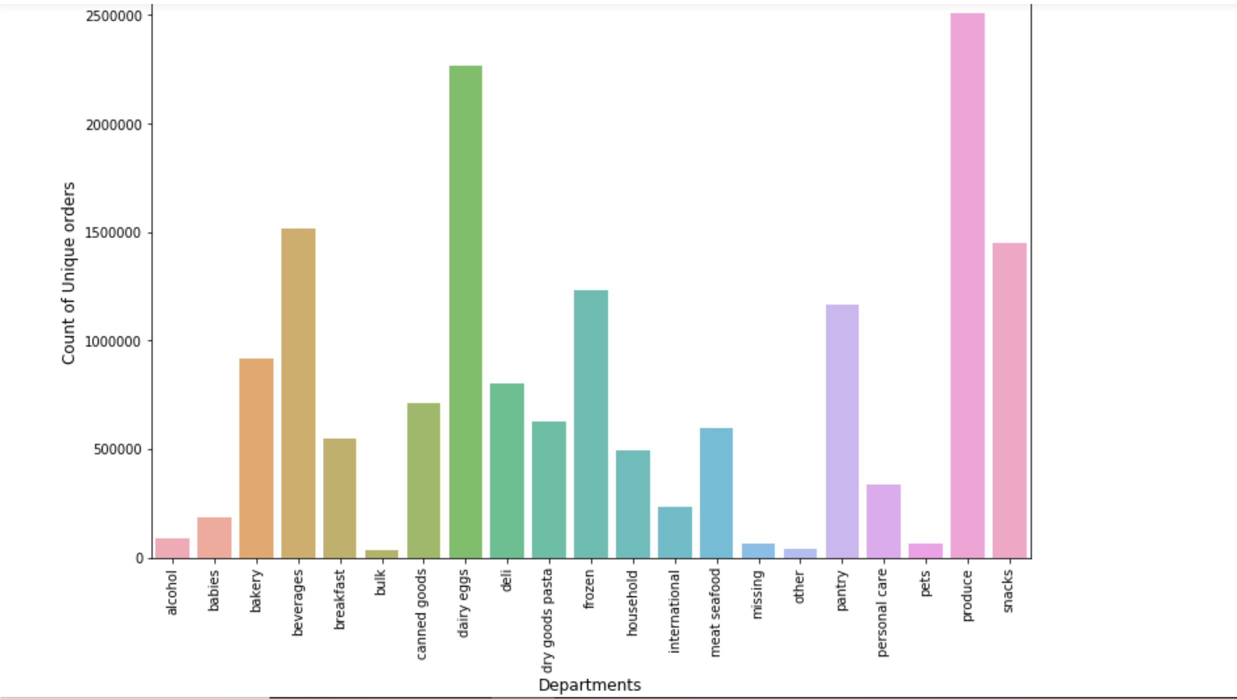
Next, we observed a pattern in the sales by Day of the Week.



Analysis : More than 35% of orders are placed in just two days. This summary metric can help Instacart on their restocking capability. This also helps us study a buying patter, by day.

Orders placed by Department

Next, we observed if there was a pattern of orders dedicated to a particular department.



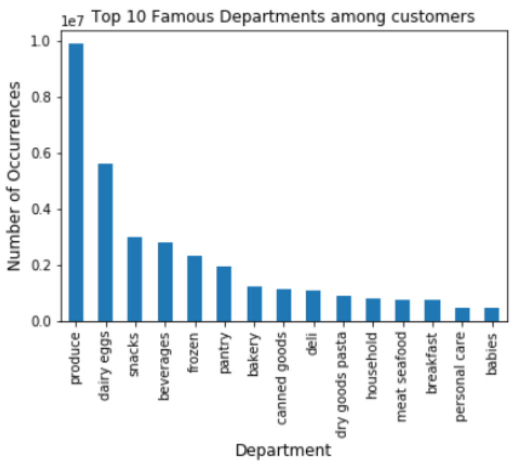
Analysis : We can clearly see that some departments have really infrequent sales. This helps us to decide the clustering as well as eliminate some of the departments from recommending

Top 10 Popular Departments

Next, we observed the most popular departments

Top 10 Famous Departments among customers

In [66]:
```
top_product_cnt=15
customerscount= recommendproddep.groupby("department")["user_id"].aggregate("count").sort_values(ascending
=False)[:top_product_cnt]
customerscount.plot(kind='bar')
plt.ylabel('Number of Occurrences', fontsize=12)
plt.xlabel('Department', fontsize=12)
plt.xticks(rotation='vertical')
plt.title("Top 10 Famous Departments among customers")
plt.show()
```



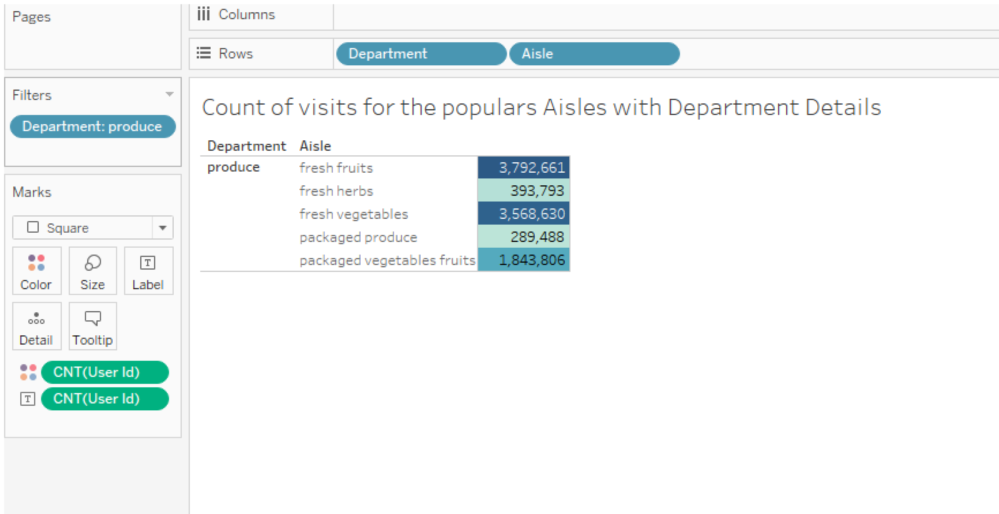
Top 3 Famous Departments among customers are:

- Produce
- dairy eggs
- snacks

This indicates that Instacart is most popular for its Produce Department which deals in fresh vegetables and fruits

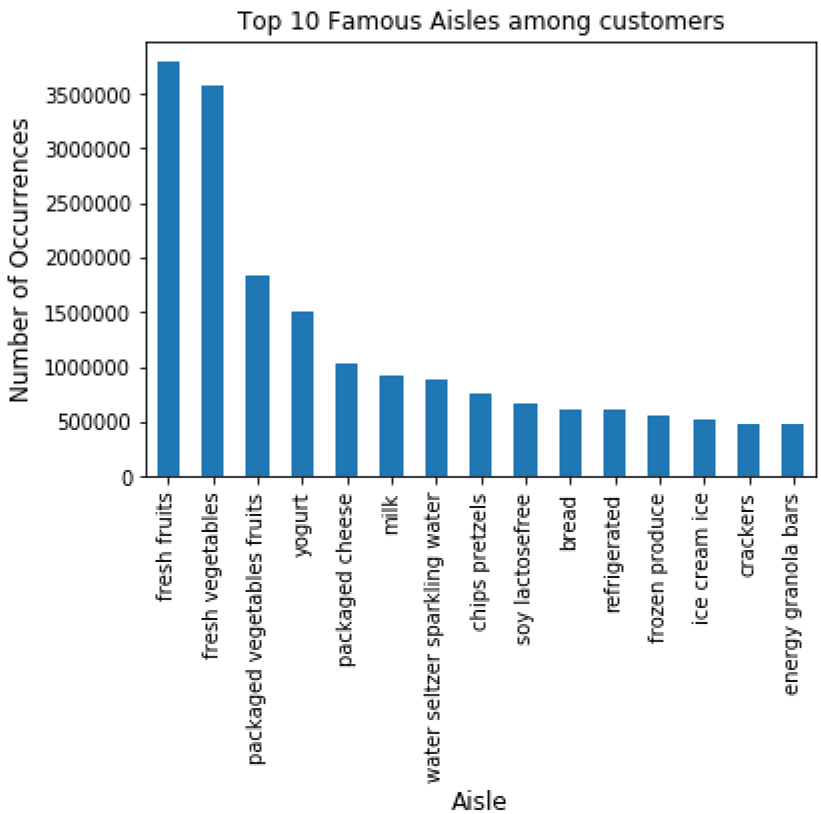
Most visited aisles among the popular departments

Aisle Details for the most popular Departments



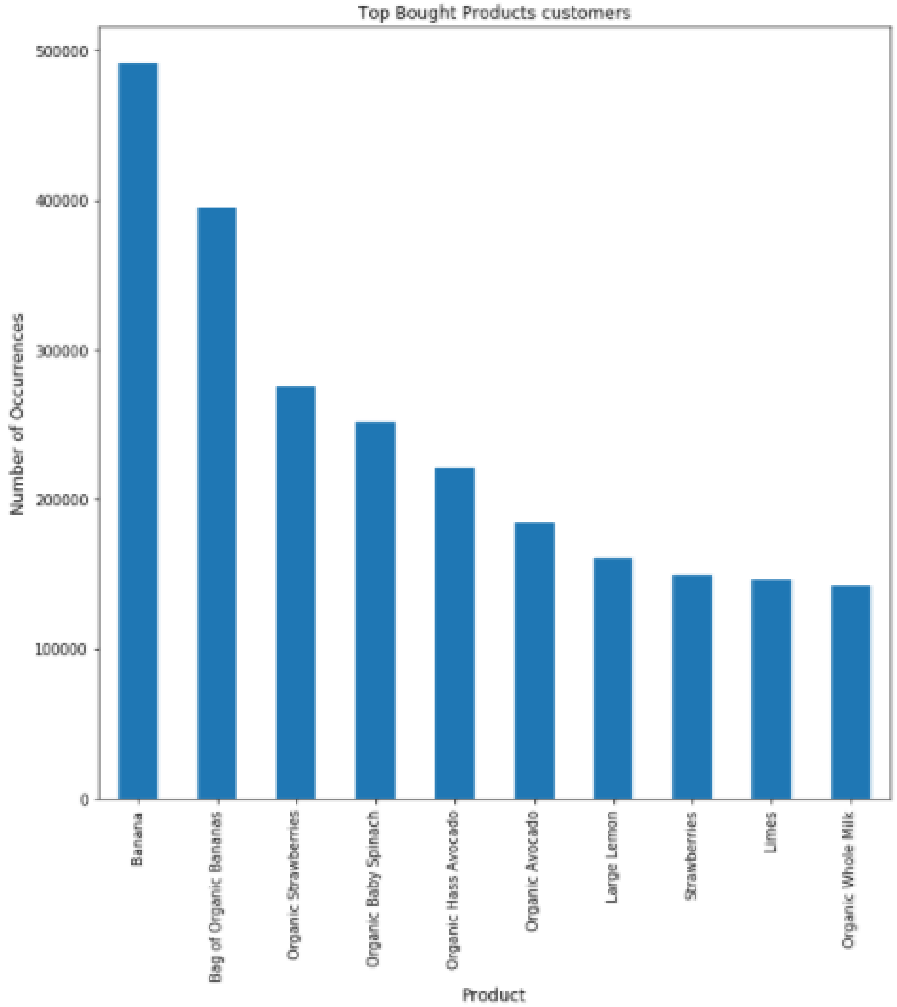
From this visualization it can be analyzed that the most popular Aisles are Fresh Fruits and Fresh Herbs in the Produce Department

Top 10 famous Aisles among Customers



10 Popular Products among Users

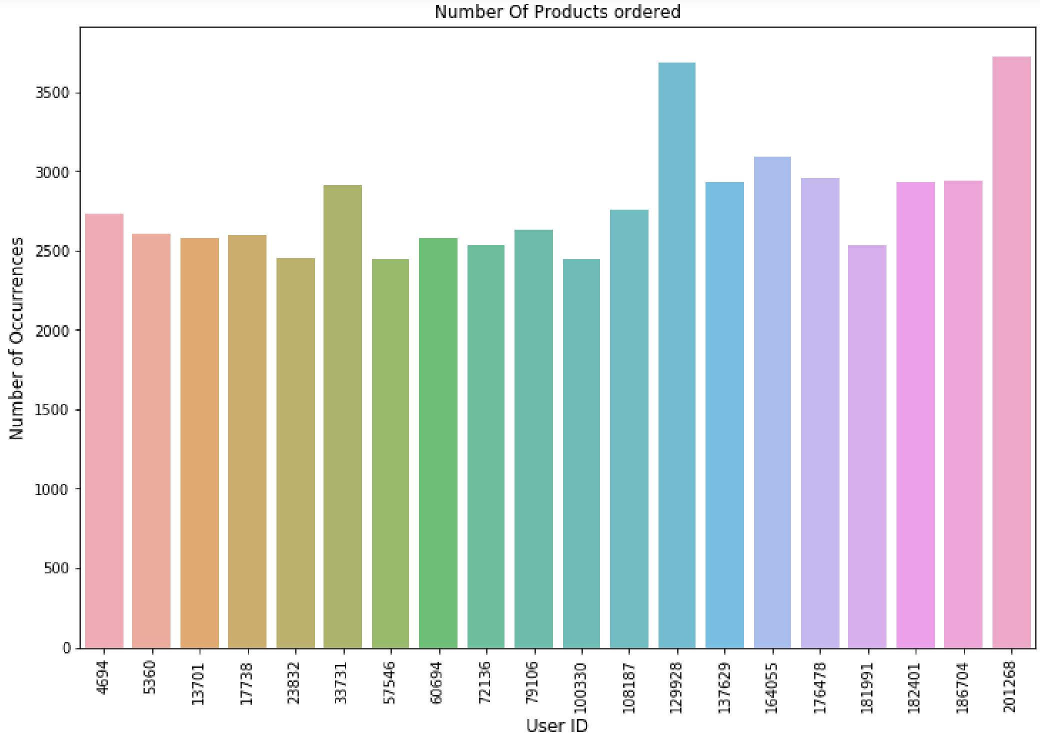
From the above visualization we observe that most popular aisle also belongs to the most popular Department – Produce



Among the top 10 products most popular products Banana, Bag of Organic Bananas Organic Strawberries. Additionally, they also belong to the most popular department and most popular aisle

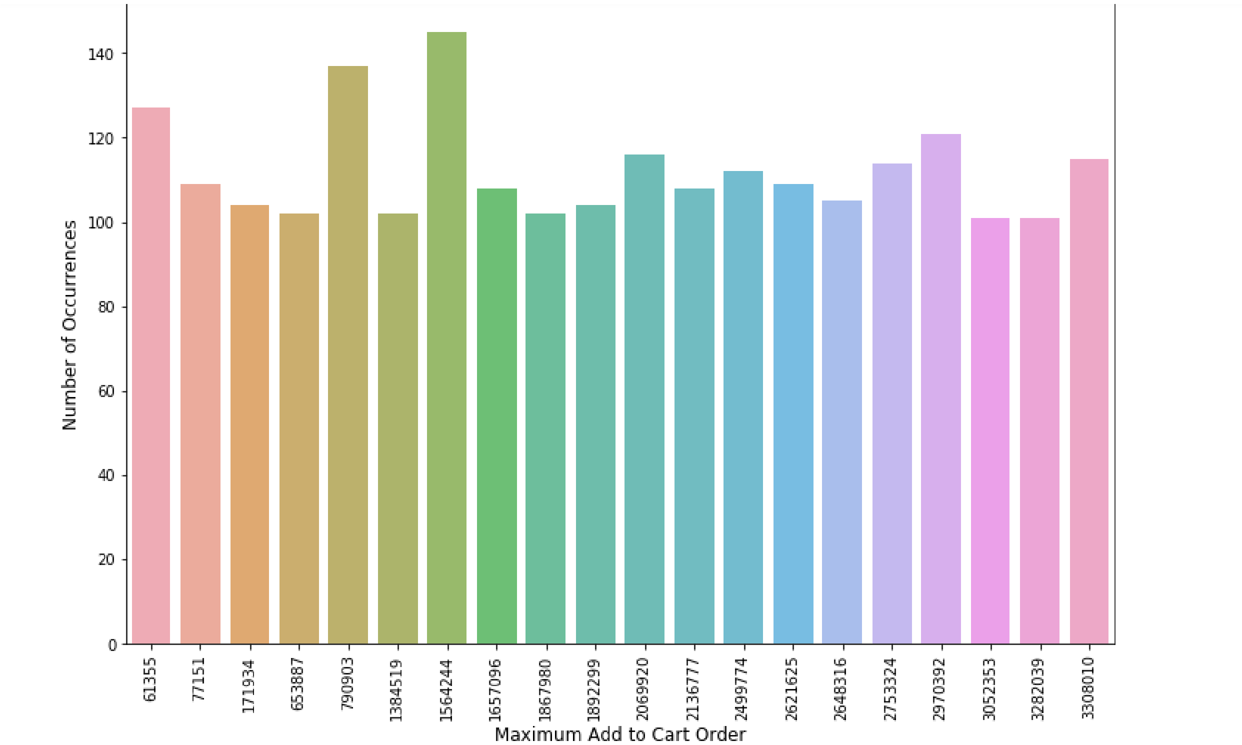
Users with the most Orders

We observed the users with the most orders



Result: We observe that userid 201268 has purchased the most number products, followed by userid 12998

Top 20 users with most Orders

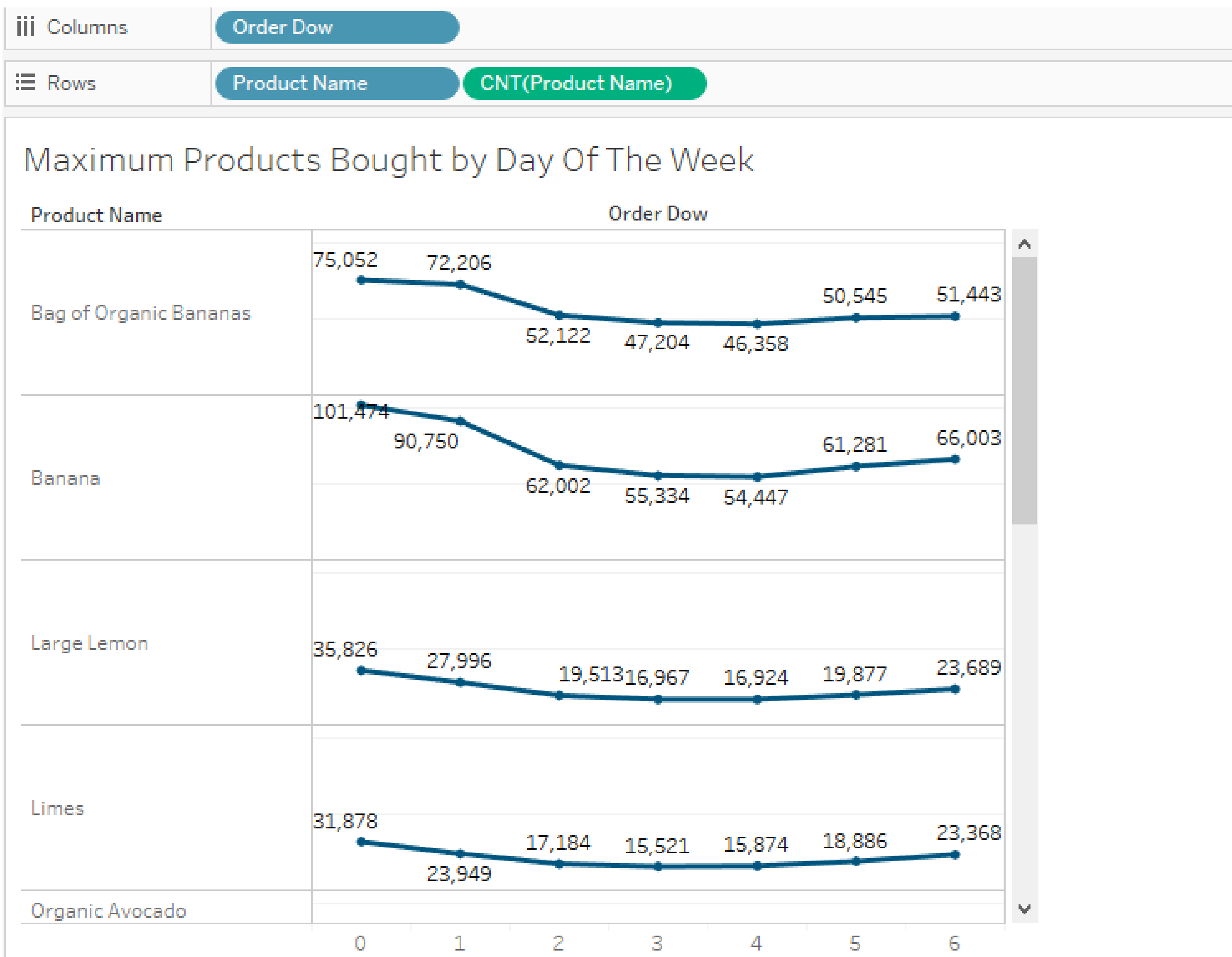


Maximum products added to cart orders are:

- Order Number 1564244
- Order Number 790903

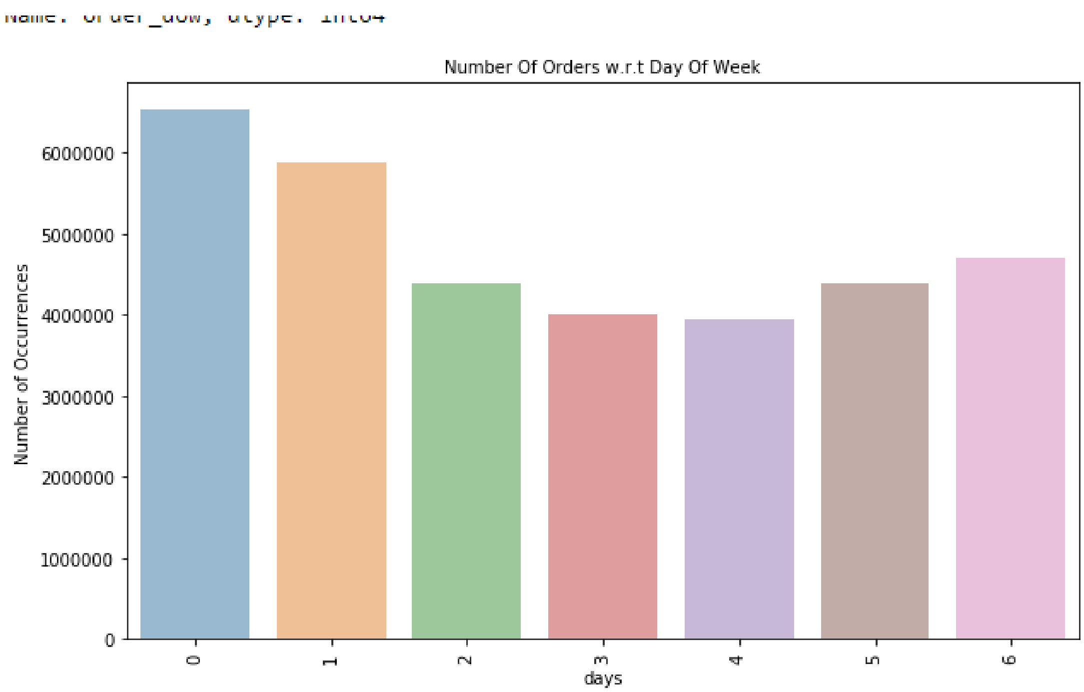
This also indicates the maximum products were bought in the above order id's specified

Trend of Top Products Bought by Day of the Week



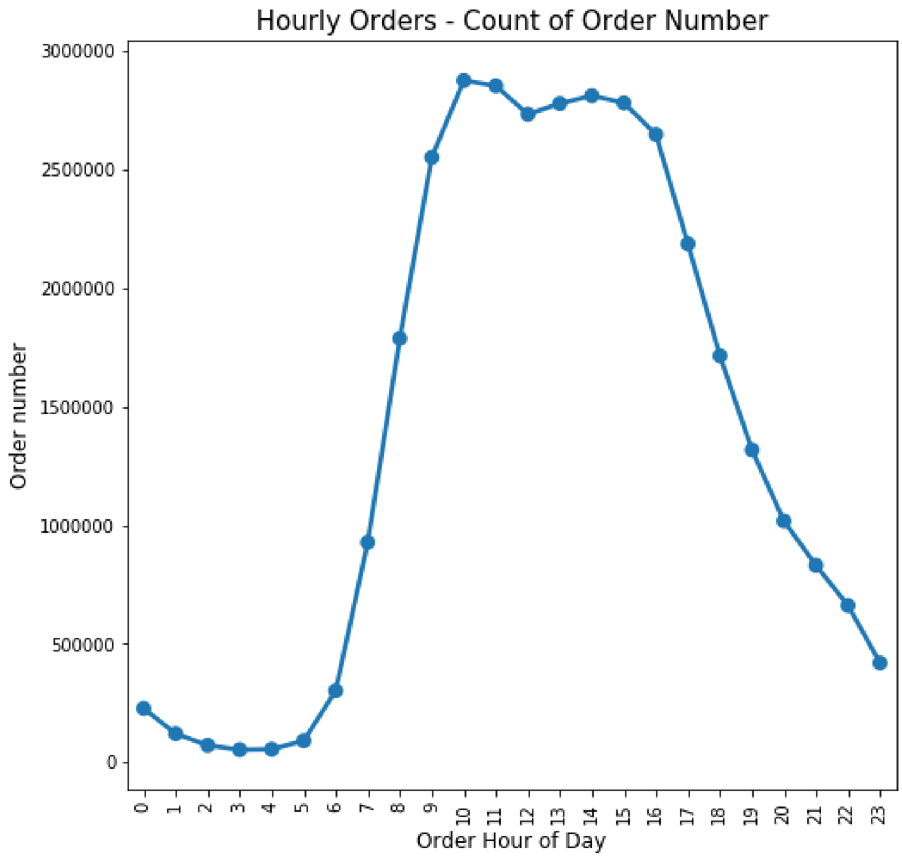
Most of the popular products are bought on Sunday. The Trend depicts they are bought the most from Sunday to Wednesday with a slight dip in purchase for the rest of the week

Count of Products Ordered on Day Of Week



Majority of the orders are placed on Day 0 and Day 1 , that is Saturday and Sunday and the least number of orders were placed on Wednesdays.

Trend of Orders by the Hour



Maximum Ratio of Weekly Orders - Count of Order Number is between :

- 11:00 am and 17:00 pm hours of the day

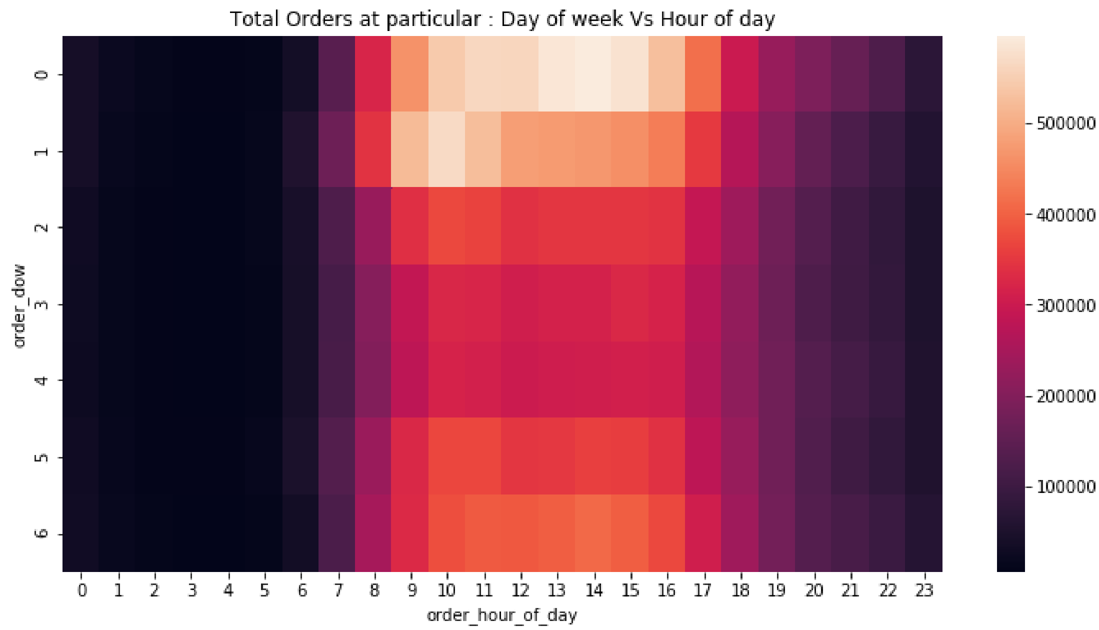
Most products bought in an order by day of the week by a particular user

Purchases with maximum Products by Users and by Day of Week

	user_id	order_dow	order_id
0	176478	1	2424
1	17738	1	2076
2	134433	5	1785
3	64719	1	1735
4	160106	0	1718
5	97899	1	1688
6	129928	1	1643
7	5360	1	1553
8	39901	5	1523
9	88996	5	1512
10	74798	4	1384
11	30910	5	1362
12	169647	4	1334
13	199743	5	1298
14	150186	5	1272
15	4694	2	1268
16	45520	1	1264
17	102282	6	1250
18	181991	1	1232
19	290	6	1212

Most products were bought in orders purchased on Sunday. This table might also suggest that users purchase more on a particular day of the week

Frequency of orders by Day of week and Hour of day



We can see very few orders are place before 8 am and most of the orders are done on Saturday and Sunday from 9 am to 5 pm

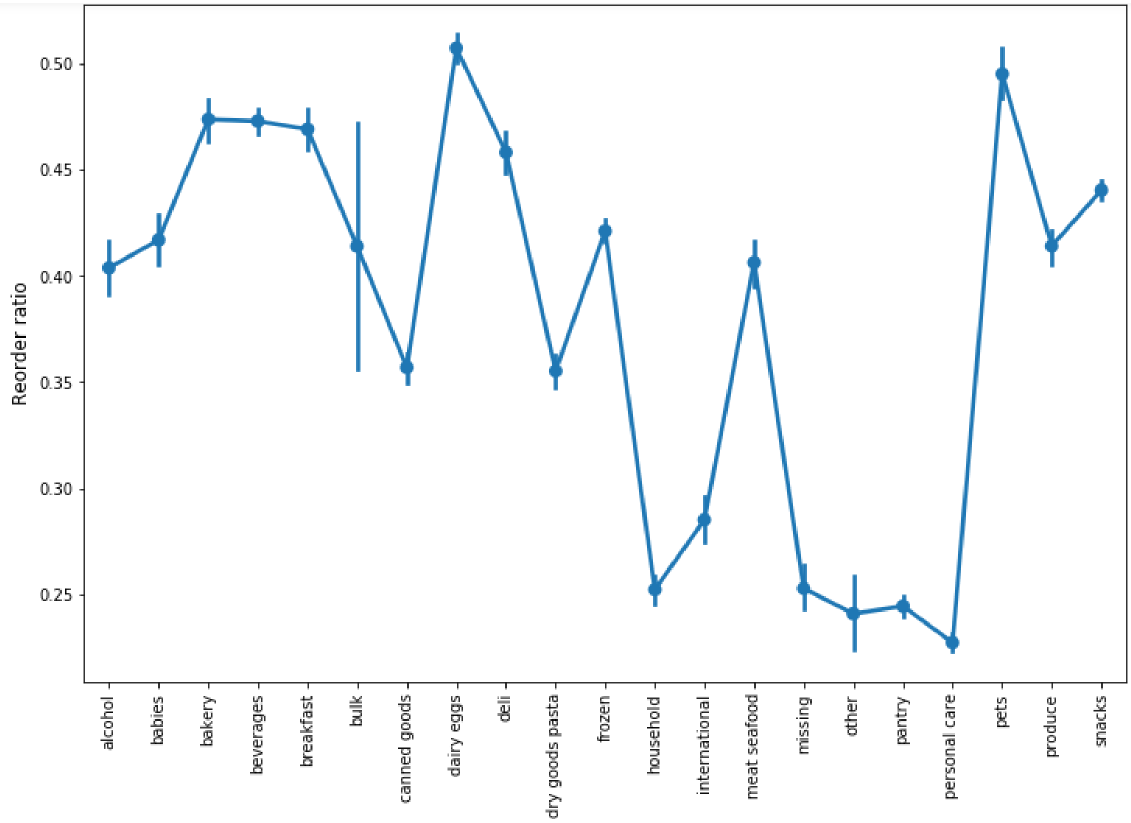
- Saturday Evenings

- Sunday Mornings are the prime time for the customer orders

This helps us understand that customers predominantly purchase on the weekends

Number of Reorders

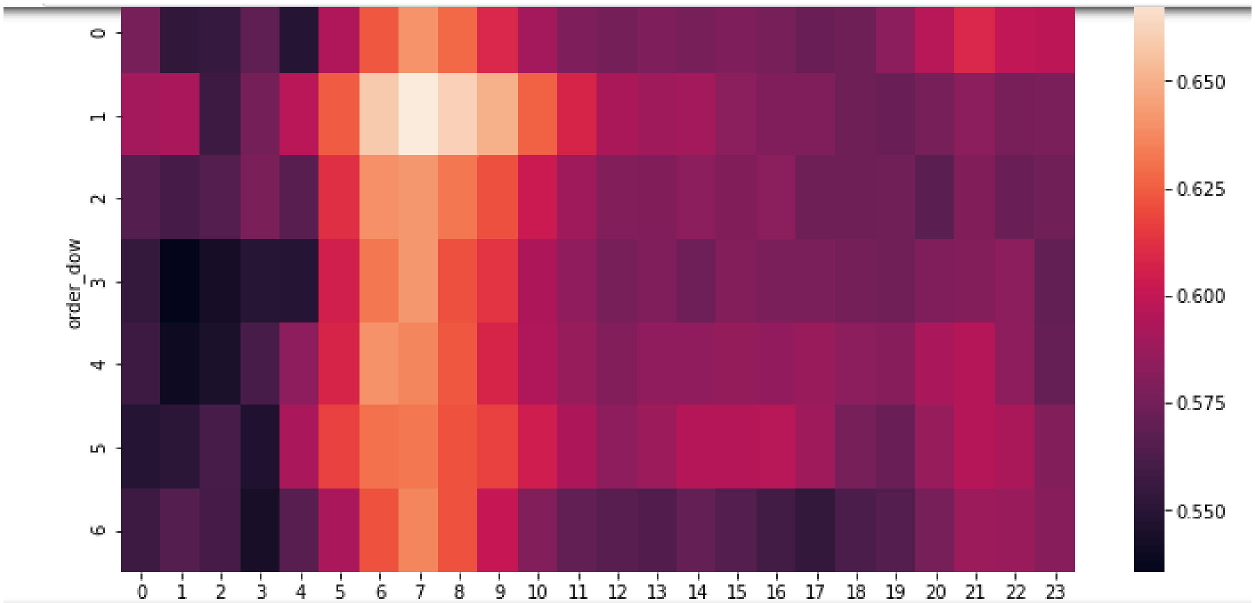
Next, we observed Number of Times any Product is Re-ordered



Using this we can use this data on department clustered recommendation algorithm and make amount of predictions accordingly. As per the above result the probability of Reorders is the highest for the Department Produce

Reorder ratio by Order of Day and Order Of Week

We, studied the reorder ratio of order of the day vs Order of the week



This indicates that maximum reorders have occured 6 am to 8 am on all days with the maximum being on Sunday

Orders_product_prior with reordered items

62 % of the items are reordered in the combined set

19955360 values are reordered in combined dataset

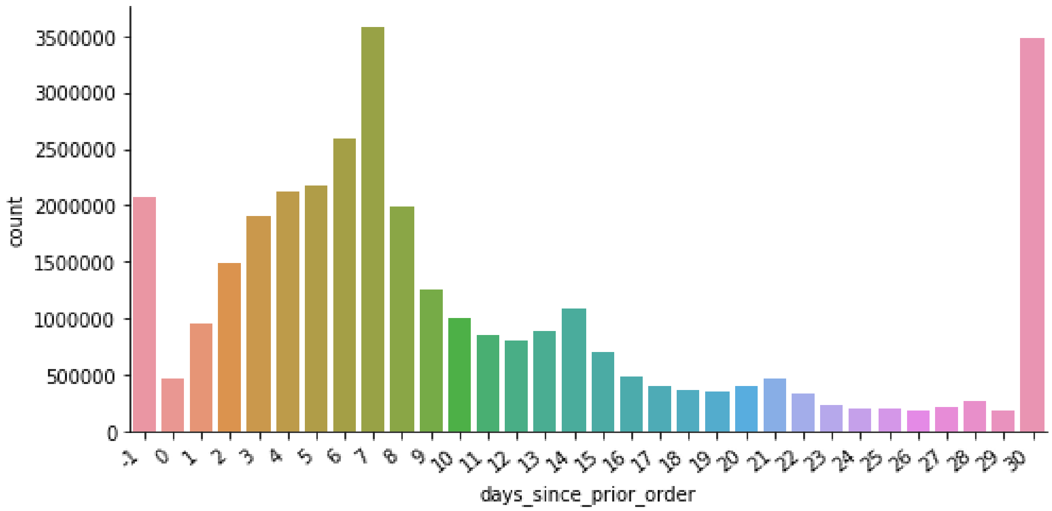
Products with the most Reorders

```
                              product_name  reorder_sum  reorder_total  \
8537                     Chocolate Love Bar           94            102
13844             Energy Shot, Grape Flavor           20             22
28415                   Orange Energy Shots           12             13
37379                   Raw Veggie Wrappers           65             69
38793   Russian River Valley Reserve Pinot Noir       27             30
39870  Serenity Ultimate Extrema Overnight Pads       84             90
40458          Simply Sleep Nighttime Sleep Aid       41             45
41723              Soy Powder Infant Formula           32             35

       Probability of being Reordered
8537                         0.921569
13844                        0.909091
28415                        0.923077
37379                        0.942029
38793                        0.900000
39870                        0.933333
40458                        0.911111
41723                        0.914286
```

The Chocolate Love Bar ,Energy Shot , Grape flavor are the items with the most reorder

Count of Days since Prior Order



We can see that customers order maximum once in every week or every month, after a sudden fall in customers order after 7 days, at 14 and 21 days gain there is a little small peak in comparison to rest of the other days.

EDA Summarization

Analysis of EDA provided us insightful evidence on the users buying pattern. It explicitly outlines that the majority of the users prefer buying on weekends. Moreover, it is observed that most orders are dedicated to a particular department. This implies that a user always returns to the department that he purchases from the department type. Additionally a user purchases almost 20% of the products from one department in a given order.

As a result, this enables us to cluster our data for recommendations and associations based on a department as well

Percentage Sales from a Department in a Transaction

Order Id	Department Id	
1	4	50.00%
	15	12.50%
	16	37.50%
	Total	100.00%
2	4	33.33%
	13	55.56%
	16	11.11%
	Total	100.00%
3	3	12.50%
	4	37.50%
	12	12.50%
	16	37.50%
	Total	100.00%
4	3	7.69%
	7	23.08%
	11	7.69%
	14	30.77%
	19	30.77%
	Total	100.00%
5	4	26.92%
	6	3.85%
	7	3.85%

Recommendation Systems Implementation for Instacart

Recommendation Systems

Business Logic for recommendations

Rating was calculated on the basis of the following:

Rating= count(Products Bought by a user))/(Count of product bought by the user most number times)*100

```
#getting only product and user data for recommendation
userandprod=mergeforprodanduser[['product_id','user_id']].copy()
recommend=pd.DataFrame(userandprod['product_id'].groupby(userandprod['user_id']).value_counts())
recommend['userprod']=recommend.index
recommend.columns=['count','userprod']
recommend.reset_index(inplace=True)
recommend.drop('userprod',axis=1,inplace=True)
recommend.to_csv('recommend.csv',index=False)
```

Converting all user data with orders and products preferences into ratings

```
convtorat=pd.DataFrame(recommend.groupby(['user_id'])['count'].max())
convtorat.rename(columns={'user_id':'max_count'},inplace=True)
convtorat.reset_index(inplace=True)
convtorat.rename(columns={'count':'max_count'},inplace=True)
normratings=recommend.merge(convtorat,how='inner',on='user_id')
normratings['rating']=normratings['count']/normratings['max_count']*100
normratings['user_id']=normratings['user_id'].astype('int32')
normratings['product_id']=normratings['product_id'].astype('int32')
normratings['rating']=normratings['rating'].astype('float32')
normratings.drop('max_count',axis=1,inplace=True)
normratings.drop('count',axis=1,inplace=True)
normratings['rating']=normratings['rating']*5/100
```

```
normratings.to_csv('recommendwithratings.csv',index=False)
```

Scope

The scope of project includes Model Based Collaborative Filtering which fetched the best results. Instacart has provided a rich dataset that allows us to study past behaviors to provide recommendations to users. We have implemented model based collaborative filtering using the **Graphlab Library Framework** because it utilizes Spark Clusters on IBM-DSX providing high computation power. We tested our dataset on 3 models:

1. Item Based Recommendation Models
2. Popularity Based Recommendation Model
3. Model Based Recommendation Models (Latent Models): Latent models are used to shrink user traits into latent features. This is based on the matrix factorization model and is widely used for recommendation systems where it can better deal with scalability and sparsity. The goal of model based collaborative filtering is learn the latent preference of users and the latent attributes of items from known ratings.

Results of the recommendation model showing the best precision and recall

Model Based Recommendation provided the best results after trying all the other models. The Dataset was used without any clustering to run Item Based, Popularity Based and Model Based recommendations to compare and select the best results.

1. **All Data**

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	0.0115626964446	0.000913421475807
2	0.0092313479748	0.00148233276347
3	0.00784375840432	0.00191872753452
4	0.00689876205891	0.00230404032193
5	0.00629513584683	0.00264680073719
6	0.00579784750374	0.00295004988483
7	0.00543910598408	0.00326441402051
8	0.00510694009276	0.00352964737372
9	0.00484253910742	0.00379207555727
10	0.0046156739874	0.00403881432145

[10 rows x 3 columns]

2. **Clustering on the basis of aisle** : Precision and Recall

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	0.132323634914	0.0753840081397
2	0.105805463966	0.121300706457
3	0.0896409290192	0.152604569126
4	0.0791038624588	0.178984412787
5	0.0720171020689	0.20359594769
6	0.06609086078	0.224188338861
7	0.0607514247411	0.239555259192
8	0.0569438657198	0.256470186652
9	0.0534308811511	0.270199281195
10	0.0507237581072	0.285379402861

[10 rows x 3 columns]

3. **Clustering on basis of particular Hour of the Day** :

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	0.00266078493155	0.000472287732029
2	0.00171550607429	0.000620204230644
3	0.00128371202838	0.000708631346169
4	0.00108532016945	0.000768013762972
5	0.000917270594825	0.000814791451279
6	0.00081107259975	0.000871474839545
7	0.000755222790723	0.000972442124895
8	0.000704582851941	0.00109579795931
9	0.00064574605064	0.00112978715279
10	0.000612680740819	0.00119594555841

[10 rows x 3 columns]

4. Clustering on the basis of Department

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	0.110030564046	0.0315781469557
2	0.0837516594116	0.0470210018995
3	0.0719397363465	0.0603287511205
4	0.0646275199901	0.0718478555194
5	0.0594955388843	0.0822278476143
6	0.0554207444455	0.0913977820511
7	0.0521908643053	0.0999577482844
8	0.0494435182612	0.10761302708
9	0.0470494204494	0.114464644203
10	0.0448408508536	0.120621192731

[10 rows x 3 columns]

5. Clustering on the basis DOW Recommendation

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	2.23249168397e-05	2.35088139448e-06
2	1.86040973664e-05	6.69183744664e-06
3	2.48054631552e-05	1.73581866034e-05
4	2.6045736313e-05	2.08206158355e-05
5	2.97665557862e-05	3.24880426124e-05
6	2.72860094707e-05	4.17900912956e-05
7	6.48485679629e-05	0.000114924299757
8	0.000597191525462	0.00107599000251
9	0.00290637343302	0.00504833966387
10	0.00585582568705	0.0112543153665

[10 rows x 3 columns]

6. **Clustering on the basis of hour of the product was bought :**

Precision and recall summary statistics by cutoff

cutoff	mean_precision	mean_recall
1	0.0108494585663	0.00243820258362
2	0.00785650447904	0.00360404695889
3	0.00654362299864	0.00459962148588
4	0.00570531872883	0.00533079189612
5	0.00517947332322	0.00607772321917
6	0.00475616430764	0.00670381618461
7	0.0044404393229	0.00731151345478
8	0.00416857190364	0.00786999940604
9	0.00398252262613	0.00849827116364
10	0.00382121256209	0.00912320325612

[10 rows x 3 columns]

Importance of heuristics

Heuristics have been used to provide basic recommendation when a new user visits in Instacart.

List of Heuristics Used:

1. Most Bought Product of all time
2. Most visited department
3. Most frequently bought products by the specific user
4. Most frequently bought products on that hour
5. Most frequently bought product on that day
6. Most frequently bought product on that hour of that day
7. Most Reordered product by that user

Model Deployment

The recommendation models are deployed in Azure using Train Matchbox Recommender

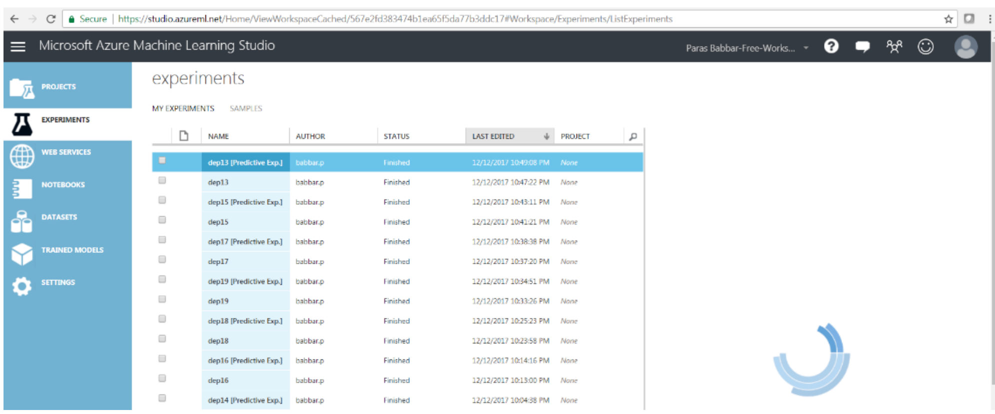
Train Matchbox Recommender

The Train Matchbox recommend uses a popular ranking matrix called NDGC (Normalized Discounted Cumulative Gain). It allows a relevant score in the form of real numbers. If there is no feedback for an item then the gain is set to 0. If not then the recommender returns some items which has a relevance score usually a non-negative number which is being added up known as Cumulative Gain. The matchbox recommender combines collaborative filtering with a content based approach. When a user is relatively new to the system, predictions are improved by making use of the future information of the user. Even, if the user item features are not available matchbox will still work in its collaborative filtering mode.

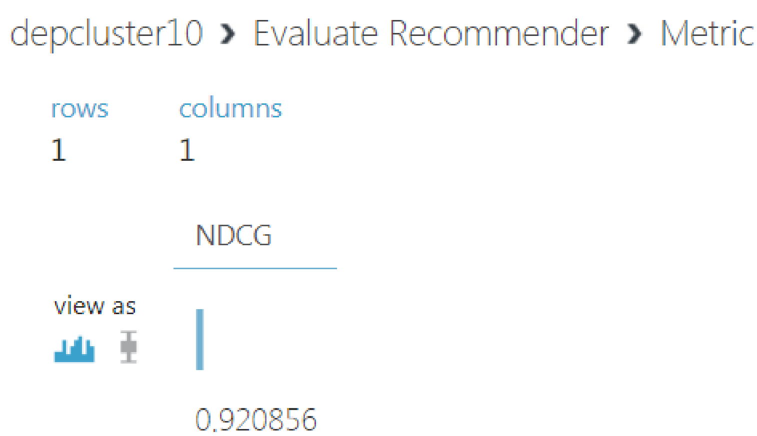
Azure Deployment Screen Shots

Model Deployment on Full Data

Model Deployment on Department Cluster



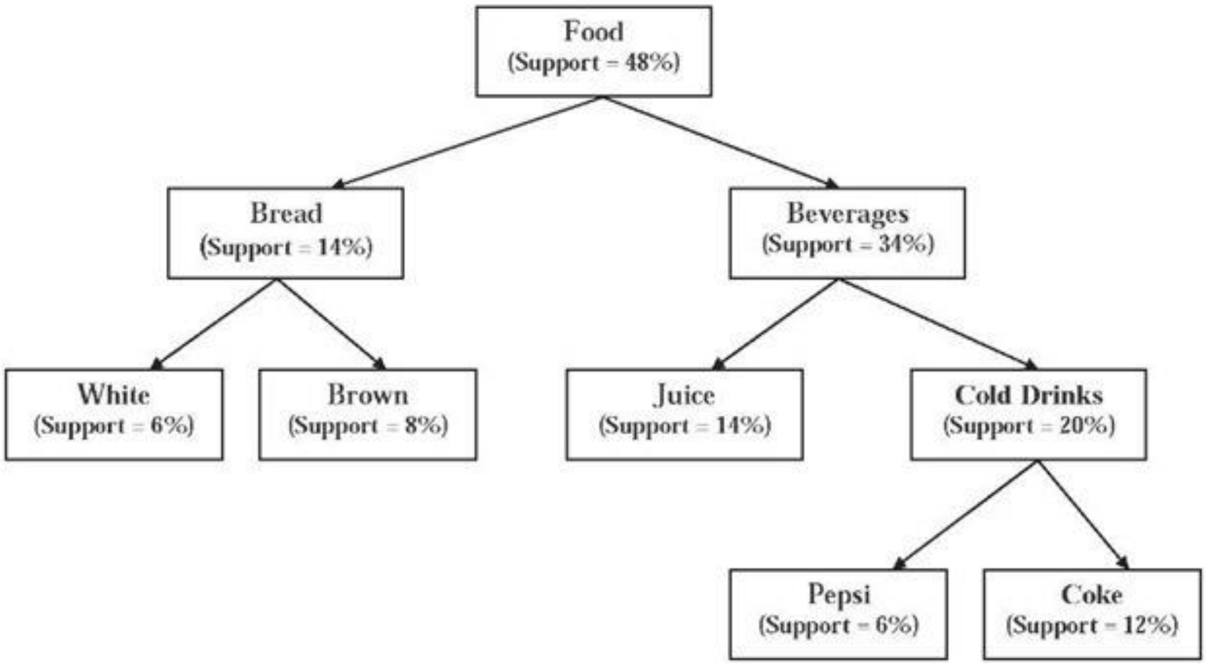

depcluster7 › Evaluate Recommender › Metric
rows
1
columns
1
NDCG
view as
0.919628


depcluster10 › Evaluate Recommender › Metric
rows
1
columns
1
NDCG
view as
0.920856

Apriori Algorithm

Significance in Instacart Recommendation Systems

Apriori Algorithm is influential algorithm for mining frequent item sets. It uses a bottom up approach where frequent subsets are extended one item at a time. The diagrammatic representation indicates the working of Apriori Algorithm



It then makes use of Association Rule Mining to get the association rules in the data.

Instacart Recommendations benefit from these models as it provides the user options on various items based on the associated items selected for purchase. This improves the user experience of a customer visiting the Instacart webpage.

Apriori Algorithm Implementation Details

In order to find associations among the various products in the dataset, we clustered the data by departments. We manually clustered by departments due to the following reasons:

1. A Transaction made by a user consists of products majorly associated with one department
2. A returning user generally visits the department that he visited in the previous orders

Manual Clustering by Departments

Manual Clustering is done by Departments

```
In [ ]: temp_df=combine_train_prior.loc[:,['order_id','product_name']]
        grouped_data=temp_df.groupby("order_id").filter(lambda x: len(x) >=4)
```

```
In [ ]: dept_produce=combine_train_prior[combine_train_prior['department'] =="produce"].reset_index()
        dept_produce.to_csv(DepartmentProduceCluster.csv,sep=',')
```

```
In [ ]: dept_dairy=combine_train_prior[combine_train_prior['department'] =="dairy eggs"].reset_index()
        dept_dairy.to_csv(DepartmentDairyCluster.csv,sep=',')
```

```
In [ ]: dept_dairy=combine_train_prior[combine_train_prior['department'] =="snacks"].reset_index()
        dept_dairy.to_csv(DepartmentSnacksCluster.csv,sep=',')
```

```
In [ ]: dept_dairy=combine_train_prior[combine_train_prior['department'] =="beverages"].reset_index()
        dept_dairy.to_csv(DepartmentBeveragesCluster.csv,sep=',')
```

Use Apriori and deriving rules using Association Rule Mining

All association rules have been derived with metrics as follows:

Minimum Support – 0.01

Lift (minimum threshold)-1

Confidence – 30%

Conviction – 40%

Department Produce

In [40]:
```
frequent_itemsets = apriori(table1, min_support=0.01, use_colnames=True)
```

In [81]:
```
frequent_itemsets
```

Out[81]:

	support	itemsets
0	0.058135	[Apple Honeycrisp Organic]
1	0.051078	[Asparagus]
2	0.011444	[Asparation/Broccolini/Baby Broccoli]
3	0.015637	[Baby Spinach]
4	0.215105	[Bag of Organic Bananas]
5	0.256616	[Banana]
6	0.025073	[Bartlett Pears]
7	0.014093	[Blackberries]
8	0.032335	[Broccoli Crown]
9	0.016331	[Brussels Sprouts]
10	0.034925	[Bunched Cilantro]
11	0.018406	[Cantaloupe]

Association Rules

In [80]:
```
final_result=rules[ (rules['lift'] > 1) &
        (rules['confidence'] >= 0.3) & rules['conviction']<1.4 ]
final_result
```

Out[80]:

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Large Lemon)	(Limes)	0.112539	0.106700	0.024059	0.213787	2.003634	0.012052	1.136207
1	(Limes)	(Large Lemon)	0.106700	0.112539	0.024059	0.225487	2.003634	0.012052	1.145831
2	(Organic Strawberries, Organic Baby Spinach)	(Bag of Organic Bananas)	0.033559	0.215105	0.010295	0.306763	1.426107	0.003076	1.132217
3	(Organic Strawberries, Bag of Organic Bananas)	(Organic Baby Spinach)	0.050377	0.169815	0.010295	0.204351	1.203375	0.001740	1.043406
4	(Organic Baby Spinach, Bag of Organic Bananas)	(Organic Strawberries)	0.041967	0.176931	0.010295	0.245300	1.386415	0.002869	1.090591
5	(Organic Strawberries)	(Organic Baby Spinach, Bag of Organic Bananas)	0.176931	0.041967	0.010295	0.058184	1.386415	0.002869	1.017219
6	(Organic Baby Spinach)	(Organic Strawberries, Bag of Organic Bananas)	0.169815	0.050377	0.010295	0.060622	1.203375	0.001740	1.010906
7	(Bag of Organic Bananas)	(Organic Strawberries, Organic	0.215105	0.033559	0.010295	0.047858	1.426107	0.003076	1.015018

Result Analysis: Large Lemon and Limes and Limes and Lemon are most frequently associate with high Lift value of 2 and small conviction and confidence of 13 % and 20 % respectively

Department – Dairy Eggs

In [15]:

```
## of the product exists then value changed to 1 else 0
def encode_units(x):
    if x == 0.0:
        return 0
    if x >= 1.0:
        return 1
basket_sets = table1.applymap(encode_units)
```

In [16]:

```
## Running Apriori to get frequent itemset with a minimum suppirt of 0.01. ie the combination must exist atleast in 1% of dataset
frequent_itemsets = apriori(table1, min_support=0.01, use_colnames=True)
```

In [17]:

```
frequent_itemsets
```

Out[17]:

	support	itemsets
0	0.012658	[1% Lowfat Milk]
1	0.012258	[2% Reduced Fat DHA Omega-3 Reduced Fat Milk]
2	0.024360	[2% Reduced Fat Milk]
3	0.014971	[2% Reduced Fat Organic Milk]
4	0.014340	[Blueberry Yoghurt]
5	0.010401	[Blueberry on the Bottom Nonfat Greek Yogurt]
6	0.010835	[Cherry Pomegranate Greek Yogurt]

In [20]:

```
## Since, ideally Lift must greater than 1 , we have selected the rules to have Lift to be greater than 1, confidence greater tha
final_result=rules[ (rules['lift'] > 1) &
        (rules['confidence'] >= 0.3) & rules['conviction']<1.4 ]
final_result
```

Out[20]:

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Icelandic Style Skyr Blueberry Non-fat Yogurt)	(Vanilla Skyr Nonfat Yogurt)	0.029076	0.025750	0.012142	0.417576	16.216821	0.011393	1.672752
1	(Vanilla Skyr Nonfat Yogurt)	(Icelandic Style Skyr Blueberry Non-fat Yogurt)	0.025750	0.029076	0.012142	0.471526	16.216821	0.011393	1.837222
2	(Total 2% with Strawberry Lowfat Greek Straine...	(Total 2% All Natural Greek Strained Yogurt wi...	0.043433	0.031430	0.010926	0.251564	8.003830	0.009561	1.294125
3	(Total 2% All Natural Greek Strained Yogurt wi...	(Total 2% with Strawberry Lowfat Greek Straine...	0.031430	0.043433	0.010926	0.347631	8.003830	0.009561	1.466297
4	(Total 2% Lowfat Greek Strained Yogurt with Pe...	(Total 2% with Strawberry Lowfat Greek Straine...	0.031283	0.043433	0.014900	0.476298	10.966245	0.013541	1.826548
5	(Total 2% with Strawberry Lowfat Greek Straine...	(Total 2% Lowfat Greek Strained Yogurt with Pe...	0.043433	0.031283	0.014900	0.343059	10.966245	0.013541	1.474587
6	(Non Fat Raspberry Yogurt)	(Icelandic Style Skyr Blueberry Non-fat Yogurt)	0.024847	0.029076	0.013137	0.528715	18.183646	0.012415	2.060160
7	(Icelandic Style Skyr Blueberry Non-fat Yogurt)	(Non Fat Raspberry Yogurt)	0.029076	0.024847	0.013137	0.451815	18.183646	0.012415	1.778876

Result Analysis: Non fat raspberry and Blueberry Non fat Yogurt are most frequently associate with high Lift value of 18.

Snacks

```
In [14]: basket_sets = table1.applymap(encode_units)

In [ ]: basket_sets

In [15]: ## Running Apriori to get frequent itemset with a minimum suppirt of 0.01. ie the combination must exist atleast in 1% of dataset
         frequent_itemsets = apriori(table1, min_support=0.01, use_colnames=True)

In [16]: frequent_itemsets
```

Out[16]:

	support	itemsets
0	0.023780	[100 Calorie Per Bag Popcorn]
1	0.023224	[Aged White Cheddar Baked Rice & Corn Puffs Gl...
2	0.013238	[Almond Nut & Rice Cracker Snacks]
3	0.021901	[Almonds & Sea Salt in Dark Chocolate]
4	0.014911	[Apple Cinnamon GoGo Squeez]
5	0.023311	[Apple Pie Fruit & Nut Food Bar]
6	0.018355	[Backyard Barbeque Potato Chips]
7	0.024248	[Baked Aged White Cheddar Rice and Corn Puffs]
8	0.010850	[Barbecue Potato Chips]
9	0.018406	[Blueberry Muffin Bar]

There were no associations found for this department

Department Frozen

```
In [15]: basket_sets = table1.applymap(encode_units)

In [21]: ## Running Apriori to get frequent itemset with a minimum support of 0.01. ie the combination must exist atleast in 1% of d
         frequent_itemsets = apriori(table1, min_support=0.009, use_colnames=True)

In [22]: frequent_itemsets
```

Out[22]:

	support	itemsets
0	0.014581	[Air Chilled Breaded Chicken Breast Nuggets]
1	0.013146	[Americone Dream® Ice Cream]
2	0.009804	[Bean & Cheese Burrito]
3	0.024423	[Berry Medley]
4	0.011346	[Birthday Cake Light Ice Cream]
5	0.010354	[Black Bean Tamale Verde]
6	0.009934	[Black Bean Vegetables Burrito]
7	0.058071	[Blueberries]
8	0.026171	[Broccoli & Cheddar Bake Meal Bowl]

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Organic Sliced Peaches)	(Frozen Organic Wild Blueberries)	0.027430	0.051708	0.009194	0.335188	6.482325	0.007776	1.426406
1	(Frozen Organic Wild Blueberries)	(Organic Sliced Peaches)	0.051708	0.027430	0.009194	0.177807	6.482325	0.007776	1.182898
2	(Organic Whole Strawberries)	(Frozen Organic Wild Blueberries)	0.062336	0.051708	0.023164	0.371603	7.186581	0.019941	1.509066
3	(Frozen Organic Wild Blueberries)	(Organic Whole Strawberries)	0.051708	0.062336	0.023164	0.447986	7.186581	0.019941	1.698622
4	(Organic Sliced Peaches)	(Organic Whole Strawberries)	0.027430	0.062336	0.011597	0.422809	6.782702	0.009888	1.624530
5	(Organic Whole Strawberries)	(Organic Sliced Peaches)	0.062336	0.027430	0.011597	0.186047	6.782702	0.009888	1.194872
6	(Organic Frozen Mango Chunks)	(Organic Whole Strawberries)	0.025843	0.062336	0.010499	0.406259	6.517204	0.008888	1.579247
7	(Organic Whole Strawberries)	(Organic Frozen Mango Chunks)	0.062336	0.025843	0.010499	0.168421	6.517204	0.008888	1.171455

Result Analysis: Organic Whole Strawberries and Organic Blueberries are most frequently associate with high Lift.

Pantry

Out[73]:

	support	itemsets
0	0.023617	[100% Pure Pumpkin]
1	0.010879	[60% Cacao Bittersweet Premium Baking Chips]
2	0.017613	[Active Dry Yeast]
3	0.027474	[All Purpose Flour]
4	0.017638	[Almond Meal/Flour]
5	0.022700	[Apple Cider Vinegar]
6	0.010125	[Apricot Preserves]
7	0.019849	[Baking Powder]
8	0.025376	[Cane Sugar]
9	0.023605	[Canola Oil]
10	0.022047	[Cinnamon Rolls with Icing]

Result Analysis: No Asscoiation Rules were found

Department Canned Goods

26	(Organic Garbanzo Beans)	(Organic Pinto Beans)	0.095757	0.063316	0.013171	0.137542	2.172300	0.007108	1.086063
27	(Organic Pinto Beans)	(Organic Garbanzo Beans)	0.063316	0.095757	0.013171	0.208014	2.172300	0.007108	1.141740
28	(Dark Red Kidney Beans No Salt Added)	(No Salt Added Black Beans)	0.031766	0.095865	0.015195	0.478335	4.989656	0.012150	1.733170
29	(No Salt Added Black Beans)	(Dark Red Kidney Beans No Salt Added)	0.095865	0.031766	0.015195	0.158502	4.989656	0.012150	1.150608
30	(Organic No Salt Added Diced Tomatoes)	(No Salt Added Black Beans)	0.048526	0.095865	0.012874	0.265295	2.767371	0.008222	1.230609
31	(No Salt Added Black Beans)	(Organic No Salt Added Diced Tomatoes)	0.095865	0.048526	0.012874	0.134291	2.767371	0.008222	1.099068

Result Analysis: RedKidneyBeans and Black Beans are always bought together as they seem to have a large Lift value

Deli

35	(Uncured Genoa Salami)	(Uncured Slow Cooked Ham)	0.182028	0.054724	0.018625	0.102321	1.869776	0.008664	1.053022
36	(Organic Uncured Sliced Black Forest Ham)	(Uncured Genoa Salami)	0.046211	0.182028	0.014145	0.306094	1.681581	0.005733	1.178794
37	(Uncured Genoa Salami)	(Organic Uncured Sliced Black Forest Ham)	0.182028	0.046211	0.014145	0.077707	1.681581	0.005733	1.034150
38	(Organic Roasted Turkey Breast)	(Organic Uncured Sliced Black Forest Ham)	0.095814	0.046211	0.015361	0.160321	3.469321	0.010933	1.135897
39	(Organic Uncured Sliced Black Forest Ham)	(Organic Roasted Turkey Breast)	0.046211	0.095814	0.015361	0.332410	3.469321	0.010933	1.354403
40	(BBQ Chopped Salad)	(Santa Fe Fiesta Salad)	0.021313	0.026882	0.013697	0.642643	23.906306	0.013124	2.723096
41	(Santa Fe Fiesta Salad)	(BBQ Chopped Salad)	0.026882	0.021313	0.013697	0.509524	23.906306	0.013124	1.995381
42	(Roasted Red Pepper Hummus)	(Original Hummus)	0.059460	0.184780	0.011777	0.198062	1.071883	0.000790	1.016563
43	(Original Hummus)	(Roasted Red Pepper Hummus)	0.184780	0.059460	0.011777	0.063734	1.071883	0.000790	1.004565
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Result Analysis : BBQ Chopped Salad and Santa Fe Fiesta Salad have high associations with high lift values of 23

Department GoodsPasta

```
In [18]: rules
```

Out[18]:

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Penne Pasta)	(Rotini Pasta)	0.063822	0.046099	0.011446	0.179339	3.890258	0.008504	1.162356
1	(Rotini Pasta)	(Penne Pasta)	0.046099	0.063822	0.011446	0.248284	3.890258	0.008504	1.245388
2	(Original Rice Pilaf Mix)	(Spanish Rice Pilaf Mix)	0.052534	0.030434	0.011182	0.212851	6.993847	0.009583	1.231744
3	(Spanish Rice Pilaf Mix)	(Original Rice Pilaf Mix)	0.030434	0.052534	0.011182	0.367418	6.993847	0.009583	1.497774
4	(Macaroni Shells & White Cheddar Cheese)	(Shells & Real Aged Cheddar Macaroni & Cheese)	0.048262	0.041300	0.014505	0.300546	7.277216	0.012512	1.370642
5	(Shells & Real Aged Cheddar Macaroni & Cheese)	(Macaroni Shells & White Cheddar Cheese)	0.041300	0.048262	0.014505	0.351213	7.277216	0.012512	1.466950
6	(Garlic Couscous)	(Original Rice Pilaf Mix)	0.041827	0.052534	0.010233	0.244641	4.656768	0.008035	1.254324

Result Analysis : Macroni Shell and Real Aged Cheddar have high associations with a lift value of 7.27

Department Meat Seafood

:

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	Organic Beef Uncured Hot Dogs	Boneless Skinless Chicken Breasts	0.068606	0.131175	0.011526	0.168000	1.280736	0.002526	1.044261
1	Boneless Skinless Chicken Breasts	Organic Beef Uncured Hot Dogs	0.131175	0.068606	0.011526	0.087866	1.280736	0.002526	1.021116
2	Ground Turkey Breast	Organic Turkey Bacon	0.107025	0.065038	0.015093	0.141026	2.168344	0.008133	1.088463
3	Organic Turkey Bacon	Ground Turkey Breast	0.065038	0.107025	0.015093	0.232068	2.168344	0.008133	1.162830
4	Ground Turkey Breast	Tilapia Filet	0.107025	0.054885	0.011526	0.107692	1.962154	0.005652	1.059181
5	Tilapia Filet	Ground Turkey Breast	0.054885	0.107025	0.011526	0.210000	1.962154	0.005652	1.130348
6	Organic Turkey Bacon	Organic Chicken & Apple Sausage	0.065038	0.060648	0.012075	0.185654	3.061191	0.008130	1.153505
7	Organic Chicken & Apple Sausage	Organic Turkey Bacon	0.060648	0.065038	0.012075	0.199095	3.061191	0.008130	1.167381
8	Organic Beef Uncured Hot Dogs	Natural Chicken & Sage Breakfast Sausage	0.068606	0.101811	0.016465	0.240000	2.357305	0.009481	1.181827
9	Natural Chicken & Sage Breakfast Sausage	Organic Beef Uncured Hot Dogs	0.101811	0.068606	0.016465	0.161725	2.357305	0.009481	1.111084
10	Natural Chicken & Sage Breakfast Sausage	Organic Beef Hot Dogs	0.101811	0.079034	0.012349	0.121294	1.534704	0.004303	1.048093
11	Organic Beef Hot Dogs	Natural Chicken & Sage Breakfast Sausage	0.079034	0.101811	0.012349	0.156250	1.534704	0.004303	1.064520

Result Analysis: Organic Turkey Bacon and Organic Turkey Breast have associations with Lift values of 2.

Department Breakfast

```
In [19]: ## Since, ideally Lift must greater than 1 , we have selected the rules to have Lift to be greater than 1, confidence greater tha
         final_result=rules[ (rules['lift'] > 1) &
                (rules['confidence'] >= 0.3) & rules['conviction']<1.4 ]
         final_result
```

50	(with Crispy Almonds Cereal)	(Cherrios Honey Nut)	0.022531	0.037775	0.017254	0.765799	20.272455	0.016403	4.108546
51	(Cherrios Honey Nut)	(with Crispy Almonds Cereal)	0.037775	0.022531	0.017254	0.456763	20.272455	0.016403	1.799341
52	(Frosted Flakes)	(Cherrios Honey Nut)	0.035681	0.037775	0.017087	0.478873	12.676868	0.015739	1.846431
53	(Cherrios Honey Nut)	(Frosted Flakes)	0.037775	0.035681	0.017087	0.452328	12.676868	0.015739	1.760760
54	(Raisin Bran)	(with Crispy Almonds Cereal)	0.024039	0.022531	0.013150	0.547038	24.279147	0.012609	2.157950
55	(with Crispy Almonds Cereal)	(Raisin Bran)	0.022531	0.024039	0.013150	0.583643	24.279147	0.012609	2.344050
56	(Peanut Butter Whole Grain Clusters, Vanilla B...	(Healthy Grains Fiber Cinnamon Oat Clusters)	0.017757	0.036016	0.010135	0.570755	15.847071	0.009495	2.245764
57	(Healthy Grains Fiber Cinnamon Oat Clusters, V...	(Peanut Butter Whole Grain Clusters)	0.019600	0.035933	0.010135	0.517094	14.390642	0.009431	1.996387
58	(Peanut Butter Whole Grain Clusters, Healthy G...	(Vanilla Blueberry Clusters With Flax Seeds Gr...	0.014490	0.046821	0.010135	0.699422	14.938102	0.009456	3.171152
59	(Vanilla Blueberry Clusters With Flax Seeds Gr...	(Peanut Butter Whole Grain Clusters, Healthy G...	0.046821	0.014490	0.010135	0.216458	14.938102	0.009456	1.257762

Result Analysis : Peanut Butter Whole Grain clusters and cinnamon clusters are bought together with a lift of 15

House Hold

```
In [109]: ## Since, ideally Lift must greater than 1 , we have selected the rules to have Lift to be greater than 1, confidence greater tha
          final_result=rules[ (rules['lift'] > 1) &
                 (rules['confidence'] >= 0.3) & rules['conviction']<1.4 ]
          final_result
```

Out[109]:

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(100% Recycled Paper Towels)	(Sustainably Soft Bath Tissue)	0.049682	0.032810	0.013978	0.281349	8.574989	0.012348	1.345841
1	(Sustainably Soft Bath Tissue)	(100% Recycled Paper Towels)	0.032810	0.049682	0.013978	0.426024	8.574989	0.012348	1.655675

```
In [110]: final_result.to_csv("Department_householdApriori.csv",sep=',')
```

```
In [111]: test1=pd.read_csv("Department_householdApriori.csv")
```

Result Analysis: Paper Towels and Soft Bath Tissue are bought together with a Lift of 8.57

Personal Care

```
test1
```

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	Vegan Nutritional Shake Sweet Vanilla Bean	Vegan Smooth Chocolate Nutritional Shake	0.003341	0.004108	0.001086	0.324955	79.095261	0.001072	1.475297
1	Vegan Smooth Chocolate Nutritional Shake	Vegan Nutritional Shake Sweet Vanilla Bean	0.004108	0.003341	0.001086	0.264234	79.095261	0.001072	1.354587
2	Lemon Verbena Hand Soap	Lavender Hand Soap	0.015420	0.018590	0.001397	0.090626	4.875059	0.001111	1.079215
3	Lavender Hand Soap	Lemon Verbena Hand Soap	0.018590	0.015420	0.001397	0.075173	4.875059	0.001111	1.064610
4	Moroccan Argan Oil + Argan Stem Cell Triple Mo...	Hair Shampoos	0.002789	0.002150	0.001014	0.363441	169.029400	0.001008	1.567568
5	Hair Shampoos	Moroccan Argan Oil + Argan Stem Cell Triple Mo...	0.002150	0.002789	0.001014	0.471409	169.029400	0.001008	1.886544

Result Analysis : Morroccon Argan Oil and Shampoo is bought with a high Lift value

Department Babies

195	(Spinach Peas & Pear Stage 2 Baby Food)	(Peach, Apricot & Banana Stage 2 Baby Food, B...	0.117706	0.030085	0.010742	0.091260	3.033427	0.007201	1.067319
196	(Peach, Apricot & Banana Stage 2 Baby Food)	(Spinach Peas & Pear Stage 2 Baby Food, Baby F...	0.076652	0.044294	0.010742	0.140138	3.163843	0.007347	1.111465
197	(Baby Food Stage 2 Blueberry Pear & Purple Car...	(Spinach Peas & Pear Stage 2 Baby Food, Peach,...	0.126838	0.021029	0.010742	0.084690	4.027280	0.008075	1.069551
198	(Mighty 4 Sweet Potato, Blueberry, Millet & Gr...	(Mighty 4 Purple Carrot Blackberry Quinoa & Gr...	0.039122	0.029706	0.016312	0.416949	14.035870	0.015150	1.664167
199	(Mighty 4 Purple Carrot Blackberry Quinoa & Gr...	(Mighty 4 Sweet Potato, Blueberry, Millet & Gr...	0.029706	0.039122	0.016312	0.549107	14.035870	0.015150	2.131057
200	(Organic Pears, Peas and Broccoli Puree Stage ...	(Stage 1 Apples Sweet Potatoes Pumpkin & Blueb...	0.022962	0.084325	0.014853	0.646865	7.671109	0.012917	2.592987
201	(Organic Pears, Peas and Broccoli Puree Stage ...	(Organic 4 Months Butternut Squash Carrots App...	0.033306	0.052971	0.014853	0.445961	8.419035	0.013089	1.709320
202	(Organic 4 Months Butternut Squash Carrots App...	(Organic Pears, Peas and Broccoli Puree Stage 1)	0.028475	0.071423	0.014853	0.521623	7.303281	0.012819	1.941100

Result Analysis: Associations have a high lift of 14

Department International

11	(Seaweed Ramen)	(Mushroom Ramen, Asian Vegetable Ramen)	0.061135	0.051092	0.020961	0.342857	6.710623	0.017837	1.443991
12	(Tikka Masala Simmer Sauce)	(Lemongrass Basil Simmer Sauce)	0.023581	0.017904	0.010044	0.425926	23.789521	0.009621	1.710748
13	(Lemongrass Basil Simmer Sauce)	(Tikka Masala Simmer Sauce)	0.017904	0.023581	0.010044	0.560976	23.789521	0.009621	2.224066
14	(Garlic Pepper Ramen)	(Seaweed Ramen)	0.084279	0.061135	0.026201	0.310881	5.085122	0.021048	1.362413
15	(Seaweed Ramen)	(Garlic Pepper Ramen)	0.061135	0.084279	0.026201	0.428571	5.085122	0.021048	1.602511
16	(Tofu Miso Ramen)	(Garlic Pepper Ramen)	0.101747	0.084279	0.052838	0.519313	6.161800	0.044263	1.905026
17	(Garlic Pepper Ramen)	(Tofu Miso Ramen)	0.084279	0.101747	0.052838	0.626943	6.161800	0.044263	2.407818
18	(Madras Lentils Indian Cuisine)	(Bombay Potatoes Vegetarian)	0.045852	0.034934	0.014410	0.314286	8.996429	0.012809	1.407387

Result Analysis : Tikka Masala Simmer Sauce and Lemon Grass Simmer is bought together with a lift of 23

Alchohol

```
In [18]: rules
```

259	(Old Vine Zinfandel)	(Cabernet Sauvignon)	0.025291	0.206586	0.010642	0.420792	2.036889	0.005418	1.369826
260	(Belgian White Wheat Ale)	(India Pale Ale)	0.069738	0.210592	0.021911	0.314183	1.491903	0.007224	1.151047
261	(India Pale Ale)	(Belgian White Wheat Ale)	0.210592	0.069738	0.021911	0.104043	1.491903	0.007224	1.038288
262	(Cabernet Sauvignon)	(Pinot Noir Wine)	0.206586	0.054839	0.024289	0.117576	2.144013	0.012960	1.071096
263	(Pinot Noir Wine)	(Cabernet Sauvignon)	0.054839	0.206586	0.024289	0.442922	2.144013	0.012960	1.424244
264	(Belgian White Beer, Beer)	(India Pale Ale)	0.014148	0.210592	0.010016	0.707965	3.361780	0.007037	2.703124
265	(Belgian White Beer, India Pale Ale)	(Beer)	0.021034	0.232753	0.010016	0.476190	2.045903	0.005121	1.464744
266	(Beer, India Pale Ale)	(Belgian White Beer)	0.092525	0.034932	0.010016	0.108254	3.099025	0.006784	1.082224
267	(Belgian White Beer)	(Beer, India Pale Ale)	0.034932	0.092525	0.010016	0.286738	3.099025	0.006784	1.272289
268	(Beer)	(Belgian White Beer, India Pale Ale)	0.232753	0.021034	0.010016	0.043034	2.045903	0.005121	1.022989
269	(India Pale Ale)	(Belgian White Beer, Beer)	0.210592	0.014148	0.010016	0.047562	3.361780	0.007037	1.035083

Result Anaysis : Belgian Beer and India Pale Ale have a high a lift value 3.36.

Department Pets

Organix Butcher & Bushel Grain-Free Turkey Din...	Organix Butcher & Bushel Grain-Free Tender Chi...	0.012675	0.023127	0.010229	0.807018	34.895749
Organix Butcher & Bushel Grain-Free Turkey Din...	Organix Butcher & Bushel Grain-Gree Turkey & C...	0.016233	0.019124	0.010229	0.630137	32.950303
Organix Butcher & Bushel Grain-Gree Turkey & C...	Organix Butcher & Bushel Grain-Free Turkey Din...	0.013342	0.023794	0.010229	0.766667	32.221495
Organix Butcher & Bushel Grain-Free Turkey Din...	Organix Butcher & Bushel Grain-Gree Turkey & C...	0.023794	0.013342	0.010229	0.429907	32.221495
Organix Butcher & Bushel Grain-Gree Turkey & C...	Organix Butcher & Bushel Grain-Free Turkey Din...	0.019124	0.016233	0.010229	0.534884	32.950303
Organix Butcher & Bushel Grain-Free Tender Chi...	Organix Butcher & Bushel Grain-Free Turkey Din...	0.023127	0.012675	0.010229	0.442308	34.895749

Result Analysis : Organix Butcher and Bushen Grain are bought together with a lift of 32

Department Missing

1027	(Organic Nondairy Strawberry Cashew Yogurt, Or...	(Organic Cashew Nondairy Blueberry Yogurt, Org...	0.075728	0.170874	0.042718	0.564103	3.301282	0.029778	1.902113
1028	(Organic Cashew Nondairy Blueberry Yogurt)	(Organic Cashew Nondairy Vanilla Yogurt, Organ...	0.238835	0.058252	0.042718	0.178862	3.070461	0.028806	1.146881
1029	(Organic Cashew Nondairy Vanilla Yogurt)	(Organic Cashew Nondairy Blueberry Yogurt, Org...	0.225243	0.060194	0.042718	0.189655	3.150723	0.029160	1.159760
1030	(Organic Nondairy Strawberry Cashew Yogurt)	(Organic Cashew Nondairy Blueberry Yogurt, Org...	0.267961	0.044660	0.042718	0.159420	3.569628	0.030751	1.136525
1031	(Organic Plain Unsweetened Nondairy Cashew Yog...	(Organic Cashew Nondairy Blueberry Yogurt, Org...	0.108738	0.163107	0.042718	0.392857	2.408588	0.024983	1.378412
1032	(Peanut Butter & Coconut Bar)	(Blueberry Cashew Bar)	0.033010	0.017476	0.011650	0.352941	20.196078	0.011074	1.518447
1033	(Blueberry Cashew Bar)	(Peanut Butter & Coconut Bar)	0.017476	0.033010	0.011650	0.666667	20.196078	0.011074	2.900971
1034	(Stage 2 Eat Your Colors Organic, Green: Pea, ...	(Organic Plain Unsweetened Nondairy Cashew Yog...	0.184466	0.108738	0.023301	0.126316	1.161654	0.003243	1.020119
1035	(Organic Plain Unsweetened Nondairy Cashew Yog...	(Stage 2 Eat Your Colors Organic, Green: Pea, ...	0.108738	0.184466	0.023301	0.214286	1.161654	0.003243	1.037952

Result Analysis : Peanut Butter and Coconut Bar and Blueberry Cashew bar Grain are bought together with a lift of 20

Other

1408	Sniffle Support Drops Alcohol Free Formula, Na...	Multivitamin, Kids Complete, Gummies, u"Childr...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1409	Multivitamin, Kids Complete, Gummies	u"Childrens Chestal Homeopathic Medicine", Bab...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1410	u"Childrens Chestal Homeopathic Medicine"	Multivitamin, Kids Complete, Gummies, Baby Vit...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1411	Baby Vitamin C tablets	Multivitamin, Kids Complete, Gummies, u"Childr...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1412	Sniffle Support Drops Alcohol Free Formula	Multivitamin, Kids Complete, Gummies, u"Childr...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1413	Nasal Aspirator	Multivitamin, Kids Complete, Gummies, u"Childr...	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1414	Multivitamin, Kids Complete, Gummies, Baby Vit...	Nasal Aspirator	0.066667	0.066667	0.066667	1.000000	15.0	0.062222
1415	Multivitamin, Kids Complete, Gummies, Baby Vit...	Sniffle Support Drops Alcohol Free Formula	0.066667	0.066667	0.066667	1.000000	15.0	0.062222

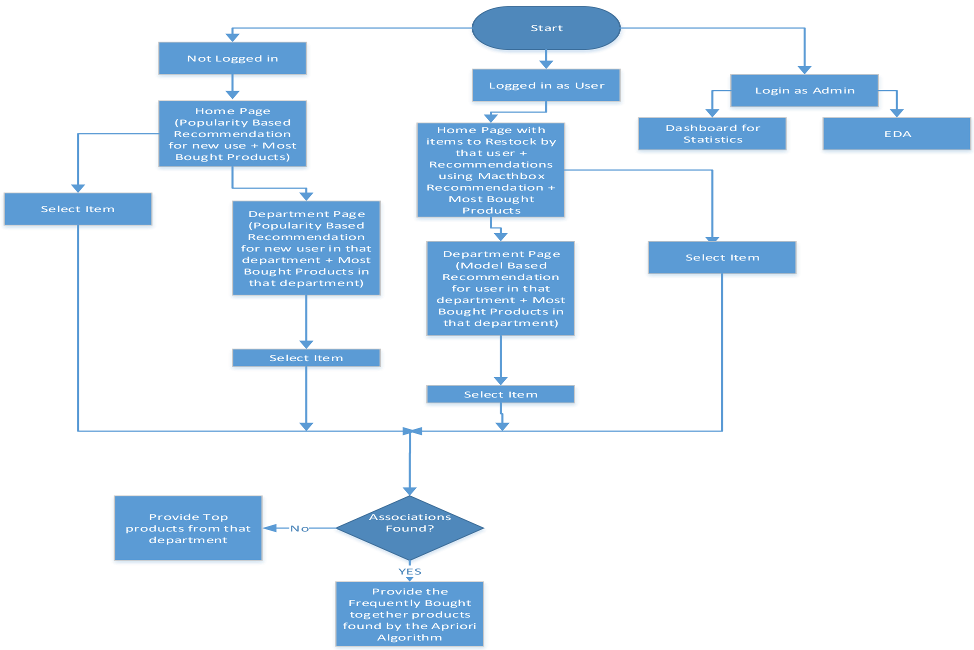
Result Analysis: All association have a Lift of 1

Department Bulk

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage
0	(Organic Rolled Oats, Dried Mango)	(Organic Pearled Barley)	0.5	0.5	0.5	1.0	2.0	0.25
1	(Organic Rolled Oats, Organic Pearled Barley)	(Dried Mango)	0.5	0.5	0.5	1.0	2.0	0.25
2	(Dried Mango, Organic Pearled Barley)	(Organic Rolled Oats)	0.5	1.0	0.5	1.0	1.0	0.00
3	(Organic Rolled Oats)	(Dried Mango, Organic Pearled Barley)	1.0	0.5	0.5	0.5	1.0	0.00
4	(Dried Mango)	(Organic Rolled Oats, Organic Pearled Barley)	0.5	0.5	0.5	1.0	2.0	0.25
5	(Organic Pearled Barley)	(Organic Rolled Oats, Dried Mango)	0.5	0.5	0.5	1.0	2.0	0.25
6	(Organic Mung Beans, Organic Royal Rainbow Qui...	(Dried Mango)	0.5	0.5	0.5	1.0	2.0	0.25
7	(Organic Mung Beans, Dried Mango)	(Organic Royal Rainbow Quinoa)	0.5	0.5	0.5	1.0	2.0	0.25

Result Analysis : All Associations have a result set of 2

Implementation on the Web Application



URL: http://35.190.167.191/

The web application handles 3 kinds of users

1. New User
2. Existing User
3. Administrator

On selecting a product from a department or from the homepage, the user is provided recommendations using the recommendation models. Once, the product is selected the user is provided options based on the association rules. If no association is found the user is prompted with the most popular product.

Conclusion

Recommendation Systems are a large scope in the industry , considering the information overload due to the internet. Recommendation Systems for Instacart improves the user experience and hence, ensures that the user to be a returning user. This also helps reduce customer churn.

Model Based recommendation provided the best **precision** of 0.28. Apriori provided the best **lift** 128

Link for Slide share : https://www.slideshare.net/TusharGoel42/final-project-ads-info7390