

Milestone Report



Recommend Products To Instacart Customers

- Oindrila Sen

[Introduction:](#)

[Data Source:](#)

[Data Wrangling:](#)

[Data Model:](#)

[Prepare Dataset:](#)

[Data Cleaning:](#)

[Feature Extraction:](#)

[Exploratory Data Analysis:](#)

[Conclusion:](#)

Introduction:

Online Shopping is a boon to many people. No more makeup, no more worries about the bad hair day and no more looking for a parking spot - The entire Retail World is just a click away. Receiving all your groceries at your doorstep on the same day is Fun. **Instacart** is the name who offers this service.

So, how does that actually work?

Customers place orders online through a Mobile app or using the website to be fulfilled from one of their grocery retail partners. A personal shopper engaged by Instacart picks the items in the store and delivers them in as little as an hour or at the time the customer selects.

Data Source:

For this project, I am using the Dataset from “The Instacart Online Grocery Shopping Dataset 2017”, Accessed from <https://www.instacart.com/datasets/grocery-shopping-2017> on 07/10/2019.

This anonymized dataset contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, the dataset provides between 4 and 100 of their orders, with the sequence of products purchased in each order. It also provides the week and hour of day the order was placed, and a relative measure of time between the orders.

Problem to Solve:

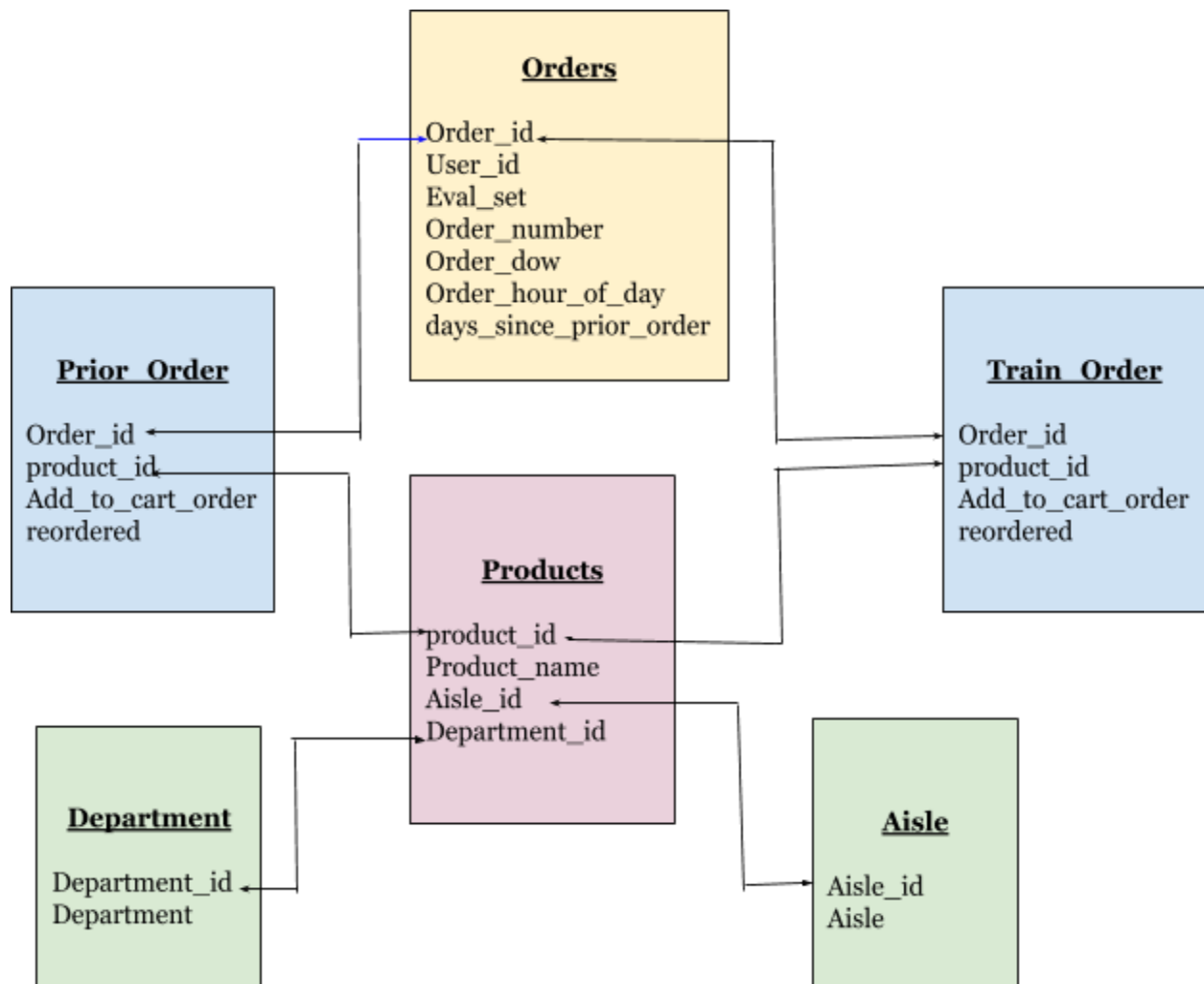
Our options are limitless and browsing through those endless choices of products are tiresome. Recommendation helps customers to find products without putting much efforts. What to eat, which movie to watch, what book to read or which product to buy are the questions that we find ourselves to answer all the time. Traditionally, these questions are answered with peer recommendations(friends, family, forums, blog posts or reviews) . But these traditional methods are limited and biased. So, how about getting an unbiased answer?

In this project, I want to explore customer shopping behavior by analyzing their previous purchases and build a shopping Recommendation Engine that could give them a tailored shopping experience. In this Dataset, we do not have any Feedback/Rating for any of the items brought by the customer. That means, we have **Implicit Data** and Recommendation Engine has to be built only focussing on user's purchase history.

Data Wrangling:

Before starting to talk about the different Data Wrangling method, let's take a look at the Data Model in Detail:

Data Model:



Now, Let's checkout the data:

1. Orders:

	order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
0	2539329	1	prior	1	2	8	NaN
1	2398795	1	prior	2	3	7	15.0
2	473747	1	prior	3	3	12	21.0
3	2254736	1	prior	4	4	7	29.0
4	431534	1	prior	5	4	15	28.0
5	3367565	1	prior	6	2	7	19.0
6	550135	1	prior	7	1	9	20.0
7	3108588	1	prior	8	1	14	14.0
8	2295261	1	prior	9	1	16	0.0
9	2550362	1	prior	10	4	8	30.0
10	1187899	1	train	11	4	8	14.0

2. Prior Order

	order_id	product_id	add_to_cart_order	reordered
24076664	2539329	196	1	0
24076665	2539329	14084	2	0
24076666	2539329	12427	3	0
24076667	2539329	26088	4	0
24076668	2539329	26405	5	0

3. Train Order

	order_id	product_id	add_to_cart_order	reordered
484420	1187899	196	1	1
484421	1187899	25133	2	1
484422	1187899	38928	3	1
484423	1187899	26405	4	1
484424	1187899	39657	5	1
484425	1187899	10258	6	1
484426	1187899	13032	7	1
484427	1187899	26088	8	1
484428	1187899	27845	9	0
484429	1187899	49235	10	1
484430	1187899	46149	11	1

4. Products

	product_id	product_name	aisle_id	department_id
0	1	Chocolate Sandwich Cookies	61	19
1	2	All-Seasons Salt	104	13

2. Department

	department_id	department
0	1	frozen
1	2	other
		bakery
3	4	produce
4	5	alcohol

3. Aisle

aisle_id		aisle
0	1	prepared soups salads
1	2	specialty cheeses
2	3	energy granola bars
3	4	instant foods
4	5	marinades meat preparation

After loading all the data, here is a snapshot:

Total Aisles: 134

Total Departments: 21

Total Products Count: 49688

Total Users/Customers: 206209

Total Orders: 3421083

Apparently, it is a huge Dataset.

Prepare Dataset:

1. Combine aisles, departments and products
2. Combine Prior Orders and Products
3. Combine Train Orders and Products
4. Combine Prior orders and Order Details
5. Combine Train Orders and Order Details
6. Merge The Train and Prior Dataset

Data Cleaning:

1. Convert NAN values to Zero
2. Change the below Column's DataTypes
 - ❖ Change Eval_set to Category

-
- ❖ Change Department name Data type to Category
 - ❖ Change Aisle name to Category
 - ❖ Change Reordered column to Boolean
3. Sort the Data

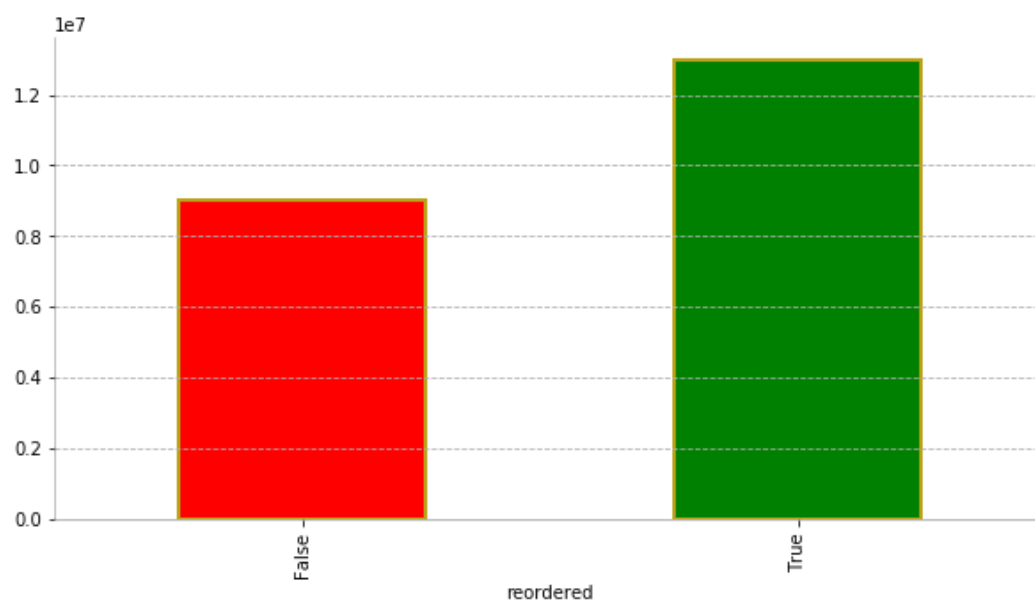
Feature Extraction:

The below new features are extracted from the existing data and added as a new column.

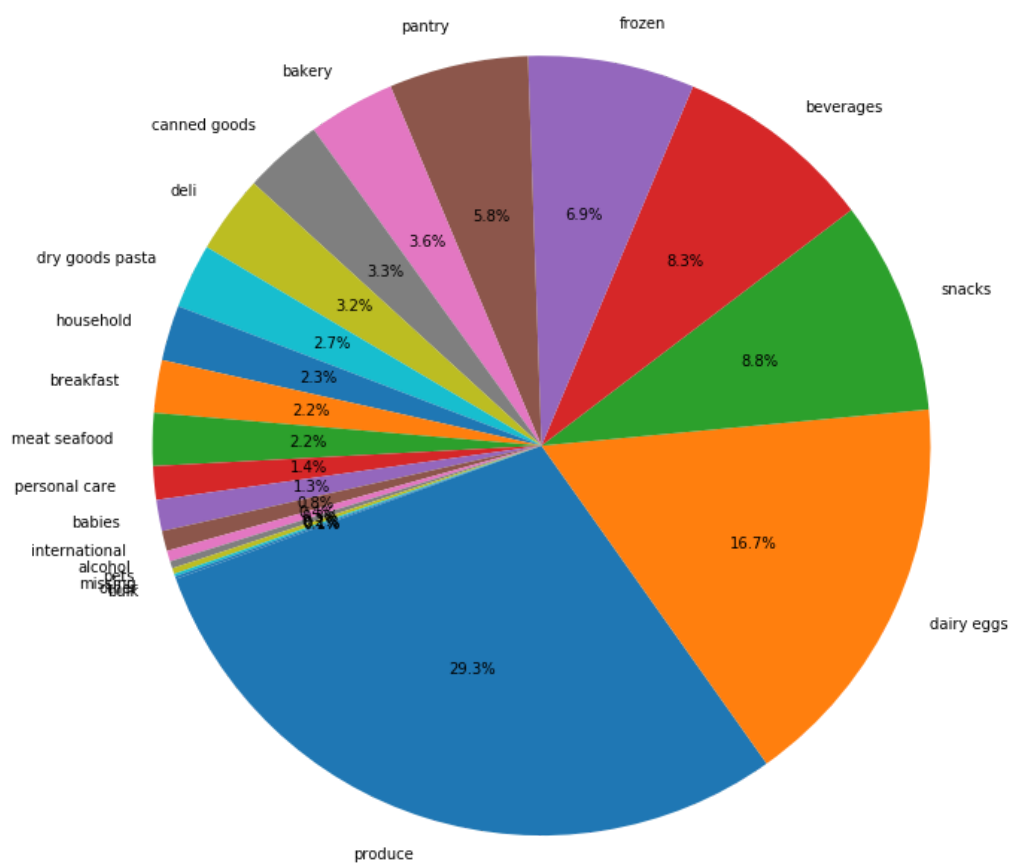
1. **Size of the Order:** How many Items are added per order?
2. **Total Orders per Customer:** How many transactions are done by the Customer?
3. **Loyal Customer:** Potential Customers are those customers who have ordered more than 32 times (The Average Order Counts)
4. **Order Span:** For how long the Customer is doing transactions?
5. **Customer Department Count:** How many times a customer bought a product from a particular department group?
6. **Customer Item Count:** How many times a customer bought a product?
7. **Weekend Customer:** Does the customer prefer to buy on weekends or a week day?
8. **Shopping Hour:** Does the customer prefers to shop at morning or evening or at mid night?
9. **Is_Organic:** The products with Organic In the Name are the Organic Products
10. **Prefers Organic:** Does the Customer prefers organic products?

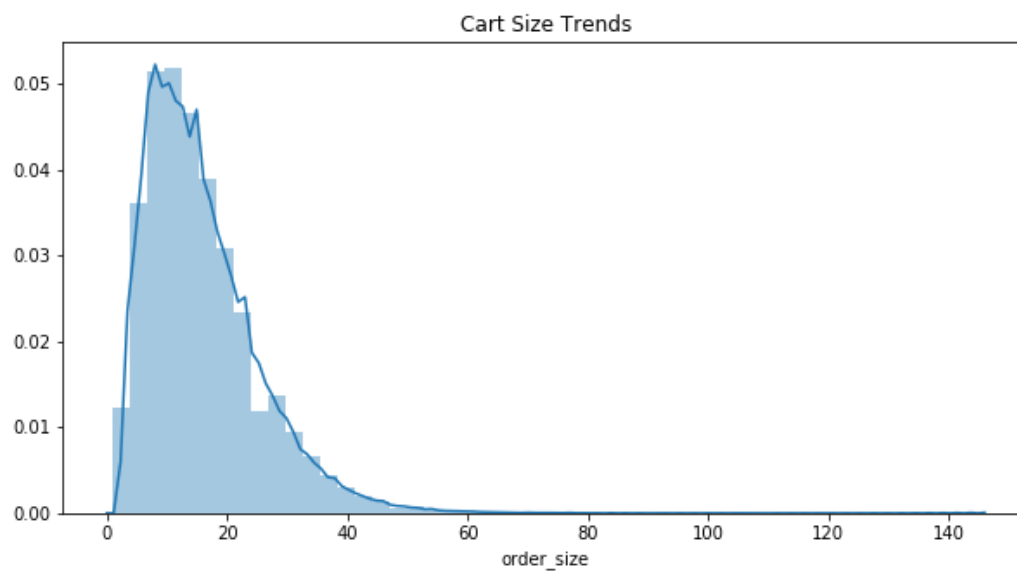
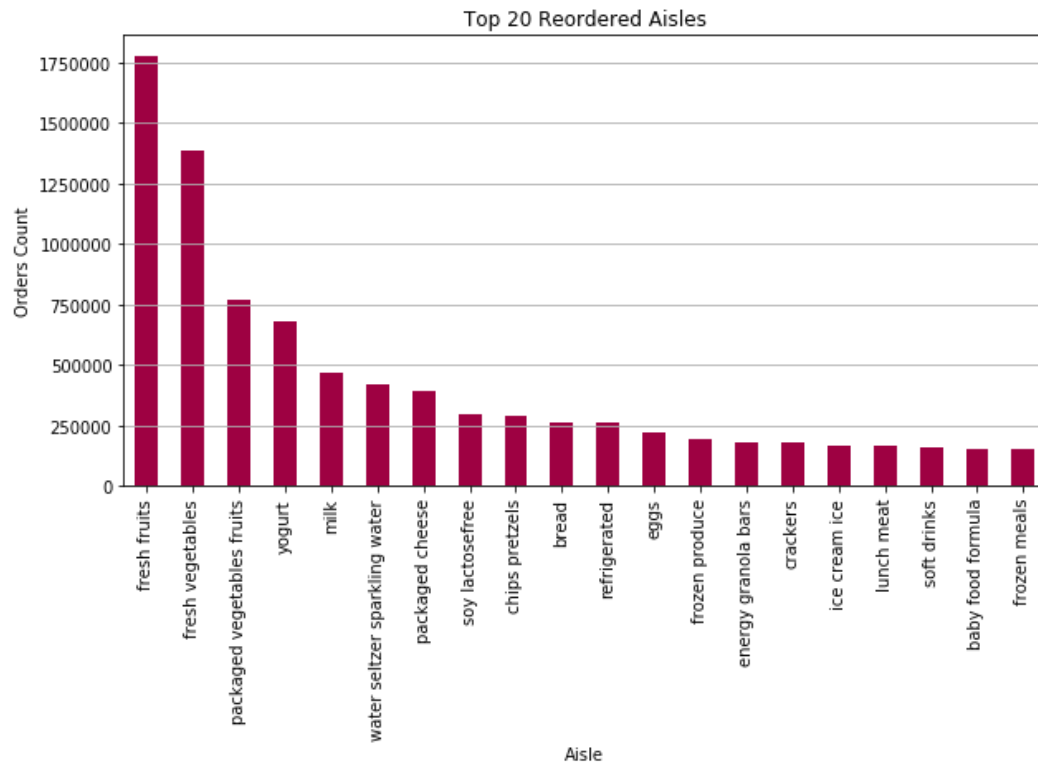
Exploratory Data Analysis:

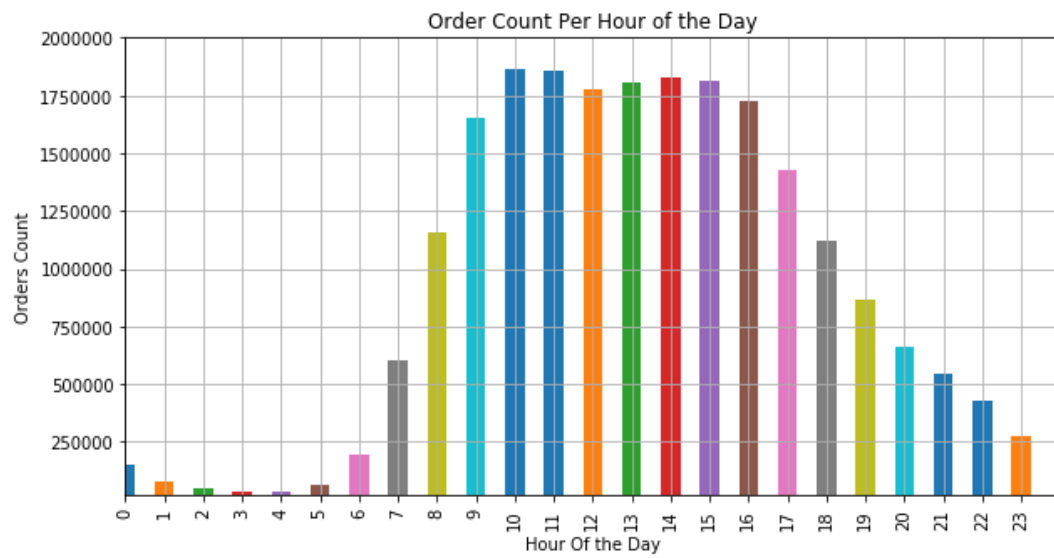
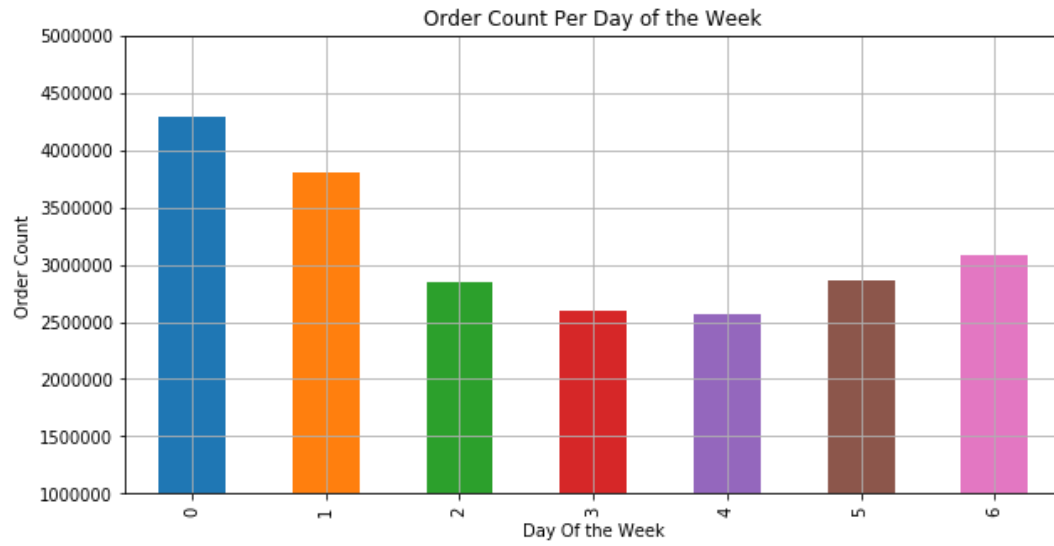
Now, let's take a look at the visual aides:

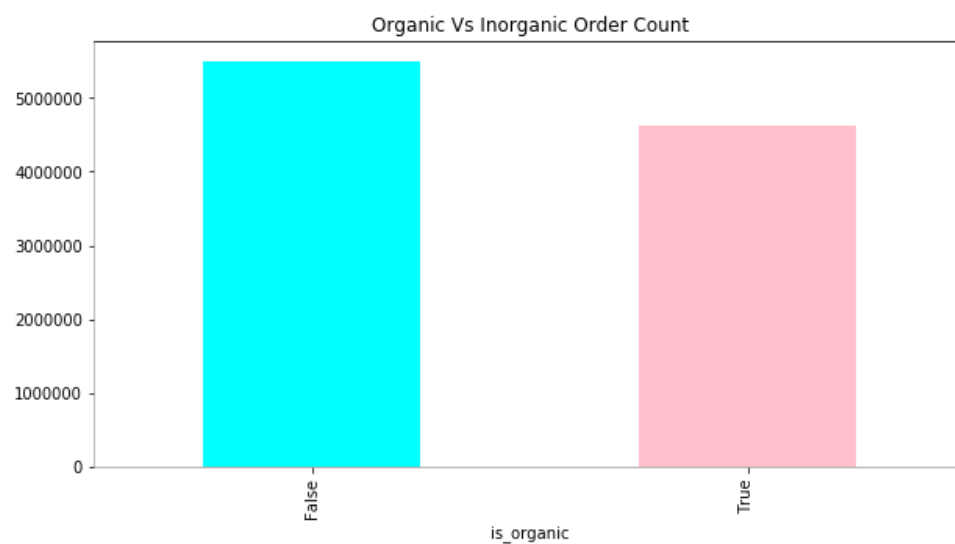
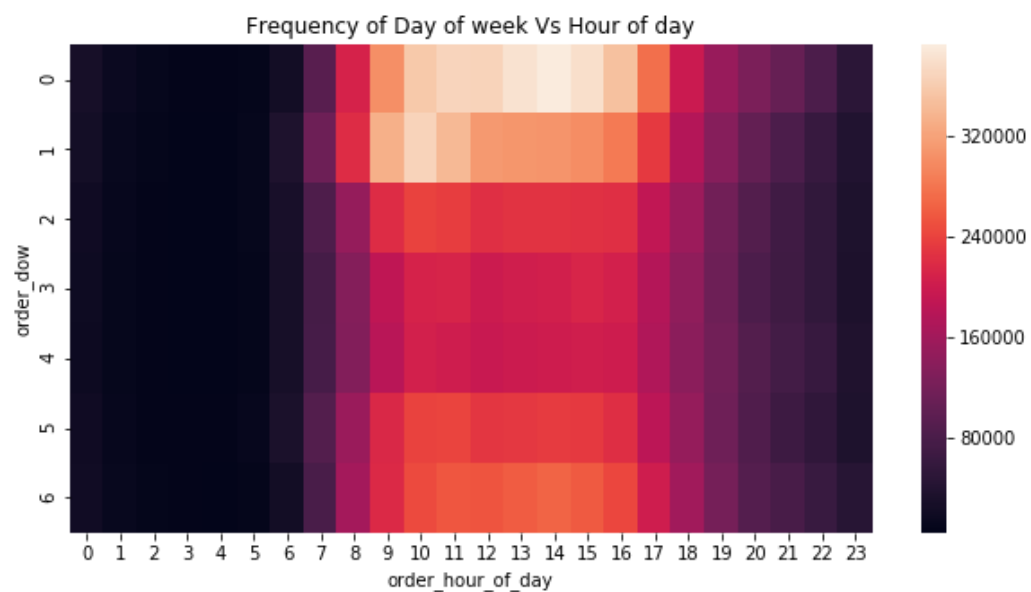


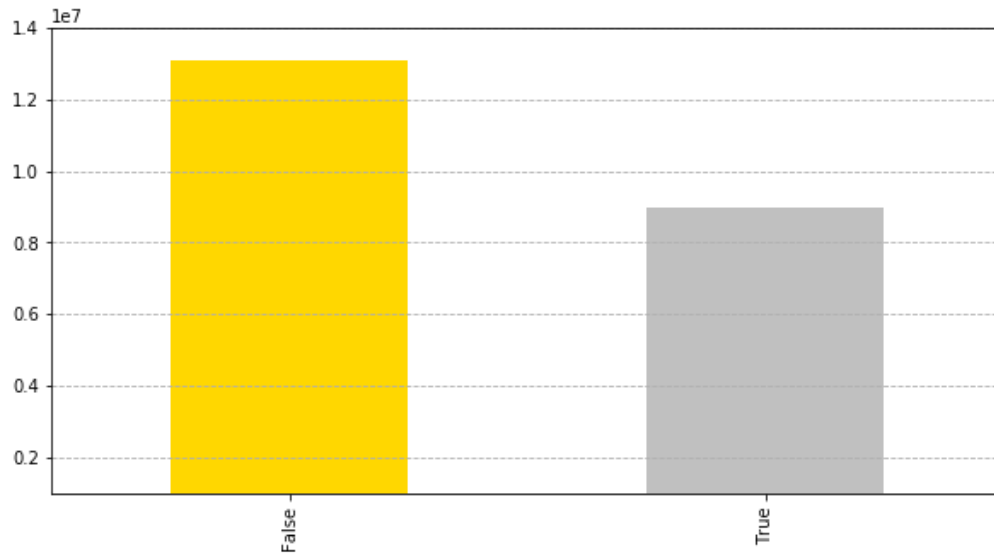
Departments distribution











Conclusion:

The objective of this project is to understand the behavior of different customer, the shopping trend and then finally recommend products to customers based on the Purchase History. I will apply different recommendation algorithm and will attempt to group Customers based on their behavior.