Capstone – Business Analysis and Recommender System(FMCG)

Ling Chong Gold

Agenda

Business Problem

Analyze and understand our customers and recommend actions to increase their sales volume

Problem Statement – Chain of Thoughts

- Gaining traction on data privacy
  - Challenge to collect customer demographic information

- Ratings from clients tends to skew towards some level of biasness
  - Circumstances leading to how the feedback is collected
  - No standard governance for ratings

- According to a study by McKinsey, 75% of what consumers watch on Netflix comes from the company's recommender system

- Amazon credit 35% of their revenue to their recommender system

- Unawareness may result in the absence of interaction between customer and a product

Source

Problem Statement – Data Analysis / Data Science

- Perform Analysis on product and customer purchase behavior to gather insights

- Perform basket analysis

- Create a recommender system with the absence of explicit feedback and customer demographic profile

Background of Data

- Data was gathered from Kaggle

- Data was split into multiple csv files with corresponding primary and foreign keys similar to a relational database

- Combined dataset about 32M rows and 15 columns
  ~ 3.4M unique orders
  ~ 200K unique users
  ~ 49K unique products

- One category of products named as missing dropped
  - 1255 unique products

Data Analysis – Purchase Behaviour

Order Counts - Day of Week and Hour of Day
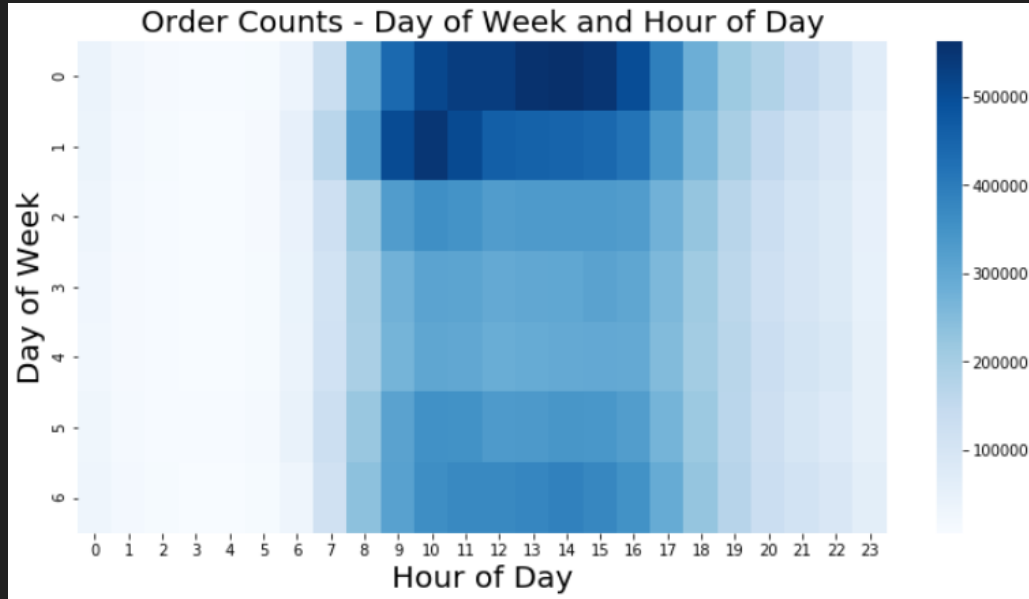
- Number of orders for day of the week and hour of the day

- Sunday (Day 0) between 9AM – 5PM

- Monday (Day 1) between 9AM – 11AM

- Marketing campaign can be strategically scheduled for maximum outreach

- System Maintenance can be performed between 1AM - 5AM

Data Analysis – Purchase Behaviour

Unique Orders Based on Days Since Prior Order
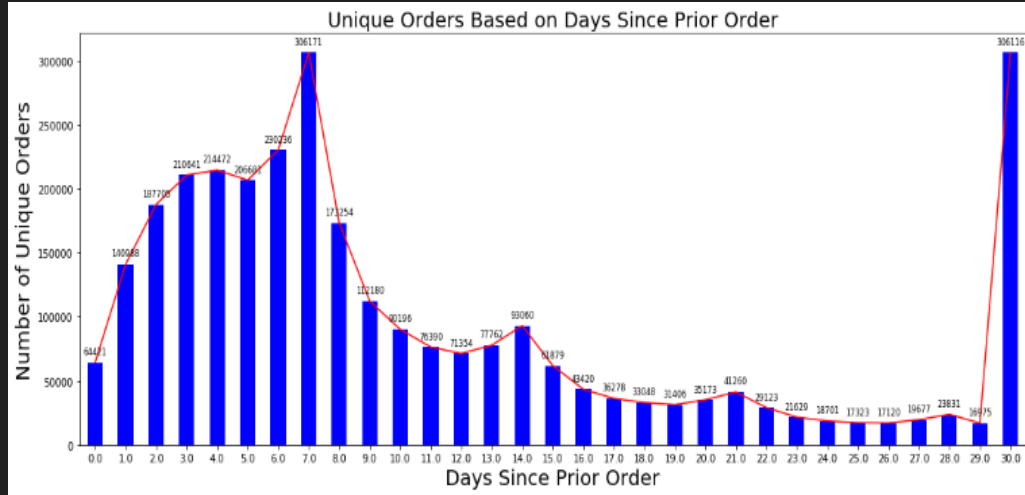
- Graph shows the number of orders and lapse in days since the last order

- Peak every 7 days

- Highest peak at day 7 and day 30

- Most repeated orders comes in within the first 7 days

- Strategic push notifications personalized to customers and be implemented

Data Analysis – Purchase Behaviour

Number of Users and Their Reordered Times
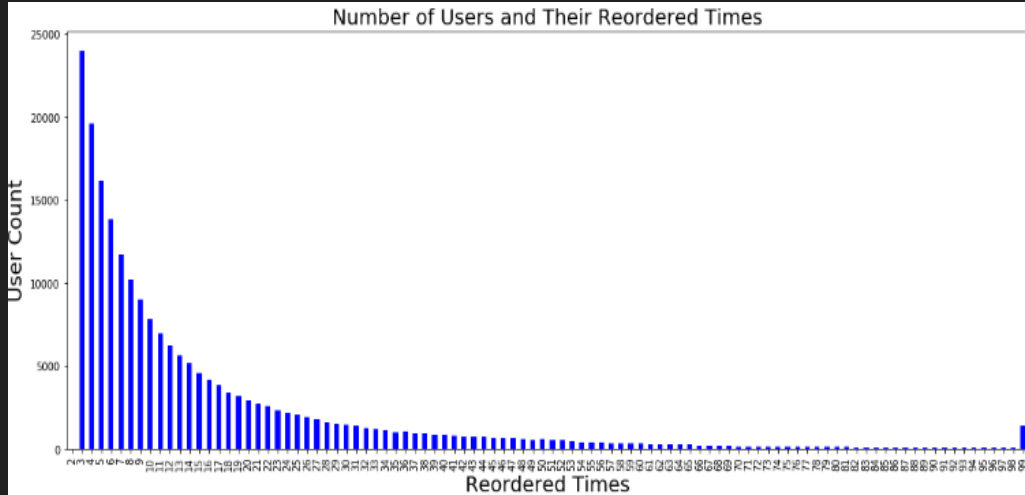
- Graph shows the number of customers and their number of reorders

- A group of loyal customers who have made 99 reorders

- Number of customers decreases as reordered times increase

- Explore on customer churn and customer retention success

Exploring Data

Relation Between Reordered and Add to Cart Sequence
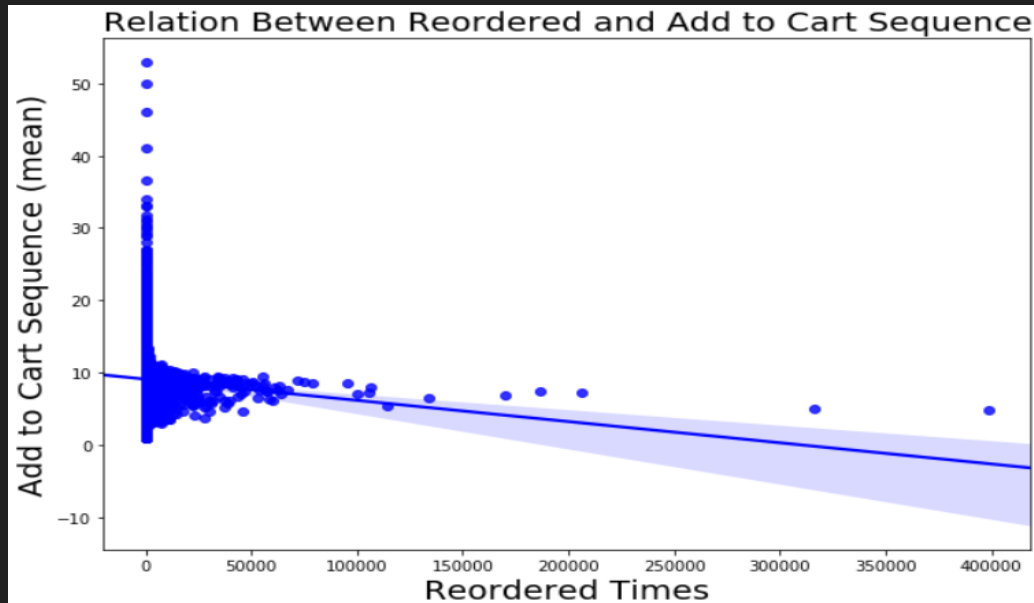
- Relation between the sequence the product is added to cart and their reordered times

- First 10 products added to cart are likely to be reordered again

- People tend to add items which they know they will buy first

- Items added later part of the cart could be to qualify for perks

Customer Segmentation - Explanation

- Relevant marketing campaigns can be targeted at individual clusters for effectiveness and cost saving

- Customers were segmented based on their purchase behavior on the sub-category of products through unsupervised learning

- PCA was utilized for feature extraction

- Elbow method utilized to determined that the best number of clusters for our dataset

- EDA was then performed on the clusters to gather insights for each cluster

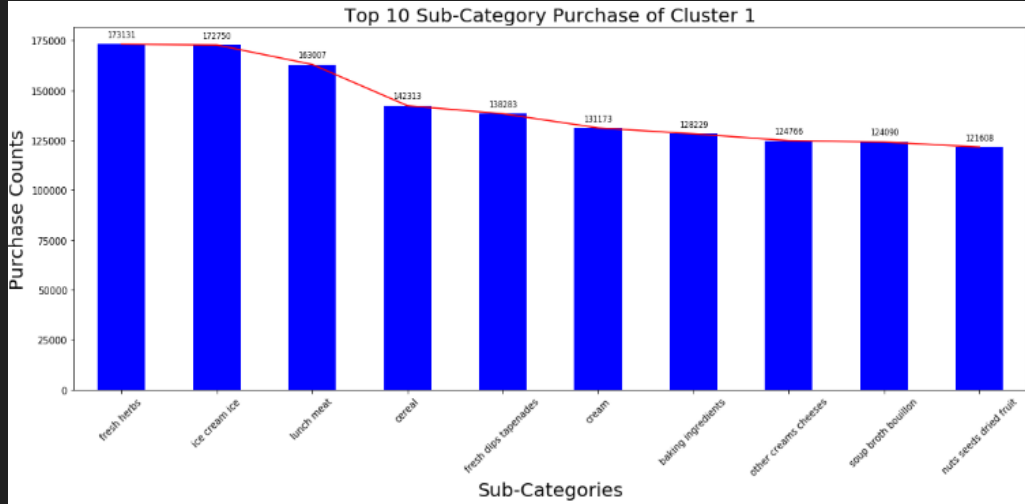# Customer Segmentation - Overview
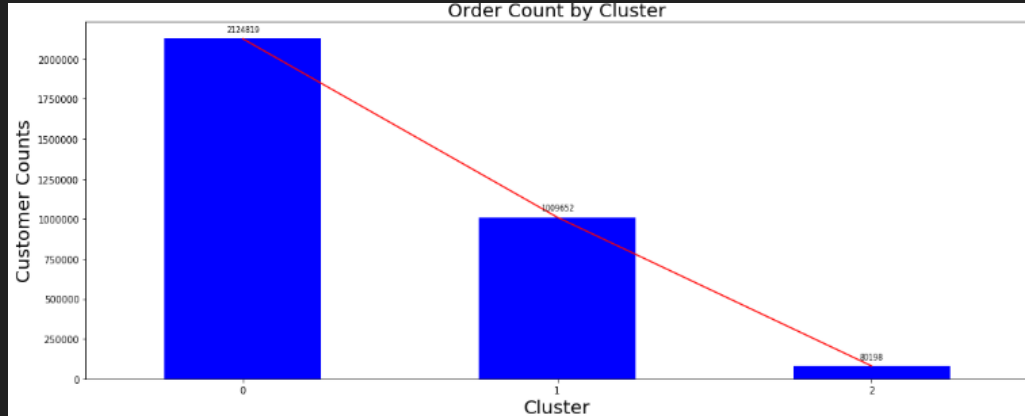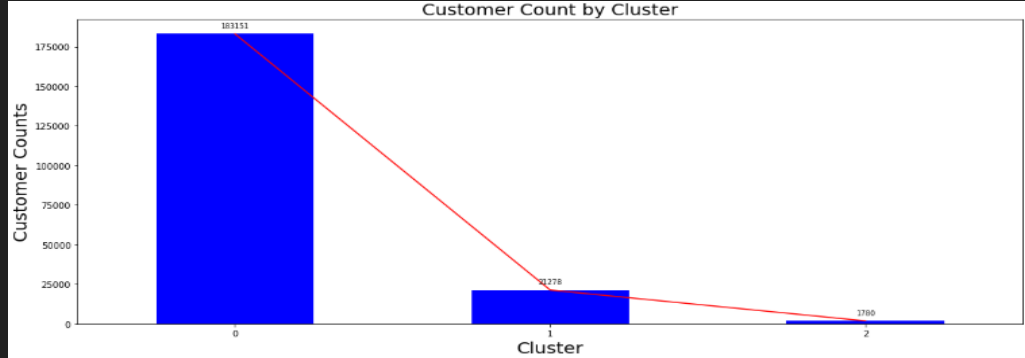
- Plot shows the number of customers for each cluster, and number of orders for each clusters.
  - Cluster 0 – 11.5 orders per customer
  - Cluster 1 – 47.6 orders per customer
  - Cluster 2 – 45 orders per customer

- Cluster 0 could contain most of the one time off customers

- Marketing and promotional efforts focusing on cluster 1 and 2 will improve customer retention

Customer Segmentation – Cluster 0

Top 10 Sub-Category Purchase of Cluster 0
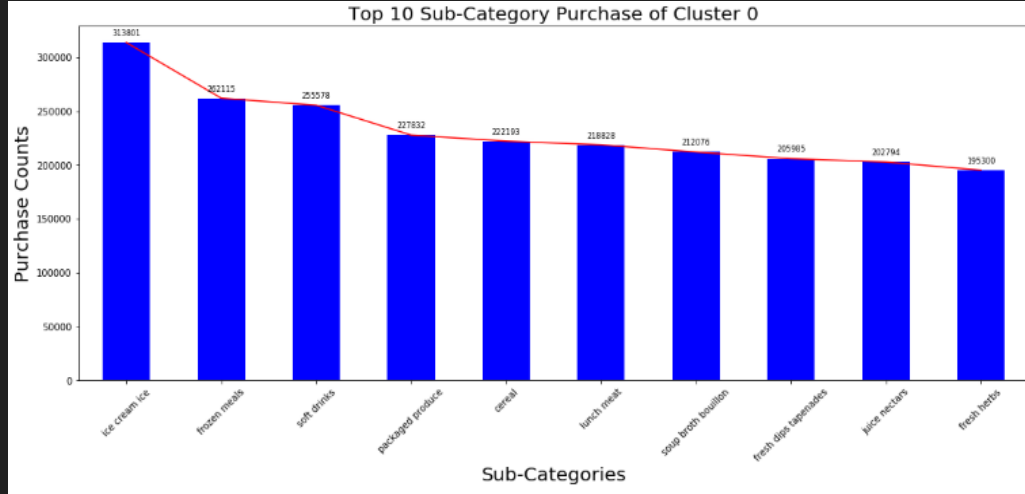
- Top sub-categories of products includes frozen meals, soft drink and packaged produces

- Promotional items related to convenience can be targeted at this cluster

- We can have easy to cook recipes with products from our grocery store made available in our site targeting at this clusters

Customer Segmentation – Cluster 1

Top 10 Sub-Category Purchase of Cluster 1

- Absence of ready-to-eat meals in this cluster's top purchase

- Fresh items are among the top purchases

- Membership perks like fixed routine delivery can be targeted at this cluster

Customer Segmentation – Cluster 2

Top 10 Sub-Category Purchase of Cluster 2
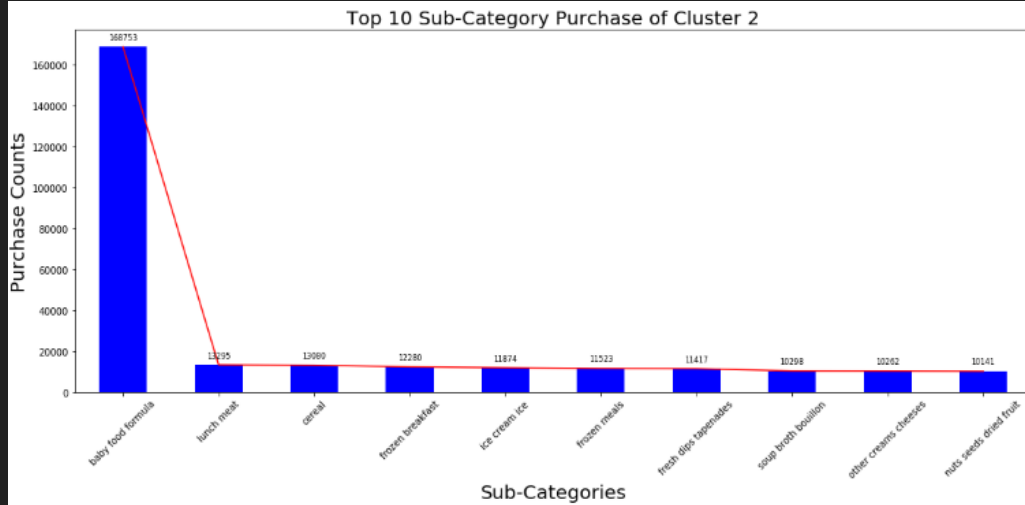
- High purchase count of baby food formula

- Baby related promotions or products can be targeted at this cluster

- Provide infant care tips with products from our stores to this cluster

Association Rules - Explanation

- Captures patterns of items appearing frequently together

- Strength of association
  - Support – Number of times the item appear out of the total transaction
  - Confidence – Likelihood that item B is bought when A is bought
  - Lift - The increase in ratio of the sale of B when A is sold

- Use cases
  - Recommender System – Recommend items upon adding to cart
  - New product – Product Y in flavors of X
  - Health Care – Symptoms and their associated disease
  - Fraud Detection – Fraudulent transaction and their associated behaviors
  - Education – Students facing difficulties in a topic will likely face the same in another

Association Rules - Evaluation

	Antecedent	Consequents	Support	Confidence	Lift
253	Bag of Organic Bananas	Organic Strawberries	0.00180	0.23579	2.73546
140	Organic Strawberries	Organic Reduced Fat Milk	0.00232	0.20025	2.32326
150	Organic Strawberries	Whipped Cream Cheese	0.00209	0.20575	2.38702
412	Organic Whole Milk	Organic Strawberries	0.00135	0.27914	3.23843
397	Organic Lemon	Organic Hass Avocado	0.00103	0.29824	4.28357

We can consider introducing Organic Reduced Fat Milk in Strawberry flavor to the store

- Organic Strawberries appeared in 0.23% of our total transaction

- People are 20% more likely to buy Organic Reduced Fat Milk when they purchase Organic Strawberries

- People are 2 times more likely to buy Organic Reduced Fat Milk and Organic Strawberries compared to just buying Organic Strawberries

Recommender System - Explanation

- Purely leveraging on implicit data gathered through customer's purchase behavior

- Interaction between customer and item is a basis of how our recommender system works

- An absence of interaction could mean that the customer do not like the item or more often, the customer do not know about the item yet

- A good recommender system is able to identify features a user like based on their past behavior and behavior of similar users and matching them with products that have these features.

Recommender System - Evaluation

```
Recommended items for User 1 are:

13517  Whole Wheat Bread 1.1968588
20063  Hazelnuts in Milk Chocolate, 33% Cocoa 1.1774435
26853  Complete Wheat 100% Whole Wheat Bread 1.1398611
15487  Raspberry English Tea Scones 1.1145719
================================================================
User 1 validation transactions are:

                     product_name  user_id
                             Soda        1
             Organic String Cheese        1
          0% Greek Strained Yogurt        1
    XL Pick-A-Size Paper Towel Rolls       1
             Milk Chocolate Almonds        1
                        Pistachios        1
             Cinnamon Toast Crunch        1
        Aged White Cheddar Popcorn        1
               Organic Whole Milk        1
               Organic Half & Half        1
                  Zero Calorie Cola        1
```

Good Recommendations:
- Recommended Hazelnut in Milk Chocolate
- Bought Milk Chocolate Almonds

Rooms for Improvements:
- Recommended bread twice

Recommender System - Evaluation

```
Recommended items for User 58144 are:

34172 Top Ramen Shrimp Flavor Instant Noodle Soup 1.1561155
39322 Caramel Almond and Sea Salt Nut Bar 1.1357532
35175 Mini Stuffers Hamburger Dill Chips 1.1336474
19604 Medium Scarlet Raspberries 1.0747831
============================================================
User 58144 validation transactions are:

                                product_name    user_id
                 Electrolyte Enhanced Water      58144
                                    Banana       58144
 Air Chilled Organic Boneless Skinless Chicken ...  58144
                    Lime Sparkling Water          58144
              Non Fat Raspberry Yogurt            58144
                         Farfalle No. 93          58144
       Total 0% Nonfat Plain Greek Yogurt         58144
                    Original Orange Juice         58144
            Best Sloppy Joe Skillet Sauce         58144
           Organic Cauliflower Florets            58144
          Grated Parmigiano Reggiano              58144
                       Whole Almonds              58144
```

Good Recommendations:
- Recommended Raspberries
- Bought Banana
- Bought Raspberry Yoghurt

Recommender System - Evaluation

```
Recommended items for User 114401 are:

44898 Organic Mac And Trees Fun Shape Macaroni & Cheese 1.1370347
35488 Organic Dry Roasted Premium Flaxseed 1.1338583
2190 Spicy Red Lentil Sauce 1.1187676
21702 Puna Coconut Pineapple 1.1060191
=============================================================
User 114401 validation transactions are:

                            product_name   user_id
                              Whole Milk    114401
No Pulp Calcium & Vitamin D Pure Orange Juice   114401
             Original Fresh Stack Crackers   114401
                  Cheddar Broccoli Rice    114401
                       Corn Pops Cereal    114401
                  Eggo Strawberry Waffles   114401
    Original 100% Pure No Pulp Orange Juice   114401
                      Orange Juice To-Go    114401
             All Natural Peach Tea Bottles   114401
                    Hickory Smoked Bacon    114401
```

Good Recommendations:
- Recommended Puna Coconut Pineapple
- Bought Juices

Recommender System - Evaluation

```
Recommended items for User 3754 are:

33502 Double Cheese Baked Snack Mix 1.1659267
45339 Men's Refresh Dandruff Shampoo 1.060647
29642 Ultra Soft Bath Tissue 1.0497061
13810 Reclosable Gallon Freezer Bags 1.0313741
==============================================================
User 3754 validation transactions are:

                                 product_name   user_id
                       Twice Baked Potatoes      3754
                       Whipped Sweet Potatoes    3754
100% Natural Skin & Hair Revitalizing Coconut Oil   3754
```

Good Recommendations:
- Shampoo
- Bought Skin and Hair product

Recommender System - Evaluation

```
Recommended items for User 200372 are:

30890 MCT Oil 1.26377
8651 Shipping Packaging Tape Heavy Duty 1.2522316
17419 Sprouted Whole Wheat Bread 1.209813
17018 Ghee Vanilla Bean 1.1393975
============================================================
User 200372 validation transactions are:

                 product_name  user_id
                    Diet Cola   200372
         Original Potato Chips   200372
   Salsa Con Queso Medium Dip   200372
          Pure Sport Body Wash   200372
           Snickers Ice Cream   200372
    Raspberry Cheesecake Gelato   200372
                      Rosemary   200372
                  Red Potatoes   200372
    2% Low Fat Cottage Cheese   200372
```

Good Recommendations:
- None

Reasons for Recommendation:
- Recommended Ghee Vanilla Bean
- Previous orders includes a variety of bread spread

- Recommended MCT Oil
- Previous orders includes supplements and energy drinks

Further Improvements

- Additional customers insights can be obtained with more data
  - Recency
  - Frequency
  - Monetary

- Evaluate the effectiveness of the recommender system after deployment and perform the next iteration of improvement

- Add features to products during the next iteration.
  e.g. organic, natural, convenient, fresh, manufacturer etc

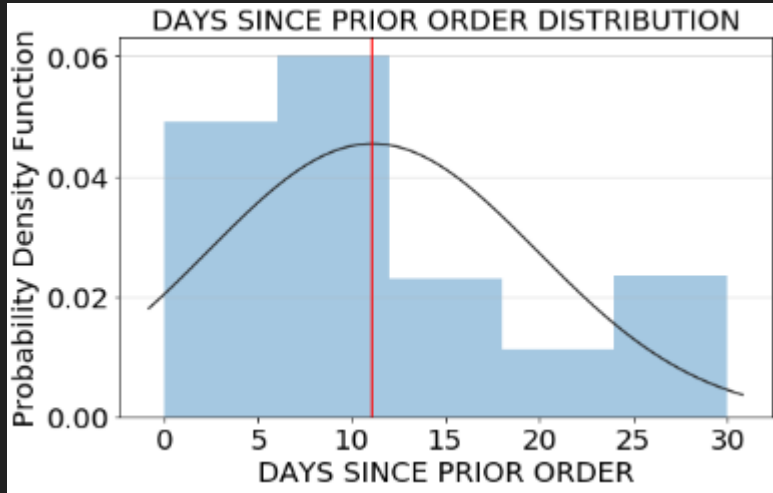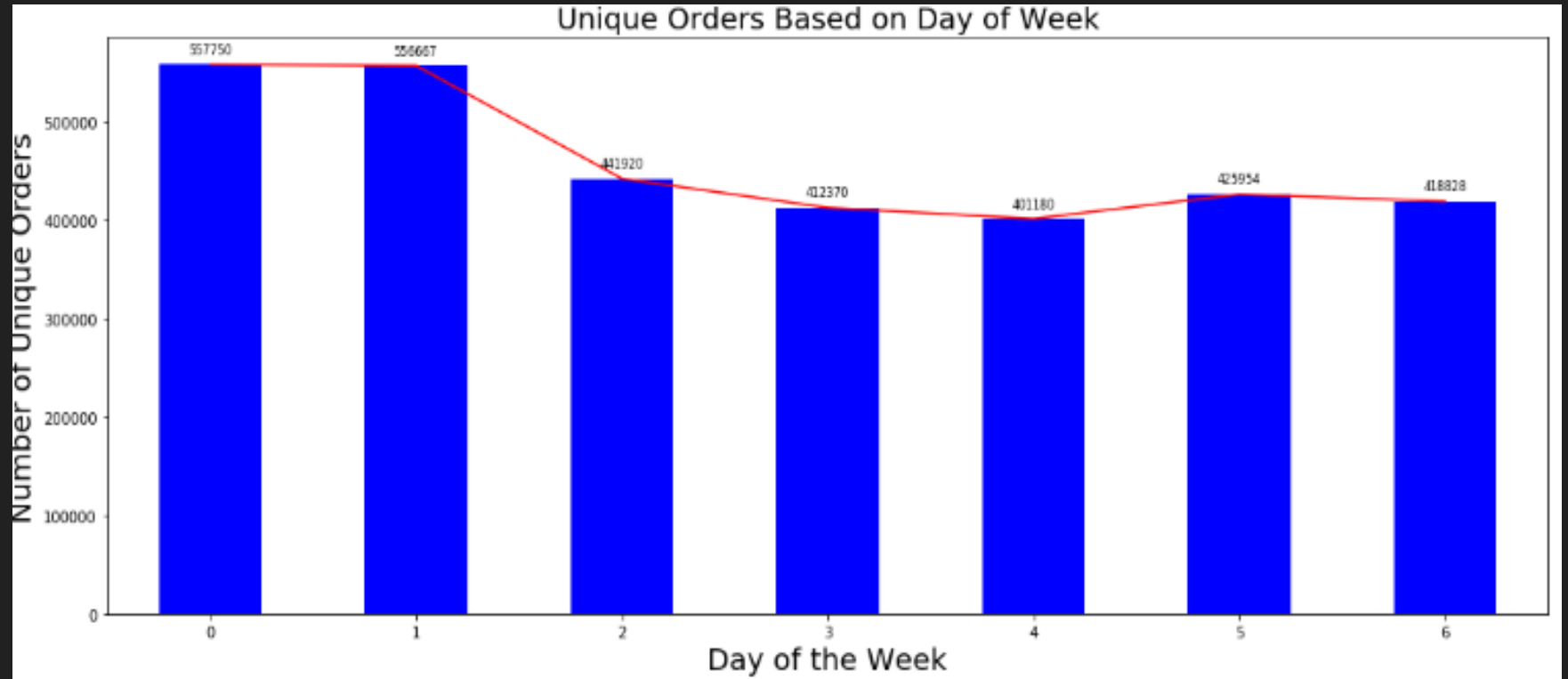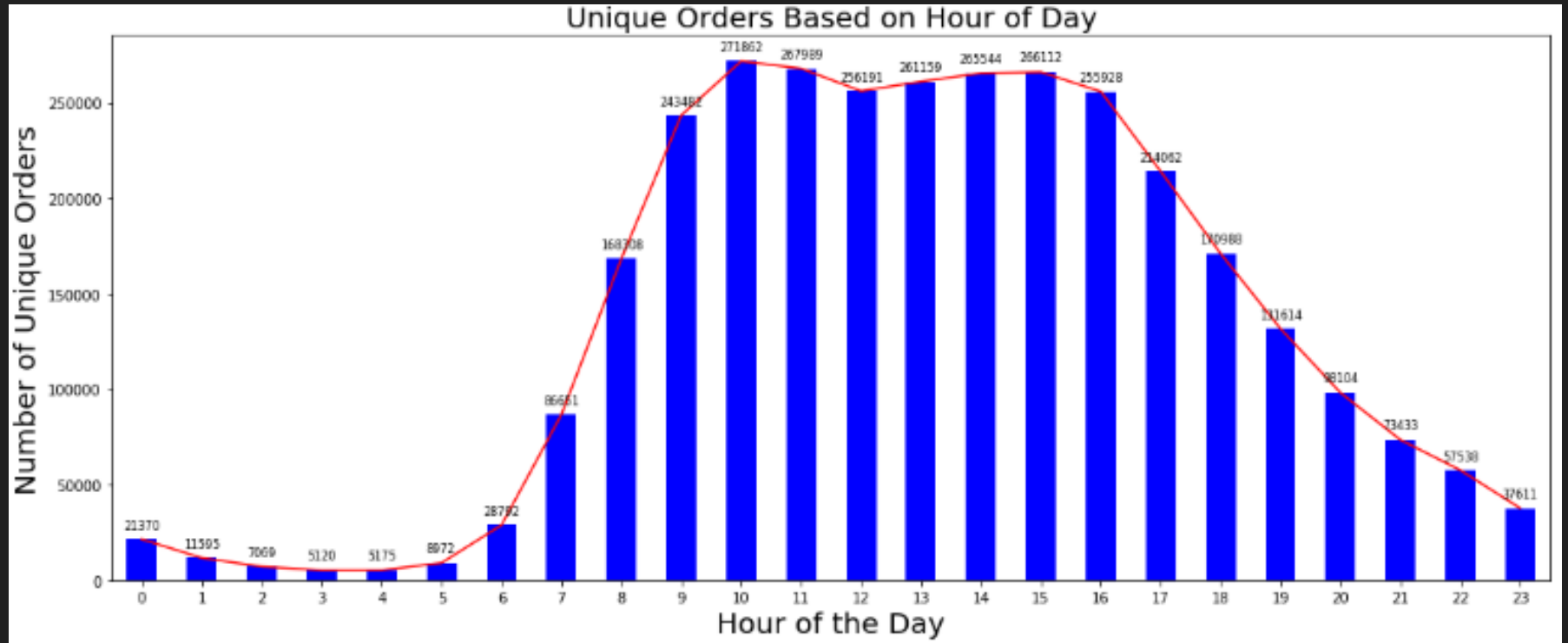- Using neural network Sequential to predict customer's next purchase and perform through association rules

Q & A

Backup Slides

Backup Slides
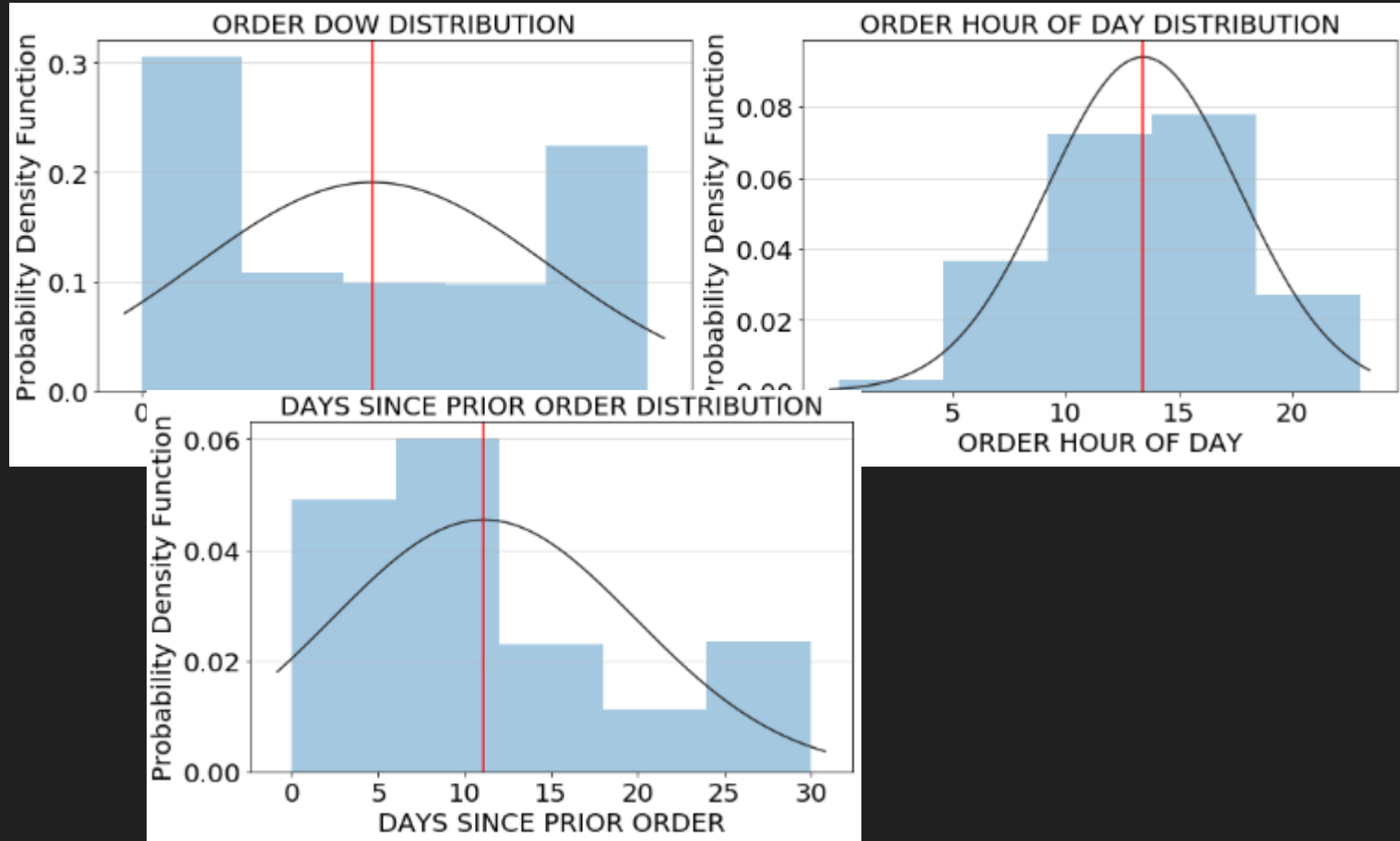
Backup Slides

DAYS SINCE PRIOR ORDER DISTRIBUTION

Good Recommendations:
- None

Reasons for Recommendation:
- Recommended Ghee Vanilla Bean
- Previous orders includes a variety of bread spread

- Recommended MCT Oil
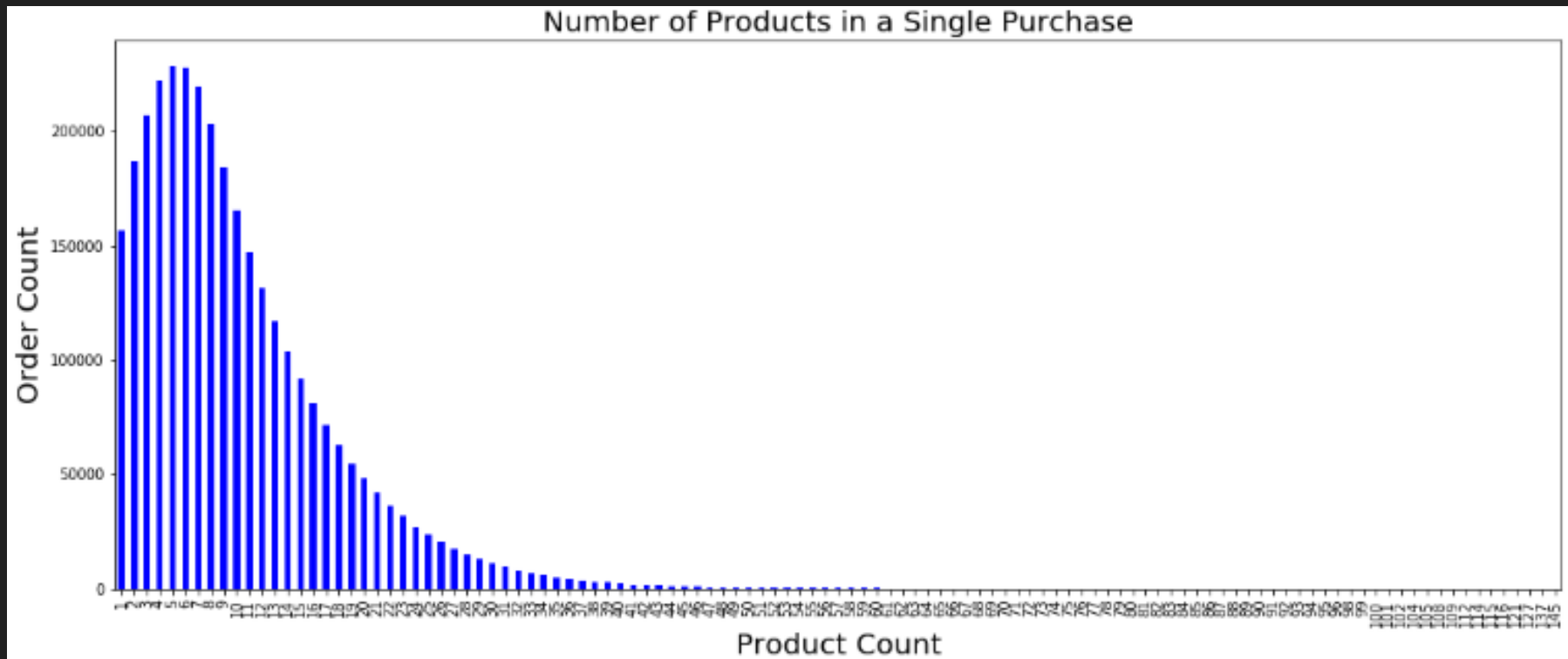- Previous orders includes supplements and energy drinks
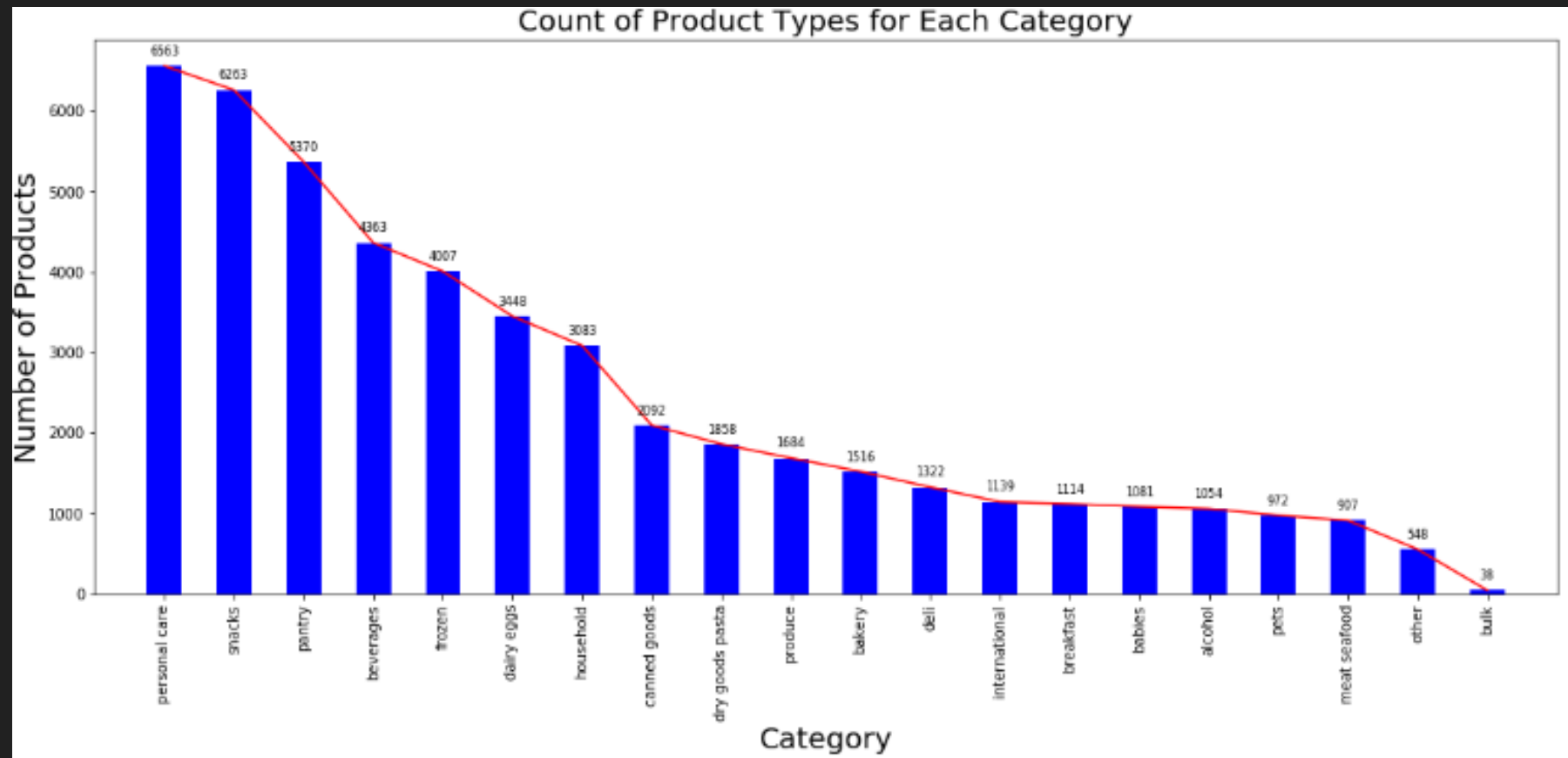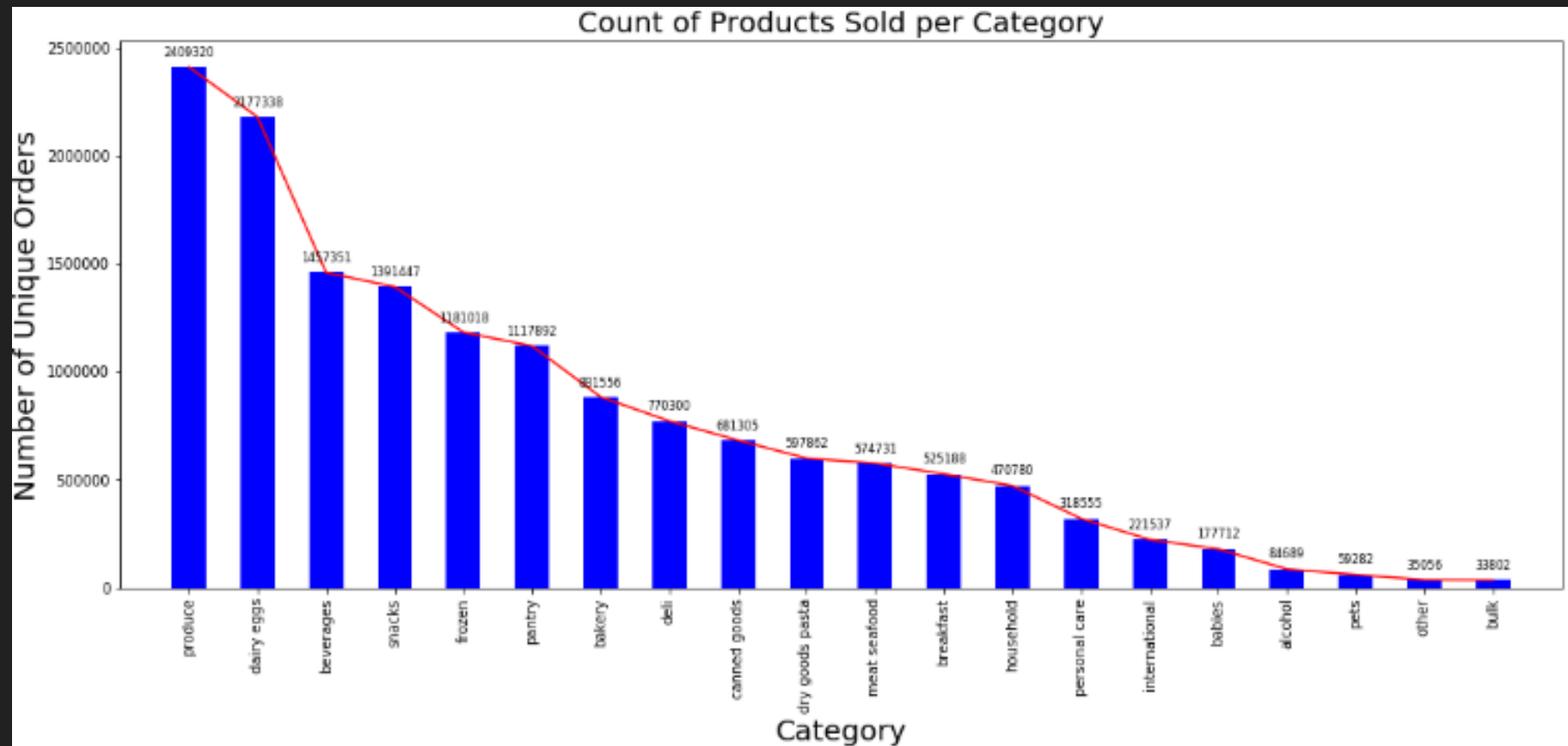
Backup Slides

Unique Orders Based on Day of Week

Backup Slides

Unique Orders Based on Hour of Day

Backup Slides

Number of Products in a Single Purchase

Backup Slides

Count of Product Types for Each Category

Backup Slides

Count of Products Sold per Category

Backup Slides

Category and Their Reordered Times

Bar chart showing reordered times by category:

Category	Reordered Times
produce	6160710
dairy eggs	3627221
beverages	1757892
snacks	1657973
frozen	1211890
bakery	739188
pantry	650301
deli	638864
canned goods	488535
meat seafood	402442
dry goods pasta	399581
breakfast	398013
household	297075
babies	243369
personal care	143584
international	99416
alcohol	87595
pets	58760
bulk	19950
other	14806

Backup Slides

Top 21 Sub-Categories Based on Product Sold

Backup Slides

Top 21 Sub-Categories Based on Reordered Times
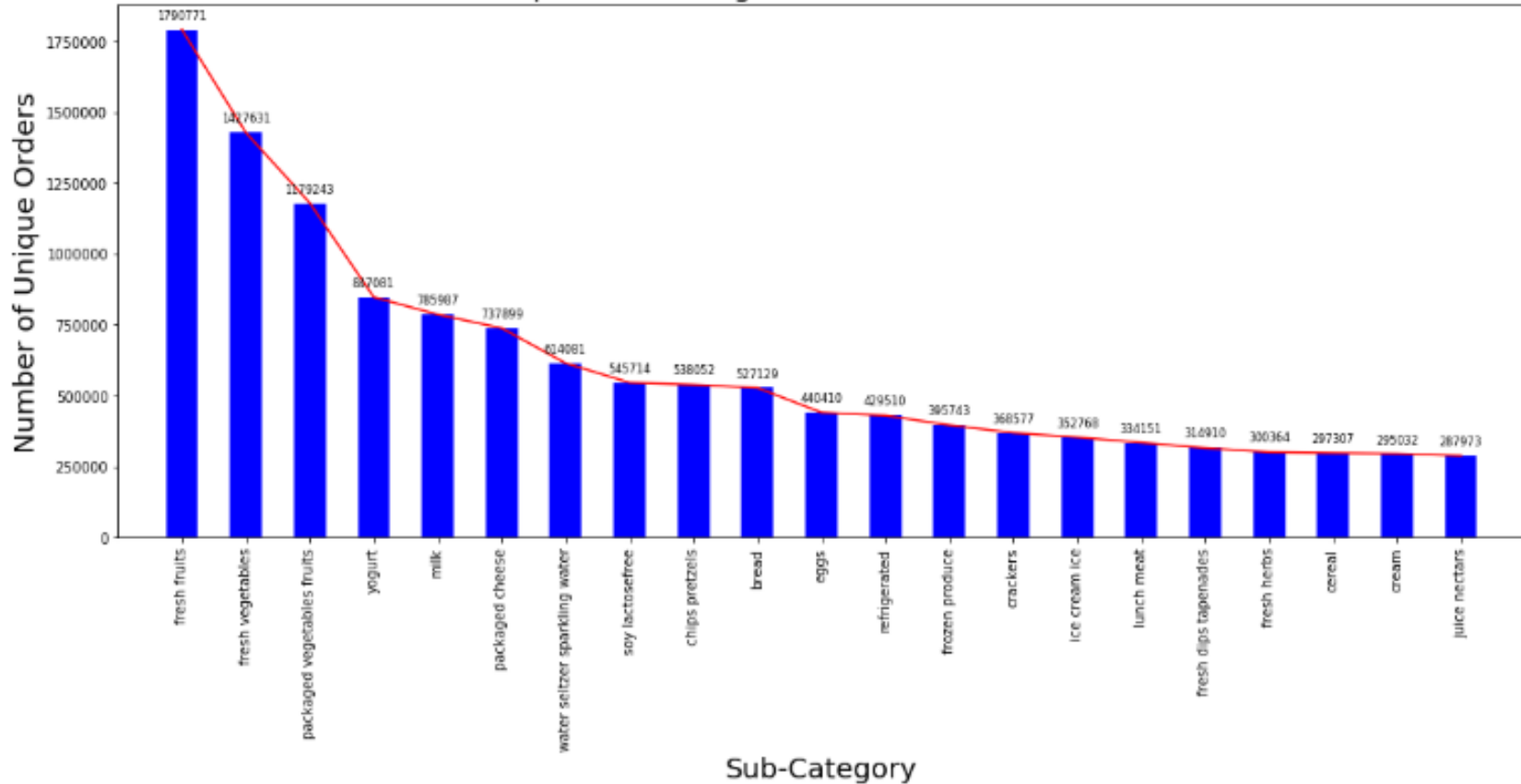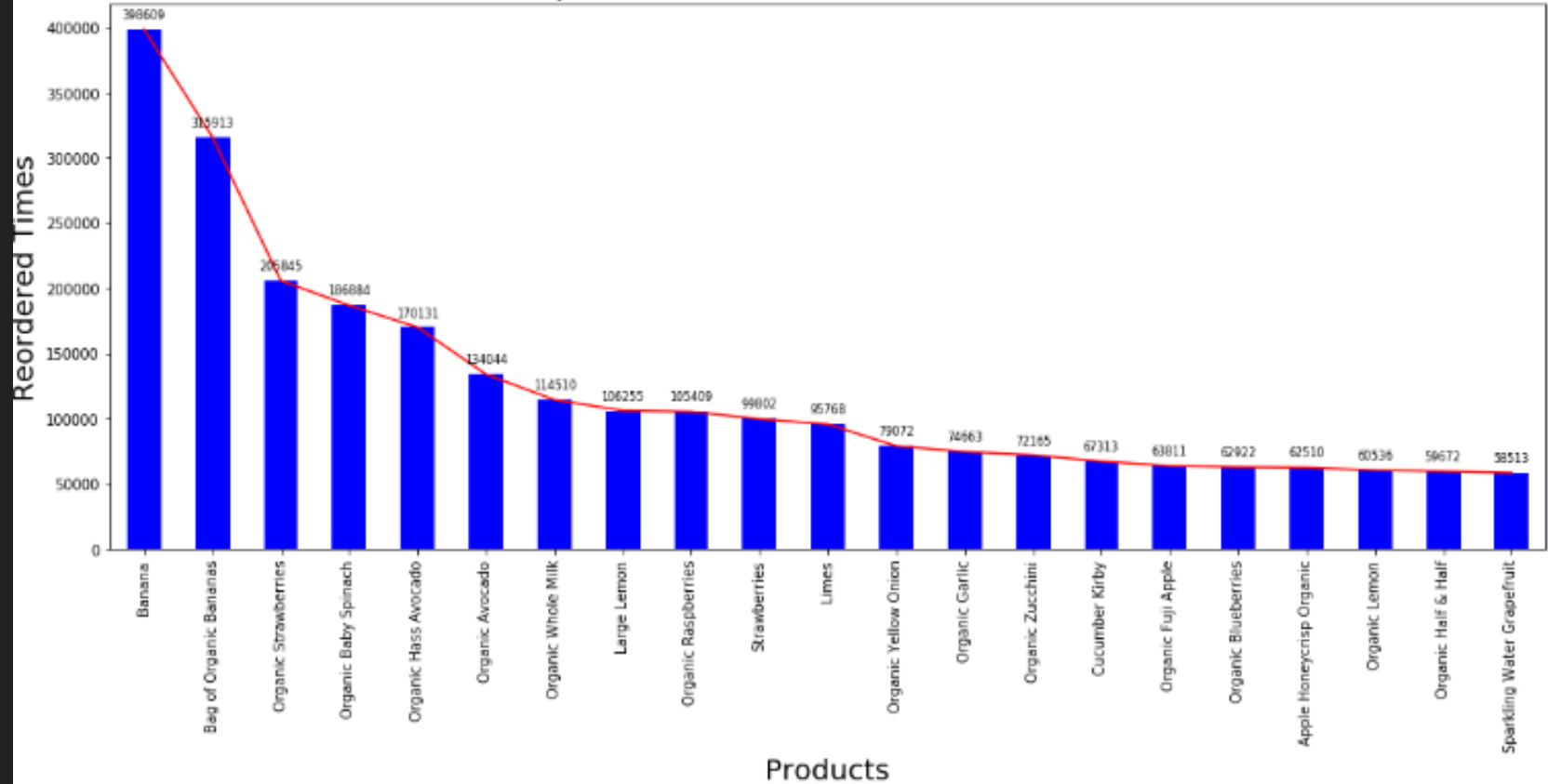
Backup Slides
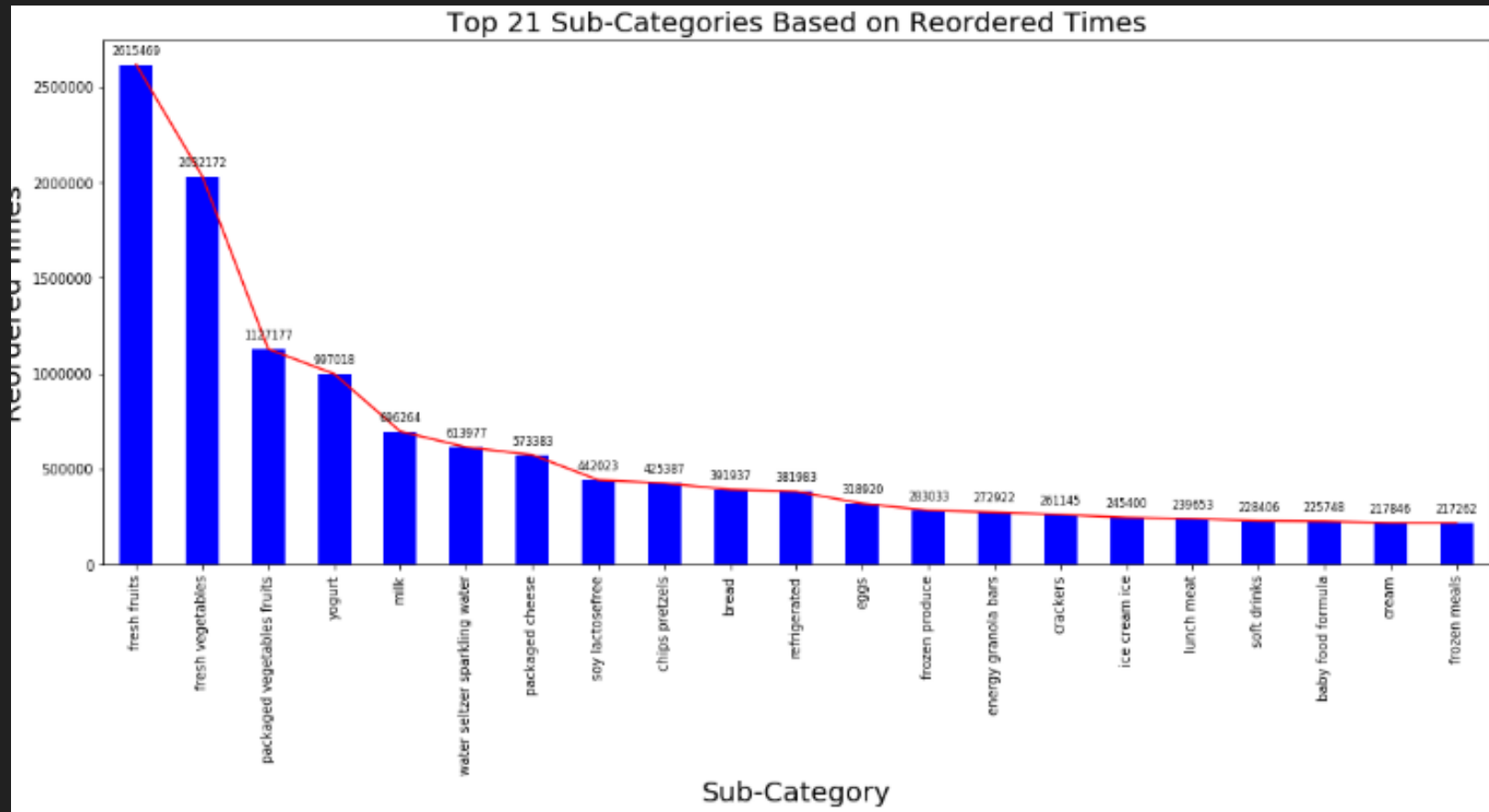
Top 21 Products Sold
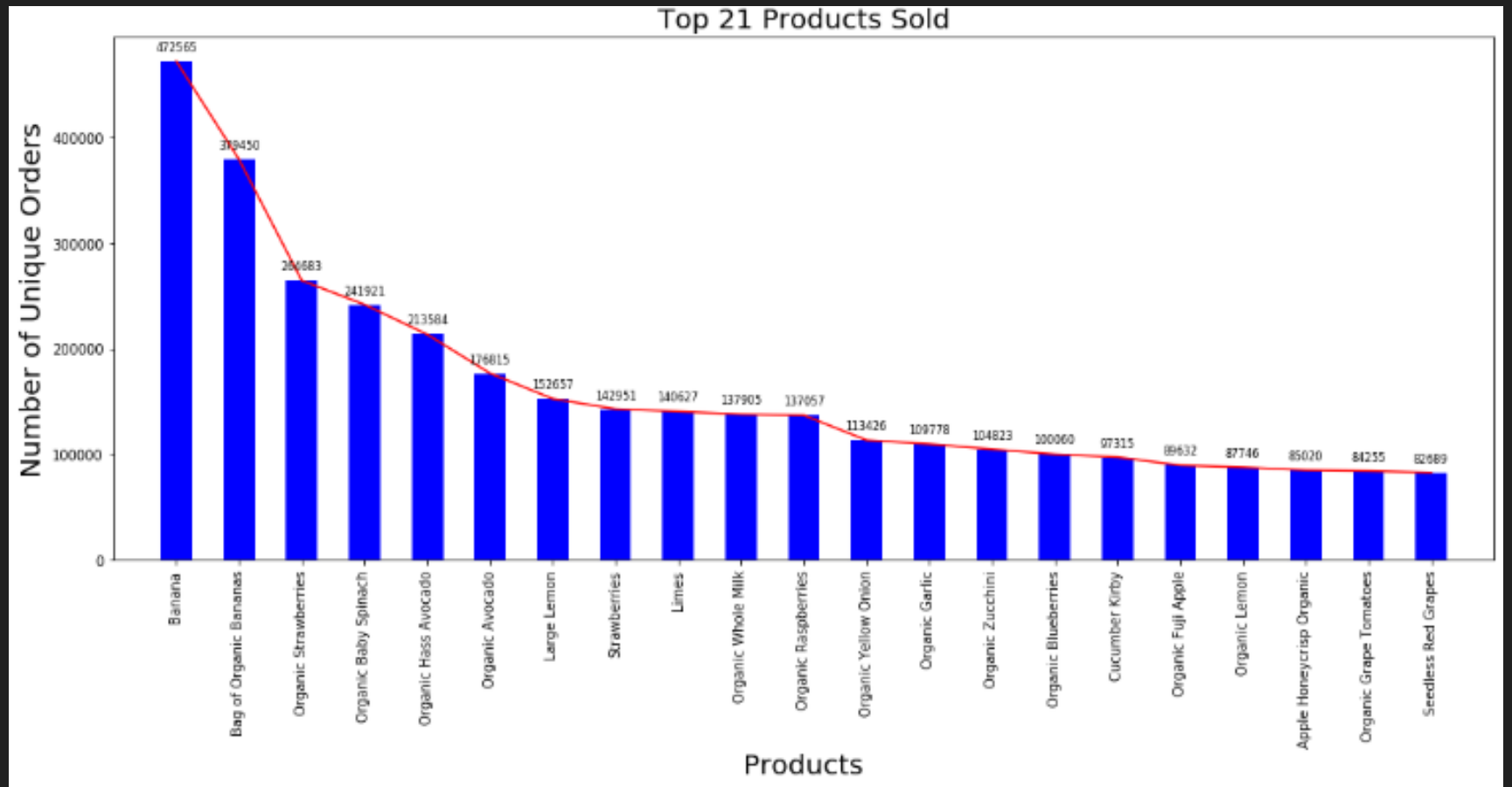
Backup Slides

Top 21 Product Based on Reordered Times

Backup Slides