**Instacart Market Basket Analysis - Kaggle**

#### Problem Description

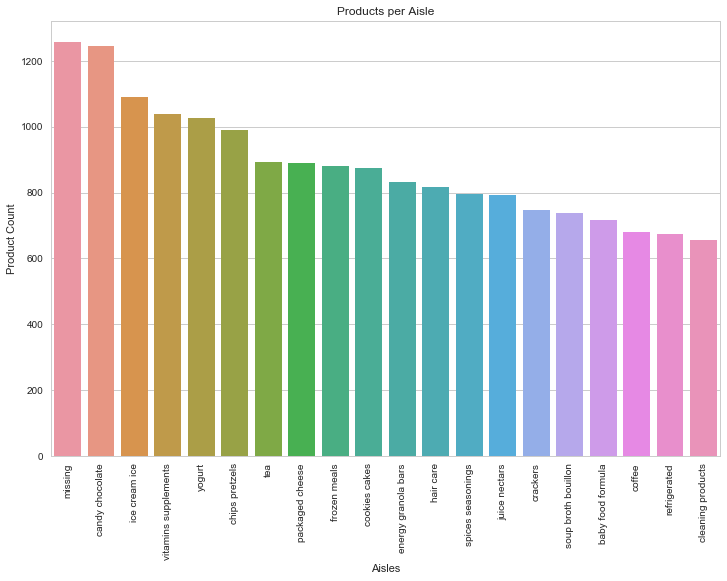
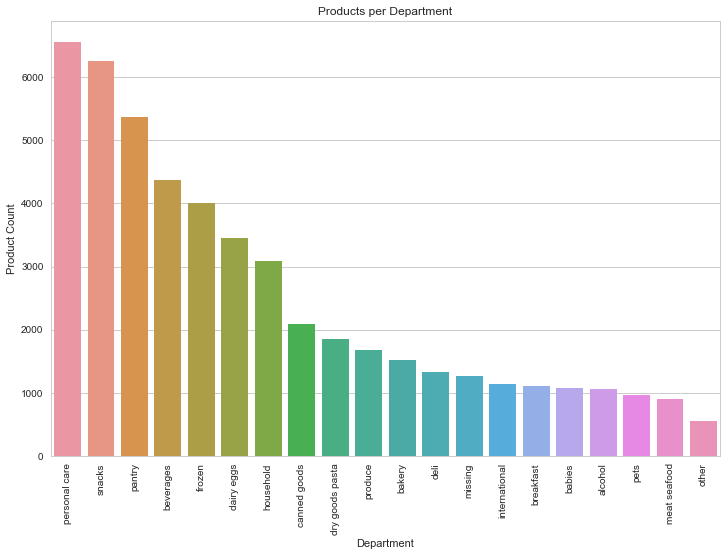
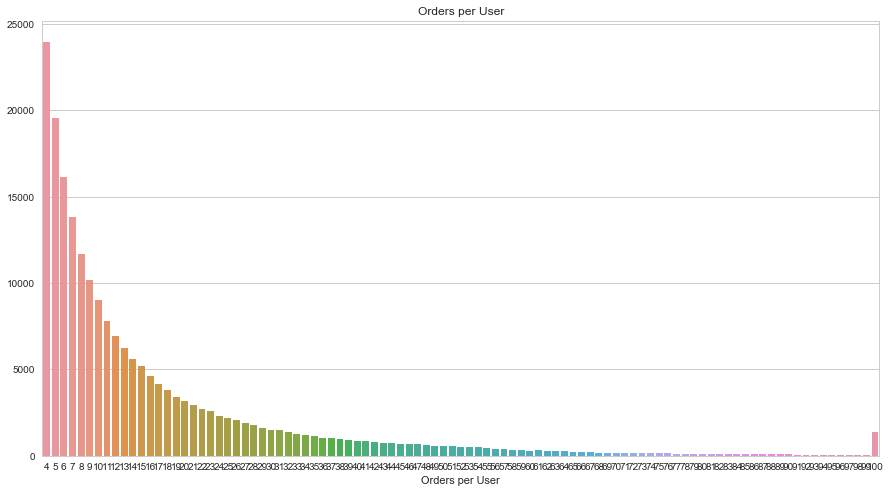
Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you.

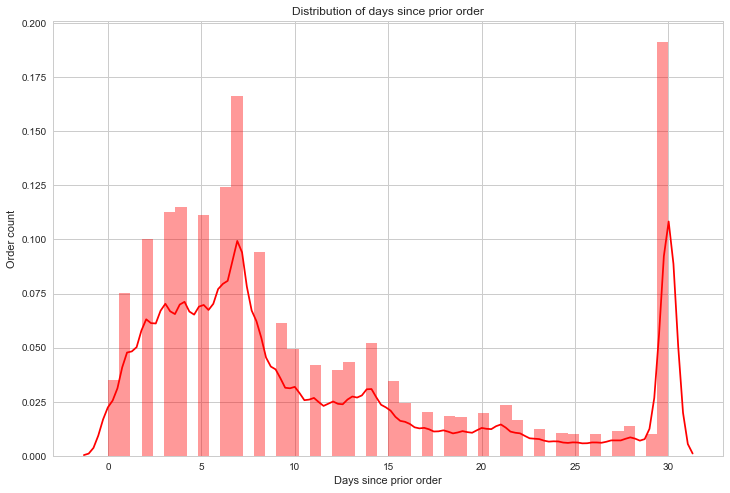
Use this anonymized data on customer orders over time to predict which previously purchased products will be in a user’s next order.

Instacart has provided over 3 million orders to analyze over 206,209 users. The dataset is split into prior orders and train-test orders. All the orders except the last order per user is in the prior orders. The last order per user is split into 131,209 training orders and 75,000 test orders.

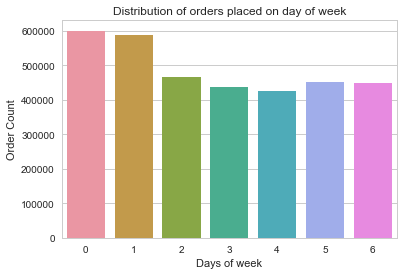
Order details include product info, department info, aisle info, order time, reorder and not reordered, item order in cart etc.

The following trends were identified on initial Exploratory Data Analysis

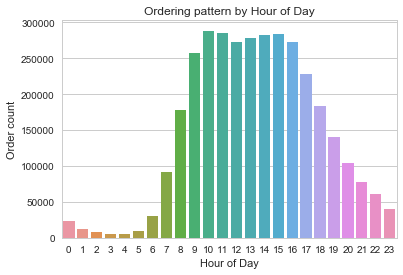
1. There is a high concentration of products in candy chocolate, ice cream ice, vitamin supplements, yogurt, chips pretzels aisle.
2. Departments Personal Care and snacks have the highest concentration of product names.
3. The number of orders per user lie in the range of 4 to 100 in this dataset. The histogram shows that the order dataset is positively skewed with users having anywhere between 4 orders per user to 100 orders per user.
4. There is a clear reordering pattern. A high number of orders are placed every 30 days. These are probably monthly items. There is also a spike at 7 days, 14 days and 21 days indicating a weekly reordering pattern.

This is an indicator of items that will be reordered.

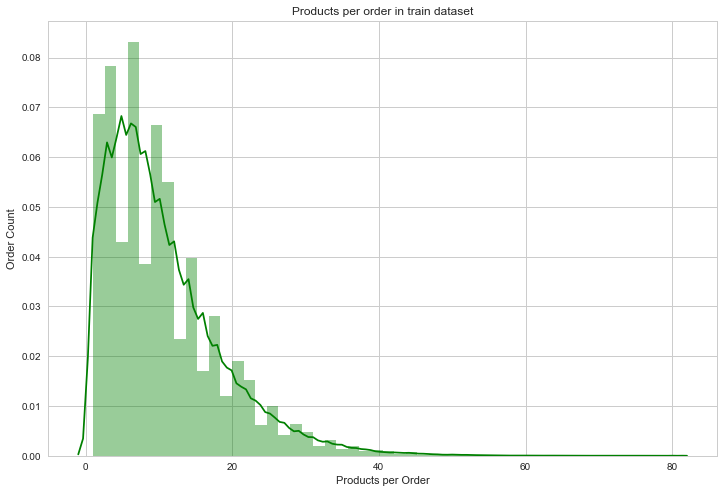
1. The plot shows a higher rate of orders are placed on Saturdays(0) and Sundays(1).

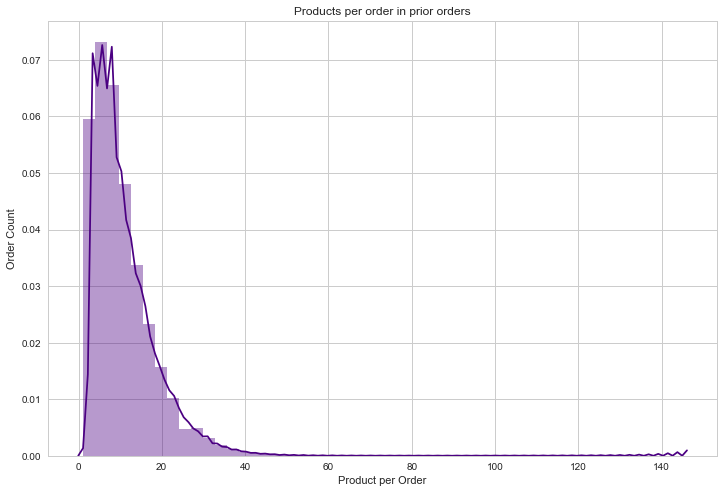


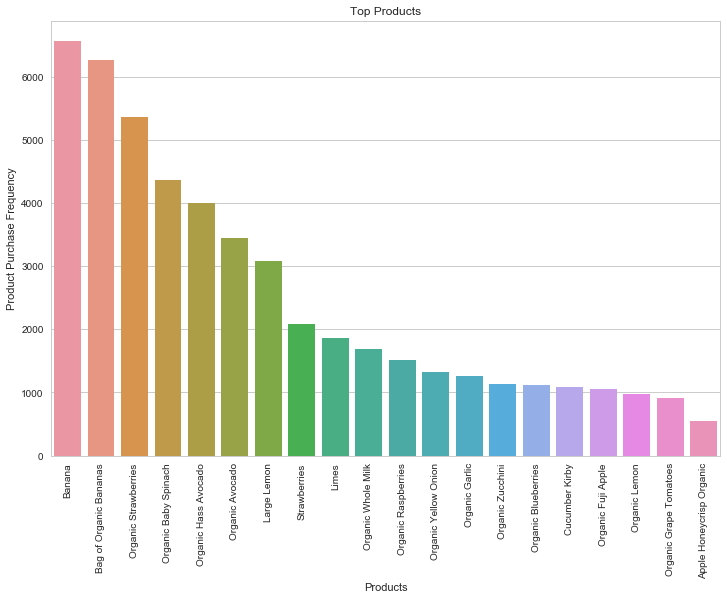
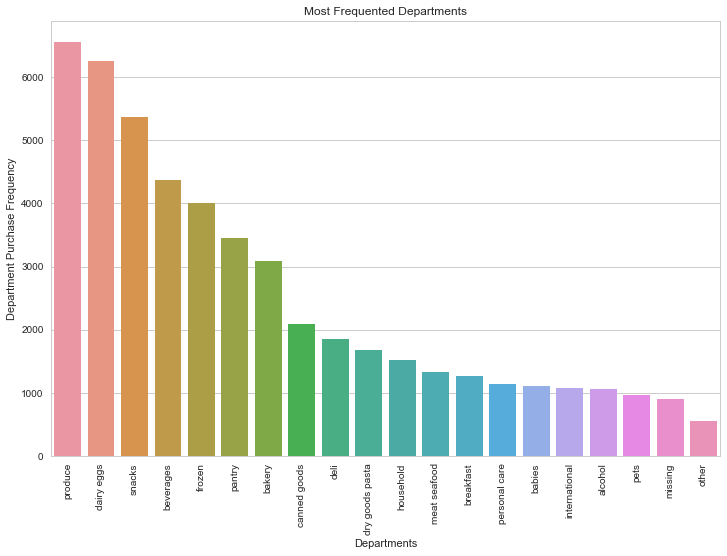
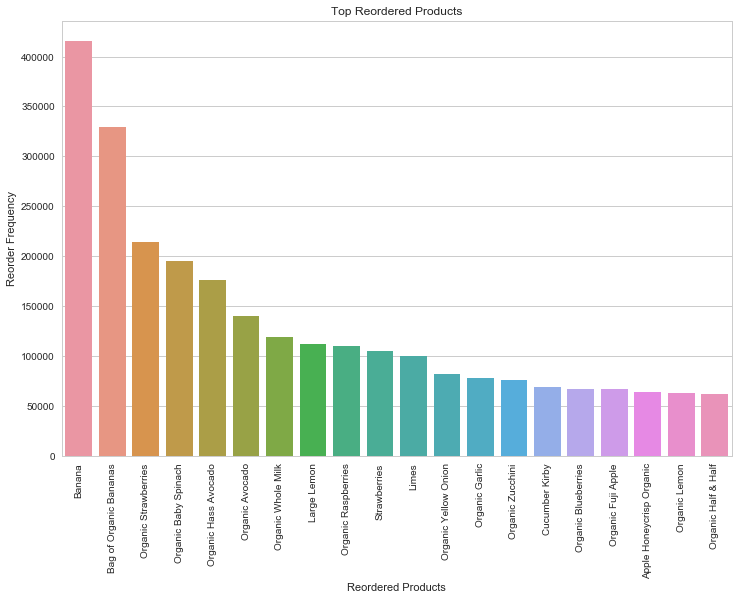
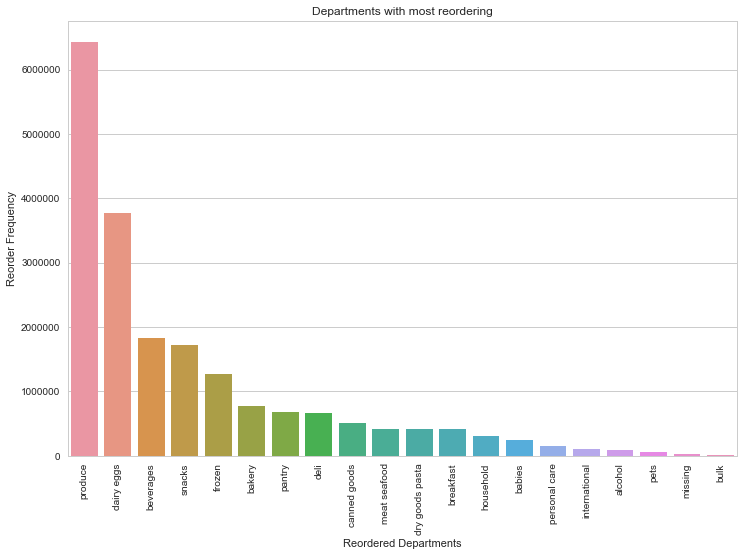
1. Most orders are placed between 8 AM and 6 PM.



1. Product count in each order shows that most orders have less than 20 products included.





1. The top products sold are
2. Top Departments are
3. Top Reordered Products are
4. Top reordered departments

This concludes the exploratory data analysis. Further patterns and correlations will be researched as part of model building.