Data Storytelling
Capstone Project 1

Raviteja Bodla
Data Science Career Track
Springboard

INTRODUCTION:

This is the second report for my capstone project as a part of 'Data Science Career track' Bootcamp with Springboard. This report describes the data story telling techniques used as a part of this project. This objective of the capstone project is to predict what the customer will buy in the next order based on previous buying behaviour using Instacart market basket analysis Kaggle competition dataset. These objectives of this particular report are to find:

- Weekday with highest number of orders.
- Hours of the day with highest number of orders.
- Aisles with highest number of orders.
- Departments with highest number of orders.
- Products with highest number of orders.
- Top reordered products.
- Is there a relationship between 'added to the cart sequence' and 'reorder percentage'?
- Departments with highest reorder rate.
- Is there a relationship between 'organic' foods and reorder rate?

METHODOLOGY:

A clear explanation of data import and data cleaning was given in first data wrangling report.
For the purpose of data storytelling, Matplotlib and Seaborn were used. All the datasets (aisles.csv, departments.csv, orders.csv, order_products_prior.csv, order_products_train.csv, products.csv) were merged using pandas.merge() function.
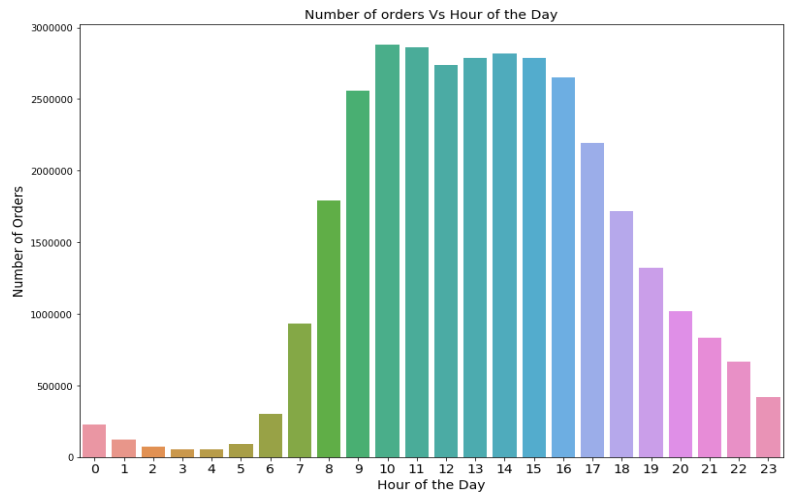
RESULTS:
1)Number of orders per day of the week:

-Number of orders are high on Sunday, which gradually decreased till Thursday and again increases towards Friday and Saturday.
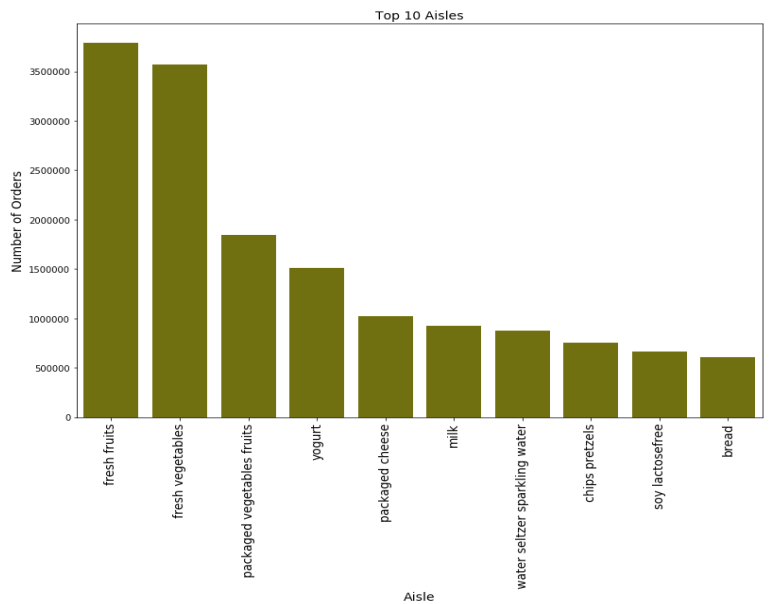

Number of orders Vs Day of the Week
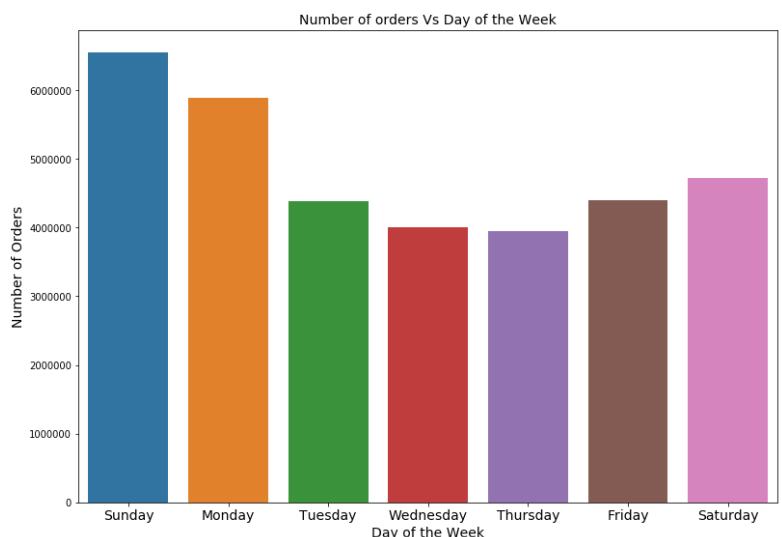
2) Number of orders per hour of the day:

-Number of orders increased from 6 AM until 10 Am and stayed high until 4PM and gradually decreased towards the night
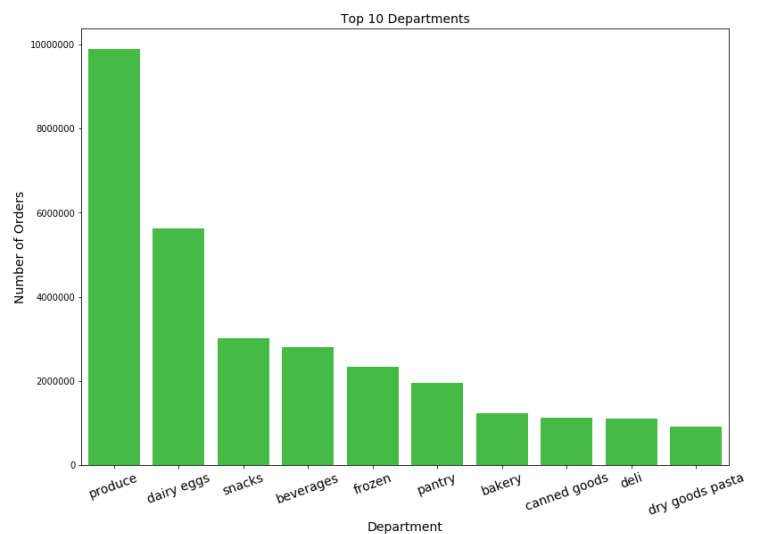

3)Top 10 Aisles:

-'Fresh fruits' is the top Aisle with highest orders followed by 'Fresh vegetables' and 'packaged vegetables fruits'.
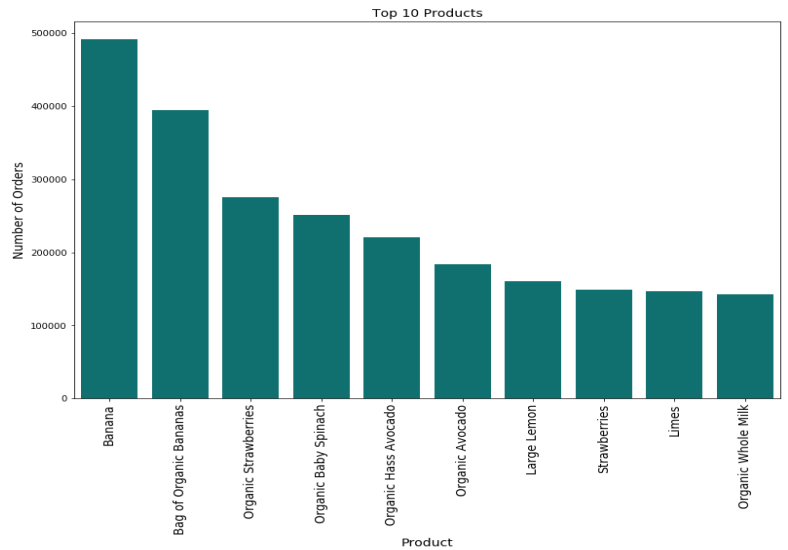

4)Top 10 Departments:

-Produce is the top department with highest number of sales followed by dairy eggs and snacks.

5)Top 10 Products:

-'Banana' is the top product with highest number of orders followed by bag of organic bananas and organic strawberries.
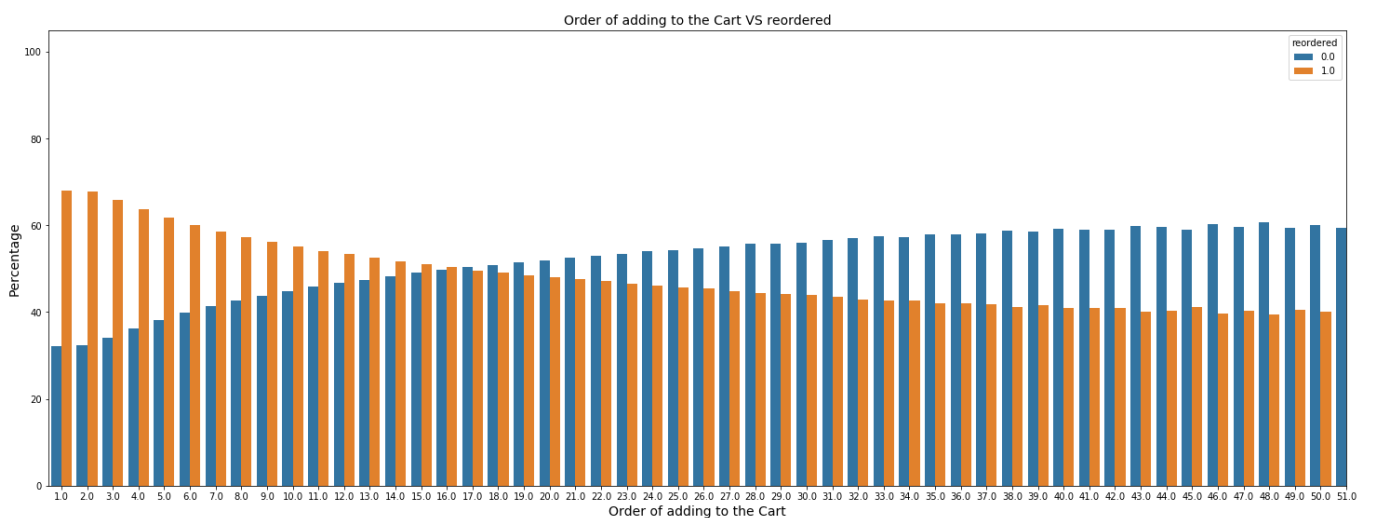

Top 10 Products

6) Top 20 reordered Products:

-'Banana' is the top reordered product with highest number of orders followed by bag of organic bananas and organic strawberries


Top 20 reordered products
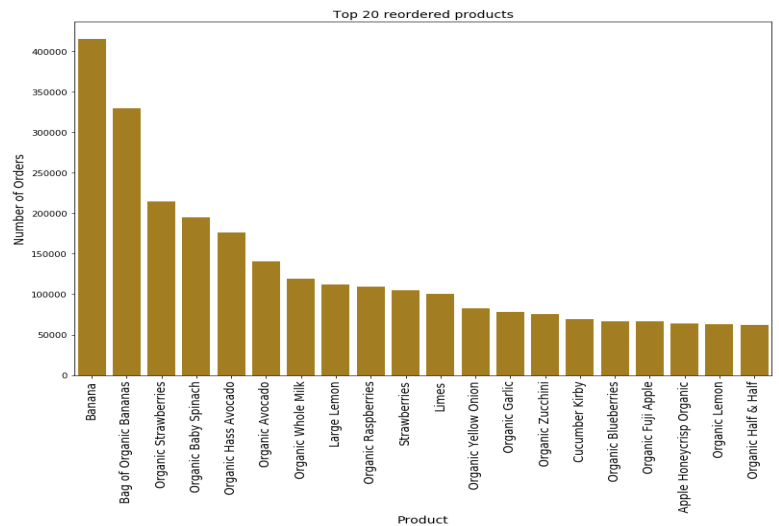
7) Added to the Cart sequence Vs reordered percentage:

Order of adding to the Cart VS reordered

-There is a clear pattern that items that were added to the cart first are reordered most. -The pattern continued until item 16 and after that the reorder percentage gradually decreased.

8) Department Vs Reordered Percentage:

Department VS reordered
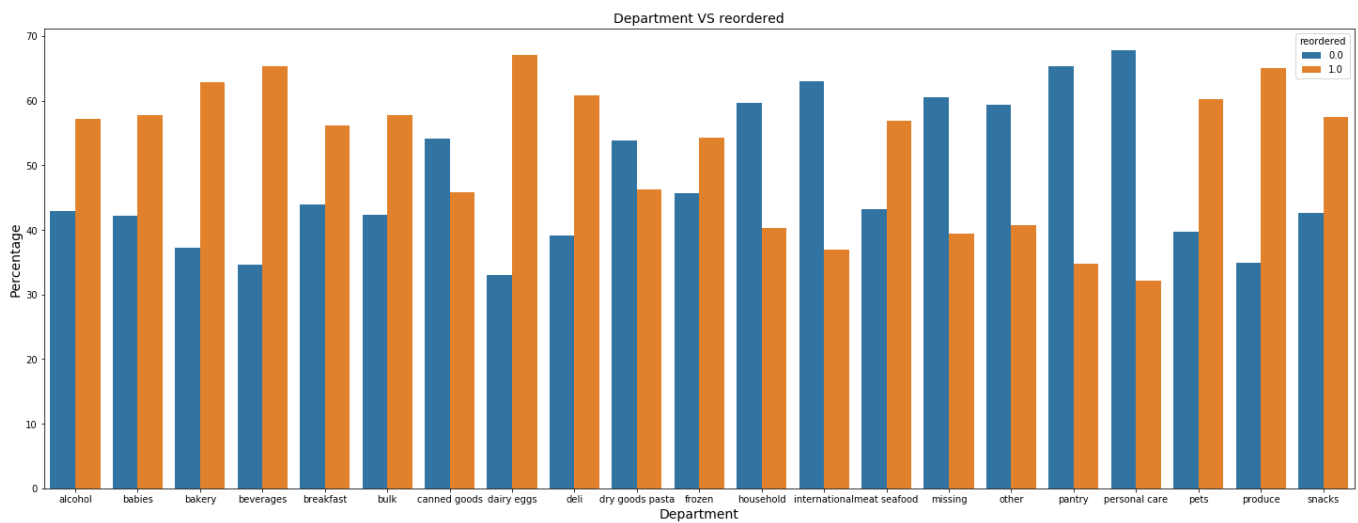
-'Dairy eggs' is the department with highest reordered rate followed by 'produce' and 'beverages'.

9) Organic Foods Vs Reordered:

-Reorder rate is more in Organic foods compared to other foods.

Organic VS reordered