

# Customer Purchase Behavior Online Shopping

---

## Introduction

Report details about the steps taken to do data analysis and visualization to determine the "Customer Purchase Behavior" for online shopping.

Our team's target for this project was to look into customer purchase behavior when shopping online for products.

In real world application, this analysis can be utilized to further increase sales and target promotional items & deals to groups of customers.

## Data Preparation

This section will detail through steps taken to prepare the data for analysis and visualization.

### Data Source:

1. We took open sourced data to understand the customer purchase behavior and shopping patterns <https://www.instacart.com/datasets/grocery-shopping-2017>.
  2. This data is from year 2017 containing ~ 3 million records.
  3. All the files were .csv and were easily downloadable.
  4. The data details are explained via data dictionary.
-

---

### Data Dictionary:

Orders	Products	Aisles	Departments	Order_product_prior
Order_id (🔑)	Product_id (🔑)	aisle_id (🔑)	department_id (🔑)	order_id (🔑)
user_id	product_name	aisle	department	product_id
eval_set	aisle_id			add_to_cart_order
order_number	department_id			reordered
order_dow				
order_hour_of_day				
days_since_prior_order				

### Data Quality:

1. Data is found to be of good quality. The dataset had ~ 3 million records and none of the files had missing data except for file "Orders.csv":
  - There were some NULL values for Column "days\_since\_prior\_order".
2. This is expected as first order will have NULL value since there was no prior order.
  - These NULL values were replaced with **Zero**.

### Data Challenges:

1. Dataset was very large for the files "**Orders**" and "**Order\_product\_prior**" and to make reading effective, **chunksize** parameter was used while reading the csv files.

---

### Data Assumptions:

1. There was only one assumption made in the entire dataset i.e. for the "order\_dow" in the file "orders".
  - **Order DOW** refers to the **Day of Week** when order was placed.
  - It had values (0,1,2,3,4,5,6). 0 was considered as Sunday, 1 being Monday and so on.

### Tools:

- **Python** language was used to do the entire coding of the project. This includes data preparation, analysis and visuals.
- Our team utilized **Jupyter** notebook to write and test the code to perform the data analysis.

---

## Data Analysis

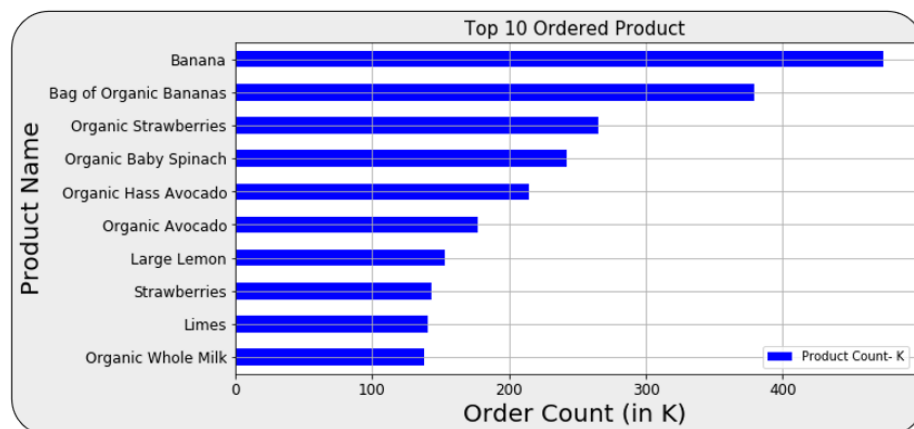
This section explains the data analysis done for determining customer purchase behavior for online shopping. This will help the owner to determine products in demand plus allow increased sales to target promotional items and deals to groups of customers.

Storytelling to describe customer behavior for online shopping.

### 1. Top 10 most **Ordered** Products

Analysis was done to find the top 10 most ordered products. It was found that the Fruits, Vegetables and Dairy products take preference over everything else.

Using “Orders, Order Products Prior & Products” dataset, we were able to determine the number of times each product was ordered.



Bananas turned out to be the item that customers always tend to place in their order.

---

Table below summarizes the top 10 most ordered products by frequency.

#	Product_name	Product Count-(in K)
1	Banana	473
2	Bag of Organic Bananas	379
3	Organic Strawberries	265
4	Organic Baby Spinach	242
5	Organic Hass Avocado	214
6	Organic Avocado	177
7	Large Lemon	153
8	Strawberries	143
9	Limes	141
10	Organic Whole Milk	138

## 2. Top 10 most **Re-ordered** Products:

Similar trend was seen for products that are often re-ordered - Fruits, Vegetables and Dairy products.

One huge variation that was seen and that was "Organic Products" takes precedence over other products when it comes to re-ordering. This means customers prefer "Organic Products" when re-ordering.

Specific example shown in the notebook & ppt is "Organic whole milk". This shows customers are becoming more health conscious and organic products are increasing in preference once used.

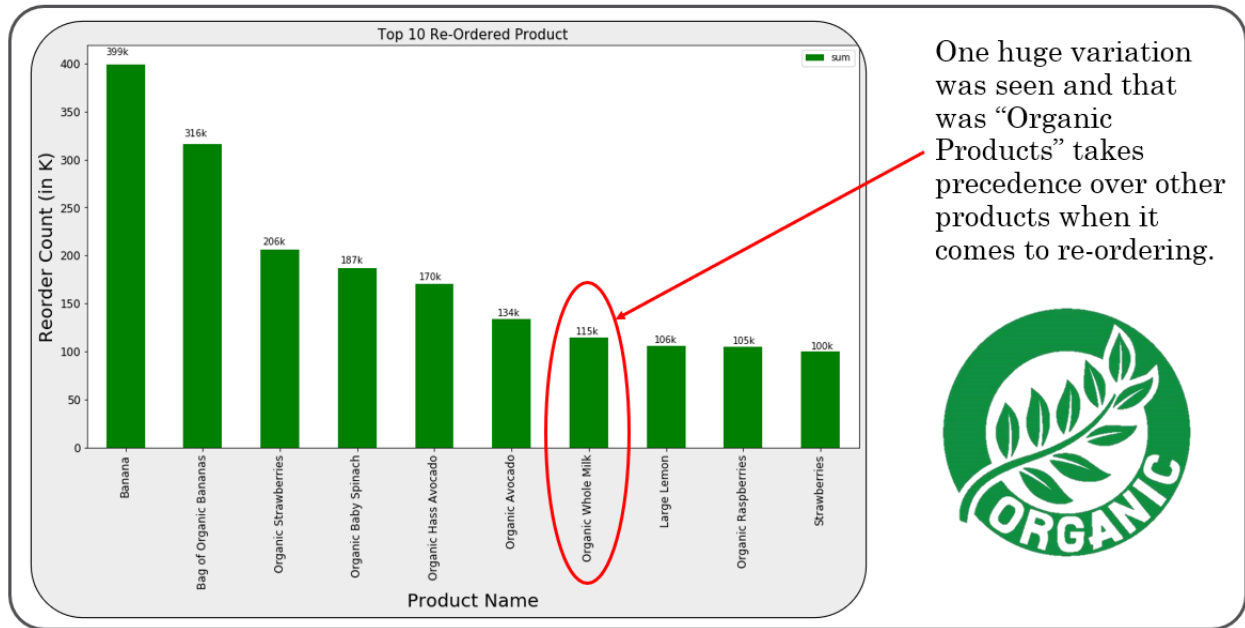
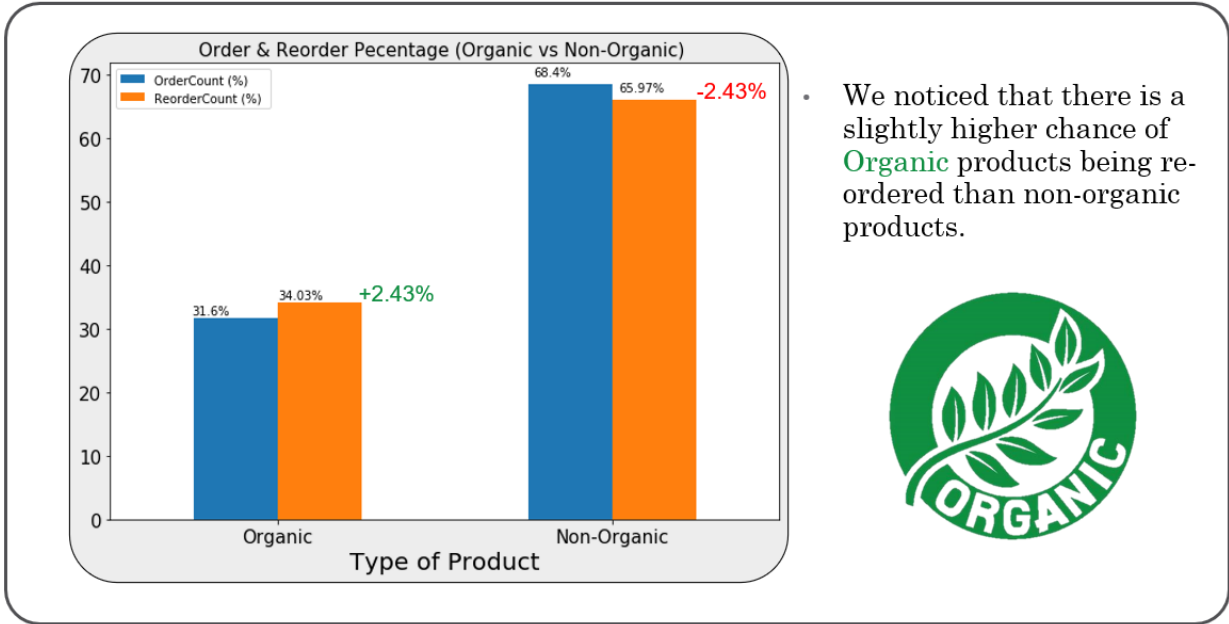


Table below summarizes the top 10 most re-ordered products by frequency.

#	product_name	Order Count-(in K)
1	Banana	399
2	Bag of Organic Bananas	316
3	Organic Strawberries	206
4	Organic Baby Spinach	187
5	Organic Hass Avocado	170
6	Organic Avocado	134
7	Organic Whole Milk	115
8	Large Lemon	106
9	Organic Raspberries	105
10	Strawberries	100

### 3. Preference for **Organic** vs **Non-Organic** products:

It's seen that the order count of "Organic Products" is **31%** of total orders whereas reorder count is **34%** of total reorders. This shows when reordering, organic products take preference over non - organic products.

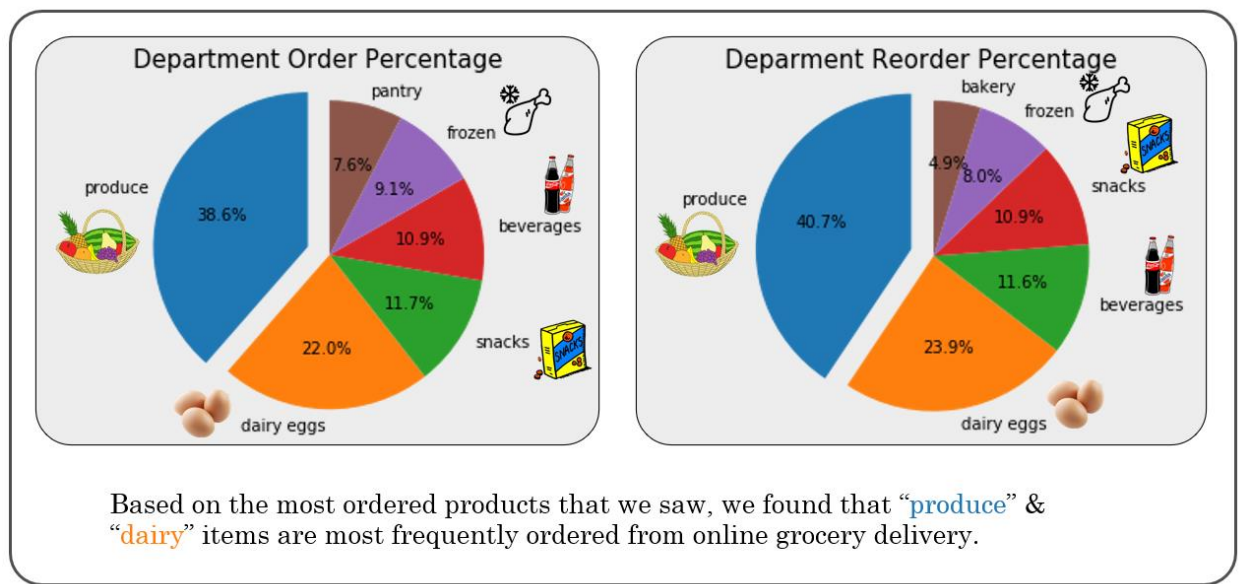


	OrderCount (%)	ReorderCount (%)
Organic	31.604367	34.028242
Non-Organic	68.395633	65.971758

#### 4. Department with most **orders** & **re-orders**:

It was seen that when reordering, the trend is usually the same, but *bakery and beverages* are slightly more frequently re-ordered.

This shows customers are more likely to reorder products such as “Beverages & Baked goods” than others such as “Snacks or Pantry”



Busiest Departments with first Orders	Busiest Departments with Re-Orders
produce	produce
dairy eggs	dairy eggs
snacks	<b>beverages</b>
beverages	snacks
frozen	frozen
pantry	<b>bakery</b>



---

## 5. Day & Time of week which has most & least Orders:

### To do this analysis we created bins:

- Values for "**day of week**" were 0 to 6:
- Bins were allocated as:
  - 0 & 6 are considered -> Sunday & Saturday (Weekend)
  - 1 to 5 are considered -> Monday to Friday (Weekday)
- Values for "**Hour of day**" were 0 to 23:
- Bins were allocated as:
  - 12am - 6am -> Early Morning (12am - 6am)
  - 7am - 12pm -> Morning (7am - 12pm)
  - 1pm - 6pm -> Afternoon (1pm - 6pm)
  - 7pm - 11pm -> Evening (7pm - 11pm)

**"Sunday"** was the busiest day and **"Thursday"** seemed to be quietest day.

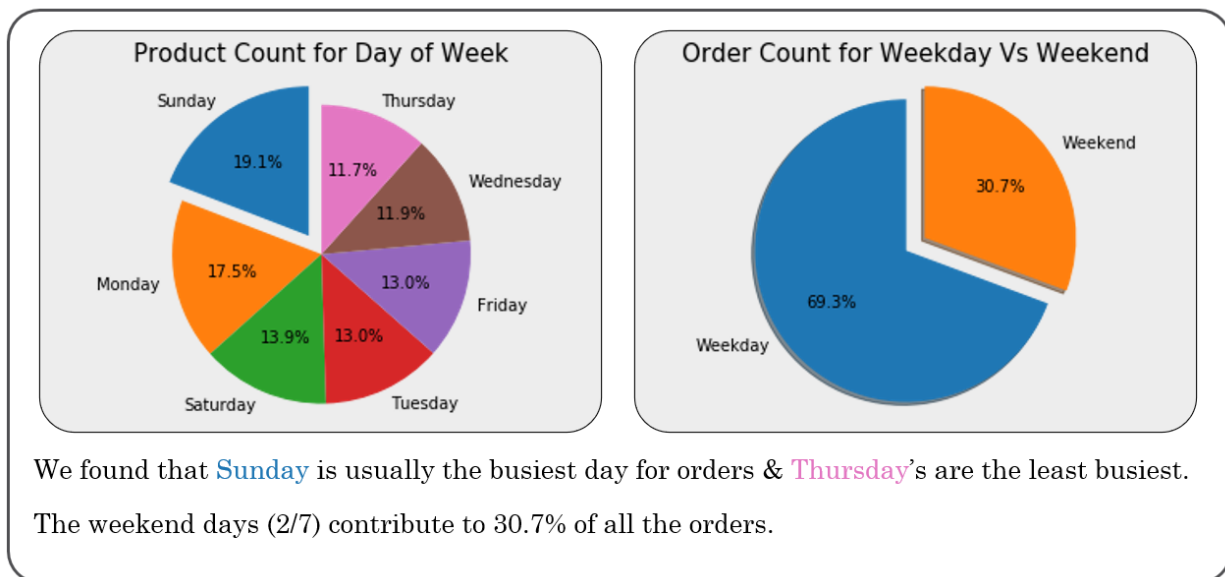
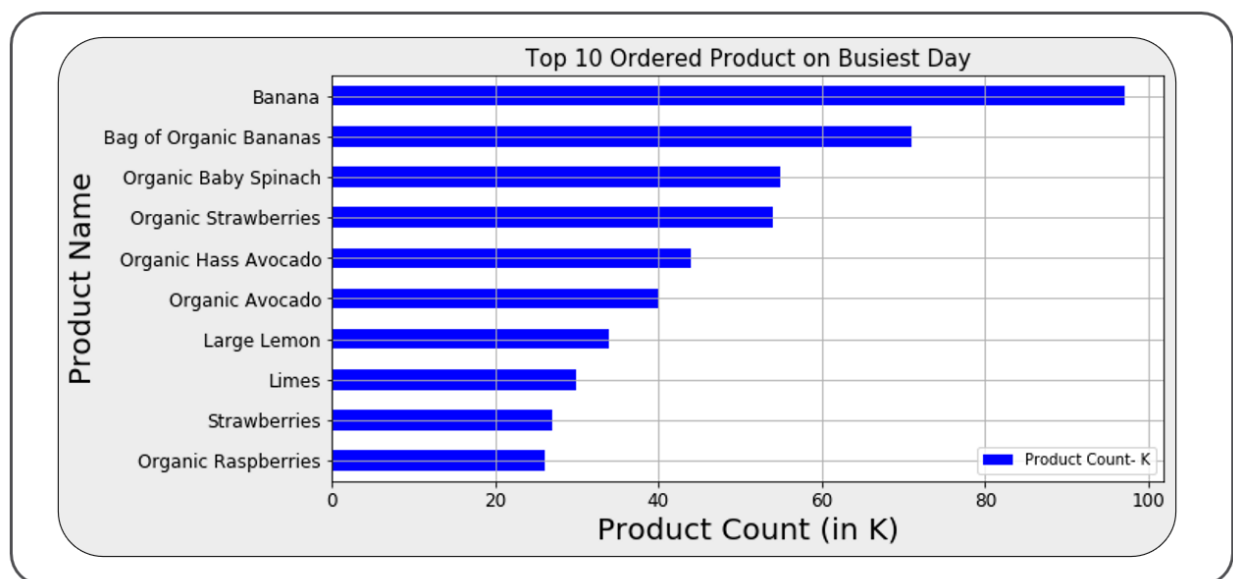


Table below summarizes percentage of orders throughout the week

Day of Week	Product Count (%)
Sunday	19.145256
Monday	17.468615
Saturday	13.875058
Tuesday	13.004053
Friday	12.97857
Wednesday	11.851943
Thursday	11.676506

- Now since we know Sunday is the busiest - it's worth knowing which products are most in demand on Sunday as that will help owners to stock items accordingly.

Top 10 products that are most in demand on busiest day:



It seems again that "Fruits & Vegetables" takes precedence as seen in previous analysis.

### Weekend vs Weekday Analysis:

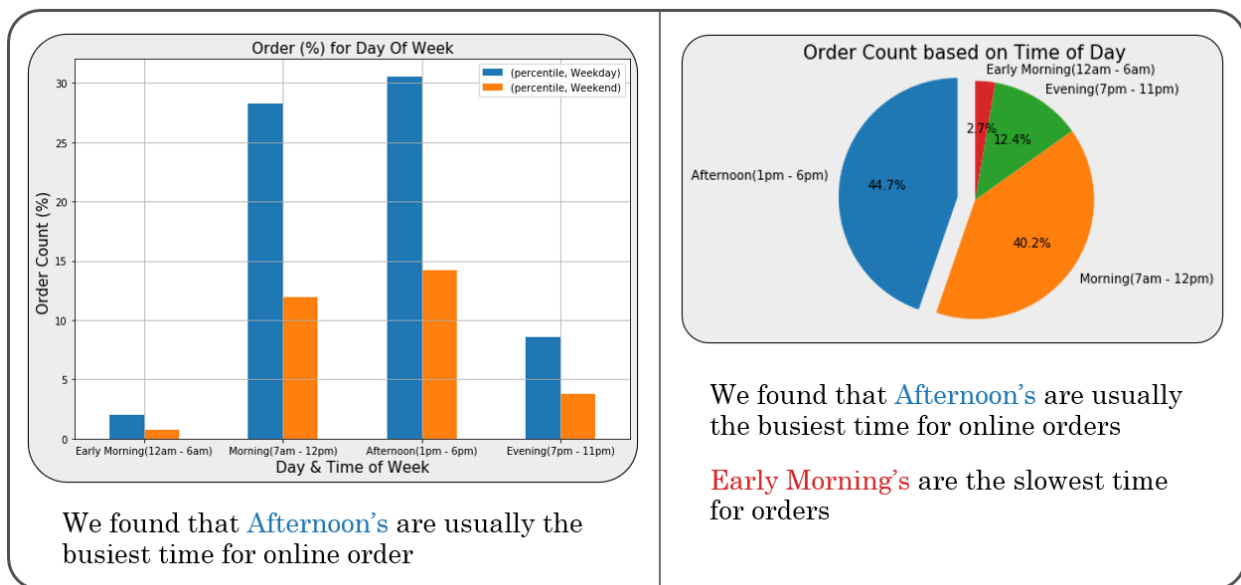
- Weekends that have **2 days contribute to 30.7% of sales.**
- Weekdays that are **5 days contribute to 69.3% of sales.**

Hence, we can conclude that there are more sales on weekends relative to weekdays.

Type of Day	% of Order
Weekday	69.31%
Weekend	30.68%

### Within a week let's see which time is busiest:

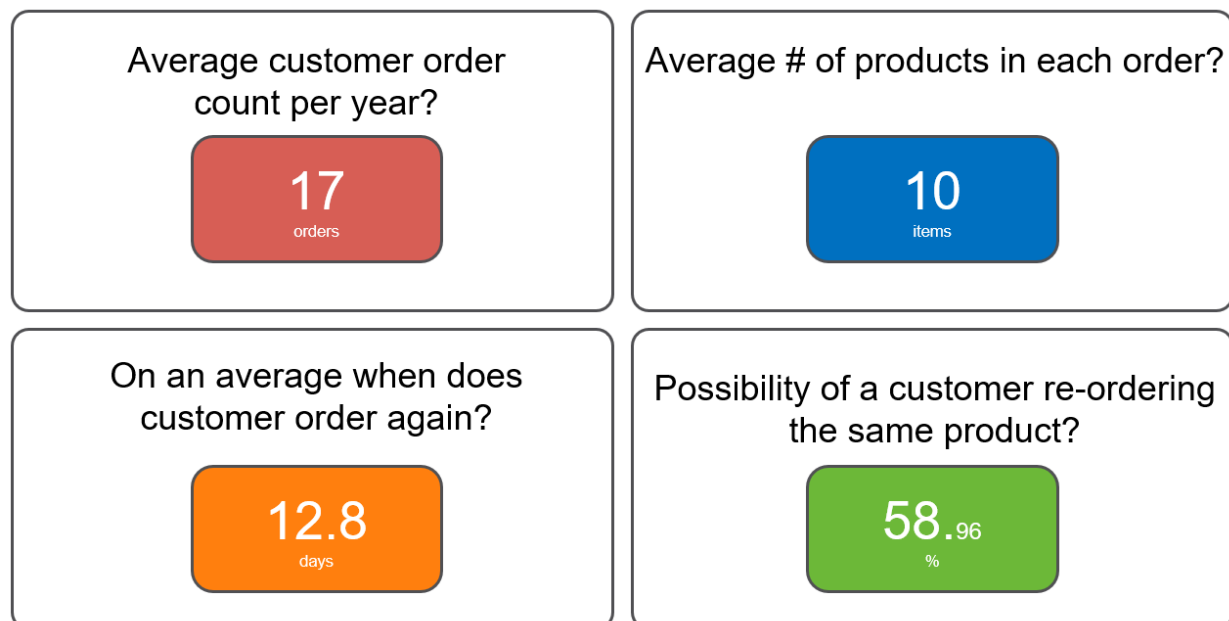
"Afternoon (1pm - 6pm)" and "Morning (7am - 12pm)" are always the busiest time for both weekends and weekdays.



Day of Week Group	Time_Of_Day	Percentage of Orders
Weekend	<b>Afternoon (1pm - 6pm)</b>	<b>14.196236</b>
	Morning (7am - 12pm)	11.901407
	Evening (7pm - 11pm)	3.810051
	Early Morning (12am - 6am)	0.774579
Weekday	<b>Afternoon (1pm - 6pm)</b>	<b>30.49546</b>
	Morning (7am - 12pm)	28.257455
	Evening (7pm - 11pm)	8.597745
	Early Morning (12am - 6am)	1.967067

Above analysis will help us consider customer purchase behavior on weekends vs weekdays and what hours are busy. Accordingly, the owner can plan stock and items and ensure enough staff is available to support.

## 6. Customer Behavior Patterns:



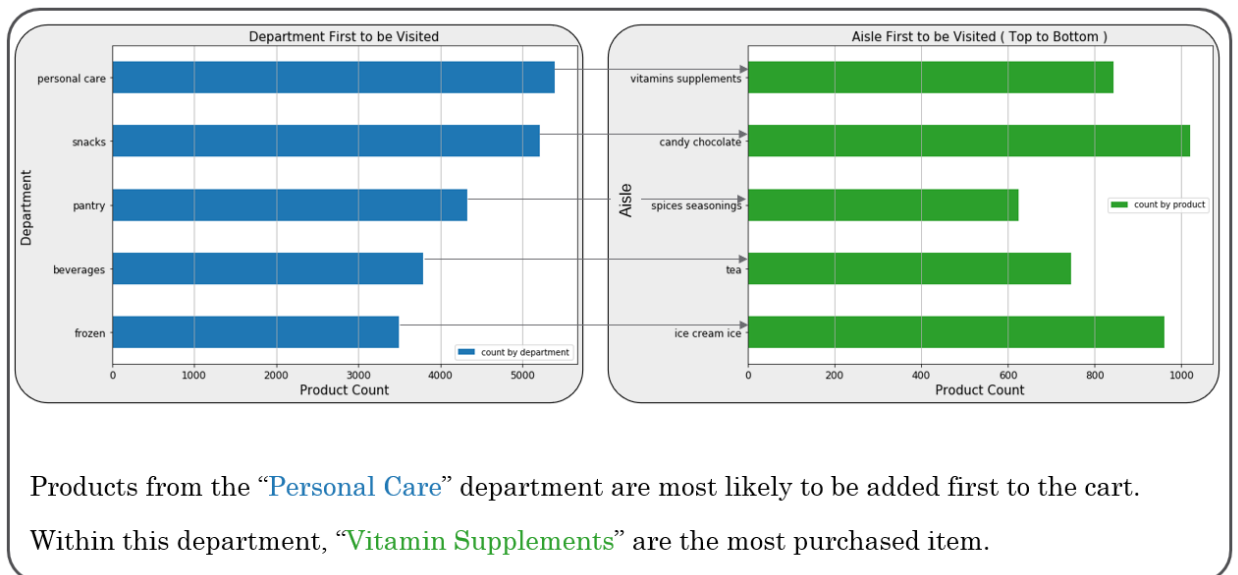
## 7. Departments & its corresponding Aisle that are first to be visited:

This section will explain human / customer mindset when they start online shopping.

*What do they prefer to pick first?*

This will allow the owner to develop the website accordingly such that the most popular departments are shown to the customer first to increase sales. Even if it is a physical store - it will be preferred to have those aisles first.

Below table shows that, **“Personal Care”** is the top department in the list and within this dept, **“Vitamins Supplements”** are first to be picked.



Department	Aisle (within dept)
personal care	vitamins supplements
snacks	candy chocolate
pantry	spices seasonings
beverages	tea
frozen	ice cream ice

---

## 8. Items that require **Promotional Deals on Weekend vs Weekday**

Based on customer purchase behavior, it's important to find which products require promotional deals so that sales can be increased.

This was analyzed based on products that are heavily sold on weekends vs weekdays and vice-versa.

We determined this using the datasets:

- Datasets for Weekends & Weekdays were merged based on "left outer join"
- Products that had "NULL" values for product count on weekday vs higher count on weekend lead to conclusion that deals on those products on weekdays can help to increase sales (vice versa for weekend deals)

*This will help to balance the product sale load across the entire week and also to allow for a better sale of items that might get spoiled or have to be thrown away (specially for perishables) because of no sales during a particular time / day of week.*

**Below list shows 5 items which can have deals on Weekdays & Weekends**

Product for Weekday Deals	Product for Weekend Deals
Red Peppers	Raspberries
Green Bell Pepper	Organic Granny Smith Apple
Boneless Skinless Chicken Breasts	Spring Water
Organic Peeled Whole Baby Carrots	Organic Unsweetened Almond Milk
Red Vine Tomato	Blueberries

Table above shows us that products such as "Red Peppers" are currently not sold during the weekdays and would make an excellent candidate for promotions during the week.

---

Similarly, "Raspberries" are not currently sold on the weekend, and deals should be created to further increase sales on weekends.

## 9. Correlation:

As per dataset reviewed via correlation function: **corr()**, minimal correlation was found within data attributes.

	order_id	user_id	order_number	order_dow	order_hour_of_day	days_since_prior_order	product_id	add_to_cart_order	reordered
order_id	1.000000	-0.000482	-0.000543	0.001205	0.000652	0.000656	-0.000082	-0.000320	-0.000253
user_id	-0.000482	1.000000	-0.000768	-0.001630	-0.000415	0.000448	0.000080	0.000729	-0.000862
order_number	-0.000543	-0.000768	1.000000	0.015209	-0.040049	-0.257864	-0.001925	-0.004228	0.306840
order_dow	0.001205	-0.001630	0.015209	1.000000	0.012176	-0.026520	-0.002350	-0.009071	-0.006949
order_hour_of_day	0.000652	-0.000415	-0.040049	0.012176	1.000000	-0.000195	0.000942	-0.014805	-0.023300
days_since_prior_order	0.000656	0.000448	-0.257864	-0.026520	-0.000195	1.000000	0.000563	0.050260	-0.024465
product_id	-0.000082	0.000080	-0.001925	-0.002350	0.000942	0.000563	1.000000	0.005529	0.003718
add_to_cart_order	-0.000320	0.000729	-0.004228	-0.009071	-0.014805	0.050260	0.005529	1.000000	-0.133024
reordered	-0.000253	-0.000862	0.306840	-0.006949	-0.023300	-0.024465	0.003718	-0.133024	1.000000

---

## Conclusion:

Our team targeted to analyze customer purchase behavior using a data set containing millions of online order details:

Using python, we were able to manipulate and reorganize the data in order to obtain valuable results.

Based on our analysis, our discovery lead to a better understanding of the customer purchase behaviors.

- [Fruits & Vegetables](#) are the most ordered & re-ordered products
- Customers tend to prefer [Organic](#) products over [Non-Organic](#) while re-ordering
- Departments that are the busiest when ordering or reordering; [Produce, Dairy, Beverages / Snacks](#)
- We found that [Sunday](#)'s are the busiest day of the week for orders, and [Thursday](#)'s are the quietest
- Customers tend to place orders most frequently during the [Morning \(7AM-12PM\)](#) & [Afternoon \(1PM-6PM\)](#)
- We were also able to determine patterns for Customer purchases such as average number of products in each order and average number of orders per year
- From our analysis, we concluded that the top three departments that are first to be visited are Personal Care, Snacks & Pantry
- We were also able to analyze what products would benefit the most from deals on Weekdays & Weekends

Ultimately, this analysis can be beneficial to online vendors to better understand what their customers are purchasing and use this data to target promotions to groups of customers.