Instacart: Strategic Marketing Analysis

Using the Instacart Online Grocery Shopping Dataset from 2017 a deep dive analysis was conducted with the purpose of analyzing behavior patterns for targeted marketing and investigating avenues that will expand the business model. Pondering question: How can we understand users better to meet their needs?

Getting started

Dataset Source: https://www.kaggle.com/c/instacart-market-basket-analysis/data
Downloaded tables in csv format

Instacart published this anonymized dataset with the hopes that the Machine Learning community will use it to predict which products a user will buy again, try for the first time, or add to their cart next during a session.

The dataset contains:

- A sample of over 3 million grocery orders (order_id)

- From more than 200,000 Instacart users (user_id)

- For each order the sequence of products purchased (product_id)

- The day of the week and hour of day the order was placed (order_dow, order_hour_of_day)

- A relative measure of time between orders (days_since_prior_order)

Prerequisites

PostgreSQL
Python 3 (Jupyter Notebook, Numpy, Pandas, Matplot.lib, Seaborn)

PostgreSQL

- Created database and uploaded tables

```
CREATE DATABASE Instacart
CREATE TABLE
COPY FROM
```

For user classification we decided to identify users by:
- Small Businesses: Retail, Food and Beverage, and Service Industry
- Offices: Corporate, Non-Corporate, Co-working Spaces
- Active: Head of Households, Housewives, Individuals
- Inactive: Trials, First-timers, Special Occasions

We assume that:
- Businesses have a higher count of orders, as their order cycle is more predictable
- Small Businesses order at least twice a week, whilst Offices might order at least

weekly
- Consumers, however, are less predictable in behavior and might order sporadically
- Inactive users have spent more than 30 days without making an order and active less than 30

```
CREATE VIEW user_segmentation AS
SELECT user_id,
COUNT(order_id),
ROUND(AVG(days_since_prior_order), 2),
CASE
    WHEN (ROUND(AVG (days_since_prior_order)) <= 3.50) AND
(COUNT(order_id) >= 60) THEN 'small business'
    WHEN (ROUND(AVG (days_since_prior_order)) > 3.50) AND
(COUNT(order_id) >= 60) THEN 'office'
    WHEN (ROUND(AVG (days_since_prior_order)) <= 30) AND
(COUNT(order_id) < 60) THEN 'active'
    WHEN (ROUND(AVG (days_since_prior_order)) > 30) AND
(COUNT(order_id) < 60) THEN 'inactive'
END
    AS user_type
FROM orders
GROUP BY user_id;
```
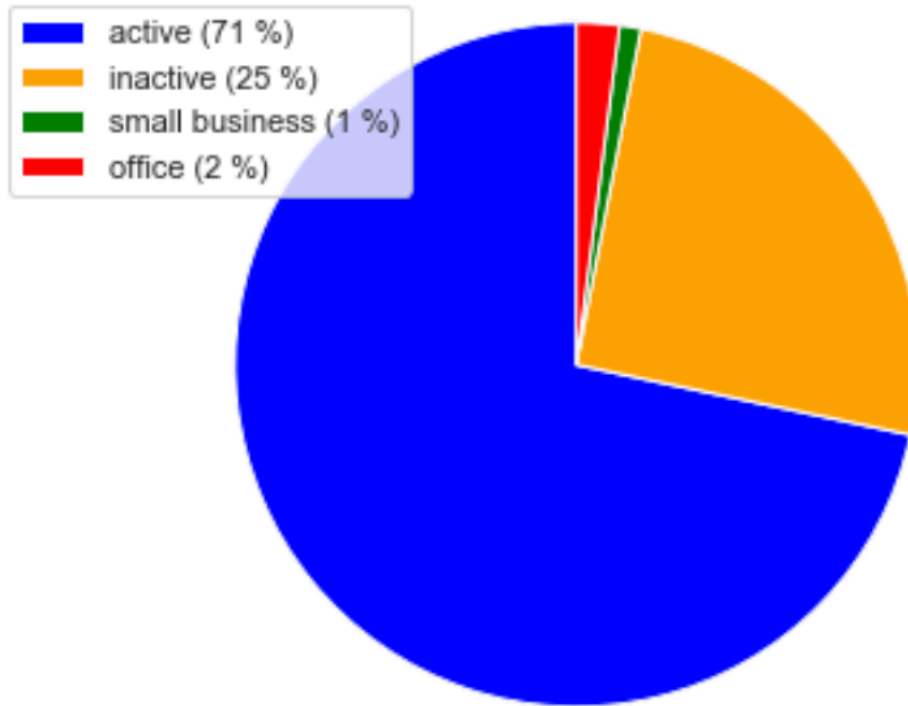
Exported view to csv table:

```
\COPY (SELECT * FROM user_segmentation) to
'final_segmentation' with csv
```

Then on, we upload it to Python.

Python

Segmentation:

The user segmentation denotes a large consumer base and a small margin of users that behave like businesses.
The most valuable sector is the smallest as it Business classifications contain a higher volume of orders and most consistent ordering cycles.
Active and Inactive users could be further segmented with the use of more data.

How exactly do they behave?

Ordering patterns based on time of day and day of the week for each segment:
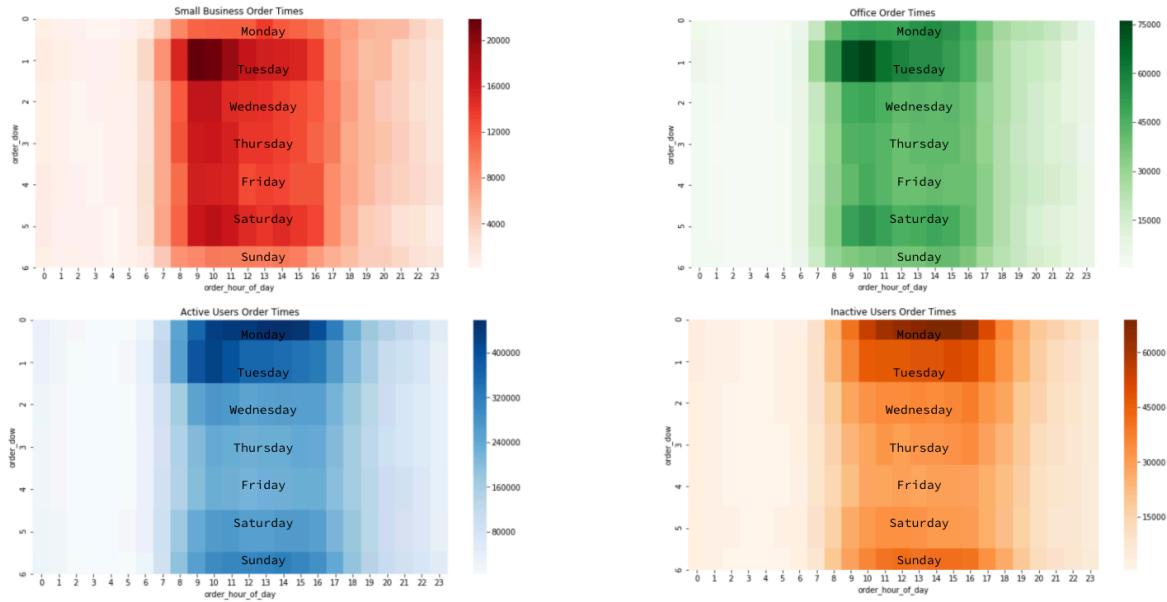
Heat maps identify the concentration of activity for each of the segments in our user classification:
- Businesses make most their orders between 9 and 10 in the morning on Tuesdays.
- Consumers concentrate their orders spread out during the day, with most traction on Sundays, Mondays, and Tuesdays.
- In general, the higher concentration of orders occurs during the day
- Businesses have a fair amount of concentration during Saturdays, implying that they are open 7 days and refilling inventory on that day.

What do our segments order?

Basket Analysis
The variance in product ordering was basically non-existent considering every segment, hour of day, and day of week pointed out as the most ordered products were consistently the same.
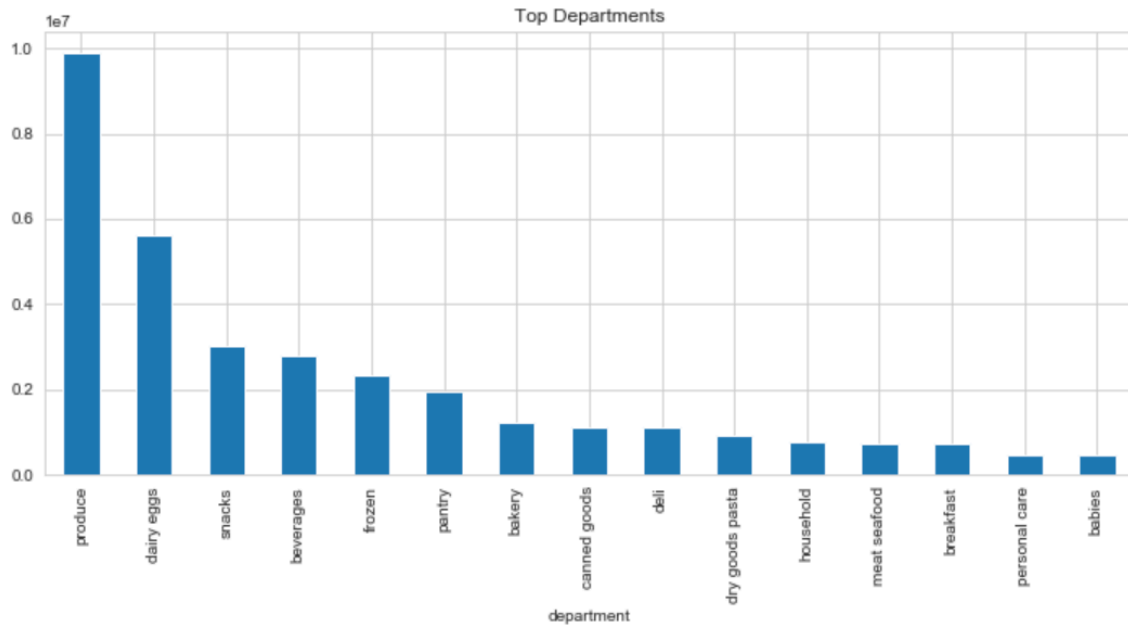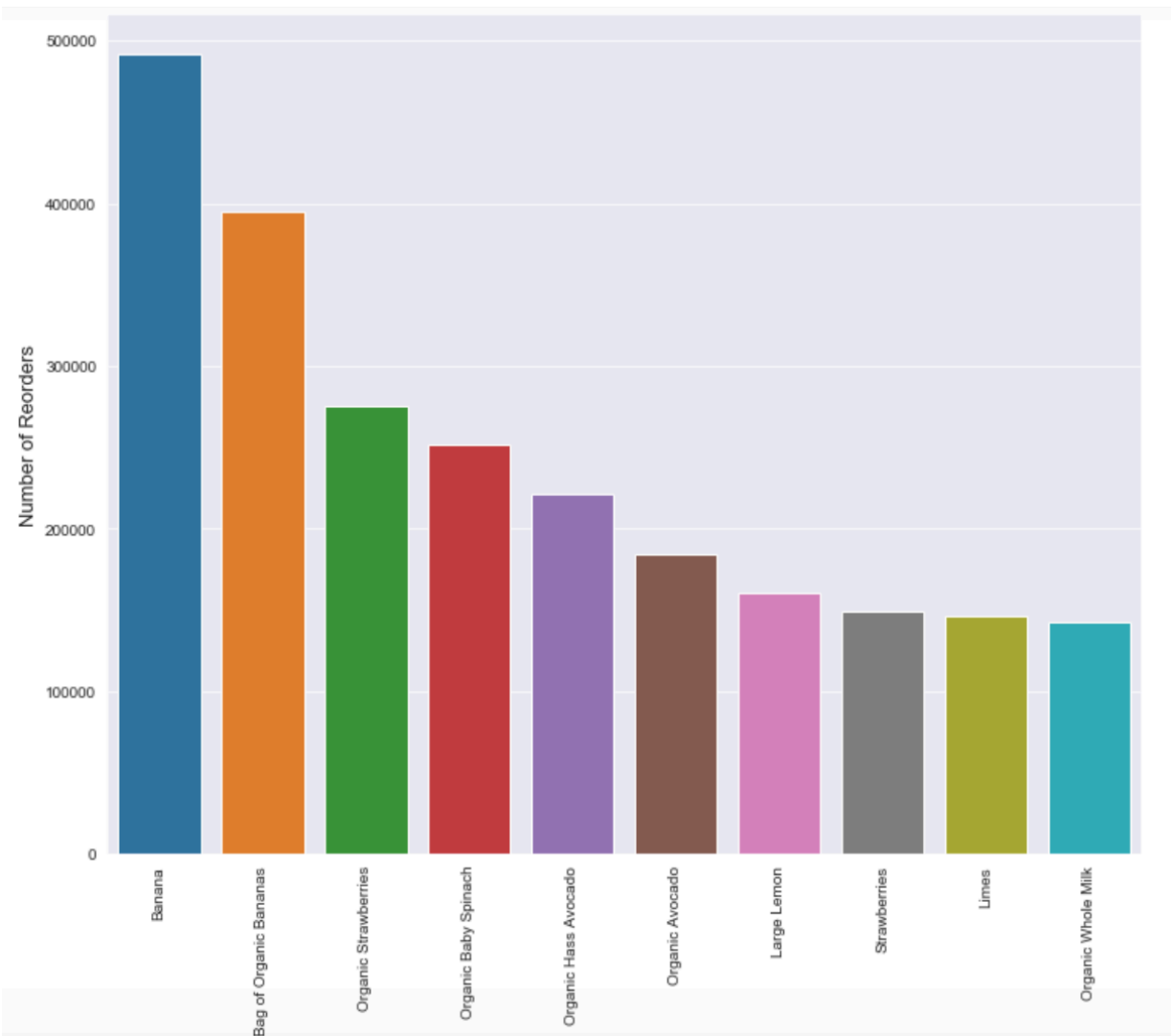
Top Departments

Department's are at the top of the product classification umbrella. The department for the dataset is produce, so the most products ordered lied within this category.

Top Aisles
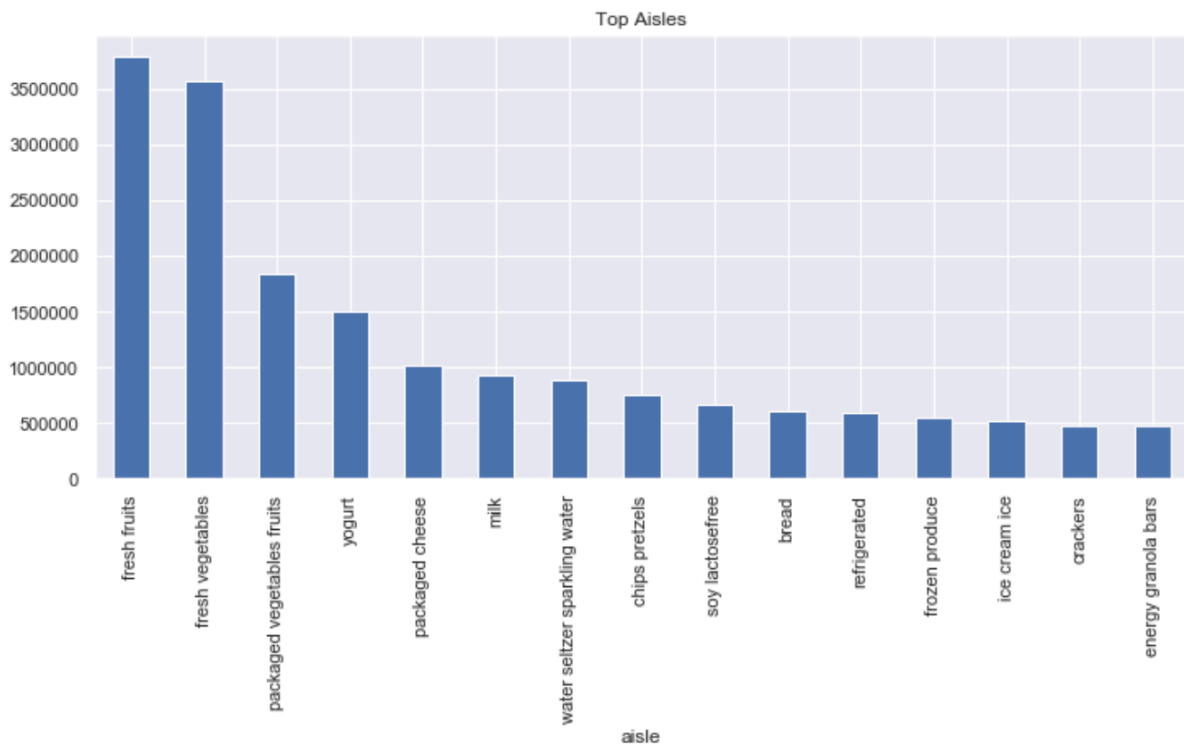
Top three aisles:
- fresh fruits
- fresh vegetables
- packaged vegetables fruits

These all fall within the product department, now let's look at the products:

All the top products are part of the produce department, and these consistently showed at the top when looking at product ordering.

How can we sell more produce?

Takeaways and Initiatives

- With a large customer base, conversion of inactive users and retention of active through marketing initiatives is a priority
    - With an Instacart retail capsule that sells produce and promotions of big box items we will promote brand awareness and customer acquisitions
    - Using a model like Amazon's Treasure Truck, we acquire vehicles and conduct a geospatial analysis to target customers, sell more produce, and make people wonder what we are about
- There is an ability to cater to the business sector by working closely with their accounts
    - Elongating or founding a business department within Instacart is key to develop

these relationships
- Hire e-mail marketers, analysts, and engineers to obtain their information, engineer features that will gear security, inventory consistency, and high potential for growth on both ends
- We can only predict our users will order more and more produce
- Focusing locally on acquiring small businesses as partners to reinforce sharing capabilities and uplift the mission to save people time and provide them with the food they love