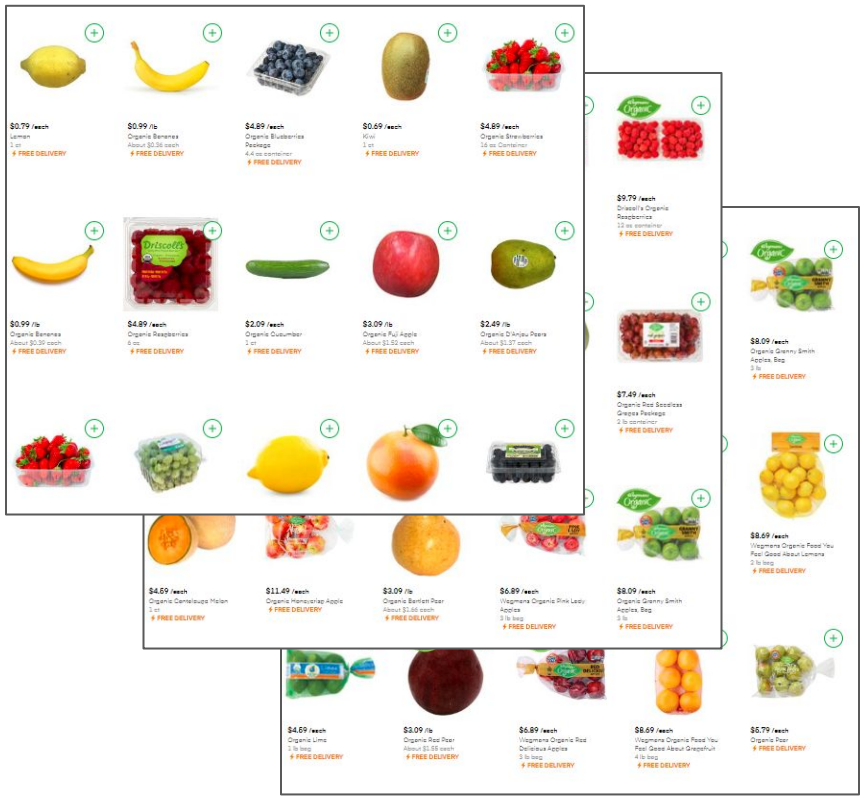
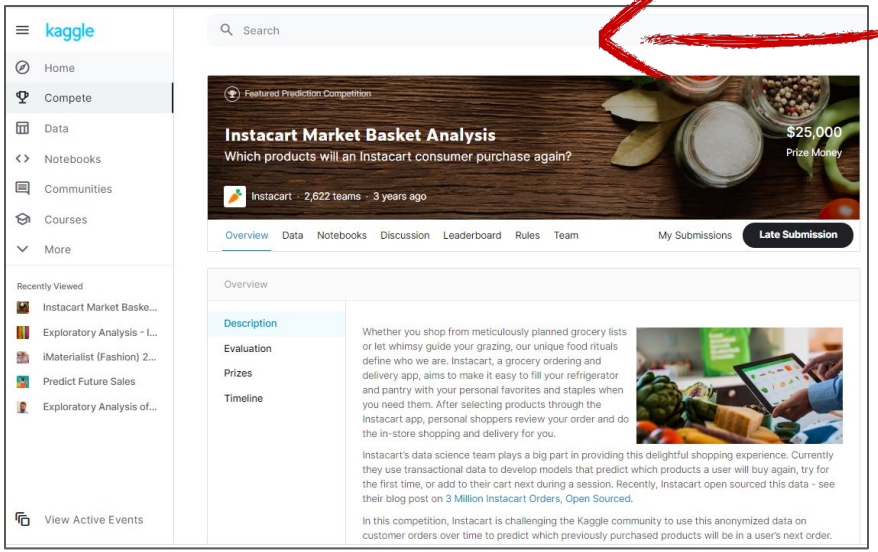
Instacart Market Basket Analysis

Faris Haddad, Greg Tully, Sam Temlock
W200, Section 1



Research Dataset



Data was sourced from an Instacart Kaggle competition launched in 2017

Key metrics from the source datasets:

Number of Orders	Number of Users	Total products ordered in prior dataset
3,421,082	206,209	32,434,489

Research Questions

1. ***What is the breakdown of customers that buy organic versus ones that never buy organic when given the option?***

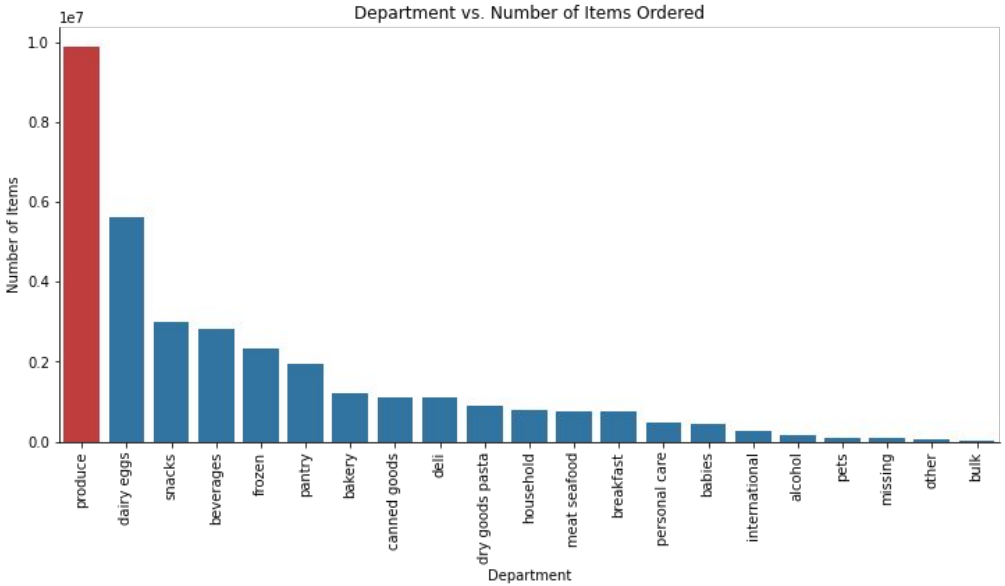
2. ***Does purchasing behavior differ between organic and never organic users?***

3. ***Does user purchasing behavior change as they make more orders on Instacart? And is buying "organic" a stable or dynamic behavior?***

Data Selection

1. Select Produce Department data because has like-for-like products

2. Produce Dept. is the largest dataset by items ordered



Data Validation & Compilation

1. Sanity checks on data, for example:
    a. ID numbers for orders, departments, aisles, and products were **checked for uniqueness**
    b. Datasets were checked to ensure that total number of values matched the amounts declared in the source and there were no NaN values

2. Joined data to create a single Produce dataset. Summary table:

Number of Orders	**Number of Users**	**Total produce items ordered**	**Total organic produce items ordered**
2,506,247	194,331	9,888,378	5,323,624

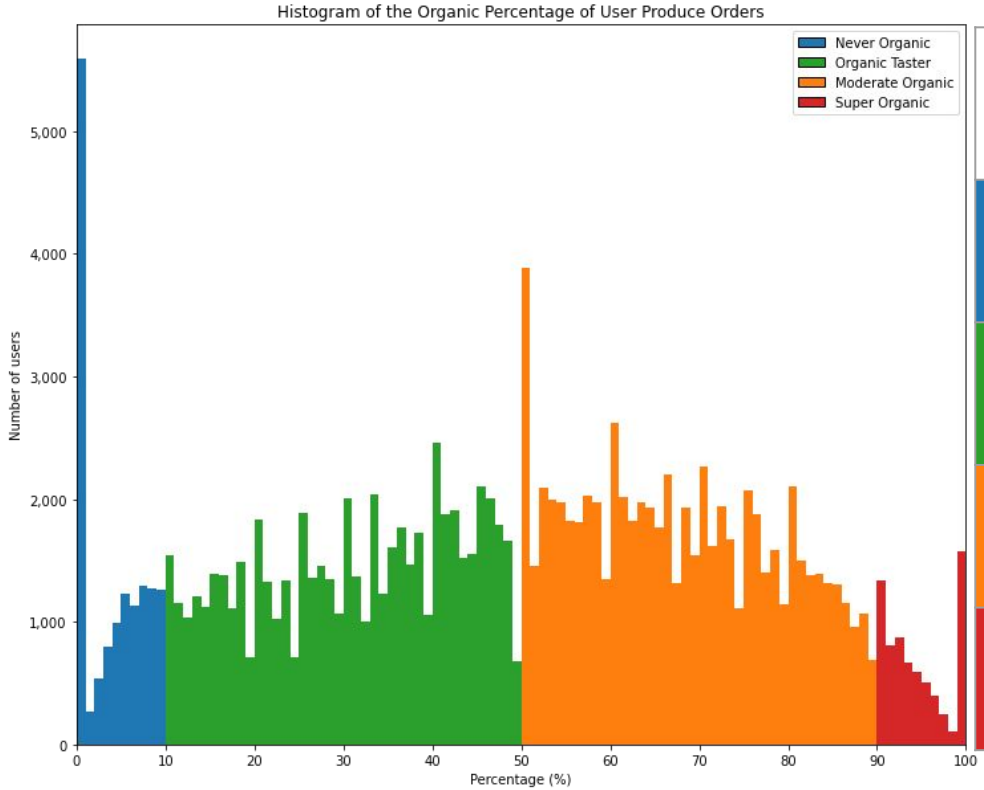
Research Questions

1. ***What is the breakdown of customers that buy organic versus ones that never buy organic when given the option?***

2. ***Does purchasing behavior differ between organic and never organic users?***

3. ***Does user purchasing behavior change as they make more orders on Instacart? And is buying "organic" a stable or dynamic behavior?***

Defined four Organic Segments based on how “organic” a customer buys produce



Segment	% of organic produce purchased	# users	# orders	# produce items purchased
Never Organic	Below 10%	14,367	188,169	527,700
Organic Taster	Between 10% and 50%	58,326	859,853	3,316,590
Moderate Organic	Between 50% and 90%	69,074	1,206,072	5,398,847
Super Organic	90% or above	7,102	117,531	427,346

Research Questions

1. *What is the breakdown of customers that buy organic versus ones that never buy organic when given the option?*

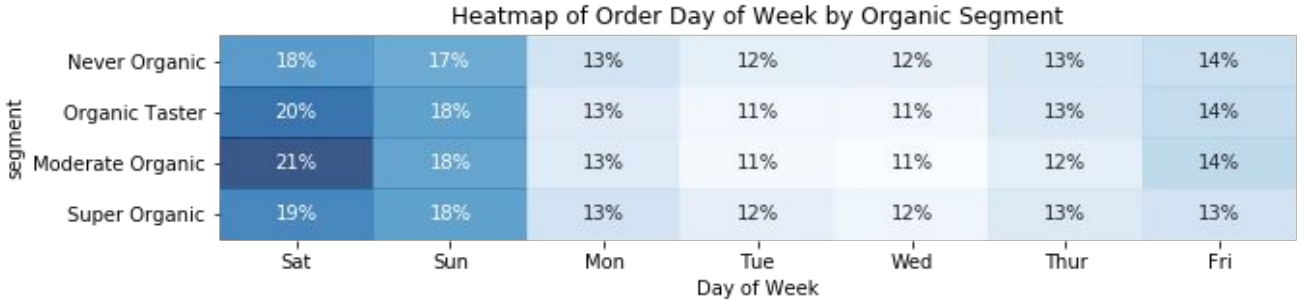
2. ***Does purchasing behavior differ between organic and never organic users?***

3. *Does user purchasing behavior change as they make more orders on Instacart? And is buying "organic" a stable or dynamic behavior?*

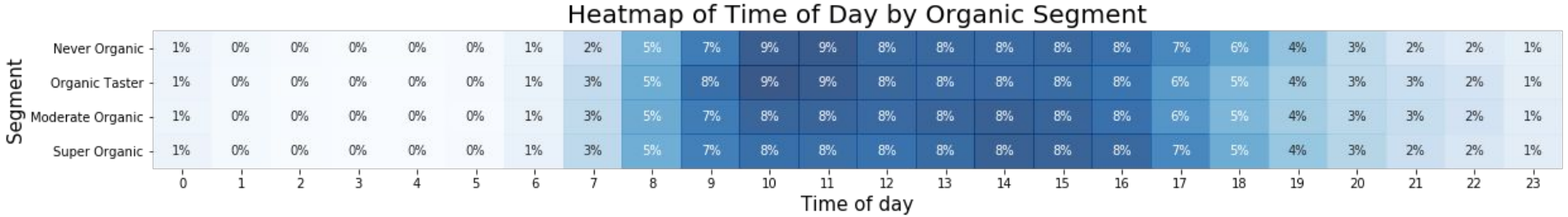
Hypothesis:
Purchasing behaviour differs between organic customer segments, especially in how often and how much produce a segment purchased

Purchase day & time does not differ among segments

All segments seem to order **more on the weekend** but during the week as well:

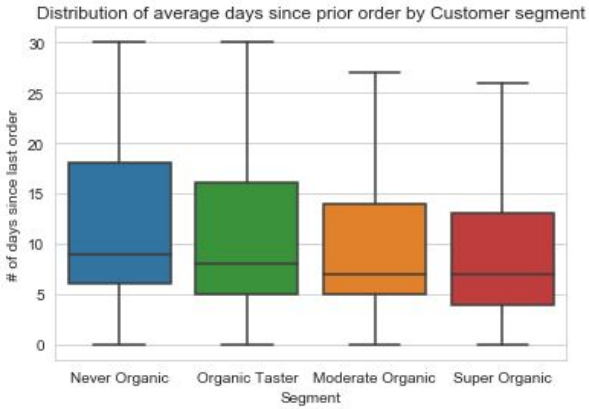


All segments seem to mostly order during **working hours**:



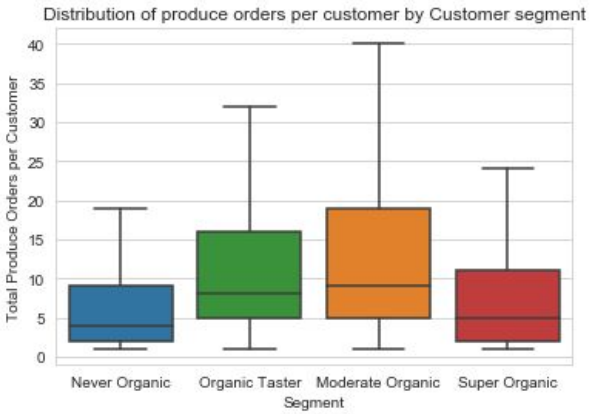
Never Organics seem to buy differently than other organic segments

Number of days since prior order per customer



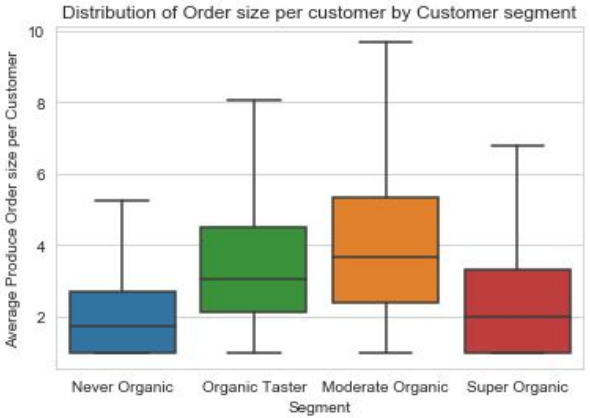
Never Organics show a <u>longer lag time</u> between orders compared to other segments

Number of Produce orders per Customer



Never Organics show the <u>lower # of orders</u> followed by Super Organics

Produce order size per customer



Never Organics show <u>lower # of produce items</u> per purchase, followed by Super Organics

Research Questions

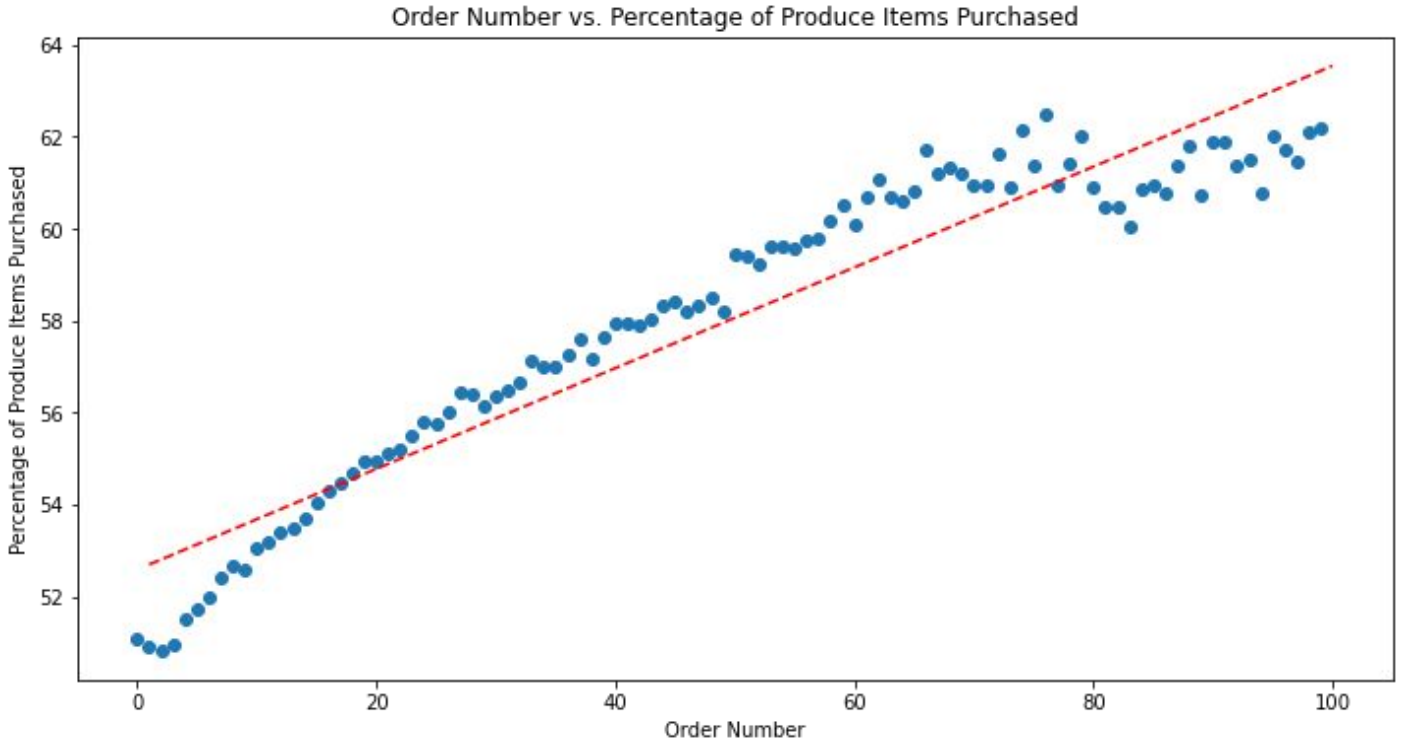
1. *What is the breakdown of customers that buy organic versus ones that never buy organic when given the option?*

2. *Does purchasing behavior differ between organic and never organic users?*

3. ***Does user purchasing behavior change as they make more orders on Instacart? And is buying "organic" a stable or dynamic behavior?***

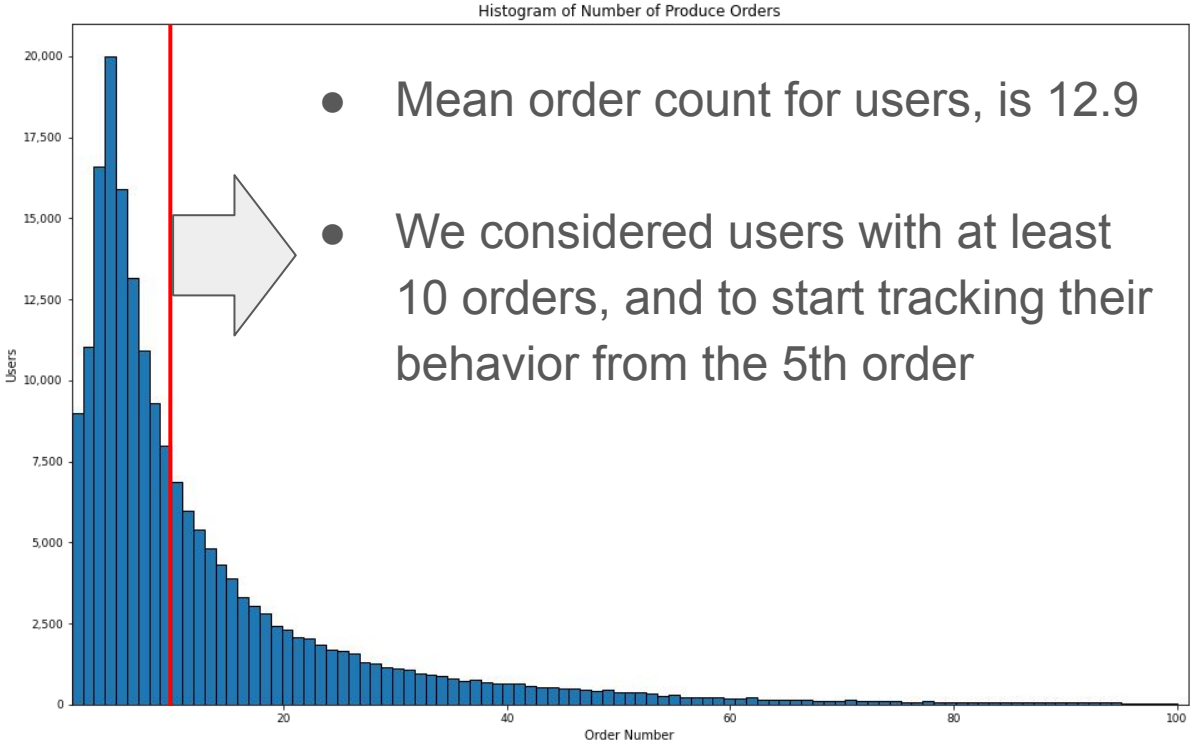
Hypothesis:
The ends of each spectrum (i.e., Never Organic and Super Organic) are the most consistent in their degree of organic buying

The percentage of organic produce among produce orders tends to increase as users order more



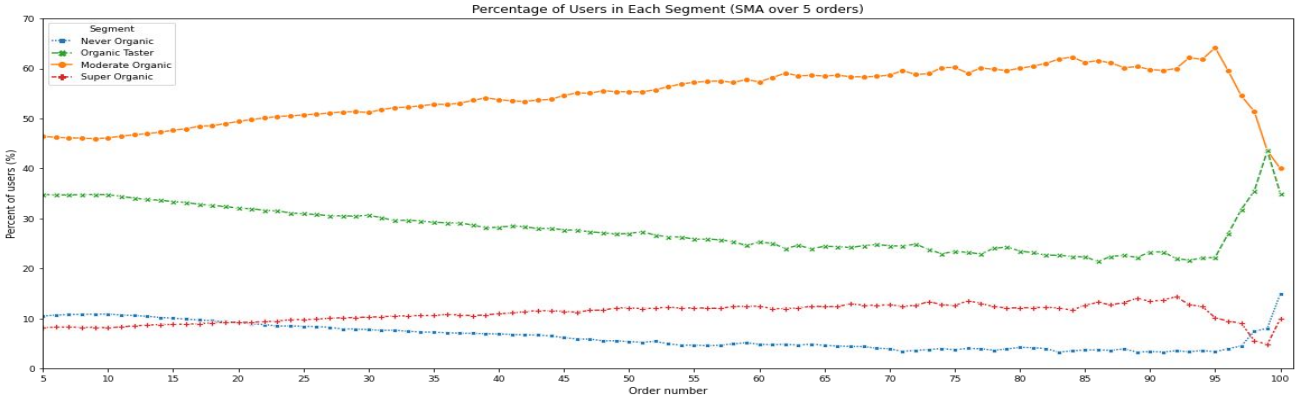
Produce order count is a heavy tailed distribution -

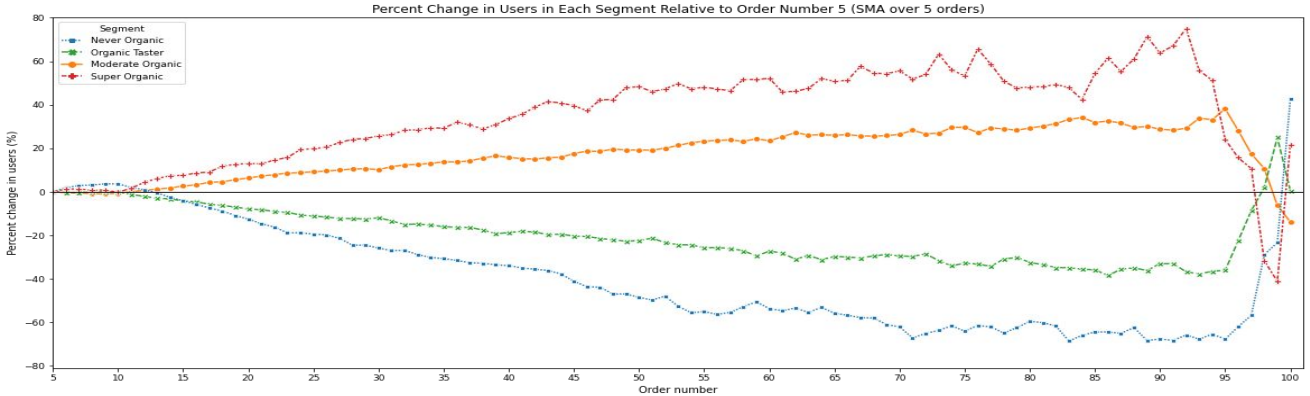
We analyzed segment stability on users with 10 or more orders



As more orders are made, the distribution of Organic users changes

- The distribution of **Moderate Organic & Super Organic users increases**
- **Never Organic & Organic Tasters users decrease**

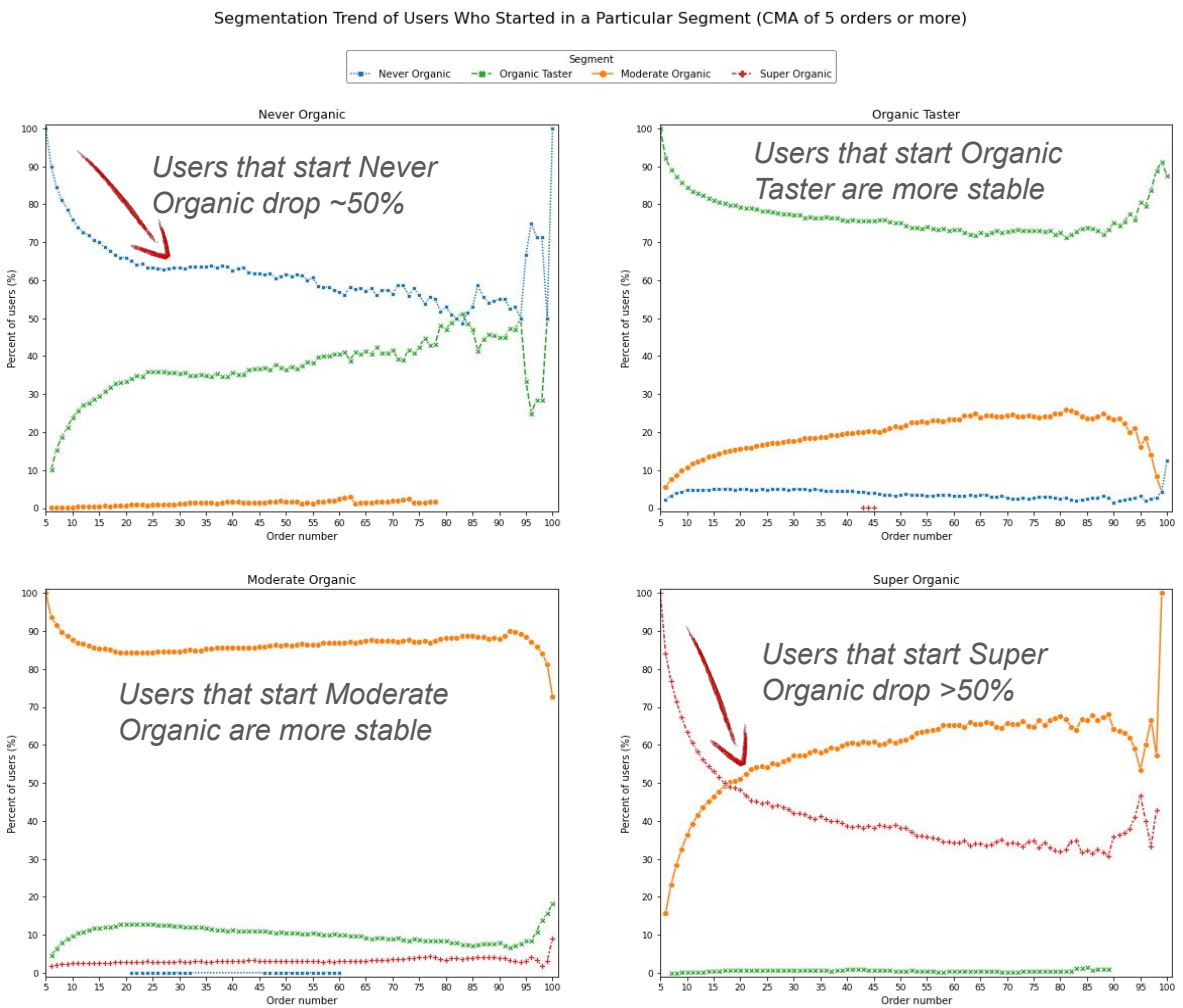




Note: Simple Moving Average (SMA) calculates the rolling average over a set "window" of 5 orders

Never Organic and Super Organic are less stable

- **Up to 50%** of the users who were initially **Super Organic and Never Organic** fall out of those segments as they put in more orders
- Conversely, users who start as **Moderate Organic and Organic Taster tend to stay that way** (roughly 90% and 80% respectively)



Note: Cumulative Moving Average (CMA) calculates the expanding average from order number 5 (includes all previous orders)

Key Takeaways

- **Organic customers make up a large portion** of users (>50% of purchased produce is organic)
- Users that **never buy organic produce (i.e., Never Organics) seem to shop differently** than users that buy organic; they order less, more sporadically and when they do order, they buy less produce than organic buyers
    - Super Organic users also order less produce per order, we hypothesize that they might be “pickier” buyers
- As users make more orders, they **seem to include produce items more often** and **to shift towards organic produce** as well
- Users in the **Super Organic and Never Organic segments appear less stable** than those in the Moderate Organic and Organic Taster segments. Super Organics migrate towards Moderate while Never Organics migrate towards Tasters

Thank You!

Organic segment behavior seen in Produce is consistent in full Instacart dataset

Produce only

~194K users

segment	Average # of orders	Average # of days since last order	Average order size
Never Organic	7.6	12.5	2.6
Organic Taster	12.8	11.8	3.8
Moderate Organic	15.4	10.6	4.4
Super Organic	10.0	10.2	3.4

Full dataset

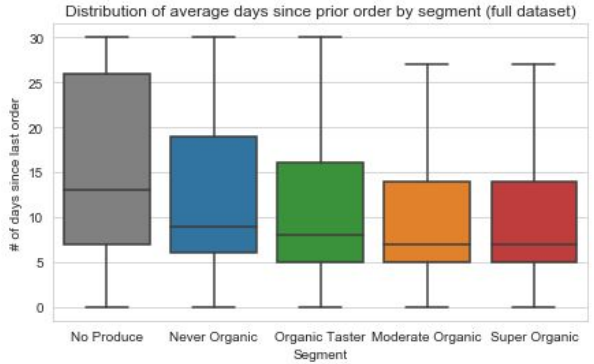
~206K users

segment	Average # of orders	Average # of days since last order	Average order size
No Produce	9.1	15.0	4.5
Never Organic	13.1	12.7	8.6
Organic Taster	16.5	11.8	10.0
Moderate Organic	18.5	10.7	11.1
Super Organic	14.6	10.9	8.8

Note: "No Produce" is 11,878 users (or ~ 6% of users)

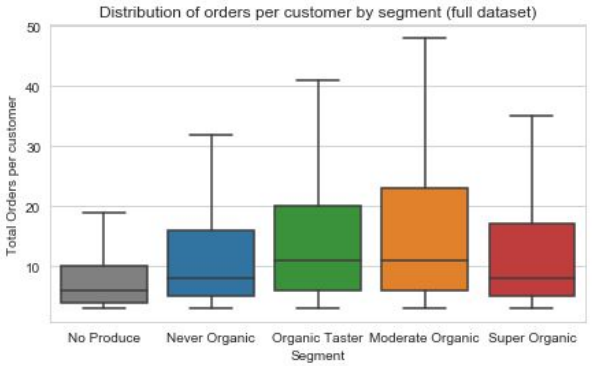
Never Organics seem to buy differently than other organic segments - *full dataset analysis*

Number of days since prior order per customer



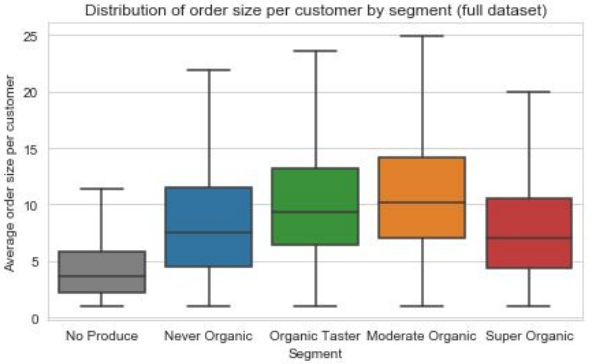
Never Organics (& No Produce) show <u>longer lag time</u> between orders

Number of orders per Customer



Never Organics (& No Produce) show the <u>lower # of orders</u> followed by Super Organics

Item order size per customer



Never Organics (& No Produce) show <u>lower # of items</u> per order, followed by Super Organics