INSTACART 2017 ANALYSIS

BY
WINRICH SY

About

What is Instacart?

Instacart is a same-day grocery delivery and pick-up service in the U.S. Customers shop for groceries through their app and or website and have their groceries delivered or ready for pick-up at local groceries. They're partnered with major grocery chains such as Costco, ALDI, Sam's Club, Sprouts, and etc. Concurrently serve more than 20,000 different grocery stores across more than 5,500 cities in North America.

Why Instacart?

Customers use Instacart to save trips to and from the groceries and time from searching for their products within the stores. With its growing popularity, it is useful to understand what items are being bought and at what rate. That information can be used to help retailers to know what and when to stock up on certain products. Data uncovered can help lower unnecessary costs on stocking unpopular items and boost net profits.

CSV Files & Schema

aisles.csv rows: 134

- **aisle_id**: integer | aisle identifier
- **aisle**: string | the name of the aisle

departments.csv rows: 21

- **department_id**: integer | department identifier
- **department**: string | the name of the department

order_products_prior.csv rows: 32434489

- **order_id**: integer | foreign key
- **product_id**: integer | foreign key
- **add_to_cart_order**: integer | order in which each product was added to cart
- **reordered**: integer | 1 if product has been ordered by this user in the past, 0 otherwise

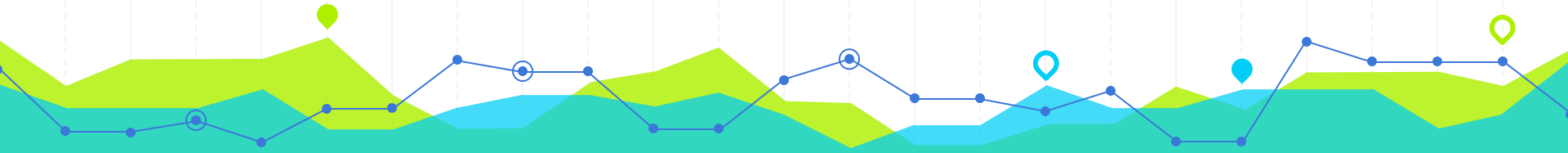
orders.csv rows: 3421083

206209 unique users!

- **order_id**: integer | order identifier
- **user_id**: integer | customer identifier
- **order_number**: integer | the order sequence number for this user (1=first, n=nth)
- **order_dow**: integer | the day of the week the order was placed on
- **order_hour_of_day**: integer | the hour of the day the order was placed on
- **days_since_prior_order**: double | days since the last order, capped at 30

products.csv rows: 49688

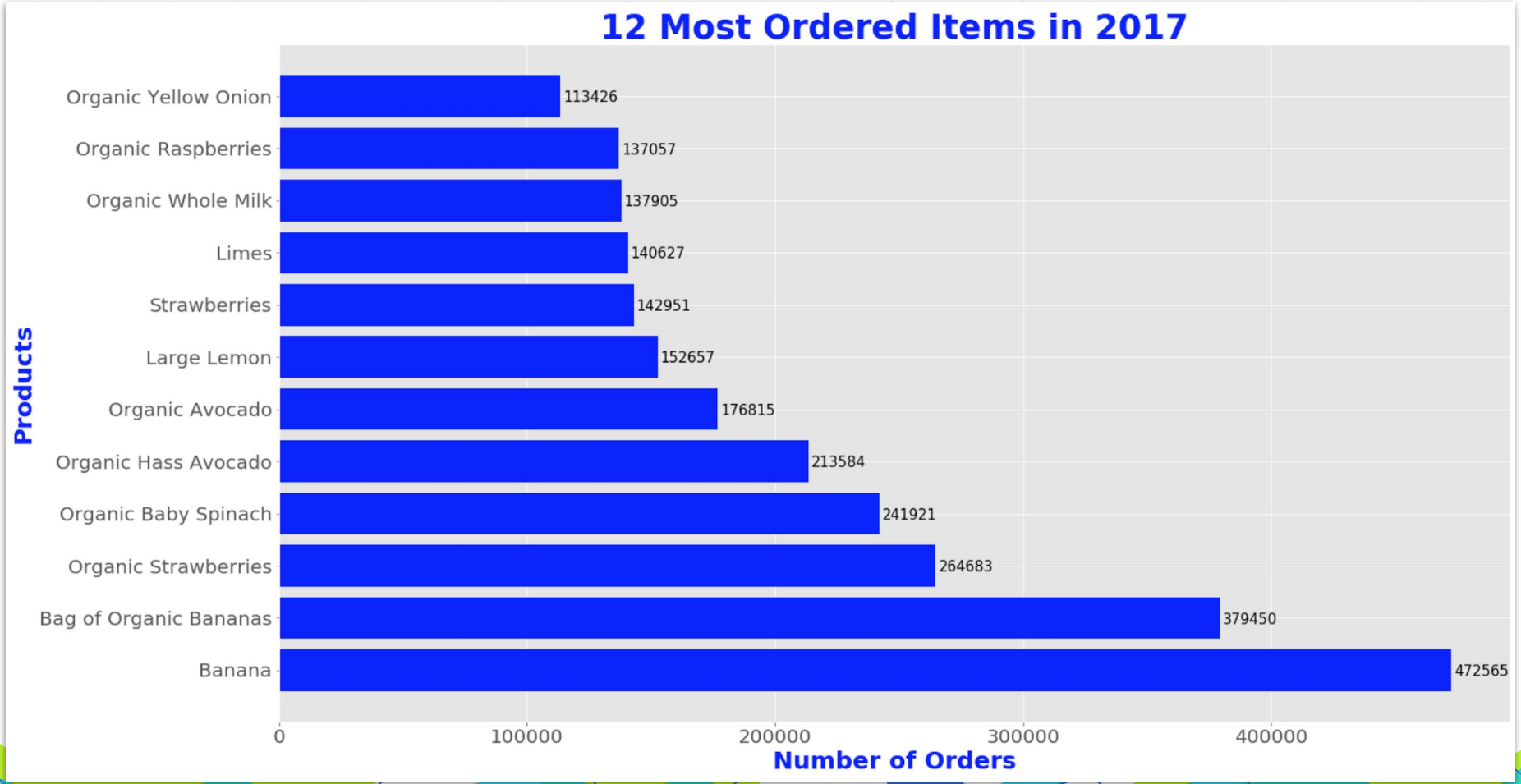
- **product_id**: integer | product identifier
- **product_name**: string | name of the product
- **aisle_id**: string | foreign key
- **department_id**: string | foreign key

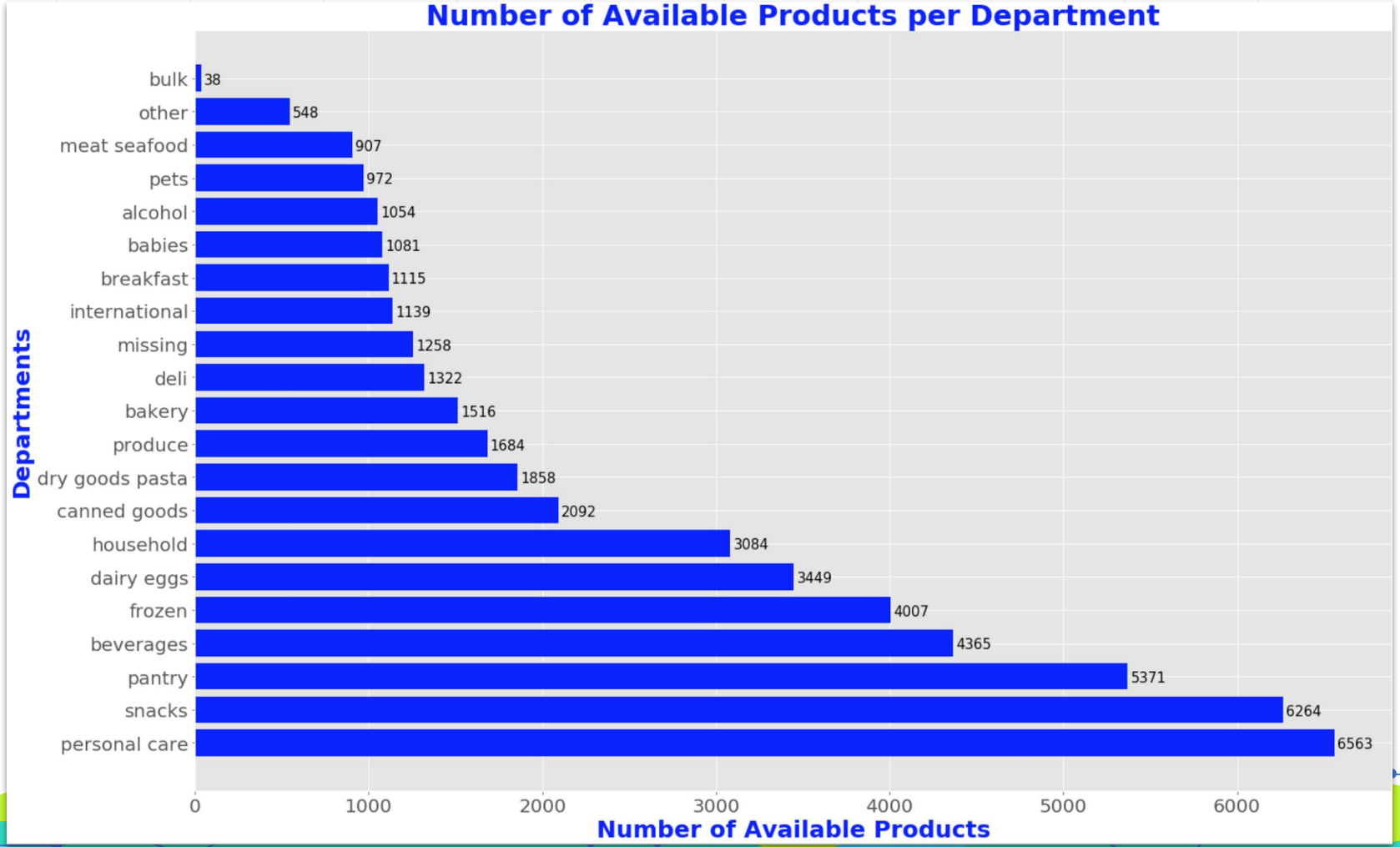


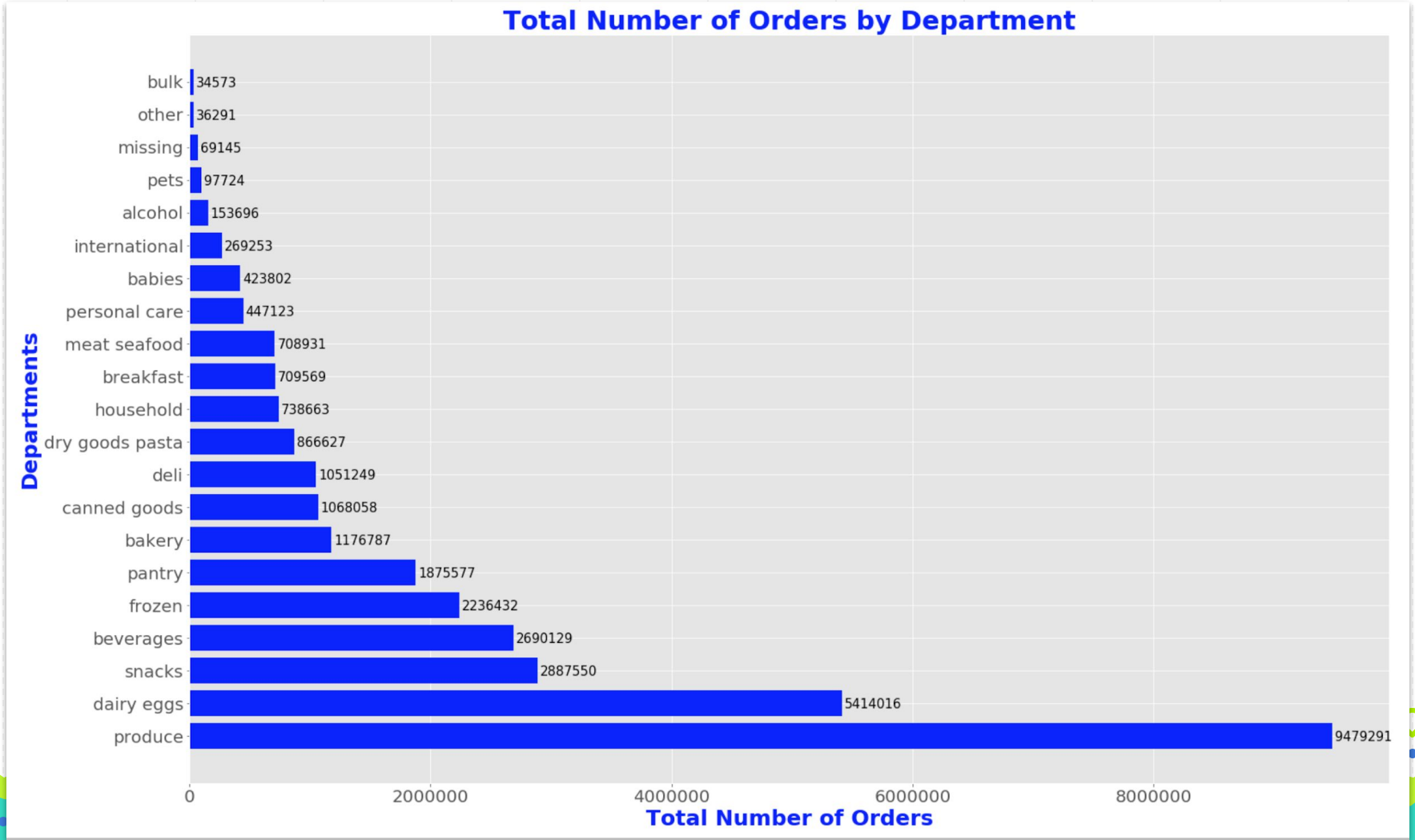
Interesting Data

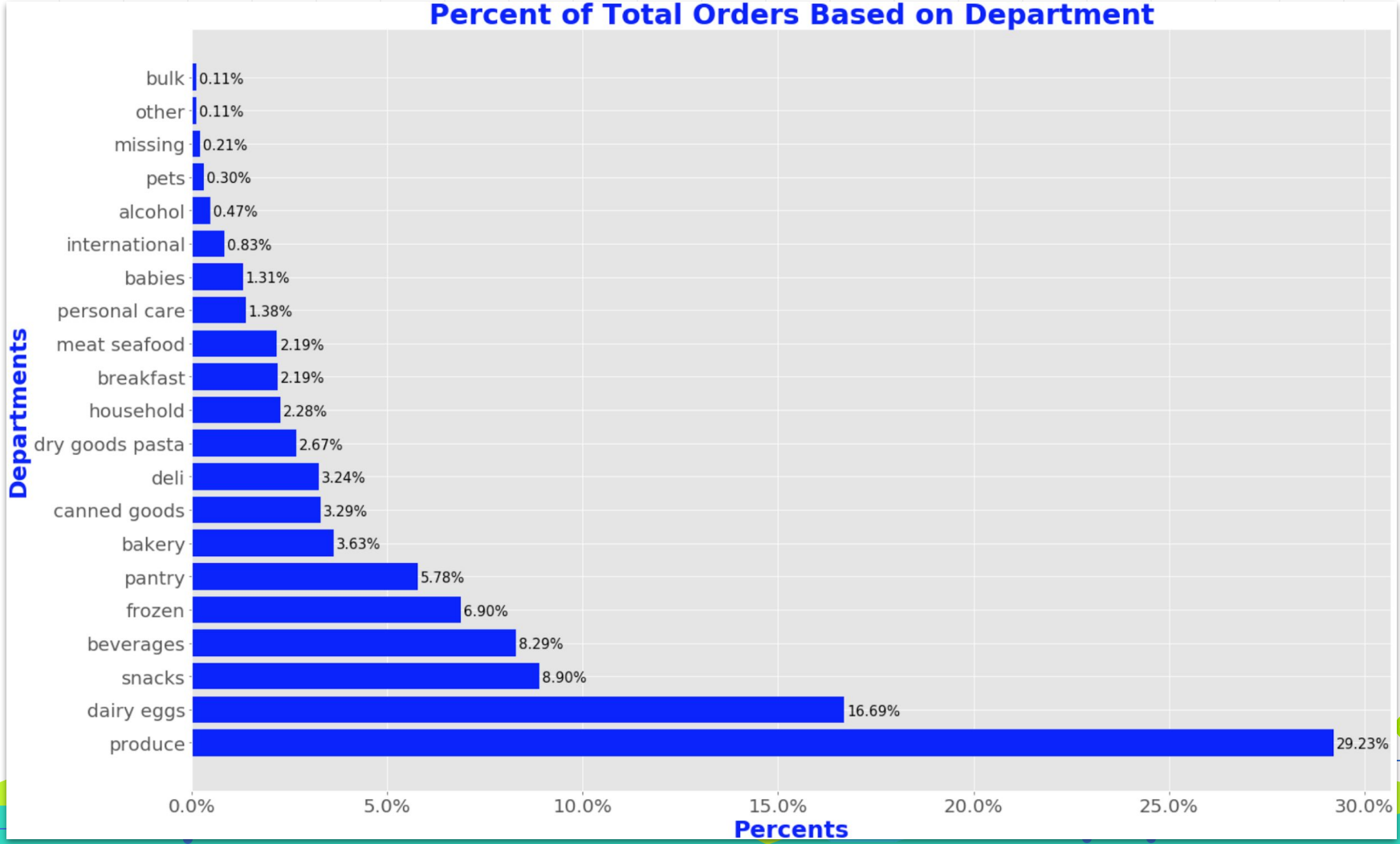
1

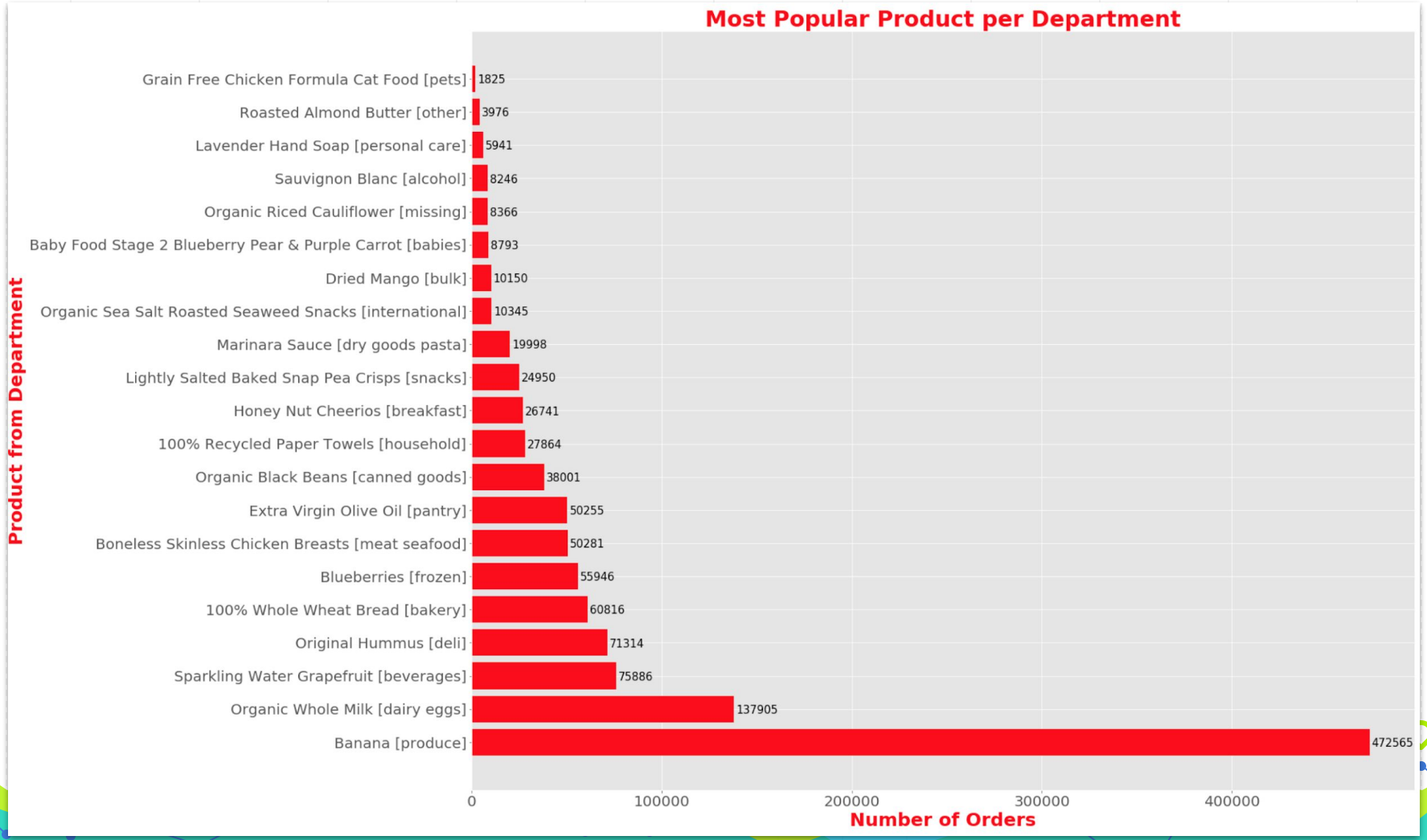
Interesting Information on Products Ordered

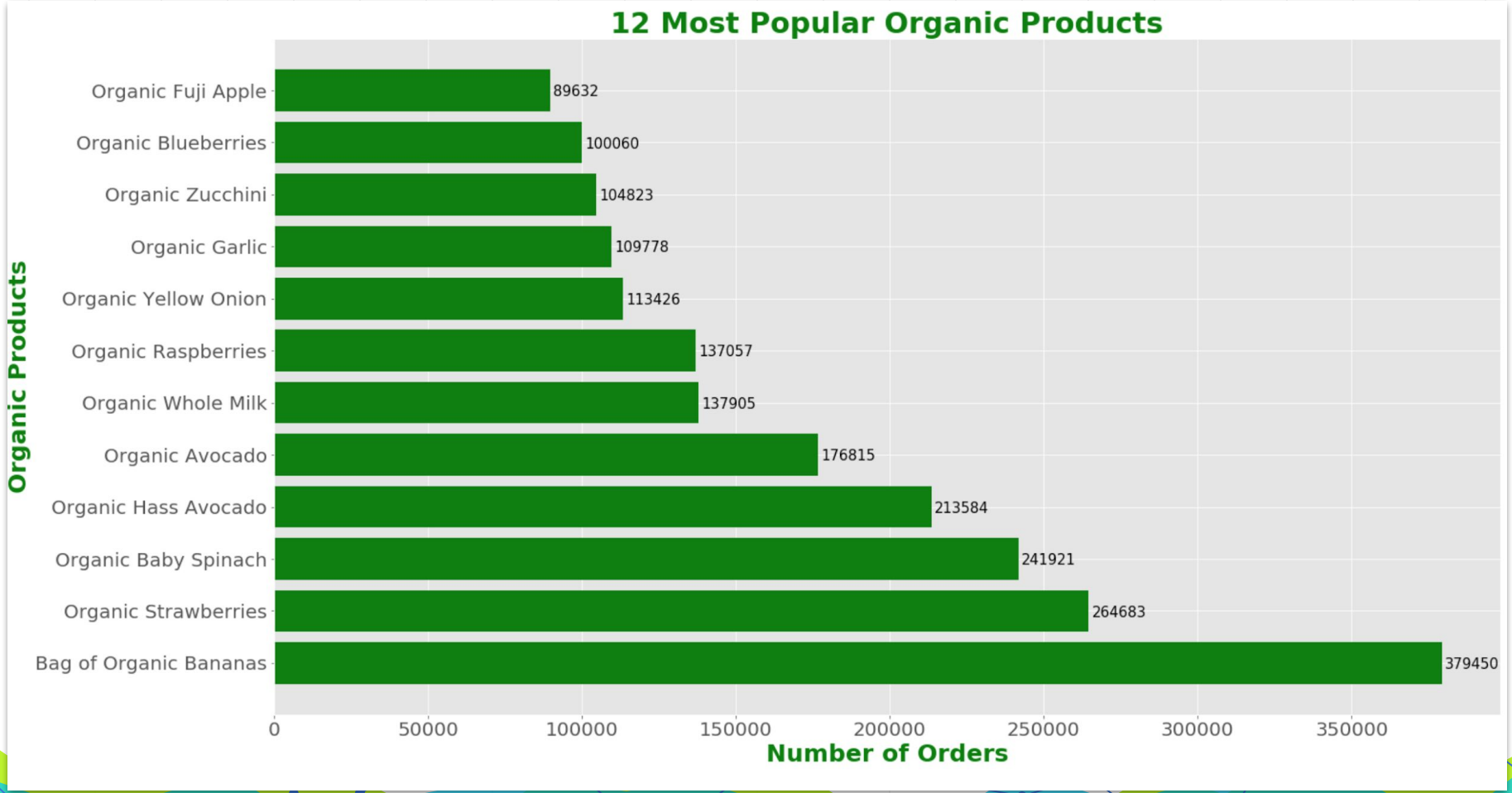

12 Most Ordered Items in 2017
Organic Yellow Onion
Organic Raspberries
Organic Whole Milk
Limes
Strawberries
Large Lemon
Organic Avocado
Organic Hass Avocado
Organic Baby Spinach
Organic Strawberries
Bag of Organic Bananas
Banana
113426
137057
137905
140627
142951
152657
176815
213584
241921
264683
379450
472565
Products
0
100000
200000
300000
400000
Number of Orders

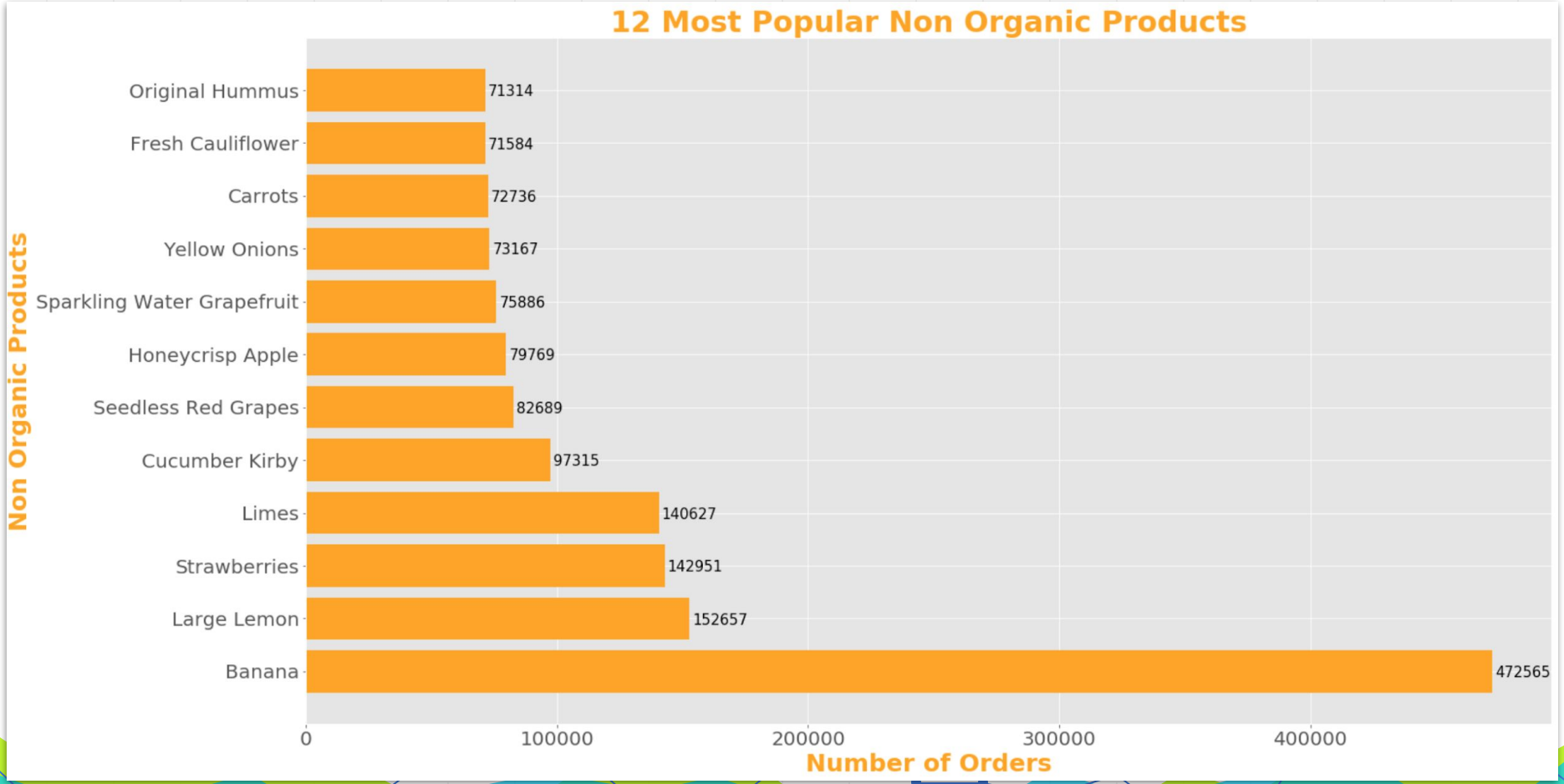

Number of Available Products per Department
Departments
bulk 38
other 548
meat seafood 907
pets 972
alcohol 1054
babies 1081
breakfast 1115
international 1139
missing 1258
deli 1322
bakery 1516
produce 1684
dry goods pasta 1858
canned goods 2092
household 3084
dairy eggs 3449
frozen 4007
beverages 4365
pantry 5371
snacks 6264
personal care 6563
0
1000
2000
3000
4000
5000
6000
Number of Available Products

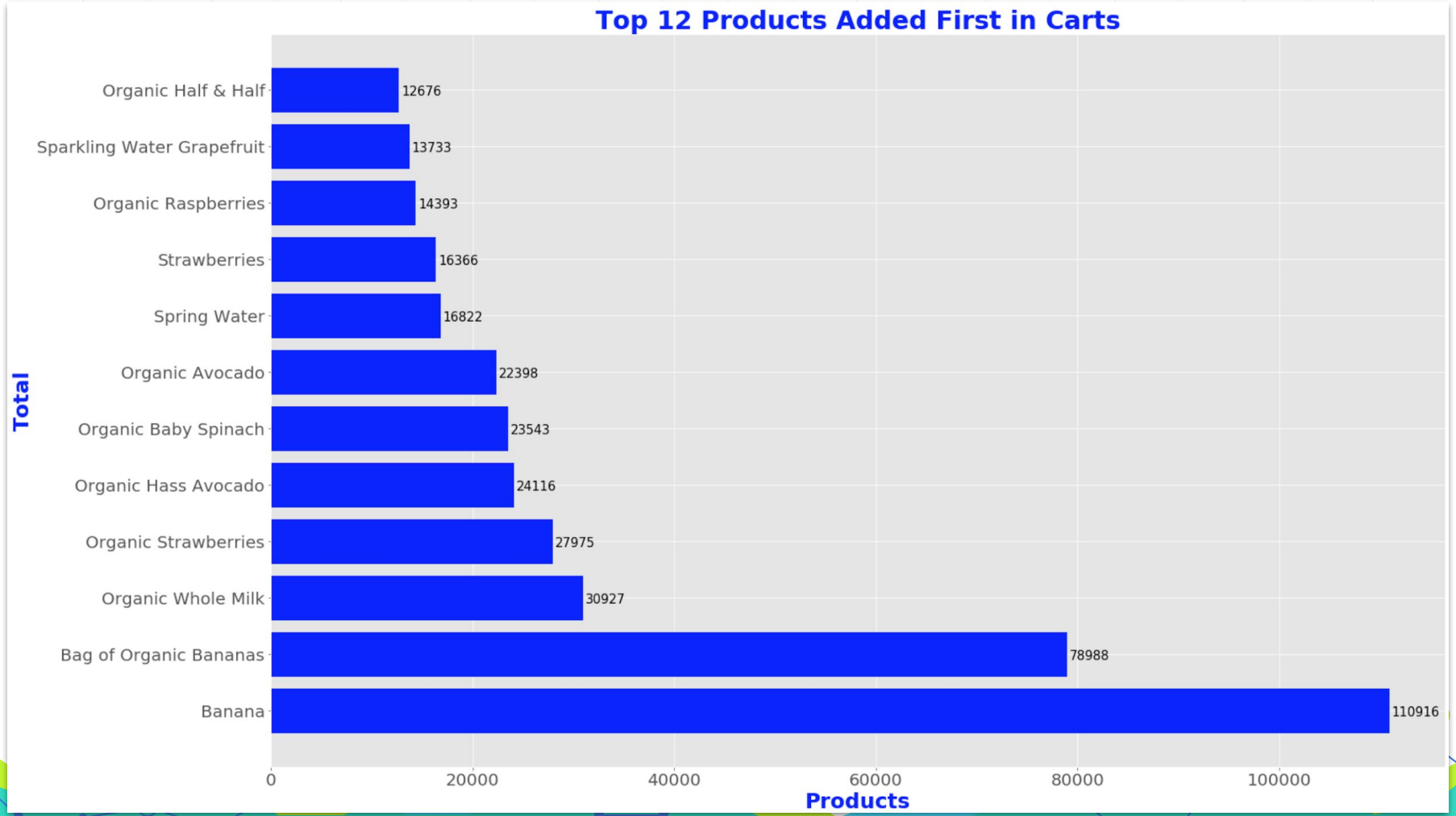

Total Number of Orders by Department
Departments
bulk 34573
other 36291
missing 69145
pets 97724
alcohol 153696
international 269253
babies 423802
personal care 447123
meat seafood 708931
breakfast 709569
household 738663
dry goods pasta 866627
deli 1051249
canned goods 1068058
bakery 1176787
pantry 1875577
frozen 2236432
beverages 2690129
snacks 2887550
dairy eggs 5414016
produce 9479291
0
2000000
4000000
6000000
8000000
Total Number of Orders

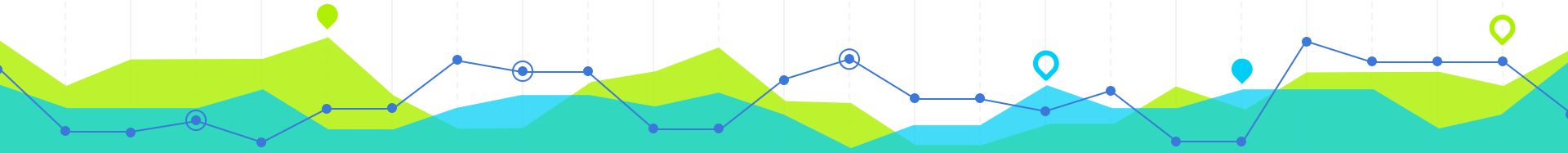

Percent of Total Orders Based on Department
Departments
bulk 0.11%
other 0.11%
missing 0.21%
pets 0.30%
alcohol 0.47%
international 0.83%
babies 1.31%
personal care 1.38%
meat seafood 2.19%
breakfast 2.19%
household 2.28%
dry goods pasta 2.67%
deli 3.24%
canned goods 3.29%
bakery 3.63%
pantry 5.78%
frozen 6.90%
beverages 8.29%
snacks 8.90%
dairy eggs 16.69%
produce 29.23%
0.0% 5.0% 10.0% 15.0% 20.0% 25.0% 30.0%
Percents


Most Popular Product per Department
Product from Department
Grain Free Chicken Formula Cat Food [pets]
Roasted Almond Butter [other]
Lavender Hand Soap [personal care]
Sauvignon Blanc [alcohol]
Organic Riced Cauliflower [missing]
Baby Food Stage 2 Blueberry Pear & Purple Carrot [babies]
Dried Mango [bulk]
Organic Sea Salt Roasted Seaweed Snacks [international]
Marinara Sauce [dry goods pasta]
Lightly Salted Baked Snap Pea Crisps [snacks]
Honey Nut Cheerios [breakfast]
100% Recycled Paper Towels [household]
Organic Black Beans [canned goods]
Extra Virgin Olive Oil [pantry]
Boneless Skinless Chicken Breasts [meat seafood]
Blueberries [frozen]
100% Whole Wheat Bread [bakery]
Original Hummus [deli]
Sparkling Water Grapefruit [beverages]
Organic Whole Milk [dairy eggs]
Banana [produce]
1825
3976
5941
8246
8366
8793
10150
10345
19998
24950
26741
27864
38001
50255
50281
55946
60816
71314
75886
137905
472565
0
100000
200000
300000
400000
Number of Orders


12 Most Popular Organic Products
Organic Products
Organic Fuji Apple
Organic Blueberries
Organic Zucchini
Organic Garlic
Organic Yellow Onion
Organic Raspberries
Organic Whole Milk
Organic Avocado
Organic Hass Avocado
Organic Baby Spinach
Organic Strawberries
Bag of Organic Bananas
89632
100060
104823
109778
113426
137057
137905
176815
213584
241921
264683
379450
0
50000
100000
150000
200000
250000
300000
350000
Number of Orders


12 Most Popular Non Organic Products
Non Organic Products
Original Hummus
Fresh Cauliflower
Carrots
Yellow Onions
Sparkling Water Grapefruit
Honeycrisp Apple
Seedless Red Grapes
Cucumber Kirby
Limes
Strawberries
Large Lemon
Banana
71314
71584
72736
73167
75886
79769
82689
97315
140627
142951
152657
472565
0
100000
200000
300000
400000
Number of Orders


Top 12 Products Added First in Carts
Organic Half & Half
12676
Sparkling Water Grapefruit
13733
Organic Raspberries
14393
Strawberries
16366
Spring Water
16822
Organic Avocado
22398
Organic Baby Spinach
23543
Organic Hass Avocado
24116
Organic Strawberries
27975
Organic Whole Milk
30927
Bag of Organic Bananas
78988
Banana
110916
Total
0
20000
40000
60000
80000
100000
Products

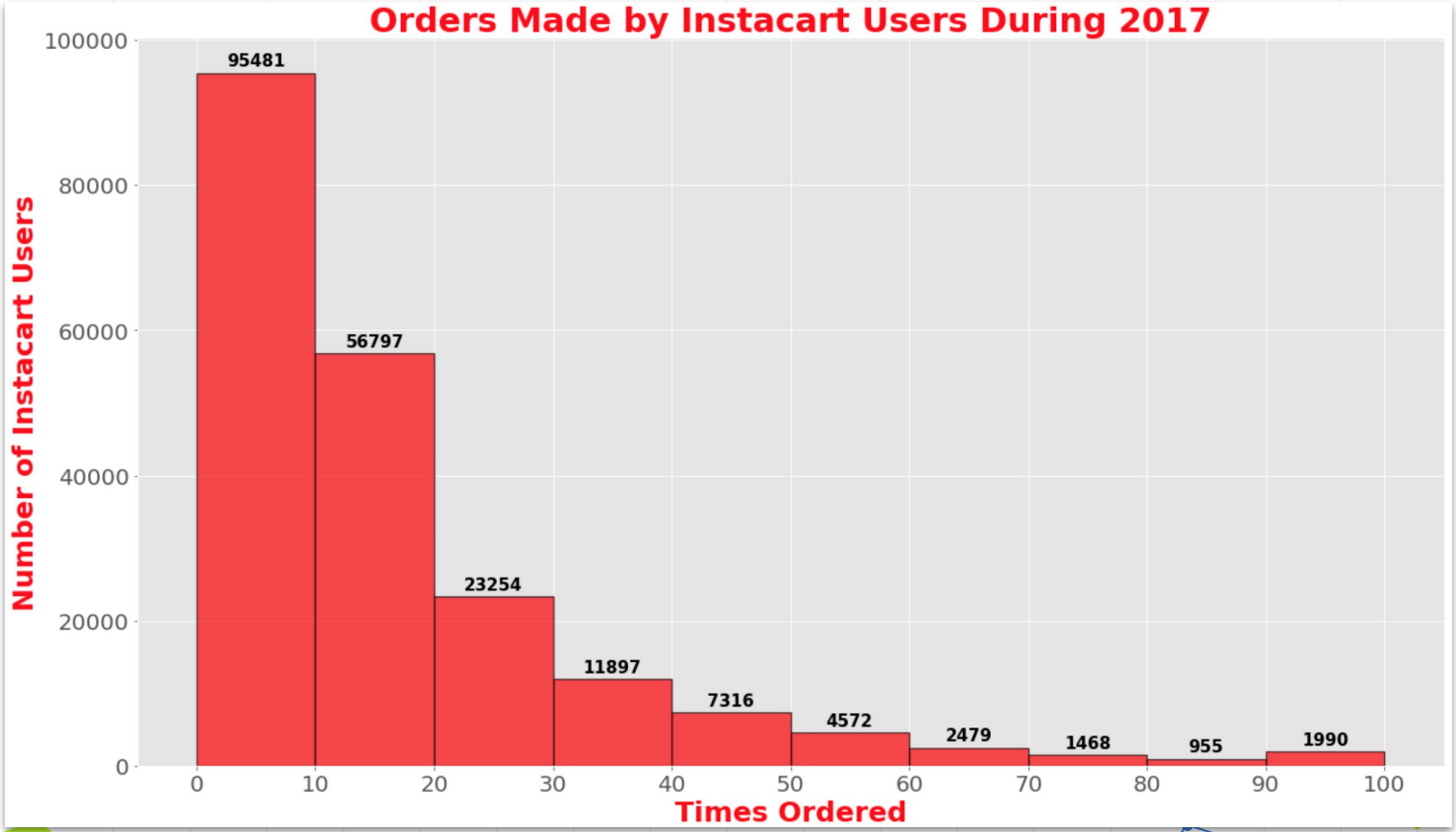


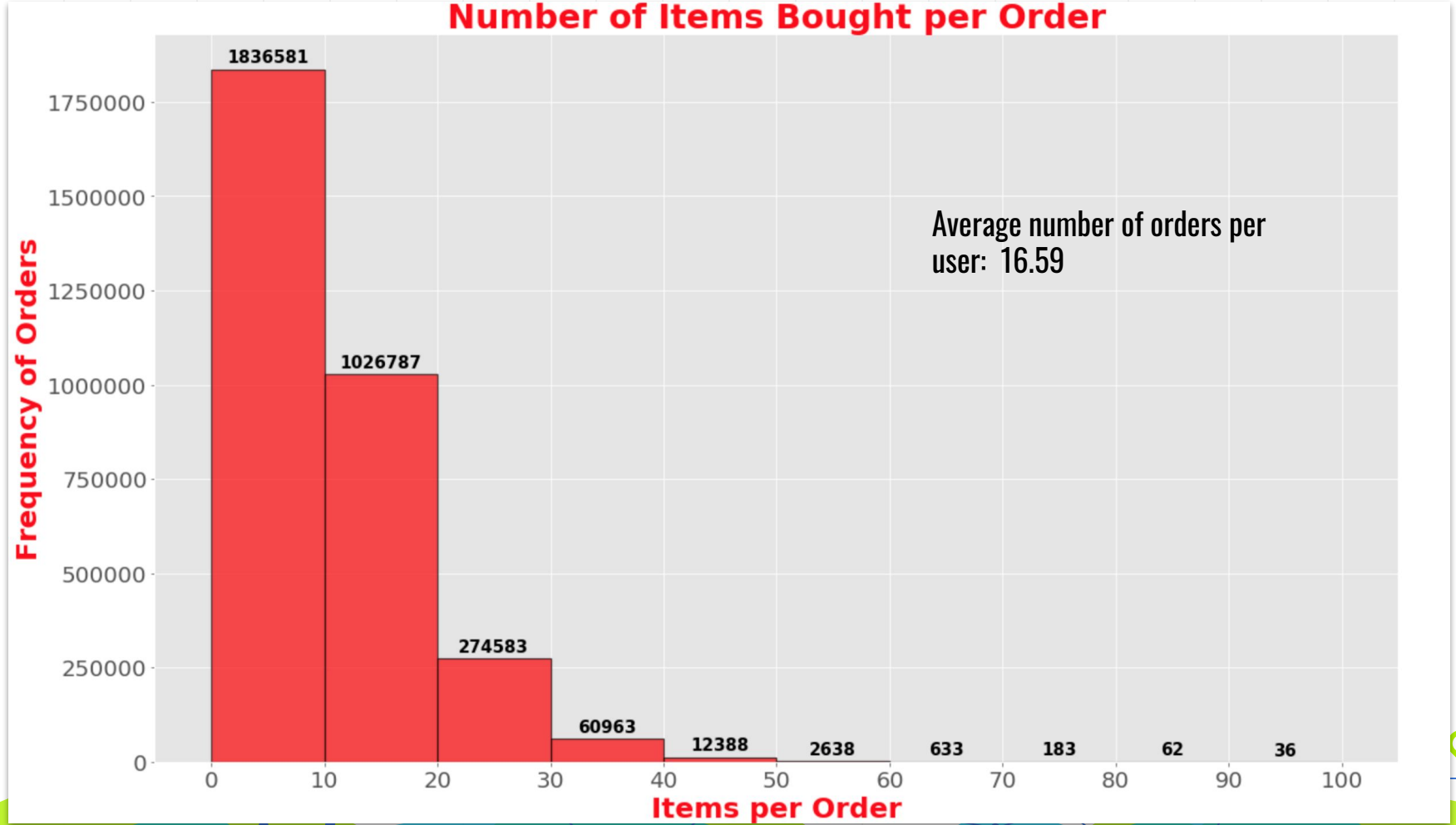
Frequency of Orders

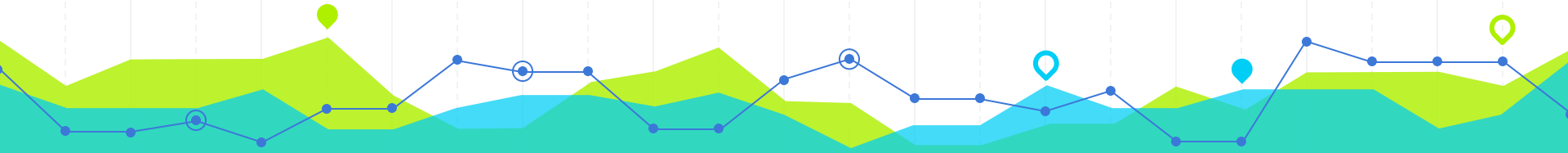
2

How many and how often do users order?

- Number of orders are throughout 2017
- Don’t have information on when users were created, hence a high 0-10 bin



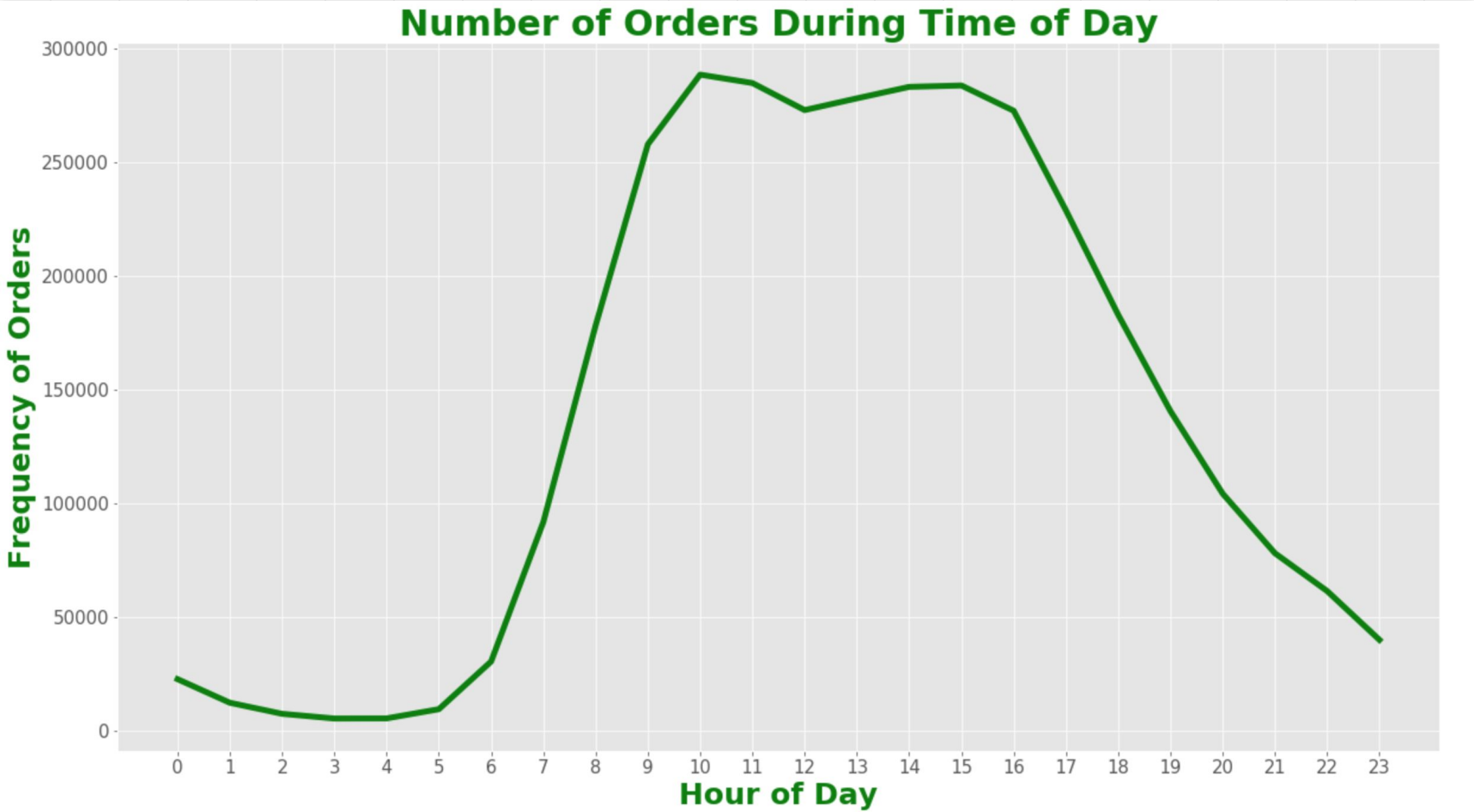

Number of Items Bought per Order
Frequency of Orders
Items per Order
1836581
1026787
274583
60963
12388
2638
633
183
62
36
0
250000
500000
750000
1000000
1250000
1500000
1750000
0
10
20
30
40
50
60
70
80
90
100
Average number of orders per user: 16.59



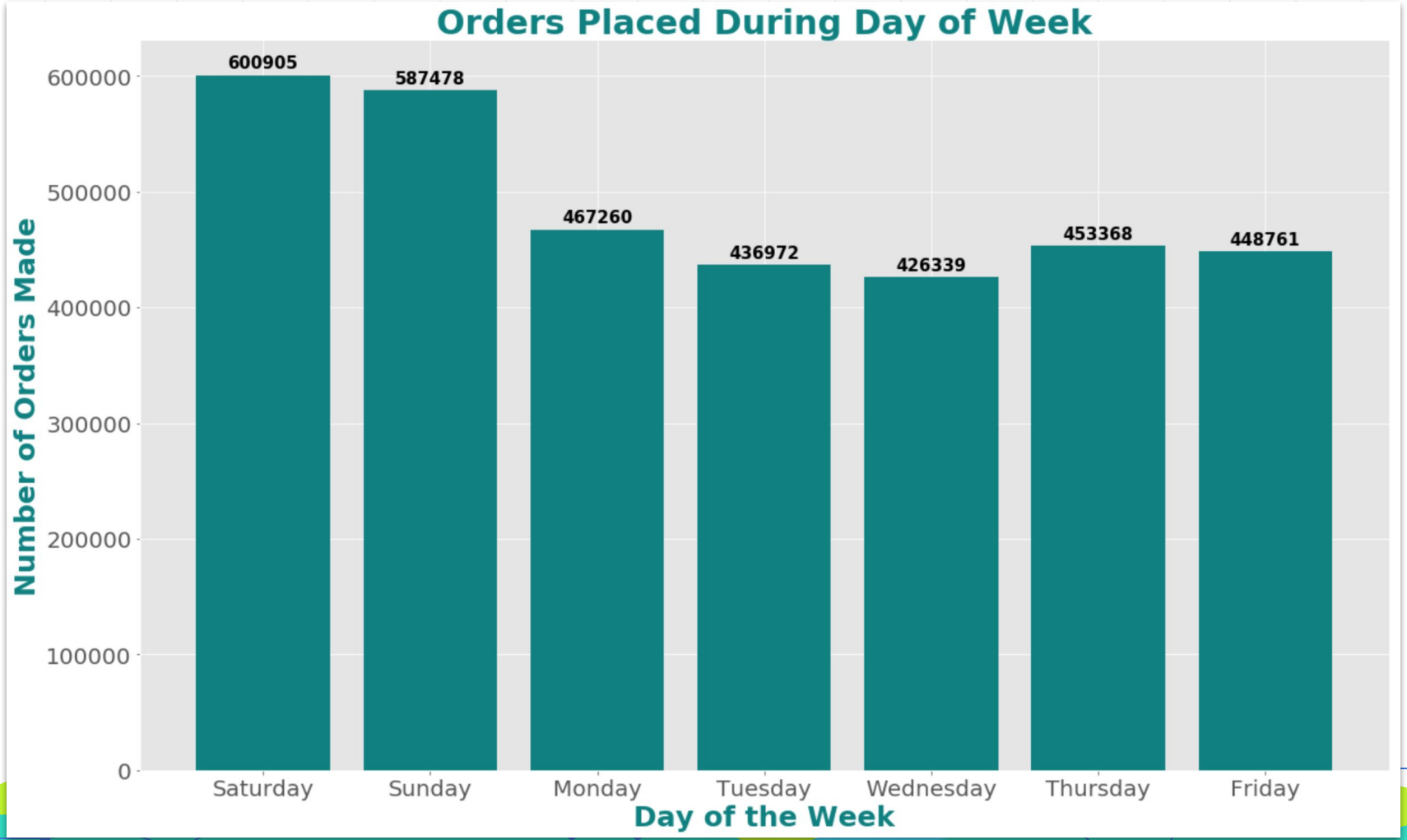
Time and Day for Ordering

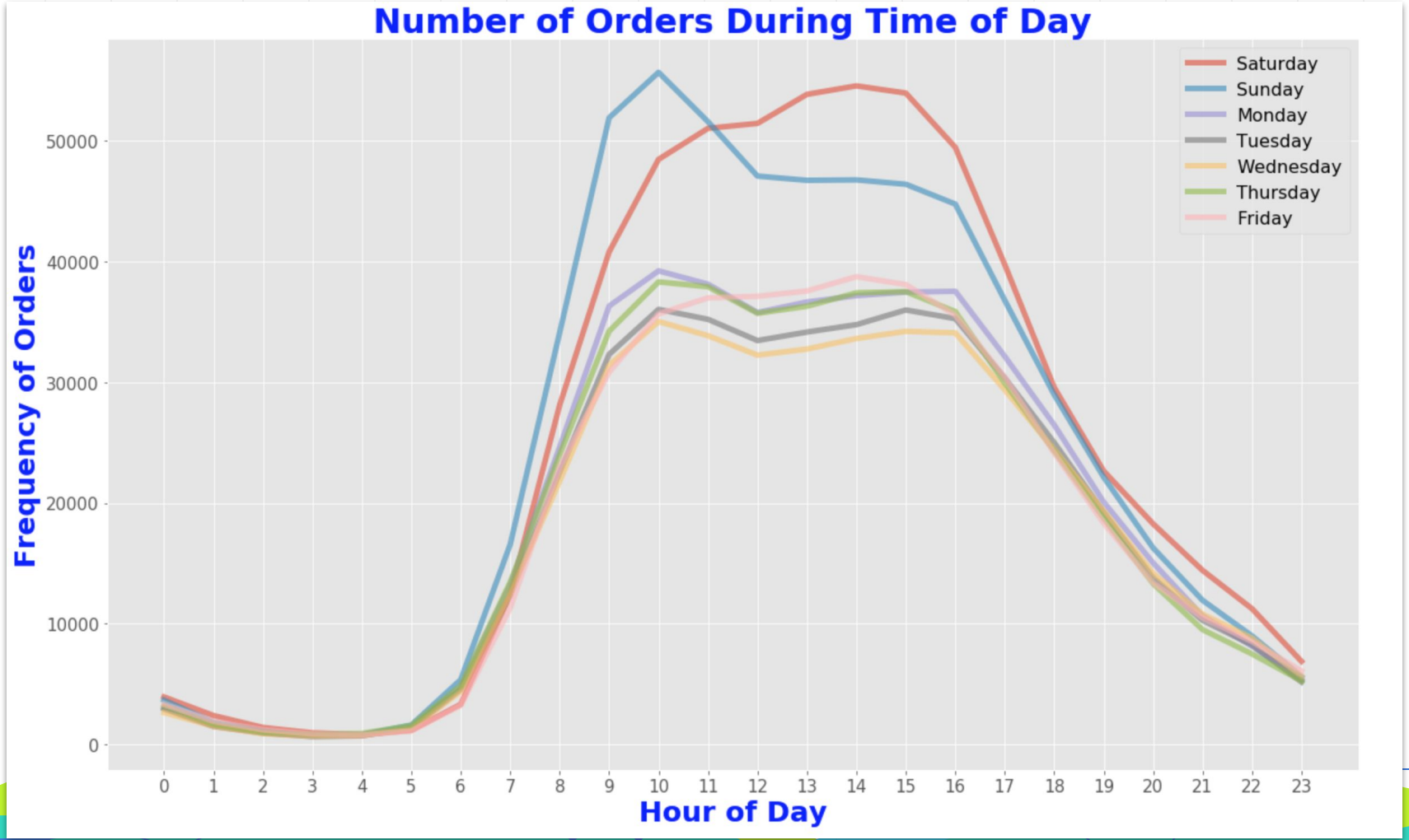
3

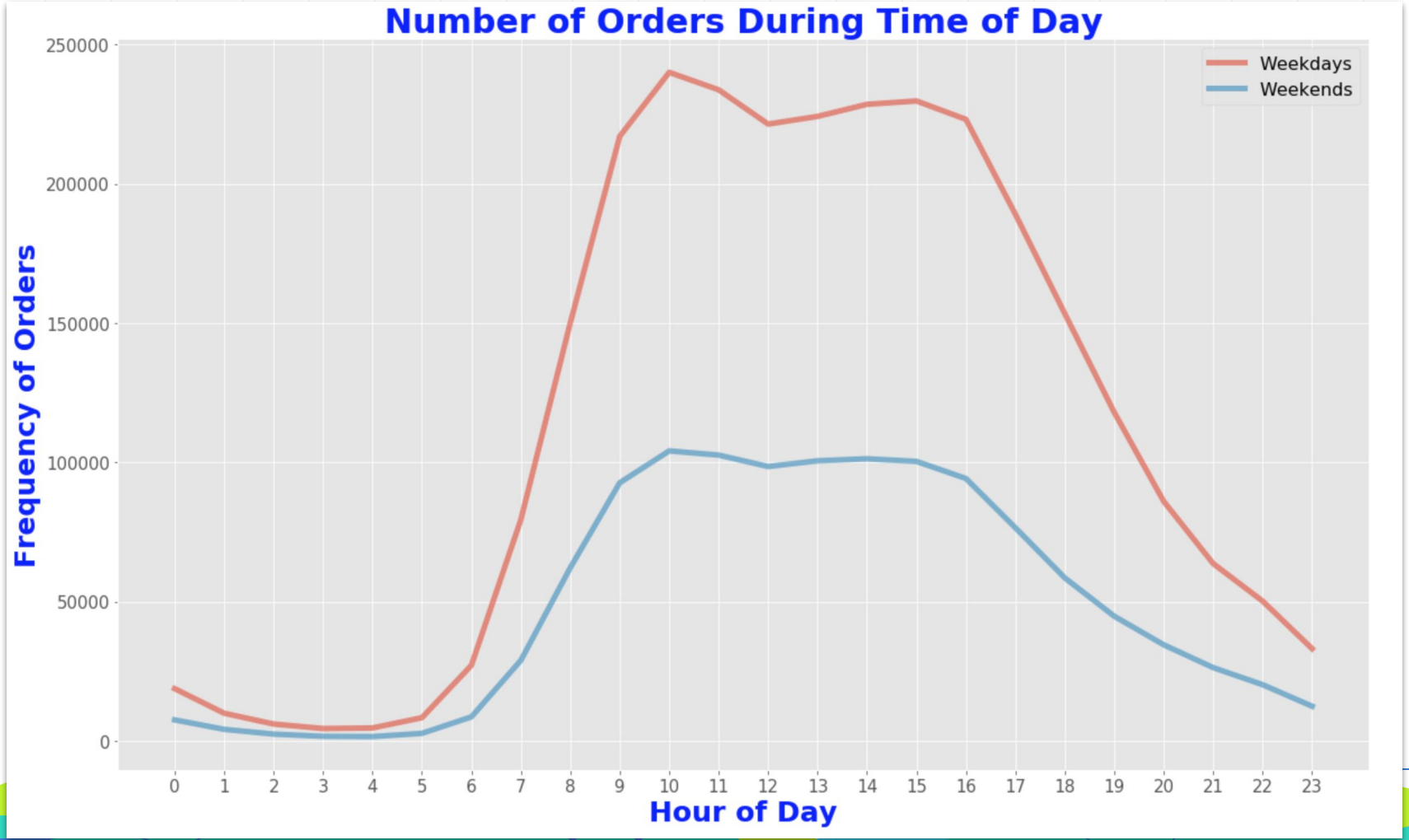
When do users do their ordering?

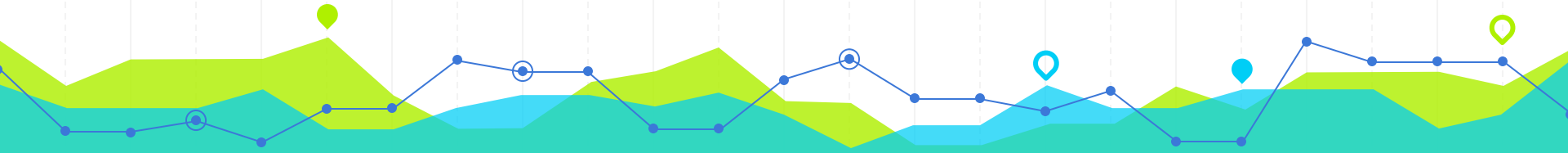


Most Idle Time of Day: 3:00 with a total of 5474 orders placed
Busiest Time of Day: 10:00 with a total of 288418 orders placed


Orders Placed During Day of Week
Number of Orders Made
600000
500000
400000
300000
200000
100000
0
600905
587478
467260
436972
426339
453368
448761
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Day of the Week


Number of Orders During Time of Day
Frequency of Orders
Hour of Day
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
0
10000
20000
30000
40000
50000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23


Number of Orders During Time of Day
Frequency of Orders
250000
200000
150000
100000
50000
0
Weekdays
Weekends
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day

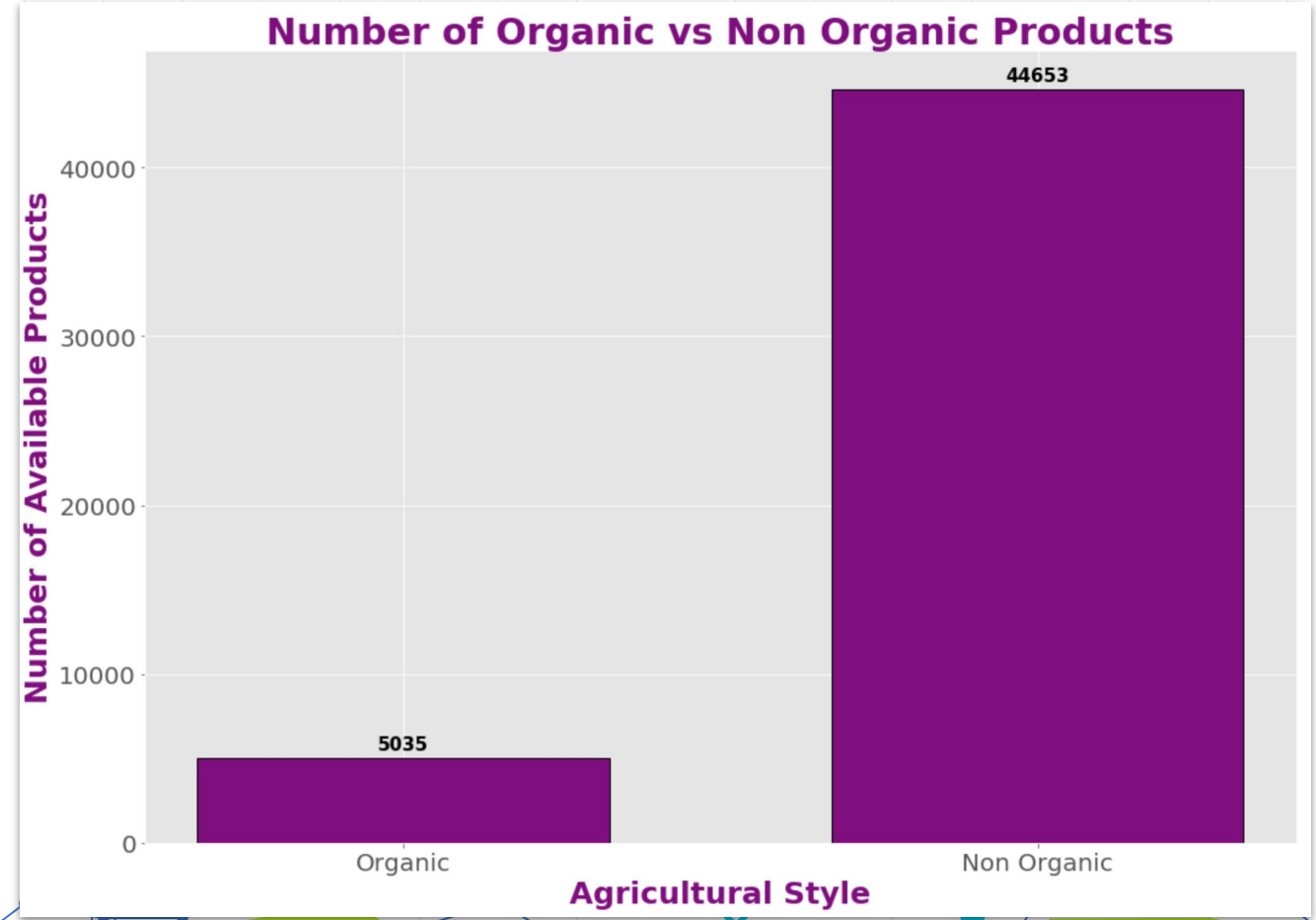


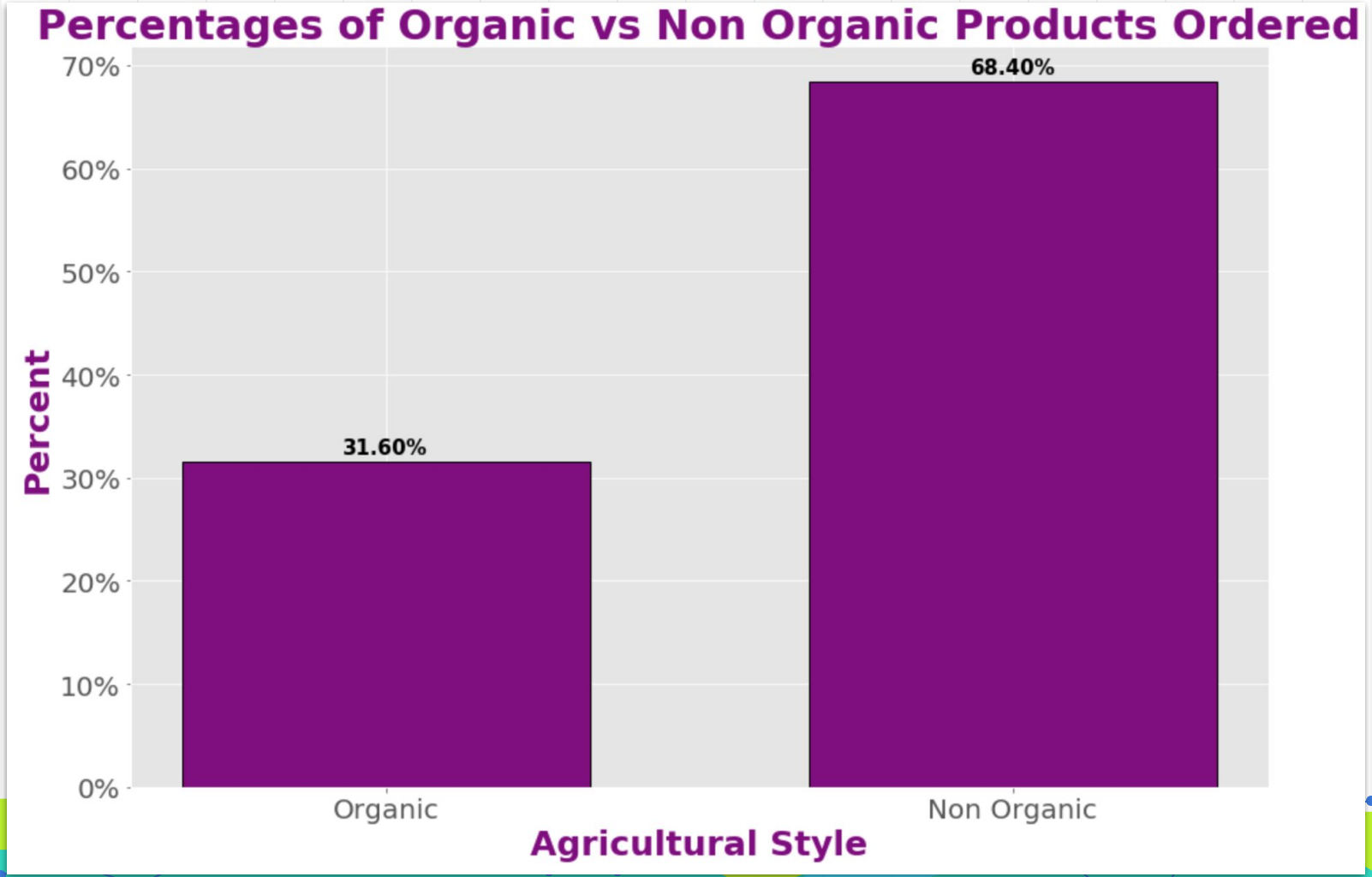
ORGANIC VS NON-ORGANIC

4

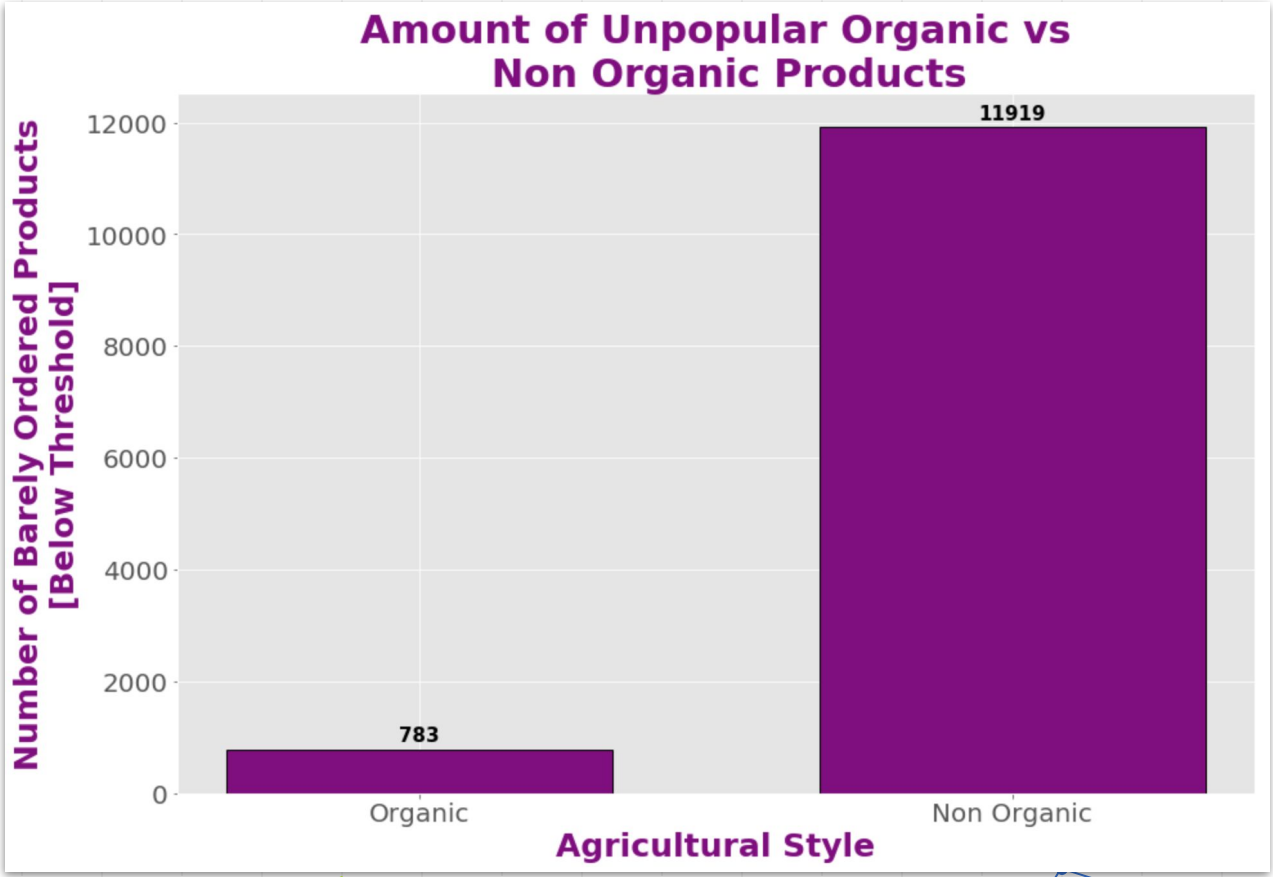
Which do customers prefer?

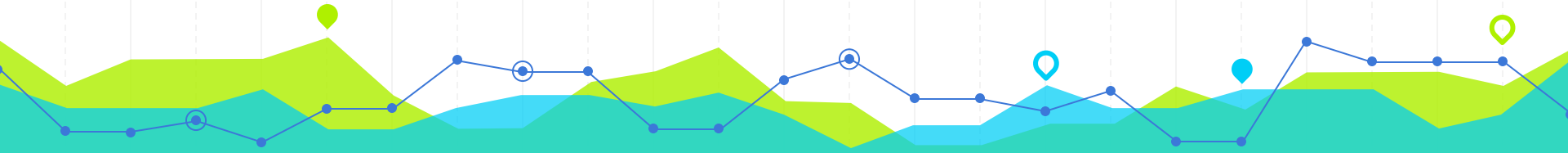
- Organic items are specifically worded with the keyword ‘organic’
- Non Organic includes every product without keyword ‘organic’




Percentages of Organic vs Non Organic Products Ordered
70%
60%
50%
40%
30%
20%
10%
0%
Percent
31.60%
68.40%
Organic
Non Organic
Agricultural Style

- Curious about the least popular items offered.
- Some items were ordered 0 times
- Don’t know when items were added for availability
- Added the products, took first quartile
    - Any item that were ordered less than quartile were considered unpopular





Hypothesis 5

Hypothesis

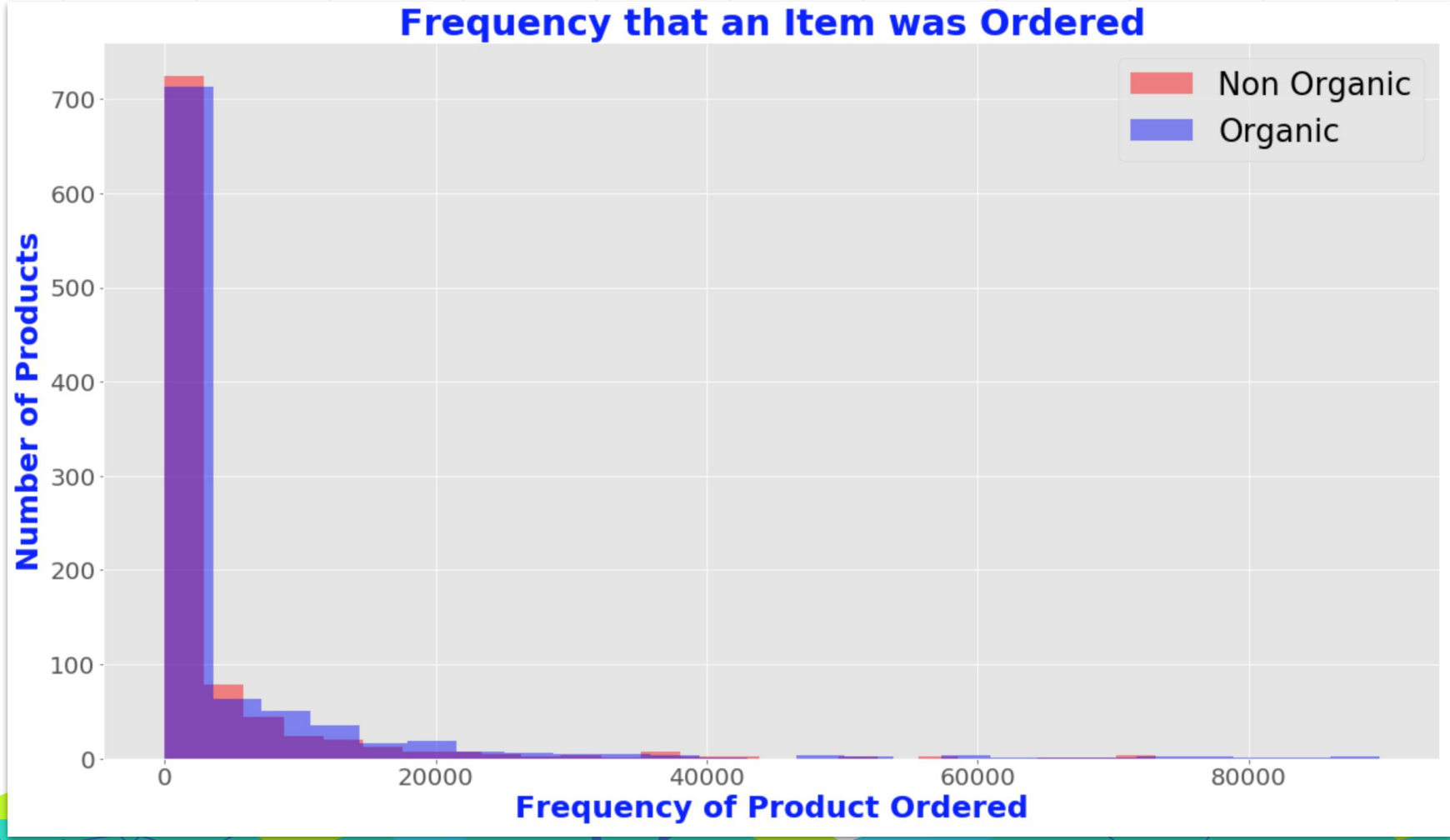
H_0:	People do NOT prefer ordering organic products
H_a:	People DO prefer ordering organic products

Alpha value = 0.05

Visualization of Filtering Process

Problems and Process:

- Finding items that are offered in both non-organic and organic.
- Too many nuances in words to properly go through:
  - Ex. Bag of Organic Bananas and Banana
- Had to filter and filter and filter again.


Frequency that an Item was Ordered
Non Organic
Organic
Number of Products
700
600
500
400
300
200
100
0
0
20000
40000
60000
80000
Frequency of Product Ordered

Hypothesis Testing

- Total of Organic Products Ordered: 5959541
- Total of NON Organic Products Ordered: 4169490

- Organic products ordered: 58.83%
  - Mean: 6253.45
- NON organic products ordered: 41.16%
  - Mean: 4375.12

Based on T-Test:

- T-Statistic: 2.1266
- P-value: 0.03358

Since p<a:

- Reject the Null Hypothesis
  - Therefore, people DO prefer organic products!

THANKS!

Any questions?

You can find me at

github.com/winrichsy