NOVA IMS Information Management School

Instacart:
Market Basket Analysis

Group Y:

João Alves Henriques;

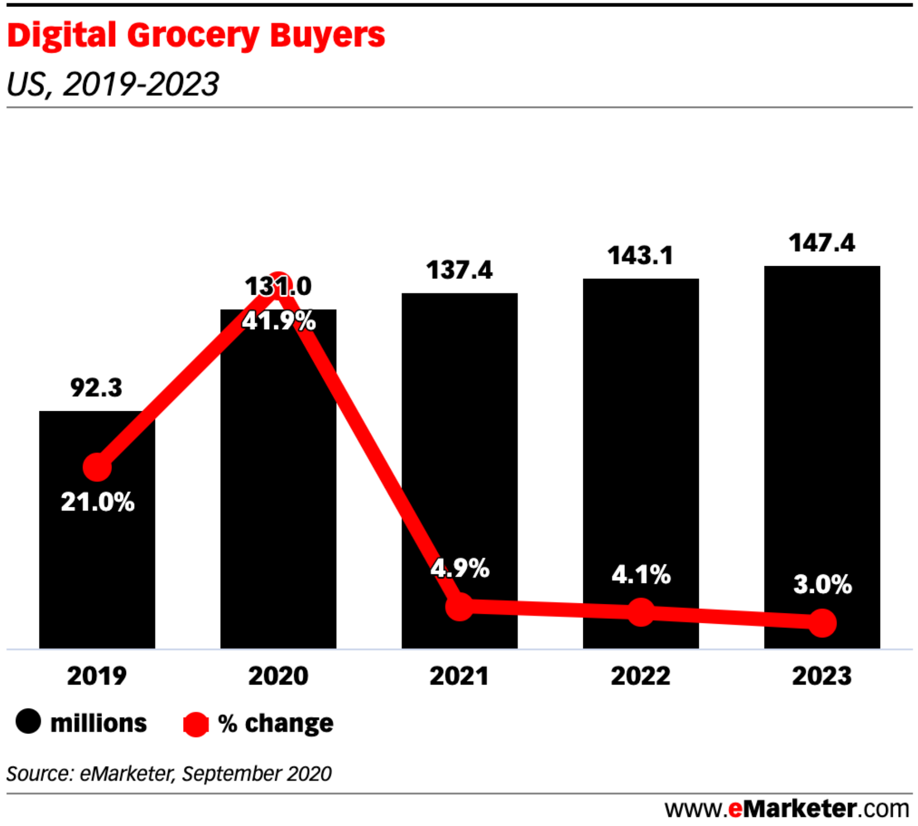
João Paulo César

Pedro Sancho

Vilmar Adriano Bussolaro

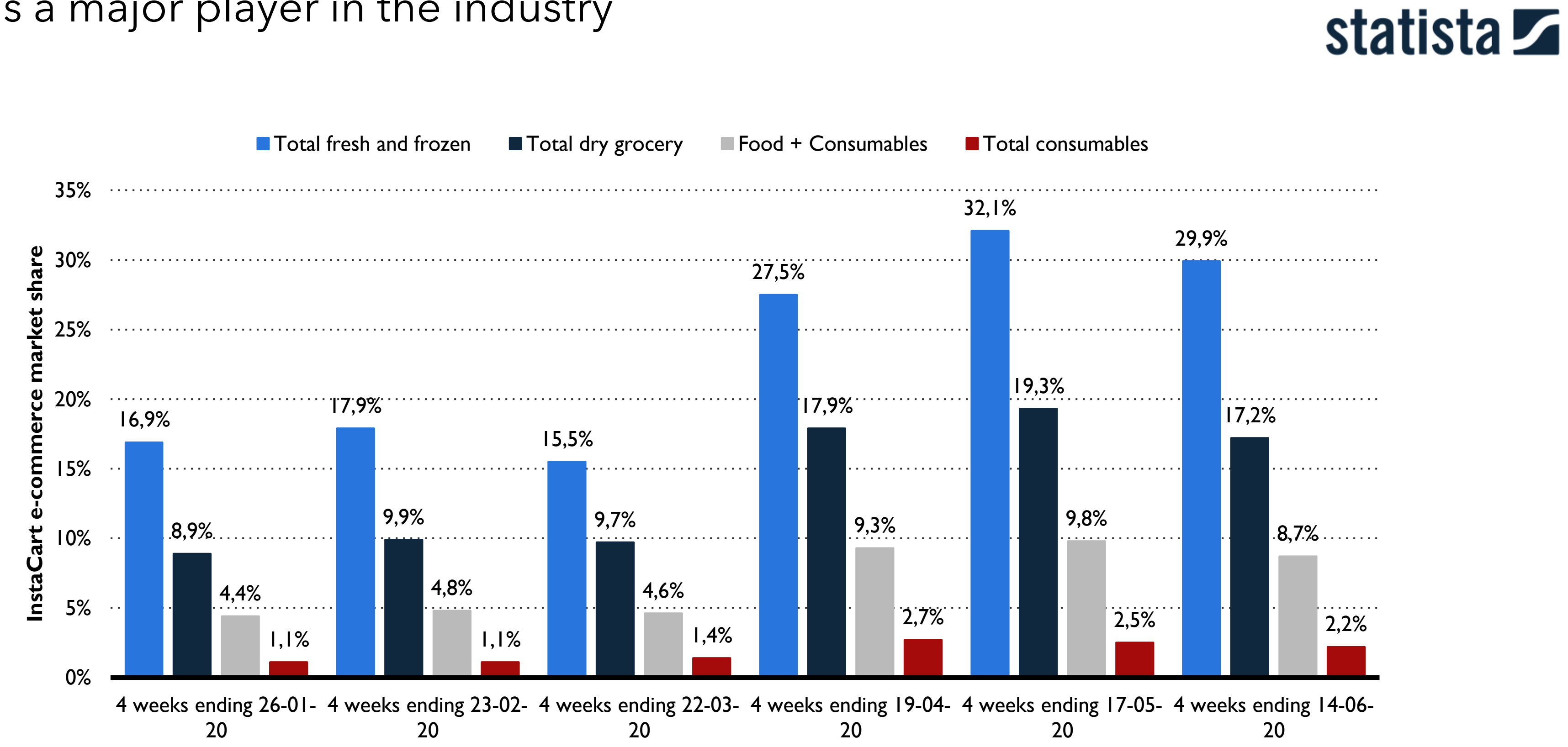
Context

- COVID-19 has accelerated the trend on e-grocery



Context

- COVID-19 has accelerated the trend on e-grocery
- Instacart is a major player in the industry



Instacart e-commerce market share during the coronavirus pandemic in the US from January to April 2020

Context

- ***Situation:***

Instacart wants to better understand consumer buying behavior:

Market Basket Analysis with apriori algorithm

- Which types of products should have an extended amount of product offerings?
- Which types of products can be seen as substitutes?
- Which items are complementary?

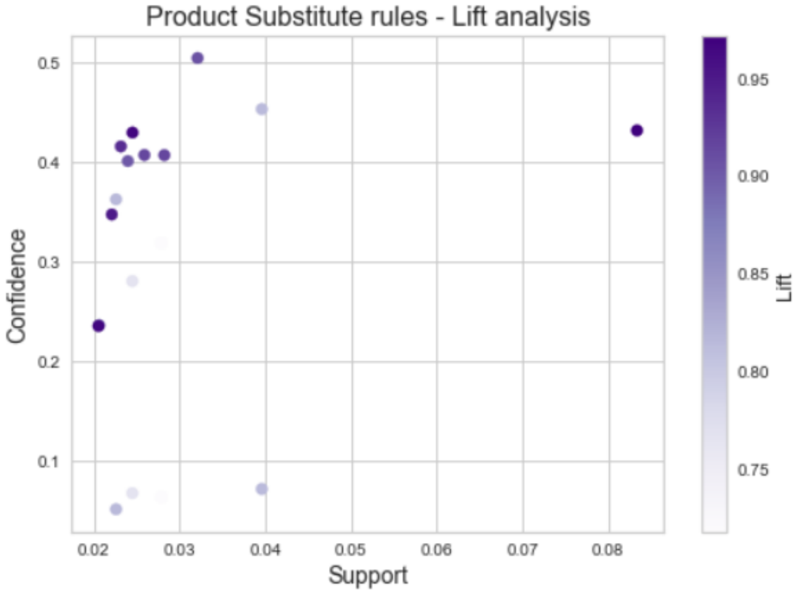
Clustering with k-means

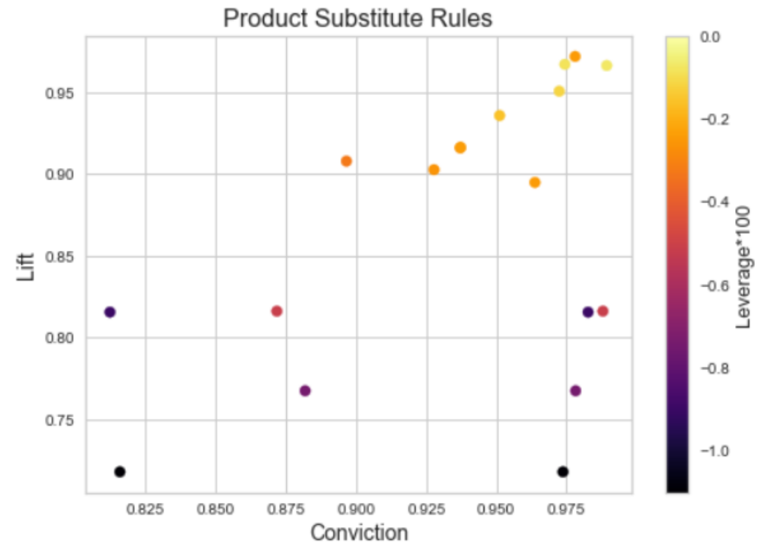
- What are the main types of consumer behavior in the business?

SUBSTITUTE PRODUCTS – Rules for fine-tuning the algorithm

<u>Lift < 1.0</u>
<u>Conviction < 0.95</u>

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
460	(fresh vegetables)	(soft drinks)	0.444360	0.087310	0.027845	0.062663	0.717709	-0.010952	0.973705
461	(soft drinks)	(fresh vegetables)	0.087310	0.444360	0.027845	0.318921	0.717709	-0.010952	0.815823
643	(soft drinks)	(packaged vegetables fruits)	0.087310	0.365415	0.024480	0.280380	0.767293	-0.007424	0.881834
642	(packaged vegetables fruits)	(soft drinks)	0.365415	0.087310	0.024480	0.066992	0.767293	-0.007424	0.978223
376	(fresh fruits)	(soft drinks)	0.555995	0.087310	0.039585	0.071197	0.815447	-0.008959	0.982652

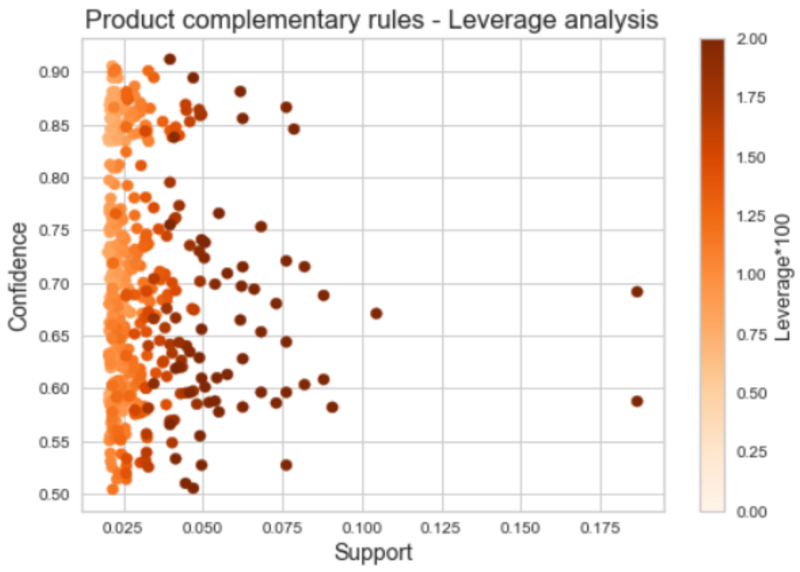


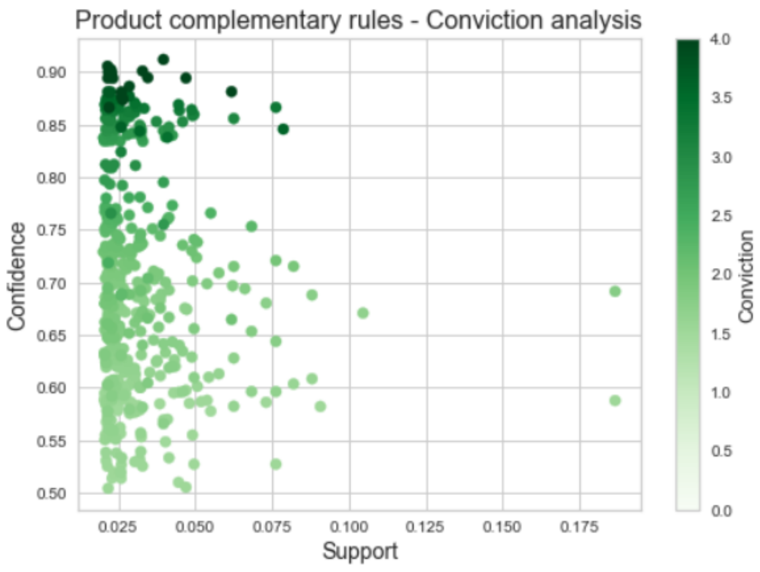


COMPLEMENTARY PRODUCTS – Rules for fine-tuning the algorithm

Confidence > 0.7
Lift > 1.5
Conviction > 1.5

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	ant_lenght	con_lenght
104	(packaged vegetables fruits, energy granola bars)	(fresh fruits)	0.039500	0.555995	0.032945	0.834051	1.500105	0.010983	2.675545	2	1
268	(eggs, fresh vegetables, milk)	(fresh fruits)	0.030660	0.555995	0.025580	0.834312	1.500574	0.008533	2.679762	3	1
261	(crackers, packaged vegetables fruits, package...	(fresh fruits)	0.024475	0.555995	0.020420	0.834321	1.500590	0.006812	2.679907	3	1
347	(packaged vegetables fruits, packaged cheese, ...	(fresh fruits)	0.025970	0.555995	0.021675	0.834617	1.501123	0.007236	2.684705	3	1
235	(bread, packaged cheese, yogurt)	(fresh fruits)	0.027480	0.555995	0.022945	0.834971	1.501760	0.007666	2.690465	3	1

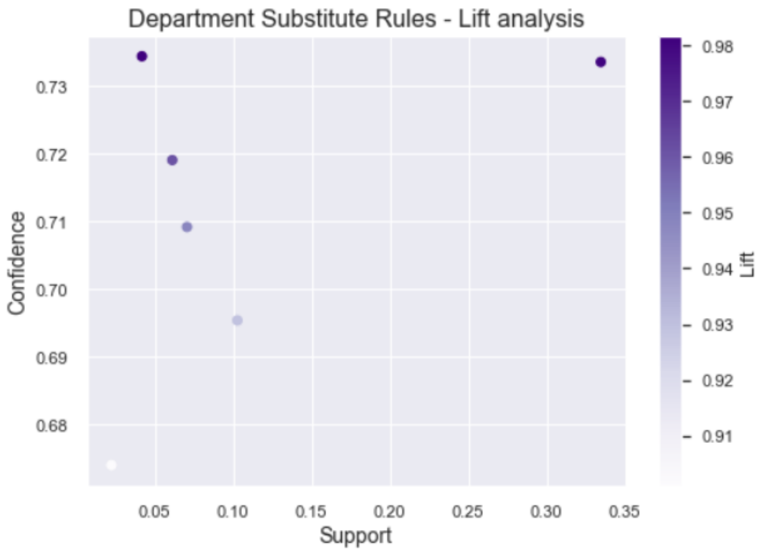


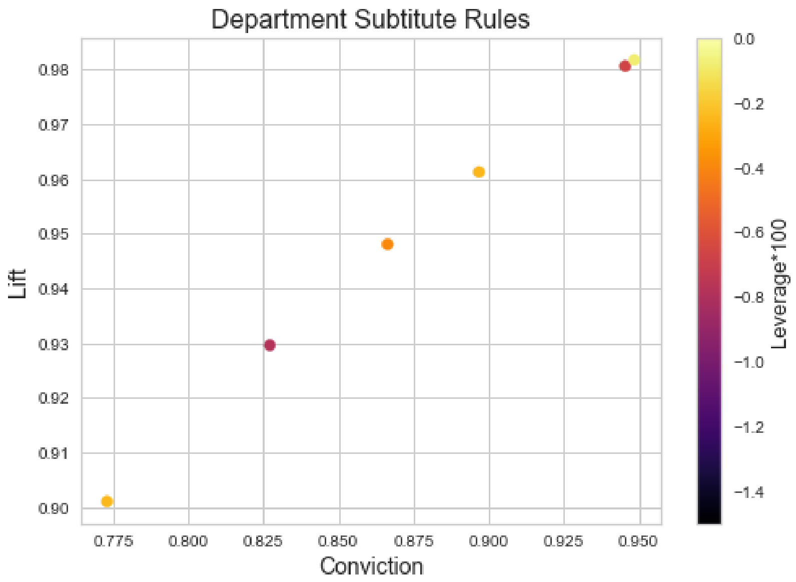


SUBSTITUTE DEPARTMENTS – Rules for fine-tuning the algorithm

Lift < 1.0
Conviction < 0.99

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2016	(personal care, household)	(produce)	0.033120	0.748065	0.022325	0.674064	0.901077	-0.002451	0.772959
187	(household)	(produce)	0.147675	0.748065	0.102695	0.695412	0.929615	-0.007775	0.827134
209	(personal care)	(produce)	0.099480	0.748065	0.070550	0.709188	0.948030	-0.003868	0.866315
953	(household, beverages)	(produce)	0.085035	0.748065	0.061145	0.719057	0.961222	-0.002467	0.896747
65	(beverages)	(produce)	0.456190	0.748065	0.334630	0.733532	0.980573	-0.006630	0.945461
1014	(personal care, beverages)	(produce)	0.056865	0.748065	0.041760	0.734371	0.981694	-0.000779	0.948446



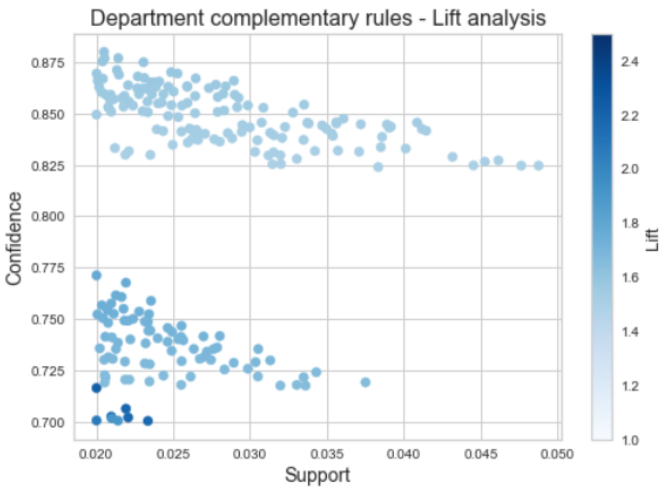


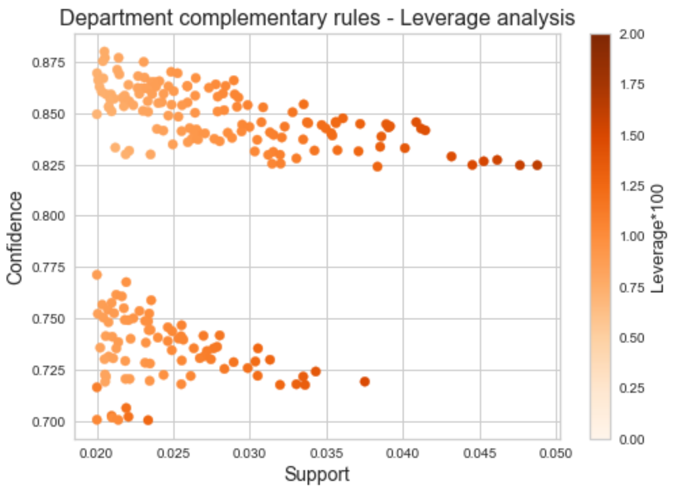
Data Mining

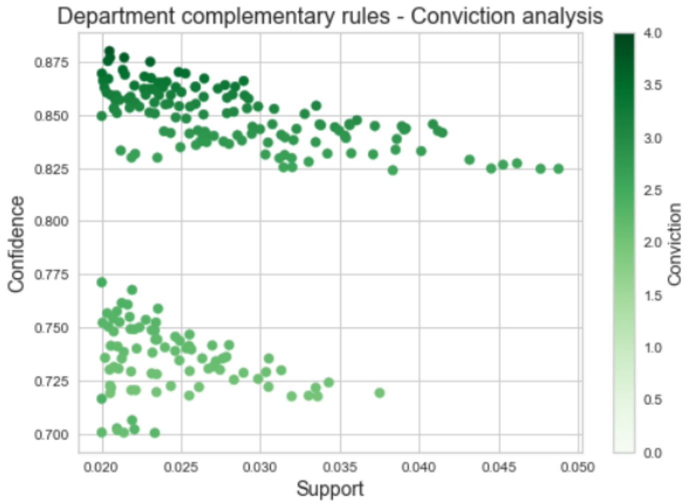
COMPLEMENTARY DEPARTMENTS – Rules for fine-tuning the algorithm

<u>Confidence > 0.5</u>
<u>Lift > 1.5</u>
<u>Conviction > 1.5</u>

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
4149	(deli, produce, breakfast, beverages)	(dairy eggs, snacks)	0.027930	0.322820	0.020010	0.716434	2.219299	0.010994	2.388086
2560	(deli, breakfast, beverages)	(dairy eggs, snacks)	0.031040	0.322820	0.021925	0.706347	2.188051	0.011905	2.306052
4694	(deli, beverages, produce, frozen, bakery)	(dairy eggs, snacks)	0.029850	0.322820	0.020970	0.702513	2.176174	0.011334	2.276331
3941	(produce, breakfast, bakery, pantry)	(dairy eggs, snacks)	0.031435	0.322820	0.022070	0.702084	2.174846	0.011922	2.273054
3928	(produce, breakfast, bakery, frozen)	(dairy eggs, snacks)	0.033345	0.322820	0.023355	0.700405	2.169645	0.012591	2.260317





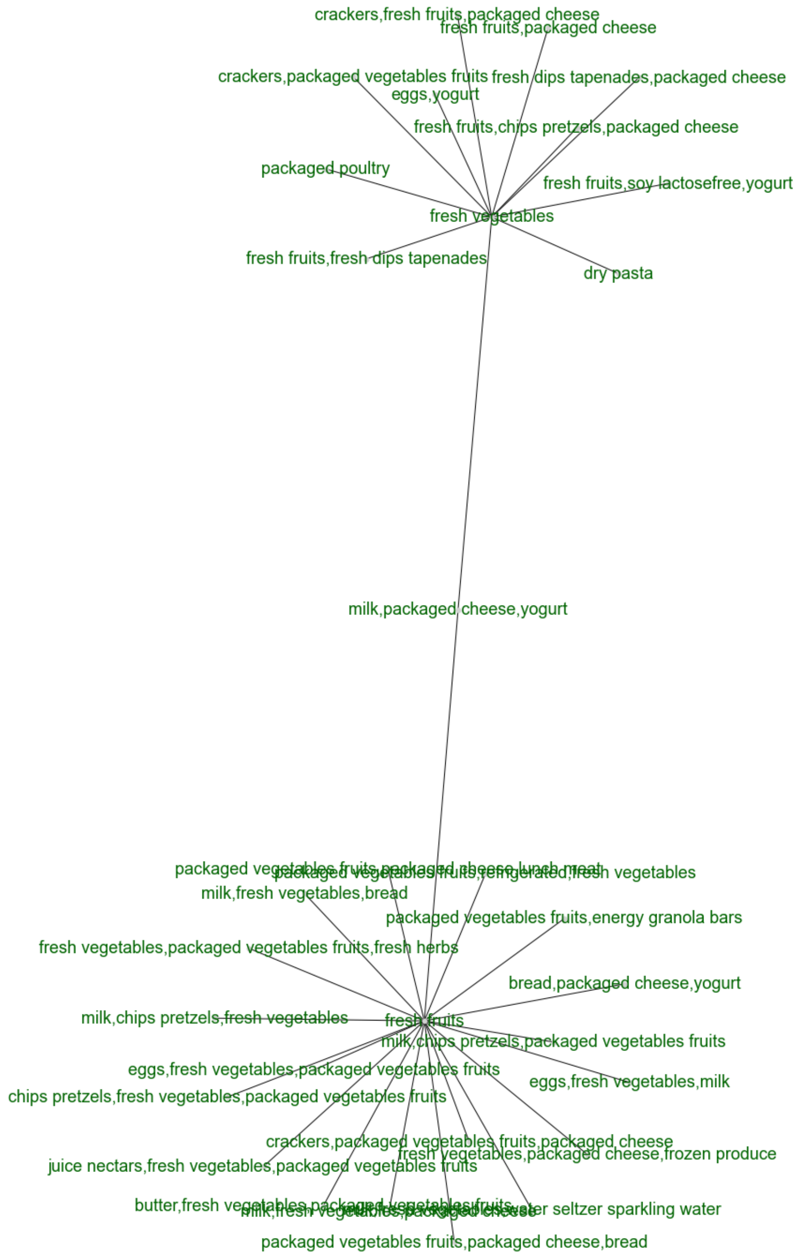


NOVA IMS
Information Management School

Results

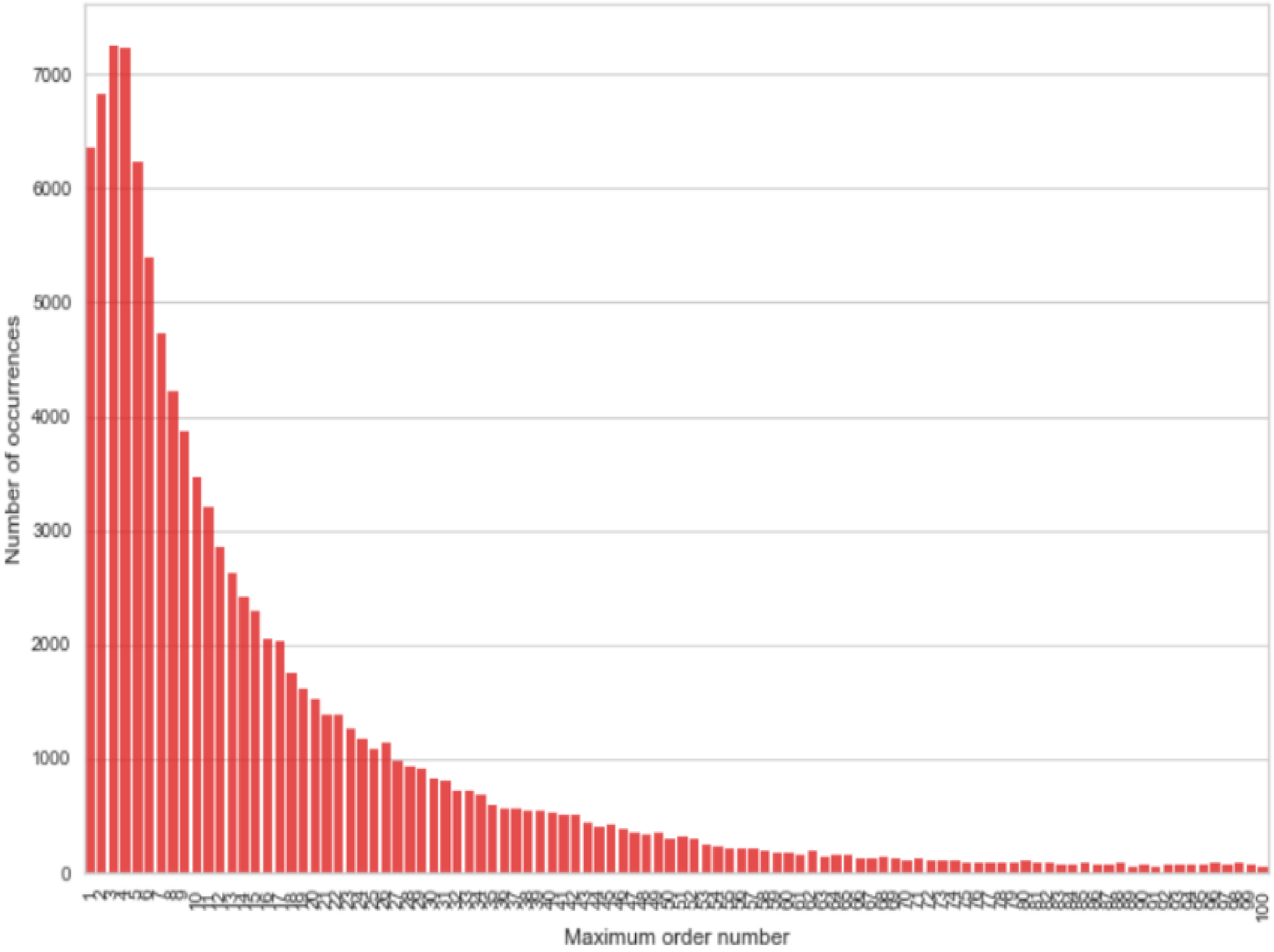
Which types of products should have an extended amount of product offerings?

- Fresh products, like:
  - Fresh fruits
  - Packaged vegetables & fruits
  - Bakery
  - Eggs and milk derivatives



NOVA IMS Information Management School

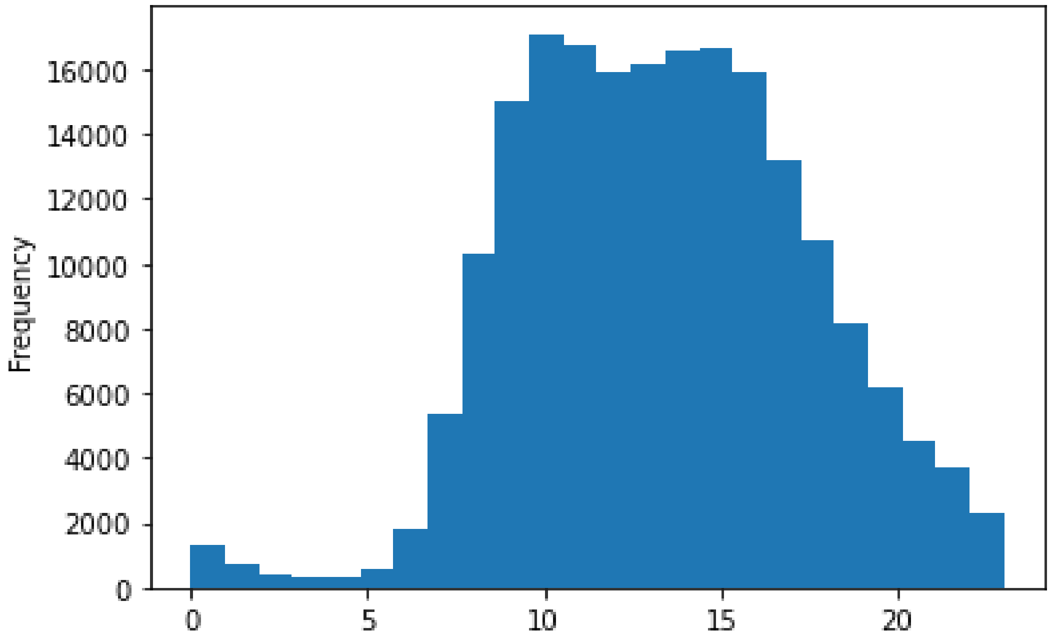
Our costumers



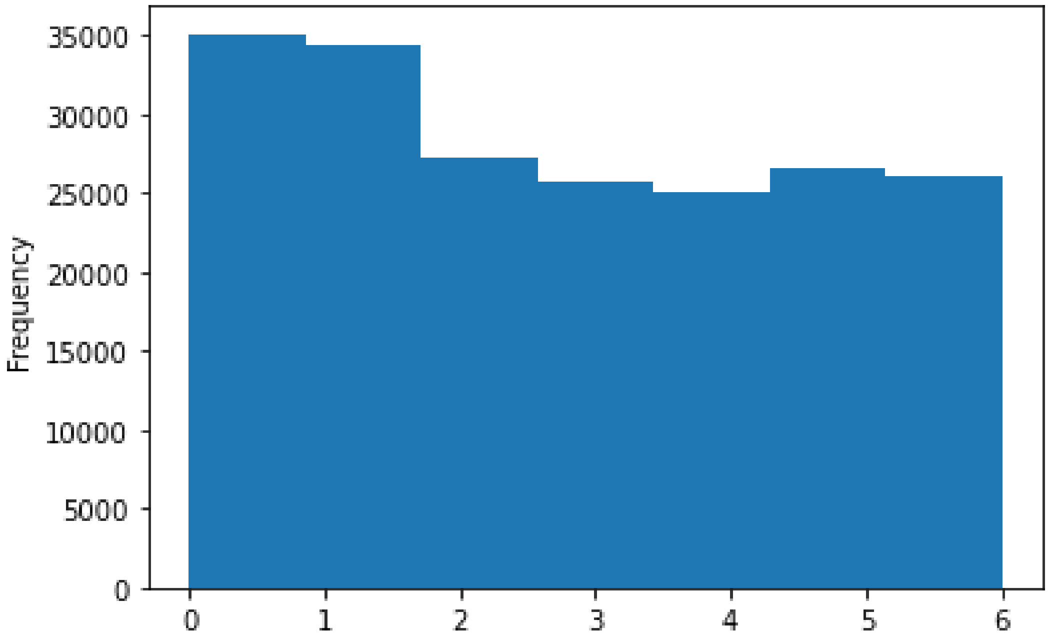
The majority of costumers are new, after the 4th order frequency seems to fall. This may be because some type of promotion in the app ends after the 4th order. Although it is also noticeable that we have loyal costumer

Our costumers

Hours of the day

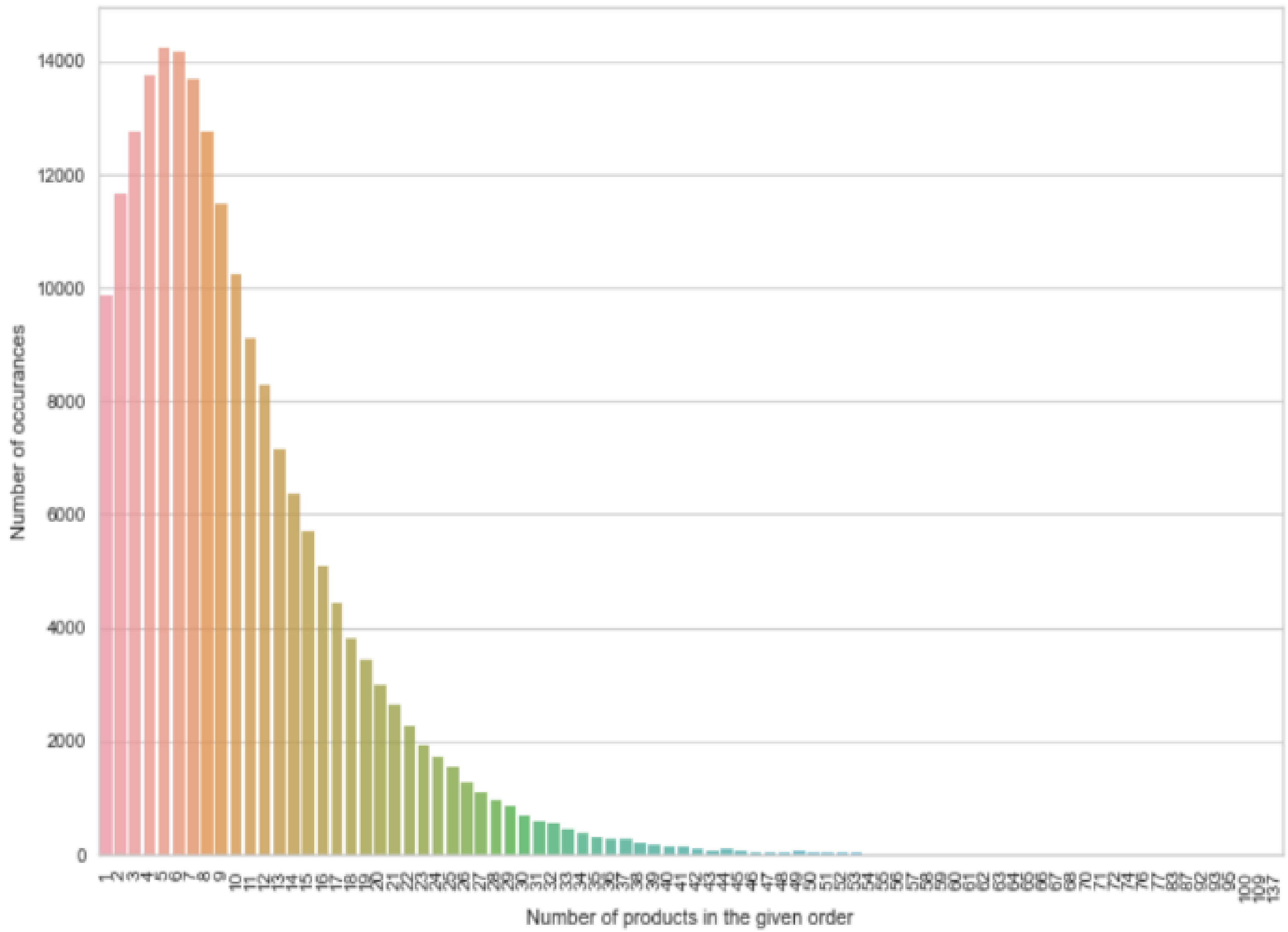


Days of the week



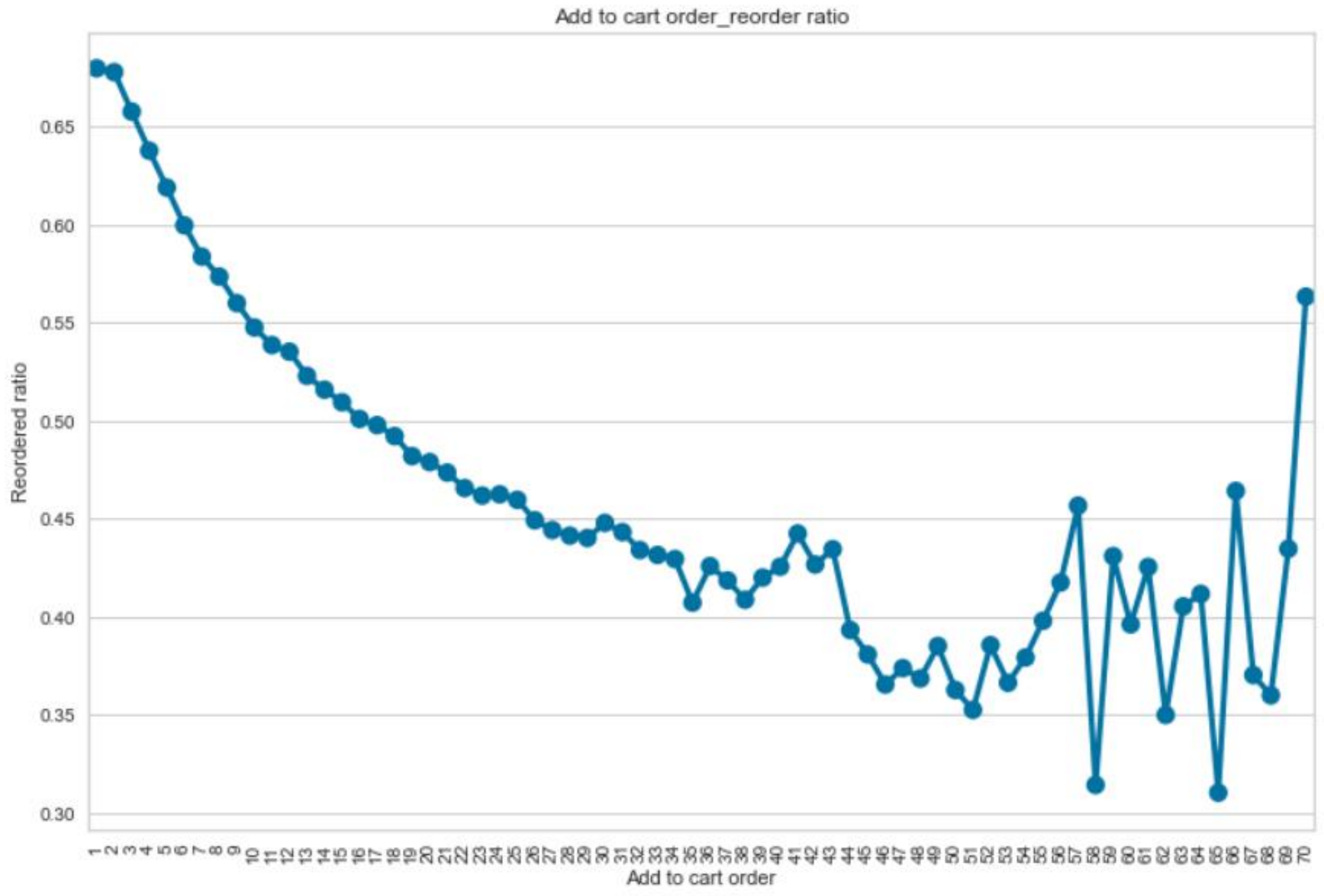
NOVA IMS Information Management School

Our costumers



Orders tend to not be that big, averaging around 8 products. Just the essencials that are missing around the house.

Our costumers



After trying the app for the 1st time, when making his next orders is likely to find clients re-ordering products.

How did we do it?

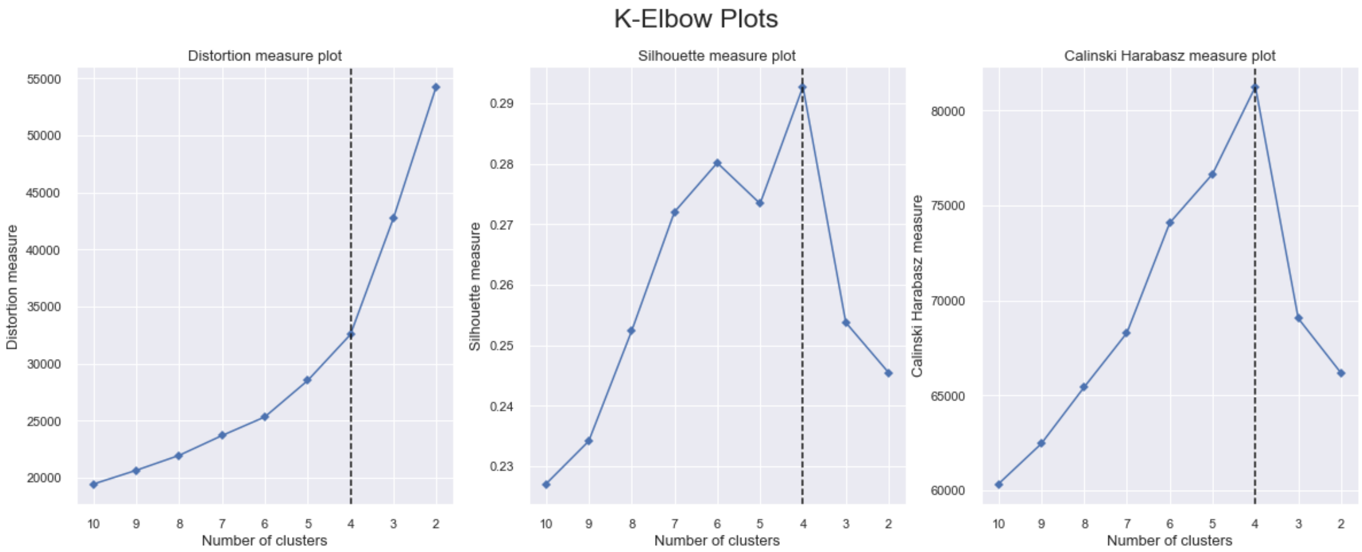
Preprocessing:

- Feature engineering to create better clusters.

Data Mining:

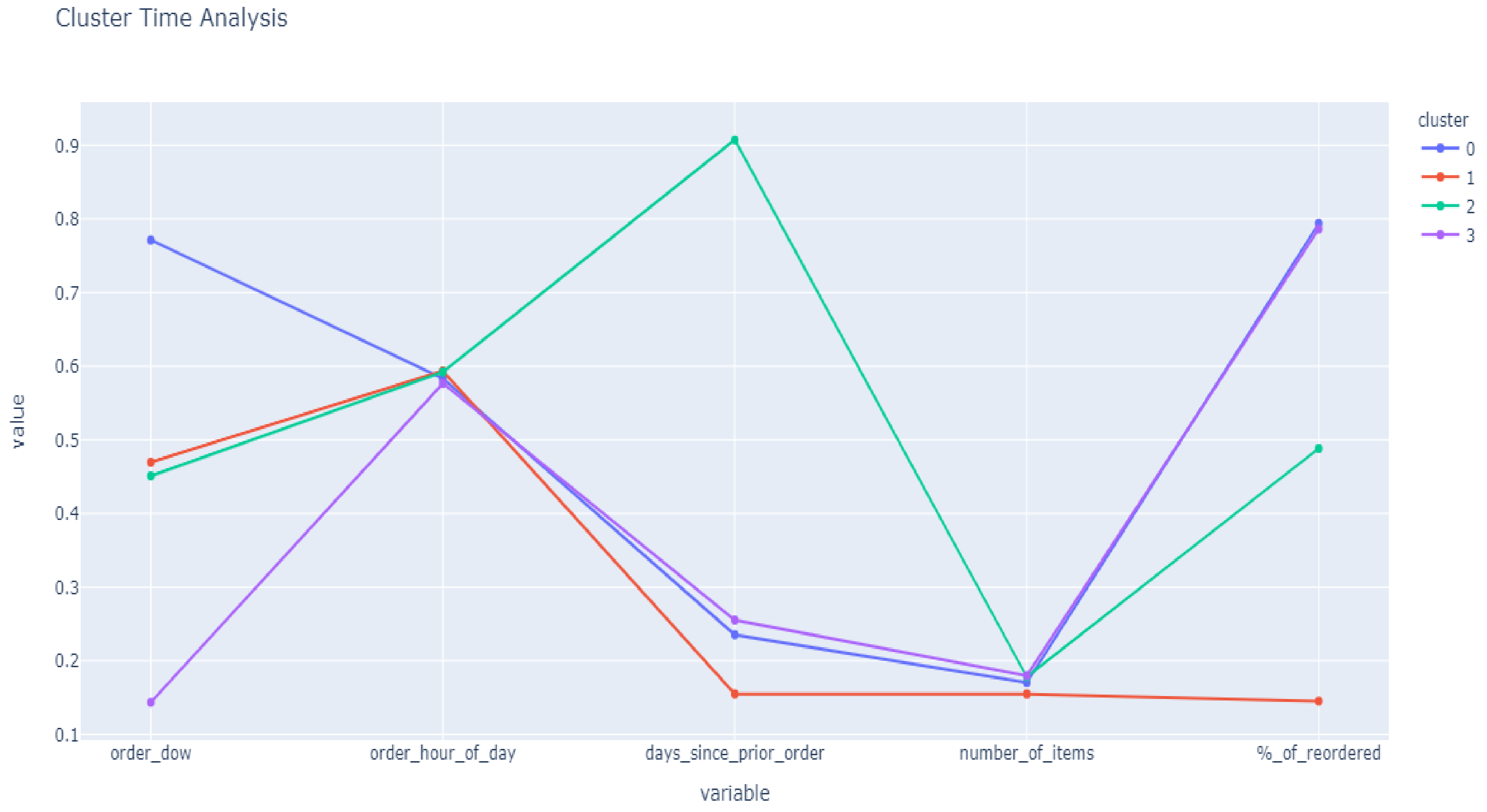
<u>K-means clustering</u>: partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean (cluster centroid).

Variables:
['order_dow',
'order_hour_of_day',
'days_since_prior_order',
'number_of_items',
'%_of_reordered']



'rsq is:'
0.53

Consumer behavior



Cluster 0 "*Weekend Usual Purchase*":
- Low to medium number of products
- End of the week purchases
- Buys mostly the same products

Cluster 1 "*First Try Purchase*":
- New costumers
- Middle of the week purchases
- Lowest number of items purchased.

Cluster 2 "*Last Purchase*":
- Churned costumers
- Middle of the week purchases
- low to medium number of products

Cluster 3 "*Monday Usual Purchase*":
- Low to medium number of products
- beginning of the week purchases
- buys mostly the same items


instacar
Avocados&
Quinoa&
Blueberries&
Chocolate.
instacart

Thank you!