

# Getting to know Instacart's users

Customer segmentation & Basket Analysis

A close-up photograph of a person's hands holding a white smartphone. The screen displays a simple graphic of a red heart with a small green sprout on top. The background is a blurred image of various grocery items, including a bag of broccoli, a bottle of olive oil, and several boxes of food, suggesting a grocery store or delivery scene.

**“Our mission is to create  
a world where everyone  
has access to the food  
they love and more time  
to enjoy it together”**

# The questions

**Who are  
our clients?**

**Which are the  
buying patterns?**

# The project

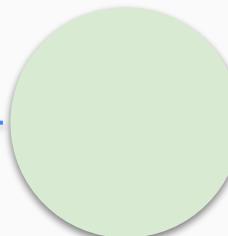
## EDA

Sample of 50k users  
and 750k transactions



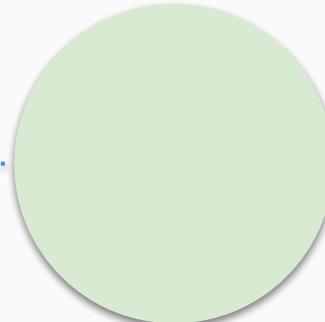
## K-Means

Dimensionality  
reduction with PCA

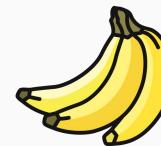
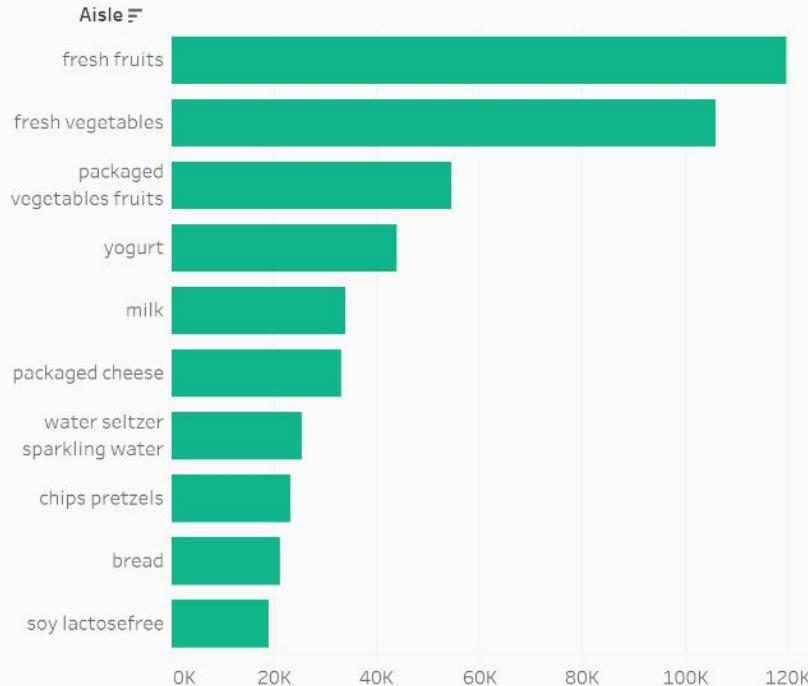


## FP Growth

Detailed and generic  
product version  
implemented.



# Top 10 aisles



**Bananas**  
4% of purchases



**Strawberries**  
1,8% of purchases



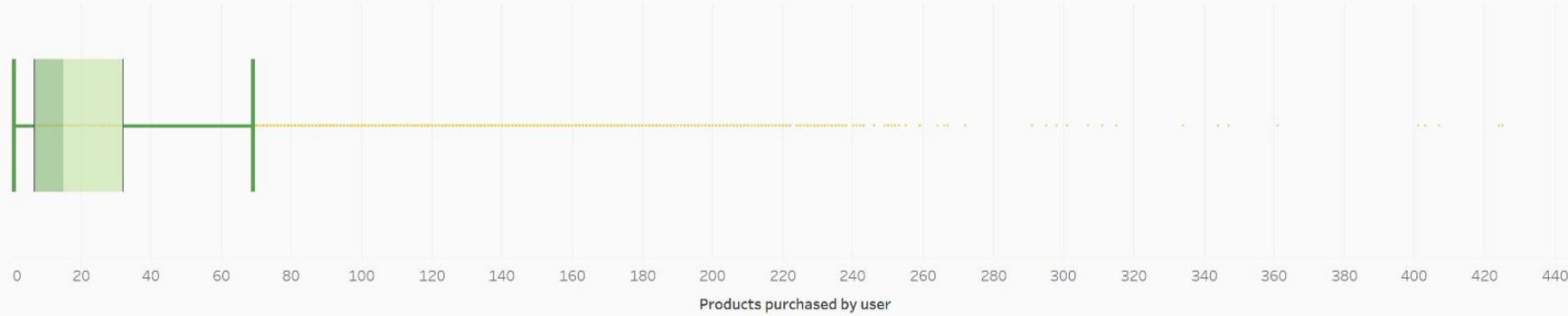
**Baby Spinach**  
1,1% of purchases

# Busiest days and hours

- Most busy hours from 9am to 17pm
- Activity decreases on Wednesday, Thursday and Friday

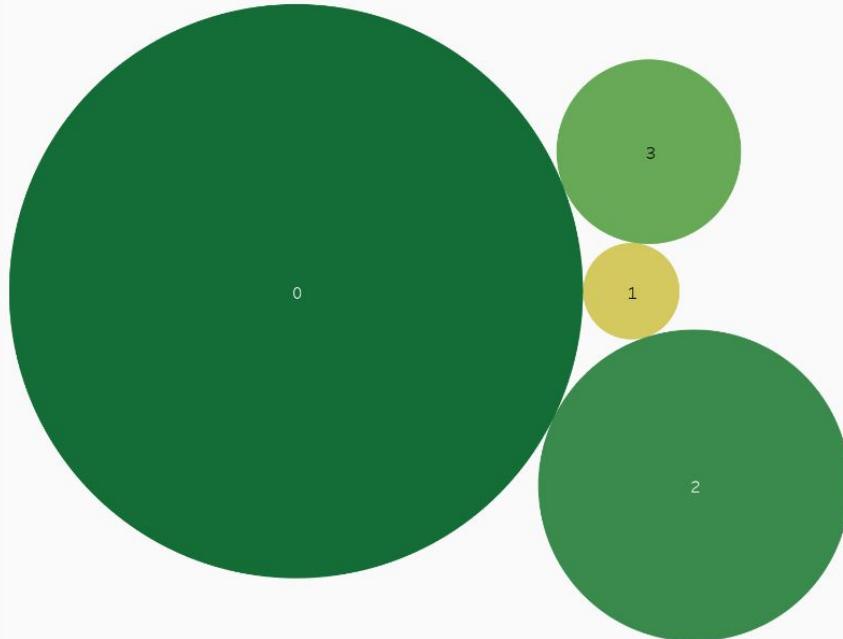
Order Hour Of Day	Order Dow						
	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday	Sunday
0	1.113	1.147	1.046	874	736	891	855
1	718	498	321	335	432	337	681
2	442	217	224	301	268	269	357
3	289	269	188	159	204	286	262
4	337	259	236	178	235	214	321
5	363	460	406	362	328	486	415
6	1.142	1.447	1.573	1.388	1.308	1.695	1.010
7	4.311	5.743	4.006	3.329	3.798	4.349	4.234
8	9.993	11.266	7.610	7.133	6.474	6.975	7.945
9	13.697	15.940	10.626	9.209	8.755	9.631	10.541
10	16.742	17.428	11.806	10.602	9.923	11.772	11.675
11	16.265	16.408	10.775	10.104	9.570	11.196	11.911
12	17.249	15.214	11.038	9.895	9.022	11.056	12.104
13	18.050	15.283	10.989	10.120	9.895	10.732	11.759
14	18.183	14.255	10.711	9.806	9.740	11.475	12.468
15	18.238	14.038	10.606	10.081	9.897	11.295	11.511
16	16.229	13.269	10.852	9.674	9.361	10.787	11.536
17	12.176	10.807	8.776	8.395	8.806	9.177	9.621
18	9.092	8.321	7.737	7.109	6.720	7.071	6.625
19	6.798	6.224	5.662	5.414	5.284	5.454	5.652
20	5.802	4.939	4.505	3.948	3.921	4.299	4.568
21	5.239	3.834	3.381	3.150	3.400	3.284	3.554
22	3.822	2.756	2.682	2.839	2.882	2.471	3.206
23	2.277	1.817	1.618	1.743	1.755	1.740	2.180

## Basket size and average orders



# Customer segmentation

# Clusters



Main challenge:

- Lack of data
- Silhouette Score high but no homogenous distribution

# Everything looks the same... Does it?

Aisle	Cluster 4 0	Aisle	Cluster 4 1	Aisle	Cluster 4 2	Aisle	Cluster 4 3
	Count		Count		Count		Count
fresh fruits	83.816	fresh fruits	2.490	fresh fruits	24.794	fresh fruits	8.695
fresh vegetables	73.696	fresh vegetables	2.193	fresh vegetables	21.932	fresh vegetables	8.032
packaged vegetables ..	38.372	packaged vegetable...	986	packaged vegetables ..	11.267	packaged vegetables ..	4.005
yogurt	30.694	yogurt	867	yogurt	9.216	yogurt	3.272
milk	23.727	milk	746	milk	6.889	packaged cheese	2.491
packaged cheese	23.224	packaged cheese	636	packaged cheese	6.867	milk	2.439
water seltzer sparklin..	17.916	water seltzer sparkl..	468	water seltzer sparklin..	5.279	water seltzer sparklin..	1.822
chips pretzels	16.341	chips pretzels	451	chips pretzels	4.873	chips pretzels	1.669
bread	14.773	bread	402	bread	4.388	bread	1.557
soy lactosefree	13.470	frozen produce	399	frozen produce	3.888	soy lactosefree	1.455
frozen produce	13.309	refrigerated	383	soy lactosefree	3.867	refrigerated	1.373
refrigerated	12.909	ice cream ice	366	refrigerated	3.775	frozen produce	1.342
eggs	11.372	soy lactosefree	329	eggs	3.337	energy granola bars	1.245
ice cream ice	10.790	crackers	301	ice cream ice	3.227	eggs	1.245
crackers	10.133	lunch meat	298	crackers	3.202	crackers	1.114
energy granola bars	9.911	eggs	293	energy granola bars	3.125	ice cream ice	1.068
lunch meat	9.514	energy granola bars	274	lunch meat	2.834	lunch meat	1.015
fresh dips tapenades	9.132	soft drinks	251	cereal	2.767	cereal	877
cereal	9.005	soup broth bouillon	242	fresh dips tapenades	2.707	fresh dips tapenades	854
soft drinks	8.614	fresh herbs	238	soup broth bouillon	2.481	canned jarred vegeta..	845
soup broth bouillon	8.607	baby food formula	234	soft drinks	2.479	soup broth bouillon	841
baby food formula	7.860	cereal	232	canned jarred vegeta..	2.413	fresh herbs	801
baking ingredients	7.838	fresh dips tapenades	224	fresh herbs	2.368	soft drinks	782
fresh herbs	7.831	packaged produce	222	nuts seeds dried fruit	2.360	baby food formula	782
canned jarred vegeta..	7.825	frozen meals	211	baby food formula	2.352	nuts seeds dried fruit	771
nuts seeds dried fruit	7.759	baking ingredients	208	frozen meals	2.269	baking ingredients	769
frozen meals	7.671	oils vinegars	207	baking ingredients	2.218	frozen meals	729
juice nectars	7.005	canned jarred veget..	207	packaged produce	2.154	spreads	689
hot dogs bacon sausa..	6.709	nuts seeds dried fruit	198	juice nectars	1.974	dry pasta	668
dry pasta	6.695	spreads	193	hot dogs bacon sausa..	1.958	packaged produce	666
packaged produce	6.616	cream	193	spreads	1.893	hot dogs bacon sausa..	664
oils vinegars	6.492	canned meals beans	193	dry pasta	1.877	paper goods	632
spreads	6.423	hot dogs bacon saus..	183	oils vinegars	1.870	coffee	625
canned meals beans	6.225	juice nectars	176	canned meals beans	1.782	other creams cheeses	621
cream	6.126	dry pasta	176	frozen breakfast	1.763	canned meals beans	619
	5.051		176				

Well, it does not.

Cluster 4		Cluster 4		Cluster 4		Cluster 4	
Aisle	F	Aisle	F	Aisle	F	Aisle	F
energy granola bars	9.911	eggs	293	energy granola bars	3.125	ice cream ice	1.068
lunch meat	9.514	energy granola bars	274	lunch meat	2.834	lunch meat	1.015
fresh dips tapenades	9.132	soft drinks	251	cereal	2.767	cereal	877
cereal	9.005	soup broth bouillon	242	fresh dips tapenades	2.707	fresh dips tapenades	854
soft drinks	8.614	fresh herbs	238	soup broth bouillon	2.481	canned jarred vegeta..	845
soup broth bouillon	8.607	baby food formula	234	soft drinks	2.479	soup broth bouillon	841
baby food formula	7.860	cereal	232	canned jarred vegeta..	2.413	fresh herbs	801
baking ingredients	7.838	fresh dips tapenades	224	fresh herbs	2.368	soft drinks	782
fresh herbs	7.831	packaged produce	222	nuts seeds dried fruit	2.360	baby food formula	782
canned jarred vegeta..	7.825	frozen meals	211	baby food formula	2.352	nuts seeds dried fruit	771
nuts seeds dried fruit	7.759	baking ingredients	208	frozen meals	2.269	baking ingredients	769
frozen meals	7.671	oils vinegars	207	baking ingredients	2.218	frozen meals	729
juice nectars	7.005	canned jarred veget..	207	packaged produce	2.154	spreads	689
hot dogs bacon sausa..	6.709	nuts seeds dried fruit	198	juice nectars	1.974	dry pasta	668
dry pasta	6.695	spreads	193	hot dogs bacon sausa..	1.958	packaged produce	666
packaged produce	6.616	cream	193	spreads	1.893	hot dogs bacon sausa..	664
oils vinegars	6.492	canned meals beans	193	dry pasta	1.877	paper goods	632
spreads	6.423	hot dogs bacon saus..	183	oils vinegars	1.870	coffee	625
canned meals beans	6.225	juice nectars	176	canned meals beans	1.782	other creams cheeses	621
cream	6.126	dry pasta	176	frozen breakfast	1.763	canned meals beans	619
frozen breakfast	5.954	breakfast bakery	171	cream	1.757	juice nectars	596
candy chocolate	5.927	candy chocolate	167	other creams cheeses	1.710	oils vinegars	585
paper goods	5.864	tea	163	paper goods	1.704	frozen breakfast	583
other creams cheeses	5.863	frozen breakfast	160	breakfast bakery	1.689	cream	583
breakfast bakery	5.813	other creams cheeses	159	pasta sauce	1.660	breakfast bakery	577
tea	5.769	coffee	152	candy chocolate	1.628	candy chocolate	566
coffee	5.526	pasta sauce	151	tea	1.491	pasta sauce	562
pasta sauce	5.350	condiments	141	coffee	1.485	tea	510
condiments	4.822	paper goods	137	condiments	1.405	tortillas flat bread	490
frozen appetizers sides	4.760	asian foods	136	tortillas flat bread	1.386	butter	484
tortillas flat bread	4.620	popcorn jerky	127	popcorn jerky	1.350	popcorn jerky	481
butter	4.502	spices seasonings	126	cookies cakes	1.316	frozen appetizers sides	481
popcorn jerky	4.481	cookies cakes	121	frozen appetizers sides	1.308	condiments	480
frozen pizza	4.318	frozen appetizers si..	119	butter	1.270	cookies cakes	449
cookies cakes	4.281	tortillas flat bread	110	frozen pizza	1.261	spices seasonings	441

# Who is who in Instacart dataset

## 0: The families

---

Biggest cluster.

Most active days:

Monday  
Tuesday  
*Sunday*

Preferred aisles:

- Energy granola bars,
- Fresh dips tapenades
- Cereal
- Soft Drinks
- Soup Broth
- Baby Food Formula
- Nuts, Seeds, Dried fruit
- Juices, pasta, candy, chocolate
- Canned meals beans
- Paper goods

## 1: The exotics

---

No big fans of eggs.

Most active days:

Monday  
Tuesday  
*Saturday*

Preferred aisles:

- Soft Drinks
- Energy Granola Bars
- Canned Meals Beans
- Candy, Chocolate
- Package produced
- Asian foods
- Tea

# Who is who in Instacart dataset

## 2: The healthiest

---

Almost a mediterranean diet.

Most active days:

Monday  
Tuesday  
*Sunday*

Preferred aisles:

- Energy granola bars
- Lunch meat
- Canned jarred vegetables
- Packaged produce (meat, fish, etc)
- Frozen meals

## 4: The 'no time to cook'

---

Frozen food is their friend.

Most active days:

Monday

Preferred aisles:

- Ice cream
- Instant foods
- Lunch meat
- Frozen meals

# Basket Analysis

# FP Growth by specific product

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
10	(Strawberries)	(Banana)	0.045778	0.151427	0.013132	0.286857	1.894362	0.006200	1.189906
11	(Bag of Organic Bananas)	(Organic Raspberries)	0.121460	0.044625	0.012948	0.106603	2.388838	0.007528	1.069373
12	(Organic Raspberries)	(Bag of Organic Bananas)	0.044625	0.121460	0.012948	0.290147	2.388838	0.007528	1.237637
13	(Organic Raspberries)	(Organic Strawberries)	0.044625	0.083993	0.011062	0.247882	2.951239	0.007314	1.217904
14	(Organic Strawberries)	(Organic Raspberries)	0.083993	0.044625	0.011062	0.131700	2.951239	0.007314	1.100282
15	(Organic Fuji Apple)	(Banana)	0.027756	0.151427	0.010427	0.375656	2.480779	0.006224	1.359144
16	(Limes)	(Banana)	0.044675	0.151427	0.010147	0.227129	1.499924	0.003382	1.097949

# FP Growth by aisle

antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
(yogurt)	(fresh fruits)	0.197828	0.482386	0.125096	0.632350	1.310880	0.029667	1.407900
(fresh vegetables)	(packaged vegetables fruits)	0.363032	0.277644	0.149676	0.412294	1.484973	0.048882	1.229111
(packaged vegetables fruits)	(fresh vegetables)	0.277644	0.363032	0.149676	0.539092	1.484973	0.048882	1.381987
(fresh fruits)	(packaged vegetables fruits)	0.482386	0.277644	0.179835	0.372802	1.342734	0.045903	1.151720
(packaged vegetables fruits)	(fresh fruits)	0.277644	0.482386	0.179835	0.647716	1.342734	0.045903	1.469310
(fresh fruits, fresh vegetables)	(packaged vegetables fruits)	0.223479	0.277644	0.105623	0.472629	1.702283	0.043575	1.369729
(fresh fruits, packaged vegetables fruits)	(fresh vegetables)	0.179835	0.363032	0.105623	0.587333	1.617855	0.040337	1.543540
(fresh vegetables, packaged vegetables fruits)	(fresh fruits)	0.149676	0.482386	0.105623	0.705677	1.462888	0.033421	1.758660
(packaged vegetables fruits)	(fresh fruits, fresh vegetables)	0.277644	0.223479	0.105623	0.380425	1.702283	0.043575	1.253312

# FP Growth by generic products

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
12	(cheese, milk)	(yogurt)	0.073970	0.176336	0.023278	0.314703	1.784682	0.010235	1.201909
16	(crackers)	(cheese)	0.062031	0.199964	0.020765	0.334755	1.674077	0.008361	1.202619
14	(milk, yogurt)	(cheese)	0.070429	0.199964	0.023278	0.330523	1.652912	0.009195	1.195016
5	(sauce)	(cheese)	0.076894	0.199964	0.025168	0.327307	1.636830	0.009792	1.189303
9	(bread)	(cheese)	0.131231	0.199964	0.041045	0.312767	1.564117	0.014803	1.164141
13	(cheese, yogurt)	(milk)	0.050625	0.294560	0.023278	0.459822	1.561048	0.008366	1.305940
24	(cereal)	(milk)	0.062585	0.294560	0.027695	0.442515	1.502295	0.009260	1.265399
28	(raspberries)	(milk)	0.056510	0.294560	0.023324	0.412750	1.401242	0.006679	1.201260
11	(yogurt)	(milk)	0.176336	0.294560	0.070429	0.399405	1.355937	0.018488	1.174568