

EDA on Instacart Dataset

Gaurang Swarge

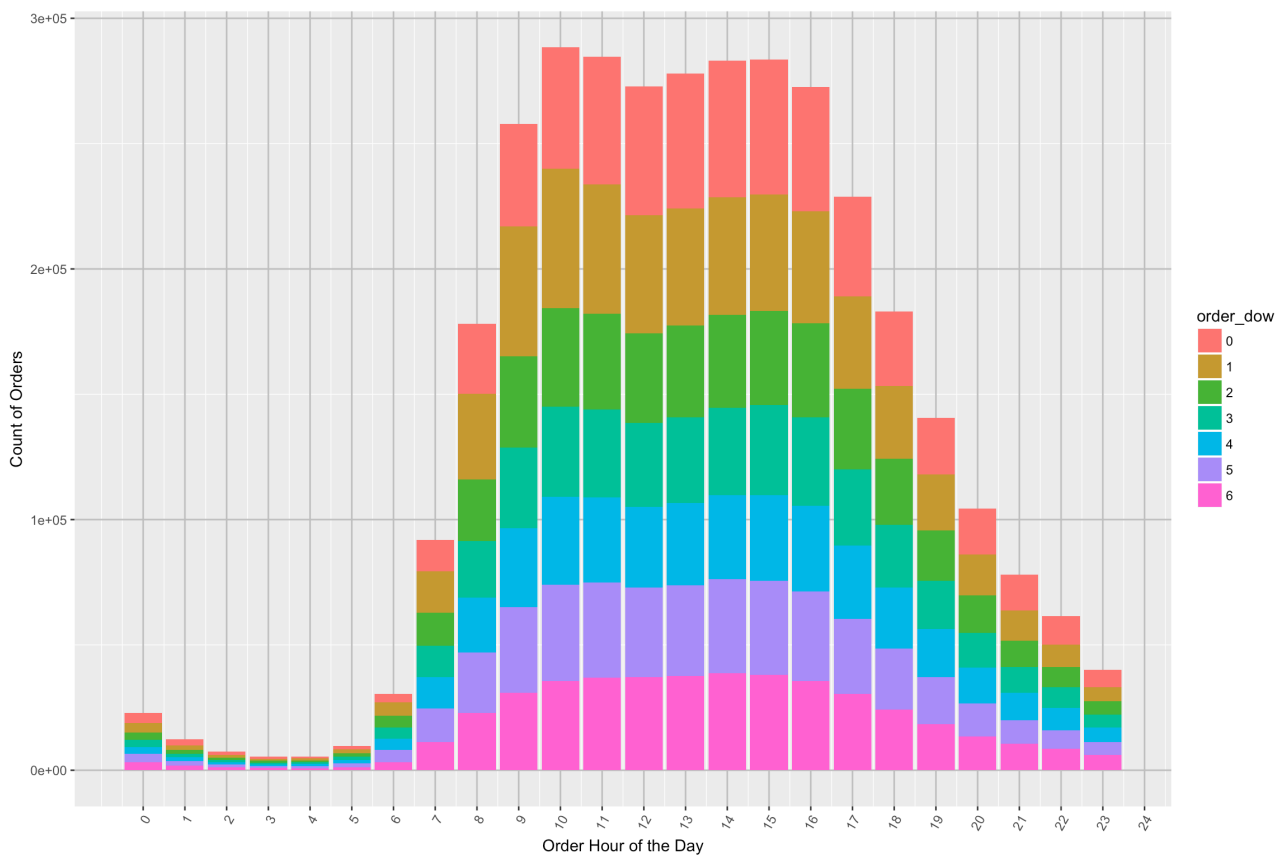
May 6th, 2018

1. Starting with Orders Table

Order Hour of the day vs no. of Orders: Checking What is the time when maximum people order:

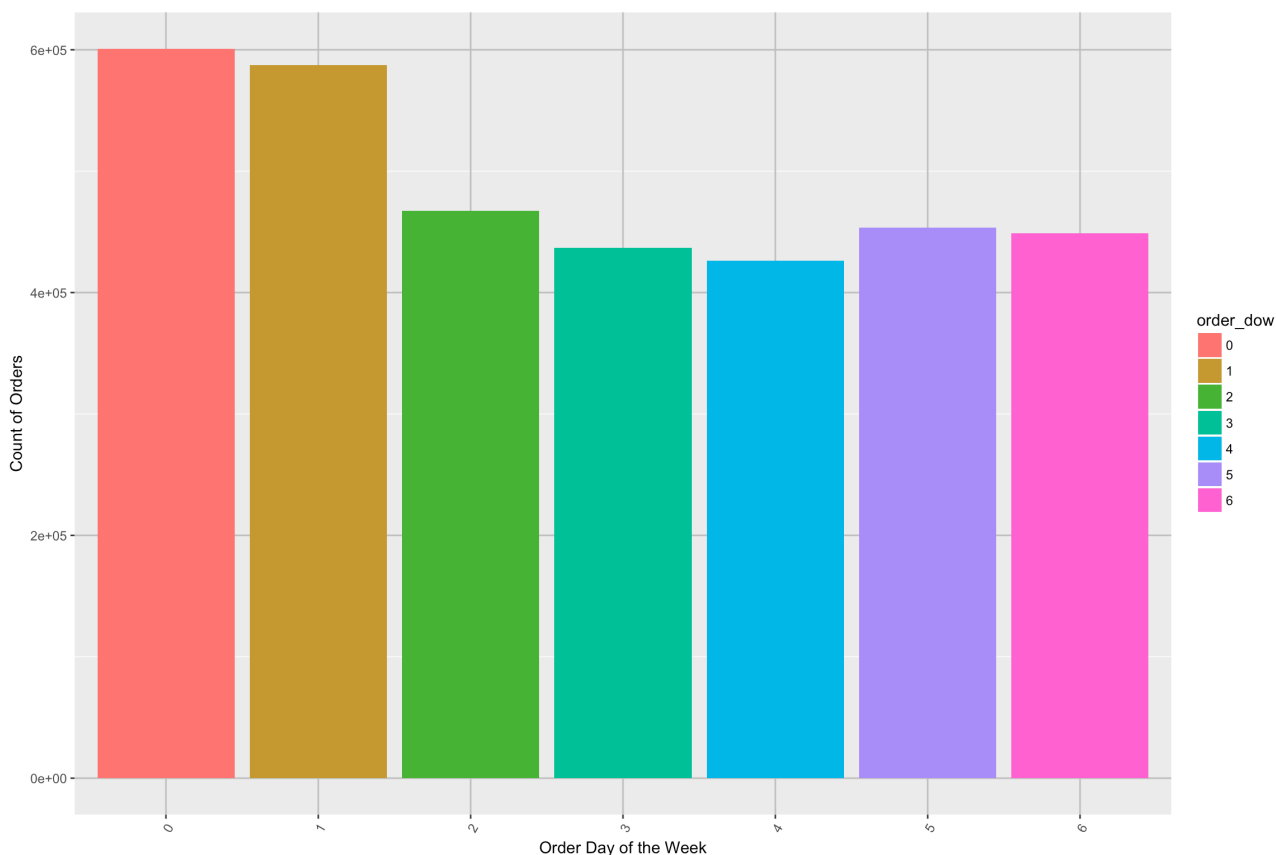
It seems like people order the most between 9:00 AM to 4:PM on Day 0 of the week presumably Saturday

order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
2539329	1	prior	1	2	8	NA
2398795	1	prior	2	3	7	15
473747	1	prior	3	3	12	21
2254736	1	prior	4	4	7	29
431534	1	prior	5	4	15	28
3367565	1	prior	6	2	7	19



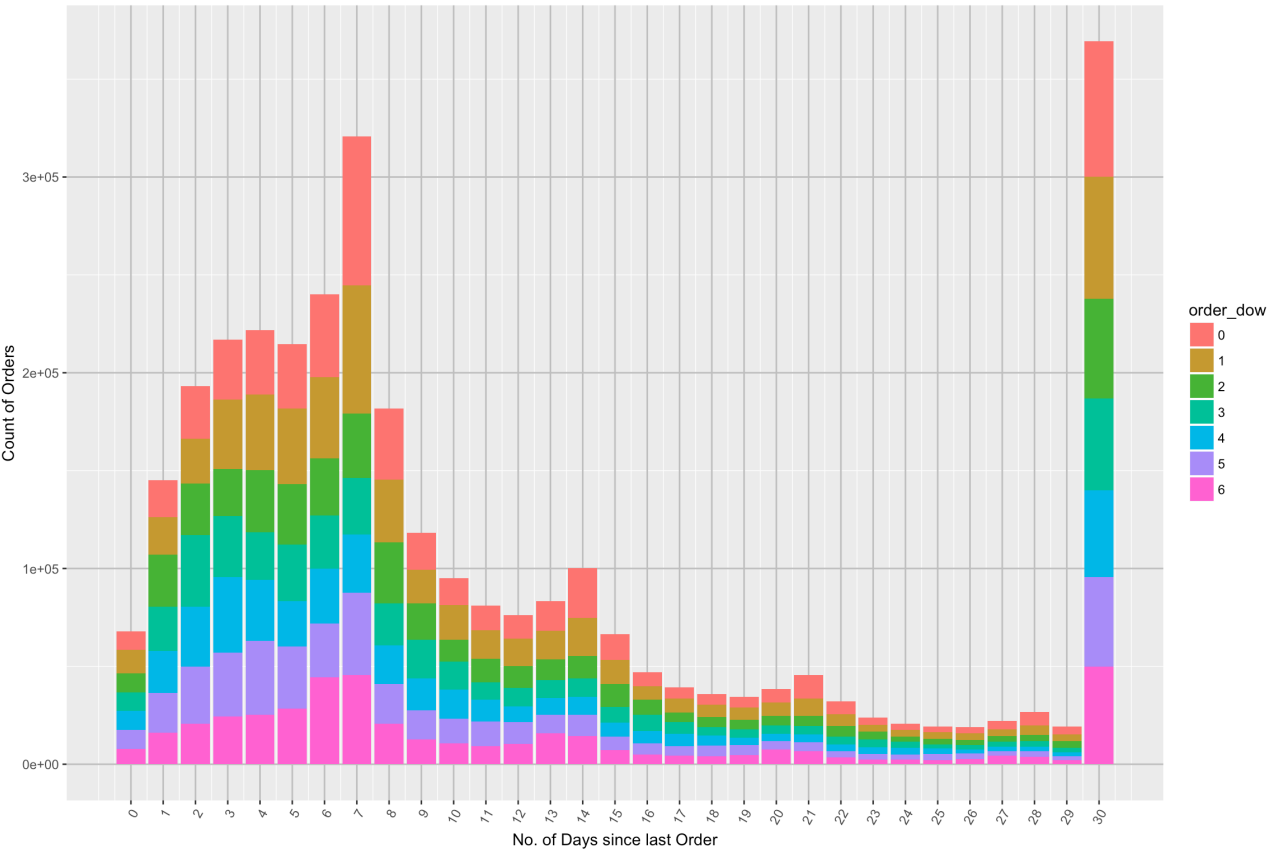
2. Checking for Order Day of Week: When do people order the most:

It seems like Day 0 & Day 1 presumable Saturday and Sunday is when people order the most



3. Checking how many days it takes to people for ordering again.

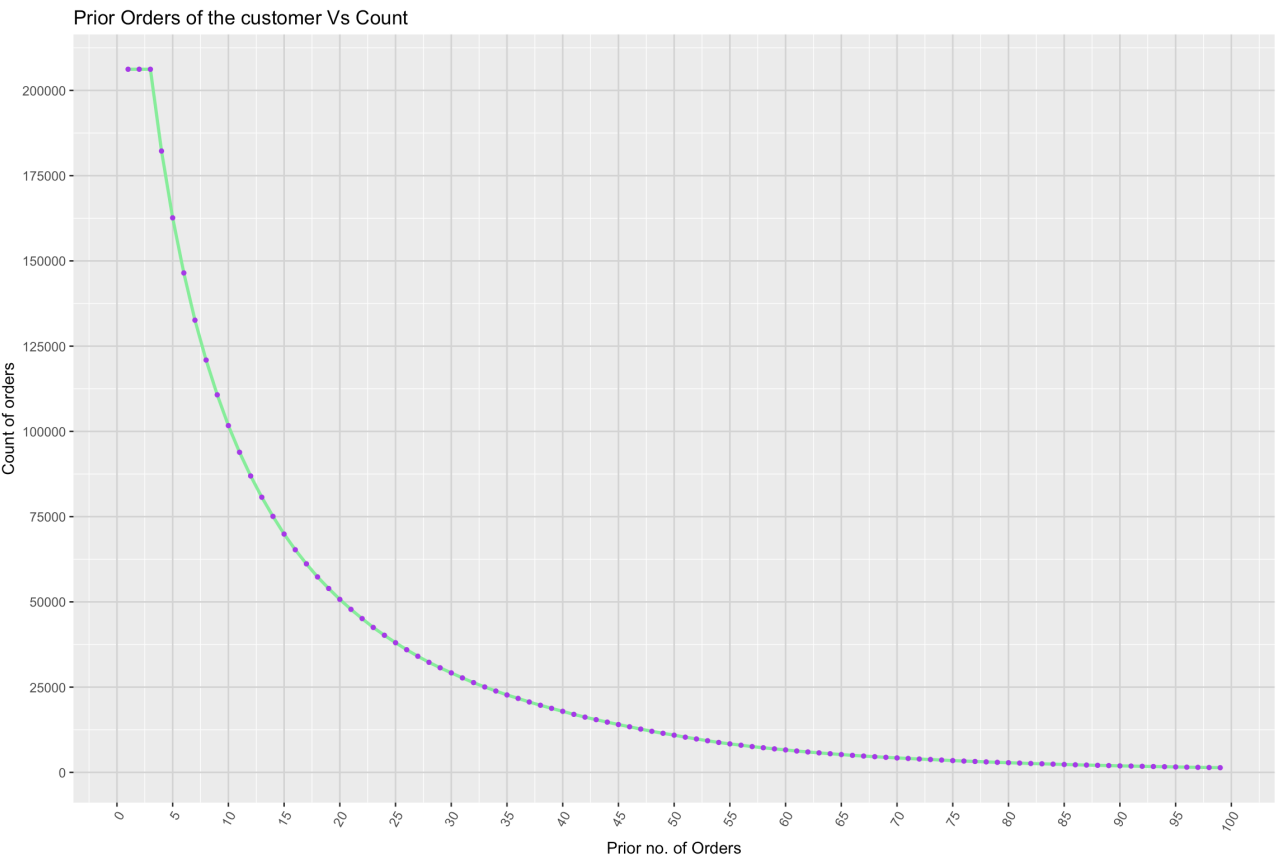
Days since prior Order vs Count of Order numbers. So seems like people order again maximum on 7th Day and 30th day



4. prior orders of the customer vs Order Number

We can see that more than 50% of the orders tend to have ordered previously upto 10 times

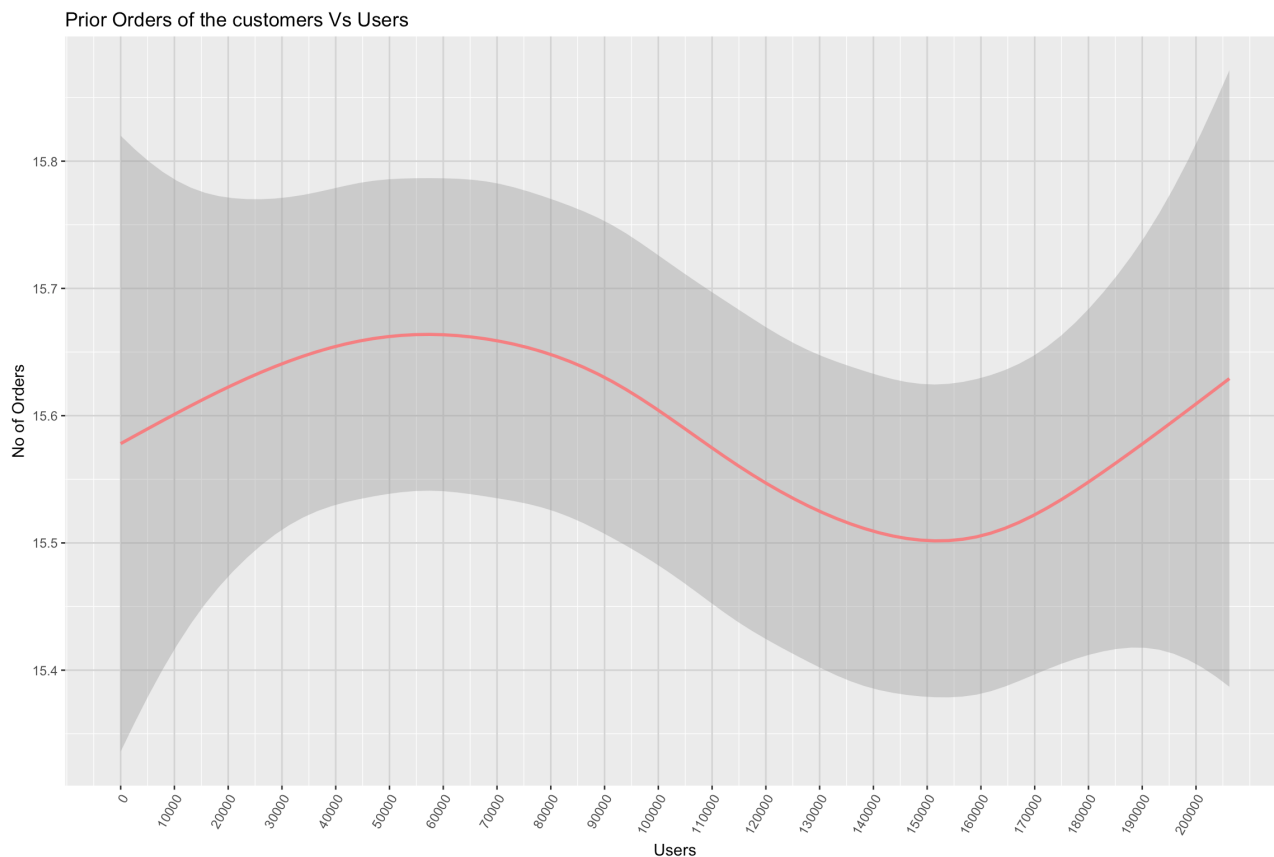
order_number	n
1	206209
2	206209
3	206209
4	182223
5	162633
6	146468
7	132618
8	120918
9	110728
10	101696



5. For all the customers who has prior order history

This graph shows customer vs No of orders
We can see the average no of orders changing as the no of users increase

user_id	n
1	10
2	14
3	12
4	5
5	4
6	3
7	20
8	3
9	3
10	5



6.Order_hour_of the day vs mean(number of orders of customers)

Trying to check how many orders are placed at what time of the day

order_hour_of_day	order_number	count
15	3	17657
15	4	17528
14	2	17318
15	2	17312
14	4	17308
15	1	17264
14	3	17223
16	4	17157
16	1	17094
16	2	17082

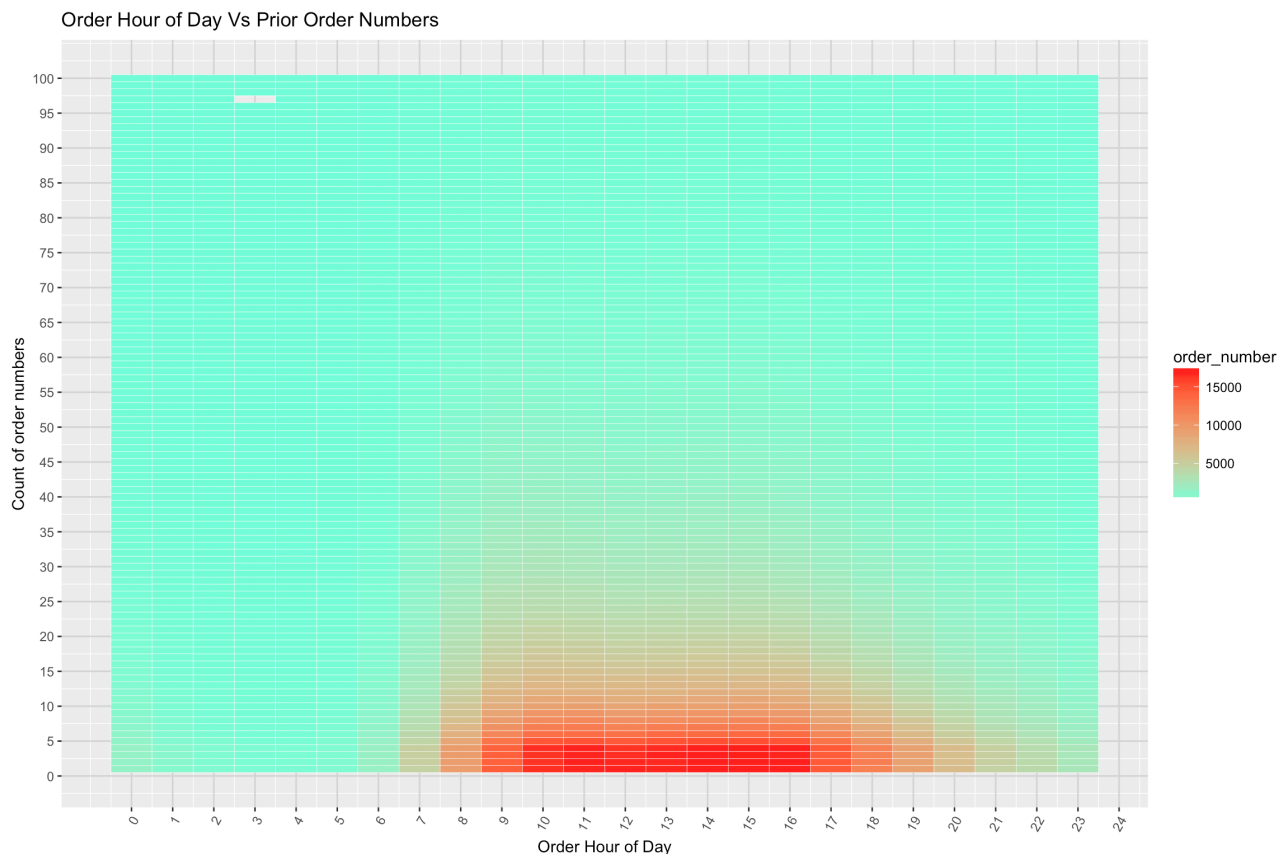


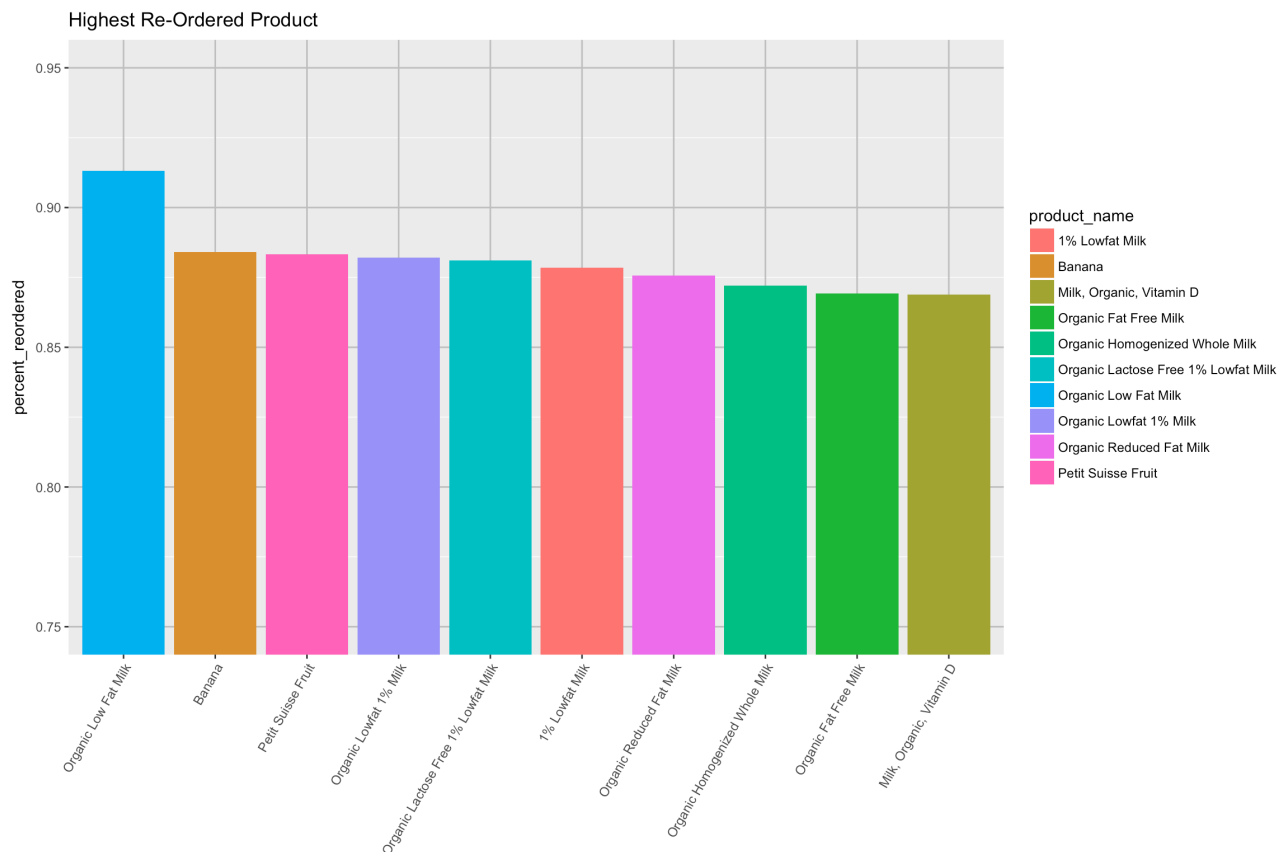
Figure 6

7. Trying to find out Highest reordered Product

Organic LowFat Milk as expected is the highest reordered product in the dataset and the second highest is the Banana

product_id	percent_reordered	n	product_name	aisle_id	department_id
20940	0.9130435	368	Organic Low Fat Milk	84	16
24852	0.8841717	18726	Banana	24	4
117	0.8833333	120	Petit Suisse Fruit	2	16
39180	0.8819876	483	Organic Lowfat 1% Milk	84	16
12384	0.8810409	269	Organic Lactose Free 1% Lowfat Milk	91	16
24024	0.8785249	461	1% Lowfat Milk	84	16
38689	0.8756219	1206	Organic Reduced Fat Milk	84	16
5514	0.8720000	125	Organic Homogenized Whole Milk	84	16
36011	0.8692972	811	Organic Fat Free Milk	84	16
29447	0.8688811	572	Milk, Organic, Vitamin D	84	16

8. Trying to check, which is the Highest Re-Ordered Department



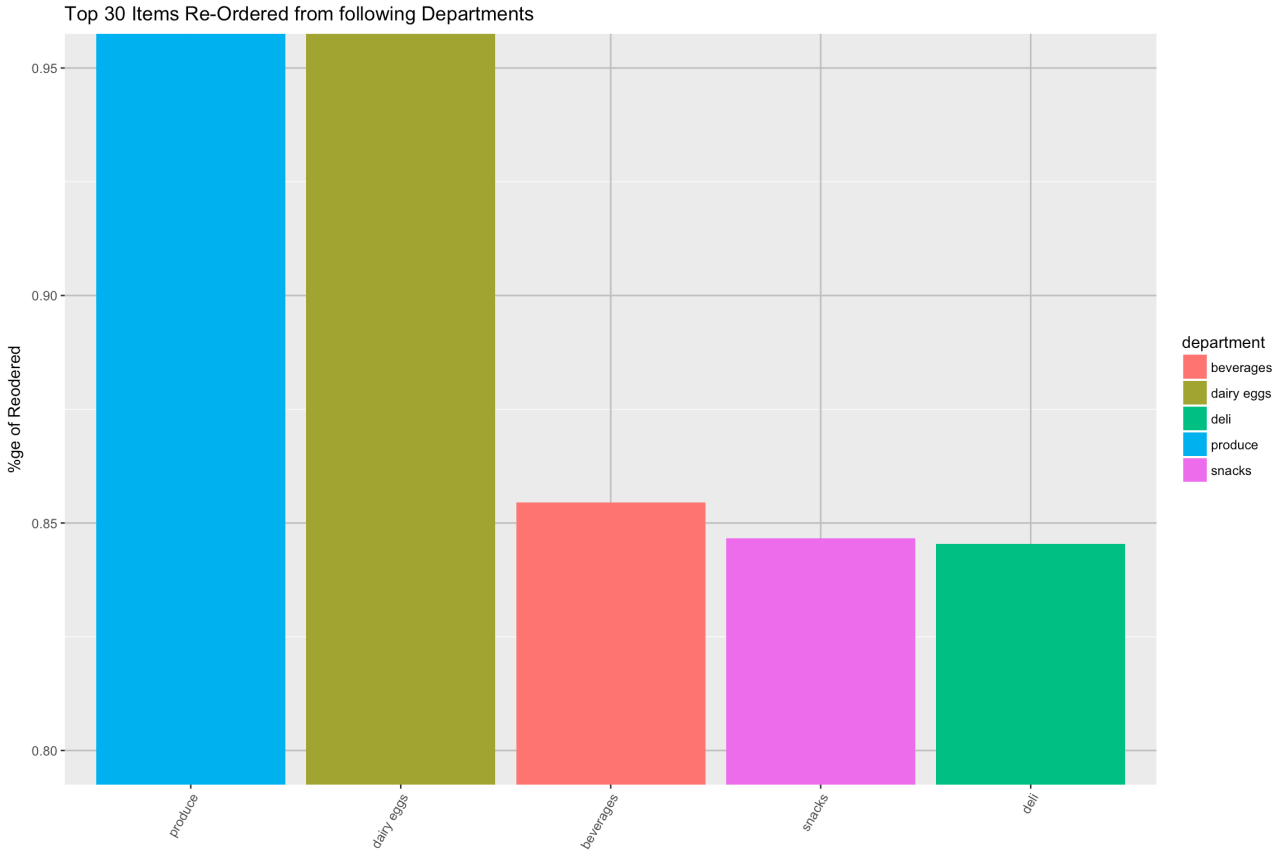
8. Trying to check, which is the Highest Re-Ordered Department

As you can see products from Produce and Dairy Eggs are reordered more than 95% of the time

product_id	percent_reordered	n	product_name	aisle_id	department_id	department
20940	0.9130435	368	Organic Low Fat Milk	84	16	dairy eggs
24852	0.8841717	18726	Banana	24	4	produce
117	0.8833333	120	Petit Suisse Fruit	2	16	dairy eggs
39180	0.8819876	483	Organic Lowfat 1% Milk	84	16	dairy eggs
12384	0.8810409	269	Organic Lactose Free 1% Lowfat Milk	91	16	dairy eggs
24024	0.8785249	461	1% Lowfat Milk	84	16	dairy eggs
38689	0.8756219	1206	Organic Reduced Fat Milk	84	16	dairy eggs
5514	0.8720000	125	Organic Homogenized Whole Milk	84	16	dairy eggs
36011	0.8692972	811	Organic Fat Free Milk	84	16	dairy eggs

product_id	percent_reordered	n	product_name	aisle_id	department_id	department
29447	0.8688811	572	Milk, Organic, Vitamin D	84	16	dairy eggs
22610	0.8688525	122	Reduced Fat 2% Lactose-Free Milk	91	16	dairy eggs
43394	0.8641509	265	Organic Lactose Free Whole Milk	91	16	dairy eggs
44799	0.8639798	397	Fresh CA Grown Eggs	86	16	dairy eggs
29118	0.8632479	117	French Vanilla Soy Creamer	91	16	dairy eggs
13176	0.8631783	15480	Bag of Organic Bananas	24	4	produce
17630	0.8617021	188	YoKids Squeeze! Organic Strawberry Flavor Yogurt	120	16	dairy eggs
2748	0.8571429	140	Organic Reduced Fat Omega-3 Milk	84	16	dairy eggs
34197	0.8571429	168	Goat Milk	84	16	dairy eggs
43643	0.8549223	193	Organic Strawana Probugs Kefir	120	16	dairy eggs
19731	0.8545455	110	Organic Orange Juice With Calcium & Vitamin D	31	7	beverages
32403	0.8524096	332	Fat Free Skim Milk	84	16	dairy eggs
26346	0.8516484	182	Organic Skim Milk	84	16	dairy eggs
37919	0.8506494	154	Organic 2% Chocolate Milk	84	16	dairy eggs
27845	0.8504482	4908	Organic Whole Milk	84	16	dairy eggs
40852	0.8502994	167	Lactose Free Fat Free Milk	91	16	dairy eggs
31981	0.8493590	312	1% Low Fat Milk	84	16	dairy eggs
45504	0.8476821	302	Whole Organic Omega 3 Milk	84	16	dairy eggs
45051	0.8466258	163	Pub Mix	107	19	snacks
14702	0.8454545	110	Maple Honey Ham	96	20	deli
6347	0.8436019	211	Unsweetened Almond Milk	91	16	dairy eggs

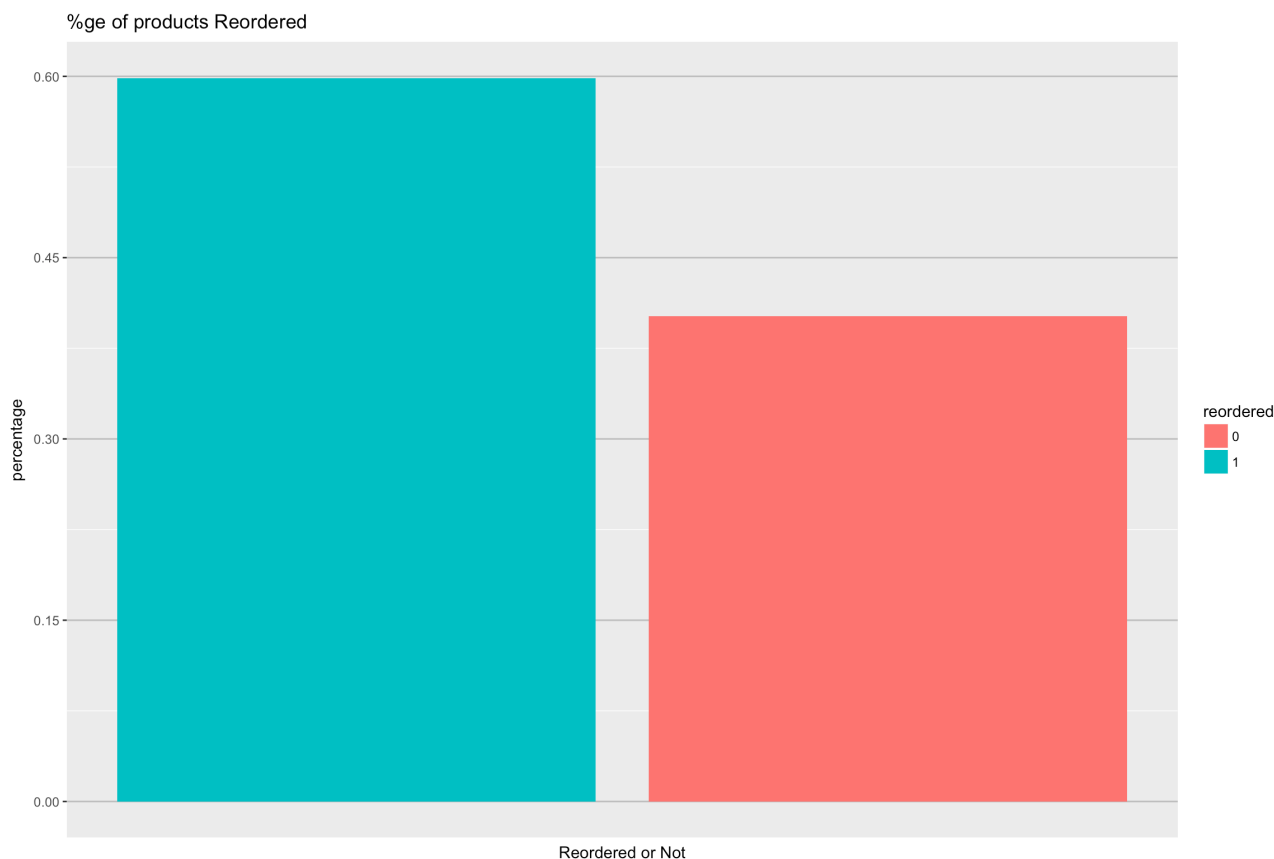
9. Finding out total %ge of products which are reordered



9. Finding out total %ge of products which are reordered

As you can se about 60% products are generally reordered

reordered	count	percentage
0	555793	0.4014056
1	828824	0.5985944



10. Trying to find out Max number of product Variety in the Aisles

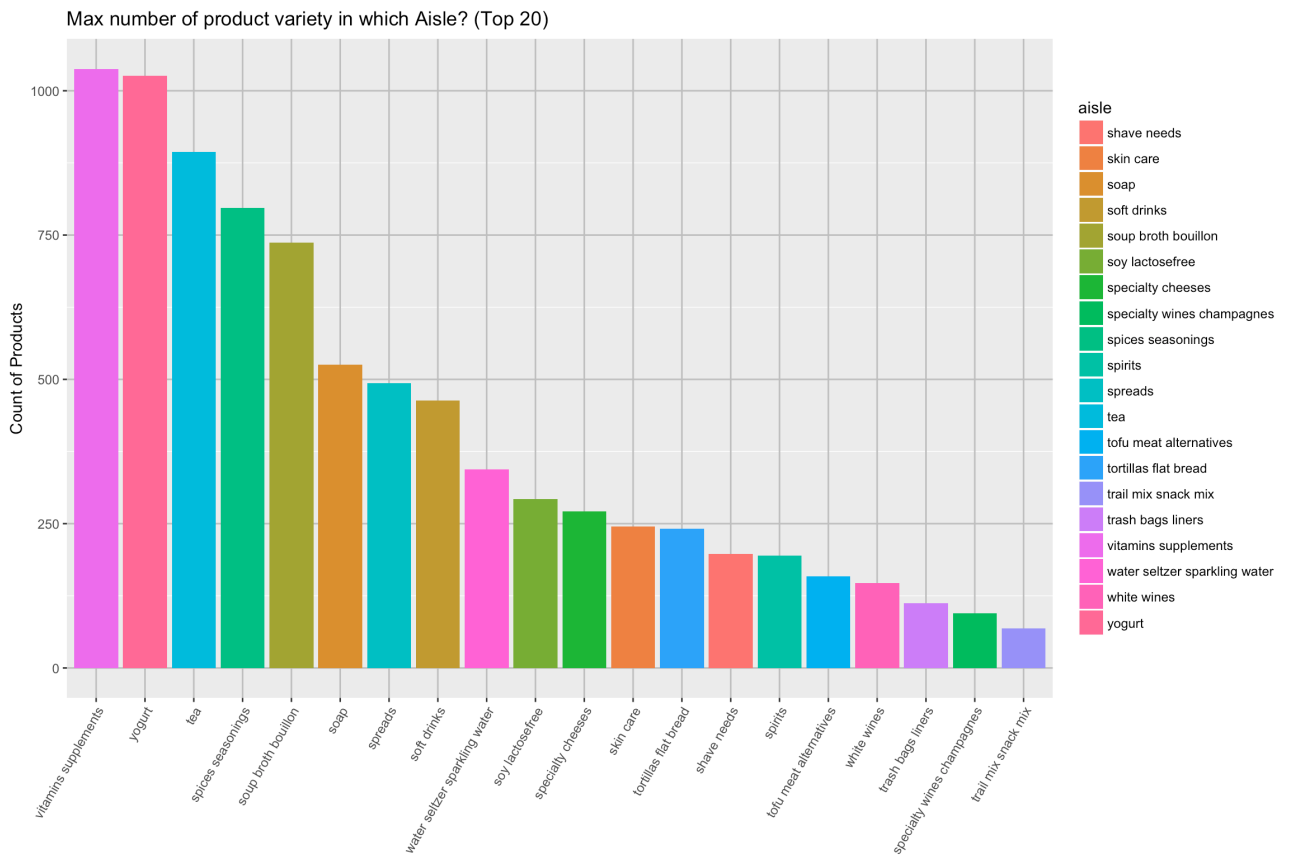
As we can see Vitamis SUpplements and yogurt aisles has the maximum product varieties

aisle_id	count	aisle
47	1038	vitamins supplements
120	1026	yogurt
94	894	tea
104	797	spices seasonings
69	737	soup broth bouillon
25	525	soap
88	493	spreads
77	463	soft drinks
115	344	water seltzer sparkling water
91	293	soy lactosefree
2	271	specialty cheeses
109	245	skin care
128	241	tortillas flat bread
55	198	shave needs

aisle_id	count	aisle
124	195	spirits
14	159	tofu meat alternatives
62	147	white wines
60	112	trash bags liners
134	95	specialty wines champagnes
125	69	trail mix snack mix

11. Trying to find out Max number of product Variety in the Departments
. Trying to find out Max number of product Variety in the Aislies

>

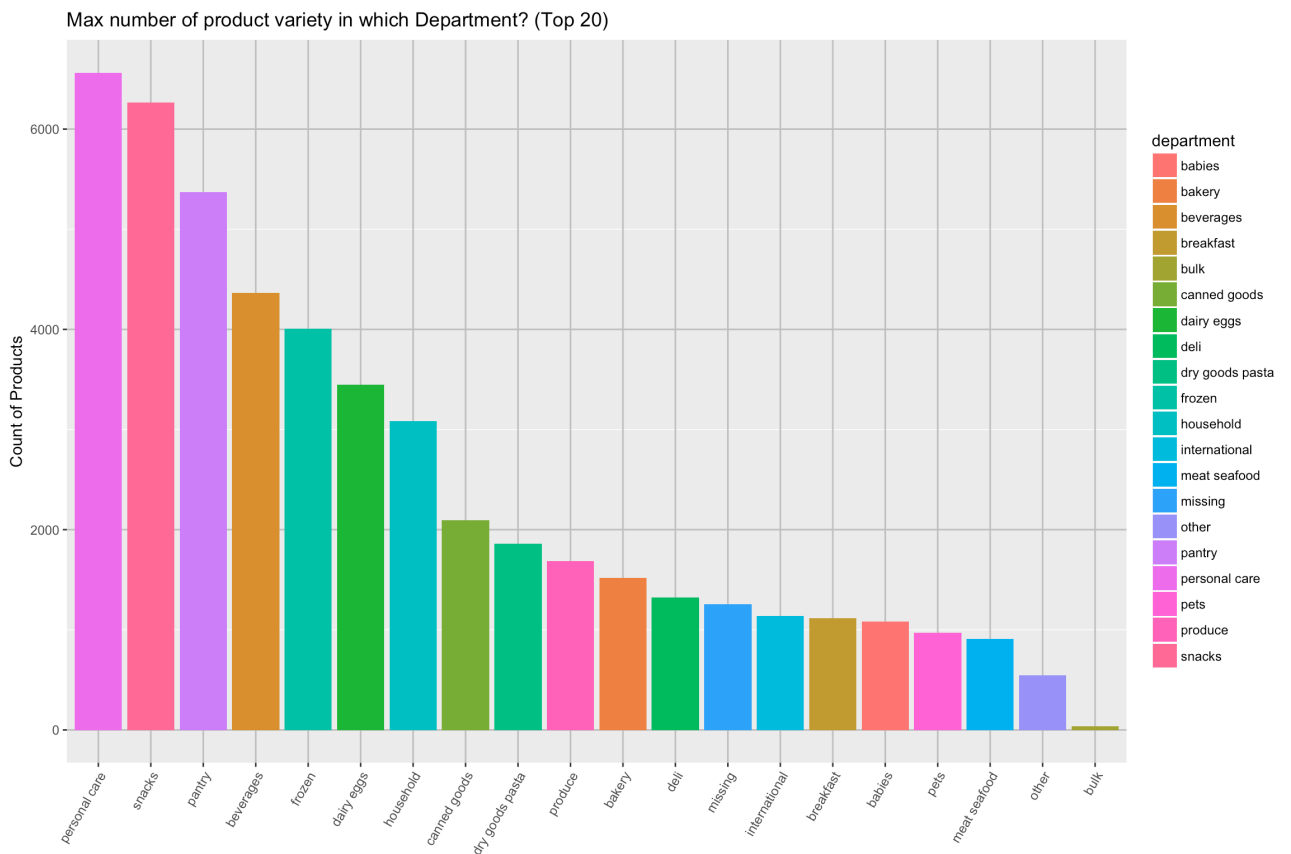


11. Trying to find out Max number of product Variety in the Departments

Trying to find out Max number of product Variety in the Departments

department_id	count	department
11	6563	personal care
19	6264	snacks
13	5371	pantry
7	4365	beverages

department_id	count	department
1	4007	frozen
16	3449	dairy eggs
17	3085	household
15	2092	canned goods
9	1858	dry goods pasta
4	1684	produce
3	1516	bakery
20	1322	deli
21	1258	missing
6	1139	international
14	1115	breakfast
18	1081	babies
8	972	pets
12	907	meat seafood
2	548	other
10	38	bulk

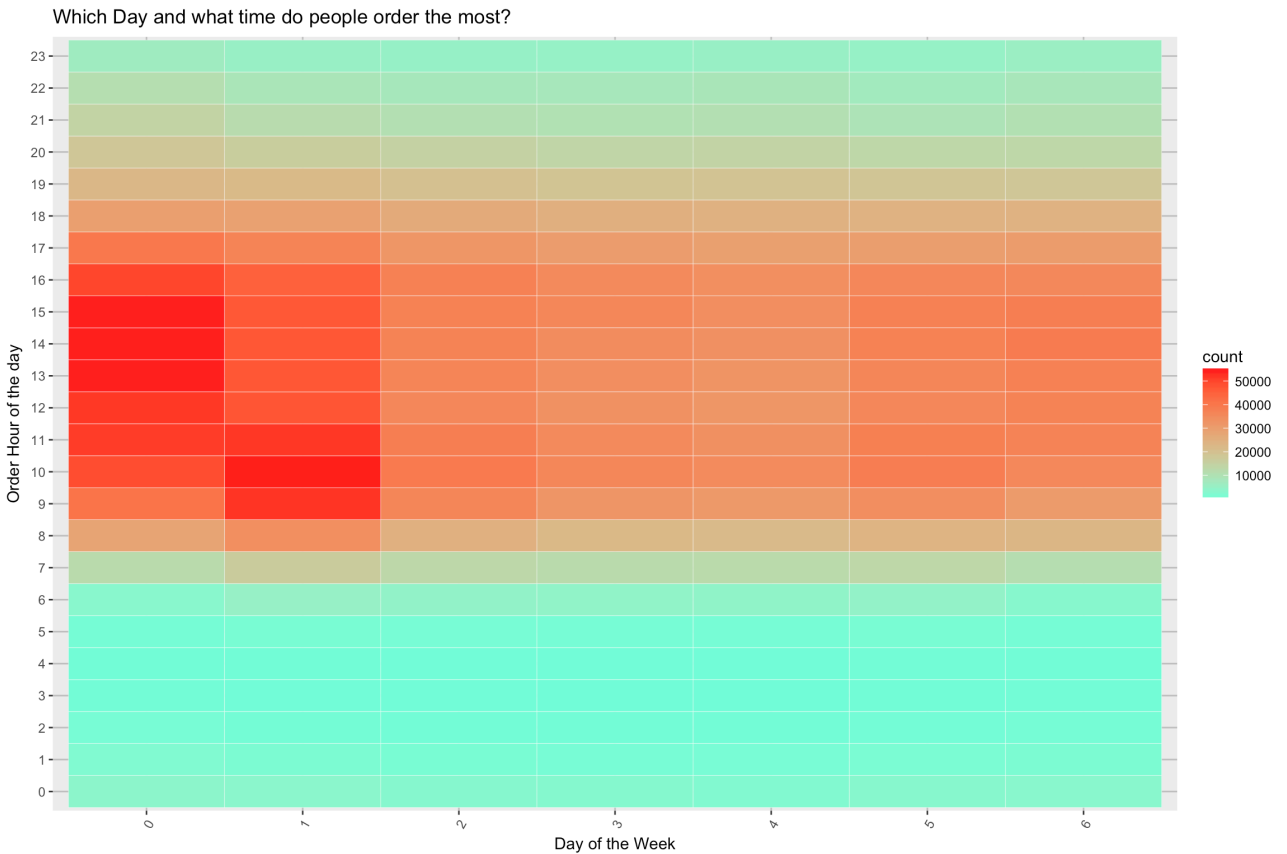


12. Heat map of Order hour vs order day

Trying to Find out WHich Day and what time do people order the most, #'People order the most on Day 0 & Day 1 between 9:00Am to 16:PM

order_dow	order_hour_of_day	count
1	10	55671
0	14	54552
0	15	53954
0	13	53849
1	9	51908
1	11	51584

13. No of Orders for Departments?



13. No of Orders for each Departments?

order_id	department
1	dairy eggs
1	produce
1	canned goods
36	dairy eggs

order_id department	
36	beverages
36	produce

[1] 640391 2

