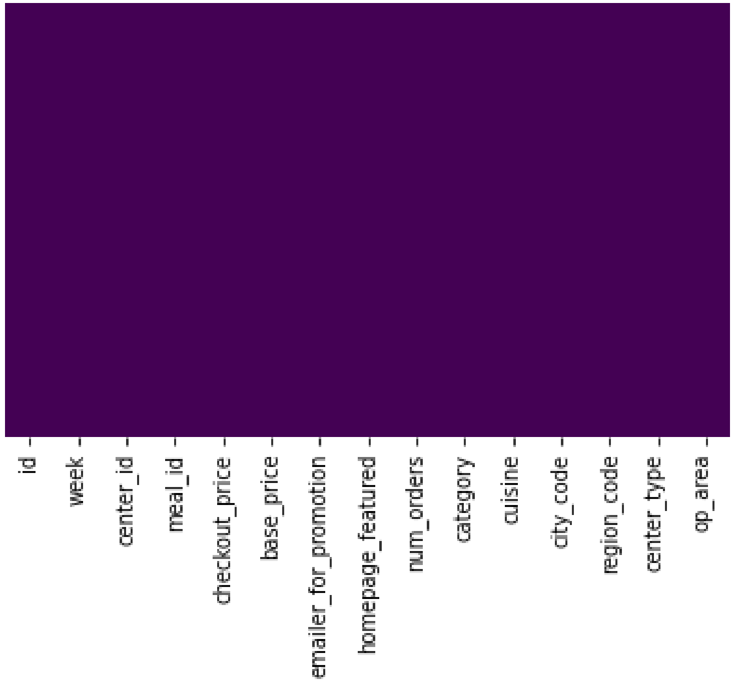
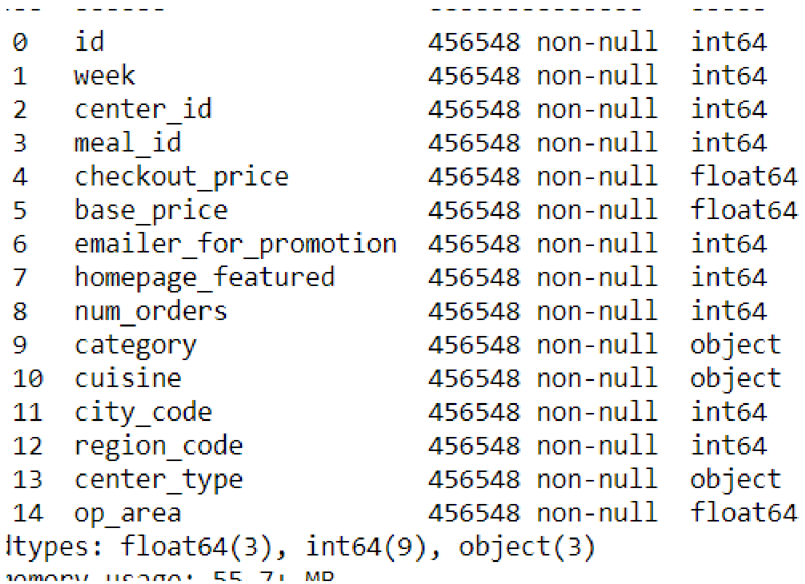
Basic Data Preprocessing



Here there are no null values in this dataset

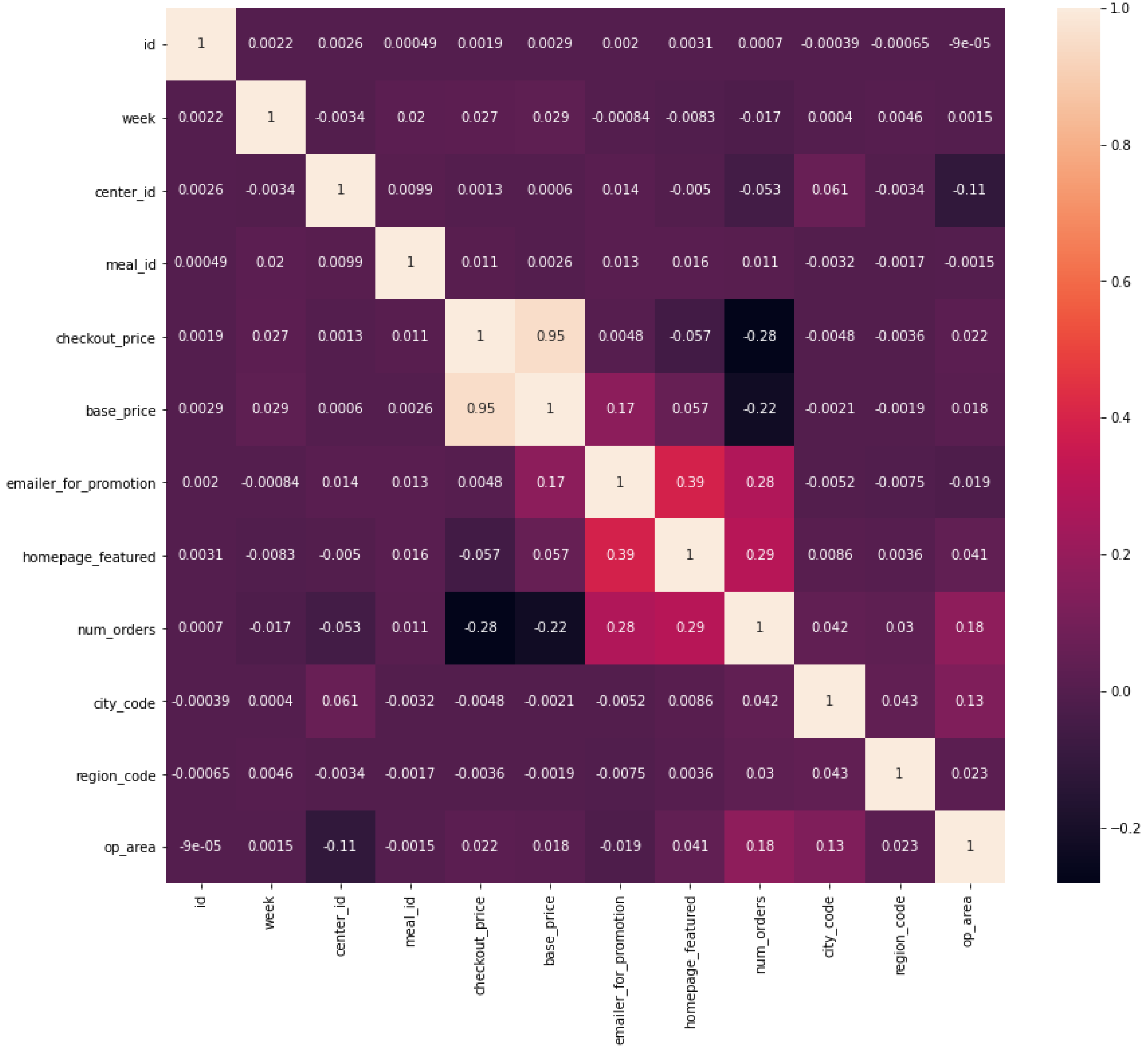


```
---  ------                 --------------  -----
 0   id                     456548 non-null  int64
 1   week                   456548 non-null  int64
 2   center_id              456548 non-null  int64
 3   meal_id                456548 non-null  int64
 4   checkout_price         456548 non-null  float64
 5   base_price             456548 non-null  float64
 6   emailer_for_promotion  456548 non-null  int64
 7   homepage_featured      456548 non-null  int64
 8   num_orders             456548 non-null  int64
 9   category               456548 non-null  object
 10  cuisine                456548 non-null  object
 11  city_code              456548 non-null  int64
 12  region_code            456548 non-null  int64
 13  center_type            456548 non-null  object
 14  op_area                456548 non-null  float64
ltypes: float64(3), int64(9), object(3)
memory usage: 55.7+ MB
```

We don’t have any date time converstion and duplicate values

Credits: https://github.com/bhanubokkasam

Correlation



	id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	num_orders	city_code	region_code	op_area
id	1	0.0022	0.0026	0.00049	0.0019	0.0029	0.002	0.0031	0.0007	-0.00039	-0.00065	-9e-05
week	0.0022	1	-0.0034	0.02	0.027	0.029	-0.00084	-0.0083	-0.017	0.0004	0.0046	0.0015
center_id	0.0026	-0.0034	1	0.0099	0.0013	0.0006	0.014	-0.005	-0.053	0.061	-0.0034	-0.11
meal_id	0.00049	0.02	0.0099	1	0.011	0.0026	0.013	0.016	0.011	-0.0032	-0.0017	-0.0015
checkout_price	0.0019	0.027	0.0013	0.011	1	0.95	0.0048	-0.057	-0.28	-0.0048	-0.0036	0.022
base_price	0.0029	0.029	0.0006	0.0026	0.95	1	0.17	0.057	-0.22	-0.0021	-0.0019	0.018
emailer_for_promotion	0.002	-0.00084	0.014	0.013	0.0048	0.17	1	0.39	0.28	-0.0052	-0.0075	-0.019
homepage_featured	0.0031	-0.0083	-0.005	0.016	-0.057	0.057	0.39	1	0.29	0.0086	0.0036	0.041
num_orders	0.0007	-0.017	-0.053	0.011	-0.28	-0.22	0.28	0.29	1	0.042	0.03	0.18
city_code	-0.00039	0.0004	0.061	-0.0032	-0.0048	-0.0021	-0.0052	0.0086	0.042	1	0.043	0.13
region_code	-0.00065	0.0046	-0.0034	-0.0017	-0.0036	-0.0019	-0.0075	0.0036	0.03	0.043	1	0.023
op_area	-9e-05	0.0015	-0.11	-0.0015	0.022	0.018	-0.019	0.041	0.18	0.13	0.023	1

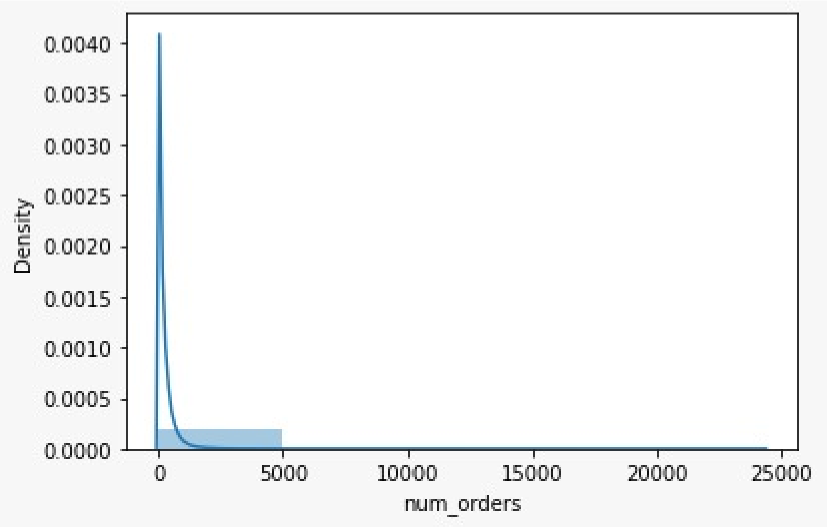
Our output (target) variable is num_orders
If we see correlation with it, main features which has high correlation are:

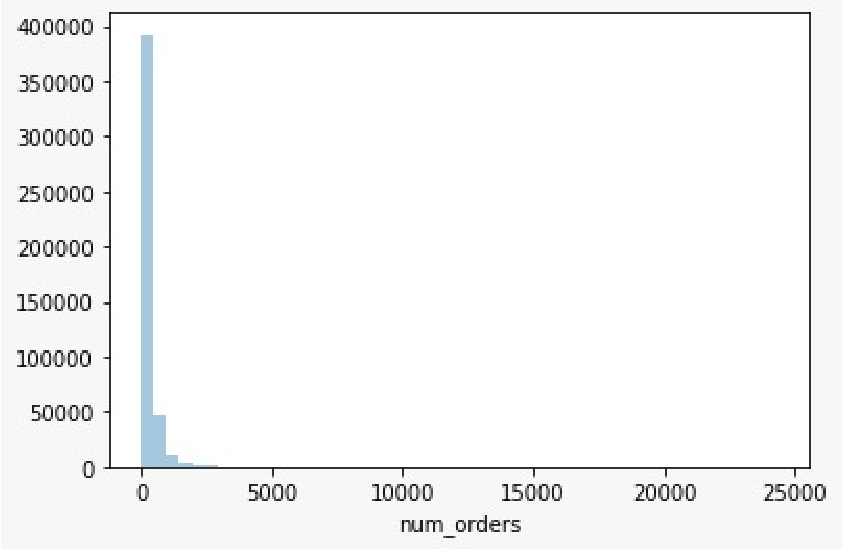
- Checkout_price
- base_price
- Emailer_promotion
- Homepage_freatured
- Op_area

Let's se all features comparison.

Credits: https://github.com/bhanubokkasam

Target column analysis

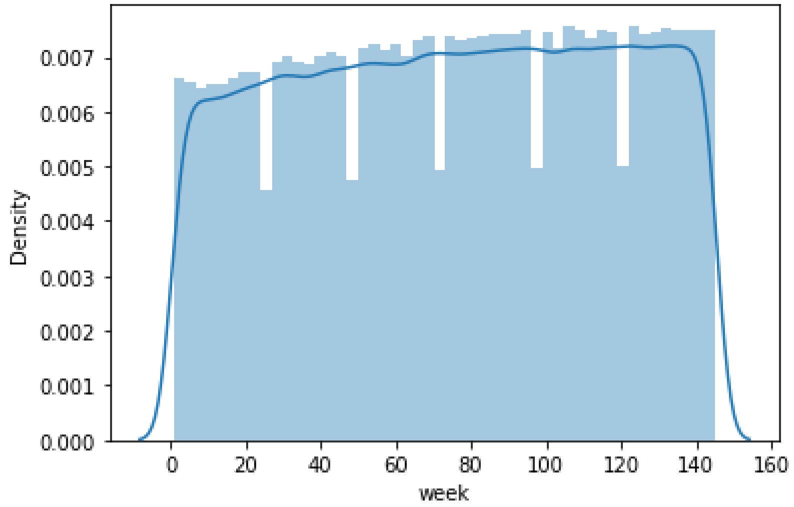


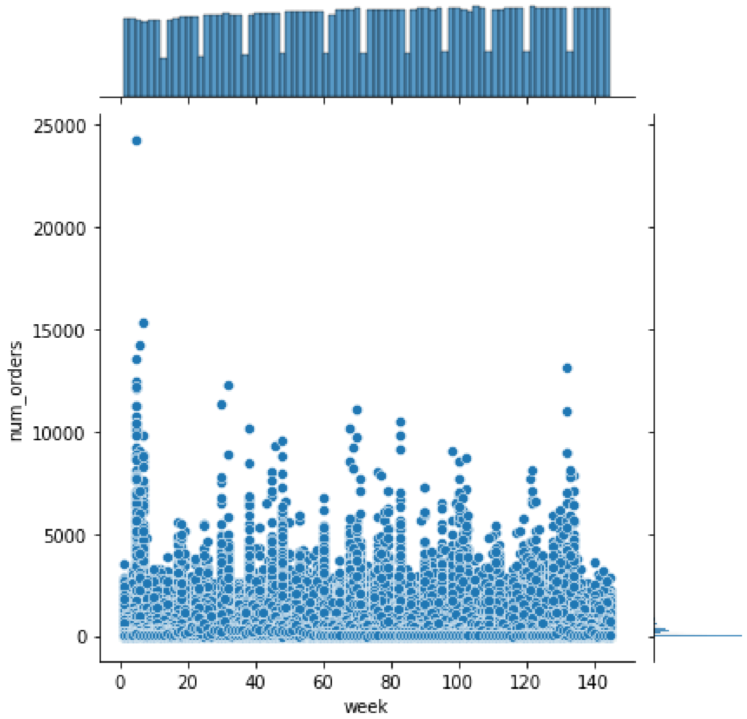


Here we got right skewed. Mode is at peak.

Credits: https://github.com/bhanubokkasam

week

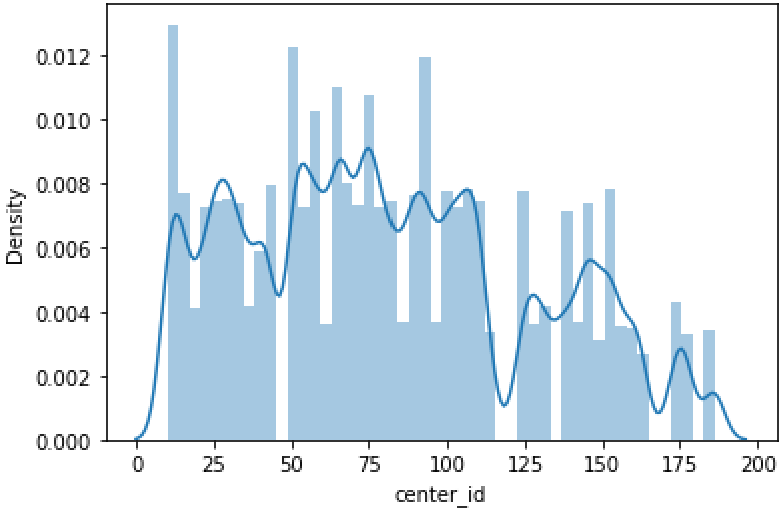


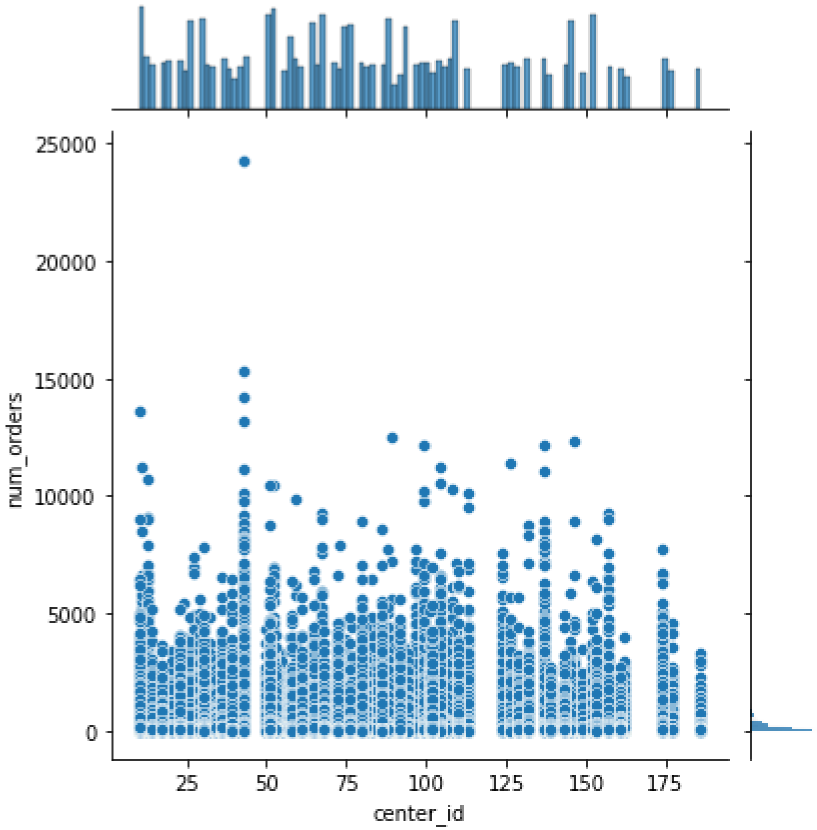


If we see week column. In some weeks (i.e 22,44,66,88,12) we are getting very less orders compared to other . This means after some weeks they are getting less orders. Maximum weeks they are getting less than 5000.

Credits: https://github.com/bhanubokkasam

Center id

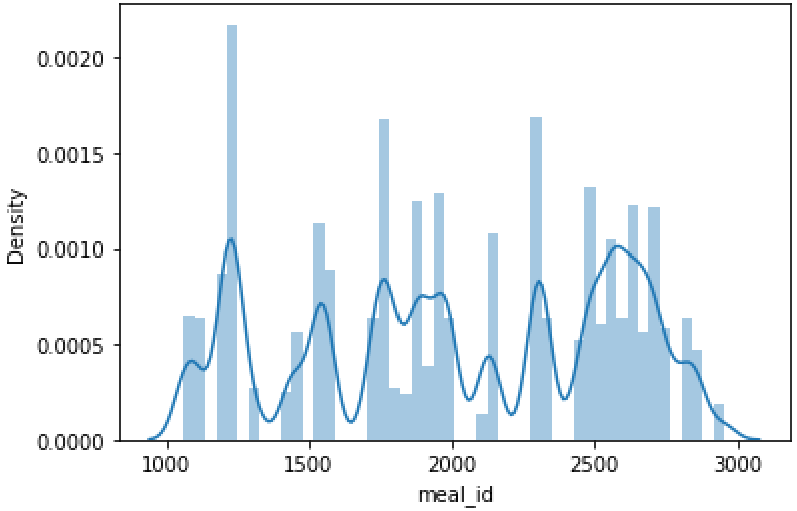


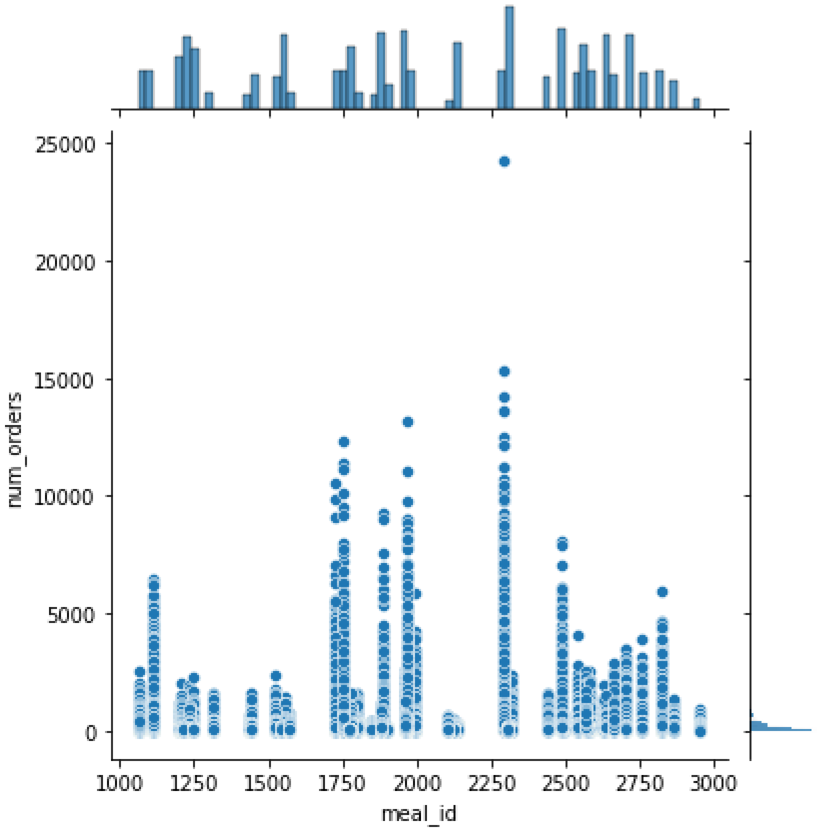


Center id means the id of distribution center.
Some centers has highest orders by seeing the area they are present. If we see some orders like 48 center we are getting high no of orders. Here around 5000 orders are placed in maximum centers.

Credits: https://github.com/bhanubokkasam

Meal Id

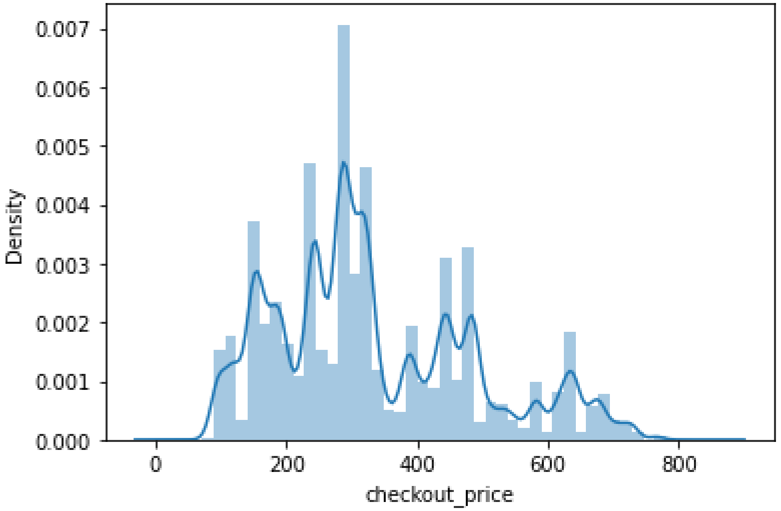


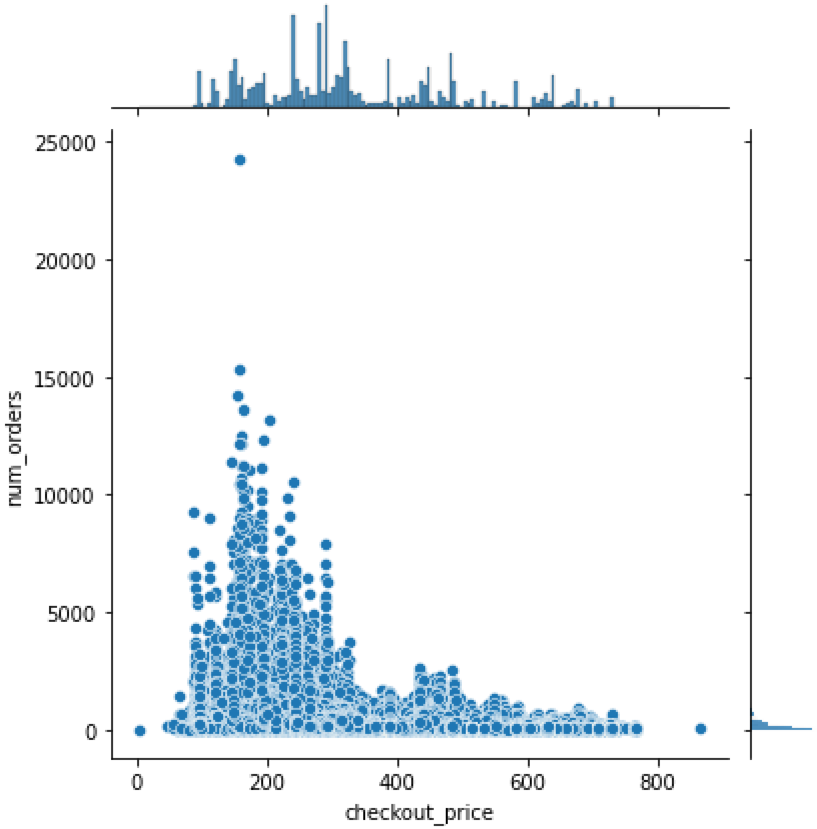


There are 51 types of meal id. For some of meals like 2300 (some type of meal) we are getting more number of orders. This means many people like that food and more orders may come for that type of food.

Credits: https://github.com/bhanubokkasam

Checkout Price

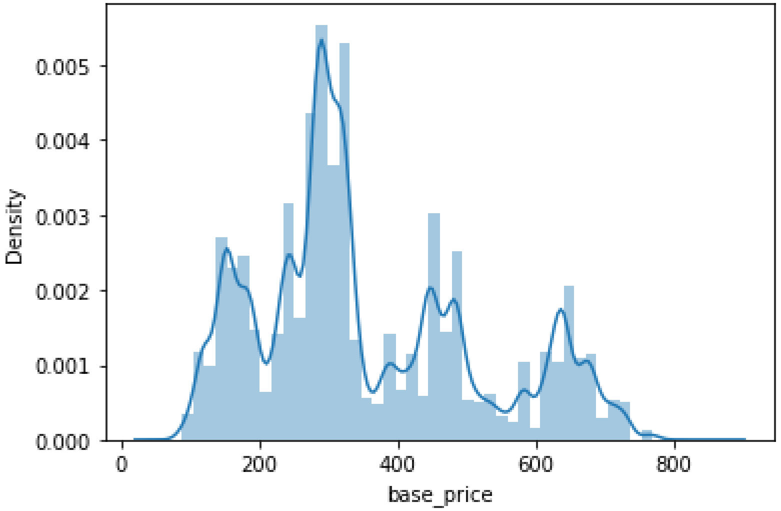


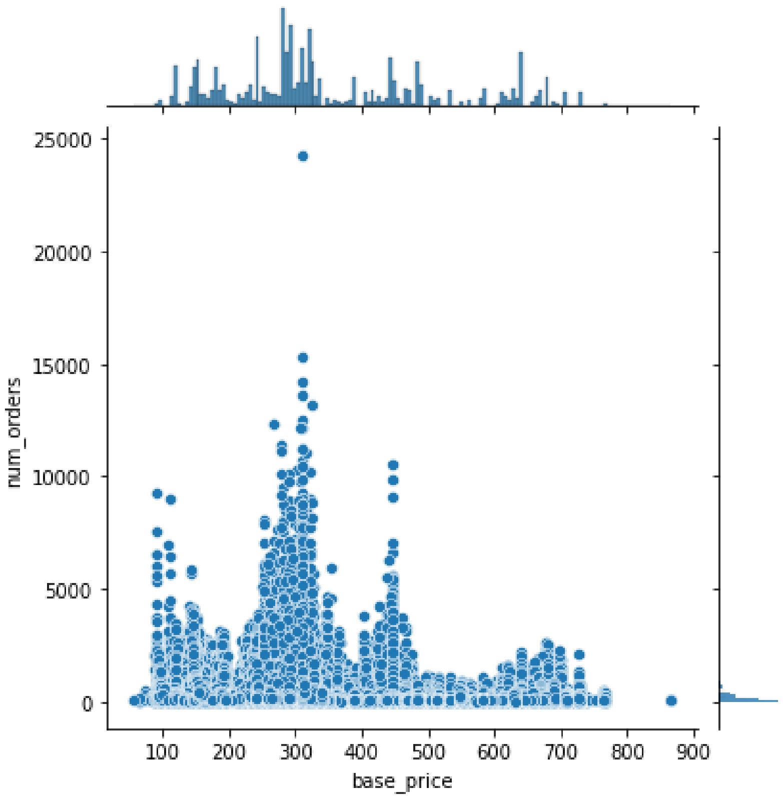


Here, checkout price has high significance. Because, maximum number of orders are placed by people acoording to price. If we see joint plot all values less than 300 are having high number of orders. This means if cost is high we may have less orders.

Credits: https://github.com/bhanubokkasam

Base Price

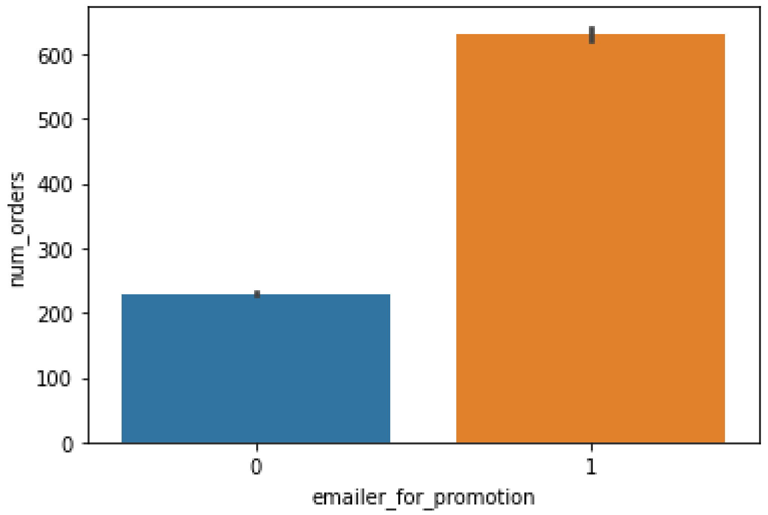




Here, base price has good significance. If we see joint plot all values less than 500 are having high number of orders. This means if cost is high we may have less orders. If we see PDF and CDF maximum values are before 500. Around 90% orders are having cost less than 500 price.

Credits: https://github.com/bhanubokkasam

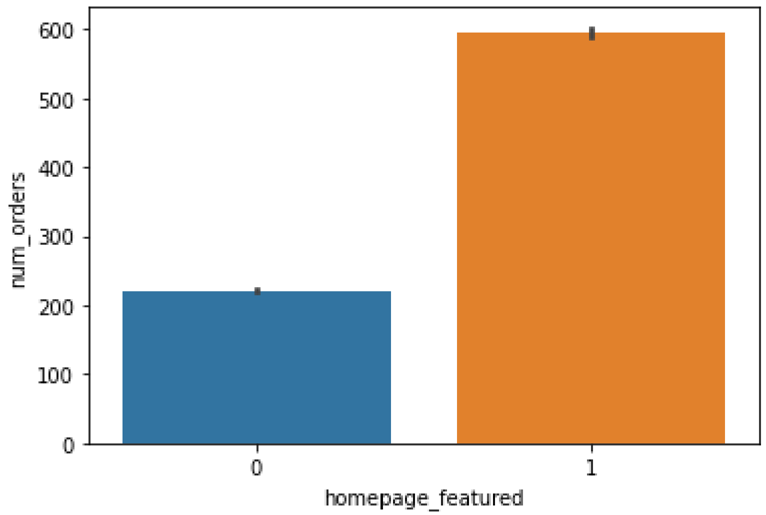
Emailer for promotion



Email has showing good role. If the mail has sent for promotion the number of orders are high. Because they are doing promotions.

Credits: https://github.com/bhanubokkasam

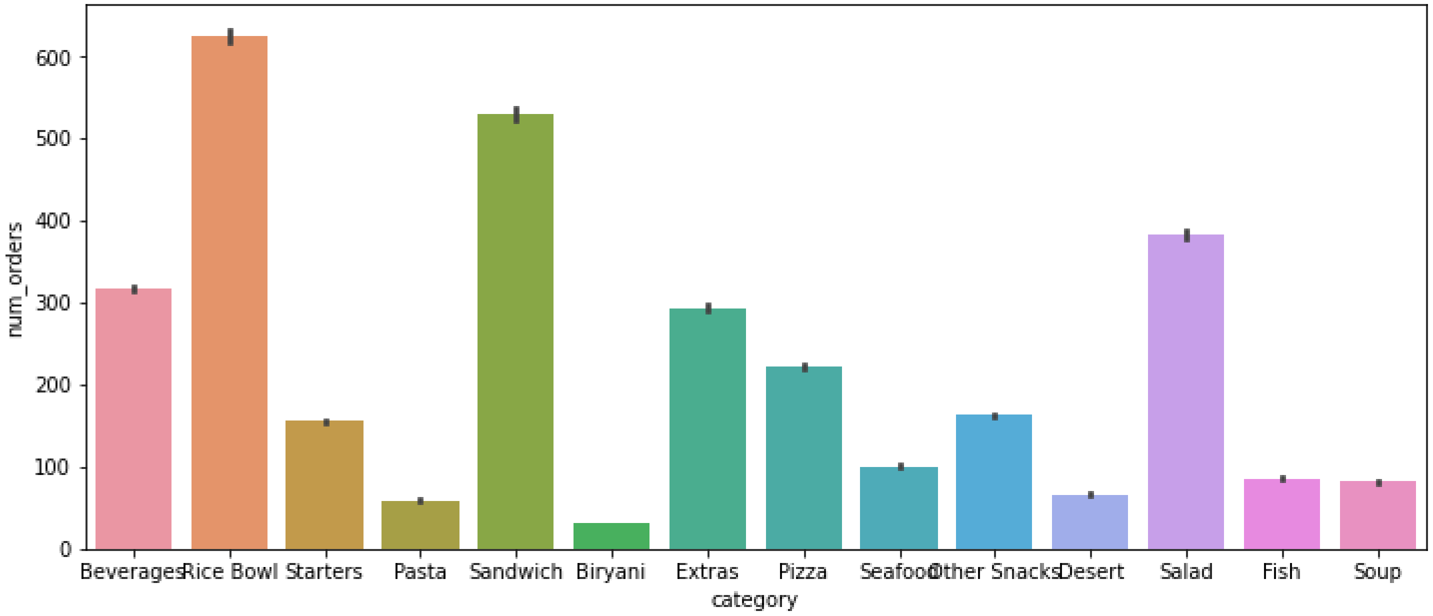
Homepage_Featured



If the product is featured in home page we can say more number of people can see it and may place order for it. So if they are featuring a meal in homepage they are getting more orders.

Credits: https://github.com/bhanubokkasam

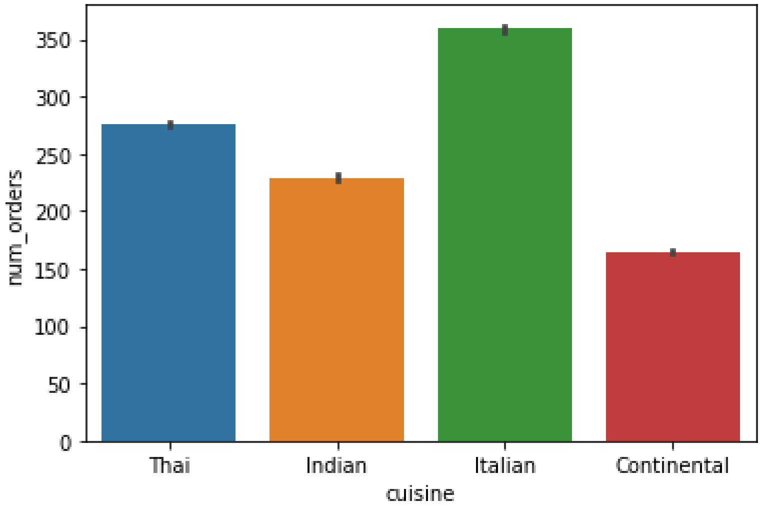
Category



Here based on type of food their sales are changing. For example rice bowl has more number of orders compared to biryani.

Credits: https://github.com/bhanubokkasam

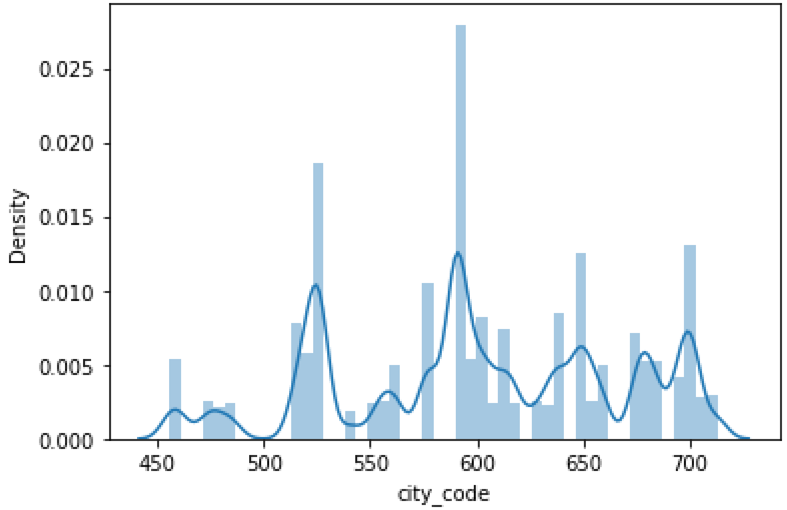
Cuisine

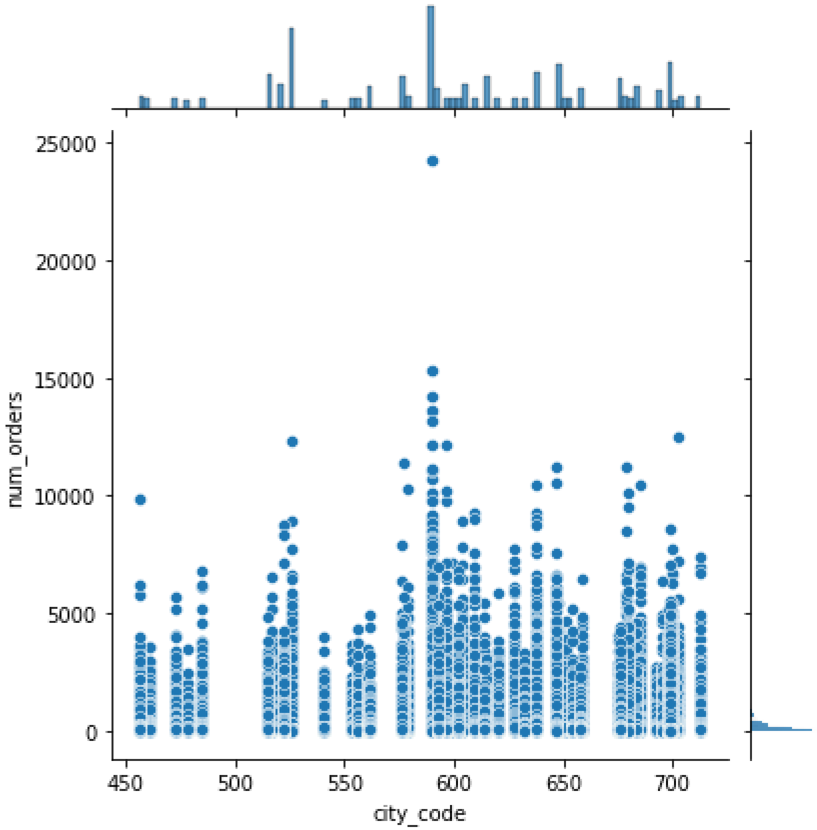


Here in cuisine Italian is more liked cuisine. This means based on the type of cuisine there may be change in number of orders.
So we can say more orders for Italian so all items required for making this cuisine should be ready.

Credits: https://github.com/bhanubokkasam

City Code

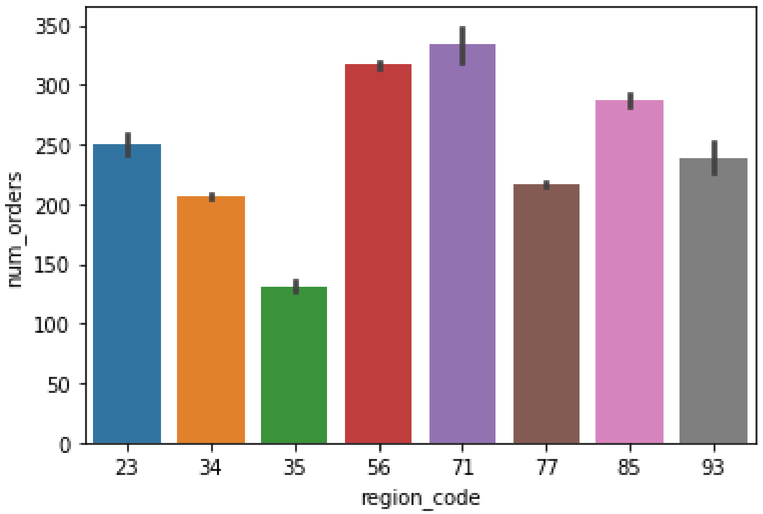




By seeing city code we can say in some cities there are more orders. For example in big cities the orders are more in number than small cities. In joint plot maximum orders are from cities which has code between 600 and 650.

Credits: https://github.com/bhanubokkasam

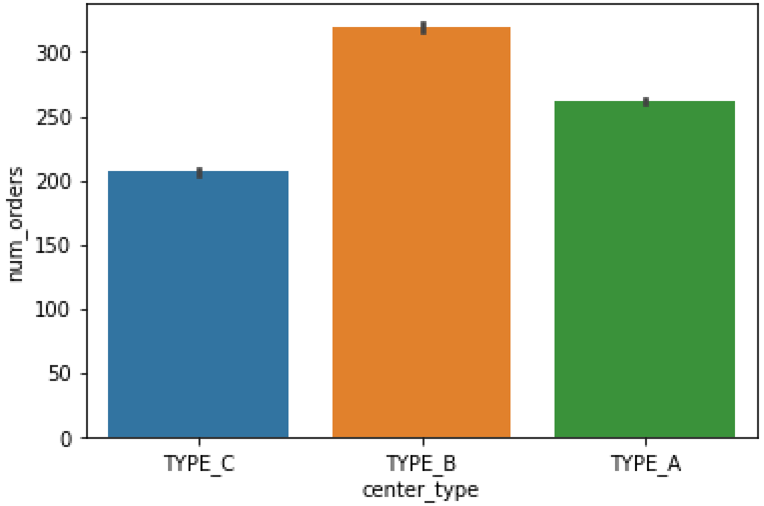
Region code



Here based on region code we can say people in different region are going to keep orders differently. For example in region number 71 and 56 there are more sales.

Credits: https://github.com/bhanubokkasam

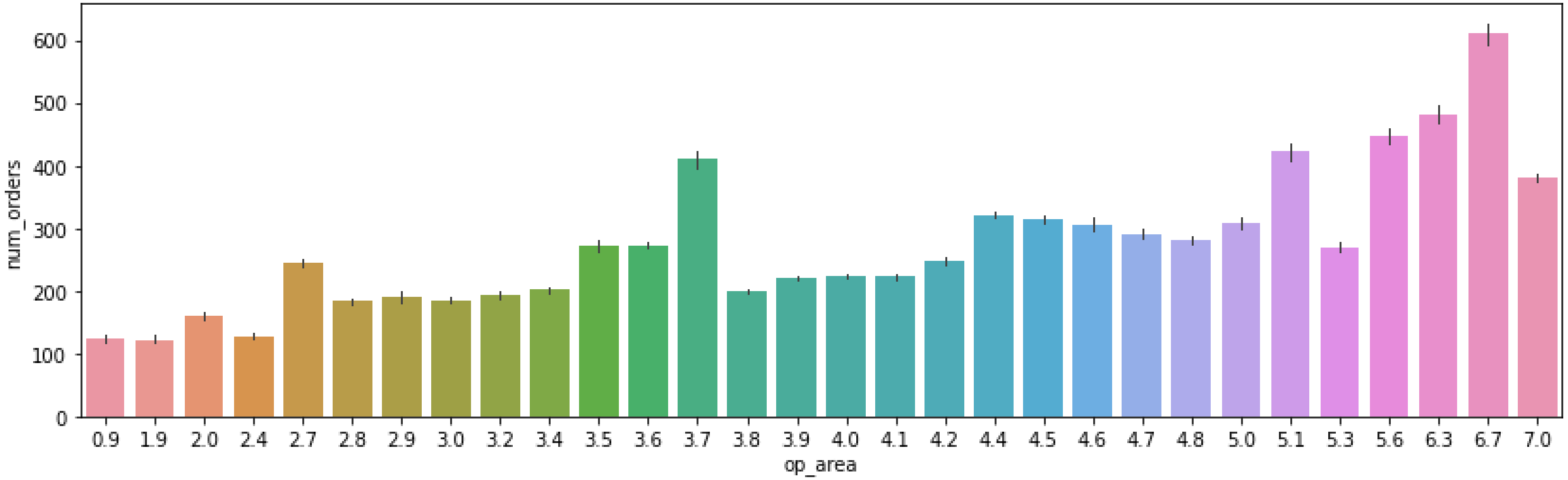
Center type



Here in type B there are more number of orders compared to A and C. This shows the orders from type B center may get high number of orders.

Credits: https://github.com/bhanubokkasam

Op Area



Here it is more correlated with target column. In above graph we can see area 6.7 more number of orders are placed. This shows the people in that area may place more number of orders.

Credits: https://github.com/bhanubokkasam