

## Preprocess, clean, and analyze the given dataset

Initially given attributes are->

```
Index(['url', 'address', 'name', 'online_order', 'book_table', 'rate', 'votes',  
      'phone', 'location', 'rest_type', 'dish_liked', 'cuisines',  
      'approx_cost(for two people)', 'reviews_list', 'menu_item',  
      'listed_in(type)', 'listed_in(city)'],  
      dtype='object')
```

Drop the following columns which are of no use or which less affect what we want like url, address, phone , etc.

```
Index(['url', 'address', 'name', 'online_order', 'book_table', 'rate', 'votes',  
      'phone', 'location', 'rest_type', 'dish_liked', 'cuisines',  
      'approx_cost(for two people)', 'reviews_list', 'menu_item',  
      'listed_in(type)', 'listed_in(city)'],  
      dtype='object')
```

Dropping the duplicates

Now filter each column

Clean Rate column

```
array(['4.1/5', '3.8/5', '3.7/5', '3.6/5', '4.6/5', '4.0/5', '4.2/5',  
      '3.9/5', '3.1/5', '3.0/5', '3.2/5', '3.3/5', '2.8/5', '4.4/5',  
      '4.3/5', 'NEW', '2.9/5', '3.5/5', nan, '2.6/5', '3.8 /5', '3.4/5',  
      '4.5/5', '2.5/5', '2.7/5', '4.7/5', '2.4/5', '2.2/5', '2.3/5',  
      '3.4 /5', '-', '3.6 /5', '4.8/5', '3.9 /5', '4.2 /5', '4.0 /5',  
      '4.1 /5', '3.7 /5', '3.1 /5', '2.9 /5', '3.3 /5', '2.8 /5',  
      '3.5 /5', '2.7 /5', '2.5 /5', '3.2 /5', '2.6 /5', '4.5 /5',  
      '4.3 /5', '4.4 /5', '4.9/5', '2.1/5', '2.0/5', '1.8/5', '4.6 /5',  
      '4.9 /5', '3.0 /5', '4.8 /5', '2.3 /5', '4.7 /5', '2.4 /5',  
      '2.1 /5', '2.2 /5', '2.0 /5', '1.8 /5'], dtype=object)
```

## Removing "NEW", "-" and "/5" from Rate Column

As /5 is of no use and model wants clean data or moreover it is easy for analyzing

## Filling Null Values in Rate Column with Mean

Listed in(city) and location, both are there, let's keep only one

So I drop listed\_in(city) column

Change the name of approx\_cost(for 2 people) with Cost2Plates as it's difficult to deal with big name and write again and again

Now Cost2Plates have values like

```
array(['800', '300', '600', '700', '550', '500', '450', '650', '400',  
      '900', '200', '750', '150', '850', '100', '1,200', '350', '250',  
      '950', '1,000', '1,500', '1,300', '199', '80', '1,100', '160',  
      '1,600', '230', '130', '50', '190', '1,700', '1,400', '180',  
      '1,350', '2,200', '2,000', '1,800', '1,900', '330', '2,500',  
      '2,100', '3,000', '2,800', '3,400', '40', '1,250', '3,500',  
      '4,000', '2,400', '2,600', '120', '1,450', '469', '70', '3,200',  
      '60', '560', '240', '360', '6,000', '1,050', '2,300', '4,100',  
      '5,000', '3,700', '1,650', '2,700', '4,500', '140'], dtype=object)
```

So comma should be removed for proper functioning

Let's simply define a function which filters the Cost2Plates

## Cleaning Rest Type Column

There are many rest\_types so rest\_types which have less count by this total no of rest\_types increasing which creates a lot of confusion

So, all rest\_types whose value is less than 1000(let) categorized as 1 column i.e. Others

So total other count as 9003

Quick Bites	19010
Casual Dining	10253
others	9003
Cafe	3682
Delivery	2574
Dessert Parlor	2242
Takeaway, Delivery	2008
Bakery	1140
Casual Dining, Bar	1130

## Cleaning Location Column

Same thing is done with the Location Column as there are many different locations.

So, Location which have less than 300(let) restaurants all categorized into 1 column i.e Other

## Cleaning Cuisines Column

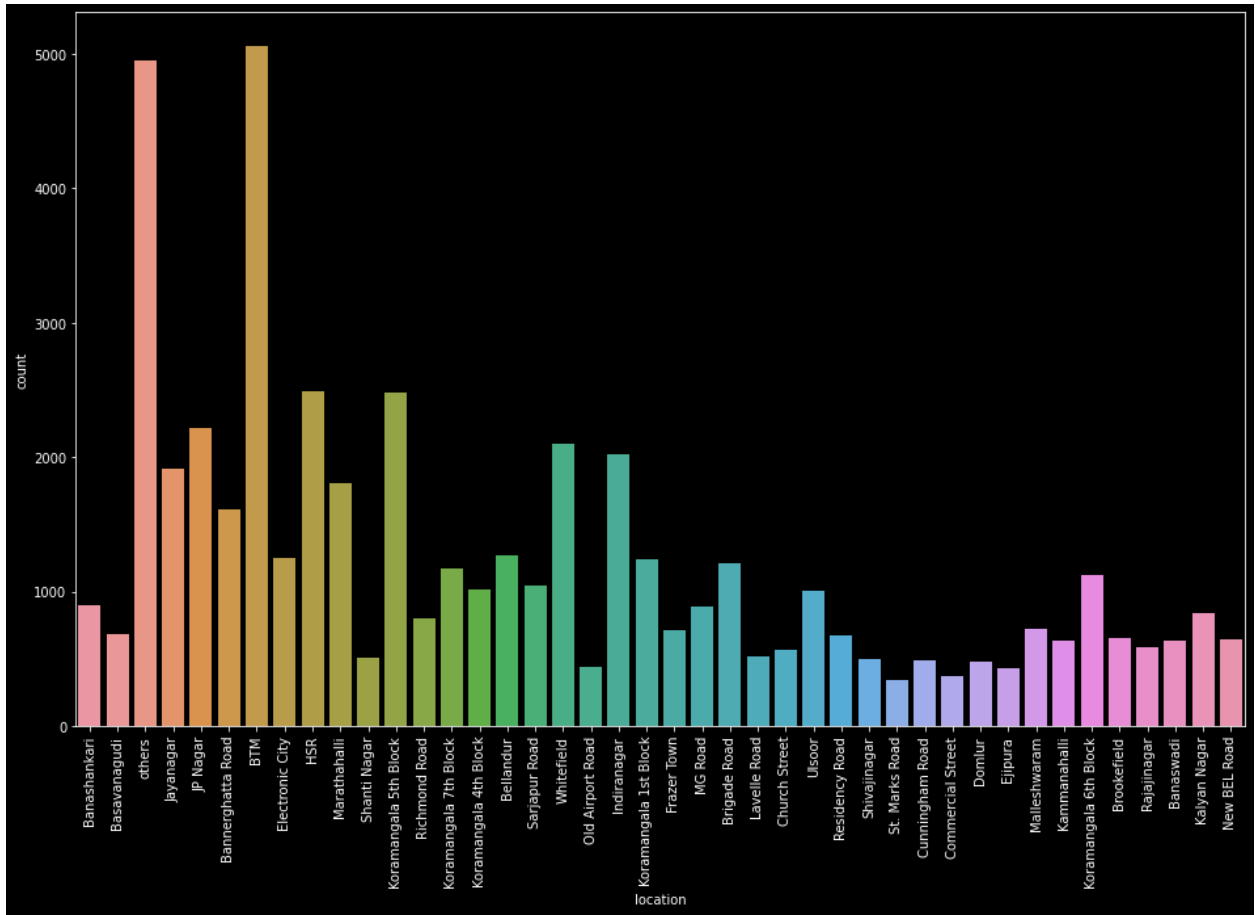
Same with Cuisine Column as there are many different Cuisine.

So, Cuisine which have less than 100(let) count all categorized into 1 column i.e Other

NOW ,

## Data is Clean, Lets jump to Visualization

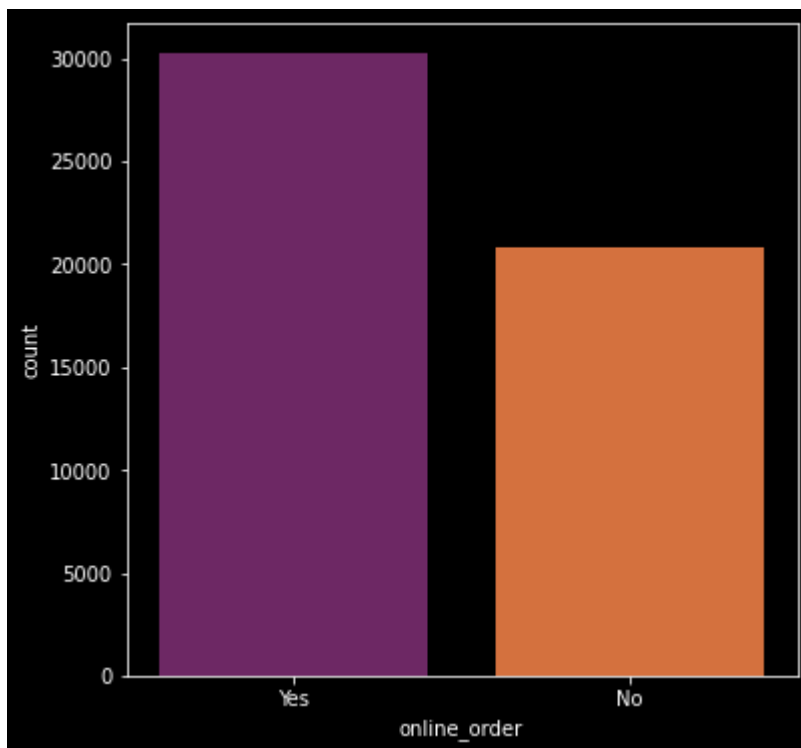
### Count Plot of Various Locations



Now by Observing, I saw that Location BTM has highest no of restaurants open  
So, it's not good to open a restaurant in BTM, If I open a restaurant in the  
Location which have less no of restaurants then the probability of success  
increases;

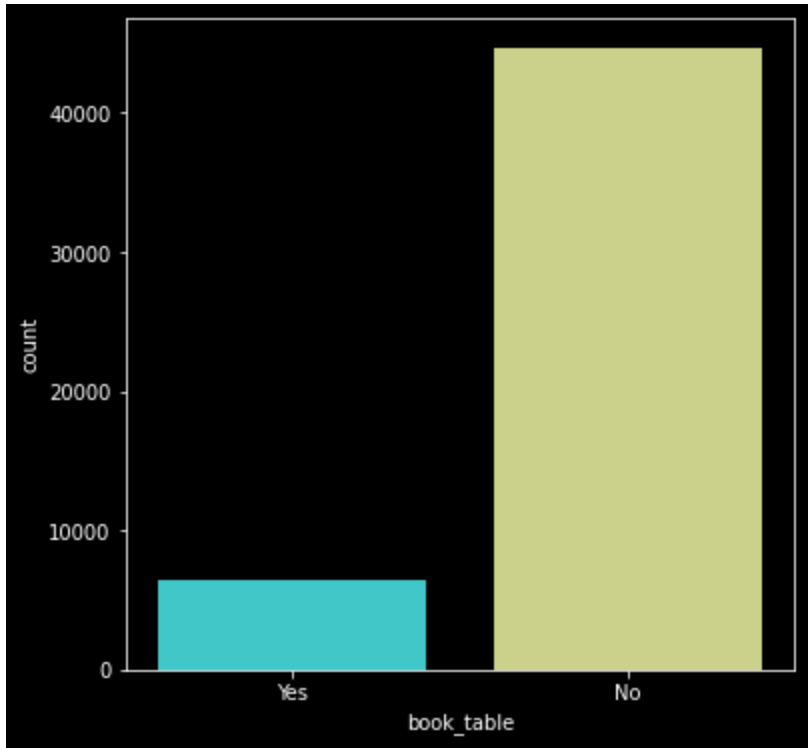
Like -Shanti Nagar, Old Airport Road, Lavelle Road, Church Street, ShivajiNagar,  
St. Markers Road, Commercial Street, Domlur, etc. ... The following are possible  
locations where chances of Success rate is high as there are less no of  
restaurants.

## Visualizing Online Order



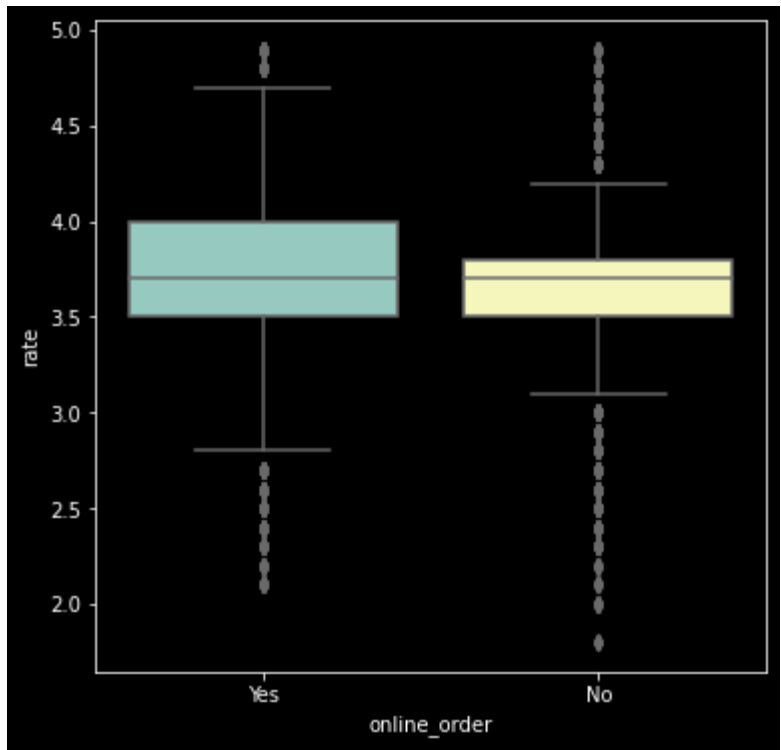
Depict More no of restaurants have Online\_order facility

## Visualizing Book Table



Book\_table facility offered by less no of restaurants  
So, if I took advantage and provide this facility then success rate of the restaurant will increases

## Visualizing Online Order vs Rate

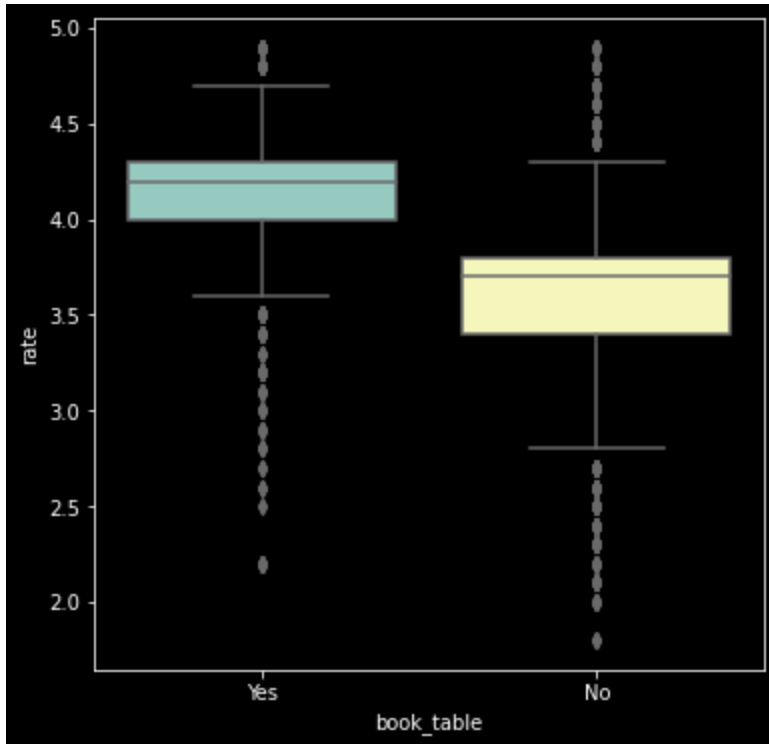


Here I compare the rating of restaurant which have online order facility and which are not having online  
So the average rating (rate) is 4.5 which have online order facility and average rating(rate) is 4.2 which are not having online order facility.

Online_order	Yes	No
Rate	4.5	4.2

So, If I want to open restaurant so i will make sure that it also have online order facility

## Visualizing Book Table vs Rate

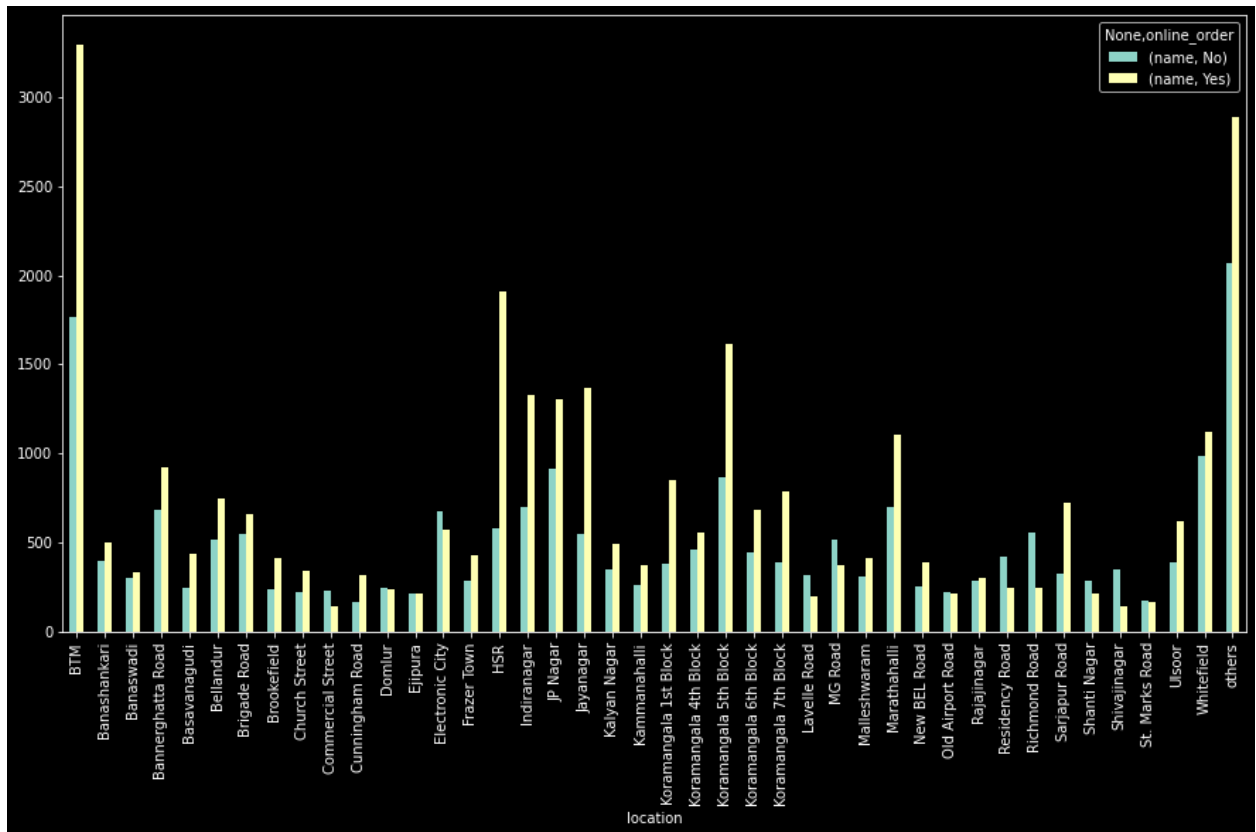


Here I compare the rating of restaurant which have Book\_table facility and which are not having Book\_table facility  
So the average rating (rate) is 4.5 which have Book\_table facility and average rating(rate) is 4.2 which are not having online order facility.

Book_table	Yes	No
Rate	4.4	3.8

So, If I want to open restaurant so i will make sure that it also have Book\_table facility

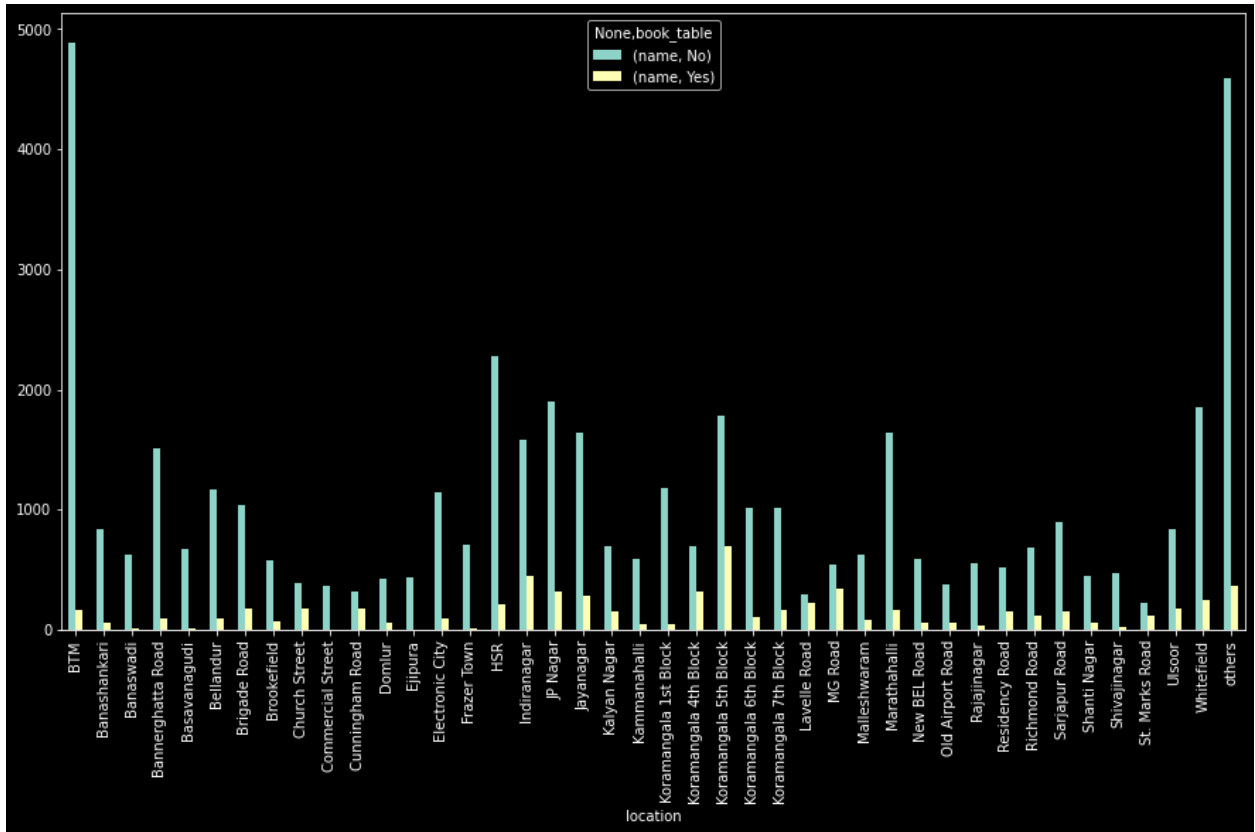
## Visualizing Online Order Facility, Location Wise



I visualize that the BTM yellow line is reaching high, already there are many restaurants who provide online\_order facility , so go for others like... Lavelle Road so there are more chances of success

## Visualizing Book Table Facility, Location Wise



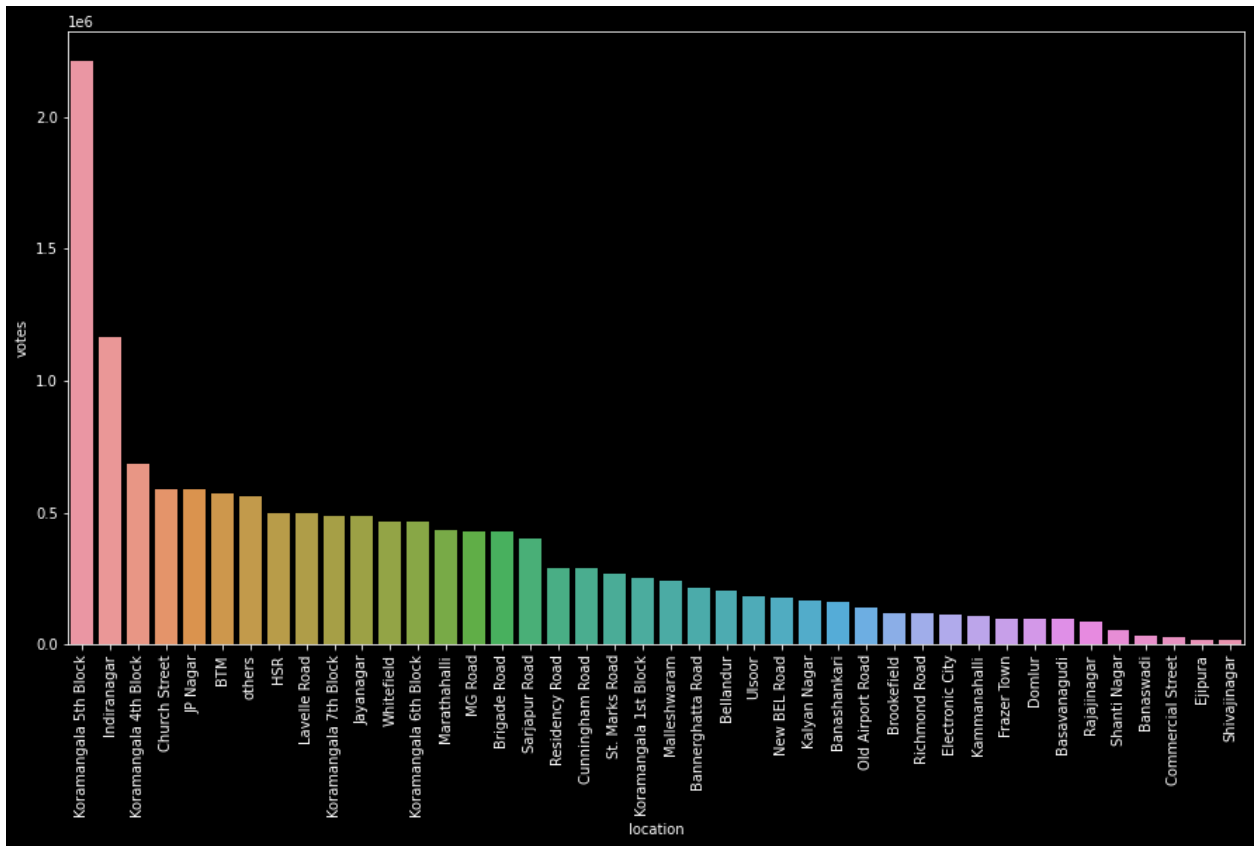


In Location BTM there are highest no of restaurants which not provide book\_table facility , but BTM location already have larger no of restaurants, so I go for HSR

## Visualizing Types of Restaurants vs Rate

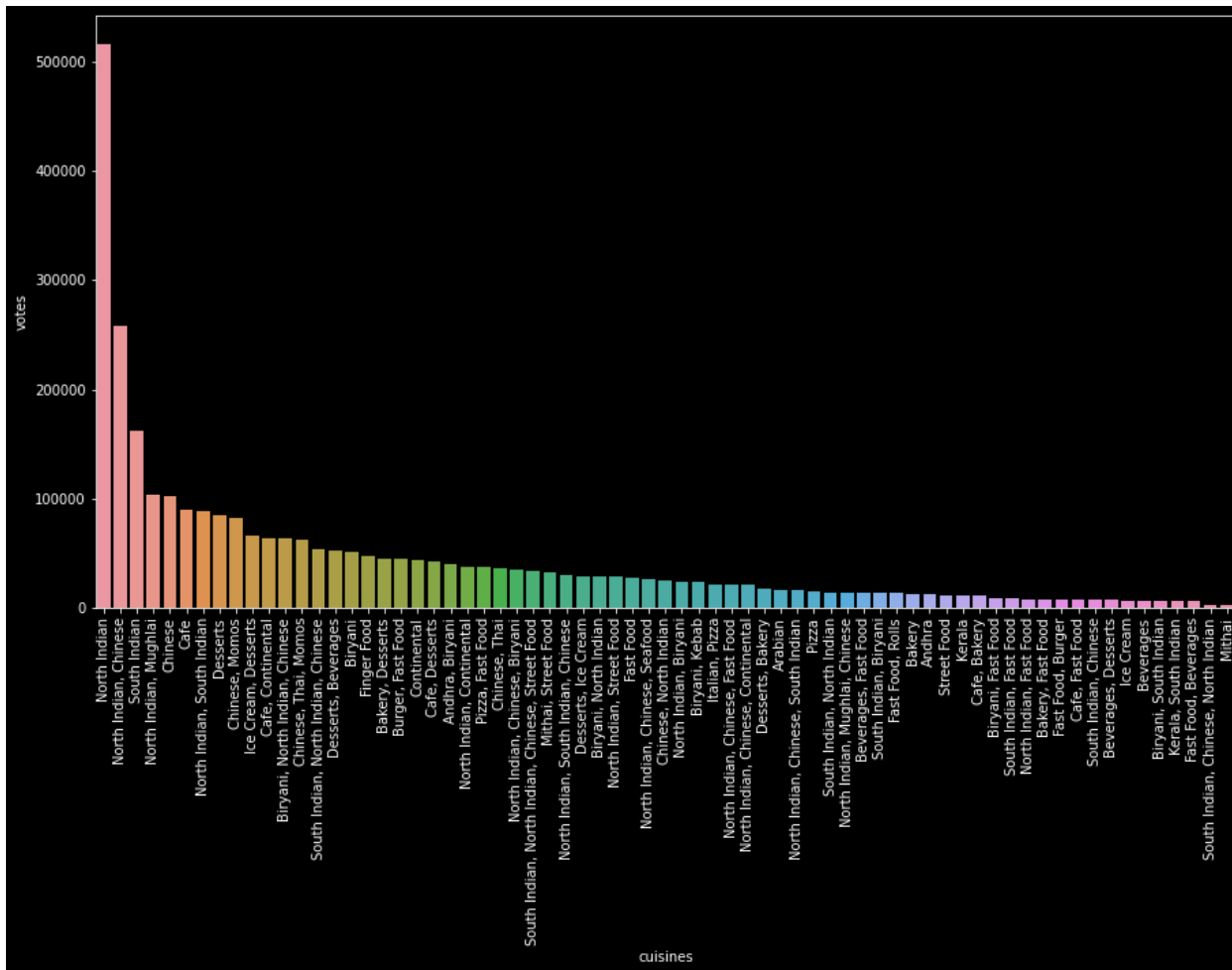


# No. of Votes, Location Wise



In Location Koramangala 5th Block people are interested in voting. They come and vote . Here i got very good customer feedback

# Visualizing Top Cuisines



North Indian food is getting the highest votes

After Visualizing many factors above which consider the restaurant to be successful i.e.

1. Location
2. Book\_table
3. Online\_order
4. Rate
5. Type\_of\_restaurant
6. No of votes
7. Top Cuisines

Like -Shanti Nagar, Old Airport Road, Lavelle Road, Church Street, ShivajiNagar, St. Markers Road, Commercial Street, Domlur, etc. ... The following are possible locations where chances of Success rate is high as there are less no of restaurants.

Book_table	Yes	No
Rate	4.4	3.8

Online_order	Yes	No
Rate	4.4	3.8

Here after visualizing above the average rating of Drinks & nightlife type of restaurant or Pubs And bars rating is also more.

Top 2 Type_of_Restaurants having highest Rating	Rating
Drinks & nightlife	4.3
Pubs and Bars	4.4

AND In Shivaji nagar pubs and bars have less no in this location, Shanti Nagar also have less number count.

Top 3 cuisines
North Indian
North Indian, Chinese
South Indian

The combination of drinks & nightlife with the North Indian cuisine with book\_table and online\_order facility and in the location where there are less no of restaurants have more chances of success. As bangalore is an IT hub so people are more comfortable in online order and booking table facility.If we open a restaurant near IT park Companies or in the residential area chances of success will be high.Moreover there are many outsiders which are from north india so the cuisine North Indian food has highest rating.