

# Exploratory Data Analysis



-By Akshay Gujare  
7741841945

# INTRODUCTION

ZOMATO

- In the age of digital dining, understanding the dynamics of restaurant choices and customer sentiments is pivotal. This project revolves around a thorough analysis of data from Zomato, a leading platform for restaurant discovery and food delivery.
- Zomato's dataset provides a goldmine of information on restaurants, user reviews, and ratings. Our goal is to dig into this data, unveil meaningful insights, and offer practical recommendations for restaurants and Zomato.
- Through our exploration, we aim to uncover patterns, correlations, and key trends in user behavior, ultimately contributing to a better understanding of what makes a restaurant successful on Zomato.
- This report showcases our journey into the world of Zomato data analytics, emphasizing the actionable insights gained and their potential impact on the culinary industry.



# OBJECTIVES

ZOMATO

**1: Business Improvement:** - Identify areas for improvement in restaurant offerings, - service, and overall customer experience to enhance business performance.

**2. Market Insights:** - Gain insights into market trends, popular cuisines, and customer preferences to stay competitive in the food industry.

**3. Strategic Decision -Making:** - Provide data -driven insights to guide strategic decisions for both individual restaurants and Zomato as a platform.

**4. Customer Retention:** -Understand customer sentiments and preferences to implement strategies that enhance customer satisfaction and loyalty.

**5. Optimization of Resources:** -Optimize menu offerings, pricing strategies, and operational efficiency to make better use of resources and maximize profitability.



**6. Platform Enhancement:** - Contribute insights to improve the Zomato platform, making it more user-friendly and valuable for both customers and restaurant partners.

**7. Marketing and Promotions:** - Develop targeted marketing and promotional campaigns based on seasonal trends and user behavior to attract and retain customers.

**8. Competitive Analysis:** - Benchmark restaurant performance against competitors to identify strengths, weaknesses, and opportunities for differentiation.

**9. Revenue Growth:** Identify strategies for revenue growth, such as upselling, cross-selling, and attracting new customers.

**10. Operational Efficiency:** - Enhance operational efficiency by identifying areas for improvement and resource optimization.

**10. Operational Efficiency:** Enhance operational efficiency by identifying areas for improvement and resource optimization.



## **1. NumPy ('import numpy as np'): Numerical Operations:**

NumPy provides support for large, multi-dimensional arrays and matrices, along with a collection of high level mathematical functions, enabling efficient numerical operations in Python.

## **2. Pandas ('import pandas as pd'): Data Manipulation and Analysis:**

Pandas simplifies data manipulation and analysis with data structures like Series and DataFrame, making it a powerful tool for tasks such as cleaning, filtering, and exploring structured data.

## **3. Seaborn ('import seaborn as sns'): Statistical Data Visualization:**

Seaborn simplifies the creation of complex statistical visualizations, providing a high-level interface for producing informative and aesthetically pleasing plots.

## **4. Matplotlib ('import matplotlib.pyplot as plt'): 2D Plotting Library:**

Matplotlib is a versatile 2D plotting library for Python, offering a wide range of customizable static visualizations, including charts, graphs, and figures.



## **5. Plotly Express ('import plotly.express as px'):-**

Interactive Visualizations: Plotly Express is a high-level interface for generating interactive visualizations with minimal code, making it suitable for creating web-based interactive plots and charts.



# Data Overview

ZOMATO

## Loading Dataset-

ZOMATO CAPSTONE PROJECT

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

Here we import python most used libraries for making a valuable Report

```
[2] df=pd.read_csv("/content/drive/MyDrive/Indian-Restaurants.csv")
```

I've load dataset named 'Indian\_restaurants.csv' into a variable named 'df'. In order to import the dataset, I've used Pandas Libraries in which I've used 'pd.read\_csv' command to import the respective dataset.

## Viewing top rows -

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	...	price_range	currency	highlights	aggregate_rating	rating_text	votes	photo_count	opentable_support	delivery	takeway
0	3400299	Bikanervala	[Quick Bites]	https://www.zomato.com/agra/bikanervala-khand...-khand...	Kalyani Point, Near Tuti Cinema, Bypass Road...	Agra	34	Khandari	27.211450	78.002381	...	2	Rs.	["Lunch", "Takeaway Available", "Credit Card"]	4.4	Very Good	814	154	0.0	-1	-1
1	3400005	Mama Chicken	[Quick Bites]	https://www.zomato.com/agra/mama-chicken-mama...-mama...	Main Market, Sadar Bazaar, Agra Cantt, Agra	Agra	34	Agra Cantt	27.160569	78.011583	...	2	Rs.	["Delivery", "No Alcohol Available", "Dinner"]	4.4	Very Good	1203	161	0.0	-1	-1
2	3401013	Bhagat Halwai	[Quick Bites]	https://www.zomato.com/agra/bhagat-halwai-2-sh...	62/1, Near Easy Day, West Shivaji Nagar, Gaop...	Agra	34	Shahganj	27.182938	77.979684	...	1	Rs.	["No Alcohol Available", "Dinner", "Takeaway"]	4.2	Very Good	801	107	0.0	1	-1
3	3400290	Bhagat Halwai	[Quick Bites]	https://www.zomato.com/agra/bhagat-halwai-civil...	Near Aranya Cinema, Nehru Nagar, Civil Lines, ...	Agra	34	Civil Lines	27.205668	78.004799	...	1	Rs.	["Takeaway Available", "Credit Card", "Lunch"]	4.3	Very Good	693	157	0.0	1	-1
4	3401744	The Salt Cafe Kitchen & Bar	[Casual Dining]	https://www.zomato.com/agra/the-salt-cafe-kitchen...	1C, 3rd Floor, Fatehabad Road, Tajganj, Agra	Agra	34	Tajganj	27.157709	78.052421	...	3	Rs.	["Lunch", "Serves Alcohol", "Cash", "Credit Ca..."]	4.9	Excellent	470	291	0.0	1	-1
5	3400275	Domino's Pizza	[Quick Bites]	https://www.zomato.com/agra/domino-s-pizza-civil...	11423 G, Deep Shikha Complex, Sanjay Place, C...	Agra	34	Civil Lines	27.201516	78.007556	...	2	Rs.	["Credit Card", "Lunch", "Delivery", "Dinner"]	4.0	Very Good	707	62	0.0	-1	-1
6	3400296	Honeydew Restaurant	[Quick Bites]	https://www.zomato.com/agra/honeydew-restaurant...	Opposite Soami Bagh Temple, Dayal Bagh, Agra	Agra	34	Dayal Bagh	27.222175	78.010174	...	2	Rs.	["Dinner", "Delivery", "Lunch", "Cash", "Takeaway"]	4.2	Very Good	647	46	0.0	1	-1
7	3400368	Domino's Pizza	[Quick Bites]	https://www.zomato.com/agra/domino-s-pizza-sik...	Plot C-16, Sector 13, Sikandra, Agra	Agra	34	Sikandra	27.203930	77.954260	...	2	Rs.	["Lunch", "Delivery", "Credit Card", "No Alco..."]	3.8	Good	617	18	0.0	-1	-1
8	3401284	Cake House	[Bakery]	https://www.zomato.com/agra/cake-house-2-civil...	23/301, Wazirpura Rd, Judge Agra	Agra	34	Civil Lines	27.204148	78.009025	...	2	Rs.	["Takeaway Available", "Credit Card", "Cash", "Indoor"]	3.4	Average	322	14	0.0	1	-1

In order to import the dataset. and then checking top rows by using pandas libraries

## Viewing last rows of data -

res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	...	price_range	currency	highlights	aggregate_rating	rating_text	votes	photo_count	opentable_support	delivery	takeaway
211934	3200763	Swad	["Quick Bites"]	https://www.zomato.com/vadodara/swad-karelibau... Complex, Opposite Annapali Complex...	Vadodara	32	Karelibaug	22.320823	73.199167	...	1	Rs.	["Dinner", "Takeaway Available", "Delivery"]	4.0	Very Good	365	9	0.0	-1	-1
211935	3201351	Mummys Pizza	["Casual Dining"]	https://www.zomato.com/vadodara/mummys-pizza-d... Top Floor 323 - 327 Siddhivinayak Central Mall, D...	Vadodara	32	Dwailpura	22.280378	73.149108	...	2	Rs.	["Dinner", "Cash", "Takeaway Available", "Lunch"]	4.3	Very Good	344	86	0.0	1	-1
211936	3202169	Red Dot Nation	["Casual Dining"]	https://www.zomato.com/vadodara/red-dot-nation... Vinayak Heights, Beside Bharat Petrol Pump, Wag...	Vadodara	32	Suryanagar	22.281816	73.232252	...	2	Rs.	["Cash", "Delivery", "Credit Card", "Dinner"]	3.6	Good	381	19	0.0	-1	-1
211937	18855810	Biryani aur Baatish	["Casual Dining"]	https://www.zomato.com/vadodara/biryani-aur-ba... Shop 14, Atlantic K... 10, A Wing, Genda Circle R...	Vadodara	32	Alikapuri	22.317746	73.168043	...	2	Rs.	["Dinner", "Cash", "Takeaway Available", "Debt"]	4.1	Very Good	154	96	0.0	-1	-1
211938	18662583	Wok On Fire	["Casual Dining"]	https://www.zomato.com/vadodara/wok-on-fire-la... Ground Floor, 1... Residence Building, Opposite Se...	Vadodara	32	Fatehgarj	22.323357	73.187461	...	3	Rs.	["Dinner", "Cash", "Debt Card", "Lunch", "Tak..."]	4.0	Very Good	301	126	0.0	1	-1
211939	3202251	Kali Mirch Cafe And Restaurant	["Casual Dining"]	https://www.zomato.com/vadodara/kali-mirch-caf... Manu Smriti Complex, Near Navaratri School, Gl... Mahalaxmi	Vadodara	32	Fatehgarj	22.336931	73.192356	...	2	Rs.	["Dinner", "Cash", "Lunch", "Takeaway", "Indo..."]	4.1	Very Good	243	40	0.0	-1	-1

In order to import the dataset. and then checking last rows by using pandas libraries.

## Viewing Columns -

```
list(df.columns)
```

```
['res_id',
 'name',
 'establishment',
 'url',
 'address',
 'city',
 'city_id',
 'locality',
 'latitude',
 'longitude',
 'zipcode',
 'country_id',
 'locality_verbose',
 'cuisines',
 'timings',
 'average_cost_for_two',
 'price_range',
 'currency',
 'highlights',
 'aggregate_rating',
 'rating_text',
 'votes',
 'photo_count',
 'opentable_support',
 'delivery',
 'takeaway']
```

I've load dataset named 'Indian\_restaurants.csv' into a variable named 'df'. In order to import the dataset, I've used Pandas Libraries in which I've used 'pd.read\_csv' command to import the respective dataset and then using pandas libraries for viewing only columns

## Viewing shape of data-

```
f"ROWS :{df.shape[0]} COLUMNS : {df.shape[1]}"  
... 'ROWS :211944 COLUMNS : 26'
```

After loading dataset viewing the shape of data so in data set as we can see here total rows of data 211944 and columns 26.

## Checking informations-

```
df.info()  
  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 211944 entries, 0 to 211943  
Data columns (total 26 columns):  
 #   Column           Non-Null Count  Dtype     
 ---  --  
 0   res_id            211944 non-null  int64    
 1   name              211944 non-null  object    
 2   establishment     211944 non-null  object    
 3   url               211944 non-null  object    
 4   address            211810 non-null  object    
 5   city               211944 non-null  object    
 6   city_id            211944 non-null  int64    
 7   locality           211944 non-null  object    
 8   latitude            211944 non-null  float64  
 9   longitude           211944 non-null  float64  
 10  zipcode             48757 non-null  object    
 11  country_id         211944 non-null  int64    
 12  locality_verbose    211944 non-null  object    
 13  cuisines            210553 non-null  object    
 14  timings              208070 non-null  object    
 15  average_cost_for_two 211944 non-null  int64    
 16  price_range          211944 non-null  int64    
 17  currency             211944 non-null  object    
 18  highlights            211944 non-null  object    
 19  aggregate_rating      211944 non-null  float64  
 20  rating_text           211944 non-null  object    
 21  votes                211944 non-null  int64    
 22  photo_count           211944 non-null  int64    
 23  opentable_support      211896 non-null  float64  
 24  delivery              211944 non-null  int64    
 25  takeaway              211944 non-null  int64  
dtypes: float64(4), int64(9), object(13)  
memory usage: 42.0+ MB
```

I've obtained information about the dataset using df.info(). This method provides a concise summary, including the total number of entries, the data types of each column, and the count of non-null values. and also here we can check dtypes:  
float64(4),int(64),  
object(13) memory usage :  
42.0+mb

## Viewing numerical data of data set-

df.describe()														
	res_id	city_id	latitude	longitude	country_id	average_cost_for_two	price_range	aggregate_rating	votes	photo_count	opentable_support	delivery	takeaway	
count	2.119440e+05	211944.000000	211944.000000	211944.000000	211944.0	211944.000000	211944.000000	211944.000000	211944.000000	211944.000000	211944.000000	211944.0	211944.0	
mean	1.349411e+07	4746.785434	21.499758	77.615276	1.0	595.812229	1.882535	3.395937	378.001864	256.971224	0.0	-0.255907	-1.0	
std	7.883722e+06	5568.766386	22.781331	7.500104	0.0	606.239363	0.892989	1.283642	925.333370	867.668940	0.0	0.964172	0.0	
min	5.000000e+01	1.000000	0.000000	0.000000	1.0	0.000000	1.000000	0.000000	-18.000000	0.000000	0.0	-1.000000	-1.0	
25%	3.301027e+06	11.000000	15.496071	74.877961	1.0	250.000000	1.000000	3.300000	16.000000	3.000000	0.0	-1.000000	-1.0	
50%	1.809573e+07	34.000000	22.514494	77.425971	1.0	400.000000	2.000000	3.800000	100.000000	18.000000	0.0	-1.000000	-1.0	
75%	1.881297e+07	11306.000000	26.841667	80.219323	1.0	700.000000	2.000000	4.100000	362.000000	128.000000	0.0	1.000000	-1.0	
max	1.915979e+07	11354.000000	10000.000000	91.832769	1.0	30000.000000	4.000000	4.900000	42539.000000	17702.000000	0.0	1.000000	-1.0	

Now here we can check numerical data of data set like :-  
count,mean,std,min,25%,50%,75%,max by using pandas libraries  
df.describe().

df.isnull().sum()	
...	0
res_id	0
name	0
establishment	0
url	0
address	134
city	0
city_id	0
locality	0
latitude	0
longitude	0
zipcode	163187
country_id	0
locality_verbose	0
cuisines	1391
timings	3874
average_cost_for_two	0
price_range	0
currency	0
highlights	0
aggregate_rating	0
rating_text	0
votes	0
photo_count	0
opentable_support	48
delivery	0
takeaway	0
dtype: int64	

df.isnull().sum()/df.shape[0]*100	
...	0
res_id	0.000000
name	0.000000
establishment	0.000000
url	0.000000
address	0.063224
city	0.000000
city_id	0.000000
locality	0.000000
latitude	0.000000
longitude	0.000000
zipcode	76.995338
country_id	0.000000
locality_verbose	0.000000
cuisines	0.656305
timings	1.827841
average_cost_for_two	0.000000
price_range	0.000000
currency	0.000000
highlights	0.000000
aggregate_rating	0.000000
rating_text	0.000000
votes	0.000000
photo_count	0.000000
opentable_support	0.022647
delivery	0.000000
takeaway	0.000000
dtype: float64	

Now here we will do very valuable work like EDA we will check null values in numerical and percentage and we will remove duplicates easily and fill null values and remove unnecessary null columns also.

## Checking & Removing null values:-

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	zipcode	country_id	locality_verbose	cuisines	timings	average_cost_for_two	price_range	currency	highlight
0	3400299	Bikanervala	[Quick Bites]	https://www.zomato.com/agra/bikanervala-khanda...	Kalyani Point, Near Tuli Cinema, Bypass Road,...	Agra	34	Khandari	27.211450	78.002381	Nan	1	Khandari, Agra	North Indian, South Indian, Mithai, Street Foo...	8:30am - 10:30pm (Mon-Sun)	700	2	Rs.	[Lunch, Takeaway Available, 'Credit Card',...
4	3401744	The Salt Cafe Kitchen & Bar	[Casual Dining]	https://www.zomato.com/agra/the-salt-cafe-kit...	1C 3rd Floor, Fatehabad Road, Tagari, Agra	Agra	34	Tajganj	27.157709	78.052421	Nan	1	Tajganj, Agra	North Indian, Continental, Italian	11:30 AM to 11:30 PM	1000	3	Rs.	[Lunch, Services Available, 'Cash On Delivery', 'Credit Card',...
5	3400275	Domino's Pizza	[Quick Bites]	https://www.zomato.com/agra/domino-s-pizza-cm...	114/23 G, Deep Shikha Complex, Sangay Place, G...	Agra	34	Civil Lines	27.201516	78.007556	Nan	1	Civil Lines, Agra	Pizza, Fast Food	10:57 AM to 11 PM	400	2	Rs.	[Credit Card, 'Delivery', 'Lunch Delivery', 'Dinner',...
7	3400368	Domino's Pizza	[Quick Bites]	https://www.zomato.com/agra/domino-s-pizza-sika...	Plot C-16, Sector 13, Sikandra, Agra	Agra	34	Sikandra	27.203930	77.954260	Nan	1	Sikandra, Agra	Pizza, Fast Food	10:57 AM to 11 PM	400	2	Rs.	[Lunch Delivery, 'Credit Card', 'No Alcohol',...
8	3401284	Cake House	[Bakery]	https://www.zomato.com/agra/cake-house-2-civil...	22/01, Wazarpura Rd, Judge Compound Chowraha...	Agra	34	Civil Lines	27.204148	78.009025	Nan	1	Civil Lines, Agra	Bakery, Fast Food	9 AM to 9 PM	500	2	Rs.	[Takeaway Available, 'Cash On Delivery', 'Indoor Seating',...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
211937	18855810	Biryani aur Baattein	[Casual Dining]	https://www.zomato.com/vaddodara/biryani-aur-ba...	Shop 14, Atlantic K-10, A Wing, Genda	Vadodara	32	Alikapuri	22.317746	73.168943	Nan	1	Alikapuri, Vadodara	Biryani, Mughlai, Chinese	11:30am - 3:30pm, 7:30pm - 1am	500	2	Rs.	[Dinner, Cash On Delivery, Takeaway Available]

Checking null column by using pandas libraries and viewing the total null value for what is impacting on dataset and zipcode column have most null rows so it's important to remove the column.

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	country_id	locality_verbose	cuisines	timings	average_cost_for_two	price_range	currency	highlights	aggregate_rating	rating_text
0	3400299	Bikanervala	[Quick Bites]	https://www.zomato.com/agra/bikanervala-khanda...	Kalyani Point, Near Tuli Cinema, Bypass Road,...	Agra	34	Khandari	27.211450	78.002381	1	Khandari, Agra	North Indian, South Indian, Mithai, Street Foo...	8:30am - 10:30pm (Mon-Sun)	700	2	Rs.	[Lunch, Takeaway Available, 'Credit Card',...	4.4	Very Good
1	3400065	Mama Chicken Mama Franky House	[Quick Bites]	https://www.zomato.com/agra/mama-chicken-mama-f...	Main Market, Sada Bazaar, Agra Cantt, Agra	Agra	34	Agra Cantt	27.160569	78.011583	1	Agra Cantt, Agra	North Indian, Mughlai, Rolls, Chinese, Fast Fo...	12:30PM to 12AM (Mon, Wed, Thu, Fri, Sat,...	600	2	Rs.	[Delivery, 'No Alcohol Available', 'Dinner',...	4.4	Very Good
2	3401013	Bhagat Hawali	[Quick Bites]	https://www.zomato.com/agra/bhagat-hawali-2-th...	62/1, Near Easy Day, West Side, Agra	Agra	34	Shahganj	27.182938	77.979684	1	Shahganj, Agra	Fast Food, Mithai	9:30 AM to 11 PM	300	1	Rs.	[No Alcohol Available, 'Dinner', 'Takeaway A...	4.2	Very Good
3	3400290	Bhagat Hawali	[Quick Bites]	https://www.zomato.com/agra/bhagat-hawali-civil...	Near Aranya Cinema, Nehru Nagar, Civil Lines, A...	Agra	34	Civil Lines	27.205668	78.004799	1	Civil Lines, Agra	Desserts, Bakery, Fast Food, South Indian	8am - 11pm (Mon-Sun)	300	1	Rs.	[Takeaway Available, 'Credit Card', 'Lunch',...	4.3	Very Good
4	3401744	The Salt Cafe Kitchen & Bar	[Casual Dining]	https://www.zomato.com/agra/the-salt-cafe-kit...	1C 3rd Floor, Fatehabad Road, Tagari, Agra	Agra	34	Tajganj	27.157709	78.052421	1	Tajganj, Agra	North Indian, Continental, Italian	11:30 AM to 11:30 PM	1000	3	Rs.	[Lunch, Services Available, 'Cash On Delivery', 'Credit Card',...	4.9	Excellent

Checking null column by using pandas libraries and viewing the total null value for what is impacting on dataset and zipcode column have most null rows so it's important to remove the column permanently.

## Checking null values & filling them:-

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	country_id	locality_verbose	cuisines	timings	average_cost_for_two	price_range	currency	highlights	aggregate_rating
126	3400957	Anise	[Fine Dining]	https://www.zomato.com/agra/anise-tajganj?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=anise&utm_content=listing	Courtyard by Marriott Agra, Taj Nagar, Phase ...	Agra	34	Tajganj	27.156386	78.059525	1	Tajganj, Agra	North Indian	Nan	2000	4	Rs.	["Cash", "Debit Card", "Dinner", "Credit Card", "Takeaway Available"]	3.8
214	19032611	Om Sai Lassi And Juice Corner	[Beverage Shop]	https://www.zomato.com/agra/om-sai-lassi-and-juice-corner?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=om-sai-lassi-and-juice-corner&utm_content=listing	Boda Chaurasi, Shahganj, Agra	Agra	34	Shahganj	27.188775	77.962173	1	Shahganj, Agra	Juices	Nan	100	1	Rs.	["Cash", "Takeaway Available", "Indoor Seating"]	3.9
220	19050396	Agarwal Lassi	[Shack]	https://www.zomato.com/agra/agarwal-lassi-civil-lines?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=agarwal-lassi&utm_content=listing	22/156 chandi ghat kothi near pallival park...	Agra	34	Civil Lines	27.199784	78.017700	1	Civil Lines, Agra	Street Food	Nan	100	1	Rs.	["Cash", "Takeaway Available"]	3.6
297	19031666	chai ki tari	[Cafe]	https://www.zomato.com/agra/chai-ki-tari-sikandra?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=chai-ki-tari&utm_content=listing	shop no 2 hno 216 sec 6a vasav colony st...	Agra	34	Sikandra	27.199757	77.958770	1	Sikandra, Agra	Cafe	Nan	100	1	Rs.	["Outdoor Seating"]	0.0
383	3400322	Filmy Zayka - Hotel Alleviate	[Casual Dining]	https://www.zomato.com/agra/filmy-zayka-hotel-alleviate?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=filmy-zayka-hotel-alleviate&utm_content=listing	627, Yamuna Krupa, Belanagar, Civil Lin...	Agra	34	Civil Lines	27.189854	78.024778	1	Civil Lines, Agra	Chinese, North Indian	Nan	700	2	Rs.	["Takeaway Available", "Lunch", "Cash", "Credit Card"]	4.1
207778	19065540	Bake N Break Bakery	[Bakery]	https://www.zomato.com/vizakapatnam/bake-n-br...?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=bake-n-br&utm_content=listing	Door 21-2-1314, Sector 9, MVP Colony, Visakha...	Vizag	28	MVP Colony	17.716327	83.311777	1	MVP Colony, Vizag	Bakery	Nan	300	1	Rs.	["Cash", "Indoor Seating", "Desserts and Bakes"]	0.0
208022	19065540	Bake N Break Bakery	[Bakery]	https://www.zomato.com/vizakapatnam/bake-n-br...?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=bake-n-br&utm_content=listing	Door 21-2-1314, Sector 9, MVP Colony, Visakha...	Vizag	28	MVP Colony	17.716327	83.311777	1	MVP Colony, Vizag	Bakery	Nan	300	1	Rs.	["Cash", "Indoor Seating", "Desserts and Bakes"]	0.0
208465	19065540	Bake N Break Bakery	[Bakery]	https://www.zomato.com/vizakapatnam/bake-n-br...?utm_source=zomato&utm_medium=referral&utm_campaign=listing&utm_term=bake-n-br&utm_content=listing	Door 21-2-1314, Sector 9, MVP Colony, Visakha...	Vizag	28	MVP Colony	17.716327	83.311777	1	MVP Colony, Vizag	Bakery	Nan	300	1	Rs.	["Cash", "Indoor Seating", "Desserts and Bakes"]	0.0

Checking null column by using pandas libraries and viewing the total null value for what is impacting on dataset and timing column have null rows so it's important to fill the column by timing mode permanently.

```
df["timings"].mode()
timings
0 11 AM to 11 PM
dtype: object

df["timings"]=df["timings"].fillna(df["timings"].mode()[0])

df["cuisines"].mode()
...
cuisines
0 North Indian
dtype: object
```

Checking null column by using pandas libraries and viewing the total null value for what is impacting on dataset and cuisine column have null rows so it's important to fill the column by cuisine mode permanently.

```
df["cuisines"]=df["cuisines"].fillna(df["cuisines"].mode()[0])
```

## Filling null & drop duplicate values-

	df.isnull().sum()
...	0
res_id	0
name	0
establishment	0
url	0
address	134
city	0
city_id	0
locality	0
latitude	0
longitude	0
country_id	0
locality_verbose	0
cuisines	0
timings	0
average_cost_for_two	0
price_range	0
currency	0
highlights	0
aggregate_rating	0
rating_text	0
votes	0
photo_count	0
opentable_support	48
delivery	0
takeaway	0
dtype: int64	

There is only address & Opentable support column which have null values we will fill them.

```
df['address'].fillna('Unknown', inplace=True)
```

```
df['address'].isnull().sum()
np.int64(0)

df["opentable_support"].isnull().sum()
np.int64(48)

df["opentable_support"].median()
0.0

df["opentable_support"]=df["opentable_support"].fillna(df["opentable_support"].median())
```

There is only address & Opentable support column which was null and after we fill them by using fillna which is a part of code in pandas

```
df.duplicated().sum()
np.int64(151533)

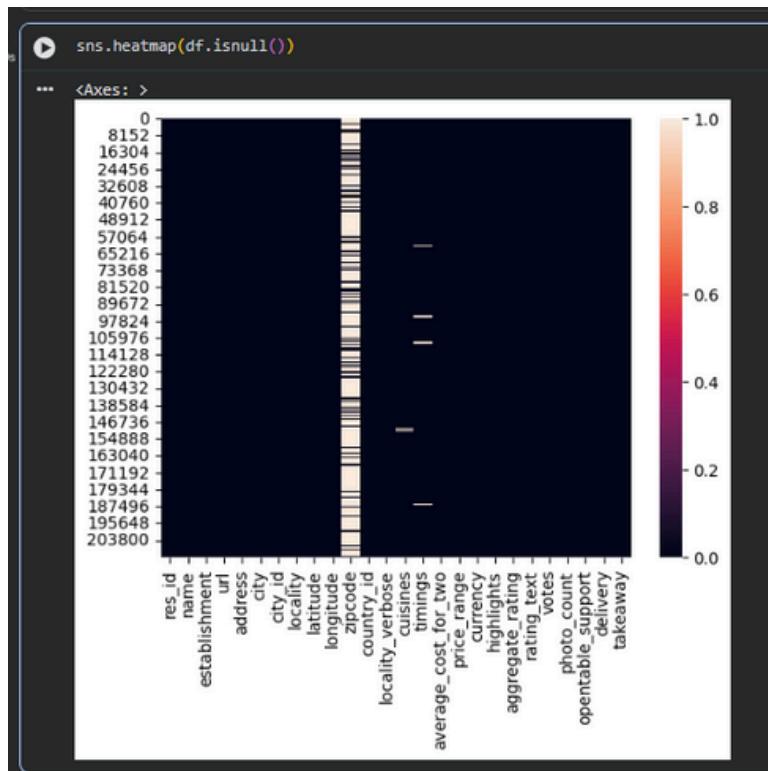
df.drop_duplicates(inplace=True)
```

In the dataset there's most of duplicates rows available which is impacting dataset .so it's important to drop duplicate rows

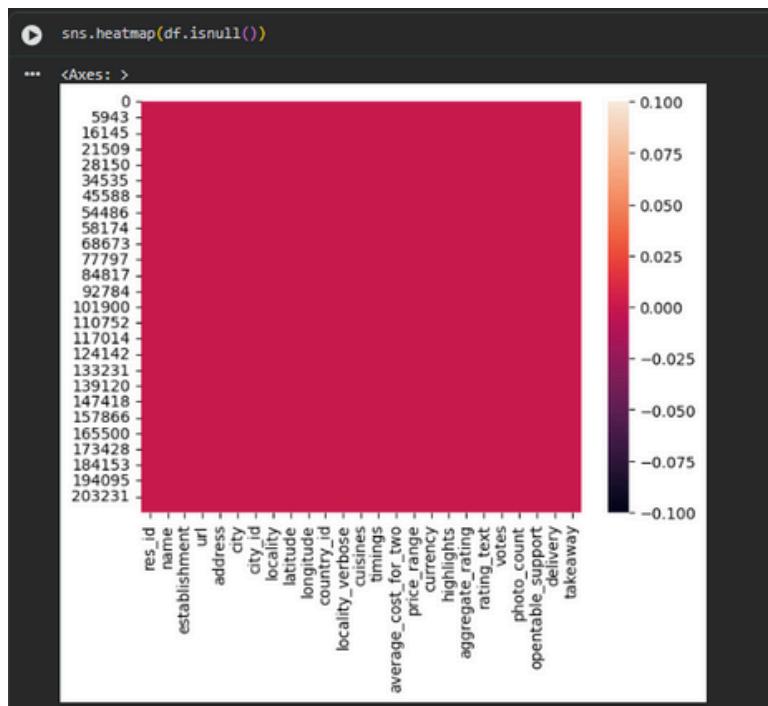
# Data visualization

ZOMATO

## Before cleaning data in heatmap-



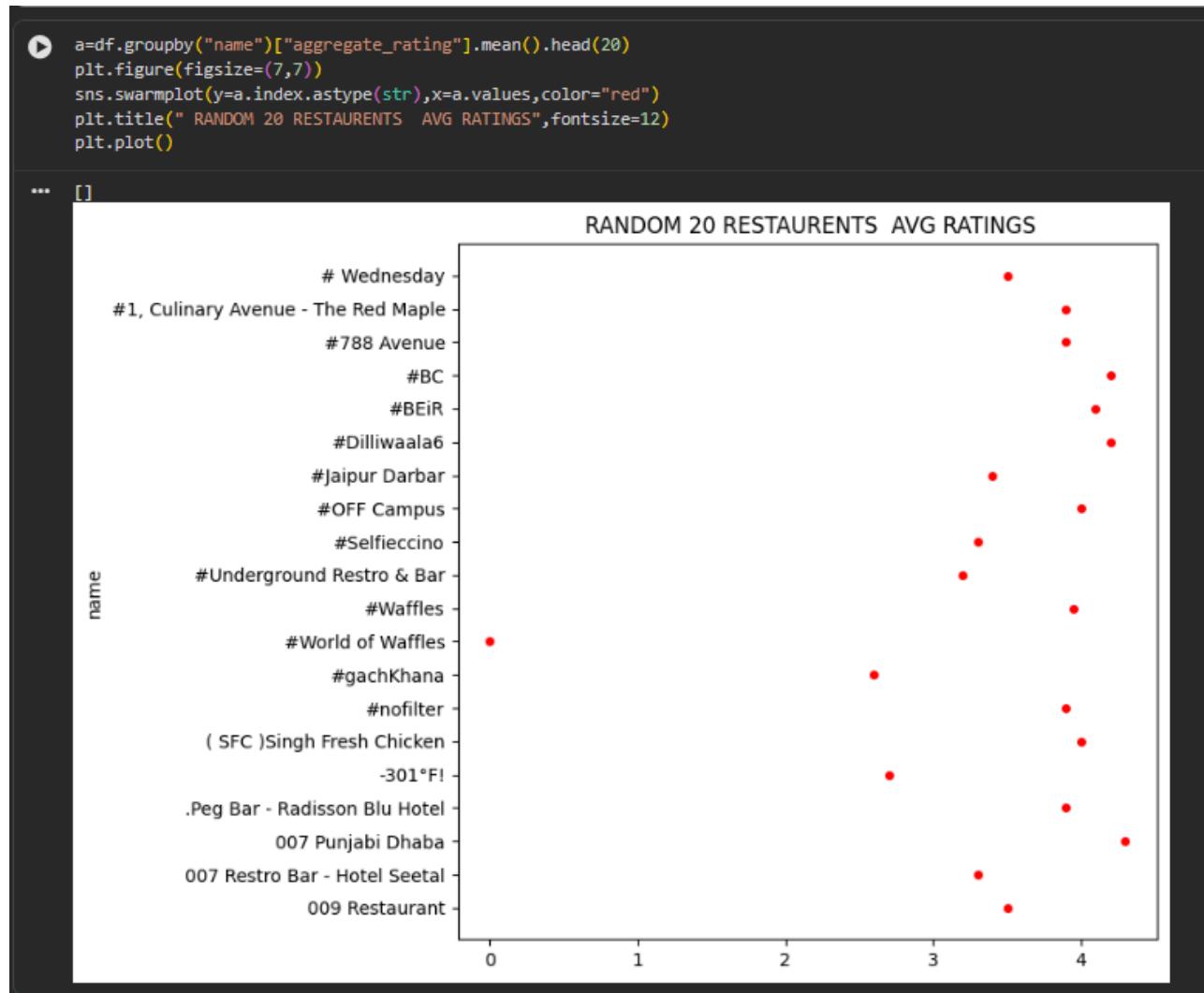
## After cleaning data in heatmap-



This is data look like before cleaning it the image is fully black and zipcode,timing,cuisine etc, multiple of data as image shows null values.

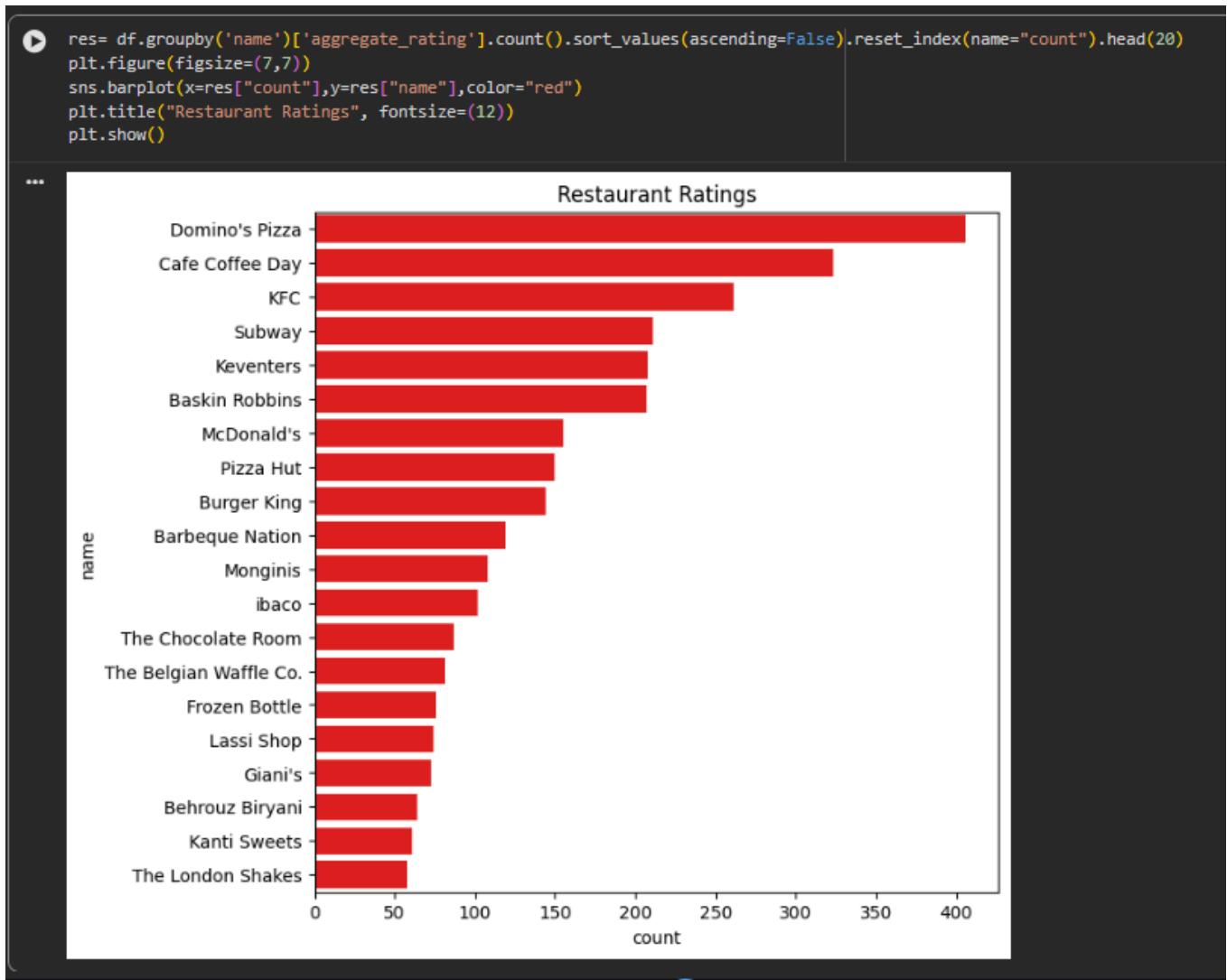
After cleaning and filled all null values and removed duplicate value the data looks like hot pink and there's no spaces in this heatmap which shows the cleaned data only.

## Avg rating of Restaurants:-



I have visualized average ratings of restaurants by using Seaborn's 'swarmplot'. The x-axis represents the average rating of restaurants, and the y-axis represents the names of the restaurants. As we can clearly see the distribution of average ratings of restaurants.

## Most rated Restaurants:-



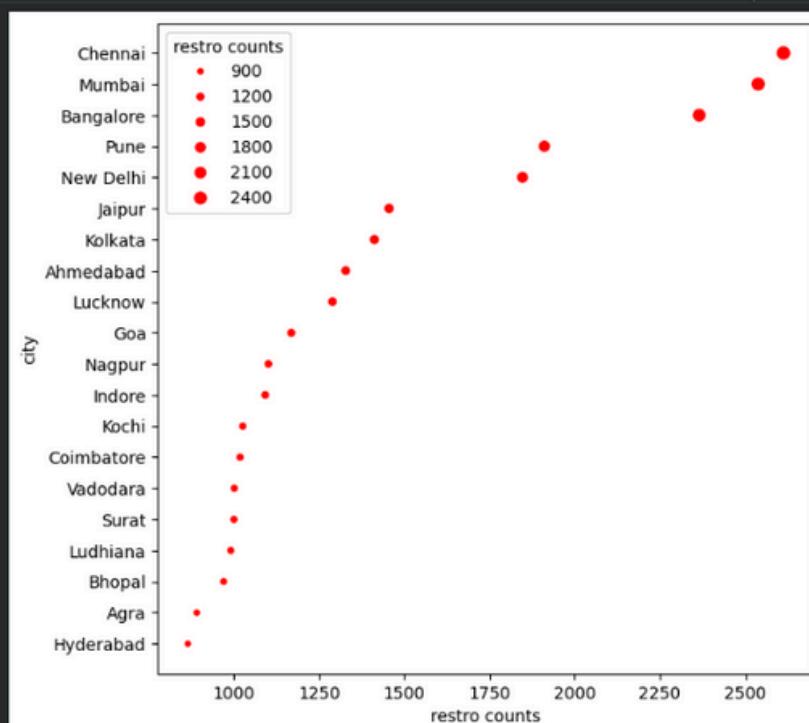
I have visualized most rated top 20 restaurants by using a Seaborn barplot ('sns.barplot'). The x-axis represents the count of rating of restaurants, and the y-axis represents the names of restaurants. As we can clearly see the distribution of most rated restaurants.

## Max restaurants count in cities:-

```
top_city=df.groupby("city")["name"].count().reset_index(name="restro counts").sort_values("restro counts", ascending=False).head(20)

plt.figure(figsize=(7,7))
sns.scatterplot(y=top_city["city"],x=top_city["restro counts"],size=top_city["restro counts"],color="red")

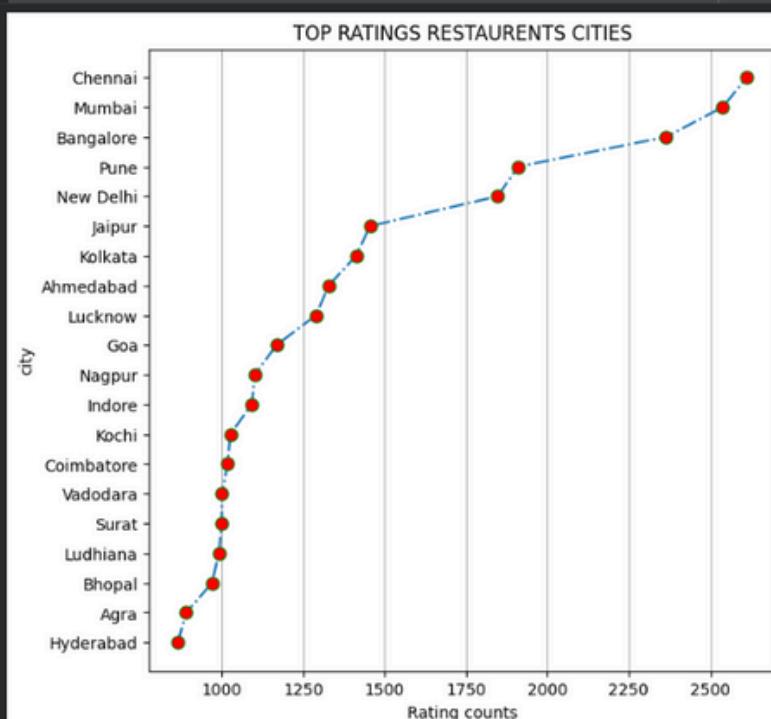
plt.show()
```



I've visualized the top 20 cities with the most number of Restaurants outlets using a seaborn scatterplot (sns.scatterplot). The x-axis represents the number of outlets, and the y-axis displays the respective cities name. This visualization shows a clear comparison of the outlet counts for different cities, providing insights into the distribution and popularity of these establishments based on the available data.

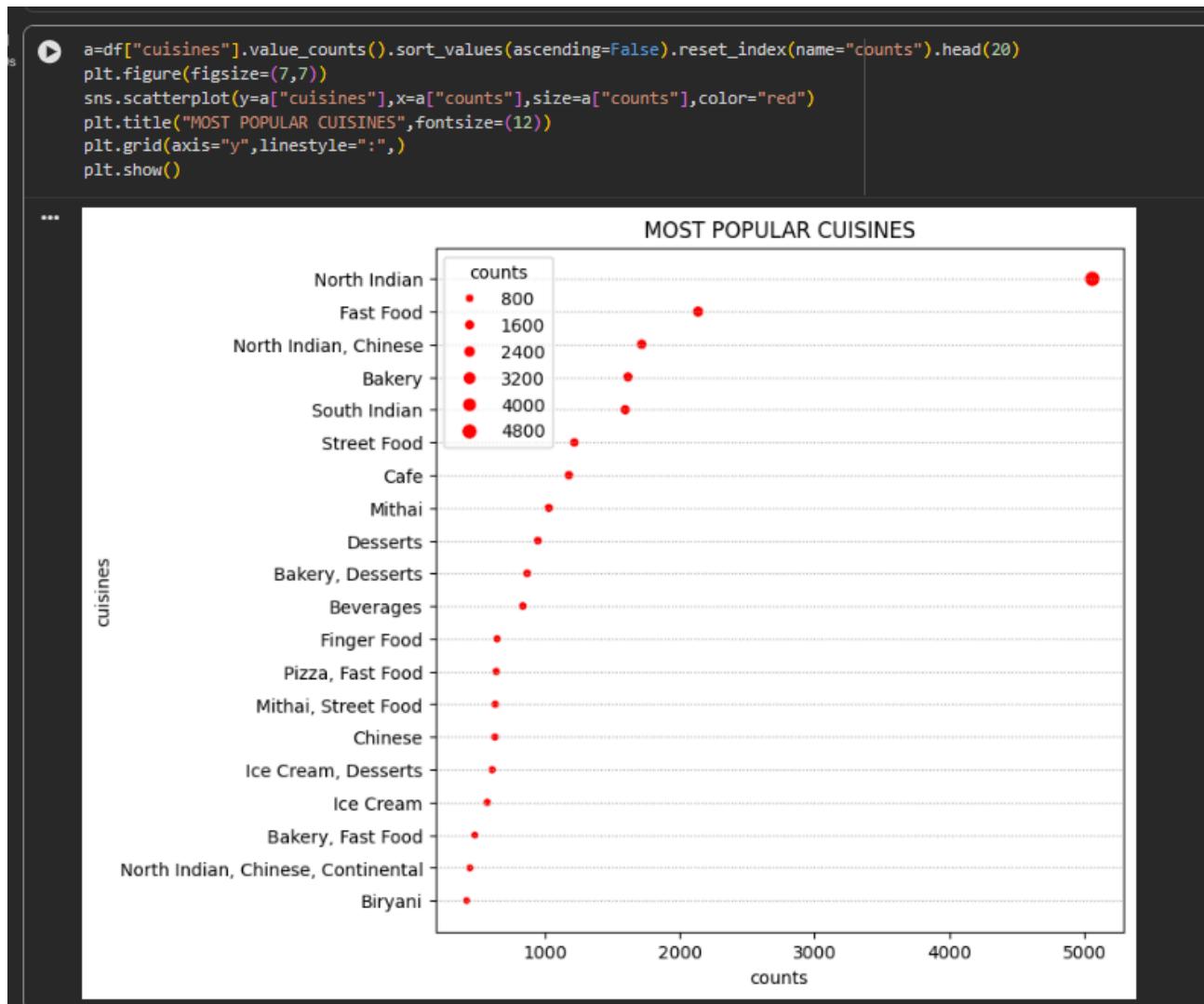
## Most rated restaurants count in cities:-

```
a=df.groupby(["city"])["aggregate_rating"].count().reset_index(name="Rating counts").sort_values("Rating counts",ascending=False).head(20)
plt.figure(figsize=(7,7))
sns.lineplot(x=a["Rating counts"],y=a["city"],marker="o",ms=8,mfc="r",mec="green",linestyle="dashdot")
plt.title("TOP RATINGS RESTAURENTS CITIES", fontsize=(12))
plt.grid(axis="x")
plt.show()
```



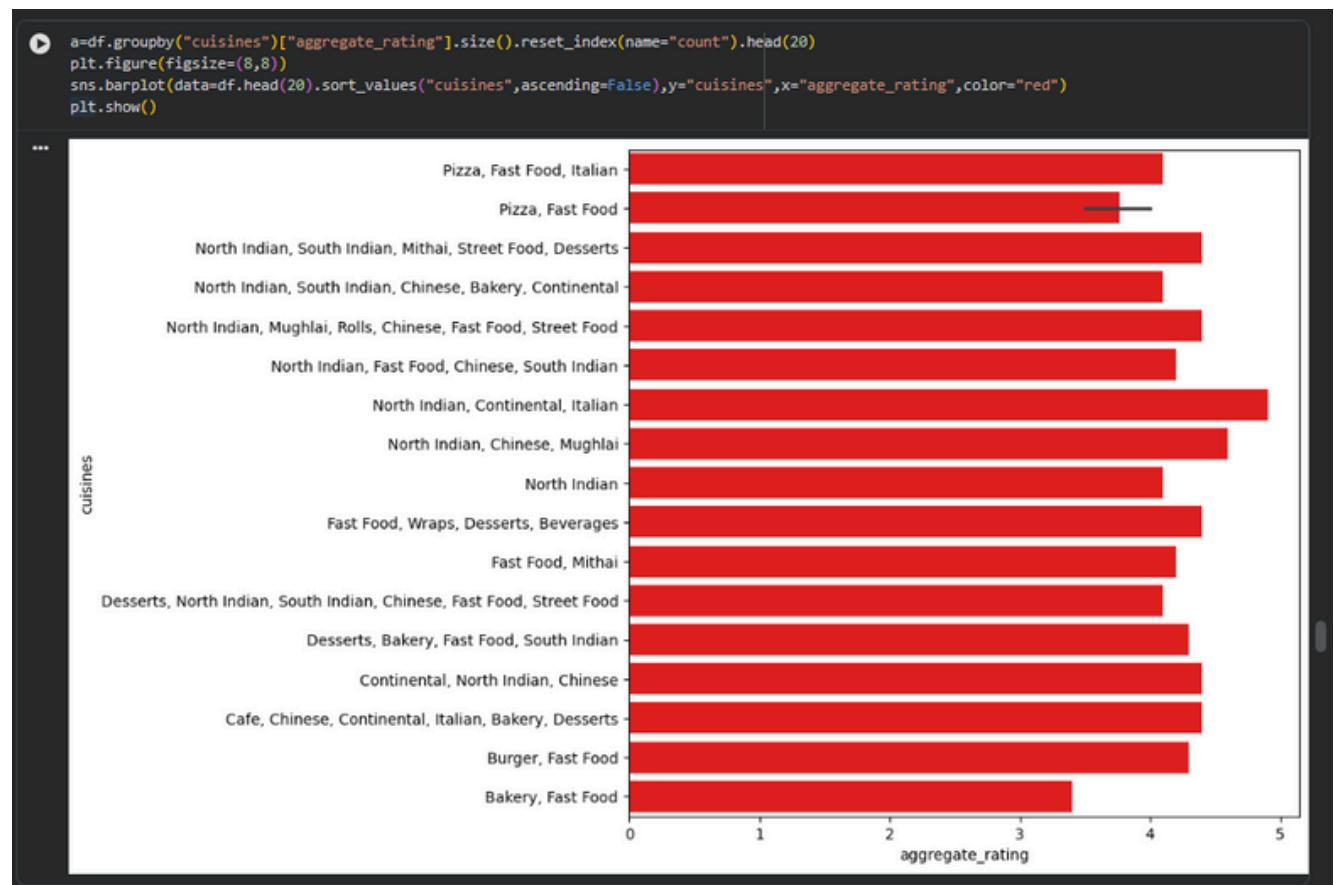
I've visualized the top 20 cities with the highest number of Restaurants rating using a seaborn lineplot (sns.lineplot). The x-axis represents the counts of ratings, and the y-axis displays the respective cities names. This visualization shows a clear comparison of the most rated cities of restaurants providing insights into the distribution and popularity of these establishments based on the available data.

## Most popular cuisines :-



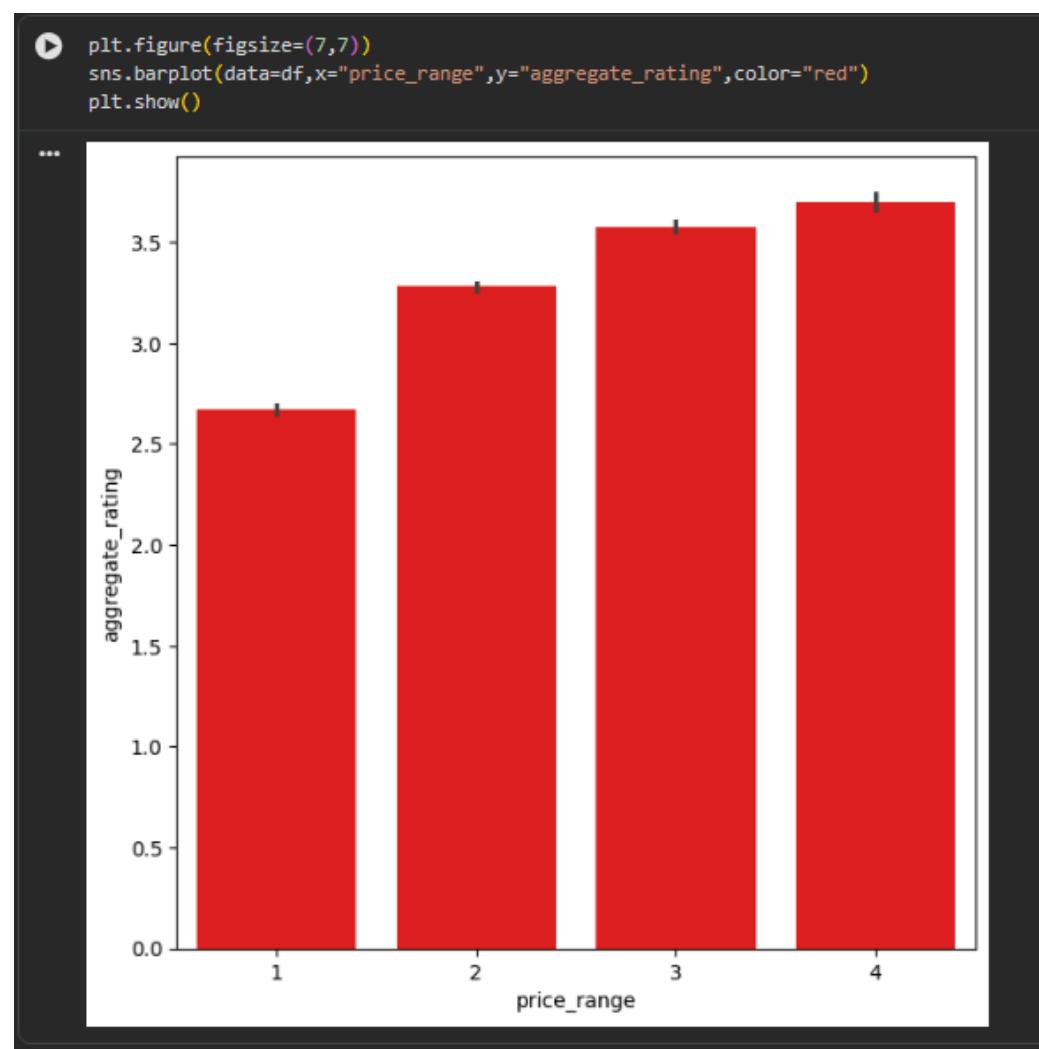
I've visualized the top 20 cuisines .which is very in demand and they are most popular , I used seaborn scatter (sns.scatterplot). The x-axis represents the counts of cuisines, and the y-axis displays the respective cuisines names. as we can see here north indian food and fast food is on top and biryani and ice cream comes at last as per on available data

## Correlation between cuisines and restaurant ratings.



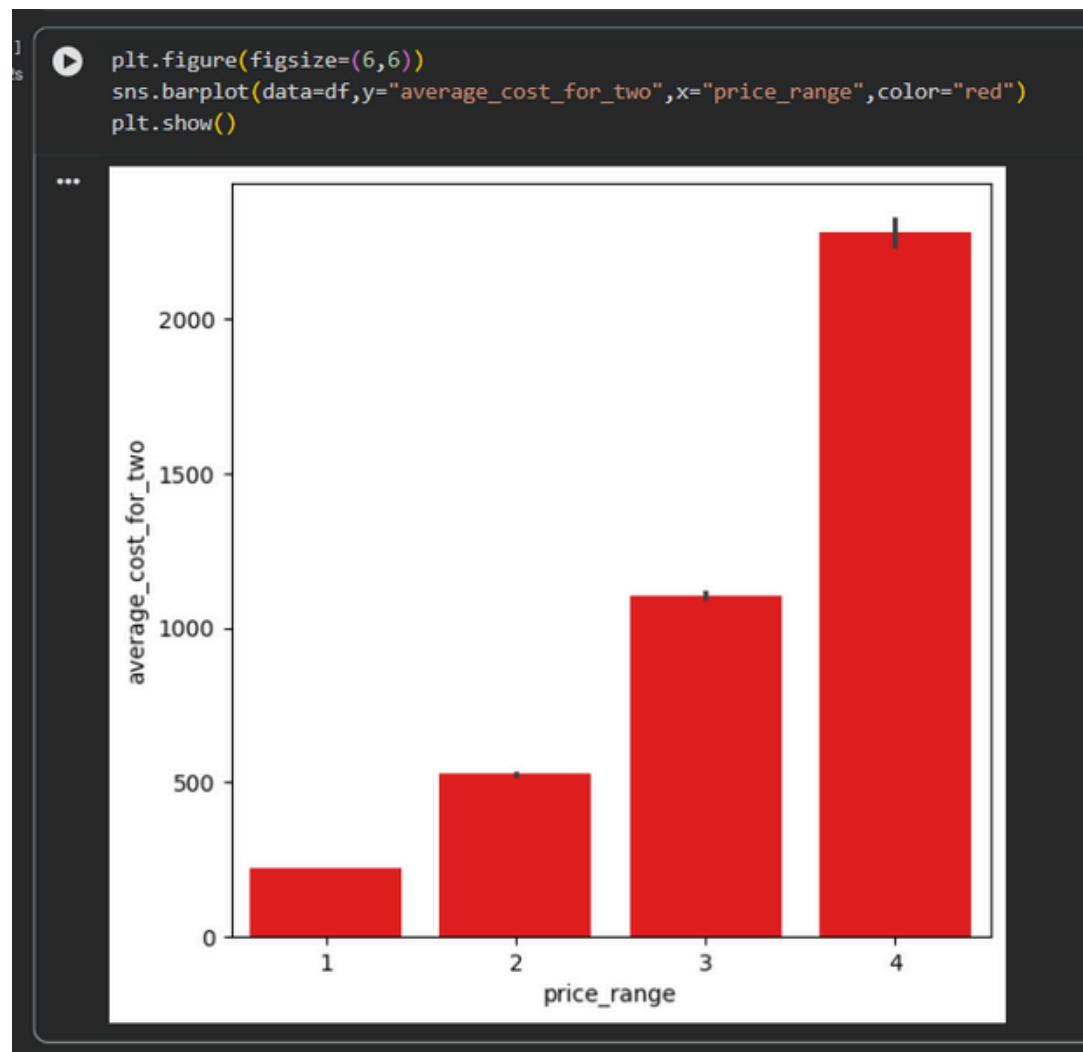
I've visualized the top 20 cuisines .with their ratings so here we can see the correlation between cuisines and their rating the most rated cuisine is noth indian,chinese,mughlai and the lowest of rating is bakery and fast food here as we can see and analysis the most demanded and rated cuisines. here we use the seaborn barplot (sns.barplot).

## Analysis between price range & rating:



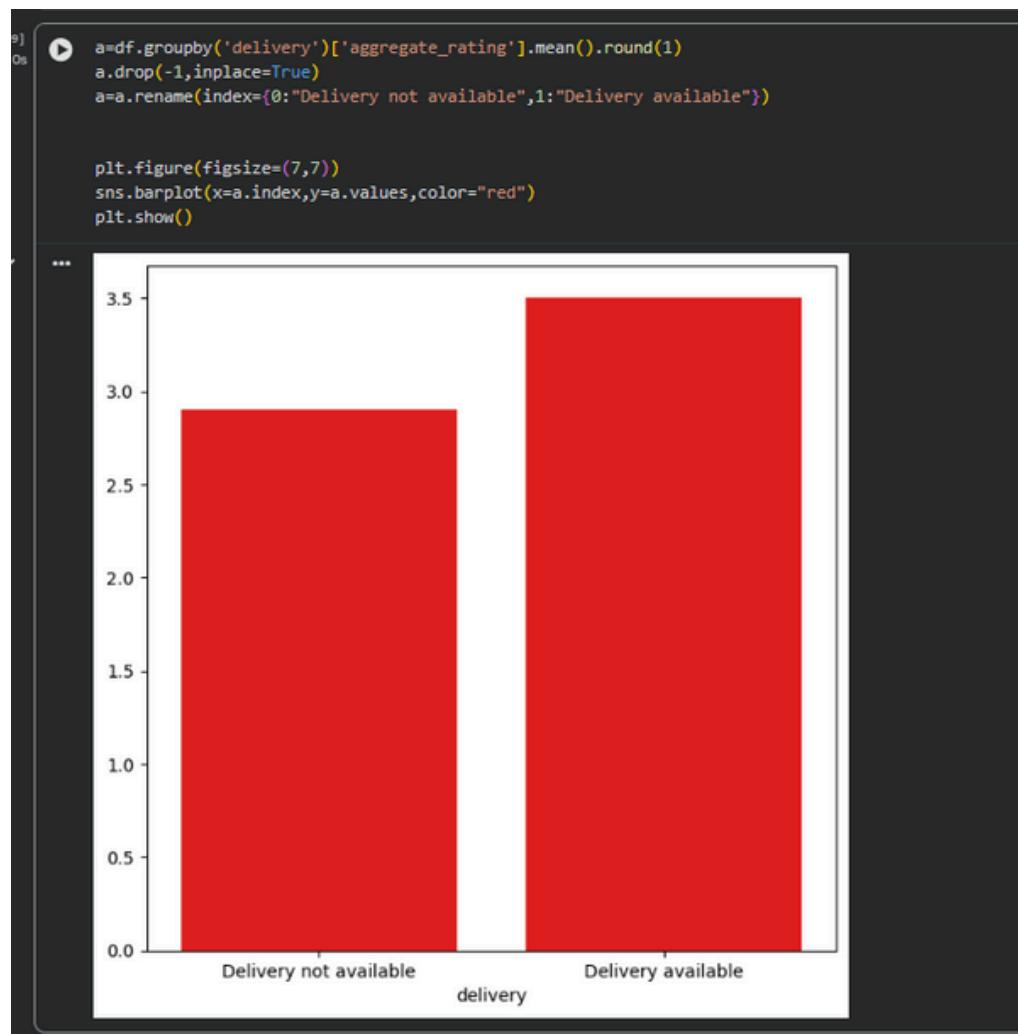
I have visualized price range and rating as we can analyse the rating with price range as we can also see here highest price range going its's peek level of rating so we can understand and lowest price range is lowest of rating so we use seaborn barplot (sns.barplot).

## Analysis between price range & cost for two people:



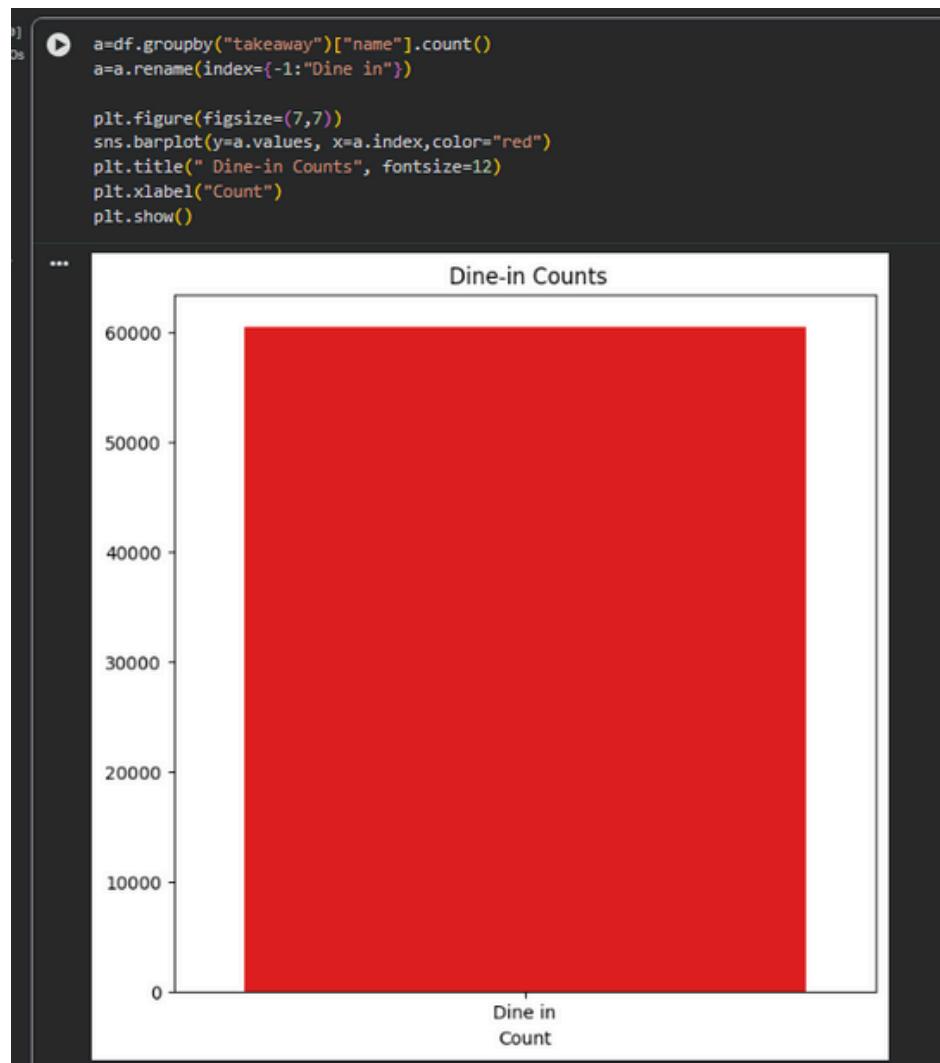
I have visualized price range and avg cost for two here we can analyse the avg cost of two people & price range so we use here boxplot from seaborn library and we use here (sns.boxplot) so here we can see in different price range have different avg cost of two people.

## Analysis online orders impact in rating:



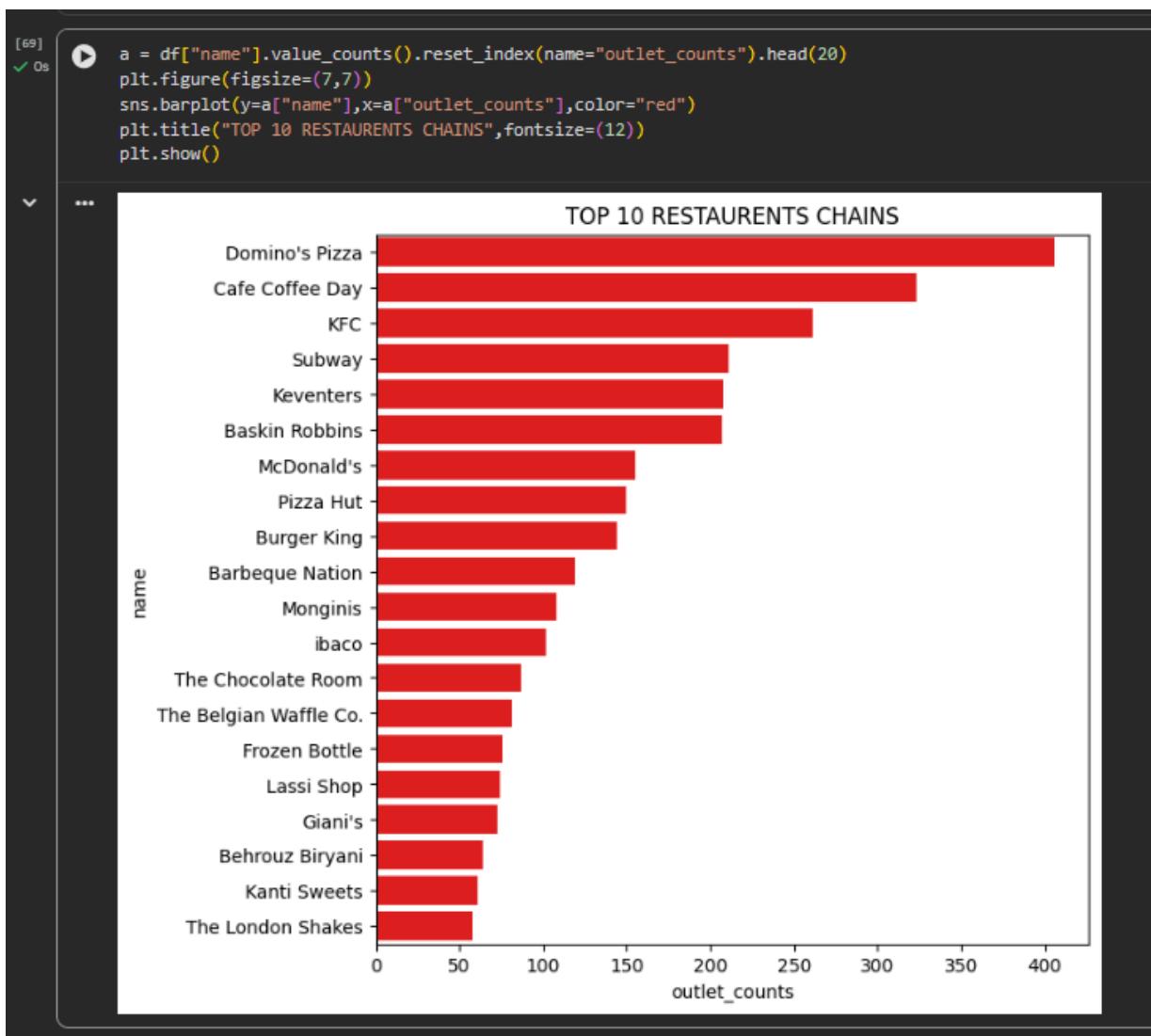
I have visualized the bar graph here we can analyse between what is impact of online order and rating so here we can understand easily the most ratings given by online orders here we use seaborn bar plot(sns.barplot).

## Analysis Dine in:



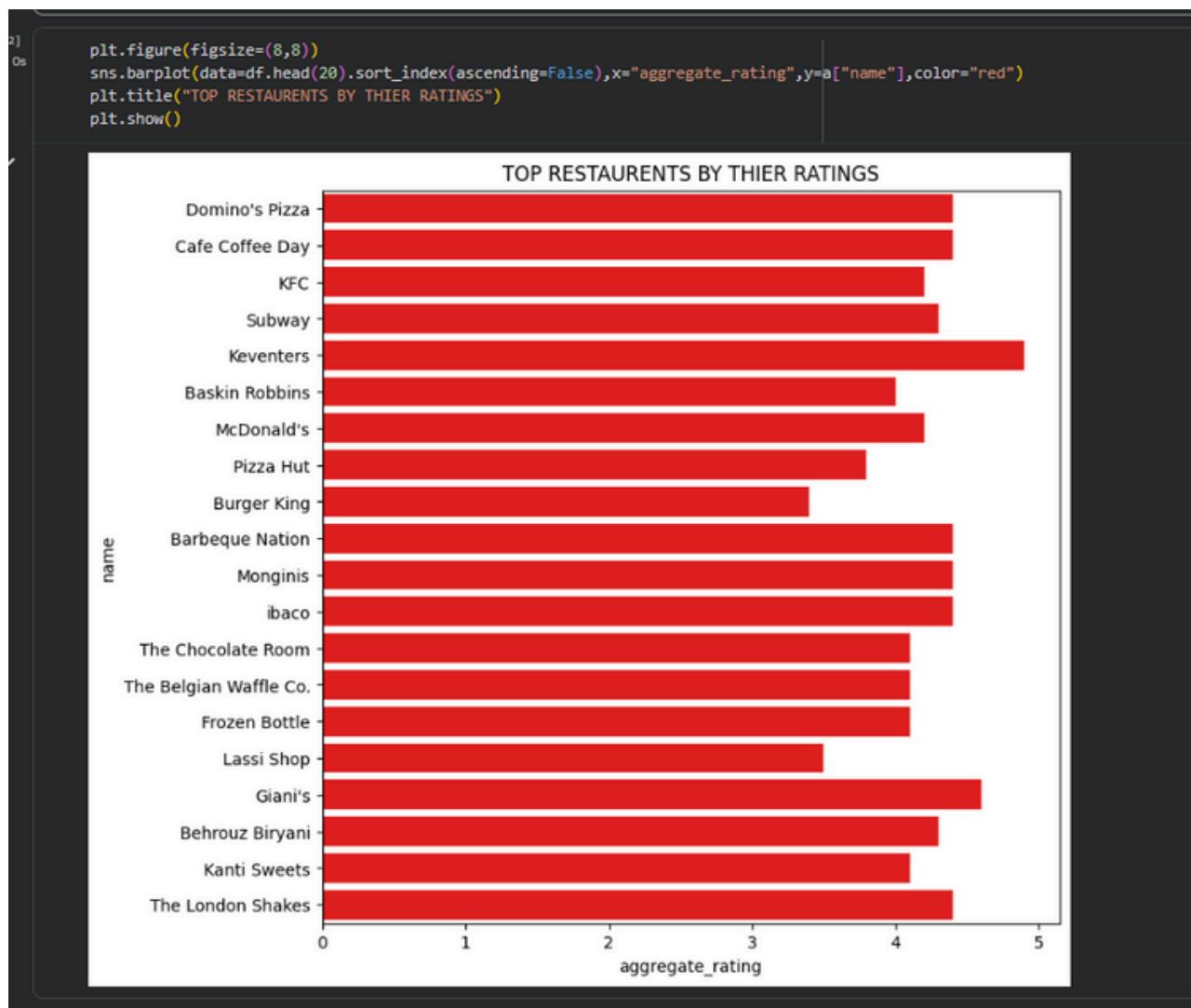
I have visualized the bar graph here we can analyse & here we use seaborn bar graph and here we see the count of dine in count as per available data here 6k booking done so as we use here seaborn bar plot(sns.barplot).

## Top restaurant chains:



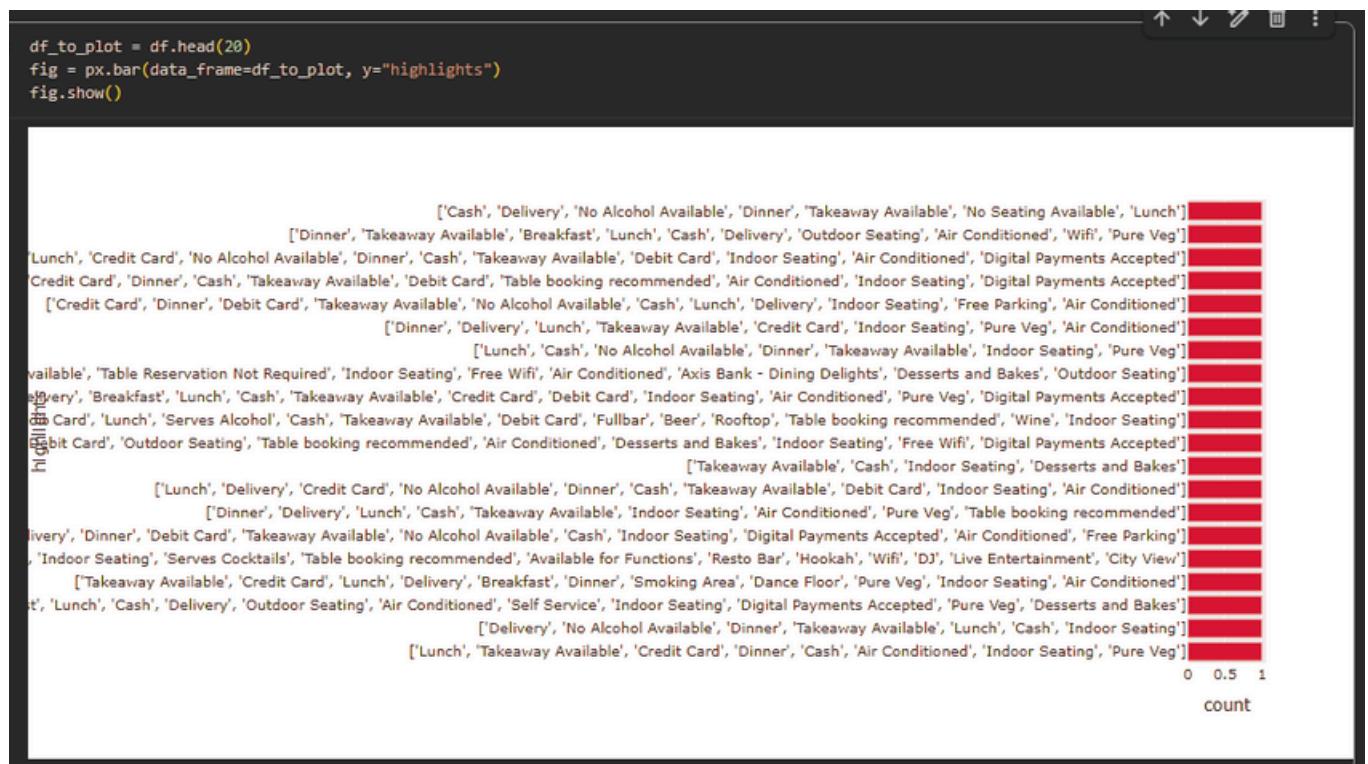
I have visualized the bar graph here we can analyse & here we use seaborn bar graph and here we see the total counts of outlets of restaurants as we can see the top and bottom of total count of outlets chain on the top domino's pizza and cafe coffee day and the bottom the london shakes & kanti sweets and here as usually we use here seaborn bar plot(sns.barplot).

## Analysis Dine in:



Here i want to show you as above i show you the top and bottom of chain of restaurants here i will show you the ratings of those restaurant outlets as you can see i already visualized the graph you see and analyse rating of those restaurants like as keventers have max rating and burger king & lassi shop have lowest rating . here we use seaborn bar graph (sns.barplot).

## Restaurant Features:



Here we can see all the types of restaurant features, including the types of facilities and cuisines served by most restaurants. Based on the available data, we can see the highlights and counts, which allow us to analyze the information & get major insights from data here we can see most of highlights is showing alcohol, wifi, indoor seating etc. Here we use plotly graph (px.bar).

## Restaurant Features with their ratings:



Here we can see all the types of restaurant features, including the types of facilities and cuisines served by most restaurants. Based on the available data, we can see the highlights and counts, which allow us to analyze the information & get major insights from data here we can see most of highlights is showing alcohol, wifi, indoor seating etc. But when it comes on rating we analyze the comfort & taste se here we visualized the highlights of restaurants with their ratings Here we use plotly graph (px.bar).

## word cloud based on customer reviews:

```
from wordcloud import WordCloud

# Join all the highlights into a single string
highlights_text = " ".join(df['highlights'].astype(str))

# Create the word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(highlights_text)

# Display the word cloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

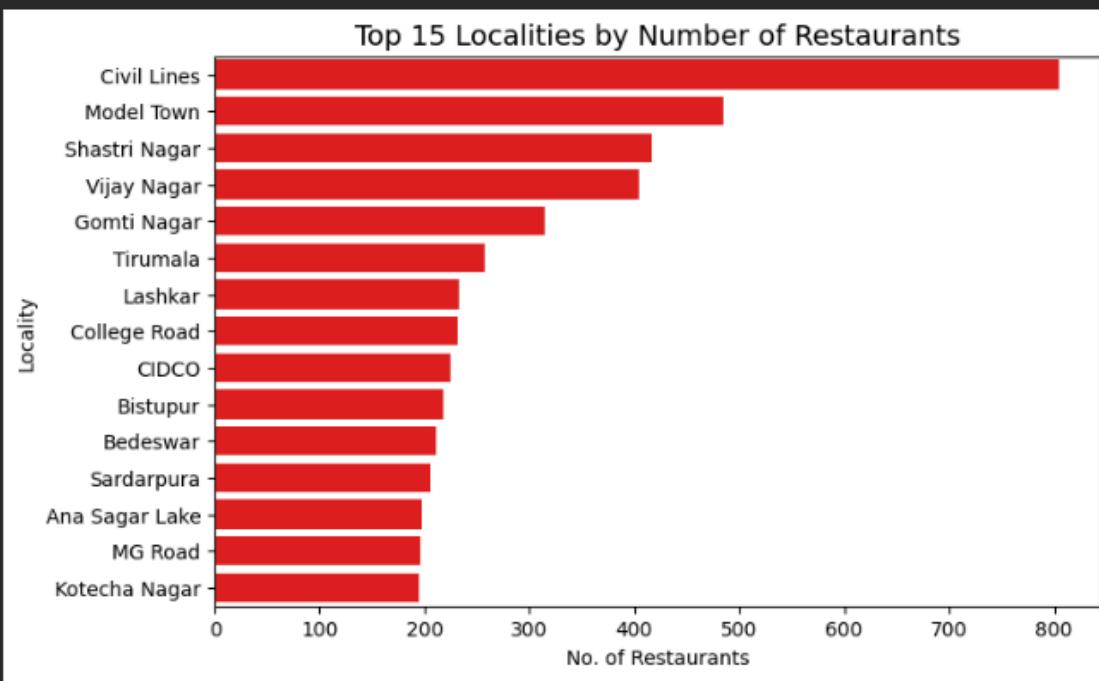


Here we can see customer reviews based on a word cloud. We majorly see the most highlighted word, which is taking up the most space, and its size is also very big in comparison to others, as we have imported the word cloud first and then visualized it.

## Analyze restaurants outlets in different locality:

```
loc = df['locality'].value_counts().head(15)

plt.figure(figsize=(8,5))
sns.barplot(x=loc.values, y=loc.index, color="red")
plt.xlabel("No. of Restaurants")
plt.ylabel("Locality")
plt.title("Top 15 Localities by Number of Restaurants", fontsize=14)
plt.show()
```

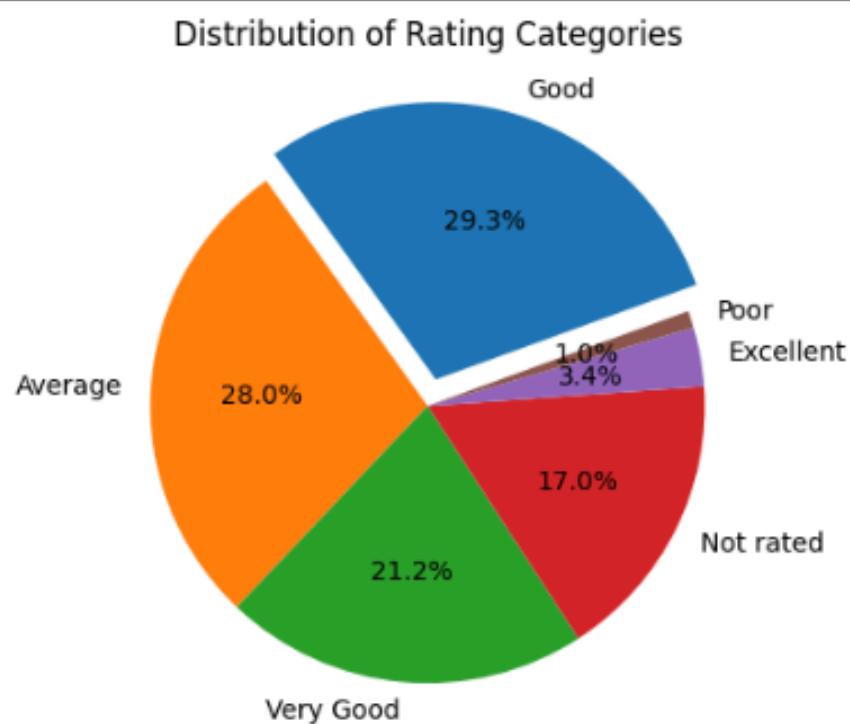


Here I put the top 15 locality data. We can see, as per the data I have analyzed, that the Civil Lines locality has the most outlets of restaurants, while MG Road and Kotecha Nagar have the lowest outlets of restaurants. Here I used a Seaborn bar graph (sns.barplot).

## Analyze restaurants outlets in different locality:

```
a=df["rating_text"].value_counts().head(6)

fig=plt.pie(x=a.values,labels=a.index,autopct="%1.1f%%",startangle=20,explode=ex)
plt.title("Distribution of Rating Categories")
plt.show()
```



Here I have visualized the top 6 rating text data. We can analyze 6 different types of text ratings: good, average, very good, poor, not rated, and excellent. Here I have visualized the counts of text ratings. I used the Matplotlib visualization library (plt.pie).

# Recommendation

ZOMATO

**1. Enhance Delivery Services:-** Restaurants with delivery options tend to have higher ratings. Encourage and support restaurants to improve and expand their delivery services, ensuring a seamless and positive customer experience.

**2. Focus on Peak Hours:-** Identify and optimize resources during peak hours, especially for restaurants operating from 11 AM to 11 PM. Efficient staff management and service optimization during busy periods can enhance customer satisfaction.

**3. Leverage Popular Features:-** Highlight popular features such as outdoor seating, delivery, and credit card acceptance. Encourage restaurants to emphasize and promote these features to attract a broader customer base.

**4. Improve Quality in Low-Rated Chains:-** For chains with lower average ratings, consider implementing quality improvement initiatives. Analyze customer reviews to identify specific areas for enhancement and prioritize actions to boost overall satisfaction.

**5.Engage with Customer Reviews:-** Actively engage with customer reviews to understand specific feedback and address concerns promptly. Building a positive and responsive online presence can contribute to improved customer perceptions.

**6.Strategic Marketing for Top Chains:-** Leverage the popularity of top restaurant chains based on outlets for strategic marketing. Highlight the extensive presence of these chains to attract more customers and reinforce brand visibility.

**7.Optimize Online Ordering:-** Recognize the positive correlation between online ordering availability and higher ratings. Encourage and support restaurants in optimizing their online ordering systems for amore user-friendly experience.

**8.Cuisine-Specific Strategies:-** Tailor strategies based on the most popular cuisines in specific restaurants. Consider unique promotions, events, or menu additions that align with the preferences of the local customer base.

# Conclusion

ZOMATO

In conclusion, the analysis of the Zomato dataset has yielded valuable insights into various aspects of the restaurant industry, catering to both business and user perspectives. Key findings include:

**1. Top Restaurant Chains and Ratings:** Identified the top restaurant chains based on outlet count and their corresponding average ratings. Barbeque Nation emerged as the highest - rated among the top chains.

**2. Customer Sentiments:** Explored customer sentiments through reviews, creating a word cloud and pie chart to visually represent frequently mentioned words and sentiments. 'Good' and 'Average' were predominant sentiments, constituting a significant portion of reviews.

**3. Delivery Services and Ratings:** Established a positive correlation between the availability of delivery services and higher ratings, suggesting the importance of efficient and reliable delivery options. Peak Hours and Operational.

**4. Peak Hours and Operational Timings:** Analyzed peak hours and operational timings, providing insights for businesses to strategically manage resources during busy periods, especially in the 11 AM to 11 PM timeframe.

**5. Cuisine Preferences:** Explored the most popular cuisines and recommended tailoring strategies based on cuisine specific insights to enhance customer satisfaction.

**6. Photo Counts and Brand Visibility:** Investigated the impact of photo counts on brand visibility, highlighting the significance of creating visually appealing experiences to attract customers.

**7. Platform and User Engagement:** Provided actionable recommendations for both restaurants and the platform to optimize services, engage with customer feedback, and enhance overall user experiences. This analysis effectively addresses the project objectives by offering a comprehensive understanding of the Zomato dataset.