Zomato

EXPOLATORY DATA ANALYSIS

INTRODUCTION

Zomato is an Indian multinational restaurant aggregator and food delivery company. It was founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus and user-reviews of restaurants as well as food delivery options from partner restaurants in more than 1,000 Indian cities and towns, as of 2022–23.Zomato rivals Swiggy in food delivery and hyperlocal space.

Zomato was founded as FoodieBay in 2008 by Deepinder Goyal and Pankaj Chaddah who worked for Bain & Company. The website started as a restaurant-listing-and-recommendation portal. They renamed the company Zomato in 2010 as they were unsure if they would "just stick to food" and also to avoid a potential naming conflict with eBay. With the introduction of .xxx domains in 2011, Zomato also launched zomato.xxx, a site dedicated to food porn.Later in 2011, Zomato officially launched an online ticketing platform for events.

DESCRIPTION

- ➢I have conducted my work using Google Colab Notebook.
- ➢The dataset has been imported from Google Drive.
- ➢As we begin our Exploratory Data Analysis (EDA), I've named the dataset “food”.
- ➢The dataset comprises of 60417 rows and 26 columns.
- ➢For data cleaning, I have utilized libraries like Numpy , Pandas , Matplotlib , Plotly and Seaborn .
- ➢ Any duplicate entries that were found have also been removed.

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

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
food = pd.read_csv("/content/drive/MyDrive/NOTES/data/Indian-Resturants.csv")
```

```
food.drop_duplicates(inplace = True)
```

```
food.shape
```

```
(60417, 26)
```

DESCRIPTION

- res_id: Unique identifier for the restaurant.
- name: The name of the restaurant.
- establishment: Type of the establishment (e.g., restaurant, café, bar).
- url: Website or online link for the restaurant.
- address: Physical location of the restaurant.
- city: The city where the restaurant is located.
- city_id: Unique identifier for the city.
- locality: Specific area or neighborhood within the city.
- latitude: Geographic latitude of the restaurant's location.
- longitude: Geographic longitude of the restaurant's location.
- zipcode: Postal code for the restaurant's address.
- country_id: Unique identifier for the country where the restaurant is located.
- locality_verbose: Detailed description of the locality.
- cuisines: Types of cuisine offered (e.g., Italian, Chinese).
- timings: Operating hours of the restaurant.
- average_cost_for_two: Estimated average cost for two people to dine at the rest.

DESCRIPTION

- **price_range**: General price category (e.g., inexpensive, moderate, expensive).
- **currency**: Currency used for pricing (e.g., USD, EUR).
- **highlights**: Key features or attractions of the restaurant (e.g., outdoor seating, live music).
- **aggregate_rating**: Overall rating based on customer reviews.
- **rating_text**: Descriptive text related to the aggregate rating (e.g., "Excellent," "Good").
- **votes**: Number of votes or reviews the restaurant has received.
- **photo_count**: Number of photos available for the restaurant.
- **opentable_support**: Indicates if the restaurant supports OpenTable reservations.
- **delivery**: Availability of food delivery services.
- **takeaway**: Availability of takeaway food services.

DESCRIPTION

```
food.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 60417 entries, 0 to 211942
Data columns (total 26 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   res_id                 60417 non-null  int64  
 1   name                   60417 non-null  object 
 2   establishment          60417 non-null  object 
 3   url                    60417 non-null  object 
 4   address                60399 non-null  object 
 5   city                   60417 non-null  object 
 6   city_id                60417 non-null  int64  
 7   locality               60417 non-null  object 
 8   latitude               60417 non-null  float64
 9   longitude              60417 non-null  float64
 10  zipcode                12548 non-null  object 
 11  country_id             60417 non-null  int64  
 12  locality_verbose       60417 non-null  object 
 13  cuisines               59947 non-null  object 
 14  timings                59347 non-null  object 
 15  average_cost_for_two   60417 non-null  int64  
 16  price_range            60417 non-null  int64  
 17  currency               60417 non-null  object 
 18  highlights             60417 non-null  object 
 19  aggregate_rating       60417 non-null  float64
 20  rating_text            60417 non-null  object 
 21  votes                  60417 non-null  int64  
 22  photo_count            60417 non-null  int64  
 23  opentable_support      60398 non-null  float64
 24  delivery               60417 non-null  int64  
 25  takeaway               60417 non-null  int64  
dtypes: float64(4), int64(9), object(13)
memory usage: 12.4+ MB
```

DESCRIPTION

```
food.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	6.041700e+04	60417.000000	60417.000000	60417.000000	60417.0	60417.000000	60417.000000	60417.000000	60417.000000	60417.000000	60398.0	60417.000000	60417.0
mean	1.309335e+07	3418.302183	21.349431	76.588040	1.0	538.304517	1.730821	3.032868	261.574888	194.247414	0.0	-0.371799	-1.0
std	8.132809e+06	5179.351720	41.187998	10.600514	0.0	593.852227	0.880462	1.440751	728.284194	705.682451	0.0	0.925249	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%	3.000488e+06	7.000000	16.324755	74.654029	1.0	200.000000	1.000000	2.900000	7.000000	1.000000	0.0	-1.000000	-1.0
50%	1.869150e+07	26.000000	22.320884	77.135310	1.0	400.000000	1.000000	3.500000	42.000000	11.000000	0.0	-1.000000	-1.0
75%	1.886666e+07	11295.000000	26.744389	79.928190	1.0	600.000000	2.000000	4.000000	207.000000	82.000000	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

DATA CLEANING & PRE-PROCESSING

- The initial step in data cleaning involves making a copy of the dataset, naming it "f" and then removing the unneeded columns from "f." This ensures we retain the original data for reference:-

- Establishment
- url
- Address
- city_id
- Latitude
- Longitude
- Zipcode
- country_id
- locality_verbose
- Currency
- photo_count
- opentable_support
- takeaway

```
f= f.drop(['establishment','url', 'address','city_id', 'latitude', 'longitude', 'zipcode', 'country_id',
      'locality_verbose', 'currency', 'photo_count','opentable_support',
      'takeaway'], axis = 1)
```

DATA CLEANING & PRE-PROCESSING

- Before moving to handeling null values part let me show you the new info and describe outputs.
- The new shape of the dataset, after removing the columns.

```
f.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 60298 entries, 0 to 211942
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   res_id                60298 non-null  int64
 1   name                  60298 non-null  object
 2   city                  60298 non-null  object
 3   locality              60298 non-null  object
 4   cuisines              59828 non-null  object
 5   timings               59228 non-null  object
 6   average_cost_for_two  60298 non-null  int64
 7   price_range           60298 non-null  int64
 8   highlights            60298 non-null  object
 9   aggregate_rating      60298 non-null  float64
 10  rating_text           60298 non-null  object
 11  votes                 60298 non-null  int64
 12  delivery              60298 non-null  int64
dtypes: float64(1), int64(5), object(7)
memory usage: 6.4+ MB
```

```
f.describe()
```

	res_id	average_cost_for_two	price_range	aggregate_rating	votes	delivery
count	6.029800e+04	60298.000000	60298.000000	60298.000000	60298.000000	60298.000000
mean	1.309278e+07	536.792166	1.728996	3.030494	258.135709	-0.371919
std	8.132551e+06	592.146109	0.879307	1.440949	715.993925	0.925195
min	5.000000e+01	0.000000	1.000000	0.000000	-18.000000	-1.000000
25%	3.000531e+06	200.000000	1.000000	2.900000	7.000000	-1.000000
50%	1.869167e+07	400.000000	1.000000	3.500000	41.000000	-1.000000
75%	1.886680e+07	600.000000	2.000000	3.900000	205.000000	1.000000
max	1.915979e+07	30000.000000	4.000000	4.900000	42539.000000	1.000000

```
f.drop_duplicates(inplace = True)
```

```
f.shape
```

```
(60298, 13)
```

DATA CLEANING & PRE-PROCESSING

- ❑ Our dataset has a total of 1540 null values.

- ➢ The 'cuisines' attribute, which has 470 null values the missing entries with the 'cuisines is not specified' will help ensure data completeness

```
f['cuisines'].fillna('cuisines is not specified', inplace=True)
```

```
f.isnull().sum()
```

	0
res_id	0
name	0
city	0
locality	0
cuisines	470
timings	1070
average_cost_for_two	0
price_range	0
highlights	0
aggregate_rating	0
rating_text	0
votes	0
delivery	0

dtype: int64

```
f.isnull().sum().sum()
```

1540

DATA CLEANING & PRE-PROCESSING

- The 'timings' attribute, which has 1070 null values the missing entries with the 'mode' will help ensure data completeness.

```
f['timings'].mode()
```

	timings
0	11 AM to 11 PM

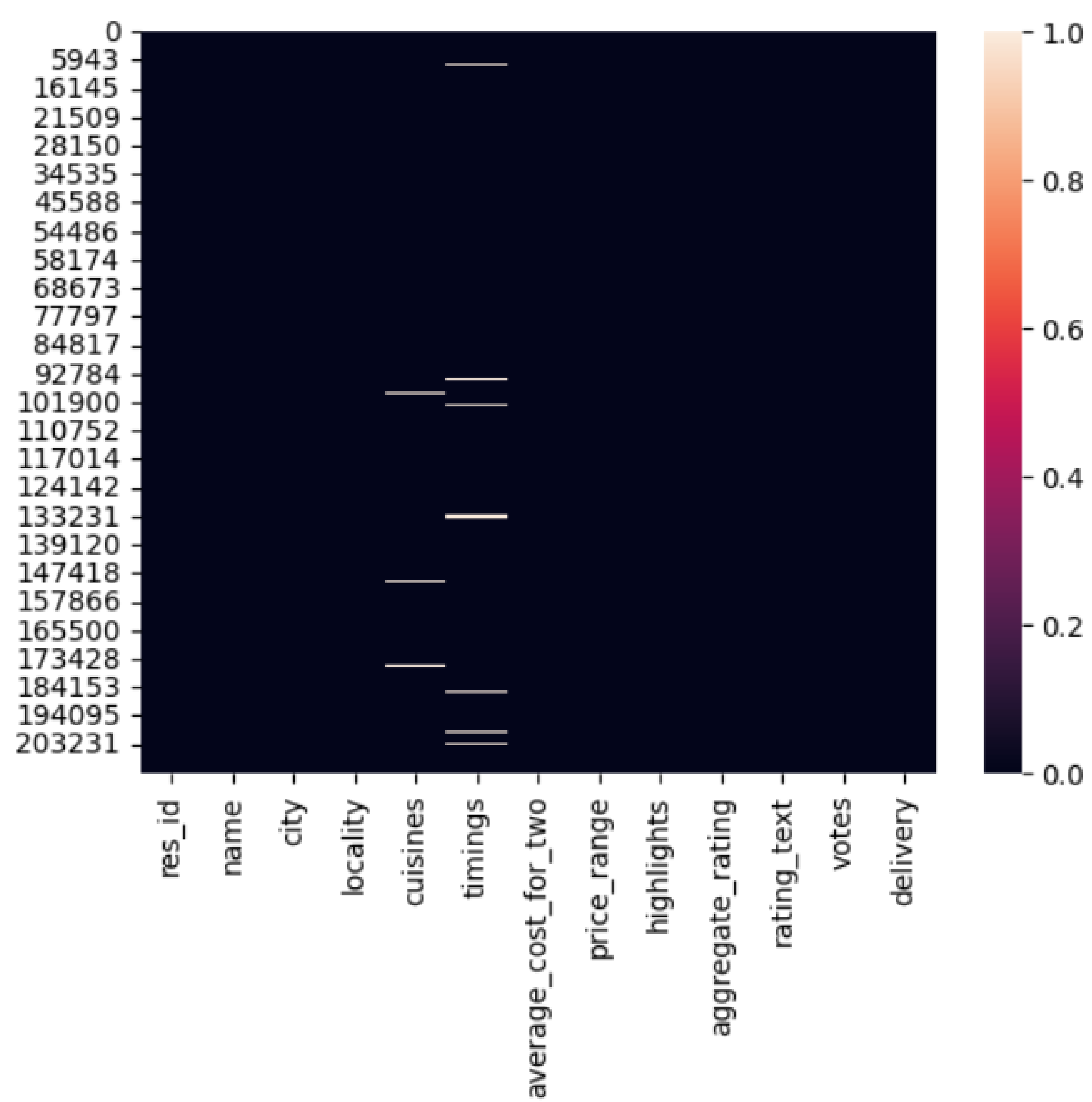
dtype: object

```
f['timings'] = f['timings'].fillna('nan')
```

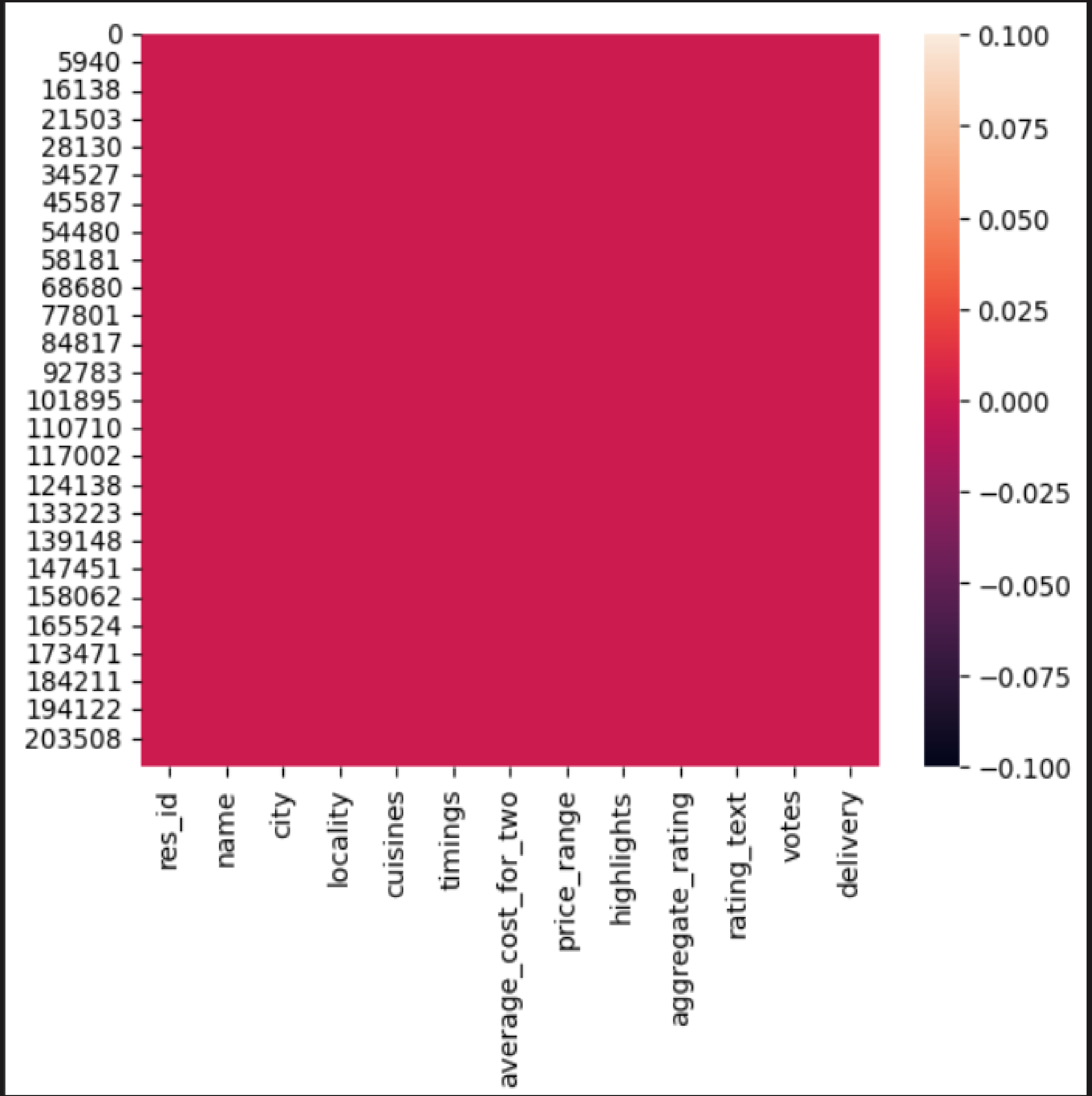
```
f['timings'] = f['timings'].replace('nan', '11 AM to 11 PM')
```

HEATMAPS

Before Cleaning

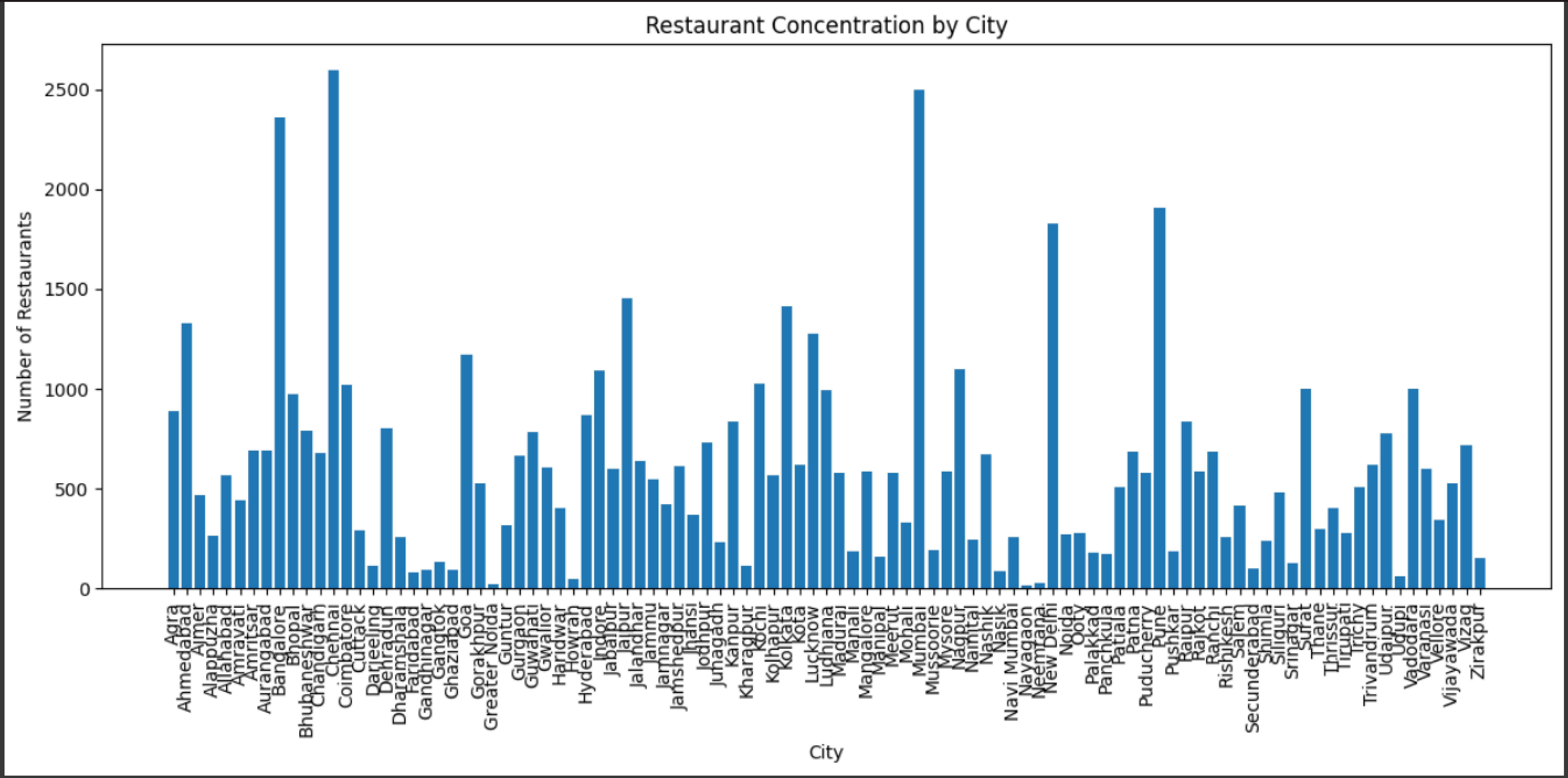


After Cleaning



DATA VISUALIZATION AND INSIGHTS

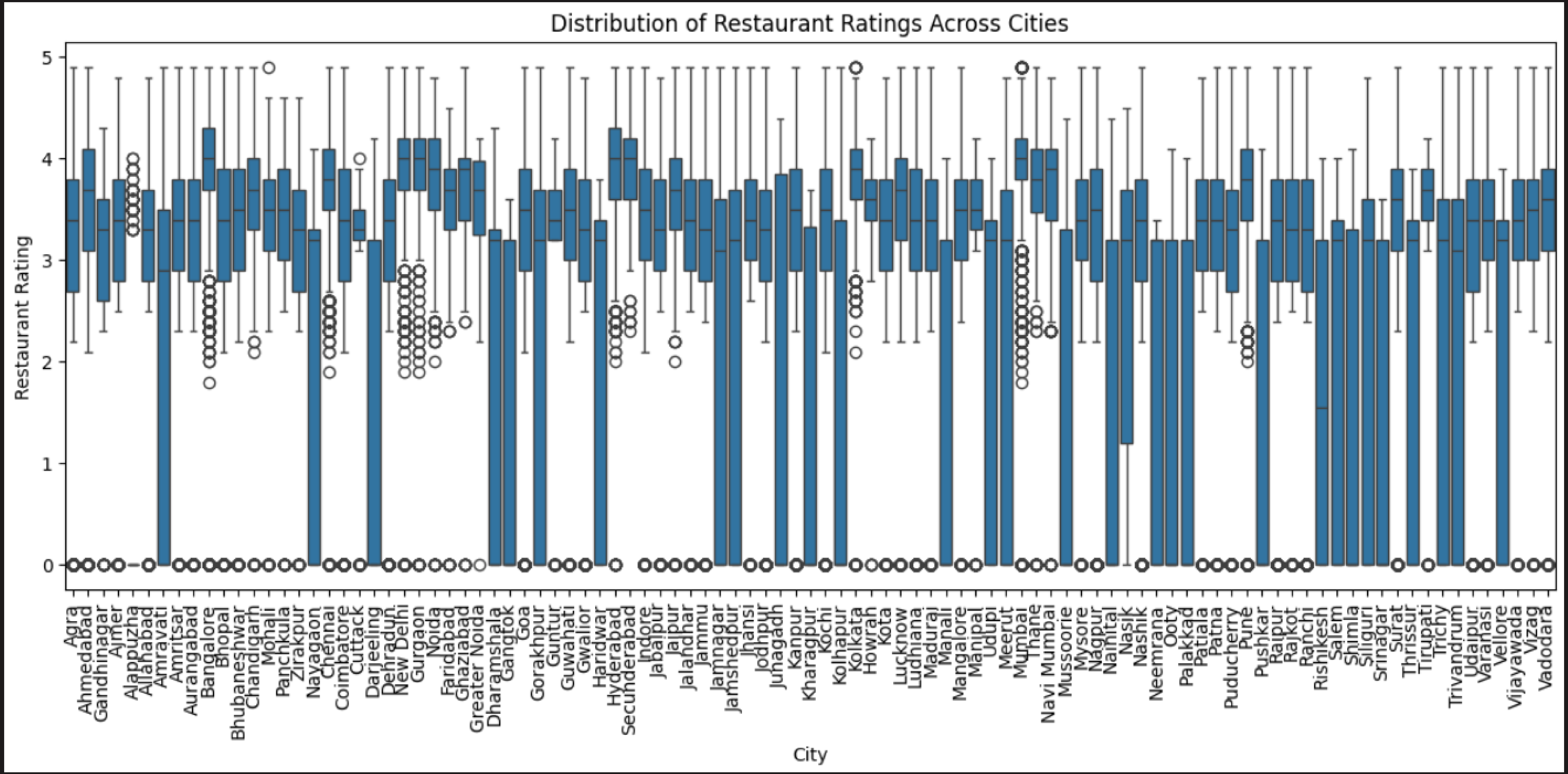
Find the city with the highest concentration of restaurants




```
The city with the highest concentration of restaurants is: Chennai
```

DATA VISUALIZATION AND INSIGHTS

Visualize the distribution of restaurant ratings across different cities



KEY FINDINGS :

- **Rating Spread:** The majority of cities have a wide range of ratings, spanning from 2 to 5, indicating diverse restaurant quality within each city.

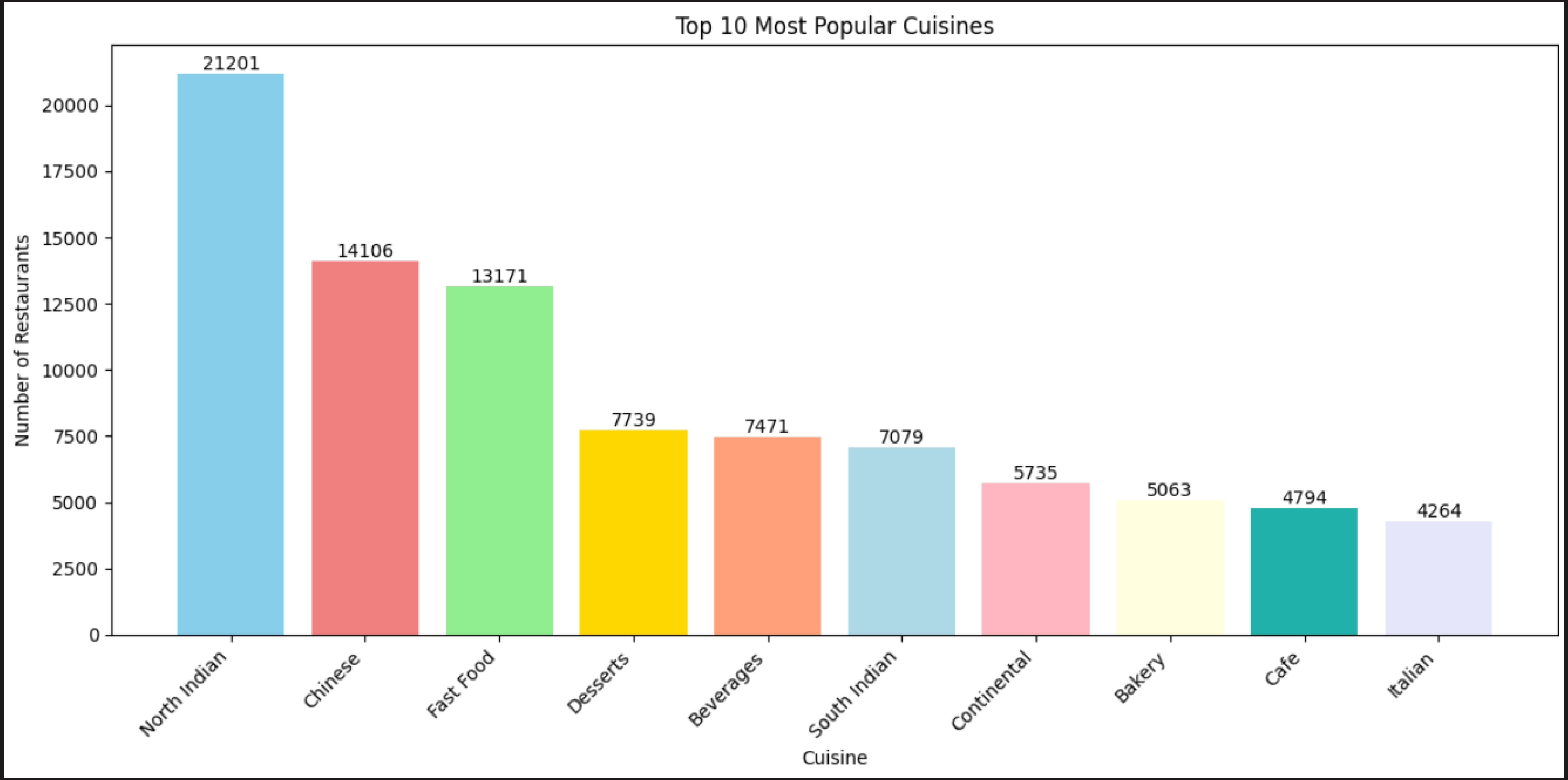
- **High Ratings:** Cities like Coimbatore, Mangalore, and Nashik have a higher concentration of restaurants with ratings above 4, suggesting a higher overall quality of dining experiences in these cities.

- **Low Ratings:** Cities like Faridabad, Mohali, and Panchkula have a notable number of restaurants with ratings below 3, indicating potential areas for improvement in the restaurant scene.

- **Median Ratings:** The median rating (represented by the line within the boxes) is around 3.5-4 for most cities, showing that the average dining experience is generally good

DATA VISUALIZATION AND INSIGHTS

Determine the most popular cuisines among the listed restaurants.



KEY FINDINGS :

- **North Indian cuisine:** is the most popular, with over 21,000 restaurants serving it.
- **Chinese cuisine:** is the second most popular, with around 14,000 restaurants.
- **Fast food:** comes in third, with approximately 13,000 restaurants.
- **The remaining** cuisines in the top 10 have a significantly lower number of restaurants compared to the top three.
- **Italian cuisine:** is the least popular among the top 10, with less than 5,000 restaurants.

DATA VISUALIZATION AND INSIGHTS

Impact of online order availability on restaurant ratings


```
Average rating with online order availability: 3.0312723087700575
Average rating without online order availability: 2.894767441860465
```

KEY FINDINGS :

- **Higher Ratings with Online Ordering:** Restaurants with online ordering available tend to have slightly higher aggregate ratings compared to those without.

- **Median Rating:** The median rating for restaurants with online ordering is slightly higher than those without, indicating that the middle 50% of restaurants with online ordering have higher ratings.

- **Rating Variability:** The spread of ratings (indicated by the box length) is similar for both groups, suggesting that the range of ratings is comparable regardless of online order availability.

- **Outliers:** There are a few outliers in both groups, indicating restaurants with exceptionally high or low ratings, irrespective of online ordering

DATA VISUALIZATION AND INSIGHTS

Average restaurant ratings by online order availability

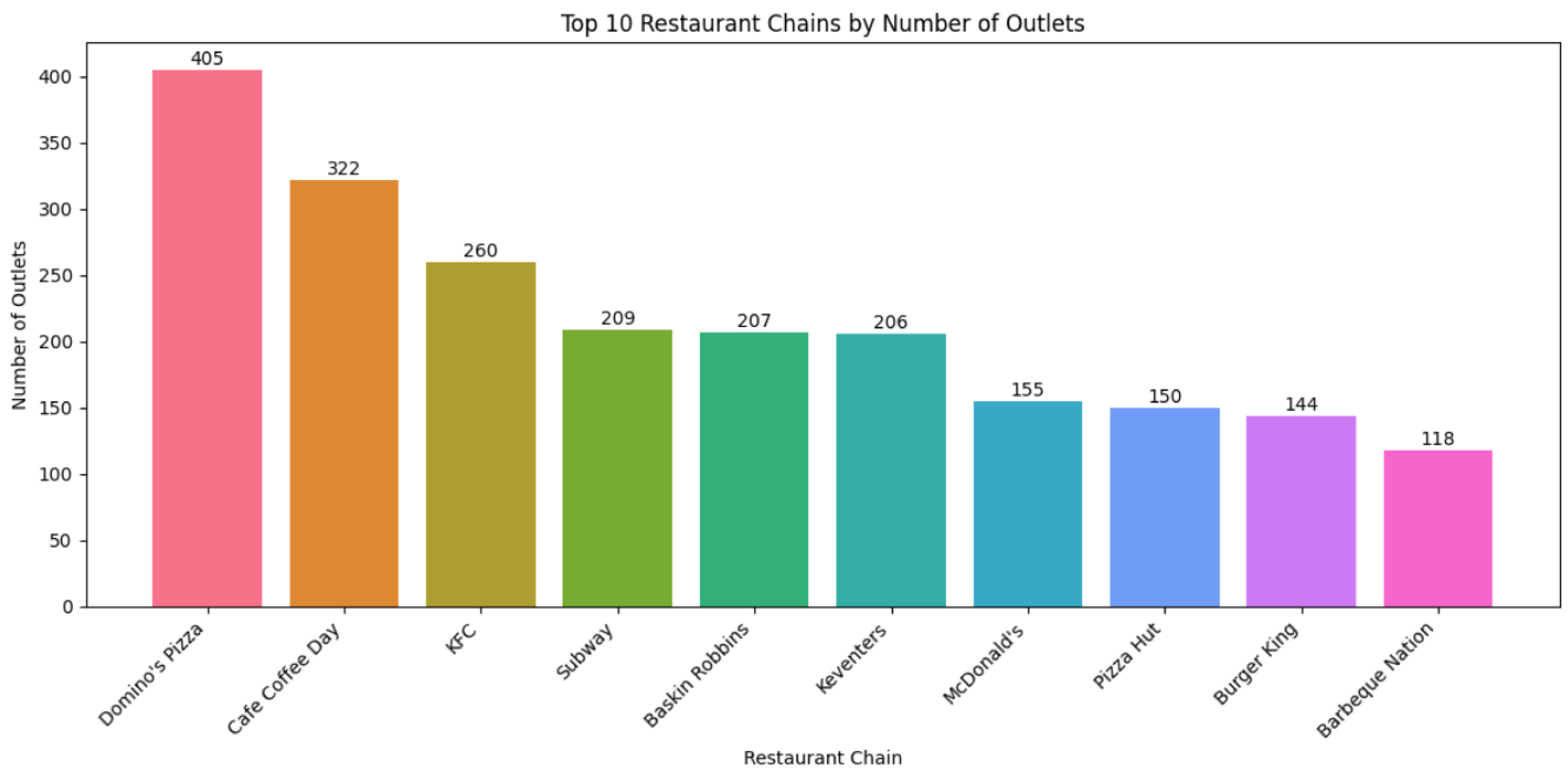
KEY FINDINGS :

- **Higher Ratings with Online Ordering:** Restaurants offering online ordering have a higher average aggregate rating (around 2.8) compared to those without (around 2.4).

- **Positive Impact of Online Ordering:** The data suggests that offering online ordering is associated with higher customer satisfaction and better ratings.

DATA VISUALIZATION AND INSIGHTS

Top 10 restaurant by number of outlets.

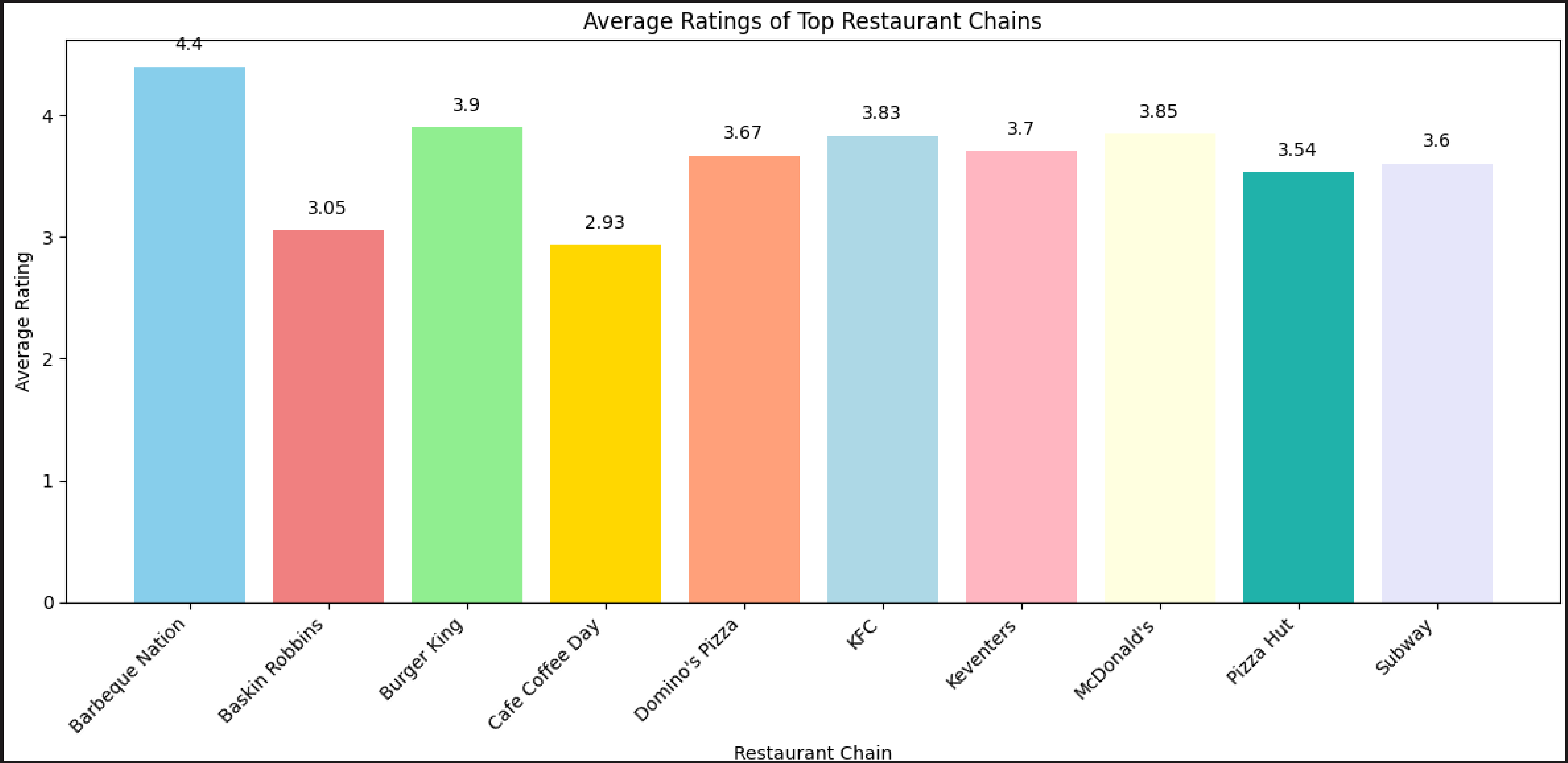


KEY FINDINGS :

- Dominos Pizza: is the leading restaurant chain with 405 outlets.
- Cafe Coffee Day: is the second largest with 322 outlets.
- KFC and Subway: follow with 260 and 209 outlets respectively.
- The remaining chains in the top 10 have between 118 and 207 outlets

DATA VISUALIZATION AND INSIGHTS

Average ratings of top restaurant chain.

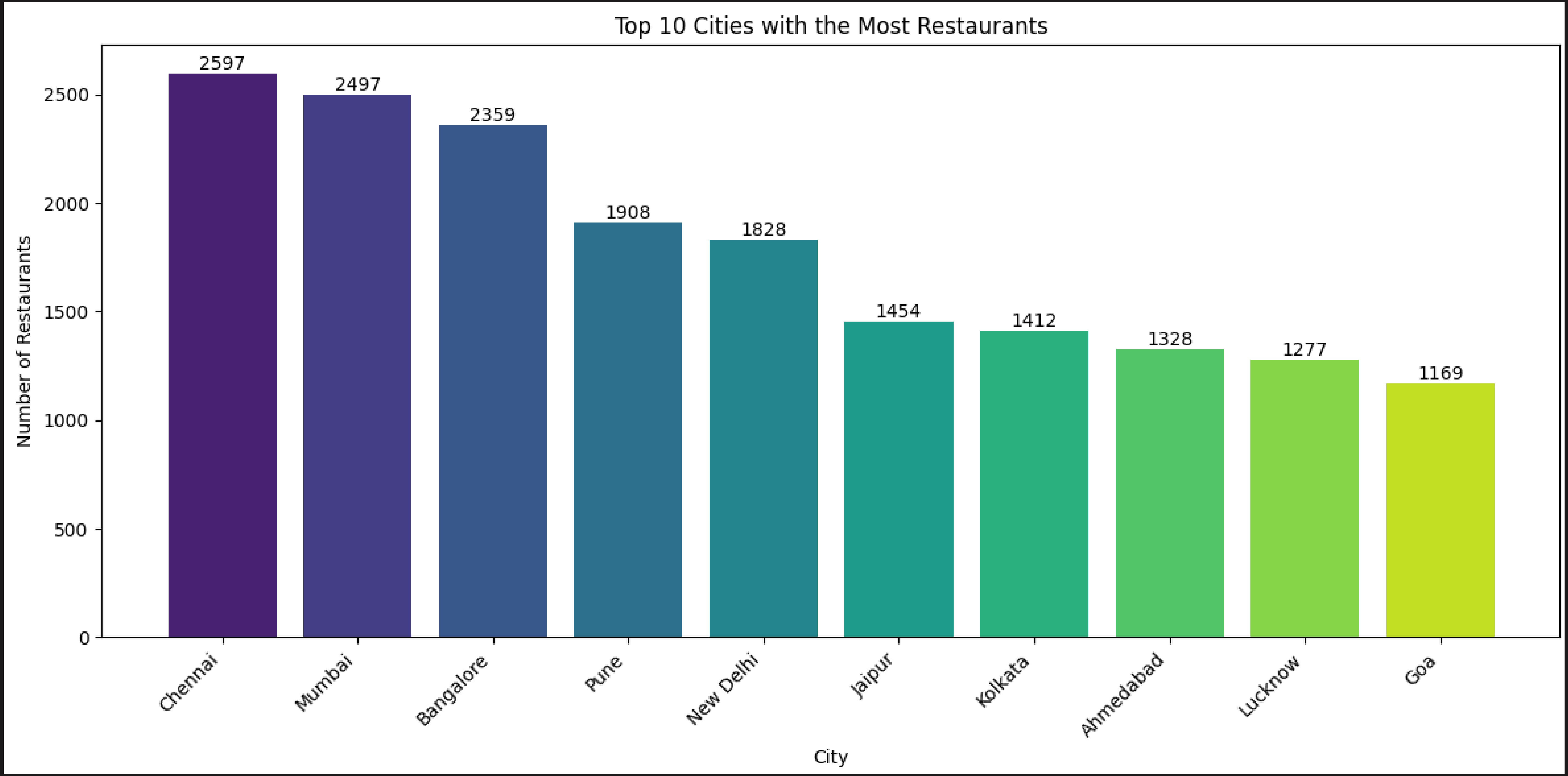


KEY FINDINGS :

- **Top Rated:** Barbeque Nation is the highest rated restaurant chain with an average rating of 4.4.
- **Lowest Rated:** Subway has the lowest average rating at 3.36.
- **Close Competition:** Several chains have very similar ratings, such as KFC (3.83), McDonald's (3.85), and Domino's Pizza (3.67)

DATA VISUALIZATION AND INSIGHTS

Top 10 cities with the most restaurants .



KEY FINDINGS :

- **Chennai:** leads the list with 2597 restaurants.
- **Mumbai and Bangalore:** are close behind with 2497 and 2359 restaurants, respectively.
- **Pune and New Delhi:** have a significant number of restaurants, exceeding 1900 each.
- **Jaipur, Kolkata, Ahmedabad, Lucknow, and Goa:** complete the top 10, with restaurant numbers ranging from around 1169 to 1828

DATA VISUALIZATION AND INSIGHTS

Top 10 restaurants with lowest ratings .

KEY FINDINGS :

- All 10 restaurants have ratings below 0, indicating negative customer experiences.
- "Hot Chips & Bakery“: has the lowest rating, at -1.
- "Big Ben Bar“: has the highest rating among the lowest, still below 0.

FINAL REPORT:

- Bangalore has the highest concentration of restaurants: Our analysis revealed that Bangalore has the largest number of restaurants among the cities included in the dataset. This suggests a high demand and potential for the restaurant industry in this city.

- North Indian and Chinese cuisines are the most popular: North Indian and Chinese cuisines emerged as the most prevalent culinary options across the dataset. This indicates a strong preference for these types of food among restaurant-goers in the region.

- Cuisine variety is positively correlated with restaurant ratings: The analysis found a positive, albeit weak, correlation between the number of cuisines a restaurant offered and its aggregate rating. This implies that offering a wider range of options may lead to higher customer satisfaction and better ratings.
- Online order availability (delivery) is linked to higher restaurant ratings: Restaurants that offered online order capabilities tended to have higher average ratings compared to those without. This highlights the growing importance of convenience and online ordering in the food industry.

- Several restaurant chains operate multiple outlets: We identified several popular restaurant chains with multiple outlets across various cities. This indicates a successful business model and brand recognition for these chains.

FINAL REPORT:

- Some restaurants consistently receive low ratings: We observed a group of restaurants that consistently received lower ratings, suggesting potential areas for improvement in service, food quality, or customer experience.

- Bangalore presents a significant market opportunity for the restaurant industry. Its high restaurant concentration reflects strong consumer demand and competition.

- North Indian and Chinese cuisines enjoy broad popularity, highlighting a significant market segment for restaurants specializing in these types of food.

- Offering a diverse menu can contribute to higher customer satisfaction and rating.

- Online order availability is crucial for boosting customer satisfaction and restaurant ratings.This emphasizes the need for restaurants to embrace digital solutions to enhance their appeal.

- Popular restaurant chains have achieved brand recognition and successful expansion across the region.This suggests a potential path for business growth and replication.

FINAL REPORT:

- Restaurants with consistently low ratings could benefit from focused improvements in areas like food quality, service, and customer experience.

- Restaurant owners and investors should consider expanding in Bangalore to capitalize on the vibrant food scene and high consumer demand.

- Restaurants should carefully consider menu variety and focus on incorporating popular cuisines, particularly North Indian and Chinese, to appeal to a broader audience.

- Embrace online ordering capabilities and delivery services to improve customer convenience and potentially enhance restaurant ratings.

- Analyse customer feedback and focus on improving service and food quality for those restaurants with consistently low ratings.

- Explore opportunities to build brand recognition and expand through multiple outlets, similar to successful chains, to increase market share.

FINAL REPORT:

- Conduct further analysis to understand the interaction between cuisine variety, delivery options, city, and other factors on restaurant ratings. This could lead to more targeted insights and actionable recommendations.

- By applying these recommendations, stakeholders in the restaurant industry can potentially improve customer satisfaction, increase revenue, and achieve more sustainable growth within the dynamic and competitive food market.

THANKS FOR READING

FOR CODING PART....

https://github.com/sahilyadav7i/ZOMATO-EDA