

ZOMATO_DATA_ANALYSIS

Analysis By Vivek Chauhan

11-04-2025

About Dataset

This dataset is a collection of restaurants that are registered on Zomato in Bengaluru City. In this dataset, we have more than 50000 rows and 17 columns, a fairly large dataset.

I was be able to get hands-on experience while performing the following tasks and will be able to understand how real-world problem statement analysis is done.

- Explore the Data**
- Data Cleaning**
- Data Visualization**
- Recomendations**

```
[50]: # Let's analyse the each and every words in the address column and then we are going to count it so we get in which place most of orders.
```

```
# import the counter library first
```

```
from collections import Counter
import nltk
from nltk.corpus import stopwords
```

```
# Download stopwords (only needed once)
nltk.download('stopwords')
```

```
# Get English stopwords
stop_words = set(stopwords.words('english'))
```

```
all_address_words = [] # create an empty list
```

```
for dish in new_data['address']: # loop through address column
    all_address_words.extend(dish.split())
```

```
# Filter out stopwords
filtered_address_words = [word for word in all_address_words if word.lower() not in stop_words]
```

```
# Count the words
address_word_freq = Counter(filtered_address_words)
```

```
# Print top 5 most common words
print(address_word_freq.most_common(5))
```

```
[('Bangalore', 48442), ('Road,', 35506), ('Block,', 11901), ('Main', 10447), ('1st', 8996)]
```

```
[nltk_data] Downloading package stopwords to C:\Users\VIVEK
[nltk_data] CHAUHAN\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Most of the order placed location is 'Bangalore' cause whole the dataset belongs to Bangalore.

```
[47]: # Let's analyse the each and every words in the cuisines column and then we are going to count it.  
# import the counter library first
```

```
from collections import Counter  
import nltk  
from nltk.corpus import stopwords  
  
# Download stopwords (only needed once)  
nltk.download('stopwords')  
  
# Get English stopwords  
stop_words = set(stopwords.words('english'))  
  
all_cuisines_words = [] # create an empty list  
  
for dish in new_data['cuisines']: # loop through cuisines column  
    all_cuisines_words.extend(dish.split())  
  
# Filter out stopwords  
filtered_cuisines_words = [word for word in all_cuisines_words if word.lower() not in stop_words]  
  
# Count the words  
cuisines_word_freq = Counter(filtered_cuisines_words)  
  
# Print top 5 most common words  
print(cuisines_word_freq.most_common(5))
```

```
[('North', 21187), ('Indian,', 20157), ('Indian', 9774), ('South', 8664), ('Chinese,', 8469)]
```

```
[nltk_data] Downloading package stopwords to C:\Users\VIVEK  
[nltk_data] CHAUHAN\AppData\Roaming\nltk_data...  
[nltk_data] Package stopwords is already up-to-date!
```

You can see clearly the winner is in the category of cuisines is 'North'.

```
[46]: # Let's analyse the each and every words in the dish_liked column and then we are going to count it.
# import the counter library first

from collections import Counter
import nltk
from nltk.corpus import stopwords

# Download stopwords (only first time)
nltk.download('stopwords')

# Get English stopwords
stop_words = set(stopwords.words('english'))

all_dish_words = []

for dish in new_data['dish_liked']:
    all_dish_words.extend(dish.split())

# Filter out stopwords
filtered_words = [word for word in all_dish_words if word.lower() not in stop_words]

# Count the words
dish_word_freq = Counter(filtered_words)

# Print top 5 most common words
print(dish_word_freq.most_common(5))
```

```
[nltk_data] Downloading package stopwords to C:\Users\VIVEK
[nltk_data] CHAUHAN\AppData\Roaming\nltk_data...
[nltk_data] Unzipping corpora\stopwords.zip.
[('mentioned', 28078), ('Chicken', 10845), ('Biryani,', 6058), ('Pizza,', 3945), ('Chicken,', 3927)]
```

Here, you can see the 'Chicken', 'Biryani', 'Pizza' is a most frequent words while placed order or you can say Most Favourite dishes of the Customers.

```
[49]: # Let's analyse the each and every words in the location column and then we are going to count it so we get in which place most of orders.  
# import the counter library first
```

```
from collections import Counter  
import nltk  
from nltk.corpus import stopwords  
  
# Download stopwords (only needed once)  
nltk.download('stopwords')  
  
# Get English stopwords  
stop_words = set(stopwords.words('english'))  
  
all_location_words = [] # create an empty list  
  
for dish in new_data['location']: # Loop through location column  
    all_location_words.extend(dish.split())  
  
# Filter out stopwords  
filtered_location_words = [word for word in all_location_words if word.lower() not in stop_words]  
  
# Count the words  
location_word_freq = Counter(filtered_location_words)  
  
# Print top 5 most common words  
print(location_word_freq.most_common(5))
```

```
[('Road', 9301), ('Koramangala', 7782), ('Block', 7734), ('BTM', 5145), ('Nagar', 4833)]
```

```
[nltk_data] Downloading package stopwords to C:\Users\VIVEK  
[nltk_data] CHAUHAN\AppData\Roaming\nltk_data...  
[nltk_data] Package stopwords is already up-to-date!
```

You can see the most of the placed order location is 'Koramangala', & second position is 'BTM'.

```
[48]: # Let's analyse the each and every words in the menu_item column and then we are going to count it.
# import the counter library first
from collections import Counter
import nltk
from nltk.corpus import stopwords

# Download stopwords (only needed once)
nltk.download('stopwords')

# Get English stopwords
stop_words = set(stopwords.words('english'))

all_menu_item_words = [] # create an empty list

for dish in new_data['menu_item']: # loop through menu_item column
    all_menu_item_words.extend(dish.split())

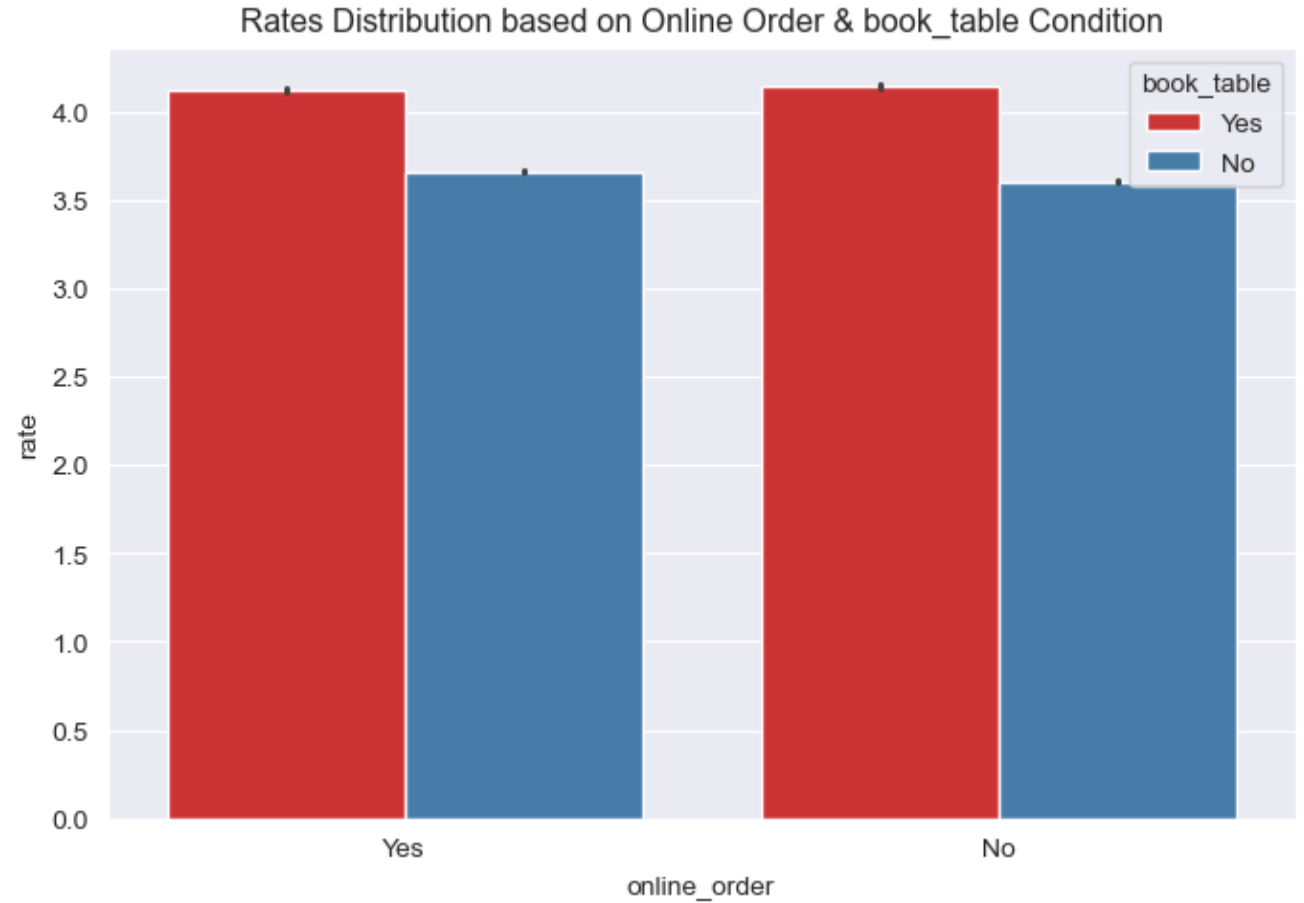
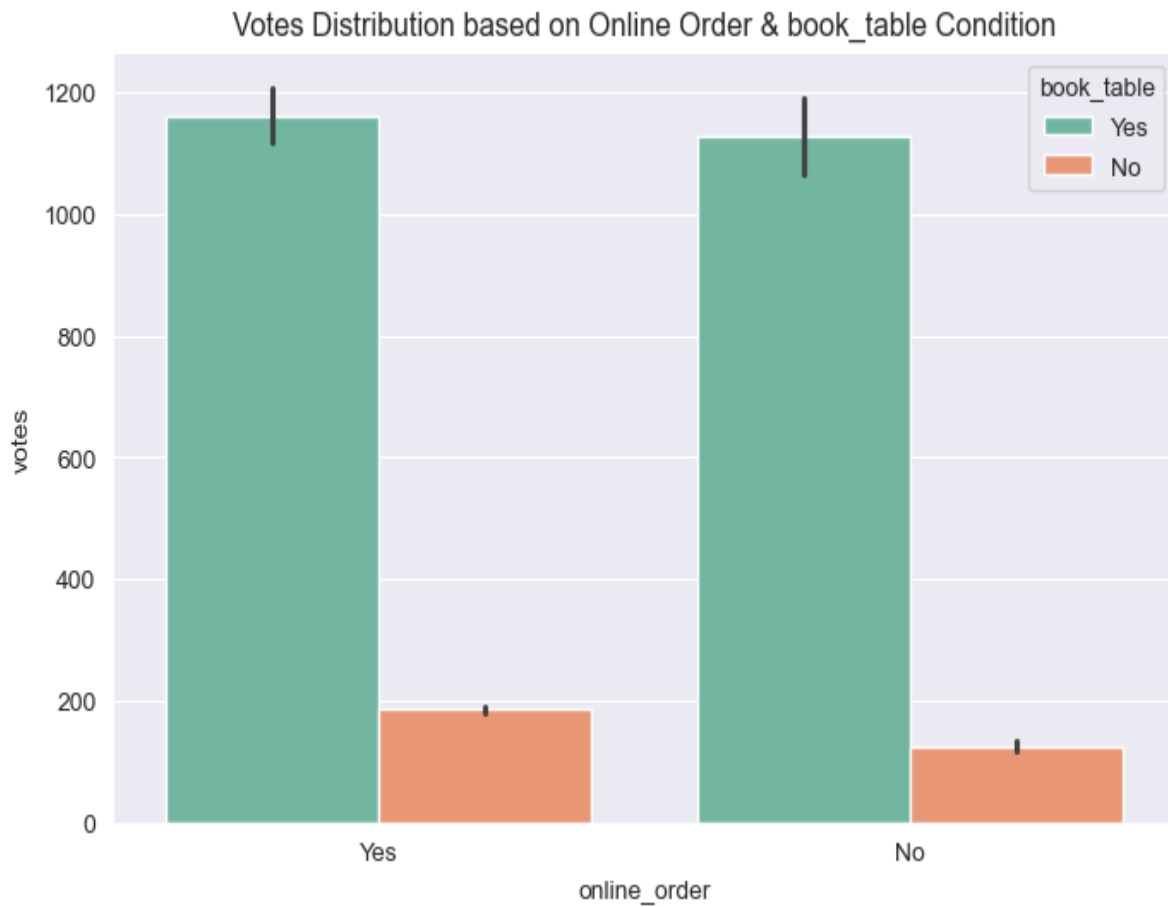
# Filter out stopwords
filtered_menu_item_words = [word for word in all_menu_item_words if word.lower() not in stop_words]

# Count the words
menu_item_word_freq = Counter(filtered_menu_item_words)

# Print top 5 most common words
print(menu_item_word_freq.most_common(5))
```

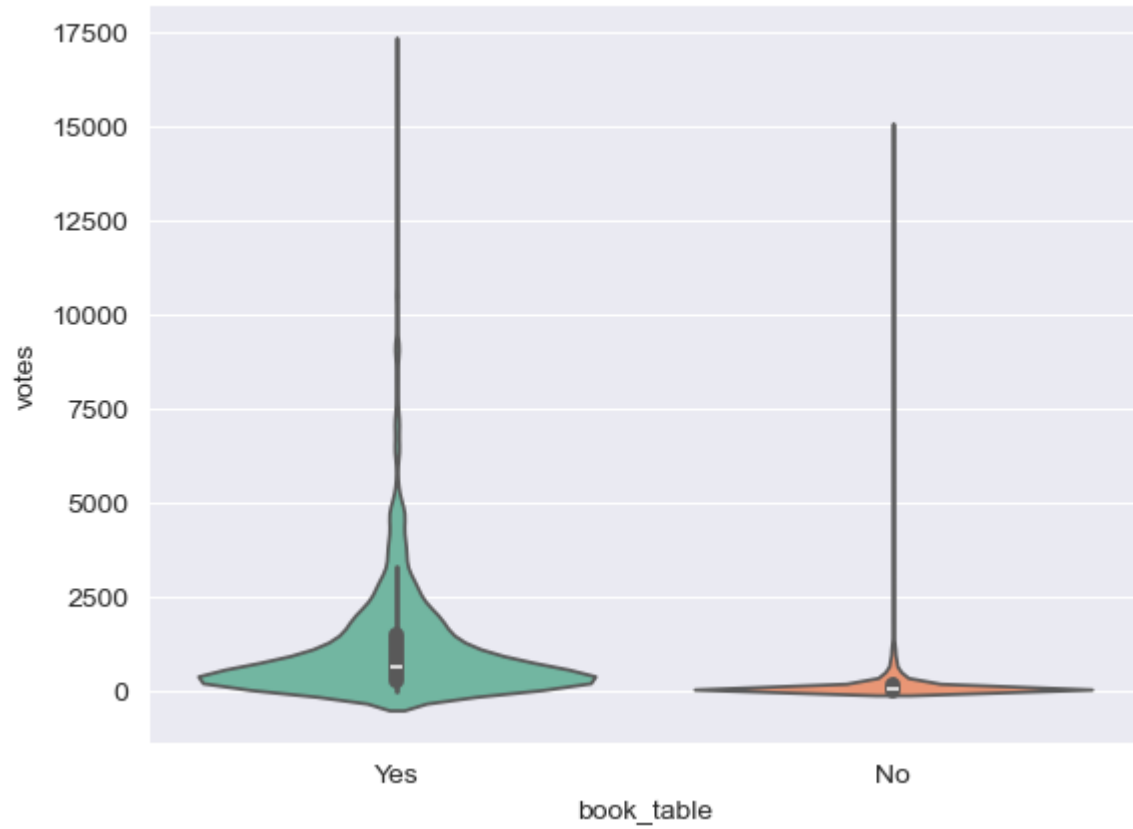
```
[nltk_data] Downloading package stopwords to C:\Users\VIVEK
[nltk_data]   CHAUHAN\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[('Chicken', 132408), ('Veg', 77055), ('Rice', 66644), ('Chicken', 65715), ('Fried', 50917)]
```

You can see clearly the most favourite menu item is 'Chicken'.

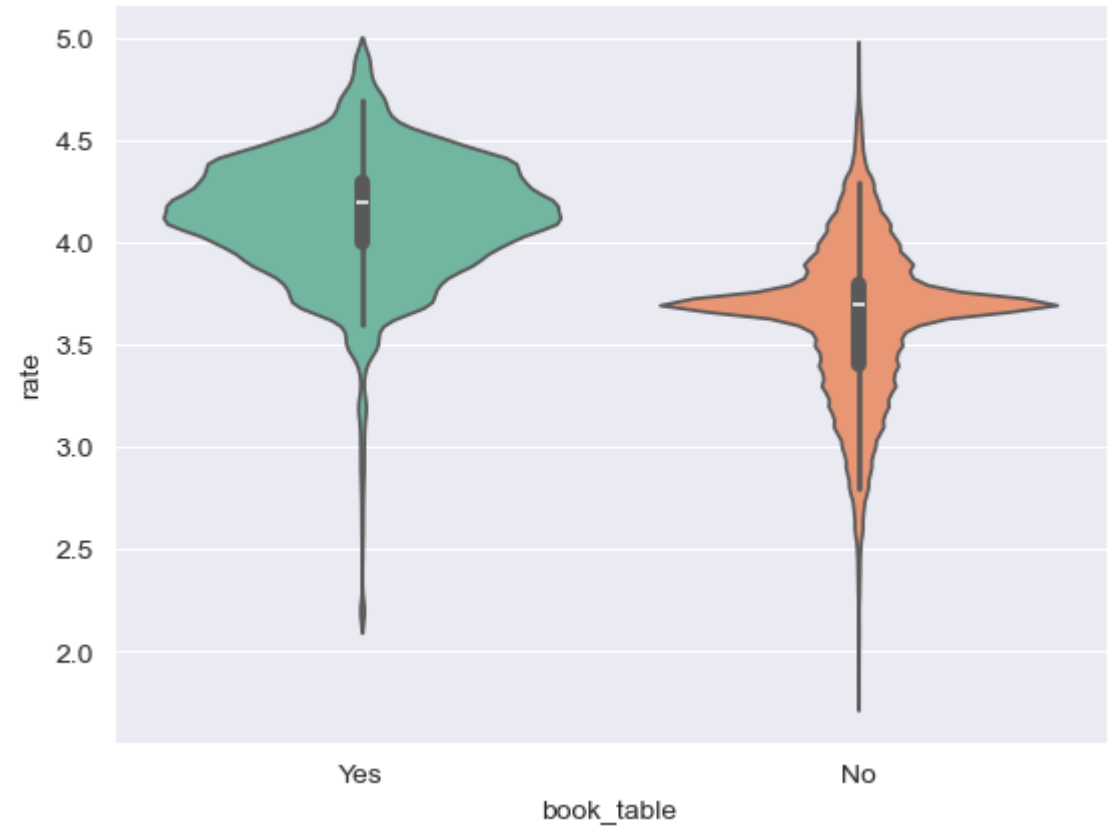


In the chart above, you can see that restaurants which received the highest number of online orders also got the highest votes and ratings.

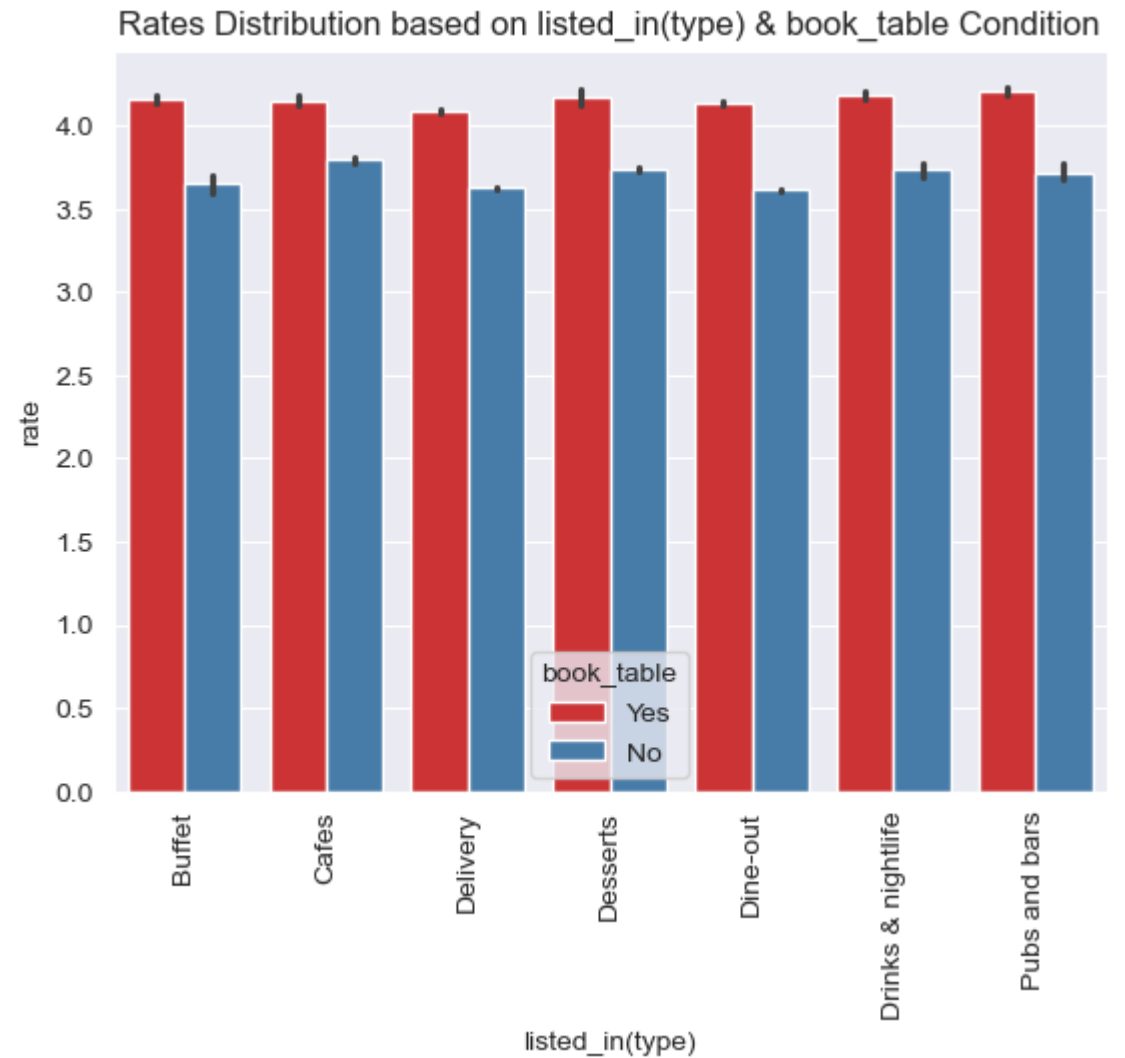
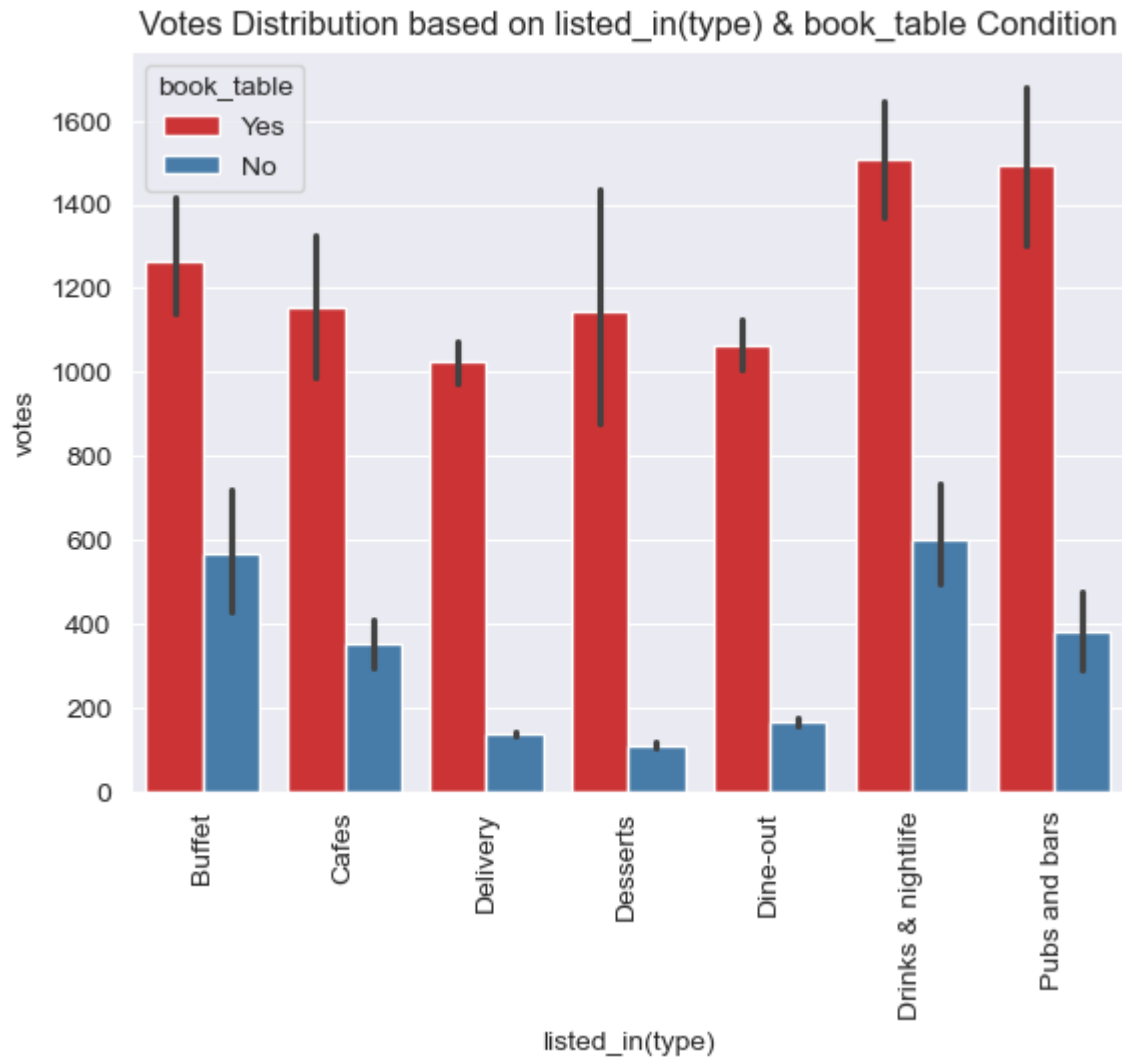
Votes Distribution based on book_table



Rates Distribution based on book_table

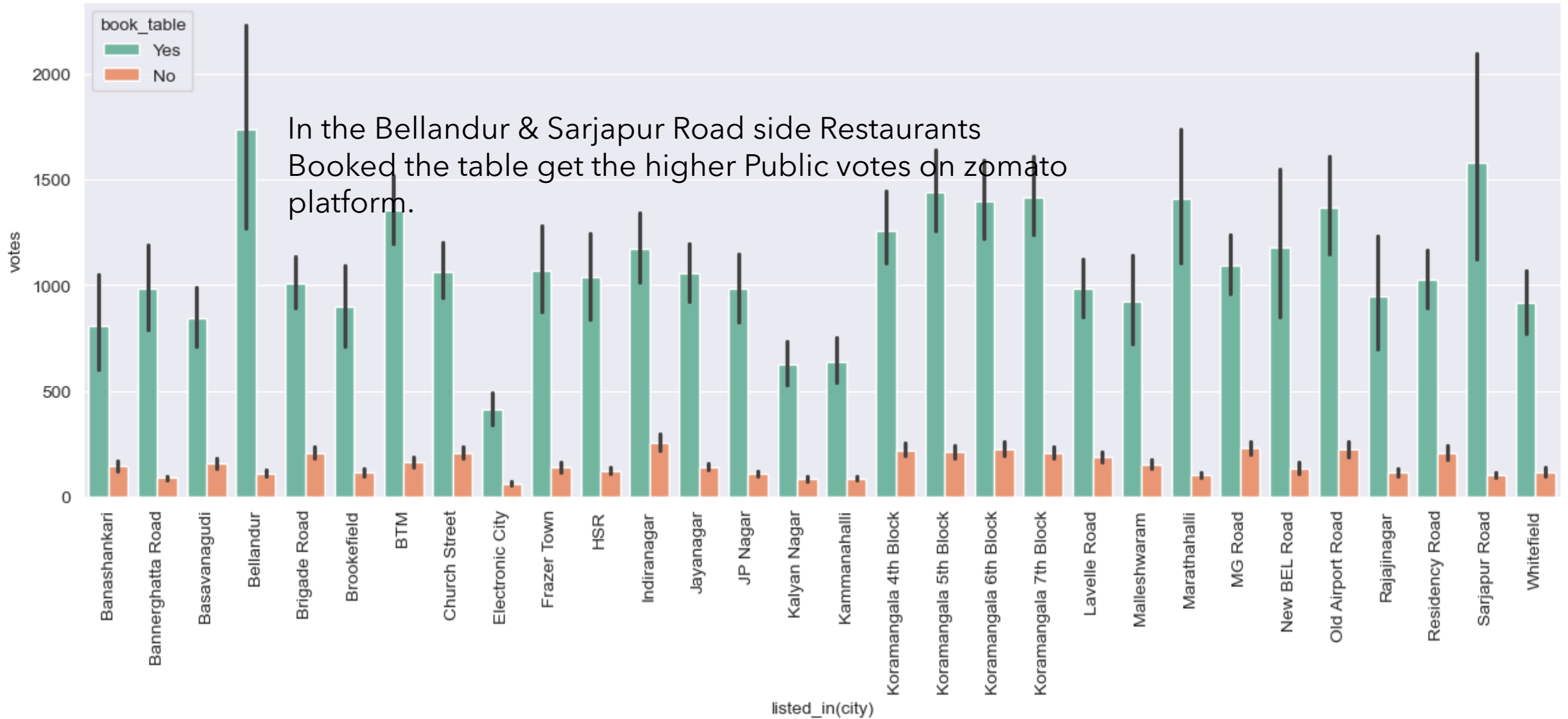


In the chart above, you can see that restaurants which received the highest number of Booked Table also got the highest votes and ratings.

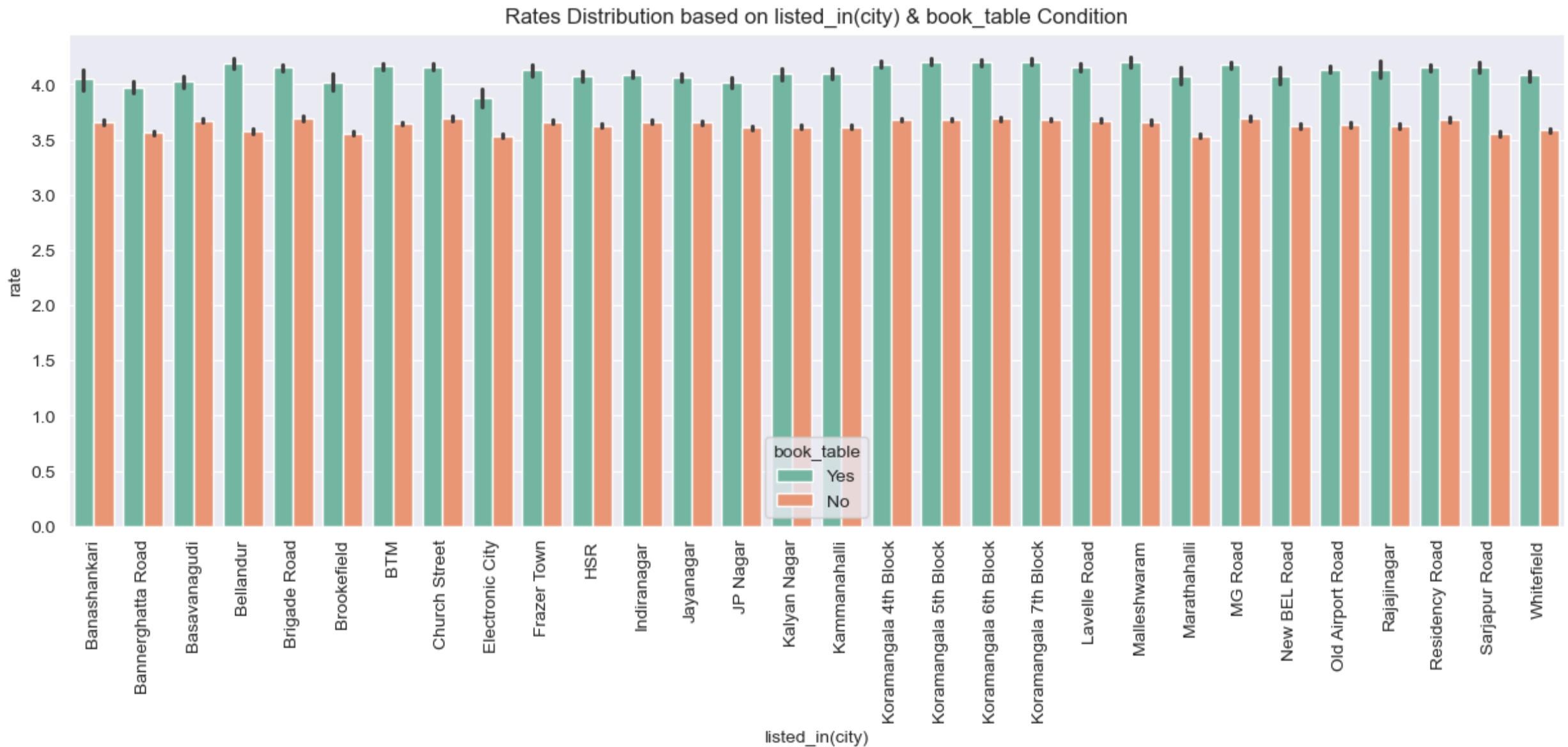


In the chart above, you can see that restaurants which received the highest number of Booked Table also got the highest votes for Drinks & Nightlife, Pubs & Bars Type restaurants and All the type restaurants Get highest ratings.

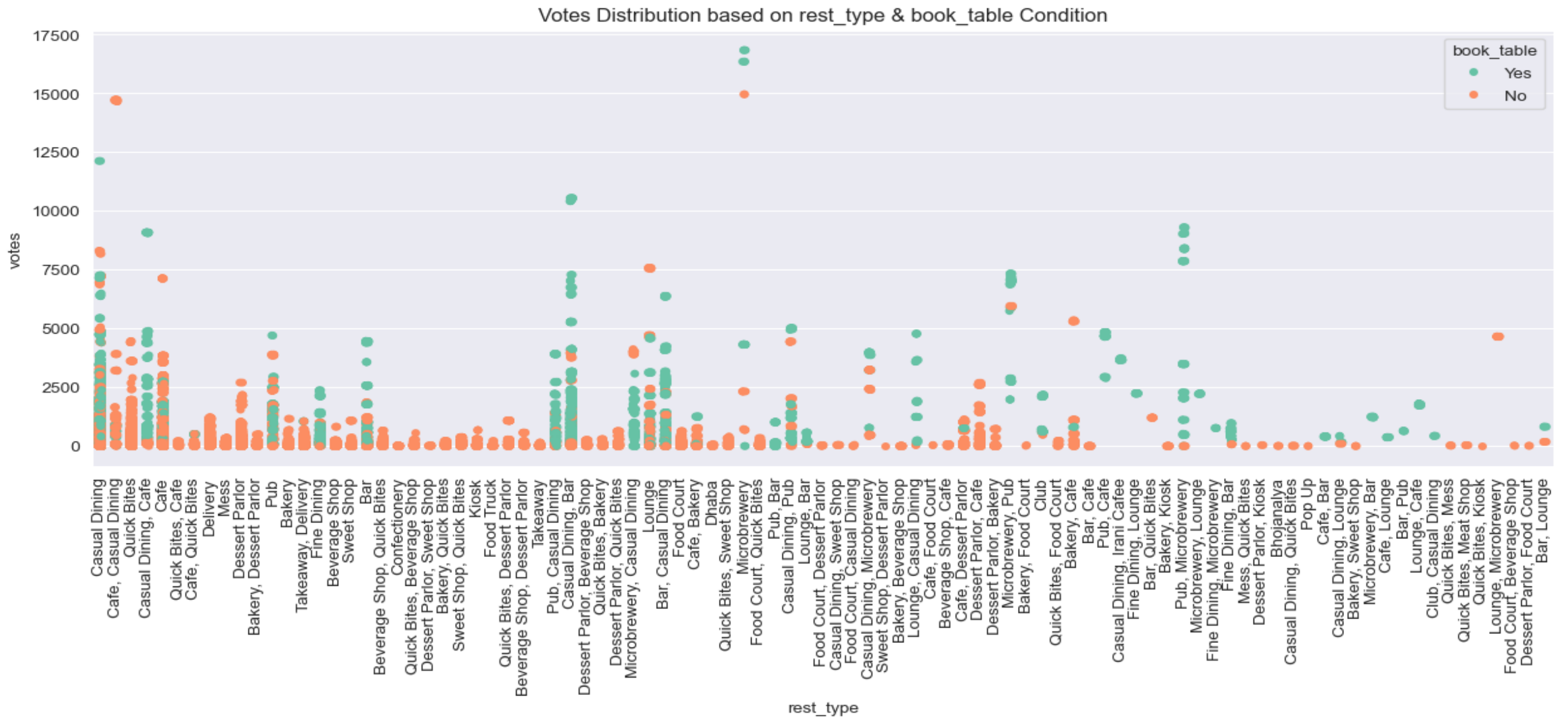
Votes Distribution based on listed_in(city) & book_table Condition



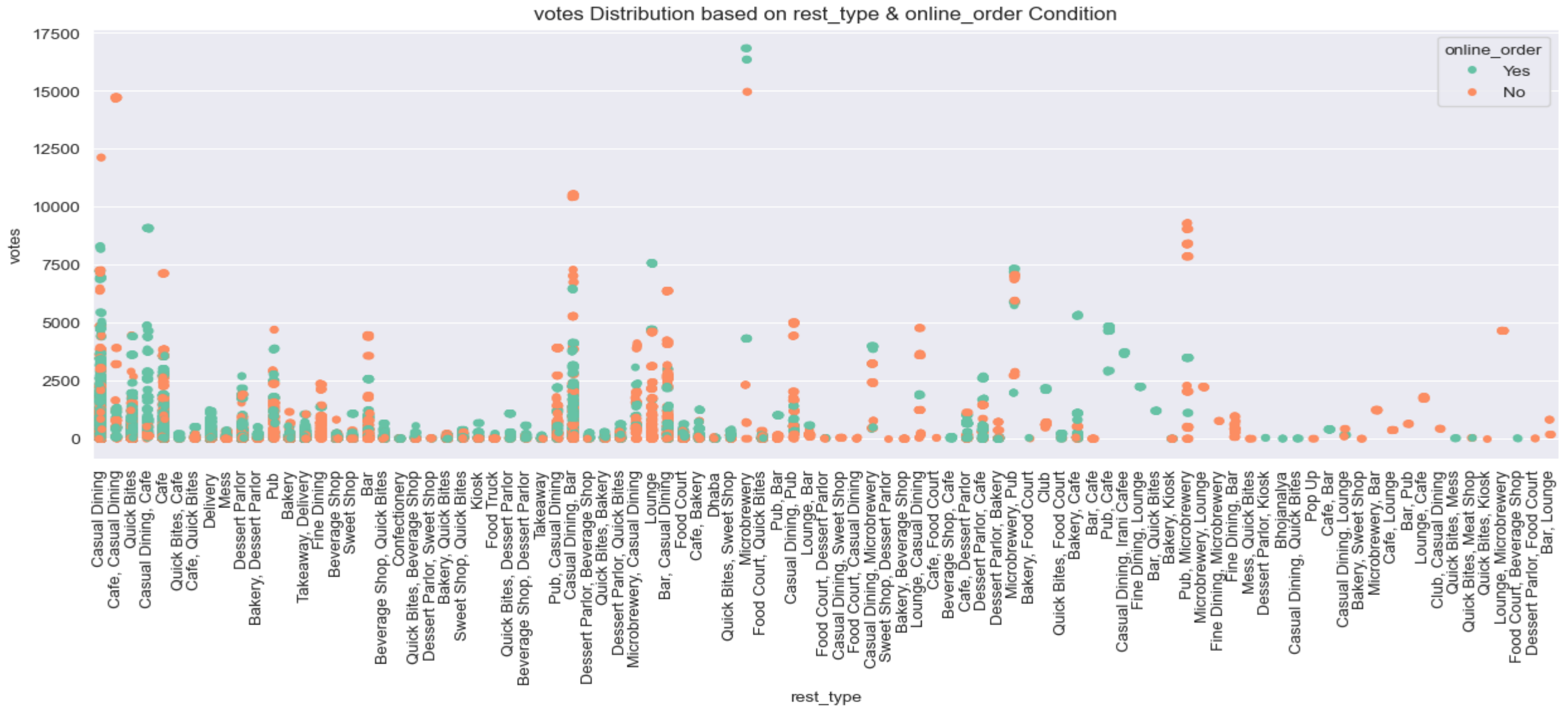
In the chart above, you can see that the restaurants with the highest number of table bookings also received the most votes. These restaurants are located in the Bellandur and Sarjapur Road areas on the Zomato platform.



In the above chart as you can see the most of the city Restaurants and booked the table online and get above 4.0 ratings and those who not booked the table online get more than or nearest 3.5 ratings on zomato platform.



Most of the type of restaurant booked the table and get the below 7500 votes on zomato platform.

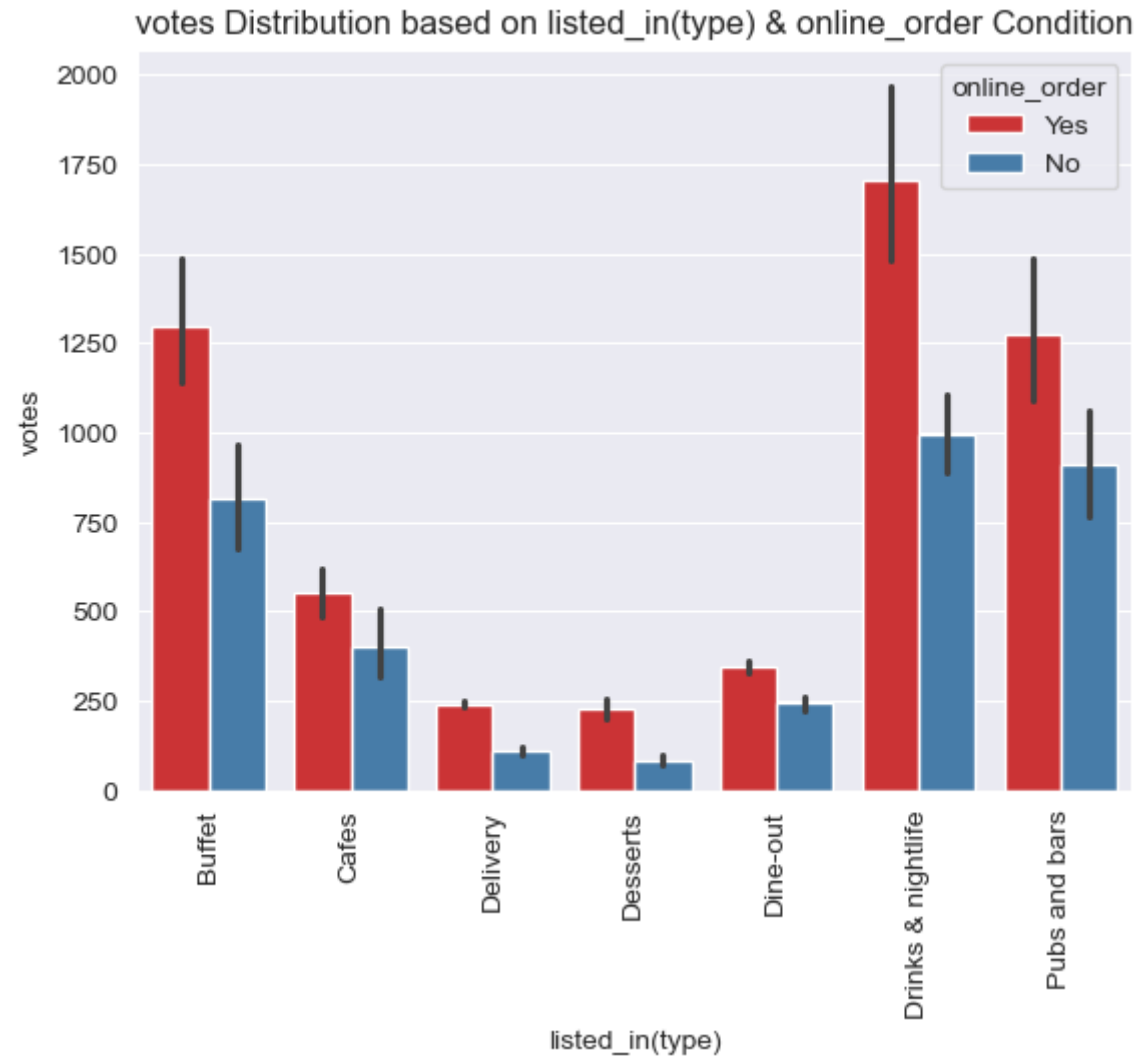


Most of the restaurant type take the online order or not get below 7500 votes on zomato platform.

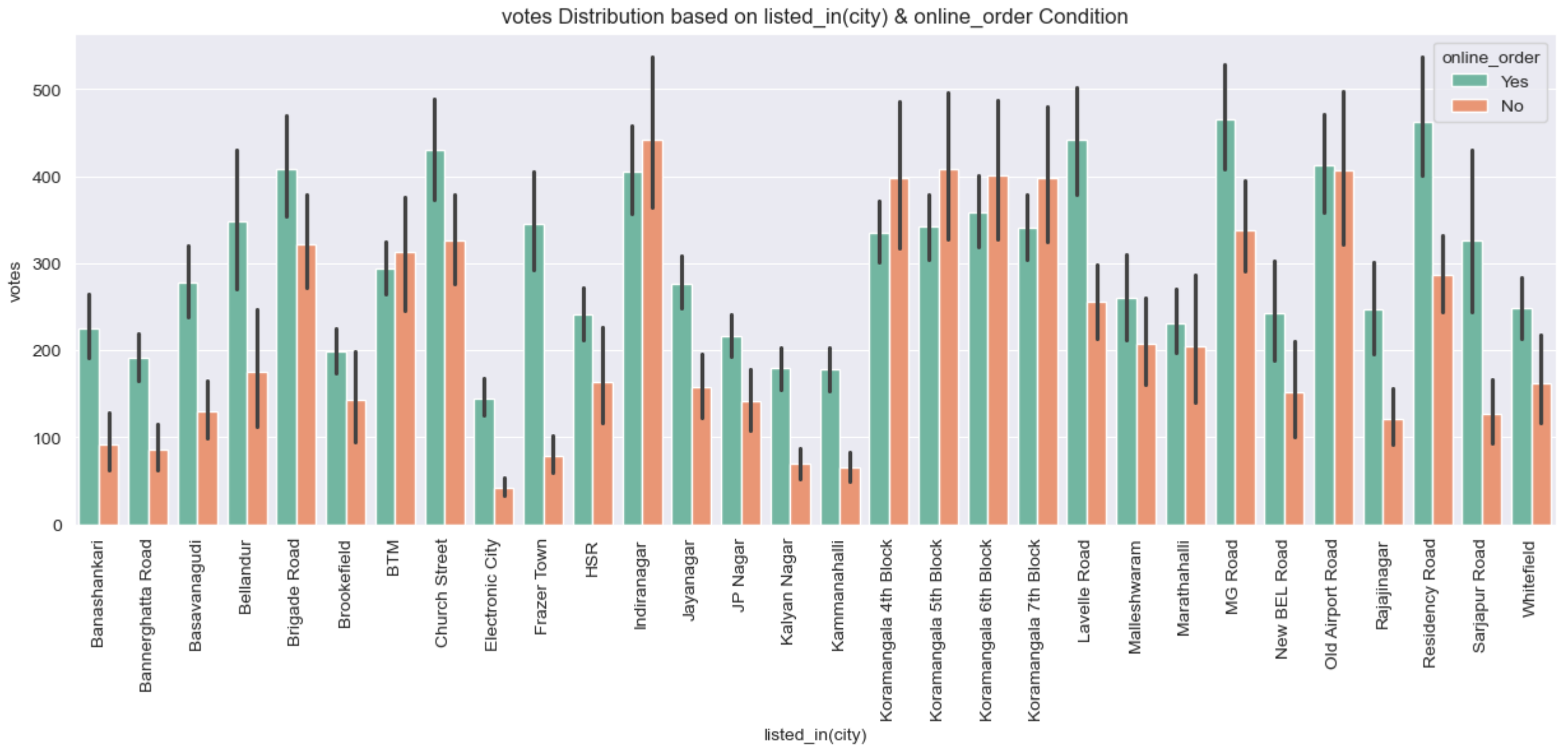
Rates Distribution based on rest_type & book_table Condition



As you can see the most of the restaurants getting the highest ratings cause they take the online orders on zomato platform.



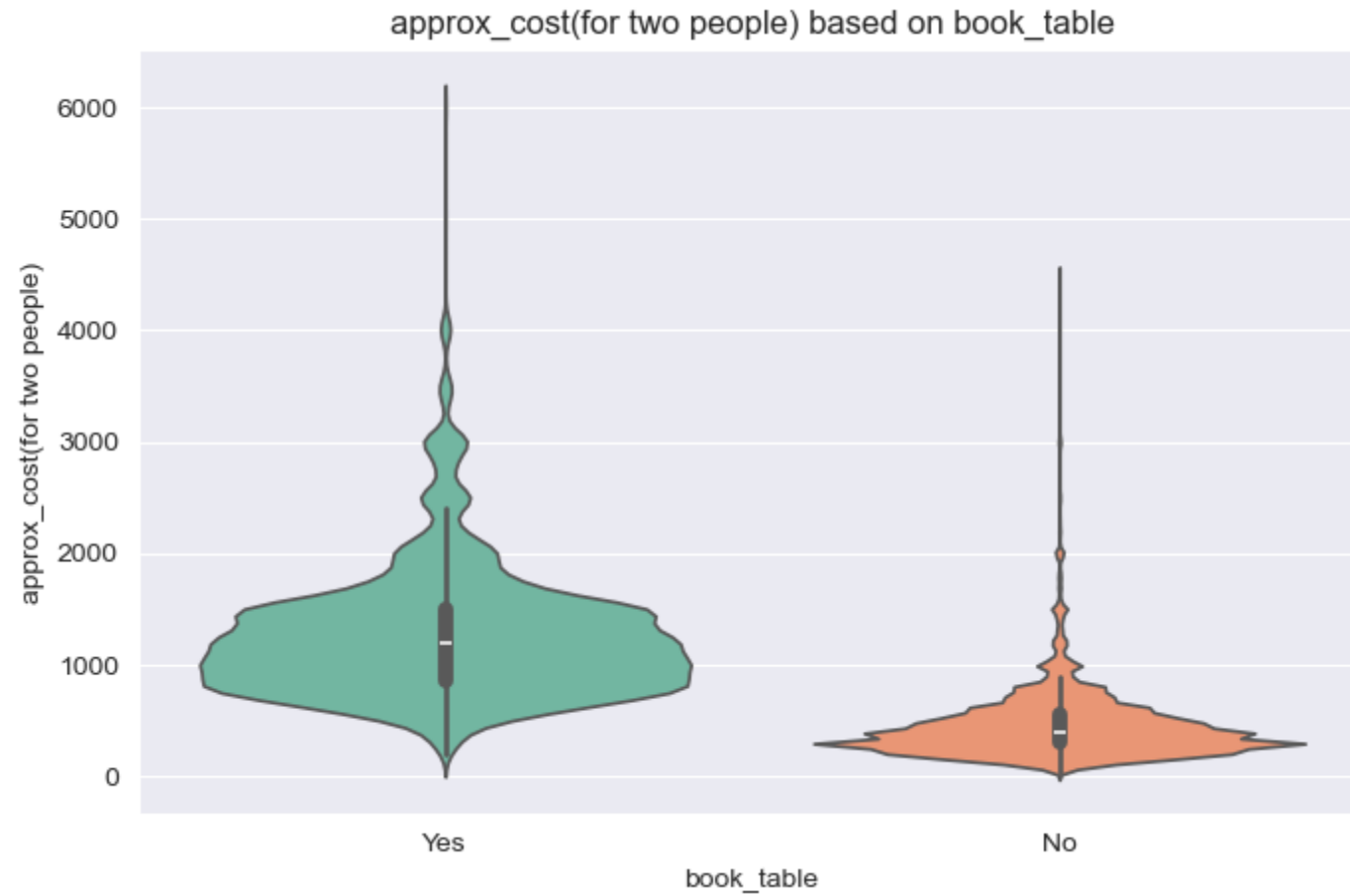
Drinks & Nightlife type Restaurant get the highest Votes cause they taked the online order & second position winner is Pubs and Bars type restaurants on zomato platform.



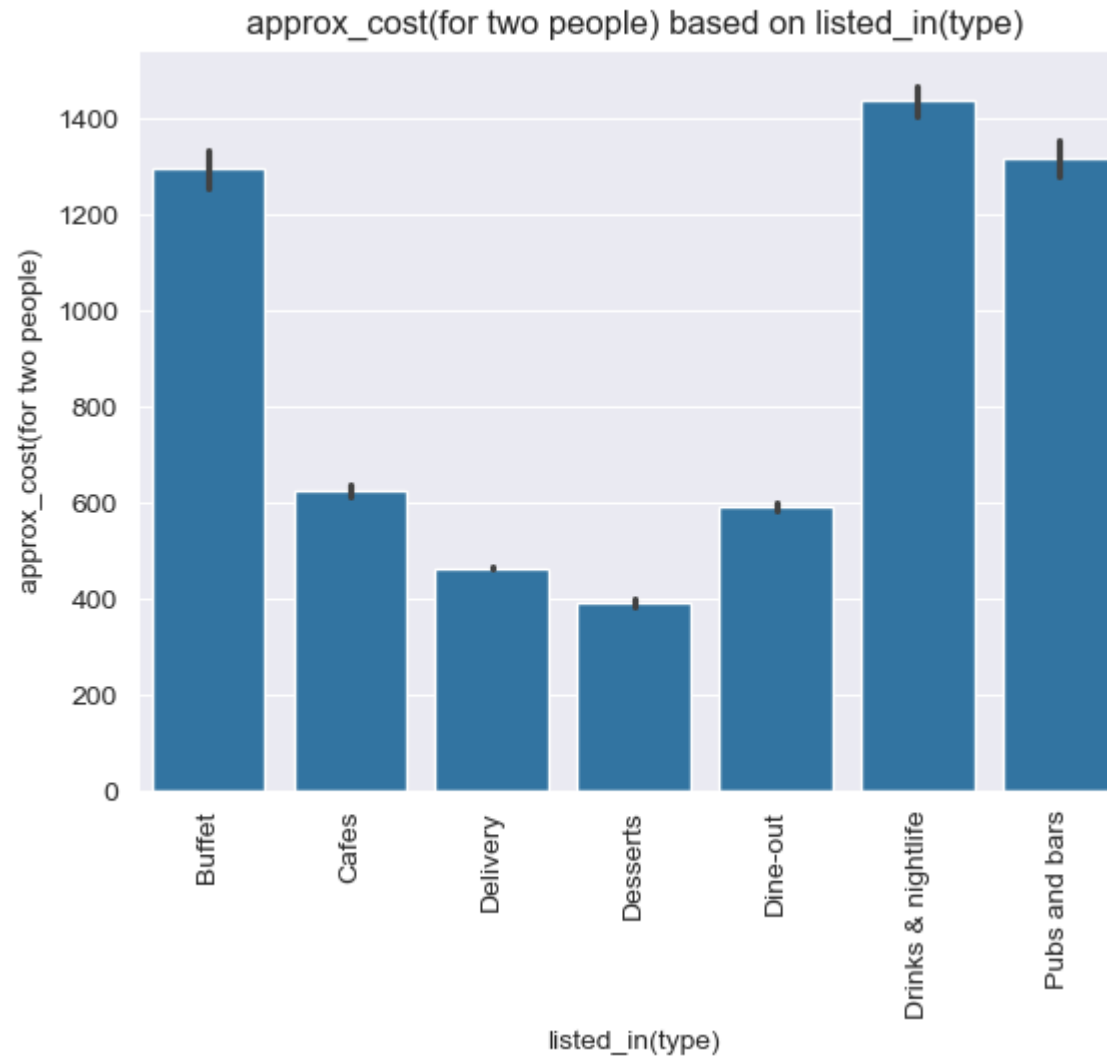
In the above chart you can see Lavelle Road, MG Road & Residency Road side restaurants get the high votes cause they take the online order but the opposite side interesting thing is Indiranagar side restaurants get the higher votes even not the online order.



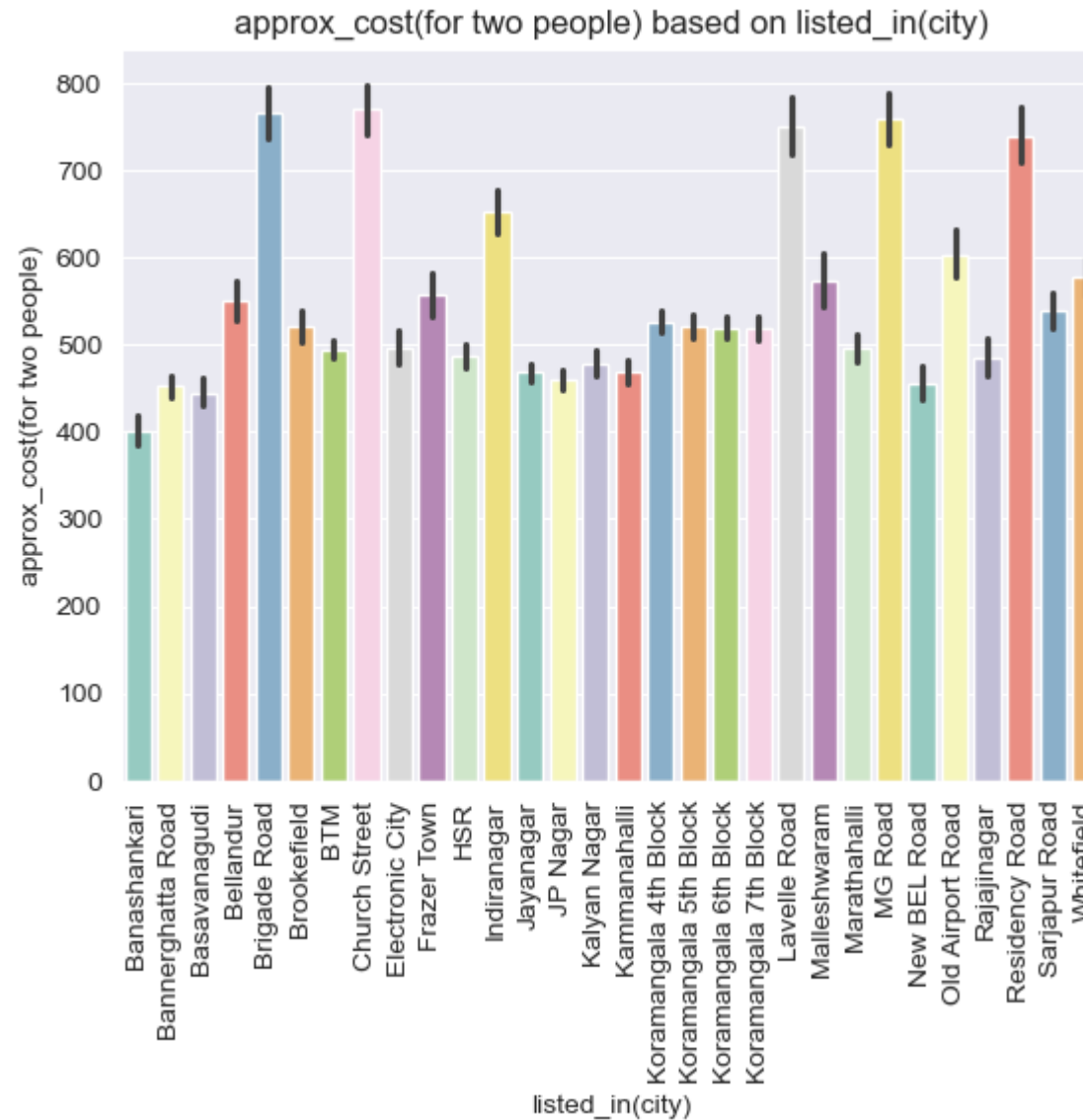
In the above chart you can see the average cost for two people is very high for those who are not placed the order online and that's why they did not get any offeres.



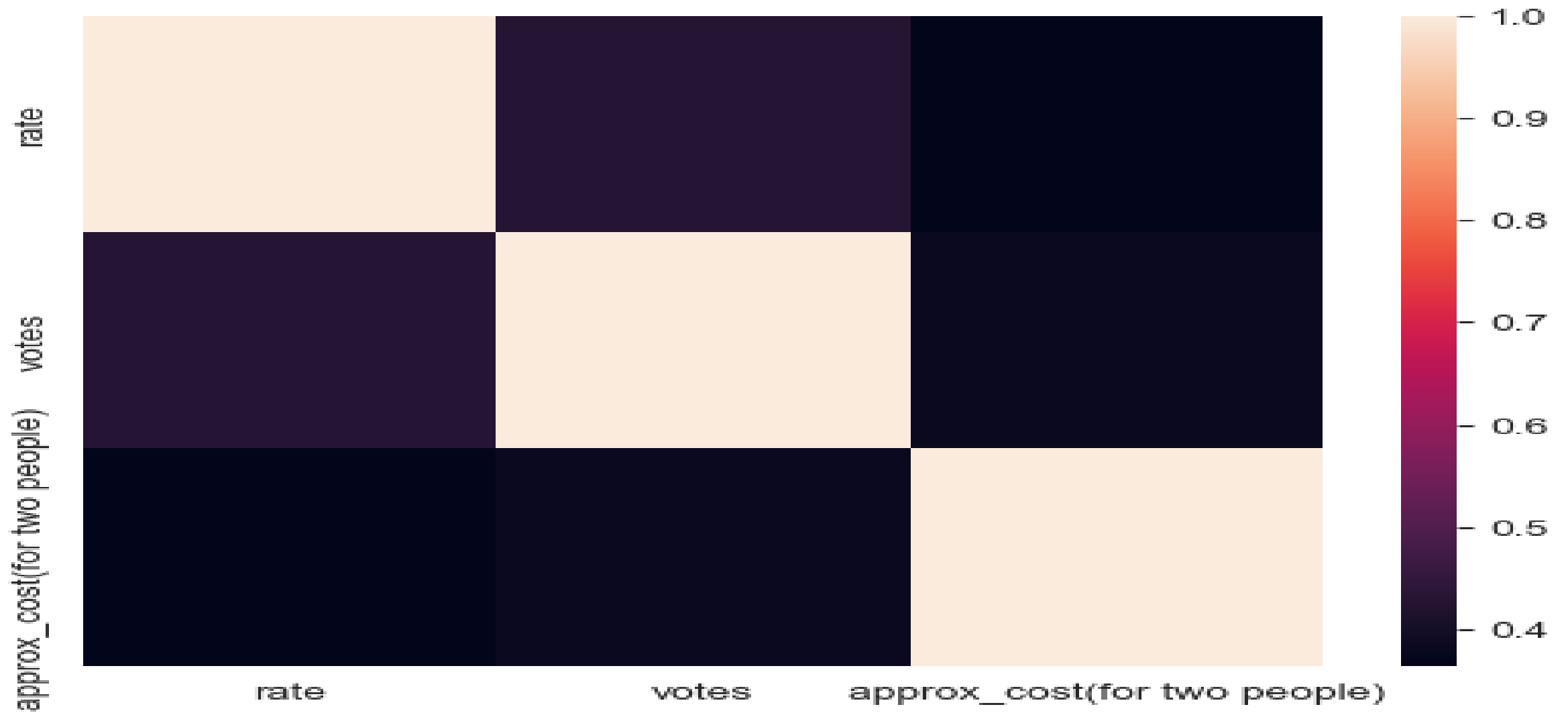
In above chart as you can see those who booked the table and pay the higher approx cost for two peoples.



Obeviosely In the Drinks & Nighlife type restaurants Customers payed the approx cost for two people is very high.



In above chart you can see the Brigade Road,Church Street,Lavelle Road,MG Road,Residency Road side Restaurants Regular/Visited Customer paid a approx cost for two people above 700.



In above Heatmap you can see the not much strong relationships between the two numerical variables.

"Zomato Restaurant Dataset – Key Insights & Recommendations":-

- 1."Online order is 'Yes'" → Gets "higher votes" and "higher ratings".**
- 2."Book table is 'Yes'" → Gets "higher votes", "higher ratings", and "higher average cost" for two people.**
- 3."listed_in(type) = 'Drinks & Nightlife'" → Gets "higher votes", "higher ratings", and "higher average cost".**
- 4."Listed cities like 'Indiranagar' and 'Old Airport Road'" → Have "higher votes", "ratings", and "cost for two people".**
- 5.If "online order = No", then "approx cost for two people is high" (indicating fine-dine or premium restaurants).**
- 6."MG Road" and "Residency Road" → Show "higher online orders" and also get "higher votes".**
- 7.Locations like "Brigade Road", "Church Street", "Lavella Road", "MG Road", and "Residency Road" → Show "higher cost for two people" (Premium dining).**
- 8."Most restaurants have below 7500 votes" — only few exceed this range, indicating top performers.**
- 9.For "restaurant types like 'Drinks & Nightlife', 'Pubs & Bars'", if "book_table = Yes", then "votes are high".**
- 10.In "Bellandur" and "Sarjapur" areas → Votes are "higher" because "book_table = Yes" is more common.**

"Zomato Restaurant Dataset – Key Insights & Recommendations":-

11.Restaurants with "online order enabled" → Tend to get "better ratings".

12.If "table booking = Yes", then "ratings > 3.0" are common; If "table booking = No", most restaurants have ratings "above 2.0", but a few fall "below 2.0".

13.Here,you can see the 'Chicken','Biryani',"Pizza" is a most frequent words while placed order or you can say Most Favourite dishes of the Customers.

14You can see clearly the winner is in the category of cuisines is 'North'.

15 You can see clearly the most favourite menu item is 'Chicken'.

16You can see the most of the placed order location is 'Koramangala',& second position is 'BTM'.

17)Most of the order placed location is 'Bangalore' cause whole the dataset belongs to Bangalore.[1](#)

Thank You