ZOMATO

STRATEGIC GOAL

*Improve customer experience.

*Focused item promotions.

*Business opportunities for expansion.

OBJECTIVE

*Recommend top ranked restaurants to the customer

*Recommend the restaurants to the customer for their menu choices

*Analysis of menuitems across various location and recommend less/more competitive location for promotion of menu items

*Recommend the location needs on menu item choices, restarurant choices, ... etc

User and Data Customers

Customers, Restaurant Managers, Business unit Heads

Explaratory Data Analysis

Field Definition

Each row represents the restaurant of Bangalore city in which it shows the following information about the restaurants by column wise. Given below,

URL- Website of restaurants.

Address- Address of restaurants.

Name - Name of the restaurants.

Online_order - Online_order facility present or not.

Book_table - Booking table facility present or not.

Rate - Rating of the restaurants by people.

Votes - Votes given by people for the restaurants.

Phone - Contact number of the restaurants.

Location - Location of the restaurants.

Rest_type - Type of restaurants.

Dish_liked - Dishes which are best in the restaurants.

Cuisines - Dishes by area wise whether it is north indian or sounth indian etc,.

Approx_cost(for two people) - Approximate cost for two people in the restaurants.

Reviews_list - Reviews given by the people for the restaurants.

Menu_items - Dishes present in the restaurants.

Listed_in(type) - Type of Dish present in the restaurants.

Listed_in(city) - City of the restaurant.

Data Cleanup

Identified the fields that are not needed for the analysis to meet the objectives. Removed Fields: URL, Address, Phone, Review list and Menu items.

Removed duplicate restaurant records. Removed empty records.

After Cleanup

IMPORTING PACKAGES

In [40]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use('ggplot')
```

IMPORTING DATASET

In [41]:
```
df = pd.read_excel('Downloads\zom.xlsx')
```

In [42]:
```
df.isnull().sum()
```

Out[42]:
```
Unnamed: 0                      0
address                         0
name                            0
online_order                    0
book_table                      0
rate                            0
location                        0
rest_type                       0
dish_liked                      0
cuisines                        0
approx_cost(for two people)     0
meal_type                       0
city                            0
votes                           0
dtype: int64
```

Describing the data

In [43]:
```
df.describe()
```

Out[43]:

	Unnamed: 0	approx_cost(for two people)	votes
count	19233.000000	19233.000000	19233.000000
mean	9616.000000	708.809858	420.982582
std	5552.233199	503.524302	699.775971
min	0.000000	40.000000	0.000000
25%	4808.000000	400.000000	94.000000
50%	9616.000000	600.000000	199.000000
75%	14424.000000	800.000000	462.000000
max	19232.000000	6000.000000	16345.000000

In [44]:
```
df.info
```

Out[44]:
```
<bound method DataFrame.info of       Unnamed: 0
address  \
0              0  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
1              1  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
2              2  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
3              3  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
4              4  #10061 Tower 10, Prestige Shanti Niketan., Whi...
...          ...                                                ...
19228      19228  Zone By The Park, 26/A, Electronic City, Banga...
19229      19229  Zone by The Park, 26/A, Phase 1, Electronic Ci...
19230      19230  Zone by The Park, 26/A, Phase 1, Electronic Ci...
19231      19231  c/o Foodism, 2nd A Main Road, 445/31, 7th Cros...
19232      19232  c/o Foodism, 2nd A Main Road, 445/31, 7th Cros...

                           name online_order book_table   rate  \
0                       Vinny's          Yes         No  4.1/5
1                       Vinny's          Yes         No  4.1/5
2                       Vinny's          Yes         No  4.1/5
3                       Vinny's          Yes         No  4.1/5
4                 Cupcake Couture           No         No  3.8/5
...                           ...          ...        ...    ...
19228                      Z-Bar           No        Yes  4.1/5
19229   Bazaar - Zone By The Park           No        Yes  4.0/5
19230   Bazaar - Zone By The Park           No        Yes  4.0/5
19231  Artinci Artisanal Ice Cream          Yes         No  4.2/5
19232  Artinci Artisanal Ice Cream          Yes         No  4.2/5

             location       rest_type  \
0              Domlur   Casual Dining
1              Domlur   Casual Dining
2              Domlur   Casual Dining
3              Domlur   Casual Dining
4          Whitefield        Delivery
...               ...             ...
19228  Electronic City             Bar
19229  Electronic City   Casual Dining
19230  Electronic City   Casual Dining
19231     Marathahalli  Dessert Parlor
19232     Marathahalli  Dessert Parlor

                                            dish_liked  \
0      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
1      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
2      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
3      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
4                          Cup Cake, Chocolate Truffle
...                                                ...
19228  Cocktails, Tandoori Chicken, Wine, Ice Tea, Fr...
19229  Salads, Chicken Tikka, Roti, Gulab Jamun, Lunc...
19230  Salads, Chicken Tikka, Roti, Gulab Jamun, Lunc...
19231        Coffee, Vanilla Ice Cream, Almond Ice Cream
19232        Coffee, Vanilla Ice Cream, Almond Ice Cream

                                          cuisines  \
0             Burger, Desserts, Italian, Pizza, Salad
```

```
1              Burger, Desserts, Italian, Pizza, Salad
2              Burger, Desserts, Italian, Pizza, Salad
3              Burger, Desserts, Italian, Pizza, Salad
4                                             Desserts
...                                                ...
19228                         Finger Food, North Indian
19229  Continental, South Indian, North Indian, Asian
19230  Continental, South Indian, North Indian, Asian
19231                              Desserts, Ice Cream
19232                              Desserts, Ice Cream

       approx_cost(for two people)      meal_type                   city  \
0                              600       Delivery  Koramangala 6th Block
1                              600       Delivery       Old Airport Road
2                              600       Dine-out            Indiranagar
3                              600       Dine-out       Old Airport Road
4                              300       Delivery             Whitefield
...                            ...            ...                    ...
19228                         1500  Pubs and bars        Electronic City
19229                         1400         Buffet        Electronic City
19230                         1400       Dine-out        Electronic City
19231                          300       Delivery           Marathahalli
19232                          300       Desserts           Marathahalli

       votes
0        366
1        367
2        360
3        366
4         40
...      ...
19228    410
19229    415
19230    415
19231     44
19232     44

[19233 rows x 14 columns]>
```

Columns present in the Data

In [45]:
```
df.columns
```

Out[45]:
```
Index(['Unnamed: 0', 'address', 'name', 'online_order', 'book_table', 'rate',
       'location', 'rest_type', 'dish_liked', 'cuisines',
       'approx_cost(for two people)', 'meal_type', 'city', 'votes'],
      dtype='object')
```

Shape of the Data

In [46]:
```
df.shape
```

Out[46]:
```
(19233, 14)
```

Count of the Data

In [47]:
```
df.count
```

Out[47]:
```
<bound method DataFrame.count of       Unnamed: 0
address  \
0              0  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
1              1  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
2              2  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
3              3  # 31, 7th Cross, Opposite Canara Bank, Domlur,...
4              4  #10061 Tower 10, Prestige Shanti Niketan., Whi...
...          ...                                                ...
19228      19228  Zone By The Park, 26/A, Electronic City, Banga...
19229      19229  Zone by The Park, 26/A, Phase 1, Electronic Ci...
19230      19230  Zone by The Park, 26/A, Phase 1, Electronic Ci...
19231      19231  c/o Foodism, 2nd A Main Road, 445/31, 7th Cros...
19232      19232  c/o Foodism, 2nd A Main Road, 445/31, 7th Cros...

                           name online_order book_table   rate  \
0                       Vinny's          Yes         No  4.1/5
1                       Vinny's          Yes         No  4.1/5
2                       Vinny's          Yes         No  4.1/5
3                       Vinny's          Yes         No  4.1/5
4                 Cupcake Couture           No         No  3.8/5
...                           ...          ...        ...    ...
19228                      Z-Bar           No        Yes  4.1/5
19229   Bazaar - Zone By The Park           No        Yes  4.0/5
19230   Bazaar - Zone By The Park           No        Yes  4.0/5
19231  Artinci Artisanal Ice Cream          Yes         No  4.2/5
19232  Artinci Artisanal Ice Cream          Yes         No  4.2/5

             location       rest_type  \
0              Domlur   Casual Dining
1              Domlur   Casual Dining
2              Domlur   Casual Dining
3              Domlur   Casual Dining
4          Whitefield        Delivery
...               ...             ...
19228  Electronic City             Bar
19229  Electronic City   Casual Dining
19230  Electronic City   Casual Dining
19231     Marathahalli  Dessert Parlor
19232     Marathahalli  Dessert Parlor

                                            dish_liked  \
0      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
1      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
2      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
3      Pizza, Pasta, Mocktails, Tiramisu, Garlic Brea...
4                          Cup Cake, Chocolate Truffle
...                                                ...
19228  Cocktails, Tandoori Chicken, Wine, Ice Tea, Fr...
19229  Salads, Chicken Tikka, Roti, Gulab Jamun, Lunc...
19230  Salads, Chicken Tikka, Roti, Gulab Jamun, Lunc...
19231        Coffee, Vanilla Ice Cream, Almond Ice Cream
19232        Coffee, Vanilla Ice Cream, Almond Ice Cream

                                          cuisines  \
0             Burger, Desserts, Italian, Pizza, Salad
```

```
1             Burger, Desserts, Italian, Pizza, Salad
2             Burger, Desserts, Italian, Pizza, Salad
3             Burger, Desserts, Italian, Pizza, Salad
4                                           Desserts
...                                              ...
19228                        Finger Food, North Indian
19229  Continental, South Indian, North Indian, Asian
19230  Continental, South Indian, North Indian, Asian
19231                             Desserts, Ice Cream
19232                             Desserts, Ice Cream

       approx_cost(for two people)      meal_type                   city  \
0                              600       Delivery  Koramangala 6th Block
1                              600       Delivery       Old Airport Road
2                              600       Dine-out            Indiranagar
3                              600       Dine-out       Old Airport Road
4                              300       Delivery             Whitefield
...                            ...            ...                    ...
19228                         1500  Pubs and bars        Electronic City
19229                         1400         Buffet        Electronic City
19230                         1400       Dine-out        Electronic City
19231                          300       Delivery           Marathahalli
19232                          300       Desserts           Marathahalli

       votes
0        366
1        367
2        360
3        366
4         40
...      ...
19228    410
19229    415
19230    415
19231     44
19232     44

[19233 rows x 14 columns]>
```

Datatype of each column

In [48]:
```
df.dtypes
```

Out[48]:
```
Unnamed: 0                      int64
address                        object
name                           object
online_order                   object
book_table                     object
rate                           object
location                       object
rest_type                      object
dish_liked                     object
cuisines                       object
approx_cost(for two people)     int64
meal_type                      object
city                           object
votes                           int64
dtype: object
```

Count of the ratings

Removing the null values present in the rating column.

In [49]:
```
df['rate'].value_counts()
df['rate'].isnull().sum()
df['rate'] = df['rate'].apply(lambda x: str(x).split('/')[0])
df['rate'].value_counts()
```

Out[49]:
```
3.9    3012
4.0    2629
4.1    2463
3.8    2288
4.2    1698
3.7    1448
4.3    1148
3.6     762
4.4     598
3.5     424
3.4     320
2.8     286
2.9     261
4.5     259
3.1     237
3.0     236
3.2     209
3.3     190
2.7     162
NEW     135
4.6     119
2.6     112
4.7      53
2.5      44
2.4      44
2.3      33
2.2      23
2.1      11
2.0      11
4.8      10
4.9       8
Name: rate, dtype: int64
```

In [50]:
```
g23 = list(df['rate'])
for i in range(0, len(g23)):
    if g23[i] == 'nan':
        g23[i] = 'unrated'
    elif g23[i] == '-':
        g23[i] = 'unrated'
    elif g23[i] == 'NEW':
        g23[i] = 'unrated'

df['rate'] = g23

df['rate'].value_counts()

g23 = list(df['rate'])
for i in range(0, len(g23)):
    if g23[i] == 'unrated':
        g23[i] = None
    else :
        g23[i] = float(g23[i])
df['rate'] = g23
df['rate'].value_counts()
```

Out[50]:
```
3.9    3012
4.0    2629
4.1    2463
3.8    2288
4.2    1698
3.7    1448
4.3    1148
3.6     762
4.4     598
3.5     424
3.4     320
2.8     286
2.9     261
4.5     259
3.1     237
3.0     236
3.2     209
3.3     190
2.7     162
4.6     119
2.6     112
4.7      53
2.5      44
2.4      44
2.3      33
2.2      23
2.0      11
2.1      11
4.8      10
4.9       8
Name: rate, dtype: int64
```

Count of city

In [51]:
```
df['city'].value_counts()
```

Out[51]:
```
BTM                    1225
Koramangala 7th Block  1132
Koramangala 4th Block  1122
Koramangala 5th Block  1047
Koramangala 6th Block   979
Jayanagar               925
Indiranagar             879
Church Street           802
Brigade Road            779
MG Road                 734
JP Nagar                689
Lavelle Road            684
Residency Road          625
HSR                     611
Old Airport Road        599
Basavanagudi            562
Brookefield             522
Whitefield              517
Marathahalli            496
Bannerghatta Road       482
Frazer Town             477
Kalyan Nagar            453
Malleshwaram            436
Kammanahalli            434
Bellandur               424
Sarjapur Road           404
Banashankari            347
Rajajinagar             314
Electronic City         288
New BEL Road            245
Name: city, dtype: int64
```

In [52]:
```
df['location'].isnull().sum()
```

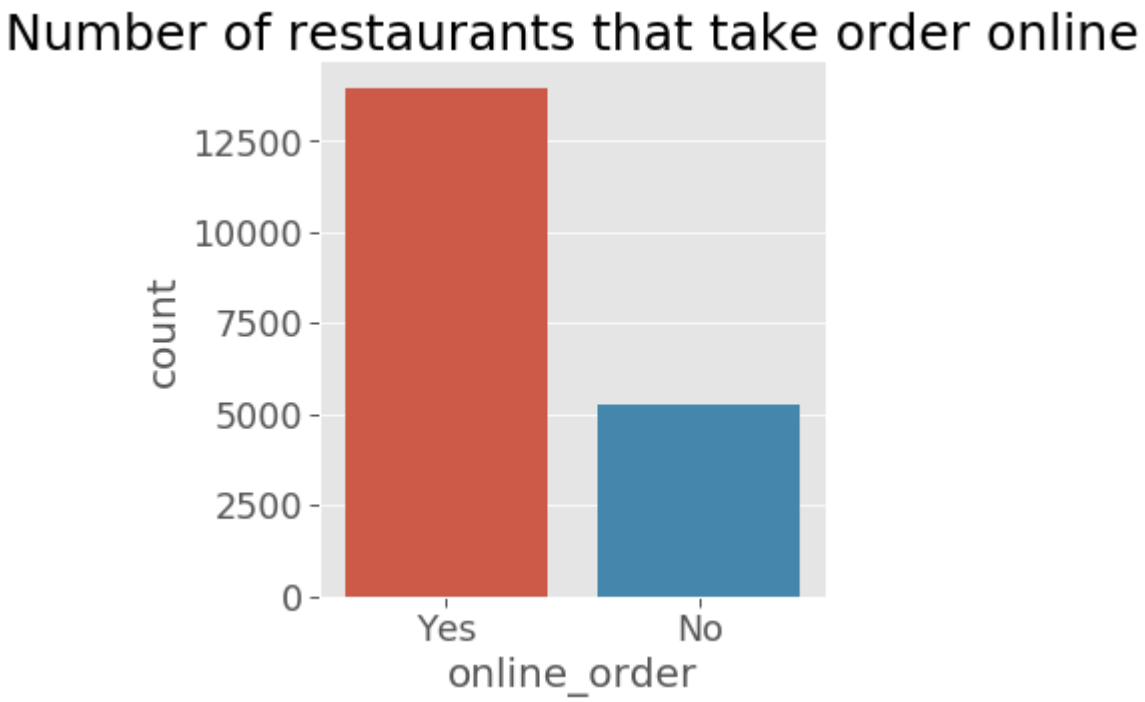
Out[52]:
```
0
```

In [53]:
```
len(df['city'].value_counts())
```

Out[53]:
```
30
```

UNIVARIATE ANALYSIS

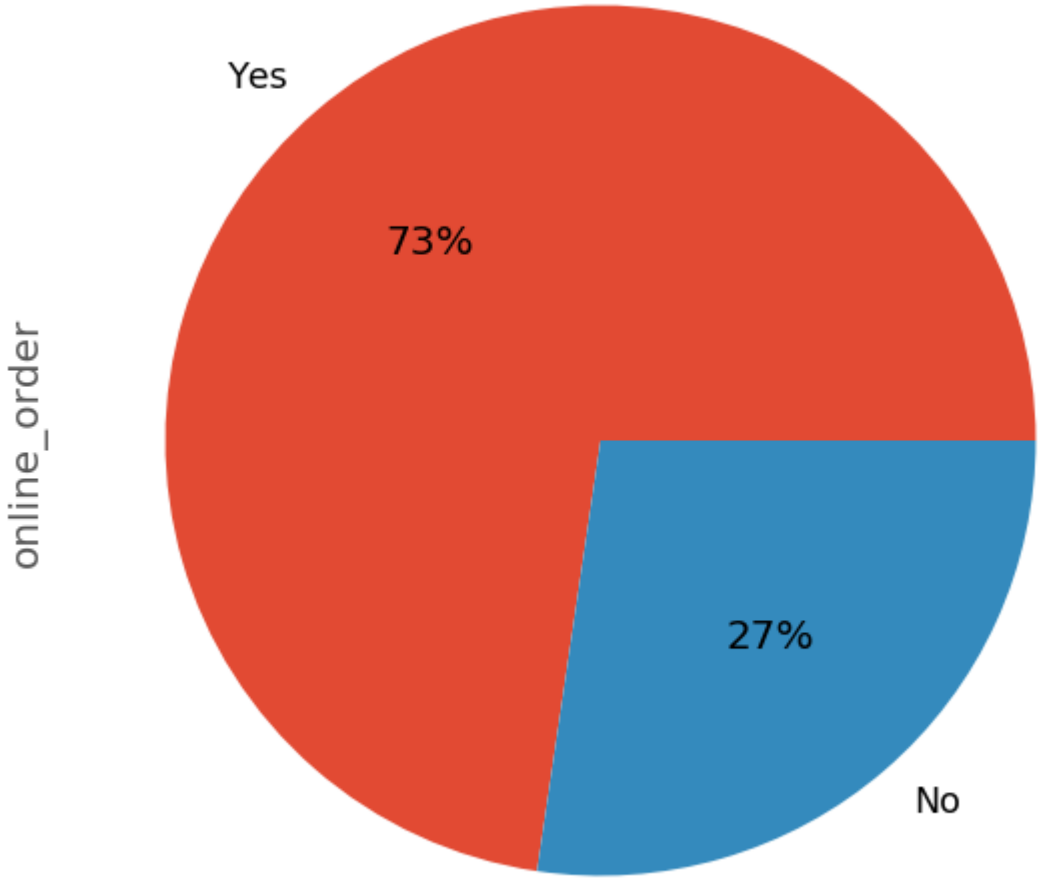
In [54]:
```
# plotting number of restaurtants that take online_order
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
sns.catplot(data = df, kind = 'count', x = 'online_order')
plt.title('Number of restaurants that take order online')
plt.show()
```



Type *Markdown* and LaTeX: α^2

In [55]:
```
df['online_order'].value_counts().plot(kind='pie', figsize=(10,10),autopct='%1.0
```

Out[55]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e418378208>
```

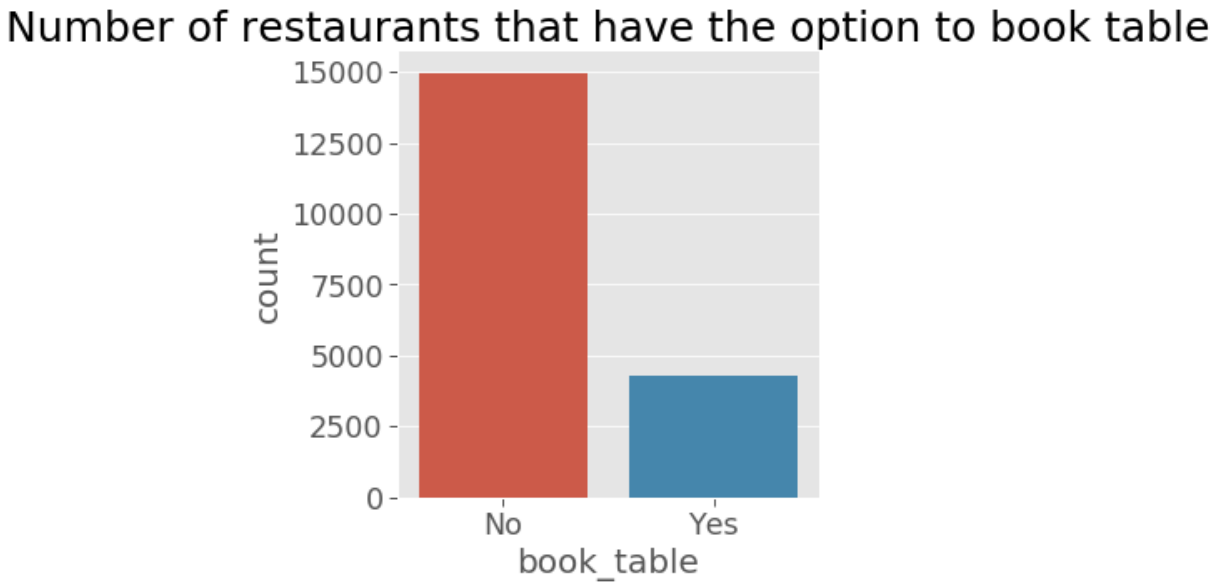


Conclusion

The IT field has the most value in developing an application for the hotels which are not taking online orders.

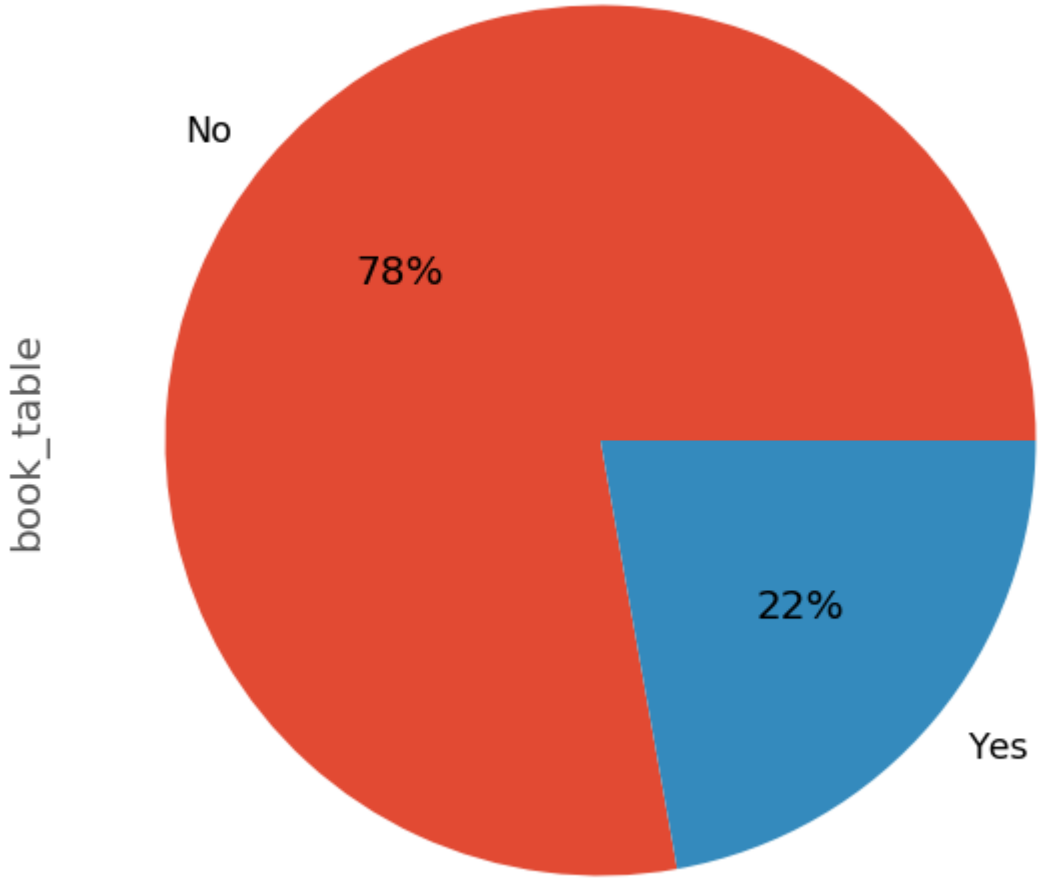
and mostly Bangalore people takes online order.

In [56]:
```
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20, "axes.titlesize"
sns.catplot(data = df, kind = 'count', x = 'book_table')
plt.title('Number of restaurants that have the option to book table')
plt.show()
```



In [57]:
```
df['book_table'].value_counts().plot(kind='pie', figsize=(10,10),autopct='%1.0f%%
```

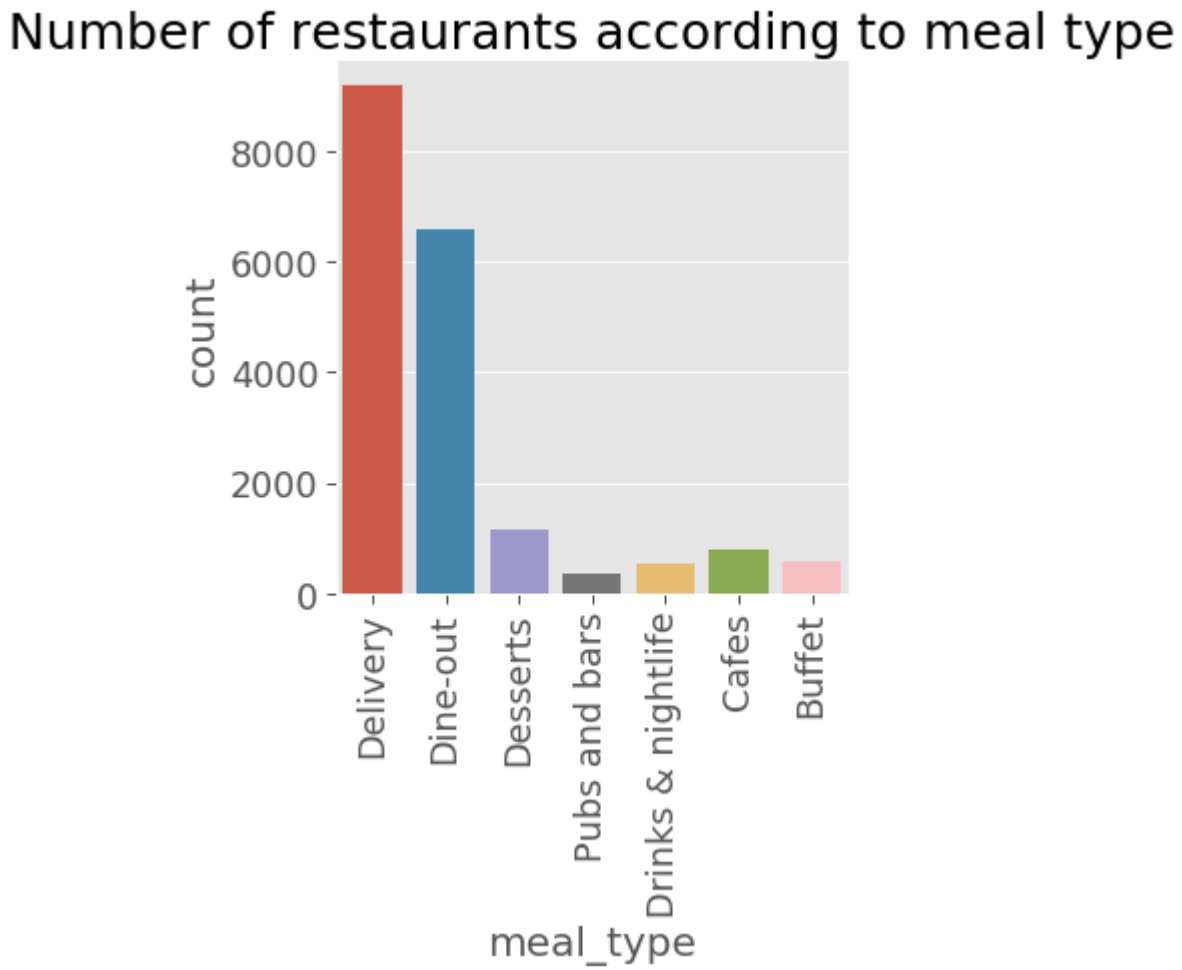
Out[57]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e415de1a48>
```



Conclusion

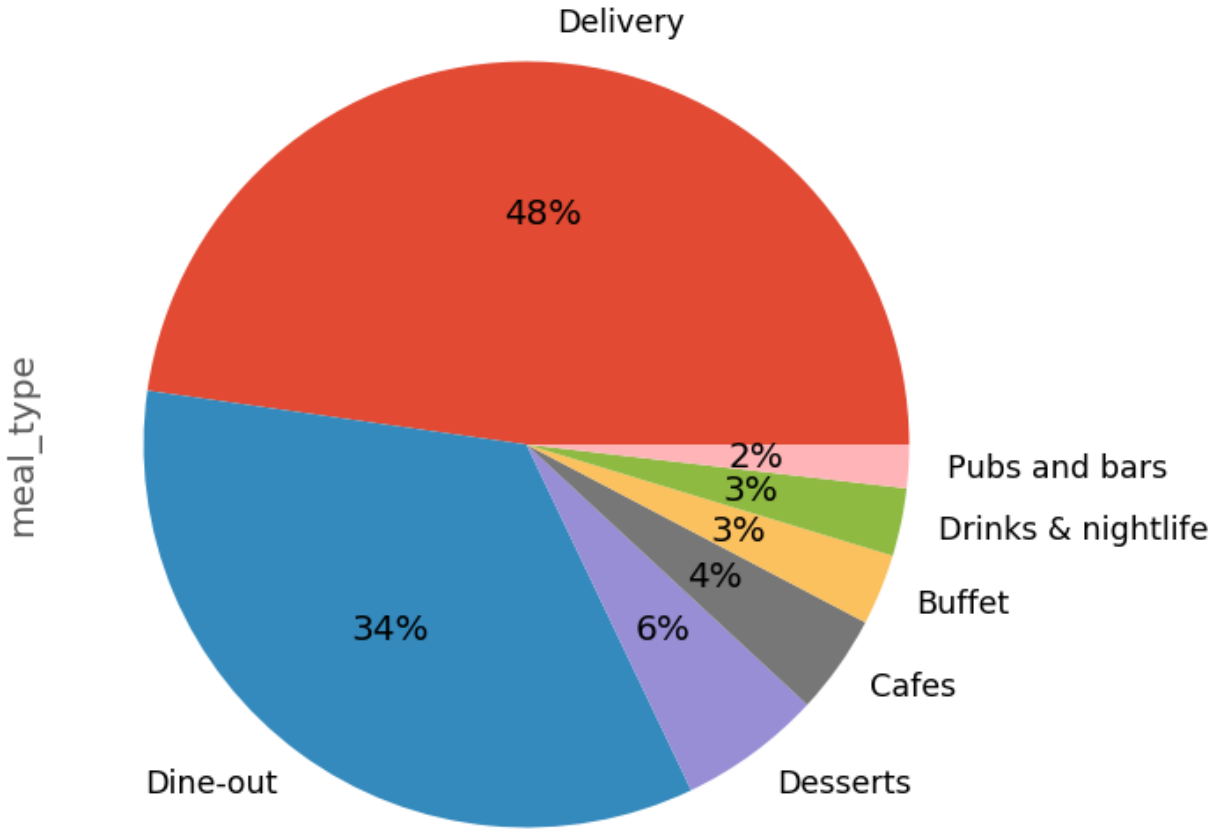
We see that very high number of restaurants does not offer the service to book table, and very less do.This means that people at Bangalore prefer to eat at their homes or prefer fast food(snacks — quick bites)

In [58]:
```
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
sns.catplot(data = df, kind = 'count', x = 'meal_type')
plt.title('Number of restaurants according to meal type')
plt.xticks(rotation=90)
plt.show()
```



In [59]:
```
df['meal_type'].value_counts().plot(kind='pie', figsize=(10,10),autopct='%1.0f%%
```

Out[59]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e4198e2908>
```



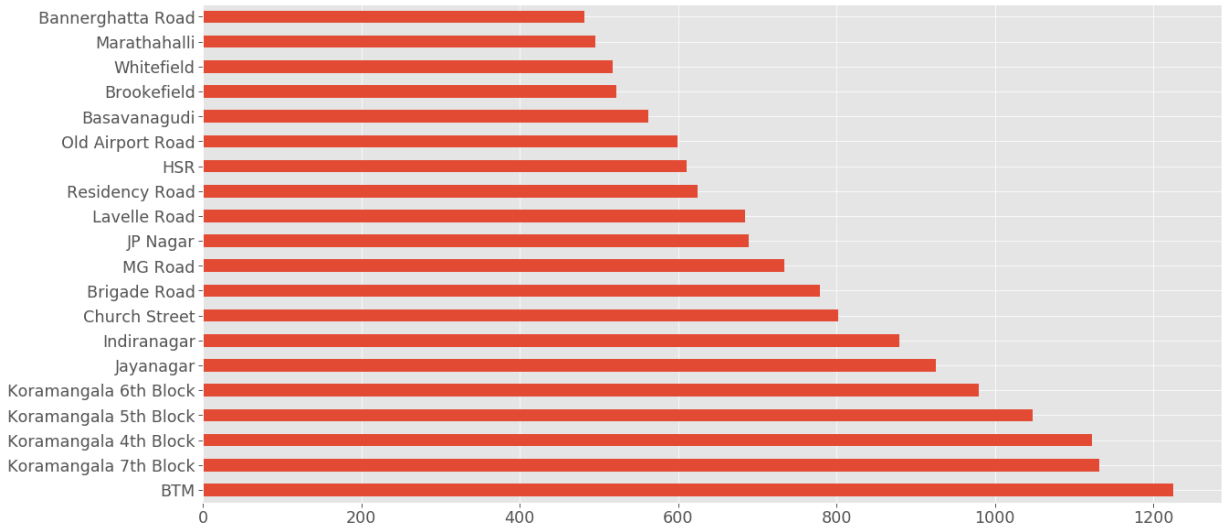
Conclusion

***We see that the majority of restaurants offer delivery, which might account for the above conclusion that people at Bangalore prefer to eat at home.**

***Pubs and Bars and Buffet are offered by very less number of restaurants.**

In [60]:
```
df['city'].value_counts().head(20).plot(kind='barh', figsize=(20,10))
```

Out[60]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e419812a48>
```

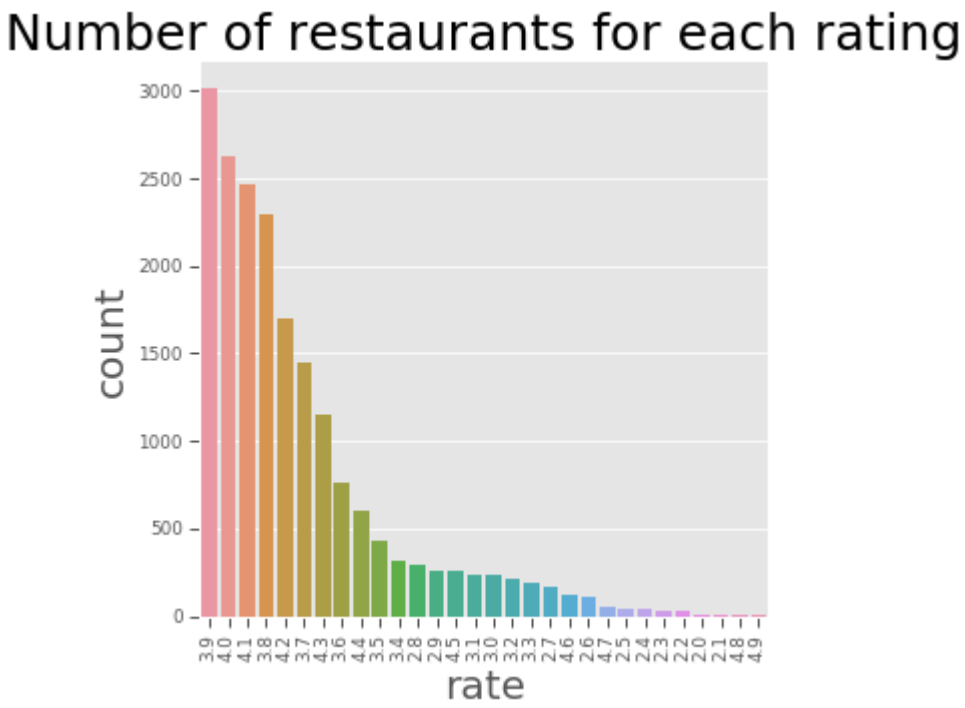


Conlcusion

***BTM has the highest number of restaurants, followed by Koramangala 7th Block.New BEL Road has the least number of restaurants, followed by Banashankari.**

***It seems that the main foodies live in BTM and Koramangala.**

In [61]:
```
sns.set_context("paper", font_scale = 1, rc = {"font.size": 20,"axes.titlesize":
g23 = sns.catplot(data = df, kind = 'count', x = 'rate', order = df['rate'].value
plt.title('Number of restaurants for each rating')
g23.set_xticklabels(rotation = 90)
plt.show()
```



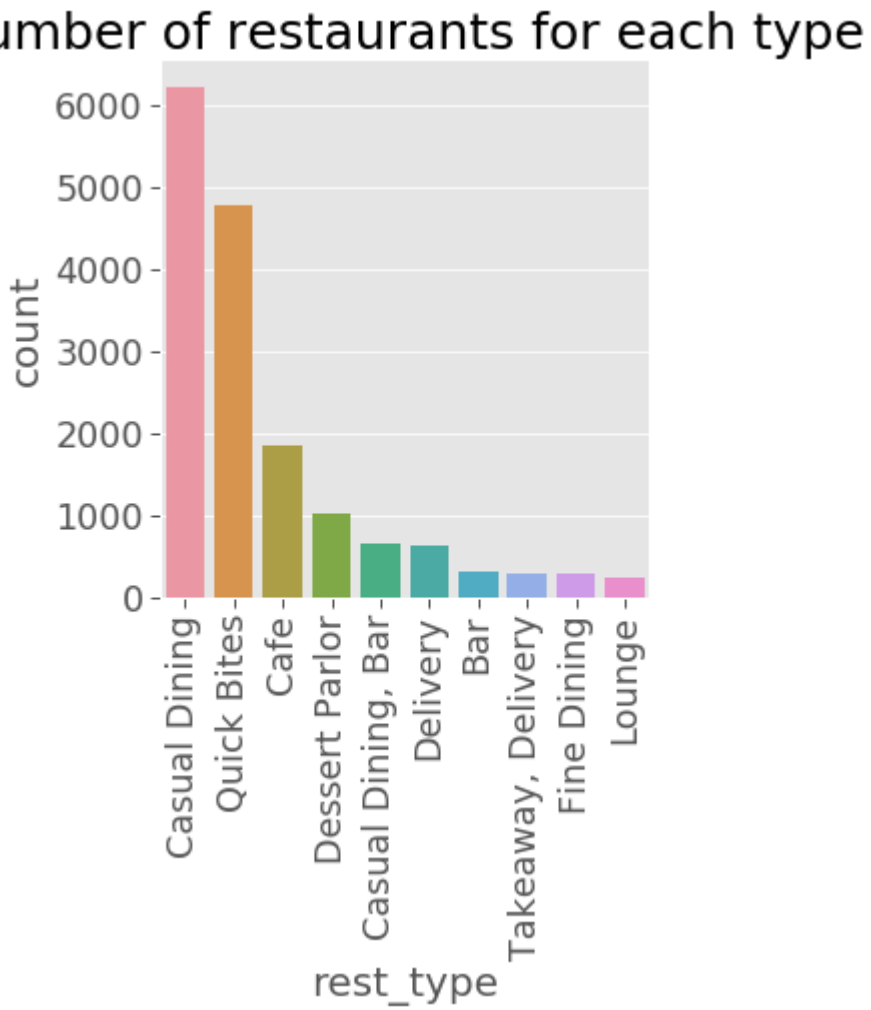
Conlcusion

***The majority of restaurants are rated 3.9(out of 5), followed by 3.8 and 3.7, which are decent ratings.**

***This means that most of the restaurants at Bangalore are liked by the citizens and hence rated above average**

***This shows that due to high rise in the number of restaurants in Bangalore, the competition is at its best, as every restaurant is trying to excel in quality and services to increase the number of customers, and hence enhance profit**

In [62]:
```
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
b = sns.catplot(data = df, kind = 'count', x = 'rest_type', order = df['rest_type
plt.title('Number of restaurants for each type')
b.set_xticklabels(rotation = 90)
plt.show()
```



Conlcusion

***We see that, the most number of restaurants are of the type casual dining.**

***From the above plot it is clear that people at Bangalore prefer ‘quick bites’ more often.**

***This supports our above conclusion, which was drawn from the count plot of book_table feature,that citizens of Bangalore prefer fast food-quick bites more often,this is why many restaurants does not offer table booking facility**

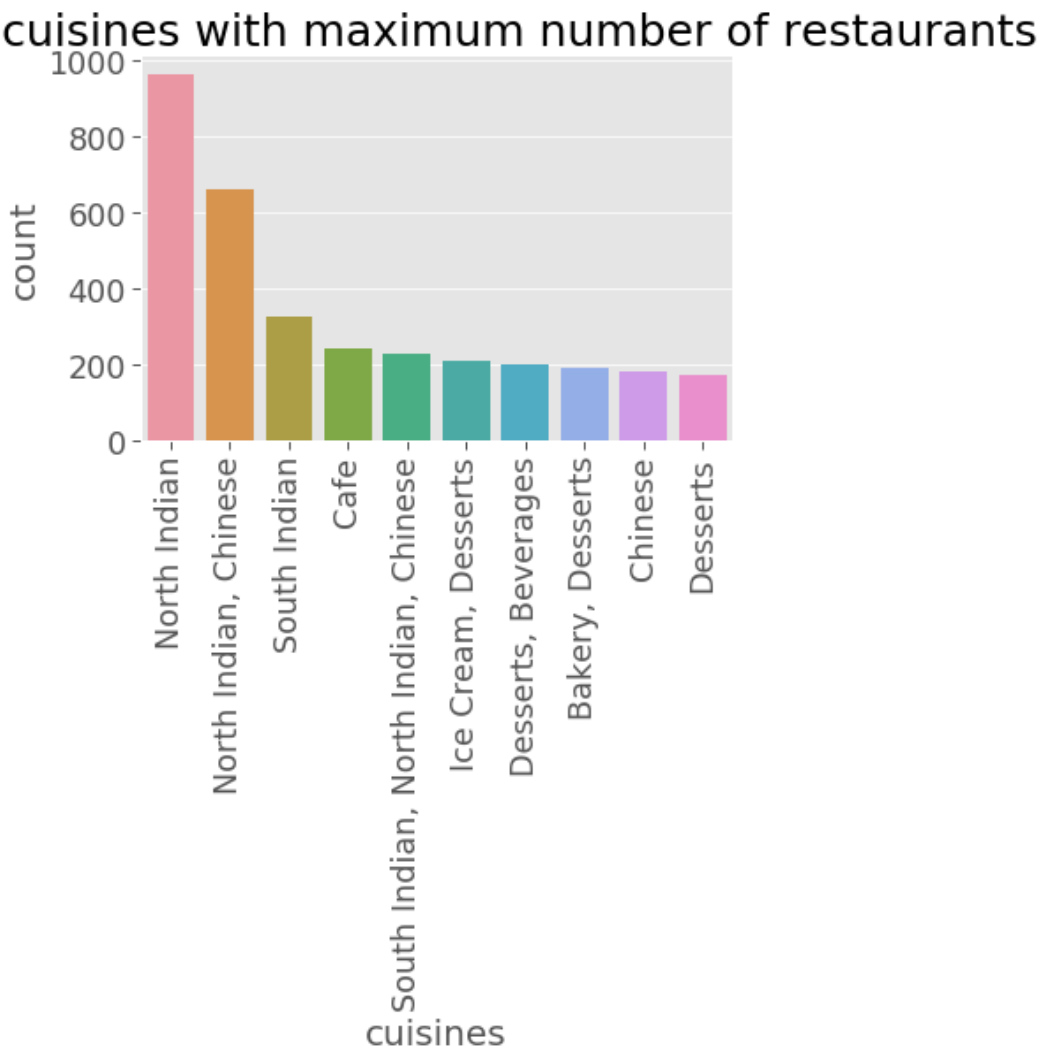
In [63]:
```
df['rest_type'].value_counts()
```

Out[63]:
```
Casual Dining               6221
Quick Bites                 4779
Cafe                        1839
Dessert Parlor              1005
Casual Dining, Bar           641
                            ...
Dessert Parlor, Kiosk          2
Microbrewery, Bar              2
Cafe, Lounge                   2
Dhaba                          1
Food Court, Casual Dining      1
Name: rest_type, Length: 73, dtype: int64
```

In [64]:
```
df['cuisines'].value_counts()
```

Out[64]:
```
North Indian                                                          964
North Indian, Chinese                                                 658
South Indian                                                          325
Cafe                                                                  243
South Indian, North Indian, Chinese                                   227
                                                                     ...
Chinese, Vietnamese, Thai, Malaysian                                    1
Coffee, South Indian                                                    1
Biryani, Andhra, North Indian, Seafood                                  1
Biryani, Kebab, Continental                                             1
Cafe, Healthy Food, North Indian, Biryani, Continental, Desserts        1
Name: cuisines, Length: 1602, dtype: int64
```

In [65]:
```
b = sns.countplot(data = df, x = 'cuisines', order = df.cuisines.value_counts().
b.set_xticklabels(b.get_xticklabels(),rotation = 90)
plt.title('Top 10 cuisines with maximum number of restaurants')
plt.show()
```

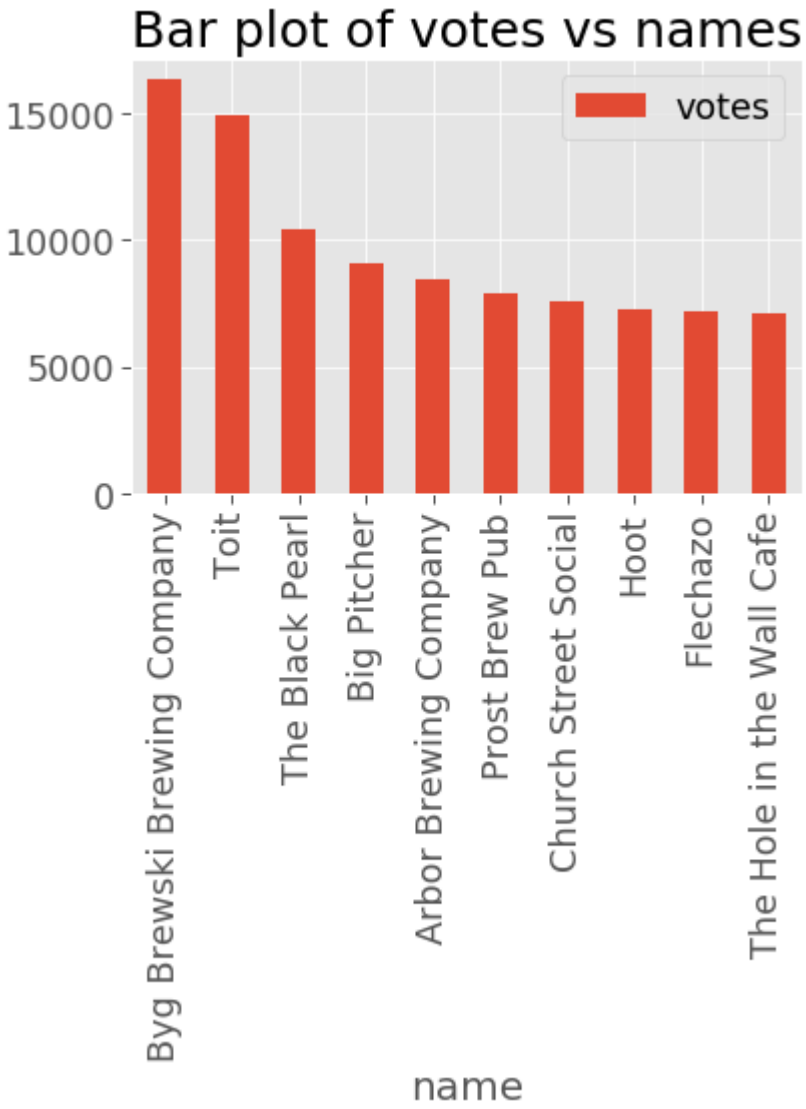


Conlcusion

*We see that North Indian cuisine is offered by most number of restaurants.

*This is worth discussing, as Bangalore is situated in South India and yet people there prefer to eat North Indian cuisine

In [66]:
```
df[['votes', 'name']].groupby(['name']).median().sort_values("votes", ascending =
plt.title('Bar plot of votes vs names')
plt.show()
```



Conlcusion

*We see that Byg Brewski Brewing Company restaurant has the maximum number of upvote.

*We see that this restaurant has outstanding rating, 4.9 out of 5.0.

*The restaurant is a Microbrewery type of restaurant, and offers the above mentioned cuisines.

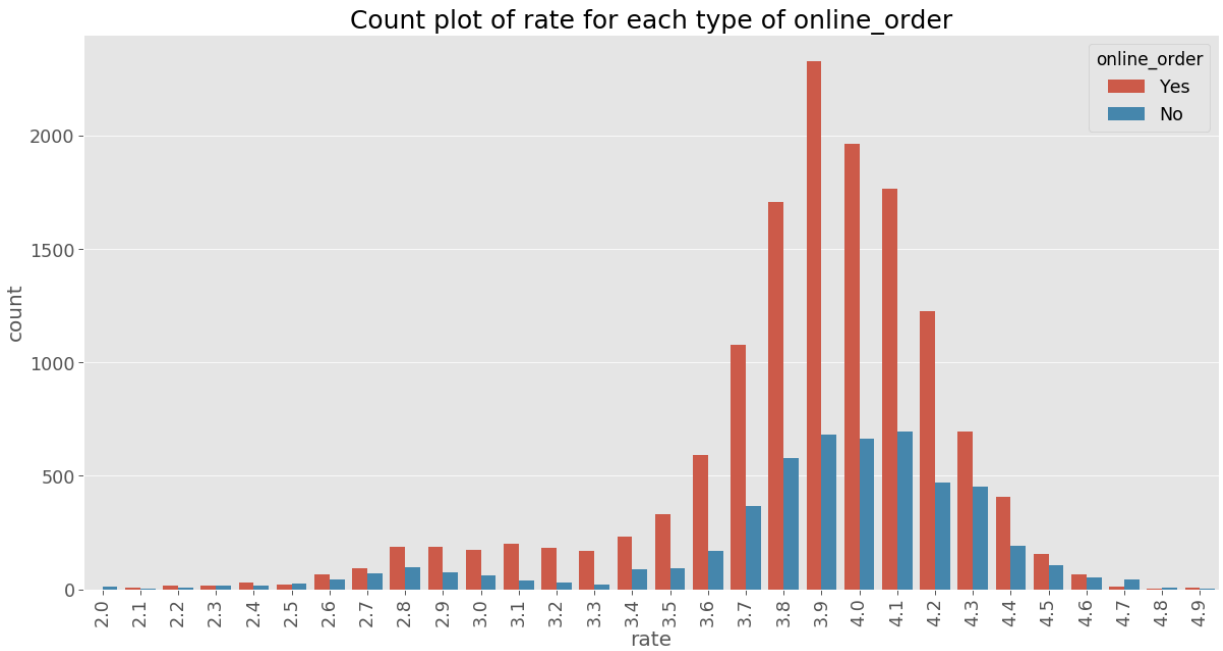
*The dishes liked by the people are also mentioned.

*We see that the meal type for different branches of the restaurant include Delivery, Dine-out and Drinks & nightlife,this means that Byg Brewski Brewing Company restaurant is one of those strong competitors in these types of meals.

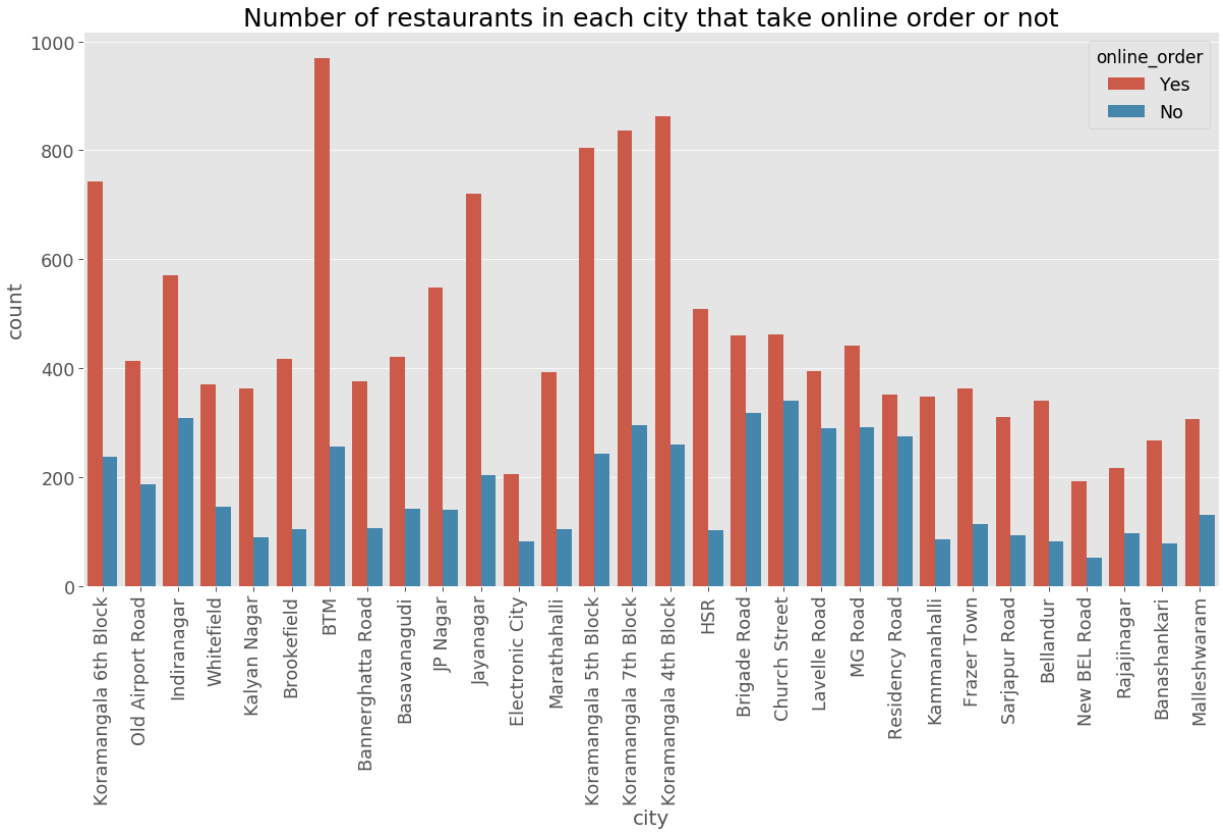
In [67]:
```
b=[]
a = df.iloc[:, :].values
for i in range(0, len(a)):
    if a[i][0] == 'Byg Brewski Brewing Company' or a[i][0] == 'Byg Brewski Brewi
        b.append(list(a[i]))
```

In [68]:
```
a = pd.DataFrame(df['rate'])
a['name'] = df['name']
a = a.dropna(axis = 0, how ='any')
a[['rate', 'name']].groupby(['name']).median().sort_values("rate", ascending = Fa
plt.title('Bar plot of rate vs names for 10 least rated restaurants')
plt.show()
```

In [69]:
```
plt.figure(figsize=(20,10))
b=sns.countplot(data = df, hue = 'online_order', x = 'rate')
plt.title('Count plot of rate for each type of online_order')
plt.xticks(rotation = 90)
plt.show()
```



In [70]:
```
plt.figure(figsize=(20,10))
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
b = sns.countplot(data = df, x = 'city', hue = 'online_order')
plt.title('Number of restaurants in each city that take online order or not')
b.set_xticklabels(b.get_xticklabels(),rotation = 90)
plt.show()
```

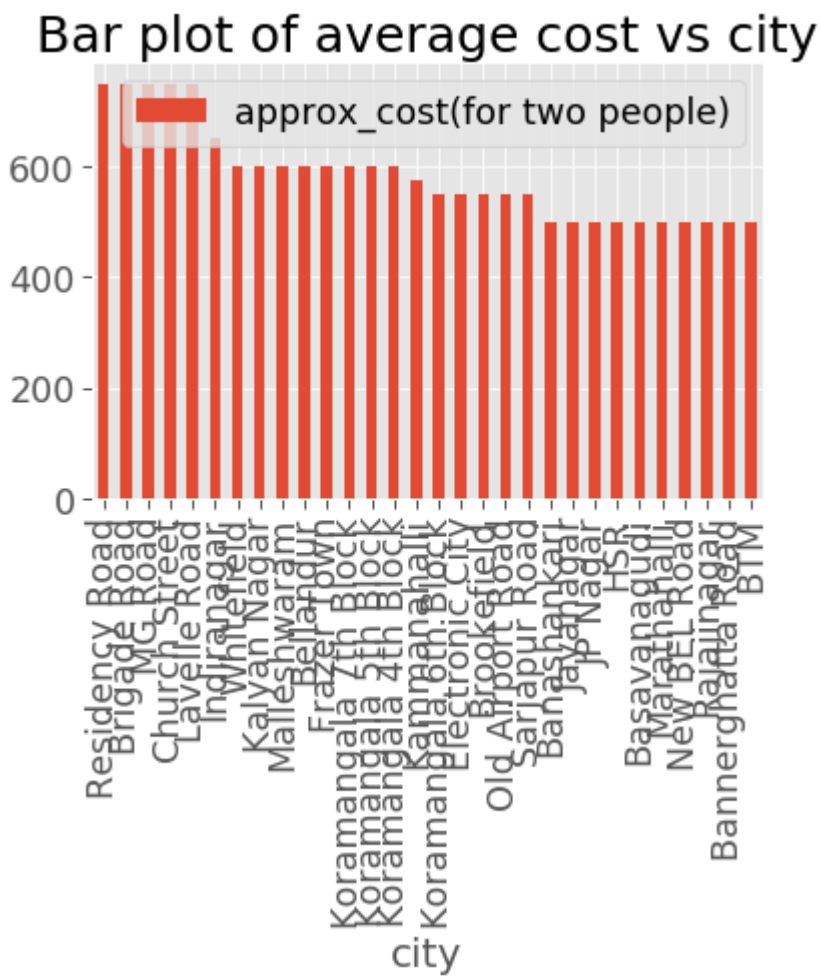


We see that BTM has the highest number of restaurants where you can or can not book table.

In [71]:
```
sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
b = sns.countplot(data = df, x = 'online_order', hue = 'book_table')
plt.title('Number of restaurants in online order in which you can book a table or
b.set_xticklabels(b.get_xticklabels(),rotation = 90)
plt.show()
```

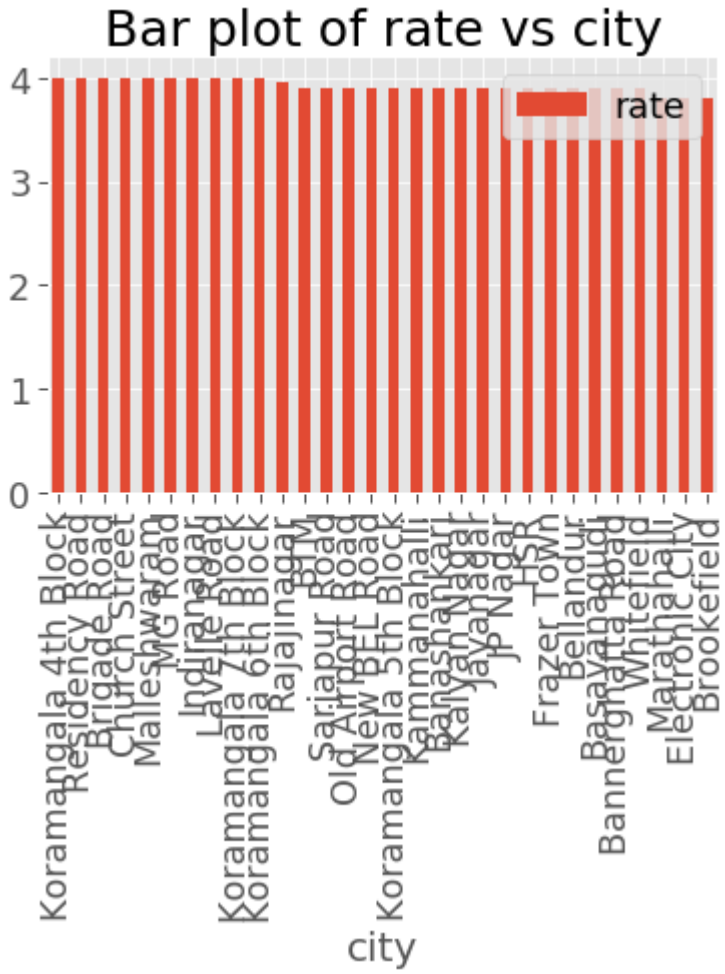
In [72]:
```
plt.figure(figsize=(20,10))
df[['approx_cost(for two people)', 'city']].groupby(['city']).median().sort_value
plt.title('Bar plot of average cost vs city')
plt.show()
```

```
<Figure size 1440x720 with 0 Axes>
```



In [73]:
```
plt.figure(figsize=(20,10))
df[['rate', 'city']].groupby(['city']).median().sort_values("rate", ascending = F
plt.title('Bar plot of rate vs city')
plt.show()
```

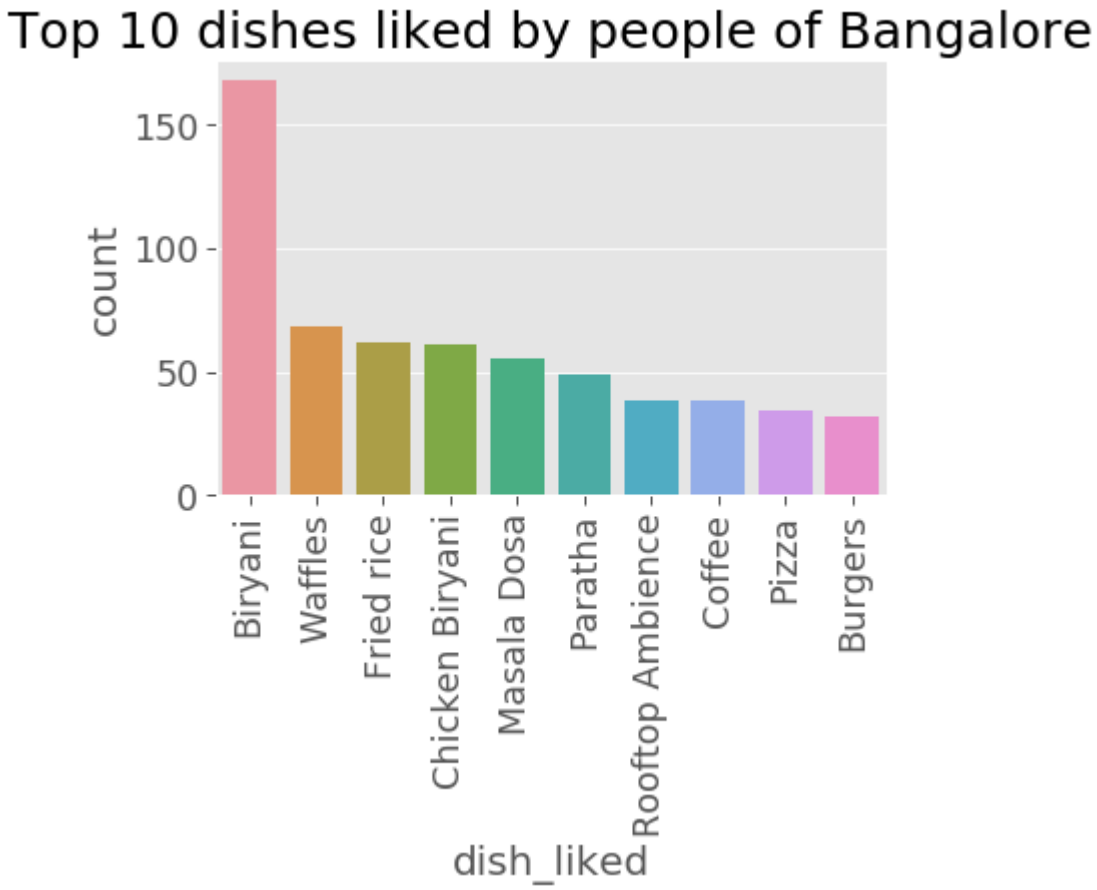
```
<Figure size 1440x720 with 0 Axes>
```



In [74]:
```
a = df.iloc[:, :].values
for i in range(0, len(a)):
    if a[i, 6] == 'Fried rice':
        print(a[i, 6])
        a[i, 6] = None

a = pd.DataFrame(a)
a.columns = df.columns

sns.set_context("paper", font_scale = 2, rc = {"font.size": 20,"axes.titlesize":
b = sns.countplot(data = a, x = 'dish_liked', order = a.dish_liked.value_counts(
b.set_xticklabels(b.get_xticklabels(),rotation = 90)
plt.title('Top 10 dishes liked by people of Bangalore')
plt.show()
```



Conlcusion

*We see that people at Bangalore love Biryani the most,then comes Chicken Biryani and then Fried rice

In [ ]:

In [ ]: