Marketing

August 31, 2024

```python
[2]: import pandas as pd
     from matplotlib import pyplot as plt
     import seaborn as sns
```

```
/usr/local/lib/python3.10/site-packages/scipy/__init__.py:155: UserWarning: A
NumPy version >=1.18.5 and <1.26.0 is required for this version of SciPy
(detected version 1.26.4
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

```python
[3]: df=pd.read_excel("data.xlsx")
```

```python
[3]: df.head()
```

```
[3]:    Restaurant ID              Restaurant Name  Country Code     City  \
     0        7402935                         Skye            94  Jakarta
     1        7410290      Satoo - Hotel Shangri-La            94  Jakarta
     2        7420899                   Sushi Masa            94  Jakarta
     3        7421967                3 Wise Monkeys            94  Jakarta
     4        7422489  Avec Moi Restaurant and Bar            94  Jakarta

                                              Address  \
     0  Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri…
     1            Hotel Shangri-La, Jl. Jend. Sudirman
     2               Jl. Tuna Raya No. 5, Penjaringan
     3             Jl. Suryo No. 26, Senopati, Jakarta
     4  Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta

                            Locality                    Locality Verbose  \
     0  Grand Indonesia Mall, Thamrin  Grand Indonesia Mall, Thamrin, Jakarta
     1     Hotel Shangri-La, Sudirman       Hotel Shangri-La, Sudirman, Jakarta
     2                   Penjaringan                   Penjaringan, Jakarta
     3                      Senopati                      Senopati, Jakarta
     4                       Thamrin                       Thamrin, Jakarta

         Longitude  Latitude                    Cuisines  Average Cost for two  \
     0  106.821999 -6.196778        Italian, Continental                800000
     1  106.818961 -6.203292  Asian, Indonesian, Western                800000
     2  106.800144 -6.101298             Sushi, Japanese                500000
```

1

```
3  106.813400 -6.235241                    Japanese                 450000
4  106.821023 -6.196270               French, Western              350000

                Currency Has Table booking Has Online delivery  Price range  \
0  Indonesian Rupiah(IDR)               No                  No            3
1  Indonesian Rupiah(IDR)               No                  No            3
2  Indonesian Rupiah(IDR)               No                  No            3
3  Indonesian Rupiah(IDR)               No                  No            3
4  Indonesian Rupiah(IDR)               No                  No            3

   Aggregate rating Rating color Rating text  Votes
0              4.1        Green   Very Good   1498
1              4.6   Dark Green   Excellent    873
2              4.9   Dark Green   Excellent    605
3              4.2        Green   Very Good    395
4              4.3        Green   Very Good    243
```

[ ]: 1. Data Preliminary analysis:
- Perform preliminary data inspection **and** report the findings **as** the structure↪
  ↪of the data,
missing values, duplicates etc.
- Based on the findings **from the** previous questions identify duplicates **and**↪
  ↪remove them.

[5]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 19 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Restaurant ID        9551 non-null   int64
 1   Restaurant Name      9550 non-null   object
 2   Country Code         9551 non-null   int64
 3   City                 9551 non-null   object
 4   Address              9551 non-null   object
 5   Locality             9551 non-null   object
 6   Locality Verbose     9551 non-null   object
 7   Longitude            9551 non-null   float64
 8   Latitude             9551 non-null   float64
 9   Cuisines             9542 non-null   object
 10  Average Cost for two 9551 non-null   int64
 11  Currency             9551 non-null   object
 12  Has Table booking    9551 non-null   object
 13  Has Online delivery  9551 non-null   object
 14  Price range          9551 non-null   int64
 15  Aggregate rating     9551 non-null   float64
```

```
16  Rating color        9551 non-null   object
17  Rating text         9551 non-null   object
18  Votes               9551 non-null   int64
dtypes: float64(3), int64(5), object(11)
memory usage: 1.4+ MB
```

[7]: `df.shape`

[7]: (9551, 19)

[8]: `df.describe()`

[8]:
	Restaurant ID	Country Code	Longitude	Latitude
count	9.551000e+03	9551.000000	9551.000000	9551.000000
mean	9.051128e+06	18.365616	64.126574	25.854381
std	8.791521e+06	56.750546	41.467058	11.007935
min	5.300000e+01	1.000000	-157.948486	-41.330428
25%	3.019625e+05	1.000000	77.081343	28.478713
50%	6.004089e+06	1.000000	77.191964	28.570469
75%	1.835229e+07	1.000000	77.282006	28.642758
max	1.850065e+07	216.000000	174.832089	55.976980

	Average Cost for two	Price range	Aggregate rating	Votes
count	9551.000000	9551.000000	9551.000000	9551.000000
mean	1199.210763	1.804837	2.666370	156.909748
std	16121.183073	0.905609	1.516378	430.169145
min	0.000000	1.000000	0.000000	0.000000
25%	250.000000	1.000000	2.500000	5.000000
50%	400.000000	2.000000	3.200000	31.000000
75%	700.000000	2.000000	3.700000	131.000000
max	800000.000000	4.000000	4.900000	10934.000000

[9]: `df.isnull().sum()`

[9]:
```
Restaurant ID         0
Restaurant Name       1
Country Code          0
City                  0
Address               0
Locality              0
Locality Verbose      0
Longitude             0
Latitude              0
Cuisines              9
Average Cost for two  0
Currency              0
Has Table booking     0
```

```
Has Online delivery      0
Price range              0
Aggregate rating         0
Rating color             0
Rating text              0
Votes                    0
dtype: int64
```

[13]: `df.dropna(inplace=True)`

[ ]: `#Clean variable names`

[7]: `df.columns = df.columns.str.replace(' ','_')`

[12]: `df.head()`

[12]:
```
   Restaurant_ID              Restaurant_Name  Country_Code     City  \
0       7402935                         Skye            94  Jakarta
1       7410290       Satoo - Hotel Shangri-La          94  Jakarta
2       7420899                   Sushi Masa            94  Jakarta
3       7421967                3 Wise Monkeys            94  Jakarta
4       7422489  Avec Moi Restaurant and Bar           94  Jakarta

                                        Address  \
0  Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamri…
1            Hotel Shangri-La, Jl. Jend. Sudirman
2                 Jl. Tuna Raya No. 5, Penjaringan
3             Jl. Suryo No. 26, Senopati, Jakarta
4  Gedung PIC, Jl. Teluk Betung 43, Thamrin, Jakarta

                           Locality                     Locality_Verbose  \
0  Grand Indonesia Mall, Thamrin  Grand Indonesia Mall, Thamrin, Jakarta
1     Hotel Shangri-La, Sudirman     Hotel Shangri-La, Sudirman, Jakarta
2                    Penjaringan                    Penjaringan, Jakarta
3                       Senopati                       Senopati, Jakarta
4                        Thamrin                        Thamrin, Jakarta

     Longitude  Latitude                  Cuisines  Average_Cost_for_two  \
0  106.821999 -6.196778       Italian, Continental                800000
1  106.818961 -6.203292  Asian, Indonesian, Western               800000
2  106.800144 -6.101298            Sushi, Japanese                500000
3  106.813400 -6.235241                   Japanese                450000
4  106.821023 -6.196270            French, Western                350000

               Currency Has_Table_booking Has_Online_delivery  Price_range  \
0  Indonesian Rupiah(IDR)                No                  No            3
1  Indonesian Rupiah(IDR)                No                  No            3
```

```
2  Indonesian Rupiah(IDR)                  No                   No         3
3  Indonesian Rupiah(IDR)                  No                   No         3
4  Indonesian Rupiah(IDR)                  No                   No         3

   Aggregate_rating Rating_color Rating_text  Votes
0              4.1        Green    Very Good   1498
1              4.6   Dark Green    Excellent    873
2              4.9   Dark Green    Excellent    605
3              4.2        Green    Very Good    395
4              4.3        Green    Very Good    243
```

[14]:
```python
df.duplicated().sum()
```

[14]: 0

[15]:
```python
df["City"].value_counts()
```

[15]:
```
New Delhi         5473
Gurgaon           1118
Noida             1080
Faridabad          251
Ghaziabad           25
                  ...
Mayfield             1
Vernonia             1
Dicky Beach          1
Inverloch            1
Lakes Entrance       1
Name: City, Length: 140, dtype: int64
```

[16]:
```python
plt.figure(figsize=(10,4))
df["City"].value_counts().head().plot(kind="bar")
plt.title("Count of Restaurants in city", color="b", fontsize=15)
plt.xlabel("City")
plt.ylabel("No. Of Restaurants")
plt.show()
```

Count of Restaurants in city
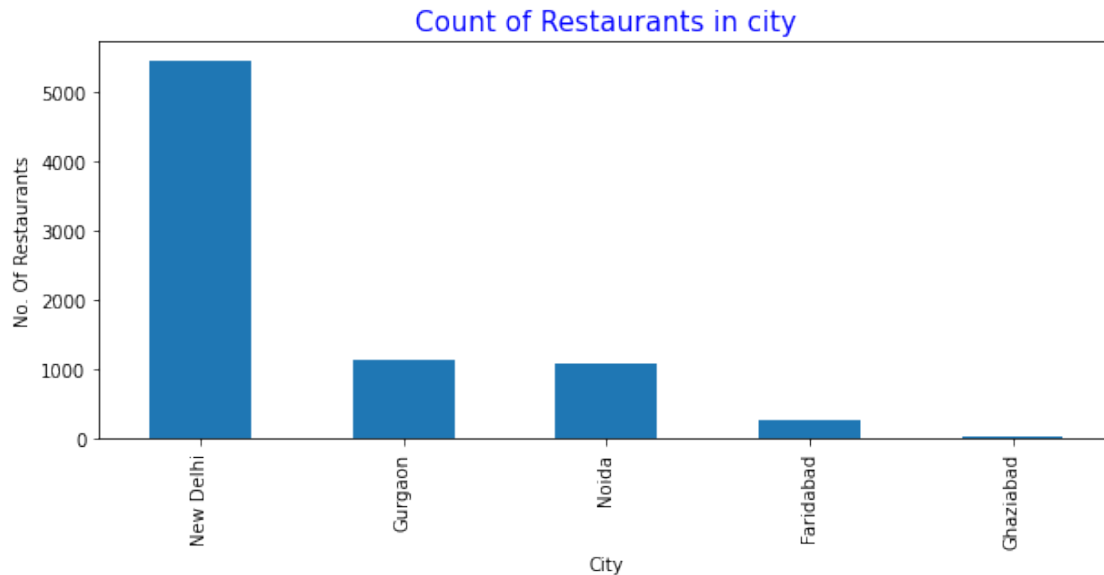
```
[ ]: From above bar graph we can see that New Dellhi has most number of Restaurants␣
     ↪present
```

```
[6]: df["country_name"]=df["Country_Code"]
```

```
[7]: df["country_name"].replace(191,"Sri Lanka",inplace=True)
     df["country_name"].replace(1,"India",inplace=True)
     df["country_name"].replace(14,"Australia",inplace=True)
     df["country_name"].replace(30,"Brazil",inplace=True)
     df["country_name"].replace(37,"Canada",inplace=True)
     df["country_name"].replace(94,"Indonesia",inplace=True)
     df["country_name"].replace(148,"New Zealand",inplace=True)
     df["country_name"].replace(162,"Phillipines",inplace=True)
     df["country_name"].replace(166,"Qatar",inplace=True)
     df["country_name"].replace(184,"Singapore",inplace=True)
     df["country_name"].replace(189,"South Africa",inplace=True)
     df["country_name"].replace(191,"Sri Lanka",inplace=True)
     df["country_name"].replace(208,"Turkey",inplace=True)
     df["country_name"].replace(214,"UAE",inplace=True)
     df["country_name"].replace(215,"United Kingdom",inplace=True)
     df["country_name"].replace(216,"United States",inplace=True)
```

```
[19]: plt.figure(figsize=(10,4))
      df["country_name"].value_counts().head().plot(kind="bar")
      plt.title("Count of Restaurants of country", color="b", fontsize=15)
      plt.xlabel("Country")
      plt.ylabel("No. Of Restaurants")
      plt.show()
```

6

Count of Restaurants of country
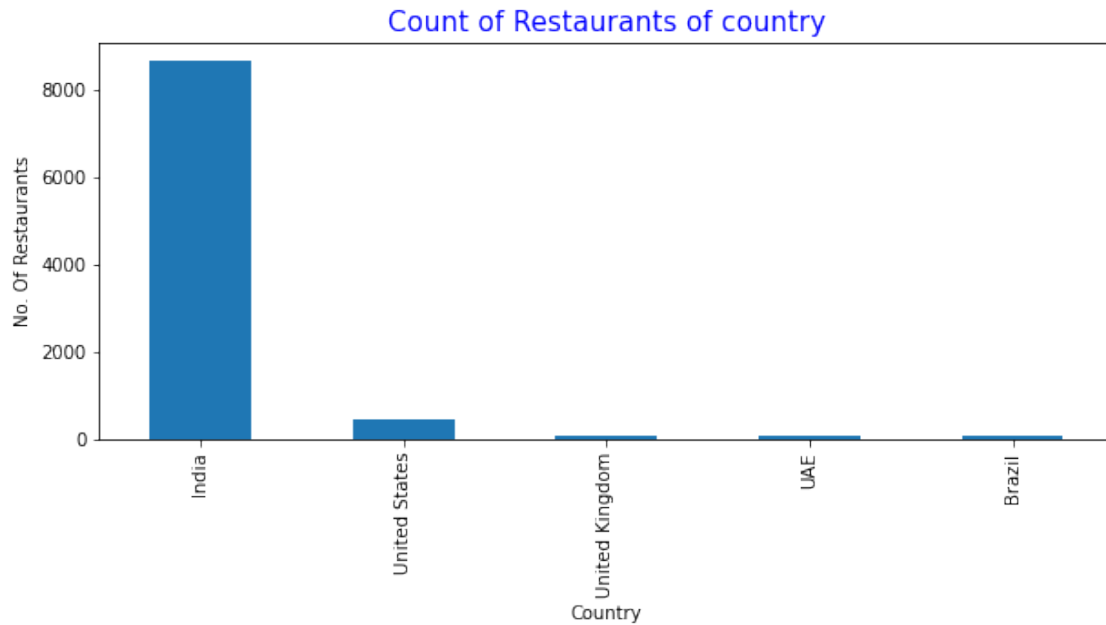
[ ]: From above bar graph we can see that India has most number of Restaurants␣
     ↪present compare to all other country's.

[21]: #4
      df["Has_Table_booking"].value_counts()

[21]: No   8393
      Yes   1158
      Name: Has_Table_booking, dtype: int64

[31]: Ratio = 8393/1158

[32]: Ratio

[32]: 7.247841105354059

[ ]: Ratio of allow table booking and dont allow table booking is 7:1, we need to␣
     ↪allow more table booking service in various reaturants

[26]: #Sol5
      df["Has_Online_delivery"].value_counts()

[26]: No   7100
      Yes   2451
      Name: Has_Online_delivery, dtype: int64

```
[28]: df.shape
```

```
[28]: (9551, 20)
```

```
[29]: Total  = 7100/9551*100
```

```
[30]: Total
```

```
[30]: 74.33776567898649
```

```
[ ]: 74% of restaurants providing online delivery which is high in number
```

```
[35]: df.groupby("Has_Online_delivery")["Votes"].sum()
```

```
[35]: Has_Online_delivery
      No      980731
      Yes     517914
      Name: Votes, dtype: int64
```

```
[ ]: people are not bias towards deliver factor as there is almost 50% of difference␣
      ↪in votes and majority votes towrds restuarants which are not deliver
```

```
[5]: plt.figure(figsize=(10,4))
     df["Cuisines"].value_counts().head(10).plot(kind="bar")
     plt.title("Top 10 cuisines served across cities", color="b", fontsize=15)
     plt.xlabel("Cuisines")
     plt.ylabel("Count of Cuisines")
     plt.show()
```

Top 10 cuisines served across cities
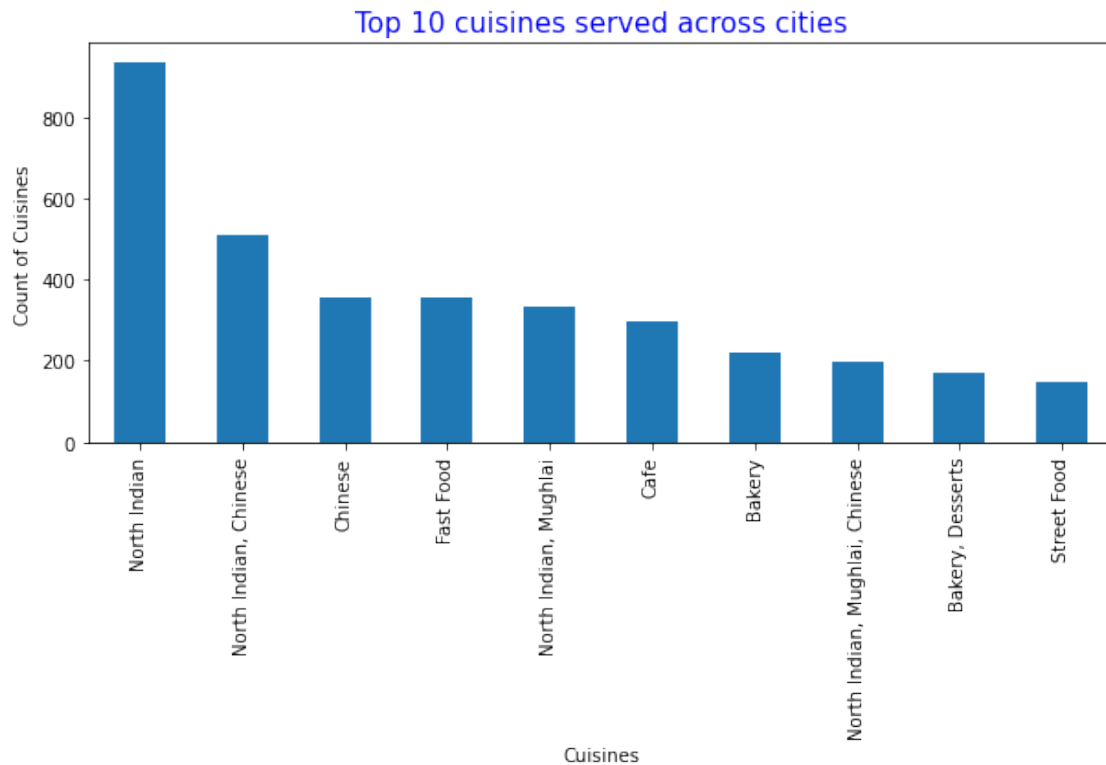

```
[8]: print("Maximum No. of Cuisines",df.groupby("Restaurant_Name")["Cuisines"].
      ↪count().max())
     print("Minimum No. of Cuisines",df.groupby("Restaurant_Name")["Cuisines"].
      ↪count().min())
```

```
Maximum No. of Cuisines 83
Minimum No. of Cuisines 0
```

```
[ ]: Maximum number of cuisines that a restaurant serve is : 83
     Minimum number of cuisines that a restaurant serve is : 0
```

```
[9]: df.describe()
```

[9]:

	Restaurant_ID	Country_Code	Longitude	Latitude	\
count	9.551000e+03	9551.000000	9551.000000	9551.000000	
mean	9.051128e+06	18.365616	64.126574	25.854381	
std	8.791521e+06	56.750546	41.467058	11.007935	
min	5.300000e+01	1.000000	-157.948486	-41.330428	
25%	3.019625e+05	1.000000	77.081343	28.478713	
50%	6.004089e+06	1.000000	77.191964	28.570469	
75%	1.835229e+07	1.000000	77.282006	28.642758	
max	1.850065e+07	216.000000	174.832089	55.976980	

```
        Average_Cost_for_two  Price_range  Aggregate_rating        Votes
count           9551.000000  9551.000000       9551.000000  9551.000000
mean            1199.210763     1.804837          2.666370   156.909748
std            16121.183073     0.905609          1.516378   430.169145
min                0.000000     1.000000          0.000000     0.000000
25%              250.000000     1.000000          2.500000     5.000000
50%              400.000000     2.000000          3.200000    31.000000
75%              700.000000     2.000000          3.700000   131.000000
max           800000.000000     4.000000          4.900000 10934.000000
```

[16]:
```python
df.hist(column='Average_Cost_for_two',grid=False , color='g' )
plt.show()
```
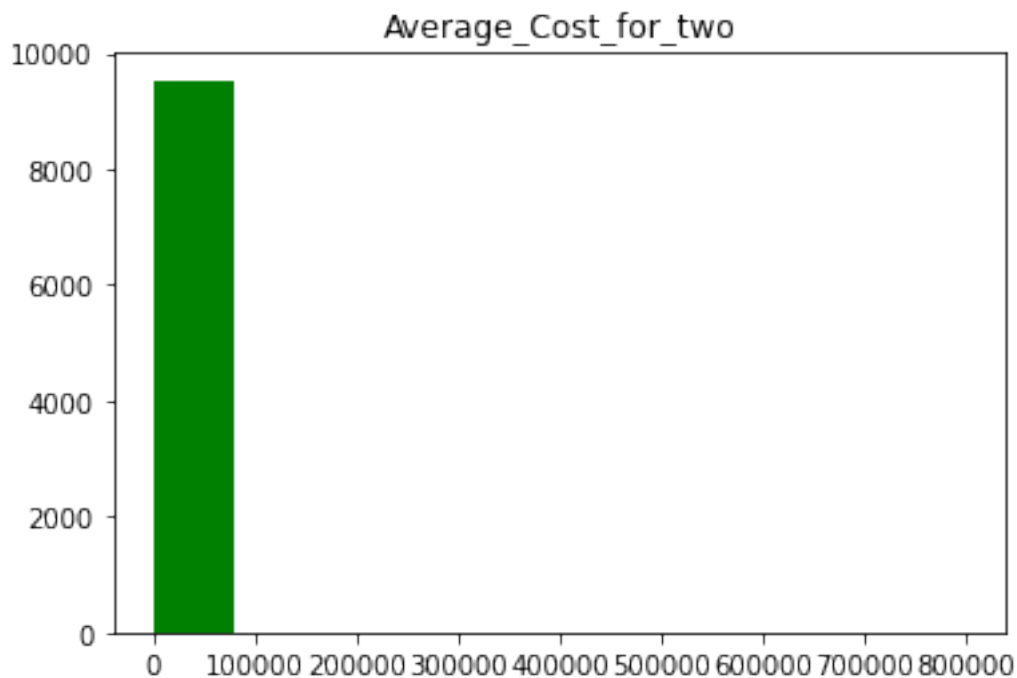

[ ]:
```
There are outliers present in the feature of average cost for two, due to that
↪there is no even distribution we get
```

[17]:
```python
df[df["Aggregate_rating"]==0]
```

[17]:
```
     Restaurant_ID                          Restaurant_Name  Country_Code  \
117       18466422                              Playboy Club             1
128       18415386               Aqua Grill - Pride Plaza Hotel         1
159       18441707   Courtyard Grill - Courtyard by Marriott          1
245       18317498       The Butler & The Chef - Jaypee Greens        1
```

248	17977757	Coffee to Cocktail Bar - Hyatt Place	1
…	…	…	…
9175	6600060	Sandubas Cafì©	30
9257	18453427	Frick's Tap	216
9258	17793744	Los Agaves	216
9320	18491935	Nosh Mahal	216
9389	18273002	Damascena Coffee House	215

	City	Address \
117	New Delhi	Samrat Hotel, 50-B, Kautilya Marg, Chanakyapur…
128	New Delhi	Pride Plaza Hotel, 5A, Hospitality District, A…
159	Gurgaon	Courtyard by Marriott, Plot 27 B, Sushant Lok …
245	Noida	Jaypee Greens, Surajpur Kasna Road, Greater No…
248	Gurgaon	Hyatt Place,15/1, Old Delhi-Gurgaon Road, Sect…
…	…	…
9175	Brasì_lia	Edifì_cio Josì© Severo, SCS 6, Bloco A, Loja 9…
9257	Davenport	1402 W 3rd Street, IA 52802
9258	Davenport	4882 Utica Ridge Rd, Davenport, IA 52807
9320	Pocatello	303 E Alameda Road, ID 83201
9389	Birmingham	133 Alcester Road, Moseley, Birmingham

	Locality \
117	Chanakyapuri
128	Pride Plaza Hotel, Aerocity
159	Courtyard by Marriott, Sushant Lok
245	Jaypee Greens, Greater Noida, Noida
248	Hyatt Place Gurgaon
…	…
9175	Asa Sul
9257	Davenport
9258	Davenport
9320	Pocatello
9389	Moseley

	Locality_Verbose	Longitude	Latitude \
117	Chanakyapuri, New Delhi	77.197445	28.595791
128	Pride Plaza Hotel, Aerocity, New Delhi	77.122890	28.552732
159	Courtyard by Marriott, Sushant Lok, Gurgaon	77.080367	28.460925
245	Jaypee Greens, Greater Noida, Noida, Noida	77.521526	28.464167
248	Hyatt Place Gurgaon, Gurgaon	77.065978	28.500845
…	…	…	…
9175	Asa Sul, Brasì_lia	-47.890167	-15.797000
9257	Davenport, Davenport	-90.594665	41.522536
9258	Davenport, Davenport	-90.515431	41.570943
9320	Pocatello, Pocatello	-112.448530	42.891174
9389	Moseley, Birmingham	-1.888555	52.446302

	Cuisines	Average_Cost_for_two
117	Finger Food	3000
128	Continental, North Indian	3000
159	Chinese, North Indian, Italian	2500
245	Finger Food	2200
248	Drinks Only	2100
…	…	…
9175	Brazilian, Cafe	30
9257	American, Bar Food, BBQ	25
9258	Mexican	25
9320	Indian, Persian	25
9389	Greek, Mediterranean, Middle Eastern	20

	Currency	Has_Table_booking	Has_Online_delivery	Price_range
117	Indian Rupees(Rs.)	No	No	4
128	Indian Rupees(Rs.)	No	No	4
159	Indian Rupees(Rs.)	No	No	4
245	Indian Rupees(Rs.)	Yes	No	4
248	Indian Rupees(Rs.)	Yes	No	4
…	…	…	…	…
9175	Brazilian Real(R$)	No	No	1
9257	Dollar($)	No	No	2
9258	Dollar($)	No	No	2
9320	Dollar($)	No	No	2
9389	Pounds(£)	No	No	2

|	Aggregate_rating	Rating_color	Rating_text	Votes
117	0.0	White	Not rated	3
128	0.0	White	Not rated	0
159	0.0	White	Not rated	3
245	0.0	White	Not rated	1
248	0.0	White	Not rated	0
…	…	…	…	…
9175	0.0	White	Not rated	2
9257	0.0	White	Not rated	2
9258	0.0	White	Not rated	3
9320	0.0	White	Not rated	1
9389	0.0	White	Not rated	3

[2148 rows x 19 columns]

```
[20]: df.hist(column='Aggregate_rating',  grid=False,bins=10,figsize=(10,5),
      ↪color='y')
      plt.show()
```

Aggregate_rating
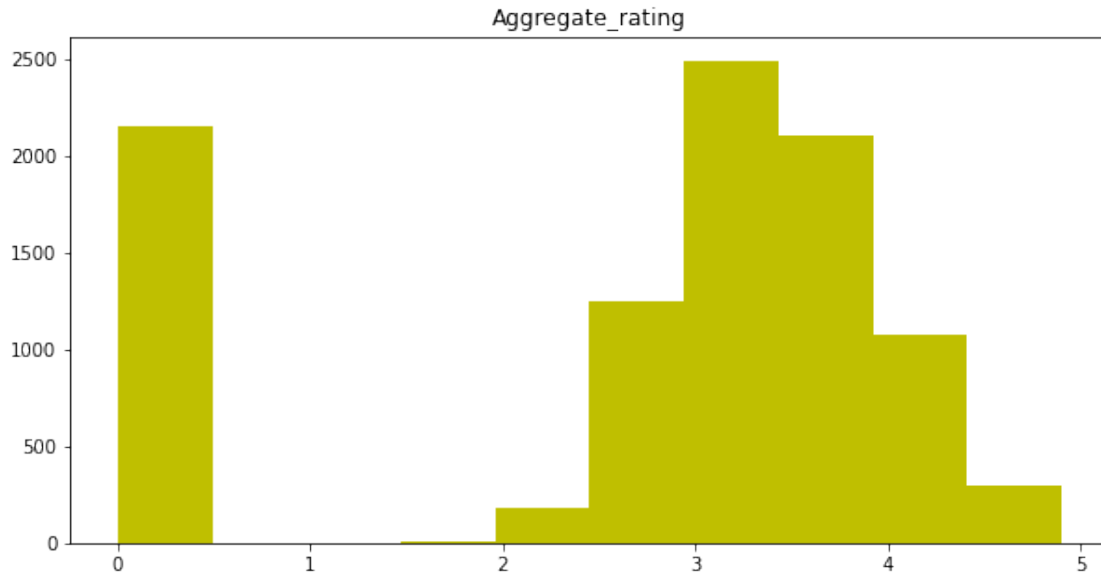
[ ]: Most of the data distributed besides the average ratinng, but also above 2000␣
     ↪restaurants have 0 rating on which organization should take decision to␣
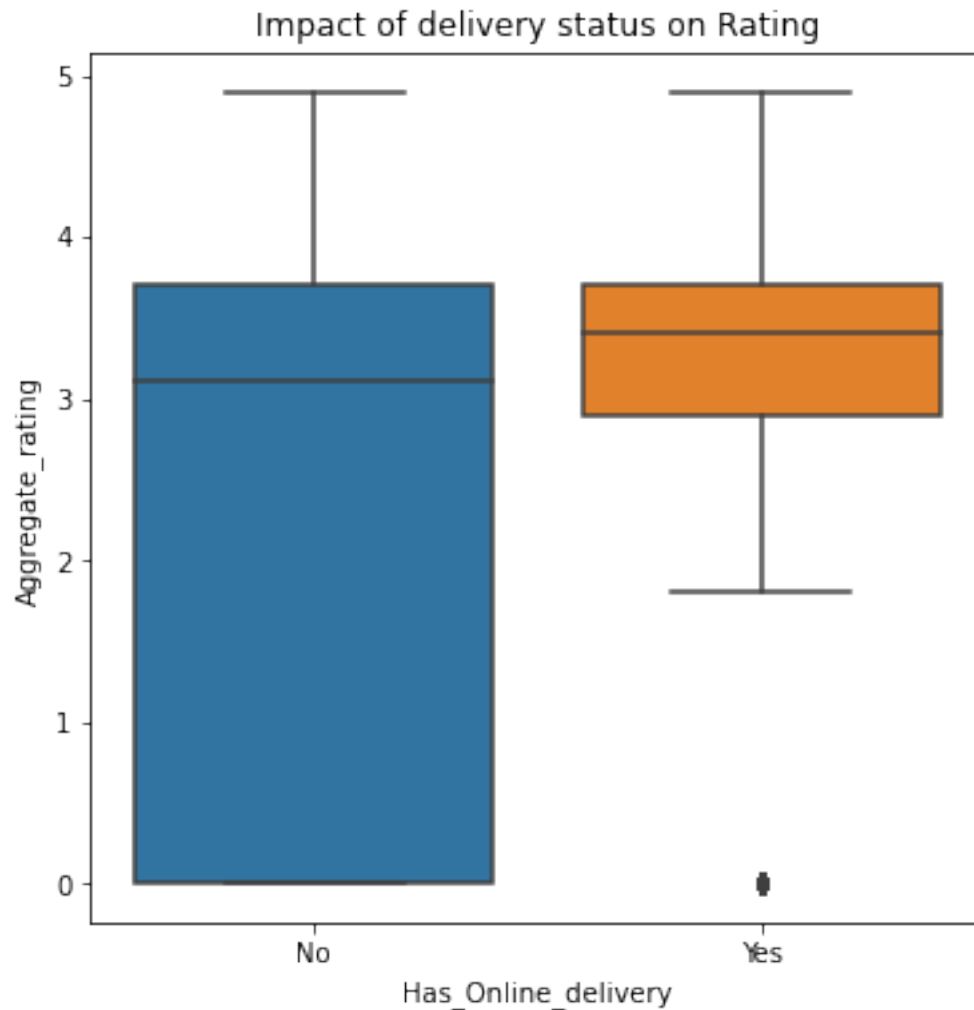   ↪improve ratings

[ ]: Now explaining the factors whos affects on ratings
     Visualising Book Table vs Rating

[22]:
```python
plt.figure(figsize=(6,6))
sns.boxplot(x="Has_Table_booking", y="Aggregate_rating", data=df)
plt.title("Impact of hotel booking status on Rating")
plt.show()
```
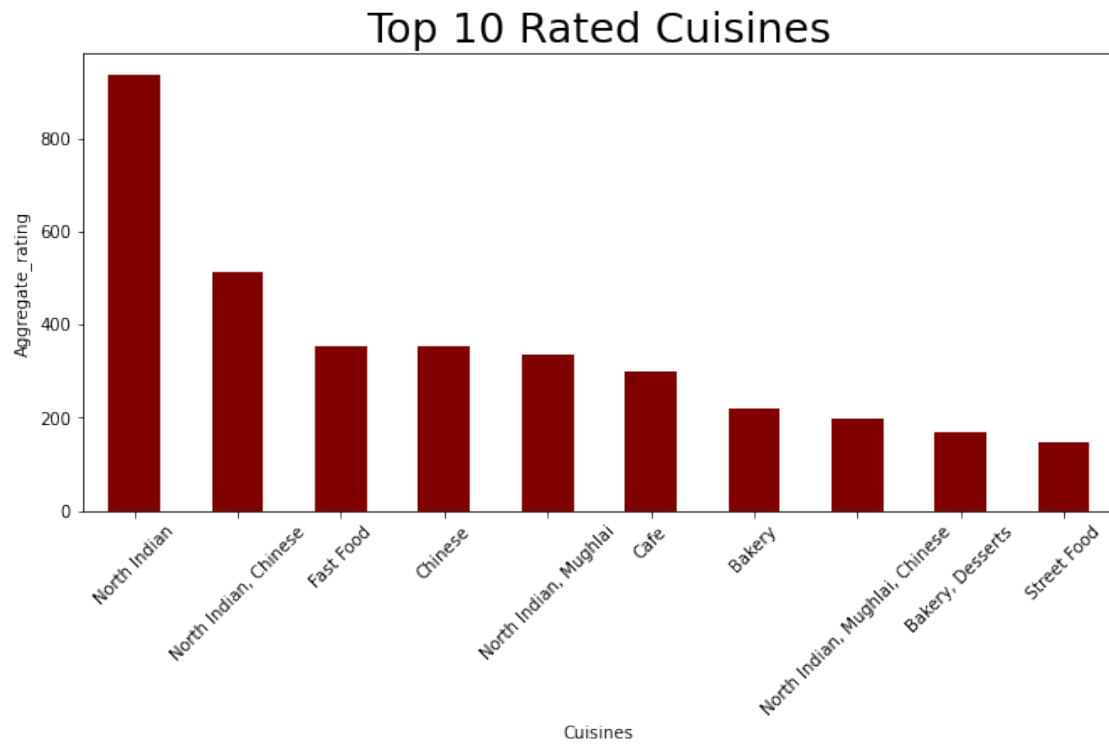
Impact of hotel booking status on Rating
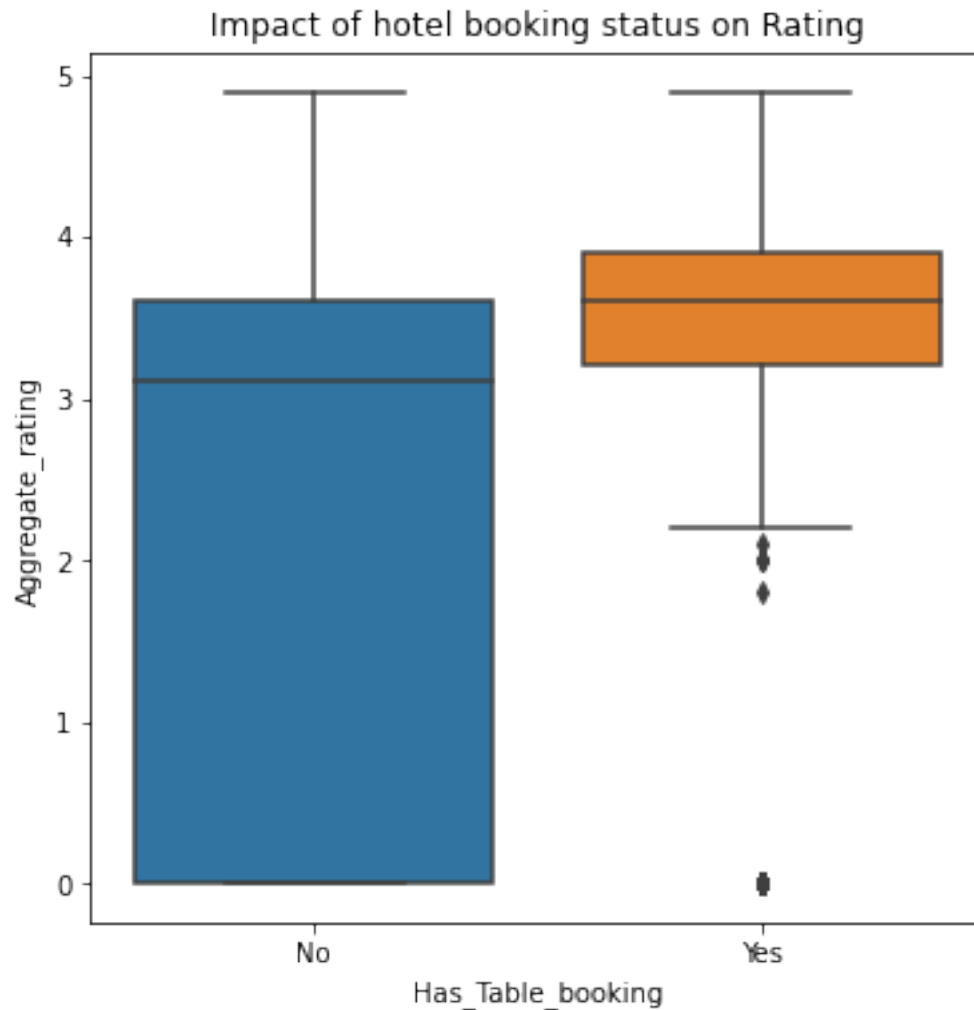
[ ]: Here we can see that the restaurants who having table booking facility has high
     ↪average rating than restaurnt don't have table booking facility

[23]: ```python
plt.figure(figsize=(6,6))
sns.boxplot(x="Has_Online_delivery", y="Aggregate_rating", data=df)
plt.title("Impact of delivery status on Rating")
plt.show()
```

## Impact of delivery status on Rating



[ ]: Here we can see that the restaurants who having online delivery service has␣
     ↪high average rating than restaurnt don't have online delivery service

```
[43]: df2 = df.groupby("Cuisines")["Aggregate_rating"].count().
      ↪sort_values(ascending=False).head(10)
      plt.figure(figsize=(11, 5))
      df2.plot(kind='bar', color='maroon')
      plt.title("Top 10 Rated Cuisines", fontsize=25)
      plt.xlabel("Cuisines")
      plt.ylabel("Aggregate_rating")
      plt.xticks(rotation=45)
      plt.show()
```

Top 10 Rated Cuisines


```
[ ]: Here from the above bar graph we can see that customers have highly prefered␣
      ↪Nort Indian Cuisines
```