In [1]:

```python
import pandas as pd

df = pd.read_csv('zomato.csv')

df.head(2)
```

Out[1]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latit
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553

2 rows × 21 columns

◄ |████| ► |

Checking if dataset contains any null

In [2]:

```python
## Checking if dataset contains any null

nan_values = df.isna()
nan_columns = nan_values.any()

columns_with_nan = df.columns[nan_columns].tolist()
print(columns_with_nan)
```

```
['Cuisines']
```

- Cuisines seems to contain null values. Hence any further analysis involving Cuisines the NaN values has to be considered.
- There is an other file which is also available along with this dataset

In [3]:

```
1  df1 = pd.read_excel('Country-Code.xlsx')
2  df1.head()
```

Out[3]:

	Country Code	Country
0	1	India
1	14	Australia
2	30	Brazil
3	37	Canada
4	94	Indonesia

Let us merge both the datasets. This will help us to understand the dataset country wise.

In [4]:

```
1  df2 = pd.merge(df,df1,on='Country Code',how='left')
2  df2.head(2)
```

Out[4]:

rant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisin
7637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535	14.565443	Fren Japane Desse
4287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101	14.553708	Japane

2 columns

◄ |▬▬▬▬▬▬▬▬▬▬▬▬▬▬| ►

Exploratory Analysis and Visualization

Before we ask question on the dataset, it would be helpful to understand the restaurants geographical spread, understanding the rating, Currency, Online Delivery, City coverage...etc.

List of countries the survey is spread across

In [5]:

```
1  print('List of counteris the survey is spread accross - ')
2  for x in pd.unique(df2.Country): print(x)
3  print()
4  print('Total number to country', len(pd.unique(df2.Country)))
```

```
List of counteris the survey is spread accross -
Phillipines
Brazil
United States
Australia
Canada
Singapore
UAE
India
Indonesia
New Zealand
United Kingdom
Qatar
South Africa
Sri Lanka
Turkey

Total number to country 15
```

The survey seems to have spread across15 countries. This shows that Zomato is a multinational company having actives business in all those countries.

In [6]:

```python
from plotly.offline import init_notebook_mode, plot, iplot

labels = list(df2.Country.value_counts().index)
values = list(df2.Country.value_counts().values)

fig = {
    "data":[
        {
            "labels" : labels,
            "values" : values,
            "hoverinfo" : 'label+percent',
            "domain": {"x": [0, .9]},
            "hole" : 0.6,
            "type" : "pie",
            "rotation":120,
        },
    ],
    "layout": {
        "title" : "Zomato's Presence around the World",
        "annotations": [
            {
                "font": {"size":20},
                "showarrow": True,
                "text": "Countries",
                "x":0.2,
                "y":0.9,
            },
        ]
    }
}

iplot(fig)
```

```
---------------------------------------------------------------------------
ModuleNotFoundError                       Traceback (most recent call last)
<ipython-input-6-4462544319cb> in <module>
----> 1 from plotly.offline import init_notebook_mode, plot, iplot
      2
      3 labels = list(df2.Country.value_counts().index)
      4 values = list(df2.Country.value_counts().values)
      5

ModuleNotFoundError: No module named 'plotly'
```

As Zomato is a startup from India hence it makes sense that it has maximum business spread across restaurants in India

Understanding the Rating aggregate, color and text

In [ ]:

```
1  df3 = df2.groupby(['Aggregate rating','Rating color', 'Rating text']).size().reset_inde
2  df3
3  df3
```

The above information helps us to understand the realation between Aggregate rating, color and text. We conclude the following color assigned to the ratings:

- Rating 0 — White — Not rated
- Rating 1.8 to 2.4 — Red — Poor
- Rating 2.5 to 3.4 — Orange — Average
- Rating 3.5 to 3.9 — Yellow — Good
- Rating 4.0 to 4.4 — Green — Very Good
- Rating 4.5 to 4.9 — Dark Green — Excellent

Let us try to understand the spread of rating across restaurants

In [ ]:

```
1  import seaborn as sns
2  import matplotlib
3  import matplotlib.pyplot as plt
4  %matplotlib inline
5
6  sns.set_style('darkgrid')
7  matplotlib.rcParams['font.size'] = 14
8  matplotlib.rcParams['figure.figsize'] = (9, 5)
9  matplotlib.rcParams['figure.facecolor'] = '#00000000'
10
11  plt.figure(figsize=(12,6))
12  # plt.xticks(rotation=75)
13  plt.title('Rating Color')
14  sns.barplot(x=df3['Rating color'], y=df3['Rating Count']);
```

Interesting, Maximum restaurants seems to have gone No ratings. Let us check if these restaurants belong to some specific country.

In [ ]:

```
1  No_rating = df2[df2['Rating color']=='White'].groupby('Country').size().reset_index().r
2  No_rating
```

India seems to have maximum unrated restaurants. In India the culture of ordering online food is still gaining momentum hence most of the restaurants are still unrated on Zomato as people might be preferring to visiting the restaurant for a meal.

Country and Currency

In [ ]:

```
1  country_currency = df2[['Country','Currency']].groupby(['Country','Currency']).size().r
2  country_currency.sort_values('Currency').reset_index(drop=True)
```

Above table display country and the currency they accept. Interestingly four countries seems to be accepting currency in dollars.

Online delivery distribution

In [ ]:

```
1  plt.figure(figsize=(12,6))
2  plt.title('Online Delivery Distribution')
3  plt.pie(df2['Has Online delivery'].value_counts()/9551*100, labels=df2['Has Online deli
```

Only 25% of restaurants accepts online delivery. This data might be biased as we have maximum restaurants listed here are from India. Maybe analysis over city wise would be more helpful.

Let us try to understand the coverage of city

In [ ]:

```python
from plotly.offline import init_notebook_mode, plot, iplot
import plotly.graph_objs as go
plt.figure(figsize=(12,6))
# import plotly.plotly as py

labels = list(df2.City.value_counts().head(20).index)
values = list(df2.City.value_counts().head(20).values)

fig = {
    "data":[
        {
            "labels" : labels,
            "values" : values,
            "hoverinfo" : 'label+percent',
            "domain": {"x": [0, .9]},
            "hole" : 0.6,
            "type" : "pie",
            "rotation":120,
        },
    ],
    "layout": {
        "title" : "Zomato's Presence Citywise",
        "annotations": [
            {
                "font": {"size":20},
                "showarrow": True,
                "text": "Cities",
                "x":0.2,
                "y":0.9,
            },
        ]
    }
}
iplot(fig);
```

The data seems to be skewed towards New Delhi, Gurgaon and Noida. I see minimal data for other cities. Hence I would do my analysis predominantly on New Delhi.

Asking and Answering Questions

We've already gained several insights about the restaurants present in the survey. Let's ask some specific questions and try to answer them using data frame operations and visualizations.

Q1: From which Locality maximum hotels are listed in Zomato

In [ ]:

```
1  Delhi = df2[(df2.City == 'New Delhi')]
2  plt.figure(figsize=(12,6))
3  sns.barplot(x=Delhi.Locality.value_counts().head(10), y=Delhi.Locality.value_counts().h
4
5  plt.ylabel(None);
6  plt.xlabel('Number of Resturants')
7  plt.title('Resturants Listing on Zomato');
```

Connaught place seems to have high no of restaurants registered with Zomato, Let us understand the cuisines the top rated restaurants has to offer

Q2: What kind of Cuisine these highly rates restaurants offer

In [ ]:

```
1  # I achieve this by the following steps
2
3  ## Fetching the resturants having 'Excellent' and 'Very Good' rating
4  ConnaughtPlace = Delhi[(Delhi.Locality.isin(['Connaught Place'])) & (Delhi['Rating text
5
6  ConnaughtPlace = ConnaughtPlace.Cuisines.value_counts().reset_index()
7
8  ## Extracing all the cuisens in a single list
9  cuisien = []
10 for x in ConnaughtPlace['index']:
11   cuisien.append(x)
12
13 # cuisien = '[%s]'%', '.join(map(str, cuisien))
14 cuisien
```

In [ ]:

```python
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
import pandas as pd

comment_words = ''
stopwords = set(STOPWORDS)

# iterate through the csv file
for val in cuisien:

    # typecaste each val to string
    val = str(val)

    # split the value
    tokens = val.split()

    # Converts each token into lowercase
    for i in range(len(tokens)):
        tokens[i] = tokens[i].lower()

    comment_words += " ".join(tokens)+" "

wordcloud = WordCloud(width = 1500, height = 1500,
                background_color ='white',
                stopwords = stopwords,
                min_font_size = 10).generate(comment_words)

# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = 'orange')
plt.title('Resturants cuisien -  Top Resturants')
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```

Top rated restaurants seems to be doing well in the following cuisine

- North Indian
- Chinese
- Italian
- American

# Q3: How many of such restaurants accept online delivery

In [ ]:

```python
top_locality = Delhi.Locality.value_counts().head(10)
sns.set_theme(style="darkgrid")
plt.figure(figsize=(12,6))
ax = sns.countplot(y= "Locality", hue="Has Online delivery", data=Delhi[Delhi.Locality.
plt.title('Resturants Online Delivery');
```

- Apart from Shahdara locality, restaurants in other locality accepts online delivery.
- Online Delivery seems to be on higher side in Defence colony and Malviya Nagar

Q4: Understanding the Restaurants Rating localities.

- Apart from Malviya nagar, Defence colony in rest of the locality people seems to prefer visiting the restaurants rather ordering food online.
- I would now like to understand the rating of these restaurants that are providing online delivery in Malviya nagar, Defence colony.
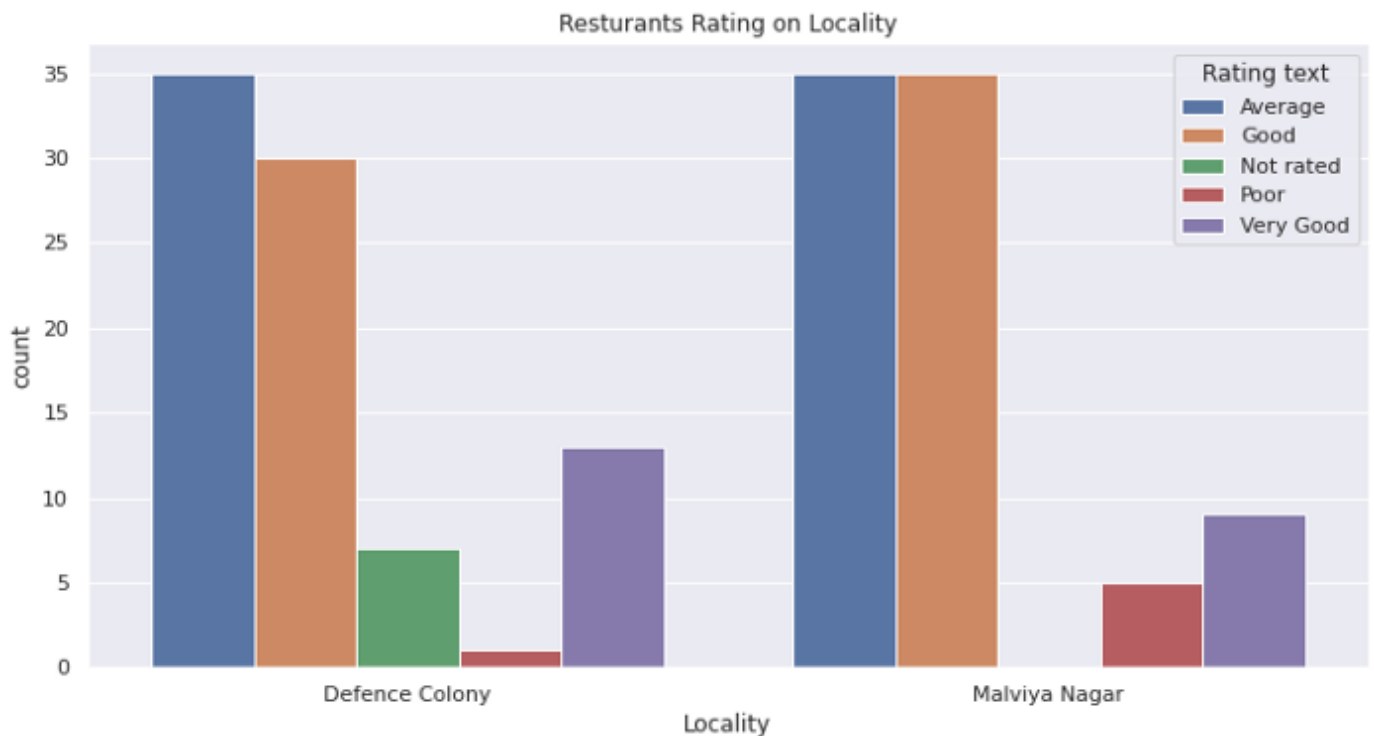
- Defence colony seems to have high no of highly rated restaurants but Malviya Nagar seems to done better in terms of Good and Average restaurants.
- As restaurants with 'Poor' and 'Not Rated' is far lesser that 'Good', 'Very Good' and 'Excellent' restaurants. Hence people in these localities prefer online ordering

Q5: Rating VS Cost of dinning

In [ ]:

```
1  plt.figure(figsize=(12,6))
2  sns.scatterplot(x="Average Cost for two", y="Aggregate rating", hue='Price range', data
3
4  plt.xlabel("Average Cost for two")
5  plt.ylabel("Aggregate rating")
6  plt.title('Rating vs Cost of Two');
```

I observe there is no linear relation between price and rating. For instance, Restaurants with good rating (like 4–5) have restaurants with all the price range and spread across the entire X axis

Q6: Location of Highly rated restaurants across New Delhi

In [ ]:

```python
1  Delhi['Rating text'].value_counts()
```

In [ ]:

```python
1  import plotly.express as px
2  Highly_rated = Delhi[Delhi['Rating text'].isin(['Excellent'])]
3
4  fig = px.scatter_mapbox(Highly_rated, lat="Latitude", lon="Longitude", hover_name="City
5                          color_discrete_sequence=["fuchsia"], zoom=10, height=300)
6  fig.update_layout(mapbox_style="open-street-map")
7  fig.update_layout(margin={"r":0,"t":0,"l":0,"b":0})
8  fig.update_layout(title='Highle rated Resturants Location',
9                    autosize=True,
10                   hovermode='closest',
11                   showlegend=False)
12 fig.update_layout(
13     autosize=False,
14     width=800,
15     height=500,)
16
17 fig.show()
```

The aforementioned four cities represent nearly 65% of the total data available in the dataset. Apart from the higly rated local restaurants, it'd be intersting to know where the known-eateries that are commonplace. The verticles across which these can be located are -

- Breakfast
- American Fast Food
- Ice Creams, Shakes & Desserts

Q7: Common Eateries

1: Breakfast and Coffee locations

In [ ]:

```python
1  types = {
2      "Breakfast and Coffee" : ["Cafe Coffee Day", "Starbucks", "Barista", "Costa Coffee"
3      "American": ["Domino's Pizza", "McDonald's", "Burger King", "Subway", "Dunkin' Donu
4      "Ice Creams and Shakes": ["Keventers", "Giani", "Giani's", "Starbucks", "Baskin Rob
5  }
6
7  breakfast = Delhi[Delhi['Restaurant Name'].isin(types['Breakfast and Coffee'])]
8  american = Delhi[Delhi['Restaurant Name'].isin(types['American'])]
9  ice_cream = Delhi[Delhi['Restaurant Name'].isin(types['Ice Creams and Shakes'])]
```

In [ ]:

```python
breakfast = breakfast[['Restaurant Name','Aggregate rating']].groupby('Restaurant Name'
breakfast
```

In [ ]:

```python
import plotly.express as px

df= breakfast
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating', ti
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)
fig.show()
```

Chaayos outlets are doing better. We need more of those in Delhi. Café coffee day seems to be performing poorly in avg rating. They are required to improve their services.

2: Fast Food Restaurants

In [ ]:

```python
american = american[['Restaurant Name','Aggregate rating']].groupby('Restaurant Name').
american
```

In [ ]:

```python
import plotly.express as px

df= american
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating', ti
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)

fig.show()
```

3: Ice Cream Parlors

In [ ]:

```python
ice_cream = ice_cream[['Restaurant Name','Aggregate rating']].groupby('Restaurant Name'
ice_cream
```

In [ ]:

```python
import plotly.express as px

df= ice_cream
fig = px.bar(df, y='Aggregate rating', x='Restaurant Name', text='Aggregate rating', ti
fig.update_traces(texttemplate='%{text:.3s}', textposition='outside')
fig.update_layout(
    autosize=False,
    width=800,
    height=500,)
fig.show()
```

Foreign brands seems to be doing better than the local brands

Inferences and Conclusions

We've drawn many inferences from the survey. Here's a summary of a few of them:

- The dataset is skewed towards India and doesn't represent the complete data of restaurants worldwide.
- Restaurants rating is categorized in categories
  - Not Rated
  - Average
  - Good
  - Very Good
  - Excellent
- Connaught Palace have maximum restaurants listed on Zomato but in terms of online delivery acceptance Defence colony and Malviya nagar seems to be doing better.
- The top rated restaurants seems to be getting better rating on the following cuisine
  - North Indian
  - Chinese
  - American
  - Italian
- There is no relation between cost and rating. Some of the best rated restaurants are low on cost and vice versa.
- On common Eateries, For Breakfast and Coffee location Indian restaurants seems to be better rated but for Fast food chain and Ice cream parlors American restaurants seems to be doing better.

In [ ]:

```
1
```