In [ ]:

```
# import all the libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [6]:

```
zomato= pd.read_csv('C:/Users/dabir/Downloads/EDA FILE/zomato_EDA.csv', encoding='latin-1
```

In [7]:

```
zomato.head() # Read the zomato dataset
```

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **2** | 6300002 | Heat - Edsa Shangri-La | 162 | Mandaluyong City | Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal... | Edsa Shangri-La, Ortigas, Mandaluyong City | Edsa Shangri-La, Ortigas, Mandaluyong City, Ma... | 121.056831 | 14.581 |
| **3** | 6318506 | Ooma | 162 | Mandaluyong City | Third Floor, Mega Fashion Hall, SM Megamall, O... | SM Megamall, Ortigas, Mandaluyong City | SM Megamall, Ortigas, Mandaluyong City, Mandal... | 121.056475 | 14.585 |
| **4** | 6314302 | Sambo Kojin | 162 | Mandaluyong City | Third Floor, Mega Atrium, SM Megamall, Ortigas... | SM Megamall, Ortigas, Mandaluyong City | SM Megamall, Ortigas, Mandaluyong City, Mandal... | 121.057508 | 14.584 |

In [8]:

```
zomato.columns
```

Out[8]:

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Addres
s',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisine
s',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

In [9]:

```
#given the basic info of dataset
zomato.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   Restaurant ID         9551 non-null   int64  
 1   Restaurant Name       9551 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  Is delivering now     9551 non-null   object 
 15  Switch to order menu  9551 non-null   object 
 16  Price range           9551 non-null   int64  
 17  Aggregate rating      9551 non-null   float64
 18  Rating color          9551 non-null   object 
 19  Rating text           9551 non-null   object 
 20  Votes                 9551 non-null   int64  
dtypes: float64(3), int64(5), object(13)
memory usage: 1.5+ MB
```

In [10]:

```
zomato.describe()
```

Out[10]:

| | Restaurant ID | Country Code | Longitude | Latitude | Average Cost for two | Price range | A |
|---|---|---|---|---|---|---|---|
| **count** | 9.551000e+03 | 9551.000000 | 9551.000000 | 9551.000000 | 9551.000000 | 9551.000000 | 955 |
| **mean** | 9.051128e+06 | 18.365616 | 64.126574 | 25.854381 | 1199.210763 | 1.804837 | |
| **std** | 8.791521e+06 | 56.750546 | 41.467058 | 11.007935 | 16121.183073 | 0.905609 | |
| **min** | 5.300000e+01 | 1.000000 | -157.948486 | -41.330428 | 0.000000 | 1.000000 | |
| **25%** | 3.019625e+05 | 1.000000 | 77.081343 | 28.478713 | 250.000000 | 1.000000 | |
| **50%** | 6.004089e+06 | 1.000000 | 77.191964 | 28.570469 | 400.000000 | 2.000000 | |
| **75%** | 1.835229e+07 | 1.000000 | 77.282006 | 28.642758 | 700.000000 | 2.000000 | |
| **max** | 1.850065e+07 | 216.000000 | 174.832089 | 55.976980 | 800000.000000 | 4.000000 | |

In [11]:

```
zomato.isnull().sum()
```

Out[11]:

```
Restaurant ID         0
Restaurant Name       0
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
Is delivering now     0
Switch to order menu  0
Price range           0
Aggregate rating      0
Rating color          0
Rating text           0
Votes                 0
dtype: int64
```

In [12]:

```
[features for features in zomato.columns if zomato[features].isnull().sum()>0]
```

Out[12]:

```
['Cuisines']
```

In [13]:

```
zomato.columns
```

Out[13]:

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Addres
s',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisine
s',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

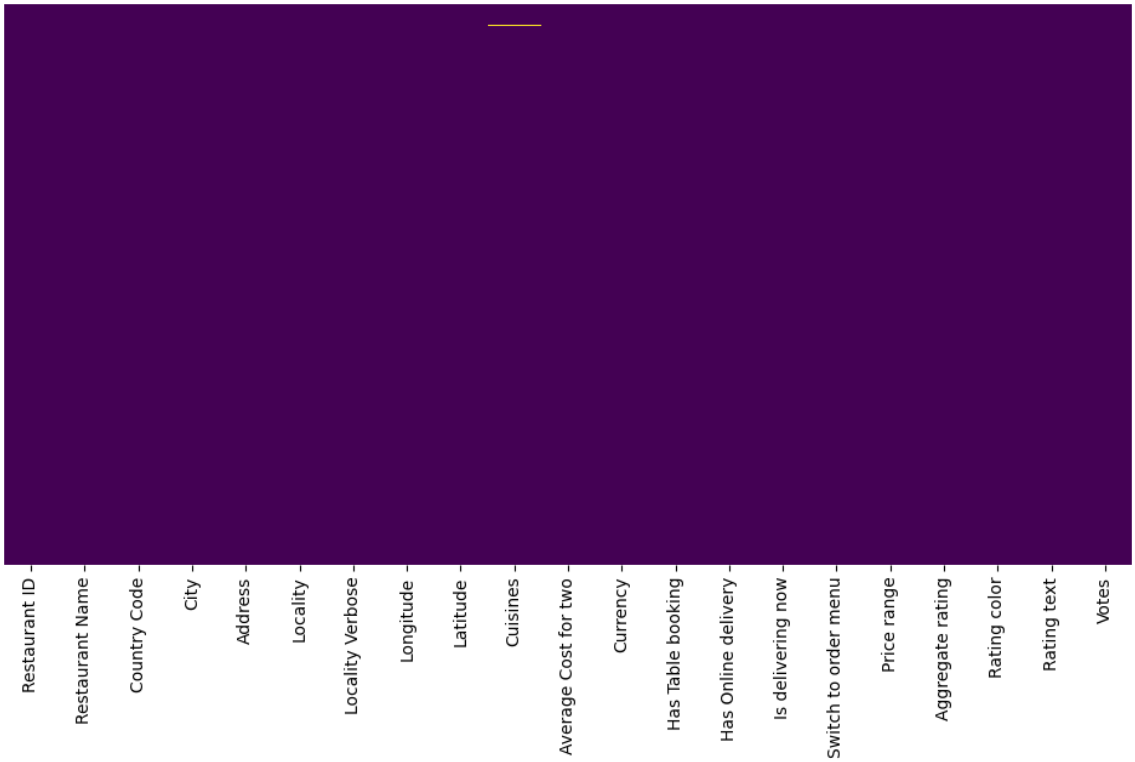
# Heatmap for zomato dataset

In [14]:

```
plt.rcParams['figure.figsize'] = (12, 6)
sns.heatmap(zomato.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

Out[14]:

```
<AxesSubplot: >
```




In [15]:

```
zomato.shape
```

Out[15]:

```
(9551, 21)
```

In [21]:

```
# Read Country-code dataset
zomato_country = pd.read_excel('C:/Users/dabir/Downloads/EDA FILE/Country-Code.xlsx')
zomato_country.head()
```

Out[21]:

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

In [17]:

```
zomato.columns
```

Out[17]:

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Addres
s',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisine
s',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes'],
      dtype='object')
```

In [23]:

```
zomato_final = pd.merge(zomato, zomato_country, on = 'Country Code', how = 'left' )
zomato_final.head(2)
```

Out[23]:

| | Restaurant ID | Restaurant Name | Country Code | City | Address | Locality | Locality Verbose | Longitude | Lat |
|---|---|---|---|---|---|---|---|---|---|
| **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 | 14.56 |
| **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 | 14.55 |

2 rows × 22 columns

In [25]:

```
zomato_final.dtypes
```

Out[25]:

```
Restaurant ID             int64
Restaurant Name          object
Country Code              int64
City                     object
Address                  object
Locality                 object
Locality Verbose         object
Longitude               float64
Latitude                float64
Cuisines                 object
Average Cost for two      int64
Currency                 object
Has Table booking        object
Has Online delivery      object
Is delivering now        object
Switch to order menu     object
Price range               int64
Aggregate rating        float64
Rating color             object
Rating text              object
Votes                     int64
Country                  object
dtype: object
```

In [26]:

```
zomato_final.columns
```

Out[26]:

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Addres
s',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisine
s',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes', 'Country'],
      dtype='object')
```

In [28]:

```
zomato_final.Country.value_counts()
```

Out[28]:

```
India             8652
United States      434
United Kingdom      80
Brazil              60
UAE                 60
South Africa        60
New Zealand         40
Turkey              34
Australia           24
Phillipines         22
Indonesia           21
Singapore           20
Qatar               20
Sri Lanka           20
Canada               4
Name: Country, dtype: int64
```

In [ ]:

```
##getting all country name seperately
country_names = zomato_final.Country.value_counts().index
country_names
```

# Get all the values seperately

In [31]:

```
country_val = final_df.Country.value_counts().values
country_val
```

Out[31]:

```
array([8652,  434,   80,   60,   60,   60,   40,   34,   24,   22,   21,
         20,   20,   20,    4], dtype=int64)
```

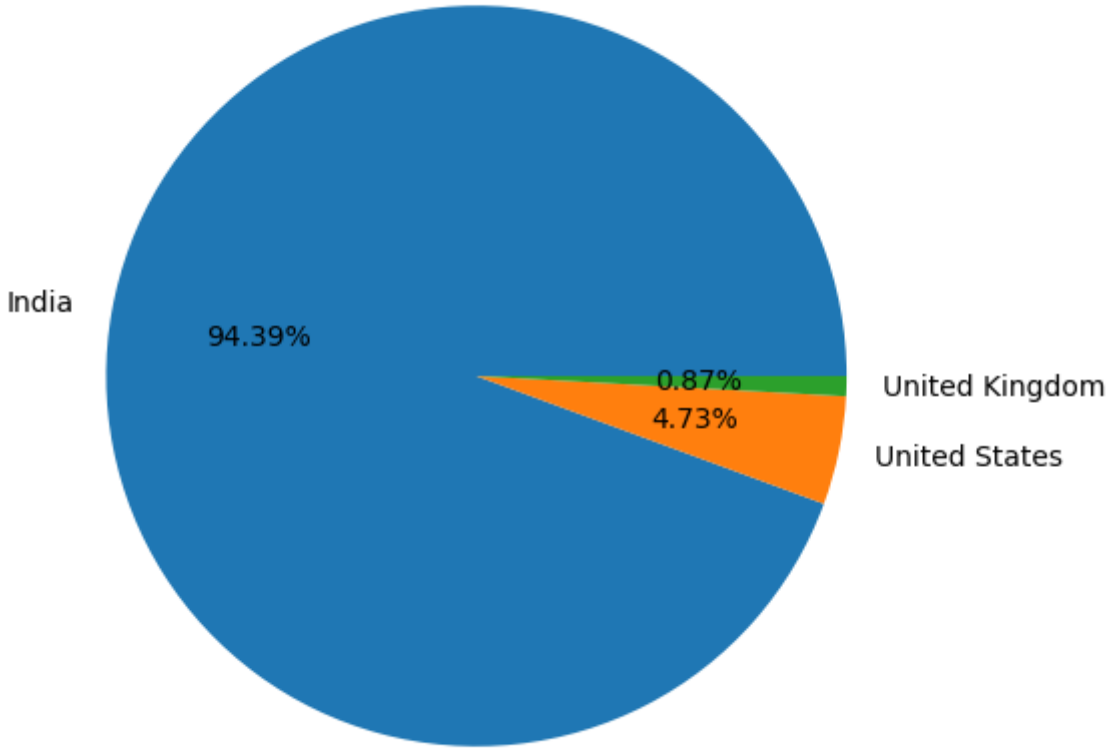
# pie chart between Top 3 country_name and country_value

In [32]:

```
plt.pie(country_val[:3], labels=country_names[:3], autopct= '%1.2f%%')
```

Out[32]:

```
([<matplotlib.patches.Wedge at 0x2358665ccd0>,
  <matplotlib.patches.Wedge at 0x2358665d1e0>,
  <matplotlib.patches.Wedge at 0x2358665d750>],
 [Text(-1.0829742700952103, 0.19278674827836725, 'India'),
  Text(1.077281715838356, -0.22240527134123297, 'United States'),
  Text(1.0995865153823035, -0.03015783794312073, 'United Kingdom')],
 [Text(-0.590713238233751, 0.10515640815183668, '94.39%'),
  Text(0.5876082086391032, -0.12131196618612707, '4.73%'),
  Text(0.5997744629358018, -0.01644972978715676, '0.87%')])
```




Observation:Zomato maximum records or transaction are from India After that USA and then United Kingdoms

In [33]:

```
zomato_final.columns
```

Out[33]:

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Addres
s',
       'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisine
s',
       'Average Cost for two', 'Currency', 'Has Table booking',
       'Has Online delivery', 'Is delivering now', 'Switch to order menu',
       'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
       'Votes', 'Country'],
      dtype='object')
```

In [38]:

```
zomato_final.groupby(['Country','Currency']).size().reset_index()
```

Out[38]:

| | Country | Currency | 0 |
|---|---|---|---|
| 0 | Australia | Dollar($) | 24 |
| 1 | Brazil | Brazilian Real(R$) | 60 |
| 2 | Canada | Dollar($) | 4 |
| 3 | India | Indian Rupees(Rs.) | 8652 |
| 4 | Indonesia | Indonesian Rupiah(IDR) | 21 |
| 5 | New Zealand | NewZealand($) | 40 |
| 6 | Phillipines | Botswana Pula(P) | 22 |
| 7 | Qatar | Qatari Rial(QR) | 20 |
| 8 | Singapore | Dollar($) | 20 |
| 9 | South Africa | Rand(R) | 60 |
| 10 | Sri Lanka | Sri Lankan Rupee(LKR) | 20 |
| 11 | Turkey | Turkish Lira(TL) | 34 |
| 12 | UAE | Emirati Diram(AED) | 60 |
| 13 | United Kingdom | Pounds(□£) | 80 |
| 14 | United States | Dollar($) | 434 |

In [39]:

```
zomato_final[zomato_final['Has Online delivery'] =="Yes"].Country.value_counts()
```

Out[39]:

```
India    2423
UAE        28
Name: Country, dtype: int64
```

In [40]:

```
zomato_final.groupby(['Has Online delivery','Country']).size().reset_index()
```

Out[40]:

| | Has Online delivery | Country | 0 |
|---|---|---|---|
| 0 | No | Australia | 24 |
| 1 | No | Brazil | 60 |
| 2 | No | Canada | 4 |
| 3 | No | India | 6229 |
| 4 | No | Indonesia | 21 |
| 5 | No | New Zealand | 40 |
| 6 | No | Phillipines | 22 |
| 7 | No | Qatar | 20 |
| 8 | No | Singapore | 20 |
| 9 | No | South Africa | 60 |
| 10 | No | Sri Lanka | 20 |
| 11 | No | Turkey | 34 |
| 12 | No | UAE | 32 |
| 13 | No | United Kingdom | 80 |
| 14 | No | United States | 434 |
| 15 | Yes | India | 2423 |
| 16 | Yes | UAE | 28 |

Observations: Online Deliveries are available in India and UAE

# Creating a pie chart for top 5 cities distribution

In [41]:

```
 city_labels = zomato_final['City'].value_counts().index
city_labels
```

Out[41]:

```
Index(['New Delhi', 'Gurgaon', 'Noida', 'Faridabad', 'Ghaziabad',
       'Bhubaneshwar', 'Amritsar', 'Ahmedabad', 'Lucknow', 'Guwahati',
       ...
       'Ojo Caliente', 'Montville', 'Monroe', 'Miller', 'Middleton Beach',
       'Panchkula', 'Mc Millan', 'Mayfield', 'Macedon', 'Vineland Statio
n'],
      dtype='object', length=141)
```

In [43]:

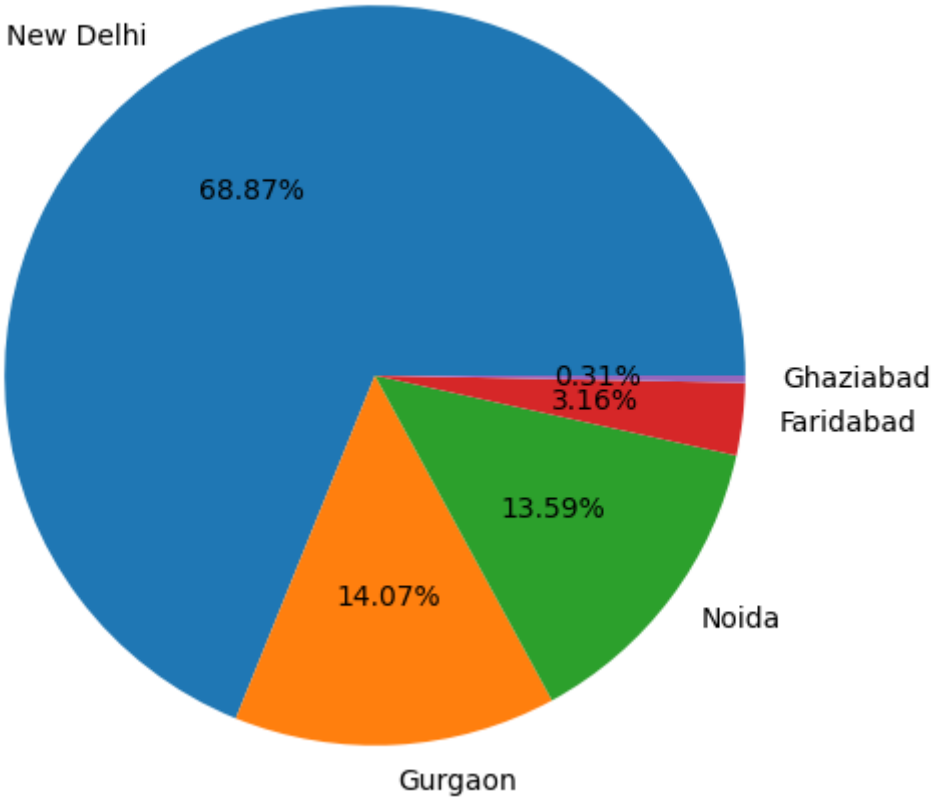
```
city_values = zomato_final['City'].value_counts().values
```

In [44]:

```
plt.pie(city_values[:5], labels = city_labels[:5],autopct='%1.2f%%')
```

Out[44]:

```
([<matplotlib.patches.Wedge at 0x235868eba90>,
  <matplotlib.patches.Wedge at 0x23586c56dd0>,
  <matplotlib.patches.Wedge at 0x23586c57460>,
  <matplotlib.patches.Wedge at 0x23586c57af0>,
  <matplotlib.patches.Wedge at 0x235875081c0>],
 [Text(-0.6145352824185932, 0.9123301960708633, 'New Delhi'),
  Text(0.0623675251198054, -1.0982305276263407, 'Gurgaon'),
  Text(0.8789045225625368, -0.6614581167535246, 'Noida'),
  Text(1.0922218418223437, -0.13058119407559224, 'Faridabad'),
  Text(1.099946280005612, -0.010871113182029924, 'Ghaziabad')],
 [Text(-0.3352010631374145, 0.497634652402289, '68.87%'),
  Text(0.0340186500653484, -0.5990348332507311, '14.07%'),
  Text(0.47940246685229276, -0.36079533641101336, '13.59%'),
  Text(0.5957573682667329, -0.07122610585941394, '3.16%'),
  Text(0.5999706981848791, -0.005929698099289049, '0.31%')])
```



In [50]:

```
zomato_final['Cuisines'].value_counts()[:10]
```

Out[50]:

```
North Indian                       936
North Indian, Chinese              511
Chinese                            354
Fast Food                          354
North Indian, Mughlai              334
Cafe                               299
Bakery                             218
North Indian, Mughlai, Chinese     197
Bakery, Desserts                   170
Street Food                        149
Name: Cuisines, dtype: int64
```

In [ ]:

```

```