zomato-eda

August 7, 2023

[1]:
```
pip install numpy
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: numpy in c:\programdata\anaconda3\lib\site-
packages (1.21.5)
Note: you may need to restart the kernel to use updated packages.
```

[2]:
```
pip install pandas
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pandas in c:\programdata\anaconda3\lib\site-
packages (1.4.2)
Requirement already satisfied: pytz>=2020.1 in
c:\programdata\anaconda3\lib\site-packages (from pandas) (2021.3)
Requirement already satisfied: python-dateutil>=2.8.1 in
c:\programdata\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: numpy>=1.18.5 in
c:\programdata\anaconda3\lib\site-packages (from pandas) (1.21.5)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[3]:
```
import numpy as np
import pandas as pd
import pandas_profiling
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
C:\Users\PC\AppData\Local\Temp\ipykernel_6600\3978922422.py:3:
DeprecationWarning: `import pandas_profiling` is going to be deprecated by April
1st. Please use `import ydata_profiling` instead.
  import pandas_profiling
```

[4]:
```
pip install ydata_profiling
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: ydata_profiling in
```

```
c:\users\pc\appdata\roaming\python\python39\site-packages (4.1.2)
Requirement already satisfied: scipy<1.10,>=1.4.1 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (1.7.3)
Requirement already satisfied: tqdm<4.65,>=4.48.2 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (4.64.0)
Requirement already satisfied: typeguard<2.14,>=2.13.2 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(2.13.3)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (2.11.3)
Requirement already satisfied: pandas!=1.4.0,<1.6,>1.1 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (1.4.2)
Requirement already satisfied: phik<0.13,>=0.11.1 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(0.12.3)
Requirement already satisfied: matplotlib<3.7,>=3.2 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (3.5.1)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (6.0)
Requirement already satisfied: visions[type_image_path]==0.7.5 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(0.7.5)
Requirement already satisfied: seaborn<0.13,>=0.10.1 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (0.11.2)
Requirement already satisfied: multimethod<1.10,>=1.4 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(1.9.1)
Requirement already satisfied: imagehash==4.3.1 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(4.3.1)
Requirement already satisfied: requests<2.29,>=2.24.0 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (2.27.1)
Requirement already satisfied: pydantic<1.11,>=1.8.1 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(1.10.7)
Requirement already satisfied: htmlmin==0.1.12 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from ydata_profiling)
(0.1.12)
Requirement already satisfied: statsmodels<0.14,>=0.13.2 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (0.13.2)
Requirement already satisfied: numpy<1.24,>=1.16.0 in
c:\programdata\anaconda3\lib\site-packages (from ydata_profiling) (1.21.5)
Requirement already satisfied: PyWavelets in c:\programdata\anaconda3\lib\site-
packages (from imagehash==4.3.1->ydata_profiling) (1.3.0)
Requirement already satisfied: pillow in c:\programdata\anaconda3\lib\site-
packages (from imagehash==4.3.1->ydata_profiling) (9.0.1)
Requirement already satisfied: attrs>=19.3.0 in
c:\programdata\anaconda3\lib\site-packages (from
```

```
visions[type_image_path]==0.7.5->ydata_profiling) (21.4.0)
Requirement already satisfied: networkx>=2.4 in
c:\programdata\anaconda3\lib\site-packages (from
visions[type_image_path]==0.7.5->ydata_profiling) (2.7.1)
Requirement already satisfied: tangled-up-in-unicode>=0.0.4 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from
visions[type_image_path]==0.7.5->ydata_profiling) (0.2.0)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\programdata\anaconda3\lib\site-packages (from
jinja2<3.2,>=2.11.1->ydata_profiling) (2.0.1)
Requirement already satisfied: python-dateutil>=2.7 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (2.8.2)
Requirement already satisfied: cycler>=0.10 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (4.25.0)
Requirement already satisfied: packaging>=20.0 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (21.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (1.3.2)
Requirement already satisfied: pyparsing>=2.2.1 in
c:\programdata\anaconda3\lib\site-packages (from
matplotlib<3.7,>=3.2->ydata_profiling) (3.0.4)
Requirement already satisfied: pytz>=2020.1 in
c:\programdata\anaconda3\lib\site-packages (from
pandas!=1.4.0,<1.6,>1.1->ydata_profiling) (2021.3)
Requirement already satisfied: joblib>=0.14.1 in
c:\programdata\anaconda3\lib\site-packages (from
phik<0.13,>=0.11.1->ydata_profiling) (1.1.0)
Requirement already satisfied: typing-extensions>=4.2.0 in
c:\users\pc\appdata\roaming\python\python39\site-packages (from
pydantic<1.11,>=1.8.1->ydata_profiling) (4.5.0)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib<3.7,>=3.2->ydata_profiling)
(1.16.0)
Requirement already satisfied: certifi>=2017.4.17 in
c:\programdata\anaconda3\lib\site-packages (from
requests<2.29,>=2.24.0->ydata_profiling) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in
c:\programdata\anaconda3\lib\site-packages (from
requests<2.29,>=2.24.0->ydata_profiling) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
c:\programdata\anaconda3\lib\site-packages (from
```

```
requests<2.29,>=2.24.0->ydata_profiling) (1.26.9)
Requirement already satisfied: charset-normalizer~=2.0.0 in
c:\programdata\anaconda3\lib\site-packages (from
requests<2.29,>=2.24.0->ydata_profiling) (2.0.4)
Requirement already satisfied: patsy>=0.5.2 in
c:\programdata\anaconda3\lib\site-packages (from
statsmodels<0.14,>=0.13.2->ydata_profiling) (0.5.2)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-
packages (from tqdm<4.65,>=4.48.2->ydata_profiling) (0.4.4)
Note: you may need to restart the kernel to use updated packages.
```

[5]:
```
import ydata_profiling as pandas_profiling
```

[6]:
```
z_data=pd.read_csv("D:\P EDA\zomato.csv")
```

[7]:
```
z_data.shape
```

[7]:
```
(7105, 12)
```

[8]:
```
z_data.head(10)
```

[8]:

	Unnamed: 0.1	Unnamed: 0	restaurant name	\
0	0	0	#FeelTheROLL	
1	1	1	#L-81 Cafe	
2	2	2	#refuel	
3	3	3	'@ Biryani Central	
4	4	4	'@ The Bbq	
5	5	5	'@99	
6	6	6	'@Italy	
7	7	7	'@North Parontha Hut	
8	8	8	1000 B.C	
9	9	9	100Ã Â Ã Â Ã Â Ã Â Ã Â Ã Â Ã Â Ã Â°C	

	restaurant type	rate (out of 5)	num of ratings	avg cost (two people)	\
0	Quick Bites	3.4	7	200.0	
1	Quick Bites	3.9	48	400.0	
2	Cafe	3.7	37	400.0	
3	Casual Dining	2.7	135	550.0	
4	Casual Dining	2.8	40	700.0	
5	Takeaway, Delivery	3.4	37	200.0	
6	Casual Dining	4.1	305	700.0	
7	Takeaway, Delivery	2.8	40	300.0	
8	Quick Bites	3.2	49	300.0	
9	Casual Dining	3.7	41	450.0	

	online_order	table booking	\
0	No	No	

```
1          Yes            No
2          Yes            No
3          Yes            No
4          Yes            No
5           No            No
6          Yes            No
7           No            No
8          Yes            No
9           No            No

                                        cuisines type  \
0                                           Fast Food
1                                 Fast Food, Beverages
2                                      Cafe, Beverages
3                             Biryani, Mughlai, Chinese
4  BBQ, Continental, North Indian, Chinese, Bever…
5              Mughlai, Biryani, Chinese, North Indian
6                                               Italian
7                                         North Indian
8                     Arabian, Sandwich, Rolls, Burger
9                                Biryani, North Indian

                              area          local address
0                         Bellandur             Bellandur
1  Byresandra,Tavarekere,Madiwala                   HSR
2                Bannerghatta Road     Bannerghatta Road
3                     Marathahalli          Marathahalli
4                        Bellandur             Bellandur
5                       Whitefield            Whitefield
6                       Banashankari    Kumaraswamy Layout
7                        Indiranagar      Old Airport Road
8  Byresandra,Tavarekere,Madiwala  Koramangala 5th Block
9  Byresandra,Tavarekere,Madiwala                   BTM
```

[9]:
```
z_data.drop(['Unnamed: 0.1','Unnamed: 0','area'],axis=1,inplace=True)
```

[10]:
```
z_data.head(10)
```

[10]:
```
                       restaurant name     restaurant type  rate (out of 5)  \
0                          #FeelTheROLL         Quick Bites              3.4
1                            #L-81 Cafe         Quick Bites              3.9
2                               #refuel                Cafe              3.7
3                   '@ Biryani Central       Casual Dining              2.7
4                           '@ The Bbq       Casual Dining              2.8
5                                '@99  Takeaway, Delivery              3.4
6                              '@Italy       Casual Dining              4.1
7                  '@North Parontha Hut  Takeaway, Delivery              2.8
```

```
8                              1000 B.C         Quick Bites              3.2
9  100Ã Ã Ã Ã Ã Ã Ã Ã Ã Ã Ã Ã Ã Ã°C       Casual Dining              3.7

   num of ratings  avg cost (two people) online_order table booking  \
0               7                  200.0           No            No
1              48                  400.0          Yes            No
2              37                  400.0          Yes            No
3             135                  550.0          Yes            No
4              40                  700.0          Yes            No
5              37                  200.0           No            No
6             305                  700.0          Yes            No
7              40                  300.0           No            No
8              49                  300.0          Yes            No
9              41                  450.0           No            No

                                    cuisines type          local address
0                                       Fast Food              Bellandur
1                               Fast Food, Beverages                  HSR
2                                    Cafe, Beverages      Bannerghatta Road
3                            Biryani, Mughlai, Chinese           Marathahalli
4  BBQ, Continental, North Indian, Chinese, Bever…               Bellandur
5              Mughlai, Biryani, Chinese, North Indian            Whitefield
6                                            Italian     Kumaraswamy Layout
7                                        North Indian       Old Airport Road
8                    Arabian, Sandwich, Rolls, Burger  Koramangala 5th Block
9                               Biryani, North Indian                   BTM
```

[11]:
```
z_data.dtypes
```

[11]:
```
restaurant name           object
restaurant type           object
rate (out of 5)          float64
num of ratings             int64
avg cost (two people)    float64
online_order              object
table booking             object
cuisines type             object
local address             object
dtype: object
```

[12]:
```
z_data.describe(include='all')
```

[12]:
```
       restaurant name restaurant type  rate (out of 5)  num of ratings  \
count             7105            7105      7037.000000     7105.000000
unique            7105              81              NaN             NaN
top      #FeelTheROLL     Quick Bites              NaN             NaN
freq                1            2840              NaN             NaN
```

```
mean                NaN             NaN         3.514253      188.921042
std                 NaN             NaN         0.463249      592.171049
min                 NaN             NaN         1.800000        1.000000
25%                 NaN             NaN         3.200000       16.000000
50%                 NaN             NaN         3.500000       40.000000
75%                 NaN             NaN         3.800000      128.000000
max                 NaN             NaN         4.900000    16345.000000

        avg cost (two people) online_order table booking  \
count              7048.000000         7105          7105
unique                     NaN            2             2
top                        NaN          Yes            No
freq                       NaN         3727          6361
mean                540.286464          NaN           NaN
std                 462.902305          NaN           NaN
min                  40.000000          NaN           NaN
25%                 300.000000          NaN           NaN
50%                 400.000000          NaN           NaN
75%                 600.000000          NaN           NaN
max                6000.000000          NaN           NaN

                cuisines type local address
count                    7105          7105
unique                   2175            90
top     North Indian, Chinese    Whitefield
freq                      421           459
mean                      NaN           NaN
std                       NaN           NaN
min                       NaN           NaN
25%                       NaN           NaN
50%                       NaN           NaN
75%                       NaN           NaN
max                       NaN           NaN
```

[13]:
```
z_data.isnull().sum()
```

[13]:
```
restaurant name           0
restaurant type           0
rate (out of 5)          68
num of ratings            0
avg cost (two people)    57
online_order              0
table booking             0
cuisines type             0
local address             0
dtype: int64
```

[14]:
```
profile=pandas_profiling.ProfileReport(z_data)
profile.to_file(output_file='zomato_before_EDA.html')
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]

Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]

Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]

Export report to file:   0%|          | 0/1 [00:00<?, ?it/s]
```

[15]:
```
avg=z_data['rate (out of 5)'].mean()
avg
```

[15]: 3.5142532329117144

[16]:
```
avg=round(avg,1)
avg
```

[16]: 3.5

[17]:
```
z_data['rate (out of 5)'].fillna(avg,inplace=True)
```

[18]:
```
z_data['rate (out of 5)'].isnull().sum()
```

[18]: 0

[19]:
```
a=z_data['avg cost (two people)'].mean()
a
```

[19]: 540.2864642451759

[20]:
```
z_data['avg cost (two people)'].fillna(a,inplace=True)
```

[21]:
```
z_data['avg cost (two people)'].isnull().sum()
```

[21]: 0

[22]:
```
z_data['avg cost (two people)']=z_data['avg cost (two people)'].round().
↪astype(int)
```

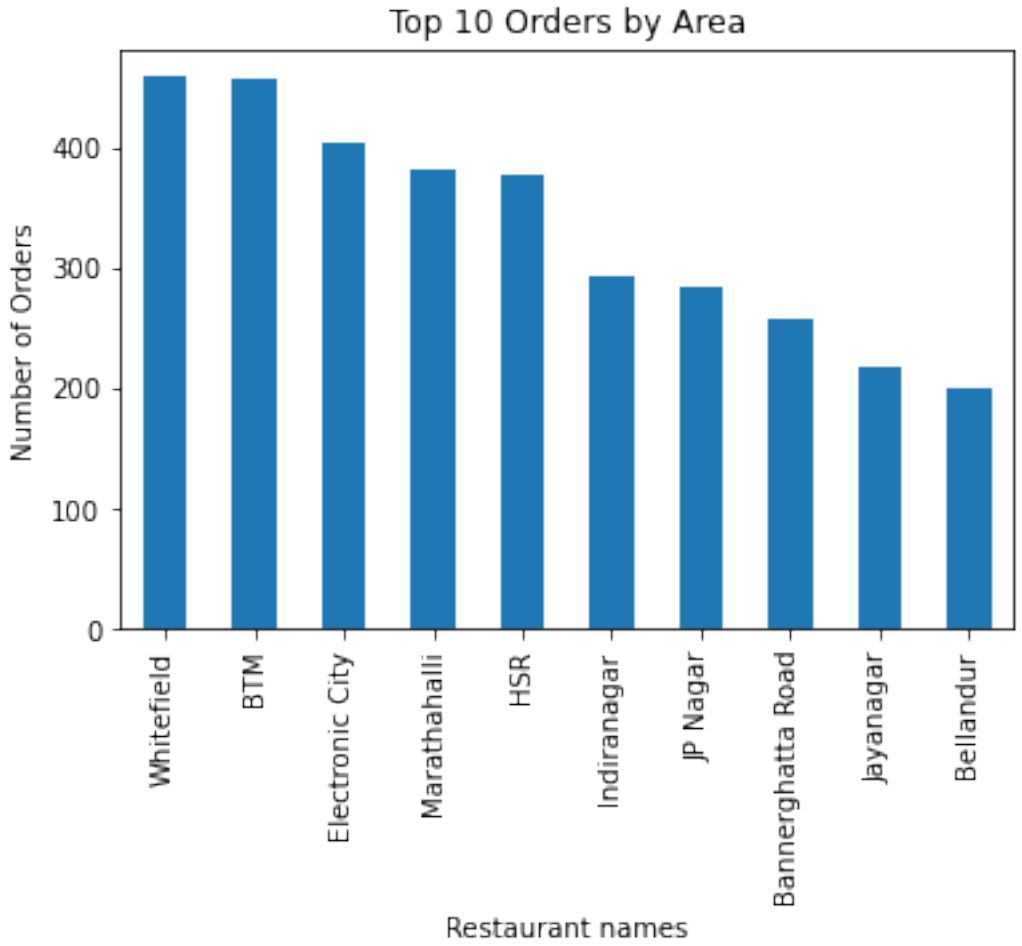
[23]:
```
z_data.duplicated().sum()
```

[23]: 0

[68]:
```
a=z_data['local address'].value_counts().head(10)
a
```

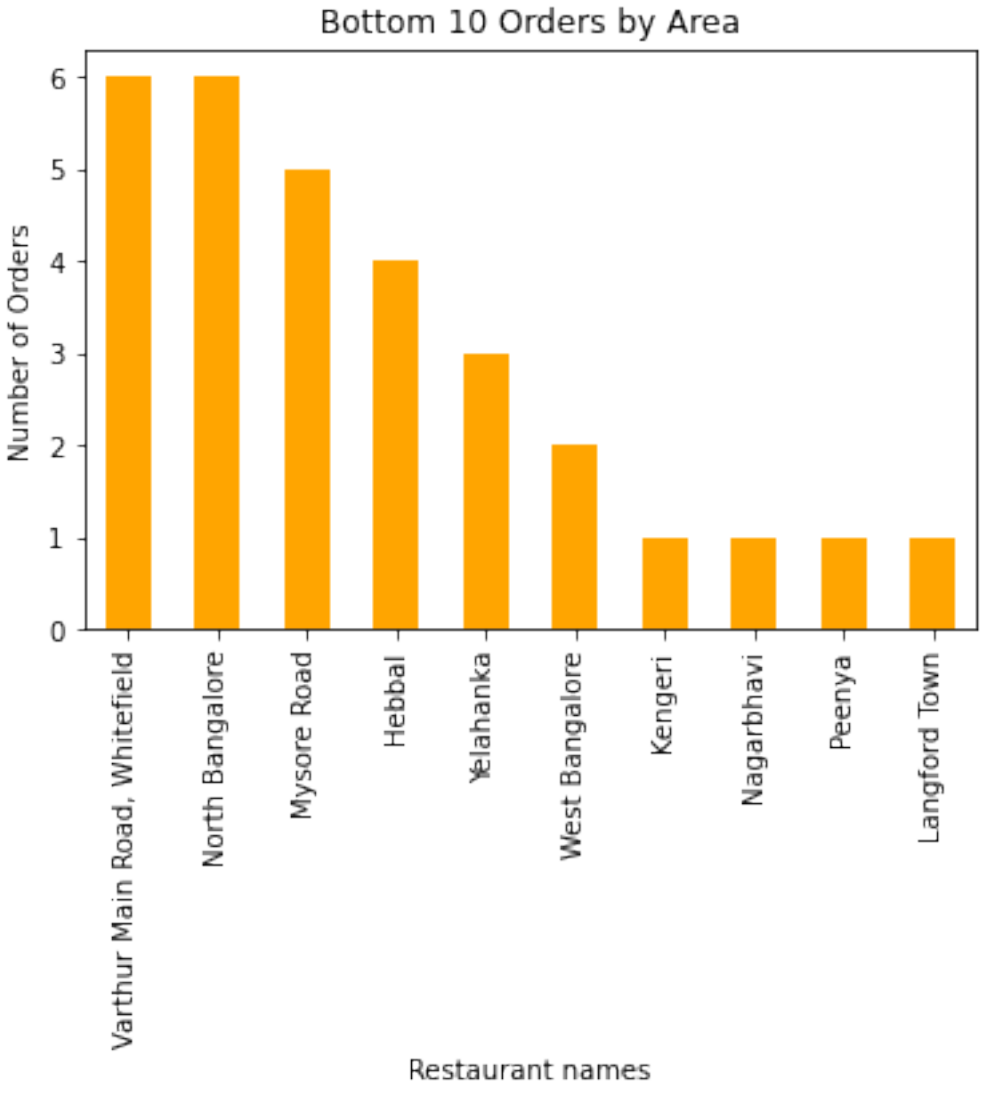
[68]:
```
Whitefield           459
BTM                  458
Electronic City      404
Marathahalli         381
HSR                  378
Indiranagar          292
JP Nagar             283
Bannerghatta Road    257
Jayanagar            218
Bellandur            199
Name: local address, dtype: int64
```

[162]:
```
a.plot(kind='bar')
plt.xlabel('Restaurant names')
plt.ylabel('Number of Orders')
plt.title('Top 10 Orders by Area')
plt.xticks(rotation=90)
plt.show()
```



[113]:
```
areas = z_data['local address'].value_counts().tail(10)
```

[222]:
```
areas.plot(kind='bar',color='orange')
plt.xlabel('Restaurant names')
plt.ylabel('Number of Orders')
plt.title('Bottom 10 Orders by Area')
plt.xticks(rotation=90)
plt.show()
```

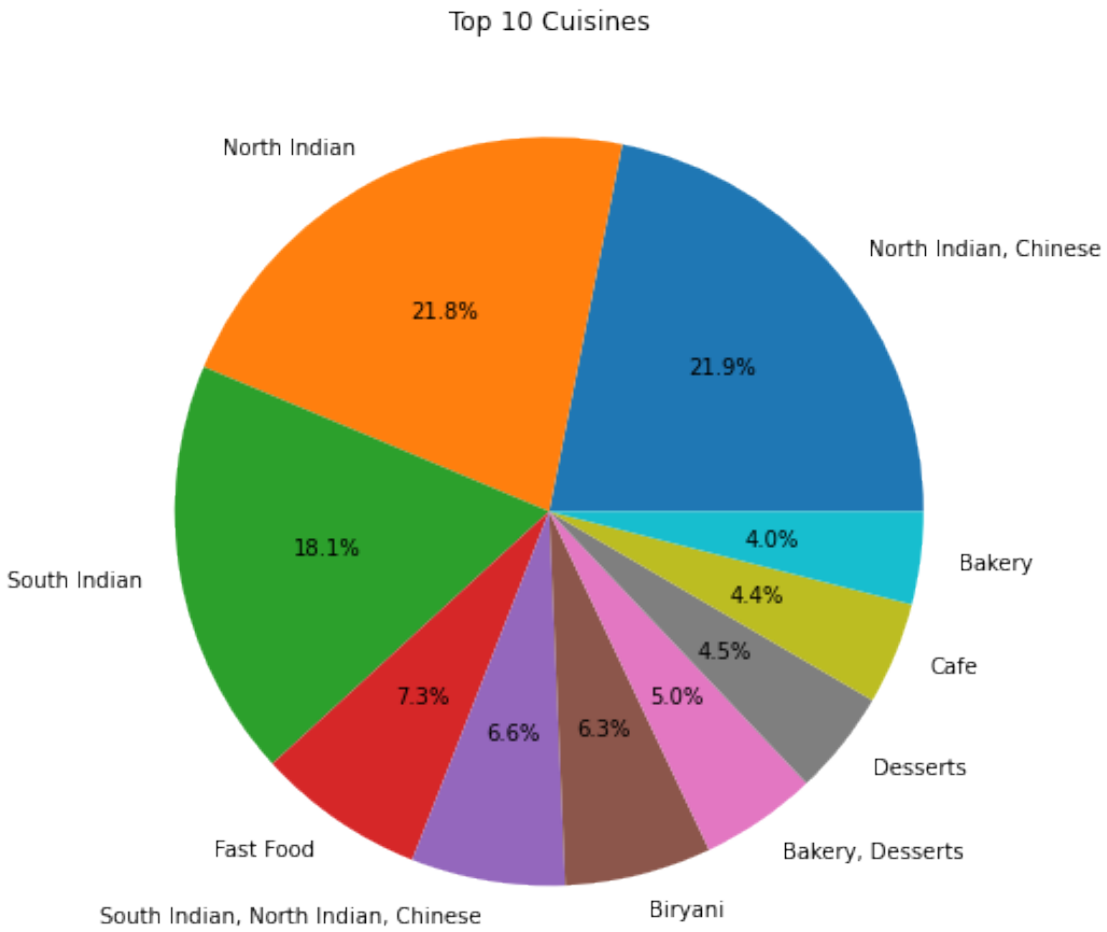


[115]:
```
Top_Cuisines=z_data['cuisines type'].value_counts().head(10)
```

[116]:
```
Cuisines_name=z_data['cuisines type'].value_counts().head(10).index
Cuisines_name
```

[116]:
```
Index(['North Indian, Chinese', 'North Indian', 'South Indian', 'Fast Food',
       'South Indian, North Indian, Chinese', 'Biryani', 'Bakery, Desserts',
       'Desserts', 'Cafe', 'Bakery'],
      dtype='object')
```

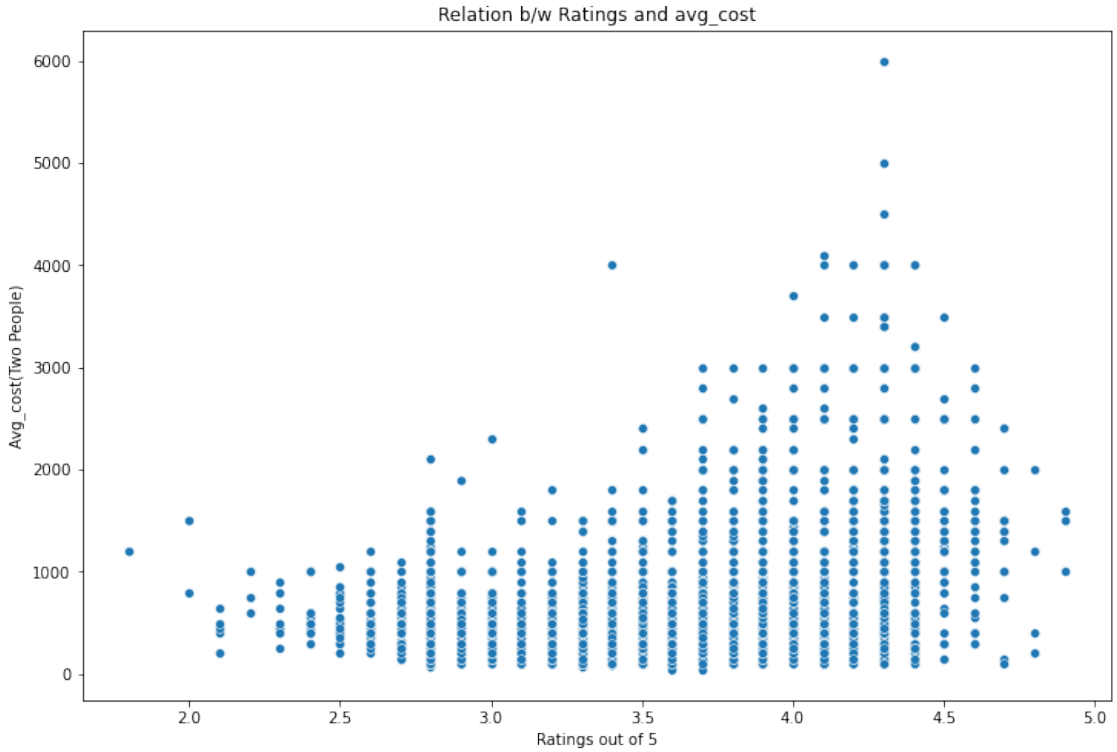
[160]:
```
Top_Cuisines.plot(kind='pie',autopct='%1.1f%%',figsize=(8,8))
plt.title('Top 10 Cuisines')
plt.ylabel('')
plt.xticks(rotation=90)
plt.show()
```



[173]:
```
plt.figure(figsize=(6, 6))
sns.countplot(data=z_data, x='online_order')
```

```
plt.title('Distribution of Online Orders')
plt.xlabel('Online Orders')
plt.ylabel('Count')
plt.show()
```

[215]:
```
plt.figure(figsize=(12,8))
sns.scatterplot(data=z_data,x='rate (out of 5)',y='avg cost (two people)')
plt.title('Relation b/w Ratings and avg_cost')
plt.xlabel('Ratings out of 5')
plt.ylabel('Avg_cost(Two People)')
plt.show()
```


Relation b/w Ratings and avg_cost
Avg_cost(Two People)
Ratings out of 5
0
1000
2000
3000
4000
5000
6000
2.0
2.5
3.0
3.5
4.0
4.5
5.0

[ ]: