

hackthon

April 19, 2025

```
[70]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import folium
from geopy.geocoders import Nominatim
from geopy.exc import GeocoderTimedOut
from IPython.display import display, HTML
import time
```

```
[72]: zomato_df = pd.read_csv('zomato_data.csv')
geo_df = pd.read_csv('Geographical Coordinates.csv')
```

```
[74]: zomato_df.info()
zomato_df.isnull().sum()
zomato_df.describe(include='all')
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   online_order                          51717 non-null  object
1   book_table                            51717 non-null  object
2   rate                                  43942 non-null  object
3   votes                                51717 non-null  int64
4   rest_type                             51490 non-null  object
5   dish_liked                           23639 non-null  object
6   cuisines                             51672 non-null  object
7   approx_costfor_two_people             51371 non-null  object
8   listed_intype                         51717 non-null  object
9   listed_incity                         51717 non-null  object
dtypes: int64(1), object(9)
memory usage: 3.9+ MB
```

```
[74]:      online_order  book_table  rate  votes  rest_type  dish_liked  \
count          51717      51717  43942  51717.000000      51490      23639
unique           2           2      64           NaN          93          5271
```

top	Yes	No	NEW	NaN	Quick Bites	Biryani
freq	30444	45268	2208	NaN	19132	182
mean	NaN	NaN	NaN	283.697527	NaN	NaN
std	NaN	NaN	NaN	803.838853	NaN	NaN
min	NaN	NaN	NaN	0.000000	NaN	NaN
25%	NaN	NaN	NaN	7.000000	NaN	NaN
50%	NaN	NaN	NaN	41.000000	NaN	NaN
75%	NaN	NaN	NaN	198.000000	NaN	NaN
max	NaN	NaN	NaN	16832.000000	NaN	NaN

	cuisines	approx_costfor_two_people	listed_intype	listed_incite
count	51672	51371	51717	51717
unique	2723	70	7	30
top	North Indian	300	Delivery	BTM
freq	2913	7576	25942	3279
mean	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN

```
[84]: zomato_df['rate'] = zomato_df['rate'].replace(['-', 'NEW'], np.nan) # Handle
      ↪ placeholders
zomato_df['rate'] = zomato_df['rate'].astype(str).str.replace('/5', '',
      ↪ regex=False) # Remove "/5"
zomato_df['rate'] = pd.to_numeric(zomato_df['rate'], errors='coerce') #
      ↪ Convert to float
zomato_df['rate'] = zomato_df['rate'].fillna(zomato_df['rate'].median()) #
      ↪ Fill NaNs with median
```

```
[88]: zomato_df['approx_costfor_two_people'] = zomato_df['approx_costfor_two_people'].
      ↪ astype(str).str.replace(',', '')
zomato_df['approx_costfor_two_people'] = pd.
      ↪ to_numeric(zomato_df['approx_costfor_two_people'], errors='coerce')
zomato_df['approx_costfor_two_people'] = zomato_df['approx_costfor_two_people'].
      ↪ fillna(zomato_df['approx_costfor_two_people'].median())
```

```
[92]: zomato_df['dish_liked'] = zomato_df['dish_liked'].fillna('Not Available')
zomato_df['cuisines'] = zomato_df['cuisines'].fillna('Other')
zomato_df['rest_type'] = zomato_df['rest_type'].fillna('Unknown')
```

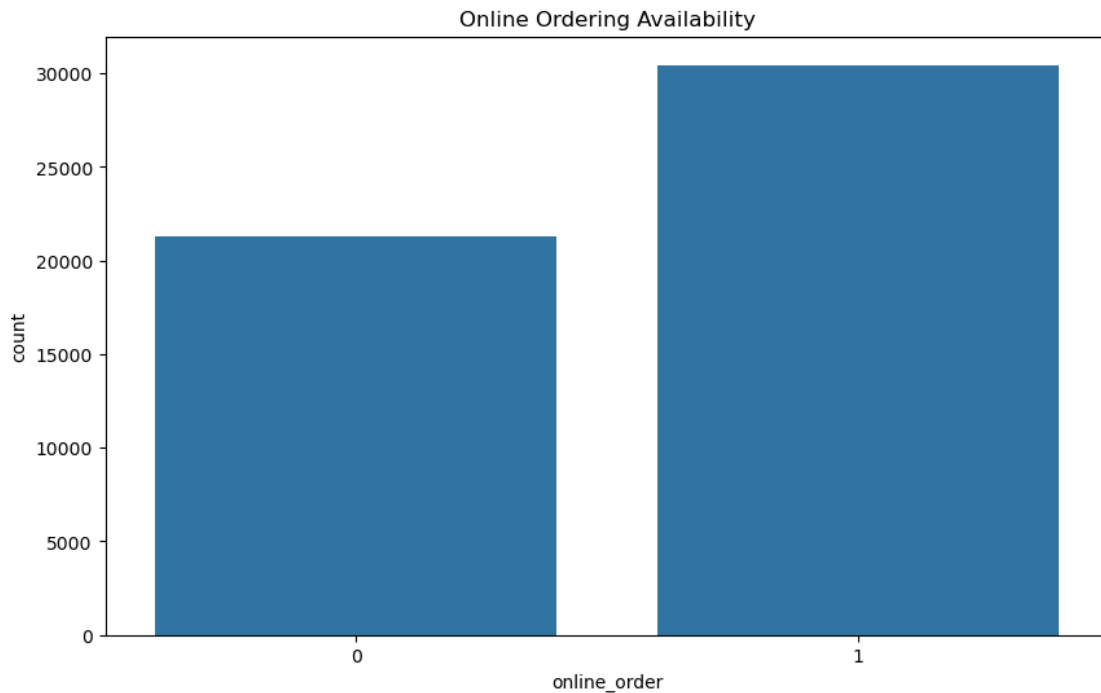
```
[96]: zomato_df['votes'] = zomato_df['votes'].fillna(zomato_df['votes'].median())
```

```
[98]: zomato_df['online_order'] = zomato_df['online_order'].map({'Yes': 1, 'No': 0})
      zomato_df['book_table'] = zomato_df['book_table'].map({'Yes': 1, 'No': 0})
```

```
[100]: zomato_df['votes'] = zomato_df['votes'].astype(int)
      zomato_df['approx_costfor_two_people'] = zomato_df['approx_costfor_two_people'].
      ↪astype(int)
      zomato_df['rate'] = zomato_df['rate'].astype(float)
```

```
[102]: merged_df = pd.merge(zomato_df, geo_df, on='listed_incity', how='left')
```

```
[104]: plt.figure(figsize=(10, 6))
      sns.countplot(data=merged_df, x='online_order')
      plt.title('Online Ordering Availability')
      plt.show()
```



```
[108]: merged_df_clean = merged_df.dropna(subset=['Latitude', 'Longitude'])
      m_density = folium.Map(location=[merged_df_clean['Latitude'].mean(),
      ↪merged_df_clean['Longitude'].mean()], zoom_start=11)
      for idx, row in merged_df_clean.iterrows():
          folium.CircleMarker(
              location=[row['Latitude'], row['Longitude']],
              radius=2,
              color='blue',
              fill=True,
```

```

        fill_color='blue',
        fill_opacity=0.5
    ).add_to(m_density)

m_density

```

[108]: <folium.folium.Map at 0x2bf538daff0>

```

[112]: italian_restaurants_clean = italian_restaurants.dropna(subset=['Latitude',
↳ 'Longitude'])
m_italian = folium.Map(
    location=[italian_restaurants_clean['Latitude'].mean(),
↳ italian_restaurants_clean['Longitude'].mean()],
    zoom_start=12
)
for idx, row in italian_restaurants_clean.iterrows():
    folium.Marker(
        location=[row['Latitude'], row['Longitude']],
        popup=row.get('name', row.get('Name', 'Restaurant'))
    ).add_to(m_italian)

m_italian

```

[112]: <folium.folium.Map at 0x2bf647a69c0>

```

[114]: merged_df.info()
merged_df.isnull().sum()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51717 entries, 0 to 51716
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   online_order                          51717 non-null  int64
1   book_table                            51717 non-null  int64
2   rate                                  51717 non-null  float64
3   votes                                 51717 non-null  int32
4   rest_type                             51717 non-null  object
5   dish_liked                           51717 non-null  object
6   cuisines                              51717 non-null  object
7   approx_costfor_two_people             51717 non-null  int32
8   listed_intype                         51717 non-null  object
9   listed_incity                         51717 non-null  object
10  Latitude                              46137 non-null  float64
11  Longitude                             46137 non-null  float64
dtypes: float64(3), int32(2), int64(2), object(5)
memory usage: 4.3+ MB

```

```
[114]: online_order          0
       book_table           0
       rate                 0
       votes                0
       rest_type            0
       dish_liked           0
       cuisines             0
       approx_costfor_two_people 0
       listed_intype        0
       listed_incitey       0
       Latitude             5580
       Longitude            5580
       dtype: int64
```

```
[ ]:
```