swiggy

April 19, 2023

1 CS 6220 Final Project

2 Analyzing Swiggy Restaurant Catalog Data

2.1 Group: Amber

2.1.1 Members: Sudhanva Narayana, Shreyas Prasad, Matthew Chan

2.2 Table of Contents

```python
[1]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
```

```python
[2]: plots_path = './plots/'
```

```python
[3]: df = pd.read_csv('swiggy.csv')
```

2.3 Quick Inspection of the Data

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

```
[4]:        id              name    city rating      rating_count  cost  \
     0  567335     AB FOODS POINT  Abohar     --  Too Few Ratings   200
     1  531342  Janta Sweet House  Abohar    4.4       50+ ratings   200
     2  158203  theka coffee desi  Abohar    3.8      100+ ratings   100
     3  187912         Singh Hut  Abohar    3.7       20+ ratings   250
     4  543530      GRILL MASTERS  Abohar     --  Too Few Ratings   250

                      cuisine          lic_no  \
     0        Beverages,Pizzas  22122652000138
     1           Sweets,Bakery  12117201000112
     2               Beverages  22121652000190
     3        Fast Food,Indian  22119652000167
     4  Italian-American,Fast Food  12122201000053
```

```
                                                             link  \
0  https://www.swiggy.com/restaurants/ab-foods-po…
1  https://www.swiggy.com/restaurants/janta-sweet…
2  https://www.swiggy.com/restaurants/theka-coffe…
3  https://www.swiggy.com/restaurants/singh-hut-n…
4  https://www.swiggy.com/restaurants/grill-maste…


                                                address               menu
0  AB FOODS POINT, NEAR RISHI NARANG DENTAL CLINI…  Menu/567335.json
1  Janta Sweet House, Bazar No.9, Circullar Road,…  Menu/531342.json
2        theka coffee desi, sahtiya sadan road city  Menu/158203.json
3    Singh Hut, CIRCULAR ROAD NEAR NEHRU PARK ABOHAR  Menu/187912.json
4  GRILL MASTERS, ADA Heights, Abohar - Hanumanga…  Menu/543530.json
```

[5]: `df.tail()`

[5]:
```
            id                     name      city rating    rating_count  \
148536  553122         The Food Delight  Yavatmal     --  Too Few Ratings
148537  562647  MAITRI FOODS & BEVERAGES  Yavatmal     --  Too Few Ratings
148538  559435           Cafe Bella Ciao  Yavatmal     --  Too Few Ratings
148539  418989               GRILL ZILLA  Yavatmal     --  Too Few Ratings
148540  447770           Lazeez kitchen  Yavatmal     --  Too Few Ratings


        cost           cuisine          lic_no  \
148536   200  Fast Food,Snacks   21522053000452
148537   300            Pizzas          license
148538   300  Fast Food,Snacks   21522251000378
148539   250       Continental   21521251000241
148540   200            Pizzas   21521251000634


                                                     link  \
148536  https://www.swiggy.com/restaurants/the-food-de…
148537  https://www.swiggy.com/restaurants/maitri-food…
148538  https://www.swiggy.com/restaurants/cafe-bella-…
148539  https://www.swiggy.com/restaurants/grill-zilla…
148540  https://www.swiggy.com/restaurants/lazeez-kitc…


                                                  address               menu
148536  The Food Delight, 94MC+X35, New Singhania Naga…  Menu/553122.json
148537  MAITRI FOODS & BEVERAGES, POLIC MITRYA SOCIETY…  Menu/562647.json
148538  Cafe Bella Ciao, SHOP NO 2 NEMANI MARKET SBI S…  Menu/559435.json
148539  GRILL ZILLA, SHO NO 2/6, POSTEL GROUND CHOWPAT…  Menu/418989.json
148540  Lazeez kitchen, 94G3+2RR, Wadgaon, Yavatmal, M…  Menu/447770.json
```

[6]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148541 entries, 0 to 148540
Data columns (total 11 columns):
 #   Column        Non-Null Count   Dtype
---  ------        --------------   -----
 0   id            148541 non-null  int64
 1   name          148455 non-null  object
 2   city          148541 non-null  object
 3   rating        148455 non-null  object
 4   rating_count  148455 non-null  object
 5   cost          148410 non-null  object
 6   cuisine       148442 non-null  object
 7   lic_no        148312 non-null  object
 8   link          148541 non-null  object
 9   address       148455 non-null  object
 10  menu          148541 non-null  object
dtypes: int64(1), object(10)
memory usage: 12.5+ MB
```

[7]: `df.describe(include='all')`

[7]:

	id	name	city	rating	rating_count
count	148541.000000	148455	148541	148455	148455
unique	NaN	112818	821	42	8
top	NaN	Domino's Pizza	Bikaner	--	Too Few Ratings
freq	NaN	442	1666	87014	87014
mean	363466.378912	NaN	NaN	NaN	NaN
std	167890.977174	NaN	NaN	NaN	NaN
min	211.000000	NaN	NaN	NaN	NaN
25%	233320.000000	NaN	NaN	NaN	NaN
50%	412628.000000	NaN	NaN	NaN	NaN
75%	502223.000000	NaN	NaN	NaN	NaN
max	581031.000000	NaN	NaN	NaN	NaN

	cost	cuisine	lic_no
count	148410	148442	148312
unique	363	2132	108763
top	200	North Indian,Chinese	license
freq	38635	6471	12861
mean	NaN	NaN	NaN
std	NaN	NaN	NaN
min	NaN	NaN	NaN
25%	NaN	NaN	NaN
50%	NaN	NaN	NaN
75%	NaN	NaN	NaN
max	NaN	NaN	NaN

```
                                                             link  \
count                                                      148541
unique                                                     148541
top     https://www.swiggy.com/restaurants/ab-foods-po…
freq                                                            1
mean                                                          NaN
std                                                           NaN
min                                                           NaN
25%                                                           NaN
50%                                                           NaN
75%                                                           NaN
max                                                           NaN

                                          address                menu
count                                      148455              148541
unique                                     148401              148541
top     Gold Star Biryani, RS Road Dindigul - 624001   Menu/567335.json
freq                                            3                   1
mean                                          NaN                 NaN
std                                           NaN                 NaN
min                                           NaN                 NaN
25%                                           NaN                 NaN
50%                                           NaN                 NaN
75%                                           NaN                 NaN
max                                           NaN                 NaN
```

[8]: `df.columns`

[8]: ```
Index(['id', 'name', 'city', 'rating', 'rating_count', 'cost', 'cuisine',
       'lic_no', 'link', 'address', 'menu'],
      dtype='object')
```

## 2.4 Important Features that could be useful

- **city**: Location can be a major aspect of how restaurants and their properties can change
- **rating**: Could be a major factor in deciding the quality of the restaurant
- **rating_count**: Could be a major factor in deciding the authenticity of the ratings
- **cost**: Important feature for deciding how to price the restaurants
- **cuisine**: Could be a major factor in categorizing the restaurants based on trends and type of food

[9]: ```python
def preprocess_city(entry):
    # Split the entry by comma separator
    values = entry.split(',')

    # Check number of values and modify as necessary
    if len(values) == 1:
```

```python
        # Prepend the same value twice
        values = [values[0], values[0], values[0]]
    elif len(values) == 2:
        # Prepend the first value to the list
        values = [values[0]] + values

    # Join the values back into a comma separated string
    return ','.join(values)

def preprocess_cuisine(entry):
    values = entry.split(',')

    if len(values) == 1:
        values = [values[0]] + values

        # Join the values back into a comma separated string
        return ','.join(values)
    return ','.join(values)
```

[10]: ```python
df['city'].value_counts()
```

[10]: 
```
Bikaner   1666
Noida-1   1428
Indirapuram,Delhi   1279
BTM,Bangalore   1161
Rohini,Delhi   1136
                          ...
Alwarpet,Chennai   1
Naharlagun   1
Mahim Dadar,Mumbai   1
Starbucks_BKC,Mumbai   1
Rangpo   1
Name: city, Length: 821, dtype: int64
```

[11]: ```python
df['city'][df['city'].isna()]
```

[11]: ```
Series([], Name: city, dtype: object)
```

### 2.4.1 Imputing data

[12]: ```python
df['city'] = df['city'].apply(preprocess_city)
```

[13]: ```python
df['city'].value_counts()
```

[13]: 
```
Bikaner,Bikaner,Bikaner   1666
Noida-1,Noida-1,Noida-1   1428
Indirapuram,Indirapuram,Delhi   1279
```

```
BTM,BTM,Bangalore                      1161
Rohini,Rohini,Delhi                    1136
                                        …
Alwarpet,Alwarpet,Chennai                 1
Naharlagun,Naharlagun,Naharlagun          1
Mahim Dadar,Mahim Dadar,Mumbai            1
Starbucks_BKC,Starbucks_BKC,Mumbai        1
Rangpo,Rangpo,Rangpo                      1
Name: city, Length: 821, dtype: int64
```

Spitting the city column into sub_area, area and city

```
[14]: df[['sub_area', 'area', 'city']] = df['city'].str.split(',', expand=True)
```

```
[15]: df
```

```
[15]:            id                      name        city rating       rating_count  \
      0       567335          AB FOODS POINT       Abohar     --     Too Few Ratings
      1       531342       Janta Sweet House       Abohar    4.4         50+ ratings
      2       158203       theka coffee desi       Abohar    3.8        100+ ratings
      3       187912               Singh Hut       Abohar    3.7         20+ ratings
      4       543530           GRILL MASTERS       Abohar     --     Too Few Ratings
      …          …                       …            …      …                  …
      148536  553122         The Food Delight     Yavatmal     --     Too Few Ratings
      148537  562647  MAITRI FOODS & BEVERAGES   Yavatmal     --     Too Few Ratings
      148538  559435         Cafe Bella Ciao     Yavatmal     --     Too Few Ratings
      148539  418989              GRILL ZILLA    Yavatmal     --     Too Few Ratings
      148540  447770          Lazeez kitchen     Yavatmal     --     Too Few Ratings

              cost                 cuisine           lic_no  \
      0        200       Beverages,Pizzas   22122652000138
      1        200          Sweets,Bakery   12117201000112
      2        100              Beverages   22121652000190
      3        250       Fast Food,Indian   22119652000167
      4        250  Italian-American,Fast Food  12122201000053
      …          …                     …                …
      148536   200       Fast Food,Snacks   21522053000452
      148537   300                 Pizzas          license
      148538   300       Fast Food,Snacks   21522251000378
      148539   250            Continental   21521251000241
      148540   200                 Pizzas   21521251000634

                                              link  \
      0         https://www.swiggy.com/restaurants/ab-foods-po…
      1         https://www.swiggy.com/restaurants/janta-sweet…
      2         https://www.swiggy.com/restaurants/theka-coffe…
      3         https://www.swiggy.com/restaurants/singh-hut-n…
```

6

```
4         https://www.swiggy.com/restaurants/grill-maste…
…                                                       …
148536    https://www.swiggy.com/restaurants/the-food-de…
148537    https://www.swiggy.com/restaurants/maitri-food…
148538    https://www.swiggy.com/restaurants/cafe-bella-…
148539    https://www.swiggy.com/restaurants/grill-zilla…
148540    https://www.swiggy.com/restaurants/lazeez-kitc…

                                              address              menu  \
0         AB FOODS POINT, NEAR RISHI NARANG DENTAL CLINI…  Menu/567335.json
1         Janta Sweet House, Bazar No.9, Circullar Road,…  Menu/531342.json
2                 theka coffee desi, sahtiya sadan road city  Menu/158203.json
3           Singh Hut, CIRCULAR ROAD NEAR NEHRU PARK ABOHAR  Menu/187912.json
4         GRILL MASTERS, ADA Heights, Abohar - Hanumanga…  Menu/543530.json
…                                                       …                 …
148536    The Food Delight, 94MC+X35, New Singhania Naga…  Menu/553122.json
148537    MAITRI FOODS & BEVERAGES, POLIC MITRYA SOCIETY…  Menu/562647.json
148538    Cafe Bella Ciao, SHOP NO 2 NEMANI MARKET SBI S…  Menu/559435.json
148539    GRILL ZILLA, SHO NO 2/6, POSTEL GROUND CHOWPAT…  Menu/418989.json
148540    Lazeez kitchen, 94G3+2RR, Wadgaon, Yavatmal, M…  Menu/447770.json

          sub_area      area
0           Abohar    Abohar
1           Abohar    Abohar
2           Abohar    Abohar
3           Abohar    Abohar
4           Abohar    Abohar
…              …         …
148536    Yavatmal  Yavatmal
148537    Yavatmal  Yavatmal
148538    Yavatmal  Yavatmal
148539    Yavatmal  Yavatmal
148540    Yavatmal  Yavatmal

[148541 rows x 13 columns]
```

2.4.2   Quick look after imputation

```
[16]: df[['sub_area','area','city']].value_counts()
```

```
[16]: sub_area      area          city
      Bikaner       Bikaner       Bikaner      1666
      Noida-1       Noida-1       Noida-1      1428
      Indirapuram   Indirapuram   Delhi        1279
      BTM           BTM           Bangalore    1161
      Rohini        Rohini        Delhi        1136
                                                 …
```

```
BBK_MayurVihar    BBK_MayurVihar    Delhi         1
Mahim Dadar       Mahim Dadar       Mumbai        1
Rangpo            Rangpo            Rangpo        1
Manali            Manali            Manali        1
Kohima            Kohima            Kohima        1
Length: 821, dtype: int64
```

2.4.3   Visualising Ratings

```python
[17]:  # Define a colormap
       cmap = sns.color_palette('tab10', len(df['rating_count'].value_counts()))

       # Create a bar plot of the rating counts
       ax = df['rating_count'].value_counts().plot(kind='bar', color=cmap,
        ↪figsize=(12, 8))

       # Set the title of the plot
       ax.set_title('Distribution of Ratings')

       # Set the label for the x-axis
       ax.set_xlabel('Number of Ratings')

       # Set the label for the y-axis
       ax.set_ylabel('Frequency')

       # Add a legend to the plot
       ax.legend(['Count'], loc='upper right')

       # Display the values on top of the bars
       for i in ax.containers:
           ax.bar_label(i, label_type='edge', fontsize=10, padding=8)

       # Display the plot
       plt.savefig(plots_path + 'distribution-of-ratings.png', bbox_inches='tight')
       plt.show()
```
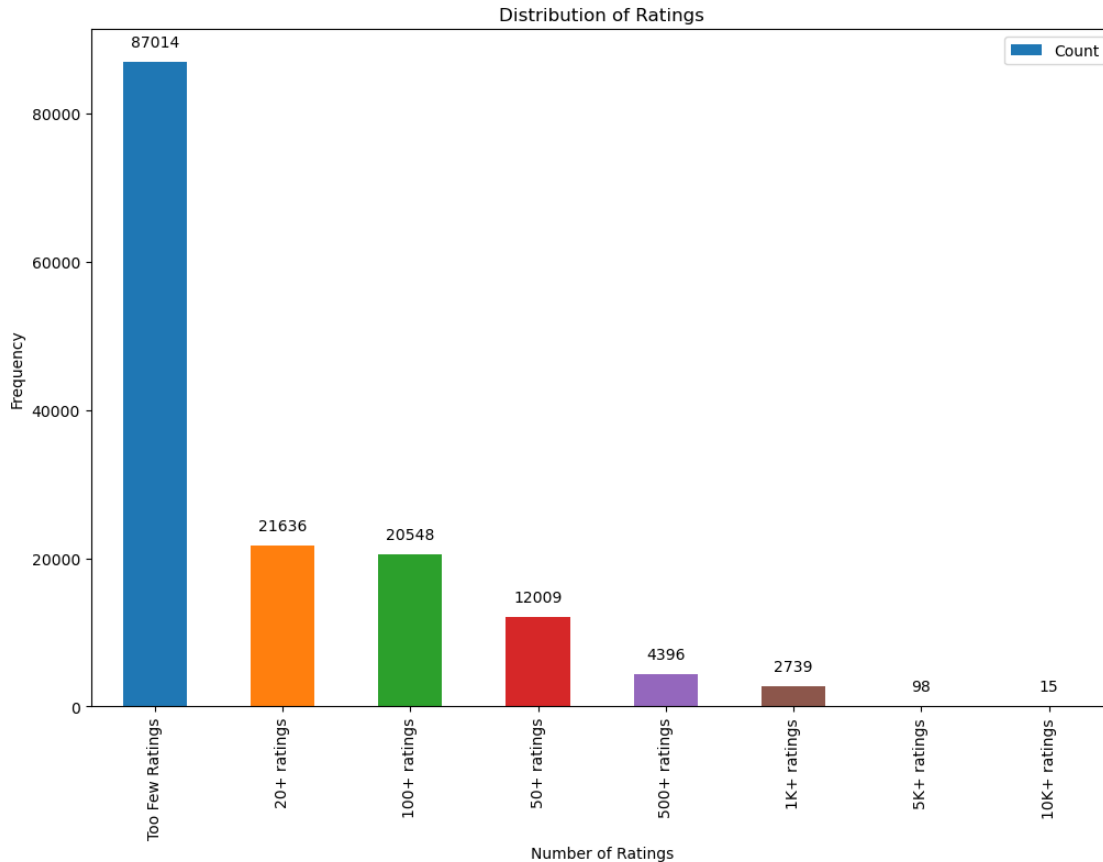
```
[18]: df['cost'] = df['cost'].str.slice(start=1).astype('float')
```

```
[19]: def filter_outliers(df, column_name):
          mean = df[column_name].mean()
          std = df[column_name].std()
          filtered = df[column_name][(df[column_name] > mean - 2 * std) &↵
      ↪(df[column_name] < mean + 2 * std)]
          return filtered
```

2.4.4 Visualising Cost

```
[20]: df['cost'] = filter_outliers(df, 'cost')
```

```
[21]: df['cost'] = df['cost'].fillna(df['cost'].median())
```

```
[22]: # Set up the histogram plot
      fig, ax = plt.subplots(figsize=(12, 8))
      n, bins, patches = ax.hist(df['cost'], bins=10)
```

```python
# Customize plot details
ax.set_title('Cost Distribution')
ax.set_xlabel('Cost')
ax.set_ylabel('Frequency')
ax.set_xticks(bins)
ax.set_xticklabels(['${:.0f}'.format(b) for b in bins])
ax.legend(['Cost'])

# Add value labels to the top of each bar
for i in range(len(patches)):
    x = patches[i].get_x() + patches[i].get_width() / 2
    y = patches[i].get_height()
    ax.annotate('{:.0f}'.format(y), (x, y), xytext=(0, 5),
                textcoords='offset points', ha='center', va='bottom')

# Display the plot
plt.savefig(plots_path + 'cost-distribution.png', bbox_inches='tight')
plt.show()
```
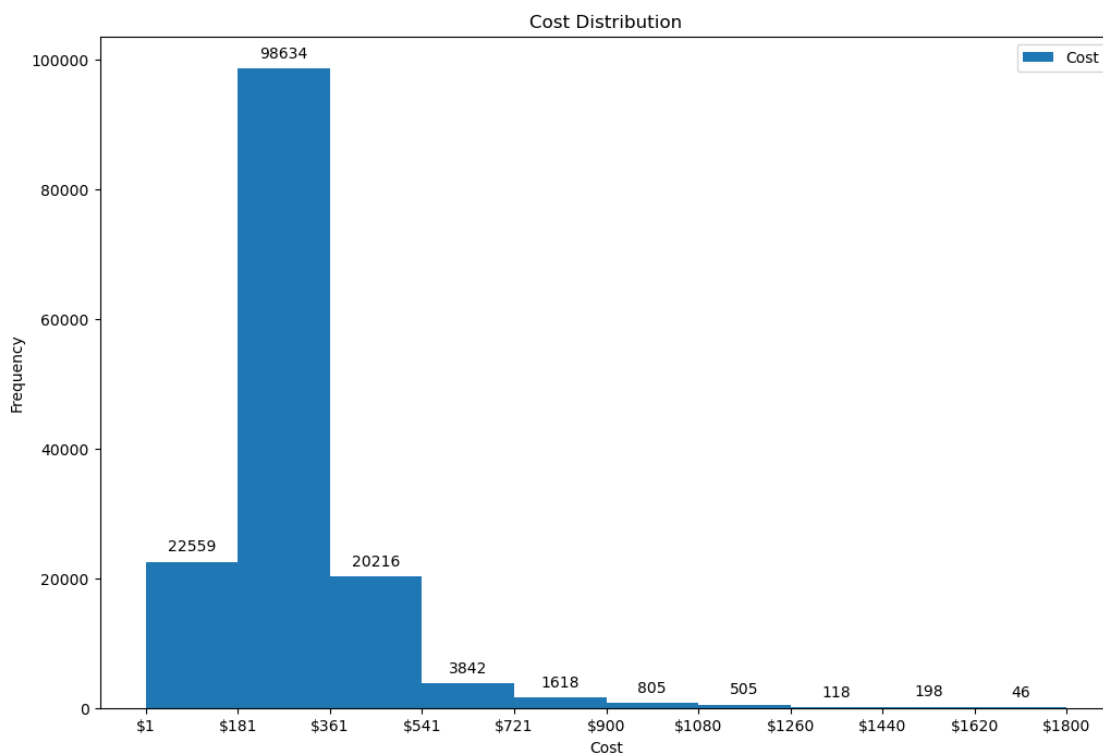
2.4.5 Analsing Cuisime

```
[23]: df['cuisine']
```

```
[23]: 0                 Beverages,Pizzas
      1                   Sweets,Bakery
      2                       Beverages
      3                Fast Food,Indian
      4         Italian-American,Fast Food
                           …
      148536          Fast Food,Snacks
      148537                    Pizzas
      148538          Fast Food,Snacks
      148539               Continental
      148540                    Pizzas
      Name: cuisine, Length: 148541, dtype: object
```

```
[24]: df['cuisine'].value_counts()
```

```
[24]: North Indian,Chinese       6471
      Indian                     6414
      Chinese                    5051
      North Indian               4775
      Indian,Chinese             4374
                                  …
      Vietnamese,Pan-Asian          1
      Healthy Food,Pan-Asian        1
      Home Food,Naga                1
      Korean,Desserts               1
      Barbecue,Italian-American     1
      Name: cuisine, Length: 2132, dtype: int64
```

```
[25]: df = df.dropna(subset=['cuisine']).copy()
```

```
[26]: df.loc[:, 'cuisine'] = df['cuisine'].apply(preprocess_cuisine)
```

```
[27]: df['cuisine']
```

```
[27]: 0                 Beverages,Pizzas
      1                   Sweets,Bakery
      2             Beverages,Beverages
      3                Fast Food,Indian
      4         Italian-American,Fast Food
                           …
      148536          Fast Food,Snacks
      148537             Pizzas,Pizzas
      148538          Fast Food,Snacks
```

```
148539          Continental,Continental
148540                    Pizzas,Pizzas
Name: cuisine, Length: 148442, dtype: object
```

[28]:
```python
cuisine_counts = df['cuisine'].str.split(',', expand=True).stack().
 ↪value_counts()


# Set up the bar plot
fig, ax = plt.subplots(figsize=(30,8))
cuisine_counts.plot(kind='bar')

# Customize plot details
ax.set_title('Cuisine Counts')
ax.set_xlabel('Cuisine')
ax.set_ylabel('Count')
ax.legend(['Count'])

# Display the plot
plt.savefig(plots_path + 'cusine_counts.png', bbox_inches='tight')
plt.show()
```
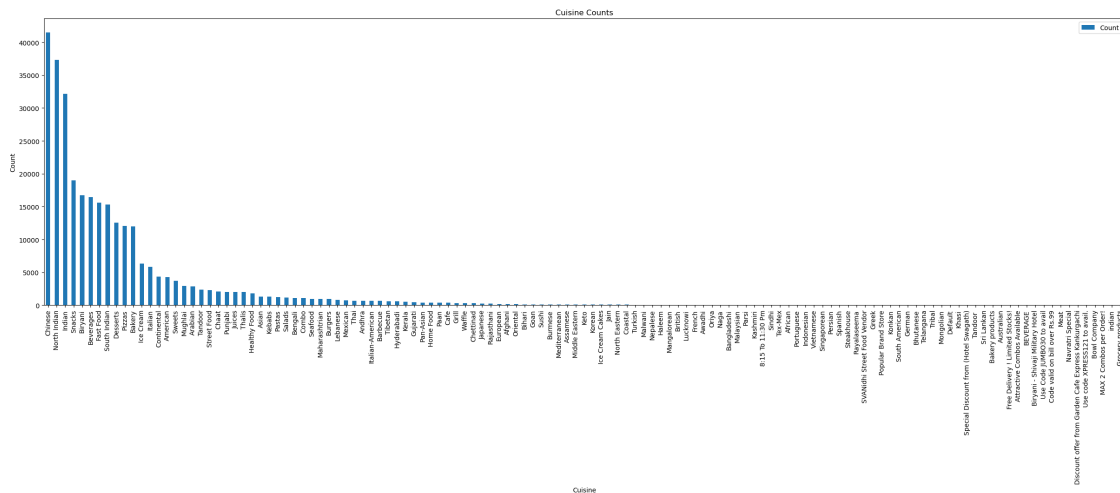

[29]:
```python
not_cuisines = [
    '8:15 To 11:30 Pm',
    'Tex-Mex',
    'SVANidhi Street Food Vendor',
    'Popular Brand Store',    'Bhutanese',
    'Default',
    'Special Discount from (Hotel Swagath)',
    'Bakery products',
    'Free Delivery ! Limited Stocks!',
```

```
    'Attractive Combos Available',
    'BEVERAGE',
    'Biryani - Shivaji Military Hotel',
    'Use Code JUMBO30 to avail',
    'Code valid on bill over Rs.99',
    'Meat',
    'Navratri Special',
    'Discount offer from Garden Cafe Express Kankurgachi',
    'Use code XPRESS121 to avail.',
    'Bowl Company',
    'MAX 2 Combos per Order!',
    'Grocery products'
]
```

[30]:
```python
# Set a minimum threshold frequency
min_frequency = 0.01

# Count the frequency of each cuisine
cuisine_counts = df['cuisine'].str.split(',', expand=True).stack().
 ↪value_counts()

# Filter out infrequent cuisines
frequent_cuisines = cuisine_counts[cuisine_counts >= min_frequency * len(df)]
```

[31]: `frequent_cuisines`

[31]:
```
Chinese          41515
North Indian     37312
Indian           32130
Snacks           19019
Biryani          16750
Beverages        16427
Fast Food        15563
South Indian     15317
Desserts         12529
Pizzas           12060
Bakery           11970
Ice Cream         6354
Italian           5802
Continental       4319
American          4279
Sweets            3690
Mughlai           2934
Arabian           2870
Tandoor           2388
Street Food       2270
Chaat             2066
```

```
Punjabi           2004
Juices            1997
Thalis            1971
Healthy Food      1769
dtype: int64
```

[32]:
```python
df = df[df['cuisine'].str.split(',', expand=True).isin(frequent_cuisines.index).
↪any(axis=1)]
```

[33]:
```python
df = df[~df['cuisine'].str.split(',', expand=True).isin(not_cuisines).
↪any(axis=1)]
```

[34]:
```python
cuisine_counts = df['cuisine'].str.split(',', expand=True).stack().
↪value_counts()

# Set up the bar plot
fig, ax = plt.subplots(figsize=(30,8))
cuisine_counts.plot(kind='bar')

# Customize plot details
ax.set_title('Cuisine Counts')
ax.set_xlabel('Cuisine')
ax.set_ylabel('Count')
ax.legend(['Count'])

# Display the plot
plt.savefig(plots_path + 'cusine-counts-2.png', bbox_inches='tight')
plt.show()
```
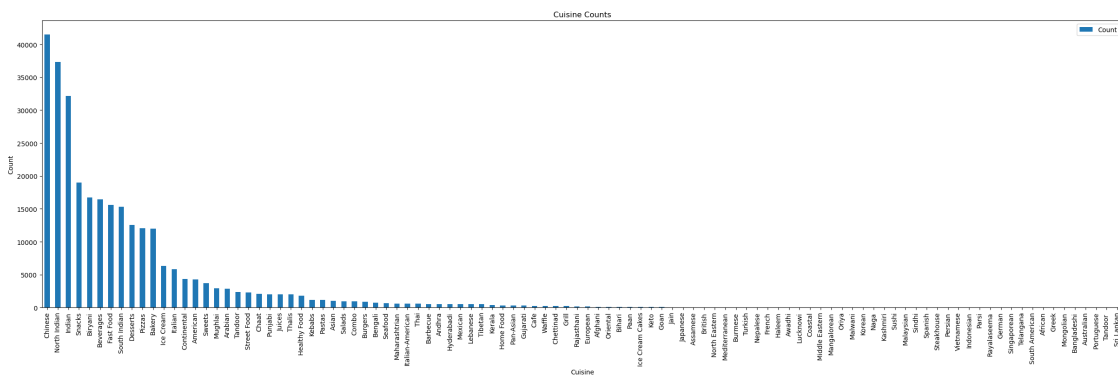

[35]:
```python
df[['cuisine1', 'cuisine2']] = df['cuisine'].str.split(',', expand=True)
df.drop(columns=['cuisine'], inplace=True)
```

[36]:
```python
df.head()
```

```
[36]:         id          name  city rating     rating_count   cost  \
      0  567335    AB FOODS POINT  Abohar     --  Too Few Ratings  200.0
      1  531342  Janta Sweet House  Abohar    4.4      50+ ratings  200.0
      2  158203  theka coffee desi  Abohar    3.8     100+ ratings  100.0
      3  187912         Singh Hut  Abohar    3.7      20+ ratings  250.0
      4  543530    GRILL MASTERS  Abohar     --  Too Few Ratings  250.0

                lic_no                                             link  \
      0  22122652000138  https://www.swiggy.com/restaurants/ab-foods-po…
      1  12117201000112  https://www.swiggy.com/restaurants/janta-sweet…
      2  22121652000190  https://www.swiggy.com/restaurants/theka-coffe…
      3  22119652000167  https://www.swiggy.com/restaurants/singh-hut-n…
      4  12122201000053  https://www.swiggy.com/restaurants/grill-maste…

                                            address           menu  \
      0  AB FOODS POINT, NEAR RISHI NARANG DENTAL CLINI…  Menu/567335.json
      1  Janta Sweet House, Bazar No.9, Circullar Road,…  Menu/531342.json
      2      theka coffee desi, sahtiya sadan road city  Menu/158203.json
      3    Singh Hut, CIRCULAR ROAD NEAR NEHRU PARK ABOHAR  Menu/187912.json
      4  GRILL MASTERS, ADA Heights, Abohar - Hanumanga…  Menu/543530.json

        sub_area    area          cuisine1   cuisine2
      0  Abohar  Abohar          Beverages     Pizzas
      1  Abohar  Abohar             Sweets     Bakery
      2  Abohar  Abohar          Beverages  Beverages
      3  Abohar  Abohar          Fast Food     Indian
      4  Abohar  Abohar  Italian-American  Fast Food
```

```python
[37]: df['rating'] = df['rating'].replace('--', '0')
```

```python
[38]: df['rating'] = df['rating'].astype('float')
```

```python
[39]: # Set up the histogram plot
      fig, ax = plt.subplots(figsize=(12, 8))
      n, bins, patches = ax.hist(df['rating'], bins=10, color='green')

      # Customize plot details
      ax.set_title('Ratings Distribution')
      ax.set_xlabel('Ratings')
      ax.set_ylabel('Frequency')
      ax.set_xticks(bins)
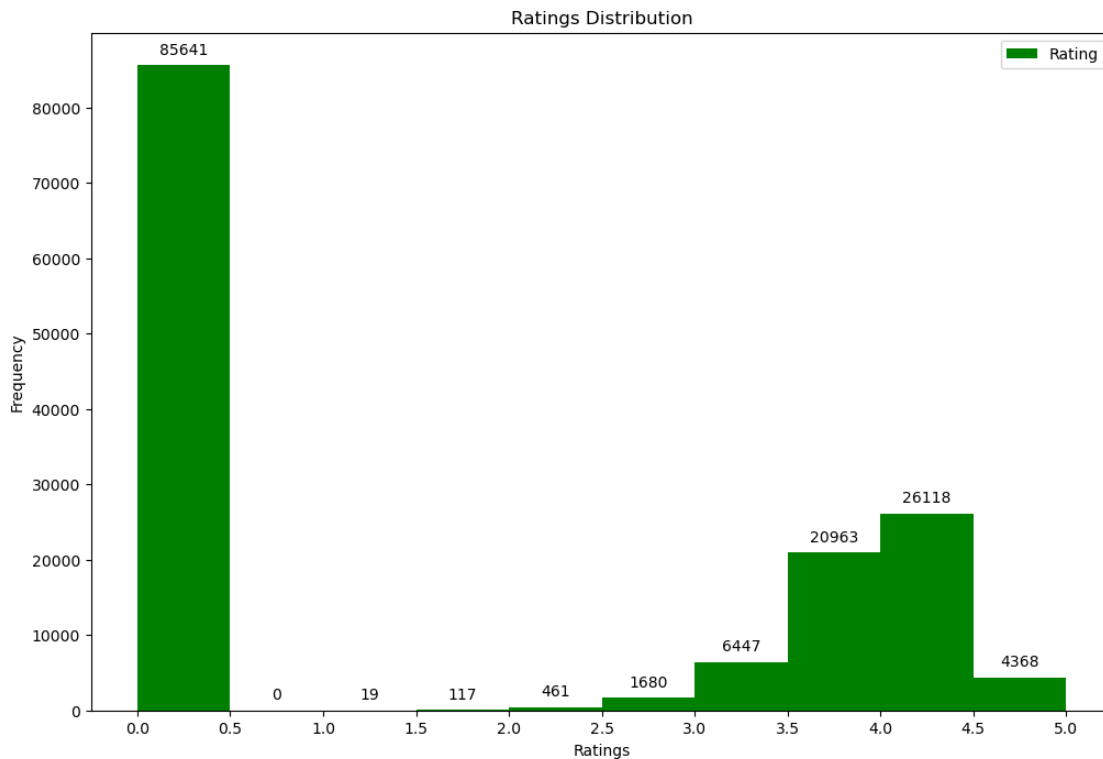      ax.legend(['Rating'])

      # Add value labels to the top of each bar
      for i in range(len(patches)):
          x = patches[i].get_x() + patches[i].get_width() / 2
          y = patches[i].get_height()
```

```
    ax.annotate('{:.0f}'.format(y), (x, y), xytext=(0, 5),
                textcoords='offset points', ha='center', va='bottom')

# Display the plot
plt.savefig(plots_path + 'ratings-distribution.png', bbox_inches='tight')
plt.show()
```



```
[40]:  # Set up the histogram plot
       fig, ax = plt.subplots(figsize=(12, 8))
       n, bins, patches = ax.hist(df['rating'][df['rating'] != 0], bins=10,␣
        ↪color='green')

       # Customize plot details
       ax.set_title('Ratings Distribution')
       ax.set_xlabel('Ratings')
       ax.set_ylabel('Frequency')
       ax.set_xticks(bins)
       ax.legend(['Rating'])

       # Add value labels to the top of each bar
       for i in range(len(patches)):
           x = patches[i].get_x() + patches[i].get_width() / 2
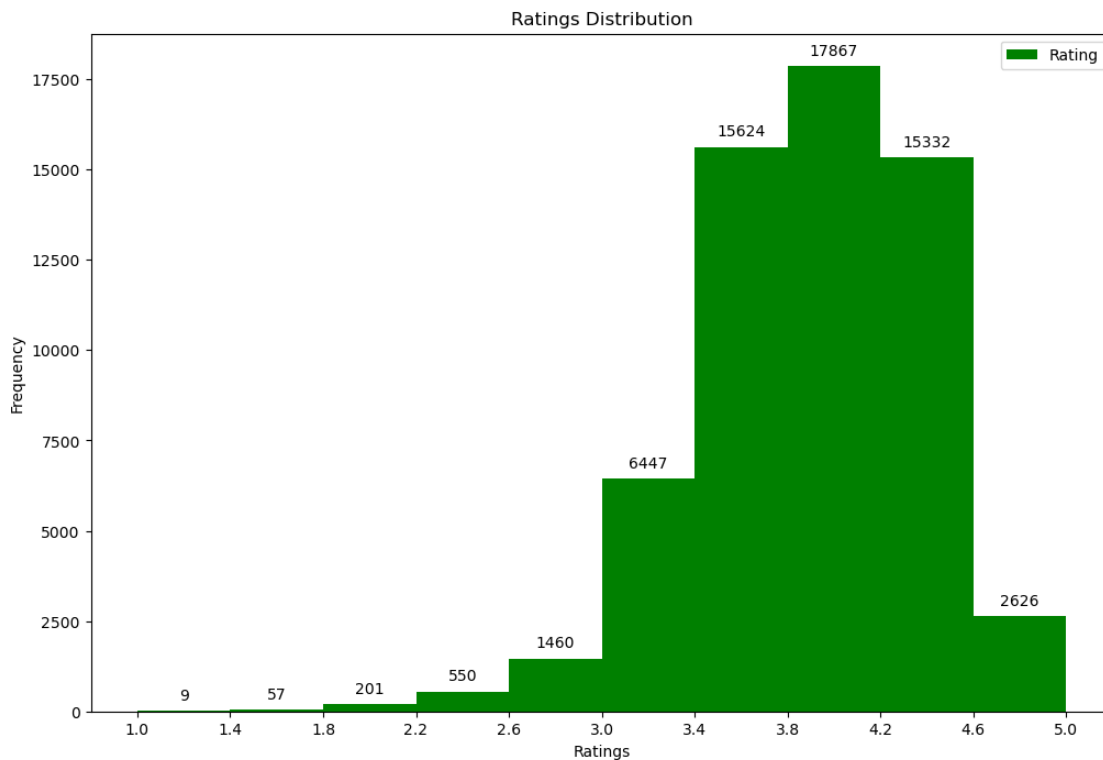```

```
        y = patches[i].get_height()
        ax.annotate('{:.0f}'.format(y), (x, y), xytext=(0, 5),
                    textcoords='offset points', ha='center', va='bottom')

# Display the plot
plt.savefig(plots_path + 'ratings-distribution-2.png', bbox_inches='tight')
plt.show()
```


[41]: `df.head()`

[41]:
```
        id              name     city  rating      rating_count   cost  \
0   567335     AB FOODS POINT   Abohar     0.0   Too Few Ratings   200.0
1   531342   Janta Sweet House  Abohar     4.4       50+ ratings   200.0
2   158203   theka coffee desi  Abohar     3.8      100+ ratings   100.0
3   187912          Singh Hut   Abohar     3.7       20+ ratings   250.0
4   543530      GRILL MASTERS   Abohar     0.0   Too Few Ratings   250.0

         lic_no                                              link  \
0   22122652000138   https://www.swiggy.com/restaurants/ab-foods-po…
1   12117201000112   https://www.swiggy.com/restaurants/janta-sweet…
2   22121652000190   https://www.swiggy.com/restaurants/theka-coffe…
3   22119652000167   https://www.swiggy.com/restaurants/singh-hut-n…
4   12122201000053   https://www.swiggy.com/restaurants/grill-maste…
```

17

```
                                       address               menu  \
0  AB FOODS POINT, NEAR RISHI NARANG DENTAL CLINI…  Menu/567335.json
1  Janta Sweet House, Bazar No.9, Circullar Road,…  Menu/531342.json
2          theka coffee desi, sahtiya sadan road city  Menu/158203.json
3    Singh Hut, CIRCULAR ROAD NEAR NEHRU PARK ABOHAR  Menu/187912.json
4  GRILL MASTERS, ADA Heights, Abohar - Hanumanga…  Menu/543530.json

  sub_area     area           cuisine1   cuisine2
0   Abohar   Abohar          Beverages     Pizzas
1   Abohar   Abohar             Sweets     Bakery
2   Abohar   Abohar          Beverages  Beverages
3   Abohar   Abohar          Fast Food     Indian
4   Abohar   Abohar  Italian-American  Fast Food
```

[42]:
```python
df.to_csv('swiggy-preprocessed.csv', index=False)
```