

major-project

May 13, 2024

```
[ ]: # Name: Pravir Mishra  
# Major Project March 2024 (1)
```

```
[1]: import pandas as pd  
  
bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')  
  
print("Missing values in Bangalore Restaurants dataset:")  
print(bangalore_restaurants.isnull().sum())  
  
bangalore_restaurants.columns = bangalore_restaurants.columns.str.lower().str.  
    ↪replace(' ', '_')
```

Missing values in Bangalore Restaurants dataset:

Restaurant_Name	0
Category	0
Pricing_for_2	0
Locality	0
Dining_Rating	8
Dining_Review_Count	8
Delivery_Rating	1412
Delivery_Rating_Count	8
Website	0
Address	0
Phone_No	0
Latitude	0
Longitude	0
dtype: int64	

```
[2]: import pandas as pd  
  
pune_restaurants = pd.read_csv('Pune_Restaurants.csv')  
  
print("Missing values in Pune Restaurants dataset:")  
print(pune_restaurants.isnull().sum())  
  
pune_restaurants.columns = pune_restaurants.columns.str.lower().str.replace(' ',  
    ↪ '_')
```

Missing values in Pune Restaurants dataset:

Restaurant_Name	0
Category	0
Pricing_for_2	0
Locality	0
Dining_Rating	0
Dining_Review_Count	0
Delivery_Rating	1571
Delivery_Rating_Count	0
Website	0
Address	0
Phone_No	0
Latitude	0
Longitude	0
Known_for1	642
Known_for2	3719

dtype: int64

```
[3]: import pandas as pd
import matplotlib.pyplot as plt

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')

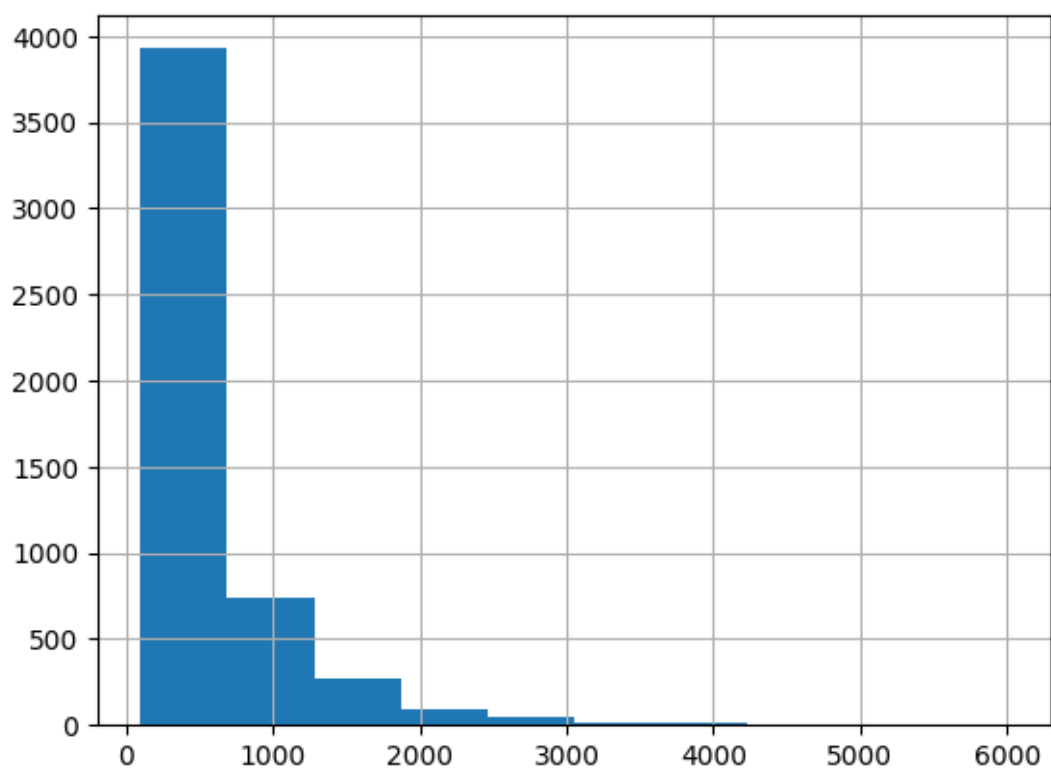
bangalore_restaurants['Category'].value_counts().plot(kind='bar')
plt.show()

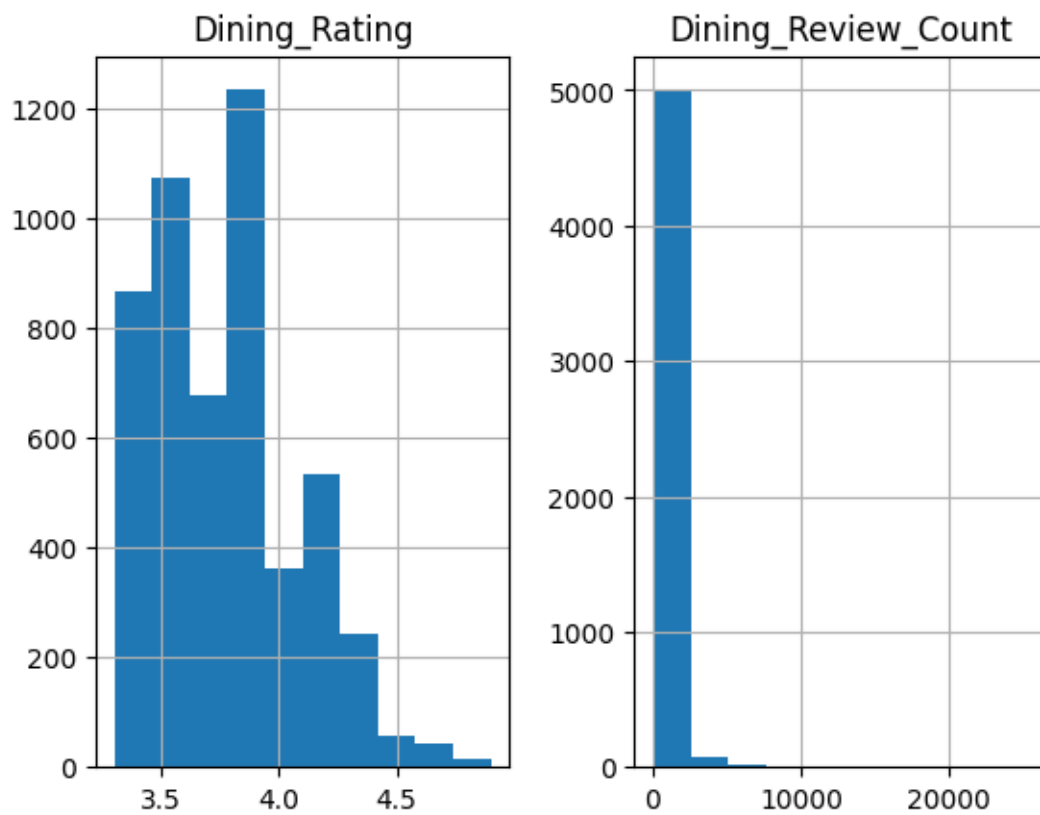
bangalore_restaurants['Pricing_for_2'].hist()
plt.show()

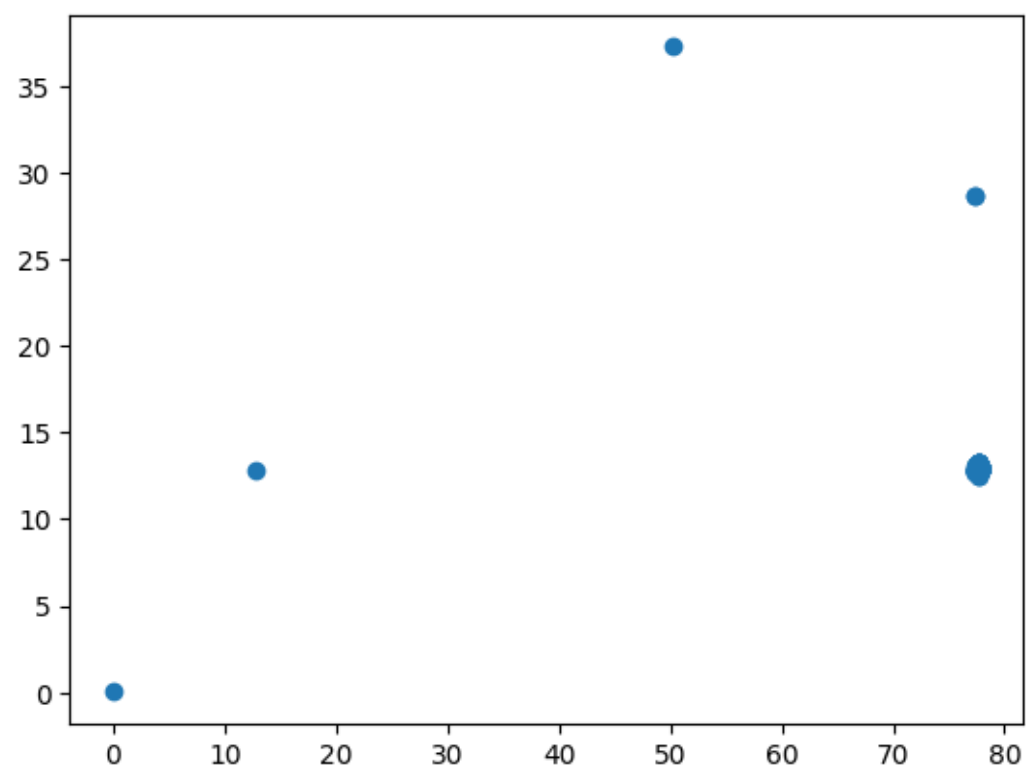
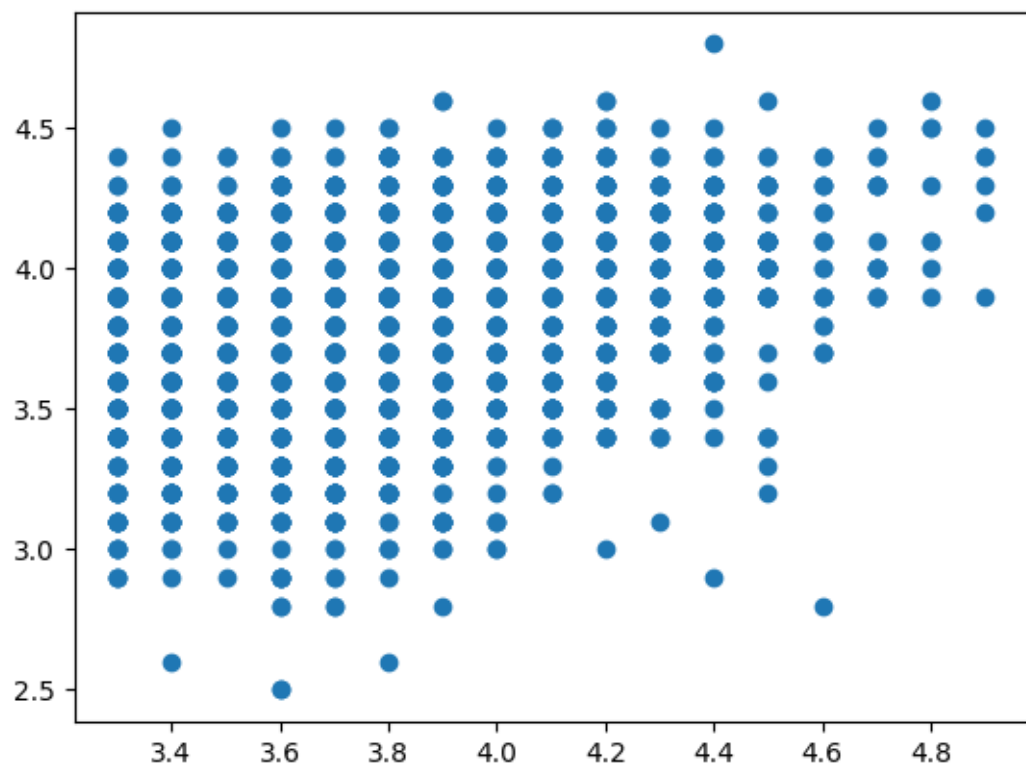
bangalore_restaurants[['Dining_Rating', 'Dining_Review_Count']].hist()
plt.show()

plt.scatter(bangalore_restaurants['Dining_Rating'],
            ↪ bangalore_restaurants['Delivery_Rating'])
plt.show()

plt.scatter(bangalore_restaurants['Longitude'],
            ↪ bangalore_restaurants['Latitude'])
plt.show()
```





```
[4]: import pandas as pd
import matplotlib.pyplot as plt

pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

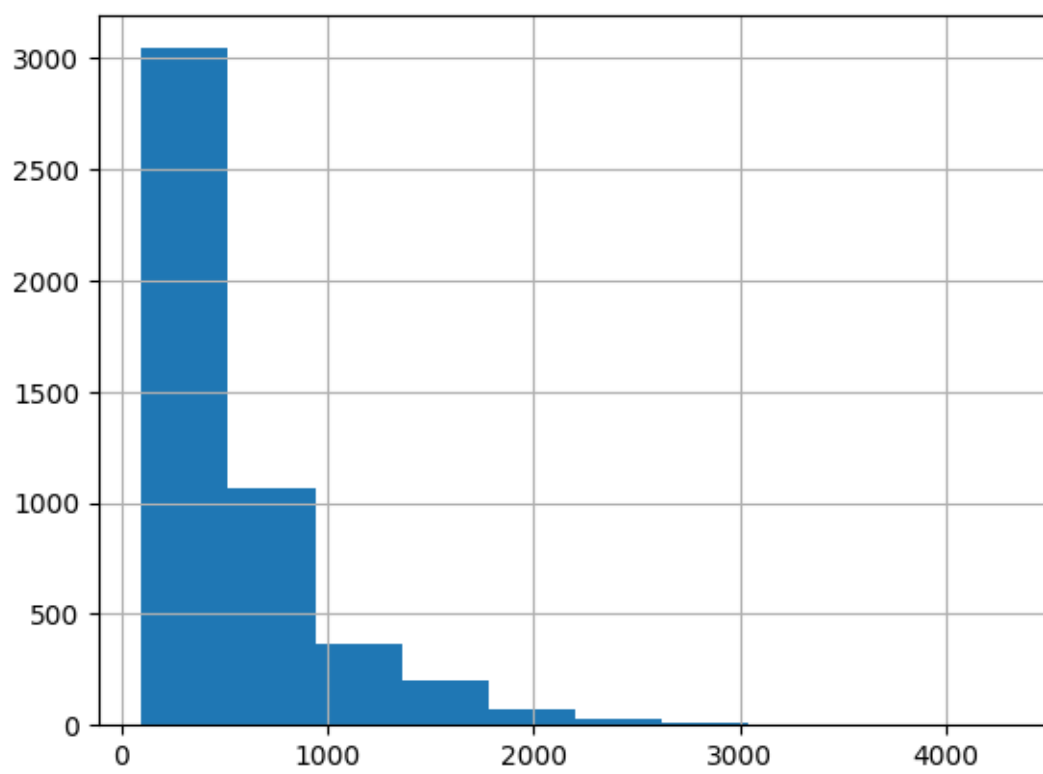
pune_restaurants['Category'].value_counts().plot(kind='bar')
plt.show()

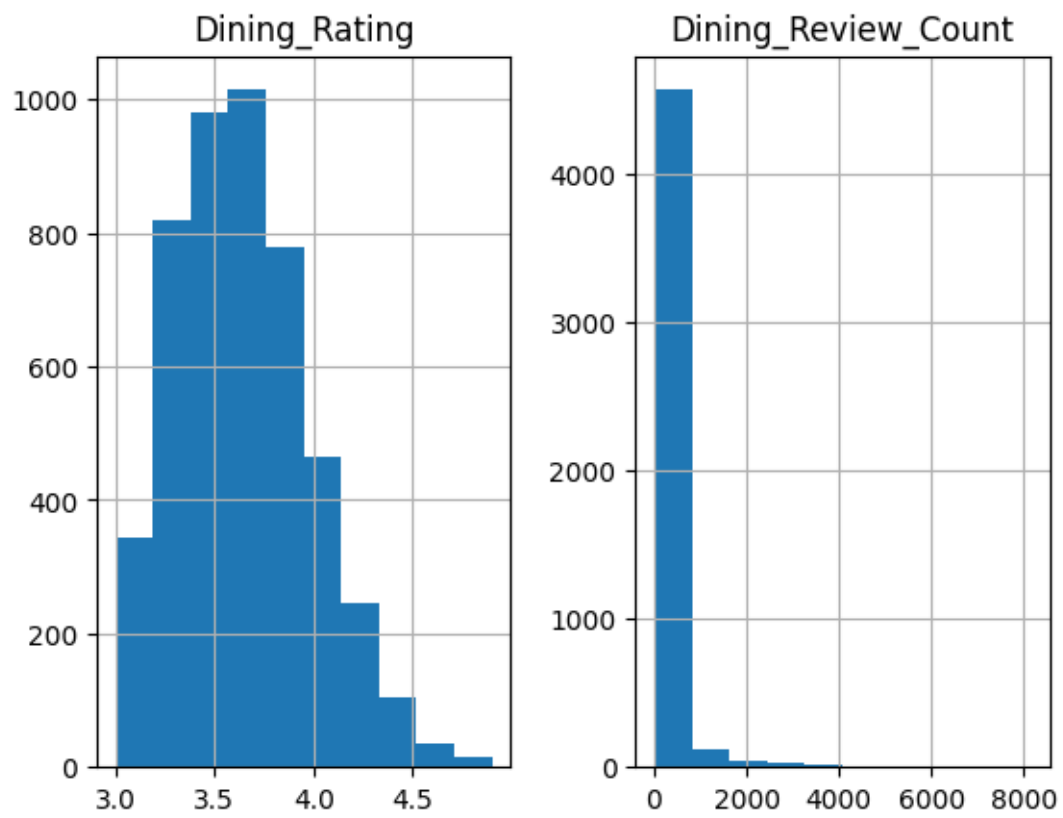
pune_restaurants['Pricing_for_2'].hist()
plt.show()

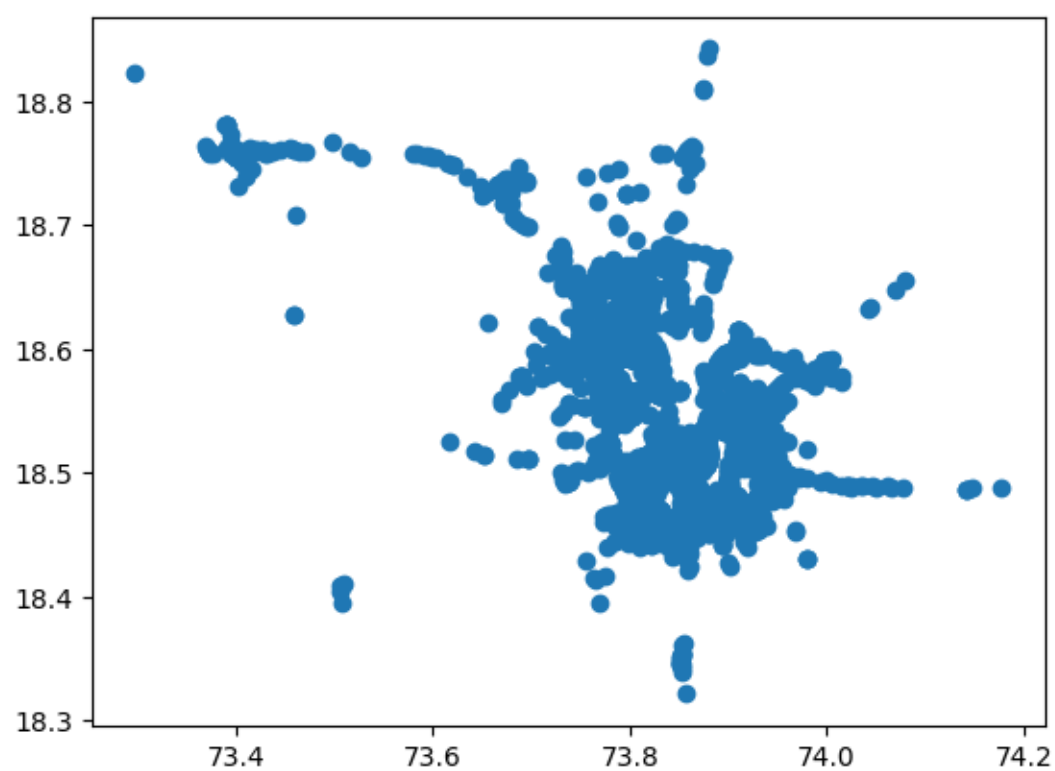
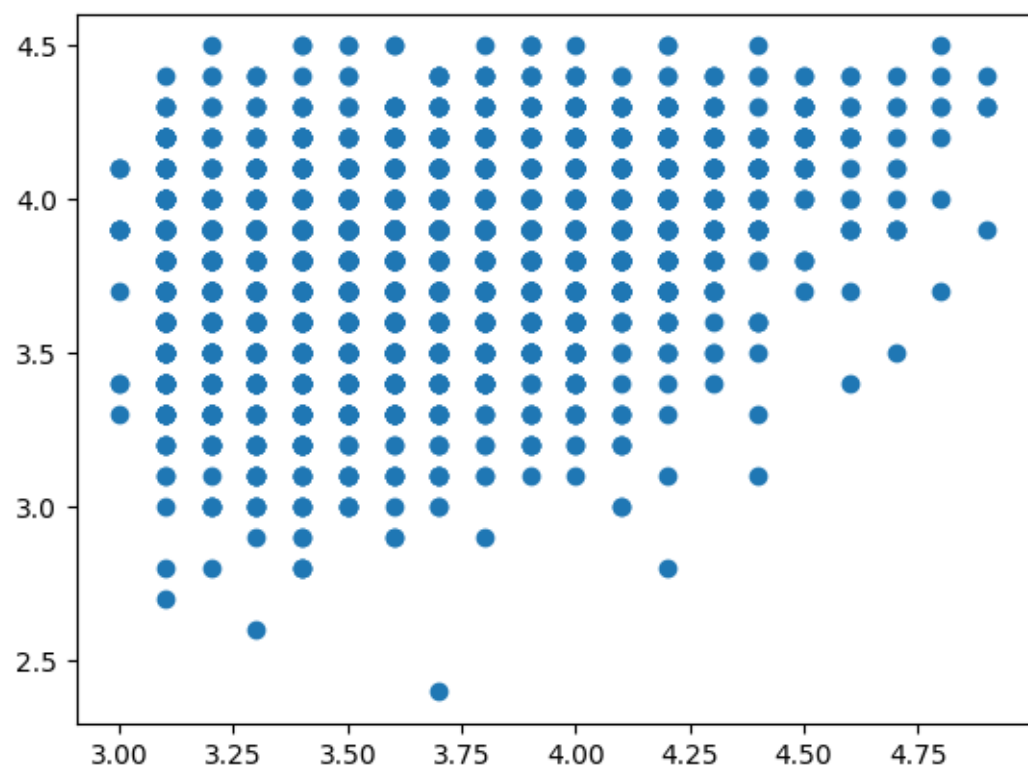
pune_restaurants[['Dining_Rating', 'Dining_Review_Count']].hist()
plt.show()

plt.scatter(pune_restaurants['Dining_Rating'],
            ↪pune_restaurants['Delivery_Rating'])
plt.show()

plt.scatter(pune_restaurants['Longitude'], pune_restaurants['Latitude'])
plt.show()
```





```
[5]: import pandas as pd

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')

top_rated_bangalore = bangalore_restaurants.sort_values(by=['Dining_Rating',
↳ 'Dining_Review_Count'], ascending=False).head(10)
print("Top rated restaurants in Bangalore:")
print(top_rated_bangalore['Restaurant_Name'])

popular_cuisines_bangalore = top_rated_bangalore['Category'].value_counts().
↳ head(10)
print("Most popular cuisines in Bangalore among top-rated restaurants:")
print(popular_cuisines_bangalore)
```

Top rated restaurants in Bangalore:

```
1    Windmills Craftworks
2          CTR Shri Sagar
3    Brahmin's Coffee Bar
0          Burma Burma
4          Milano Ice Cream
5          BelgYum
6          Lot Like Crepes
7          Truffles
10         Chianti
13         Brik Oven
```

Name: Restaurant_Name, dtype: object

Most popular cuisines in Bangalore among top-rated restaurants:

```
Category
South Indian                                     2
Continental, Fast Food, Kebab, Beverages, Italian, Desserts 1
Asian, Burmese, Bubble Tea, Salad, Tea, Desserts, Ice Cream, Beverages 1
Desserts, Ice Cream, Beverages                    1
Desserts                                           1
Desserts, Fast Food, Beverages                    1
Cafe, American, Coffee, Steak, Beverages, Fast Food 1
Italian, Pasta, Pizza                             1
Cafe, Italian, Fast Food, Desserts                 1
Name: count, dtype: int64
```

```
[6]: import pandas as pd

pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

top_rated_pune = pune_restaurants.sort_values(by=['Dining_Rating',
↳ 'Dining_Review_Count'], ascending=False).head(10)
```

```

print("Top rated restaurants in Pune:")
print(topRatedPune['Restaurant_Name'])

popularCuisinesPune = topRatedPune['Category'].value_counts().head(10)
print("Most popular cuisines in Pune among top-rated restaurants:")
print(popularCuisinesPune)

```

Top rated restaurants in Pune:

```

1          Le Plaisir
2          Gong
3  The French Window Patisserie
0          Santè Spa Cuisine
4          Savya Rasa
6  Pagdandi Books Chai Cafe
5          Le Flamington
11         Vohuman Cafe
7  Paasha - JW Marriott Pune
12         Andaground

```

Name: Restaurant_Name, dtype: object

Most popular cuisines in Pune among top-rated restaurants:

```

Category
Cafe, Italian, Continental, Salad, Sandwich, Pizza, Beverages    1
Chinese, Sushi, Asian, Momos, Beverages                          1
Cafe, Desserts, French, Bakery, European                        1
Continental, Healthy Food, Mediterranean                        1
South Indian, Mangalorean, Kerala, Chettinad, Beverages        1
Cafe, Fast Food, Beverages                                       1
Cafe, European, Fast Food, Bakery, Desserts                     1
Street Food, Beverages                                           1
North Indian, Kebab, Desserts                                    1
Fast Food, Street Food, North Indian, Beverages                 1
Name: count, dtype: int64

```

```

[2]: import pandas as pd
import matplotlib.pyplot as plt

bangaloreRestaurants = pd.read_csv('Bangalore_Restaurants.csv')

bangaloreRestaurants['Locality'].value_counts().plot(kind='bar')
plt.show()

topLocalitiesBangalore = bangaloreRestaurants['Locality'].value_counts().
    .head(10)
print("Areas in Bangalore with the highest concentration of restaurants:")
print(topLocalitiesBangalore)

```

```
bangalore_restaurants.groupby('Locality')[['Dining_Rating', 'Pricing_for_2']].  
    ↪mean().plot(kind='bar')  
plt.show()
```


Areas in Bangalore with the highest concentration of restaurants:

Locality

BTM, Bangalore	301
Indiranagar, Bangalore	200
HSR, Bangalore	197
Whitefield, Bangalore	195
Electronic City, Bangalore	178
JP Nagar, Bangalore	157
Jayanagar, Bangalore	144
Marathahalli, Bangalore	142
Yelahanka, Bangalore	125
Sarjapur Road, Bangalore	124

Name: count, dtype: int64

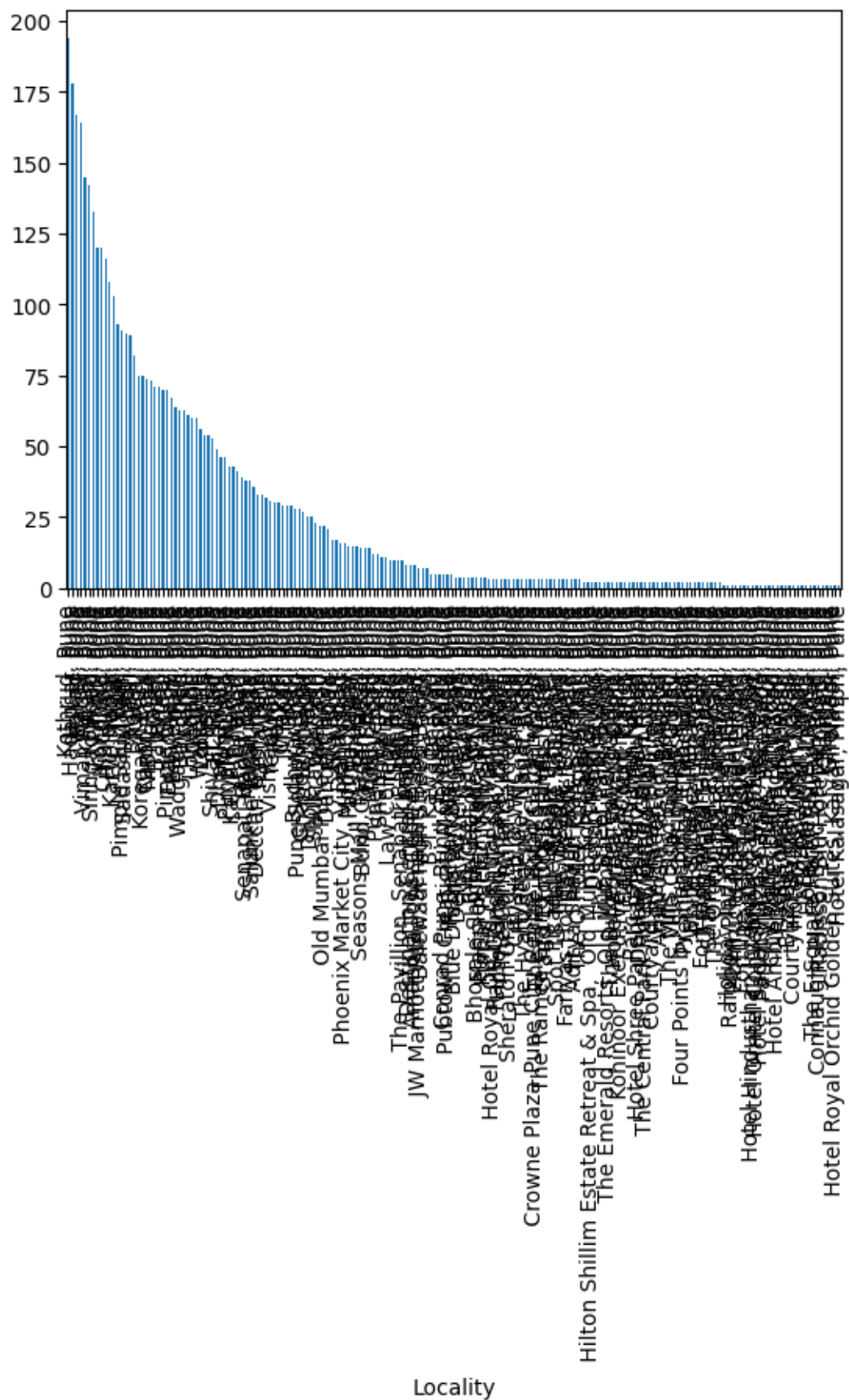

```
[3]: import pandas as pd
import matplotlib.pyplot as plt

pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

pune_restaurants['Locality'].value_counts().plot(kind='bar')
plt.show()

top_localities_pune = pune_restaurants['Locality'].value_counts().head(10)
print("Areas in Pune with the highest concentration of restaurants:")
print(top_localities_pune)

pune_restaurants.groupby('Locality')[['Dining_Rating', 'Pricing_for_2']].mean().
    .plot(kind='bar')
plt.show()
```

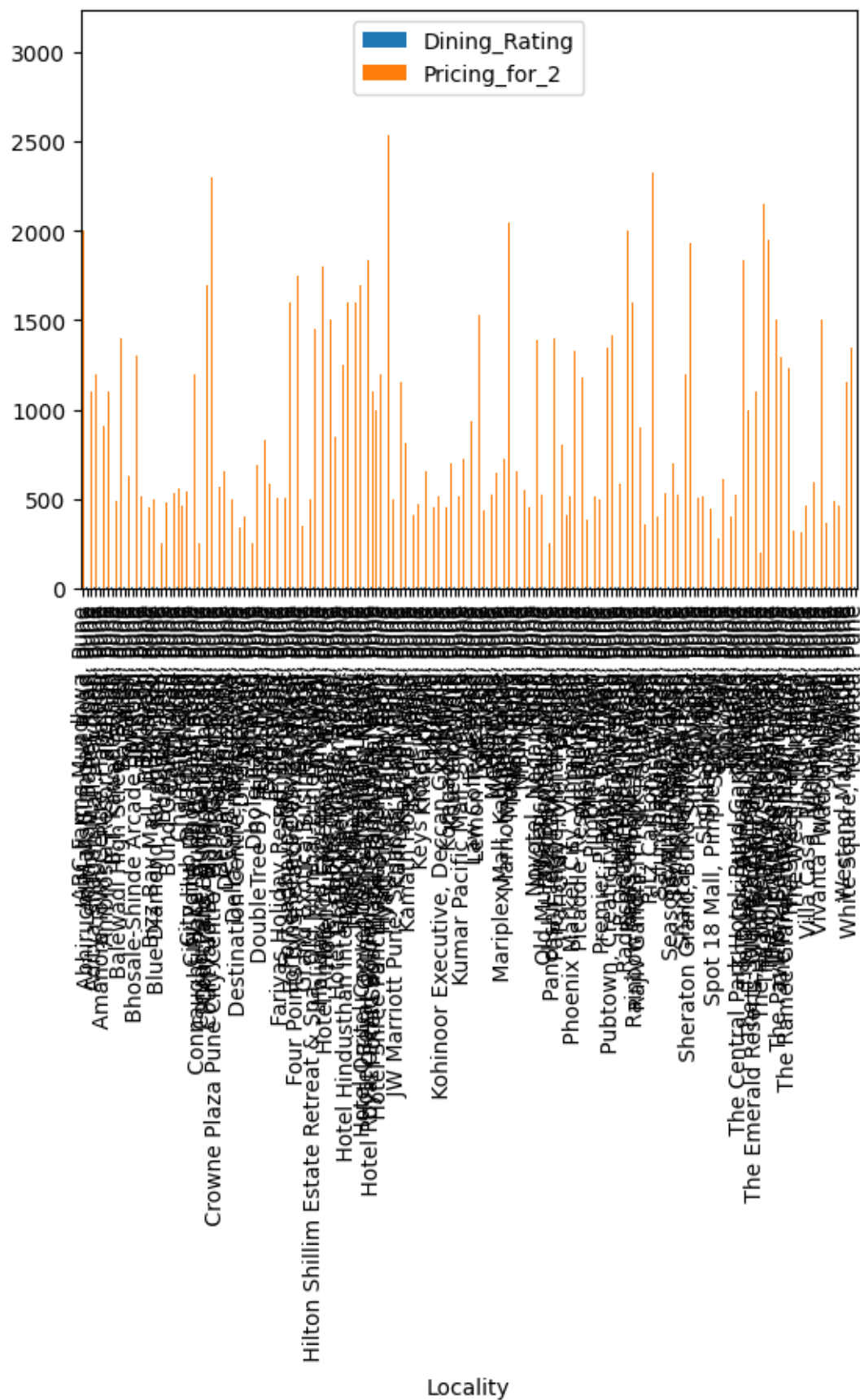


Areas in Pune with the highest concentration of restaurants:

Locality

Kothrud, Pune	194
Hadapsar, Pune	178
Wakad, Pune	167
Baner, Pune	164
Viman Nagar, Pune	145
Kharadi, Pune	142
Sinhgad Road, Pune	133
Pimpri, Pune	120
Kondhwa, Pune	120
Chinchwad, Pune	116

Name: count, dtype: int64



```
[4]: import pandas as pd

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')

popular_cuisines_bangalore = bangalore_restaurants['Category'].value_counts()
print("Most popular cuisines in Bangalore:")
print(popular_cuisines_bangalore)

average_pricing_bangalore = bangalore_restaurants.
    ↳groupby('Category')['Pricing_for_2'].mean()
print("Average pricing for two people for each category in Bangalore:")
print(average_pricing_bangalore)

topRatedBangalore = bangalore_restaurants.sort_values('Dining_Rating',
    ↳ascending=False).groupby('Category').first()
print("Top-rated restaurants for each cuisine category in Bangalore:")
print(topRatedBangalore['Restaurant_Name'])
```

Most popular cuisines in Bangalore:

Category

South Indian

220

North Indian, Chinese

191

North Indian

108

Biryani, South Indian

83

South Indian, North Indian, Chinese

61

...

Kerala, Chinese, South Indian, Beverages

1

North Indian, Chinese, South Indian, Fast Food, Street Food, Mithai

1

Kerala, Biryani, North Indian, South Indian, Seafood, Chinese

1

North Indian, South Indian, Chinese, Street Food, Beverages, Biryani, Desserts,

Ice Cream 1

North Indian, Chinese, Andhra, Fast Food

1

Name: count, Length: 2545, dtype: int64

Average pricing for two people for each category in Bangalore:

Category

African, Burger, Fast Food, Beverages, Desserts

```

1000.0
American, Burger, Fast Food, Pizza, Italian
700.0
American, Continental, Bar Food, Chinese, Fast Food, Pizza, Burger, Beverages
1300.0
American, Continental, Cafe, Italian, Sandwich
500.0
American, Continental, Fast Food, Beverages
2400.0
...
Vietnamese, Asian
1700.0
Vietnamese, Desserts, Thai
1100.0
Vietnamese, Seafood, Beverages
3200.0
Vietnamese, Thai, Burmese, Japanese
1200.0
Wraps, Fast Food, Beverages
300.0
Name: Pricing_for_2, Length: 2545, dtype: float64
Top-rated restaurants for each cuisine category in Bangalore:
Category
African, Burger, Fast Food, Beverages, Desserts
Galito's
American, Burger, Fast Food, Pizza, Italian
Wolf'ish
American, Continental, Bar Food, Chinese, Fast Food, Pizza, Burger, Beverages
Plan B
American, Continental, Cafe, Italian, Sandwich
Cafe OTW - On The Way
American, Continental, Fast Food, Beverages
Arbor Brewing Company
...
Vietnamese, Asian
Hanoi - Vietnamese Cuisine
Vietnamese, Desserts, Thai
Phobidden Fruit
Vietnamese, Seafood, Beverages
Blue Ginger - The Taj West End
Vietnamese, Thai, Burmese, Japanese
The Asian Curry House
Wraps, Fast Food, Beverages
Hodge Podge
Name: Restaurant_Name, Length: 2545, dtype: object

```

```
[5]: import pandas as pd

pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

popular_cuisines_pune = pune_restaurants['Category'].value_counts()
print("Most popular cuisines in Pune:")
print(popular_cuisines_pune)

average_pricing_pune = pune_restaurants.groupby('Category')['Pricing_for_2'].
    .mean()
print("Average pricing for two people for each category in Pune:")
print(average_pricing_pune)

top_rated_pune = pune_restaurants.sort_values('Dining_Rating', ascending=False).
    .groupby('Category').first()
print("Top-rated restaurants for each cuisine category in Pune:")
print(top_rated_pune['Restaurant_Name'])
```

Most popular cuisines in Pune:

Category	
Chinese	130
North Indian, Chinese	126
Maharashtrian	122
North Indian	110
Street Food	92

...

Cafe, Italian, Fast Food, Chinese, Beverages	1
South Indian, North Indian, Chinese, Fast Food, Beverages	1
Cafe, Sandwich, Fast Food, Beverages, Desserts	1
Maharashtrian, North Indian, Goan	1
North Indian, Continental, Chinese, Desserts, Beverages	1

Name: count, Length: 2322, dtype: int64

Average pricing for two people for each category in Pune:

Category	
Afghan, Arabian	800.0
Afghan, Chinese, Biryani, Beverages, Kebab	300.0
Afghan, Lebanese	650.0
Afghan, Mughlai	500.0
Afghan, North Indian, Kebab	500.0

...

Thai, Japanese, Asian	1500.0
Thai, Malaysian, Japanese, Burmese, Indonesian, Chinese	1200.0
Thai, Seafood	1000.0
Tibetan, Vietnamese	600.0
Wraps, Chinese, Biryani, Rolls, Fast Food, Sandwich	300.0

Name: Pricing_for_2, Length: 2322, dtype: float64

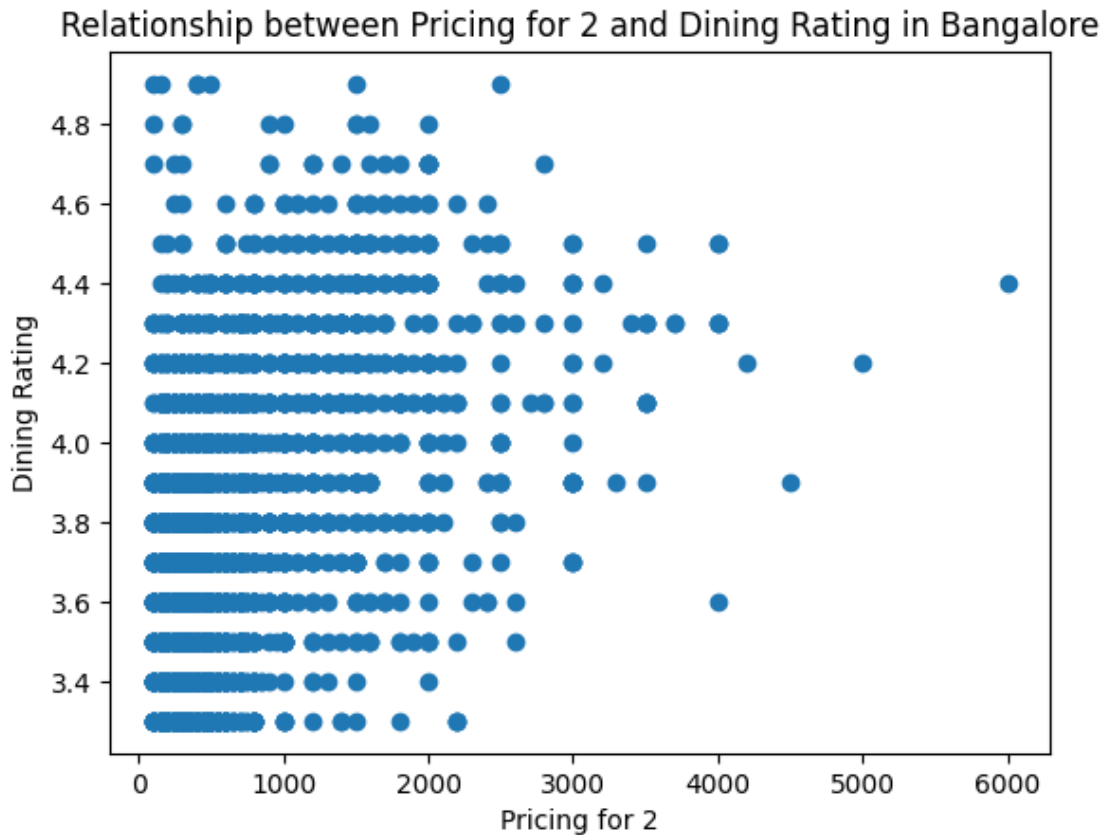
Top-rated restaurants for each cuisine category in Pune:

Category	
Afghan, Arabian	New Afghan Zaika
Restaurant	
Afghan, Chinese, Biryani, Beverages, Kebab	Hotel Afghan
Shamal	
Afghan, Lebanese	Kabul
Restaurant	
Afghan, Mughlai	Balkh Afghan
Kitchen	
Afghan, North Indian, Kebab	Afghan
Darbar	
	...
Thai, Japanese, Asian	Malaka
Spice	
Thai, Malaysian, Japanese, Burmese, Indonesian, Chinese	
Yin Yang	
Thai, Seafood	Thai
House	
Tibetan, Vietnamese	Yeti And
The Monk	
Wraps, Chinese, Biryani, Rolls, Fast Food, Sandwich	Qwik
O'Byts	
Name: Restaurant_Name, Length: 2322, dtype: object	

```
[6]: import pandas as pd
import matplotlib.pyplot as plt

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')

plt.scatter(bangalore_restaurants['Pricing_for_2'],
            bangalore_restaurants['Dining_Rating'])
plt.xlabel('Pricing for 2')
plt.ylabel('Dining Rating')
plt.title('Relationship between Pricing for 2 and Dining Rating in Bangalore')
plt.show()
```



```
[7]: import pandas as pd
import matplotlib.pyplot as plt

pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

plt.scatter(pune_restaurants['Pricing_for_2'],
            pune_restaurants['Dining_Rating'])
plt.xlabel('Pricing for 2')
plt.ylabel('Dining Rating')
plt.title('Relationship between Pricing for 2 and Dining Rating in Pune')
plt.show()
```



```
[10]: import pandas as pd
from scipy import stats
import statsmodels.api as sm

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')

asian_ratings = bangalore_restaurants[bangalore_restaurants['Category'] == 'Asian']['Dining_Rating']
italian_ratings = bangalore_restaurants[bangalore_restaurants['Category'] == 'Italian']['Dining_Rating']

t_stat, p_val = stats.ttest_ind(asian_ratings, italian_ratings)
print(f'T-statistic: {t_stat}, P-value: {p_val}')

X = bangalore_restaurants['Pricing_for_2']
y = bangalore_restaurants['Dining_Rating']

X = sm.add_constant(X)

model = sm.OLS(y, X).fit()
```

```

predictions = model.predict(X)

print_model = model.summary()
print(print_model)

```

T-statistic: -0.632455532033676, P-value: 0.5720033807006406

OLS Regression Results

```

=====
Dep. Variable:          Dining_Rating      R-squared:                  nan
Model:                  OLS                Adj. R-squared:          nan
Method:                 Least Squares       F-statistic:              nan
Date:                  Sun, 12 May 2024     Prob (F-statistic):       nan
Time:                  05:37:24            Log-Likelihood:           nan
No. Observations:      5109               AIC:                     nan
Df Residuals:          5107               BIC:                     nan
Df Model:               1
Covariance Type:       nonrobust
=====

```

```

=====
=
              coef      std err          t      P>|t|      [0.025
0.975]
-----
-
const              nan         nan         nan         nan         nan
nan
Pricing_for_2      nan         nan         nan         nan         nan
nan
=====
Omnibus:            nan    Durbin-Watson:           nan
Prob(Omnibus):      nan    Jarque-Bera (JB):           nan
Skew:               nan    Prob(JB):              nan
Kurtosis:           nan    Cond. No.                1.14e+03
=====

```

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.14e+03. This might indicate that there are strong multicollinearity or other numerical problems.

```

[11]: import pandas as pd
      from scipy import stats
      import statsmodels.api as sm

      pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

```

```

asian_ratings = pune_restaurants[pune_restaurants['Category'] == 'Asian']['Dining_Rating']
italian_ratings = pune_restaurants[pune_restaurants['Category'] == 'Italian']['Dining_Rating']

t_stat, p_val = stats.ttest_ind(asian_ratings, italian_ratings)
print(f'T-statistic: {t_stat}, P-value: {p_val}')

X = pune_restaurants['Pricing_for_2']
y = pune_restaurants['Dining_Rating']

X = sm.add_constant(X)

model = sm.OLS(y, X).fit()
predictions = model.predict(X)

print_model = model.summary()
print(print_model)

```

T-statistic: 0.7581888283278772, P-value: 0.47705210626249084

OLS Regression Results

```

=====
Dep. Variable:          Dining_Rating      R-squared:                0.128
Model:                  OLS               Adj. R-squared:           0.128
Method:                 Least Squares      F-statistic:              706.8
Date:                  Sun, 12 May 2024    Prob (F-statistic):       2.14e-145
Time:                  05:37:37           Log-Likelihood:           -1453.1
No. Observations:      4797              AIC:                     2910.
Df Residuals:          4795              BIC:                     2923.
Df Model:               1
Covariance Type:       nonrobust
=====

```

```

=====
              coef      std err          t      P>|t|      [0.025
0.975]
-----
-
const          3.4606      0.008    436.523      0.000      3.445
3.476
Pricing_for_2   0.0003    1.11e-05    26.586      0.000      0.000
0.000
=====

```

```

=====
Omnibus:            86.989    Durbin-Watson:           0.187
Prob(Omnibus):      0.000    Jarque-Bera (JB):        91.489
Skew:               0.337    Prob(JB):                1.36e-20
Kurtosis:           2.946    Cond. No.:               1.19e+03
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.19e+03. This might indicate that there are strong multicollinearity or other numerical problems.

```
[19]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

bangalore_restaurants = pd.read_csv('Bangalore_Restaurants.csv')
pune_restaurants = pd.read_csv('Pune_Restaurants.csv')

plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)
plt.scatter(bangalore_restaurants['Longitude'],
            ↪ bangalore_restaurants['Latitude'])
plt.title('Distribution of Restaurants in Bangalore')
plt.xlabel('Longitude')
plt.ylabel('Latitude')

plt.subplot(1, 2, 2)
plt.scatter(pune_restaurants['Longitude'], pune_restaurants['Latitude'])
plt.title('Distribution of Restaurants in Pune')
plt.xlabel('Longitude')
plt.ylabel('Latitude')

plt.tight_layout()
plt.show()

kmeans_bangalore = KMeans(n_clusters=5, random_state=42)
bangalore_restaurants['cluster'] = kmeans_bangalore.
    ↪ fit_predict(bangalore_restaurants[['Longitude', 'Latitude']])

kmeans_pune = KMeans(n_clusters=5, random_state=42)
pune_restaurants['cluster'] = kmeans_pune.
    ↪ fit_predict(pune_restaurants[['Longitude', 'Latitude']])

plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)
plt.scatter(bangalore_restaurants['Longitude'],
            ↪ bangalore_restaurants['Latitude'], c=bangalore_restaurants['cluster'])
plt.title('Restaurant Clusters in Bangalore')
plt.xlabel('Longitude')
```

```

plt.ylabel('Latitude')

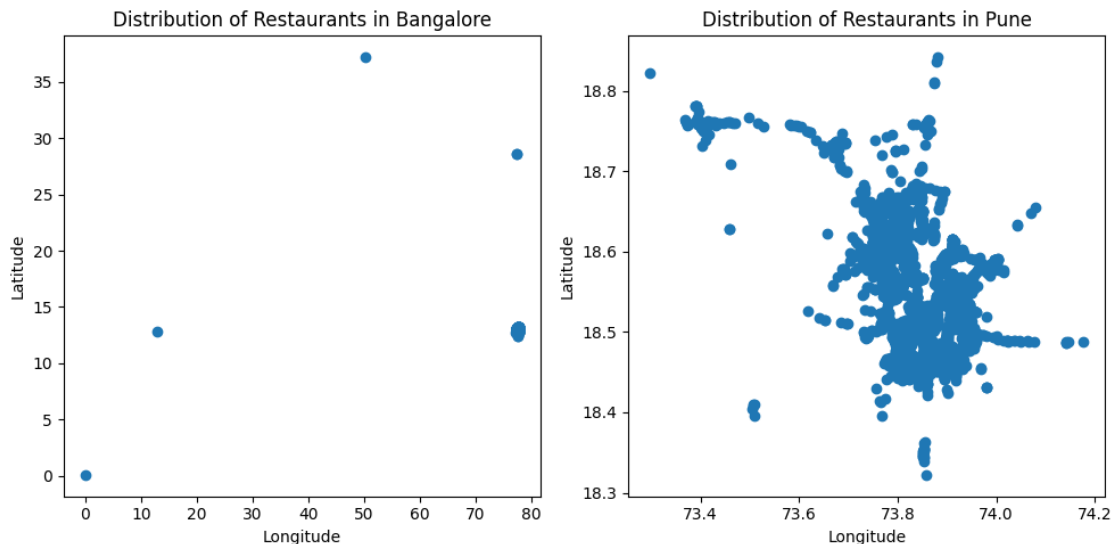
plt.subplot(1, 2, 2)
plt.scatter(pune_restaurants['Longitude'], pune_restaurants['Latitude'],
            c=pune_restaurants['cluster'])
plt.title('Restaurant Clusters in Pune')
plt.xlabel('Longitude')
plt.ylabel('Latitude')

plt.tight_layout()
plt.show()

average_ratings_bangalore = bangalore_restaurants.
    ↳groupby('Locality')['Dining_Rating'].mean().sort_values(ascending=False)
print("Localities in Bangalore with the highest average dining ratings:")
print(average_ratings_bangalore.head(10))

average_ratings_pune = pune_restaurants.groupby('Locality')['Dining_Rating'].
    ↳mean().sort_values(ascending=False)
print("Localities in Pune with the highest average dining ratings:")
print(average_ratings_pune.head(10))

```

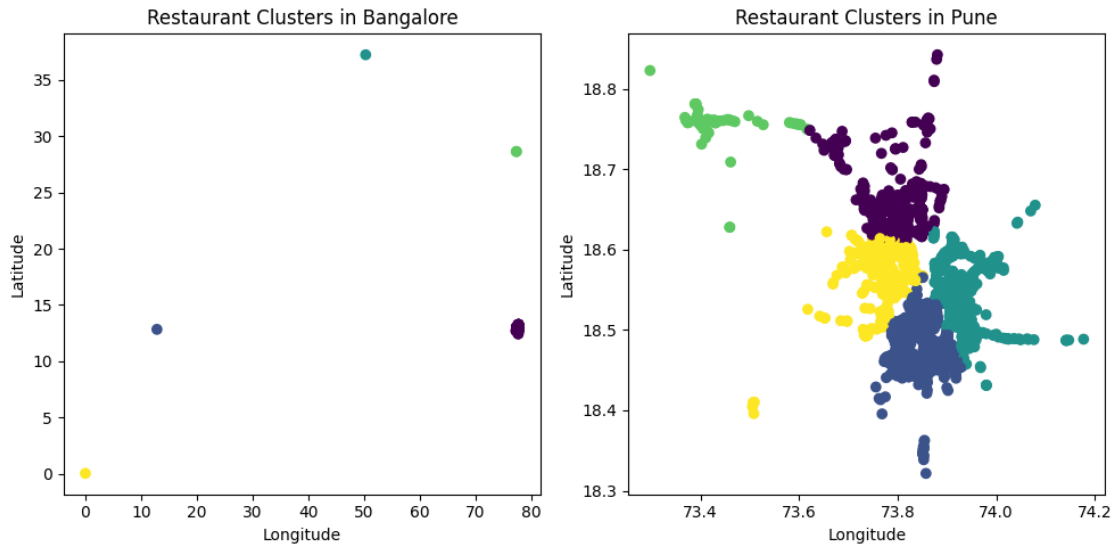


```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning

```

```
warnings.warn(
```



Localities in Bangalore with the highest average dining ratings:

Locality

Windmills Craftworks, Bangalore	4.750000
Eva Mall, Brigade Road, Bangalore	4.600000
Soul Space Arena Mall, Marathahalli, Bangalore	4.600000
Lemon Tree Premier, Ulsoor, Bangalore	4.500000
BluPetal Hotel, Koramangala, Bangalore	4.450000
Royal Meenakshi Mall, Bannerghatta Road, Bangalore	4.400000
OPUS, Sarjapur Road, Bangalore	4.400000
D and N Enterprises, Marathahalli, Bangalore	4.400000
Barton Centre, Bangalore	4.366667
Gilly's Redefined, Koramangala 4th Block, Bangalore	4.333333

Name: Dining_Rating, dtype: float64

Localities in Pune with the highest average dining ratings:

Locality

City Point, Dhole Patil Road, Pune	4.600000
The E-Square Hotel, Shivaji Nagar, Pune	4.500000
White Square, Hinjawadi, Pune	4.450000
Panchshil Tech Park, Yerawada, Pune	4.400000
ABC Farms,Mundhwa, Pune	4.400000
JW Marriott Pune, Senapati Bapat Road, Pune	4.385714
Sayaji Hotels, Wakad, Pune	4.366667
Westend Mall, Aundh, Pune	4.350000
Balewadi High Street, Baner, Pune	4.342857
Marriott Suites Pune, Pune	4.300000

Name: Dining_Rating, dtype: float64