swiggy-recsys

April 19, 2023

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import prince

from sklearn.preprocessing import StandardScaler, LabelEncoder, StandardScaler
from sklearn.neighbors import NearestNeighbors
```

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

[3]:
```
df = pd.read_csv('swiggy-preprocessed.csv', index_col=[0])
df_og = df.copy()
```

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

[4]:

id	name	city	rating	rating_count	cost	lic_no	link	address	menu
567335	AB FOODS POINT	Abohar	0.0	Too Few Ratings	200.0	22122652000138	https://www.swiggy.com/restaurants/ab-foods-po…	AB FOODS POINT, NEAR RISHI NARANG DENTAL CLINI…	Menu/567335.json
531342	Janta Sweet House	Abohar	4.4	50+ ratings	200.0	12117201000112	https://www.swiggy.com/restaurants/janta-sweet…	Janta Sweet House, Bazar No.9, Circullar Road,…	Menu/531342.json
158203	theka coffee desi	Abohar	3.8	100+ ratings	100.0	22121652000190	https://www.swiggy.com/restaurants/theka-coffe…		
187912	Singh Hut	Abohar	3.7	20+ ratings	250.0	22119652000167	https://www.swiggy.com/restaurants/singh-hut-n…		
543530	GRILL MASTERS	Abohar	0.0	Too Few Ratings	250.0	12122201000053	https://www.swiggy.com/restaurants/grill-maste…		

```
158203          theka coffee desi, sahtiya sadan road city  Menu/158203.json
187912     Singh Hut, CIRCULAR ROAD NEAR NEHRU PARK ABOHAR  Menu/187912.json
543530  GRILL MASTERS, ADA Heights, Abohar - Hanumanga…  Menu/543530.json

       sub_area    area          cuisine1   cuisine2
id
567335   Abohar  Abohar         Beverages     Pizzas
531342   Abohar  Abohar            Sweets     Bakery
158203   Abohar  Abohar         Beverages  Beverages
187912   Abohar  Abohar         Fast Food     Indian
543530   Abohar  Abohar  Italian-American  Fast Food
```

[5]:
```
cat_vars = ['sub_area', 'area', 'cuisine1', 'cuisine2', 'rating_count', 'city']
num_vars = ['rating', 'cost']
```

[6]:
```
# Calculate the mean rating for each combination of 'sub_area', 'area', 'city',␣
 ↪'cuisine1', 'cuisine2'
mean_ratings = df.groupby(['sub_area', 'area', 'city', 'cuisine1',␣
 ↪'cuisine2'])['rating'].mean()

# Define a function to replace 0 rating with the mean for its corresponding␣
 ↪combination of 'sub_area', 'area', 'city', 'cuisine1', 'cuisine2'
def replace_rating(row):
    if row['rating'] == 0:
        return mean_ratings.loc[row['sub_area'], row['area'], row['city'],␣
 ↪row['cuisine1'], row['cuisine2']]
    else:
        return row['rating']

# Apply the function to the 'rating' column
df['rating'] = df.apply(replace_rating, axis=1)
```

[7]:
```
le = LabelEncoder()
```

[8]:
```
# Encode the categorical variables using LabelEncoder
for col in cat_vars:
    df[col] = le.fit_transform(df[col])
```

[9]:
```
# import required libraries
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics.pairwise import cosine_similarity
```

[10]:
```
# Select the relevant columns from the dataset
df = df[['sub_area', 'area', 'cuisine1', 'cuisine2', 'rating_count', 'city',␣
 ↪'rating', 'cost']]
```

```
# Scale the 'rating' and 'cost' columns between 0 and 1
scaler = MinMaxScaler()
df[['rating', 'cost']] = scaler.fit_transform(df[['rating', 'cost']])
```

[11]:
```
# Encode the categorical variables using one-hot encoding
encoded_data = pd.get_dummies(df, columns=['sub_area', 'area', 'cuisine1',␣
 ↪'cuisine2', 'city', 'rating_count'])
```

[12]:
```
from sklearn.neighbors import NearestNeighbors

# Fit the NearestNeighbors model
model = NearestNeighbors(metric='cosine', algorithm='brute')
model.fit(encoded_data)
```

[12]:
```
NearestNeighbors(algorithm='brute', metric='cosine')
```

[13]:
```
def recommend_nn(id, model, data, k=5):
    # Find the index of the restaurant with the given id
    index = data.index.get_loc(id)

    # Get the indices and distances of the k-nearest neighbors
    distances, indices = model.kneighbors(data.iloc[index].values.reshape(1,␣
 ↪-1), n_neighbors=k+1)

    # Remove the index of the queried restaurant from the indices list
    indices = indices.squeeze().tolist()[1:]

    # Return the ids of the k-nearest neighbors
    return [data.iloc[i].name for i in indices]
```

[14]:
```
le_dict = {}
for var in cat_vars:
    df[var] = le.fit_transform(df[var])
    le_dict[var] = le
```

[15]:
```
# Test the recommendation function using NearestNeighbors
restaurant_id = 531342  # Janta Sweet House
recommended_ids = recommend_nn(restaurant_id, model, encoded_data)
print(recommended_ids)
```

```
/Users/sudhanva/mambaforge/envs/pt/lib/python3.10/site-
packages/sklearn/base.py:420: UserWarning: X does not have valid feature names,
but NearestNeighbors was fitted with feature names
  warnings.warn(
```

```
[156587, 156590, 161396, 530909, 327360]
```

[16]:
```
# Create a new dataframe to store recommended restaurants
recommended_df = pd.DataFrame(recommended_ids, columns=['id'])
recommended_df = pd.merge(recommended_df, df, on='id')
```

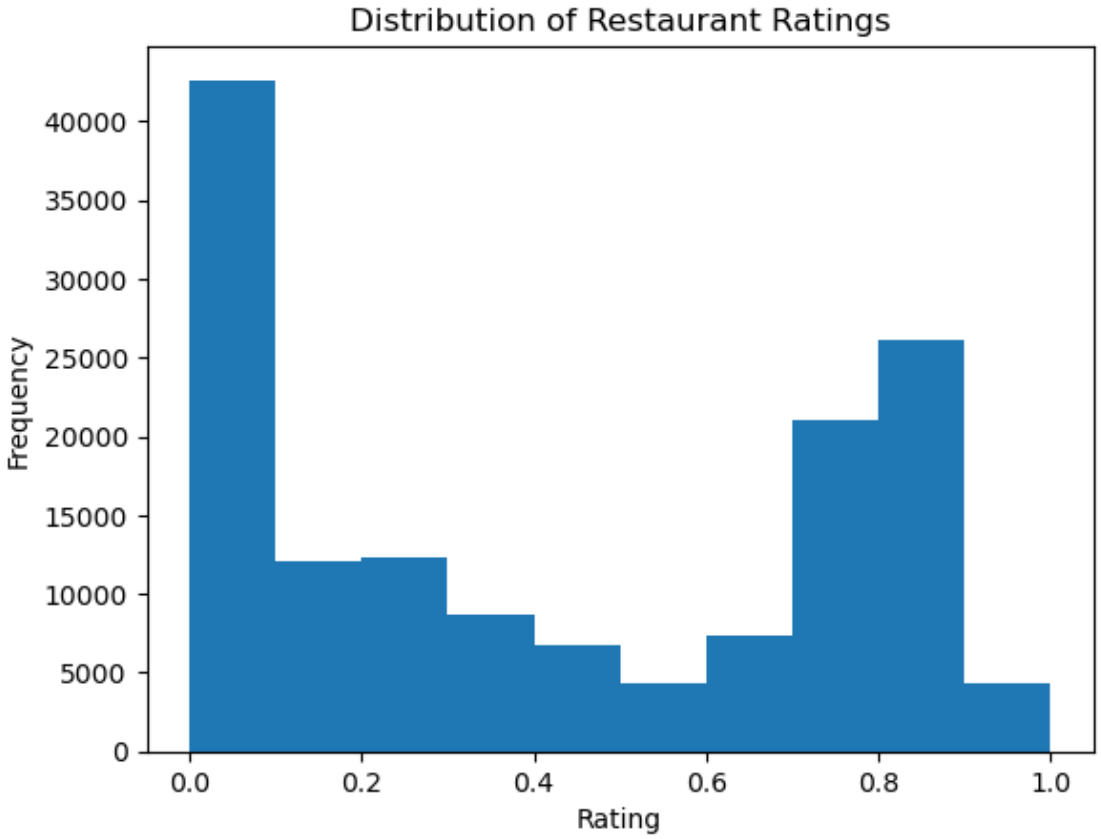
[17]:
```
# Reverse transform categorical variables
for var in cat_vars:
    le = le_dict[var]
    recommended_df[var] = le.inverse_transform(recommended_df[var].astype(int))

# Display recommended restaurants
print("Recommended restaurants based on restaurant_id = ", restaurant_id)
for id in recommended_ids:
    print(df_og.loc[id][['name', 'city', 'rating', 'rating_count', 'cost',␣
 ↪'lic_no', 'link', 'address', 'sub_area', 'area', 'cuisine1', 'cuisine2']])
    print()
```

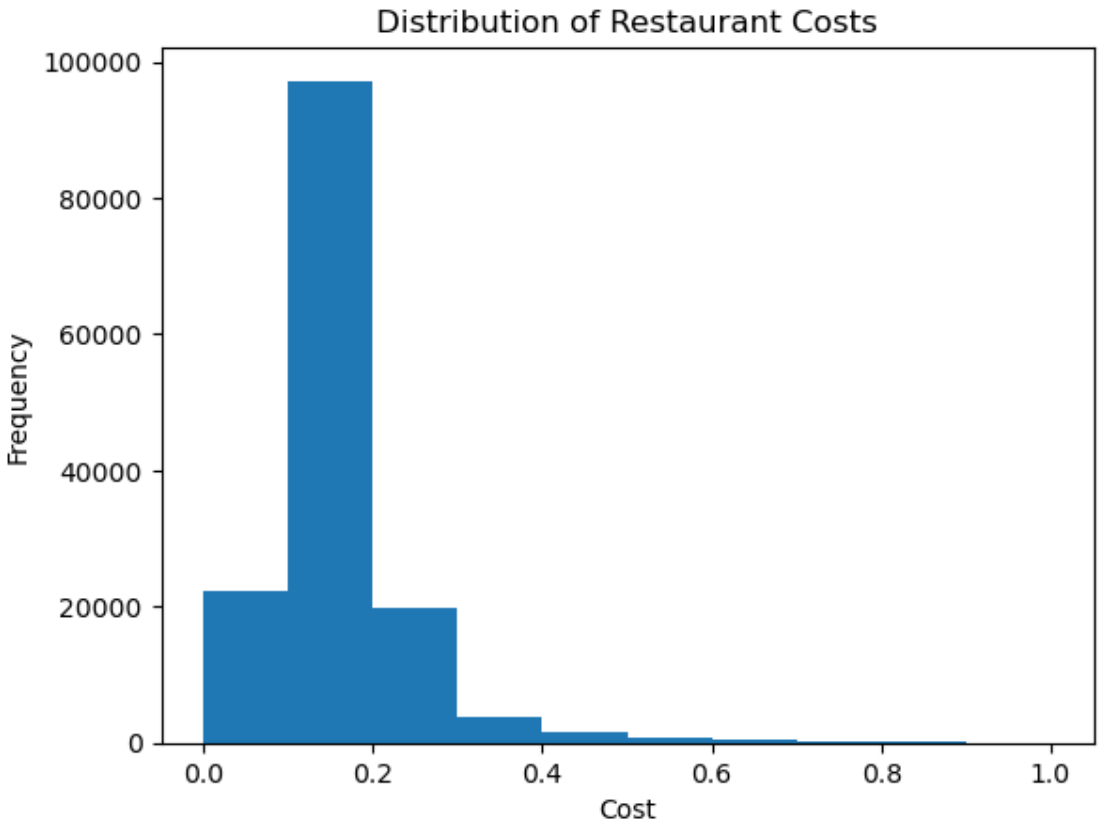
```
Recommended restaurants based on restaurant_id =  531342
name                                          Bharawan Da Dhaba
city                                                     Abohar
rating                                                      4.4
rating_count                                          50+ ratings
cost                                                      300.0
lic_no                                                  license
link            https://www.swiggy.com/restaurants/bharawan-da…
address         Bharawan Da Dhaba, rimpy bakery 12 circular ro…
sub_area                                                 Abohar
area                                                     Abohar
cuisine1                                                 Indian
cuisine2                                                 Indian
Name: 156587, dtype: object

name                                           Sethi Milk Badam
city                                                     Abohar
rating                                                      4.2
rating_count                                          20+ ratings
cost                                                      100.0
lic_no                                          22119652000039
link            https://www.swiggy.com/restaurants/sethi-milk-…
address          Sethi Milk Badam, main bazar street no 11 abohar
sub_area                                                 Abohar
area                                                     Abohar
cuisine1                                                 Sweets
cuisine2                                               Desserts
Name: 156590, dtype: object

name                                         chacha sweets house
city                                                     Abohar
```

```
rating                                                      0.0
rating_count                                    Too Few Ratings
cost                                                      200.0
lic_no                                          22119652000095
link            https://www.swiggy.com/restaurants/chacha-swee…
address         chacha sweets house, h no 2402 st no 24 gausha…
sub_area                                                Abohar
area                                                    Abohar
cuisine1                                                Sweets
cuisine2                                             Beverages
Name: 161396, dtype: object

name                                                FOODY MOOD
city                                                    Abohar
rating                                                      4.7
rating_count                                        20+ ratings
cost                                                      300.0
lic_no                                          22122652000115
link            https://www.swiggy.com/restaurants/foody-mood-…
address         FOODY MOOD, Arya nagar St no 7 Abohar, Abohar …
sub_area                                                Abohar
area                                                    Abohar
cuisine1                                            Fast Food
cuisine2                                              Chinese
Name: 530909, dtype: object

name                                  Haldirams Sweets & Namkeen
city                                                    Nagpur
rating                                                      4.7
rating_count                                        50+ ratings
cost                                                      300.0
lic_no                                                  license
link            https://www.swiggy.com/restaurants/haldirams-s…
address         Haldirams Sweets & Namkeen, Plot no. 1288/1042…
sub_area                                              Nandanvan
area                                                  Nandanvan
cuisine1                                                Sweets
cuisine2                                                Bakery
Name: 327360, dtype: object
```

[18]:
```python
# Plot the distribution of restaurant ratings
plt.hist(df['rating'], bins=10)
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.title('Distribution of Restaurant Ratings')
```

```
plt.savefig(plots_path + 'distribution-of-rest-ratings.png',␣
  ↪bbox_inches='tight')
plt.show()

# Plot the distribution of restaurant costs
plt.hist(df['cost'], bins=10)
plt.xlabel('Cost')
plt.ylabel('Frequency')
plt.title('Distribution of Restaurant Costs')
plt.savefig(plots_path + 'distribution-of-rest-costs.png', bbox_inches='tight')
plt.show()
```




Distribution of Restaurant Costs
Frequency
100000
80000
60000
40000
20000
0
0.0
0.2
0.4
0.6
0.8
1.0
Cost