Coursera CapstoneProject Report

Introduction

Osaka - Local restaurant research before travel.

Osaka shares some similarities with Tokyo, the Capital of Japan.
Inspire by https://medium.com/@radialee/capstone-project-the-battle-of-neighborhoods-in-tokyo-restaurants-45a503e65ff, the mentioned project studied Tokyo's restaurant data since it will be the city holding 2020 Olympic Games and discuss how to support increasing visitors on one of their needs - Food. As Japan's third largest city (credit to: http://www.webster.edu/study-abroad/programs/japan/osaka/facts.html), I am curious about Osaka's potential. I had visited Osaka once before in person, this time I will take a different approach to view this city, start from its local restaurant.

Further discussion on the problem

In 2019, Japan's Tourism Hits Record Numbers (credit to: https://www.tourism.jp/en/tourism-database/stats/inbound/), the amount of foreign visitors and their spending during travel also record a huge boost (credit to : https://www.prnewswire.com/news-releases/japans-tourism-hits-record-numbers-in-2019-300979556.html). Although Osaka was considered a well-developed city, visitors may encounter difficult to find desired restaurant. In this study, I hoped to apply data from Foursquare's location data, apply clustering methods (mainly K-means), breakdown Osaka as clusters and identify which cluster visitors should go for certain type of restaurants.

Data source that will be used in this project

Data Source:

1. Wards - https://en.wikipedia.org/wiki/Osaka Data will be scrapped from table "Wards of Osaka" for us have a brief view of Osaka's base distribution on Population and Population density
2. Restaurant data - Via Foursquare API Data will be import via API for us to get the restaurant location within Osaka.

Methodology

Data Preparation

We can scrape the wards of Osaka from https://en.wikipedia.org/wiki/Osaka and use Pandas to transform the table into a dataframe for further use.

```
In [5]: df = pd.read_html('https://en.wikipedia.org/wiki/Osaka')[4] #extract the fourth table in Osaka Wiki (Ward of Osaka)
```

```
In [6]: df.head()
```

Out[6]:

Wards of Osaka							
Unnamed: 0_level_1	Place Name			Unnamed: 3_level_1	Unnamed: 4_level_1	Unnamed: 5_level_1	Map of Osaka
Unnamed: 0_level_2	Rōmaji	Kanji		Population	Land area in km2	Pop. density per km2	Unnamed: 6_level_2
0	1	Abeno-ku	阿倍野区	107000	5.99	18440	NaN
1	2	Asahi-ku	旭区	90854	6.32	14376	NaN
2	3	Chūō-ku	中央区	100998	8.87	11386	NaN
3	4	Fukushima-ku	福島区	78348	4.67	16777	NaN
4	5	Higashinari-ku	東成区	83684	4.54	18433	NaN

After some clean up, we combine the table with Longitude and Latitude.

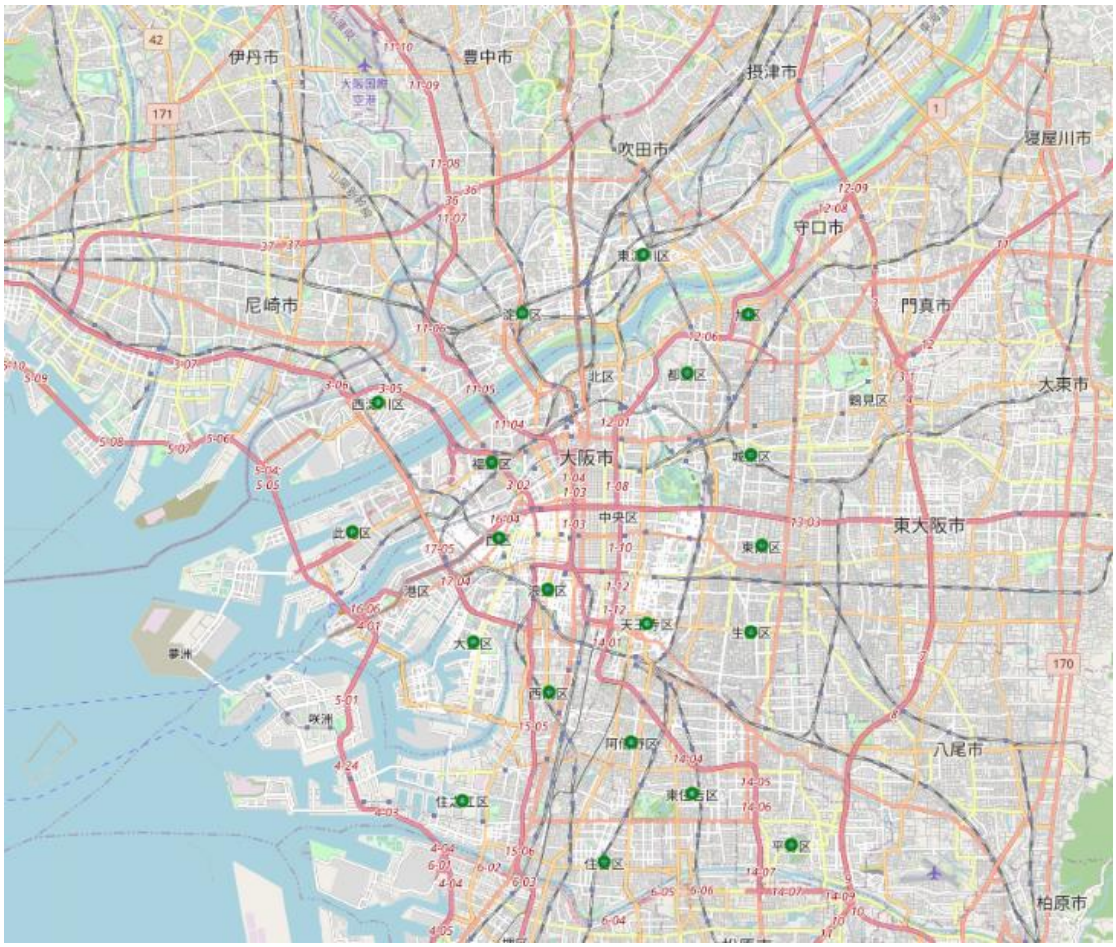
```
In [9]: geolocator = Nominatim(user_agent="Osaka_travel")
        df['Ward_Coordination']= df['Kanji'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
        df[['Latitude', 'Longitude']] = df['Ward_Coordination'].apply(pd.Series)
        df.drop(['Ward_Coordination'], axis=1, inplace=True)
```

```
In [10]: df
```

Out[10]:

	ID	Neighbourhood	Kanji	Population	Land area in km2	Pop. density per km2	Latitude	Longitude
0	1	Abeno-ku	阿倍野区	107000	5.99	18440	34.627501	135.514095
1	2	Asahi-ku	旭区	90854	6.32	14376	34.726483	135.546952
2	3	Chūō-ku	中央区	100998	8.87	11386	35.666255	139.775565
3	4	Fukushima-ku	福島区	78348	4.67	16777	34.692104	135.474812
4	5	Higashinari-ku	東成区	83684	4.54	18433	34.672912	135.550567
5	6	Higashisumiyoshi-ku	東住吉区	126704	9.75	12995	34.615662	135.531096
6	7	Higashiyodogawa-ku	東淀川区	176943	13.27	13334	34.740212	135.517432
7	8	Hirano-ku	平野区	193282	15.28	12649	34.603715	135.559027
8	9	Ikuno-ku	生野区	129641	8.37	15489	34.653003	135.547722
9	10	Jōtō-ku	城東区	167925	8.38	20039	34.693887	135.547769
10	11	Kita-ku (administrative center)	北区	136602	10.34	13211	35.755838	139.736687
11	12	Konohana-ku	此花区	65086	19.25	3381	34.676245	135.435550
12	13	Minato-ku	港区	80759	7.86	10275	35.643227	139.740055
13	14	Miyakojima-ku	都島区	107555	6.08	17690	34.712703	135.529697
14	15	Naniwa-ku	浪速区	74992	4.39	17082	34.662830	135.490485
15	16	Nishi-ku	西区	103089	5.21	19787	34.674598	135.476774
16	17	Nishinari-ku	西成区	108654	7.37	14743	34.639074	135.490813
17	18	Nishiyodogawa-ku	西淀川区	95960	14.22	6748	34.705938	135.442936
18	19	Suminoe-ku	住之江区	120629	20.61	5853	34.614132	135.466545
19	20	Sumiyoshi-ku	住吉区	153425	9.40	16322	34.599765	135.506426
20	21	Taishō-ku	大正区	62872	9.43	6667	34.650640	135.469570
21	22	Tennōji-ku	天王寺区	80830	4.84	16700	34.655043	135.518370
22	23	Tsurumi-ku	鶴見区	111501	8.17	13648	35.480132	139.693663
23	24	Yodogawa-ku	淀川区	182254	12.64	14419	34.726613	135.483397

We then can use folium library in python to create a map preview of Osaka and the Wards mentioned in above table.



Using Foursquare API to gather venue data for further analysis

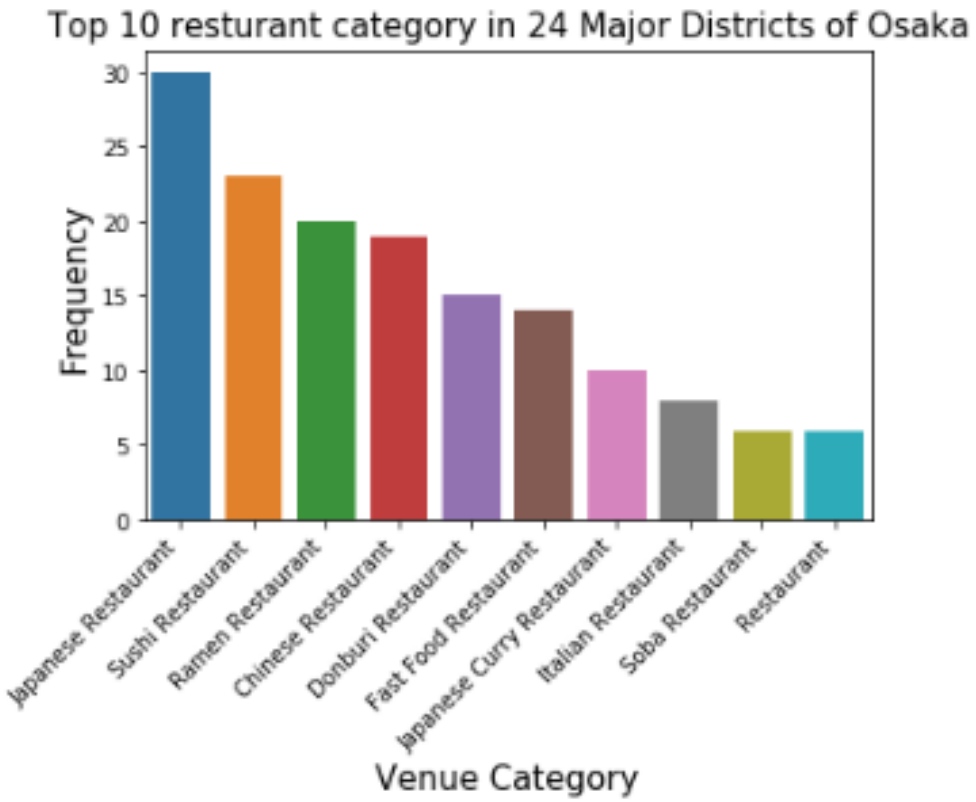
We will be using Foursquare API to get top 100 venues that are in these 24 wards within 500 meters, respectively.

```
In [20]: print (Osaka_Filter_restaurant['Venue Category'].value_counts())
```

```
Japanese Restaurant          30
Sushi Restaurant             23
Ramen Restaurant             20
Chinese Restaurant           19
Donburi Restaurant           15
Fast Food Restaurant         14
Japanese Curry Restaurant    10
Italian Restaurant            8
Soba Restaurant               6
Restaurant                    6
Udon Restaurant               5
Yakitori Restaurant           4
French Restaurant             4
Tempura Restaurant            3
Indian Restaurant             3
Dumpling Restaurant           2
Unagi Restaurant              2
Kushikatsu Restaurant         2
Seafood Restaurant            2
Spanish Restaurant            2
Sri Lankan Restaurant         1
Yoshoku Restaurant            1
Kebab Restaurant              1
Tonkatsu Restaurant           1
Halal Restaurant              1
Kosher Restaurant             1
Vietnamese Restaurant         1
Kaiseki Restaurant            1
German Restaurant             1
Korean Restaurant             1
Name: Venue Category, dtype: int64
```

```
In [21]: print('There are {} uniques restaurant categories.'.format(len(Osaka_Filter_restaurant['Venue Category'].unique())))
```

```
There are 30 uniques restaurant categories.
```



Looks like Osaka is a Japanese food focus city (Ramen, Sushi, Traditional Japanese dishes) with a few Chinese and fast food shop you could visit. Let us further dive into the data to explore more.

Here we first transform our data into labels, then group up to get the mean by each neighbour.

In [27]:
```
Osaka_onehot = pd.get_dummies(Osaka_Filter_restaurant[['Venue Category']], prefix="", prefix_sep="")

# add neighbourhood column back to dataframe
Osaka_onehot['Neighbourhood'] = Osaka_Filter_restaurant['Neighbourhood']
cols = Osaka_onehot.columns.tolist()
cols = cols[-1:] + cols[:-1]
Osaka_onehot = Osaka_onehot[cols]
Osaka_onehot.head()
```

Out[27]:

	Neighbourhood	Chinese Restaurant	Donburi Restaurant	Dumpling Restaurant	Fast Food Restaurant	French Restaurant	German Restaurant	Halal Restaurant	Indian Restaurant	Italian Restaurant
1	Abeno-ku	0	0	0	1	0	0	0	0	0
2	Asahi-ku	0	0	0	1	0	0	0	0	0
3	Asahi-ku	0	1	0	0	0	0	0	0	0
4	Asahi-ku	1	0	0	0	0	0	0	0	0
5	Asahi-ku	0	0	0	0	0	0	0	0	0

5 rows × 31 columns

Let’s start using clustering to analyze the data, here we will be using K-Means as our method.

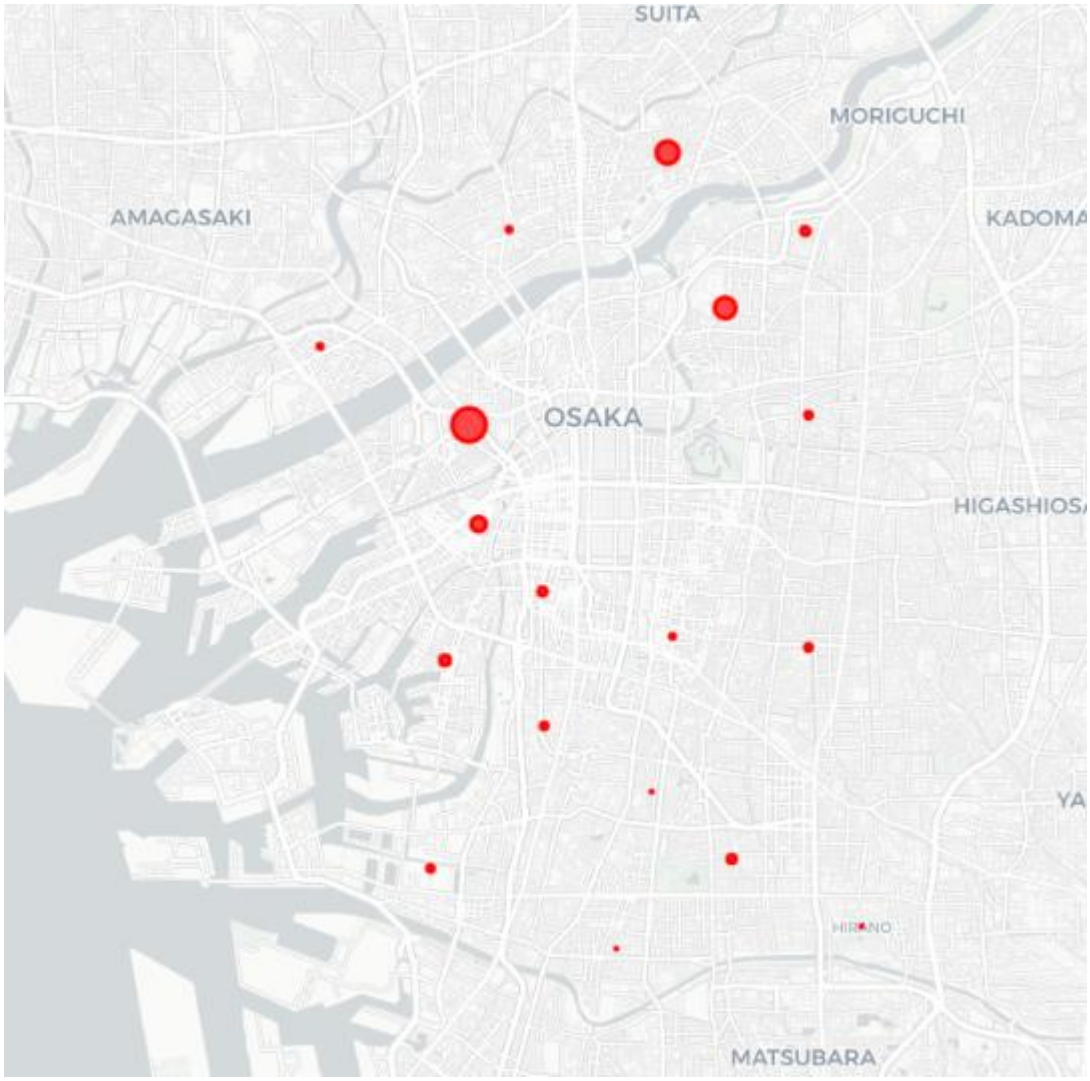
In [34]:
```
# set number of clusters
kclusters = 6

Osaka_grouped_clustering = Osaka_grouped.drop('Neighbourhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Osaka_grouped_clustering)
```

In [36]:
```
# add clustering labels
Neighbourhoods_venues_sorted.reset_index()

Neighbourhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Osaka_merged = df

Osaka_merged.rename(columns={'Name':'Neighbourhood'}, inplace=True)

# merge osaka_grouped with osaka_data to add latitude/longitude for each neighbourhood
Osaka_merged = Osaka_merged.join(Neighbourhoods_venues_sorted.set_index('Neighbourhood'), on='Neighbourhood')

Osaka_merged_exclude = Osaka_merged[Osaka_merged['Cluster Labels'].isnull()]

Osaka_merged = Osaka_merged[Osaka_merged['Cluster Labels'].notna()]

Osaka_merged.head() # check the last columns!
```

Out[36]:

	ID	Neighbourhood	Kanji	Population	Land area in km2	Pop. density per km2	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	1	Abeno-ku	阿倍野区	107000	5.99	18440	34.627501	135.514095	3.0	Fast Food Restaurant	Yoshoku Restaurant	Yakitori Restaurant
1	2	Asahi-ku	旭区	90854	6.32	14376	34.726483	135.546952	4.0	Chinese Restaurant	Donburi Restaurant	Udon Restaurant
2	3	Chūō-ku	中央区	100998	8.87	11386	35.666255	139.775565	1.0	Sushi Restaurant	Japanese Restaurant	Soba Restaurant
3	4	Fukushima-ku	福島区	78348	4.67	16777	34.692104	135.474812	1.0	Fast Food Restaurant	Chinese Restaurant	Donburi Restaurant
5	6	Higashisumiyoshi-ku	東住吉区	126704	9.75	12995	34.615662	135.531096	4.0	Chinese Restaurant	Vietnamese Restaurant	Fast Food Restaurant

After we have added the cluster label into the data, we are going to visualize the result in a map preview.



Result Discussion

Summarize our findings in this project:

- Restaurants in Osaka mainly focus on Japanese dishes, good for travelers who like to try out Japanese food.
- The largest cluster contains mainly Japanese Food restaurants eg Kita-ku (administrative center) and Yodogawa-ku
- Chūō-ku has the highest amount of restaurant (59 records) through out all 24 wards
- Sumiyoshi-ku, Abeno-ku and Hirano-ku has the least amount of restaurant (1 record in each district)

Problem of above clustering method:

First, we only picking venue categories that contains “restaurant”, ignored the restaurant types like sake bar, Café, Steak house will affect the result of clustering since the number of restaurants will be different. Hope Foursquare API could provide a category and sub-category of venues for more clear venue identification.

Second, data we used rely on Foursquare’s API data, if there were some shops that have not been recorded, our result will be inaccurate.

Third, this analysis focused only on geographic location of each restaurant in Osaka, we had ignored other factors like the detail

There were 3 wards that cannot be labeled by K-Means due to no restaurant records was found on those wards, that will also make the result bias.

Conclusion

This is a fresh look for me to explore Osaka in a data perspective. I was excited for the result and how it can help my / other's next travel, I think viewing suggestion on data is far better than reading tour guide as it may be outdated due to its published date.

We can see Osaka is indeed a Japanese food focused city, strongly suggest people who love Japanese food to give it a try. Besides restaurants, Convenience Store is also a good place to go for a quick breakfast, refills or mid night snack, I personally loved to buy a cup noodle and fried chicken when I was heading back hotel.

Hope all of you can enjoy your travel to Japan!