Discovering the Ideal Location for an Italian Café

1. Introduction

1.1 Background

Small restaurants require several elements to be successful in the modern day. A single misstep can mean the difference between success or failure given the competitive landscape. Our client is a first-time business owner looking to open an Italian themed café near the Sangenjaya district in Tokyo. He has hired our team to create a competitive edge for his restaurant by utilizing Data Science to guide his decisions.

1.2 Business Problem

Our client needs to determine the best location for his Italian themed café. He has asked us to find a location that is close to other Italian themed shops to help with the atmosphere while being far away from other cafés and coffee shops. He's hoping this will create a natural flow of customers looking for a late-night dessert or espresso after visiting one of these Italian restaurants.

2. Data acquisition and cleaning

2.1 Data sources

We will be primarily using location data provided by Foursquare and their available API. This will give our team access to a large number or relevant restaurants which includes the frequency, category and location data. Based off a one kilometer radius 50 Italian restaurants and 52 café/coffee shops were identified. A sample of this data is shown below:

	Venue Name	Venue Latitude	Venue Longitude	Venue Category	Category
0	LOVEL DINING (ラブルダイニング)	35.642739	139.667546	Italian Restaurant	Italian
1	caZé uzumaki	35.642771	139.668341	Italian Restaurant	Italian
2	ヨーロッパ食堂	35.645003	139.669534	Italian Restaurant	Italian
3	il Pizzaiolo (イル ピッツァイオーロ)	35.644332	139.668702	Italian Restaurant	Italian
4	comodo kitchen	35.642195	139.672997	Italian Restaurant	Italian

Figure 1: Example data provided by the Foursquare API

2.2 Data cleaning

Foursquare has a robust and complete dataset so little data cleaning was required. The main step was to convert the *Venue Category* to a binary value to easily distinguish which type of venue we were working with, café or Italian restaurant. An example of the resulting dataset is shown below:

	Venue Name	Venue Latitude	Venue Longitude	Venue Category	Category	Category Type
0	LOVEL DINING (ラブルダイニング)	35.642739	139.667546	Italian Restaurant	Italian	1
1	caZé uzumaki	35.642771	139.668341	Italian Restaurant	Italian	1
2	ヨーロッパ食堂	35.645003	139.669534	Italian Restaurant	Italian	1
3	il Pizzaiolo (イル ピッツァイオーロ)	35.644332	139.668702	Italian Restaurant	Italian	1
4	comodo kitchen	35.642195	139.672997	Italian Restaurant	Italian	1
5	Rainbow	35.644361	139.670262	Italian Restaurant	Italian	1
6	Vineria il Chiasso	35.640591	139.674575	Italian Restaurant	Italian	1
7	ワイン食堂 INASEYA MARE	35.643214	139.662873	Italian Restaurant	Italian	1
8	Bricca	35.640972	139.673939	Italian Restaurant	Italian	1
9	Trattoria e Pizzeria L'ARTE	35.643074	139.672357	Italian Restaurant	Italian	1

Figure 2: Cleaned data set showing Venues, Locations and Category Type

Using the above data, we were able to plot venues on a map of Sangenjaya.

Figure 3: Sangenjaya Italian restaurants (purple) and cafe (red) plotted

3. Exploratory Data Analysis

3.1 Determining venue zones

Our goal is to determine areas with high amounts of Italian restaurants and low cafés. To accomplish this the data was categorized using a k-means analysis on each type of venue, Italian restaurants, and

cafés separately. Both analyses were based off a 10 zone range repeated 12 times to minimize local minimums. Location data as inputs to determine clusters or 'zones' of venues.

Data was summarized by averaging each zone's location data and sorting by frequency of restaurants. Results of both zones are shown below:

```
Italian_Zones.sort_values('Count', ascending = False)
```

Label	Venue Latitude	Venue Longitude	Count
9	35.644661	139.669929	11
7	35.642637	139.668303	10
1	35.642774	139.671979	9
5	35.641158	139.674108	5
3	35.638236	139.667478	4
2	35.638837	139.673238	3
8	35.642781	139.661268	3
0	35.643239	139.663591	2
4	35.647326	139.669094	2
6	35.636049	139.664232	1

Figure 4 K-Means results - Italian Zones sorted by count

```
Coffee_Zones.sort_values('Count', ascending = False)
```

Label	Venue Latitude	Venue Longitude	Count
1	35.643287	139.669788	20
7	35.643146	139.671923	7
0	35.640021	139.674534	5
3	35.640385	139.670994	5
8	35.646304	139.670547	4
2	35.646256	139.660810	3
4	35.649213	139.667968	3
6	35.644106	139.665658	3
5	35.635465	139.662173	1
9	35.646011	139.675612	1

Figure 5 K-Means results – Coffee Zones sorted by count

The resulting zones were plotted to better visualize their locations. Opacity was chosen as a measure of high density zones and the colors blue and red were chosen for the Italian and Coffee zones respectively.

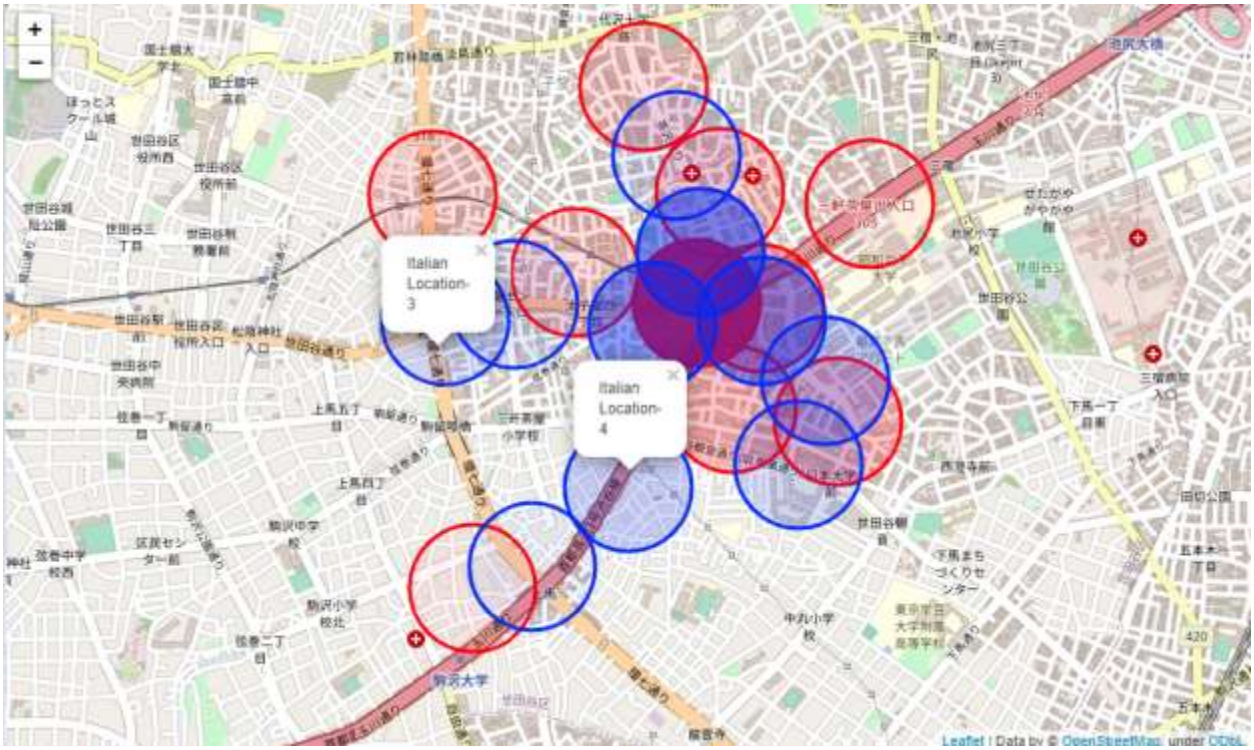

Figure 6: Italian (blue) and Coffee (red) zones mapped. Opacity shows higher frequency of restaurants

4. Results and Discussion

The clients request was to find zones in Sangenjaya, Tokyo where there many Italian restaurants exist but few cafés or coffee shops. Based on Figure 6 only Italian Zones 3 and 4 meet this criteria.

	Venue Latitude	Venue Longitude	Label
22	35.638462	139.667464	3
23	35.638473	139.668007	3
25	35.638218	139.668060	3
29	35.637790	139.666380	3
43	35.646284	139.669689	4
49	35.648367	139.668500	4

Figure 7 List of Italian venues located in Zones 3 and 4

Since there exists four venues in Zone 3 and two in Zone 4, Zone 3 best meets the original requirements.

5. Conclusion

Based on the results of this exploratory analysis it seems like the ideal location to build an Italian themed café would be in Zone 3 in Figure 6. It should be noted that other factors such as foot traffic, available locations, transportation and commercial zoning (to name a few) should also be considered before proceeding with a final decision.