Irvine new restaurant study

Table of Contents

Introduction ***1***

Data description ***1***

Methodology ***1***

Results ***2***

Discussion ***6***

Conclusion ***6***

Introduction

The scenario under analysis here is that a immigrant is moving to South California and he wants to open any kind of restaurant. He is moving specifically to a town called Irvine, which is located in Orange County. Several companies are there, over 60 malls and over 40 villages are also part of the picture. But where should he open this restaurant considering all business that are located near those villages? The main audience to this storytelling line is business persons who recently moved into town.

The data was retrieved using Foursquare API and it describes what kind of business that are located within 1 mile from each village. The goal here is to make an analysis from the distribution of venues from all over the city and specific to each village.

Data description

Next we use the Notebook to build the code to scrape the Irvine Wikipedia page, https://en.wikipedia.org/wiki/Irvine,_California, in order to obtain the data that is input for a table of Irvine villages and to transform the data into a pandas dataframe.

The main data consists of a dataframe containing all Irvine villages. Besides that, one also needs to retrieve their coordinates (longitude and latitude), not to mention all venues around them along with their type.

Methodology

At this point, all necessary data to solve this problem is already gathered. The method to get to an answer is:

1. List the number of venues per type and sort the list to get the most frequent type of venues. Those are the most frequent type of venues and they are also the type of venues which most of the population is interest in.
2. Get the most frequent type of restaurant. That would be our goal type of restaurant. Using the same assumption from step 1, one can conclude that the type de restaurant indicated by step 2 is the type of restaurant that attracts most of the population interest.
3. List the villages around which there is no restaurant from the goal type of restaurant. Those are possibly the villages around which there is no restaurant from the goal type even though it is the most demanded type of restaurant within this town.
4. But one main concern about the result form step 3 is to accidentally list a village around which there is no restaurants or just a few, which would indicate that there is a very small demand for restaurants around this area. So the next step would be to list the top villages with the largest number of restaurants around them.
5. Now, it is natural to cross data from both step 3 and step 4. For each village from step 4, check if it is contained in step 3 list and take the first occurrence of that crossed reference. This would represent a village around which there is no restaurants from goal type but also around which there are several restaurants. This would be the goal village.
6. Once it is already decided the village around which the businessperson should open the restaurant (from the goal type), it is still important to decide where exactly to open it. In order to do so, the k-means algorithm is used to cluster all restaurants around the resulting village and all types of restaurant belonging to each cluster are then analyzed.
7. The resulting area should be a place where there would be minimum to none similar restaurants in such a way that all or most of the demand is attracted toward this businessperson restaurant.

Results

According to step 1 and 2, the most frequent types of restaurants in Irvine are given by Table 1. From this table, one can conclude that the goal type of restaurant is Chinese restaurant.

Table 1: Most frequent type of restaurants in Irvine.

Type	Total
Chinese Restaurant	88
Korean Restaurant	86
Fast Food Restaurant	72
Japanese Restaurant	56
Mexican Restaurant	51
American Restaurant	25
Indian Restaurant	24
Vegetarian / Vegan Restaurant	21
Seafood Restaurant	20
Asian Restaurant	17

Then following step 3, the villages around which there is no restaurants from the goal type (Chinese) can be found in Figure 1.

```
Columbus Grove
Lambert Ranch
Northpark
Orchard Hills
Portola Springs
Quail Hill
Shady Canyon
The Colony
The Willows
Turtle Ridge
Turtle Rock
University Hills
University Park
```

Figure 1: Irvine Villages around which there are no Chinese restaurants.

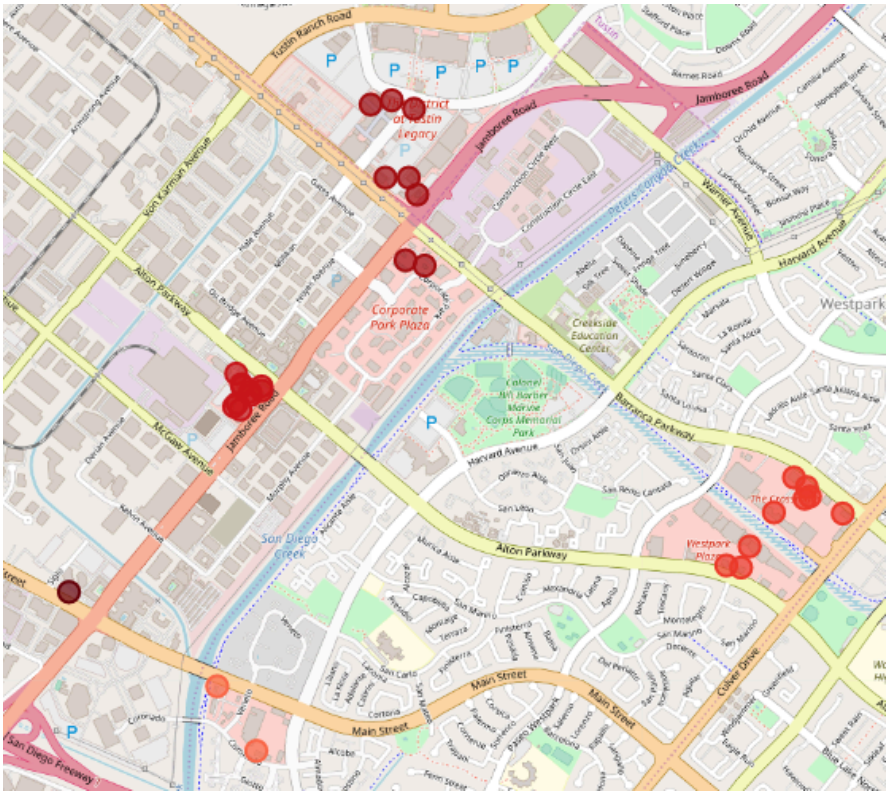
Step 4 reminds the concern about not indicating a area with poor demand toward restaurants. In order to so do, a list of the villages with the largest number of restaurants around them. The list can be seen in Table 2.

Neighborhood	Total
El Camino Real	33
The Willows	30
Northwood	28
College Park	26
Harvard Square	26
Northpark Square	25
Westpark	24
Racquet Club	24
Walnut	24
Cypress Village	23

Table 2: Irvine villages with the largest number of restaurants around them.

The next step (5) states that the resulting data from both steps 3 and 4 must be crossed to reach the village around which there is no restaurants from goal type (Chinese) but also around which there are several restaurants. This village would be: The Willows, around which there are 30 restaurants, but none of them is Chinese.

Finally, one must decide which region around the goal village would be used. The k-means algorithm divided the local restaurants into 5 clusters and can be visualized in Figure 2. This algorithm was selected because it is capable of grouping the data based on the provided data. In this case, the provided data would be the coordinates (latitude and longitude) of the restaurants that can be found within a 1 mile range from the goal village.



Labels	Venue
0	Fast Food
0	Mexican
1	New American
1	Japanese
1	Mexican
1	Italian
1	Mediterranean
1	Mexican
1	New American
1	Seafood
1	Mediterranean
2	Sushi
2	Korean
2	Korean
2	Hotpot
2	Cajun / Creole
2	Asian
2	Japanese
2	Japanese
2	Japanese
2	Japanese
3	Cantonese
3	Sushi
3	Asian
3	Halal
3	Fast Food
3	Fast Food
3	Thai
3	Middle Eastern
4	New American

Figure 2: Clusters of restaurants around the goal village.

Analyzing the list of venue types for each cluster on can conclude that:

- Even though there are no Chinese restaurants in any of these clusters, there are some restaurants in these clusters that could be classified in a super group that can be considered to contain Chinese restaurants, which is Asian restaurants.
- Clusters 0 and 4 do not contain any Asian restaurants

- Clusters 2 and 3 have several Asian restaurants, with a high percentage with respect to the total number of restaurants in each cluster.
- Cluster 1 has the fewest options of Asian restaurants

Discussion

The above results were limited to the maximum number of venues per village. This limitation is due to the free Foursquare account. A more accurate result would be reached if a larger limit was allowed.

A deeper analysis could have been done by looking at population origin, incoming status, restaurant customer records and so on. But this info could not be found on the required time.

Conclusion

So, as a result of all conclusions mentioned above, one can say that:

- If the businessperson prefers no geographical or conceptual-wise competition, the best clusters would have to be 0 or 4. But the businessperson would have to keep in mind that those are options with very few restaurants.
- If it is intended to provide a Chinese restaurant to a cluster that is already well served by Asian restaurants options but has no Chinese restaurant whatsoever, taking advantage of the fact that some customers of other restaurants may be attracted to this new restaurant, the best cluster would have to be 2.
- If the businessperson is looking for a cluster that represents some trade-off between number of restaurants and types of already stablished restaurants, the best cluster would have to be number 1.