Determining ideal location for Malaysian Restaurant in San Diego

IBM Capstone Project - Battle of the Neighbourhoods

Introduction/Business Problem

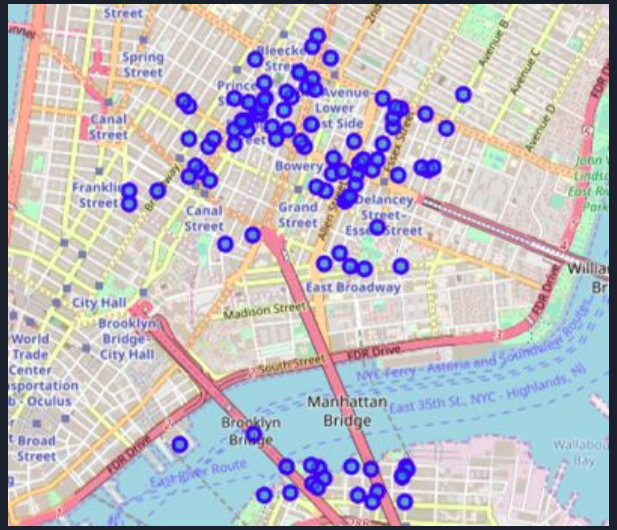
- Lack of Malaysian Restaurants in San Diego means there is an open market to enter.
- Already a large Asian population in San Diego, as well as a wide variety of Asian restaurants.
    - Restaurants of Korean, Thai, Japanese and Chinese cuisine already proven to be very popular among people in San Diego.
- Many Malaysian university students in San Diego with no place to go to have a taste of home.
- This project aims to find the ideal location within San Diego to open a Malaysian restaurant.

Data - Existing Restaurants

- Manual search of already popular Malaysian/Southeast Asian restaurants in the United States
  - To analyze the neighbourhoods/areas that they are in, and help find similar neighbourhoods in San Diego.
  - Restaurants chosen were:
    - Lukshon in Culver City, California
    - Kopitiam in New York City, New York
    - Kedai Makan in Seattle, Washington
    - Pok Pok in Portland, Oregon

- Foursquare API used to gather data on the venues around the restaurants.
  - Used to find frequencies of venue categories in each location.

Data - Existing Restaurants



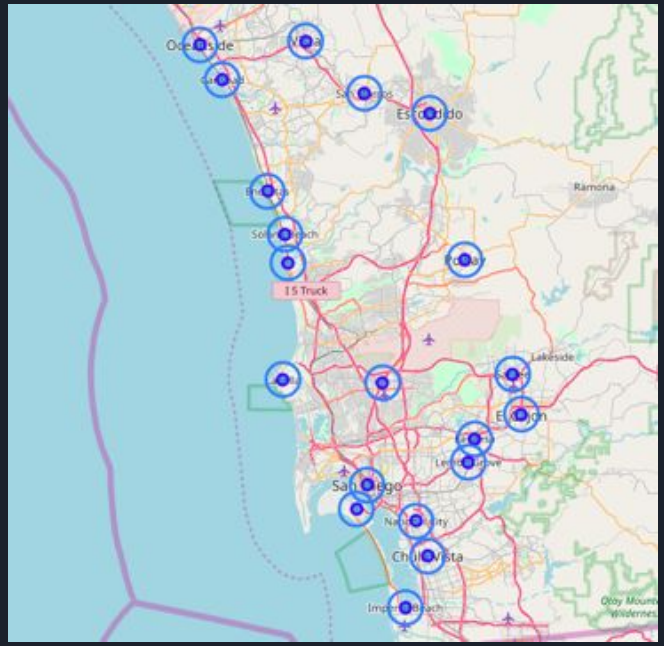
Map showing venues around Kopitiam restaurant

	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lukshon	34.029960	-118.384247	Asian Restaurant
1	Room & Board	34.030223	-118.384531	Furniture / Home Store
2	Arcana: Books on the Arts	34.030146	-118.383500	Bookstore
3	The COOLHAUS Shop	34.030329	-118.381487	Ice Cream Shop
4	H.D. Buttercup L.P.	34.030752	-118.385067	Furniture / Home Store

Table showing first 5 rows in dataframe containing venues around Lukshon

Data - San Diego

- Foursquare API was also used to gather venues in the cities of San Diego County.
- Similar analysis of venues was performed to get the details of each city, in order to perform comparison with the locations of existing restaurants.



Cities in San Diego County with 2000 meter radius

Methodology

- Neighbourhood/Cities were analyzed with a 2000 meter (2km) radius around the centers obtained through Geocoding.
  - This is done to reduce overlap between the areas analyzed.
- First, the locations around the 4 identified existing restaurants were analyzed.
  - Venue category frequencies were calculated to identify the most common categories within the 2km radius of the restaurants.
- From these frequencies, an aggregate was calculated to get an average of the 4 locations.
  - Mean of frequencies among the 4 locations were taken.
  - To be used in comparison with the cities in San Diego County.
- The same analysis is done on the cities in San Diego County, and each city is compared to the aggregate calculated.
- From the “most similar” city in San Diego to the aggregate, k-means clustering was performed to find a cluster of possible locations/cities.

Analysis - One Hot Encoding

- One hot encoding was performed in order to find the most common venue categories in each location.
- The average of the frequencies calculated were taken to get the aggregate with which to compare to the cities in San Diego.

- Similar data was collected on the cities in San Diego county.
- Tables containing the most common venues in each location were produced.

Analysis - One Hot Encoding

```
----Grouped----
                        venue  freq
0                 Coffee Shop  0.08
1                        Café  0.03
2                     Theater  0.02
3                 Cocktail Bar  0.02
4                       Hotel  0.02
5             Ice Cream Shop  0.02
6          Italian Restaurant  0.02
7                   Juice Bar  0.02
8          Mexican Restaurant  0.02
9  New American Restaurant  0.02
```

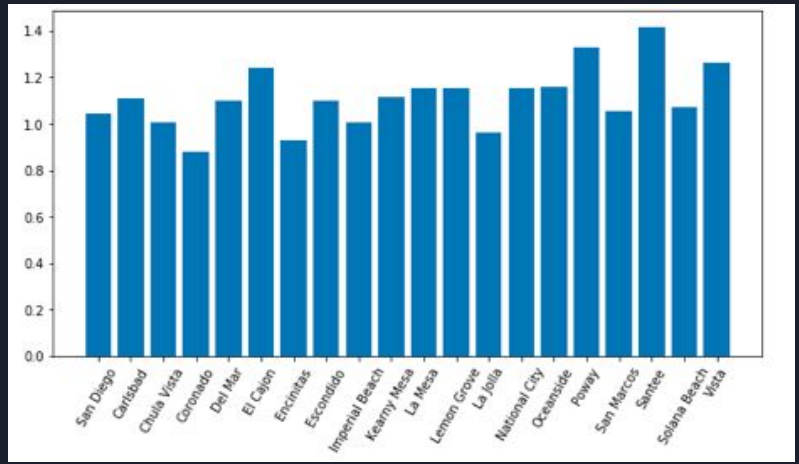
Aggregate frequencies

	Neighbourhood	ATM	Accessories Store	American Restaurant	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Vegetarian / Vegan Restaurant	Video Game Store	Video Store
0	San Diego	0.0	0.000000	0.040000	0.0	0.000000	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000
1	Carlsbad	0.0	0.000000	0.044444	0.0	0.000000	0.0	0.0	0.011111	0.0	...	0.0	0.000000	0.000000
2	Chula Vista	0.0	0.000000	0.022472	0.0	0.011236	0.0	0.0	0.000000	0.0	...	0.0	0.011236	0.011236
3	Coronado	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.010870	0.0	...	0.0	0.000000	0.000000
4	Del Mar	0.0	0.015385	0.076923	0.0	0.000000	0.0	0.0	0.000000	0.0	...	0.0	0.000000	0.000000

First 5 rows of dataframe containing frequencies of San Diego cities

Analysis - Similarity Comparison

- The similarity between the aggregate and the San Diego cities were calculated.
  - The difference between the frequencies were taken to determine the similarities
- From this comparison, it was found that **Coronado** was the city most similar to the aggregate.

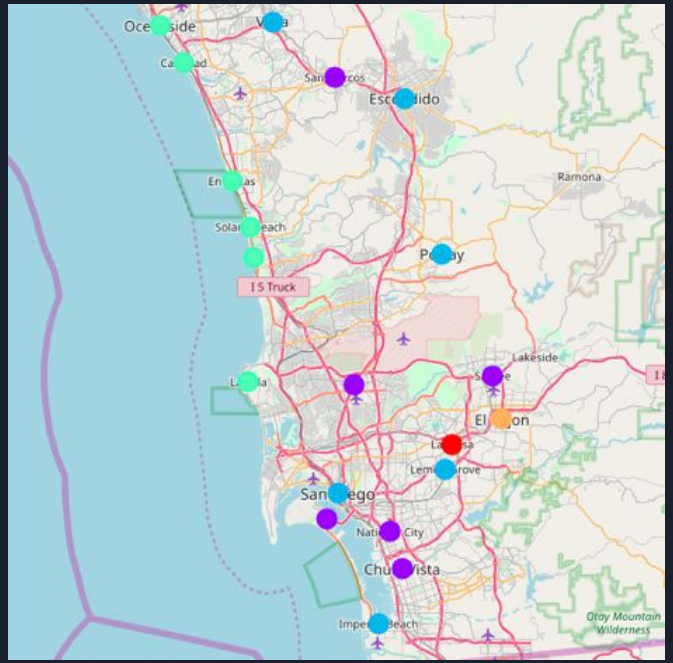


Bar chart of the calculated differences for each city in San Diego County

Analysis - Clustering

- After identifying Coronado as the most similar city, k-means clustering was performed to find other cities clustered with Coronado.
- This was done to identify a few more options for locations
- From the clustering, Chula Vista, Coronado, Kearny Mesa, National City, San Marcos, and Santee were identified as options.

Analysis - Clustering



	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Chula Vista	Mexican Restaurant	Grocery Store	Seafood Restaurant	Sandwich Place	Taco Place	Hotel	Pizza Place	Burger Joint	Pharmacy	Brewery
1	Coronado	Park	Seafood Restaurant	Hotel	Ice Cream Shop	Mexican Restaurant	Burger Joint	Pizza Place	Surf Spot	Italian Restaurant	Resort
2	Kearny Mesa	Japanese Restaurant	Sushi Restaurant	Bubble Tea Shop	Korean Restaurant	Mexican Restaurant	Sandwich Place	Vietnamese Restaurant	Chinese Restaurant	Noodle House	Ramen Restaurant
3	National City	Mexican Restaurant	Chinese Restaurant	Filipino Restaurant	Pizza Place	Park	Bakery	Sushi Restaurant	Auto Dealership	Tea Room	Pub
4	San Marcos	Mexican Restaurant	Pizza Place	Sushi Restaurant	Brewery	Baseball Field	Deli / Bodega	Gym / Fitness Center	Park	Diner	Hotel
5	Santee	Mexican Restaurant	Brewery	Breakfast Spot	Sushi Restaurant	Japanese Restaurant	Vietnamese Restaurant	Shoe Store	Deli / Bodega	Gymnastics Gym	American Restaurant

- The results of clustering can be seen here.
  - Different colored markers on the map indicating different clusters.
  - Dataframe showing the cluster containing Coronado and its similar cities.

Results, Discussion and Conclusion

- It can be seen that most of the cities in the cluster have restaurants as their more common venues.
- Especially apparent in Kearny Mesa.
  - All 10 most common venue categories are food or drink categories.
  - 8 out of 10 most common are Asian venues.
- In order to determine the ideal location, additional data would be required
  - Eg. Demographical data, ratings and popularities of existing venues, customer preference in the area, etc.
- Also would need to determine if more beneficial to open in more saturated area or one sparse of Asian restaurants.
- The study has provided 6 possible locations, all of which would be good locations nonetheless.