

## Introduction

San Diego, the city the author lives in, is a border city in the southwestern United States. Its beautiful seaside environment, pleasant climate, numerous high-tech companies and a good education system are attracting a large number of people to move in. For the person who would like to move in, finding the most suitable neighborhood according to his/her own preferences, like is there your favorite restaurant that best suits your taste, is there a grocery store nearby or is it near the beach and so on. If you live for a long time, you also need to consider the housing price level of the community.

Thus, the problem to be solved by this project is to take the city of San Diego as an example, giving a basic recommendation to the person who are preparing to move in San Diego: for different neighborhoods, what are the most common venues, what are the characteristics of this neighborhood, and what is the level of housing prices. Through this basic recommendation, people can get a general understanding of the community environment in San Diego, and then choose the most suitable place to live.

## Description of the data

We need data from reliable sources for further analysis. Focusing on this project, we will use following data:

1) Pulling the data of neighborhood of the city of San Diego out of HTML

<https://web.archive.org/web/20080509171047/http://www.sandiego.gov/neighborhoodmaps/>. The neighborhoods of the City of San Diego are divided into regions. These regions are based on the neighborhood policing and community services areas.

2) Foursquare Developers Access to venue data. Foursquare is a US tech company from New York focusing on location data. I will use foursquare data such as the venue name, location and category.

3) The house price comes from Zillow. We use Zillow home value index, which means the typical value for a geographic area, to represent house price in that neighborhood.

4) The latitude and longitude data of each neighborhood comes from Google search, due to recently Google started charging for their API.

## Methodology

In this section, I will describe the data exploration, analysis and how I used the data to yield the results.

1) Scrape and explore dataset

Firstly, I scraped the data of neighborhood of the city of San Diego from

<https://web.archive.org/web/20080509171047/http://www.sandiego.gov/neighborhoodmaps/>. For this, I used BeautifulSoup to scrape the data from HTML and the pandas to store

the data. The neighborhoods of the City of San Diego are divided into regions. Thus, the dataset has 2 columns, i.e. 'Region' and 'Neighborhood'. From the dataset, we can find there are 8 regions in the city of San Diego, i.e. Northern, Eastern, Southeastern, Northeastern, Central, Southern, Western and Mid-City. There are totally 121 neighborhoods in these 8 regions. For illustration purposes, I will focus on the Northern region. Other regions can be analyzed in exactly the same way as here.

## 2) Explore neighborhood in Northern region

In Northern region, there are 16 neighborhoods. It is necessary to get all of latitude and longitude of these neighborhoods for the further exploration and creating a map with neighborhoods. Considering we cannot use Google's API geocoder freely, and only 16 neighborhoods, I just search the locations in Google search engine. And as we said, we are also interested in the house price for each neighborhood. I collect the house price data from ZILLOW website. The dataset that expanded location (latitude & longitude) and house prices is shown below.

|    | Region   | Neighborhood         | Latitude | Longitude | Price (\$) |
|----|----------|----------------------|----------|-----------|------------|
| 0  | Northern | Bay Ho               | 32.8242  | -117.2176 | 878644     |
| 1  | Northern | Bay Park             | 32.7918  | -117.2000 | 993202     |
| 2  | Northern | Carmel Valley        | 32.9439  | -117.2308 | 1284041    |
| 3  | Northern | Clairemont Mesa East | 32.8124  | -117.1676 | 757674     |
| 4  | Northern | Clairemont Mesa West | 32.8072  | -117.1852 | 810498     |
| 5  | Northern | Del Mar Heights      | 32.9515  | -117.2529 | 1643284    |
| 6  | Northern | La Jolla             | 32.8289  | -117.2647 | 1874385    |
| 7  | Northern | La Jolla Village     | 32.8429  | -117.2320 | 784488     |
| 8  | Northern | Mission Beach        | 32.7733  | -117.2503 | 1287889    |
| 9  | Northern | Mission Bay Park     | 32.7778  | -117.2265 | 828656     |
| 10 | Northern | North City           | 32.9771  | -117.2162 | 1801601    |
| 11 | Northern | North Clairemont     | 32.8328  | -117.2000 | 790675     |
| 12 | Northern | Pacific Beach        | 32.8064  | -117.2382 | 1009741    |
| 13 | Northern | Torrey Pines         | 32.8969  | -117.2387 | 603162     |
| 14 | Northern | Torrey Preserve      | 32.9194  | -117.2461 | 1524588    |
| 15 | Northern | University City      | 32.8636  | -117.2121 | 784488     |

Table 1 The neighborhood of Northern region with location and price information

Using the folium package and dataset above, we can visualize Northern region with the neighborhoods in it.

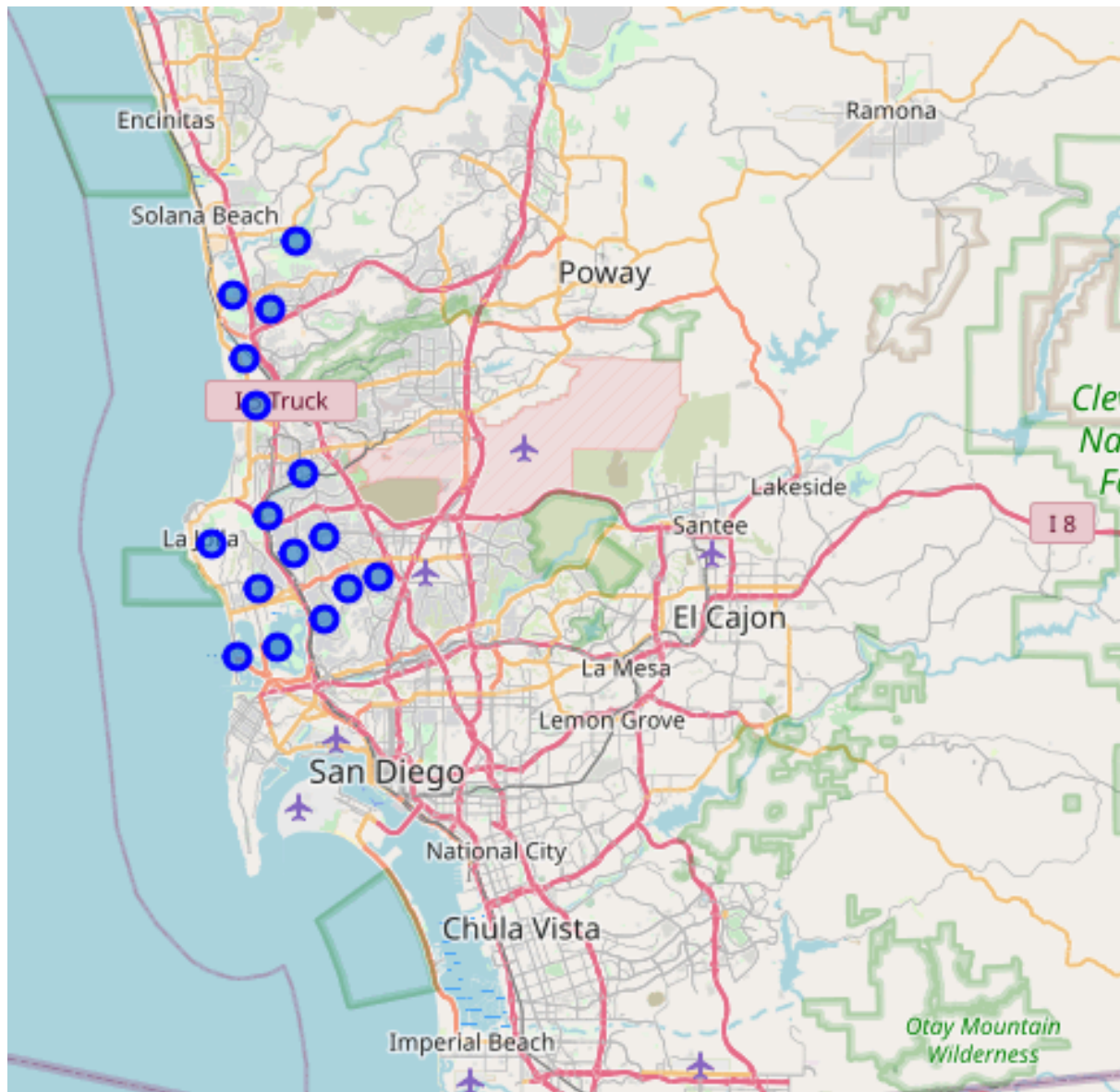


Fig. 1 The map of northern region with the neighborhoods.

In the next step, we collect venues for each neighborhood and see which venues are the most common. In this step, the venue data was collected from Foursquare via API. After collecting the data and organizing into a pandas dataframe, we have a table that looks like this (here only neighborhood of Bay Ho data is shown):

|     | name                      | categories           | lat       | lng         |
|-----|---------------------------|----------------------|-----------|-------------|
| 0   | San Diego Harley-Davidson | Motorcycle Shop      | 32.822666 | -117.226574 |
| 1   | El Rey Taco Shop          | Mexican Restaurant   | 32.823327 | -117.214070 |
| 2   | Costco Tire Center        | Automotive Shop      | 32.822858 | -117.224535 |
| 3   | Costco Food Court         | Food Court           | 32.821746 | -117.224616 |
| 4   | The Fruit Stand           | Farmers Market       | 32.816834 | -117.221380 |
| ... | ...                       | ...                  | ...       | ...         |
| 73  | SportRx                   | Optical Shop         | 32.814869 | -117.220595 |
| 74  | Frank's Starshine Market  | Convenience Store    | 32.835927 | -117.222804 |
| 75  | McDonald's                | Fast Food Restaurant | 32.833172 | -117.201923 |
| 76  | Rose Canyon               | Mountain             | 32.834580 | -117.233023 |
| 77  | Kate Sessions Canyon      | Trail                | 32.819331 | -117.238190 |

Table 2 Neighborhoods of Bay Ho

After getting all of venues including in 16 neighborhoods, we get 168 unique categories on these venues. Then the dataset is grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category. According to this, a new dataframe is created and display the top 10 venues for each neighborhood.

|    | Neighborhood         | 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  | Bay Ho               | Coffee Shop           | Mexican Restaurant    | Park                    | Grocery Store         | Beach                   | Pizza Place             | Seafood Restaurant       | Fast Food Restaurant    | Breakfast Spot        | Burger Joint            |
| 1  | Bay Park             | Park                  | Mexican Restaurant    | Pizza Place             | Beach                 | Café                    | Thai Restaurant         | Coffee Shop              | American Restaurant     | Seafood Restaurant    | Brewery                 |
| 2  | Carmel Valley        | Coffee Shop           | Beach                 | Seafood Restaurant      | Trail                 | Italian Restaurant      | American Restaurant     | Mediterranean Restaurant | New American Restaurant | Golf Course           | Restaurant              |
| 3  | Clairemont Mesa East | Japanese Restaurant   | Sushi Restaurant      | Grocery Store           | Korean Restaurant     | Café                    | Pizza Place             | Dessert Shop             | Mexican Restaurant      | Noodle House          | Fast Food Restaurant    |
| 4  | Clairemont Mesa West | Japanese Restaurant   | Grocery Store         | Sushi Restaurant        | Pizza Place           | Fast Food Restaurant    | Thai Restaurant         | Korean Restaurant        | Park                    | Coffee Shop           | Mexican Restaurant      |
| 5  | Del Mar Heights      | Coffee Shop           | Beach                 | Seafood Restaurant      | Mexican Restaurant    | American Restaurant     | Trail                   | Italian Restaurant       | Grocery Store           | Breakfast Spot        | Park                    |
| 6  | La Jolla             | Beach                 | Breakfast Spot        | Seafood Restaurant      | Coffee Shop           | Italian Restaurant      | Park                    | American Restaurant      | Mexican Restaurant      | Café                  | Surf Spot               |
| 7  | La Jolla Village     | Coffee Shop           | Seafood Restaurant    | Beach                   | Park                  | Café                    | New American Restaurant | Grocery Store            | Italian Restaurant      | Scenic Lookout        | Ice Cream Shop          |
| 8  | Mission Bay Park     | Park                  | Beach                 | Coffee Shop             | Bar                   | Brewery                 | Seafood Restaurant      | Grocery Store            | Fast Food Restaurant    | Breakfast Spot        | Theme Park              |
| 9  | Mission Beach        | Beach                 | Park                  | Bar                     | American Restaurant   | Brewery                 | Burger Joint            | Mexican Restaurant       | Coffee Shop             | Seafood Restaurant    | Café                    |
| 10 | North City           | Golf Course           | Coffee Shop           | Restaurant              | American Restaurant   | Grocery Store           | Italian Restaurant      | Seafood Restaurant       | Pizza Place             | Breakfast Spot        | New American Restaurant |
| 11 | North Clairemont     | Coffee Shop           | Grocery Store         | Mexican Restaurant      | Pizza Place           | Park                    | Fast Food Restaurant    | Japanese Restaurant      | Korean Restaurant       | Ice Cream Shop        | Sporting Goods Shop     |
| 12 | Pacific Beach        | Beach                 | Park                  | Seafood Restaurant      | Coffee Shop           | Breakfast Spot          | Mexican Restaurant      | Bar                      | Pizza Place             | Burger Joint          | Café                    |
| 13 | Torrey Pines         | Beach                 | Trail                 | New American Restaurant | Coffee Shop           | Furniture / Home Store  | Brewery                 | Burger Joint             | Cosmetics Shop          | Gym                   | Grocery Store           |
| 14 | Torrey Preserve      | Coffee Shop           | Beach                 | Trail                   | Mexican Restaurant    | Scenic Lookout          | American Restaurant     | Italian Restaurant       | New American Restaurant | Seafood Restaurant    | Hotel                   |
| 15 | University City      | Coffee Shop           | Sandwich Place        | Burger Joint            | Grocery Store         | New American Restaurant | Furniture / Home Store  | Park                     | Cosmetics Shop          | Gym                   | Clothing Store          |

Table 3 Display the top 10 venues for each neighborhood

### 3) Cluster neighborhoods

Now we run k-means to cluster the neighborhood into 5 clusters, and visualize the resulting clusters:



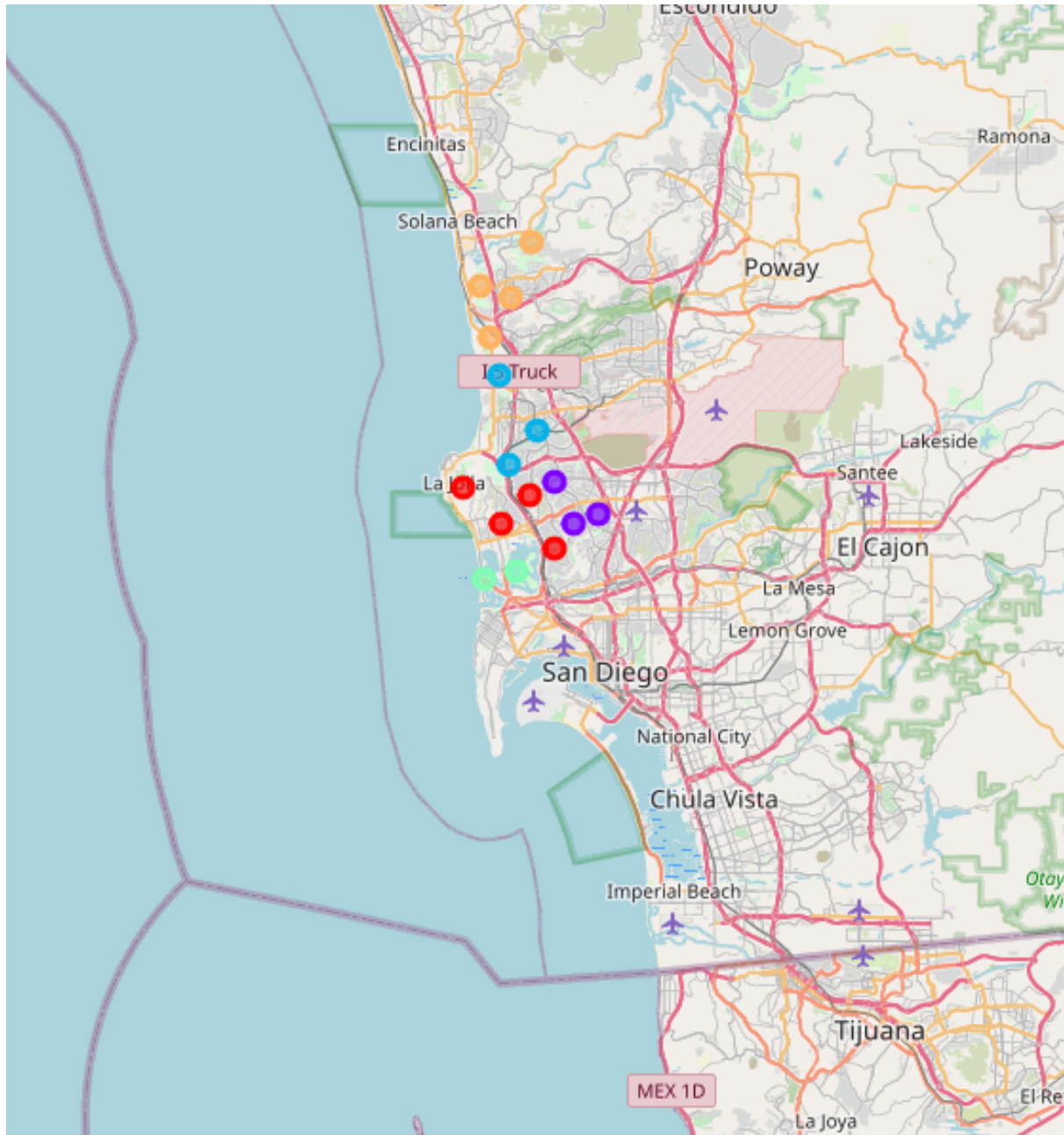


Fig.2 The map of northern region with cluster labels

## Results

By looking at the cluster data, we can get some basic ideas about living in the northern region of the city of San Diego.

The first cluster (red points) includes the neighborhoods of Bay Ho, Bay Park, La Jolla and Pacific Beach. This area is characterized by beaches, parks, Mexican restaurants and seafood restaurants. If you like living near the sea, and you like Mexican food or seafood, this will be your best living choice. The house price of this area is moderate level of the city, except La Jolla, which is very closed to sea and have the most excellent seascape in Southern California. The figure below shows the most common venues for the first cluster.

|    | Neighborhood  | Price (\$) | Cluster Labels | 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  | Bay Ho        | 878644     | 0              | Coffee Shop           | Mexican Restaurant    | Park                  | Grocery Store         | Beach                 | Pizza Place           | Seafood Restaurant    | Fast Food Restaurant  | Breakfast Spot        | Burger Joint           |
| 1  | Bay Park      | 993202     | 0              | Park                  | Mexican Restaurant    | Pizza Place           | Beach                 | Café                  | Thai Restaurant       | Coffee Shop           | American Restaurant   | Seafood Restaurant    | Brewery                |
| 6  | La Jolla      | 1874385    | 0              | Beach                 | Breakfast Spot        | Seafood Restaurant    | Coffee Shop           | Italian Restaurant    | Park                  | American Restaurant   | Mexican Restaurant    | Café                  | Surf Spot              |
| 12 | Pacific Beach | 1009741    | 0              | Beach                 | Park                  | Seafood Restaurant    | Coffee Shop           | Breakfast Spot        | Mexican Restaurant    | Bar                   | Pizza Place           | Burger Joint          | Café                   |

Table 4 The most common venues for the first cluster

The second cluster (purple points) includes the neighborhoods of Clairemont Mesa East, Clairemont Mesa West and North Clairemont. From the naming of the neighborhoods, we can also see their similarities. This area is characterized by Japanese/Korean food (and other Asian food), grocery store and pizza place. The house price is lower than other areas of cluster. This is a typical living area, and you can find many Asian food here. Life is also very convenient. Since it is not close to the sea, the house prices are also more friendly. The figure below shows the most common venues for the second cluster. And I also show all of the venues which two of neighborhoods have, and you can find more similarities of them.

|    | Neighborhood         | Price (\$) | Cluster Labels | 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 |
|----|----------------------|------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 3  | Clairemont Mesa East | 757674     | 1              | Japanese Restaurant   | Sushi Restaurant      | Grocery Store         | Korean Restaurant     | Café                  | Pizza Place           | Dessert Shop          | Mexican Restaurant    | Noodle House          | Fast Food Restaurant   |
| 4  | Clairemont Mesa West | 810498     | 1              | Japanese Restaurant   | Grocery Store         | Sushi Restaurant      | Pizza Place           | Fast Food Restaurant  | Thai Restaurant       | Korean Restaurant     | Park                  | Coffee Shop           | Mexican Restaurant     |
| 11 | North Clairemont     | 790675     | 1              | Coffee Shop           | Grocery Store         | Mexican Restaurant    | Pizza Place           | Park                  | Fast Food Restaurant  | Japanese Restaurant   | Korean Restaurant     | Ice Cream Shop        | Sporting Goods Shop    |

Table 5 The most common venues for the second cluster

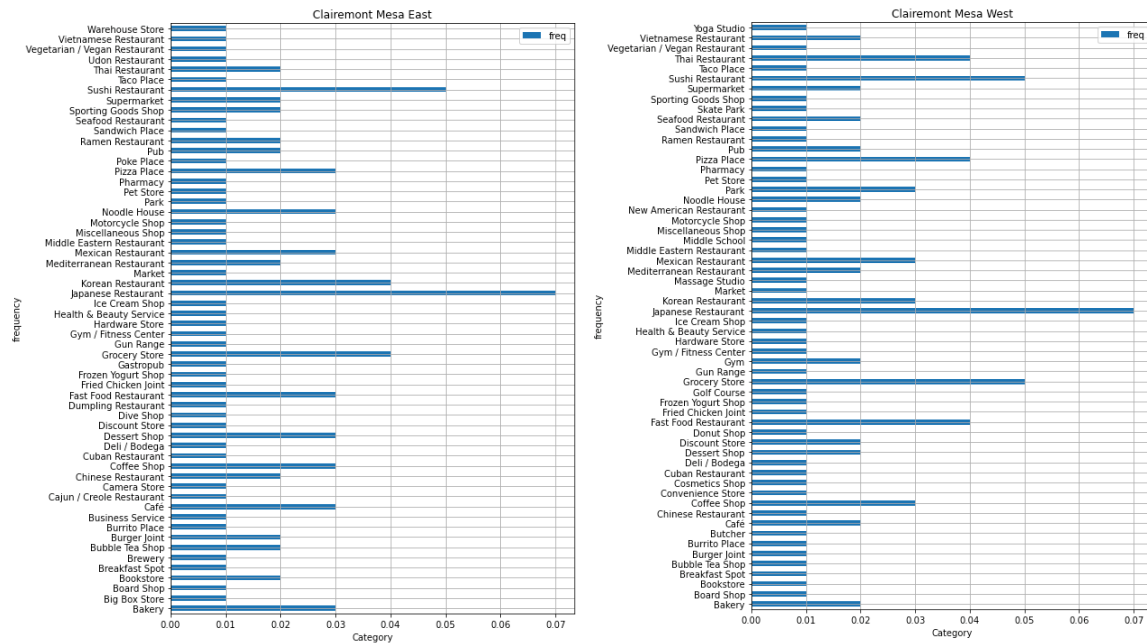


Fig.3 All of the venues which two of neighborhoods have

The third cluster (blue points) includes the neighborhoods of La Jolla Village, Torrey Pines and University City. This area is characterized by coffee shop, New American restaurant, grocery store and different kinds of shops. This area and the area of last cluster have the low level of house price in the northern part of the city of San Diego. Further research for this area, you will find there are a large number of technology companies and university campuses here, so these venues are also very suitable for the needs of the people who work and live here. If you are an employee of a nearby company or a university student, you can consider living in this area. The lower price also means that renting a house in this area is also very cost-effective. Of course, the most important thing is the convenience of life, which meets the needs of fast-paced life. The figure below shows the most common venues for the third cluster. I also show all of venues of neighborhood of University City, you can find how this neighborhood meets the requirement of modern fast-paced life.

|    | Neighborhood     | Price (\$) | Cluster Labels | 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 |
|----|------------------|------------|----------------|-----------------------|-----------------------|-------------------------|-----------------------|-------------------------|-------------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 7  | La Jolla Village | 784488     | 2              | Coffee Shop           | Seafood Restaurant    | Beach                   | Park                  | Café                    | New American Restaurant | Grocery Store         | Italian Restaurant    | Scenic Lookout        | Ice Cream Shop         |
| 13 | Torrey Pines     | 603162     | 2              | Beach                 | Trail                 | New American Restaurant | Coffee Shop           | Furniture / Home Store  | Brewery                 | Burger Joint          | Cosmetics Shop        | Gym                   | Grocery Store          |
| 15 | University City  | 784488     | 2              | Coffee Shop           | Sandwich Place        | Burger Joint            | Grocery Store         | New American Restaurant | Furniture / Home Store  | Park                  | Cosmetics Shop        | Gym                   | Clothing Store         |

Table 6 The most common venues for the third cluster



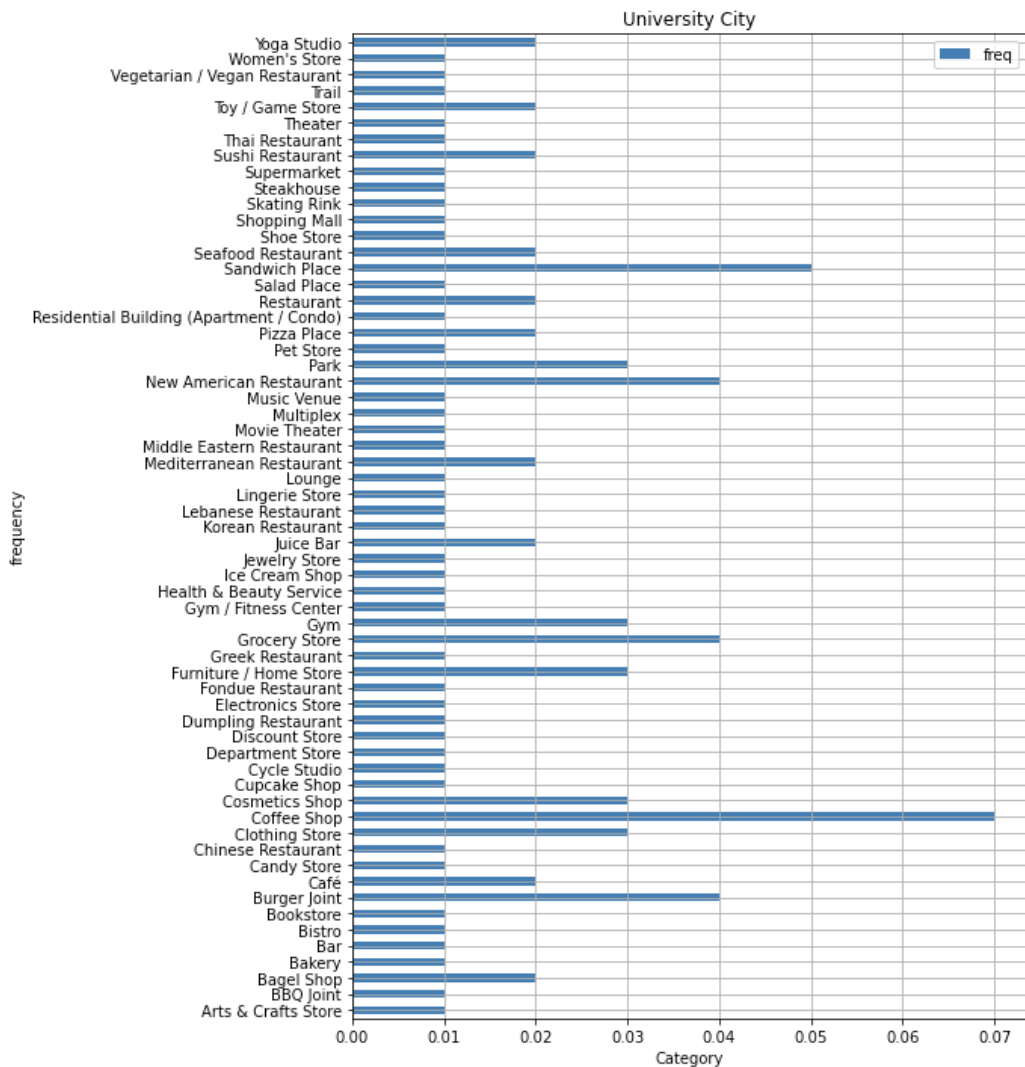


Fig.4 All of the venues in University City

The fourth cluster (green points) includes the neighborhoods of Mission Beach and Mission Bay Park. This area is characterized by beach, park, brewery, bar and seafood restaurant. This area is similar to the first area (the area in cluster 1) in that it contains lots of venues of beach, park and seafood restaurants. But the difference is that there are many bars and breweries here. The house price belongs to the moderate-high level. If you like living near the sea and drink with friends, this area will be a good choice. The figure below shows the most common venues for the fourth cluster.

|   | Neighborhood     | Price (\$) | Cluster Labels | 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 |
|---|------------------|------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 8 | Mission Beach    | 1287889    | 3              | Beach                 | Park                  | Bar                   | American Restaurant   | Brewery               | Burger Joint          | Mexican Restaurant    | Coffee Shop           | Seafood Restaurant    | Café                   |
| 9 | Mission Bay Park | 828656     | 3              | Park                  | Beach                 | Coffee Shop           | Bar                   | Brewery               | Seafood Restaurant    | Grocery Store         | Fast Food Restaurant  | Breakfast Spot        | Theme Park             |

Table 6 The most common venues for the fourth cluster

The fifth cluster (orange points) includes the neighborhoods of Carmel Valley, Del Mar Heights, North City and Torrey Preserve. The area is characterized by coffee shop, beach, all

kinds of restaurant, trail and golf course. You can easily see that this is a densely populated area, and the catering industry is very prosperous. House prices are the highest level of Northern region and even in the whole city of San Diego. Most of the people who live here are in a better financial situation, so you can find some high-end facilities here like Golf. The figure below shows the most common venues for the fifth cluster.

|    | Neighborhood    | Price (\$) | Cluster Labels | 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  |
|----|-----------------|------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|-------------------------|-----------------------|-------------------------|
| 2  | Carmel Valley   | 1284041    | 4              | Coffee Shop           | Beach                 | Seafood Restaurant    | Trail                 | Italian Restaurant    | American Restaurant   | Mediterranean Restaurant | New American Restaurant | Golf Course           | Restaurant              |
| 5  | Del Mar Heights | 1643284    | 4              | Coffee Shop           | Beach                 | Seafood Restaurant    | Mexican Restaurant    | American Restaurant   | Trail                 | Italian Restaurant       | Grocery Store           | Breakfast Spot        | Park                    |
| 10 | North City      | 1801601    | 4              | Golf Course           | Coffee Shop           | Restaurant            | American Restaurant   | Grocery Store         | Italian Restaurant    | Seafood Restaurant       | Pizza Place             | Breakfast Spot        | New American Restaurant |
| 14 | Torrey Preserve | 1524588    | 4              | Coffee Shop           | Beach                 | Trail                 | Mexican Restaurant    | Scenic Lookout        | American Restaurant   | Italian Restaurant       | New American Restaurant | Seafood Restaurant    | Hotel                   |

Table 7 The most common venues for the fifth cluster

## Discussion

Now that we have a basic understanding of neighborhoods in northern region of the city of San Diego. Based on this, we can provide some living suggestions for people who are new to this city (northern). For example:

- 1) If you like living near the sea, and you like Mexican food or seafood, the neighborhoods in Cluster 1 will be your best living choice. You also can find some house with moderate prices, except in La Jolla, which has the most wonderful seascape in Southern California.
- 2) If you like Asian food, especially Japanese/Korean food, the neighborhoods of Cluster 2 will be your dream place. Life will be very convenient and the house price is also low comparing other areas.
- 3) If the company you will join places in the neighborhoods of Cluster 3, you could choose the living place here directly. The venues in this area meet all of requirement of modern fast-paced life. You also find the house price and renting price are also friendly.
- 4) If you like living near the sea and drink with friends, this area of Cluster 4 will be a good choice. It is quite similar with the area of Cluster 1 but if you are not a big fan of Mexican food and loves drink with friend or by yourself, believe me, this is your best choice.
- 5) If you have good financial situation, choose the area of Cluster 5. This area contains all the characteristics of the above area, except for the most expensive housing price.

## Conclude

This report discusses the process of coming up with an answer for the problem we raised at the beginning: if you plan to move to San Diego (Northern region for example) to live, which area is best for you to live in. The data processing and analysis used mainly come from python, pandas, BeautifulSoup, folium and the k-means cluster method of Machine Learning. The method of thinking about similar problems can be applied to any cities or areas.