The Battle of Neighborhoods:

Finding Neighborhoods to Live in New Cities

New York vs San Diego

Mona(MoonHee) Chang

19th May, 2020

Ⅰ. Abstract

My imaginary client is **moving from New York, NY to San Diego, CA for a job**. However, she's only lived in New York all her life and is having a hard time to decide which neighborhood in San Diego she should live in. She is interested in areas that meets the following criteria:

- Similar to her current neighborhood, Chelsea, Manhattan(**venues and amenities as in my current residence**)
- Have her **preferred venue** types
- Have at least 2bed+ houses(**preferably 3bed+**) that are affordable, her maximum budget is **1.5m**
- **Safe** enough for her to be able to walk around alone whenever

Ⅱ. Introduction

i. Background

According to American Moving Statistics, 32 million people in the US moved in 2018, that is 10.1% of the population of 319 million US residents. The reasons for moving cities varied from Wanting new or better home/apartment to new job/job transfer/other job related reasons. Also, Rober Half's survey shows that 62% of workers said they would relocate for a job.

When relocating, some people struggle to figure out which area they should look into, mainly because a lot of times they don't know much about the new city they're moving to. The basis of this project is to help people, who are relocating and don't know much about the city they're moving to, easily compare cities and recommend possible locations/neighborhoods that they might be interested in. It does this by exploring venue-based city-to-city similarity measures, clustering neighborhoods within a city, and looking at housing requirements and prices as well as crime rates.

In this project, I am going to use New York, NY and San Diego, CA as my two cities to demonstrate the methodology, which can then be used for any cities that people want to compare.

ii. Problem

My client does not know San Diego well, since she's lived in NY only all her life. She has some ideas of what she wants in her new neighbor, but cannot figure out which neighbors in San Diego would meet her preference. How does she choose her new neighborhood?

iii. Target Audience

Obviously my client, who is relocating to San Diego and does not know much about the city nor the neighborhoods, is the target audience of this project. However, others who struggle to figure out which area they should look into, when relocating, may also be interested.

Ⅲ. Data Acquisition and Cleaning

i. Data Requirements

- List of Manhattan neighborhoods with latitude and longitude
- List of San Diego neighborhoods with latitude and longitude
- Geodata for current residence in Manhattan, NY with venues established using Foursquare
- Geodata for Sandiego, CA with venues established using Foursquare
- Each neighborhood's housing prices in San Diego with location, number of beds and complemented with geodata via Nominatim
- Crime rate in San Diego by neighborhoods
- Crime types by Neighborhoods in San Diego

ii. Data Sources

US Moving Statistics for 2019: This site provides the compiled data from the United States Census Bureau. We can see how many people in the US move each year, top 10 states people are moving to and moving from, the reasons for moving, etc. - 32 million people in the US moved in 2018, that is 10.1% of the population of 319 million US residents.

Robert Half survey: The online surveys developed by Robert Half and conducted by Independent research firms include responses from more than 2,800 US workers 18 years of age or older and employed in office environments and more than 2,800 senior managers at companies with 20 or more employees in 28 major US cities - 62% of workers said they would relocate for a job.

NYU 2014 New York City Neighborhood Names: GeoJSON file that was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." This includes information on boroughs, neighborhoods, and location(latitudes & longitudes) of each neighborhood.

City of San Diego Neighborhoods shapes: GeoJSON file that provides an overview of the neighborhoods in Sandiego. This includes information on neighborhoods and location(latitudes & longitudes) of each neighborhood.

Foursquare API: This collects information on venues in the selected neighborhoods.

Zillow Market Reports: This provides an excel file for the selected city's housing market reports. It includes information on neighborhood location, Zillow home value index per number of beds for each neighborhood, average zillow home value index for each neighborhood, Median value per sq.ft.($) for each neighborhood.

Sandiego.gov Crime Statistics and Maps: San Diego provides crimes by neighborhood, and I'm using 'January-August 2019 data'. This includes information on the type of crimes and neighborhood that the crimes occurred in.

iii. Data Cleaning

Group 1: Neighborhoods in New York & San Diego with coordinates from GeoJson data

The New York data was pulled from the NYU GeoJson file, using the json library, to create a dataframe (Table1). Here, the data frame was filtered into columns based on borough, neighborhood, latitude and longitude (Table1). This original data frame consisted of 5 boroughs and 306 neighborhoods, but my client was only interested in venues and amenities in her current residence, Chelsea, Manhattan. Hence, I extracted the Manhattan borough data from the original New York dataframe to create a new dataframe, which consisted of 40 neighborhoods in Manhattan only (Table2). Also, the 'Borough' column was dropped in the new data frame, because it only consisted of Manhattan borough neighborhoods and was not necessary (Table2).

Table 1.
New York dataframe from NYU GeoJson data (first 5 rows)

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

```
The dataframe has 5 boroughs and 306 neighborhoods.
```

Table 2.
Manhattan dataframe after cleaning (first 4 rows)

	Neighborhood	Latitude	Longitude
0	Marble Hill	40.876551	-73.910660
1	Chinatown	40.715618	-73.994279
2	Washington Heights	40.851903	-73.936900
3	Inwood	40.867684	-73.921210
4	Hamilton Heights	40.823604	-73.949688

```
The dataframe has 40 neighborhoods.
```

The San Diego data was pulled from the San Diego GeoJson file, also using the json library. Then it was filtered to produce a data frame with information on the city's neighborhoods, latitudes and longitudes, which consisted of 124 neighborhoods in total (Table3).

Table 3.
San Diego dataframe (first 5 rows)

	Neighborhood	Latitude	Longitude
0	Qualcomm	32.783322	-117.119738
1	Egger Highlands	32.587312	-117.100488
2	Old Town	32.754602	-117.195528
3	Morena	32.769888	-117.193218
4	Midtown	32.739864	-117.175020

```
The dataframe has 124 neighborhoods
```

Group 2: Exploring neighborhoods in New York & San Diego by venues

I got venue information of each neighborhood using the Foursquare API, for both of the cities (Table4, Table5).

Table 4.
Venue information for manhattan (first 5 rows)

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Table 5.
Venue information for San Diego (first 5 rows)

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Qualcomm	32.783322	-117.119738	SDCCU Stadium	32.783035	-117.119493	Football Stadium
Qualcomm	32.783322	-117.119738	Qualcomm Stadium Swap Meet	32.784975	-117.121750	Flea Market
Qualcomm	32.783322	-117.119738	E3 All Stars Tailgate	32.784613	-117.122946	Football Stadium
Egger Highlands	32.587312	-117.100488	Jalisco Cafe	32.583549	-117.097720	Mexican Restaurant
Egger Highlands	32.587312	-117.100488	Rally's	32.583572	-117.099726	Fast Food Restaurant

However, to further explore each neighborhood with venues, I utilized the one hot encoding to present all categorical values for the venue category section (Table6, Table7).

Table 6.
One hot encoded for manhattan (first 5 rows)

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcad
0	Marble Hill	0	0	0	0	0	0	
1	Marble Hill	0	0	0	0	0	0	
2	Marble Hill	0	0	0	0	0	0	
3	Marble Hill	0	0	0	0	0	0	
4	Marble Hill	0	0	0	0	0	0	

```
Manhattan: (3017, 330)
```

Table 7.
One hot encoded for San Diego (first 5 rows)

	Neighborhood	ATM	Accessories Store	American Restaurant	Amphitheater	Antique Shop	Argentinian Restaurant	Art Gallery
0	Qualcomm	0	0	0	0	0	0	0
1	Qualcomm	0	0	0	0	0	0	0
2	Qualcomm	0	0	0	0	0	0	0
3	Egger Highlands	0	0	0	0	0	0	0
4	Egger Highlands	0	0	0	0	0	0	0

```
San Diego: (1965, 276)
```

Then I grouped the data frame by neighborhood and calculated the mean of the frequency of occurrence of each venue category (Table8, Table9).

Table 8.
Grouped & occurence freq. - manhattan (first 5 rows)

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade
0	Battery Park City	0.0	0.0	0.0	0.000000	0.000000	0.0	0.0
1	Carnegie Hill	0.0	0.0	0.0	0.000000	0.011905	0.0	0.0
2	Central Harlem	0.0	0.0	0.0	0.065217	0.043478	0.0	0.0
3	Chelsea	0.0	0.0	0.0	0.000000	0.030000	0.0	0.0
4	Chinatown	0.0	0.0	0.0	0.000000	0.030000	0.0	0.0

```
Manhattan data shape: (40, 330)
```

Table 9.
Grouped & occurrence freq. - San Diego (first 5 rows)

	Neighborhood	ATM	Accessories Store	American Restaurant	Amphitheater	Antique Shop	Argentinian Restaurant	Ar Gallery
0	Adams North	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0
1	Allied Gardens	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0
2	Alta Vista	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0
3	Azalea/Hollywood Park	0.0	0.0	0.000000	0.000000	0.0	0.0	0.0
4	Balboa Park	0.0	0.0	0.022727	0.022727	0.0	0.0	0.0

```
San Diego data shape: (116, 276)
```

Group 3: Adding coordinates of each San Diego neighborhoods to clustered dataframe

Neighborhoods in San Diego were clustered based on venue categories each neighborhood has. But for visualization purposes, each neighborhood's longitudes and latitudes were added and matched based on the name of the neighborhoods, from the original San Diego dataframe that was imported from the San Diego GeoJSON file (Table 10).

Table 10.
San Diego neighborhood clustering with coordinates (first 5 rows)

Neighborhood	Latitude	Longitude	Cluster labels
Qualcomm	32.783322	-117.119738	21
Egger Highlands	32.587312	-117.100488	8
Old Town	32.754602	-117.195528	59
Morena	32.769888	-117.193218	19
Midtown	32.739864	-117.175020	8

Group 4: Housing prices in San Diego by neighborhoods

The housing prices data from the Zillow Market Report had a lot of information that I did not need for this project and were distributed into 4 different sheets. So I firstly imported the data of my choice - neighborhood name, zillow home value index of all homes, median value per sq.ft($) of all homes, zillow home value index of 2 bedroom houses, zillow home value index of 3 bedroom houses, and zillow home value index of 4 bedroom houses - from all 4 sheets using pandas ExcelFile and read_excel functions. Then the imported data were merged into one dataframe, using pandas concat function and columns were renamed as well (Table11).

Table 11.
San Diego home prices after data cleaning (first 5 rows)

Neighborhood	Zillow Home Value Index_Allhomes	Median Value per sq.ft($)_Allhomes	Zillow Home Value Index_2bed	Zillow Home Value Index_3bed	Zillow Home Value Index_4bed
Adams North	691900	648	628700	766300	997200
Allied Gardens	659300	505	455400	658800	711900
Alta Vista	509300	342	NaN	497600	516300
Azalea-Hollywood Park	472500	756	445900	512100	578400
Balboa Park	922200	613	783700	948200	1136300

In order to visualize the data frame, Latitudes and Longitudes were matched to each neighborhood name through merging/joining San Diego original data (Table2) and the housing price data frame (Table11) to create a new data frame that included coordinates of each neighborhood in San Diego along with housing prices (Table 12).

Table 12.
San Diego home prices with coordinates of each neighborhood (first 5rows)

Neighborhood	Latitude	Longitude	Zillow Home Value Index_Allhomes	Median Value per sq.ft($)_Allhomes	Zillow Home Value Index_2bed	Zillow Home Value Index_3bed	Zillow Home Value Index_4bed
Egger Highlands	32.587312	-117.100488	545700	1638	438600.0	529100.0	570800.0
Old Town	32.754602	-117.195528	734200	719	520900.0	1091700.0	1703700.0
Morena	32.769888	-117.193218	517500	408	458700.0	580300.0	807700.0
Midtown	32.739864	-117.175020	865700	624	678900.0	1012000.0	1386400.0
Lincoln Park	32.700617	-117.090095	481500	362	389300.0	470500.0	525800.0

Group 5: Crime Rate in San Diego by neighborhoods

Firstly, the crime data from San Diego.gov was in pdf format. I was going to use Camelot to read and import the file into my Jupyter Notebook, but the side did not allow for anyone to read the file directly from its link/website - http error 403: Forbidden. So, I converted the pdf file to excel online, using ZAMZAR, then uploaded it to my GitHub repository. The original pdf file had 4 pages, hence when it was converted to an excel file, it created 4 different sheets. I used pandas ExcelFile and read_excel functions to read and import the 4 sheets into my Notebook, where I skipped the first 6 rows of each sheet in the file. Here, each of the data's index had to be reset to fix the error that occured due to the difference in the shape of passed values (132, 68) and indices implied (131, 68). Then I merged all the selected data from 4 sheets into one dataframe using the append function and reset the new data frame's index once again (Table13).

Table 13.
San Diego crime data original (first 5 rows)

	Neighborhood	MURDER	RAPE	RAPE.1	ARMED	ST/	ASSAULT	CRM TL	RESID	NRESID	TOTAL	$400+	Unnamed: 12	TOTAL.1	THEFT	CI
0	NaN	NaN	NaN	(HISTORICAL)	NaN	ARM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	*************	NaN	NaN	NaN	NaN	NaN	NaN	NaN	****
2	Agency = SAN DIEGO	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	ADAMS NORTH	0.0	1.0	1	1.0	6	4	12	2	1	3	19	24	43	9	
4	ALLIED GARDENS	0.0	3.0	2	0.0	0	8	11	19	6	25	32	37	69	18	

As seen above, the original San Diego crime data frame had a lot of NaN values and rows that were not needed - e.g. Agency=SAN DIEGO(row3), UNKNOWN, SUBTOTAL, GRAND TOTALS, etc. Hence, I removed them all using the drop function (Table14). Also, some columns could had been combined - e.g. RAPE & RAPE.1, ARMED(armed robbery) & ST/(strong arm robbery). In order to combine the columns, all the columns' values(dtype) except for the 'Neighborhood' column were converted to float. Then the columns of the dataframe, including the ones combined, were renamed as it was hard to understand column names without an explanation (Table14).

Table 14.
San Diego Crime data cleaned (first 5rows)

	Neighborhood	MURDER	RAPE	ROBBERY	AGGRAVATED ASSAULT	TOTAL VIOLENT CRIME	RESIDENTIAL BURGLARY	NON-RESIDENTIAL BURGLARY	TOTAL BURGLARY	THEFT $400+	THEFT<$400	TOTAL THEFTS
0	ADAMS NORTH	0.0	2.0	7.0	4.0	12.0	2.0	1.0	3.0	19.0	24.0	43.0
1	ALLIED GARDENS	0.0	5.0	0.0	8.0	11.0	19.0	6.0	25.0	32.0	37.0	69.0
2	ALTA VISTA	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	4.0	5.0
3	AZALEA/HOLLYWOOD PARK	0.0	2.0	1.0	10.0	12.0	3.0	0.0	3.0	8.0	7.0	15.0
4	BALBOA PARK	0.0	5.0	7.0	10.0	21.0	1.0	6.0	7.0	69.0	23.0	92.0

For my project, I only needed Total number of crimes for each crime category, so I created a new dataframe with only neighborhood names, total violent crime(the sum of murders, rapes, robberies, and aggravated assault), total burglary(the sum of residential and non-residential burglaries), total thefts(the sum of thefts >= $400 and < $400), total property crime(the sum of total burglary, total thefts, and motor vehicle thefts), crime index total(the sum of total violent crime and total property crime) (Table15). Also, I changed the format of the neighborhood names, where the first letter of each word was capitalized (Table15), using the str.title function.

Table 15.
San Diego Crime data selected columns (first 5rows)

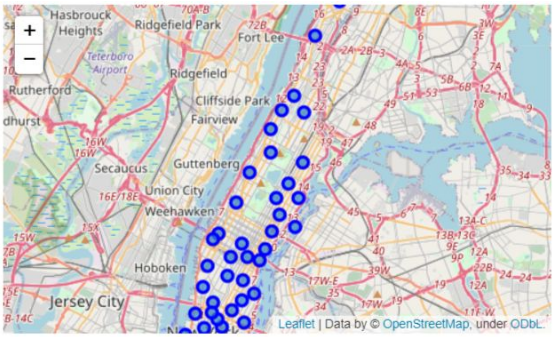
	Neighborhood	TOTAL VIOLENT CRIME	TOTAL BURGLARY	TOTAL THEFTS	TOTAL PROPERTY CRIME	CRIME INDEX TOTAL
0	Adams North	12.0	3.0	43.0	55.0	67.0
1	Allied Gardens	11.0	25.0	69.0	112.0	123.0
2	Alta Vista	0.0	1.0	5.0	7.0	7.0
3	Azalea/Hollywood Park	12.0	3.0	15.0	25.0	37.0
4	Balboa Park	21.0	7.0	92.0	110.0	131.0

Ⅳ. Methodology

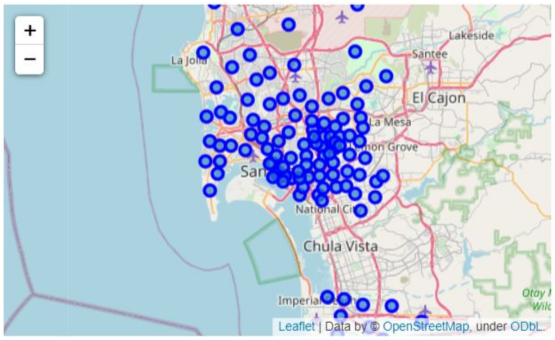
i. City overview - Manhattan & San Diego

I obtained the shape of the Manhattan and San Diego dataframe (Table2, Table3). To show the difference between two cities visually, I used Nomintim to obtain each city's latitude and longitude for folium mapping, then created markers to show every neighborhood in Manhattan and San Diego (Map1, Map2).

Map 1.
Manhattan with neighborhoods



Map 2.
San Diego with neighborhoods



ii. Venue overview - Manhattan & San Diego

I explored neighborhoods by utilizing the Foursquare API:

1. Defined my Foursquare API credentials - ID, Password, Version
2. Created a function to repeat the same process to all neighborhoods in Manhattan and San Diego
3. Created the API request URL
4. Made the GET request
5. Returned only relevant information for each nearby venue (Table4, Table5)
6. Created a dataframe that showed all the unique venue categories for each neighborhood by using One Hot encoding and the groupby() function (Table6, Table7)
7. Got the occurrence frequency of each venue category in every neighborhood in Manhattan and San Diego (Table8, Table9)

iii. Chelsea, Manhattan - Analyze Chelsea’s venues

My client is currently living in Chelsea, Manhattan and loves the neighborhood. One of her criteria was to find neighborhoods in San Diego that have similar venues and amenities as Chelsea, Manhattan does:

Extracted Chelsea data from the venue occurrence frequency dataframe og Manhattan (Table8, Table9)

1. What kind of venues does Chelsea have?

   Deleted all the venue columns that have ‘0’ values, since this means Chelsea does not have those venues (Table16).

Table 16.
Chelsea, Manhattan venue data

	American Restaurant	Art Gallery	Bagel Shop	Bakery	Bar	Beer Bar	Bookstore	Boutique	Café	Cheese Shop	Chinese Restaurant	Clothing Store	Club House	Coffee Shop	College Theater	Cosmeti Sh
Neighborhood																
Chelsea	3.0	15.0	1.0	2.0	1.0	1.0	1.0	2.0	4.0	1.0	1.0	1.0	1.0	7.0	1.0	

2. Visualize

   Converted venue columns’ values to float, in order to multiply the values by 100 (Figure1) and make it big enough for visualization (Figure1, Table17) .

Figure 1.
Graph - Chelsea, Manhattan venue data

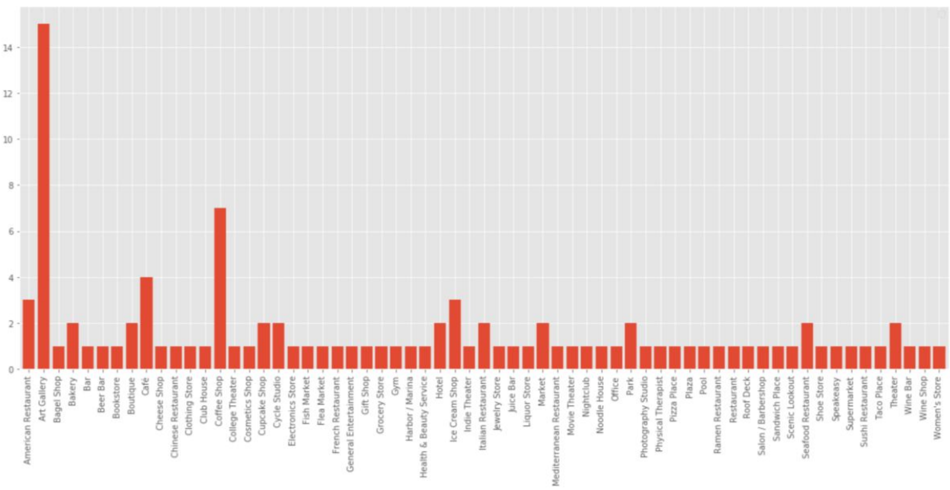


Table 17.
Chelsea’s top 10 common venues

	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
3	Chelsea	Art Gallery	Coffee Shop	Café	Ice Cream Shop	American Restaurant	Cycle Studio	Seafood Restaurant	Market	Theater	Bakery

iv. Find potential neighborhoods that fits my client's criteria

1. <u>Measure similarities between Chelsea, Manhattan and neighborhoods in San Diego</u>
    a. Based on the venues that Chelsea has

    Measured similarity score between Chelsea and San Diego neighborhoods (Table18) by:

    i. subtracting Chelsea venue occurrence frequency data frame values from San Diego venue occurrence frequency data frame values. This took into account that some columns/venue categories in the San Diego data frame do not exist in the Chelsea data frame. These columns were still included, only the values were 'NaN'.
    ii. Squaring all the values to unify positive and negative values
    iii. Square-rooting everything
    iv. Sorting ascending order

Table 18.
10 San Diego cities that are most similar to Chelsea, Manhattan

	Neighborhood	Similarity score
36	Gaslamp	0.183030
70	North Park	0.193883
23	Core-Columbia	0.196038
1	Allied Gardens	0.196469
104	Sunset Cliffs	0.196469
85	Rancho Encantada	0.196469
73	Ocean Crest	0.196469
42	Horton Plaza	0.203715
52	Little Italy	0.210613
41	Hillcrest	0.214522

2. <u>Desired venues in the new neighborhood in San Diego</u>
    a. Venues in Chelsea, Manhattan

    Displayed all the venue columns with non '0' or 'NaN' values in the Chelsea venue occurrence frequency data frame (Table16)

    'American Restaurant', 'Art Gallery', 'Bagel Shop', 'Bakery', 'Bar', 'Beer Bar', 'Bookstore', 'Boutique', 'Café', 'Cheese Shop', 'Chinese Restaurant', 'Clothing Store', 'Club House', 'Coffee Shop', 'College Theater', 'Cosmetics Shop', 'Cupcake Shop', 'Cycle Studio', 'Electronics Store', 'Fish Market', 'Flea Market', 'French Restaurant', 'General Entertainment', 'Gift Shop', 'Grocery Store', 'Gym', 'Harbor / Marina', 'Health & Beauty Service', 'Hotel', 'Ice Cream Shop', 'Indie Theater', 'Italian Restaurant', 'Jewelry Store', 'Juice Bar', 'Liquor Store', 'Market', 'Mediterranean Restaurant', 'Movie Theater', 'Nightclub', 'Noodle House', 'Office', 'Park', 'Photography Studio', 'Physical Therapist', 'Pizza Place', 'Plaza', 'Pool', 'Ramen Restaurant', 'Restaurant', 'Roof Deck', 'Salon / Barbershop', 'Sandwich Place', 'Scenic Lookout', 'Seafood Restaurant', 'Shoe Store', 'Speakeasy', 'Supermarket', 'Sushi Restaurant', 'Taco Place', 'Theater', 'Wine Bar', 'Wine Shop', "Women's Store"

    b. What does Chelsea not have that my client would like to have in her new neighborhood?

Displayed all the venue columns with ‘0’ values.

'Accessories Store', 'Adult Boutique', 'Afghan Restaurant', 'African Restaurant', 'Antique Shop', 'Arcade', 'Arepa Restaurant', 'Argentinian Restaurant', 'Art Museum', 'Arts & Crafts Store', 'Asian Restaurant', 'Athletics & Sports', 'Auditorium', 'Australian Restaurant', 'Austrian Restaurant', 'BBQ Joint', 'Baby Store', 'Bank', 'Baseball Field', 'Basketball Court', 'Basketball Stadium', 'Beer Garden', 'Beer Store', 'Bike Rental / Bike Share', 'Bike Shop', 'Bike Trail', 'Bistro', 'Board Shop', 'Boat or Ferry', 'Boxing Gym', 'Brazilian Restaurant', 'Breakfast Spot', 'Bridal Shop', 'Bridge', 'Bubble Tea Shop', 'Building', 'Burger Joint', 'Burrito Place', 'Bus Line', 'Bus Station', 'Butcher', 'Cafeteria', 'Cajun / Creole Restaurant', 'Cambodian Restaurant', 'Camera Store', 'Candy Store', 'Cantonese Restaurant', 'Caribbean Restaurant', 'Caucasian Restaurant', 'Chocolate Shop', 'Circus', 'Climbing Gym', 'Cocktail Bar', 'College Academic Building', 'College Arts Building', 'College Bookstore', 'College Cafeteria', 'Comedy Club', 'Community Center', 'Concert Hall', 'Convenience Store', 'Cooking School', 'Coworking Space', 'Creperie', 'Cuban Restaurant', 'Czech Restaurant', 'Dance Studio', 'Daycare', 'Deli / Bodega', 'Department Store', 'Dessert Shop', 'Dim Sum Restaurant', 'Diner', 'Discount Store', "Doctor's Office", 'Dog Run', 'Donut Shop', 'Drugstore', 'Dry Cleaner', 'Dumpling Restaurant', 'Duty-free Shop', 'Eastern European Restaurant', 'Egyptian Restaurant', 'Empanada Restaurant', 'English Restaurant', 'Ethiopian Restaurant', 'Event Space', 'Exhibit', 'Falafel Restaurant', 'Farmers Market', 'Fast Food Restaurant', 'Filipino Restaurant', 'Financial or Legal Service', 'Flower Shop', 'Food', 'Food & Drink Shop', 'Food Court', 'Food Stand', 'Food Truck', 'Fountain', 'Fried Chicken Joint', 'Frozen Yogurt Shop', 'Furniture / Home Store', 'Gaming Cafe', 'Garden', 'Garden Center', 'Gas Station', 'Gastropub', 'Gay Bar', 'German Restaurant', 'Golf Course', 'Gourmet Shop', 'Greek Restaurant', 'Gym / Fitness Center', 'Gym Pool', 'Gymnastics Gym', 'Hardware Store', 'Hawaiian Restaurant', 'Health Food Store', 'Heliport', 'High School', 'Historic Site', 'History Museum', 'Hobby Shop', 'Hookah Bar', 'Hostel', 'Hot Dog Joint', 'Hotel Bar', 'Hotpot Restaurant', 'Indian Restaurant', 'Indie Movie Theater', 'Intersection', 'Irish Pub', 'Israeli Restaurant', 'Japanese Curry Restaurant', 'Japanese Restaurant', 'Jazz Club', 'Jewish Restaurant', 'Karaoke Bar', 'Kids Store', 'Kitchen Supply Store', 'Korean Restaurant', 'Kosher Restaurant', 'Latin American Restaurant', 'Laundry Service', 'Lebanese Restaurant', 'Library', 'Lingerie Store', 'Lounge', 'Mac & Cheese Joint', 'Malay Restaurant', 'Martial Arts Dojo', 'Massage Studio', 'Mattress Store', 'Medical Center', 'Memorial Site', "Men's Store", 'Mexican Restaurant', 'Middle Eastern Restaurant', 'Mini Golf', 'Miscellaneous Shop', 'Mobile Phone Shop', 'Modern European Restaurant', 'Molecular Gastronomy Restaurant', 'Monument / Landmark', 'Moroccan Restaurant', 'Museum', 'Music School', 'Music Venue', 'Nail Salon', 'New American Restaurant', 'Newsstand', 'Non-Profit', 'North Indian Restaurant', 'Opera House', 'Optical Shop', 'Organic Grocery', 'Other Great Outdoors', 'Outdoor Sculpture', 'Outdoors & Recreation', 'Paella Restaurant', 'Paper / Office Supplies Store', 'Pedestrian Plaza', 'Performing Arts Venue', 'Persian Restaurant', 'Peruvian Restaurant', 'Pet Café', 'Pet Service', 'Pet Store', 'Pharmacy', 'Pie Shop', 'Pier', 'Pilates Studio', 'Playground', 'Poke Place', 'Pub', 'Public Art', 'Record Shop', 'Rental Car Location', 'Residential Building (Apartment / Condo)', 'Resort', 'Rest Area', 'River', 'Rock Club', 'Russian Restaurant', 'Sake Bar', 'Salad Place', 'Scandinavian Restaurant', 'School', 'Sculpture Garden', 'Shanghai Restaurant', 'Shipping Store', 'Shopping Mall', 'Skate Park', 'Smoke Shop', 'Snack Place', 'Soba Restaurant', 'Soccer Field', 'Soup Place', 'South American Restaurant', 'South Indian Restaurant', 'Southern / Soul Food Restaurant', 'Spa', 'Spanish Restaurant', 'Spiritual Center', 'Sporting Goods Shop', 'Sports Bar', 'Sports Club', 'Sri Lankan Restaurant', 'Stables', 'Steakhouse', 'Street Art', 'Strip Club', 'Supplement Shop', 'Swiss Restaurant', 'Szechuan Restaurant', 'Tailor Shop', 'Taiwanese Restaurant', 'Tapas Restaurant', 'Tattoo Parlor', 'Tea Room', 'Tech Startup', 'Temple', 'Tennis Court', 'Tennis Stadium', 'Thai Restaurant', 'Theme Park Ride / Attraction', 'Tiki Bar', 'Tourist Information Center', 'Toy / Game Store', 'Track', 'Trail', 'Train Station', 'Turkish Restaurant', 'Udon Restaurant', 'Used Bookstore', 'Vegetarian / Vegan Restaurant', 'Veterinarian', 'Video Game Store', 'Video Store', 'Vietnamese Restaurant', 'Volleyball Court', 'Waterfront', 'Whisky Bar', 'Wings Joint', 'Yoga Studio'

c. What does San Diego have that Manhattan does not? / Are there any other venue categories my client would like to have in her new neighborhood?

   Calculated the difference between the Manhattan venue occurrence frequency data frame columns (Table8) and San Diego venue occurrence frequency data frame columns (Table9) by using the difference() function.

'ATM', 'Amphitheater', 'Arts & Entertainment', 'Auto Garage', 'Auto Workshop', 'Automotive Shop', 'Baseball Stadium', 'Beach', 'Beach Bar', 'Bed & Breakfast', 'Big Box Store', 'Border Crossing', 'Botanical Garden', 'Brewery', 'Bus Stop', 'Business Service', 'Child Care Service', 'Comfort Food Restaurant', 'Construction & Landscaping', 'Country Dance Club', 'Credit Union', 'Cruise', 'Currency Exchange', 'Distillery', 'Dive Bar', 'Eye Doctor','Fish & Chips Shop', 'Fishing Spot', 'Fondue Restaurant','Football Stadium', 'Fruit & Vegetable Store', 'Gun Range', 'Home Service', 'Hotel Pool', 'IT Services', 'Indian Chinese Restaurant', 'Insurance Office', 'Laundromat', 'Light Rail Station', 'Locksmith', 'Marijuana Dispensary', 'Motel', 'Motorcycle Shop', 'Other Repair Shop', 'Pawn Shop', 'Piano Bar', 'Pop-Up Shop', 'Portuguese Restaurant', 'Print Shop', 'Recording Studio', 'Recreation Center', 'Rental Service', 'Roller Rink', 'Science Museum', 'Shopping Plaza', 'State / Provincial Park', 'Street Food Gathering', 'Tanning Salon', 'Theme Park', 'Theme Restaurant', 'Thrift / Vintage Store', 'Travel & Transport', 'Zoo Exhibit'

d. Venues that my client wants to have in her neighborhood

Given the lists of venues in Chelsea, venues that do not exist in Chelsea, and new venues that exist only in San Diego, my client has chosen:

i. All the venues categories in Chelsea

and

ii. Art museum, Bank, Climbing gym, Cocktail bar, Drugstore, English Restaurant, Flower shop, Frozen yogurt shop, Gas station, Gym/fitness center, Nail salon, Pet Service, Tiki bar, Automotive shop, Pharmacy

Created a new dataframe that includes all the venue categories my client desires to have in her new neighborhood in San Diego and displayed each neighborhood's venue occurrence frequency for each venue category (Table19):

Table 19.
Desired venues occurrence frequency by neighborhood

Neighborhood	American Restaurant	Art Gallery	Bagel Shop	Bakery	Bar	Beer Bar	Bookstore	Boutique	Café	Cheese Shop	Chinese Restaurant	Clothing Store	Club House	Coffee Shop	College Theater	Cosr
Adams North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Allied Gardens	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Alta Vista	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Azalea/Hollywood Park	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Balboa Park	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	

3. <u>Find Potential neighborhoods in San Diego that my client would like to look at / live in based on the venue analyzation</u>
   a. Top 10 San Diego neighborhoods based on the similarity score (Table18)
   b. Top 20 neighborhoods based on the desired venue list of my client

From the San Diego dataframe with only the desired venue categories (Table19), I counted the number of non '0' values for each neighborhood, i.e. number of venue categories of my client's preference per neighborhood (Figure2).

Figure 2.
Top 20 San Diego neighborhoods
that has most number of the desired

venue categories

```
Gaslamp            29
Hillcrest          29
Little Italy       25
Core-Columbia      25
Horton Plaza       24
Harborview         24
Pacific Beach      21
North Park         21
Grantville         19
Marina             18
Ocean Beach        17
Old Town           15
Petco Park         14
East Village       13
Park West          12
Golden Hill        11
Carmel Mountain    11
Normal Heights     11
South Park         11
dtype: int64
```

c. Neighborhoods that are in both 4a & 4b lists (Figure3) - potential neighborhoods.

   Find neighborhoods that intersect through inner joining/merging:

   Figure 3.
   San Diego Neighborhoods that are both most similar to Chelsea and have most number of desired venue categories

```
Gaslamp         29  0.183030
Hillcrest       29  0.214522
Little Italy    25  0.210613
Core-Columbia   25  0.196038
Horton Plaza    24  0.203715
North Park      21  0.193883
```

d. Analyze each potential neighborhoods by venues
   i. Looked at the venue occurrence frequency for each neighborhood per desired venue category (Table20):

   Table 20.
   Potential neighborhoods for my client - desired venue occurrence frequency

	Neighborhood	American Restaurant	Art Gallery	Bagel Shop	Bakery	Bar	Beer Bar	Bookstore	Boutique	Café	Cheese Shop	Chinese Restaurant	Clothing Store	Club House	Coffee Shop	College Theater	Cos
0	Gaslamp	3.0	2.0	0.0	3.0	3.0	0.0	0.0	2.0	5.0	0.0	0.0	3.0	0.0	6.0	0.0	
1	Hillcrest	0.0	0.0	0.0	2.0	0.0	0.0	1.0	0.0	2.0	0.0	3.0	1.0	0.0	4.0	0.0	
2	Little Italy	4.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	3.0	0.0	0.0	0.0	0.0	5.0	0.0	
3	Core-Columbia	3.0	0.0	0.0	2.0	1.0	0.0	0.0	2.0	1.0	0.0	0.0	2.0	0.0	5.0	0.0	
4	Horton Plaza	4.0	0.0	0.0	2.0	3.0	0.0	0.0	1.0	4.0	0.0	0.0	2.0	0.0	9.0	0.0	
5	North Park	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	5.0	0.0	0.0	1.0	0.0	3.0	0.0	

   ii. Looked at the 10 most common venues in each potential neighborhood (Table21):

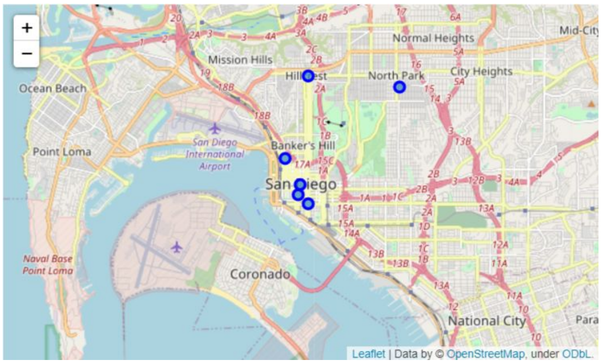
   Table 20.
   Potential neighborhoods for my client - 10 most common venues

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
Gaslamp	Hotel	Coffee Shop	Café	Gastropub	Clothing Store	Italian Restaurant	Seafood Restaurant	Sushi Restaurant	Bar	Bakery
Hillcrest	Mexican Restaurant	Pizza Place	Italian Restaurant	Sandwich Place	Coffee Shop	Restaurant	Breakfast Spot	Pharmacy	Chinese Restaurant	Greek Restaurant
Little Italy	Italian Restaurant	Wine Bar	Coffee Shop	American Restaurant	Hotel	Café	Pizza Place	New American Restaurant	Japanese Restaurant	Grocery Store
Core-Columbia	Coffee Shop	Hotel	Seafood Restaurant	Shoe Store	Accessories Store	American Restaurant	Sushi Restaurant	Mexican Restaurant	Italian Restaurant	Lingerie Store
Horton Plaza	Hotel	Coffee Shop	Mexican Restaurant	Italian Restaurant	American Restaurant	Café	Seafood Restaurant	Sushi Restaurant	Bar	Bakery
North Park	Café	Brewery	Pizza Place	Coffee Shop	Music Venue	Sushi Restaurant	New American Restaurant	Taco Place	Breakfast Spot	Italian Restaurant

e. Visualization : Map

First, added/merged latitude & longitude data for each potential neighborhood. Then by using folium mapping, I created the map visualization of the potential neighborhoods' location (Map3):

Map 3.
Potential neighborhoods for my client - location map

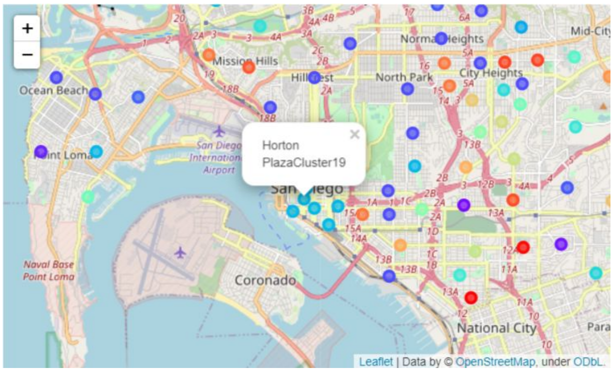


v. Find similar neighborhoods to the potential neighborhood in San Diego

I explored more neighborhoods by clustering (Table10) based on their venues. I used K-means to cluster the neighborhoods in 70 clusters, where k=70, to see what other neighborhoods my client might be interested in, so she has more options to choose from when it comes to housing prices and crime rates.

1. Visualize location - Map (Map3):

Map 3.
Location Map - Clustered neighborhoods in San Diego



2. Examine clusters that include potential neighborhoods I found earlier
    a. 1st Cluster that includes Gaslamp, Core-Columbia, Little Italy and Horton Plaza - Cluster 19

       Extracted neighborhoods data that had cluster label 19 by using the loc() function (Table22):

Table 22.
1st Clustered neighborhood area with neighborhoods and 10 common venues

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
Morena	Coffee Shop	Café	Sandwich Place	Food Truck	Massage Studio	Martial Arts Dojo	Brewery	Bakery	Bridal Shop	Pizza Place
Gaslamp	Hotel	Coffee Shop	Café	Gastropub	Clothing Store	Italian Restaurant	Seafood Restaurant	Sushi Restaurant	Bar	Bakery
Sorrento Valley	Hotel	Food	Coffee Shop	Café	Chinese Restaurant	Restaurant	Sushi Restaurant	Indian Restaurant	Mexican Restaurant	Deli / Bodega
Horton Plaza	Hotel	Coffee Shop	Mexican Restaurant	Italian Restaurant	American Restaurant	Café	Seafood Restaurant	Sushi Restaurant	Bar	Bakery
Core-Columbia	Coffee Shop	Hotel	Seafood Restaurant	Shoe Store	Accessories Store	American Restaurant	Sushi Restaurant	Mexican Restaurant	Italian Restaurant	Lingerie Store
South Park	Juice Bar	Food Truck	Italian Restaurant	Café	Brewery	Breakfast Spot	Flower Shop	Bike Shop	Coffee Shop	Street Food Gathering
Roseville / Fleet Ridge	Sporting Goods Shop	Coffee Shop	American Restaurant	Hotel	Hotel Bar	Convenience Store	Sushi Restaurant	Ethiopian Restaurant	Exhibit	Eye Doctor
Little Italy	Italian Restaurant	Wine Bar	Coffee Shop	American Restaurant	Hotel	Café	Pizza Place	New American Restaurant	Japanese Restaurant	Grocery Store
Mission Valley East	Furniture / Home Store	Salon / Barbershop	Coffee Shop	Sandwich Place	Hotel	Gym	Lingerie Store	Bar	Paper / Office Supplies Store	Lounge
East Village	Coffee Shop	Thrift / Vintage Store	Café	Beer Bar	Breakfast Spot	Sporting Goods Shop	Grocery Store	Mexican Restaurant	Gastropub	Eastern European Restaurant
Petco Park	Hotel	Brewery	Light Rail Station	Bar	Park	ATM	Clothing Store	Spa	Breakfast Spot	Burger Joint
Marina	Hotel	Seafood Restaurant	Hotel Bar	Park	Breakfast Spot	New American Restaurant	American Restaurant	Gym	Shopping Plaza	Fish Market
Cortez	Park	Café	Coffee Shop	Concert Hall	Mexican Restaurant	Taco Place	Italian Restaurant	Middle Eastern Restaurant	Convenience Store	Hotel
Harborview	Boat or Ferry	Italian Restaurant	Coffee Shop	Wine Bar	American Restaurant	Pizza Place	Rental Car Location	Hotel	Café	Japanese Restaurant

b. 2nd Cluster that includes Hillcrest and North Park - Cluster 8
Extracted neighborhoods data that had cluster label 8 by using the loc() function (Table23):

Table 23.
2nd Clustered neighborhood area with neighborhoods and 10 common venues

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
Egger Highlands	Mexican Restaurant	Fast Food Restaurant	Coffee Shop	Marijuana Dispensary	Juice Bar	Sandwich Place	Thai Restaurant	Auto Workshop	ATM	Pizza Place
Midtown	Rental Car Location	Pizza Place	Pub	Gas Station	Intersection	Lounge	Motorcycle Shop	Pet Service	Print Shop	Whisky Bar
Grantville	Coffee Shop	Mexican Restaurant	Pharmacy	Sushi Restaurant	Fast Food Restaurant	Sandwich Place	Vietnamese Restaurant	Pet Store	Brewery	Pilates Studio
Ocean Beach	Café	Brewery	Mexican Restaurant	Pizza Place	Bar	Coffee Shop	Antique Shop	Smoke Shop	Breakfast Spot	Park
Clairemont Mesa East	Bubble Tea Shop	Bookstore	Convenience Store	Fast Food Restaurant	Chinese Restaurant	Grocery Store	Breakfast Spot	Mediterranean Restaurant	Martial Arts Dojo	Board Shop
Mission Beach	Beach	Burger Joint	Sandwich Place	Pizza Place	Dessert Shop	New American Restaurant	Board Shop	Sushi Restaurant	Taco Place	Coffee Shop
Hillcrest	Mexican Restaurant	Pizza Place	Italian Restaurant	Sandwich Place	Coffee Shop	Restaurant	Breakfast Spot	Pharmacy	Chinese Restaurant	Greek Restaurant
North Park	Café	Brewery	Pizza Place	Coffee Shop	Music Venue	Sushi Restaurant	New American Restaurant	Taco Place	Breakfast Spot	Italian Restaurant
Teralta West	Mexican Restaurant	Pizza Place	Vietnamese Restaurant	Taco Place	Food Truck	Liquor Store	Sandwich Place	Fast Food Restaurant	Donut Shop	Snack Place
Rancho Bernardo	Pizza Place	Coffee Shop	Sports Bar	Nail Salon	Tennis Court	Thai Restaurant	Bakery	Spa	Pharmacy	American Restaurant
Burlingame	Pizza Place	Mexican Restaurant	Bar	Indian Restaurant	Record Shop	Coffee Shop	Nail Salon	Café	Bookstore	Gift Shop
Rancho Penasquitos	Sandwich Place	Pizza Place	Mexican Restaurant	Bank	Bubble Tea Shop	Smoke Shop	Ice Cream Shop	Donut Shop	Chinese Restaurant	Eye Doctor
Pacific Beach	Sandwich Place	Coffee Shop	Convenience Store	Shipping Store	Mobile Phone Shop	Breakfast Spot	Chinese Restaurant	Bagel Shop	Pizza Place	Gym
Normal Heights	Convenience Store	Liquor Store	ATM	BBQ Joint	Pet Store	Park	Donut Shop	Sandwich Place	Locksmith	Bookstore
Barrio Logan	Fast Food Restaurant	Mexican Restaurant	Coffee Shop	Pizza Place	BBQ Joint	Filipino Restaurant	Boat or Ferry	Marijuana Dispensary	Frozen Yogurt Shop	Sandwich Place
Point Loma Heights	Hostel	Italian Restaurant	Cupcake Shop	Pharmacy	Convenience Store	Dive Bar	Nail Salon	Burger Joint	Mexican Restaurant	Breakfast Spot
University Heights	Coffee Shop	Pizza Place	Bar	Ice Cream Shop	Breakfast Spot	Middle Eastern Restaurant	Ethiopian Restaurant	Sandwich Place	New American Restaurant	Mexican Restaurant
Kensington	Coffee Shop	Liquor Store	Spa	Café	Pizza Place	Burger Joint	Pet Store	Bar	Video Store	French Restaurant
Del Mar Heights	Martial Arts Dojo	Greek Restaurant	Breakfast Spot	Shopping Plaza	Fast Food Restaurant	Seafood Restaurant	Sandwich Place	Sushi Restaurant	Restaurant	Insurance Office
Grant Hill	Mexican Restaurant	Indie Theater	Pharmacy	Convenience Store	Comfort Food Restaurant	Discount Store	Coffee Shop	Park	Restaurant	Caribbean Restaurant
Balboa Park	Zoo Exhibit	Garden	Art Museum	Gift Shop	History Museum	Performing Arts Venue	Exhibit	Donut Shop	Museum	Fountain
Fox Canyon	Pizza Place	Asian Restaurant	Taco Place	Grocery Store	Dive Bar	Rental Service	Chinese Restaurant	Ethiopian Restaurant	Motorcycle Shop	Food
San Ysidro	Convenience Store	Fast Food Restaurant	Pet Store	Juice Bar	Taco Place	Chinese Restaurant	Marijuana Dispensary	Grocery Store	Sandwich Place	Video Store
Carmel Mountain	Fast Food Restaurant	Sandwich Place	Donut Shop	Mexican Restaurant	Italian Restaurant	Shipping Store	Mattress Store	Breakfast Spot	Flower Shop	Food
Park West	Coffee Shop	Mexican Restaurant	Sushi Restaurant	French Restaurant	Ramen Restaurant	Mediterranean Restaurant	Bridge	Brewery	Breakfast Spot	Lounge
Golden Hill	Coffee Shop	Mexican Restaurant	Liquor Store	Convenience Store	Bakery	Deli / Bodega	Diner	Pedestrian Plaza	Pawn Shop	Sandwich Place

vi. Explore & analyze neighborhoods by housing prices in San Diego

1. I first looked at the overall housing prices in San Diego (Figure4). Then focused on the 2 or more bedroom houses, since my client is only interested in houses with at least 2bedrooms, preferably 3beds, in every neighborhood in San Diego (Table11, Figure5).

Figure 4.
Housing price overview in San Diego

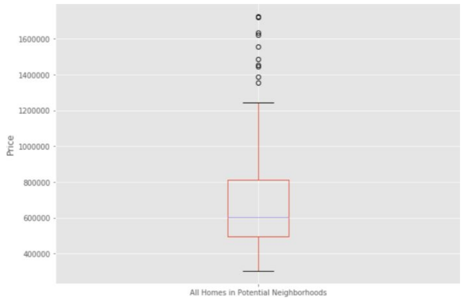
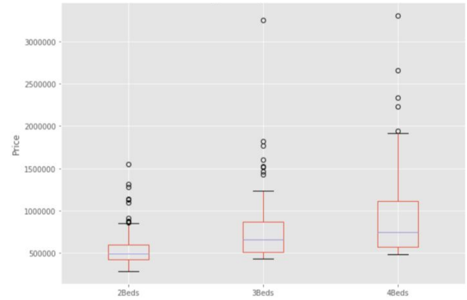


Figure 5.
Housing price overview by number of bedrooms



2. After, I extracted only the neighborhoods that I was interested in from clustered neighborhoods (Table24).

Table 24.
Housing prices of neighborhoods I was are interested in

Neighborhood	Latitude	Longitude	Zillow Home Value Index_Allhomes	Median Value per sq.ft($)_Allhomes	Zillow Home Value Index_2bed	Zillow Home Value Index_3bed	Zillow Home Value Index_4bed
Morena	32.769888	-117.193218	517500.0	408.0	458700.0	580300.0	807700.0
Gaslamp	32.711518	-117.160128	445000.0	677.0	598500.0	NaN	NaN
Sorrento Valley	32.898222	-117.190598	901800.0	452.0	597000.0	708900.0	1026100.0
Horton Plaza	32.714130	-117.163787	1388400.0	649.0	1553300.0	3252700.0	3305600.0
Core-Columbia	32.717211	-117.162952	764100.0	613.0	914700.0	1134200.0	NaN
South Park	32.723523	-117.126734	800400.0	695.0	723500.0	867600.0	987700.0
Roseville / Fleet Ridge	32.727408	-117.233350	1200000.0	2571.0	818900.0	1190500.0	1565500.0
Little Italy	32.724606	-117.168006	524100.0	571.0	590600.0	824500.0	1165100.0
Mission Valley East	32.774494	-117.140347	480300.0	1574.0	463800.0	676800.0	716600.0
East Village	32.712100	-117.152221	528500.0	1333.0	624800.0	654500.0	931500.0
Marina	32.710724	-117.167559	764400.0	607.0	820700.0	1235500.0	2658500.0
Cortez	32.721604	-117.156100	499900.0	2184.0	585900.0	771900.0	NaN
Harborview	32.724925	-117.172413	724100.0	566.0	839900.0	992400.0	NaN
Egger Highlands	32.587312	-117.100488	545700.0	1638.0	438600.0	529100.0	570800.0
Midtown	32.739864	-117.175020	865700.0	624.0	678900.0	1012000.0	1386400.0
Grantville	32.791924	-117.099716	357700.0	411.0	380800.0	532800.0	645400.0
Ocean Beach	32.748289	-117.246648	920200.0	1030.0	870000.0	1068000.0	1356200.0
Clairemont Mesa East	32.818926	-117.172351	675000.0	845.0	379900.0	672200.0	717100.0
Mission Beach	32.773229	-117.250313	1242400.0	1138.0	1138100.0	1429300.0	1869900.0
Hillcrest	32.748324	-117.160321	584700.0	535.0	567800.0	898800.0	1184500.0
North Park	32.745051	-117.128897	649400.0	715.0	590200.0	857500.0	993200.0
Teralta West	32.752399	-117.104799	442500.0	598.0	423700.0	557300.0	566800.0
Rancho Bernardo	33.032546	-117.077018	671800.0	1001.0	491600.0	708500.0	839500.0
Burlingame	32.732625	-117.127283	947800.0	633.0	840500.0	963300.0	1121900.0
Rancho Penasquitos	32.963241	-117.124901	828200.0	793.0	400600.0	741500.0	861800.0
Pacific Beach	32.803154	-117.237876	948900.0	799.0	750600.0	1090800.0	1483300.0
Normal Heights	32.759384	-117.117752	541600.0	736.0	560300.0	684800.0	790600.0
Barrio Logan	32.692421	-117.135204	469800.0	644.0	459400.0	485600.0	599100.0
Point Loma Heights	32.743653	-117.233610	818000.0	1290.0	579500.0	966600.0	1202500.0
University Heights	32.757811	-117.148354	597600.0	663.0	542600.0	867000.0	1023200.0
Kensington	32.767054	-117.104436	895000.0	2044.0	644900.0	945800.0	1176700.0
Del Mar Heights	32.950792	-117.250699	1557200.0	1766.0	1095400.0	1511700.0	1847400.0
Grant Hill	32.709764	-117.135070	505800.0	523.0	483900.0	523100.0	620100.0
Balboa Park	32.731088	-117.145920	922200.0	613.0	783700.0	948200.0	1136300.0
Fox Canyon	32.746445	-117.090505	433800.0	528.0	417000.0	517500.0	564700.0
San Ysidro	32.556028	-117.048248	496300.0	964.0	289100.0	474000.0	519000.0
Carmel Mountain	32.977222	-117.077848	774300.0	445.0	538500.0	748300.0	871300.0
Golden Hill	32.716448	-117.135596	507200.0	584.0	466900.0	753400.0	959600.0

3. Visualized each column and include all the neighborhoods for an easy comparison (Figure6, Figure7, Figure8, Figure9, Figure10).

Figure 6.
Overview

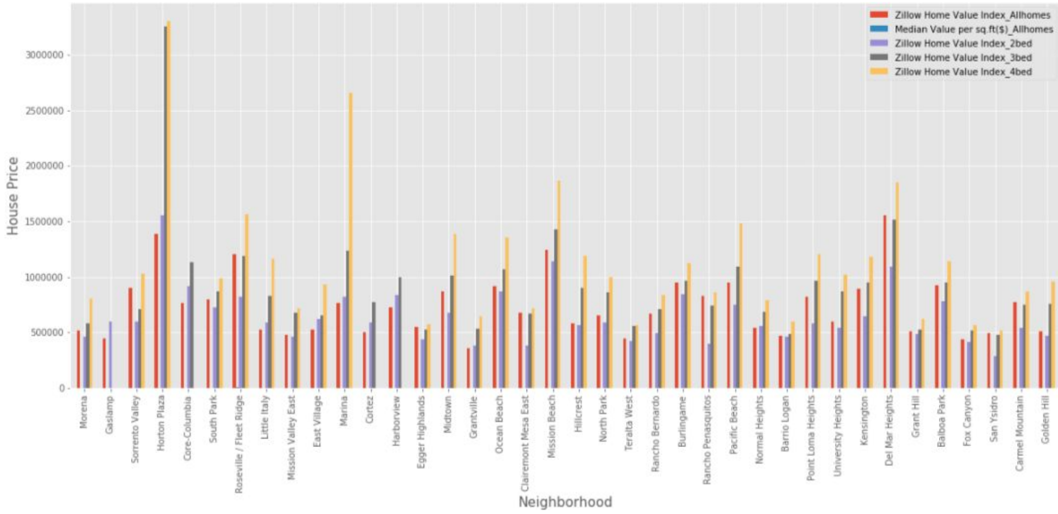


Figure 7.
All homes

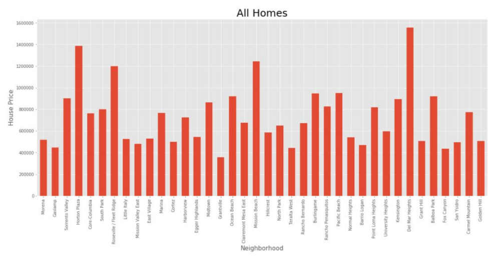


Figure 8.
2 bedrooms

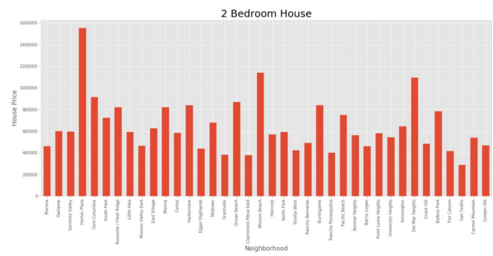


Figure 9.
3 bedrooms

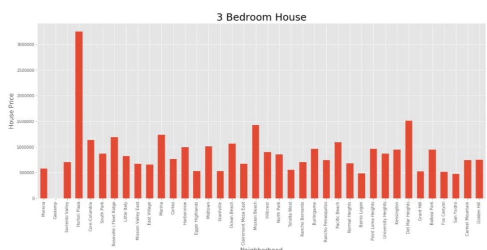
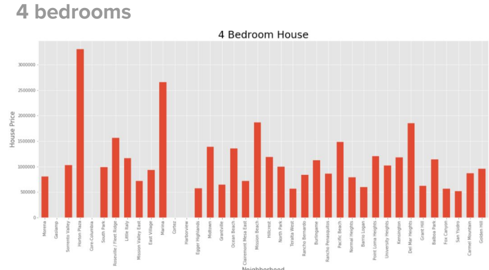
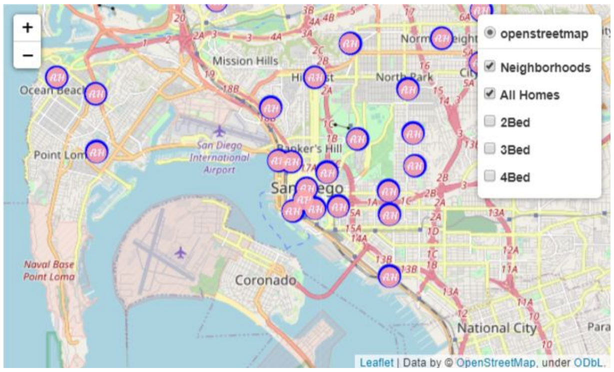


Figure 10.
4 bedrooms



4. Visualized above data with location - folium mapping (Map4). I've added 5 markers that indicate Neighborhoods, All Homes, 2Bed, 3Bed, and 4Bed, in the toggle form. All the markers had popups in which Neighborhood markers show neighborhood names and all the others show prices. This way, I could easily identify which neighborhoods had what kind of houses and prices at the same time.

Map 3.
Location Map - Clustered neighborhoods in San Diego



vii. Explore & analyze neighborhoods by crime rate in San Diego

1. From the interested neighborhoods' crime rate data frame (Table15), I visualized each crime category for every neighborhood from clustering (Figure11, Figure12, Figure13, Figure14, Figure15, Figure16)

Figure 11.
Crime Overview

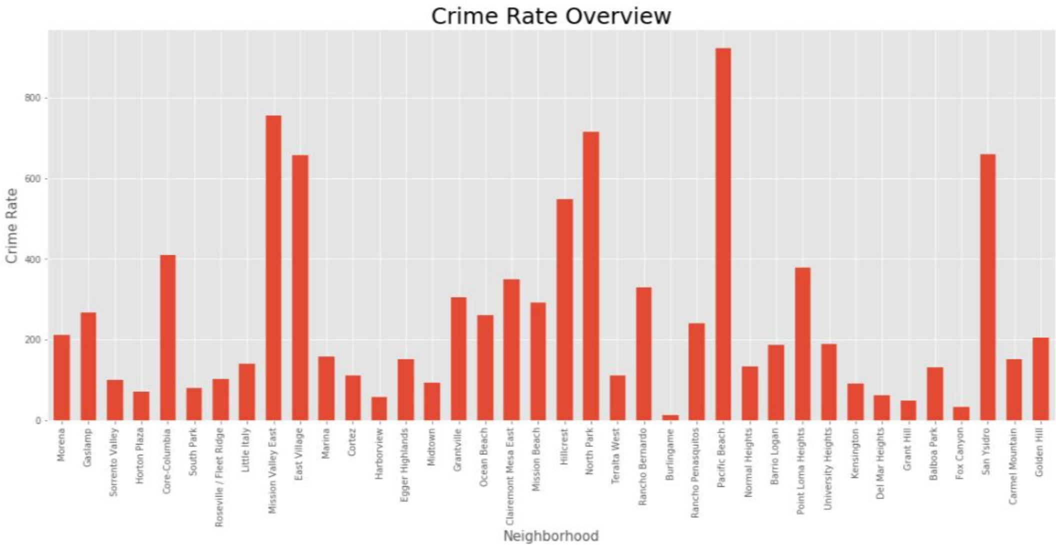


Figure 12.
Crime Overview - detailed

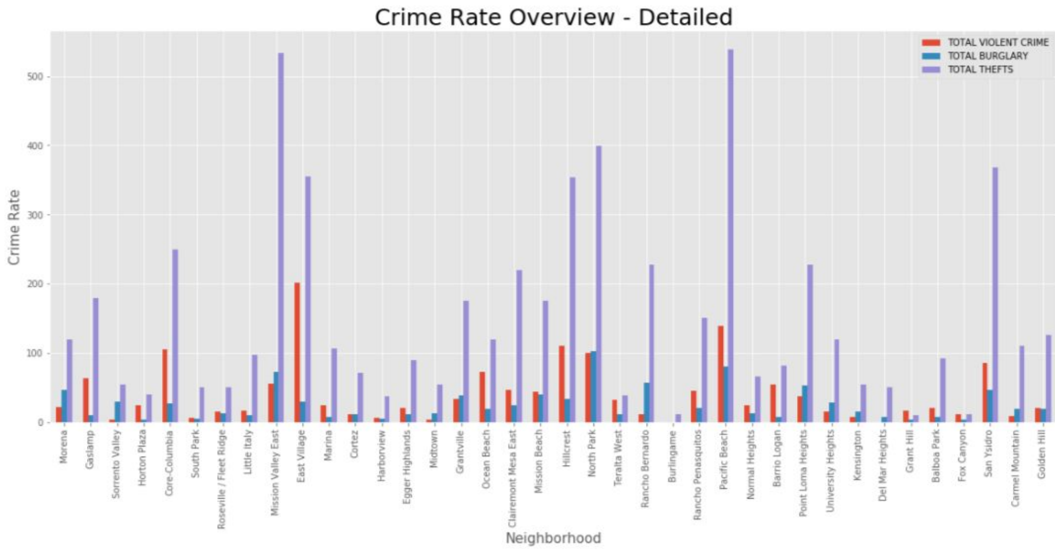


Figure 13.
All homes

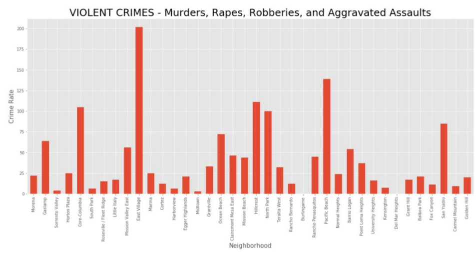


Figure 14.
2 bedrooms

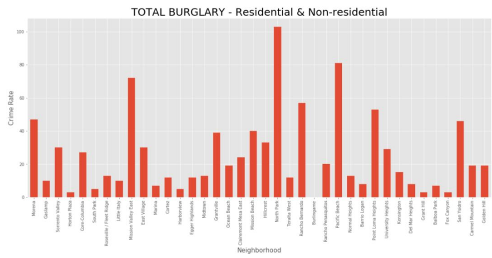


Figure 13.
All homes

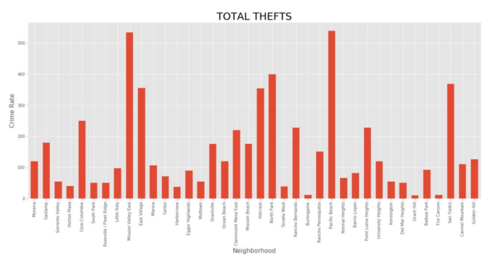
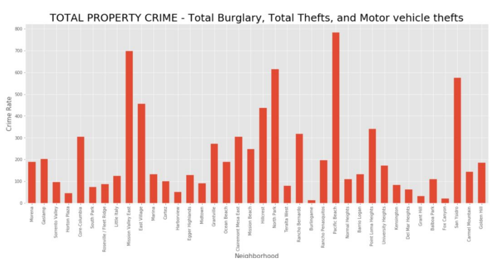


Figure 14.
2 bedrooms



Ⅴ. Results / Exploratory Data Analysis

City Overview - Manhattan & San Diego

Manhattan has 40 neighborhoods and San Diego has 124 neighborhoods

Venue Overview - Manhattan & San Diego

Manhattan has 330 unique venue categories and San Diego has 276 unique venue categories.

Chelsea, Manhattan - Analyze Chelsea's venues

Chelsea's top 15 common venues are (Table17): Art Gallery, Coffee Shop, Cafe, Ice Cream Shop, American Restaurant, Cycle Studio, Seafood Restaurant, Market, Theater, Bakery, Boutique, Cupcake Shop, Hotel, Italian Restaurant, Park

Find Potential neighborhoods that fits my client's criteria

San Diego neighborhoods that are similar to Chelsea, Manhattan by venues are (Table18): Gaslamp, North Park, Core-Columbia, Allied Gardens, Sunset Cliffs, Rancho Encantada, Ocean Crest, Horton Plaza, Little Italy, and Hillcrest.

Client's desired venues in her new neighborhood in San Diego are: American Restaurant, Art Gallery, Bagel Shop, Bakery, Bar, Beer bar, Bookstore, Boutique, Cafe, Cheese shop, Cupcake shop, Cycle Studio, Electronics store, Fish market, Flea market, French restaurant, General entertainment, Gift shop, Grocery store, Gym, Harbor/Marina, Health & Beauty service, hotel, Ice cream shop, Indie theater, Italian restaurant, Jewelry store, Juice bar, Liquor store, Market, Meditterranean restaurant, Movie theater, Nightclub, Noodle house, Office, Park, Photography studio, Physical therapist, Pizza place, Plaza, Pool, Ramen restaurant, Restaurant, Roof deck, Salon/Barbershop, Sandwich place, Scenic lookout, Seafood restaurant, Shoe store, Speakeasy, Supermarket, Sushi restaurant, Taco place, Theater, Wine bar, Wine shop, Women's store, Art museum, Bank, Climbing gym, Cocktail bar, Drugstore, English Restaurant, Flower shop, Frozen yogurt shop, Gas station, Gym/fitness center, Nail salon, Pet Service, Tiki bar, Automotive shop, Pharmacy.

Neighborhoods with most number of my client's preferred venue categories are: Gaslamp, Hillcrest, Little Italy, Core-Columbia, Horton Plaza, Harborview, Pacific Beach, North Park, Grantville, and Marina.

Neighborhoods that fit my client's preference the best are: Gaslamp, Hillcrest, Little Italy, Core-Columbia, Horton Plaza, and North Park.

Find similar neighborhoods to the potential neighborhood in San Diego

There are two potential areas in San Diego that would be suitable for my client:

- The first neighborhood area includes Gaslamp, Horton Plaza, Little Italy and Core-Columbia. And neighborhoods that are similar to them in the area are - Morena, Sorrento Valley, South Park, Roseville/Fleet Ridge, Mission Valley East, East Village, Petco Park, Marina, Cortez and Harborview
- The second neighborhood area includes North Park and Hillcrest. And neighborhoods that are similar to them in the area are - Egger Highlands, Midtown, Grantville, Ocean Beach, Clairemont Mesa East, Teralta West, Rancho Bernardo, Burlingame, Rancho Penasquitos, Pacific Beach, Normal Heights, Barrio Logan, Point Loma Heights, University Heights, Kensington, Del Mar Heights, Grant Hill, Balboa Park, Fox Canyon, San Ysidro, Carmel Mountain, Park West, Golden Hill.

Explore & analyze neighborhoods by housing prices in San Diego

San Diego's overall home price ranged from $304,100 to $1,726,200, where the data was skewed right. This means most data were concentrated on the low end of the scale, where the median was around $600,000 and the mean price of all homes in San Diego was $715,131. There were only 9 outliers out of 118, so I would say most houses in San Diego fall under my client's budget of $1.5m.

In San Diego, 3 bedroom houses are more common than 2 bedroom and 4 bedroom houses. Only 112 out of 118 neighborhoods had 2 bedroom houses and 110 out of 118 neighborhoods had 4 bedroom houses, whereas 116 out of 118 had 3 bedroom houses.

My client preferred 3 bedroom houses, which ranged from $428,900 to $3,252,700 and the data was skewed right. Also, the mean price of all 3 bedroom houses came around $773,032, meaning most 3 bedroom houses fell under my client's budget.

The budget is below $1.5m. And my client preferred 3 bedroom houses. Hence, Horton Plaza, Gaslamp, and Del Mar Heights were out of options, because Horton Plaza and Del Mar Heights 3 bedroom houses were over our budget and Gaslamp did not have 3 bedroom houses.

Based on the findings, out of our first option neighborhoods, only Hillcrest, Little Italy, Core-Columbia, and North Park were left for consideration.

Explore & analyze neighborhoods by crime rate in San Diego

From the crime rate graphs, I'd eliminated following neighborhoods from consideration: Core-Columbia, Mission Valley East, East Village, Grantville, Ocean Beach, Clairemont Mesa East, Mission Beach, Hillcrest, North Park, Rancho Bernardo, Rancho Penasquitos, Pacific Beach, Point Loma Heights, San Ysidro. All these neighborhoods had 200 or more crimes between January - August 2019. That left us with: Morena, Sorrento Valley, South Park, Roseville / Fleet Ridge, Little Italy, Marina, Cortez, Harborview, Egger Highlands, Midtown, Teralta West, Burlingame, Normal Heights, Barrio Logan, University Heights, Kensington, Grant Hill, Balboa Park, Fox Canyon, Carmel Mountain, and Golden Hill.

However, the following neighborhoods had close to and/or more than 25 violent crimes: Morena, Marina, Egger Highlands, Teralta West, Normal Heights, Barrio Logan, Balboa Park, and Golden Hill. Hence these could not fall under our consideration. Neighborhoods I could look at were now Sorrento Valley, South Park, Roseville / Fleet Ridge, Little Italy, Cortez, Harborview, Midtown, Burlingame, University Heights, Kensington, Grant Hill, Fox Canyon, and Carmel Mountain.

Then I looked at the Burglary data, and I eliminated Sorrento Valley, University Heights, and Carmel Mountain.

All of the above left me with South Park, Roseville / Fleet Ridge, Little Italy, Cortez, Harborview, Midtown, Burlingame, Kensington, Grant Hill, and Fox Canyon. These also fell under 100 thefts between January - August 2019.

The safest neighborhoods were Burlingame, Fox Canyon, and Grant Hill.

Based on the findings, out of our first option neighborhoods, Little Italy was still left for consideration.

Ⅵ. Discussion

Based on the findings, my client's first choice would be Little Italy

- Little Italy's similarity score to Chelsea, Manhattan was 0.21, 3rd most similar.
- Little Italy had 25 of the preferred venue categories, third most preferred venues, where Gaslamp and Hillcrest had the most preferred venues - 29.
- 10 Most common venues in Little Italy are: Italian restaurant, Wine bar, Coffee shop, American restaurant, Hotel, Cafe, Pizza place, New American restaurant, Japanese restaurant, Grocery store
- Little Italy's 3 bedroom houses cost around $824,500 and 4 bedroom houses cost around $1,165,100. So if my client wanted, she could get either 3 bedroom or 4 bedroom houses.
- Distance to Gaslamp, which is most similar to Chelsea and has most preferred venues - 1.4mi, 8min drive, 28min walk

Then the second choice would be South Park, Roseville/Fleet Ridge, Cortez and Harborview, since these neighborhoods fell in the same cluster as Little Italy, meaning they are similar to Little Italy.

- All of them have 3 bedroom houses and are under budget - South Park($867,600), Roseville/Fleet Ridge($1,190,500), Cortez($771,900), Harborview($992,400)
- However, both Cortez and Harborview don't have 4 bedroom houses and Roseville/Fleet Ridge's 4 bedroom houses are over $1.5m. So if my client would like to consider 4 bedroom houses as well, the only choice left would be South Park($987,700).

- 10 most common venues in South Park: Juice bar, Food truck, Italian restaurant, Cafe, Brewery, Breakfast spot, Flower shop, Bake shop, Coffee shop, Street food gathering
- 10 most common venues in Roseville/Fleet Ridge: Sporting goods shop, coffee shop, American restaurant, Hotel, Hotel bar, Convenience store, Sushi restaurant, Ethiopian restaurant, Exhibit, Eye doctor
- 10 most common venues in Cortez: Park, Cafe, Coffee shop, Concert hall, Mexican restaurant, Taco place, Italian restaurant, Middle Eastern restaurant, Convenience store, Hotel
- 10 most common venues in Harborview: Boat or Ferry, Italian restaurant, Coffee shop, Wine bar, American restaurant, Pizza place, Rental car location, Hotel, Cafe, Japanese restaurant
- Distance to Gaslamp, which is most similar to Chelsea and has most preferred venues - South Park(2.6mi, 8min drive, 51min walk), Roseville/Fleet Ridge(5.6mi, 16min drive, 1h51min walk), Cortez(9mi, 5min drive, 18min walk), Harborview(1.7mi, 8min drive, 36min walk)

The third choice would be Midtown, Burlingame, Kensington, Grant Hill and Fox Canyon.

- Burlingame, Grant Hill and Fox Canyon are the safest neighborhoods
- 10 most common venues in Midtown: Rental car location, Pizza place, Pub, Gas station, Intersection, Lounge, Motorcycle shop, Pet service, Print shop, Whisky bar
- 10 most common venues in Burlingame: Pizza place, Mexican restaurant, Bar, Indian restaurant, Record shop, Coffee shop, Nail salon, Cafe, Bookstore, Gift shop
- 10 most common venues in Kensington: Coffee shop, Liquor store, Spa, Cafe, Pizza place, Burger joint, Pet store, Bar, Video store, French restaurant
- 10 most common venues in Grant Hill: Mexican restaurant, Indie theater, Pharmacy, Convenience store, Comfort food restaurant, discount store, Coffee shop, Park, Restaurant, Caribbean restaurant
- 10 most common venues in Fox Canyon: Pizza place, Asian restaurant, Taco place, Grocery store, Dive bar, Rental service, Chinese restaurant, motorcycle shop, food
- All of them have 3 bedroom houses and are under budget - Midtown($1,012,000), Burlingame($963,300), Kensington($945,800), Grant Hill($523,100) and Fox Canyon($517,500)

- All of them have 4 bedroom houses and are under budget - Midtown($1,386,400), Burlingame($1,121,900), Kensington($1,176,700), Grant Hill($620,100) and Fox Canyon($564,700)

- Distance to Gaslamp, which is most similar to Chelsea and has most preferred venues - Midtown(2.4mi, 8min drive, 54min walk), Burlingame(3.2mi, 10min drive, 1h2min walk), Kensington(6.5mi, 11min drive, 2h7min walk), Grant Hill(1.7mi, 6min drive, 34min walk) and Fox Canyon(5.7mi, 11min drive, 2h5min walk)

Ⅶ. Conclusions

In this project, I analyzed the neighborhoods in San Diego by venues, housing price, and crime rate, in order to provide my client with a list of neighborhoods that fit her criteria and she would be interested in looking.

My research consists of similarity scores between Chelsean, Manhattan and San Diego neighborhoods, the number of preferred venue categories that each neighborhood in San Diego has, 10 most common venue categories for every neighborhood in San Diego, map of San Diego with neighborhood locations, and housing prices and crime data per category for every neighborhood in San Diego. Overall, my research showed all the data she wanted and would want about San Diego neighborhoods.

My analysis summarizes the research I carried and makes it easy for my client to look and understand the data, and decide where to live, even though she doesn't know San Diego very well.

In conclusion, my research and analysis show that my client would want to look at Little Italy as her next neighborhood to live in when she moves to San Diego.