The best location for a new restaurant

1. Introduction/Business Problem section

Hello! My theme is how to get the best location for a new Chinese restaurant in Los Angeles?

As we all know, the popularity of Chinese food is getting higher and higher in the world. In the United States, we can also see many Chinese restaurants. So, my question is, where is the best place for someone who wants to open a Chinese restaurant in Los Angeles? **This question is closely related to the people who plan to run their own restaurants and I believe that this can also be expanded to many other businesses.**

My research method is mainly studying the location and operation of existing Chinese restaurants in Los Angeles. Firstly, after studying their information, we could imitate and help us make decsions. Secondly, we can also avoid those well-managed exsiting Chinese restaurants in the same street or neighborhood to avoid direct competition. And at the same time , I added a extra factor, crime rate data, into my analysis.

Therefore, the data I need include:

1) Los Angeles street distribution,
2) Los Angeles Chinese restaurant information,
3) Los Angeles street population data etc.

The above data is mainly obtained through foursquare and the local statistical website of Los Angeles, and then data cleaning, visualization and clustering are performed on it to obtain the results I want.

When finished the project, we can answer several questions as followed:

1) Which zone has a relative better safety?
2) What streets can be considered similar from commercial angle?
3) Where should I locate my new Chinese Restaurant in LA?

2. Data Collection

Based on my target, I collected some various data as followed:

1) LA neighborhoods geojson (To divide the city into different zones) ***from Los Angeles Open Data***
2) LA neighborhoods demographic data (Include population and owner/renter quantity) ***from Los Angeles Open Data***
3) LA streets name lists ***from Los Angeles Open Data***
4) LA streets geo locations (Latitudes & Longitudes) ***by using GeoLocators***
5) LA crime data (Quantity, Location, Area, Type) ***from Los Angeles Open Data***
6) LA restaurant information ***by using Foursquare***
7) LA chinese restaurant information ***by using Foursquare***

However, the meta data quality I acquired is not high I think, so I emit some more detailed information to analyze.

3. Data Wrangling

The whole wrangling process includes:

Streets & Neighborhoods data → Streets & Neighborhoods geo data → Venue and Restaurants data

Read data including neighborhoods name, streets name, crime rate of neighborhoods and obtain their specific geo locations with help of *geocoder.*

```
strloc
```

Out[118]:

	street	council	loca
0	MALDEN ST	West Hills	(34.2264751, -118.4698491)
1	NAPA ST	West Hills	(46.0538992, 21.576302)
2	CHASE ST	West Hills	(34.2129886, -118.3921419)
3	ELKWOOD ST	Canoga Park	(34.2119885, -118.616682)
4	CAPISTRANO AVE	West Hills	(34.220179, -118.6519326)
...	...	...	...
585	BURLINGTON AVE	Westlake North	(33.7886538, -118.241847)
586	BEVERLY BLVD	Westlake North	(36.057071, -79.863219)
587	LAKE ST	Echo Park	(42.286454, -71.4009814)
588	COLTON ST	Echo Park	(40.7771416, -80.771736)
589	PATTON ST	Echo Park	(33.88688315, -117.31635904990785)

590 rows × 3 columns

	Name	TotalPopulation	Owner	Renter	Lati	Longi
1	PORTER RANCH NC	21786.70	5912.04	1470.36	34.281816	-118.561271
2	TARZANA NC	36091.73	7448.16	5930.31	34.171444	-118.542979
3	RESEDA NC	69166.59	10256.75	10025.51	34.200078	-118.536988
4	NORTHRIDGE WEST	21305.59	4797.84	2674.91	34.234561	-118.536932
5	SYLMAR NC	75158.20	13252.74	6133.06	34.307625	-118.449215
6	SHERMAN OAKS NC	66486.14	13330.17	17925.01	34.150872	-118.448987
7	VAN NUYS NC	86434.13	7417.51	20384.25	34.186619	-118.448667
8	WEST LOS ANGELES NC	32515.06	2914.16	12786.70	34.046399	-118.448135
9	PANORAMA CITY NC	69984.74	6328.07	11481.21	34.224290	-118.445374
10	NC WESTCHESTER/PLAYA	58145.15	12612.31	11986.51	33.939319	-118.440478

(street & neighborhood data)

We can see the column[loca] shows a sort of tuple things so necessarily I need to switch it into 2 independent columns. And I got this:

```
In [120]: strloc.sort_values(by=['Longi'],axis=0,ascending=True,inplace=True)
print(strloc.shape)
strloc = strloc.iloc[40:437]
strloc
```

(590, 3)

Out[120]:

	Street	Lati	Longi
240	LURLINE AVE	34.211100	-118.592871
231	TRIBUNE ST	34.249691	-118.588566
46	VENTURA BLVD	34.161291	-118.588374
239	SAN JOSE ST	34.234400	-118.584204
244	NORDOFF ST	34.193529	-118.582843
...	...	...	...
398	WILSHIRE BLVD	38.959272	-76.486797
323	BEAUMONT AVE	36.915157	-76.300565
386	AIRDROME ST	36.786115	-76.292997
273	RINDGE AVE	45.127347	-76.133867
154	BORDEN AVE	42.128051	-75.980827

397 rows × 3 columns

Ok, now I can use the street names and Lati&Longi data to explore each street.
It seems that I got over 5000 venues of these streets including geoloacations, category and streets they belong to.

```
In [682]: nearby_venues.tail(30)
```

Out[682]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
5232	AVENUE 58	33.906097	-118.010709	La Mirada Disc Golf Course	33.905737	-118.006026	Disc Golf
5233	AVENUE 58	33.906097	-118.010709	Player Development Tennis Academy	33.905463	-118.006316	Tennis Court
5234	AVENUE 58	33.906097	-118.010709	La Mirada Tennis Center	33.905614	-118.005797	Tennis Court
5235	AVENUE 58	33.906097	-118.010709	La Mirada Park	33.903122	-118.007004	Park
5236	EDGEWARE RD	33.949858	-118.009245	Rockies Frozen yogurt	33.950149	-118.010284	Frozen Yogurt Shop
5237	EDGEWARE RD	33.949858	-118.009245	Starbucks	33.948724	-118.007131	Coffee Shop
5238	EDGEWARE RD	33.949858	-118.009245	NORMS Restaurant	33.949949	-118.009490	Diner
5239	EDGEWARE RD	33.949858	-118.009245	Trader Joe's	33.949167	-118.005319	Grocery Store

And now I can do maching learning to divide the 376 streets into several different types with the venues data.

```
In [688]: la_grouped = la_onehot.groupby('Neighborhood').mean().reset_index()
          la_grouped
```

Out[688]:

	Neighborhood	ATM	Accessories Store	Airport	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Water Park	Waterfall	Waterfront	Weight Loss Center	Whisky Bar	Wi Ba
0	001 ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
1	108TH ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
2	110TH ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.027778	0.0	...	0.0	0.0	0.027778	0.0	0.0	0
3	112TH ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
4	118TH ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
371	WYANDOTTE ST	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
372	YOLANDA AVE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
373	ZELZAH AVE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
374	ZITOLA TER	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0
375	ZONAL AVE	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000	0.0	0.0	0

376 rows × 367 columns

I used onehot method to calculate the frequency one type of venues appears in each streets. And then I built a clustering model to cluster these streets by using the data shown above.

```
In [807]: # set number of clusters
          kclusters = 8

          neighborhoods_clustering = la_grouped.drop('Neighborhood',axis=1) #drop the neighborhood column

          # run k-means clustering
          kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(neighborhoods_clustering)

          # check cluster labels generated for each row in the dataframe
          kmeans.labels_[:]

Out[807]: array([5, 5, 5, 1, 1, 5, 5, 5, 2, 2, 2, 5, 5, 5, 5, 5, 2, 1, 5, 3, 5, 5,
                 5, 1, 2, 5, 5, 1, 5, 5, 5, 5, 3, 5, 1, 2, 5, 2, 5, 5, 5, 2, 5, 1,
                 5, 2, 5, 5, 3, 6, 3, 1, 1, 1, 2, 6, 5, 5, 1, 7, 5, 5, 1, 5, 5, 2,
                 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 1, 3, 5, 5, 5, 5, 5, 5, 5, 5, 2, 7,
                 5, 5, 5, 2, 2, 5, 3, 5, 6, 5, 5, 2, 1, 5, 1, 7, 1, 3, 5, 5, 1, 5,
                 5, 5, 5, 5, 5, 5, 1, 5, 5, 5, 5, 6, 5, 5, 5, 1, 5, 5, 2, 1, 5, 5,
                 5, 5, 5, 5, 1, 3, 5, 5, 5, 0, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 1, 5,
                 5, 5, 5, 5, 5, 5, 4, 5, 2, 5, 5, 5, 5, 1, 0, 5, 5, 5, 0, 5, 2, 2,
                 5, 5, 2, 5, 3, 5, 3, 2, 5, 5, 2, 5, 5, 5, 2, 2, 5, 5, 5, 5, 1, 5,
                 5, 5, 5, 5, 7, 5, 5, 5, 1, 6, 1, 5, 5, 5, 5, 1, 2, 2, 5, 5, 5, 5,
                 5, 5, 3, 5, 5, 5, 3, 1, 5, 5, 1, 5, 1, 5, 5, 5, 5, 1, 5, 5, 5, 3,
                 5, 5, 5, 5, 2, 5, 4, 5, 5, 5, 3, 5, 2, 2, 5, 5, 5, 5, 5, 3, 5, 5,
                 2, 5, 5, 5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 1, 5, 3, 5, 5, 5, 1, 5, 5,
                 5, 5, 1, 2, 2, 5, 6, 5, 3, 5, 5, 5, 3, 1, 5, 5, 2, 5, 5, 5, 5, 5,
                 1, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 5, 5, 0, 5, 5, 5, 5, 5, 5, 5, 2,
                 5, 2, 2, 5, 5, 5, 5, 5, 2, 2, 5, 5, 5, 5, 5, 3, 3, 3, 5, 5, 5, 5,
                 5, 5, 5, 5, 5, 5, 5, 5, 7, 5, 5, 5, 5, 5, 5, 5, 2, 2, 5, 2, 2, 2,
                 5, 5], dtype=int32)
```

After this, I got a table that tells me the top5 common places of a street.

	Street	Lati	Longi	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	ALAMEDA ST	34.286918	-118.711384	5.0	Coffee Shop	Grocery Store	Pharmacy	Sandwich Place	Liquor Store
2	JUSTICE ST	34.186423	-118.655813	5.0	Tennis Court	Sports Bar	Performing Arts Venue	Park	Spa
3	CAPISTRANO AVE	34.220179	-118.651933	5.0	Miscellaneous Shop	Yoga Studio	Flower Shop	Farmers Market	Fast Food Restaurant
5	HATTERAS ST	34.227583	-118.646692	5.0	Tea Room	Yoga Studio	Flower Shop	Farm	Farmers Market
6	SYLMAR AVE	34.189901	-118.644897	5.0	Fast Food Restaurant	Pizza Place	Mexican Restaurant	Chinese Restaurant	Coffee Shop
...	...	...	...	...	...	...	...	...	...
379	ALVARADO ST	33.768715	-118.189399	5.0	Brewery	Plaza	Bar	Pizza Place	Coffee Shop
381	SANTEE ST	33.773538	-118.189380	5.0	Pharmacy	Mexican Restaurant	Bakery	Pizza Place	Clothing Store
383	DOBBS ST	34.085986	-118.182518	5.0	Mexican Restaurant	Convenience Store	BBQ Joint	South American Restaurant	Asian Restaurant
394	OCEAN FRONT WALK	34.160964	-118.054681	5.0	Coffee Shop	Mexican Restaurant	American Restaurant	Cosmetics Shop	Bank
396	EDGEWARE RD	33.949858	-118.009245	5.0	Rental Car Location	Mexican Restaurant	Japanese Restaurant	Park	Gift Shop

256 rows × 9 columns

Next, we could search existing Chinese Restaurant of LA. Finally, I got 27 useful samples. Because only 14 venues of the 27 have been given tips, we can see the top 14 Chinses Restaurant. I firmly believed that the more tips a restaurant get, the better it is.

Of course, I must say that during this process, I must have emitted so much other good Chinese restaurants cause many of them do not use ‘noodle’ or ‘hotpot’ words.

Now, I sorted the result table so we can see the top1 has got 208 tips. That’s amazing.

	restaurant	categories	lat	lng	id	tips
0	Randy's Donuts	Donut Shop	33.9618	-118.37	4bafd591f964a52088243ce3	208
1	88 Chinese & Sushi	Chinese Restaurant	34.1739	-118.466	4b5fba41f964a52033ca29e3	15
2	Super Wok	Chinese Restaurant	34.1695	-118.536	4c1adf3ab9f876b0e2777946	11
3	Coral Reef Chinese Restaurant	Chinese Restaurant	34.1176	-118.261	3fd66200f964a520aeee1ee3	11
4	Coral Reef Chinese Restaurant	Chinese Restaurant	34.1176	-118.261	3fd66200f964a520aeee1ee3	11
5	Soybean Chinese Korean Restaurant	Chinese Restaurant	34.0454	-118.465	4b01d1e7f964a520b34522e3	5
6	Mom's Donuts & Chinese Food	Donut Shop	34.0788	-118.279	4b131df6f964a520529423e3	5
7	China Way	Chinese Restaurant	34.0912	-118.291	4bd9de712a3a0f47ba9ea8b6	4
8	H.H. China Express Chinese Food	Fast Food Restaurant	33.7787	-118.192	4dbb1ab24b222080d36f7e49	3
9	Chinese Taste	Asian Restaurant	34.0502	-118.218	4b8c64dbf964a5208bce32e3	2
10	Hong 2 Express Chinese Food Louisiana Fried Ch...	Chinese Restaurant	33.9168	-118.282	513952e2e4b0cff955ca9d14	1
11	Tangerine	Chinese Restaurant	34.05	-118.242	51840930498ef5dc7b535e35	1
12	Chinese Deli	Chinese Restaurant	33.8085	-118.265	4c5b67272815c9281e7daf67	1
13	Little Beijing Chinese Food Donuts	Chinese Restaurant	34.0812	-118.178	4edd69b28b8173160fefdb72	1
14	Abc express chinese food	Asian Restaurant	34.1812	-118.536	4eb34c90cc2143e8271159b6	0
15	Bamboo Chinese	Chinese Restaurant	34.0834	-118.305	59c80a372079556b588a1aae	0

Afterwards, I wrangled the table further and concated more information into that table such as cluster label and street name so we can get a better understanding of the interrelationships.

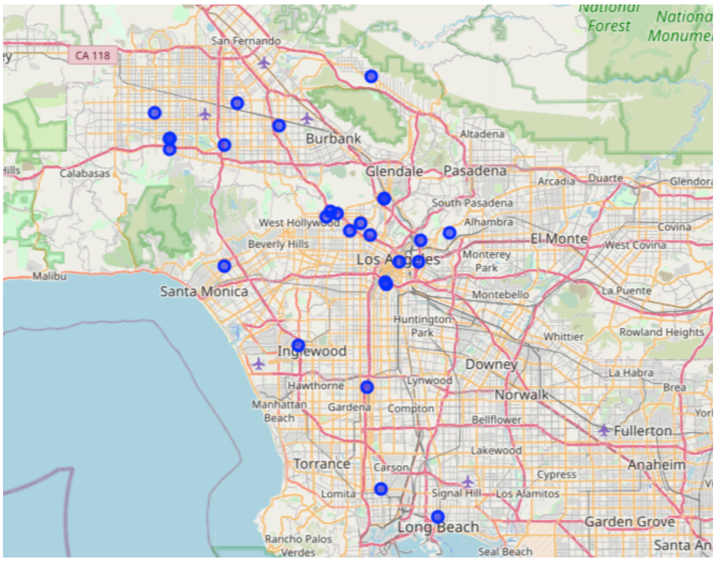
	restaurant	categories	lat	lng	id	tips	rating	street	cluster
0	Randy's Donuts	Donut Shop	33.961813	-118.370479	4bafd591f964a52088243ce3	208	8.5	98TH ST	1.0
1	88 Chinese & Sushi	Chinese Restaurant	34.173906	-118.466181	4b5fba41f964a52033ca29e3	15	6.6	LENNOX AVE	5.0
2	Super Wok	Chinese Restaurant	34.169480	-118.535642	4c1adf3ab9f876b0e2777946	11	7.7	BLYTHE ST	5.0
3	Coral Reef Chinese Restaurant	Chinese Restaurant	34.117639	-118.261024	3fd66200f964a520aeee1ee3	11	7.5	ECHO PARK AVE	5.0
4	Coral Reef Chinese Restaurant	Chinese Restaurant	34.117639	-118.261024	3fd66200f964a520aeee1ee3	11	7.5	ECHO PARK AVE	5.0
5	Soybean Chinese Korean Restaurant	Chinese Restaurant	34.045385	-118.465488	4b01d1e7f964a520b34522e3	5	0.0	MIDVALE AVE	5.0
6	Mom's Donuts & Chinese Food	Donut Shop	34.078782	-118.278669	4b131df6f964a520529423e3	5	0.0	BULLARD AVE	5.0
7	China Way	Chinese Restaurant	34.091158	-118.291349	4bd9de712a3a0f47ba9ea8b6	4	0.0	25TH ST	5.0
8	H.H. China Express Chinese Food	Fast Food Restaurant	33.778663	-118.192067	4dbb1ab24b222080d36f7e49	3	0.0	WESTMONT DR	0.0
9	Chinese Taste	Asian Restaurant	34.050243	-118.217572	4b8c64dbf964a5208bce32e3	2	0.0	4TH PL	1.0
10	Hong 2 Express Chinese Food Louisiana Fried Ch...	Chinese Restaurant	33.916839	-118.282376	513952e2e4b0cff955ca9d14	1	0.0	BARING CROSS ST	2.0

Here are my data wrangling process.

In the part4, we can analyze further by visualizing them.

4. Data Analysis and Visualization

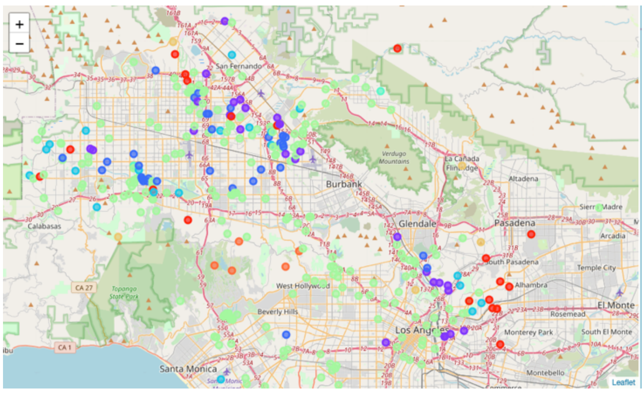
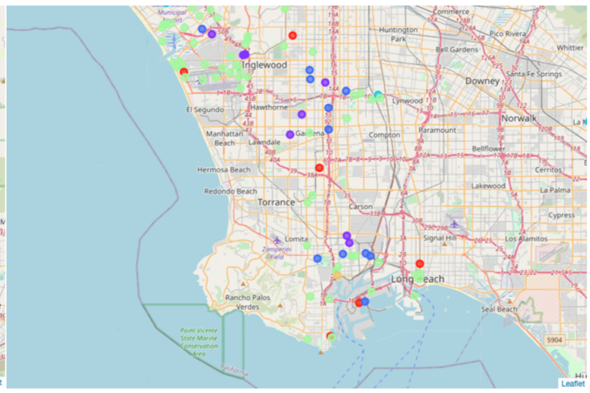
Firstly, I visualize all the restaurant spot I got.



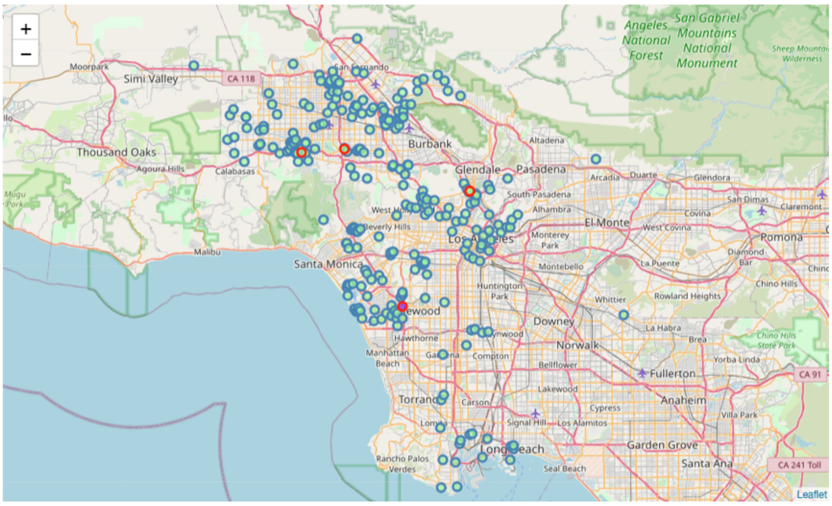
The image does not indicate any pattern but covers quite many region of LA.

Then, I used the clustered venues data of different streets to show the different streets.

From the image, we can know that the shallow green dots represent the most common clusters, which is cluster 5.

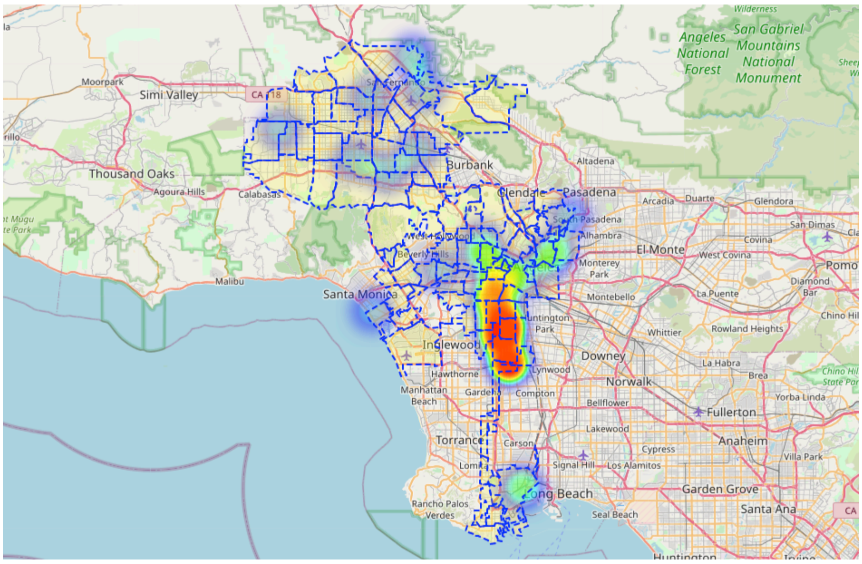
After this, Let me add the top 5 popular Chinses restaurant to the map.



OK, now the map gives us all the streets that are similar to the streets where the top5 Chinses restaurant reside in. From this map, we can know, after machine learning process, we can locate our restaurant in these blue dots area and we can clearly get far from the street where the top5 live in.

But, that’ s not enough. We need to take other factors into our considerations.

In fact, there are many more factors influencing this topic in the real world so the process should have been much more complicated. But, just like what I said, our meta data quality is not satisfying and I need to simplify this model.
Therefore, our task is to take crime rate and population as factors.



I draw a crime information data into this map. The darker the area is , the higher the crime rate is. This tells us that maybe locating our restaurant in the north of city is a better idea than in the center.

Through this process, we can actually add more factors into our considering model.

Finally, we can get our conclusion by our analysis:
Crime Rata in central Los Angeles is not ideal. At the same time, because we need to avoid several strong competitors, **Long beach, Burbank, San Fernando** in *Northwest and South Los Angeles* looks like a good choice!

So, this is my project. Thanks for watching!