Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

This submission will eventually become your Data section in your final report. So I recommend that you push the report (having your Data section) to your Github repository and submit a link to it.

1. **Data Collection**

To search for the best location to launch a new Chinese restaurant in the Los Angeles area, we need to collect two sets of data: The Chinese American distribution by City, and the Chinese restaurant spreads by City.

**3.1 Top Cities with Chinese-American in Los Angeles Area**

Ethical polulation data can be found in the national census web site as well as the institutions studying the ethical developments. Wikimedia is another source to summarize the population migrations and growth. We have exaimed all above sources.

The national census web site (https://www.census.gov/) contains many details of population data for in depth study. One can search for populaiton spread by age, ethics, region, immigration origin, etc. The tools on that website is very useful, especially the “Data via FTP” tool, making it very convenient to derive data. Even thouhg, depending the question, one may have to check into many data tables to put together a full set of information.

Immigration focused agencies such as MPI (Migration Policy Institude) can provide many meaningful insights into the ethical issues. They also collect data in their study areas, and summarize the information to provide guidance to policy makers. With MPI as an example, they study the i[mmigration over time](https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#history), [immigrants' countries and regions of birth](https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#source), [educational attainment](https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#education), etc….Had one found relevant study articles, one can quickly gain deep insights into the problems and seek for possible solutions.

Unfortunately, for our case, where we need to derive the data by city for a specific ethics, above sites have not delivered handy details to advance our study. I further searched the Wikimedia, and luckily found out the exact data with Chinese American populationvby City in the Los Angeles area, as shown in the table 1 in the 2.2 section of this article. Not only the population is presented, but also the penetration rate is listed for added insights.

It shows that the majority of the Chinese-American population of more than 400,000 in [Los Angeles County](https://en.wikipedia.org/wiki/Los_Angeles_County) lives within the [San Gabriel Valley](https://en.wikipedia.org/wiki/San_Gabriel_Valley), which is particularly noted for cities that have a large Chinese-American ethnic plurality. There are 5 cities having Chinese American population exceeds 40% of total population, including Arcadia (46%), Monterey Park (45%), Temple City (44%), San Marino (43%), and San Gabriel (42%). Alhambra has most population at 30,683 Chinese American. All those cities are potential locations for our new business.

**3.2 Chinese Restauran Business Spread**

We leverage the Foursquare to acquire the Chinese and Asian restaturant information. Foursquare is a location technology platform dedicated to provide location information for people moving through the real world. It is the location engin supporting many mobile apps to allow people to conduct data-driven media planning. The venue based search can help people quickly drill into the targeting dataset.

We will need the location information, such as the lantitude and longtitude, to execute the Foursquare searches. Geopy library will be used to retrive the data. More specificlaly, we use the geolocator to derive data. One has to be a bit patient in case it “times out”.

With the pre-acquired Client ID and Client Secret codes, I can leverage the geolocator provided locaiton data to retrive the top venue info through Foursquare for each of the top cities in Los Angeles area with high concentration of Chinese Americans.

Some data processing will be needed to clense data. Out of 185 venue obtained, we first group the top venue categories for the selected cities. Attention is eslecially paid to the Chinese restaurant and other Asian restaurants, for the latter is also a key factor affecting our locaiton decision. The final feature list includde in the study are: Chinese restaurant, Korean restaurant, Sushi restaurant, Vietname restaurant, and Indian restaurant. Based on those Asian restaurant spreads, one can use the cluster technique to group the Cities into categories to narrow down our location selection.

We use the Geo visulization to help further fine tune the location decision. We will avoid the areas that having multiple neighboring cities with high dendity of Aisan restaurantr, and look into other factores, such as overall population, etc… to identify the optimized city.

By combining the ethical poluation spread and competitive business distribution, we will derive a shortlist of citys in the Los Angeles area to launch a new Chinese restaurant business.

Appendix:

<https://www.census.gov/>

https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#source

<https://en.wikipedia.org/wiki/List_of_U.S._cities_with_significant_Chinese-American_populations>

<https://www.cnbc.com/2016/01/20/heres-the-real-reason-why-most-restaurants-fail.html>

<https://www.thebalancesmb.com/choosing-a-location-for-your-restaurant-2888635>

<https://www.latimes.com/food/la-fo-modern-chinese-food-california-20190411-story.html>

<https://nypost.com/2016/01/25/the-17-best-chinese-restaurants-in-america/>