Analyze the correlation between venues and populations in Bay Area

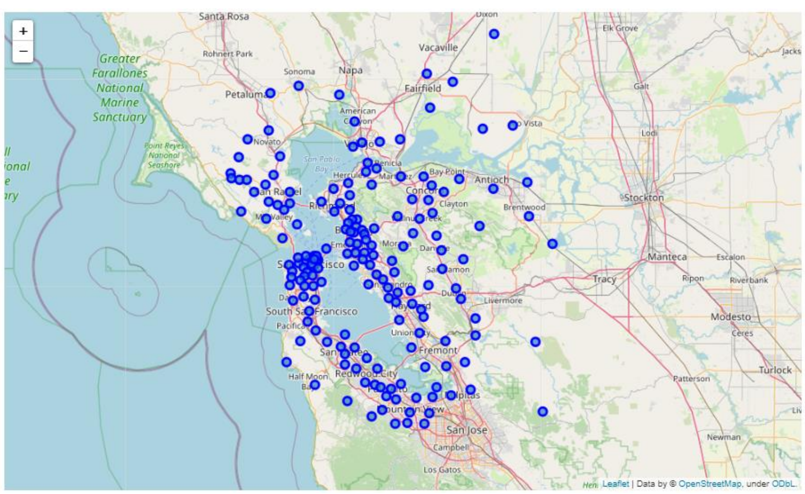
1. Introduction

This project is the final capstone project for the course Applied Data Science Capstone which is the last course of the IBM Data Science Professional Certificate program. In this project, I will use the data science skills and tools learned from the series of courses to use location data to explore a geographical location. The goal is to prove that I am capable of defining an idea or problem, looking for data in the web, using the Foursquare location data to explore different areas, effectively analyzing the data, and presenting the results.

1.1 Description of the problem and a discussion of the background

I live in a suburban city in the grand Houston area. My neighborhood is very diverse, including a large amount resident originally from China. However, when I am tired of cooking and want to dine in a Chinese restaurant, I have very limited choices in the nearby neighborhood. I have to drive 15-20 minutes (without traffic) away to get one. Based on the information from my neighbors and friends, there is actually a demand to the traditional Chinese food. Since I am equipped with skills and tools from this series courses, an idea to my mind is to explore the relation between the number of Chinese restaurants and the population distribution in an area. This idea can be used to explore the similar relation between other venues and the population distribution.

Bay Area zip codes distribution.



At the beginning I used Foursquare to search the venues information in my city but found out the search didn't work well since some restaurants I know well are not searched out. I choose to explore the San Francisco Bay Area because it is a metropolitan region with 7 million inhabitants and more than 100 cities/town in the 7000 square miles of land. The high-density population supports a sufficient number of venues to use in this project. In addition, I don't have a chance to visit the Bay Area yet and it is in my visit list in near future, so I'd like to get some impression of it in advance.

1.2 Target audience

Some data scientists might be interested in this project since it includes some real-life data and exploratory data analysis methods, and the data is able to tell a story out of it. Business personnel who want to invest a venue such as a Chinese restaurant, might be interested in this project too since its analysis results provide a comprehensive understanding on the relation between the number of a venue in an area and the population distribution. Some analysis is kind of easy to understand, such as that the number of Chinese restaurants is positively correlated to the Chinese population. Some others might be surprising.

2. Data Preparation

The goal of these data is to get a data set consisting of the venue's information and population information. For this purpose, four sets of data are used in this project as below.

2.1 Bay Area ZIP Codes from https://catalog.data.gov/dataset/bay-area-zip-codes.
The original data was downloaded as a csv file. It includes zip codes, city names, geometry locations, area information in the Bay Area as shown below. While this data set does not have the latitude and longitude information of the zip codes, the data set 2.2 has the latitude and longitude information but it includes all zip codes in California. So, this data set is used to limit the zip codes within the Bay Area.

This data set is uploaded in my github: https://github.com/xwangphy/Coursera-Applied_Data_Science_Capstone/blob/main/bayarea_zipcodes.csv

Here is an example, the first 5 rows of this data set:

```
path='D:/bayarea_zipcodes.csv'
bay_geom=pd.read_csv(path)
bay_geom.head()
```

	PO_NAME	the_geom	ZIP	STATE	Area__	Length__
0	NAPA	MULTIPOLYGON (((-122.10329200180091 38.5132829...	94558	CA	1.231326e+10	995176.225313
1	FAIRFIELD	MULTIPOLYGON (((-121.947475002335 38.301511000...	94533	CA	9.917861e+08	200772.556587
2	DIXON	MULTIPOLYGON (((-121.65335500334429 38.3133870...	95620	CA	7.236950e+09	441860.201400
3	SONOMA	MULTIPOLYGON (((-122.406843003057 38.155681999...	95476	CA	3.001414e+09	311318.546326
4	NAPA	MULTIPOLYGON (((-122.29368500225117 38.1552379...	94559	CA	1.194302e+09	359104.646602

Here only the zip codes, city names, and area of the data are used in this project. With renaming of the columns, a data frame named bay_geom is extracted from the original data as:

```
bay_geom=bay_geom[['ZIP','PO_NAME','Area__']]
bay_geom=bay_geom.rename(columns={'ZIP':'ZipCode','PO_NAME':'City','Area__':'Area'})
bay_geom.head()
```

	ZipCode	City	Area
0	94558	NAPA	1.231326e+10
1	94533	FAIRFIELD	9.917861e+08
2	95620	DIXON	7.236950e+09
3	95476	SONOMA	3.001414e+09
4	94559	NAPA	1.194302e+09

2.2 Scraping the population distribution of California from http://zipatlas.com/us/ca/zip-code-comparison/population-density.htm.

This data includes zip codes, location (latitude, longitude), city, population, people/ sq. mile, national rank data of all the zip codes in California. I need all the data except the national rank.

Here is an example, the first 5 rows of this data set:

```python
urls=['http://zipatlas.com/us/ca/zip-code-comparison/population-density.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.2.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.3.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.4.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.5.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.6.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.7.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.8.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.9.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.10.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.11.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.12.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.13.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.14.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.15.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.16.htm',
      'http://zipatlas.com/us/ca/zip-code-comparison/population-density.17.htm'
     ]

cadata=pd.DataFrame()
for url in urls:
    cadata=cadata.append(pd.read_html(url,header=0)[10].drop(['#'], axis=1), ignore_index=True)

cadata=cadata.rename(columns={'Zip Code':'ZipCode'})
cadata.head()
```

	ZipCode	Location	City	Population	People / Sq. Mile	National Rank
0	94108	37.791998, -122.408653	San Francisco, California	13716	53134.47	#48
1	90057	34.061918, -118.277939	Los Angeles, California	43986	49226.28	#54
2	94109	37.794487, -122.422270	San Francisco, California	56322	46521.46	#64
3	94102	37.779500, -122.419233	San Francisco, California	28991	44719.24	#71
4	94133	37.802071, -122.411004	San Francisco, California	26827	40117.97	#77

This set of original data need to be cleaned and refined, and then merged with the data set 2.1 to get a dataframe named bay_data, which include 'ZipCode', 'City', 'Latitude', 'Longitude', 'Area', 'Population', 'People / Sq. Mile' data within the Bay Area.

```python
bay_data=pd.merge(bay_geom.drop('City', axis=1), cadata, on='ZipCode')
bay_data['Latitude']=bay_data['Location'].apply(lambda x:float(x.split(', ')[0]))
bay_data['Longitude']=bay_data['Location'].apply(lambda x:float(x.split(', ')[1]))
bay_data=bay_data[['ZipCode','City','Latitude','Longitude','Area','Population','People / Sq. Mile']]
bay_data.head()
```

	ZipCode	City	Latitude	Longitude	Area	Population	People / Sq. Mile
0	94558	Napa, California	38.489789	-122.270110	1.231326e+10	63932	155.68
1	94533	Fairfield, California	38.287136	-122.027110	9.917861e+08	77666	2290.55
2	95620	Dixon, California	38.392821	-121.799917	7.236950e+09	18510	91.81
3	95476	Sonoma, California	38.254850	-122.461799	3.001414e+09	34310	309.44
4	94559	Napa, California	38.232389	-122.324944	1.194302e+09	26891	800.91

2.3 Download California population by race from the State of California Department of Finance, https://www.dof.ca.gov/Reports/Demographic_Reports/Census_2010/#SF1

This data set is uploaded in my github: https://github.com/xwangphy/Coursera-Applied_Data_Science_Capstone/blob/main/2010Census_DemoProfile3a_ZCTA.xls

This data set consists of the population by race in all zip codes of California.

Here is an example of the original data set:

```
path2='D:/2010Census_DemoProfile3a_ZCTA.xls'
population=pd.read_excel(path2,sheet_name='2010_3a',header=8)
population.columns=['ZipCode','Total Population','Total Population of One Race','White','Black or African American','American Inc
population
```

	ZipCode	Total Population	Total Population of One Race	White	Black or African American	American Indian and Alaska Native	Total Asian	Asian Indian	Chinese	Filipino	Japanese	Korean	Vietnamese	Otl Asi
0	California	37253956.0	35438572.0	21453934.0	2299072.0	362801.0	4861007.0	528176.0	1253102.0	1195580.0	272528.0	451892.0	581946.0	57778
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
2	ZCTA5 89010 (California part only)	31.0	31.0	18.0	0.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	ZCTA5 89019 (California part only)	69.0	64.0	55.0	6.0	0.0	2.0	0.0	0.0	0.0	0.0	2.0	0.0	
4	ZCTA5 89060 (California part only)	30.0	28.0	27.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
5	ZCTA5 89061 (California part only)	51.0	44.0	42.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
6	ZCTA5 89439 (California part only)	80.0	79.0	78.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	
7	ZCTA5 90001	57110.0	54985.0	20509.0	6133.0	410.0	145.0	24.0	19.0	38.0	14.0	6.0	4.0	4
8	ZCTA5 90002	51223.0	49284.0	14392.0	13101.0	371.0	148.0	23.0	21.0	43.0	17.0	11.0	16.0	1
9	ZCTA5 90003	66266.0	63750.0	19878.0	16181.0	516.0	149.0	44.0	1.0	70.0	3.0	6.0	12.0	1

Here we can see that this set of data has zip code information, population by race. But the data need to be cleaned, refined, and merged with data set 2.2 to get a dataframe named population, which include detailed population distribution in each zip code in the Bay Area, as well as the location data. Below is an example of the final population dataframe I got. It includes all the zip codes in the Bay Area, the corresponding city, latitude, longitude, area, people/sq. mile, total population, and all the population by races.

```
population=population.drop(['Total Population of One Race','Total Asian','Total NHOPI','Total Population of Two or More Races'],
population=population.drop([0,1,2,3,4,5,6,1770],axis=0)
population=population.reset_index(drop=True)
population['ZipCode']=population['ZipCode'].apply(lambda x:int(x[6:]))
population=pd.merge(bay_data.drop(['Population'], axis=1), population, on='ZipCode')
population.head()
```

	ZipCode	City	Latitude	Longitude	Area	People / Sq. Mile	Total Population	White	Black or African American	American Indian and Alaska Native	Asian Indian	Chinese	Filipino	Japanese	Korean
0	94558	Napa, California	38.489789	-122.270110	1.231326e+10	155.68	66830.0	51624.0	654.0	491.0	133.0	286.0	478.0	308.0	152.0
1	94533	Fairfield, California	38.287136	-122.027110	9.917861e+08	2290.55	69277.0	29074.0	12261.0	660.0	802.0	540.0	5408.0	419.0	233.0
2	95620	Dixon, California	38.392821	-121.799917	7.236950e+09	91.81	20553.0	14581.0	578.0	204.0	84.0	110.0	314.0	75.0	28.0
3	95476	Sonoma, California	38.254850	-122.461799	3.001414e+09	309.44	35394.0	28449.0	178.0	253.0	56.0	138.0	178.0	105.0	47.0
4	94559	Napa, California	38.232389	-122.324944	1.194302e+09	800.91	27184.0	20256.0	189.0	265.0	78.0	97.0	171.0	75.0	38.0

2.4 Foursquare location data.

Foursquare is a location technology platform offering business solutions and consumer products through a deep understanding of location. It gives the venues information in Bay Area including venue category, geographic location, etc.

Since I will study the venue distribution in different zip code, I need to extract the venue names, venue latitude, venue longitude, and venue category information from Foursquare. I borrowed the function getNearbyVenues() from the Upgraded External Tool lab and modified to extract the features I need.

```
def getNearbyVenues(postalcodes,names, latitudes, longitudes, radius=2000):

    venues_list=[]
    for postalcode, name, lat, lng in zip(postalcodes,names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.forr
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            postalcode,
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['ZipCode',
                  'City',
                  'City Latitude',
                  'City Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Then write the code to run the above function on each postal code/borough and create a new dataframe called bay_venues. There are a total of 10067 venues included.

Here is an example, the first 5 venues in Bay Area from Foursquare:

```
zipcodes=bay_data['ZipCode']
names=bay_data['City']
latitudes=bay_data['Latitude']
longitudes=bay_data['Longitude']

bay_venues=getNearbyVenues(zipcodes, names, latitudes, longitudes, 2000)
#bay_venues.head()
```

```
San Jose, California
Alviso, California
Redwood City, California
Sunnyvale, California
Palo Alto, California
Menlo Park, California
Palo Alto, California
Milpitas, California
Mount Hamilton, California
Redwood City, California
Mountain View, California
Palo Alto, California
Stanford, California
Palo Alto, California
Portola Valley, California
Mountain View, California
Los Altos, California
Sunnyvale, California
Los Altos, California
Sunnyvale, California
```

```
bay_venues.head()
```

	ZipCode	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	94558	Napa, California	38.489789	-122.27011	Turtle Rock	38.494408	-122.251862	Bar
1	94558	Napa, California	38.489789	-122.27011	Brown Estate	38.505196	-122.276495	Winery
2	94558	Napa, California	38.489789	-122.27011	Turtle Rock Bar & Cafe	38.496575	-122.250497	Bar
3	94533	Fairfield, California	38.287136	-122.02711	In-Shape Health Clubs	38.284730	-122.025750	Gym / Fitness Center
4	94533	Fairfield, California	38.287136	-122.02711	Raley's	38.289158	-122.033253	Grocery Store

```
len(bay_venues['ZipCode'])
```

```
10072
```

Here, one thing needed to be pointed out is that the requests sent to Foursquare include 179 zip codes. However, the returned venue data only includes 174 zip code. This means either Foursquare is lack of some venue data or there is no venue in some zip codes (94512, 94550, 94571, 95140, 95620).

```
bay_data['ZipCode'].value_counts().sum()
```

```
179
```

```
len(bay_venues['ZipCode'].unique())
```

```
174
```

```
set(bay_data['ZipCode'])-set(bay_venues['ZipCode'].unique())
```

```
{94512, 94550, 94571, 95140, 95620}
```

3. Data Analysis

3.1 Cluster zip codes

In this part, I will cluster the zip codes based on the venues (category and quantity) in each zip code. The goal is to find out whether the Chinese restaurants are located in specific zip code group.

The data set bay_venues from 2.4 will be used for the clustering. This data set contains 10072 venues and these venues belong to 439 categories.

```
len(bay_venues['Venue Category'].unique())
```

```
439
```

3.1.1 One hot encoding

One hot encoding technique is used to quantify each venue.

```
# one hot encoding
bay_onehot = pd.get_dummies(bay_venues[['Venue Category']], prefix="", prefix_sep="")

# add postalcode column back to dataframe
bay_onehot['ZipCode'] = bay_venues['ZipCode']

# move postal code column to the first column
fixed_columns = [bay_onehot.columns[-1]] + list(bay_onehot.columns[:-1])
bay_onehot = bay_onehot[fixed_columns]

bay_onehot.head()
```

	ZipCode	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Alternative Healer	American Restaurant	Amphitheater	Andhra Restaurant	Animal Shelter	Antiqu Sho
0	94558	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	94558	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	94558	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	94533	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	94533	0	0	0	0	0	0	0	0	0	0	0	0	0	

```
bay_onehot.shape
```

```
(10072, 440)
```

Then group the venues by zip code and by taking the mean of the frequency of occurrence of each category to get a data set called bay_group. In this way we can get the relative occurrence frequency of the venues in each zip code.

```
bay_grouped = bay_onehot.groupby('ZipCode').mean().reset_index()
bay_grouped.head()
```

	ZipCode	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Alternative Healer	American Restaurant	Amphitheater	Andhra Restaurant	Animal Shelter	Antiqu Sho
0	94002	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.00	0.0	0
1	94005	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.00	0.0	0
2	94010	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.041667	0.0	0.00	0.0	0
3	94014	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.00	0.0	0
4	94015	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.01	0.0	0

After removing the ZipCode column from this data set, another data set called bay_grouped_clustering is generated for the clustering. The unsupervised clustering method, KMeans will be used here.

3.1.2 Find optimal k

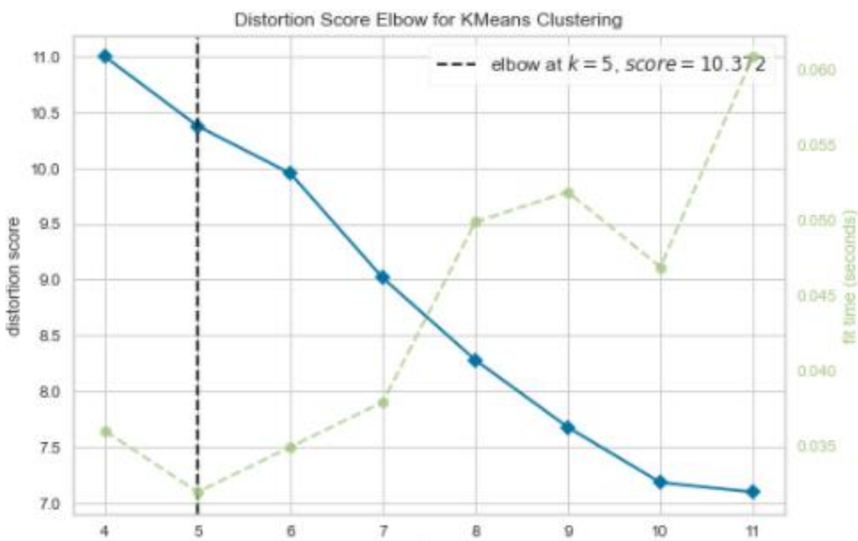
For KMeans, the number of groups need to be assigned first. So, in order to find the optimal k for KMeans, I used the yellowbrick package to visualize the Elbow curve.

```
! pip install yellowbrick

# Import ElbowVisualizer
from yellowbrick.cluster import KElbowVisualizer

model = KMeans()
# k is range of number of clusters.
visualizer = KElbowVisualizer(model, k=(4,12), metric='distortion', timings=True)

visualizer.fit(bay_grouped_clustering)        # Fit the data to the visualizer
visualizer.show()
# Finalize and render the figure
```



As a result, k=5 is the optimal k for my data set.

3.2.3 KMeans clustering

```
# set number of clusters
kclusters = 5

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([0, 1, 1, 1, 1, 1, 0, 1, 1, 1])
```

Kmeans clustering generates the cluster labels for each row in the dataframe.

The clustering results shows most (152 of 174) of the zip codes belong to cluster 1, 19 of 174 belong to cluster 0, cluster 2, 3, or 4 all has only 1 of 174 of the zip codes.

Bay Area Zipcodes Clustering

	Cluster Labels	Counts
0	1	152
1	0	19
2	4	1
3	3	1
4	2	1

3.2.4 Display the top 10 venues for zip code

To make it easier to find the features of each clustering group, I created a new dataframe called zipcodes_venues_sorted to display the top 10 venues for each zip code.

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['ZipCode']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
zipcodes_venues_sorted = pd.DataFrame(columns=columns)
zipcodes_venues_sorted['ZipCode'] = bay_grouped['ZipCode']

for ind in np.arange(bay_grouped.shape[0]):
    zipcodes_venues_sorted.iloc[ind, 1:] = return_most_common_venues(bay_grouped.iloc[ind, :], num_top_venues)

zipcodes_venues_sorted.head()
```

	ZipCode	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	94002	Trail	Scenic Lookout	Theater	Sandwich Place	Grocery Store	Park	Salon / Barbershop	Chinese Restaurant	Bagel Shop	Gas Station
1	94005	Harbor / Marina	Food Truck	Vietnamese Restaurant	Intersection	Mexican Restaurant	Chinese Restaurant	Donut Shop	Train Station	Coffee Shop	Trail
2	94010	Coffee Shop	Breakfast Spot	Italian Restaurant	Pizza Place	American Restaurant	Clothing Store	Ice Cream Shop	Grocery Store	Historic Site	Trail
3	94014	Trail	Fast Food Restaurant	Chinese Restaurant	Grocery Store	Playground	Flower Shop	Rental Car Location	Filipino Restaurant	Golf Course	Elementary School
4	94015	Clothing Store	Bakery	Sandwich Place	Cosmetics Shop	Filipino Restaurant	Gym / Fitness Center	Discount Store	Supermarket	Supplement Shop	Pharmacy

3.2.5 Merge clustering labels with zip codes

By merging the clustering labels, the location data, and top 10 venues in each zip code, a detailed dataframe called bay_merged is generated for convenient analysis.

```
# add clustering labels
zipcodes_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
zipcodes_venues_sorted.head()
```

```
# merge bay_grouped with bay_data to add latitude/longitude for each zip code
bay_merged = bay_data.join(zipcodes_venues_sorted.set_index('ZipCode'), on='ZipCode',how='inner')

bay_merged # check the last columns!
```

	ZipCode	City	Latitude	Longitude	Area	Population	People / Sq. Mile	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	94558	Napa, California	38.489789	-122.270110	1.231326e+10	63932	155.68	4	Bar	Winery	Zoo Exhibit	Fish & Chips Shop	Eye Doctor	Fa
1	94533	Fairfield, California	38.287136	-122.027110	9.917861e+08	77666	2290.55	1	Fast Food Restaurant	Pizza Place	Pharmacy	Grocery Store	Video Store	F
3	95476	Sonoma, California	38.254850	-122.461799	3.001414e+09	34310	309.44	1	Garden Center	Farm	Wine Shop	Beer Garden	Event Space	Z
4	94559	Napa, California	38.232389	-122.324944	1.194302e+09	26891	800.91	1	Vineyard	Winery	Food	Harbor / Marina	Boat or Ferry	
5	94954	Petaluma, California	38.235021	-122.557332	2.006544e+09	35400	540.11	1	Farm	Vineyard	Zoo Exhibit	Ethiopian Restaurant	Exhibit	E
7	94535	Travis Afb, California	38.265899	-121.939461	3.029397e+08	9966	1131.27	1	Sandwich Place	Fast Food Restaurant	Pizza Place	Fried Chicken Joint	Convenience Store	D

3.2.6 Visualize the resulting clusters

Folium mapping is used to visualize the resulting clusters. The different colors of the dots indicate different clusters: red (cluster 0), purple (cluster 1), blue (cluster 2), green (cluster 3), and orange (cluster 4).

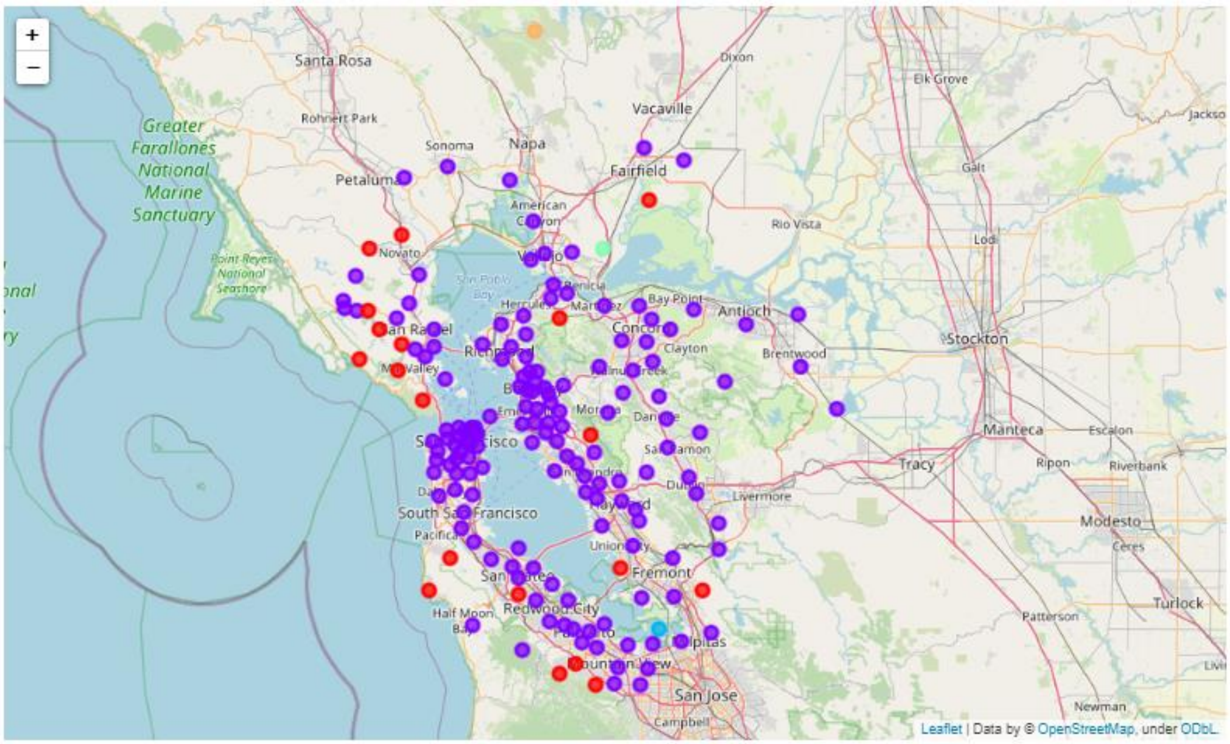
```python
# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(bay_merged['Latitude'], bay_merged['Longitude'], bay_merged['City'], bay_merged['Cluster Labels
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```



From the Folium map, we can easily see that the majority of the zip codes belong to cluster 1, and only very few zip codes belong to cluster 2, 3 or 4.

3.2.7 Examine clusters

The category features are quite distinguishing for each cluster. In cluster 0, trail is the 1st most common venue for each zip code. So, Cluster 0 is an outdoor living zone. Cluster 1 is more like a restaurant zone. Cluster 2, 3,4 are some special zones.

```
bay_merged.sort_values(by='Cluster Labels')
```

	ZipCode	City	Latitude	Longitude	Area	Population	People / Sq. Mile	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	
135	94044	Pacifica, California	37.573633	-122.455167	4.027890e+08	38869	889.83	0	Trail	Park	Gym / Fitness Center	Gym	Inf
28	94547	Hercules, California	37.991259	-122.214945	1.607516e+08	22460	1152.00	0	Trail	Home Service	Café	Golf Driving Range	Zc
67	94970	Stinson Beach, California	37.921137	-122.657562	3.265383e+07	772	132.23	0	Trail	Beach	Lake	Zoo Exhibit	Fis
144	94555	Fremont, California	37.555473	-122.080312	3.345923e+08	33863	2836.78	0	Trail	Intersection	Park	Baseball Field	
37	94930	Fairfax, California	37.971751	-122.611873	7.954870e+08	8476	681.30	0	Trail	Park	Golf Course	Food	
38	94973	Woodacre, California	38.005839	-122.638155	9.758632e+07	1434	709.74	0	Trail	Post Office	Deli / Bodega	Fish Market	Ey
118	94124	San Francisco, California	37.731505	-122.384532	1.369976e+08	33170	6630.97	1	Coffee Shop	Park	Vietnamese Restaurant	Southern / Soul Food Restaurant	
120	94577	San Leandro, California	37.717196	-122.159338	2.247702e+08	41867	5254.35	1	Pizza Place	Mexican Restaurant	Coffee Shop	Burger Joint	Cloth
124	94568	Dublin, California	37.713926	-121.928425	2.198047e+08	29633	2573.26	1	Coffee Shop	Mexican Restaurant	Bakery	Indian Restaurant	Re
125	94578	San Leandro, California	37.704384	-122.126703	1.253342e+08	36565	9709.74	1	Sandwich Place	Pizza Place	Mexican Restaurant	Fast Food Restaurant	Con
126	94015	Daly City, California	37.680844	-122.481310	1.594835e+08	63317	11003.66	1	Clothing Store	Bakery	Filipino Restaurant	Cosmetics Shop	Gym
116	94603	Oakland, California	37.739113	-122.175602	7.798013e+07	31389	11528.22	1	Grocery Store	Hotel	Fast Food Restaurant	Fried Chicken Joint	Con
127	94005	Brisbane, California	37.684694	-122.407120	1.323574e+08	3615	680.80	1	Vietnamese Restaurant	Food Truck	Harbor / Marina	Garden	
160	95002	Alviso, California	37.449537	-121.994813	1.240214e+08	2128	458.40	2	Bike Trail	Zoo Exhibit	Fishing Spot	Fabric Shop	
12	94510	Benicia, California	38.113533	-122.119260	8.567942e+08	25573	784.88	3	Beach	Zoo Exhibit	Fish Market	Fabric Shop	
153	94560	Newark, California	37.504133	-122.032347	5.489937e+08	42471	1340.22	4	Shipping Store	Zoo Exhibit	Ethiopian Restaurant	Exhibit	Ey

4. Results and discussion

Here I use Chinese restaurant as an example to analyze the data. The same method will be extended for other venues.

4.1 Extract the Chinese restaurants in Bay Area

With the combination of the venue data and the clustering result, the detailed data of the Chinese restaurants, named bay_chinese, is extracted. The data includes the location of the venues and the corresponding zip code information.

There is a total of 186 Chinese restaurants from 75 zip codes. Considering that there are 174 zip codes in Bay Area, we can see that either there is no Chinese restaurant in 99 zip codes or Foursquare does not contains Chinese restaurant data from these zip codes.

```
bay_chinese=bay_venues[bay_venues['Venue Category']=='Chinese Restaurant']
bay_chinese=pd.merge(bay_merged[['ZipCode','Cluster Labels']],bay_chinese, on='ZipCode', how='inner' )
bay_chinese['Number of Chinese Restaurant']=1
bay_chinese.sort_values(by=['Cluster Labels'], inplace=True)
bay_chinese.head()
```

	ZipCode	Cluster Labels	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Number of Chinese Restaurant
163	94002	0	Belmont, California	37.509366	-122.306132	Gin Mon Chinese Cuisine	37.510888	-122.293122	Chinese Restaurant	1
0	94533	1	Fairfield, California	38.287136	-122.027110	Panda Express	38.289116	-122.033499	Chinese Restaurant	1
117	94541	1	Hayward, California	37.674002	-122.076796	Panda Express	37.672728	-122.085242	Chinese Restaurant	1
118	94541	1	Hayward, California	37.674002	-122.076796	Great River Resturant	37.680682	-122.085624	Chinese Restaurant	1
119	94542	1	Hayward, California	37.656695	-122.048361	China Best	37.655671	-122.048794	Chinese Restaurant	1

```
bay_chinese['ZipCode'].shape
```

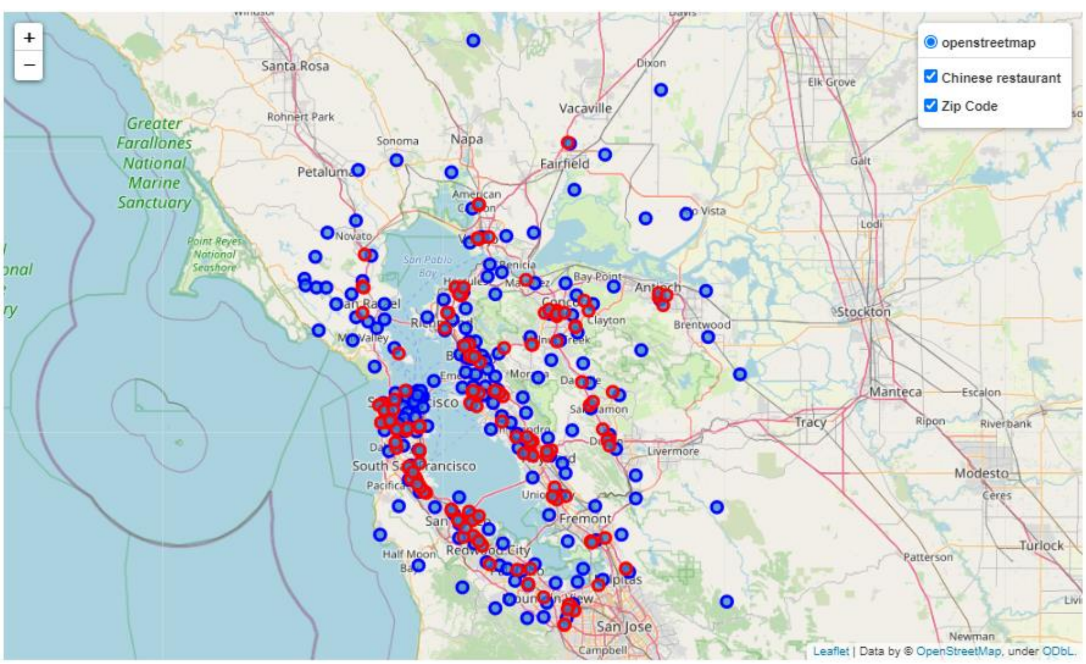
```
(186,)
```

```
bay_chinese['ZipCode'].unique().shape
```

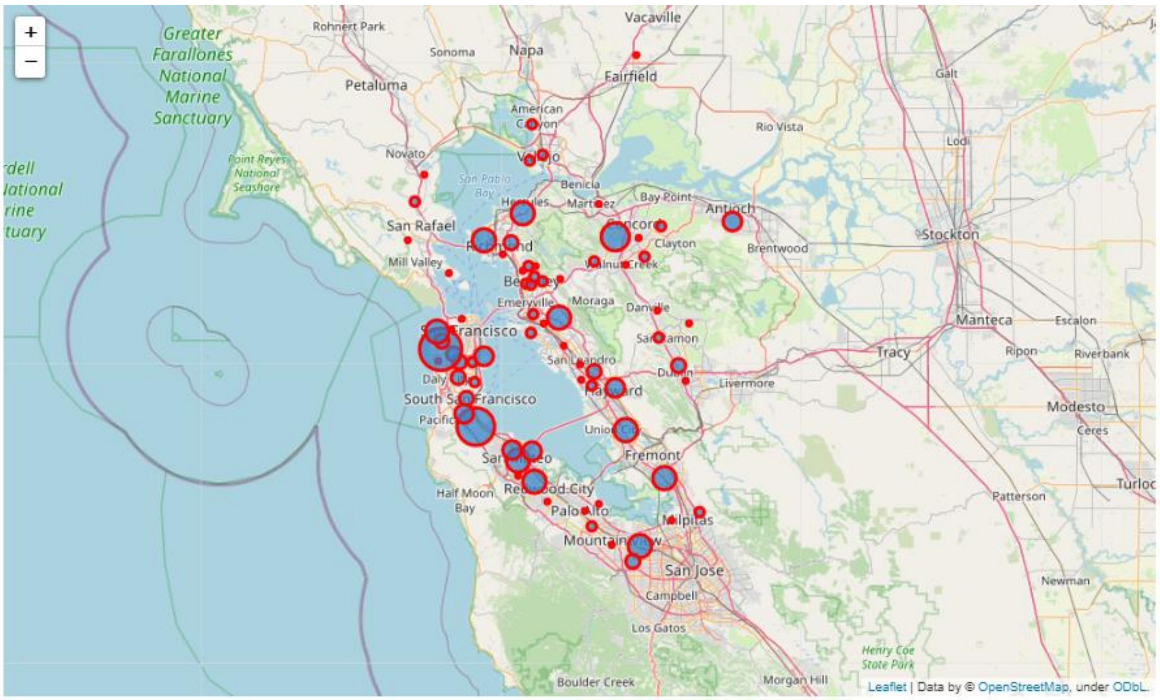
```
(75,)
```

4.2 Folium mapping

The folium map directly displays the distribution of Chinese restaurants in all of the zip codes. Each red circle represents a Chinese restaurant and each blue circle represents a zip code.



We can see that the Chinese restaurants' distribution is not even in Bay Area. There are more Chinese restaurants around the San Francisco Bay, where there are much fewer around the San Pablo Bay and the Susun Bay. Since the blue and red circle markers are crowded on the Folium map, it is hard to say which zip codes have more dense Chinese restaurants. I make a further step to display the distribution more intuitively by grouping the Chinese restaurants by the zip code, and on the Folium map, using the size of the circle to represent the number of the restaurants in the corresponding zip code. On the map, it is obvious that some zip codes, like 94116 and 94030, have much larger circle markers, which means much more Chinese restaurants in these zip code.



4.3 Clusters of the zip codes with Chinese Restaurants

As shown in 3.2.3, the clustering result of the zip codes in Bay Area based on their venue categories shows that 87% of the zip codes belong to cluster 1, 11% of the zip codes belong to cluster 0, and only around 2% belong to clusters 2, 3, and 4

.

Bay Area Zipcodes Clustering

	Cluster Labels	Counts
0	1	152
1	0	19
2	4	1
3	3	1
4	2	1

Clusters of the zip codes with Chinese Restaurants

	Cluster Labels	Counts
0	1	73
1	0	1

The virtual distribution pattern of the Chinese restaurants can be seen on the Folium map. Here the purpose to cluster the zip codes is to find out the quantitively distribution pattern. The statistics result shows that 98.6% (73 of 74) of the zip codes where the Chinese restaurants are located belong to cluster 1, 1.4% (1 of 74) belong to cluster 0, and none of them belong to cluster 2, 3, or 4. We can see that the zip codes with Chinese restaurants have a common feature: nearly all of them (98.6%) belong to cluster 1.

4.4 Relation of the Chinese restaurant distribution to population

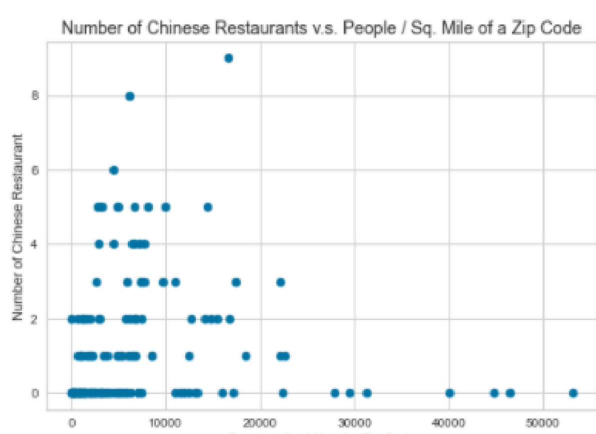
There are a lot of things to consider for an investor to decide the location of a venue, such as the rent, surroundings, population, etc. For a restaurant, the population is definitely an important factor to be considered. The point is how the population affects the location distribution of the venues and whether there is any quantitative relation between them.

4.4.1 Merge the Chinese restaurant data and the population data

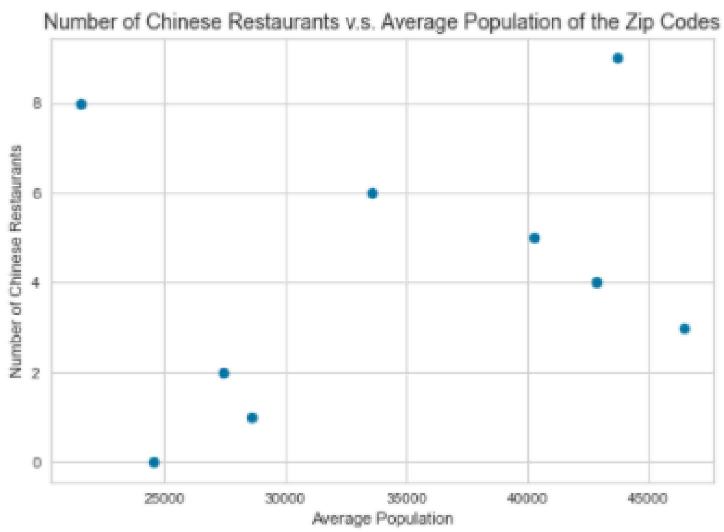
The resulting data set contains zip codes, number of Chinese restaurants in each zip code, total population, population density, and population of different races. I will analyze the relation between the number of Chinese restaurants and the population.

	ZipCode	Number of Chinese Restaurant	City	Latitude	Longitude	Area	People / Sq. Mile	Total Population	White	Black or African American	American Indian and Alaska Native	Asian Indian	Chinese	Filipino	Japa
0	94002	1	Belmont, California	37.509366	-122.306132	1.632100e+08	3405.90	25992.0	17555.0	424.0	72.0	979.0	2234.0	791.0	
1	94005	2	Brisbane, California	37.684694	-122.407120	1.323574e+08	680.80	4282.0	2578.0	80.0	21.0	52.0	456.0	335.0	
2	94010	0	Burlingame, California	37.570280	-122.365778	3.590162e+08	2860.17	40737.0	27523.0	403.0	82.0	768.0	5056.0	1437.0	
3	94014	3	Daly City, California	37.691561	-122.445202	1.815553e+08	7788.15	47014.0	11703.0	1625.0	276.0	191.0	6408.0	14610.0	
4	94015	0	Daly City, California	37.680844	-122.481310	1.594835e+08	11003.66	60927.0	14622.0	2155.0	169.0	476.0	9846.0	20792.0	

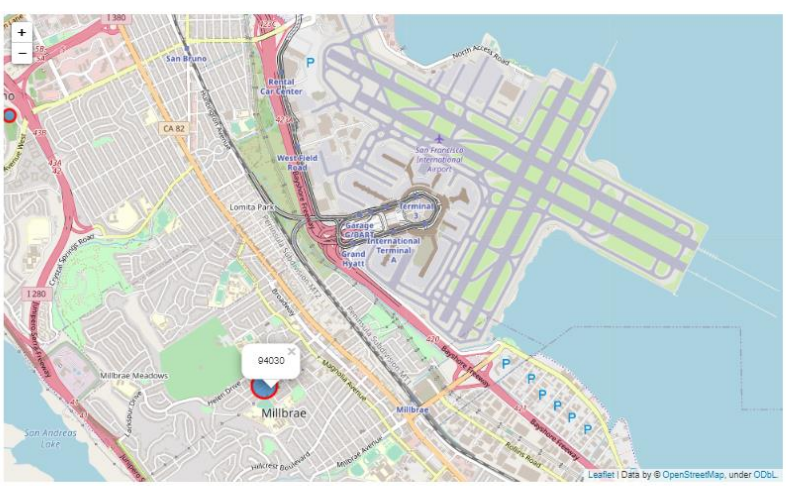
4.4.2 Total population



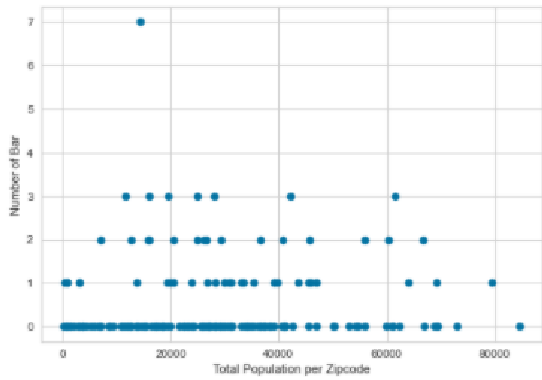
Above are two plots: the number of Chinese restaurants versus the total population of each zip code and versus the population density (People / Sq. Mile) of each zip code. However, these neither of these two plots shows a direct relation. Some zip codes with a large population have none or few Chinese restaurants, while some zip codes with a small population have rather large number of Chinese restaurants. In another words, the population of the zip codes with the same number of Chinese restaurants may vary significantly.

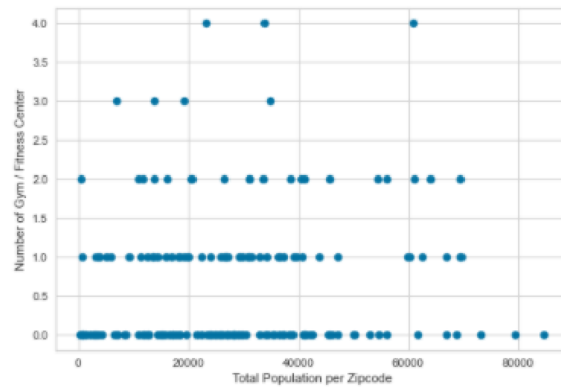


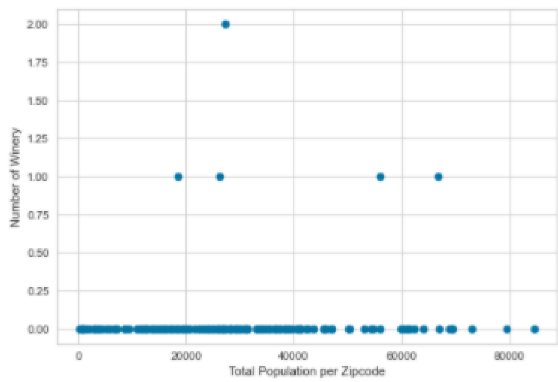
However, the number of Chinese restaurants has a loose proportional dependence on the average population of the zip codes with the same number of Chinese restaurants, as shown above. There is an obvious odd point on this plot where the number of Chinese restaurants is 8. This point has the least average population but almost the largest number of Chinese restaurants. I checked the data and find out that there is only one zip code (94030) that has 8 Chinese restaurants. This zip code is actually where the San Francisco International Airport is located. This can explain why this zip code (94030) has a small population but a large number of Chinese restaurants.

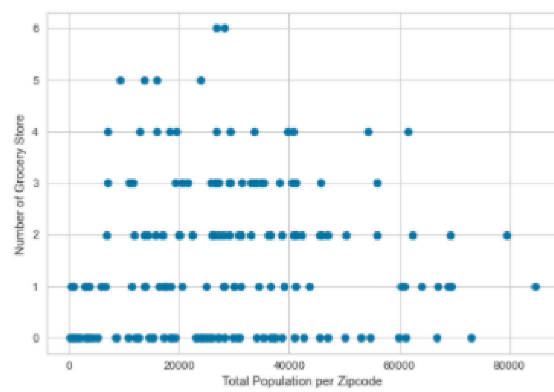


Other venues have the similar results. Below are a few examples:



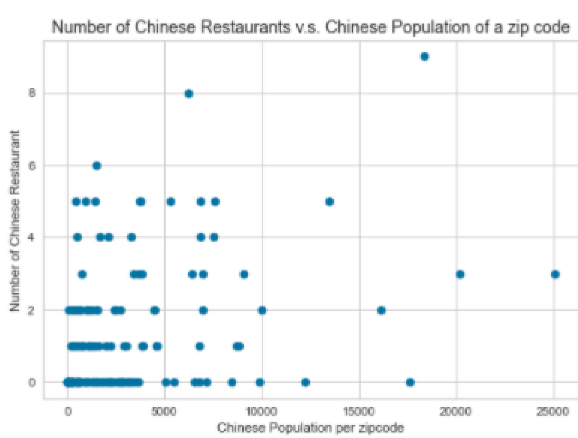


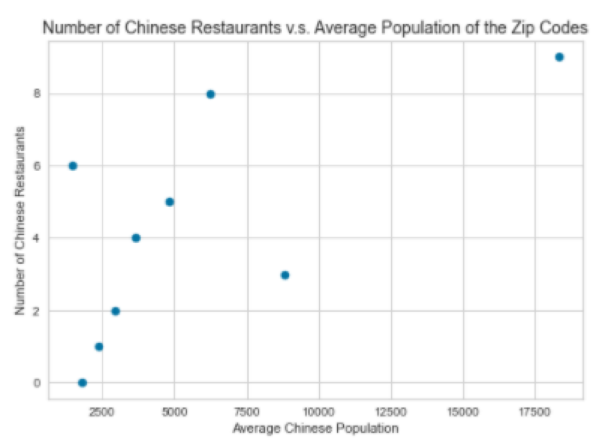




4.4.3 Chinese population

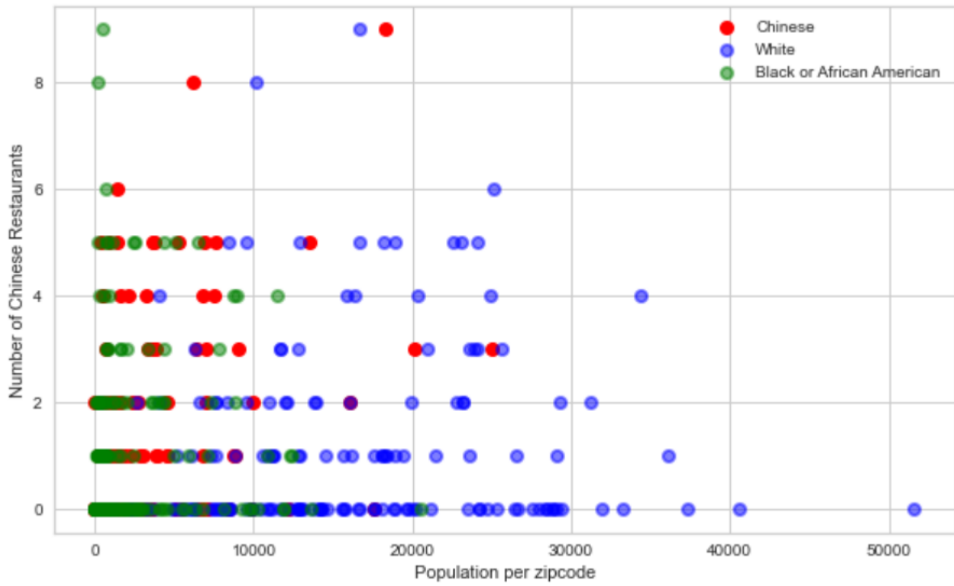
To explore the relation between the number of Chinese restaurants and the Chinese population is the original intention of this project. The result is actually very similar to that for the total population in 4.4.2, only a tighter proportional dependence on the average Chinese population of the zip codes with the same number of Chinese restaurants. The result is reasonable. Generally, if an area has a large population of Chinese, there should be sufficient restaurants to supply the dining demand. However, the locations of the restaurants are affected by many factors, not only the population.





4.4.4 Population of different races

I plot the number of Chinese restaurants vs. the population of Chinese, white, and Black/African American of each zip code. Here white and black/African American races are chosen because their large population in USA. We can see similar results to those in 4.4.2 and 4.4.3.

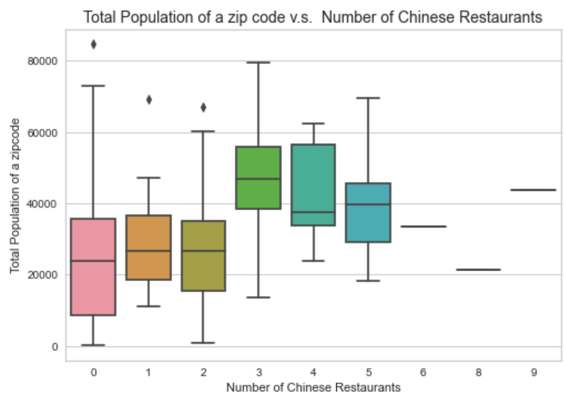


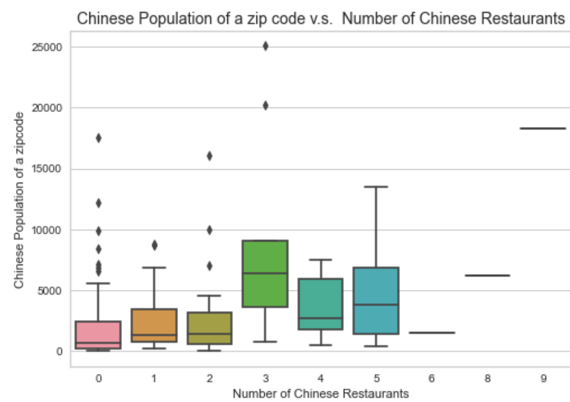
4.5 Statistics analysis

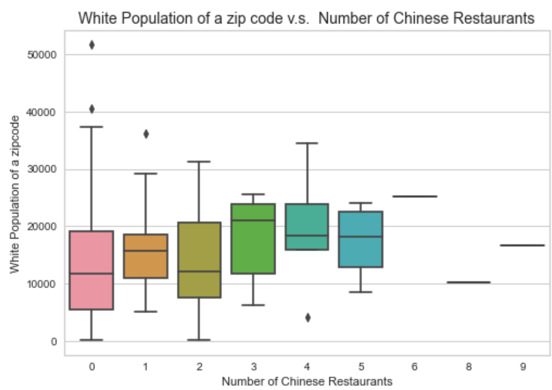
Due to the large variation of the population of the zip codes with the same number of Chinese restaurants, statistics method is desired to analyze the data. Here I will use box plot for graphically depicting groups of the data through their quartiles, and use Pearson Correlation to find out the strength of the correlation between the number of Chinese restaurants and the population.

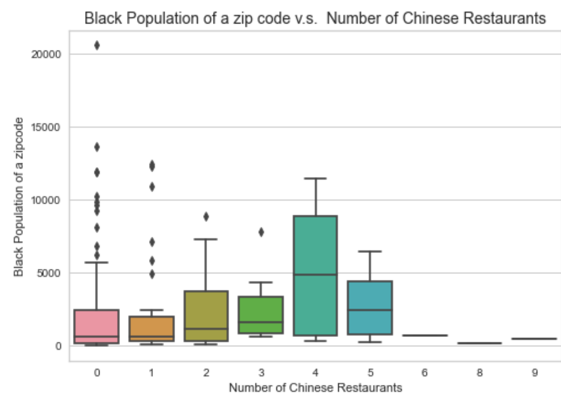
4.5.1 Box plot

In descriptive statistics, boxplot is a method for graphically depicting groups of numerical data through their quartiles. The box plots below on the total population, Chinese population, white population, and African American population, graphically display the similar results in 4.4. From the median values of each plot, we can see a subtle positive relation between the number of Chinese restaurants and the population, although they look kind of random. Although for each plot the population corresponding to each number of Chinese restaurants has a large variation, this variation is larger in the total population plot and the white population plot, and is relatively smaller in the Chinese population plot.









4.5.2 Pearson Correlation

Pearson correlation is a measure of linear correlation between two sets of data. This method will give us two values: the correlation coefficient and the P-value. For the correlation coefficient, a value close to 1 implies a large positive correlation, while a value close to negative 1 implies a large negative correlation, and a value close to zero implies no correlation between the variables. The P-value will tell us how certain we are about the correlation that we calculated. For the P-value, a value less than.001 gives us a strong certainty about the correlation coefficient that we calculated. A value between 0.001 and 0.05 gives us moderate certainty. A value between 0.05 and 0.1 will give us a weak certainty. And a P-value larger than 0.1 will give us no certainty of correlation at all. We can say that there is a strong correlation when the correlation coefficient is close to 1 or negative 1, and the P-value is less than.001. Below are the correlations between the number of Chinese restaurants and population of different races. The result is sorted by the correlation coefficients.

Race	Pearson Correlation Coefficient	P-value
Chinese	0.389104	1.119018e-07
Japanese	0.351877	1.923586e-06
Other Asian	0.287259	1.214346e-04
Total Population	0.273408	2.619307e-04
Korean	0.261904	4.812362e-04
Native Hawaiian	0.234785	1.818263e-03
Samoan	0.184477	1.481648e-02
Asian Indian	0.182008	1.623211e-02
Filipino	0.178221	1.863173e-02
Vietnamese	0.167707	2.696989e-02
White	0.142055	6.150780e-02
American Indian and Alaska Native	0.097619	2.000317e-01
Some Other Race	0.095911	2.080510e-01
Guamanian or Chamorro	0.083007	2.761872e-01
Other Pacific Islander	0.080477	2.911376e-01
People / Sq. Mile	0.075353	3.230481e-01
Black or African American	0.044992	5.555229e-01

In the above list, the first on is the correlation with Chinese population, with a Pearson Correlation coefficient around 0.4 and P-value less than 0.001, which means a relatively large positive correlation and a strong certainty about the correlation coefficient. We can see that the number of Chinese restaurants indeed has the strongest correlation with Chinese population. The second strongest correlation is with Japanese population. The correlation is slightly weaker for the total population. White population, as the main population in USA, has an even weaker correlation. And African American, as one of the main populations, has the weakest correlation.

4.6 Correlation of the venues and population

In the same way as that to get the correlation between the Chinese restaurants and the population distribution, the correlation for all the venues can be obtained. The results might be interesting for researchers and investors.

4.6.1 Define a function to get the Pearson Correlation of a venue category on the races

```python
def pearson_venue(venue_category):
    category=venue_category
    number_of_venue='Number of '+category
    zip_venue=bay_onehot[['ZipCode',category]].groupby('ZipCode').sum()
    zip_venue=zip_venue.rename(columns={category:number_of_venue})
    bay_venue_population=pd.merge(zip_venue, population, on='ZipCode')

    pearson_venue=pd.DataFrame(columns=['Race','Pearson Correlation Coefficient','P-value'])

    for race in bay_venue_population.columns.values[8:]:
        pearson_coef, p_value = stats.pearsonr(x=bay_venue_population[race],y=bay_venue_population[number_of_venue])
        pearson_venue=pearson_venue.append({'Race':race, 'Pearson Correlation Coefficient':pearson_coef,'P-value':p_value},
                                           ignore_index=True)

    return pearson_venue.set_index('Race').sort_values(by='Pearson Correlation Coefficient',ascending=False)
```

As an example, the correlation between the Sushi restaurant and the populations is generated as below:

```python
pearson_venue('Sushi Restaurant')
```

Race	Pearson Correlation Coefficient	P-value
Korean	0.295263	0.000076
Japanese	0.229815	0.002284
White	0.196312	0.009427
Chinese	0.119663	0.115777
Native Hawaiian	0.059342	0.436679
Other Asian	0.036585	0.631743
Asian Indian	0.032393	0.671327
American Indian and Alaska Native	0.005829	0.939154
Vietnamese	-0.004753	0.950372
Filipino	-0.025333	0.740029
Guamanian or Chamorro	-0.053009	0.487250
Other Pacific Islander	-0.061592	0.419456
Samoan	-0.065511	0.390424
Some Other Race	-0.098485	0.196048
Black or African American	-0.127114	0.094633

The result is a little bit surprising. The population with the strongest correlation with sushi restaurants is Korean, instead of Japanese.

4.6.2 Correlation of all the venues and population

Here I extend the correlation results for each venue to list the most correlated population, the correlation coefficient, P-value, and the least correlated population. I list them here for an extension of this project.

	Most correlated Population	Pearson Correlation Coefficient	P-value	Least correlated Population
Cambodian Restaurant	American Indian and Alaska Native	0.233371	1.941169e-03	Asian Indian
Argentinian Restaurant	American Indian and Alaska Native	0.021159	7.816880e-01	Other Asian
Performing Arts Venue	American Indian and Alaska Native	0.144479	5.716299e-02	Other Pacific Islander
Jewish Restaurant	American Indian and Alaska Native	0.255452	6.689998e-04	Other Pacific Islander
South American Restaurant	American Indian and Alaska Native	0.255452	6.689998e-04	Other Pacific Islander
Adult Boutique	American Indian and Alaska Native	0.103691	1.733273e-01	Other Pacific Islander
Sausage Shop	American Indian and Alaska Native	0.255452	6.689998e-04	Other Pacific Islander
Pedestrian Plaza	American Indian and Alaska Native	0.028580	7.081382e-01	White
Cocktail Bar	American Indian and Alaska Native	0.118128	1.205612e-01	Other Pacific Islander
Opera House	American Indian and Alaska Native	0.137317	7.078548e-02	Other Pacific Islander
Motorcycle Shop	American Indian and Alaska Native	0.137317	7.078548e-02	Other Pacific Islander
Event Space	American Indian and Alaska Native	0.235416	1.765758e-03	Asian Indian
Mexican Restaurant	American Indian and Alaska Native	0.399923	4.571615e-08	Asian Indian
Jazz Club	American Indian and Alaska Native	0.139905	6.558755e-02	Other Pacific Islander
Marijuana Dispensary	American Indian and Alaska Native	0.207504	6.006921e-03	Asian Indian
North Indian Restaurant	Asian Indian	0.328370	9.708771e-06	Black or African American
Trail	Asian Indian	0.024582	7.474724e-01	American Indian and Alaska Native
Boxing Gym	Asian Indian	0.218744	3.733583e-03	Other Pacific Islander
Middle Eastern Restaurant	Asian Indian	0.132040	8.242427e-02	Black or African American
Cafeteria	Asian Indian	0.098699	1.950743e-01	Black or African American
Volleyball Court	Asian Indian	-0.006714	9.299380e-01	White
Warehouse Store	Asian Indian	0.145050	5.617594e-02	Samoan
Mountain	Asian Indian	0.113385	1.363098e-01	Black or African American
Afghan Restaurant	Asian Indian	0.187493	1.323483e-02	Chinese
Climbing Gym	Asian Indian	0.179240	1.795753e-02	Other Pacific Islander
Train	Asian Indian	0.195303	9.806592e-03	Black or African American
Metro Station	Asian Indian	0.285406	1.349024e-04	Chinese
Big Box Store	Asian Indian	0.256841	6.236423e-04	Chinese
Bank	Asian Indian	0.180732	1.700845e-02	Other Pacific Islander
Falafel Restaurant	Asian Indian	0.293349	8.550356e-05	Samoan
Hardware Store	Asian Indian	0.236595	1.671389e-03	Chinese
Residential Building (Apartment / Condo)	Asian Indian	0.178601	1.837787e-02	Black or African American
River	Asian Indian	0.185443	1.429304e-02	Black or African American
Mobile Phone Shop	Asian Indian	0.244059	1.172973e-03	Guamanian or Chamorro
Comedy Club	Asian Indian	0.523342	1.272752e-13	Black or African American
Caribbean Restaurant	Black or African American	0.192120	1.109565e-02	Asian Indian
Bike Trail	Black or African American	-0.041762	5.842798e-01	White
Restaurant	Black or African American	0.024510	7.481851e-01	Native Hawaiian
Fondue Restaurant	Black or African American	-0.007479	9.219792e-01	White
Speakeasy	Black or African American	0.069755	3.603933e-01	Native Hawaiian
Bar	Black or African American	0.143101	5.960147e-02	Asian Indian
Lounge	Black or African American	0.088597	2.450191e-01	Other Pacific Islander
Multiplex	Black or African American	-0.040553	5.952140e-01	Other Asian

Food Truck	Black or African American	0.143739	5.846254e-02	White
Camera Store	Black or African American	0.135901	7.376888e-02	White
Photography Lab	Black or African American	-0.017481	8.189161e-01	White
Lawyer	Black or African American	0.128213	9.179158e-02	Korean
Theme Park Ride / Attraction	Black or African American	0.385223	1.530831e-07	White
Canal	Black or African American	0.146025	5.452624e-02	Korean
Veterinarian	Black or African American	0.019659	7.968159e-01	Some Other Race
BBQ Joint	Black or African American	0.215073	4.371997e-03	Guamanian or Chamorro
Stadium	Black or African American	0.243393	1.211180e-03	White
Scandinavian Restaurant	Black or African American	0.176680	1.969256e-02	White
Print Shop	Black or African American	0.240920	1.363114e-03	White
Roller Rink	Black or African American	0.226017	2.710628e-03	Native Hawaiian
Soba Restaurant	Black or African American	0.163046	3.158496e-02	White
Mac & Cheese Joint	Black or African American	0.101151	1.841567e-01	Chinese
Medical Center	Black or African American	0.180265	1.730100e-02	Korean
Music Venue	Black or African American	0.204841	6.700560e-03	White
Theme Park	Black or African American	0.195143	9.868047e-03	Native Hawaiian
Brewery	Black or African American	0.240061	1.419879e-03	Asian Indian
Shoe Store	Black or African American	0.230427	2.221486e-03	Japanese
Basketball Stadium	Black or African American	0.098200	1.973541e-01	White
Fish & Chips Shop	Black or African American	0.207340	6.047667e-03	Asian Indian
Fried Chicken Joint	Black or African American	0.281095	1.718127e-04	Korean
Automotive Shop	Black or African American	0.125130	9.993778e-02	Asian Indian
Beer Garden	Black or African American	0.081526	2.848749e-01	White
Boat or Ferry	Black or African American	0.145379	5.561442e-02	White
Intersection	Black or African American	0.166277	2.832040e-02	Japanese
Diner	Black or African American	0.241677	1.314853e-03	Asian Indian
Bus Stop	Black or African American	0.198553	8.628702e-03	White
Cajun / Creole Restaurant	Black or African American	0.087360	2.516928e-01	Native Hawaiian
Truck Stop	Black or African American	0.213728	4.629615e-03	Japanese
Harbor / Marina	Black or African American	0.178843	1.821737e-02	Japanese
Baseball Stadium	Black or African American	0.153538	4.310293e-02	Japanese
Burrito Place	Chinese	0.175954	2.021097e-02	Guamanian or Chamorro
Lake	Chinese	-0.013544	8.592107e-01	American Indian and Alaska Native
Spanish Restaurant	Chinese	-0.024486	7.484273e-01	White
Trattoria/Osteria	Chinese	0.173010	2.243352e-02	Some Other Race
Hunan Restaurant	Chinese	0.165585	2.899365e-02	Asian Indian
Monument / Landmark	Chinese	0.222709	3.139342e-03	Some Other Race
Cheese Shop	Chinese	0.147762	5.168376e-02	Other Pacific Islander
College Gym	Chinese	0.241138	1.349076e-03	Other Pacific Islander
Smoothie Shop	Chinese	0.211053	5.182335e-03	Other Pacific Islander
Church	Chinese	0.147306	5.241740e-02	Some Other Race
Field	Chinese	0.274155	2.515516e-04	Black or African American

Trade School	Chinese	0.086099	2.586298e-01	Some Other Race
Dessert Shop	Chinese	0.272764	2.711976e-04	Black or African American
Gastropub	Chinese	0.216987	4.027888e-03	Guamanian or Chamorro
Men's Store	Chinese	0.041003	5.911302e-01	White
Boutique	Chinese	0.070636	3.543457e-01	Some Other Race
Exhibit	Chinese	0.050638	5.069672e-01	White
Outdoor Sculpture	Chinese	0.067379	3.770277e-01	Black or African American
Szechuan Restaurant	Chinese	0.210760	5.246374e-03	Some Other Race
Shoe Repair	Chinese	0.168666	2.609590e-02	Black or African American
Russian Restaurant	Chinese	0.122744	1.066272e-01	American Indian and Alaska Native
Scenic Lookout	Chinese	0.078080	3.057946e-01	American Indian and Alaska Native
Bubble Tea Shop	Chinese	0.524951	1.039009e-13	Other Pacific Islander
Science Museum	Chinese	0.182046	1.620935e-02	Black or African American
Dumpling Restaurant	Chinese	0.540246	1.429075e-14	Asian Indian
Art Museum	Chinese	0.165105	2.946968e-02	Some Other Race
Beer Bar	Chinese	0.172031	2.321755e-02	Other Pacific Islander
Disc Golf	Chinese	0.294436	8.024680e-05	Black or African American
Wine Bar	Chinese	0.148623	5.032118e-02	Some Other Race
Vietnamese Restaurant	Chinese	0.457345	2.237466e-10	White
Board Shop	Chinese	0.332089	7.581073e-06	Asian Indian
Tourist Information Center	Chinese	0.104848	1.685543e-01	Some Other Race
History Museum	Chinese	0.096970	2.030507e-01	Black or African American
Pool	Chinese	0.228463	2.428461e-03	Black or African American
Bakery	Chinese	0.482491	1.564816e-11	Guamanian or Chamorro
Sports Club	Chinese	0.390412	1.005952e-07	Some Other Race
Art Gallery	Chinese	0.139466	6.644651e-02	Filipino
Alternative Healer	Chinese	0.288117	1.156339e-04	Korean
Pizza Place	Chinese	0.207514	6.004302e-03	Other Pacific Islander
Dive Bar	Chinese	0.235026	1.798040e-03	Asian Indian
Salvadoran Restaurant	Chinese	0.269594	3.214106e-04	Asian Indian
Café	Chinese	0.230131	2.251623e-03	Guamanian or Chamorro
Yoga Studio	Chinese	0.170858	2.418786e-02	Other Pacific Islander
Zoo Exhibit	Chinese	0.161814	3.291071e-02	White
Waterfall	Chinese	0.399497	4.738539e-08	Some Other Race
Soccer Field	Chinese	0.377456	2.831973e-07	Other Pacific Islander
Deli / Bodega	Chinese	0.387564	1.267840e-07	Asian Indian
Pakistani Restaurant	Chinese	0.216851	4.051481e-03	Black or African American
Liquor Store	Chinese	0.150879	4.689137e-02	Asian Indian
Bike Rental / Bike Share	Chinese	0.390412	1.005952e-07	Some Other Race
Chinese Restaurant	Chinese	0.393966	7.513882e-08	Black or African American
Poke Place	Chinese	0.324971	1.213733e-05	Black or African American
Empanada Restaurant	Chinese	0.224637	2.882566e-03	Some Other Race
Playground	Chinese	0.436498	1.732829e-09	Asian Indian
Zoo	Chinese	0.176920	1.952422e-02	Asian Indian

Ethiopian Restaurant	Chinese	0.087173	2.527147e-01	Some Other Race
Hookah Bar	Filipino	0.472713	4.517449e-11	Black or African American
General College & University	Filipino	0.327954	9.979193e-06	Korean
Auto Dealership	Filipino	0.217519	3.936690e-03	Black or African American
Lingerie Store	Filipino	0.215047	4.376907e-03	Black or African American
Discount Store	Filipino	0.357305	1.298801e-06	Asian Indian
Supplement Shop	Filipino	0.258192	5.822488e-04	Black or African American
Andhra Restaurant	Filipino	0.135570	7.448222e-02	Guamanian or Chamorro
Casino	Filipino	0.213983	4.579608e-03	White
Pool Hall	Filipino	0.279019	1.927682e-04	Black or African American
Supermarket	Filipino	0.403048	3.508946e-08	Asian Indian
Filipino Restaurant	Filipino	0.580696	4.500352e-17	Asian Indian
Athletics & Sports	Filipino	0.165962	2.862551e-02	Black or African American
Golf Course	Filipino	0.225357	2.791719e-03	Asian Indian
Kebab Restaurant	Filipino	0.378983	2.512420e-07	White
Tex-Mex Restaurant	Filipino	0.311328	2.895913e-05	White
Rental Service	Guamanian or Chamorro	0.398007	5.369350e-08	Korean
Credit Union	Guamanian or Chamorro	0.680489	5.261395e-25	Chinese
ATM	Guamanian or Chamorro	0.145060	5.615961e-02	Other Asian
Light Rail Station	Guamanian or Chamorro	0.467090	8.187087e-11	Chinese
Post Office	Guamanian or Chamorro	-0.008299	9.134536e-01	Japanese
Pharmacy	Guamanian or Chamorro	0.301926	5.146398e-05	Asian Indian
Video Store	Guamanian or Chamorro	0.357368	1.292801e-06	Korean
Home Service	Guamanian or Chamorro	0.115343	1.296267e-01	Japanese
Escape Room	Guamanian or Chamorro	-0.023592	7.573219e-01	American Indian and Alaska Native
Halal Restaurant	Guamanian or Chamorro	0.680489	5.261395e-25	Chinese
Fast Food Restaurant	Guamanian or Chamorro	0.410397	1.863320e-08	Chinese
Fishing Spot	Guamanian or Chamorro	-0.009697	8.989435e-01	White
Swim School	Guamanian or Chamorro	-0.030971	6.849750e-01	White
Paintball Field	Guamanian or Chamorro	-0.033975	6.562757e-01	White
Airport Terminal	Guamanian or Chamorro	0.011112	8.842917e-01	White
Sports Bar	Guamanian or Chamorro	0.188052	1.295866e-02	Asian Indian
Israeli Restaurant	Japanese	0.056978	4.551926e-01	Native Hawaiian
Churrascaria	Japanese	0.097638	1.999436e-01	American Indian and Alaska Native
Swiss Restaurant	Japanese	0.300036	5.763455e-05	American Indian and Alaska Native
Bookstore	Japanese	0.183798	1.519485e-02	Some Other Race
South Indian Restaurant	Japanese	0.191298	1.145161e-02	Black or African American
Bike Shop	Japanese	0.182711	1.581744e-02	Some Other Race
Peruvian Restaurant	Japanese	0.111185	1.441257e-01	Black or African American
National Park	Japanese	0.063102	4.081226e-01	Filipino
Smoke Shop	Japanese	0.199864	8.190264e-03	Vietnamese
Cycle Studio	Japanese	0.065897	3.876340e-01	Native Hawaiian
Eye Doctor	Japanese	0.301972	5.132363e-05	Some Other Race
Ramen Restaurant	Japanese	0.178446	1.848106e-02	American Indian and Alaska Native
Japanese Curry Restaurant	Japanese	0.329401	9.068025e-06	Black or African American

Vegetarian / Vegan Restaurant	Japanese	0.086143	2.583859e-01	Guamanian or Chamorro
Creperie	Japanese	0.123268	1.051295e-01	Black or African American
Physical Therapist	Japanese	0.205375	6.555871e-03	Some Other Race
Planetarium	Japanese	0.205375	6.555871e-03	Some Other Race
Hotel Bar	Japanese	-0.022187	7.713666e-01	Native Hawaiian
Synagogue	Japanese	-0.019481	7.986169e-01	American Indian and Alaska Native
Neighborhood	Japanese	0.099446	1.917006e-01	Some Other Race
Hotpot Restaurant	Japanese	0.438952	1.371603e-09	Black or African American
Tapas Restaurant	Japanese	0.040709	5.937987e-01	Native Hawaiian
Food & Drink Shop	Japanese	0.194929	9.950756e-03	Other Pacific Islander
Taiwanese Restaurant	Japanese	0.472119	4.812319e-11	Other Pacific Islander
Portuguese Restaurant	Japanese	0.060114	4.307193e-01	Some Other Race
Gourmet Shop	Japanese	-0.025843	7.349973e-01	Native Hawaiian
Windmill	Japanese	0.254554	6.999432e-04	Some Other Race
Surf Spot	Japanese	0.254554	6.999432e-04	Some Other Race
Pastry Shop	Japanese	0.085744	2.606034e-01	Native Hawaiian
Pub	Japanese	0.133344	7.941495e-02	American Indian and Alaska Native
Garden	Japanese	0.225884	2.726718e-03	Some Other Race
Toy / Game Store	Japanese	0.185620	1.419866e-02	American Indian and Alaska Native
Comic Shop	Japanese	0.289905	1.043645e-04	Some Other Race
Cuban Restaurant	Japanese	0.138348	6.867662e-02	Black or African American
Music Store	Japanese	0.121043	1.116057e-01	Black or African American
Moroccan Restaurant	Japanese	0.222925	3.109648e-03	Some Other Race
Antique Shop	Japanese	0.254554	6.999432e-04	Some Other Race
Buffet	Japanese	0.175977	2.019384e-02	Vietnamese
Medical Supply Store	Japanese	0.058693	4.417162e-01	American Indian and Alaska Native
Bagel Shop	Japanese	0.206009	6.388047e-03	Some Other Race
Market	Japanese	0.136146	7.324686e-02	Guamanian or Chamorro
Korean Restaurant	Japanese	0.367591	6.046415e-07	Some Other Race
Wine Shop	Japanese	0.241473	1.327678e-03	Some Other Race
Travel Agency	Japanese	0.102417	1.786968e-01	Black or African American
Burmese Restaurant	Japanese	0.136626	7.222867e-02	Some Other Race
Financial or Legal Service	Japanese	0.241909	1.300359e-03	Black or African American
Miscellaneous Shop	Japanese	0.082367	2.799185e-01	White
Seafood Restaurant	Japanese	0.194819	9.993757e-03	Some Other Race
Park	Japanese	0.444639	7.919675e-10	American Indian and Alaska Native
Beach	Japanese	0.202818	7.274326e-03	Some Other Race
Health & Beauty Service	Japanese	0.226557	2.645845e-03	Black or African American
Business Service	Japanese	0.062940	4.093284e-01	Some Other Race
Theater	Japanese	0.099036	1.935456e-01	Filipino
Farmers Market	Japanese	0.108272	1.549973e-01	Some Other Race
Frozen Yogurt Shop	Japanese	0.450036	4.659140e-10	Some Other Race
Japanese Restaurant	Japanese	0.304568	4.387366e-05	Some Other Race
Steakhouse	Japanese	0.213253	4.723713e-03	Black or African American

Grocery Store	Japanese	0.309781	3.187733e-05	Filipino
Gym / Fitness Center	Japanese	0.223581	3.020754e-03	Black or African American
Indie Movie Theater	Japanese	0.070881	3.526768e-01	Some Other Race
Italian Restaurant	Japanese	0.130858	8.523129e-02	Black or African American
Flower Shop	Japanese	0.108405	1.544902e-01	Asian Indian
Tennis Court	Japanese	0.178834	1.822325e-02	Some Other Race
Community Center	Japanese	0.218312	3.804153e-03	Black or African American
Mediterranean Restaurant	Japanese	0.253287	7.458139e-04	Some Other Race
Massage Studio	Japanese	0.349859	2.222032e-06	Other Pacific Islander
Dog Run	Japanese	0.309431	3.257443e-05	Some Other Race
Taco Place	Japanese	0.225135	2.819408e-03	Native Hawaiian
Breakfast Spot	Japanese	0.236912	1.646804e-03	Guamanian or Chamorro
Tiki Bar	Korean	0.094122	2.167017e-01	Some Other Race
Tea Room	Korean	0.118617	1.190194e-01	Other Pacific Islander
Campground	Korean	0.060606	4.269523e-01	White
Health Food Store	Korean	0.028306	7.108087e-01	Black or African American
College Science Building	Korean	-0.013418	8.605068e-01	Native Hawaiian
College Arts Building	Korean	0.286834	1.244031e-04	Native Hawaiian
Botanical Garden	Korean	0.223385	3.047148e-03	Some Other Race
Rugby Pitch	Korean	0.133338	7.942840e-02	White
Tattoo Parlor	Korean	0.119133	1.174131e-01	Native Hawaiian
Udon Restaurant	Korean	0.062749	4.107601e-01	American Indian and Alaska Native
College Theater	Korean	0.117561	1.223662e-01	Native Hawaiian
Amphitheater	Korean	0.131841	8.289262e-02	Native Hawaiian
Resort	Korean	0.099220	1.927182e-01	American Indian and Alaska Native
Fish Market	Korean	0.097859	1.989195e-01	Guamanian or Chamorro
Tibetan Restaurant	Korean	0.101984	1.805535e-01	Some Other Race
Baseball Field	Korean	0.104824	1.686498e-01	Other Pacific Islander
Chocolate Shop	Korean	-0.002245	9.765484e-01	Other Asian
Bay	Korean	0.092406	2.252357e-01	Some Other Race
College Library	Korean	0.185423	1.430363e-02	Native Hawaiian
Shipping Store	Korean	0.219813	3.564111e-03	Black or African American
Movie Theater	Korean	0.084652	2.667508e-01	Guamanian or Chamorro
Indonesian Restaurant	Korean	0.146318	5.403684e-02	American Indian and Alaska Native
Tailor Shop	Korean	0.156505	3.918078e-02	Some Other Race
College Rec Center	Korean	0.074372	3.294043e-01	White
Road	Korean	-0.000951	9.900647e-01	White
College Basketball Court	Korean	0.035858	6.385369e-01	Some Other Race
Rest Area	Korean	-0.017558	8.181314e-01	American Indian and Alaska Native
College Baseball Diamond	Korean	0.035858	6.385369e-01	Some Other Race
College Soccer Field	Korean	0.035858	6.385369e-01	Some Other Race
College Football Field	Korean	0.035858	6.385369e-01	Some Other Race
Indian Restaurant	Korean	0.294974	7.775947e-05	Some Other Race
Coffee Shop	Korean	0.240864	1.366779e-03	Guamanian or Chamorro

Coffee Shop	Korean	[illegible]	[illegible]	Guamanian or Chamorro
Mongolian Restaurant	Korean	0.144900	5.643364e-02	Filipino
American Restaurant	Korean	0.180351	1.724665e-02	Some Other Race
Gym	Korean	0.221923	3.249972e-03	Native Hawaiian
Sporting Goods Shop	Korean	0.076217	3.175122e-01	White
Organic Grocery	Korean	0.175527	2.052108e-02	Some Other Race
Hill	Korean	0.128092	9.210035e-02	Some Other Race
Stables	Korean	0.023222	7.610111e-01	Other Asian
Nightlife Spot	Korean	0.022425	7.689769e-01	White
Hotel	Korean	0.117461	1.226857e-01	White
Ice Cream Shop	Korean	0.339809	4.489347e-06	Some Other Race
Picnic Area	Korean	0.039542	6.044302e-01	White
Rental Car Location	Korean	0.101943	1.807297e-01	White
Mini Golf	Korean	0.292315	9.080198e-05	Some Other Race
Burger Joint	Korean	0.232157	2.052599e-03	Some Other Race
Sushi Restaurant	Korean	0.298676	6.249937e-05	Black or African American
Pet Store	Korean	0.294724	7.890593e-05	Other Pacific Islander
Juice Bar	Korean	0.192317	1.101152e-02	Black or African American
Shopping Mall	Korean	0.092417	2.251768e-01	Some Other Race
New American Restaurant	Korean	0.076352	3.166578e-01	Native Hawaiian
Distillery	Korean	0.060866	4.249700e-01	Some Other Race
Furniture / Home Store	Korean	0.092065	2.269558e-01	Samoan
Tree	Korean	0.073802	3.331346e-01	Black or African American
Spa	Korean	0.319842	1.691514e-05	Guamanian or Chamorro
Women's Store	Korean	0.028532	7.086103e-01	American Indian and Alaska Native
Sculpture Garden	Korean	0.110639	1.461187e-01	Some Other Race
Spiritual Center	Korean	0.204078	6.911904e-03	Other Pacific Islander
Concert Hall	Korean	0.136554	7.238166e-02	Some Other Race
Street Food Gathering	Korean	0.043165	5.717010e-01	Some Other Race
Historic Site	Korean	0.099240	1.926272e-01	Some Other Race
Souvlaki Shop	Korean	0.143107	5.958955e-02	Other Pacific Islander
Kitchen Supply Store	Korean	0.016396	8.299717e-01	American Indian and Alaska Native
Train Station	Korean	0.011182	8.835687e-01	Black or African American
Candy Store	Korean	0.186100	1.394611e-02	Some Other Race
Salad Place	Korean	0.233757	1.906869e-03	Other Pacific Islander
Brazilian Restaurant	Korean	0.116299	1.264586e-01	Some Other Race
French Restaurant	Korean	0.199304	8.375118e-03	Some Other Race
Gift Shop	Korean	0.232674	2.004434e-03	Guamanian or Chamorro
Perfume Shop	Korean	0.216296	4.149300e-03	Some Other Race
Accessories Store	Korean	0.231850	2.081647e-03	Other Pacific Islander
Plaza	Korean	0.209616	5.503166e-03	Guamanian or Chamorro
Arts & Crafts Store	Korean	0.208262	5.821468e-03	Other Pacific Islander
Fountain	Korean	-0.039829	6.018128e-01	American Indian and Alaska Native
Tennis Stadium	Korean	0.060606	4.269523e-01	White
Parking	Korean	0.060606	4.269523e-01	White

Rock Club	Korean	0.165604	2.897557e-02	Some Other Race
Airport Service	Korean	0.060606	4.269523e-01	White
Butcher	Korean	0.090835	2.332505e-01	Other Pacific Islander
Martial Arts School	Korean	0.234432	1.848283e-03	Some Other Race
Airport Lounge	Korean	0.097062	2.026224e-01	American Indian and Alaska Native
Molecular Gastronomy Restaurant	Korean	0.146318	5.403684e-02	American Indian and Alaska Native
Beer Store	Korean	0.244476	1.149613e-03	Other Pacific Islander
Cosmetics Shop	Korean	0.291341	9.606976e-05	Guamanian or Chamorro
Record Shop	Korean	0.216554	4.103602e-03	Some Other Race
Convenience Store	Native Hawaiian	0.379781	2.359601e-07	Japanese
Salon / Barbershop	Native Hawaiian	0.196909	9.208158e-03	Asian Indian
Gun Shop	Native Hawaiian	0.224101	2.952051e-03	Other Pacific Islander
Go Kart Track	Native Hawaiian	0.224101	2.952051e-03	Other Pacific Islander
City Hall	Native Hawaiian	0.296759	7.000803e-05	Chinese
Hot Dog Joint	Native Hawaiian	0.239100	1.485846e-03	Japanese
Food Stand	Native Hawaiian	0.109354	1.508921e-01	Japanese
Motorsports Shop	Native Hawaiian	0.296759	7.000803e-05	Chinese
Fabric Shop	Native Hawaiian	0.190096	1.199081e-02	Japanese
Gymnastics Gym	Native Hawaiian	0.284027	1.458130e-04	Black or African American
Cupcake Shop	Native Hawaiian	0.043956	5.646729e-01	Chinese
Optical Shop	Other Asian	0.105094	1.675487e-01	Other Pacific Islander
Flea Market	Other Asian	0.245349	1.102121e-03	White
Auto Garage	Other Asian	0.204785	6.715752e-03	Native Hawaiian
Snack Place	Other Asian	0.265141	4.066390e-04	Guamanian or Chamorro
Department Store	Other Asian	0.310310	3.084861e-05	Black or African American
Video Game Store	Other Asian	0.307926	3.573841e-05	Japanese
Herbs & Spices Store	Other Asian	0.173714	2.188330e-02	Asian Indian
Community College	Other Asian	0.226351	2.670417e-03	White
Kids Store	Other Asian	0.136801	7.186079e-02	Black or African American
Other Great Outdoors	Other Asian	0.192367	1.099043e-02	White
Office	Other Asian	-0.004268	9.554310e-01	Native Hawaiian
Outdoors & Recreation	Other Asian	0.295495	7.541657e-05	White
Basketball Court	Other Asian	0.077760	3.077880e-01	Native Hawaiian
Nail Salon	Other Asian	0.162757	3.189135e-02	Black or African American
Asian Restaurant	Other Asian	0.301099	5.408457e-05	Other Pacific Islander
Food Court	Other Asian	0.099866	1.898225e-01	Samoan
Sandwich Place	Other Asian	0.306634	3.868466e-05	Other Pacific Islander
Museum	Other Asian	0.077233	3.110880e-01	White
German Restaurant	Other Asian	0.114623	1.320550e-01	Guamanian or Chamorro
Thai Restaurant	Other Asian	0.234601	1.833827e-03	Samoan
Wings Joint	Other Asian	0.339597	4.555396e-06	White
Clothing Store	Other Asian	0.154198	4.220205e-02	Guamanian or Chamorro
Airport	Other Pacific Islander	0.100066	1.889313e-01	Other Asian
Food	Other Pacific Islander	0.023820	7.550468e-01	Other Asian

Tanning Salon	Other Pacific Islander	-0.028361	7.102725e-01	American Indian and Alaska Native
Hobby Shop	Other Pacific Islander	0.096657	2.045187e-01	Guamanian or Chamorro
Mattress Store	Other Pacific Islander	0.199689	8.247753e-03	White
Bridge	Other Pacific Islander	0.246033	1.066148e-03	White
Golf Driving Range	Other Pacific Islander	0.186156	1.391670e-02	Guamanian or Chamorro
Recording Studio	Other Pacific Islander	0.233680	1.913619e-03	White
Skate Park	Other Pacific Islander	0.160418	3.447027e-02	Filipino
Gym Pool	Other Pacific Islander	0.133745	7.850816e-02	Black or African American
Nature Preserve	Other Pacific Islander	0.073868	3.327040e-01	White
Garden Center	Samoan	0.108260	1.550427e-01	Asian Indian
Donut Shop	Samoan	0.353808	1.674214e-06	Asian Indian
Costume Shop	Samoan	-0.017834	8.153207e-01	White
Social Club	Samoan	-0.017834	8.153207e-01	White
Electronics Store	Samoan	0.107320	1.586821e-01	American Indian and Alaska Native
Library	Samoan	0.335919	5.856172e-06	White
Waterfront	Samoan	0.603261	1.260515e-18	White
Piercing Parlor	Samoan	0.693195	3.012845e-26	White
Building	Samoan	0.603261	1.260515e-18	White
Dim Sum Restaurant	Samoan	0.432065	2.631487e-09	White
Hawaiian Restaurant	Samoan	0.179610	1.771768e-02	Asian Indian
State / Provincial Park	Samoan	0.327403	1.034800e-05	Asian Indian
Dance Studio	Samoan	0.191834	1.121842e-02	Guamanian or Chamorro
Latin American Restaurant	Samoan	0.410255	1.886438e-08	Korean
Pet Service	Samoan	0.328687	9.506879e-06	White
Island	Samoan	-0.042162	5.806768e-01	White
Bus Station	Samoan	0.180506	1.714941e-02	White
Thrift / Vintage Store	Samoan	0.249897	8.825175e-04	Asian Indian
Southern / Soul Food Restaurant	Samoan	0.379569	2.399303e-07	Asian Indian
Tunnel	Samoan	-0.041806	5.838828e-01	White
Motel	Samoan	0.158641	3.654633e-02	Asian Indian
Cantonese Restaurant	Samoan	0.400397	4.392501e-08	Guamanian or Chamorro
Non-Profit	Samoan	0.403882	3.268199e-08	Guamanian or Chamorro
African Restaurant	Samoan	0.349469	2.284471e-06	Asian Indian
Nightclub	Samoan	0.525702	9.447643e-14	White
General Entertainment	Samoan	0.132460	8.144449e-02	White
Recreation Center	Samoan	0.351215	2.016996e-06	Asian Indian
Storage Facility	Some Other Race	0.233900	1.894290e-03	Japanese
Skating Rink	Some Other Race	0.183454	1.538911e-02	Filipino
Jewelry Store	Vietnamese	0.205196	6.604034e-03	Native Hawaiian
Dentist's Office	Vietnamese	0.057119	4.540807e-01	White
Noodle House	Vietnamese	0.207781	5.938538e-03	Black or African American
Government Building	Vietnamese	-0.032677	6.686145e-01	White
City	Vietnamese	-0.034281	6.533832e-01	White
Fruit & Vegetable Store	Vietnamese	-0.048501	5.250765e-01	White

Karaoke Bar	Vietnamese	0.491585	5.661512e-12	White
Gas Station	Vietnamese	0.389007	1.127821e-07	Korean
Pier	Vietnamese	0.053082	4.866533e-01	White
Shopping Plaza	Vietnamese	0.276405	2.225451e-04	Black or African American
Paper / Office Supplies Store	Vietnamese	0.253233	7.478211e-04	Black or African American
Factory	Vietnamese	0.018889	8.046070e-01	White
Weight Loss Center	Vietnamese	0.403633	3.338485e-08	Black or African American
Greek Restaurant	Vietnamese	0.119572	1.160570e-01	Some Other Race
High School	White	-0.012109	8.739939e-01	American Indian and Alaska Native
Vape Store	White	0.070208	3.572761e-01	Black or African American
School	White	0.055851	4.641761e-01	American Indian and Alaska Native
Mobility Store	White	0.070208	3.572761e-01	Black or African American
Winery	White	0.184060	1.504756e-02	Black or African American
Turkish Restaurant	White	0.165965	2.862204e-02	Black or African American
Other Repair Shop	White	0.097178	2.020788e-01	Some Other Race
Leather Goods Store	White	-0.012522	8.697421e-01	American Indian and Alaska Native
Track Stadium	White	0.207096	6.108851e-03	Chinese
Insurance Office	White	-0.012109	8.739939e-01	American Indian and Alaska Native
Gelato Shop	White	0.044620	5.588013e-01	Some Other Race
Dam	White	-0.031749	6.774938e-01	American Indian and Alaska Native
Bowling Alley	White	0.157191	3.831740e-02	Chinese
Vineyard	White	0.050174	5.108691e-01	Other Asian
Shop & Service	White	0.041224	5.891326e-01	Samoan
Animal Shelter	White	-0.013459	8.600854e-01	Other Asian
Eastern European Restaurant	White	0.155460	4.052584e-02	Some Other Race
Newsstand	White	0.040603	5.947627e-01	American Indian and Alaska Native
Farm	White	0.165587	2.899222e-02	Japanese
Construction & Landscaping	White	0.137202	7.102407e-02	Korean
Comfort Food Restaurant	White	0.006196	9.353257e-01	Chinese
Track	White	0.059824	4.329548e-01	Some Other Race
Knitting Store	White	0.098059	1.979997e-01	American Indian and Alaska Native
Laundromat	White	0.130812	8.534214e-02	Asian Indian
Bridal Shop	White	0.060339	4.289972e-01	Other Asian
Toll Plaza	White	0.175495	2.054428e-02	Chinese
College Bookstore	White	0.086502	2.564006e-01	Black or African American
Arcade	White	0.123819	1.035699e-01	Some Other Race
Persian Restaurant	White	-0.013459	8.600854e-01	Other Asian
Photography Studio	White	0.087666	2.500333e-01	Some Other Race
Entertainment Service	White	-0.005563	9.419261e-01	Other Asian

5. Conclusion

Data science provides the tools to use data to solve real-life problems or gain insights into complex situation. This project is such an example. After data analysis and virtualization, I have better understanding why I it takes me some efforts to get to my favorite Chinese restaurants even though there is a relatively large Chinese population in my neighborhood. Although the above data is for Bay Area, the results can be applied to explain my complaint at the beginning of this project. The relatively strong positive correlation between the number of Chinese restaurants and the Chinese population can explain the relation between the large number of Chinese restaurants in the Houston area and a large Chinese population. However, for an individual area (zip code), there might be a large variation. There are a lot of factors to affect the real-life problems. If we can pull more data about more factors, we can move further to the truth.