THE BATTLE OF NEIGHBORHOODS:

BEST PLACE TO OPEN A DINE-OUT PLACE
IN LOS ANGELES

Lina Sakiene

2021-03-24

TABLE OF CONTENTS:

1. INTRODUCTION

Los Angeles city, the heart of California, is the second-most populous city and metropolitan area after New York city in the United States. At the start of the 20th century, it was considered merely "a large village." Yet it overcame natural deficiencies and established itself as an important center of commerce, agriculture, tourism, and industry.

In such a big and diverse city there are a lot of opportunities for those who want to start a business as well as risks of doing so. Thus, a business case analysis should be done before considering any business idea.

1.1. Problem description

Let's imagine that someone wants to open a dine-out place in Los Angeles. It could be either a restaurant or a café, a pizzeria and so on. First what he needs to do, is to get a better understanding of what kind of venues are the most popular in Los Angeles and where could be the best place to open one.

I am going to analyze this problem based on certain criteria which I consider to be significant to results.

1.2. Target audience

The objective of this study is to locate and recommend the best neighborhood **to anyone who wants to open a dine-out place** in Los Angele and specify what kind of place it should be.

2. DATA

Since the target of this study is to find the best neighborhood in Los Angeles to open a dine-out place, we need to define first, what criteria we are going to use when evaluating neighborhoods.

First of all, I am going to look after the neighborhoods where dine-out places are among the most popular venues. Also, there is a high chance that people with high income are more likely to go for a dine-out, I will check this hypothesis too. And of course, a safe neighborhood is always a more attractive place to start a business.

Besides, we will find out what kind of dine-out place it would be the best to open as we have not determined this yet.

2.1. Neighborhoods in LA

First, we need to get a list of Los Angeles neighborhoods. We can find it on Wikipedia page: http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data. It contains 110 Los Angeles neighborhoods' names as well as some additional measures about each of them.

Luckily, there are geographical coordinates already included so I am not going to need to search for them additionally. Also, I am going to use the Income information. In this case, it is a median household income, which reports the amount of money earned by the household that falls exactly in the middle of the pack. As I mentioned before, I presume that people, who live in a high-income rate neighborhood are more likely to go for a dine-out.

This is a table fragment from Wikipedia page with columns selected I am going to use:

Will use information about Los Angeles neighborhoods from Wikipedia.

```
In [3]: url1 = 'http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data'
df1 = pd.read_html(url1, flavor='bs4')[2]
df1.head()
```

Out[3]:

	LA_Nbhd	Income	Schools	Diversity	Age	Homes	Vets	Asian	Black	Latino	White	Population	Area	Longitude	Latitude
0	Adams_Normandie	29606	691	0.6	26	0.26	0.05	0.05	0.25	0.62	0.06	31068	0.8	-118.300270	34.030970
1	Arleta	65649	719	0.4	29	0.29	0.07	0.11	0.02	0.72	0.13	31068	3.1	-118.430015	34.240603
2	Arlington_Heights	31423	687	0.8	31	0.31	0.05	0.13	0.25	0.57	0.05	22106	1.0	-118.320109	34.043611
3	Atwater_Village	53872	762	0.9	34	0.34	0.06	0.20	0.01	0.51	0.22	14888	1.8	-118.265808	34.124908
4	Baldwin_Hills/Crenshaw	37948	656	0.4	36	0.36	0.10	0.05	0.71	0.17	0.03	30123	3.0	-118.366700	34.019090

Figure 1. Los Angeles' neighborhoods from Wikipedia page

2.2. Los Angeles map

I am going to use a geopy library to get the latitude and longitude values of Los Angeles. Then I will put them in Folium to create a map of Los Angeles with all its neighborhoods marked down:
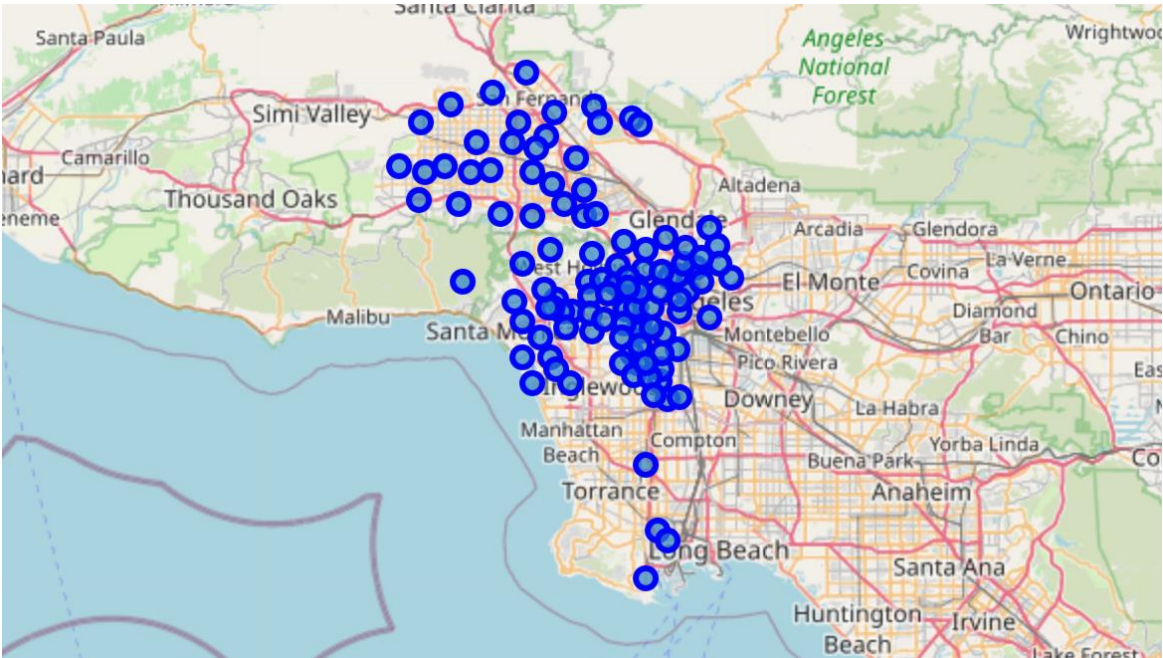

Figure 2. Map of Los Angeles' neighborhoods

2.3. Safety in neighborhoods

Los Angeles Times has a project named Mapping L.A. where certain statistics about Los Angeles' neighborhoods are provided and showed on a map. I am going to use Property Crime Ranking information which indicates the rate of property crimes per 10,000 people from Dec. 30, 2019, to June 28, 2020, the most recent six months for all areas patrolled by the LAPD and L.A. County Sheriff's Department.

I will get the table from Los Angeles Times page by using BeautifulSoup:

Let's get Los Angeles neighborhoods Crime data.

```
In [5]: url2 = 'http://maps.latimes.com/neighborhoods/property-crime/neighborhood/list/'
        df2 = pd.read_html(url2, flavor='bs4')[3]
        df2.head()
```

Out[5]:

	Rank	Neighborhood	Per Capita	Total
0	1	Elysian Park	263.2	70
1	2	Fairfax	251.5	336
2	3	Beverly Grove	243.3	556
3	4	Playa Vista	231.3	139
4	5	Rancho Park	199.7	91

Figure 3. Los Angeles' crime data

2.4. Popular places to eat

I am going to return all venues in 500m. from Foursquare to all Los Angeles' neighborhoods, based on their longitude and latitude. I will be concentrating on the Venue category since a more detailed level about venues is not needed here. Then I will take one hot encoding approach to build a table of frequencies of occurrence of each venue category. This will help me to create a data frame with TOP10 most common venues for each neighborhood:

Out[16]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adams Normandie	Sushi Restaurant	Gas Station	Park	Grocery Store	Playground	Taco Place	Latin American Restaurant	Women's Store	Ethiopian Restaurant	Event Service
1	Arleta	Bakery	Flower Shop	Convenience Store	Historic Site	Video Store	Farm	Escape Room	Ethiopian Restaurant	Event Service	Fabric Shop
2	Arlington Heights	Seafood Restaurant	Shop & Service	Grocery Store	Rental Car Location	Restaurant	Donut Shop	Café	Escape Room	Women's Store	Event Service
3	Atwater Village	Food Truck	Mobile Phone Shop	Chinese Restaurant	Fast Food Restaurant	Ice Cream Shop	Diner	Steakhouse	Spa	Shoe Store	Coffee Shop
4	Baldwin Hills / Crenshaw	Flower Shop	Clothing Store	Women's Store	Escape Room	Food Stand	Food Service	Food Court	Food	Filipino Restaurant	Fast Food Restaurant

Figure 4. TOP10 most common venues in Los Angeles' neighborhoods

3. METHODOLOGY

In this section, I am going to analyze the data I have collected above. The methods I am going to use are:
- Elbow method – to find out how many clusters I should choose for k-means analysis;

- K-means clustering – to cluster Los Angeles' neighborhoods and find where the dine-out places are the most popular;
- Word Cloud – to find out what kind of dine-out places are the most popular;
- str.contains method – to count all the venues in the neighborhood, which are restaurant type, in purpose to calculate the Restaurants' preference rate;
- Pearsons' correlation coefficient – to check if there is a relationship between the median income of the neighborhood and its Restaurants' preference rate;
- Bubble chart – to examine in what neighborhoods dining out in the restaurants is the most popular thing.

3.1. Elbow method

Since I have a TOP10 most common venues list for each neighborhood, now I can cluster them to see, which neighborhoods are popular for dining out. I am going to use a k-means clustering method. But first what I need to do, is to find out how many clusters I should use. I am going to use an Elbow method, which will indicate an optimal number of clusters. It means that adding another cluster will not give much better modeling of the data.
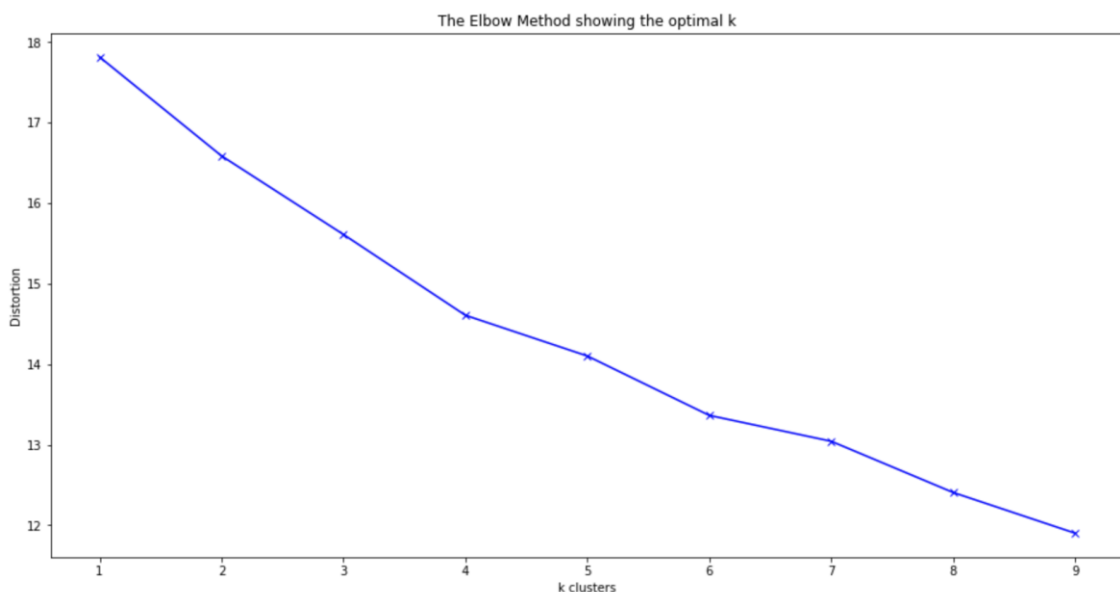

Figure 5. Elbow method chart

3.2. K-means clustering

Since the elbow of the curve is where k cluster is equal to 4, I will run k-means to cluster Los Angeles' neighborhoods to 4 clusters. I will add Cluster Labels to the neighborhoods' table and with help of Folium will visualize the clustered neighborhoods on the Los Angeles map:
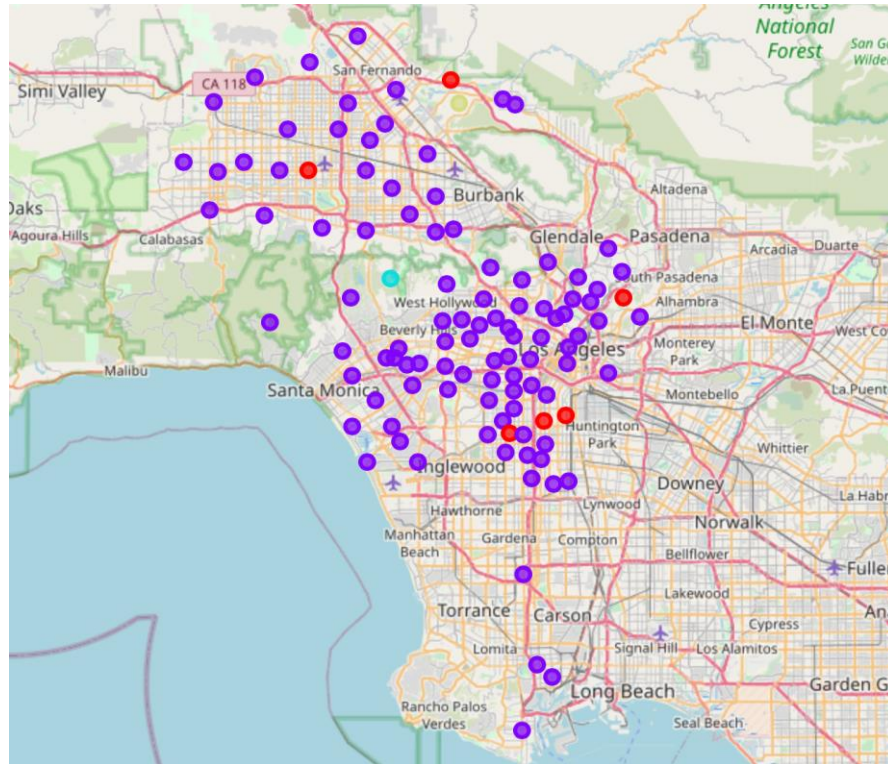
Figure 6. Clustered neighborhoods of Los Angeles

Then I will examine each cluster to determine the discriminating venue categories that distinguish each cluster. It is visible, that the neighborhoods with Cluster Label equal to 1 are popular for dining out. I am going to examine the neighborhoods of this cluster more properly.

3.3. Word Cloud

Let's put all TOP10 most common venues of neighborhoods in cluster 1 together, to see what categories are the most popular in there. I will make a list of the venues categories' names and generate a Word Cloud. It will give us a good visual understanding:

Figure 7. Most popular venues in Word Cloud

3.4. Restaurants' preference rate

On Word Cloud, we can see, that the most popular venue category consists of the word "Restaurant". It can be any kind of restaurant, from a Mexican Restaurant to a Fast-food Restaurant. So, first of all, here we have an answer, what kind of dine-out place would be the best to open – a Restaurant. Second, I am going to stick to this venue category for my further analysis.

I will add an additional column to my clustered neighborhoods table. It will show how many venues, among the TOP10 most common venues in that neighborhood, are a restaurant type. In other words, it is going to show how strong people prefer restaurants in that neighborhood among other dine-out places. For simplicity, let's call it "Restaurants' preference rate".

3.5. Pearson's correlation coefficient

As I have mentioned in the introduction section, there is a high chance that people with high income are more likely to go for a dine-out. Let's check this hypothesis and look for a relationship between neighborhoods' Income rate and their Restaurants' preference rate.

First, I need to normalize data, because Income data varies between ten-to-hundred-thousands, and the Restaurants' preference rate varies between 0-10. I will use Sklearn preprocessing for this purpose. After the data is normalized, we can calculate a Pearson's correlation coefficient:

Out[37]:

	restaurants	income
restaurants	1.000000	-0.012711
income	-0.012711	1.000000

Figure 8. Relationship between Income and Restaurants' preference rates

As the Pearson's correlation coefficient is equal to –0.012711, it shows, that there is no relationship between the median income people earn in that neighborhood and their preference of going to the restaurants.

3.6. Bubble chart

A safe neighborhood is always an attractive place to start a business. So, let's check which neighborhoods could be our target by plotting a bubble chart. The size of the bubble will show the neighborhoods' Restaurants' preference rate.
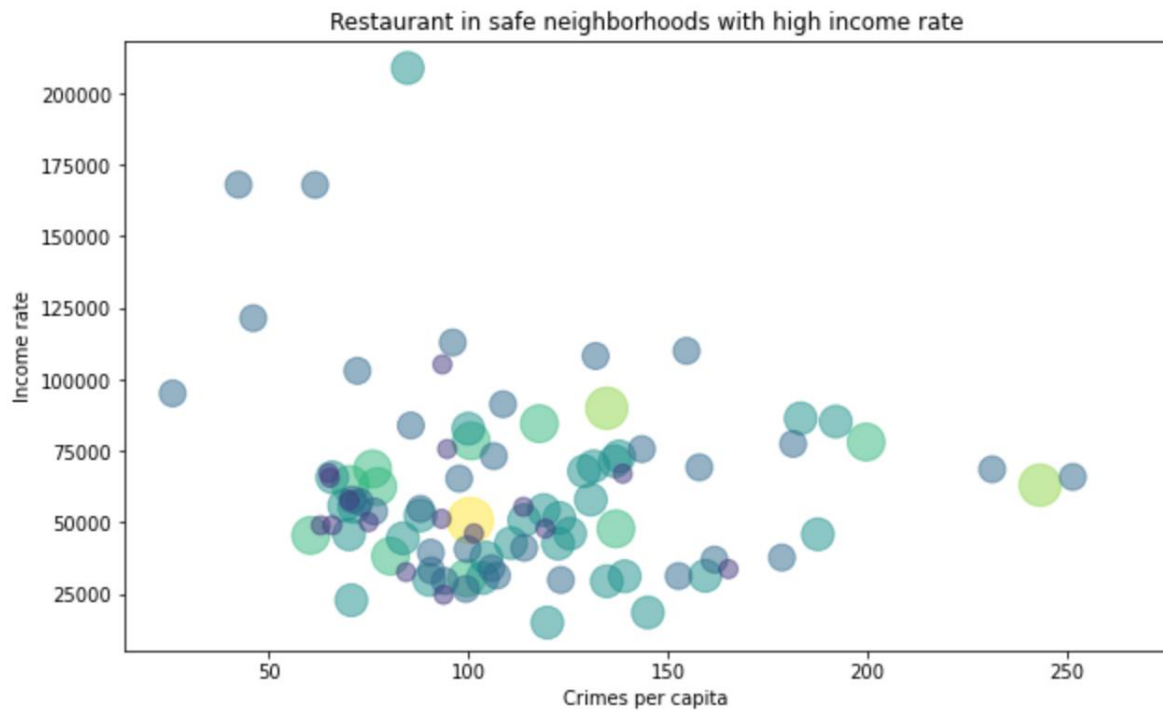
Figure 9. Bubble chart of restaurants' popularity

This chart confirms once again that there is no relationship between the neighborhoods' income rate and their Restaurants' preference rate. But we can definitely see that the Restaurants are more popular in safe neighborhoods, where crimes per capita are smaller.

4. RESULTS

Based on the analysis above, we can state, that the best dine-out place in Los Angeles would be a restaurant. Also, here is a list of 12 Los Angeles' neighborhoods, where could be the best place to open one:

	Neighborhood	Income	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	Palms	50684	Thai Restaurant	Coffee Shop	Mediterranean Restaurant	Mexican Restaurant	Gym	Burger Joint	Theater	Indian Restaurant	Cuban Restaurant	Italian Restaurant
1	Woodland Hills	89946	Sushi Restaurant	Ice Cream Shop	Mediterranean Restaurant	Burger Joint	Shipping Store	Chinese Restaurant	Pharmacy	Bagel Shop	Indian Restaurant	Mexican Restaurant
2	Chatsworth	84549	Mexican Restaurant	Gym / Fitness Center	Hotel	Beer Bar	Thai Restaurant	Sushi Restaurant	Fast Food Restaurant	Dry Cleaner	Supermarket	BBQ Joint
3	Encino	78529	Shipping Store	American Restaurant	Mexican Restaurant	Clothing Store	Salon / Barbershop	Coffee Shop	Sushi Restaurant	Seafood Restaurant	Bank	Bakery
4	Sunland	68720	Fast Food Restaurant	Asian Restaurant	Convenience Store	Grocery Store	Sandwich Place	Mexican Restaurant	Pizza Place	Coffee Shop	Farmers Market	Filipino Restaurant
5	Del Rey	63317	Mexican Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Sushi Restaurant	Sandwich Place	Coffee Shop	Latin American Restaurant	Bus Station	Bakery	Donut Shop
6	Winnetka	62535	Ice Cream Shop	Latin American Restaurant	Food	Fried Chicken Joint	South American Restaurant	Filipino Restaurant	Mexican Restaurant	Grocery Store	Bar	Home Service
7	Larchmont	47780	Spa	Park	Indie Movie Theater	American Restaurant	Korean Restaurant	Farmers Market	Ethiopian Restaurant	Event Service	Fabric Shop	Falafel Restaurant
8	Highland Park	45478	Pizza Place	Fast Food Restaurant	Romanian Restaurant	Breakfast Spot	Gourmet Shop	Asian Restaurant	Liquor Store	Grocery Store	Vietnamese Restaurant	Arcade
9	West Adams	38209	Seafood Restaurant	Wine Bar	Sandwich Place	Pizza Place	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Dumpling Restaurant
10	Koreatown	30558	Korean Restaurant	Coffee Shop	Asian Restaurant	Hotel	Ice Cream Shop	Bakery	Seafood Restaurant	Mobile Phone Shop	Café	Restaurant
11	Bel-Air	208861	Scenic Lookout	Dive Bar	Women's Store	Farm	Ethiopian Restaurant	Event Service	Fabric Shop	Falafel Restaurant	Farmers Market	Empanada Restaurant

Figure 10. List of recommended neighborhoods

5. DISCUSSION

When making a final list of recommendation of where to open a restaurant, I was concentrating on the densest part of a bubble chart and picked the biggest bubbles on it:
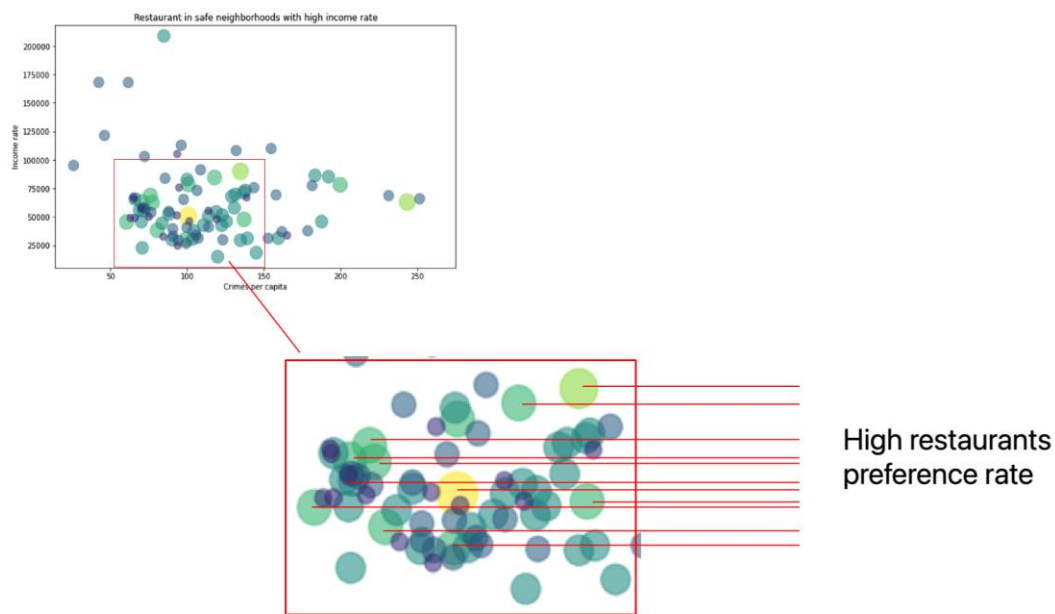

Figure 11. Bubble chart analysis

It represents the criteria applied to evaluate the neighborhoods of Los Angeles. The neighborhood, recommendable to open a restaurant, should:

- be from cluster 1, where dine-out places are popular;
- have a high restaurants' preference rating (at least 4 out of 10 most popular venues in that neighborhood are restaurant type);
- be a quiet safe neighborhood (crimes per capita are less than 150).
- ★ I also added to this list a neighborhood, which has a bit smaller Restaurants' preference rating (equal to 3), but is in a safe neighborhood and has an extremely high-income rate.

Other approaches can be considered and applied in the further analysis as well, since not the highest Restaurants' preference rate could also mean that there are just no good enough restaurants in that neighborhood. And it is a good opportunity to open one there.

Also, the analysis showed that there is no relationship between median income in the neighborhood and the its Restaurants' preference rate. It could be because of the fact, that people from other neighborhoods, cities, or even countries can go to the certain restaurants, so it totally does not matter, what households' median income is in the neighborhood where that restaurant is.

6. CONCLUSION

If someone wants to start a business in such a big and diverse city like Los Angeles, a comprehensive analysis should be done at the beginning. To open a dine-out place, first, you have to find an answer to two main questions: what kind of dine-out place it should be and where could be the best place to open one?

I have analyzed Los Angeles neighborhoods trying to answer these questions. With help of Foursquare and Python libraries' capabilities, I succeed to determine the neighborhoods in LA, where are popular to go for a dine-out. Also, I had a chance to find out, that the Restaurants are the most popular among other dine-out venues.

After applying some criteria, like the safety of the neighborhood and its Restaurants' preference rate, I simply could generate a list of the 12 most recommended neighborhoods in Los Angeles to open a Restaurant.