Best Location to Open a Chinese Restaurant in Amsterdam

Capstone Project - Final Assignment

Table of Contents

1. Introduction: The Challenge

A good friend of mine, Linda (a nickname), whose dream is to open a Chinese restaurant one day to share the joy of good food with others. Now, everything is ready, she can finally realize her dream. She has chosen Amsterdam to be the place that her dream starts. Not only because Amsterdam is one of the most populous and visited cities in Europe but also because the diverse culture the city embraces.

My challenge as a data scientist is to help her to find the ideal location in Amsterdam using whatever data I can access.

1.1. A Cooperative Iterative Approach

It's important to emphasize that even though I might know how to deal with data better, Linda definitely knows food business a lot more. Throughout this project, we will work closely together to take full advantage of our expertise.

The process is also iterative. Not all decisions made at the beginning of the project remains the best ones. We both fully understand that when we learn more about Amsterdam and the its residences, we need to continuously re-examine and fine tune our decisions.

1.2. About Amsterdam

Amsterdam is the capital city and most populous municipality of the Netherlands. Here are some facts about its residences:

- Amsterdam has a population of **854,047** within the city proper
- Amsterdam city proper has **4,457** inhabitants per square kilometer and **2,275** households per square kilometer.
- Amsterdam has more than **100 kilometers (60 miles)** of canals, most of which are navigable by boat.

As to tourism, Amsterdam is one of the most popular tourist destinations in Europe

- Number of international tourists per year: **20.63 million**.
- Out of which, the number of day-trippers is: **16 million**.

Reference: https://en.wikipedia.org/wiki/Amsterdam

1.3. Business Questions

To find the ideal location for the restaurant, we must first seek answers to a few questions.

Question 1: How many restaurants already exist?

If this new restaurant would be the only one in a neighborhood, there will be more profit for Linda. So, the number of existing restaurants in the neighborhood must be taken into consideration. Again, this question can be answered, hopefully, by using Foursquare API.

Question 2: How popular will Chinese food be in the neighborhood?

For Linda, it's important to serve traditional Chinese food the way she knows. Even though Chinese food is widely loved, it makes sense to double check how existing Chinese restaurants (or Asian restaurants) are perceived. This question can be answered, hopefully, by using Foursquare API.

Question 3: Who are the target customers and where do they live?

It is going to be a small restaurant (5 to 7 tables) due to the limited investment. The primary income would be takeout, and orders made online. From past experience, Linda knows that people who live alone are more likely to buy takeout or use online food ordering apps such as Uber Eats. They are the ideal target customers for her new restaurant. So, we will look for an area with a relatively high density of one-person household. We need demographic information to answer this question.

2. Data Requirements

We have the big question, where to open a Chinese restaurant in Amsterdam. Now, we need to collect data that can help us answer the questions. We need to collect data from at least two sources:

- Data of the surroundings (density of similar restaurants nearby)
- Demographic data (per area in Amsterdam)

2.1. Data of Surroundings

Foursquare API. Foursquare API Endpoints that we need:

- **Search**: Returns a list of venues near the current location, optionally matching a search term.
- **Category ID**: Crucial information to accurately query for one specific type of places.
- **Trending**: Returns a list of venues near the current location with the most people currently checked in.

Information isn't available for a free account

- **Number of Likes**: Assumption: the more voting, the more the restaurant is visited. It can be used to see how often
    - Unfortunately, this information needs a premium account.

2.2 Demographic Data

In order to know which neighborhoods are more interesting to investigate further, we need to look into demographic data to pick the neighborhoods that have more target customers in terms of quantity and density.

The Central Bureau of Statistics of the Netherlands, **CBS** in short, provides a large number of demographic data regarding who live and work in the Netherlands.
In this page, you can choose whatever feature you need to solve the problem. The list of features is quite comprehensive. It is important, therefore, to define exactly what the data can be used.

Features that can be used to answer the question

- **Regional specifics** *(Regioaanduiding)*: This information can help me to link relative details to a specific area.
- **Total Households** *(Particulier huishouden)*: Number of households in a neighborhood.
- **One-person Households** *(Eenpersoonshuishoudens)*: Number of the households that with only one person.
- **Population density** *(Bevolkingsdichtheid)*: A more densely populated area means more customers for a restaurant. The unit of population density is **number of people per square kilometer**.

Information that isn't available

- **Origin of birth** *(Personen met een migratieachtergrond)*: This information can help us to determine the types of food the restaurant should offer.
    - Unfortunately, the categorization is not detailed enough. I am not able to single out people come from China in the data.
- **Income per household** *(Inkomen van huishoudens)*: Since restaurants are usually quite expensive comparing to home-cooking. Only people who have sufficient income can afford to go to restaurants often.
    - Unfortunately, this information is missing consistently from the database.

Features deliberately excluded

- **Civil status**: Since we can distinguish one-person households, it's not necessary to understand why people live alone (single or divorced does not seem to link to food strategy directly).
- **Gender**: Reason, in the Netherlands, there isn't a big difference between men and women in terms of the likeliness of cooking at home.
- **Type of house**: This has no direct correlation to people's choice of food.

3. Analyze Demographic Data

3.1. Load and Prepare Data

I selected necessary data from CBS, as mentioned in chapter 2.2. The data is in CSV format.

I had done the following to prepare the data in the CVS file to be ready to use in further analysis:

- Load CSV file to a dataframe.
- Sort the neighborhoods by the population density (The most densely populated neighborhood at the top)
- Drop unnecessary data (such as Total Residences, since it is useful to answer any of the three questions mentioned earlier)

The Result Dataframe contains the following information:

- Neighborhood
- ID (of the neighborhood)
- Total Households
- One-person Households
- Population Density
- Lat, Lon

	Neighborhood	ID	Total Households	One-person Households	Population Density	Lat	Lon
14	Staatsliedenbuurt	WK036314	8105	4860	28139	52.380287	4.870951
19	Van Lennepbuurt	WK036319	4535	3005	28005	52.365144	4.867845
31	Indische Buurt West	WK036331	7060	3930	26985	52.361625	4.938813
21	Overtoomse Sluis	WK036321	4840	2910	26482	52.359468	4.860689
18	Kinkerbuurt	WK036318	3950	2460	26135	52.369167	4.866649

Figure 1: The first 5 rows of the dataframe

3.2. Observe Data using a Bar Chart

In order to better decide what to do with the data, I want to take a good look at the data. Visualizing the data will help a lot. I choose to use a horizontal bar chart, because I want the neighborhood names to be very easy to read. Due to the number of neighborhoods (65), the vertical bar might not offer enough room to show all the bars.

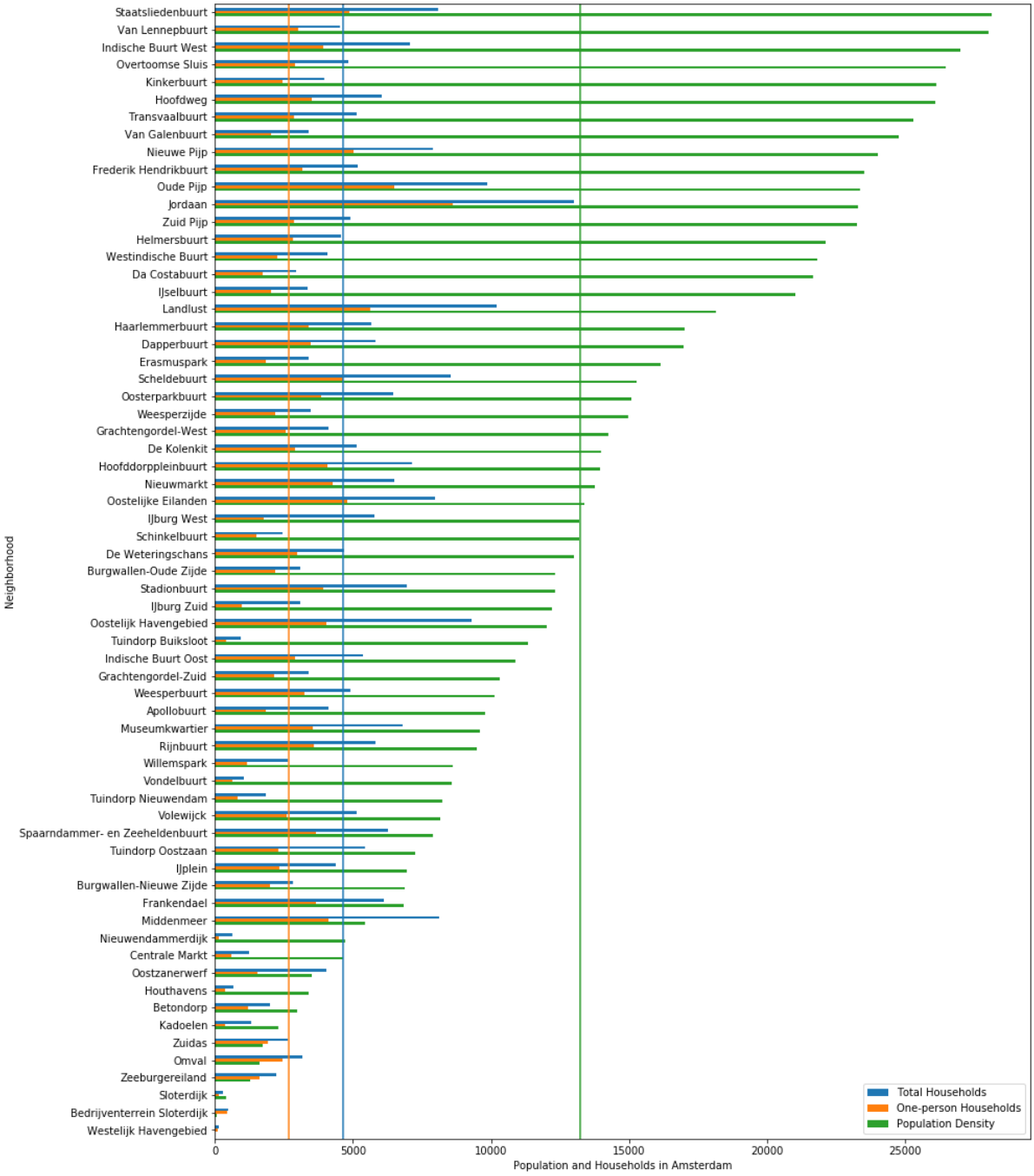

Figure 2: Bar Chart of Total Households, One-person Households and Population Density per neighborhood

3.3.1. Learnings from the Above Bar Chart

1. One thing becomes evident is that many neighborhoods in Amsterdam has much lower population density than the average value. These neighborhoods are very unlikely to be ideal

location to open the restaurant. Therefore, we should remove them to focus on the neighborhoods that are more densely populated. We will do that in the next chapter.

2. The second learning is that amongst the more densely populated neighborhoods, not all of them have above average total number of households and one-person households. We will do further analysis in chapter 3.4 to filter out the neighborhoods that do not have enough one-person households.

3.3. Remove Neighborhoods that have below average population density

Average population density of Amsterdam city proper is: 13233

Number of neighborhoods that has higher than average population density: 29

3.4. Neighborhoods that Have More One-person Households

In order to do so I did the following:

Step 1: Calculate the percentage of one-person households of every neighborhood and include this information in a new column “Percentage of one-person households”

	Neighborhood	ID	Total Households	One-person Households	Population Density	Lat	Lon	Percentage of One-person Households
14	Staatsliedenbuurt	WK036314	8105	4860	28139	52.380287	4.870951	59.96
19	Van Lennepbuurt	WK036319	4535	3005	28005	52.365144	4.867845	66.26
31	Indische Buurt West	WK036331	7060	3930	26985	52.361625	4.938813	55.67
21	Overtoomse Sluis	WK036321	4840	2910	26482	52.359468	4.860689	60.12
18	Kinkerbuurt	WK036318	3950	2460	26135	52.369167	4.866649	62.28

Figure 3: The first 5 rows of the result

Step 2: Sort the dataframe by the Percentage of One-person Households. (the highest percentage on top)

Step 3: Extract only the necessary information to create a bar chart.

	Neighborhood	Percentage of One-person Households
6	Jordaan	66.42
19	Van Lennepbuurt	66.26
4	Nieuwmarkt	66.08
24	Oude Pijp	65.92
25	Nieuwe Pijp	63.44

Figure 4: The first 5 rows of the result

Step 4: Plot the bar chart that shows the percentage of one-person households per neighborhood in descendance order.

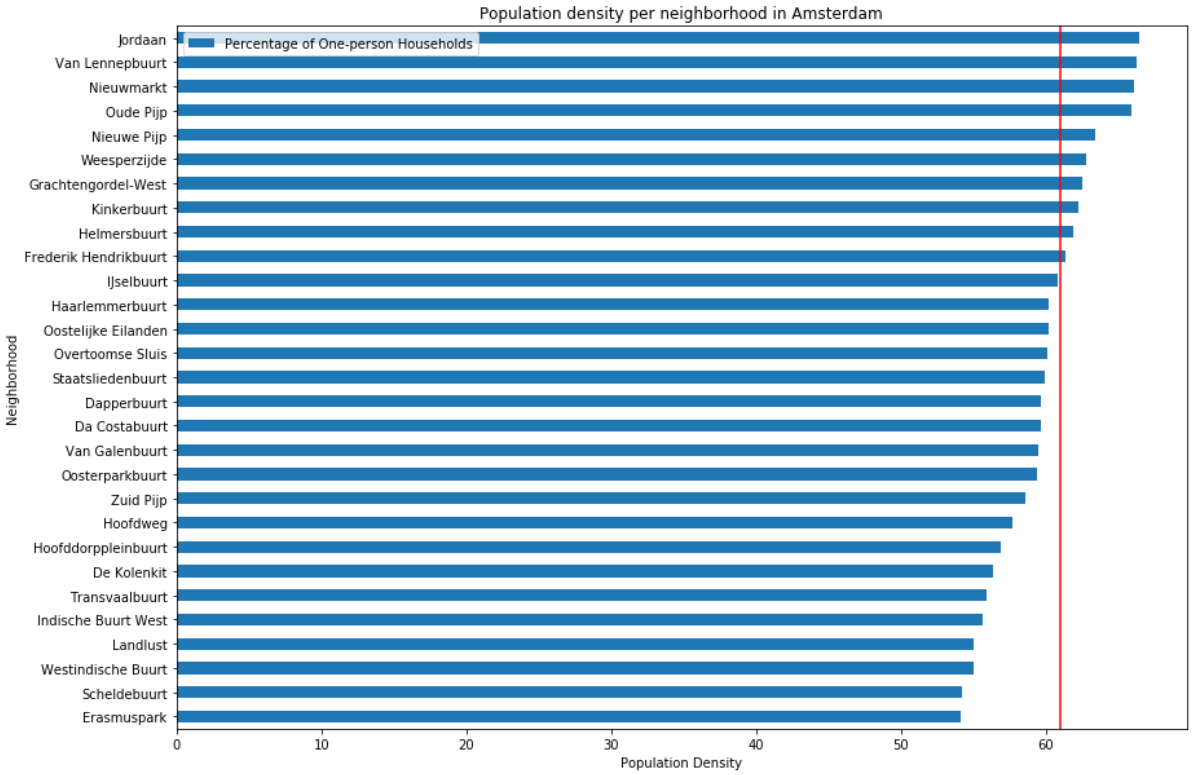

Figure 5: The bar chart of one-person households per neighborhood

3.5.1. Learnings from the Above Bar Chart

As mentioned after we examine the data in a bar chart (chapter 3.2). It seems that the percentage of one-person households (in comparison to the total number of households) seems to be rather consistent for the neighborhoods that have above average population density.

Now, in the above bar chart, the consistency becomes quite clear.

However, we observe there are roughly 3 ranges of the percentage:

1. High: 66% ~ 67%. The first four neighborhood
2. Medium: 59% ~ 63.5%.
3. Low: 53% ~ 59%. The last 10 neighborhood

We decided to focus on the **top 10**, the neighborhoods that have **over 61 percent of one-person households** (Marked by the red line in *Figure 5*).

3.6. View Candidate Neighborhoods on a Map

The next step is to examine the geospatial information of the remaining 10 neighborhood and visualize the following information on the map:

- Total households
- One-person households
- Population density

The result map is:

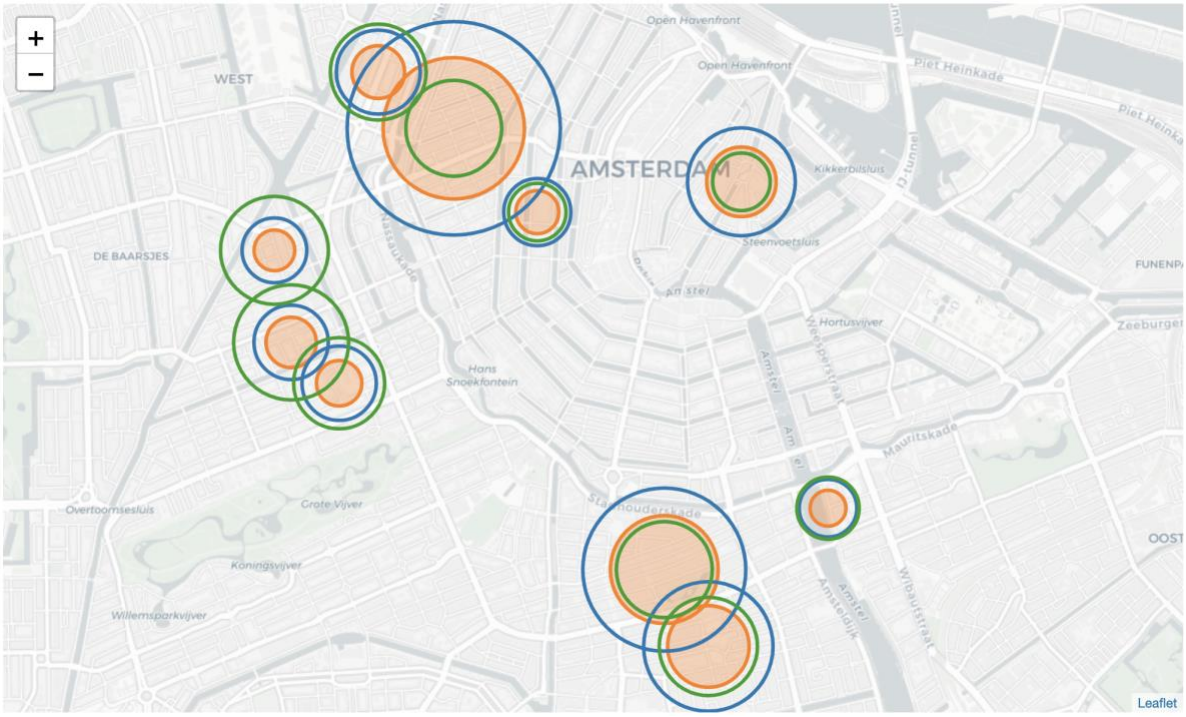

Figure 6: Examine (geospatial) information on the map

The above map gives us an impression of how the numbers of one-person households per neighborhood looks like.

1. **Orange circles** represent the number of one-person households.
2. **Blue circles** represent the number of households in total.
3. **Green circles** represent the population density.

As to the circles:

- The center of the orange, green, and blue circles is the center of the neighborhood. Click the center of the circles to see the name of the neighborhood.
- The radius of each circle represents the number of each feature.

3.6.1. Learnings from the above data visualization

As you can see, when we choose an ideal location to open the Chinese restaurant:

- The bigger the green circles the better.
- The less difference between the size of the blue circles and the orange circles the better.

3.7. Conclusion of Demographic Data Analysis

Through the above analysis, we have chosen **10** out of **65** neighborhoods in Amsterdam city proper as our candidate neighborhood to investigate further.

In the next chapter we will further analyze the **10** neighborhoods by looking into the number of restaurants and the density of Chinese restaurants to further narrow down to **3 ~ 5** neighborhoods for future analysis.

4. Explore the Surroundings

As mentioned earlier, we are going to use **Foursquare API** to continue our exploration of the remaining **10** candidate neighborhoods.

Foursquare API Endpoints that we need:

- **Search**: Returns a list of venues near the current location, optionally matching a search term.
- **Category ID**: Crucial information to accurately query for one specific type of places.
- **Trending**: Returns a list of venues near the current location with the most people currently checked in.

4.1. Show Chinese Restaurants in Candidate Neighborhoods on the Map

One very important criterion when deciding where to open a new Chinese restaurant is, how many existing Chinese restaurants are there already. Therefore, it's very important for us to see the number and distribution of existing Chinese restaurants in the candidate neighborhoods.

4.1.1. Get Information from Foursquare API and convert it to a Dataframe

I have done the following to achieve the goal:

- Define all necessary information needed for the GET request:
  - Foursquare credentials
  - Version
  - Radius
  - Limits (The maximum search result for a free account is 50)
  - GPS coordinates
  - Category ID
- Send the GET Request and examine the results.
- Get relevant part of JSON and transform it into a pandas dataframe.
- Transform venues into a dataframe.

The result dataframe is as following:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhc
0	Sapporo Ramen Sora	Ramen Restaurant	Ceintuurbaan 49 H	NL	Amsterdam	Nederland	NaN	492	[Ceintuurbaan 49 H, Amsterdam, Nederland]	[{'label': 'display', 'lat': 52.3525348486028,...	52.352535	4.888807	Oude I
1	Sushi Op De Markt	Sushi Restaurant	Albert Cuypstraat 172	NL	Amsterdam	Nederland	NaN	60	[Albert Cuypstraat 172, 1072 CN Amsterdam, Ned...	[{'label': 'display', 'lat': 52.3557292653159,...	52.355729	4.894302	N

Figure 7: Dataframe of Foursquare information

4.1.2. Visualize Chinese Restaurants on the Map

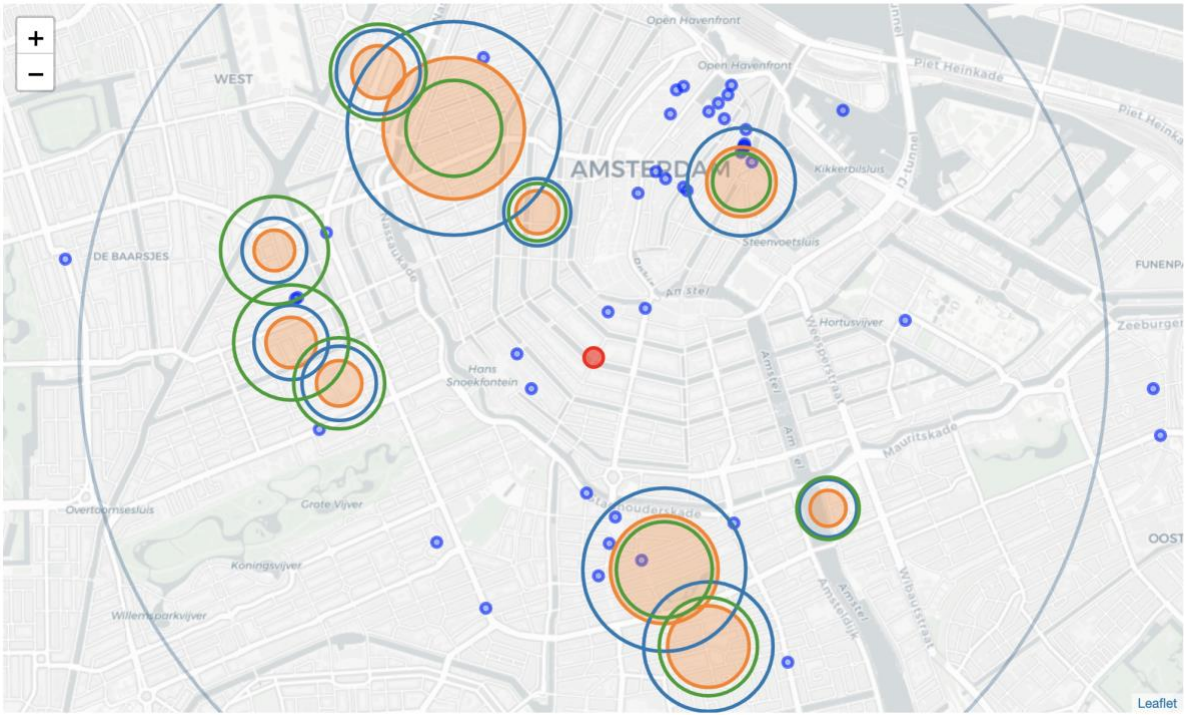

Figure 8: The top 50 Chinese restaurants in Amsterdam city proper

Legends of the above map

1. **Orange circles**: the number of one-person households.
2. **Blue circles**: the number of households in total.
3. **Green circles**: the population density.
4. **Red circle** in the center: the center of the map.
5. **Grey circle**: the search radius.

4.1.3. Learnings from the above map

Our criteria to further narrow down our choice of neighborhoods are as such:

- There must be at least one existing Chinese restaurant in or near the neighborhood. Because if there isn't at least one Chinese restaurant, it might mean that there isn't such demand. Opening a Chinese restaurant without understanding why there isn't any existing ones would be riskier.
- There cannot more than 10 existing Chinese restaurants.
  The more existing Chinese restaurants means more competition.

If we apply the criteria, from the above map, we can exclude these neighborhoods:

- Too many existing Chinese restaurants
  - Nieuwmarkt
- No existing Chinese restaurant in or near the neighborhood
  - Weesperzijde
  - Frederik Hendrikbuurt
  - Grachtengordel-West
  - Nieuwe Pijp

4.2. In-depth Analysis of a Neighborhood

Let's use Oude Pijp as an example to show how we look into one particular candidate neighborhood. To achieve the goal, I had done the following:

- Get the coordinates of the center of Oude Pijp
- Query for all Asian and Chinese restaurants within reachable distance

  We want to extend our query to restaurants that serve all sorts of Asian food. And observe how many Chinese restaurants in comparison of all Asian restaurants. The reachable distance means that people live within this distance are more likely to walk to dine in a restaurant or get takeout food. For this project, we define the reachable distance as **500 meters (about 0.3 mile)**.

- Try to get the Most Popular Venues (via Trending offered by Foursquare)
  - The Foursquare "Get Trending Venues" returns a list of more liked places within the given radius
  - Unfortunately, the Get Trending Venues around Oude Pijp returns 0 result.

- Draw all the Asian and Chinese restaurants on the map

The result map is:

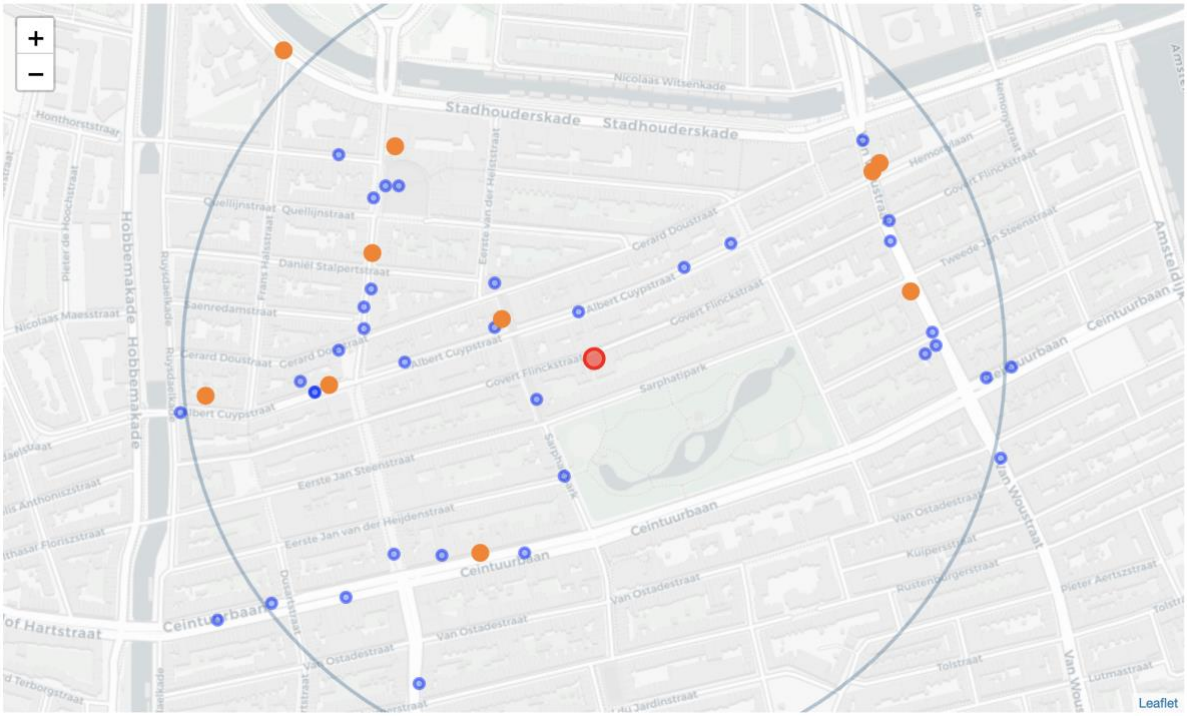

Figure 9: Asian restaurants and Chinese restaurants in Oude Pijp

Legends of the above map

- **Orange dots**: Chinese restaurants.
- **Blue circles**: Asian restaurants.
- **Red circle**: the center of the map.
- **Grey circle**: the search radius.

4.3.1. Learnings of the map above

1. According to Foursquare there are 10 Chinese Restaurants within the range of 500 meters in the neighborhood Oude Pijp.
2. As to Asian restaurants. Due to the 45 limits, we can only show 45 Asian restaurants here. The actual number could be higher.
3. From the map we learn that in the North and Southeast of Oude Pijp seems to be a void of all restaurants. If we open a Chinese restaurant there, we will likely have enough customers.

4.3. Repeat the In-depth Analysis

We can see the value of the in-depth analysis. Later on, we will repeat this analysis to all other remaining candidate neighborhoods. We will compare more categories of venues. Such as, Chinese restaurants and all other restaurants, café, snack bars, etc.

5. Conclusion and Future Work

5.1. The limitation of this project

Only focus on residential information

This project is limited by the lack of crucial information. So far we have been focused quite a lot on residences information and one-person households. However, customers can also come from nearby business venues. We are unable to validate any assumption or answer any questions, because the information of business venues in Amsterdam is not as available as demographic information.

Rent of a venue is not taken into consideration

Due to lack of information, we are unable to include rental price as part of the analysis. The rent is a large part of costs for a restaurant. In order to be able to predict the potential profit, it is crucial to include potential rental price.

Lack of understanding of how popular existing Chinese restaurants are

For a free Foursquare account, information that helps us understand how popular a venu is is not accessible. Information such as:

- **Likes**: how many people like a place
- **Rating**: star rating of a place

5.2. Next Step

- Continue to do more in-depth analysis as mentioned in chapter 4.3.
- Include rental price of each neighborhood in future analysis.