

## Opening a new chinese restaurant in London, United Kingdom



Tim Walther

## **Introduction**

London is the capital of England with lots of tourism. Also, there are a lot of restaurants in the city for all kinds of food. To open a chinese restaurant in London is a tough challenge. Therefore, we need to find a good location for the restaurant.

## **Business Problem**

The goal of this project is to find the best location for our new chinese restaurant.

We use data science methodology and machine learning techniques to find the best location and conclude with a recommendation for the best locations.

## **Which data do we use?**

To achieve or business goal, we use the following data:

1. List of neighbourhoods in London, Germany
2. Latitude and longitude of the neighbourhoods
3. Data related to restaurants in London, UK

## Data sources

We use the wikipedia page

“[https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)” to scrape the neighbourhoods of London. We use the python packages request and beautifulsoup to webscrape the wiki page.

We get the Latitude and longitude of the neighbourhoods by using the geocoder package, which gives us the coordinates of the neighbourhoods.

Then, we use the Foursquare API to get data related to restaurants in London.

## Methodology

We use the data from the wikipedia page

[https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London) to get the list of neighborhoods in London. We will webscrape this list with the python packages beautifulsoup and request.

When we have the list, we need the latitude and longitude of the neighborhoods. We get them by using the geocoder package for python.

After we gathered the data, we will convert the data into a python dataframe and create a map, where we visualize the neighborhoods.

For the map, we use the folium package.

Then, we need to know the exact positions of other chinese restaurants in London. We get this data by using the Foursquare API.

Foursquare data comes in JSON format. We extract the venue name, venue category, venue latitude and venue longitude from it.

At least, we use k-means with  $k=5$  to examine the different locations of chinese restaurants, and conclude based on this examination, which locations are best for opening a new chinese restaurant in London.

## Results:

We examined 5 clusters. Most competition is in cluster 0 and cluster 2. In cluster 4, there is low competition, and in cluster 1 and 3 no competition at all. So I recommend opening a new chinese restaurant within cluster 1, 3, or 4.

London map with clusters:

