1 Introduction

People in Italy love to dine in restaurants and enjoy the moment of being in touch with other ones. But since the COVID-19 pandemic striked into the country things started to get out of control. Fake news were popping out ever more often. Sadly, most of them were about Chinese people. Demonizing and putting them on the edge. To the point were, not only shops, but even restaurants started to close [1] and have fewer and fewer reservations [2] hitting a -43% decrease due to all these fake news.

In Italy there was an 18% increase in people coming from Mainland China from 2013 to 2019 [3]. So the interest in buying properties, and to spend their life in Italy, is actually rising. Therefore we can bet that there will still be people that want to open a new restaurant in a wonderful city like Milan. Opening a business in such a big city can be quite burdensome, given that position is crucial for two reasons. Firstly, it gives visibility to the place in such a way to keep up with the competition. Secondly, since Milan it's one of the biggest city in Italy, with approximately 1.4 million of residents, and an average density of about 7000 ab/km^2, the housing costs can skyrocket the closer the place is to the city center. Thus the location is actually one of the key parameters to look after.

1.1 Business Problem & Target Audience

The objective of this project is to find the best location in the city of Milan to open a new Chinese restaurant. We are going to use some machine learning techinques, such as K-means for clustering, and some Data Science methods to clean up and prepare the dataset. Thus this project is aimed not only to business owners, but to investor too that could ask themselves *"Where should I open my traditional chinese cusine restaurant? "*

1.2 Data

In order to solve this problem we need:

1. List of neighborhoods in Milan. This is the focus of the project. Milan is one of the most multi-cultural cities across Italy, situated in Nothern part;

2. Latitude and longitude of these neighborhoods to plot a map and get the venues data;

3. Venues data about the actual number of chinese restaurant in the city. This will allow us to cluster the neighborhoods and get a classification based on the presence or not of similar restaurants.

We will use some web scraping techniques to get the neighborhoods from a Wikipedia page [4]. Using the beatifulsoup library we will handle the request and then clean it. Afterwards we are going to use the Python Geocoder package to associate to each neighborhood a precise latitude and longitude. Finally we will use Foursquare APIs to get the venues data that we need proceding thus to use K-means to cluster the data. All of this will be represented graphically using Folium.