

Asian Restaurant Borough Business Problem

HAZEM KAMEL

March 05 , 2020

1. Introduction

1. Problem Description:

In this project, the problem attempted to solve will be to find the best possible location or the most optimal, for an Asian restaurant in the city of London, England. To achieve this task, an analytical approach will be used, based on advanced machine learning techniques and data analysis, concretely clustering and perhaps some data visualization techniques.

During the process of analysis, several data transformations will be performed including the number of asians per borough , working hours per borough , Asian restaurants already opening per borough and earnings per borough, in order the find the best possible data format for the machine learning model to ingest. Once the data is set up and prepared, a modeling process will be carried out, and this statistical analysis will provide the best possible places to locate the Asian restaurant.

2. Data Presentation

The Foursquare Api: This data will be accessed via Python, and used to obtain the most common venues per neighborhood in the city of London. This way, it is possible to have a taste of how the city's Asian restaurants are distributed, what are the most common places for an Asian restaurant to open, and in general, it will provide an idea of what people's likes are.

data.london.gov Ethnic groups: This site provides information about ethnicity of population in London which is of great utility to solve this problem. The data is uploaded as an excel file and the Asian population of London is converted into DataFrame. The data contains information about the immigrant population per borough and per nationality. This data will be analyzed in such a way that one could determine the best location of a new Asian restaurant based on people's nationalities. For the sake of simplicity, it will be assumed for this exercise that people's likes vary according to their nationality, and that people from one specific country will be more attracted to places that match the environment and culture of their own countries, rather than the ones from foreign countries.

data.london.gov Earnings per hour data: This site provides information about earnings per hour of the population in London which is of great utility to solve this problem. The data is uploaded as an excel file and the earning per hour of London's population is converted into DataFrame. The data contains information about the earnings per hour and per borough. This data will be analyzed in such a way that one could determine the best location of a new Asian restaurant based on people's earnings. For the sake of simplicity, it will be assumed that people could spend more money in eating outside if their earning is higher than people who have a less earning job per hour.

data.london.gov Working hours per borough data: This site provides information about working hours per borough of the population in London which is of a great utility to solve this problem. The data is uploaded as an excel file and the working hours per borough of London's population is converted into DataFrame. The data contains information about the working hours per borough. This data will be analyzed in such a way that one could determine the best location of a new Asian restaurant based on people's working hours. For the sake of simplicity, it will be assumed that people would eat outside in a restaurant if they have more working hours rather than a less working hour job which they will have more flexibility and time to eat at home.

3. Target Audience

The target audience of this project could be any business owner that is planning to open a new business local, restaurant, real estate agency, shops, etc... Since this approach could be applicable not only to Asian food restaurants but to other kinds of businesses, anybody who is considering placing a new business local or even relocating it, could benefit from this project's approach.

4. Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of wrong values from Geopy. I changed the wrong values with the right geo-location values from google maps, I removed all Nan values found in any dataframe.