Coursera Capstone Project

# Weeks 4 & 5

# Korean Restaurants in Los Angeles

## Dougie Scott

## 06/01/2021

# Contents

Table of Contents

**Introduction and Business Problem3**

**The Data and our Approach to the Problem4**

Suggested Outputs4-5

# Introduction & Business Problem

We work for a firm of freelance Data Scientists and we have been approached by a firm called “Kor That’s Tasty” (KTC), who operate a chain of Korean Restaurants. They offer a full Korean menu, including all variants of Korean food as well as standard alcoholic and non-alcoholic beverages. They operate both sit-in and delivery services from their restaurants.

For now, KTC are not active in the United States, but they have plans to ambitiously expand into the US. They have identified Los Angeles as the city they would like to start in, given its size and convenience for them in relation to current operations and supply chains.

They have approached our firm of Data Scientists to deliver for them a report on the existing situation with regards restaurants in Los Angeles. They have asked us to return to them with some recommendations on which Districts/Neighbourhoods to open their first restaurant(s), and wish us to keep the following in mind:

* They do not want strong competition from other nearby Korean Restaurants, and would therefore like to set up in an area with few Korean restaurants nearby
* They are not overly concerned if there are other restaurants nearby, provided these are not Korean restaurants
* They want to set up in an area with reasonable population density
* They would like a reasonable level of other non-restaurant amenities nearby, since these can attract footfall into the location of their restaurant.

They would like the deliverable back as a series of maps and other appropriate presentation mechanisms that demonstrate the best locations for them to set up their restaurants. They would like us to segment (cluster) the neighbourhoods based on the best prospects for them, highlighting both good and bad neighbourhoods.

# The Data and our Approach to a Solution

There are some key data that we clearly need in order to begin to work on this problem:

1. **A list of Los Angeles districts, together with their location data**

I have identified a list of Los Angeles districts and zip codes in the PDF at the following location

<http://file.lacounty.gov/SDSInter/lac/1031552_MasterZipCodes.pdf>

This will serve as a base for beginning work on Los Angeles. Only those districts listed as belonging to “City of LA” will be considered.

Using FourSquare, we can determine Latitude/Longitude data via the zip code. This will be useful when determining nearby restaurants and amenities, as well as when relating the population density back.

1. **The population counts of each Los Angeles district, from census data**

I have identified data from the 2010 Census that can assist us with this. It will be found here:

<http://www.laalmanac.com/population/po24la.php>

We can use this data to determine population densities in each district, and work out how well the population is served by the existing restaurants nearby, as well as whether our client should set up in that district.

1. **Location data of Korean Restaurants throughout Los Angeles**

This is clearly essential, and is data that will be provided by FourSquare, via their API. It is essential that we can gain a measure of neighbourhoods based on how many Korean Restaurants are already present, perhaps in the form of a ratio vs. population size. We will need the Latitude and Longitude of each such restaurant.

1. **Location data on non-restaurant amenities throughout Los Angeles**

This will again be provided by FourSquare. We will need to determine the nearby non-restaurant amenities and find a suitable way to measure this in terms of closeness to the given restaurant. Again, the latitude and longitude will be required, and we must remove non-restaurant locations from this part of the analysis.

## Suggested Outputs

This is subject to change as I explore the data, but I intend to do the following:

* Obtain all raw data around LA districts, populations, and location data of all amenities throughout the city of Los Angles. This data must be appropriately wrangled.
* Familiarise myself with this data via simple visualisations, and checking its accuracy
* Determine simple measures of neighbourhood fitness such as population density, density of Korean Restaurants, and density of non-restaurant amenities
* Output some simple visualisations for our client based on the measures listed above
* Output a score of “Neighbourhood Suitability” based on the above measures
* Use these “Neighbourhood Suitability” scores to segment the neighbourhoods into ones that are of similar potential, and provide these back to the client