1. Introduction and business problem

1.1 Background:

Arica is a small city located at the North border of Chile, it has a estimated population of 202.131 hab. My brother is a No1 Coffee fan and in the last year he has been investigating about it in the region and found out that there are a very small amount of café in the city and almost none of them sells real quality Coffee. I decided to help him in his goal and apply all the knowledge that I had received in the IBM Data Science Course to find the perfect location to place a Café.

1.2 Problem to solve

The problem to solve is to find areas that fulfill the next conditions with the idea of find the best place to locate a Cafe:

- Be highly populated, more people mean more possible clients
- Low amount of cafe nearby, to fulfill an empty place and avoid competition

1.3 Target

The target is to help people who wants to start their own Cafe to find the best place in Arica, Chile to locate their shop, They will really care about this because is well known that the income that this kind of shop receive are highly related to the place in which they are located. So, this work will at least find the best place based on statistics and may increase the possible revenue that any owner could have.

1.4 Interest

This work will be of high interest for those people who wants to start its own Cafe in the region, and it may be of interest for people who want to make the same research in their city.

2. Data

In chile the biggest source of demographic information is the National Census that is held by the “Instituto Nacional de Estadisticas” (INE), In this case the 2002 census will be used because is the one that’s has the highest accuracy.

2.1 Data Requirements

The relevant data needed is the following:

- City separated by sectors, unfortunately unlike the US in Chile the INE does not separate the Cities by Neighborhoods or Postal code. The most similar to that are the Districts (Clusters of highly populated areas in each city).
- Geographical References to place each district.
- Population of each District.
- Amount of Cafe near each District.

2.2 Data Sources.

The INE has all the information that is needed but it isn’t highly detailed so there is a lack of information in some cases, as an habitant of the city I will use the INE data and fill the empty space with what I know about the city, for Ex. The districts don’t have the Latitude and longitude but there is a map with the borders of each one and their names are based on relevant places in each district so it will be easy to fulfill those empty spaces, I will locate the relevant places on Google Maps, Wikipedia and fuse the data with each district. The population of each district is found in this same data source.

The Cafe data will be obtained from the Foursquare API.

Example of INE District information:

Distrito censal	Población (2002)	Superficie (km²)
Puerto	2 744	1,2
Regimiento	3 880	0,7
Chinchorro	12 816	13,3
San José	13 216	1,2
Población Chile	9 086	17,3

Example of District map: Green Lines represents each district area.