

1. A description of the problem and a discussion of the background. **(15 marks)**

## **Opening a food business in the Toluca Valley**

Toluca Valley is a zone located 50km from the Mexico City (Mexico's capital and largest city in Latin America), it is an area that has grown a lot in the last 15 years in terms of population and economical activity.

The economical zone is comprised of several municipalities surrounding the capital of the state, Toluca, which is one of the largest cities in the country, and the valley has had an exponential growth in population in the last 20 years, due to the better salaries, services, life quality and its proximity with the country capital.

One of the main activities that has grown the most is the food industry, both food production & processing as well as small food businesses.

There has been a perceived increase in the creation of new food business, small ones, informal, restaurant franchises and so on.

Unfortunately the small business do not last long in the market, this for several reasons: people does not like the food, location, food quality, etc. The business that last longer are the typical mexican food (we mexicans, love mexican food). Foreign food does not last long, exception would be Japanese (Sushi) and American food (hamburguers, sandwiches, hot dogs).

Typical local food seems to prevail in municipalities with low income population, due to its lower price and requires less infrastructure and resources to be cooked/served. In some other municipalities there is an increase in food franchises with different types of food, these are the municipalities with higher salaries and income, but even in those still prevails typical food.

The problem for an entrepreneur is the election of the correct location of his/her business depending of the type of food that's going to sell. As said before, many food business are not seen for more than three months in a location and are replaced by a different business.

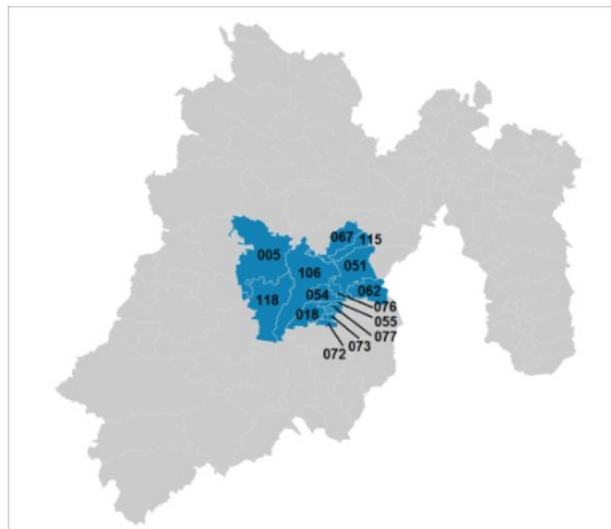
In Mexico there is not a culture of making factibility tests before opening a business, so it would be a nice opportunity to provide a scientific analysis for a potential entrepreneur and provide help to choose a nice location for a new food business.

2. A description of the data and how it will be used to solve the problem. (15 marks)

## Data

A list of the municipalities that comprise the Toluca Valley zone, this can be found in Wikipedia with the population number for each municipality.

[https://es.m.wikipedia.org/wiki/Zona\\_metropolitana\\_del\\_Valle\\_de\\_Toluca](https://es.m.wikipedia.org/wiki/Zona_metropolitana_del_Valle_de_Toluca)



Código INEGI	Municipio	Población (2010) <sup>[2]</sup>
106	Toluca	489,333
054	Metepec	206,005
118	Zinacantepec	136,167
005	Almoloya de Juárez	147,653
051	Lerma de Villada	134,799
067	Villa Cuauhtémoc	78,146
076	San Mateo Atenco	72,579
062	Ocoyoacac	61,805
018	Calimaya	47,033
115	Xonacatlán	46,331
090	Tenango del Valle	21,765
073	San Antonio la Isla	22,152
072	Santa María Rayón	12,748
055	Mexicaltzingo	11,712
027	Chapultepec	9,676
	<b>Total</b>	1,846,116

Next, obtain the coordinates for every Municipality, these can be found in another web page:

<https://www.coordenadas.com.es/mexico/pueblos-de-mexico/15/1>

<b>Coordenadas</b> <i>Consultar Coordenadas</i>		Coordenadas del Mundo	España	México	Argentina	Buscador
--	--	-----------------------	--------	--------	-----------	----------

* -Coordenadas geográficas provincia de Mexico	
- ¿Cuales son las coordenadas geográficas, longitud y latitud de los pueblos de Mexico? - A continuación se relacionan todos los pueblos y lugares geográficos de la provincia de Mexico, ríos, montañas, playas, sitios turísticos, lugares para visitar con información de las coordenadas geográficas para planificar sus viajes y localizar puntos exactos en el mapa:	
Coordenadas Mexico.....Más Pueblos por orden alfabético, página : 1 2 Siguiente >	
Ciudad	Coordenadas
- Acambay	19.95389,-99.84306
- Acolman	19.64146,-98.88145
- Aculco	20.09820,-99.82810
- Almoloya-de-Alquisiras	18.55000,-100.30833
- Almoloya-de-Juarez	19.37314,-99.83481
- Almoloya-del-Rio	19.15861,-99.48861
- Amanalco	19.86389,-99.71222
- Amatepec	18.67563,-100.24290
- Amecameca	19.11666,-98.75546
- Apaxco	19.56326,-99.31564
- Atenco	19.44735,-98.94447
- Atizapan	19.17556,-99.48861
- Atizapan-de-Zaragoza	19.59387,-99.25025

The data can be scraped for both data sources, the objective is to obtain a data frame like with the following structure:

Municipality	Population	Latitude	Longitude
Toluca	489333	19.288762	-99.655247
Metepec	206005	19.2564	-99.6048
Zinacantepec	136167	19.286518	-99.727837

The data sources contains coordinates for all main populations in the State of Mexico, but we are only interested in populations that are located in the Toluca Valley.

With the previous data frame we can make a request to Foursquare to obtain the venues in a radius of 2 or 3 km around the center of the municipalities.

We will obtain a data frame with the following structure:

Field	Description
Municipality	Municipality name
Municipality Latitude	Municipality Latitude Center
Municipality Longitude	Municipality Longitude Center
Venue	Name of the Venue
Venue Latitude	Venue Latitude
Venue Longitude	Venue Longitude
Venue Category	Foursquare Category

From the Foursquare data set we will filter all those venues that do not seem to be Food Venues, like:

- Parks
- Gyms
- Plazas
- Forests
- Churches
- etc.

A segmentation algorithm (KNN) will be used to cluster the venues by Municipality and from then the clusters will be examined to decide in which municipality is best to open a certain kind of food business. Maybe further segmentation will be necessary but this is an initial exercise.