

# Coursera Capstone Project: The Battle of Neighborhoods

## IBM Data Science Professional Certificate

Luis Avila<sup>1</sup>  
April, 2019

### I. INTRODUCTION

#### A. Background

Growing cities suppose challenges that many disciplines have to overcome. Construction space, type of land use, mobility, accessibility, streets and highways with enough space for transit, supermarkets and stores to supply basic necessities.

Every time people moves from other localities to their nearest city for many reasons, mainly being for a higher lifestyle, a higher paying job, access to basic services, supermarkets and amenities. When people consider moving to a city they need to know what is the best place to move, according to their profile.

#### B. Problem

Workers from medium and big companies are often required to relocate in another city and sometimes with no much time in advance in order to make a proper research to know what is the best place to live in the new city. The lack of time to make a research about what is the best place to live can lead to have an unpleasant stay while living in this new city.

An analysis is performed for the case of the company worker Luis, who is moving from Monterrey, Mexico to Austin, Texas due to his company requirements. We need to identify what are the current characteristics of the place where Luis lives, and to compare if there is any similar place in Austin, Texas.

#### C. Objective

Analyse and identify the neighborhood with the highest percentage of similarity compared with the origin city.

Neighborhood segmentation according to venue information acquired using Foursquare API. Data obtaining and cleaning using web scrapping techniques. Obtaining of latitude and longitude of each one of the cities using the geocoder package. Feature engineering and clustering using k-means algorithm.

The final deliverable is a recommendation based on the percentage of affinity for both areas in Monterrey and Austin.

### II. DATA AND METHODS

#### A. Data Sources

The description of the data to be used in this project is as follows:

- **Foursquare API:** Provides location information of nearby venues given an address. The Foursquare API will be used to extract nearby places for every neighborhood in Monterrey and in Austin. The neighborhood segmentation will take place using this information to extract particular characteristics of each neighborhood.
- **List of Monterrey Postal Codes (1):** Monterrey postal codes are used to extract the main database that are used for this project.
- **List of Austin Postal Codes (2):** Austin postal codes will be used to extract the main database that will be used for this project.

#### B. Methodology

Web scrapping is used to transform web lists that contains the postal codes for Monterrey and Austin. The creation of a database in form of a pandas dataframe provide the structure required to apply the algorithms proposed. Data cleaning and feature preprocessing is required to avoid duplicates in data, null values and correct data formatting and type.

Figure 1 is an example of the postal codes for Monterrey, Mexico. The list consist of Postal Code, Borough and Neighborhood. In Figure 2 shows a sample of the zip codes available for Austin, Texas. Similar as the Postal codes list from Mexico, the Austin list includes zip code, type and county, also the population for that county is included.



🏠 México » Nuevo León » Monterrey						
Listado de todos los Códigos Postales de Monterrey, Nuevo León						
Buscar en esta tabla (Opcional)						
Asentamiento ▼	Tipo de Asentamiento	Código Postal	Municipio	Ciudad	Zona	Mapa
1 de Mayo (F-97)	Colonia	64220	Monterrey	Monterrey	Urbana	Mapa
10 de Junio	Colonia	64268	Monterrey	Monterrey	Urbana	Mapa
10 de Marzo	Colonia	64488	Monterrey	Monterrey	Urbana	Mapa
13 de Junio	Colonia	64780	Monterrey	Monterrey	Urbana	Mapa

Fig. 1. Example of Monterrey Postal codes list on internet

As described before, the information obtained from the two data sources scraped from the web contains the data required to build a database for both cities. A python package for obtaining the latitude and longitude of each

<sup>1</sup>This work was not supported by any organization

Austin, TX Covers 74 ZIP Codes				
ZIP Code	Type	County	Population	Area Code(s)
<a href="#">ZIP Code 73301</a>	Unique	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 73344</a>	Unique	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78681</a>	Standard	<a href="#">Williamson</a>	50,606	<a href="#">512</a>
<a href="#">ZIP Code 78701</a>	Standard	<a href="#">Travis</a>	6,841	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78702</a>	Standard	<a href="#">Travis</a>	21,334	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78703</a>	Standard	<a href="#">Travis</a>	19,690	<a href="#">512</a>
<a href="#">ZIP Code 78704</a>	Standard	<a href="#">Travis</a>	42,117	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78705</a>	Standard	<a href="#">Travis</a>	31,340	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78708</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78709</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78710</a>	Standard	<a href="#">Travis</a>	0	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78711</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78712</a>	Standard	<a href="#">Travis</a>	860	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78713</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78714</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78715</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78716</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78717</a>	Standard	<a href="#">Williamson</a>	22,538	<a href="#">512</a>
<a href="#">ZIP Code 78718</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78719</a>	Standard	<a href="#">Travis</a>	1,764	<a href="#">512</a>
<a href="#">ZIP Code 78720</a>	R.O. Box	<a href="#">Travis</a>	0	<a href="#">512</a>
<a href="#">ZIP Code 78721</a>	Standard	<a href="#">Travis</a>	11,425	<a href="#">512</a>
<a href="#">ZIP Code 78722</a>	Standard	<a href="#">Travis</a>	5,901	<a href="#">512 / 737</a>
<a href="#">ZIP Code 78723</a>	Standard	<a href="#">Travis</a>	28,330	<a href="#">512 / 737</a>

Fig. 2. Example of Austin Postal codes list on internet

neighborhood is used to complement the information for the analysis.

The Geocoder package is used to transform postal codes and boroughs into Latitude and Longitude coordinates to be used while creating the maps for the two cities. Once the creation of the latitude and longitude fields of the cities is done the exploratory data analysis is performed and the search for venues using the Foursquare API takes place.

Code example for obtaining San Francisco, California coordinates:

```
import geocoder

g = geocoder.arcgis('San Francisco , CA')
g.latlng
```

After obtaining the venues for each city's neighborhoods they are displayed in a map to visually inspect the elements to proceed to implement the K-Means Clustering algorithm.

The K-Means clustering labels the neighborhoods according to their venues characteristics. The algorithm is applied for each city and the results are presented in a map and in the form of a table, to extract useful patterns. After this step it is selected the Monterrey neighborhood that the test subject lives, extract the venue pattern created from the K-Means labeling and proceed to find the Austin neighbor with a percentage of similarity of 70% or above.

### C. Data Cleaning

(Coming soon)

### D. Feature Engineering

(Coming soon)

## III. RESULTS

(Coming Soon)

## IV. DISCUSSION

(Coming Soon)

## V. CONCLUSIONS

(Coming Soon)

## ACKNOWLEDGMENT

Thanks to the IBM Team for offering the Data Science courses available through Coursera. The courses syllabus, videos and practice labs are high quality resources that help introduce us to the Data Science world.

## REFERENCES

- [1] Monterrey List of Postal Codes. (n.d.). Retrieved April 11, 2019, from <https://micodigopostal.org/nuevo-leon/monterrey/>
- [2] List of Austin, Texas Postal Codes. (n.d.). Retrieved April 11, 2019, from <https://www.zip-codes.com/city/tx-austin.asp>