CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

DAVID VÁZQUEZ

November 2020

Table of Contents

INTRODUCTION

The aim of this project is to focus on one of the main issues the city of Los Angeles is facing these days. That is crime rate. Even though historically, the amount of crimes has descended over the years (the end of the 20th century was particularly challenging), the sense of insecurity has risen in the city, mostly due to an unseen rate of homelessness in the county.

According to the Los Angeles Mayor office, the causes for this social phenomenon are numerous, including the steep increase in housing prices, the lack of resources for veterans, the cuts to mental health care or the stagnant salaries.

This situation affects every citizen in the city of Los Angeles (specially the people living in the most iconic neighborhoods) and every person thinking about moving to the largest city in California. That is where this project will focus, in newcomers to the city of Los Angeles, and in their decision on where they should locate their residency.
However, and going back to the housing prices mentioned above, people moving to Los Angeles need to be financially stable in order to thrive in the city. That is why this capstone project will provide the best area to open an spanish restaurant.

This project will deploy a tool (in this case, a Jupyter Notebook) that provides the optimum location in terms of neighborhood for a family business to deploy, taking into account the number and trending venues in the neighborhood as well as the crime rate in the area.

The outcome that this project provides will have a high versatility since the venue is easily interchangeable and can deliver solutions for different kinds of businesses. The model can become very helpful for agencies trying to attract entrepreneurs into the county or the government for parallel reasons.

DATA

Given that this model will account for crime rate and for venue density in the area, the data collection will come from two well-differentiated strings:

- Crime Data
- Venues per Neighborhood Data

Crime Data

For this information, we need to check the Los Angeles Mayor website. Inside the site, there is a large number of repositories full of data and available to download.
We are going to select a dataset containing all crimes committed in the city of Los Angeles during this year 2020.

This dataset contains several information of no use to us. For example, since we are going to look for the total number of crimes committed, columns explaining the type of crime, the type of weapon or the criminal's physical description will be dropped.

	DR_NO	AREA NAME	LOCATION	LAT	LON
0	10304468	Southwest	1100 W 39TH PL	34.0141	-118.2978
1	190101086	Central	700 S HILL ST	34.0459	-118.2545
2	191921269	Mission	14400 TITUS ST	34.2198	-118.4468
3	200100501	Central	700 S BROADWAY	34.0452	-118.2534
4	200100504	Central	200 E 6TH ST	34.0448	-118.2474

Furthermore, the dataset locates the crimes in 21 different areas, corresponding to the jurisdiction of the 21 Los Angeles Police Community Stations. Given that we are going to provide the location of our venue per neighborhood, it is not coherent that we use the areas stipulated in this dataset. Instead, we will extract the crime's latitude and longitude and assign it to the nearest neighborhood given its own coordinates.

Venues per Neighborhood Data

This second string of data comes from two different sources. First, will be a dataset online containing all neighborhoods in the city of Los Angeles along with their location (coordinate wise).

	Name	sqmi	Longitude	Latitude
0	Acton	39.339109	-118.169810	34.497355
1	Adams-Normandie	0.805350	-118.300208	34.031461
2	Agoura Hills	8.146760	-118.759885	34.146736
3	Agua Dulce	31.462632	-118.317104	34.504927
4	Alhambra	7.623814	-118.136512	34.085539

Second will be a request running through the Foursquare API. The request will have the neighborhood's location and "restaurants" as inputs, and it will provide the general information about the venues as outputs.

This information will be refined in order to maintain the type of restaurant and its longitude and latitude. Then an analysis will be performed in order to obtain the most popular venue types per neighborhoods in order to open our Spanish restaurant in an area where that type of restaurants is not predominant.

METHODOLOGY

RESULTS AND DISCUSSION

CONCLUSIONS