Ángel Guardián Expansion
Finding the Best Location for a New Base

Daniel Zachrisson

July 8, 2020

1. Introduction

1.1 Background

Ángel Guardián is an organization in Guatemala City that provides a service that consists in connecting people who are under the influence of alcohol or drugs with a volunteer driver (off duty volunteer firefighter). The volunteer driver takes the person under the influence on their own vehicle to their home. The purpose of this service is to reduce the amount of people out in the streets driving under the influence of alcohol or drugs.

1.2 Problem

The purpose of the project is to analyze location data of Guatemala City to find the most strategic location for Ángel Guardián to expand throughout the city so that waiting times for clients are reduced.

1.3 Interest

The organization Ángel Guardián will have interest in this project if they are looking to expand, or any other organization that wants to start a similar service may use the insights found on this analysis for their benefit.

2. Data Acquisition and Data Processing

2.1 Data Sources

The postal codes for each 'zona' was obtained from https://worldpostalcode.com/guatemala/ciudad-de-guatemala. The dataset was combined with the longitude and latitude for approximately the center of each 'zona' obtained from https://www.gps-coordinates.net. The location information was obtained by using the Foursquare API by using the 'zonas' number to retrieve the venue information for each zona. The radius was set to be 800 meters, with a limit of 100 venues per call.

2.2 Cleaning the Data

The data from both sources were combined into a single data frame using the Pandas package. The Foursquare API was used to retrieve information such as the names of the venues close to each center of each 'zona' and to obtain the latitude and longitude of such venues. The information collected from the Foursquare API was saved into a separate data frame. The venues were grouped by their respective 'zona' and later counted to have an idea of how many venues are in each 'zona'.

A new data frame was created using the 'get_dummies()' function on the 'Venues Category' column to get the categorical values of 'category' into a numerical value that could be used for the analysis later. The dummy values of 0s and 1s was grouped again by 'zona'. The mean was calculated by each 'zona' to find the most popular venue.

The data was placed into a new data frame where the first column has the 'Zona' and the following columns are the ranking of popularity of each kind of venue, starting with the most popular and to the least popular.

	Zona	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Zona 1	Café	Bar	Restaurant	Bakery	Coffee Shop	Pharmacy	Latin American Restaurant	Performing Arts Venue	Park	College Auditorium
1	Zona 10	Café	Hotel	Italian Restaurant	Steakhouse	Breakfast Spot	Bar	Seafood Restaurant	American Restaurant	Restaurant	Shopping Mall
2	Zona 11	Bus Station	Arts & Crafts Store	Sandwich Place	Athletics & Sports	Taco Place	Soccer Stadium	Dive Bar	Donut Shop	Diner	Department Store
3	Zona 12	Bar	Café	Bike Rental / Bike Share	Seafood Restaurant	Restaurant	Athletics & Sports	Yoga Studio	Factory	Flea Market	Fast Food Restaurant
4	Zona 13	Airport Lounge	Airport	Bar	Café	Fast Food Restaurant	Chinese Restaurant	Rental Car Location	Breakfast Spot	Fried Chicken Joint	Bed & Breakfast

Figure 1. Clean Data