Capstone Project: Taco Shops in San Antonio

Benoit DARTIGUEMALLE, 2020

1. Introduction

1.1. Background

Moving to San Antonio in a while and I realized I did not know anything about the city. I found a study on San Antonio about food and ratings of Taco Shops. The study reports a database of Taco Shops to evaluate the burritos of the city and that contains the following information: Name, location, address, neighborhood, rating by Google, by Yelp, etc…

1.2. Business Problem

My problem in this study will be to decide around which neighborhood I will need to live to be closer to a maximum of taco shops and with good ratings.

Every person who lives or travel through the city will be interested in that study.

2. Data Processing

2.1. Data sources

I have two sources of data:

1. Taco Shop study in San Antonio (csv format)
2. Neighborhood and Perimeter Plans of the city of San Antonio (geojson format)

The study is stored on my IBM cloud account alongside my python program. The geojson is available on opensource from the following website: data.sanantonio.gov

2.2. Data Cleaning

The study of taco is a table of 385x66 (385 rows and 66 columns). I obviously do not need all that information. My first step will be to "clean" the data in order to keep only the relevant information.

For example, each line correspond to a particular burrito from a taco. It means that a taco shop, which sell multiple types of burrito, will appear more in the dataset. To evaluate the number of taco shops in the city, we have to get rid of this.

In the dataset, we also have information about the ingredient present or not in the burrito. We do not need this information for the study.

Finally, we came up with a new database of 83x5.

Data Analysis and Visualization

Target Analysis

During this project, I will analyze a study done on taco shops in San Antonio on different burritos. I will conduct a statistic analysis to find out the neighborhood with the most taco shops during a first step. Then I will give the list of the best taco shops rated by Google from the study, and the cheapest ones. I will also conduct an analysis to find out if there is a correlation between price and grades.

In a second part, we will analyze via a map in which neighborhood we will find out the most a taco shops in the Town. Another geographical analysis of the best burritos will be provided as well.

So that, the customer will have the choice to find out where are the biggest choices of shop in town or in which area he can find the best burritos.

The customer can also base his choice by combining and analyzing both maps.

Data Visualization

TOP10 Neighborhood

In this section, we will give the list of the neighborhood that get the most taco shops location in their area. The result give be find in *Figure 1*.
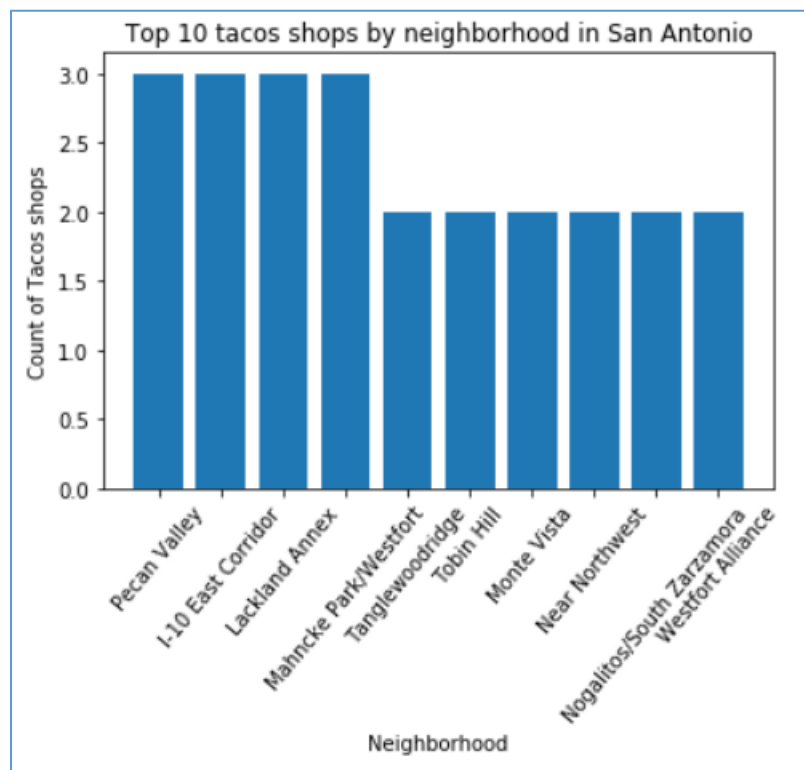

Figure 1 : TOP10 neighborhood with the most taco shops location

Best ratings and best value

In this part, we evaluate the neighborhood with the customer feedback given via Google. We calculate the mean value for the grade and the cost by grouping the data. Results shown in *Figure 2* and *Figure 3*.

	Neighborhood	Mean Google Grade
39	United Southwest	4.900000
5	Eastern Triangle	4.700000
36	Stinson Airport Vicinity	4.650000
33	River Road	4.600000
38	Tobin Hill	4.550000
16	Lackland Annex	4.433333
30	Pecan Valley	4.433333
15	Lackland AFB	4.400000
19	Mahncke Park	4.400000
0	Arena District	4.350000

Figure 2 : Best ratings per neighborhood

	Neighborhood	Mean Google Grade	Mean Cost
25	Nogalitos/South Zarzamora	4.000000	5.220000
21	Meadow Village	4.300000	5.490000
37	Tanglewoodridge	4.300000	5.890000
18	Lone Star	4.100000	5.935000
12	Huebner/Leon Creeks	3.400000	6.090000
23	Monte Vista	4.000000	6.095000
2	Camp Bullis	3.950000	6.100000
13	I-10 East Corridor	4.133333	6.163333
19	Mahncke Park	4.400000	6.250000
10	Guadalupe Westside	4.350000	6.270000

Figure 3 : Best value per neighborhood

Best Choice and Rating

In this part, we combined the results for the most choices per neighborhood and the rating per neighborhood to find out if there is certain tendency between those parameters. It can be useful to

eliminate some neighborhood from the list or to select a top list on them. The result can be found in the Figure 4 below.

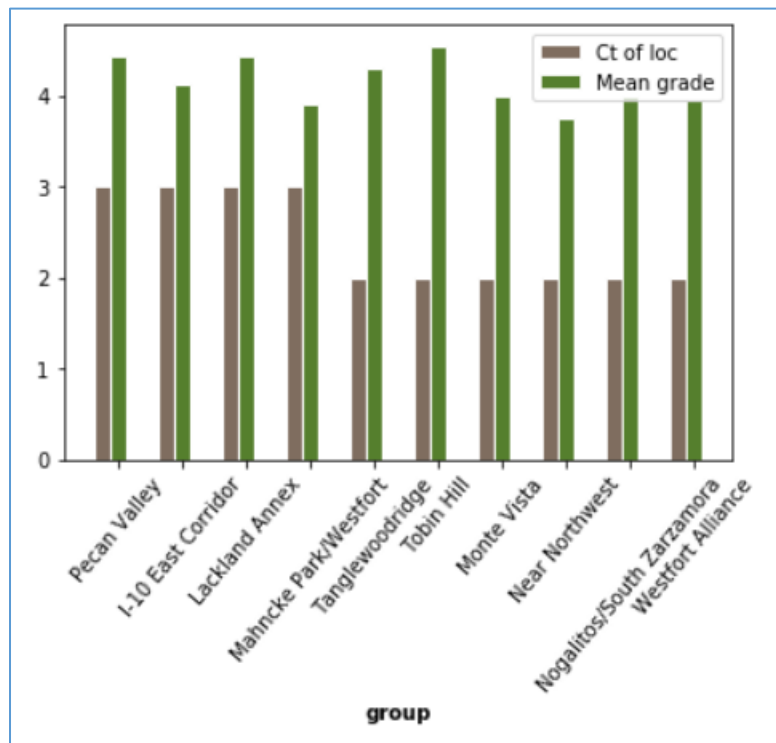(Note : The x axis has been filled in with the TOP10 neighborhood values)

Figure 4 : Balance Choice/Rating

Extra : Best Burrito in Town

Without taking care of the location and the neighborhood, the following table will give the list the best burrito of SA according to the ratings in Google.

	Location	Neighborhood	Google
341	Los Tacos 2	Stinson Airport Vicinity	5.0
50	Mikes Taco Club	United Southwest	4.9
25	Lola's 7 Up Market & Deli	River Road	4.9
362	Burrito Factory	Arena District	4.8
239	La Morena Taco Shop and Seafood	Tobin Hill	4.7
180	El Pueblo Mexican Food	Nogalitos/South Zarzamora	4.7
322	Lourdes	Eastern Triangle	4.7
52	La Perla Cocina	Lackland Annex	4.7
113	Taco Surf PB	Woodlawn Hills/ Ingram Hills	4.6
300	Mister Falafel	Pecan Valley	4.6

Figure 5 : Best Burrito in Town

Correlation between Rating and Price

Now we want to know if there is any correlation between the price and the rating. To do that, I simply use the *regplot* function to visualize it.
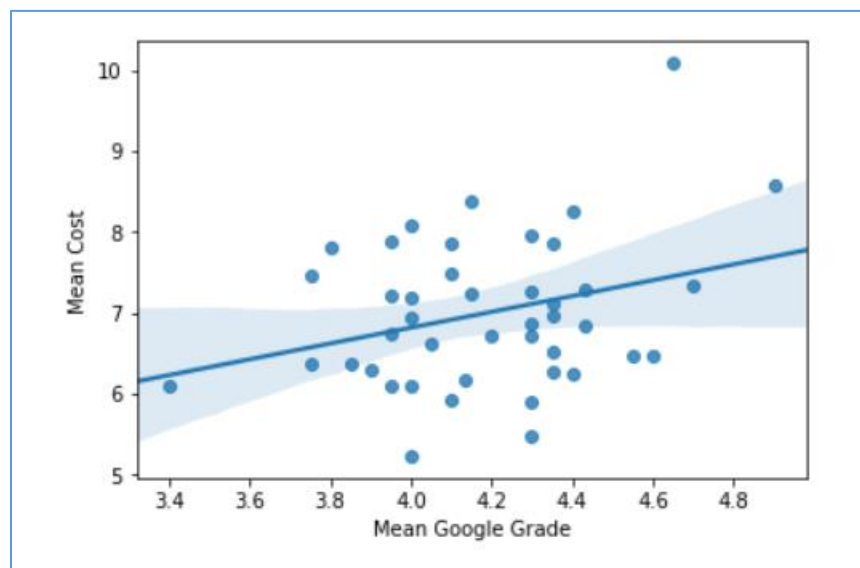

Figure 6 : Correlation Price/Rating

We can say that there is no evident correlation between those parameters.
A simple linear equation cannot give a good approximation or reflect the tendency of this study.

Mapping – Visualization of results

Count per neighborhood

The following map has been build given the neighborhood in the original dataset, mapped with the geojson file from the city of San Antonio.

The result give a map with neighborhood clearly delimited and color shades depending on the count of taco shops in that particular neighborhood.
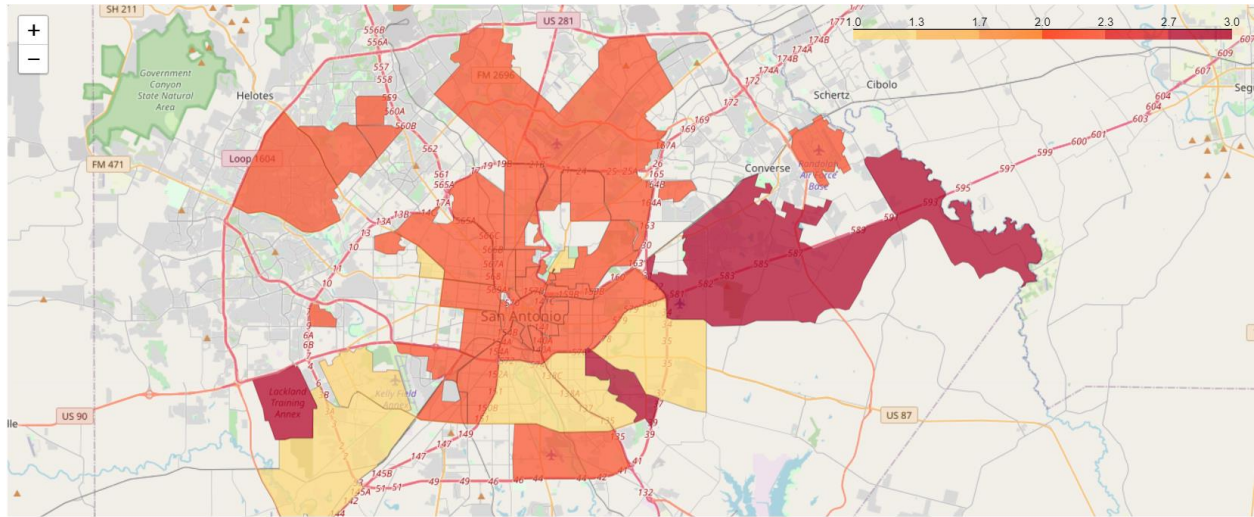

Figure 7 : Count of shops per neighborhood

Rating per neighborhood

This map gives the mean rating per neighborhood. The legend and the shades of red can be used to find out the neighborhood of your choice.
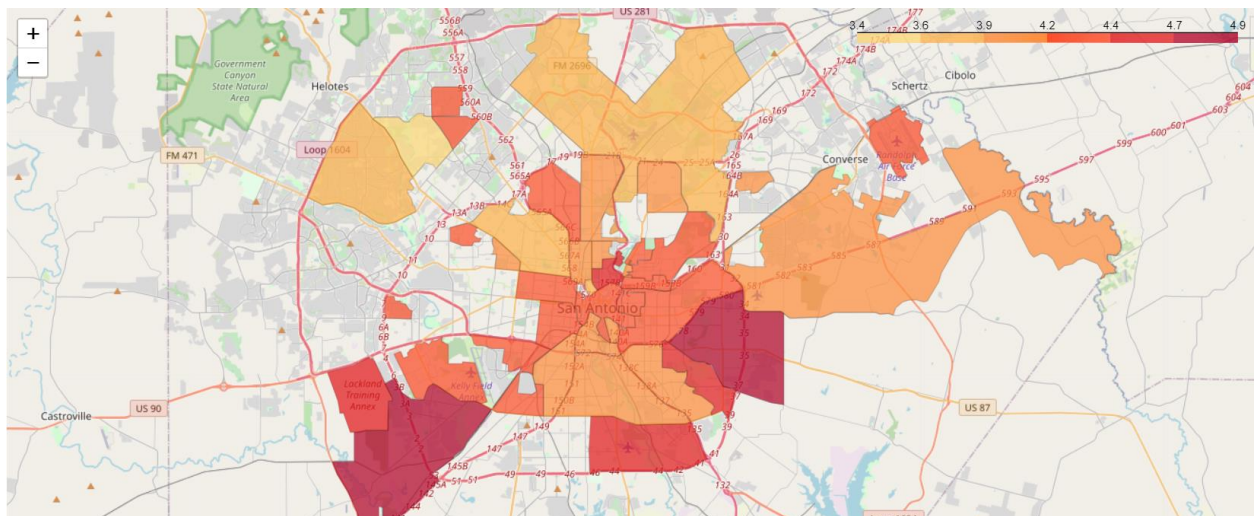

Figure 8 : Ratings per Neighborhood

Results

Analysing both map to find out the best recommendation, I would recommend the **_Lakeland Training Annex neighborhood_**.

Discussion

This study is based on a dataset from 2016. Knowing that certain shops might have closed or that the quality of some could have increased/decreased with the time, I would recommend combining this study with more researches on the web.

A refresh of the database can be useful to give a refresh to the results.

For recommendation and based on this study, I would recommend avoiding the north of the city (lack of choice and poor ratings).

Conclusion

This conclude the Capstone Project and the analysis of Taco Shop in San Antonio.