Applied Data Science Capstone: The Battle of Neighborhoods *

Baeza Olvera Daniel Jesus
dbaeza@ibm.com

May 28, 2020

Abstract

As part of the IBM & Coursera Applied Data Science Specialization, it is required to do a small research using the taught data science methodologies and data sources. The data science methodologies and data sources that had more emphasis were, Geographical Data with the Python package Folium, and the usage of third party APIs for retrieval of location data.

The problem that this little research tackles is about Food Venues, especially Mexican Food venues in the city of Naha, Okinawa Japan which is the place I wish to live in the not so distant future. One of my plans while living there is Co-owning a Mexican Food restaurant.

Keywords: *Data Science; Foursquare; GIS; Google Maps API; Capstone, Python.*

1 Introduction

In this project we will try to find an optimal location for a restaurant.

Specifically, this report will be targeted to stakeholders (mainly Myself) interested in opening a Mexican Taqueria in Naha City, Okinawa, Japan. The focus of this is looking for locations that are not already crowded with restaurants. We are also particularly interested in areas with no Mexican restaurants in vicinity. We would also prefer locations as close to the IT COLLEGE as possible, this because IT people love tacos and because I'd love to work as a teacher there and have tacos after class, but only assuming that first two conditions are met. We will use a data science approach to select the most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that taking a decision using the top neighborhoods will be simplyfied.

All of this is part of the final project for the Applied Data Science Specialization offered by IBM Through Coursera.

Applied Data Science Specialization, IBM, and Coursera (2020)

*Presentation for the Capstone project, part of Applied Data Science Specialization, Coursera.

2 Data

Based on definition of our problem, factors that will influence our decission are:

- Number of existing restaurants in the neighborhood (any type of restaurant)
- Number of and distance to Mexican restaurants in the neighborhood, if any
- Distance of neighborhood from IT College

Following the project example, I decided to also use regularly spaced grid of locations, centered around city center, to define our neighborhoods. In fact I will mostly use the same approach and step by step analysis, due to time constraint and not having any other particulary interesting problem to solve using the foursquare API which is required for this project.

(Applied Data Science Capstone Example Notebook, IBM, & Coursera, 2020)

Following data sources will be needed to extract/generate the required information:

- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding.
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API.
- Coordinates of the IT college will be obtained using Google Maps API geocoding.

(Foursquare Development Sources , 2020)
(Google Maps Development Sources , 2020)

3 Results (WIP)

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

4 Discussion and Conclusions (WIP)

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

References

IBM, & Coursera Applied Data Science Specialization
IBM, & Coursera Capstone Example Project Notebook
Foursquare Developer Portal
Google Maps Developer Portal

Table 1: Example table of descriptive statistics of the main variables.

Variables	Categories	Unit	Rep	Mean	St. Dev.	Min	Max
Variable 1	Category A	$	8	0	0	0	0
	Category B	lb	8	22,411.20	6,325.90	13,819	31,201
	Category C	$	8	5,869.60	4,609.90	-464.1	12,744.10
Variable 2	Category A	$	8	1,777.40	144.5	1,642.30	1,912.60
	Category B	lb	8	21,444.80	5,146.90	15,096	28,032
	Category C	$	8	4,138.50	2,644.10	22.2	7,932.70
Variable 3	Category A	$	8	2,346.80	190.8	2,168.30	2,525.20
	Category B	lb	8	18,343.30	2,460.70	15,269.00	21,524.10
	Category C	$	8	3,699.20	2,549.80	1,299.10	8,709.80
Variable 4	Category A	$	8	2,288.80	186.1	2,114.80	2,462.90
	Category B	lb	8	23,450.40	4,172.50	20,045.00	32,363.00
	Category C	$	8	6,619.80	1,918.40	4,479.70	10,633.90
	CASE #1			14	6.61	6.9	27.9
	CASE #2			22.8	7.73	10.2	31.4

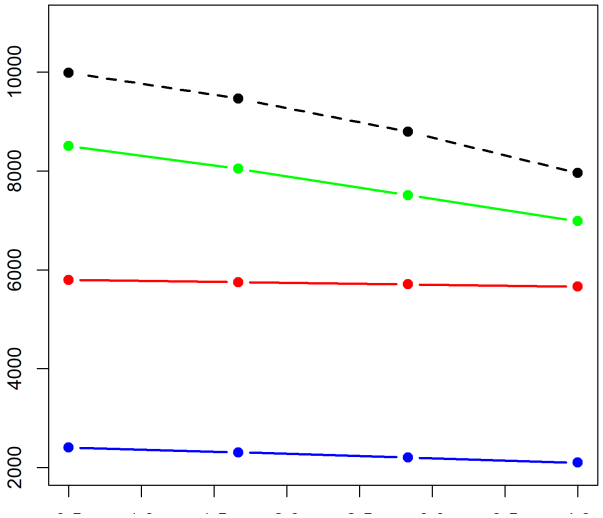

Figure 1: Example figure.