Capstone Project - "Zero Waste Company"

Applied Data Science Capstone by IBM/Coursera

Table of contents

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

1. Introduction: Business problem

A Uruguaian brand named "Zero Waste Company" (ZWC) is a small start up that grow as an idea developed between three friends. The core of their business and culture is that they offer a unique experience to their customers, by providing healthy products and without impact in the environment. This will be achivied offering re-usable packages and incentiving customers to do that by providing recyclable trash stations. The start up is planning to install operations in Montevideo, Uruguay, where these three friends are from. To better evaluate the best neighborhood to install the Zer Waste Company Shop, a analysis will be performed.

First, to be sure that you are familiar with Uruguay and its capital, the following maps will help you. Uruguay is in the South of South America, between Argenitna and Brazil.

Now that you are familiar with the Uruguaian country, we will focus in the objetive of this article. Below are some pre-requisites of the ZWC members to install their operations in Montevideo. To be considerer:

- The ZWC will sell medium to high price products. Therefore, it must be located in a neighborhood of medium to high economic social level and high density area.
- In the other hand, the shop shall be located near similar shops as healthy restaurants, coffe shops, take and go, etc. This is to be closer and exposed to more customers (or potential customers).
- Also, part of the positive environment impact is to be a station of re-usable and recyclable packagest, compostable garabage. This must be considerer as a part of the location selection.

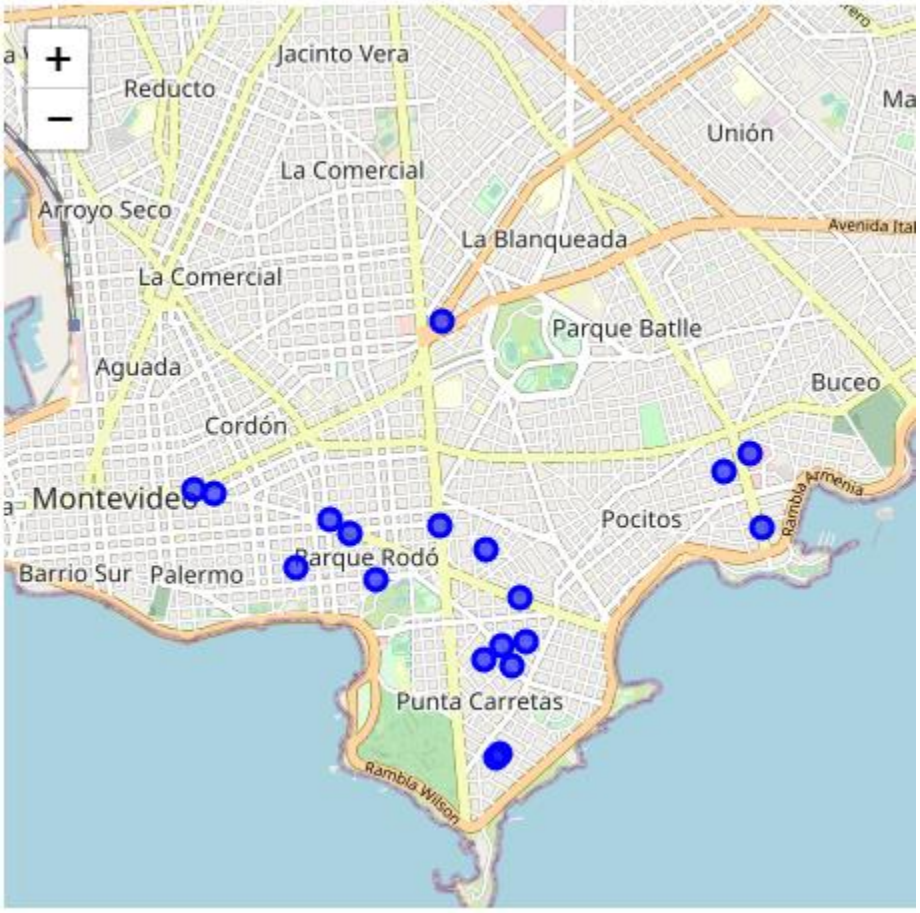
2. Data

As stablished in section 1, there are some pre-requisites that we need to consider to determine the best set of locations to provide to ZWC. In this section, the reader will understand how to obtain valuable data conserning all the pre-requisites and how to integrate and analyze it. The Foursquare location data will be used as the principal source of data for this analysis, anyway, other data source or tools could be used.

Ok, now that you are introduced, lets run some code to find the healthy shops in Montevideo. Run the following code to install and import some libraries and then search for all shops that has something to do with the word "healthy" (this is "saludable" in spanish).

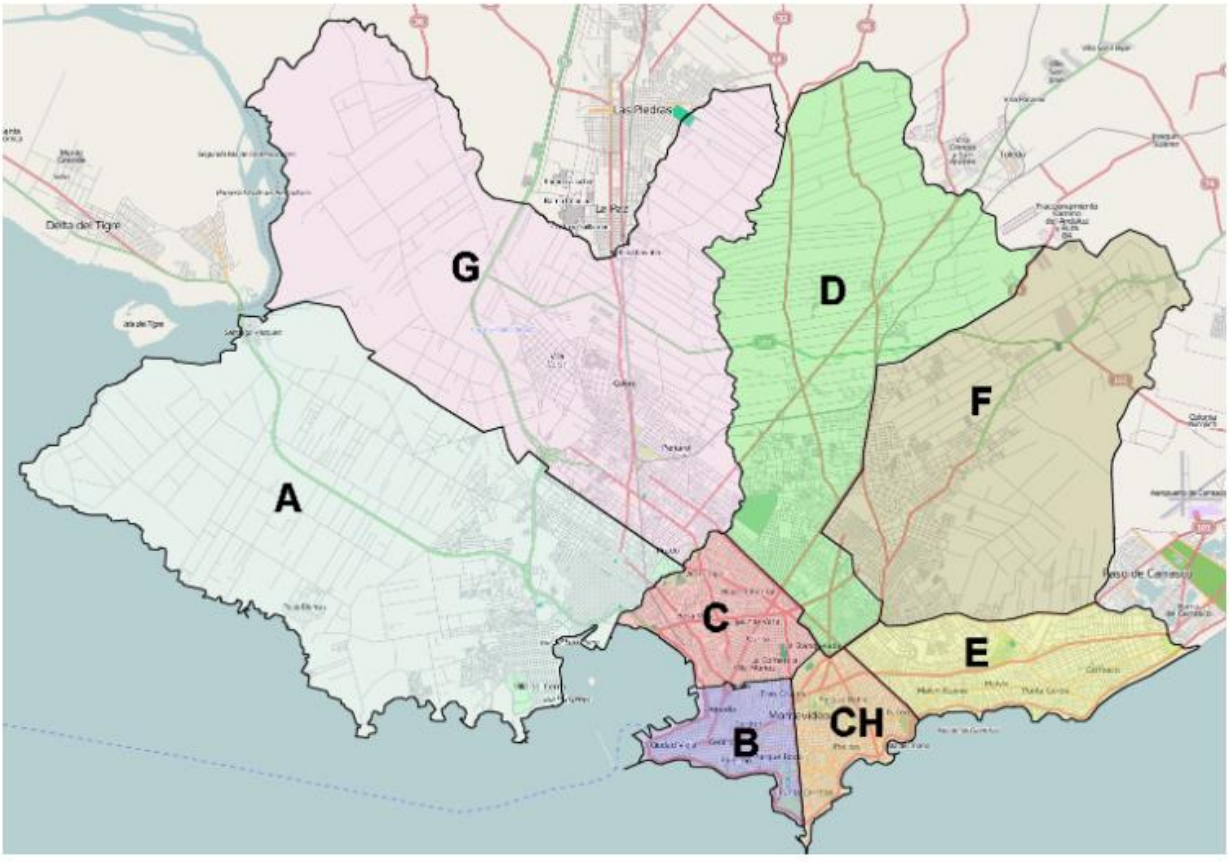
Get data from Foursquare

Name	Lat	Long	Location street
Club Natural y Popular	-34.912986	-56.156655	Luis de la Torre
Benicio Deli & Coffee House	-34.908079	-56.135901	Echaverriarza
Camelia	-34.917726	-56.157381	NaN
Mercado Verde	-34.916334	-56.158203	21 De Setiembre
Dorothea	-34.904057	-56.139085	Plácido Ellauri



For demography data, the Uruguaian goverment has some websites to find that information. For easy upload of that information, the following data was obtained from Ministery of Uruguay. The following data frame was defined.

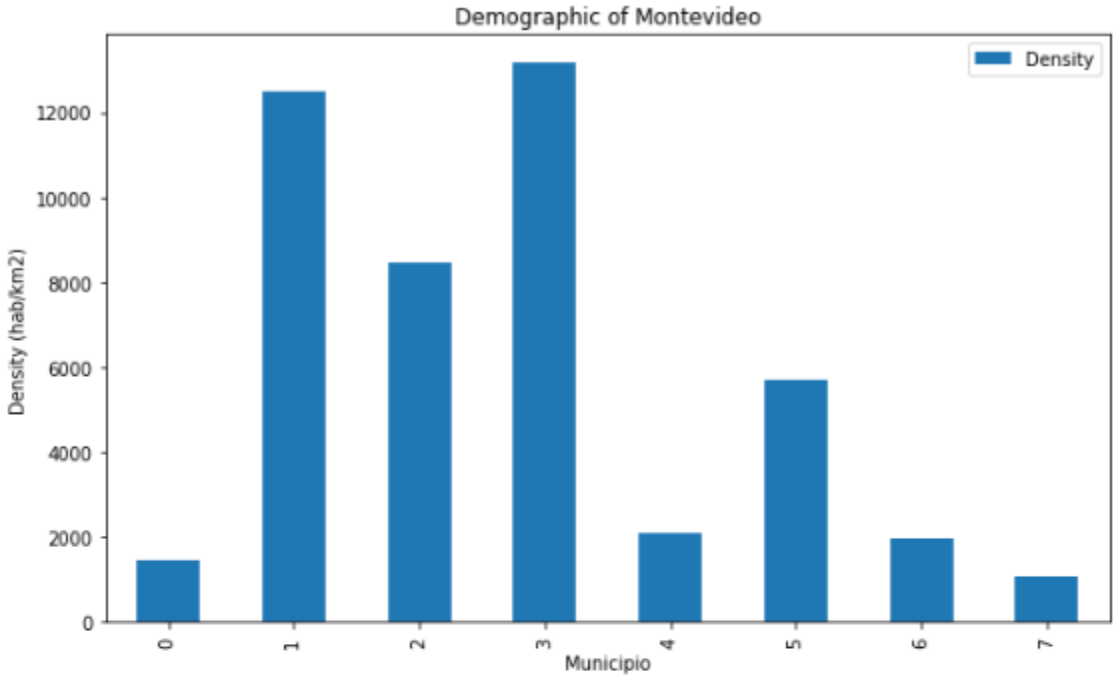
And the map of Montevideo with the Municipios

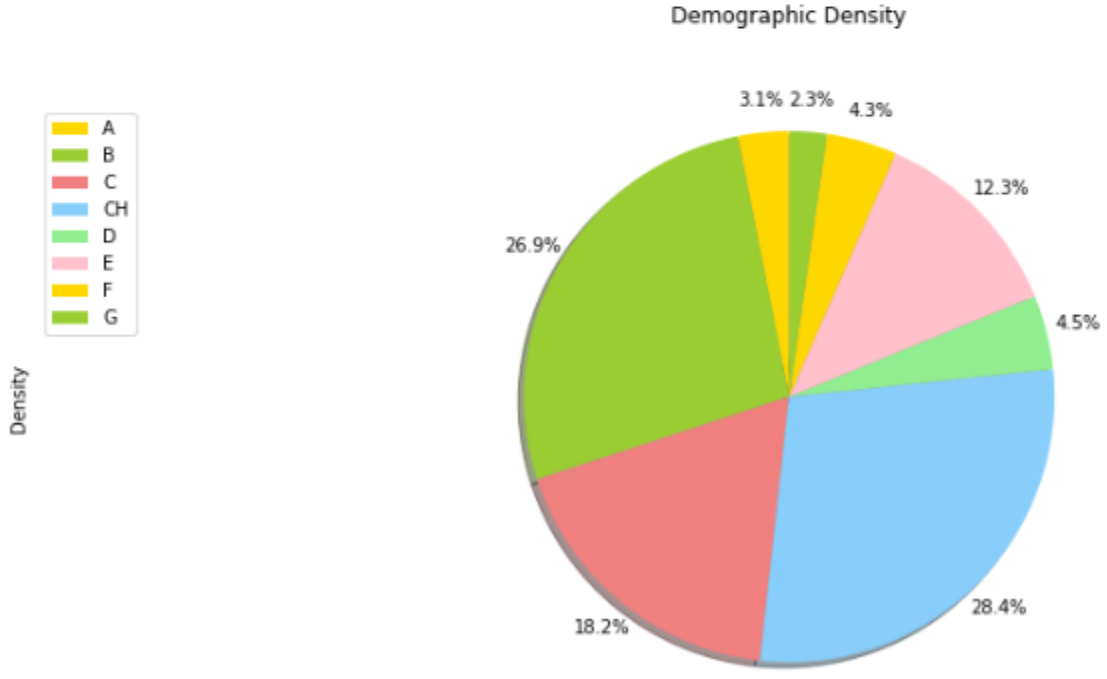


3. Methodology

Through this chapter, the reader will understand how the data defined in chapter 2 will be analized and other data generated if it is required. Results of this chapter will be presented and discussed in chapter 4.

First, lets plot the density (people per sq km - hab/km2) by Municipio to explore some estimation for the best place to locate the shop





Now we have to cross this data with the current healthy shops to better understand the best location for the ZWC shop. For this purpose, we will add some extra data to combine both data frames with Foreing and Primary Keys. This will be a column containing the Municipio of each shop.

Name	Lat	Long	Location street	Municipio
Club Natural y Popular	-34.912986	-56.156655	Luis de la Torre	B
Benicio Deli & Coffee House	-34.908079	-56.135901	Echaverriarza	CH
Camelia	-34.917726	-56.157381	NaN	C
Mercado Verde	-34.916334	-56.158203	21 De Setiembre	CH
Dorothea	-34.904057	-56.139085	Plácido Ellauri	CH
Don Lemon	-34.907862	-56.163510	Maldonado	B
Adolfo Café	-34.917376	-56.159640	NaN	C
Buena Costumbre	-34.910857	-56.175873	San salvador e isla de flores	B
La Claraboya	-34.902782	-56.136934	NaN	CH
Fans café	-34.905650	-56.182871	Vázquez	CH
ThePutaMadre Bar	-34.907458	-56.172913	Jackson	B
La Cocina Del Parque	-34.911628	-56.168998	NaN	B

So, we could say that the best location (y) is a function of most frequented customer street (s) and the Municipio (m) with higher density.

location = f(street, municipio)

y = f(s, m)

As an Analytical Approach, find the most congruent Municipio and frequented streets (in therms of zone):

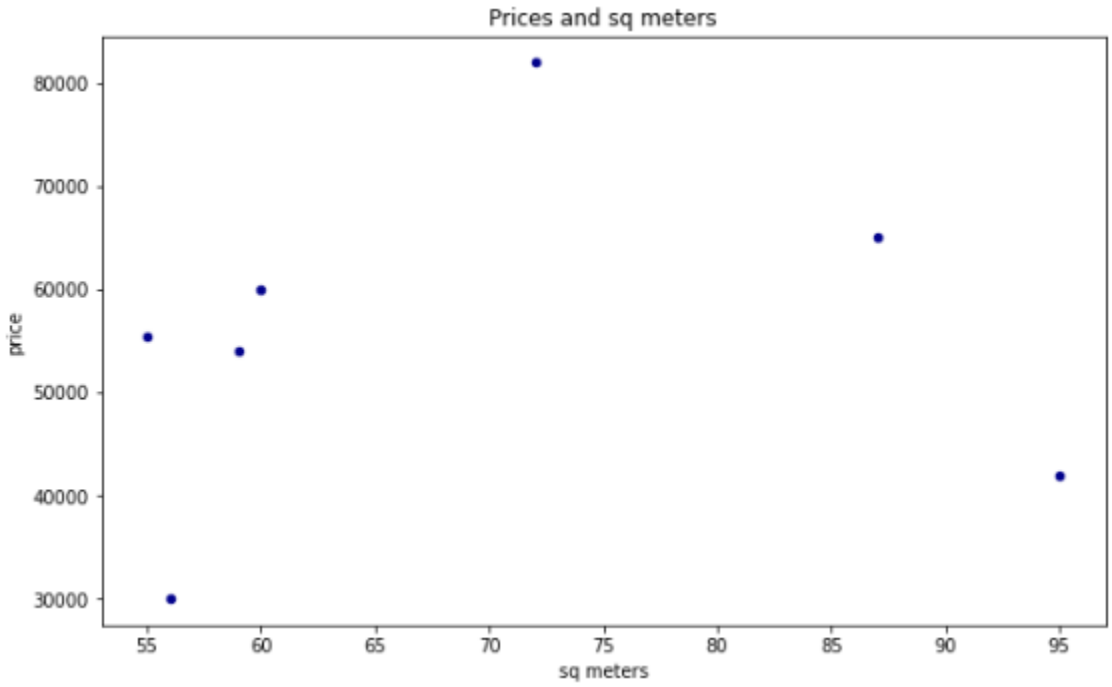
```
The maximum density is 13177 - Municipio CH
Municipios at CH:                          Name        Lat       Long  Location street  \
1   Benicio Deli & Coffee House -34.908079 -56.135901    Echaverriarza
3                 Mercado Verde -34.916334 -56.158203  21 De Setiembre
4                      Dorothea -34.904057 -56.139085  Plácido Ellauri
8                   La Claraboya -34.902782 -56.136934              NaN
9                      Fans café -34.905650 -56.182871          Vázquez
12                        Subway -34.893518 -56.163377  Avelino Miranda
15                        Sésamo -34.909638 -56.159490              NaN
```

It looks that we found the best Municipio to locate the shop, but, are there any available any shop for rent in this Area and with ZWC members requirments? As we don´t have credentials to access to API for databases of rental locations, let´s make the example with the following data frame:

Municipio	Lat	Long	price	sqmeters
B	-34.912986	-56.156655	49990	40
CH	-34.908079	-56.135901	60000	60
C	-34.917726	-56.157381	78000	80
CH	-34.916334	-56.158203	55500	55
CH	-34.904057	-56.139085	82000	72
B	-34.907862	-56.163510	50000	160
C	-34.917376	-56.159640	45000	133
B	-34.910857	-56.175873	20000	120
CH	-34.902782	-56.136934	30000	56
CH	-34.905650	-56.182871	42000	95
B	-34.907458	-56.172913	78000	65
B	-34.911628	-56.168998	100000	87
CH	-34.893518	-56.163377	54000	59
E	-34.916074	-56.156043	63000	63
E	-34.923914	-56.158326	78500	78
CH	-34.909638	-56.159490	65000	87
B	-34.924196	-56.158624	55000	82
B	-34.908433	-56.171200	30000	54
B	-34.905391	-56.184597	68000	65

Municipio	Lat	Long	price	sqmeters
CH	-34.908079	-56.135901	60000	60
CH	-34.916334	-56.158203	55500	55
CH	-34.904057	-56.139085	82000	72
CH	-34.902782	-56.136934	30000	56
CH	-34.905650	-56.182871	42000	95
CH	-34.893518	-56.163377	54000	59
CH	-34.909638	-56.159490	65000	87

As we discussed before, the best combination of parameters are those that maximize sqmeters, minimize price and it should be in the Municipio with higher density. So, we now that the Municipio is CH and the location is a function that depends on price and sq meters:



As seen before, the coorelation is postivibe but is not a good correlation. Lets try with polinomial regression:

```
        2
-97.4 x + 1.461e+04 x - 4.674e+05
```

But, how this polynomial model fits to our data?

```
The R-square value is:  0.731328020823206
```

Since it is not a perfect fit, it is a good value of Rsquare, so we can assume that is a good aproximation for the relation between both variables.

4. Results & Discussion

As shown in the previous section, the analysis of the data show that the best palce to locate the shop is in Municipio CH. According to the locations founded in Municipio CH, the best option (considering price and sq meters and ZWC members pre-requisites) are the ones that minimize the funcion y=p(x). In other words, minimize the price but maximizing the area. The exploratory and analysis data suggest that the best option is:

first, the following code shows around what y_hat (predicted price) should be considerer as a part of the ZWC project

```
y_hat (predicted price): 41506.408249764296
```

If we search for some location near this price, we will find that is:

```
Municipio         CH
Lat          -34.9057
Long         -56.1829
price           42000
sqmeters           95
Name: 9, dtype: object
```

5. Conclusion

Well, we have defined the business problem and some data was defined to explore some solutions for the presented problem. During the route, it was found the relation between dempgraphy density and shops related with heathy products. Also, a relation was found for price and sq meters for available rental shops. Now, the ZWC have the best option to locate its shop and also a methodology to follow in case they need/want to move or expand operations along the city, country or the same method could be applied for other locations. This study is also applicable for other applications related with the location of a shop, even the topic that it is about.