Applied Data Science Capstone Project: Battle of the Neighborhoods

By Magdalena Sapag

February 2, 2019
Santiago, Chile

Table of Contents

Table Index

Figure Index

1. Introduction

Data Science is an important tool for decision making in almost every business. It allows managers and analysts to take advantage of the data available to make intelligent decisions. The data used for these decisions can come either from within the company and its records or can be obtained from an outside provider. This project will focus specifically on location data made available by Foursquare and the analysis that can be made based on it to help a business grow its profits.

Objective

The objective of this project is to identify the best corners in the city of Santiago, Chile to place on-field sellers for a flower shop named Crocus. According to the company, the average daily sales of a flower shop show seasonality within a week, increasing by 200% on the weekends. Additionally, profits increase to 9 times the average daily profit for special dates, like Valentine's Day or Mother's Day. The company is interested in capturing more of the market in those dates, and to achieve that they plan on placing on-field sellers in 5 corners of certain neighborhoods of Santiago for weekends and special dates.

To decide which corners are better than others, Crocus has provided a list of 30 potential corners of the city that have shown high foot traffic. The company also requested that the choice is made based on nearby complementary products, such as chocolate shops, restaurants and movie theaters, and competitor location. The final criteria for the selection of corners will be explained in detail in the methodology section.

Audience

The specific client for this project is a flower shop named Crocus that has online operations on the city of Santiago and has been in operation for 5 years. The company is a small business; hence the scope of this specific project is small. However, the basic logic and methodology can be scaled to satisfy similar needs for larger companies.

There is a large amount of businesses in the city of Santiago that rely on on-field sales force or promoters that are located on specific streets or corners. The results of this project have the potential to help these types of businesses make decisions that would grow their profits and help them capture a higher percent of the market. Any business interested in deciding the placement of on-field sales force can replicate this model and add to it if necessary, in order to make the decision based on Foursquare location data, or any other location data provider.

2. Data

For the purpose of this project there are two types of data that will be used: The table containing the potential corners and their coordinates and Foursquare location data.

Potential corners coordinates

The company requesting this project has provided a list of 30 potential corners in the city of Santiago, specifically the upper east side of the city, that have high foot traffic. The data set has been delivered as a csv file with 31 rows and 5 columns. The columns of the data set are: an id number for each corner, the address, the neighborhood, the latitude and the longitude of its location. The complete data set can be seen in the following table.

Table 1: Potential Corners in Santiago

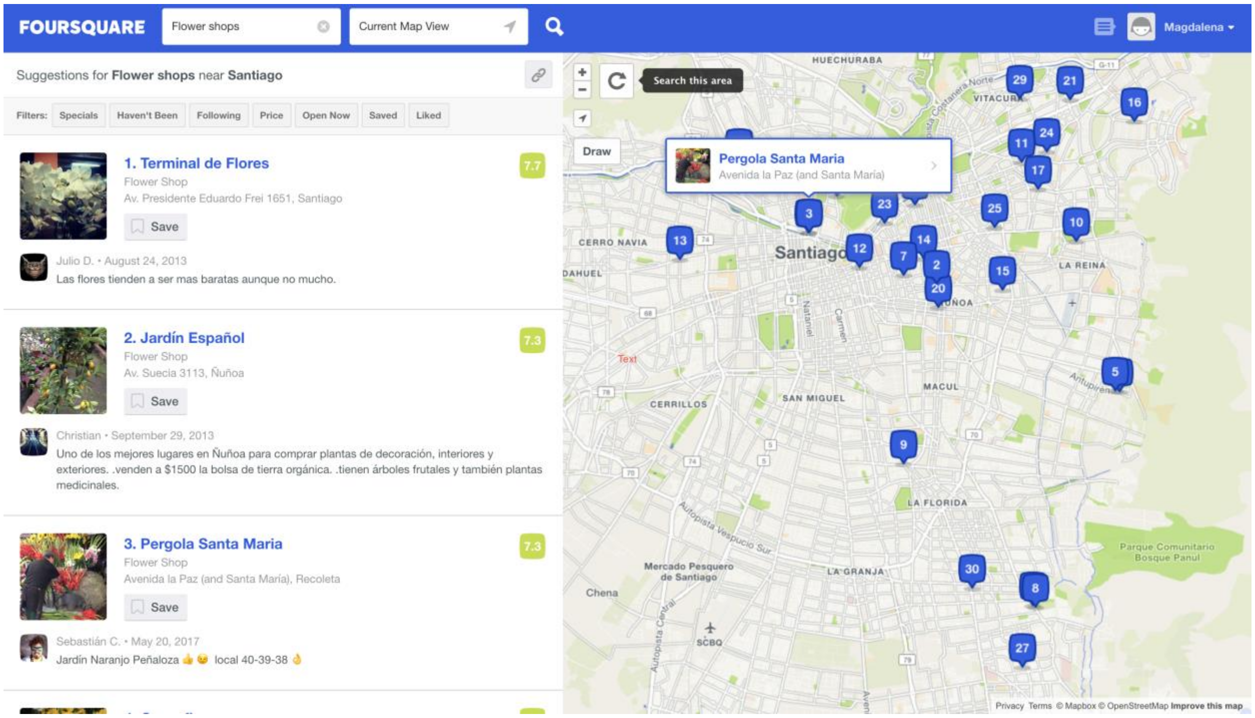
Corner	Address	Neighborhood	Latitude	Longitude
1	Plazuela Los Leones	Providencia	-33.419869	-70.605912
2	Avenida Nueva Providencia 2200	Providencia	-33.422751	-70.60956

3	Avenida Pedro de Valdivia 101	Providencia	-33.424649	-70.611985
4	Avenida Suecia 780	Providencia	-33.427532	-70.605505
5	Avenida El Bosque 963	Providencia	-33.428427	-70.596128
6	Avenida Nueva Providencia 1398	Providencia	-33.428874	-70.618359
7	Metro Francisco Bilbao	Providencia	-33.430808	-70.586816
8	Latadia 4141	Las Condes	-33.431166	-70.578576
9	Avenida Ossa 1552	Ñuñoa	-33.439474	-70.572611
10	El Alcalde 15	Las Condes	-33.416337	-70.595227
11	Avenida Apoquindo 3898	Las Condes	-33.415298	-70.590077
12	Avenida Apoquindo 4483	Las Condes	-33.413435	-70.583082
13	Centro Comercial Apumanque	Las Condes	-33.409423	-70.567761
14	Parroquial Los Dominicos	Las Condes	-33.407847	-70.541797
15	Avenida Presidente Kennedy 5413	Las Condes	-33.402279	-70.578125
16	Avenida Cuarto Centenario 1001	Las Condes	-33.417755	-70.55362
17	Avenida Padre Hurtado Sur 875	Las Condes	-33.416215	-70.53963
18	Avenida Francisco Bilbao 8464	Las Condes	-33.429683	-70.54508
19	Avenida Larrain 5862	La Reina	-33.453282	-70.570271
20	Salvador Izquierdo 1777	La Reina	-33.438958	-70.556324
21	Plaza Pedro de Valdivia 1731	Providencia	-33.438743	-70.608079
22	Santa Isabel 400	Providencia	-33.446353	-70.626189
23	Avenida Salvador 42	Providencia	-33.433757	-70.626389
24	Avenida Ricardo Lyon 1146	Providencia	-33.430869	-70.606272
25	Avenida Suecia 181	Providencia	-33.418879	-70.609346
26	Avenida Cristobal Colon 4431	Las Condes	-33.423352	-70.578923
27	Parque Bicentenario	Vitacura	-33.399924	-70.602398
28	Avenida Vitacura 3744	Vitacura	-33.402006	-70.594212
29	Mall Alto Las Condes	Las Condes	-33.390066	-70.547686
30	Manuel de Salas 39	Ñuñoa	-33.454969	-70.593477

Foursquare location data

Foursquare is a provider of accurate location data, currently used by over 100,000 developers, including services like Apple Maps, Uber, Twitter, among others. Foursquare allows access to information about venues and users in different locations. Each venue has different attributes, such as name, category, address, working hours, menu, tips and images. There is also the possibility of getting information about the users, such as name, profile picture, tips and images she has uploaded.

For the purpose of this project only information about venues will be taken into consideration and for each venue, only the category and exact location will matter. An example of how Foursquare categories work can be seen bellow, where a search for Flower Shops in Santiago is carried out and the results given by Foursquare are shown.



FOURSQUARE
Flower shops
Current Map View
Magdalena
Filters: Specials Haven't Been Following Price Open Now Saved Liked
Save
Draw
Search this area
HUECHURABA
VITACURA
CERRO NAVIA
Santiago
LA REINA
PUDAHUEL
CERRILLOS
SAN MIGUEL
MACUL
LA FLORIDA
Mercado Pesquero de Santiago
LA GRANJA
Chena
SCBQ
Parque Comunitario Bosque Panul
Privacy Terms © Mapbox © OpenStreetMap Improve this map

Suggestions for **Flower shops** near **Santiago**

1. Terminal de Flores — 7.7
Flower Shop
Av. Presidente Eduardo Frei 1651, Santiago
Julio D. • August 24, 2013
Las flores tienden a ser mas baratas aunque no mucho.

2. Jardín Español — 7.3
Flower Shop
Av. Suecia 3113, Ñuñoa
Christian • September 29, 2013
Uno de los mejores lugares en Ñuñoa para comprar plantas de decoración, interiores y exteriores. .venden a $1500 la bolsa de tierra orgánica. .tienen árboles frutales y también plantas medicinales.

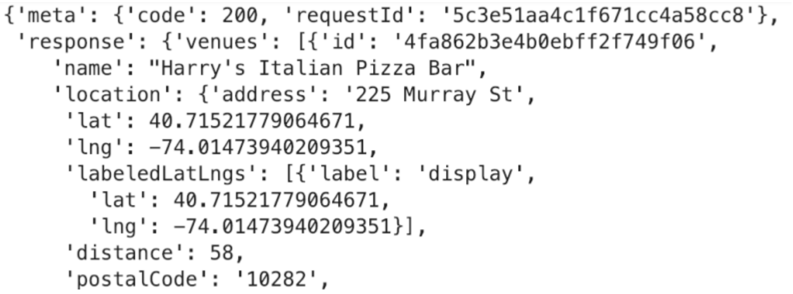
3. Pergola Santa Maria — 7.3
Flower Shop
Avenida la Paz (and Santa María), Recoleta
Sebastián C. • May 20, 2017
Jardín Naranjo Peñaloza 👍 😉 local 40-39-38 👌

Pergola Santa Maria
Avenida la Paz (and Santa María)

Figure 1 Foursquare search for flower shops in Santiago

Having an account in Foursquare will provide credentials which allow Foursquare information about venues to be accessed through a Python script. This will be the second main source of information for the project, which will have to be cleaned and processed first, in order to use it for the analysis required.

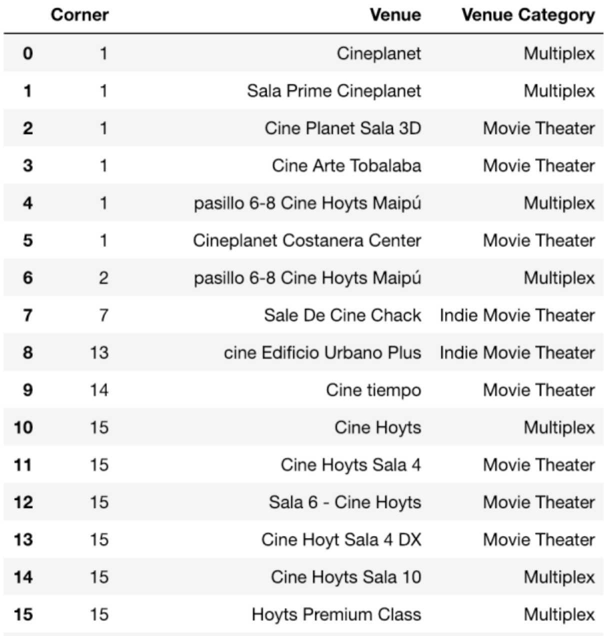
Specifically, the information that will be used for this project is obtained through a search request in Foursquare. For the request a query will be given to find each type of venues, such as 'Flower Shop' or 'Movie Theater'. Another argument that will be used in order to narrow the search is categoryID, which allows the user to give a specific category in which the result must be, for example the category ID for flower shops is '4bf58dd8d48988d11b951735', so that will be given as an argument when the request for flower shops is made. The result of this request is a json file with all the venues that were found such as the one shown in the following figure, this example comes from a different project in which the search was made in New York for Italian food.



```
{'meta': {'code': 200, 'requestId': '5c3e51aa4c1f671cc4a58cc8'},
 'response': {'venues': [{'id': '4fa862b3e4b0ebff2f749f06',
    'name': "Harry's Italian Pizza Bar",
    'location': {'address': '225 Murray St',
     'lat': 40.71521779064671,
     'lng': -74.01473940209351,
     'labeledLatLngs': [{'label': 'display',
       'lat': 40.71521779064671,
       'lng': -74.01473940209351}],
     'distance': 58,
     'postalCode': '10282',
```

Figure 2 Example of json file obtained from a search request in Foursquare

This json file will be transformed into a data frame using the keys for name and category of the resulting venues. No other key will be used since the final analysis only requires a count of resulting venues for each category. The data frame obtained for this project will have 3 columns: Corner, Venue and Venue Category. The resulting data frame will look like the example shown in the next figure, in this example the data frame shows the result for the search for movie theaters for different corners in Santiago.



	Corner	Venue	Venue Category
0	1	Cineplanet	Multiplex
1	1	Sala Prime Cineplanet	Multiplex
2	1	Cine Planet Sala 3D	Movie Theater
3	1	Cine Arte Tobalaba	Movie Theater
4	1	pasillo 6-8 Cine Hoyts Maipú	Multiplex
5	1	Cineplanet Costanera Center	Movie Theater
6	2	pasillo 6-8 Cine Hoyts Maipú	Multiplex
7	7	Sale De Cine Chack	Indie Movie Theater
8	13	cine Edificio Urbano Plus	Indie Movie Theater
9	14	Cine tiempo	Movie Theater
10	15	Cine Hoyts	Multiplex
11	15	Cine Hoyts Sala 4	Movie Theater
12	15	Sala 6 - Cine Hoyts	Movie Theater
13	15	Cine Hoyt Sala 4 DX	Movie Theater
14	15	Cine Hoyts Sala 10	Multiplex
15	15	Hoyts Premium Class	Multiplex

Figure 3 Example for data frame resulting from Foursquare search

For this project a search will be carried out for the following types of venues: Flower shop, chocolate shop, candy shop (since some chocolate shops are under this category), restaurants and movie theaters. Each request will result in a different data frame. The 5 data frames will then be combined to carry out the analysis. These steps will be explained in the next section of the report.

3. Methodology

In order to do the required analysis and select the 5 corners, a series of 3 main steps must be followed:

a. Extract the information
b. Process the information to count number of venues by category on each corner
c. Sort corners according to criteria

Every step taken and the python code used in the notebook will be explained in this section, as will be the criteria used for decision making at the end.

a. Extract the information

The data comes from two sources as was explained above. First, there is a csv file containing every corner, their addresses and coordinates. Second, there is the location data from Foursquare. The logical sequence is to first import the data from the csv file, since then it will be used to search for the venues in Foursquare.

CSV File

In order to get the information from the csv file, the pandas library needs to be imported into the notebook. Then, the pandas.read_csv() function can be used. When looking at the notebook on github with the code, refer to inputs 1 and 2 for the exact code. The resulting data frame contains the following columns: Corner, Address, Neighborhood, Latitude, Longitude. The last two columns will then be used to search Foursquare information for venues relevant to the analysis.

Foursquare location data

The first step to access Foursquare data is to define the credentials that will be used, this includes client id, client secret and foursquare version. It is also useful to define at the beginning the radius for the search and the limit of maximum requests made. For the purpose of this project, the radius was set to 300m, since the business is looking to attract people who stumble upon them near either a complimentary product, such as chocolate, or a final destination such as a restaurant or movie theater. The company considered that 300m is the maximum distance a person would walk to get flowers as last-minute decision.

Then a request is carried out in Foursquare, this means building the url for the request using the search request. Then the credentials for the Foursquare account are given in the url and the coordinates and a search query need to be defined. In order to do the requests efficiently a function is defined. The function takes as argument a list of coordinates and number id for each corner to search, it also receives the search query, a radius (with 300m set as default) and an additional argument for the request: categoryID, which will allow to narrow the search results for a specific category. The searches that will be carried out for this project are shown in the following table.

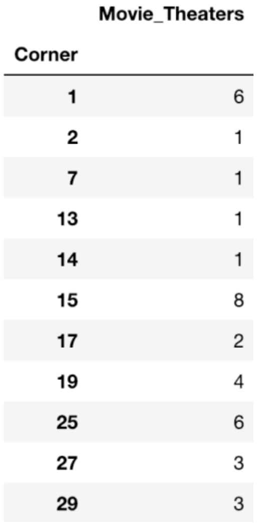
Table 2 Searches made to Foursquare for each corner

Search	Search query	Category ID restriction	Description
Flower shops	Flower Shop	4bf58dd8d48988d11b951735	Competition
Candy shops	Candy Shop	4bf58dd8d48988d117951735	Complimentary product
Chocolate shops	Chocolate	52f2ab2ebcbc57f1066b8b31	Complimentary product
Restaurants	Restaurant	4d4b7105d754a06374d81259	Final destination
Movie theaters	Cine	4bf58dd8d48988d17f941735	Final destination

The search for chocolate shops and candy shops refers to the exact same product and will be joined into one result further on for the purpose of the analysis. The function gives as a result a data frame with the venues for each corner found for the requested category and search query. An example can be seen in Figure 3 above.

b. Process the information to count number of venues by category on each corner

After the search is carried out, the information obtained needs to be polished in order to do the analysis. First the venues obtained for each search and each corner need to be counted, since the client only cares about the number of venues nearby and not any qualitative information about them. In order to achieve this the group by function will be used set to count the number of rows in each data frame by each corner. The resulting data frame will look like the figure shown bellow.



```
       Movie_Theaters
Corner
     1              6
     2              1
     7              1
    13              1
    14              1
    15              8
    17              2
    19              4
    25              6
    27              3
    29              3
```

Figure 4 Result after counting movie theaters by corner

Then the 5 separate data frames will be merged into one using the join function 5 times. The resulting data frame will have 30 rows corresponding to each corner and 10 columns: Corner, Address, Neighborhood, Latitude, Longitude, Flower_Shops, Chocolate_Shops, Candy_Shops, Restaurants and Movie_Theaters.

If a type of venue was not found for a corner, the corresponding cell will show a NaN result. In order to add the chocolate shop results and the candy shops results together the NaN need to be turned into zeros. That is achieved using the fillna() method. After that step is carried out, the columns for chocolate shop and candy shop can be added into a column named Chocolate_Stores.

c. Sort corners according to criteria

Finally, the corners need to be sorted according to the results, so since there are 4 columns to be sorted a priority needs to be defined. According to the business owners, the most important restriction is to avoid other flower shops, since any established flower shop near a corner will absorb most of the market, since the potential customers will prefer it or will think about them first when deciding to buy flowers. The second column to be sorted is Chocolate Stores, since the pull given by a complementary product is stronger than the one resulting from final destinations. Finally, between restaurants and movie theaters, the client considers restaurants have a stronger relationship with the purchase of flowers than movie theaters have, so the priority for sorting the corners will be:

1. Flower shops (Ascending)
2. Chocolate Stores (Descending)
3. Restaurants (Descending)
4. Movie Theaters (Descending)

The resulting data frame after the columns have been sorted can be seen in the following figure.

Table 3 Result after sorting corners

Corner	Latitude	Longitude	Flower_Shops	Restaurants	Movie_Theaters	Chocolate_Stores
1	-33.4199	-70.6059	0	31	6	1
3	-33.4246	-70.612	0	31	0	1
6	-33.4289	-70.6184	0	26	0	1
25	-33.4189	-70.6093	0	22	6	1
22	-33.4464	-70.6262	1	8	0	1
29	-33.3901	-70.5477	1	4	3	1
19	-33.4533	-70.5703	0	2	4	1
2	-33.4228	-70.6096	0	43	1	0
23	-33.4338	-70.6264	0	10	0	0
10	-33.4163	-70.5952	0	9	0	0
30	-33.455	-70.5935	0	9	0	0
28	-33.402	-70.5942	0	8	0	0
15	-33.4023	-70.5781	0	7	8	0
12	-33.4134	-70.5831	0	4	0	0
11	-33.4153	-70.5901	0	4	0	0
13	-33.4094	-70.5678	1	3	1	0
24	-33.4309	-70.6063	0	3	0	0
16	-33.4178	-70.5536	0	3	0	0
9	-33.4395	-70.5726	0	3	0	0
21	-33.4387	-70.6081	1	3	0	0
8	-33.4312	-70.5786	0	2	0	0

27	-33.3999	-70.6024	0	1	3	0
7	-33.4308	-70.5868	0	1	1	0
26	-33.4234	-70.5789	0	1	0	0
20	-33.439	-70.5563	0	1	0	0
18	-33.4297	-70.5451	0	1	0	0
4	-33.4275	-70.6055	0	1	0	0
17	-33.4162	-70.5396	0	0	2	0
14	-33.4078	-70.5418	0	0	1	0
5	-33.4284	-70.5961	0	0	0	0

From this resulting table the selection of the first 5 corners can be made, if the client decided to select fewer or more corners the decision can be made using the same table.

4. Results

Considering the sorted columns obtained as a final result of the processing of the information, the 5 corners selected to place on-field flower sellers are corners 1,3,6,25 and 22. In the following table the address, neighborhood and coordinates for these corners can be seen.

Table 4 Selected corners

Corner	**Address**	**Neighborhood**	**Latitude**	**Longitude**
1	Plazuela Los Leones	Providencia	-33.4199	-70.6059
3	Avenida Pedro de Valdivia 101	Providencia	-33.4246	-70.612
6	Avenida Nueva Providencia 1398	Providencia	-33.4289	-70.6184
25	Avenida Suecia 181	Providencia	-33.4189	-70.6093
22	Santa Isabel 400	Providencia	-33.4464	-70.6262

A different visualization of the results is as points on a map. The following figure shows the selected corners in blue and the rest of the corners in red in a map of the upper east side of Santiago.

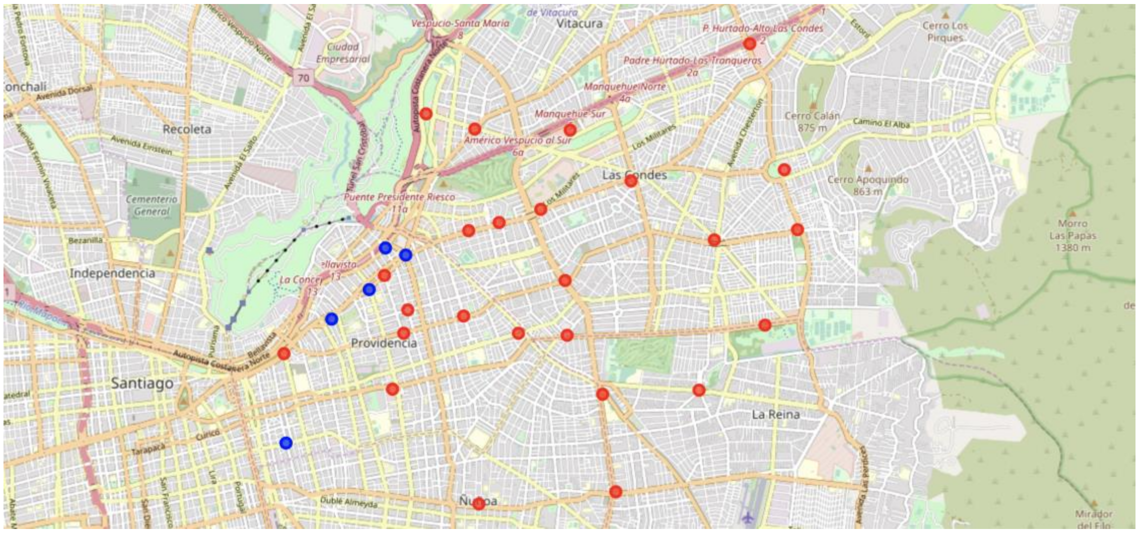


Figure 5 Map of selected corners

The map shows that the selected corners are mostly placed along the same street and for the most part nearby each other. These corners are also near the center of the city an no corners near the edges of the city were selected. All of the corners are in the neighborhood of Providencia. The implications of the results will be discussed in the following section.

5. Discussion

The discussion of the results obtained can be divided into two categories: Observations of the results and recommendations made for future related projects.

Observations

The results obtained by this project are a set of 5 corners, selected out of a pool of 30 potential highly frequented corners. After the analysis of the venues nearby the selection was made and the results show that the corners are all in the same neighborhood, and 60% of them are even in the same street. The following table shows the reasons why these corners where the highest in the list.

Table 5 Count of venues for selected corners

Corner	Address	Flower_Shops	Restaurants	Movie_Theaters	Chocolate_Stores
1	Plazuela Los Leones	0	31	6	1
3	Avenida Pedro de Valdivia 101	0	31	0	1
6	Avenida Nueva Providencia 1398	0	26	0	1
25	Avenida Suecia 181	0	22	6	1
22	Santa Isabel 400	1	8	0	1

A possible reason for that result is that the Providencia neighborhood is one of the most popular neighborhoods in Santiago, busy with commercial activity and where most of Santiago’s best restaurants are located. It makes sense that those corners had the highest amount of chocolate shops, restaurants and movie theaters nearby. The fact that just one of the corners had a flower shop nearby makes less sense. It is possible that Foursquare doesn’t have the information regarding flower shops of a more informal setting, that most of the time take the form of kiosks in the middle of the sidewalk.

Another important observation based on the results is the fact that the selections are sometimes nearby each other. It is the case between corners 1 and 25, located 550m from each other, these corners can be seen in figure 6. In this case the result is still admissible because it is considered that the maximum distance the target audience will walk to get flowers is 300m, but for a different kind of business this result might not be admissible.

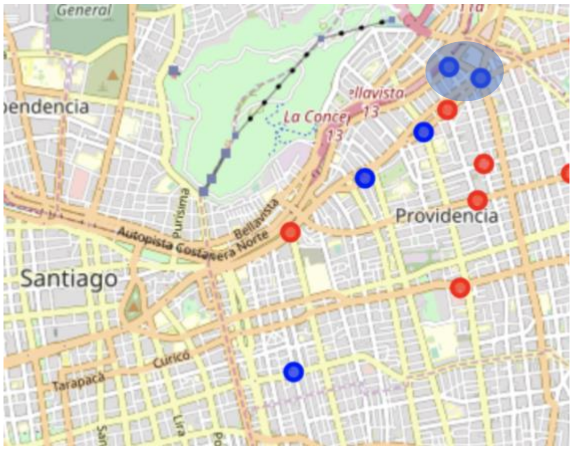


Figure 6 Close up of selected corners

Recommendations

Considering the results obtained and the observations made based on them, three main recommendations can be made:

a. Including more venues than the ones offered by Foursquare

In order to improve the results for flower shops nearby the corners, an additional data set could be included showing flower shops in the form of kiosks placed in the neighborhoods included. This information could be obtained by hiring an outside research consultant or by replacing the location data provider with one that includes more venues.

b. Restricting distance between the selected corners

The distance restriction was not an issue for this particular project but for a larger project including more potential corners it would be important to restrict that the selected corners are distanced for more than 300m from one another. This could be achieved using a more sophisticated code that selected the corners based on a score, which could be lowered if a different corner within that distance had a higher score.

c. Including more corners in different neighborhoods of Santiago.

The current project only considered five neighborhoods in the upper east side of Santiago, but if more corners were included a better result could be achieved. More corners could be considered either by including more corners within the same neighborhoods, or by covering more neighborhoods in Santiago.

6. Conclusion

The objective of this project was to select a total of 5 corners from a set of 30 potential corners in the upper east side of Santiago that would be best for the placement of on-field flower sellers for a flower shop that is looking to increase its profits in the weekends and special occasions, such as Valentine's Day and Mother's Day. To carry out the analysis the project included a data set with the coordinates of the 30 original corners and the results from search requests made in the Foursquare API for flower shops, chocolate shops, restaurants and movie theaters nearby each corner.

The analysis of the information consisted of retrieving the relevant venues for each corner, counting the amount of venues in each category and then sorting the corners, considering the four categories in the following priority: Least flower shops nearby, most chocolate shops nearby, most restaurants nearby and lastly most movie theaters nearby.

The five corners selected where corners 1,3,6,25 and 22 that can be seen in figure 5. These corners where all in the same neighborhood and were all placed relatively nearby each other. Suggestions for scaling this project include: Including venues like kiosks in the analysis, restricting distance between the selected corners and including more corners in the original pool.

In conclusion, python and its libraries are an important tool that can help businesses retrieve information and process the data in order to carry out analysis that will help them make intelligent decisions. In this case the client can increase its profit by placing on-field sellers in the five corners selected for weekends and special dates.