Location of a new business near a subway station

Mauricio R. Bustos

March 25th, 2021

Table of Contents

Executive Summary

This study defines the type of a new business and it's best location with the only restriction to be near a subway station in the city of Santiago de Chile.

For the definition of the type of business, the information of the venues around subway stations of the city of New York was used. This data was contrasted with the same for the Metro de Santiago (Metro) and differences were encountered.

A location for this business was then defined using the quantity of evaluations received by the actual venues and considering the top evaluated venue.

The business chosen to set up was a *Pizza place* and the location is at 150 m or less around the *Bellas Artes* Metro station in Santiago de Chile.

Information of venues around subway stations, in both cities, their geographic coordinates and quantity of evaluations received was obtained from the Foursquare API.

Detailed Contents

Introduction

Subway stations are places with a high transit rate of people, each one with their associated consumption needs, which reflects in the high density of commerce and services in the nears.

In places like Santiago de Chile, which is a growing city, the subway network (Metro) is still expanding, projecting new stations and commerce for products and services are establishing in the nears.

This project aims to study the opportunity of set up a new business near the actual and/or the future Metro stations, taking into account what kind of commerce is now established in a consolidated city.

Subway networks in mega cities could give valuable information in order to pick the best business to locate near stations in subways of growing cities, such as Santiago.

It is intended to analyze what kind of businesses are located near each metro station in New York City, and use this info to compare it with the info in a growing city (Santiago) to identify which businesses are not established right now and which ones is to be good to set up in the future in the next stations to be built.

It is expected that this study should be of interest for small local investors that want to develop a small business in the city of Santiago de Chile.

Literature Review

For the development of this study and also for the elaboration of this report, it was necessary to review extensively various modules of the Coursera - IBM Data Science courses:

- What is Data Science?: structure of a final Data Science report
- Data Analysis with Python: exploratory data analysis
- Data Visualization with Python: graphs and maps
- Applied Data Science Capstone: guidances for this project

Also, it was necessary to review the Pandas documentation for various functions uses.

It was required the review of numerous web sources with available datasets in order to obtain the adequate data with the info of geolocalization required.

Methodology

The analysis was made using databases based on geolocalization provided by Foursquare and datasets containing coordinates of each subway station from both cities.

Data

The data collected and analyzed could be grouped into two sets:

- Big city analysis:
  - NYC subway stations geolocalization: collected from a data source on the web. The city is New York as it is the economic center of the country. Encountered the latitude and longitude (in decimal format) for each subway station.
  - Venues around each subway station: obtained via the Foursquare API within a certain radius (150m) around each station. From these info was used the venues classification.

- Growing city analysis:
  - Actual and future Metro stations geolocalization: the same info as for the NYC subway stations, plus the info of projected Metro stations. All info from data sources from the web.
  - Venues around each Metro station (actual and future): idem as for NYC. Also used the Foursquare API to obtain venues and frequency of evaluations received by each venue.

In both cases, the results of the venues has been ranked and compared to identify the differences between them and then make an analysis to identify which businesses appears as the most attractive to locate near the Metro of Santiago stations.

Results

In order to know the frequency of venues around each subway station, this study begins with the download of the geographic coordinates of the subway stations in the cities of New York and of Santiago de Chile.

The dataset for the NYC subway was obtained from the web of the Metropolitan Transportation Authority (MTA) in their Developers section.

Data was downloaded and verified that included the name of each station and their geographic coordinates (latitude and longitude) in decimal format. An extract of this information is shown in Table 1.

	Station ID	Complex ID	GTFS Stop ID	Division	Line	Stop Name	Borough	Daytime Routes	Structure	GTFS Latitude	GTFS Longitude	North Direction Label	South Direction Label	ADA	ADA Notes
0	1	1	R01	BMT	Astoria	Astoria-Ditmars Blvd	Q	N W	Elevated	40.775036	-73.912034	NaN	Manhattan	0	NaN
1	2	2	R03	BMT	Astoria	Astoria Blvd	Q	N W	Elevated	40.770258	-73.917843	Ditmars Blvd	Manhattan	1	NaN
2	3	3	R04	BMT	Astoria	30 Av	Q	N W	Elevated	40.766779	-73.921479	Astoria - Ditmars Blvd	Manhattan	0	NaN
3	4	4	R05	BMT	Astoria	Broadway	Q	N W	Elevated	40.761820	-73.925508	Astoria - Ditmars Blvd	Manhattan	0	NaN
4	5	5	R06	BMT	Astoria	36 Av	Q	N W	Elevated	40.756804	-73.929575	Astoria - Ditmars Blvd	Manhattan	0	NaN

Table 1: NYC subway stations dataframe

This info is then visualized in a map, with each point representing a station. We can see in Fig. 1 this map and we can observe a wide distribution of this stations across the city.

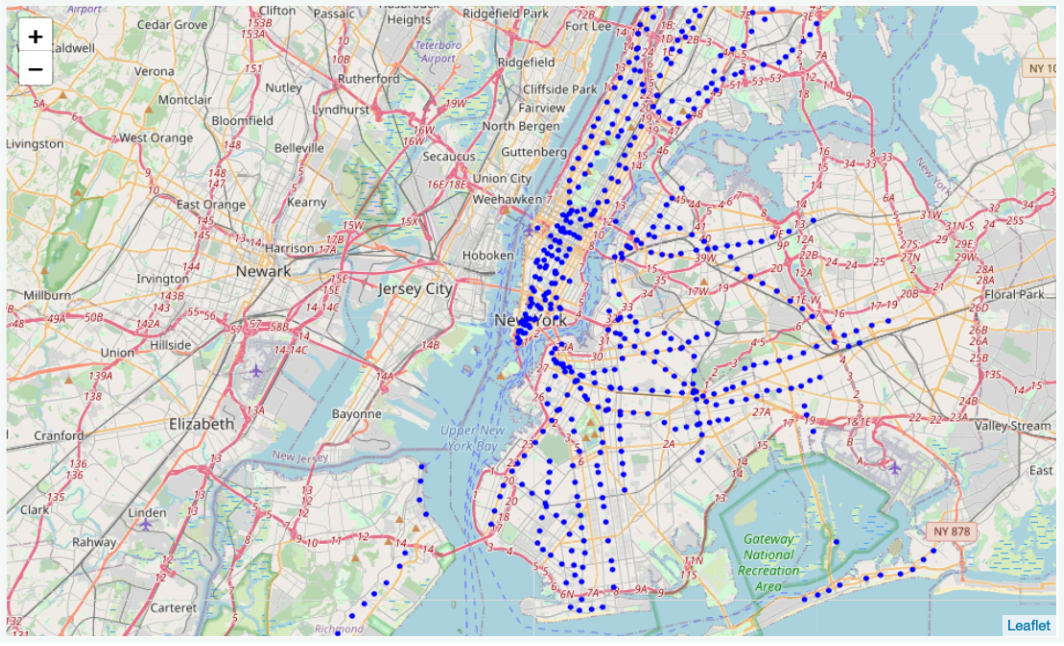

Fig. 1: Subway stations in NYC

As we have the NYC info, we downloaded the same info but for the Metro de Santiago. The source of the data set was obtained from the Observatorio de Ciudades UC, maintained by the Universidad Católica de Chile in an ArcGIS site.

The data was transformed into a data frame and an extract can be seen in Table 2.

	X	Y	FID	objectid	id	especial	estacion	nombre	temporal	f1	tipo	linea
0	-70.723289	-33.444190	1	1	1	EXISTENTE INTERMODAL	EXISTENTE	SAN PABLO	INTERMODAL	1	ESTACION METRO	Linea 1
1	-70.722538	-33.451933	2	2	2	EXISTENTE	EXISTENTE	NEPTUNO	NaN	1	ESTACION METRO	Linea 1
2	-70.713403	-33.457826	3	3	3	EXISTENTE INTERMODAL	EXISTENTE	PAJARITOS	INTERMODAL	1	ESTACION METRO	Linea 1
3	-70.706475	-33.457238	4	4	4	EXISTENTE	EXISTENTE	LAS REJAS	NaN	1	ESTACION METRO	Linea 1
4	-70.680073	-33.440344	5	5	5	CONSTRUCCION INTERMODAL	CONSTRUCCION	QUINTA NORMAL	INTERMODAL	1	ESTACION METRO	Linea 5

Table 2: Metro stations dataframe

We had also the name and coordinates in decimal format. In Fig. 2 is the map visualization of this info.

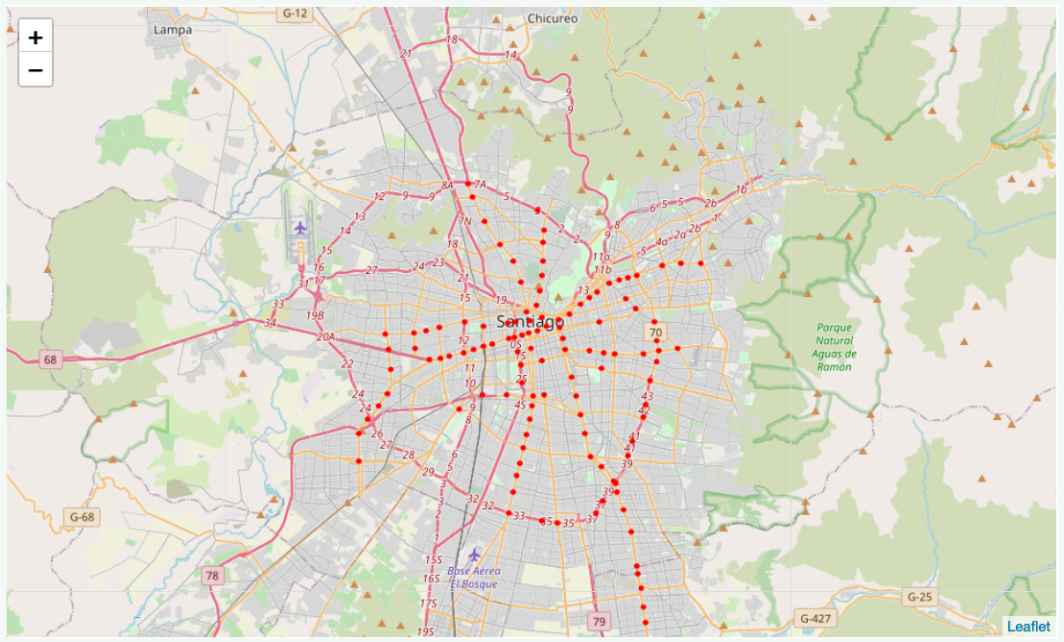

Fig. 2: Metro stations

The number of NYC station was around 500 and in Metro was about 120. So, we confirm that the New York subway has a much bigger network than the Santiago Metro and we can obtain valuable information for a growing city.

The next step was to obtain the actual venues around each station, in NYC and in Santiago, and then, compare this two sets.

The data of venues downloaded considers a radius of 150 m around the station as this study is centered in consumers as pedestrians.

The venues obtained were sorted and the top 20 were selected for the analysis. Venues like Bus station or Metro station were dropped as there is no possibility of establish a business like these.

The frequency of the venues in Santiago are presented in the chart at Fig. 3.

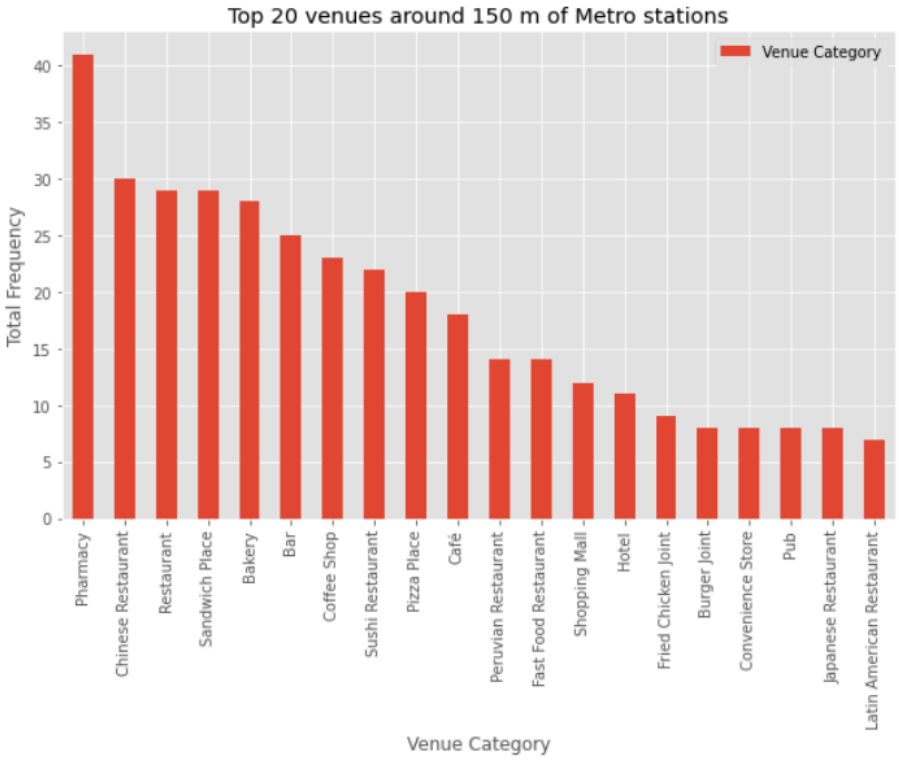

Fig. 3: Venues around Metro stations

It can be seen that, with some distance, the top venue around metro stations are Pharmacies.

This situation has absolute sense, since it is in Chileans culture the high dependence of medication and auto-medication is a wide spread habit.

In the rest top 5 venues, we see restaurants and bakeries, which is very real as Chile is one of the top consumers of bread in the region.

Doing the same analysis for the venues around 150 m of each NYC subway station (Fig. 4), we see similitudes and some differences.

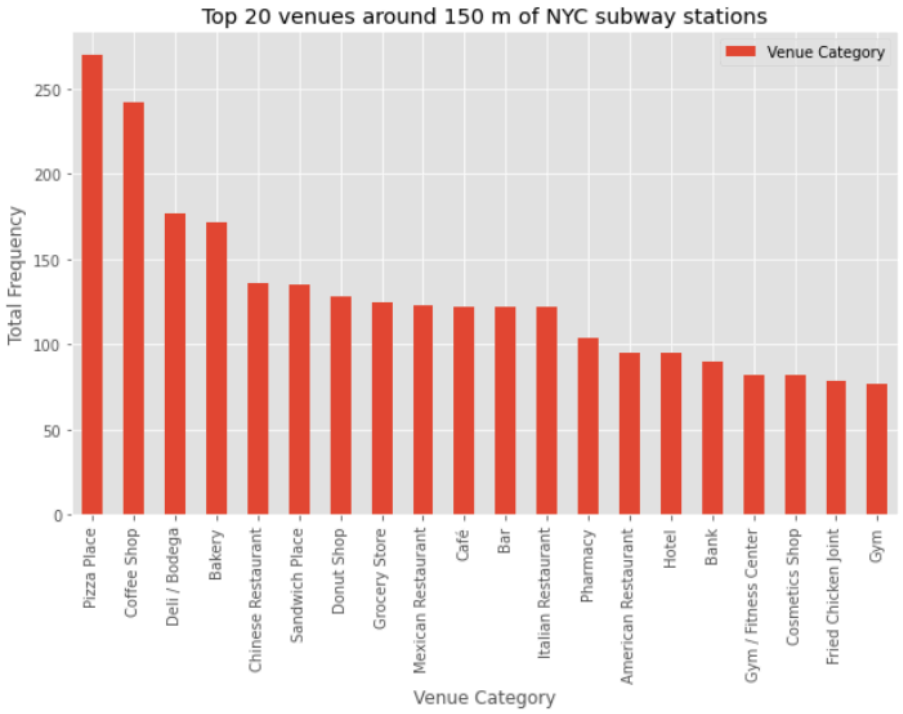

Fig. 4: Venues around NYC subway stations

In this case we have, in the top 5 venues, the bakeries and Chinese restaurants, same as in Chile. The coffee shops here are top 2 and in Chile top 7. The deli/bodega is not seen in Chile (maybe because of different concept) but we see a notable difference in the **pizza places**.

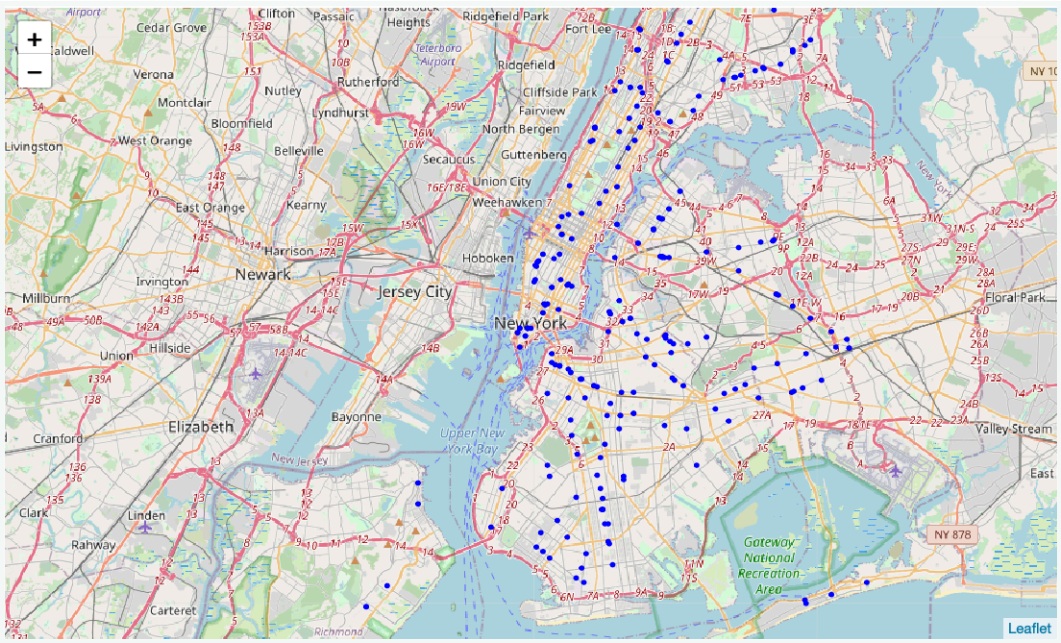

Fig. 5: Pizzerias near NYC subway stations

Pizza places are the top 1 in NYC and top 9 in Santiago. So, we choose the pizza places to deepen the analysis.

Then, we filter the venues result for NYC to see in a map where there are located. And as we see in the map in Fig. 5, the pizza places are evenly distributed in the city, it doesn't show any concentration of this venues in some particular neighborhood, they are all around the city.

For Metro de Santiago, we did the same analysis, and we can see on the map of Fig. 6, that around Metro stations there are very few pizzerias. In Table 3 we can see all the venues returned by Foursquare with their corresponding Metro station.

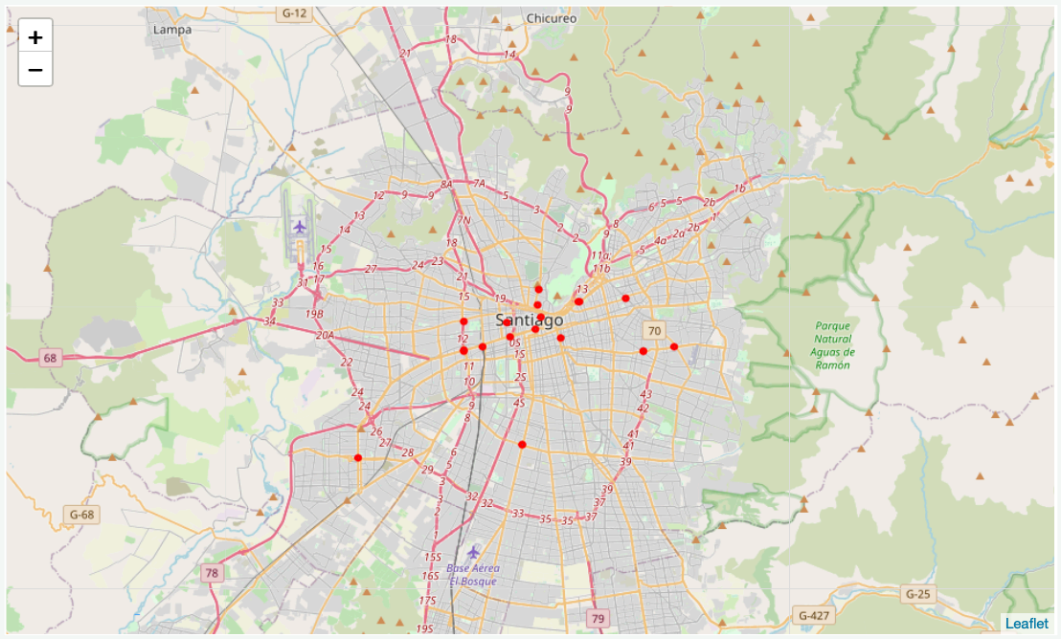

Fig. 6: Pizzerias near Metro stations

There is something particular for this venues in Santiago. The very few pizza places surrounding Metro stations are somewhat concentrated in the center of the city.

So we decided to continue with the analysis only with the data from Santiago since the distribution in NYC seems so spread.

We choose a criteria to define a location for a new pizzeria based upon the number of evaluations received by each venue, under the hypothesis that as more signals received, more people visiting, and more people around looking for pizzas.

The second hypothesis considered is that businesses that are alike, tends to be close from each other. So we will choose a location near the actual pizzeria with the most signals received.

From Foursquare, we can obtain the number of signals received by each venue, so, for pizza places near Metro stations, the top 5 are shown in Table 4.

	Station	Station Latitude	Station Longitude	Venue ID	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	RICARDO CUMMING	-33.438779	-70.664495	4c15929977cea5936a7fd260	Pizzas a La Piedra Benvenutto	-33.439290	-70.664703	Pizza Place
1	PILA DEL GANSO	-33.454256	-70.692291	5963f11d813488243d205a9b	Little Caesars Pizza	-33.454383	-70.691868	Pizza Place
2	PILA DEL GANSO	-33.454256	-70.692291	59f3d0b06c08d16511147091	Papa John's	-33.453789	-70.691787	Pizza Place
3	ESTACION CENTRAL	-33.450894	-70.679965	4da483467ccc816e3867937b	Pizza Hut	-33.452204	-70.679861	Pizza Place
4	REPUBLICA	-33.446896	-70.664245	4b6b0235f964a5204fec2be3	Papalino's Pizza	-33.446922	-70.662667	Pizza Place
5	BELLAS ARTES	-33.436299	-70.643274	4b69aa48f964a52020ac2be3	Pizzería Verace	-33.436691	-70.643457	Pizza Place
6	SANTA LUCIA	-33.442985	-70.646459	4bc4af26920eb713ee191f2c	Da Dino	-33.442810	-70.646727	Pizza Place
7	DEPARTAMENTAL	-33.503434	-70.654956	52213a9011d202a823354e0c	Pizzeria Fugazza (nueva)	-33.503033	-70.654843	Pizza Place
8	DEPARTAMENTAL	-33.503434	-70.654956	4f9f41d6e4b0aa6a18911c9c	Fugazza	-33.502756	-70.654875	Pizza Place
9	SANTA ISABEL	-33.447152	-70.630264	58acd56a9900e670e34bd157	Peyuco	-33.447303	-70.631404	Pizza Place
10	MANUEL MONTT	-33.428974	-70.619388	5a1eed553b83076e7674c4af	Papa John's	-33.428499	-70.619325	Pizza Place
11	MANUEL MONTT	-33.428974	-70.619388	4b6855c4f964a52049722be3	Voraz Pizza	-33.428774	-70.619999	Pizza Place
12	COLON	-33.426530	-70.591209	50ee042de4b024b73ad28a93	Domino's Pizza	-33.426828	-70.590597	Pizza Place
13	PATRONATO	-33.429663	-70.647044	4e4016e418a83d5b2866ea69	La Pizarra	-33.430153	-70.645631	Pizza Place
14	CERRO BLANCO	-33.422182	-70.644948	56c26ae3498ea94e3fcbb1ff	Papa John's	-33.422211	-70.645111	Pizza Place
15	PLAZA DE MAIPU	-33.510491	-70.757166	504e60dee4b05f0232fe5897	Telepizza	-33.509666	-70.757671	Pizza Place
16	DIAGONAL ORIENTE	-33.454750	-70.580188	53a34335498edce8c30abab6	Papa John's	-33.454638	-70.579589	Pizza Place
17	DIAGONAL ORIENTE	-33.454750	-70.580188	50303c27f1366d6c2973056f	Da Cesare&Paolo Pizzeria Ristorante	-33.454467	-70.579695	Pizza Place
18	LARRAIN	-33.452187	-70.558958	58cabc35ef787f1c205dc2c8	Pizzeria Roma	-33.452153	-70.560092	Pizza Place
19	GRUTA DE LOURDES	-33.438088	-70.691187	4cdefcffabfb2c0f802f9b5a	PizzaPizza Santiago Poniente	-33.439096	-70.691608	Pizza Place

Table 3: Actual pizzerias near Metro stations

So, we see that Pizzería Verace is the top signaled venue, the index 5 in the dataset. This Pizzeria is at less than 150 m from the Bellas Artes Metro station, as we see in table 3.

	Nombre	Rating signals
5	Pizzería Verace	346
6	Da Dino	283
11	Voraz Pizza	138
17	Da Cesare&Paolo Pizzeria Ristorante	118
16	Papa John's	71

Table 4: Top 5 pizzerias by rating signals

Then, our recommendation for a new business near a Metro Station is to place a ***Pizzeria*** at 150 m or less around the ***Bellas Artes*** Metro Station.

In the map of Fig. 7 can be seen an area highlighted which shows the 150m radius circle where the venue selected could be placed.

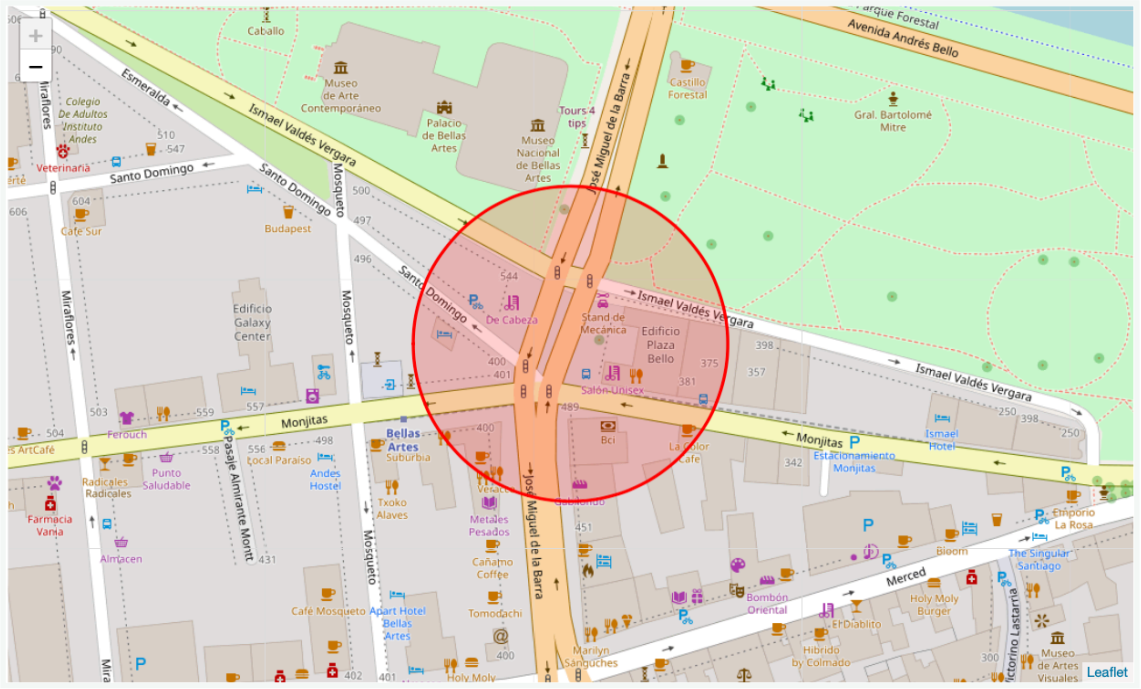

Fig. 7: Location for the business to set up

Discussion

The data obtained from Foursquare seems very reliable since the results of the frequency of venues in Santiago is very consistent with the culture of the Chileans, in terms of cuisine preferences (Chinese food in first place), high dependence of pharmacies and medications and the high consumption of bread (high rank for bakeries)..

The same can be said for the results from the venues obtained for NYC around subway stations. Pizza places everywhere makes all sense, given the high rate of Italian immigration in the city.

Also makes sense that the chosen place is a high transit place, just near the financial center of the city, with high rate of food places around.

Conclusion

This study generated a surprising result, since before the beginning of the work there were not even a clue about what kind of venue could be interesting to locate near a Metro de Santiago station based upon the info obtained from the NYC subway.

The potential of the data available in services like Foursquare is amazing and in combination with the power of languages like Python or similars and the specific libraries needed makes a set of tools for amazing discoveries.

About this study, a deep one can be done considering not only the metro station but with each entrance/exit of a station, adding a radius even smaller than the 150 m considered here, maybe 50 m around each station entrance/exit.

In the case of Santiago, the info containing the coordinates of each station entrance is a much harder dataset to find. For NYC is complete and available.