Where to open a new pharmacy on Gustavo A. Madero, Mexico City

1. Introduction

Location of a new business is crutial to its success. A new Pharmacy needs a constant stream of clients and a constant flow of supplies.The present project aims to recommend locations for a new pharmacy within the county of Gustavo A. Madero in Mexico City. Pharmacy owners looking forward to open a new sucursal will find usefull the results of the analisys.

2.- Data section

2.1- Data sources

The map of the county of Gustavo A. Madero and each of its boroughs can be found on the open data page of the Mexico City Government. https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/export/ Fourthsquare API is used to get the location of general venues and pharmacies within 500 meters of the center of each borough.

Data about the population, access to healthcare services and sanity level of each borough was found on census reports of Mexico City.

http://www.sideso.cdmx.gob.mx/documentos/2010/indice_desarrollo_social/ Indice_Desarrollo_Social_UT.xls
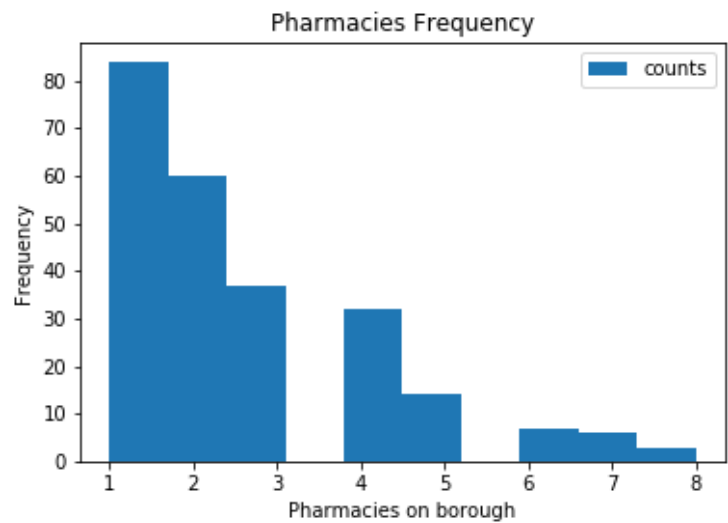
2.2 Data processing

The center of each borough was obtained from the map of the county. Forthsquare data was transformed into a dataframe counting the amount of venues per type for each borough. Socioeconomic data was merged into the previous dataframe using the name of the boroughs. Since some boroughs had a slightly different name on each data set text cleaning was applied to normalize the names.

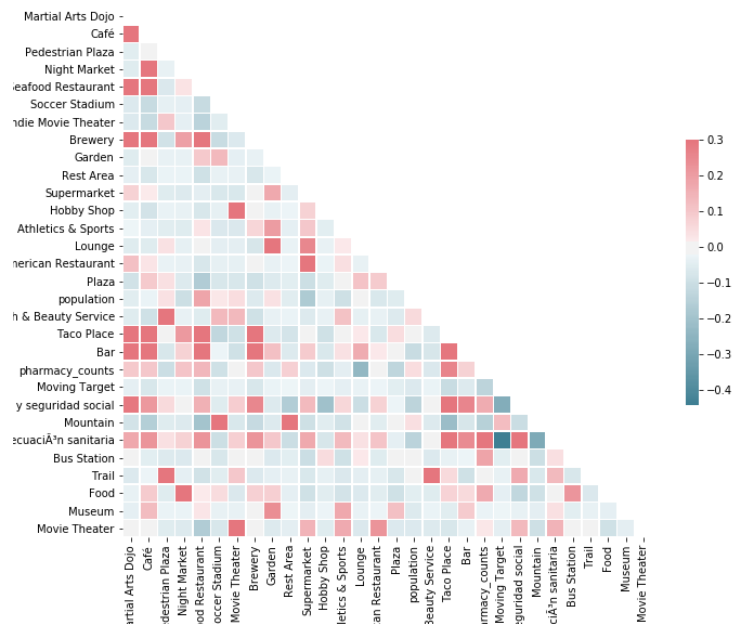|	name_colonia	population	AdecuaciÃ³n sanitaria	Acceso a salud y seguridad social	center	id_colonia	Longitude	Latitude	counts	Moving Target	...	Bus Station	Athletics & Sports	Movie Theater
0	15 DE AGOSTO	3099	0.8857	0.5638	POINT (-99.10795687889645 19.48457404030296)	NaN	-99.107957	19.484574	2	0	...	0	0	0
1	6 DE JUNIO	3241	0.6080	0.4329	POINT (-99.1258656186826 19.55060820667433)	NaN	-99.125866	19.550608	1	1	...	0	0	0
2	7 DE NOVIEMBRE	3869	0.9652	0.6237	POINT (-99.12046074465488 19.46048775211773)	NaN	-99.120461	19.460488	3	0	...	0	0	1
3	ACUEDUCTO DE GUADALUPE	11011	0.9955	0.7201	POINT (-99.15113662198358 19.52716602351677)	NaN	-99.151137	19.527166	1	0	...	0	0	0
4	AHUEHUETES	4945	0.6335	0.5283	POINT (-99.1313543450819 19.54595844315317)	NaN	-99.131354	19.545958	1	0	...	0	0	0

3.- Methodology
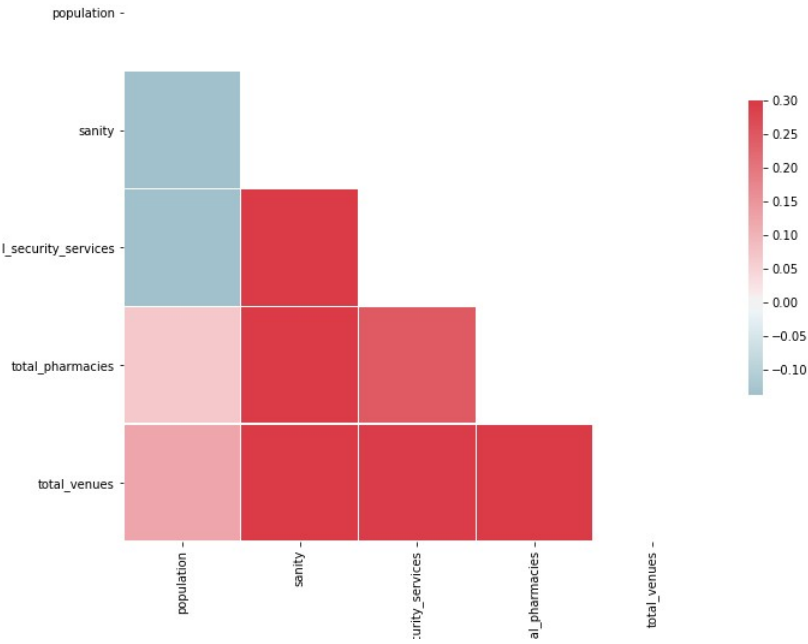## 3.1 Exploratory analisis

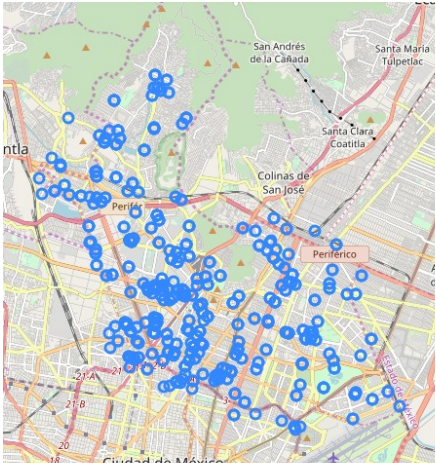The frequency of pharmacy counts per borough is shown next

Correlation between the amount of each venue per borough and the amount of pharmacies per borough was calculated.
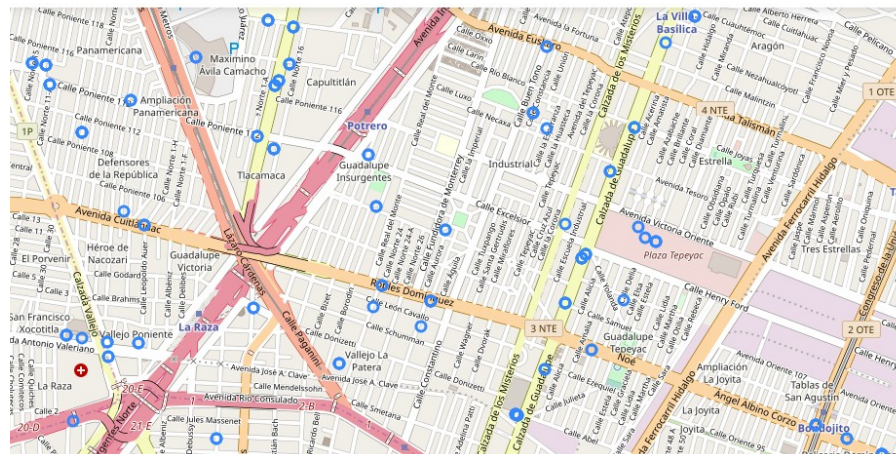
Correlation between socioeconomic variables and the amount of pharmacies per borough was calculated.

Plot of the position of each pharmacy.

It is observed that pharmacies tend to cluster together
and are located along medium sized avenues or small streets near medium size avenues.

3.2 Feature selection
Venue types were ordered by venue counts and only the 100 with the most venues were kept.
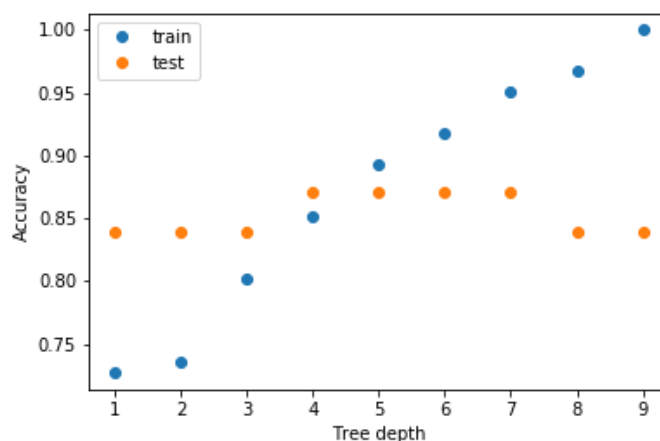Venues with a correlation between -0.05 and 0.3 were discarded.
Using the data about pharmacycounts per borough two target classes were created, those boroughs with just one pharmacy and those with more than one pharmacies.

3.3 Machine learning models
Since this is a classification problem a random forest was used to classify each borough into the categories "one pharmacy" which corespond to label 0 and "More than one pharmacies" which corresponds to the label 1.
The depth of the random forest was chosen by selecting the poing where the test set accuracy begins to diminish.

4. Results

The confusion matrix of the trained model is represented in the next table.

Actual 0	26	12
Actual 1	0	114
	Predicted 0	Predicted 1

5. Dicussion

There are some points which belong the class 0 but where classified on the class 1 which means this boroughs can host more than one pharmacies. Thus it is recommended to select a location within this boroughs and along a medium sized avenue or near one. The name of the boroughs are:

- AMPLIACION PROGRESO NACIONAL
- BENITO JUAREZ
- GUADALUPE INSURGENTES
- LA JOYITA
- SAN JOSE DE LA ESCALERA
- SAN PEDRO EL CHICO
- TRIUNFO DE LA REPUBLICA
- VALLE DE MADERO
- ZONA ESCOLAR ORIENTE
- EX-EJIDO SAN JUAN DE ARAGON SECTOR 32
- PRADERA
- PUEBLO SANTIAGO ATZACOALCO

.

The venues that are most often found near a pharmacy (the ones that fulfilled conditions stated on section 3.2) are Moving Target, American Restaurant, Museum, Garden, Indie Movie Theater, Hobby Shop, Night Market, Health & Beauty Service, Soccer Stadium, Trail, Pedestrian Plaza, Rest Area, Lounge, Food, Martial Arts Dojo, Supermarket, Mountain, Plaza, Bus Station, Athletics & Sports, Movie Theater, Brewery, Café, Bar, Seafood Restaurant and Taco Place.

6. Conclusion

In this study boroughs of Gustavo A. Madero, Mexico City where analyzed to find the ones that can have more than pharmacy but actuallly have just one. A model was built to detect the models wich can host more than one pharmacy. A list of the names of these boroughs is provided.