The Battle of the Neighborhoods

How to improve Neighborhood security & life quality

1. Introduction & Business Problem

Today architects, urban planners and public administrators are faced with the need to redesign large cities to guarantee citizens, in continuous increase and often with a progressive concentration in a few neighborhoods, greater safety and quality of life. The large investments in "smart cities" that are being made are addressed in this sense. Unfortunately one of the most serious problems that many large cities are facing is that of widespread crime especially in some neighborhoods: a lot of work has been done in recent years (think of the redevelopment of the Bronks neighborhood in New York) but much remains to be done do.

Our goal is to provide an analytical tool to study a possible correlation between crime and neighborhood characteristics so that targeted redevelopment interventions can be carried out by optimizing the available resources.

1.1 Target Audience

Public administration with particular reference to urban planners, mayors and all those who are dealing with the modernization of cities and whose are invloved in “smart cities” projects.

2. Data

One city will be analysed in this project : San Francisco.

We'll be using the following datasets for analyzing this case.

Data 1 : San Francisco Crime Dataset available at URL
https://cocl.us/sanfran_crime_dataset

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	Neighborhood	Resolution	Address	X	Y	Location
0	120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)
1	120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.403404791479)
2	141059263	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST / SHAFTER AV	-122.388856	37.729981	(37.7299809672996, -122.388856204292)
3	160013662	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES ST / OFARRELL ST	-122.412971	37.785788	(37.7857883766888, -122.412970537591)
4	160002740	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST / MISSION ST	-122.419672	37.765050	(37.7650501214668, -122.419671780296)
5	160002869	ASSAULT	BATTERY	Friday	01/01/2016 12:00:00 AM	21:35	NORTHERN	NONE	1700 Block of BUSH ST	-122.426077	37.788019	(37.788018555829, -122.426077177375)

Data 2 : San Francisco Neighborhood

We'll scrape the following page containing San Francisco zipcode, http://www.healthysf.org/bdi/outcomes/zipmap.htm.

	Zip Code	Neighborhood	Latitude	Longitude
1	94102	Hayes Valley/Tenderloin/North of Market	37.780	-122.420
2	94103	South of Market	37.780	-122.410
3	94107	Potrero Hill	37.770	-122.390
4	94108	Chinatown	37.791	-122.409
5	94109	Polk/Russian Hill (Nob Hill)	37.790	-122.420
6	94110	Inner Mission/Bernal Heights	37.750	-122.420
7	94112	Ingelside-Excelsior/Crocker-Amazon	37.720	-122.440
8	94114	Castro/Noe Valley	37.760	-122.440
9	94115	Western Addition/Japantown	37.790	-122.440
10	94116	Parkside/Forest Hill	37.740	-122.480
11	94117	Haight-Ashbury	37.770	-122.440
12	94118	Inner Richmond	37.780	-122.460
13	94121	Outer Richmond	37.800	-122.700
14	94122	Sunset	37.760	-122.480
15	94123	Marina	37.800	-122.440
16	94124	Bayview-Hunters Point	37.730	-122.380
17	94127	St. Francis Wood/Miraloma/West Portal	37.730	-122.460
18	94131	Twin Peaks-Glen Park	37.750	-122.440
19	94132	Lake Merced	37.720	-122.480
20	94133	North Beach/Chinatown	37.800	-122.440
21	94134	Visitacion Valley/Sunnydale	37.720	-122.410

Neighborhhod geographical coordinates provieded by previous dataset will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We'll use the Foursquare API to explore neighborhoods in San Francisco.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Herbst Theater	37.779548	-122.420953	Concert Hall
1	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	War Memorial Opera House	37.778601	-122.420816	Opera House
2	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	San Francisco Ballet	37.778580	-122.420798	Dance Studio
3	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	Louise M. Davies Symphony Hall	37.777976	-122.420157	Concert Hall
4	Hayes Valley/Tenderloin/North of Market	37.78	-122.42	War Memorial Court	37.779042	-122.420971	Park

...

...

Fig. 1 - Top 100 venues that are in each neighborhood within a radius of 500 meters

Next step will be clustering Neighborhoods based on previous venues information. So we'll study correlation between clusters and crime dataset.

3. Methodology

Business Understanding

Our goal is to study correlation between Neighborhood Clusters venues and crime diffusion in San Francisco.

Analytic Approach

In this project we collect and analyze data as following steps :

- San Francisco crime events ;
- Collect data about San Francisco neighborhood ;
- Qualitative superposition of neighborhood clusters and crime events map :
- Collect crime events frequency for the most common venues type ;
- Study correlation between data

This is done in the following Exploratory data analysis section.

Exploratory Data Analysis

Data 1 - San Francisco crime events.

Data, collected through https://cocl.us/sanfran_crime_dataset, are grouped by Latitude & Longitude in order to facilitate joining with other data set. So we have the following data :

	Longitude	Latitude	Count
0	-122.51	37.73	7
1	-122.51	37.74	98
2	-122.51	37.75	112
3	-122.51	37.76	444
4	-122.51	37.77	498
5	-122.51	37.78	587
6	-122.50	37.72	81
7	-122.50	37.73	70
8	-122.50	37.74	331
9	-122.50	37.75	237
10	-122.50	37.76	312

...

Fig. 2 - Latitude & Longitude Crime Events Count

Data 2 - San Francisco Neighborhood.

Data are collected through http://www.healthysf.org/bdi/outcomes/zipmap.htm

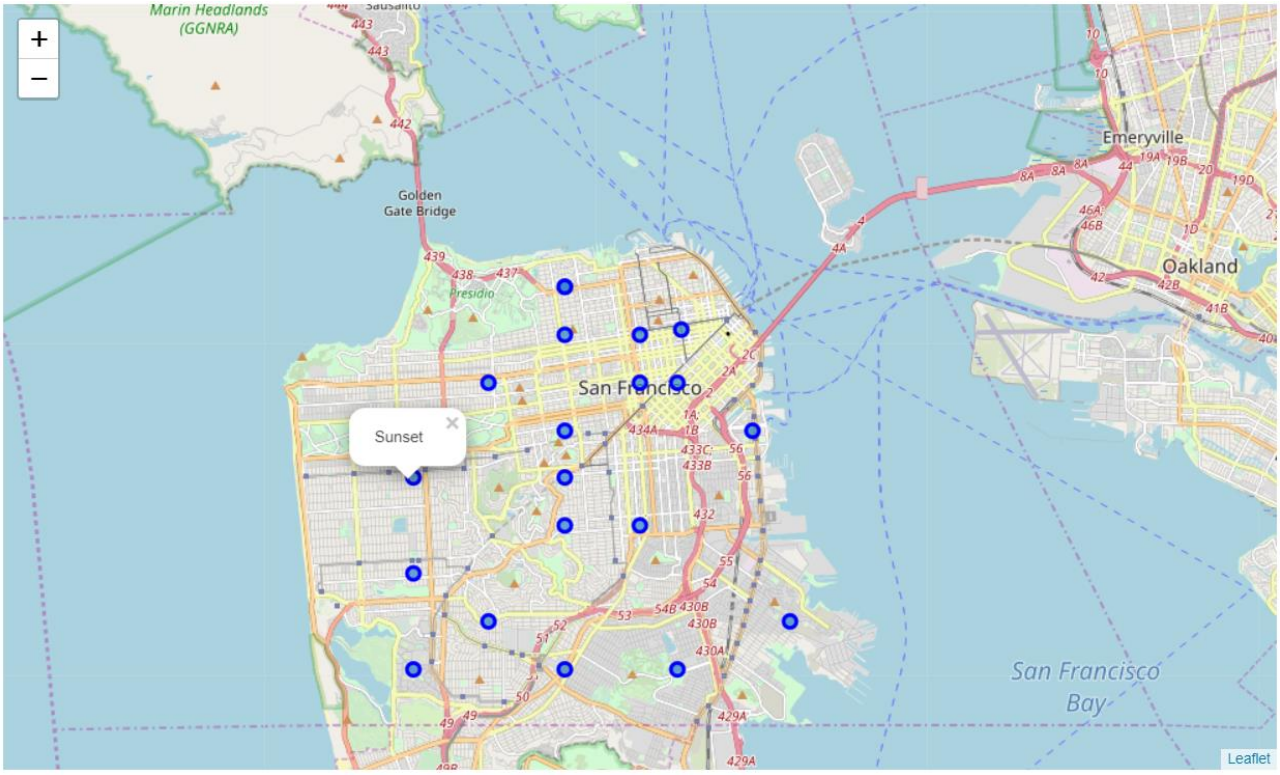

Fig. 3 - San Francisco Neighborhood

Data 3 : Cluster San Francisco neighborhoods

Venues data are collected from Foursquare

Map offer a qualitative superposition of clusters and crime events

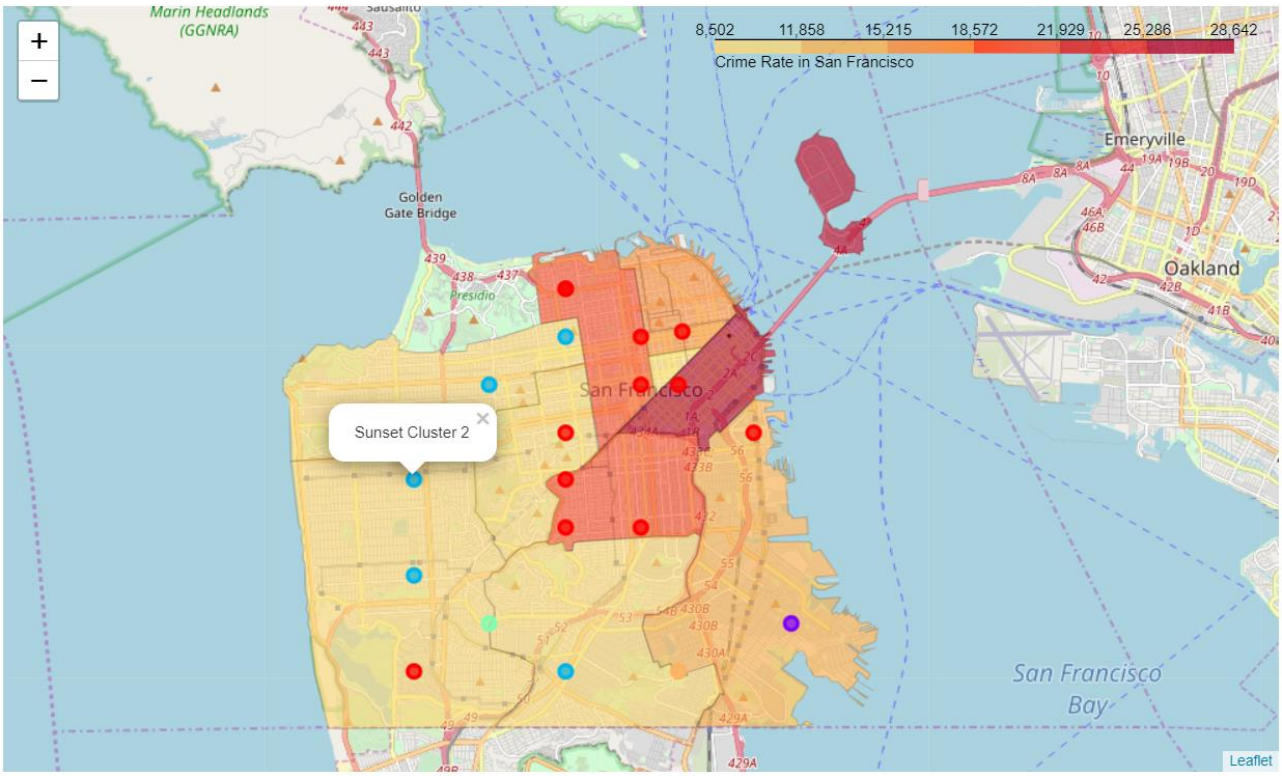

Fig. 4 - Qualitative superposition of neighborhood clusters and crime events map

Data 4 : Collect crime events frequency for the most common venues type

	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Count	1st_frequency	2nd_frequency	3rd_
0	0	Coffee Shop	Vietnamese Restaurant	Cocktail Bar	Performing Arts Venue	Theater	Café	Beer Bar	French Restaurant	Mexican Restaurant	Juice Bar	7738	5	4	
1	0	Coffee Shop	Sandwich Place	Theater	Vietnamese Restaurant	American Restaurant	Wine Bar	Café	Bakery	Music Venue	Marijuana Dispensary	13189	9	4	
2	0	Food Truck	Coffee Shop	Pharmacy	Gym	Park	Sandwich Place	Bank	Pizza Place	Café	Harbor / Marina	566	9	5	
3	0	Sushi Restaurant	Bar	Grocery Store	Thai Restaurant	Mexican Restaurant	Coffee Shop	Diner	Bakery	Massage Studio	Vietnamese Restaurant	5250	4	4	
4	0	Mexican Restaurant	Italian Restaurant	Dive Bar	Grocery Store	Pizza Place	Massage Studio	Café	Cocktail Bar	Coffee Shop	New American Restaurant	2280	7	3	
5	2	Pizza Place	Mexican Restaurant	Sandwich Place	Coffee Shop	Bar	Chinese Restaurant	Vietnamese Restaurant	Bus Station	Burrito Place	Filipino Restaurant	1509	5	3	
			Clothing		Yoga	Coffee			Thai	Grocery					

10th Most Common Venue	Count	1st_frequency	2nd_frequency	3rd_frequency	4th_frequency	5th_frequency	6th_frequency	7th_frequency	8th_frequency	9th_frequency	10th_frequency
Juice Bar	7738	5	4	4	3	3	3	3	3	2	2
Marijuana Dispensary	13189	9	4	4	3	3	3	3	3	2	2
Harbor / Marina	566	9	5	3	3	3	2	2	2	2	2
Vietnamese Restaurant	5250	4	4	4	3	3	3	3	3	3	3
New American Restaurant	2280	7	3	3	3	2	2	2	2	2	2
Filipino Restaurant	1509	5	3	2	2	2	2	2	1	1	1

Data 5 : Correlation Matrix

	Cluster Labels	Count	1st_frequency	2nd_frequency	3rd_frequency	4th_frequency	5th_frequency	6th_frequency	7th_frequency	8th_frequency	9th_frequency	10th_frequency
Cluster Labels	1	-0.37	-0.39	-0.4	-0.45	-0.4	-0.51	-0.47	-0.49	-0.61	-0.58	-0.54
Count	-0.37	1	0.27	0.17	0.25	0.19	0.26	0.28	0.29	0.47	0.26	0.27
1st_frequency	-0.39	0.27	1	0.81	0.72	0.75	0.79	0.77	0.71	0.72	0.72	0.72
2nd_frequency	-0.4	0.17	0.81	1	0.96	0.96	0.96	0.94	0.94	0.89	0.92	0.92
3rd_frequency	-0.45	0.25	0.72	0.96	1	0.97	0.96	0.96	0.97	0.93	0.94	0.94
4th_frequency	-0.4	0.19	0.75	0.96	0.97	1	0.96	0.96	0.96	0.9	0.93	0.95
5th_frequency	-0.51	0.26	0.79	0.96	0.96	0.96	1	0.98	0.97	0.91	0.93	0.93
6th_frequency	-0.47	0.28	0.77	0.94	0.96	0.96	0.98	1	0.99	0.93	0.94	0.94
7th_frequency	-0.49	0.29	0.71	0.94	0.97	0.96	0.97	0.99	1	0.94	0.95	0.94
8th_frequency	-0.61	0.47	0.72	0.89	0.93	0.9	0.91	0.93	0.94	1	0.97	0.95
9th_frequency	-0.58	0.26	0.72	0.92	0.94	0.93	0.93	0.94	0.95	0.97	1	0.98
10th_frequency	-0.54	0.27	0.72	0.92	0.94	0.95	0.93	0.94	0.94	0.95	0.98	1

4. Results

Qualitative results

Data visualization (fig. 4) suggest that the greater number of crime event is located in San Francisco north-east cluster 0 neighborhood

Clusters 1,2,3,4 has a low crime rate.

Quantitative results

Correlation matrix calculated on all available metrics add no further information due to limited data but show a strange correlation between 8th_frequency venue type and crime rate : perhaps a specific venue type in this class is more related with crime rate.

It would be necessary a deep study on a more complete and updated data.

5. Discussion

- It’s useful to study in deep qualitative results : maps suggests correlation between metrics involved in this analysis ;
- This is an “abstract” analysis : perhaps a more “contestualized” analysis with urban experts would’be useful ;
- It's necessary to study also the correlation between social context and single clusters and neighborhoods ;
- Compare analysis with other cities ;
- This analysis would be a starting point for a more complete project.

6. Conclusion & Further Investigations

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available and updated there is scope to come up with better results.

Surely this is an investigation that is useful to complete because there is an evidence, at now qualitative, that one o more venue type is strictly related with crime rate.

This would be a good project for San Francisco municipality in order to qualify neighborhoods or update urban developement policy.