Use Case - Exercise 1

The purpose of this exercise is to reproduce the same case explained in class, but now changing the kind of incident.

I have selected “Assault” incident as study subject.

Although comments have been made in the Jupyter notebook file, here I’m going to summarize the most important parts of the task step by step.

STEP 1 & 2: Importing packages and glancing the dataset

After importing packages required, we have had a look at the dataset using bar graphs, showing number of crimes reported by districts (fig 1) and by categories (fig 2)

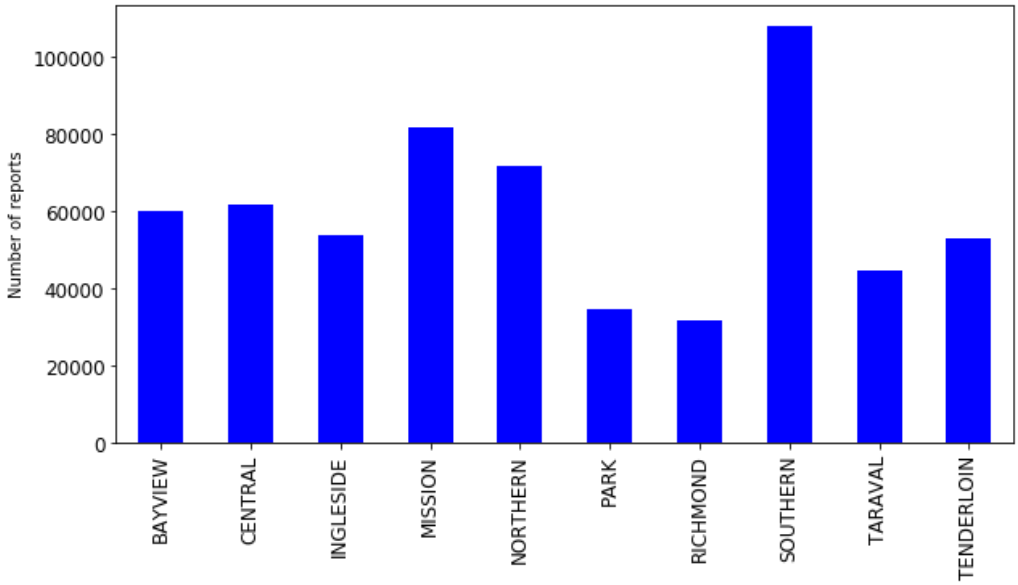

Fig 1

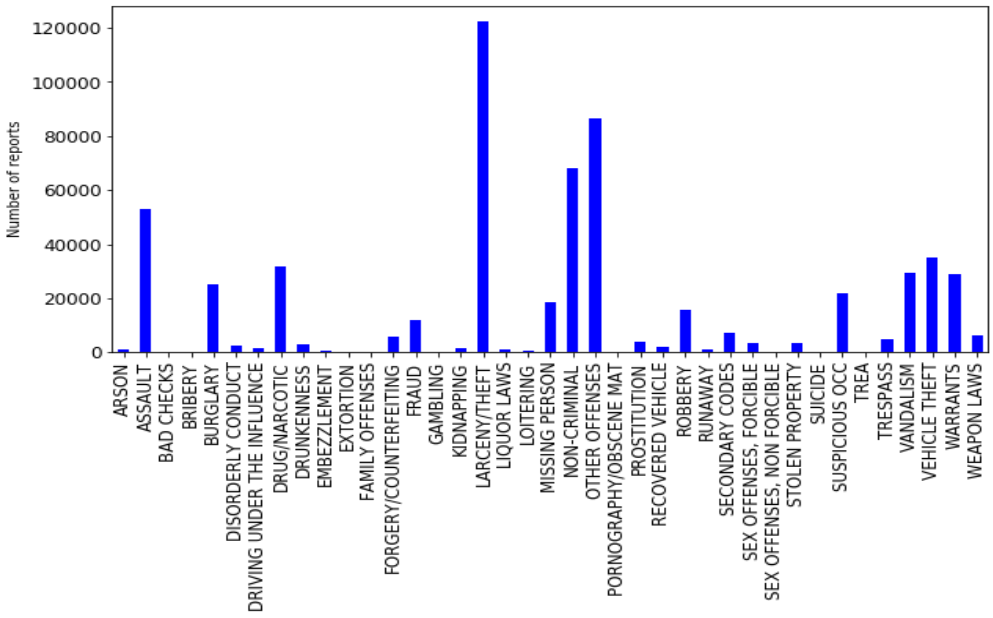

Fig 2

Then, we have printed cluster map making a relationship between districts, categories and number of crimes reported (fig 3).

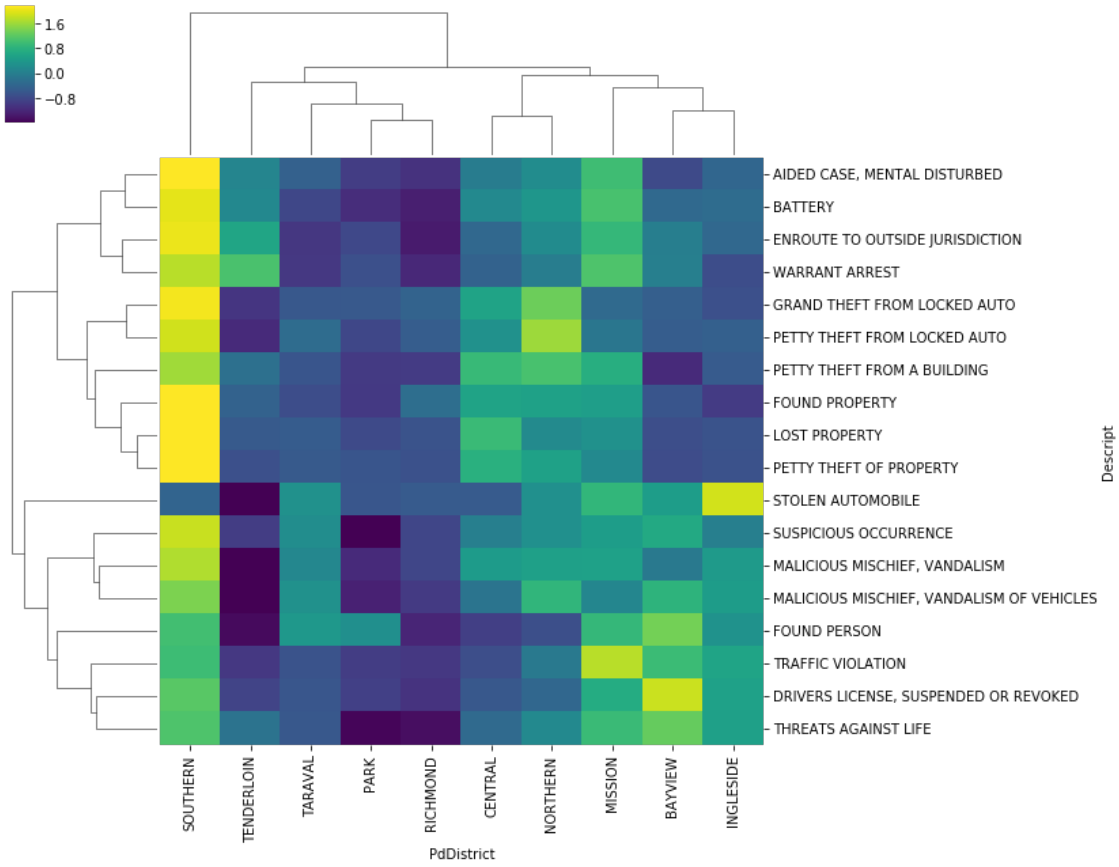

Fig 3

STEP 3: Connecting and loading to MongoDB. Filtering by "Assault" crimes

First I have connected to MongoDB server and then I have filtered by Assault

The output shown is:

```
BATTERY                                              17764
THREATS AGAINST LIFE                                  9267
INFLICT INJURY ON COHABITEE                           4932
AGGRAVATED ASSAULT WITH A DEADLY WEAPON               4510
AGGRAVATED ASSAULT WITH BODILY FORCE                  3639
BATTERY, FORMER SPOUSE OR DATING RELATIONSHIP         2708
AGGRAVATED ASSAULT WITH A KNIFE                       1734
CHILD ABUSE (PHYSICAL)                                 980
BATTERY OF A POLICE OFFICER                            900
BATTERY WITH SERIOUS INJURIES                          747
Name: Descript, dtype: int64
```

Later I have made a relationship between districts and the assault type by their description and put it in a cluster map (fig 4).

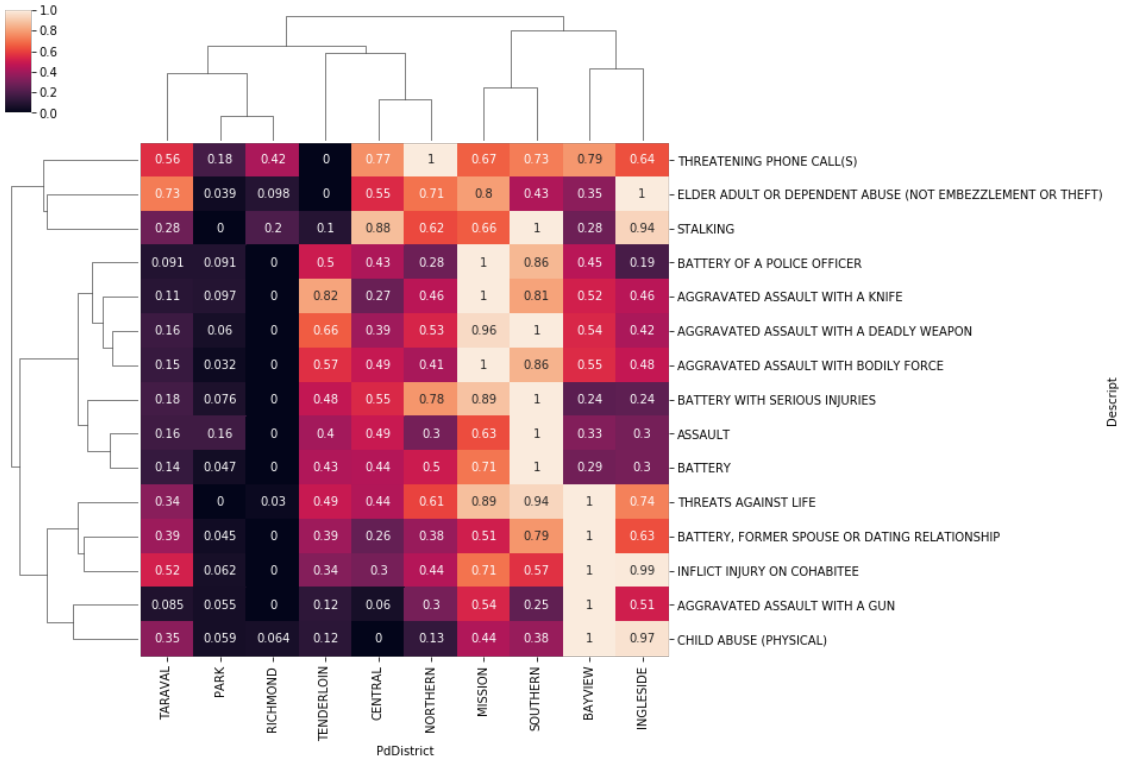

Fig 4

STEP 4: Time series analysis

After grouping category by month with a function, I have filtered Assaults description crimes by "Battery". This is the result:

```
Descript
BATTERY                                          17764
BATTERY DURING LABOR DISPUTE                         4
BATTERY OF A POLICE OFFICER                        900
BATTERY WITH SERIOUS INJURIES                      747
BATTERY, FORMER SPOUSE OR DATING RELATIONSHIP     2708
dtype: int64
['BATTERY', 'BATTERY DURING LABOR DISPUTE', 'BATTERY OF A POLICE OF
FICER', 'BATTERY WITH SERIOUS INJURIES', 'BATTERY, FORMER SPOUSE OR
DATING RELATIONSHIP']
```

Attending to that, I have differentiated between "Battery situation". I called that "situation" because it could be because of a relationship, a labor dispute or an argument with a police officer and it could involve a serious injury or not either of them.

So I've defined these four situations: "labor", "police", "injuries" and "dating".

Analyzing per situation, I have obtained this graph (fig 5):

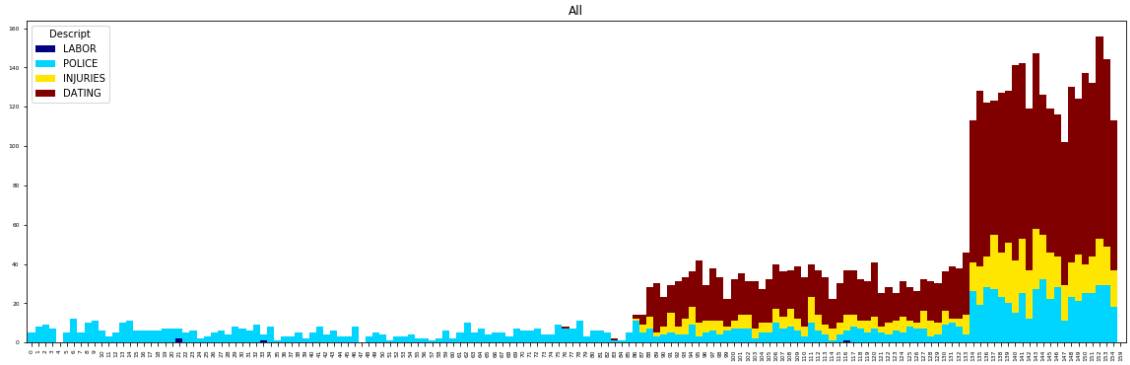

Fig 5

Putting it into a plot graph, the result is (fig 6):

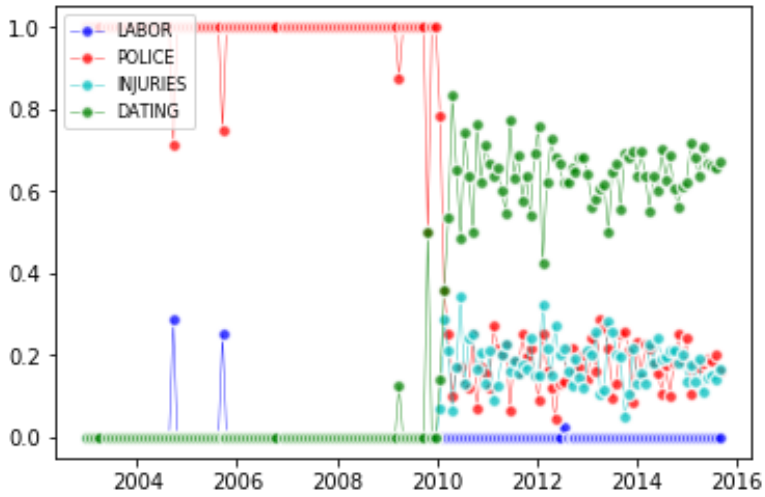

Fig 6

As we can see, there are significant outliers until 2010, so I filtered and the result is (fig 7):

Fig 7

It is nonsense to stablish relationships between categories I selected, but I did anyway for the purpose of the exercise to show what it can be done and replicate what we did in class. It is possible this part would be more interesting choosing any other category rather than assault.

That said, I have try to make a correlation by Police Officer and a serious injury involved (fig 8). The result doesn't show any correlation as expected.

Fig 8

STEP 5: Correlation analysis

I have try to make a relationship between different districts for the same Assault crime type. To do that, first I have done a heatmap with "situations" and districts (fig 9), and then I have compared between number of dating or former spouse relationships situation in batteries assault crimes (fig 10).

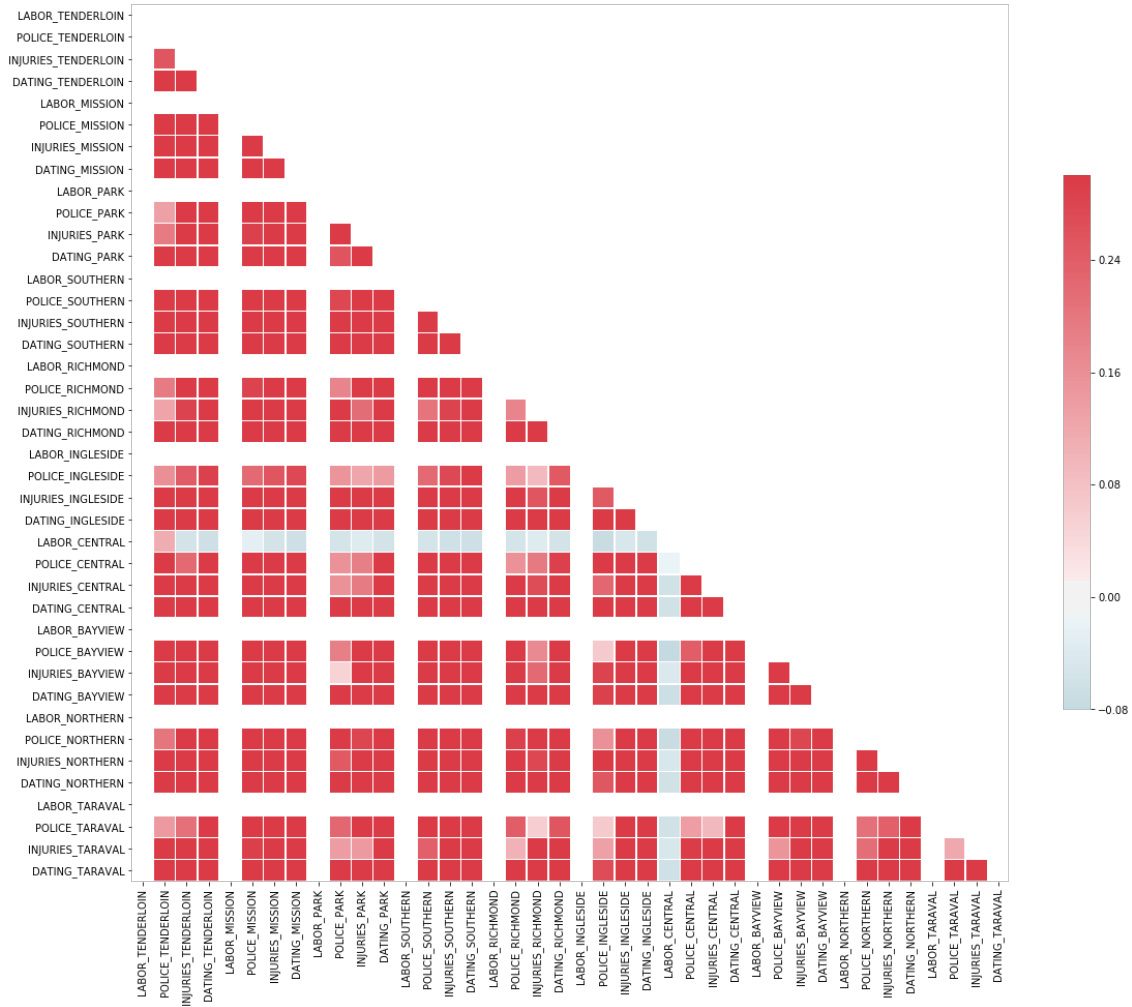
LABOR_TENDERLOIN
POLICE_TENDERLOIN
INJURIES_TENDERLOIN
DATING_TENDERLOIN
LABOR_MISSION
POLICE_MISSION
INJURIES_MISSION
DATING_MISSION
LABOR_PARK
POLICE_PARK
INJURIES_PARK
DATING_PARK
LABOR_SOUTHERN
POLICE_SOUTHERN
INJURIES_SOUTHERN
DATING_SOUTHERN
LABOR_RICHMOND
POLICE_RICHMOND
INJURIES_RICHMOND
DATING_RICHMOND
LABOR_INGLESIDE
POLICE_INGLESIDE
INJURIES_INGLESIDE
DATING_INGLESIDE
LABOR_CENTRAL
POLICE_CENTRAL
INJURIES_CENTRAL
DATING_CENTRAL
LABOR_BAYVIEW
POLICE_BAYVIEW
INJURIES_BAYVIEW
DATING_BAYVIEW
LABOR_NORTHERN
POLICE_NORTHERN
INJURIES_NORTHERN
DATING_NORTHERN
LABOR_TARAVAL
POLICE_TARAVAL
INJURIES_TARAVAL
DATING_TARAVAL
0.24
0.16
0.08
0.00
−0.08

Fig 9

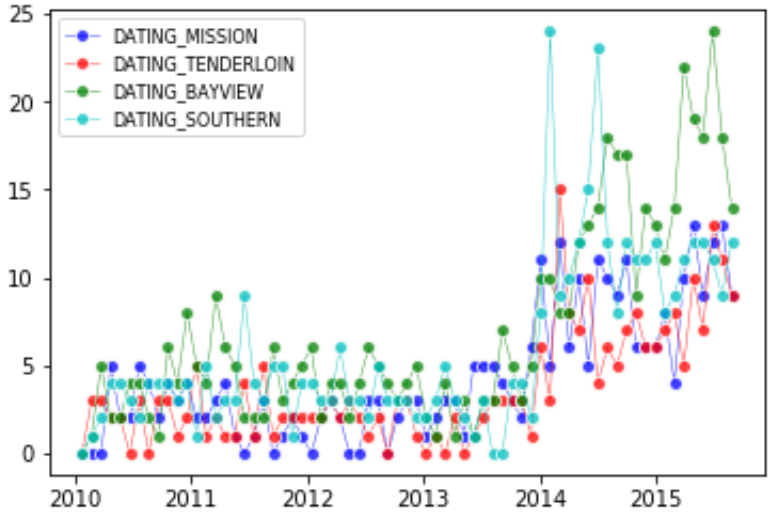
DATING_MISSION
DATING_TENDERLOIN
DATING_BAYVIEW
DATING_SOUTHERN
25
20
15
10
5
0
2010
2011
2012
2013
2014
2015

Fig 10

STEP 6: Mapping relationships

For this last step, it has been stablished two different regimes according to the number of police battery assault recorded over time. I have used a geojson map imported from https://data.sfgov.org/Geographic-Locations-and-Boundaries/Analysis-Neighborhoods/p5b7-5n3h

It has been created two maps:

Map 1: contain every battery assault that involved police officer, in which green is for new regimes and red is for old ones (fig 11).

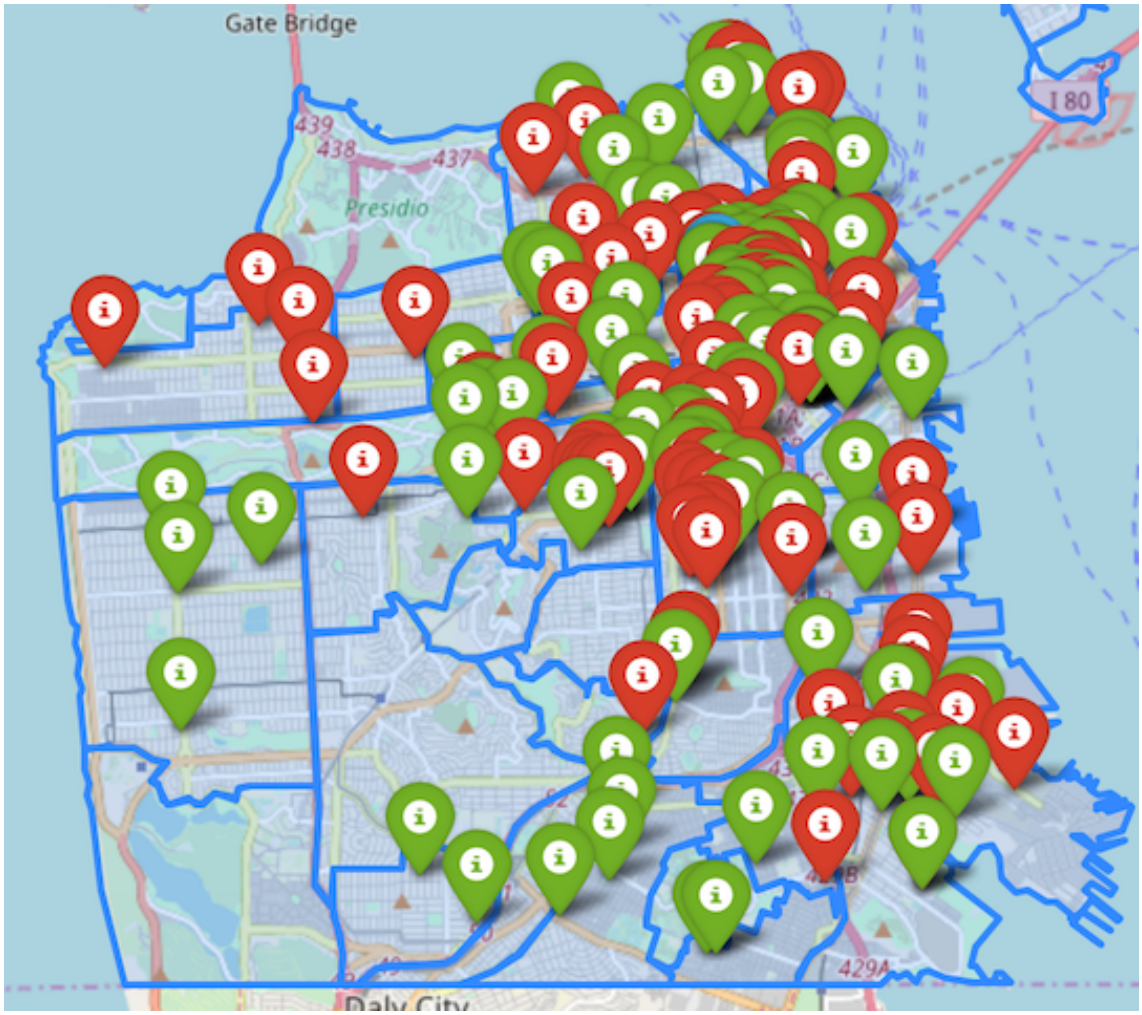

Fig 11

Map 2: selected point into the map imported (fig 12).

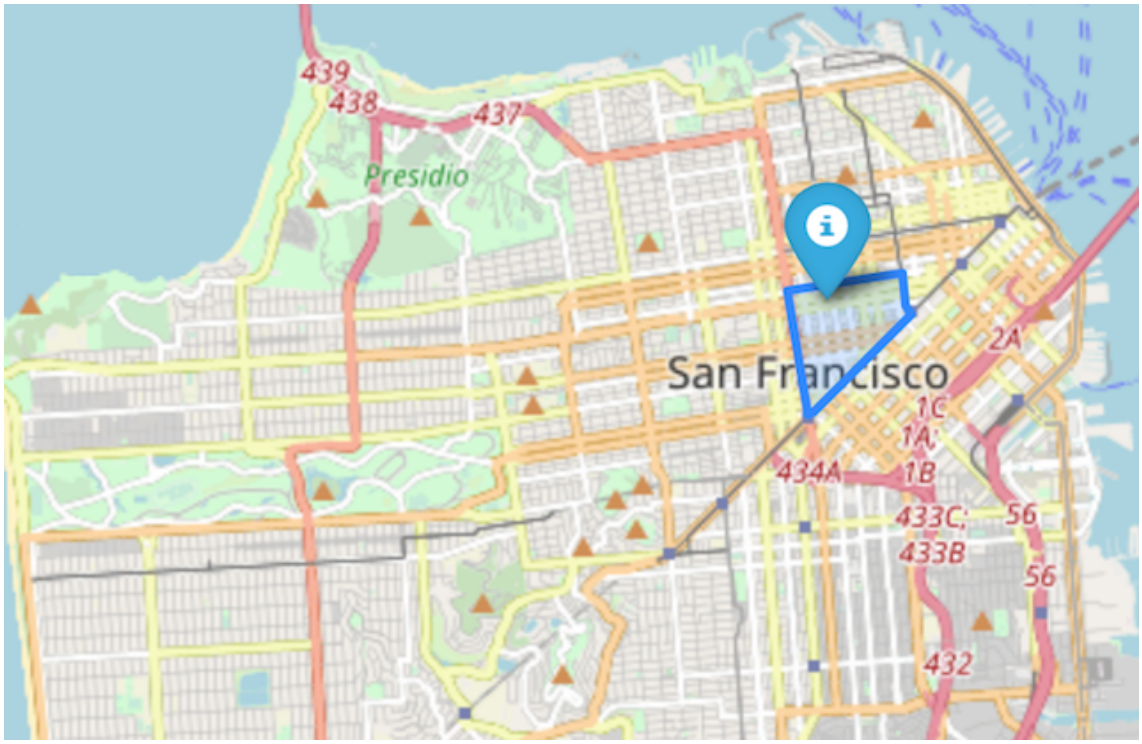
439
438
437
Presidio
San Francisco
2A
1C
1A
1B
434A
433C
433B
56
56
432

Fig 12