Neighborhoods security & life quality

How machine learning can provide a support to build secure smart cities

Abstracts

- Introduction
- Data
- Exploratory Data Analysis
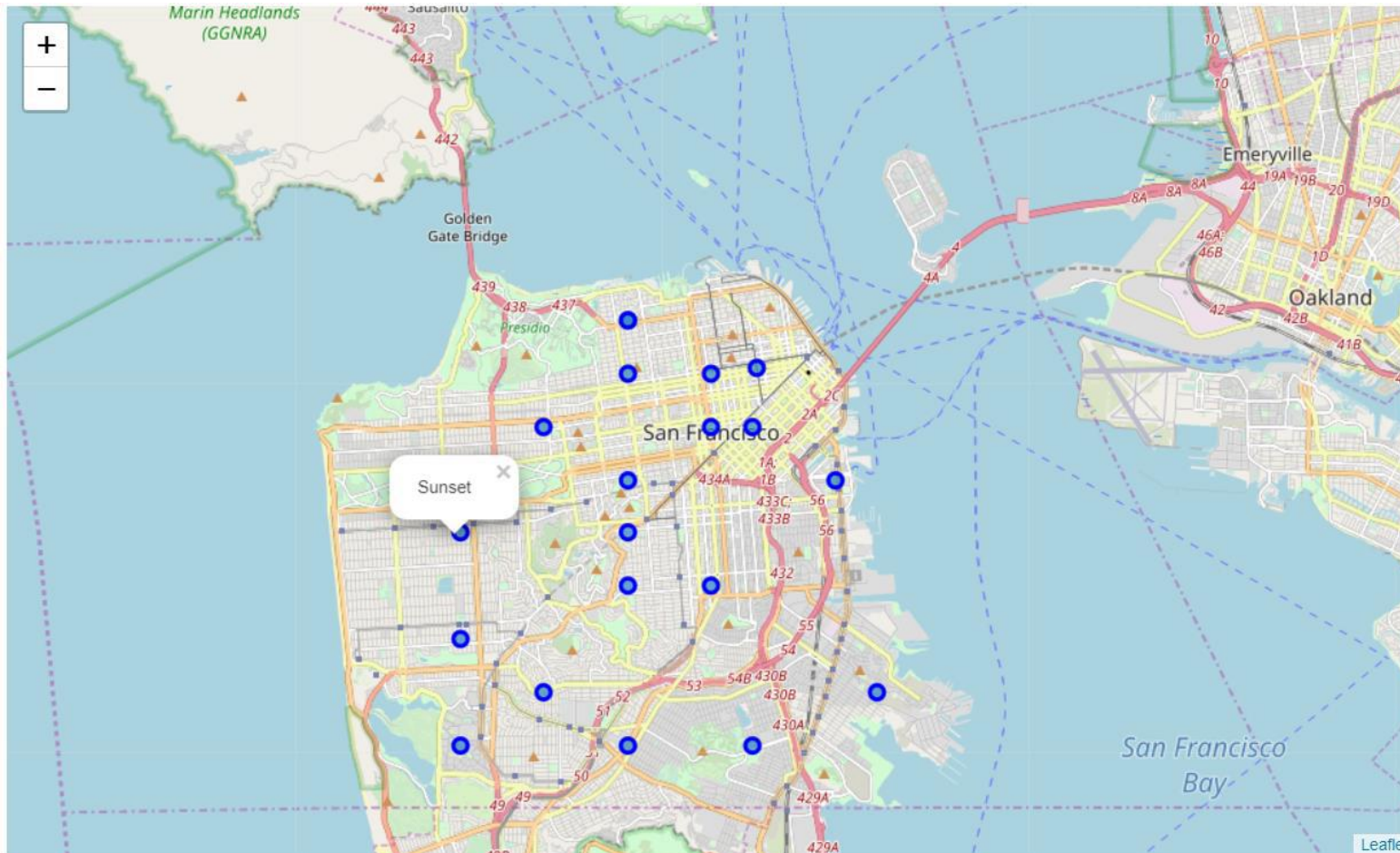- Results
- Conclusion and future directions

Introduction & Goals

- Improve security & life quality in «smart» cities ;

- Optimize available resource ;

- Get new directions for urban development.

Data involved in analysis

- San Francisco Crime Dataset available at URL https://cocl.us/sanfran_crime_dataset
- San Francisco zipcode http://www.healthysf.org/bdi/outcomes/zipmap.htm
- San Francisco neighborhood venues through Foursquare API at : https://api.foursquare.com/v2/venues/explore
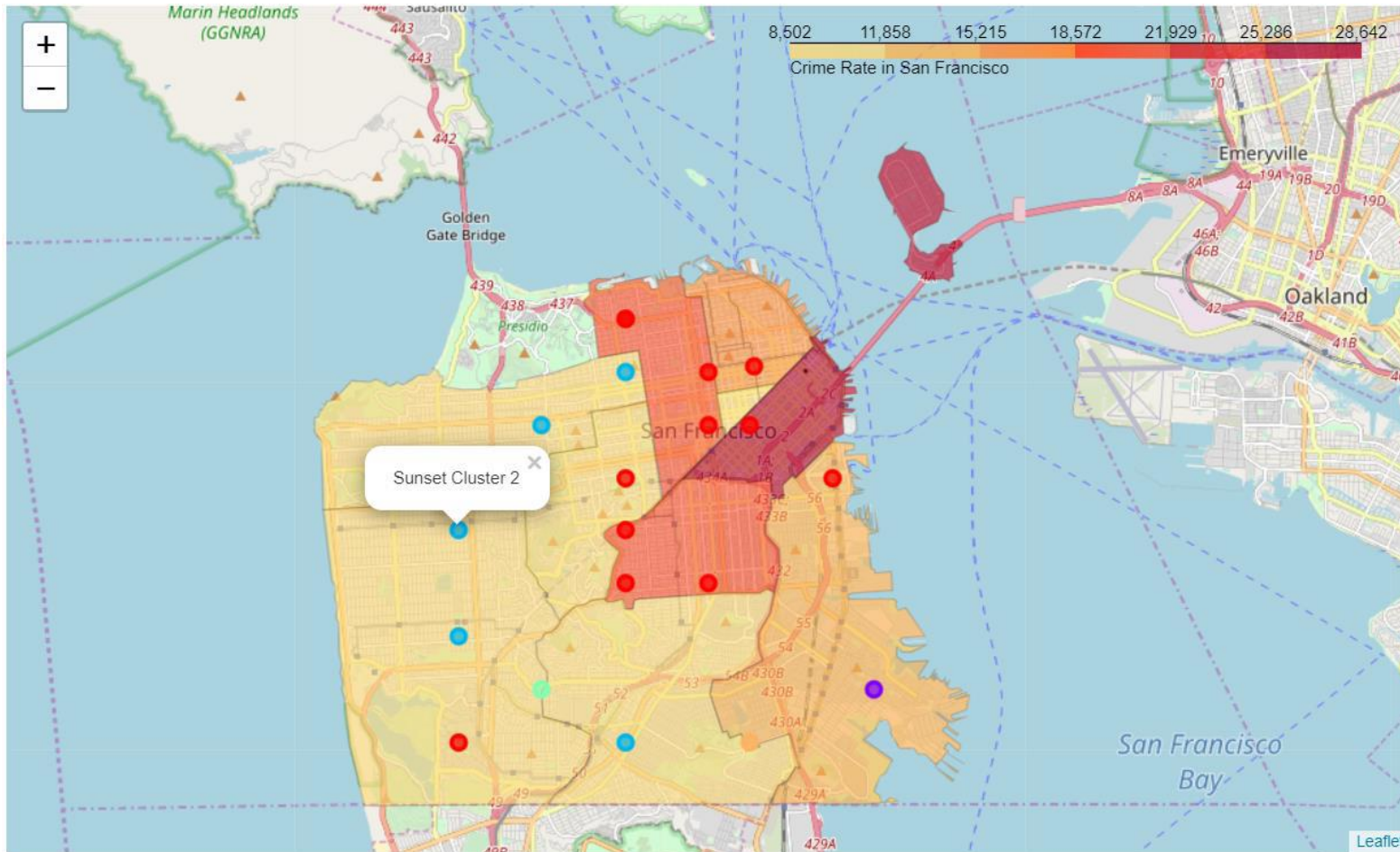
Exploratory Data Analysis (1)

San Francisco Neighborhoods

Exploratory Data Analysis (2)

San Francisco Neighborhoods Clusters superpositioned with crime rate map

Exploratory Data Analysis (3)

	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

Correlation Matrix between all referenced metrics

Results

- **Qualitative results**
  - Data visualization 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.

Conclusion and future directions

➢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.