RELATIONSHIP BETWEEN NUMBER OF POLICE INCIDENTS AND VENUES CATEGORY IN THE NEIGHBORHOODS OF SAN FRANCISCO

Fernando Luiz Trazzi Junior

February 05, 2021

Introduction

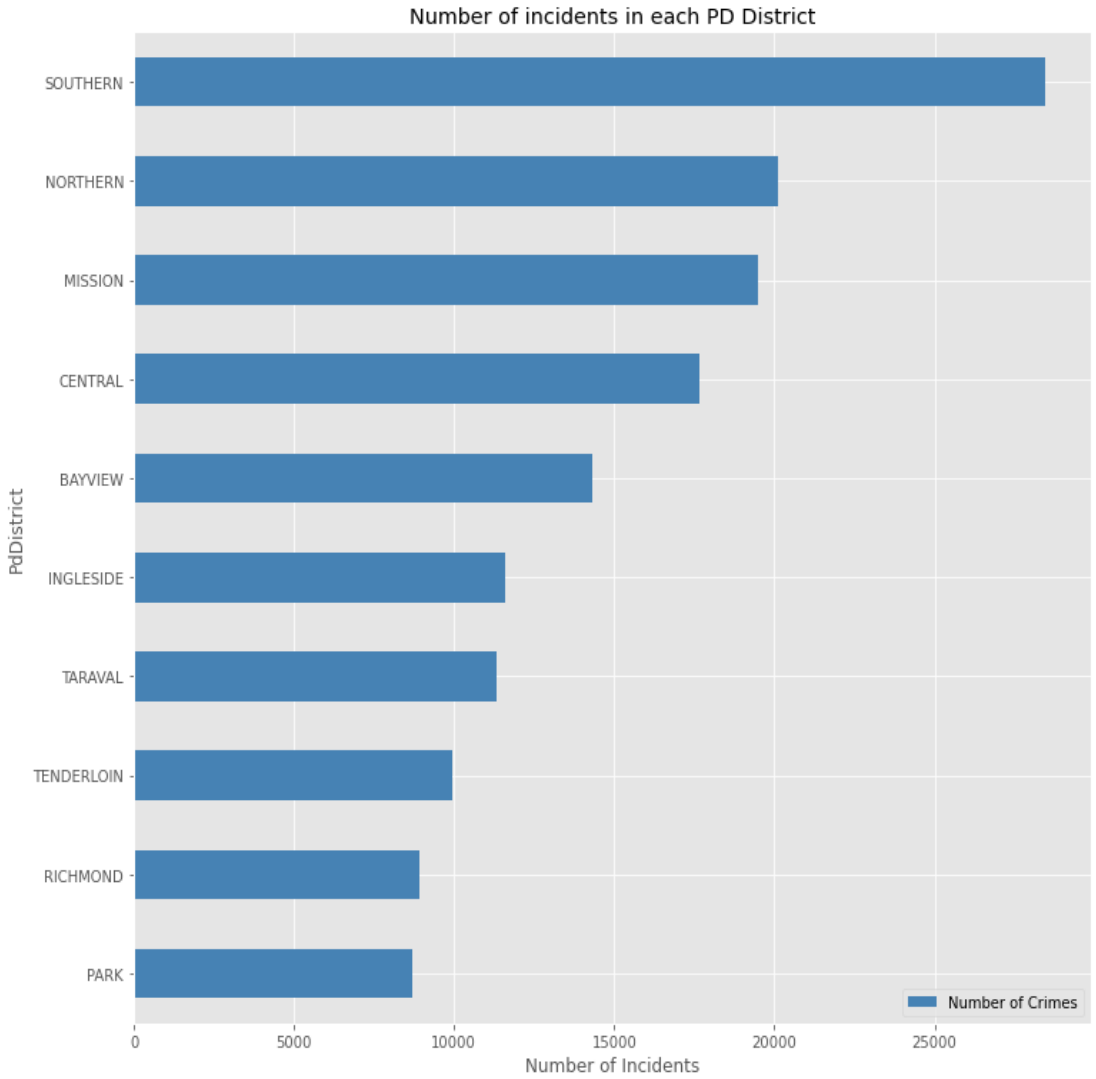
- Background:
  - Throughout the Data Visualization course, we used some information from the city of San Francisco to generate a map and present police statistics on each of the city's 10 police districts in 2016. In addition, in the introduction to this Capstone Project, we learned to search for information about different places, such as stores, restaurants, parks, etc., of a neighborhood through the Foursquare API. From this data, we learn to generate clusters, each grouping similar neighborhoods.
- Problem:
  - Is it possible to associate the number of police incidents to neighborhoods with similar venues categories?
- Interests:
  - People who are choosing a place to live in San Francisco could make a tradeoff between neighborhoods and choose a neighborhood that has many locations such as shops, restaurants and bars, in other words, having more facilities nearby, but with more police occurrences, or a neighborhood more residential and with fewer police occurrences.

Data

- Datasets:
  - Dataset of police events in the city of San Francisco, divided among the various police districts, which occurred during the year 2016.
  - A geojson file from the city of San Francisco
  - Table with the neighborhoods and their zipcodes
  - To get the most common venues of given neighborhood of San Francisco was used the Foursquare API.

Data

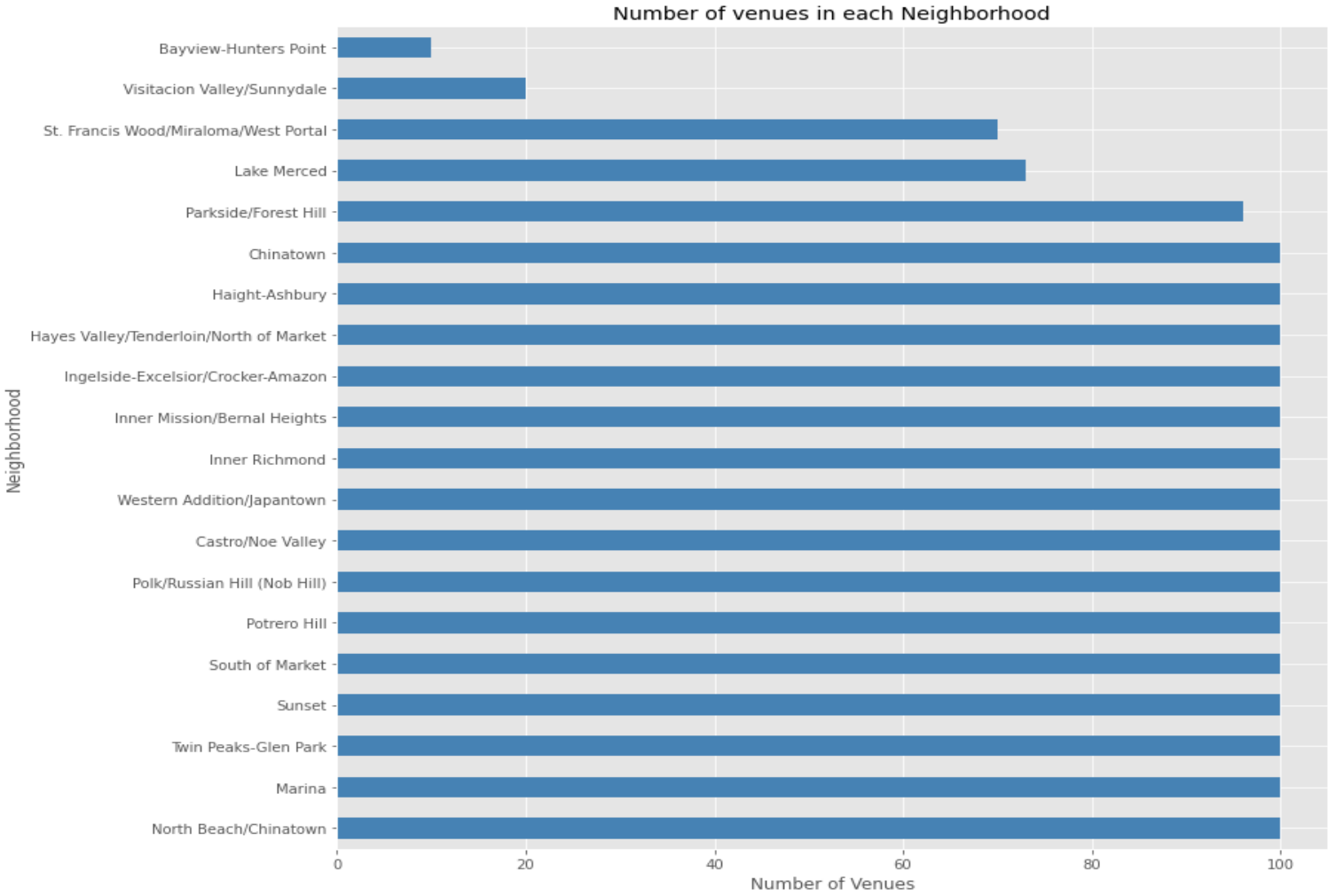
- Data Cleaning:



	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

Methodology

- Neighborhood Analysis

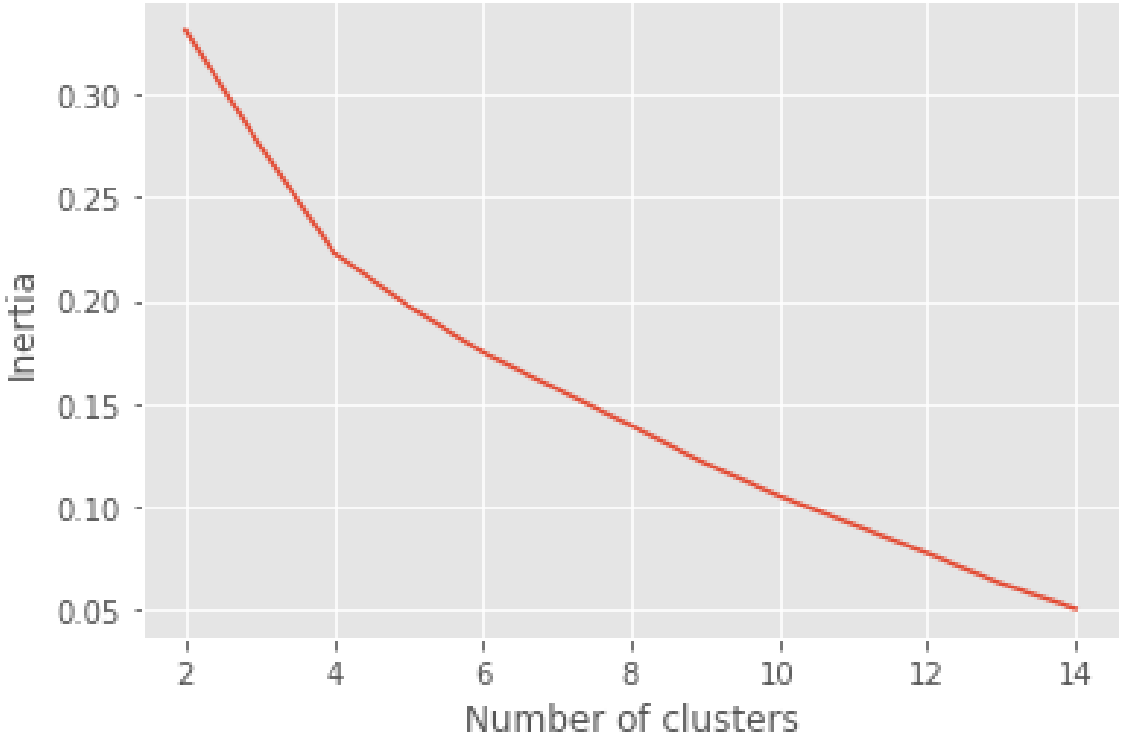


Methodology

- Clustering

  - Elbow Method to define the number of cluster: 4 clusters

  - The neighborhoods was divided into 4 clusters according to their top 10 venue categories. So, it was run K-means to cluster the neighborhood into 4 clusters. K-means will then partition our neighborhoods into 4 groups, where the neighborhoods in each cluster are similar to each other in terms of the top 10 venues categories included in the dataset.



Results

- Examining Clusters

  - Cluster 0 (red in the map below) consists of one neighborhood and the most common venue's categories are related to outside recreation.

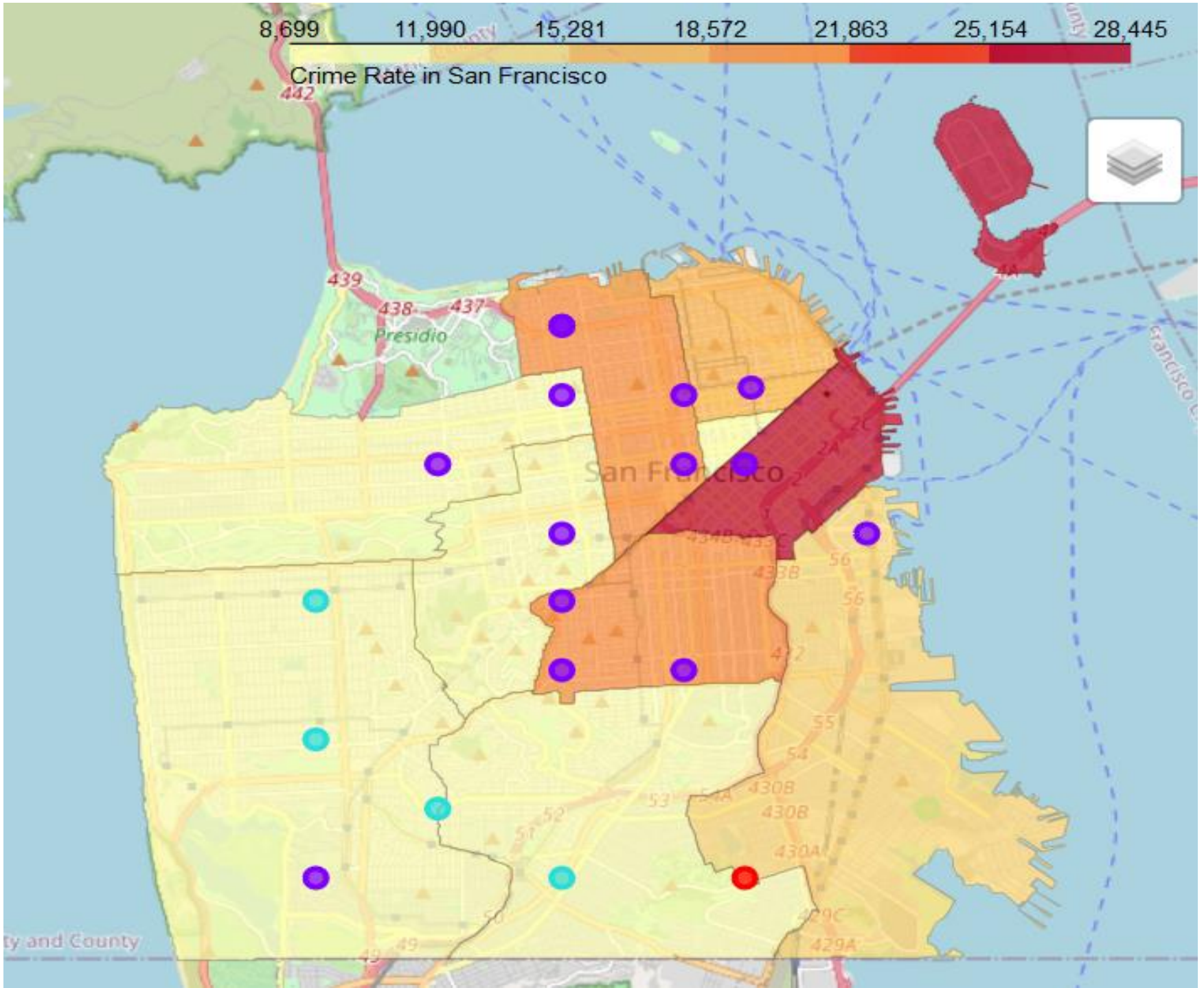
  - Cluster 1 (purple) consists of 14 neighborhoods and the most common venue's categories are related to different types of bars, restaurants and general stores.

  - Cluster 2 (cyan) consists of 4 neighborhoods and this cluster is formed by different kinds of restaurants.

  - Cluster 3 (green) consists of 1 neighborhood and this cluster a mixed venue's categories, from park and gym to coffee shop and brewery.

Results

- Map
  - Cluster 0 (red).
  - Cluster 1 (purple).
  - Cluster 2 (cyan).
  - Cluster 3 (green).



Discussion

- As we can see on the map, the police districts with the highest number of police incidents are located mainly in the neighborhoods of cluster 1 (purple), mostly composed of venue’s categories of bars, restaurants and shops in general, that is, districts with greater number of people both day and night.

- On the other hand, neighborhood clusters number 0 (red), 2 (cyan) and 3 (green), are located mainly in districts with fewer police incidents, perhaps because of the neighborhood's characteristic there is less movement of people both during both day and night.

- Only from this map it is not possible to determine that there is a cause and effect between the type of neighborhood and the number of police incidents, however it indicates that there may be this type of relationship between the two variables.

Conclusion

- There may be a relationship between the number of police incidents and the type of neighborhood in a neighborhood. However, this project is not intended to be a definitive analysis on this subject, it only shows a perception when these two datasets are compared on the same map.

- Others can deepen this study in the city of San Francisco or even take this project to other cities where this data can be easily available.