An investigation into the correlation between crime rates and police station presence in San Francisco

Coursera Capstone Project – by Theresa Hughes

4th September 2021

Who is interested?

- Police Department of San Francisco:
    - Should we set up a new police station to tackle crime (will it work)?
    - Where is there a concentration of crime? We should open a new police stations here.

- Prospective buyers/renters:
    - What area should we move to? Which is safest, and which should be avoided.
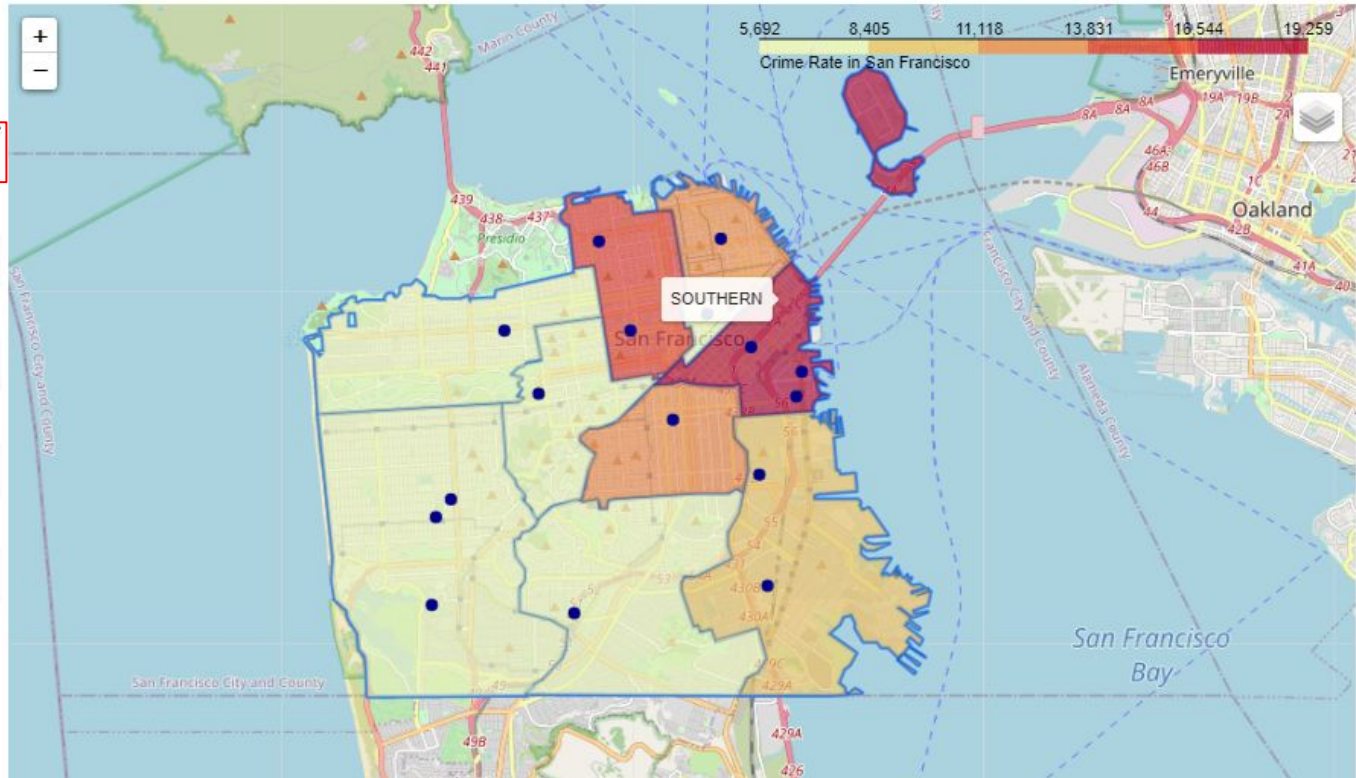
Data

- San Francisco Crime Data
  - Source: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DV0101EN-SkillsNetwork/Data%20Files/Police_Department_Incidents_-_Previous_Year__2016_.csv
  - Raw data contains 150,500 observations, 13 fields
  - Cleansed data (dropping non-criminal incidents and unnecessary fields) contains: 100,517 observations, 7 fields
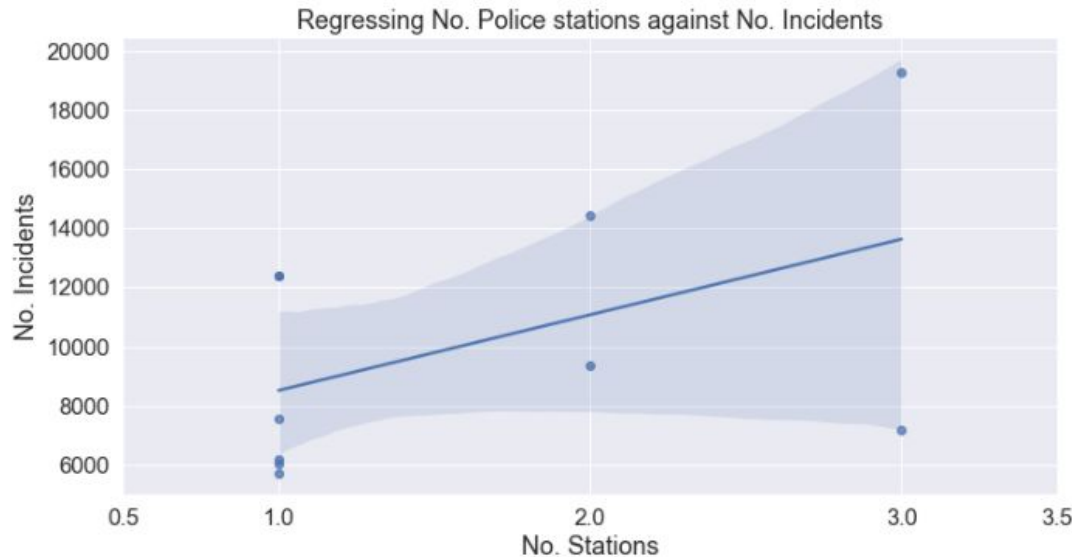
- Foursquare Police Station Data
  - Source: https://api.foursquare.com (using relevant GET request)
  - Extracted data contains 30 observations, 8 fields
  - Cleansed data (drop police stations outside of San Francisco, non-functional police stations, and unnecessary fields) contains: 16 observations, 5 fields

Which district of San Francisco is least safe?

	PdDistrict	No. Incidents
0	SOUTHERN	19258
1	NORTHERN	14425
2	CENTRAL	12404
3	MISSION	12394
4	BAYVIEW	9347
5	INGLESIDE	7578
6	TARAVAL	7187
7	TENDERLOIN	6190
8	RICHMOND	6041
9	PARK	5692

Crime Rate in San Francisco

5,692   8,405   11,118   13,831   16,544   19,259

SOUTHERN

Regression modelling for correlation between number of stations & number of crimes?

Regressing No. Police stations against No. Incidents

R-squared: 23.25%

No clear correlation between the number of police stations and number of criminal incidents. A very weak positive relationship - model is not very predictive

Limitations:
- Very small range of values of dependent variables (largest number of stations is 3, smallest number is 1)
- Crime rates not scaled per unit of area or per share of population

Key stats of the analysis…

19,258

Number of crimes reported in the Southern District of San Francisco in 2016

5,692

Number of crimes reported in the Park District of San Francisco in 2016

23.25%

R-Squared of the regression model between police stations and crime rates in San Francisco

Conclusions

- Southern district is least safe with highest crime rate.
- Park district is the safest area with lowest crime rate.
- No clear correlation between number of police stations and number of crimes.
- There is a very weak positive correlation between the 2 variables (unexpected).
- Western San Francisco is safer than east San Francisco (less crime).

Recommendations

- San Francisco PD should open new stations in Southern district to tackle high crime rate.
- No guarantee that an additional station will rescue crime (based on weak correlation).
- Prospective buyers/movers should look at properties in west San Francisco.
- Prospective buyers/movers should consider other factors (facilities, average salaries, etc.) that may offset the negatives of high crime rates in eastern districts.