Family-Friendly Neighborhoods in San Francisco, USA

(IBM Applied Data Science Capstone Project)

1. Introduction

1.1 Problem

Settling in a new city is a very arduous and sometimes an extremely complicated task. A person coming to a new city might not be familiar with the amenities around the city and as to which neighborhoods are safe for them and their families.

While looking for a family-friendly neighborhood one has to take into accounts things such as whether the neighborhood is safe or not and what kind of amenities are available around the neighborhood such as – parks, libraries, etc.

This project aims to help determine family friendly neighborhoods in the city of San Francisco, USA. San Francisco is a huge city with a population of over 3 million people and as such it can be quite difficult to find family-friendly neighborhoods in the city. This project uses location data and machine learning to determine which neighborhoods in the city have low crime rates and good basic amenities so that they are family-friendly.

1.2 Stakeholders

This project would be useful to families moving to San Francisco, Real Estate agents who sell and rent houses to such families and to Real Estate developers in determining which neighborhoods are more family-friendly and thus more lucrative to potential customers.

In addition, Police Forces and Law enforcement officials can also benefit from this project as it also determines neighborhoods which have a higher number of reported criminal incidents which in turn can help determine as to which neighborhoods need more resources in order to maintain law and order.

2. Data

2.1 Data Used

For this project the following data was used:

- Crime statistics of neighborhoods in San Francisco city
- List of public venues in the neighborhoods
- Number of such public venues
- Neighborhood boundary data to generate a choropleth map

2.2 Data Sources

Crime Statistics of San Francisco city were obtained from: https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

San Francisco Neighborhood Boundaries data was obtained from: https://data.sfgov.org/Geographic-Locations-and-Boundaries/Analysis-Neighborhoods/p5b7-5n3h

Data regarding public amenities in various neighborhoods was obtained from Foursquare using the ‘request’ library in Python

3. Methodology

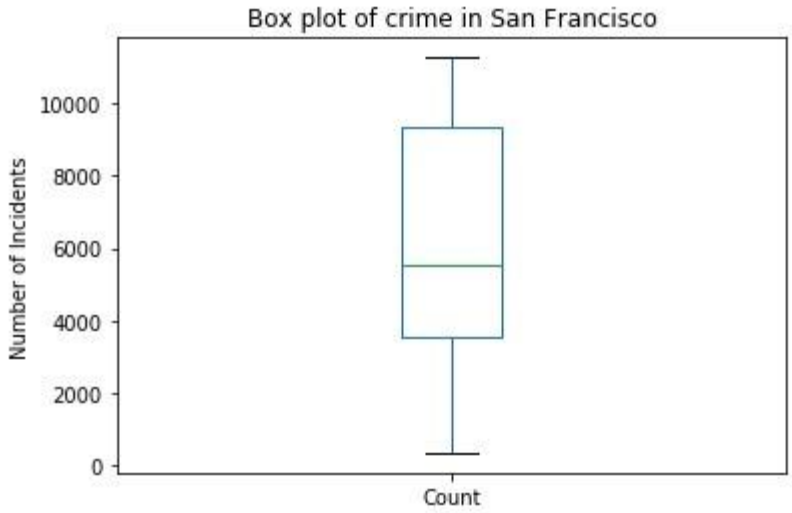
To identify which neighborhoods were the most family friendly K-Means clustering was used as it would allow clustering neighborhoods into 3 segments – Family-Friendly, Unsafe, Dangerous.

3.1 Data Processing

- The criminal statistics data was obtained from the source in form of a '.csv'('Comma Separated Values') file. This data was loaded into a Pandas Dataframe in Python using the 'pandas' library. The unnecessary fields were dropped from the data and the data was merged according to the neighborhoods so that the total number of reported incidents could be obtained. The final data looked as shown below

	Neighborhood	Latitude	Longitude	Count
0	Bayview Hunters Point	37.732943	-122.390954	20046
1	Bernal Heights	37.741410	-122.416379	5972
2	Castro/Upper Market	37.763186	-122.432706	10353
3	Chinatown	37.796527	-122.407235	6767
4	Excelsior	37.720130	-122.433183	5476

- The data was then plotted in form of a box plot as shown below

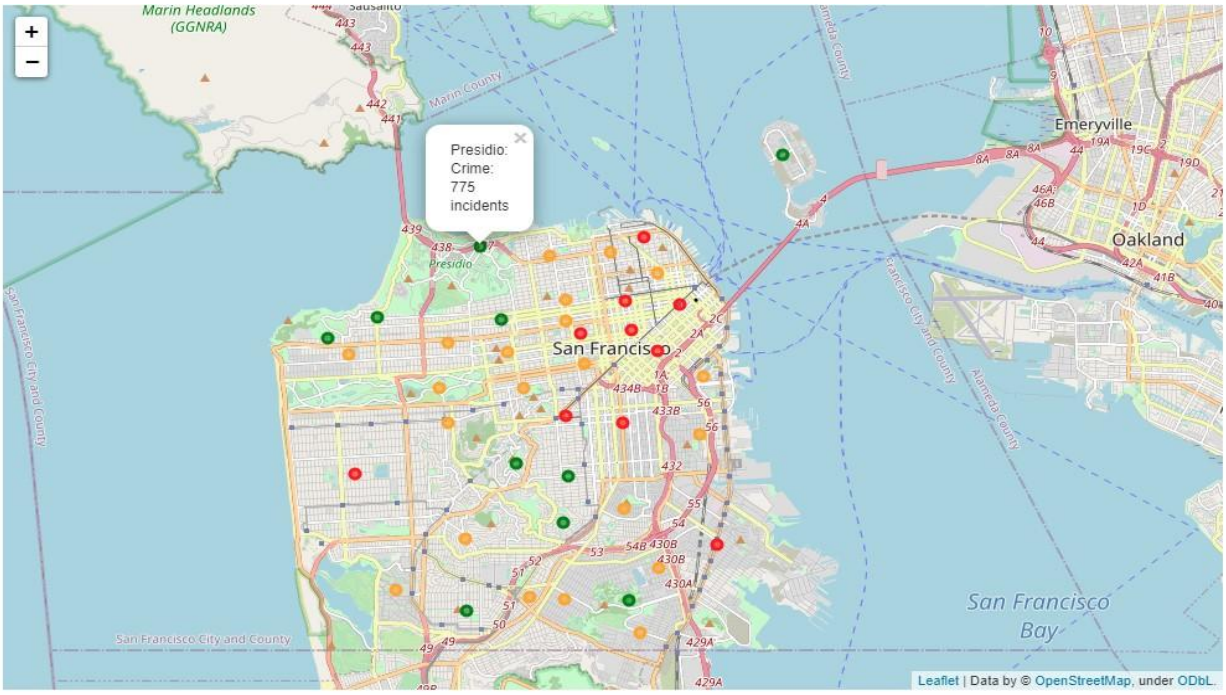


and it was determined that neighborhoods whose number of criminal incidents fell in the interquartile range of the boxplot were average neighborhoods, neighborhoods with less reported incidents than in the interquartile range were considered safe neighborhoods and ones above the interquartile range were considered dangerous neighborhoods. Below is a detailed description of the data:

	Latitude	Longitude	Count
count	41.000000	41.000000	41.000000
mean	37.765557	-122.435187	8252.951220
std	0.028849	0.030042	9310.265767
min	37.712351	-122.498417	349.000000
25%	37.741410	-122.456210	3553.000000
50%	37.772392	-122.432706	5521.000000
75%	37.785546	-122.414493	9350.000000
max	37.824155	-122.372732	38921.000000

- The neighborhoods were then plotted on a map of San Francisco generated using ‘Folium’ library in Python and color coded according to the number of crime reported in that neighborhood as shown below.

  (Green:Low Orange:Average Red:Dangerous)



- Number of parks, libraries and playgrounds in a Neighborhood were then obtained using Foursquare API and ‘request’ library in Python and this data was added into the criminal dataframe and the resulting dataframe was as shown below

	Neighborhood	Latitude	Longitude	Count	Park	Libraries	Playgrounds
0	Bayview Hunters Point	37.732943	-122.390954	20046	8	2	5
1	Bernal Heights	37.741410	-122.416379	5972	17	2	12
2	Castro/Upper Market	37.763186	-122.432706	10353	42	5	22
3	Chinatown	37.796527	-122.407235	6767	45	20	18
4	Excelsior	37.720130	-122.433183	5476	8	3	6

3.2 Clustering

The data obtained was then clustered using KMeans clustering method and the neighborhoods were clustered into 3 segments – Family-Friendly, Unsafe and Dangerous. The cluster labels were then added into the dataframe as shown below

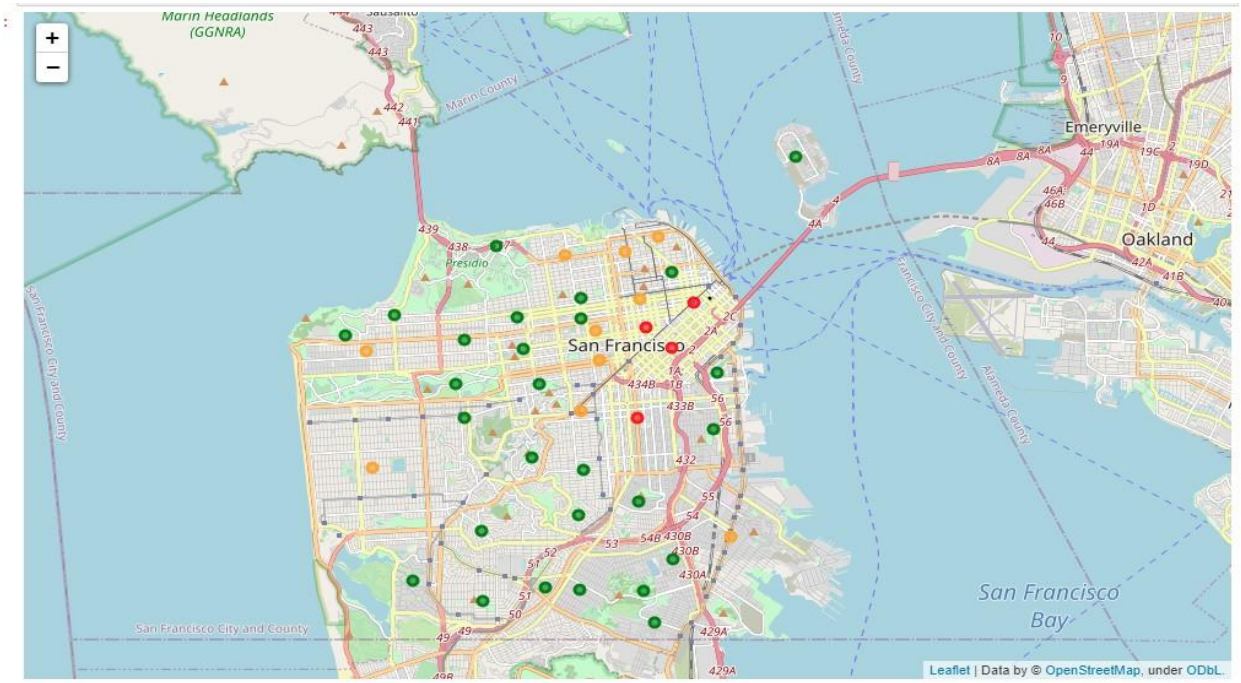
	Cluster_Labels	Neighborhood	Latitude	Longitude	Count	Park	Libraries	Playgrounds
0	2	Bayview Hunters Point	37.732943	-122.390954	20046	8	2	5
1	1	Bernal Heights	37.741410	-122.416379	5972	17	2	12
2	2	Castro/Upper Market	37.763186	-122.432706	10353	42	5	22
3	1	Chinatown	37.796527	-122.407235	6767	45	20	18
4	1	Excelsior	37.720130	-122.433183	5476	8	3	6

3.3 Visualizing

The data was then visualized in forms of graphs and maps using ‘Matplotlib’ and ‘Folium’ libraries of Python

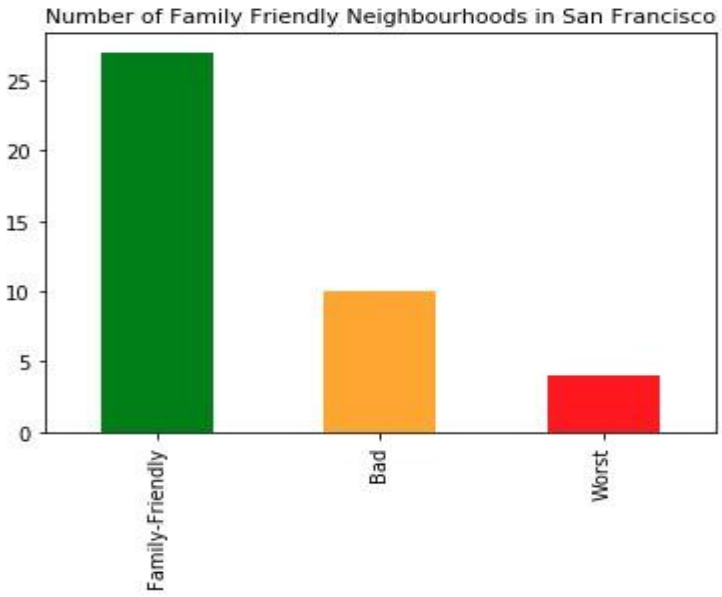
4. Results

The clustered data was plotted on a map representing different neighborhoods.



As seen in the above map there are a lot more green dots as compared to red and orange dots; this represents that there are a lot of family friendly neighborhoods around the San Francisco City.

This can also be seen in the bar graph below



It was found that out of 41 neighborhoods in San Francisco city, 27 were determined to be Family-Friendly by our machine learning model; 10 neighborhoods were considered unsafe and 4 neighborhoods were classified as Dangerous.

5. Discussion

Since the number of family-friendly neighborhoods is much more than the number of Unsafe and Dangerous neighborhoods it should be easy for families to find the perfect neighborhood as suited to their needs. Following is a detailed discussion regarding different stakeholders

5.1 Families

Families looking for safe places with easy access to amenities should have no problem in finding a neighborhood as a lot of neighborhoods are deemed family-friendly and these neighborhoods are scattered across the city instead of being in one side of the city. The following neighborhoods should be considered by families:

- Japantown
- Golden Gate Park
- Glen Park

These neighborhoods don’t have the lowest crime but they have a good number of amenities which will be a plus point for kids.

5.2 Developers

It should be noted that neighborhoods with least number of criminal incidents also have a lower number of amenities. Developers can focus on development projects in these neighborhoods as these can be quite lucrative for new families.
Following is a list of Neighborhoods with lowest crime rates and low number of amenities:

- Seacliff
- Lincoln Park
- McLaren Park

5.3 Law Enforcement

Law enforcement should note that some neighborhoods in the city have very high reports of crime. City council can allocate more funds to police districts in such neighborhoods.
Neighborhoods with the highest number of crime in the city are

- Financial District
- Tenderloin
- Mission
- South of Market

Following is a choropleth map showing which areas are family friendly and which are not.

(Blue:Family-Friendly Green:Unsafe White:Dangerous)

5.4 Improvement Ideas

- Data of school performance in each neighborhood can be added to determine which neighborhoods have better schools.
- More amenities can be added and analyzed based on user preference and what amenities they most likely prefer in a neighborhood
- Crime data can be standardized as the neighborhoods with higher crime number are also more densely populated.
- House pricing data can also be added to determine which neighborhood will be best for different pay groups.

6. Conclusion

After processing crime data it was determined which neighborhoods offer the most amenities for families and then used clustering to determine which neighborhoods in the San Francisco are the most family-friendly and which ones should be avoided.

It was found that 27 neighborhoods in the city were family friendly and thus relevant stakeholders should have no problem in finding a neighborhood that suits their needs.

It was also found that some neighborhoods with low crime rates also have less number of amenities which some stakeholders can benefit from.