

San Francisco Crime Analysis: 2003-2017



By: Nathan Nguyen

Table of Contents

Introduction	3
Purpose.....	3
Data Collection	3
Data Wrangling	4
Exploratory Data Analysis.....	5
How has crime change from 2003 to 2017?	5
Is crime really increasing in 2011 to 2016?.....	7

Table of Figures

Figure 1: Approach Flow Diagram	3
Figure 2: Dataset Columns	4
Figure 3: 2003 - 2017 Crime Rate Heat map	6
Figure 4: Crime Rate vs. Year.....	7

Introduction

San Francisco is located in the northern part of California with a population of 870,000 people and covering a land area of roughly about 46.87 square miles. SF is filled with beautiful attractions such as the Golden Gate Bridge, Alcatraz, and even the Golden Gate park. This beautiful city is surrounding by water and is blessed with the beautiful California weather, thus making San Francisco a top choice for tourism and a place you can call home. However, not everything is always as it seems. There are areas of San Francisco that is infested with crimes of all types of nature, which leads us to our purpose of this study.

Purpose

The main purpose of this study is to get a deeper understanding of crimes located in different districts of San Francisco. We want to be able to answers questions such as when and where do crimes occur? What types of crime do residence of San Francisco experience depending on a certain area? We want to see if there's any trends or triggers that cause crimes. This is all valuable information to help spread awareness. Not only will residence of San Francisco be more knowledgeable about the types of crimes that's occurring out there, but this study will ultimately also give better insight for the local SF Police Department. Knowing our purpose and target audience, we need to take the right approach in order to successfully complete the project.



Figure 1: Approach Flow Diagram

Data Collection

Thanks to the help of open data sources such as DataSF, we are able to find valuable information that will help contribute to this study. Combining all sources of information, we are able to accumulate a dataset consisting of over two million rows spanning across 13 columns. The datasets can be found listed below:

- DataSF: Police Department Incidents 2017:
 - <https://data.sfgov.org/Public-Safety/Police-Department-Incidents-Current-Year-2017-/9v2m-8wqu>
- DataSF: Police Department Incidents 2016:
 - <https://data.sfgov.org/Public-Safety/Police-Department-Incidents-Previous-Year-2016-/ritf-b9ki>
- DataSF: Police Department Incidents 2015-2003:
 - <https://data.sfgov.org/Public-Safety/Police-Department-Incidents/tmnf-yvry/data>

However, not all the data was easily accessible. Some data needed to be scraped in order to obtain the information. Web scrape data is listed below:

- Statistical Atlas:
 - <https://statisticalatlas.com/neighborhood/California/San-Francisco/Mission/Population#figure/neighborhood-in-san-francisco/total-population>

Data Wrangling

After collecting and compiling down all of our datasets and data source into one repository, the data needs to be examined and cleaned. Knowing that our dataset consists of over 2 million rows, we use pandas to locate missing information from the dataset.

```

IncidntNum    False
Category      False
Descript      False
DayOfWeek     False
Date          False
Time          False
PdDistrict    True
Resolution    False
Address       False
X             False
Y             False
Location      False
PdId          False
dtype: bool

```

Figure 2: Dataset Columns

We can see that our data has missing information in the PdDistrict column and that our dataset isn't perfect to start with. With the use of other building functions, the data was

successfully cleaned and rendered down to a total of 2,123,167 rows and 13 columns and spanning over the years of 2003 to 2017. Now that our data is clean, it's ready to be put into good use, which leads us to the start of our exploratory data analysis.

Exploratory Data Analysis

How has crime change from 2003 to 2017?

Since our dataset is based on incidents of crimes reported throughout San Francisco, a good way to help visualize how the crime changed from 2003 to 2017 is to use a heat map.

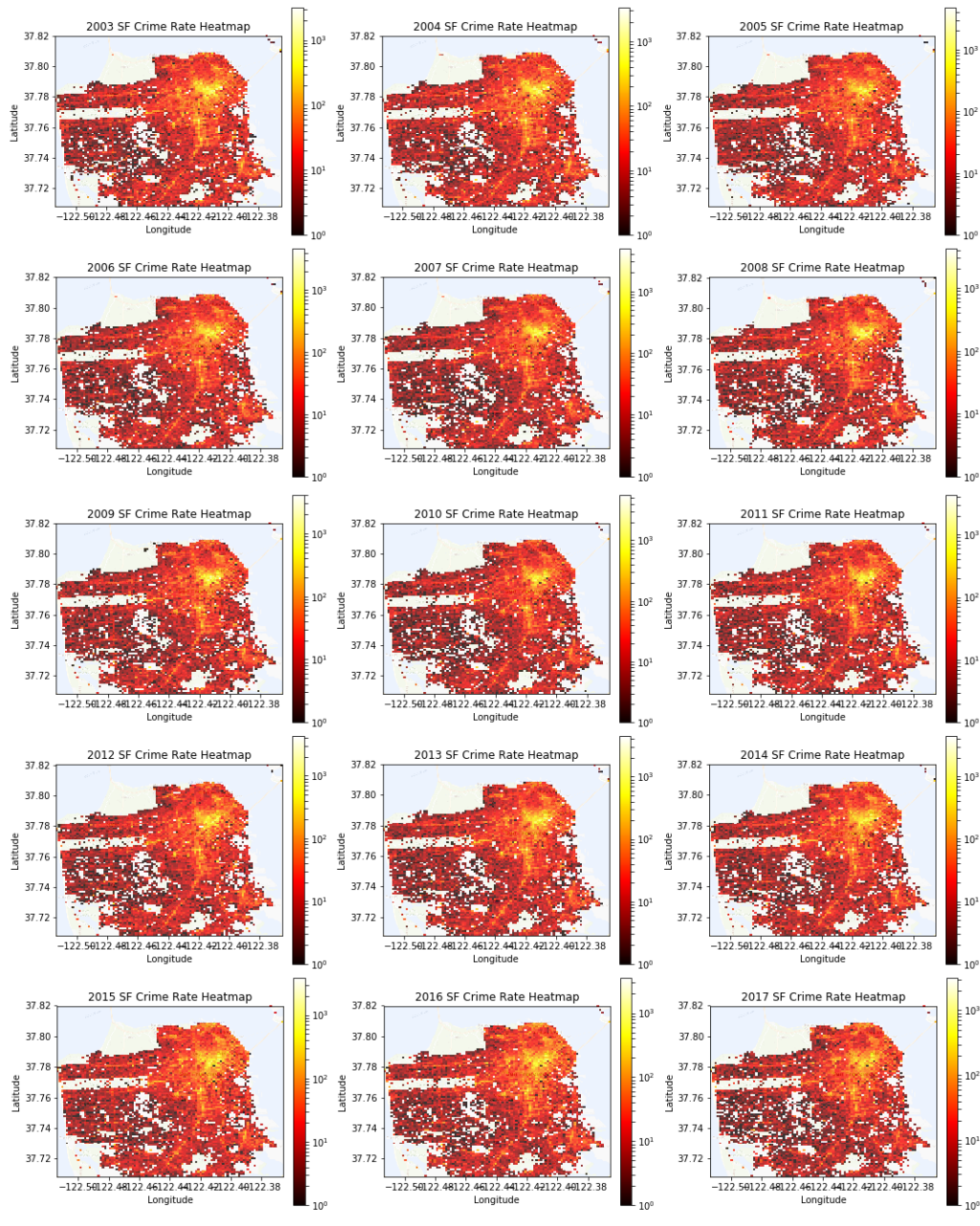


Figure 3: 2003 - 2017 Crime Rate Heat map

* Crime rate data for 2017 is not fully completed from Jan-Dec

There is clearly a reoccurring trend. There is a major hot spot around the Tenderloin area and the hot spot is consistent for all 15 years. We also see that there is a high crime rate running along the Bart station that leads use from South San Francisco into the heart of SF. Starting off in 2003, we can see that the hot spot is clearly having a wider spread. As we progress through the years this hot spot became tamer. However, starting in 2011, we can see that the crime rate starts to ramp up once again.

Is crime really increasing in 2011 to 2016?

Heat maps helps us see the bigger picture and identify trends and frequency of a crime occurring. How can we really tell if there indeed a decrease of crime rate from 2003 to 2011 and then an increase from 2011 to 2016? Simple, we can map crime

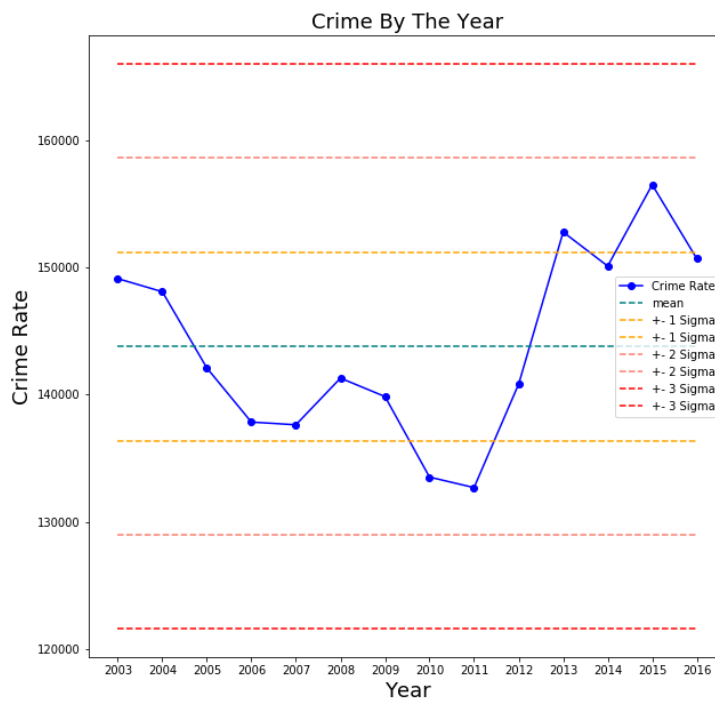


Figure 4: Crime Rate vs. Year