San Francisco Crime Analysis
By Jay Lam

Introduction

The crime rates in the Bay Area's largest cities have been increasing faster than the national average in the past few years. This research will be analyzing San Francisco's crime records from 2003 to 2015 and identifying the patterns of this city's crime history. This study will address the questions: Which area of San Francisco had the most and least crimes, what is the most committed crime in the city, and when do most crimes happen?

This analysis will alert us the trends of recent felonies and educate us on what to look out for around the Bay Area and in the city. It will inform us on the improvements needed in San Francisco to ensure a safer environment such as security, police enforcement, and other crime preventions. It will also teach us to be cautious around specific area and time period in the city. Furthermore, this research can potentially predicts values of estates in different neighborhood by knowing the crime rates in those areas.

The dataset used for this analysis is called San Francisco Crime Classification from Kaggle[1]. The data is obtained from SFPD Crime Incident Reporting System during 1/1/2003 to 5/13/2015 which is nearly 12 years of recorded felonies in San Francisco. The content of each data record contains:

- Dates timestamp of the crime incident
- Category - category of the crime incident
- Descript - detailed description of the crime incident
- DayOfWeek - the day of the week
- PdDistrict - name of the Police Department District
- Resolution - how the crime incident was resolved
- Address - the approximate street address of the crime incident
- X - Longitude
- Y - Latitude

Methodology

This analysis is performed by utilizing various Python libraries such as Pandas, Seaborn, and Matplotlib. The complete list of imports is shown in the image below.

```
import pandas as pd
from pandas import DataFrame
from datetime import datetime
from collections import Counter
import seaborn as sns
import matplotlib.pyplot as plt
```

[1] https://www.kaggle.com/kaggle/san-francisco-crime-classification

First, in order to process the data, we have to read the source csv file and save it as a dataframe using Pandas' read_csv function.

To find the amount of misdemeanors listed under each district, we have to find the value counts of the column "PdDistrict" and save it as a new dataframe. Thus, the new dataframe will have two columns: the district name and the count of each time the district appeared meaning a crime had happened in that district. Since the dataframe is already sorted by the value_counts() function, we can easily see which district has the most and the least crimes. The same procedure can be used on the "Descript" column to find the amount of times each type of crime were committed.

The datetime library was used to process the dates column and find the total occurrences of specific years, months, and time in the dataset. For every date in the "Dates" column, we apply the date value to a function that converts the value into a datetime object using datetime's strptime method with the format, "%Y-%m-%d %H:%M:%S". This will return a list of datetime objects which allows us to quickly access each value of the dates. For each datetime object, we append their attributes such as year, month, and hour to a new list respectively. From there, we can use the collection library's counter tool to count the amount of times each attributes appeared in the dataset and place each counter's items into a new dataframe. Ultimately, we will have dataframes with each date attributes and their sum of occurences for analysis.

Results

From examining the source data with 878,049 crime records and nine column of information, the answers to this research's core questions were revealed. Chart 1 shows that the Southern district has significantly more crimes than the rest of the districts. Compared to the district with the second most crimes, there is about a 31% increase of felonies in the Southern district. Additionally, Southern has about 110,000 more crimes over the past 12 years than Richmond which has nearly four times less. The Southern District is near the outskirts of downtown SF which is the most populated area in the city where there are more victims. On the other hand, Richmond has one of the lowest population density in the city which results in lesser crimes than the other districts. From looking at this chart, we can safely conclude that Richmond is the safest area and Southern is the most dangerous.

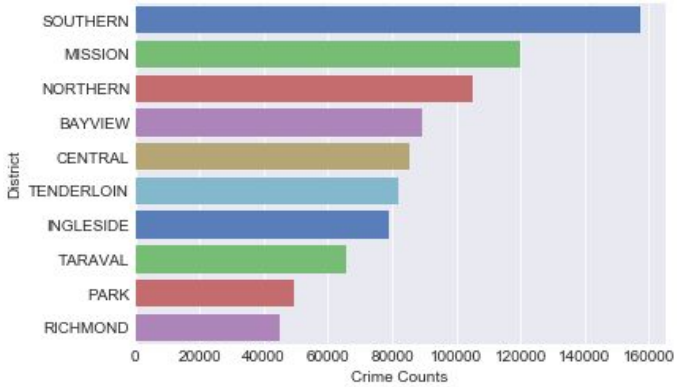

Chart 1: Crime rate of each San Francisco District

Furthermore, the most committed felony in San Francisco is grand theft from locked auto which is stealing items from a locked vehicle or the vehicle itself. We can assume that this is the cause of the city's high population density with many of the residents' cars parked on the streets instead of private areas such as a parking lot or a garage. From Chart 2, the most committed crime has almost double the amount than the runner-up, lost property and the rest of the felonies are fairly well-distributed.

Chart 2: Top ten most committed crimes in SF

This data was recorded from 1/1/2003 to 5/13/2015 which explains why the year 2015 is an outlier; 2015 was only recorded for less than half a year so there's less than 30,000 crimes recorded. From Graph 1, we see a steady decrease in crime rate from 2003 to 2007 and continues to decrease after 2008 until 2011 which was the lowest amount of crimes since the 1960s. From 2011 to 2013, the amount of felonies have been increasing every year by about 5-7% which paused after 2013. One possible explanation to this surge of crime rate is the transition from Major Gavin Newsom to Major Ed Lee. Ed Lee was elected in January of 2011 which correlates to this sudden increase of crimes in 2011 and onward.

Graph 1: Yearly crime occurrences from 2003 to 2015

According to the Graph 2, there is a dip in crimes during June to September which illustrates that there are less crimes around summer time. One assumption is that there is more police enforcements during the holidays which reduces the amount of criminal activities. Surprisingly, after the period of minimal misdemeanors, October is the month with the highest crime rates. Then the trend starts to plummet during November and December, making it the month with the least crime at 65,006 compared to September's 80,274 crimes. This furthers the assumption that there are less felonies committed during holiday periods.

Graph 2: Crime count of each month throughout 12 years

The graph below shows the amount of crimes committed throughout every hour of the day from 2003 to 2015. After midnight, we can see that there is a significant decrease in felonies due to the fact that people tend to be indoors at that time. At 5:00 AM, the amount of crimes is less than 10,000 which is the lowest throughout the day because people would most likely be asleep at that time. However, there is a gradual increase of crimes after 5:00 AM until a spike of almost 52,000 crimes at noon. After 12:00 PM is where most people would be outside of their home which

increases the amount of potential victims and crimes during that time. At 6PM, the crime rate peaks at 55,104 felonies during that hour and steadily decreases throughout the night. One explanation of this instance is that 6PM is in the midst of rush hour where everyone is either getting off work or going home.

Graph 3: Hourly crime occurrences throughout 12 years

Conclusion

This analysis on crimes from 2003 to 2015 in San Francisco has shown the amount of felonies in each district of the city, the top ten most committed crimes, and the year, month, and time where most crimes happened. We can conclude that, out of the ten districts in San Francisco, Southern has the most crimes and Richmond has the least due to the different in population density. Additionally, the most committed felony in the city is theft from parked vehicles because of the high amount of cars parked on the streets of San Francisco. Finally, 2013, October, and 6:00 PM were the time period with the highest amount of crimes committed respectively. This research will help us become more aware of the trends of criminal activities in the Bay Area to ensure a safer environment around us.