CS 235

Assignment #4
San Francisco Crime Data Visualization

Under the guidance of Prof. Ronald Mak

UX-Plorers:

Neha Rajkumar

Poonam Mehetrey

Poonkodi Ponnambalam

Shruti Sharma

Table of Contents

1. Purpose of the Application

The San Francisco Police Department is famous across the world and serves an estimated population of 1.2 million in the second most densely populated large city in North America. It is the 11th largest police department in the United States.

The high populations surely come with the adages of the increased number of Crime incidents which have plagued San Francisco for several decades.

- With this dashboard application, we intend to help the Chief of the San Francisco Police department and law enforcement officials visualize and device strategies to decrease the number of the crimes in the city of San Francisco.

- Present a statistical analysis to study the past crime statistics and present a picture for potential future residents for them to make informed choices about crime-free zones in the city.

- To predict the future crime activities based on past data.

2. Users of the Application

We have created the Dashboard for the Police Chief of San Francisco and the Law Enforcement officers who want to know the crimes happening at each station in San Francisco. This application helps the Police Chief to find out the quarterly updates, weekly updates and the hourly updates of the crimes happening in San Francisco for the year 2014. The Police Chief can keep track of the previously logged crime records in each station from the year 2007 to 2014. The real time updates helps the Police Chief to analyze the crimes happening per minute at each station.

3. Goals of the Application

The primary goal of the application is to help the Police Chief to keep San Francisco a safe place to live. We have five main goals:

Goal 1:

Description: The user, here the Police Chief can get an overall view of the number of crimes happening at each station in San Francisco.

Visualization Pattern Used:
Geo maps.

How did we achieve:
We chose a **Geo map** because these are a suitable way to find the concentration of crimes in a particular area. We have given a **Category** dropdown list and a **District** checklist near to the map to allow the user to search for a particular crime in any station. A **Monthly** slider is also provided to check for monthly updates.

Goal 2:

Description: The user can find out the quarterly, weekly and the hourly updates of the crimes happening at each station in San Francisco for the year 2014. This helps the Police Chief to devise strategies and deploy more force to stations with more number of reported crimes.

Visualization Pattern Used:
Geo maps with Bar graphs.

How did we achieve:
We chose a **bar graph** since these are the best way to show comparison of data. The user can make a comparative study on the quarterly, weekly and the hourly updates of the crimes happening at different stations in San Francisco.

Goal 3:

Description: The user can keep track of the history of crimes that happened in the previous years from 2007 to 2014.

Visualization Pattern Used:
Location maps and Heat maps

How did we achieve:
We chose a **heat map** since it conveys the number of crimes happening on a location basis. We gave varying color levels like **Green** for safer regions, **Yellow** for the lesser safer regions and

Red for the most unsafe regions. With the help of the **location map**, the user can get the street level view of a particular station on the map.

Goal 4:

Description: The user can make a comparative study on the types of crimes happening at San Francisco and the resolution taken for these crimes over the previous years from 2007 to 2014.

Visualization Pattern Used:
Bar graphs

How did we achieve:
We chose a **bar graph** since it conveys to the user a comparative study on the crimes that happened over the past years along with the resolution taken. A checklist with all the **Stations** was given so that the Police Chief can analyze on the **resolutions** taken for the crimes at a particular station.

Goal 5:

Description: The user can analyze the real time visualization of the crimes happening per minute at San Francisco.

Visualization Pattern Used:
Spline graph

How did we achieve:
We chose a **spline graph** since it conveys the trend of the crimes happening per minute as real time. Separate real time visualizations were created for **High Priority Crimes** and **Low Priority Crimes** happening per minute at San Francisco. The high priority crimes include murder, larceny/theft while the low priority crimes include loitering and trespass. The Police Chief can analyze the spline graph for predictive analysis as well.

4. Data Sources

We worked on the San Francisco Police Department Crime dataset. The dataset we chose was large and we started off with analyzing the data from 2003 to 2014. But later on we narrowed down the analysis to 2013 - 2014 (till mid November 2014) crime dataset which is about 300MB so that the Police Chief can analyze the current trend of crimes happening. The data file is in csv

format and it can be downloaded from the below link: *"https://data.sfgov.org/Public-Safety/SFPD-Reported-Incidents-2003-to-Present/dyj4-n68b"*

The dataset has 11 columns and the descriptions are given below:

Column name	Description
IncidntNum	Incident number for the crime
Category	Type of crime
Descript	Any additional details about the crime
DayOfWeek	Day, Crime occurred
Date	Date, crime occurred
Time	Time, the crime occurred
PdDistrict	Police district in San Francisco
Resolution	Action taken on the crime reported
Location	Street where the crime occurred
X	Geo coordinate
Y	Geo coordinate

Sample row from the dataset:

```
IncidntNum,Category,Descript,DayOfWeek,Date,Time,PdDistrict,Resolution,Location,X,Y
120000499,ROBBERY,"ROBBERY, BODILY FORCE",Sunday,01/01/2012,01:50,CENTRAL,JUVENILE BOOKED,VALLEJO ST / POWELL ST,-122.41045954645,37.7983714450285
120000938,NON-CRIMINAL,"DEATH REPORT, CAUSE UNKNOWN",Sunday,01/01/2012,06:40,NORTHERN,NONE,800.0 Block of OFARRELL ST,-122.418260926248,37.7851558590144
120001936,SUICIDE,SUICIDE BY JUMPING,Sunday,01/01/2012,14:45,NORTHERN,NONE,1200.0 Block of GOUGH ST,-122.424447811597,37.7843603833708
120002235,VEHICLE THEFT,STOLEN AUTOMOBILE,Sunday,01/01/2012,17:00,MISSION,NONE,1600.0 Block of BRYANT ST,-122.410532734085,37.765622799413
```

5. Working with Data and the Visualizations

We used Splunk Enterprise and Pig to query the data for analytics. Multiple queries were written around the data to analyze interesting trends from the dataset. Working with datasets from the year 2003 to 2014 helped us derive certain surprising results from the dataset. Graphs were plotted for these results to further help in predictive analysis.

Queries were run for the following and the corresponding visualizations were made:

1. **Analysis of Crimes over the Past Decade:**

   We found a dip in crimes at 2011 and then an increasing trend of crimes from 2011 to 2013. We chose a **bar graph** since it was the best way to show comparisons of crimes over a decade.
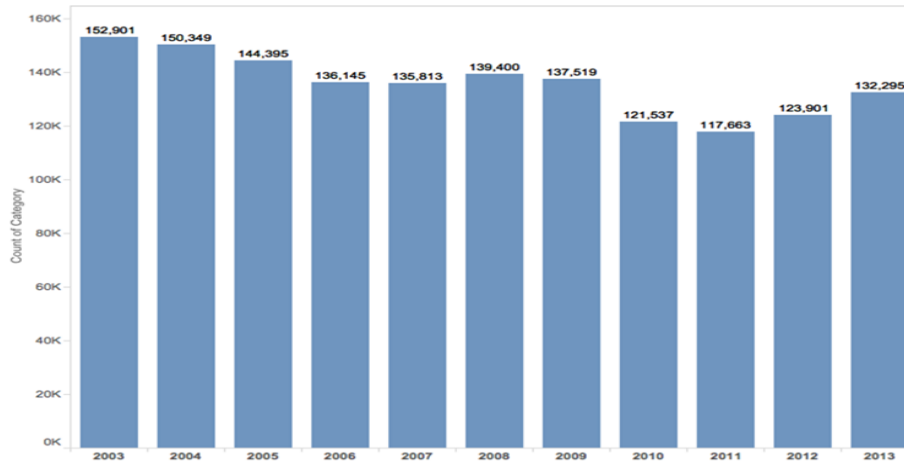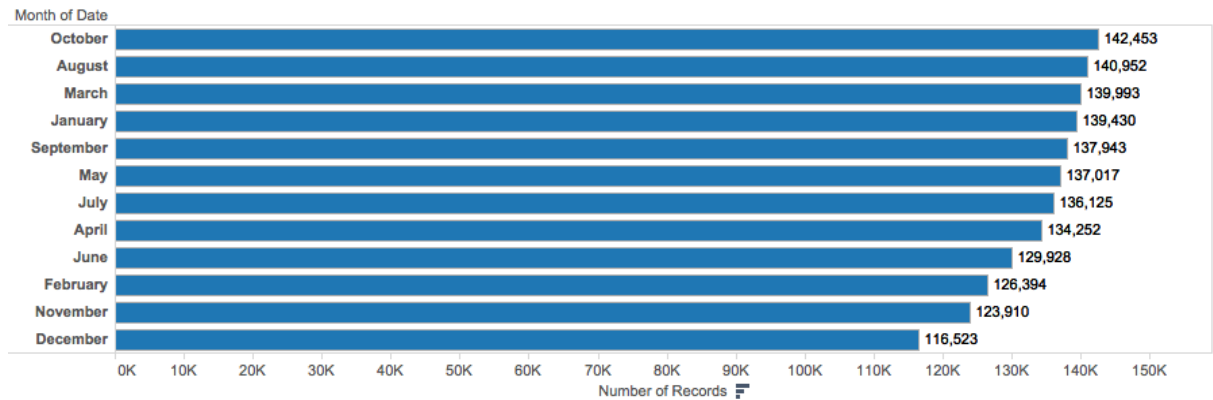
2. Major Crimes that have plagued San Francisco

We chose a **Word Cloud** to depict the major crimes happening in San Francisco. We found that larceny/theft and vehicular theft accounted for about 27% of the City's reported crimes.
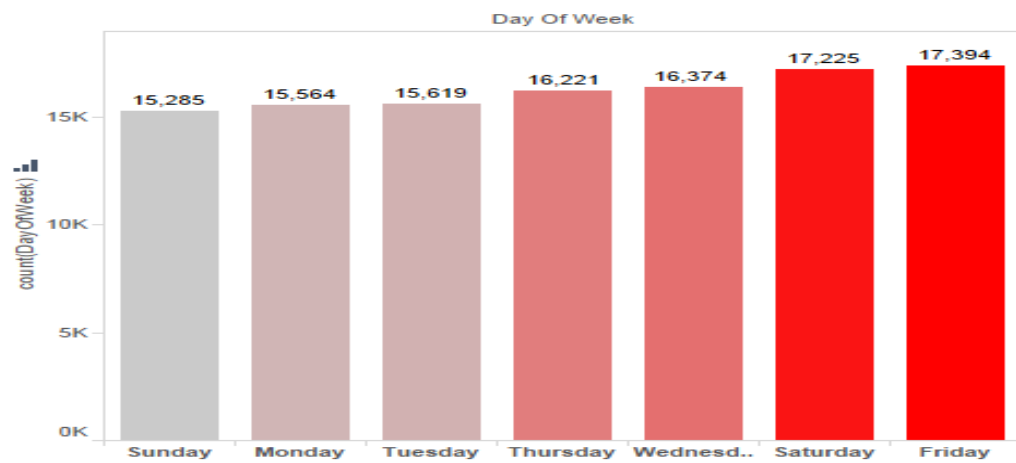
3. Month wise analysis of crimes

We chose a **bar graph** to analyze the comparison of crimes happening each month. We found that **October** had the maximum number of crimes reported.

Month of Date

	Number of Records
October	142,453
August	140,952
March	139,993
January	139,430
September	137,943
May	137,017
July	136,125
April	134,252
June	129,928
February	126,394
November	123,910
December	116,523

4. Weekly analysis of crimes

We chose a **bar graph** to analyze the comparison of crimes happening each week. We found that **Friday and Saturday** had the maximum number of crimes reported.

Day Of Week

	count
Sunday	15,285
Monday	15,564
Tuesday	15,619
Thursday	16,221
Wednesd..	16,374
Saturday	17,225
Friday	17,394

5. Hourly analysis of crimes

We chose a **line graph** to display the trend of crimes happening per hour. We found that the time intervals **12:00pm** and **4:00pm to 8:00pm** had the maximum number of crimes reported.

8
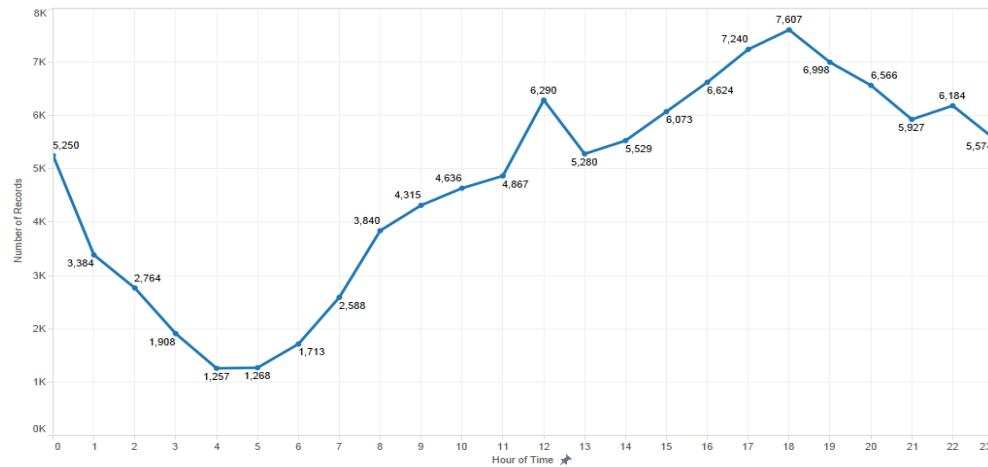
6. Crime resolution rates in 2013
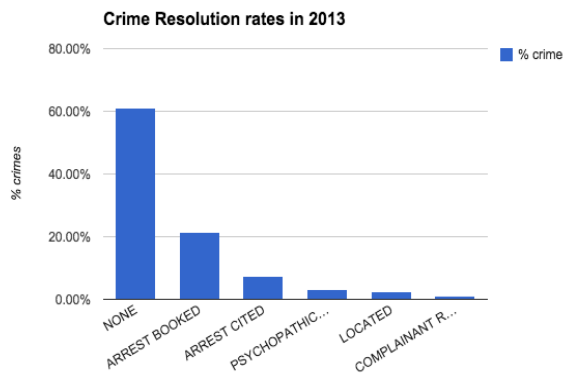
Resolution	% crime
NONE	61.02%
ARREST BOOKED	21.29%
ARREST CITED	7.47%
PSYCHOPATHIC CASE	3.26%
LOCATED	2.54%
COMPLAINANT REFUSES TO PROSECUTE	1.11%

Throughout the analysis we found that Bryant Street had the maximum number of crimes reported. We found that 800 block of Bryant Street had over 30,000 documented instances of crime in the last 10 years, which is equivalent to 1 in every 3 hours.

6. Tools used in creating the Visualizations

The tools we used for creating the visualization are Tableau, HighCharts and Google API.

Name of page	Tool used
Index.html	**Tableau public**
Maps.html	**Google API**
Page2.html	**HighCharts**

7. Dashboard

The main aim of our project was to create a dynamic and interactive dashboard which can help the Police Chief to view and analyze the crimes happening in San Francisco in a single shot.

Page 1:

Page 1, the "Dashboard page" shows **the crimes happening at different stations in San Francisco.** The quarterly, weekly and the hourly reports for the year 2014 are displayed in the form of bar graphs. The map and the bar graphs are dynamic and they change with respect to the selected station, crime category and the month selected by the user.

Page 2:

Page 2, the "History tab" **shows the history of crimes that happened in San Francisco** from the year 2007 to 2014(mid November). The heat map shows the color coding for the safest area to the least safe area. The location map allows the Police Chief to look at the street view of the area where the crime was reported. Bar graphs shows the resolution taken for the crimes reported over the previous years. The bar graphs are dynamic and they change according to the crimes reported.
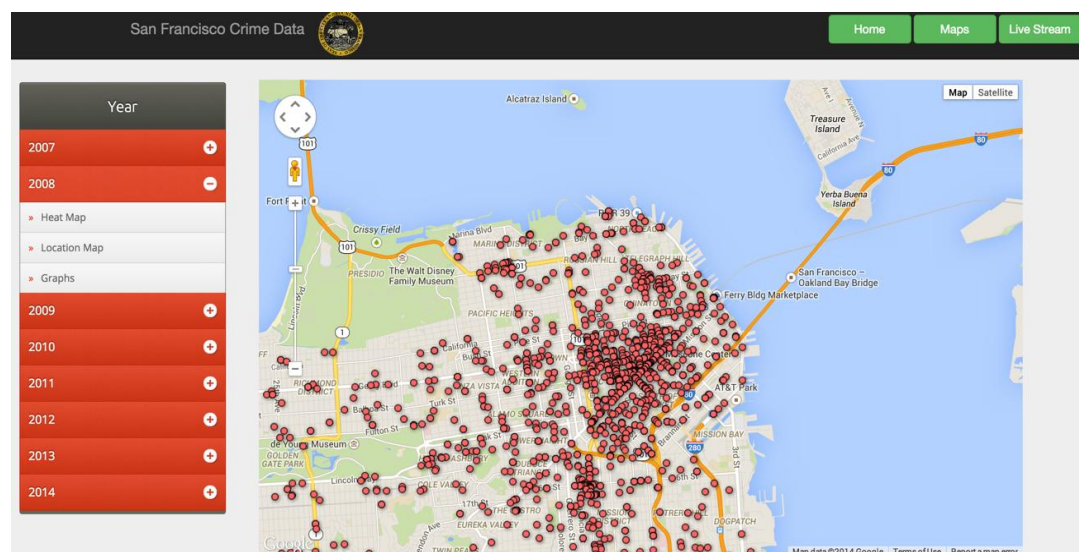
Page 3:

Page 3, the "Live Stream" shows the **real time visualization of the crimes happening** at San Francisco per minute. Real time graphs for the "high priority crimes" and the "low priority crimes" are provided.
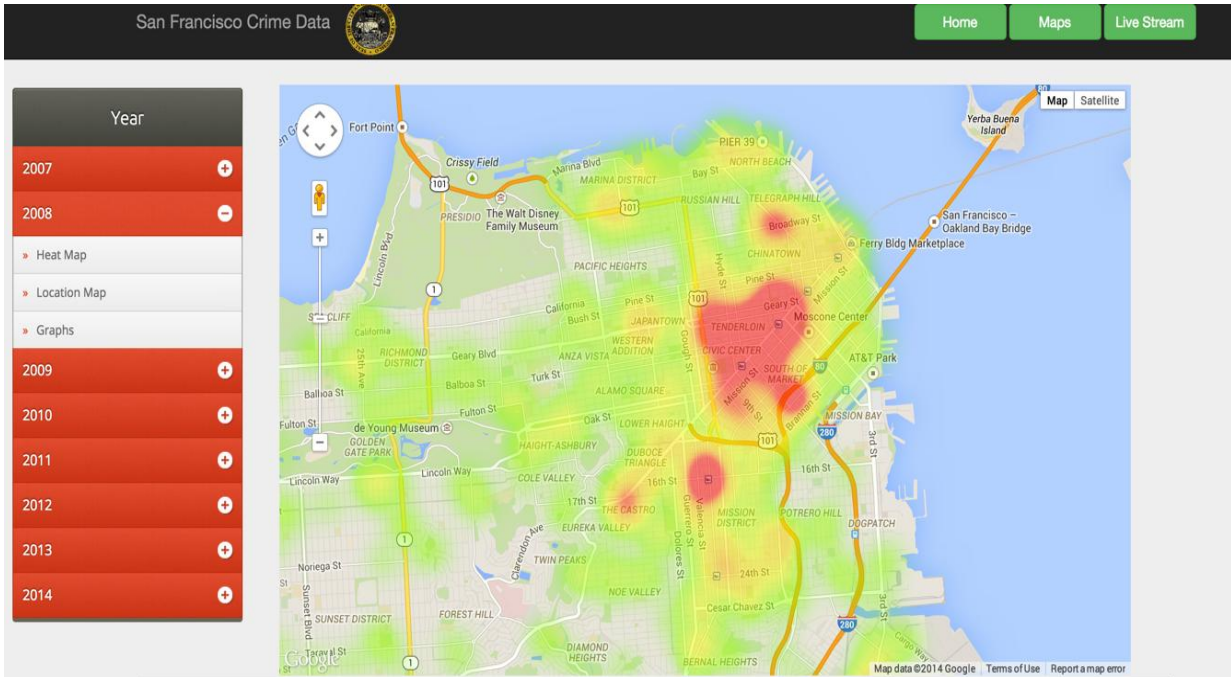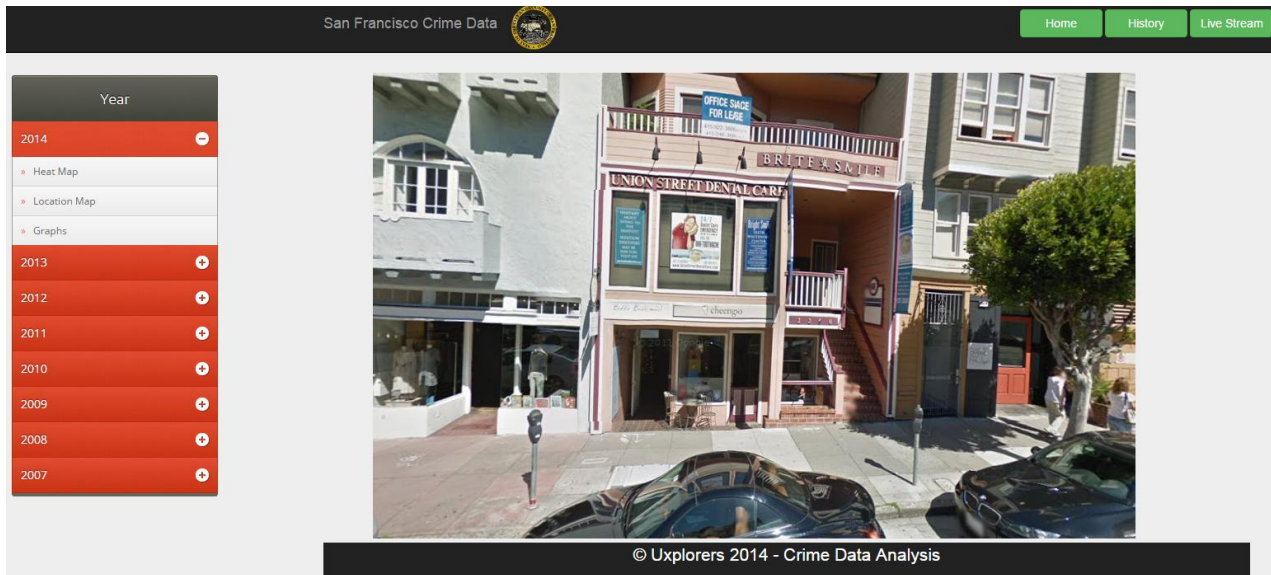
8. Screenshots

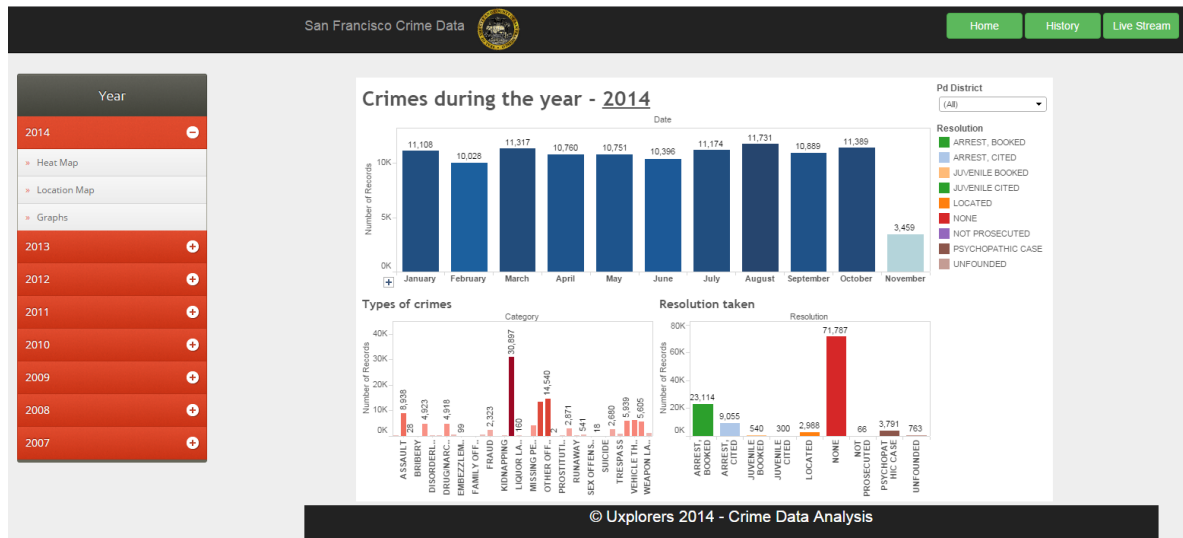## Screenshot 1: Dashboard
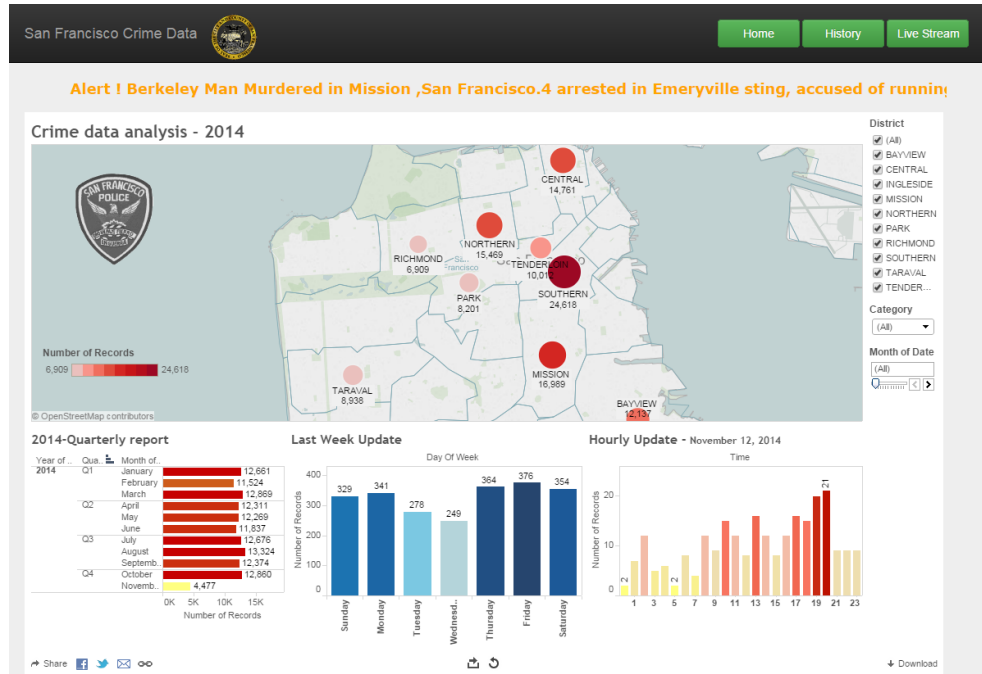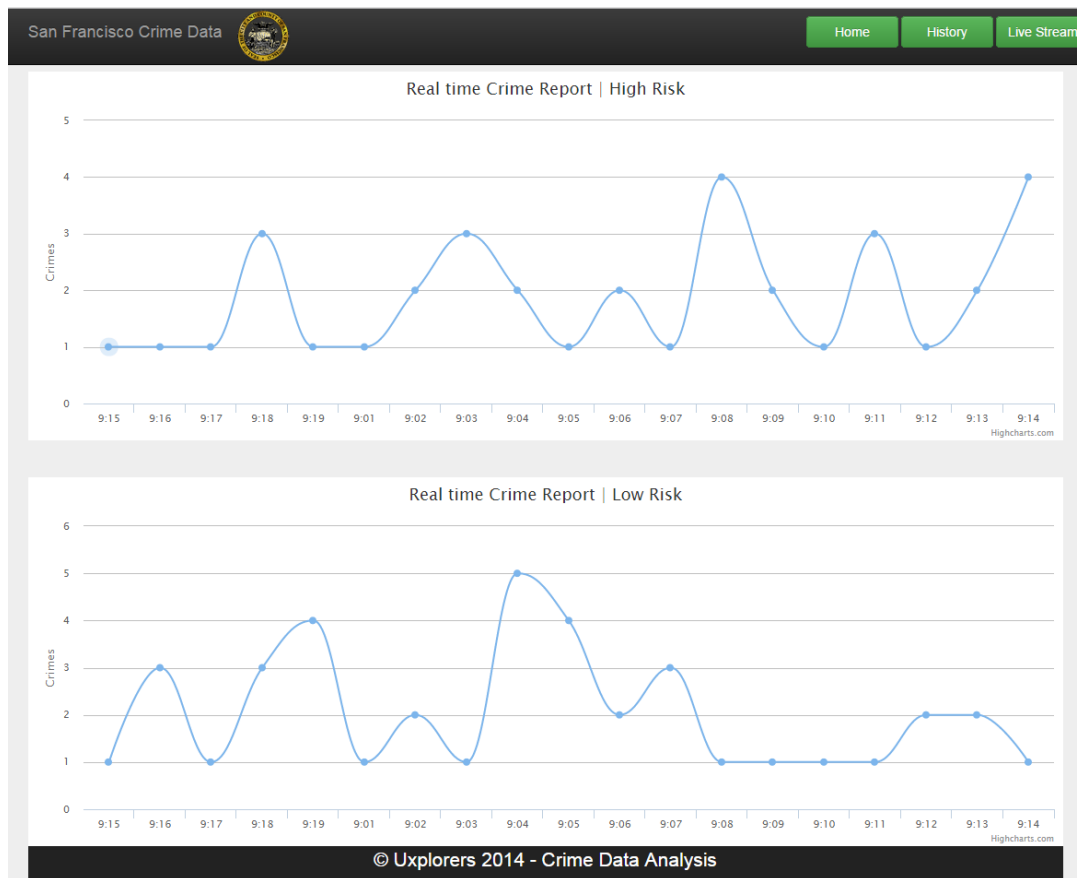

Screenshot 2: Map Screen

Screenshot 3: Heat Map

Screenshot 4: Street View

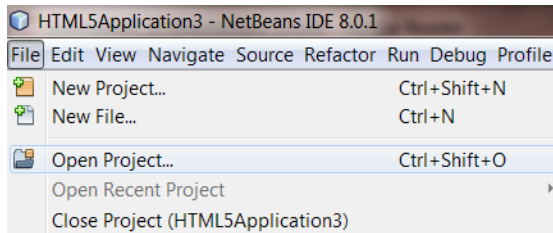**Screenshot 5: Resolutions taken**

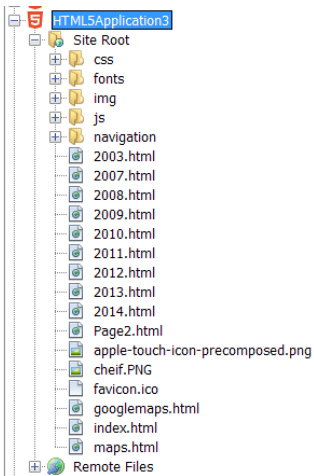**Screenshot 6: Live Simulation**

9. Running the Application

Tool Used: NetBeans IDE

1) Unzip the attached package and paste the folder "HTML5Application3" in NetBeans Workspace

2) Go to NetBeans IDE , click on File , Open Project

3) Open the Folder path from Step 1

4) Open "Index.html" and Run