Visualization and Prediction Of Property Crime Data

Nimmi CV, Priyanka Shah, Shruthi Raghunath and Veenu Agarwal

San Jose State University, SanJose, CA, USA

e-mails: nimmi.varakey@sjsu.edu, priyanka.shah@sjsu.edu, shruthi.raghunath@sjsu.edu, veenu.agarwal@yahoo.com

Abstract

The idea behind the project is to visualize and compare all kinds of criminal activities between any two selected cities in Santa Clara County. To do this, we collected data from authentic sources like FBI and other government websites and designed a model that could utilize these data sets to accomplish the goal.

The first step to design a model was to narrow down the geographical area we wanted to concentrate on. We chose our own local areas, that is, the cities that belonged to Santa Clara County, since the data patterns that we were getting suited these areas the best. The next step was to collect volumes of data from different sources, so that we have an accurate depiction of the crime status in all these areas.

After the collection of the data, we came up with a model to analyze the data we collected to find out relationships and patterns between the data. Liner regression model was used to visualize the relationship between the data collected. This data was then used for predictions for the years 2014 to 2018. We used R-tool for the statistical analysis of our data using a linear model. Google Motion charts was integrated into r-tool for a seamless visualization of both the data, and the predictions that were provided. The project was then hosted on to the cloud using Amazon EC2

I. Introduction

Crime has effect on all walks of life; naturally, it shapes a neighborhood and the area we live in. Therefore the knowledge of all kind of criminal activities including home invasion, property related crime and robbery is necessary to determine the safety of a community. Along with the knowledge of current situations in the cities, knowledge about the future is necessary to make important decisions related to property purchase, or even renting an apartment in a neighborhood. Property related crime like motor vehicle theft can also have an impact on insurance rates for that particular city. Having complete knowledge about criminal activities of any given city is not important for the safety of an individual, but helps in reducing the costs of living.

The crime statistics of a selected area should be comparable to other cities, at any given time. There is a pressing need for a system that can not only show this data in an understandable format, but also to predict the future of any chosen city for the coming years. This not only helps in important property purchase decisions, but may also help in taking precautions in keeping the community safe. The model that we propose can not only visualize different criminal activities like robbery, property theft, rape, and so on, but also can provide the prediction for these criminal activities in selected area or city. Comparison between any chosen crime between any two selected cities is also possible using our visualization model.

II. Data Collection

The data Collection for our models were the most tome consuming part of our project. The priority was to work on real and authentic data sets, to obtain a precise working model, which is as accurate as possible in both visualizing and predicting. Therefore a lot of time and energy was spent in collecting the data. Our sources were very authentic government websites, FBI websites, several local Police Department websites and the likes.

We started our web search for property related crime, home invasion and burglary data. For our project to be a success, the volume of the data we collected was very important. We needed the data at least for the last ten years (from the year 2000) to train our system. Data collection process began even before the geographical area for our project was narrowed down. We started collecting data for every city in California, and found that the data for the cities in Santa Clara County were accurate, consistent and continuous. We have covered the following Cities in the Santa Clara County for the version 1 of our project:

- Campbell
- Gilroy
- Los Altos
- Los Gatos
- Milpitas
- Morgan Hill
- Mountain View
- Palo Alto
- San Jose
- Santa Clara
- Sunnyvale

Although initially we started with the collection of robbery and home invasion data, we felt the need to incorporate all criminal activities, since that would provide a complete knowledge about the city chosen, and does not limit our system to property crime. We included different criminal activities like motor vehicle theft, burglary, robbery, rape, arson, aggravated assault, larceny and murder for the above specified areas.

After we narrowed down the geographical area and the type of criminal activities to include, the next step in data collection was to include the parameters that affected these criminal activities. We needed this data to study the criminal activity pattern, for a given area, and to analyze the factors affecting it. We narrowed down some common factors affecting the crime rate at any given point of time, like Unemployment rate, population, Law enforcement officials employed for the chosen area. Even for this data, our source was authentic websites and data sites like FBI websites, Police department websites (Santa Clara and San Jose Police Departments) and Labor Statistics government website.

After collecting all the necessary data, we then had to bring this data into a form which R-tool would accept. We pooled all the data into excel (.csv format) and fed into the linear model we designed to analyze these data. The criminal activities of any city can be compared against various affecting factors like unemployment, Population, Number of Law Officials, which all can be pulled from a single data source (the formatted excel sheet).

III. Technologies Used

1. **Google Visualization API (Motion Charts)** – This is a dynamic chart to explore several indicators over time. The chart is rendered within the browser using Flash. It lets us make a selection between bubble, bar and line charts.

2. **R Studio** – RStudio is an Integrated Development Environment (IDE) for R which works with the standard version of R available from CRAN. RStudio includes a wide range of productivity enhancing features. RStudio is an open source tool. It can be run on desktop (Windows, Mac, or Linux) or even over the web using RStudio Server. It can also be integrated with Git.

3. **Shiny** – R studio comes with a built in Shiny Package to enable hosting the project as a web application. Shiny makes it very simple for R users to turn analysis into interactive web applications that anyone can use. Shiny application is simply a directory containing a user-interface definition, a server script, and any additional data, scripts, or other resources required to support the application. Shiny provides Attractive default UI theme based on Twitter Bootstrap.

4. **Shiny Server** - Shiny Server is a server program that makes Shiny applications available over the web.

5. **Twitter Bootstrap** –Twitter Bootstrap is a free collection of tools for creating websites and web applications. It contains HTML and CSS-based design templates for typography, forms, buttons, charts, navigation and other interface components, and optional JavaScript extensions.

6. **Amazon EC2** – Amazon Elastic Compute Cloud is designed to make web-scale computing easier for developers. It is a web service that provides resizable compute capacity in the cloud. Amazon EC2's interface allows you to obtain and configure capacity with minimal friction. It provides users complete control of computing resources and allows running on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing user to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing user to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

7. **Node JS** – Node.js is a platform for building fast scalable network applications easily. Node.js uses an event-driven, non blocking I/O model that makes it lightweight and efficient, perfect for data-intensive real-time applications that run across distributed devices.

8. **Ubuntu** –Ubuntu is a complete desktop Linux operating system and freely available i.e. that software should be available free of charge. Ubuntu is suitable for both desktop and server use. Ubuntu includes more than 1000 pieces of software, starting with the Linux kernel version 3.2 and GNOME 3.4, and covering every standard desktop application from word processing and spreadsheet applications to internet access applications, web server software, email software, programming languages and tools and of course several games.

IV. Visualization

Visualization helps to represent abstract data to reinforce human cognition. The abstract data include both numerical and non-numerical data, such as text and geographic information.
Visualization allows users to see, explore, and understand large amounts of information at once. It focuses on the creation of approaches for conveying abstract information in intuitive ways. Data analysis is an indispensable part of all applied research and problem solving in industry.

Google Visualization API provides a platform that can be used to create, share and reuse visualizations written by the developer community at large, you can create reports and dashboards as well as analyze and display your data through the wealth of available visualization applications.

Google Visualization API establishes two conventions:

- A common interface to expose data on the web.
- A common interface to provide data to visualizations.

The syntax of the data format accepted by Google motion charts is as follows:

- **The first column** must be of type 'string' and contain the entity names (e.g., "City", "Location" etc).
- **The second column** must contain time values. Time can be expressed in any of the following formats:
    1. *Year* - Column type: '*number*'. Example: 2008.
    2. *Month, day and year* – Column type '*date*'
    3. *Week number*- Column type: '*string*'; values should use the pattern *YYYYWww*, example: '2008W03'.
    4. *Quarter* - Column type: '*string*'; the values should have the pattern *YYYYQq*, Example: '2008Q3'.
- **Subsequent columns** can be of type '*number*' or *'string'*. Number columns will show up in the dropdown menus for X, Y, Color and Size axes. String columns will only appear in the dropdown menu for Color.

googleVis Package-- `googleVis` is an R package providing an interface between R and the Google Chart Tools. The functions of the package allow the user to visualize data with the Google Chart Tools without uploading their data to Google.

The output of `googleVis` functions is html code that contains the data and references to `JavaScript` functions hosted by Google. To view the output a browser with Flash and Internet connection is required; the actual chart is rendered in the browser.

In order to get googleVis package working we need install *RJSONIO* package. After importing these two packages in the R interface, the Command *library (googleVis)* is used to make these packages available for use.

Plotting the sets of data which are available in the .csv format using the Google motion charts is simple. The command to accomplish this task is

gvisMotionChart(data, idvar = "id", timevar = "time")

The data can also be viewed on our website in a neat tabular format. The command *gvisTable(data)* is used to enable this feature. The data that we used for this model is also available for download, from the website directly.

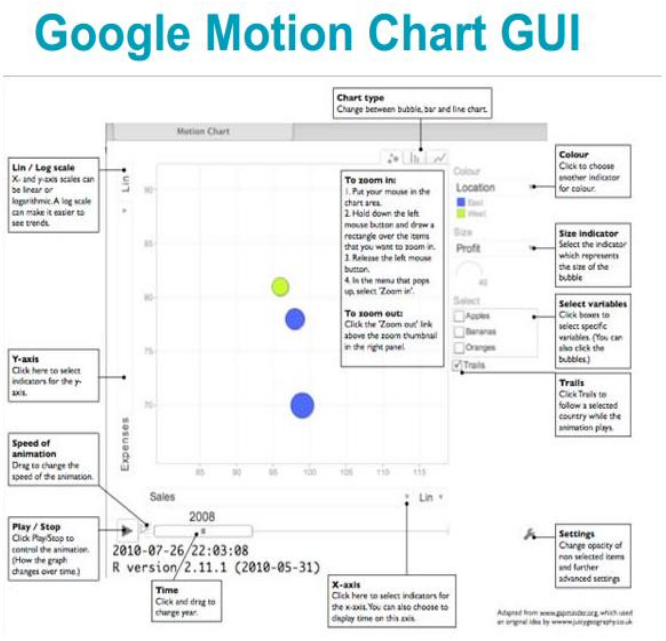

Fig 1: Parts of the Google Motion Chart GUI explained in detail

V. Predictive Analysis of Crime

Predictive analysis deals with observing a pattern of the existing data and identifying the future trends of the same.

We used R predictive analysis models to predict the Population, Crime and Unemployment for next 5 years in various cities of Santa Clara County. Linear Regression models in R are used to do the predictive analysis. Following are the list of models we used for analysis.

```
Call:
lm(formula = Population ~ Year + City, data = test.frame)

Residuals:
   Min     1Q Median     3Q    Max
-34832  -2488   -432   1787  37826

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1758964.4   356438.4  -4.935 2.61e-06 ***
Year                896.2      177.7   5.042 1.65e-06 ***
CityGilroy         7859.7     2877.7   2.731 0.007261 **
CityLos Altos     -10265.7     2877.7  -3.567 0.000519 ***
CityLos Gatos      -9329.0     2877.7  -3.242 0.001538 **
CityMilpitas       26898.8     2877.7   9.347 6.17e-16 ***
CityMorgan Hill    -2414.8     2877.7  -0.839 0.403049
CityMountain View  33159.5     2877.7  11.523  < 2e-16 ***
CityPalo Alto      21028.0     2877.7   7.307 3.30e-11 ***
CitySan Jose      890072.5     2877.7 309.301  < 2e-16 ***
CitySanta Clara    69335.2     2877.7  24.094  < 2e-16 ***
CitySunnyvale      94200.5     2877.7  32.735  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7049 on 120 degrees of freedom
Multiple R-squared:  0.9993, Adjusted R-squared:  0.9992
F-statistic: 1.525e+04 on 11 and 120 DF,  p-value: < 2.2e-16
```

Fig 2: Regression Model for Population Prediction

```
Call:
lm(formula = propertyCrime ~ Population + Year + City, data = test.frame)

Residuals:
     Min       1Q   Median       3Q      Max
-2763.34  -147.42     6.78   172.67  2810.95

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.855e+02  3.158e+04   0.025  0.98020
Population      3.199e-02  7.373e-03   4.338 3.03e-05 ***
Year           -3.624e-01  1.580e+01  -0.023  0.98174
CityGilroy      1.435e+02  2.395e+02   0.599  0.55025
CityLos Altos  -6.469e+02  2.444e+02  -2.646  0.00924 **
CityLos Gatos  -3.810e+02  2.424e+02  -1.572  0.11864
CityMilpitas   -3.307e+01  3.055e+02  -0.108  0.91400
CityMorgan Hill -3.069e+02  2.331e+02  -1.317  0.19049
CityMountain View -4.410e+02  3.373e+02  -1.307  0.19365
CityPalo Alto  -2.004e+02  2.794e+02  -0.717  0.47452
CitySan Jose   -6.828e+03  6.567e+03  -1.040  0.30058
CitySanta Clara -3.174e+02  5.616e+02  -0.565  0.57304
CitySunnyvale  -1.668e+03  7.324e+02  -2.277  0.02457 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 569.3 on 119 degrees of freedom
Multiple R-squared:  0.9924, Adjusted R-squared:  0.9916
F-statistic:  1295 on 12 and 119 DF,  p-value: < 2.2e-16
```

Fig 3: Regression Model for Property crime Prediction based on predicted Population

```
Call:
lm(formula = Unemployment ~ propertyCrime + Year + City, data = test.frame)

Residuals:
   Min     1Q Median     3Q    Max
-32416  -1843   -137   1845  43217

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.298e+06  4.254e+05  -3.051  0.00281 **
propertyCrime    1.169e+00  1.242e+00   0.941  0.34869
Year             6.475e+02  2.122e+02   3.051  0.00281 **
CityGilroy       2.175e+03  3.424e+03   0.635  0.52654
CityLos Altos   -1.678e+02  3.599e+03  -0.047  0.96290
CityLos Gatos   -1.692e+02  3.492e+03  -0.048  0.96143
CityMilpitas     1.457e+03  3.541e+03   0.411  0.68156
CityMorgan Hill  1.436e+03  3.422e+03   0.420  0.67553
CityMountain View  6.540e+02  3.475e+03   0.188  0.85105
CityPalo Alto   -6.331e+02  3.439e+03  -0.184  0.85428
CitySan Jose     4.421e+04  2.709e+04   1.632  0.10537
CitySanta Clara  2.408e+03  4.130e+03   0.583  0.56102
CitySunnyvale    4.076e+03  3.779e+03   1.079  0.28292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8301 on 119 degrees of freedom
Multiple R-squared:  0.8632, Adjusted R-squared:  0.8494
F-statistic: 62.59 on 12 and 119 DF,  p-value: < 2.2e-16
```

Fig 4: Regression Model for Unemployment Prediction

VI. Deployment steps

The CrazyRacoon application for visualization and prediction of crime of santa clara county is hosted in amazon ec2. A detailed description of the process of deploying the application in cloud is given below.

Step 1 – Create an amazon ec2 instance with Ubuntu12.10, using the key pair provided by amazon.

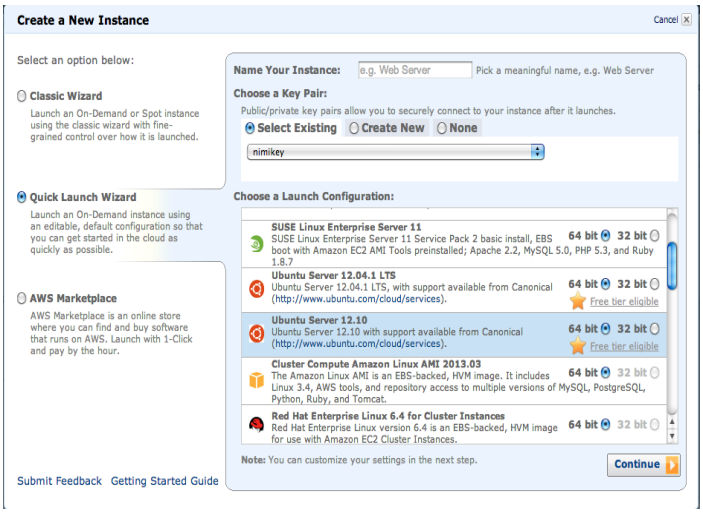

Fig 5: Creating a Amazon Instance

Step 2 - Edit the default security settings of the instance to make port 3838 available for public access. 3838 is the default port used by shiny server.

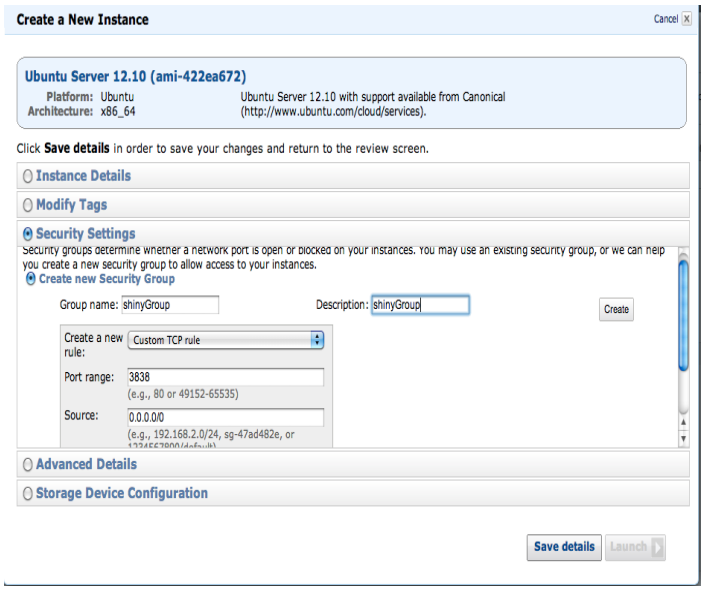

Fig 6: Making Port 3838 available for Shiny Server

Step 3 – Once the ec2 instance is up and running we can connect to the instance using an ssh client

ssh –i 'key.pem' ubuntu@'amazon instance name'

Step 4 – After connecting to the running ubuntu instance, start by installing nodejs which is a pre-requisite for shiny server.

Sudo apt-get install nodejs npm

Step 5 – The next step is tot install R from Cran-R repository http://lib.stat.cmu.edu/R/CRAN/bin/linux/ubuntu/

Step 6 – Sequentially install Shiny and googleVis packages and Shiny-server
Sudo su - -c "R –e \"install.Packages('shiny' , repos='http://cran.rstudio.com/')\""
Sudo su - -c "R –e \"install.Packages('googleVis' , repos='http://cran.rstudio.com/')\""
Sudo npm install –g shiny-server

Step 7 – Make a directory for the app CrazyRacoon and pull files from github

Sudo mkdir –p /var/shiny-server/www/CrazyRacoon
git pull
https://github.com/Nimmicv/CmpE272ProjectTeam2.git

Once the server.R, ui.R and data files are pulled to the app directory, we are ready to access the app using the public URL

VII. Solution in Detail

1. *Overview*

Any user of the application can visualize data, view table data and view the predicted data using motion charts, bar charts or line charts. A Google Motion chart gives a dynamic representation of crime statistics of various cities of Santa Clara County. Tables are used to dynamically show the data of any two cities selected by the user. These table data can be downloaded as a comma separated file at any point of time. Any two cities are selected and comparative crime statistics can be visualized. User can also see the predicted crime statistics with population and unemployment for a selected city for following years.

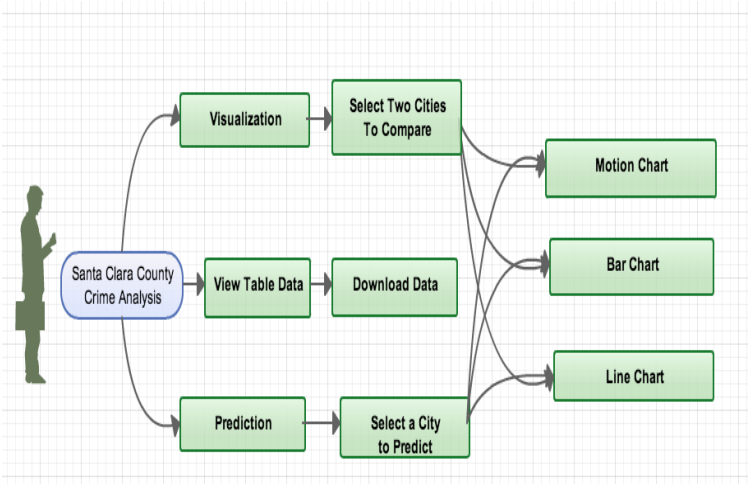

Fig 7: Overview of the UI functionality

2. *Architectural Details*

Ubuntu 12.0 instance of amazon ec2 is used to host the CrazyRacoon application. Shiny Server is listening to port 3838 of the ubuntu instance which is being accessed by a public user. The shiny server reads the UI and Server files of the shiny application in its web directory.
The server files incorporates all the server logic of reading data from CSV files, manipulating data and passing it on to googleVis packages for visualization of data on google motion charts and tables. Server uses linear prediction models to predict the data for following years, which is generated and pushed to google motion charts for visualization. The UI logic incorporates Twitter Bootstrap for default design templates.

The user connects to the application from his browser using the URL http://ec2-54-244-185-232.us-west-2.compute.amazonaws.com:3838/CrazyRacoon/ .
This HTTP request from the user is processed by shiny server hosted in the amazon cloud.

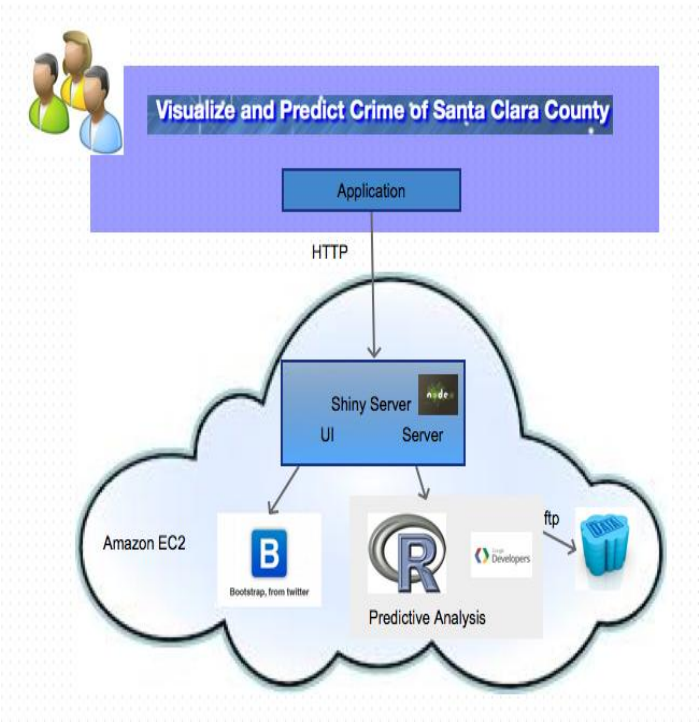

Fig 8: Architecture Diagram for Visualization and Prediction

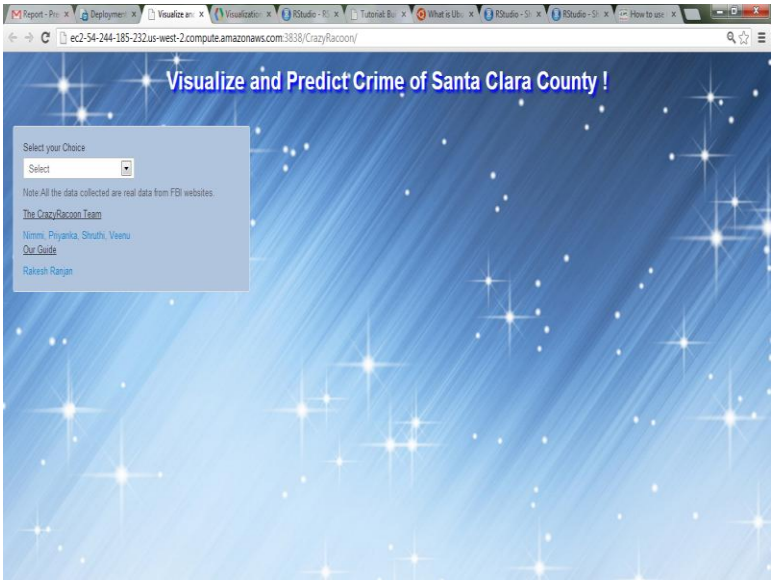

Fig 9: HomePage for CrazyRacoon Website

When the user selects visualization, the user needs to select two cities from the drop down menus of the city list provided. The screen will by default display a Bubble motion chart for the selected two cities for the years 2000 to 2011

VIII. Application in Detail

Our project opens with the home screen, where the user can select from three options: Visualization, Prediction and the plain Data Download option.

Following are the screen shots of the visualization and prediction application that we developed.

Visualization option is useful when the user just wants to compare different criminal activities between any two chosen cities. The city option can be chosen from a drop down box on the visualization screen. Visualization option has the visualization for the chosen cities from the Year 2000 to 2011. Visualization screen has the option of viewing the data in three forms: Bar Chart, Bubble Chart and Line graph. The user can make the choice and press the play button after choosing the three variables that should be compared. The bubble chart and the Bar chart provide an animated visualization, while the line graph is quick way of plotting the basics.

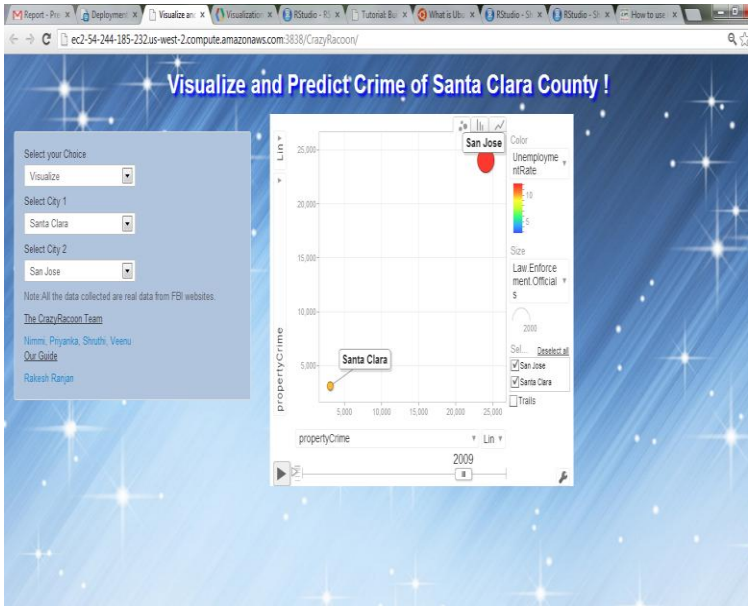

Fig 10: Visualization using Bubble Chart for San Jose and Santa Clara

The selection of bar chart gives the user a time series display for bar chart for the years 2000 to 2011

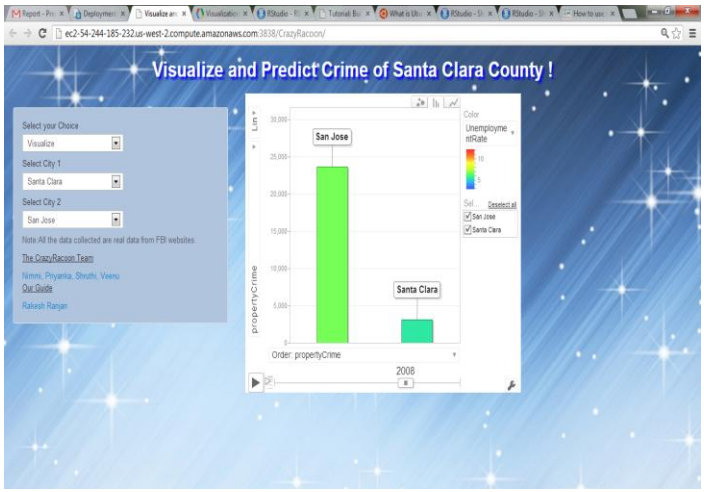

Fig 11: Visualization using Bar Chart for San Jose and Santa Clara

The figure below shows the line graph showing visualization for years 2000 -- 2011

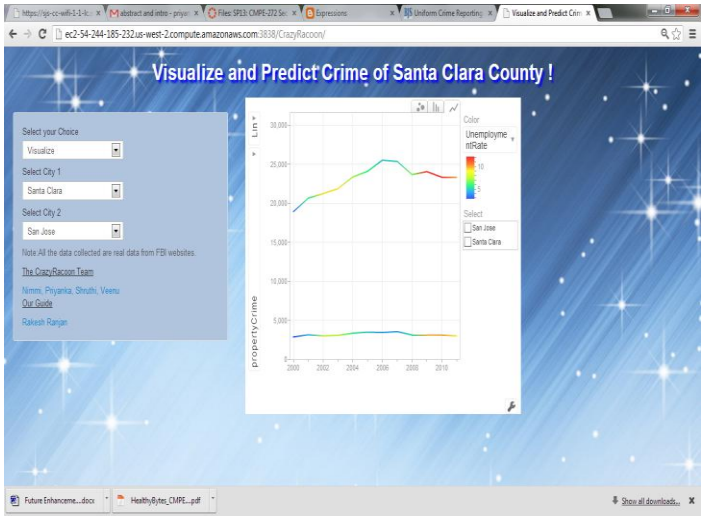

Fig 12: Visualization using Line Chart for San Jose and Santa Clara

The user can view the data in a clean tabular format on the website too. When the user makes a selection to View Table Data, he is allowed to choose two cities for which the data should be displayed. The figure below shows data for the selected two cities. This option is for people who want to utilize the collected data for their research or other purposes, or simply want a comparison in plain table format.

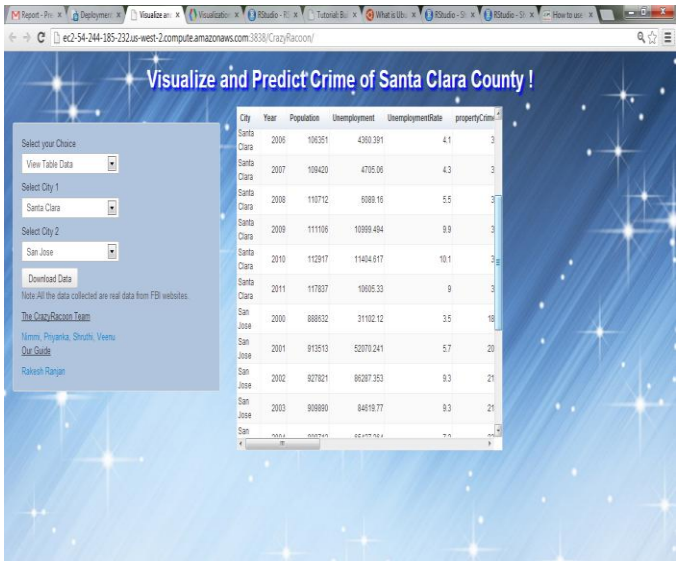

Fig 13: Data table for San Jose and Santa Clara

Prediction screen in our project has a drop down box where the user can choose a city for which the prediction is required. When the choice selected is prediction, the user can view the prediction in one of the three graph formats available on the website. The prediction provided right now if for the years 2014-2018. The user can either view the animated prediction using the bubble charts, or the plain simple plotting using the line graph. Bar chart provides an animation for the prediction too.

During the prediction process, we used linear regression models to train the system for prediction. The following predictions were applied to obtain the data for the years 2014-2018:

1.Population~Year+City – to give us the population prediction keeping Year and City constant.

2. PropertyCrime ~Population+Year+City – to give us the prediction for property crime, feeding in the population data from the previous prediction.

3. Unemployment~PropertyCrime+Population+City+Year to give us the unemployment prediction utilizing the data from population and property crime predictions.

The predicted values are dynamically written to a .csv file as and when the choices are made, and read out from the same file.

Below is the screenshot showing prediction of property crime in Santa Clara city for the years 2014-2018

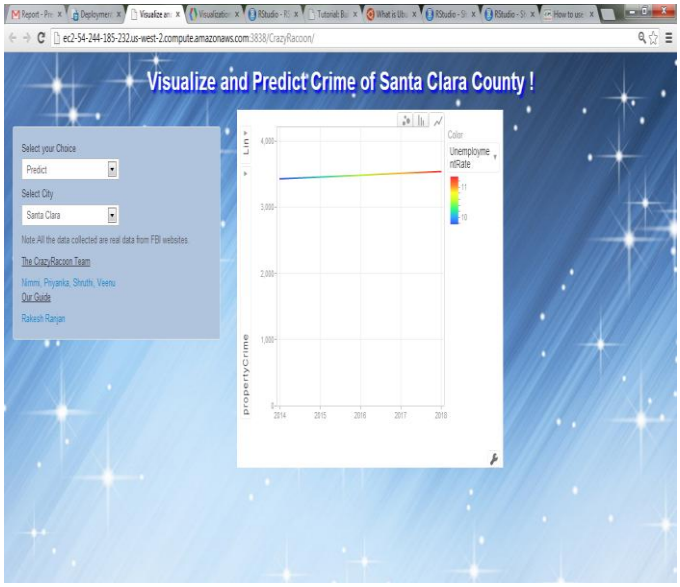

Fig 14: Prediction for 2014-2018 for Santa Clara

IX. Real Life Scenarios for Target users

In the current scenario, safety of people persists to be major issue in the society. The Crazy Racoon Application shows the recent trend of crime data in various cities in Santa Clara, comparison of crime rate between two cities, and predict crime rate for upcoming years. Thus, this application can be beneficial to public and assist them to decide on which area is safe enough to live, and in which city crime rate is expected to rise or fall. Some of the targeted users are police officials, data scientists, real estate seekers or property dealers. Below are the real life scenarios where the targeted user can utilize this application to solve problems.

1. Police Officials: Police Crazy Racoon Application can be useful for police officials. Our visualization feature aides police officials to visualize the various crime activities in various cities within Santa Clara County. Also, they could easily visualize the comparison of recent trend of criminal activities of two cities in Santa Clara County. This will help them to decide on whether they should increase the number of law enforcement officials in a city on viewing the recent trend of criminal activities in a city.

   The prediction feature of criminal activity will also help to take decisions on various fronts such as to increase the patrolling in a particular area.

2. Any Common Man: Every common man ‘s major concern is the security. Due to this, the Crazy Racoon Application provides them with visualization and prediction of criminal activities in their own city or neighboring area. Thus, this will help them decide on which city is safe to live. The prediction feature predicts the crime rate for upcoming years.

3. Data Scientist: Data scientists can access the crime data used to perform visualization and prediction. So, data analysts can fetch crime data and can be used it perform their own analysis on other factors according to their requirements.

4. Real Estate Seeker: The Cazy Racoon application performs predictive analysis to predict the property crime in various cities in Santa Clara County. The real state seekers can decide on which city can be a good fit to buy a property, which means that they could decide on whether they should sell or buy a property.

X. Future Enhancements

Completing the project in a short period of time was challenging but the model is capable of several future enhancements that can make this a more comprehensive system.

1. Prediction of crime for all states in USA

Collecting data for all the states of the USA, and mapping the model to predict the future of crime and other variables can make our system more comprehensive. The same linear regression model can be applied on the new data collected and can be applied to predictive analysis. The data collected however, should be continuous, consistent and accurate.

2. Immediate updating of recent data

The data we collected for our project reflected the crime and other parameters for the last ten years (2000-2011). For future work, the model can be enhanced to include the real time update for the most current data. If the system is being used in 2014, then the predictions may include the years that follow, and the data visualize may be up to date until the year 2013.

3. Identify new factors affecting Robbery and its prediction

The data that we have collected reflect three major factors - Unemployment, Population, and Number of Law Officials, influencing the robbery data. Several other factors may influence the crime pattern depending on the geographical location, and the demographics of the chosen city. The model can be further enhanced to include other factors affecting the crime pattern, which will give a better prediction and a more competent system.

XI. Conclusion

The CrazyRacoon application for visualization and prediction of crime is primarily aimed at visualizing and comparing crime statistics in different cities across Santa Clara County. Any two cities can be compared and contrasted for population, Murder, Forcible Rape, Robbery, Aggravated assault, Burglary, Larceny theft, Motor Vehicle theft, Arson, Total Crime, Law Enforcement officials and Unemployment Rate over the past 11 years starting from year 2000. A Linear regression analysis is done to predict the Population, Crime and Unemployment for the next 5 years in all cities of Santa Clara County.

Consider a potential user who is interested in renting an apartment in Santa Clara County, but he is worried of the increasing crime rates. CrazyRacoon comes to his rescue where he can find the crime rates of all cities and compare between cities using graphical charts. He can also view our prediction of population, crime and unemployment for next 5 years across various cities of Santa Clara County. With all the necessary information he can now make his best choice. The state officials can use this application to get a graphical overview of crime statistics and can take appropriate actions like deploying more law enforcement officials, boosting up job opportunities etc in cities with more crime rates. A Data scientist can use our application to download the data he is interested in and do further manipulations.

'Prevention is better than cure'. A realization of crime statistics in the city you live in can alert you on the safety measures you need to be bound to. This can help you save your valuable properties and above all your life which in immeasurably precious.

ACKNOWLEDGMENT

This project was completed under the guidance of Professor Rakesh Ranjan, in May 2013 for Department of Computer Engineering at San Jose State University. We are thankful to our Professor for being so inspiring and giving us a valuable learning experience.

REFERENCES

[1] "FBI Website", Data retrieved on 20th March, 2013, from http://www2.fbi.gov/ucr/cius2009/offenses/violent_crime/robbery.html

[2] "Government of USA data website", Data retrieved on 25th March, 2013, from http://www.data.gov/

[3] "Bureau of Labor Statistics website", Data retrieved on 2nd April, 2013, from http://www.bls.gov/lau/

[4] "Santa Clara Police Department website", Data retrieved on 2nd April, 2013 from http://scpd.org/index.aspx?page=9

[5] "San Jose Police Department website" Data retrieved on 3rd April, 2013 from http://www.sjpd.org/CrimeStats/

[6] "Shiny Server" Data Retrieved on 4th April, 2013, from http://www.rstudio.com/shiny/

[7] "googleVis" Data retrieved on 30th March, 2013 from http://cran.rproject.org/web/packages/googleVis/googleVis.pdf

[8] "State OF California Department of Justice Website", Data retrieved on April 2nd, 2013 from http://oag.ca.gov/crime

[9] "William B. King, PhD, Coastal Carolina University", Data Retrieved on April 3rd, 2013 from http://ww2.coastal.edu/kingw/statistics/R-tutorials/simplenonlinear.html

[10] "Using R for Linear Regression", Data retrieved on April3rd, 2013, from http://www.montefiore.ulg.ac.be/~kvansteen/GBIO00091/ac20092010/Class8/Using%20R%20for%20linear%20regression.pdf

[11] "Milpitas City data Website", Data retrieved on April 5th, 2013, from http://www.city-data.com/city/Milpitas- California.html

[12] Yvonnick Noel. *Psychologie statistique avec R*. Pratique R. Springer, Paris, 2013. [bib | http://www.springer.com/psychology/book/978-2- 8178-0424-8]

[13] Robert J Knell. *Introductory R: A Beginner's Guide to Data Visualisation and Analysis using R*. March 2013.ISBN978-0-9575971-0-5. [bib | http://www.introductoryr.co.uk]

[14] Paul Teetor. *R Cookbook*, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, published in 2011.