Safe Cities: An interactive crime rate visualization tool for prospective homeowners renters

Maryam Rahmanpour
mrahmanpour3@gatech.edu

Bharti Malik
bmalik30@gatech.edu

Joe Laniado
jlaniado3@gatech.edu

Kaige Lindberg
klindberg6@gatech.edu

1 INTRODUCTION

When purchasing or renting, especially in an area the buyer is unfamiliar with, insight into localized crime rates can be a key factor in a consumer's decision making. Many police departments provide geospatial visualizations of crime occurrences, but not aggregated statistical analysis. We are developing a web application that provides a localized breakdown of crime rates in city of Los Angeles, in a visual format a way that is easy to parse for prospective home buyers/renters.

The app we're developing should be most useful for consumers looking to rent or buy in an area that they're unfamiliar with. Since we want to provide a high degree of control over the final product, it could also be useful to anyone looking to research crime rates or correlations between crime other socio-economic factors.

2 PROBLEM DEFINITION

Using the LA Crime dataset from Kaggle[7], which reflects the incidents of crime in city of Los Angeles in the period of 2010 to 2021 and consists of more than two millions of rows crime data with the criminal code, date/time of occurrence, and latitude/longitude of occurrence, we wanted to perform the following tasks:

- Group the crimes by LA neighborhoods
- Assign each point a "crime weight" by its criminal code, and date of occurrence
- Take the summation of crime weights for each neighborhood, divide the total by the population of the neighborhood to derive an overall crime score by neighborhood.
- Design an intuitive GUI with user customization options to allow exploration of the data.

3 RELATED WORKS

We can draw on several existing works when deciding how best to assign crime indices on a geo-spatial level. Sathyadevan[13] specifically explores making an application for predicting what regions have high rates of criminal occurrence but limits itself to a Naïve Bayes classification algorithm. In "Big Data Analytics and Mining for Crime Data Analysis" [2] Holt-Winters exponential smoothing is used to create a time series model to show the crime trend in three big cities (San Francisco, Philadelphia, Chicago), then four different classification methods are used to group future crime predictions based on time and location, Naïve Bayes, K-NN, Gradient Boosted trees, and Random Forest are the applied models in this paper. There are many good and informative visualization in this paper which can be applied in our works too, however; no intractable ones are put into work. Telugu Maddileti et al. [8] use Logistic Regression, Decision Tree Classification Random Forest Classification to predict crime occurrences at a specific location which helps the law enforcement in speeding up the classification of criminal cases and proceed accordingly. While Shermila[9] actually explores predicting details concerning crime perpetrators using Mulitilinear Regression, K-Neighbors Classifier Neural Networks to detect crime patterns and predict the description of the perpetrator who is likely suspected to commit the crime. The methods used on the database in this paper can be useful for our project. Grubesic[3] specifically explores the time element of crime prediction, which could give us insight in how to present our data on a more granular timescale, e.g. the difference between crime in an area at night, and the crime in an area during the day. While many of these papers are concerned with prediction, and not description, we can actually use a prediction model to predict a region's immediate crime rate to assign an index, and it's possible these prediction methods may uncover seasonal effects.

We also want to think about the best kind of data to use, not just how to use it. Ramos et al.[12] explores levels of granularity for reporting crime data. Trading off between "internal uniformity" and "robustness to error", they ultimately recommend a level of granularity which is larger than street level, but smaller than neighborhood. This paper gives us good insight since finding the best granularity in geo-spatial studies is one of the key important criteria as well as in our study. Elizabeth[11] tries to prove that assigning crime indices to neighborhoods is poor methodology by exploring the high level of variance in crime inside of any given neighborhood, recommending that stats be collated on a street level. Lisowska-Kierepka[6] assigns crime indexes to grid squares on a map, roughly using a model that could be defined the portion of length of streets with crimes (weighted by number of crimes) over the total length of streets in a specific grid square. despite an innovative method of indexing, this paper does not account for population in each part of the city and all the crimes are weighted evenly. [5] uses regression models to connect socio-economic factors crimerates, which could be important for us to examine the possible unintended ramifications of our project, as well as exploring the ability to give better insight to the end user on what the crime indices in an area are ultimately indicative of. [1] explores a possible supplementary data source for our project, looking into the use of social media for analyzing crime data. We don't think this data will be useful for our stated purpose, but could provide an interesting secondary function.

Finally, we need to explore the best possible way to present our data. Srivas[14] explores several different GIS analytic tools for presenting geographical data; none of the techniques used interactivity, but we should be able to supplement their work with interactivity without too much issue. Mohammed[10] explores using Getis-Ord Gi statistics to create "hot spots", areas with unusually high levels of criminal activity, as well as "cold spots", not necessarily areas where crime is low, but where crime is low compared to the surrounding regions.

4 PROPOSED METHOD

One key innovation we have brought to our project over similar existing programs is to give the user the ability to customize the results to the types of crimes they specifically are more concerned about; e.g. if a user is more concerned about property crime, they should be able to increase the relative importance of property crimes in determining the overall "crime level" of a neighborhood compared to others. Our methodology can be divided roughly into two portions: data preparation and visualization.

4.1 Data Preparation:

The two tasks we had to handle in data preparation was assigning crime points to neighborhoods based on their latitude/longitude and calculating the "crime score" for every point. For each task we gathered and used complementary datasets like census data and latitude/longitude of center of each neighborhood to accomplish our defined goals.

To assign the points we found the latitude/longitude of the center of every neighborhood in Los Angeles, then treating the neighborhoods as clusters we assigned every data point to nearest cluster center using the Haversine distance[4]. With points assigned to neighborhoods, we then address the issue of each point's "crime weight". Rather than allowing the user to customize the weight of the more than 143 criminal codes in the dataset, we grouped all the crimes into 10 different categories: Drug, Sex, Theft, Human Trafficking, Violent, White Collar, Statutory, Domestic, Property, and Incholate crimes. Figure1 shows the Total number of crime for each category from 2010 till now.

The user can modify the weights for the overall crime categories, but the individual crimes that make up the categories each have a fixed weight set beforehand by us; for instance, both "Assault with Deadly Weapon, Aggravated Assault" and "Homicide" fall into the "Violent" crime category, but Assault has a backend rating of 1, while Homicide has a backend rating of 3. If a user leaves their customized weight for violent crime at 1, then the ratings for assault and homicide remain 1 and 3 respectively; if the user changes their weight for violent crime to 0.5, then the ratings for assault and homicide drop to 0.5 and 1.5. After setting the base

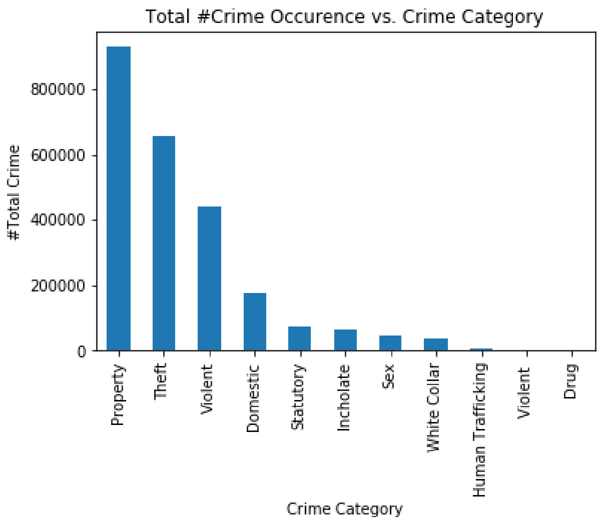

Figure 1: Total number of crimes for each crime category

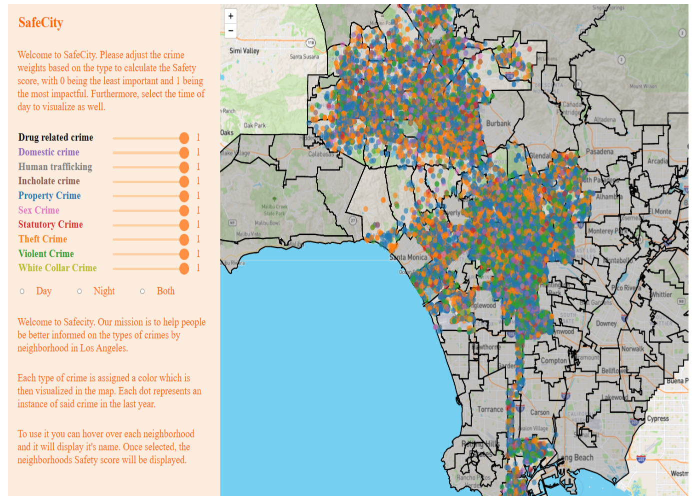

Figure 2: A total view of application

crime weight for a point, e.g. if the weight multiplier for violent crimes was set to 0.75, then the base crime weight for a homicide would be 2.25, the base weight is then adjusted based on how long ago it occurred to obtain the final weight; we want the model to prioritize more recent crimes, as well as allow neighborhoods which had a significant amount of crime in the past to get a better rating as time passes and time continues to drop. To achieve this, we reduce the crime weight for a given point by 0.1% every day using **geometric progression** with:

$$r = 0.999$$

$$c = \text{Category Weight(defined by user)}$$

$$a = \text{Base crime weight (defined with us)}$$

$$t = \text{\#days passing since the crime was committed}$$

The final equation for a given point's crime weight then is:

$$\text{Final Crime Weight} = a \times c \times r^t$$

and the overall crime index for a given neighborhood A is:

$$\text{Crime Index for Neighborhood A} = \frac{\sum \text{Final Crime Weights for all points in Neighborhood A}}{\text{Population of Neighborhood A}/1000}$$

Including the population of the neighborhood in the denominator was important because otherwise the actual effect of crime trends by neighborhoods will be drowned out by the fact that having a higher population naturally means more crime. Since the app is about addressing people's concerns over living in an area, we wanted to address that having the same amount of crime among a higher population means a person is less likely to be a victim of crime in an area; likewise having the same amount of crime among a lower population means a person is more likely to be a victim of crime, and we want our neighborhood crime index to reflect that.

4.2 Data Visualization:

We used D3 to create a visualization of LA, broken down by neighborhoods, with points plotted based on their lat/long (Figure2).

Points are color coded based on the crime category they fall into, and selecting a neighborhood displays "safety index", which is the order of that neighborhood when compared to other neighborhoods in LA based on their crime index; so the neighborhood with the lowest crime index will have safety index=100 which means

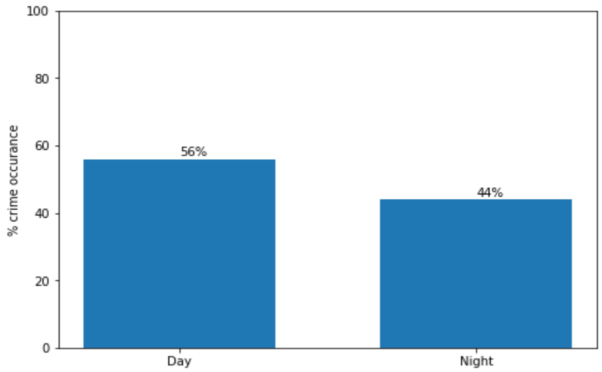

Figure 3: Day/Night crime occurance distribution

Neighborhoods	#Crimes
Downtown	104,861
Hollywood	74,151
Westlake	69,469
Van Nuys	62,825
Koreatown	52,936
San Pedro	49,068
Boyle Heights	48,373
Canoga Park	46,541
East Hollywood	46,295
Valley Glen	45,550

Table 1: Neighborhoods with most number of crimes

10 Most Dangerous Neighbors		10 Safest Neighbors	
Neighborhoods	Safety Score	Neighborhoods	Safety Score
Carthay	0	Brentwood	100
Chesterfield	1	Shadow Hills	99
Larchmont	2	Granada Hills	98
Leimert Park	3	Beverly Crest	97
Toluca Lake	4	Cheviot Hills	96
Chinatown	5	Westwood	95
Downtown	6	Mount Washington	94
Fairfax	7	Pacific Palisades	93
Beverlywood	8	Lake View Terrace	92
Vermont Knolls	9	Bel-Air	91

Table 2: List of most dangerous and safest Neighborhoods

selected neighborhood is safer than 100% of all other neighborhoods. On the left-hand side, the user can interact with sliders to change the category weights, at which point the crime indices and safety indices for every neighborhood is recalculated to meet user preferences.

Also there are Day/night/Both options in left side of the application. Selecting each of those options will visualize correspond crime distribution. If time occurrence of a crime is between 6 am and 6 pm it is considered in day time, otherwise it is considered as night time. However day and night time distribution of crimes are close to each other and is shown in Figure3.

5 EXPERIMENTS/ EVALUATION

Table1 and Table2 show some results of our experiments. As Table1 shows the neighborhoods with most number of crimes, After calculating crime indexes with time and population considerations we can see that safety score is not only depend on number crimes and neighborhoods have different ranking based on our indexing approach which we believe indicate more reliable index. It should be mentioned that these results are obtained with equal crime weights for all categories. As it is mentioned users can input their preferred weights for each crime category and receive different indexing results.

There is another experiment which shows that decaying crime score based on number of days that have been passed since the crime occurrence, is making difference in neighborhoods ranking. For example, if we do not consider the effect of time, "Bel Air" would be 8th safest neighborhood and "Mount Washington" would be 9th safest neighborhood. However, in our application and by considering time effect "Bel Air" becomes 10th safest neighborhood and "Mount Washington" become 7th safest neighborhood. Figure 4 depicts this time effect and shows crime rate in "Bel Air" has increased in recent years compare to crime rate in "Mount Washington" while in total number of crimes in "Mount Washington" is higher than "Bel Air" during 2010-2021.

In figure 5 and 6 there are some screenshots of the application, there is a bar in left side of the application and 10 different crime categories are listed in different color codes. Below the crime categories, there is a safety score for selected neighborhood on the map. Users can change the weights of each crime category with slider and see the updated score based on new weights. So they can use the application interactively and find their

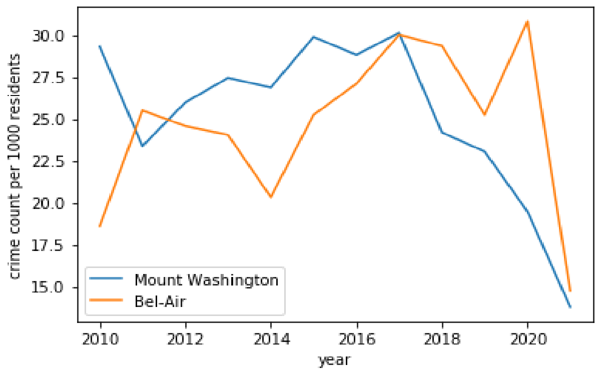

Figure 4: crime trend for "Bel Air" and "Mount Washington"

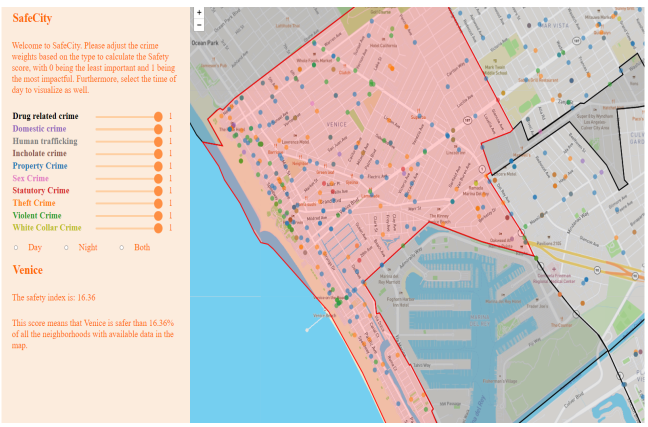

Figure 5: Safety score of "Venice" neighborhood with all weights equal to one

desire neighborhood to live. The colored dot on the map show the exact location of specific crime (according to color code) occurrence in last year which gives and intuitive insights about safety of different streets in a neighborhood.

We have shown the application to 15 individuals and asked them to work with that for 5-6 minutes and then fill a questionnaires. Each question has a range between 1-10. Figures 7, 8, and 9 show the results.

Users also included some comments on improvements they would like to see made to the app, such as the ability to click on a point and have a pop-up display data pertaining to that specific crime.

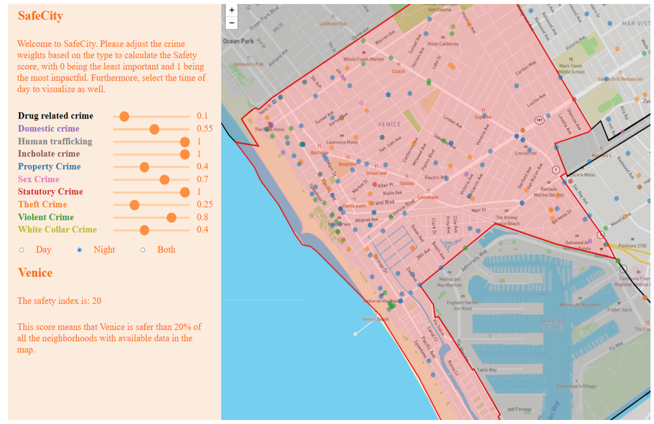

Figure 6: Safety score of "Venice" neighborhood with different weights for each category

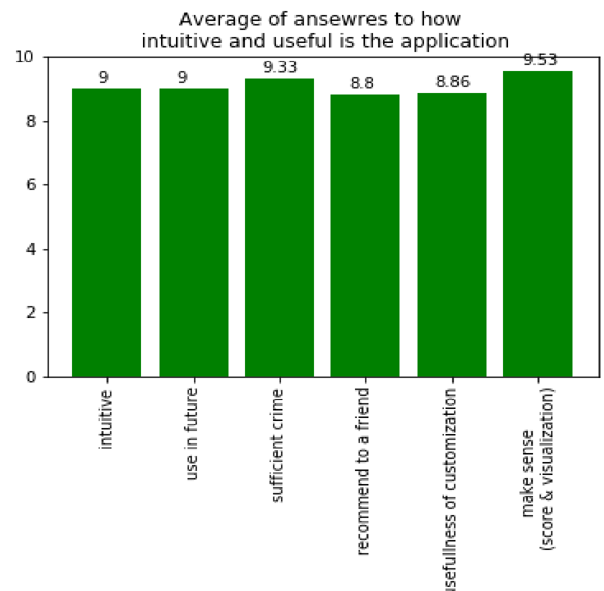

Figure 7: Average of answers of participants

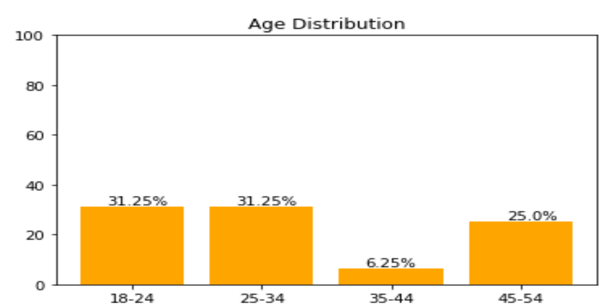

Figure 8: Age distribution of participants

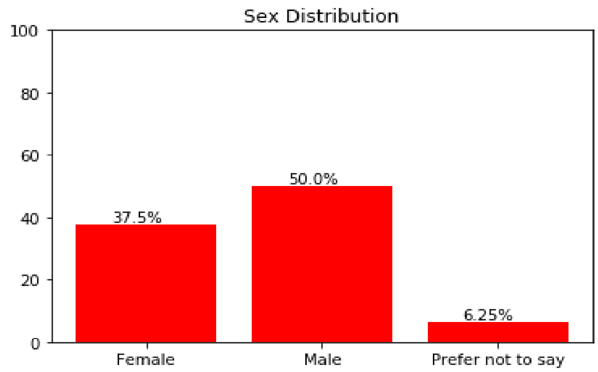

Figure 9: Sex distribution of participants

6 CONCLUSIONS AND DISCUSSION

SafeCity accomplished what we set out to do; it allows users to compare contrast neighborhoods not just by their level of criminal activity, but gives them the freedom to customize the application to focus specially on the types of crimes the user is most concerned about. Additionally, we ended up including the ability for the user to filter data based on when it occurred, providing an additional layer of customization we didn't originally have in mind.

The individuals we asked for feedback on the actual interface seemed pleased as a whole, most importantly they overwhelmingly felt that the app made sense and they could understand the information that was being conveyed, the most important goal of any data visualization project.

To take the app further, the most obvious thing we would need to do is expand beyond LA; finding a way to automatically aggregate data from many cities process it for use by our application would be essential if we wanted to handle more than just a few major cities.

We'd also want to host the app and the data on a server; currently the application needs to be ran locally, with the data also on the user's machine.

In comparison to established commercial alternatives to our project, we lack the level of polish or access to data that they have, but we do provide a reasonable alternative thanks to customization. We'd also likely want to explore further ways we can customize the experience to the user's preference, such as exploring the demographics of the victims of the crimes or perhaps exploring the level of police responsiveness in a given neighborhood (assuming any of this information is available).

7 PLAN OF ACTIVITIES

all team members have contributed similar amount of effort and work together to find the right methodology and also discuss all the steps twice a week in teams meeting. Other than that tasks were assigned to and accomplished by members as follow:

- Maryam and Kaige: Mostly focused on data gathering preparations, implementations of calculating index, and write up the report.
- Joe and Bharti: Mostly worked on D3 visualization and designing poster .

REFERENCES

[1] Mohamed Badawy et al. "Analysis and visualization of city crimes". English (US). In: *Proceedings of the International Conference on Industrial Engineering and Operations Management* 2018.SEP (2018). 3rd North American IEOM Conference. IEOM 2018 ; Conference date: 27-09-2018 Through 29-09-2018, pp. 1136–1145. ISSN: 2169-8767.

[2] Mingchen Feng et al. "Big Data Analytics and Mining for Crime Data Analysis, Visualization and Prediction". In: *Advances in Brain Inspired Cognitive Systems*. Ed. by Jinchang Ren et al. Cham: Springer International Publishing, 2018, pp. 605–614. ISBN: 978-3-030-00563-4.

[3] Grubesic et al. "Spatio-Temporal Interaction of Urban Crime". In: Sept. 2008, pp. 285–306. DOI: 10.1007/s10940-008-9047-5.

[4] *Haversine formula Wikipedia*. URL: https://en.wikipedia.org/wiki/Haversine_formula (visited on 12/01/2010).

[5] Weimin Li and John D. Radke. "Geospatial data integration and modeling for the investigation of urban neighborhood crime". In: *Annals of GIS* 18.3 (2012), pp. 185–205. DOI: 10.1080/19475683.2012.691903. eprint: https://doi.org/10.1080/19475683.2012.691903. URL: https://doi.org/10.1080/19475683.2012.691903.

[6] Agnieszka Lisowska-Kierepka. "How to analyse spatial distribution of crime? Crime risk indicator in an attempt to design an original method of spatial crime analysis". In: *Cities* (2021), p. 103403. ISSN: 0264-2751. DOI: https://doi.org/10.1016/j.cities.2021.103403. URL: https://www.sciencedirect.com/science/article/pii/S0264275121003024.

[7] *Los Angeles Crime Data 2010-2020*. URL: https://www.kaggle.com/sumaiaparveenshupti/los-angeles-crime-data-20102020?select=Crime_Data_from_2020_to_Present.csv (visited on 10/14/2010).

[8] Telugu Maddileti et al. “Crime Data Analysis Using Machine Learning Models”. In: *International Journal of Advanced Science and Technology* 29.9s (May 2020), pp. 3260–3268. URL: http://sersc.org/journals/index.php/IJAST/article/view/15887.

[9] A Mary Shermila, Amrith Basil Bellarmine, and Nirmala Santiago. “Crime Data Analysis and Prediction of Perpetrator Identity Using Machine Learning Approach”. In: *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. 2018, pp. 107–114. DOI: 10.1109/ICOEI.2018.8553904.

[10] Abbas F. Mohammed and Wadhah R. Baiee. “Analysis of Criminal Spatial Events in GIS for predicting hotspots”. In: *IOP Conference Series: Materials Science and Engineering* 928 (Nov. 2020), p. 032071. DOI: 10.1088/1757-899x/928/3/032071. URL: https://doi.org/10.1088/1757-899x/928/3/032071.

[11] Groff Elizabeth R., David Weisburd, and Sue-Ming Yang. “Is it Important to Examine Crime Trends at a Local “Micro” Level?: A Longitudinal Analysis of Street to Street Variability in Crime Trajectories”. In: *Journal of Quantitative Criminology* 26 (2010), pp. 7–32.

[12] Rafael G. Ramos et al. “Too Fine to be Good Issues of Granularity Uniformity and Error in Spatial Crime Analysis”. In: *Journal of Quantitative Criminology* 37 (June 2021), pp. 419–443. DOI: https://doi.org/10.1007/s10940-020-09474-6.

[13] Shiju Sathyadevan, Devan S., and Surya Gangadharan. “Crime Analysis and Prediction Using Data Mining”. In: Aug. 2014. DOI: 10.1109/CNSC.2014.6906719.

[14] Sanjay Srivas and PG Khot. “GIS-based Computational Tools & Techniques for Multidimensional Data Analysis & Visualization”. In: *International Journal of Applied Engineering Research* 13.15 (2018), pp. 11770–11775.