The GIS based Analysis Criminal Events using Analysing Crimes using Machine Learning technique

Abbas F. Mohammed and Wadhah R. Baiee

February 9, 2020

The GIS based Analysis Criminal Events using Analysing Crimes using A Machine Learning method

Abbas F. Mohammed[1], Wadhah R. Baiee[1]

[1] College of Information Technology, University of Babylon, Babylon, Iraq.

* Corresponding author.
E-mail address: wadhahbaiee@gmail.com

Abstract

Crime is an alarming aspect in our culture which is a crucial challenge to avoid this. the crime has been recorded with spatially distributions spread on the map it will be difficult for the local government and formal agencies to handle because of numerical crime it needs to concentrate on the specific location and reducing the coast and reduce the places of patrolling allocation and saving time and efforts. Police old fashion methods a pin placed on the map in the map on the wall to identify the crime and their detail it has to developed and merging the advanced technique to help the security do their job in practical and efficient way and sufficient to reduce time and effort. Crime evaluation is a methodical way in which patterns and trends in crime are detected and investigated. In this study, using different data mining clustering method to examine Baltimore, Maryland's crime information. Crime statistics are derived from the standardized resources of the U.S. government. It took the form of criminal data about the state of Maryland, such as the city of Baltimore, with 340,924 cases and 16 attributes to reflect the cases from 2012-2018. Density-Based Spatial Clustering with Noise (DBSCAN) algorithms are utilized to cluster crimes incidents focused on certain predefined events and the outcome of these clusters employed to find an appropriate cluster for the identification of crimes pattern recognition. The clustering findings are visualized by geographical information system (GIS) that combines to make crime distributions on map of the city and on the real-life to the law enforcement for interactive and easy understanding. This work to support and help law enforcement authorities to reliably predict and identify crime pattern recognition in Baltimore, thus reducing the location with high density of crimes and applying analysis on the spatial-temporal dataset will resize the location of police patrols that will concentrate on high density crime places. helping redistributions of the security forces in real life and high crime level neighbourhoods.

Keyword – Crime, clustering, GIS, Pattern recognition, DBSCAN algorithm, geographical dataset information, spatial, temporal, Geospatial, Hotspot.

Introduction

Crime is a social disorder behaviour it's a part of justice that is commonly prosecuted and punishable by law. Criminology is crime research and it is an informatics science that collects and examines the quality of crime data. Recently, criminal activity has grown and it is the police department's responsibility to manage and reduce illegal activity. Crime research is critical to giving police department patterns on the probability of crime in the future and related data such as crime area and possible tactics, crime category, etc. Crime pattern recognition of type shooting in during 2012–2018 and Criminal recognition is a major concern for the police force since there are large amounts of crimes. It, therefore, requires methodologies and improvements of certain techniques to help authorities identify crime trends and avoid crime in the future by supporting the relocation of security forces in certain areas that are somewhat dense and reliant on Machin Learning. For this function, Machine Learning offers clustering technique and geographical information system (GIS). Clustering has been used DBSCAN to analyses the crime attribute. A cluster points to a local crime set that could be viewed to use the map's geospatial plot while using geographical information system (GIS) the crime pattern recognition. It is a set of clusters which determines the cluster relation. The highly populated areas are labelled density location based on the attribute of crimes. Crime clustering is used primarily to classify criminal trends and also to anticipate crime. clustering is a data processing methodology used to identify and predict future information patterns dependent on similarity measures. The aim of this study is to examine crime in Baltimore county using clustering methodology for identification of crime trends and to classify the behaviour of criminals utilizing DBSCAN methods. Over the past ten years, the advent of remote sensing and survey technologies has dramatically increased our ability to gather millions of bytes of geographic data on a regular basis. Nevertheless, while information becomes difficult to discern, the abundance of spatial data can not be fully realised. This addresses GIS scientists with the emergency requirements for advanced techniques and methods to convert geographic data "intelligently and automatically" into data and also synthesise geographic knowledge [5]. This calls for new geographical representation, spatial modelling and visualisation techniques [6]. As a consequence of the digital revolution, information researchers face the same challenge that accelerates the output of millions of bytes of credit card transaction data, medical exams, phone calls, stock values, and numerous other human activities. The intervals between the sites of the crime are usually not difficult to access by the police in the unplanned areas. The badly planned areas are better represented by separating them into areas of clusters then the cluster-based analysis is performed. Therefore, the approaches that promote clustering are ideal for the badly planned configurations of crime investigation. GIS replaced the use of pins on maps. Spatial crime point patterns are focused on event coordinates such as incidence locations and may also include the occurrence time. It is possible to chart all or a sample of point pattern on the map. The purpose of the study of Spatial Crime Points trends is to determine whether the point pattern is randomly distributed, clustered or frequently. Patterns of spatial crime points are typically interpreted as clustering analysis. Usually, a dot map is used to represent patterns of spatial crime points. This method efficiently used to analysis clustering by evaluating crime incidences clustering in the detected hot spots where time and space relation analysis is required. The methods need the creation of distance matrices regarding the spatial attribute's relation between crimes. Crime hotspots are high-intensity regions on a map. They are designed to examine geographic areas in relation to the crime for researchers and analysts. Researchers investigate the frequency of hotspots in certain places and why it must be linked at this venue, and analysts study the techniques used to execute the method [1]. Developing maps containing hotspots is now a vital and powerful technique of policing; it helps to build knowledge and awareness of different areas of the city and possibly why crime occurs there and crime prevention is a try to reduce and discourage criminals activities. It analyses the crimes specific to government efforts to limit crime, enforce the law, and preserve criminal justice. It is necessary to study crime in both space and time in order to understand it. Nevertheless, the majority of crime trend analysis and related activities has been studying the spatial nature of crime ignoring the temporal aspect. Crime pattern analyses in this paper, focus on the discovering crime issue and proposes a clustering dataset of criminal activity recorded by police department of Baltimore city. Crime Pattern Discovery recognition in this paper depends on machine learning and clustering techniques.

Related Work

It will be more easily and less costly to reduce crime using machine learning, quantitative analysis to test the assumption of spatial-temporal dissimilarity locations and the low of crime concentration where the criminal commented crime whether its time changing due to temporal dimensions, or spatial dimension, that suppling a significant explication of present direction of violent crimes [2]. Mining information from vast amounts of spatial data is known as "spatial data mining". Because vast amounts of temporal-spatial data become a very pressing field collected in a huge range of applications ranging from geospatial data to biomedical information. The amount of spatial data collected increases exponentially and rapidly. Therefore, the human ability to analyse has exceeded its capacity. Recently, aggregation techniques have been identified as a primary means of extracting data to temporarily discover knowledge in spatial data sets. The development of aggregation algorithms got so much attention in the past ten years and proposed the development of a new aggregation technique of algorithms. DBSCAN is a leading density-based spatial group for noise algorithm [3]. In particular, we note that the number of crimes in one specific location is not limited only to its own historical records. Suggested two factors: temporal correlation Area-based spatial correlation, to measure such patterns of the crime data set it predicts [4]. It is quite difficult to achieve a large database because the information on the crime is classified. Crime analysis typically addresses information at the macro-level, like the regularity of crimes that occur on specific geography, rather than processing data at the incident level. It is suggested to deduce similar incidents using the aggregation (clustering) technique with mixed equivalent measures. Crime data are generated at the incident level using a geographic information system (GIS) parametric model [5]. Aggregation is a way to aggregate data in which each set of properties is identical without utilizing a recognized structure in identical data. An aggregation is a simple data extraction method which blends a set of items and topics in a manner that is more like an entity in the same category than in other classes. Recently, the Geographic Information System (GIS) has become effective and useful in dealing and displaying geographic information and its characteristics and to assess patterns of spatial crime. To determine the activity of criminals and hot spot crime. A set of geographic information systems (GIS) and data mining techniques have been employed as a tool for spatial and temporal detection to characterize where criminal activities are taking place [6]. Investigated patterns of crime and related activities in the spatial distribution of crime, ignoring the dimension of time. The crime, in particular, is classified by time and space. Analyzes of crime patterns and regular activity queries can now explore patterns using many methods. Using data on different crimes, it was found that crime hotspots quickly turn in response to the structure of life to understand spatial and temporal crime patterns. To create effective police strategies and input, both spatial and temporal aspects of crime incidents are required. This paper investigates a small point in time Use pattern-based techniques may help police and researchers discover new patterns of criminal activity. Thus, the investigation continues in this field. This paper reviews the patterns of growth of recurrent patterns of extraction of crime patterns [7]. Evaluation of spatial group is an essential way of extracting spatial data. Divides objects into groups as per commonalities in both aspects of the site and attributes. It plays a key role in determining density distribution and detecting hot spots. Spatial aggregation algorithms are found in spatial space, while algorithms in network space are less well researched [8]. Use pattern-based strategies has the potential to help crime researchers find new patterns of criminal activity. Research in this area also continues This paper reconsiders the expansion of recurrent criminal pattern mining patterns. In general, many types of research that have explored spatial patterns of crime kinds have described separate points for singular crime types separately, and have been studied if hotspot sites intersect or if crime kinds relate to hotspot sites; provide conclusions from a range of available data (e.g. Hotspots only). In comparison, a multivariate method uses this information from all regions to model correlation frameworks between types of crime and enables the generation of small-scale threats for each sort of crime from several popular, general and unique variables of crime at the same time. Aggregating facts and visibly displaying them into groups of facts, heterogeneous data is divided into several homogeneous groups. Prediction Discover patterns and information that may lead to sensible assumptions. The future value is estimated based on the record type. It complies with results of constant value. Visualizations-enables researchers to locate critical information quickly and efficiently. This also relates to the presentation of data in a manner that allows users to display complex data patterns. It is used to provide a clear understanding of the trends or interactions that are observed in accordance with other data mining methods. A crime hotspot is referred to as the focus areas for criminal activities that are usually verified by analysing crime data in geographical information system (GIS) [9]. As a result, crime hotspots map analysis assists law enforcement agencies to identify places with high crimes rates, the sorts of crimes undertaken and the best way to react to crime. Hotspot maps are an effective means of high-density visualisation of crime-prone areas and the use of GIS tools for spatial and temporal analysis has been shown to be successful in identifying the implied relation between events [10]. Provides an idea of how criminals are spatially distributed and can, therefore, help tackle them. GIS is an essential tool to deal with spatial and temporal analysis, display it on a map and fight crime [11]. A vast number of complaints and media about incidents, which are committed almost hourly, have made it harder, if not complicated, to identify these crimes. The need to track and report these crimes is, therefore, growing as a critical way to monitor and identify these signs of crime on the media. Recorded clustering has become progressively an important objective with unsupervised methods of learning to achieve good outcomes. This seeks to use various types of extractions and cluster algorithms to dynamically combine related documentation in one cluster. Continuous work is being done to improve document clustering methods such as extraction and clustering strategies to resolve the complexity of designing a general-purpose document clustering for criminal investigation and the unplaced issue of extraction and clustering.

Methodology

Data Set

The crime dataset includes a lot of information on crime Baltimore City Police Services, Maryland collects the crime datasets. Specifically, in 2012-2018. The set of crime data is formed of 340,924 cases of crime and 16 attributes. The sort of crime attribute consists of 15 distinct types of crime. Shooting 3822, Aggravated. assault (11,038), Common assault (21,074), Larceny from car (theft from vehicle) (21,352), Auto theft (11,458), Burglary 51303, Larceny 73859, homicide 73859, Assault by threat 4447, Robbery – street 22598, Robbery – residence 3550, Robbery- carjacking 2174 , Robbery – commercial 5433, Rape 2122, Arson (27,148). Just the first of the categories will be used in the study since the last classification will be a collection of various sorts of crime with the same attribute and in the potential. To analyzing the attribute using DBSCAN methods and Getis-Ord Gi*, 15 different types of crime were used in this paper will analyze. The data came from data.gov from the Police dept of and communicate with public safety expert on Make decision the City of Baltimore (Maryland) and it is available on the web. It identified in a specific address, and the offences were particular location by longitude and latitude. In order to produce average positions from 2012 to 2018, the offences are geo-coded into the map of Baltimore Maryland using geographical information system, there were 340924 reported sites of crimes in Baltimore. Table 1 shows section of crime dataset in raw spreadsheet form**.** The proposed method dissociation the attribute of the criminals action crime and configure the mentalism of criminality action depending on the crimes records spreads on the map and their feature analyzing it, and it powerfully believes that finding relationships between crimes components might extremely facilitate in detecting crimes patterns recognition and potentially hazardous hotspots in the long term at time possible happened, the probable time of occurrence and high crime cases in this time Thus, this work plan method aimed to focus on three key pieces of information on crime, which are the description of crime it will be separate the types in to categorise, the prevalence time and the location of the crime, and will focus on the spatial analysis of the data. This method tried to extract all possible attention-grabbing frequent patterns supported the crime variables by clustering methods. Then, this paper applied clustering technique using the data set features and concentrate on the key features (the description of the crime, the time, crime, date and the location of the crime) based on their longitude and latitude occurrences. Preprocessing the dataset and extraction feature do not have null value and deleting the null records do not influence the clustering accuracy method. Displaying the dataset on the ArcMap **G**eographical **I**nformation **S**ystem (**GIS**) before analysing the data in the proposed method and after using the clustering technique helping to detect the crime pattern depending on the attribute of the crime. To extract the most criminal density from the dataset records using geographical information system (GIS). This study is using the DBSCAN algorithm on the dataset. The purpose of the clustering is to detect the potential crime pattern in a specific location within a particular time in the probability. The goal of the analysis is to give the best precision in identification of crime trends. DBSCAN algorithm is among the basic algorithms for clustering mining patterns and is one of the important data mining building blocks. DBSCAN since its capability of discovering clusters with arbitrary shapes using machine learning algorithm helping to reduce time and effort, the characteristic of this algorithm that it does not need the number of clusters or the distance to be separated between the clusters and do not need centres of each cluster, this is what made it more dependable in dealing with such cases of data and spatially temporally attribute to determine the most significant hotspot in crime mapping and appropriate spatial dataset using spatial clustering analysing it to detect the hot spot crime area. DBSCAN resorts two parameters to characterize “density”, namely " (a positive real number) and MinPTS (a positive integer). Given a d-dimensional data object p, the ball cantered at p with the radius of clusters is considered as dense if it covers at least MinPTS Objects. Then the clusters formed by considering all the objects in a dense cluster cantered at p should be added to the same cluster as p. Furthermore, two clusters can be merged when the centre object of a dense cluster is added to another cluster. The merging will be performed to the effect’s fullest extent until no more clusters can be merged. DBSCAN suffers from time-intensive computations since it needs to perform n " range queries and cluster labelling propagation for all the objects. This searches the datasets to retrieve all satisfactory element sets. The aim of using this method is to identify regular trends of all potential crime depending on the type of criminal activity and to collect the outer point spread over the field which underestimates the significance and risks of these points in the specific area. This method aims to establish a dataset of all crime hotspots along with their regular period correlated with it.

using the places and time attribute in DBSCAN algorithm and omitted the noise points in the result of DBSCAN crimes attribute form before using the Getis-Ord Gi* statistics analysing to determine hot spot location. Public safety and crime prediction operations are among both residents and government's most essential worries. Such practises are dedicated to vast amounts of money, human resources, equipment and supplies. Besides, there is a massive concern to validate the allocation of police resources. If the police could anticipate, with an acceptable accuracy, when and where a specific type of criminal activity will take place, a double gain would be achieved. First, it would also be doable to focus the required logistical practises and assets to combat this specific type of criminal activity in the geographic area and the timeframe forecast, and the comparison between the number of resources allocated to police forces and the results they have achieved can lead towards a more sufficient basis for preparing and distributing public security.

Table 1: Shows the section of crime dataset in raw spreadsheet form.

t	CrimeDate	Crime Time	Crime Code	Location	Description	Inside/ Outside	Weapon	Post	District
1	12/8/2018	23:20:00	4E	100 S EUTAW ST	COMMON ASSAULT	I		113	CENTRAL
2	12/8/2018	23:00:00	6D	900 S CATON AVE	LARCENY FROM AUTO	O		832	SOUTHWESTERN
3	12/8/2018	23:00:00	6D	2600 HUDSON ST	LARCENY FROM AUTO	O		232	SOUTHEASTERN
4	12/8/2018	22:50:00	7A	3800 MARY AVE	AUTO THEFT	O		425	NORTHEASTERN
5	12/8/2018	22:49:00	4E	300 S CATHERINE ST	COMMON ASSAULT	I		842	SOUTHWESTERN
6	12/8/2018	22:15:00	3AF	NORTH AV & N MOUNT ST	ROBBERY - STREET	O	FIREARM	733	WESTERN
7	12/8/2018	22:00:00	4E	2900 ROCKROSE AVE	COMMON ASSAULT	I		612	NORTHWESTERN
8	12/8/2018	22:00:00	4E	2900 ROCKROSE AVE	COMMON ASSAULT	I		612	NORTHWESTERN
9	12/8/2018	21:48:00	4B	1400 KUPER ST	AGG. ASSAULT	O	KNIFE	935	SOUTHERN
10	12/8/2018	21:30:00	3AF	0 ALBEMARLE ST	ROBBERY - STREET	O	FIREARM	211	SOUTHEASTERN
11	12/8/2018	21:13:00	4E	2700 W FRANKLIN ST	COMMON ASSAULT	I		721	WESTERN
12	12/8/2018	20:30:00	6C	4400 HARFORD RD	LARCENY	I		422	NORTHEASTERN
13	12/8/2018	20:00:00	6D	PAUL ST & E 31ST ST	LARCENY FROM AUTO	O		512	NORTHERN
14	12/8/2018	20:00:00	7A	4700 ERDMAN AVE	AUTO THEFT	O		433	NORTHEASTERN
15	12/8/2018	19:35:00	6J	1200 SAINT PAUL ST	LARCENY	O		141	CENTRAL
16	12/8/2018	19:00:00	4E	1900 W LEXINGTON ST	COMMON ASSAULT	I		714	WESTERN
17	12/8/2018	18:55:00	9S	200 N CASTLE ST	SHOOTING	O	FIREARM	212	SOUTHEASTERN
18	12/8/2018	18:55:00	9S	200 N CASTLE ST	SHOOTING	O	FIREARM	212	SOUTHEASTERN
19	12/8/2018	18:55:00	4A	N CASTLE ST	AGG. ASSAULT	O	FIREARM	212	SOUTHEASTERN

Figure 1 shows the Baltimore crime spatially distributed on the map using a geographical information system (GIS).

Neighborhood	Longitude	Latitude	Location 1	Premise	crimeCaseNumber	Total Incidents
Downtown West	-76.62083	39.28724	(39.287240000000, -76.620830000000)	HOTEL/MOTE		1
Violetville	-76.67137	39.27355	(39.273550000000, -76.671370000000)	PARKING LO		1
Canton	-76.57891	39.28211	(39.282110000000, -76.578910000000)	STREET		1
Glenham-Belhar	-76.54639	39.34841	(39.348410000000, -76.546390000000)	STREET		1
Shipley Hill	-76.65599	39.28388	(39.283880000000, -76.655990000000)	ROW/TOWNHO		1
Sandtown-Winchester	-76.64499	39.30995	(39.309950000000, -76.644990000000)	STREET		1
Park Circle	-76.66197	39.32895	(39.328950000000, -76.661970000000)	ROW/TOWNHO		1
Park Circle	-76.66197	39.32895	(39.328950000000, -76.661970000000)	ROW/TOWNHO		1
New Southwest/Mount Clare	-76.63975	39.28505	(39.285050000000, -76.639750000000)	STREET		1
Jonestown	-76.60422	39.28909	(39.289090000000, -76.604220000000)	STREET		1
Rosemont Homeowners/Tenan	-76.66158	39.29327	(39.293270000000, -76.661580000000)	ROW/TOWNHO		1
Beverly Hills	-76.5726	39.3397	(39.339700000000, -76.572600000000)	CHAIN FOOD		1
Charles Village	-76.6159	39.32583	(39.325830000000, -76.615900000000)	STREET		1
Armistead Gardens	-76.56099	39.30815	(39.308150000000, -76.560990000000)	STREET		1
Mid-Town Belvedere	-76.61449	39.30365	(39.303650000000, -76.614490000000)	BUS/AUTO		1
Penrose/Fayette Street Ou	-76.64805	39.29056	(39.290560000000, -76.648050000000)	ROW/TOWNHO		1
CARE	-76.58848	39.29455	(39.294550000000, -76.588480000000)	Street		1
CARE	-76.58848	39.29455	(39.294550000000, -76.588480000000)	Street		1
CARE	-76.58848	39.29455	(39.294550000000, -76.588480000000)	STREET		1

The DBSCAN algorithm evolves regions with adequately high cluster density and sees arbitrary shape clusters with noise in spatial data sets. This identifies a cluster as a total series of points correlated with distance.

The procedure is:

Algorithm: DBSCAN

Input: N attributes to be clustered as well as Eps, MinPts global variables.

Output: Clusters of Crime

Method:

1. Select a point P arbitrarily.
2. From P, recover all density-accessible points from P wrt Eps and MinPts.
3. If P is a core point, there will be a cluster.
4. When P is a boundary point, no points can be reached from P and the next point of the list is visited by DBSCAN.
5. Process all the other points in the same way.

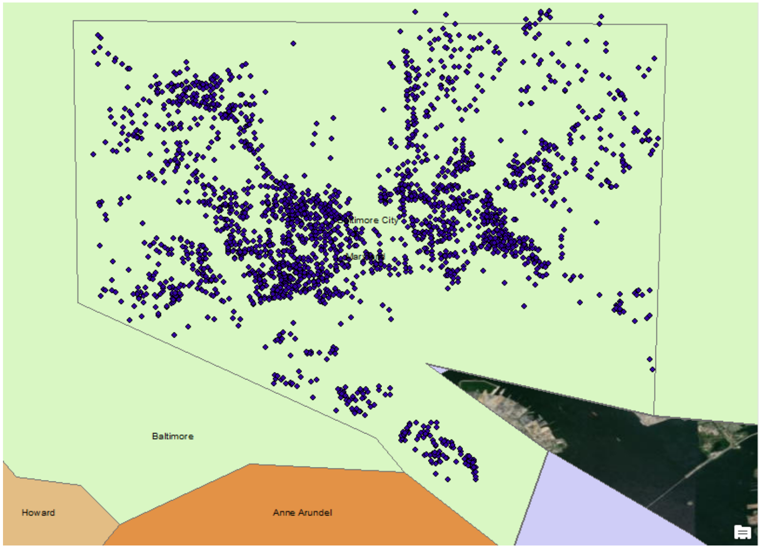
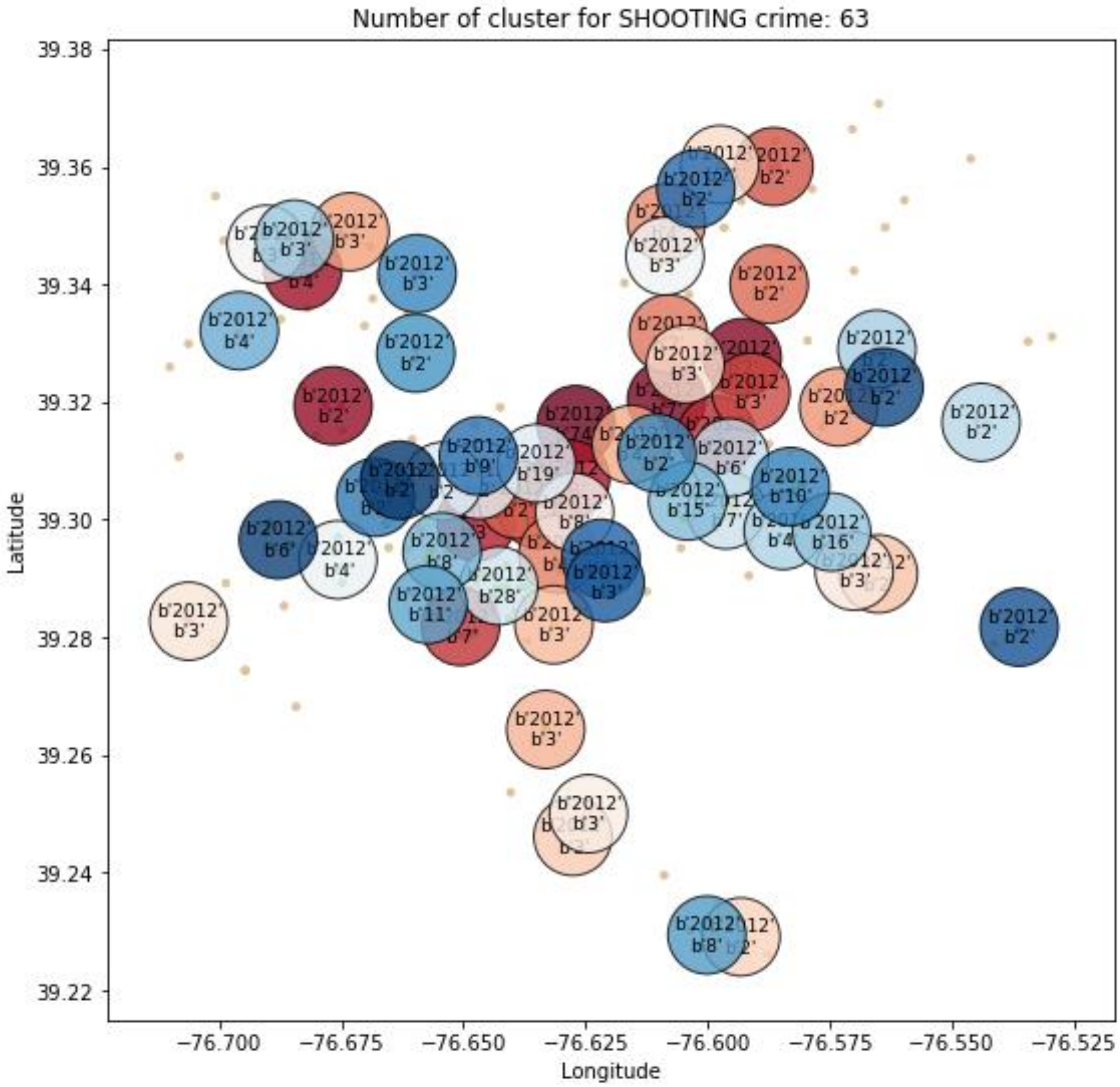


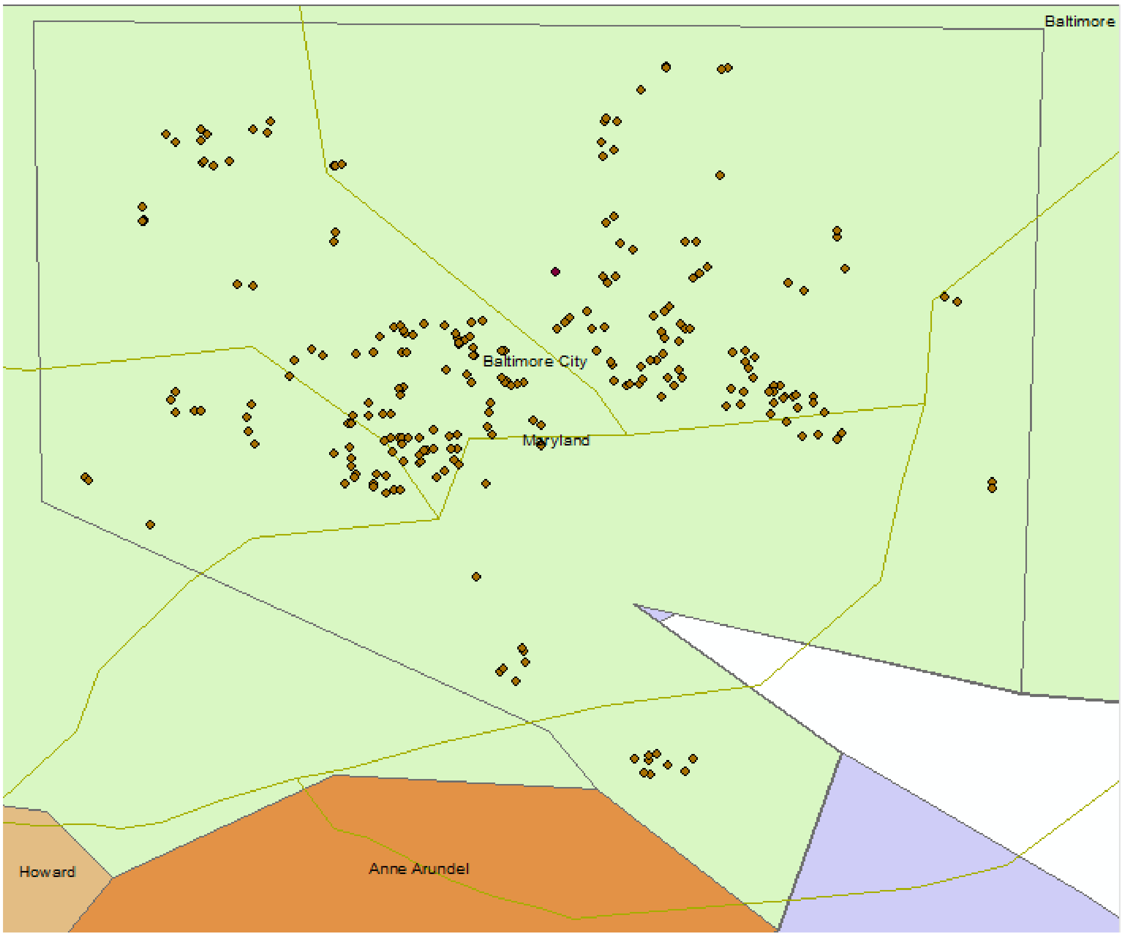
Figure 1: Baltimore Stat boundaries and displaying crime Shooting on the stat map using GIS.

This algorithm is given the crime dataset as an input. Crime date, time of the crime. The DBSCAN clustering is performed result, the Crime pattern recognition of type shooting in Baltimore during 2012–2018.

Result and Discussion

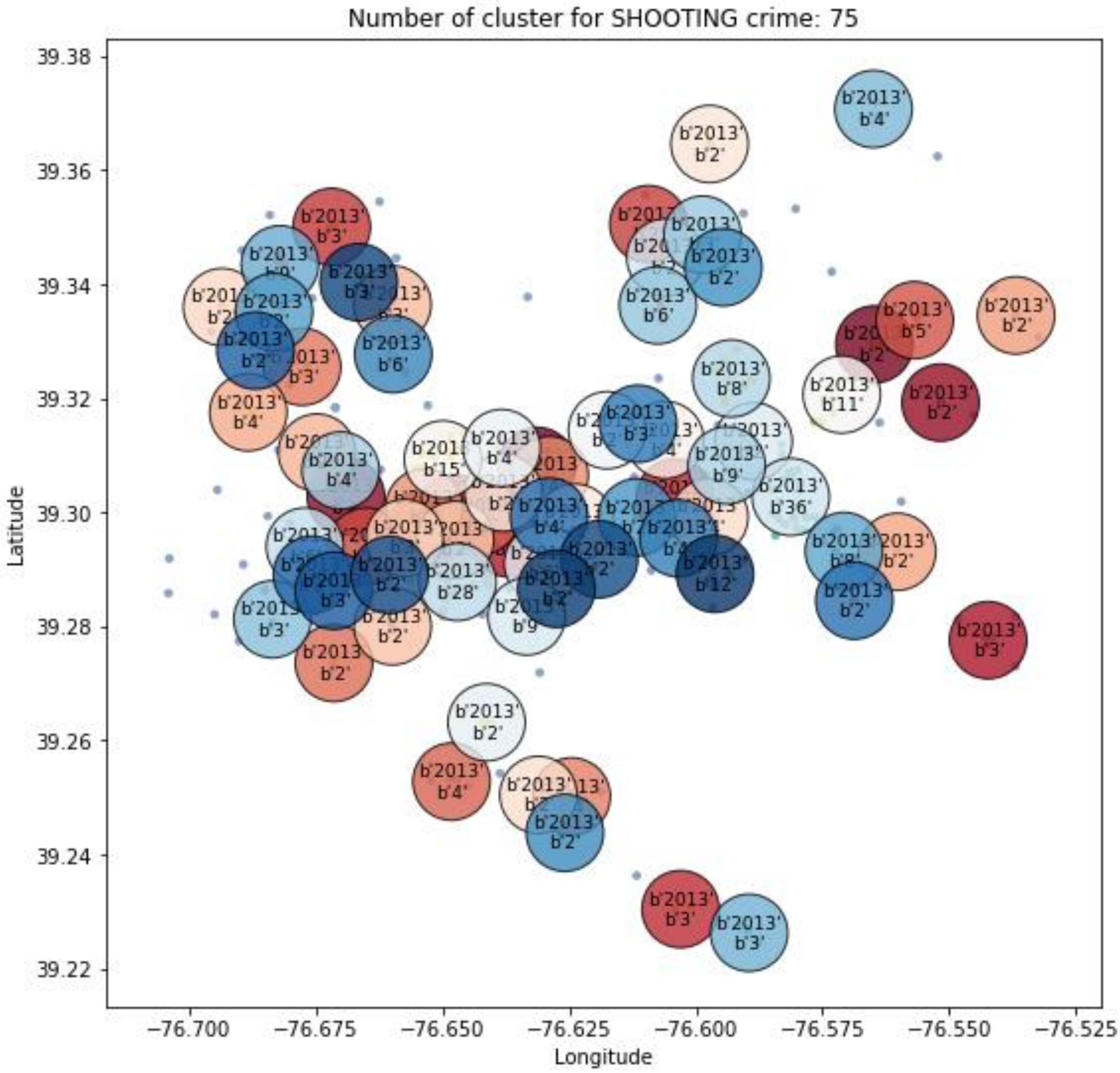


Completeness: 0.142
Estimated number of clusters: 63
Estimated number of noise points: 74
Homogeneity: 1.000
V-measure: 0.249
Adjusted Rand Index: 0.058
Adjusted Mutual Information: 0.210
Silhouette Coefficient: 0.275
precision_score: 1.000

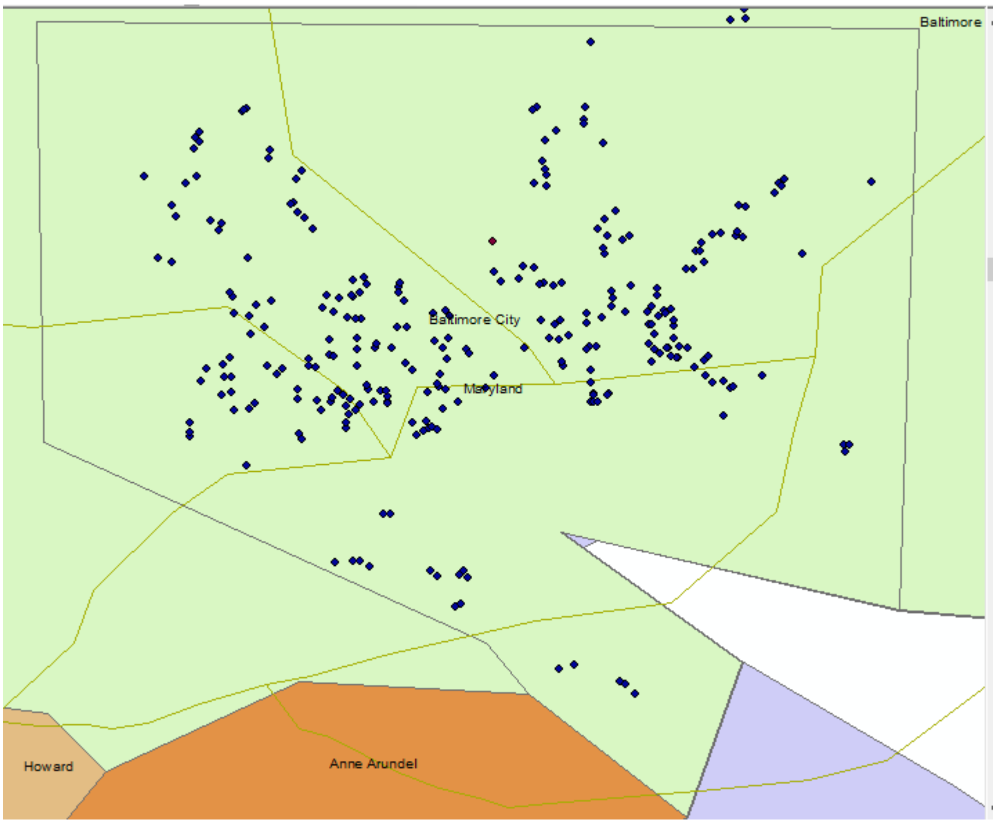


Clustering Crime shooting Baltimore mayland (2012).

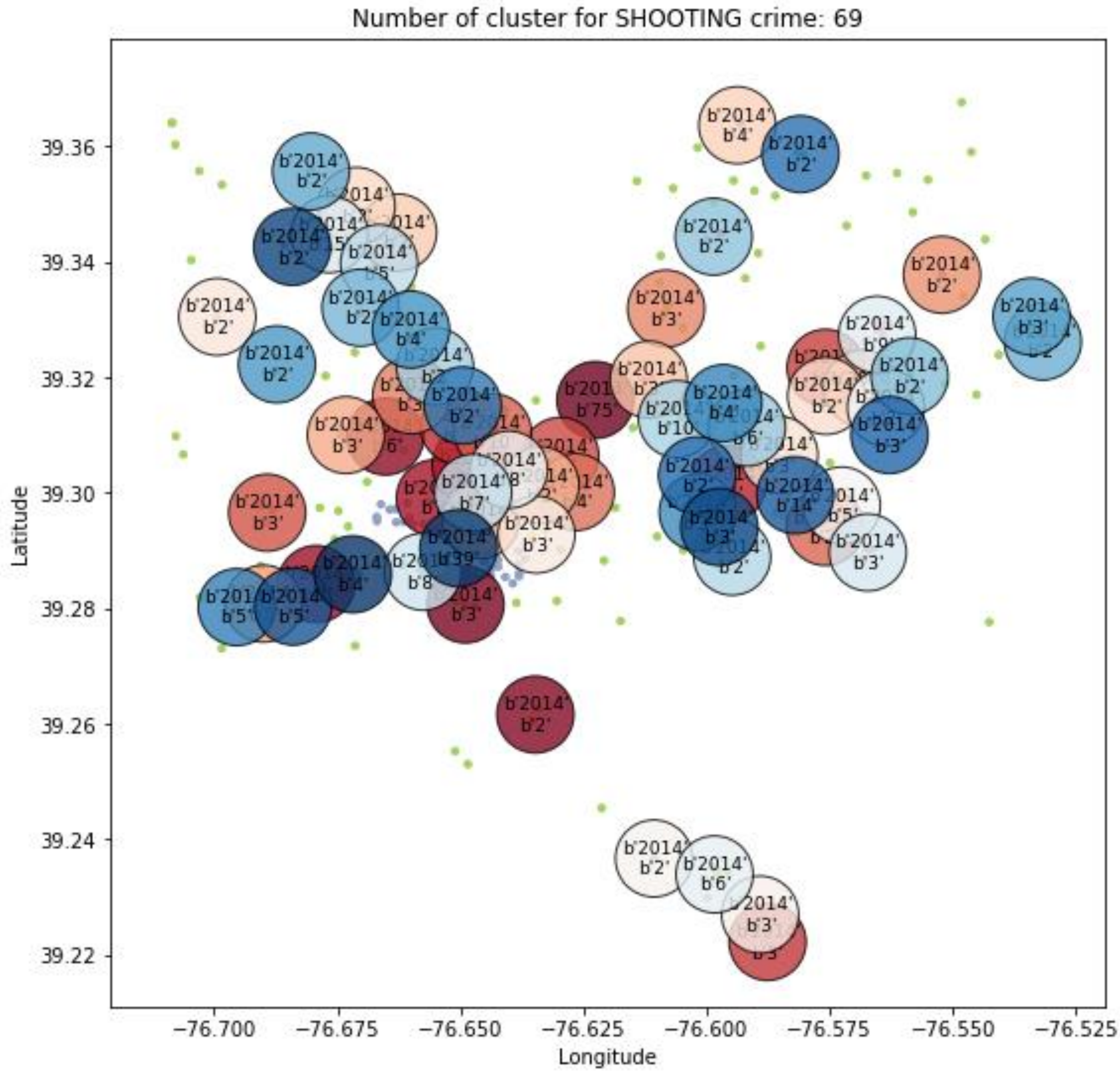
Completeness: 0.142
Estimated number of clusters: 63
Estimated number of noise points: 74
Homogeneity: 1.000
V-measure: 0.249
Adjusted Rand Index: 0.058
Adjusted Mutual Information: 0.210
Silhouette Coefficient: 0.275
precision_score: 1.000



length after 401
Completeness: 0.106
Estimated number of clusters: 75
Estimated number of noise points: 56
Homogeneity: 1.000
V-measure: 0.192
Adjusted Rand Index: 0.026
Adjusted Mutual Information: 0.153
Silhouette Coefficient: 0.354
precision_score: 1.000

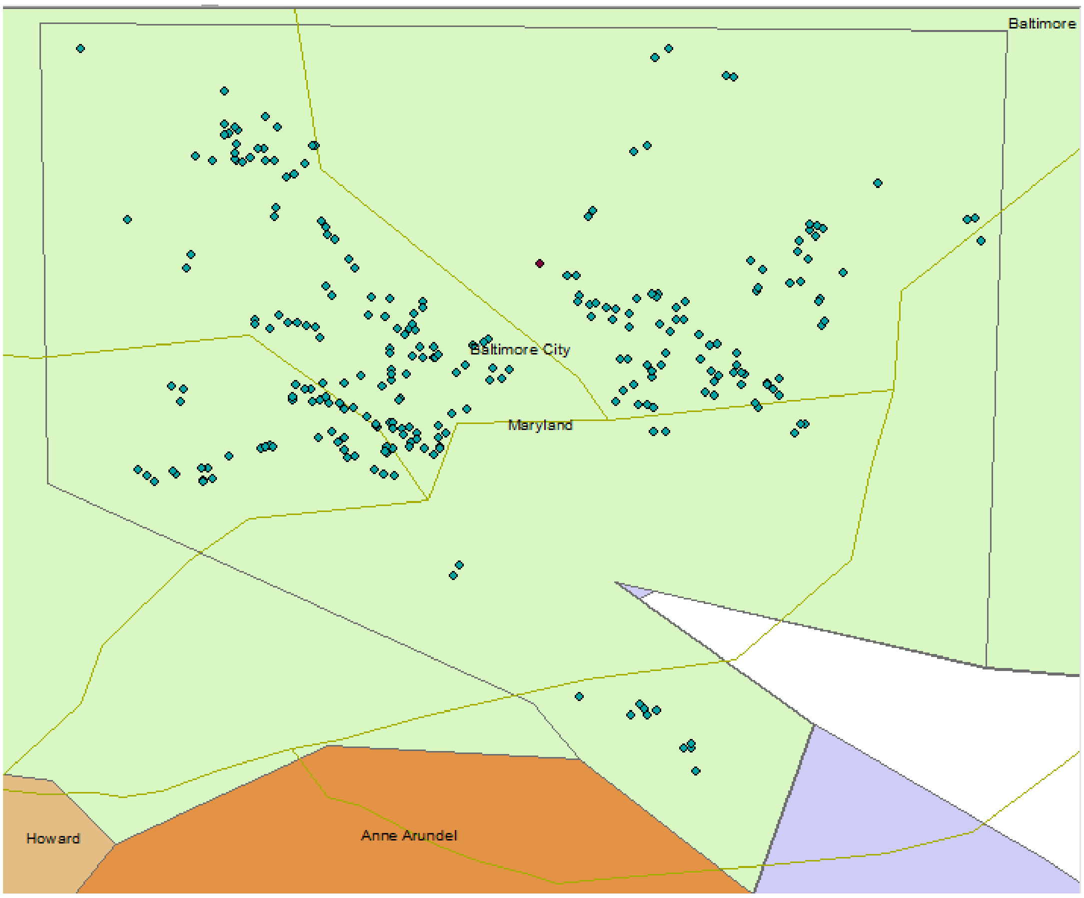


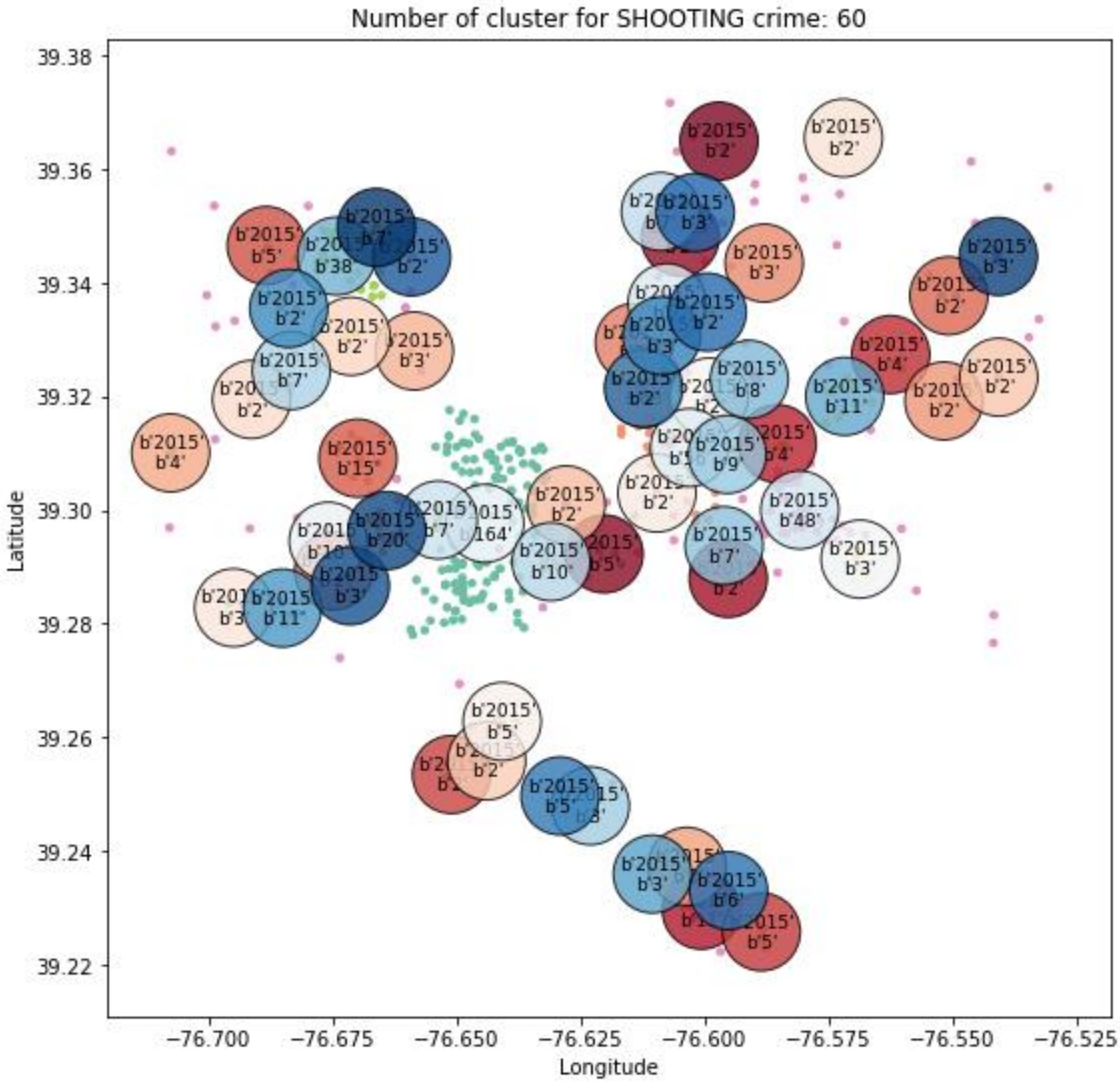
Baltimore
Baltimore City
Maryland
Howard
Anne Arundel



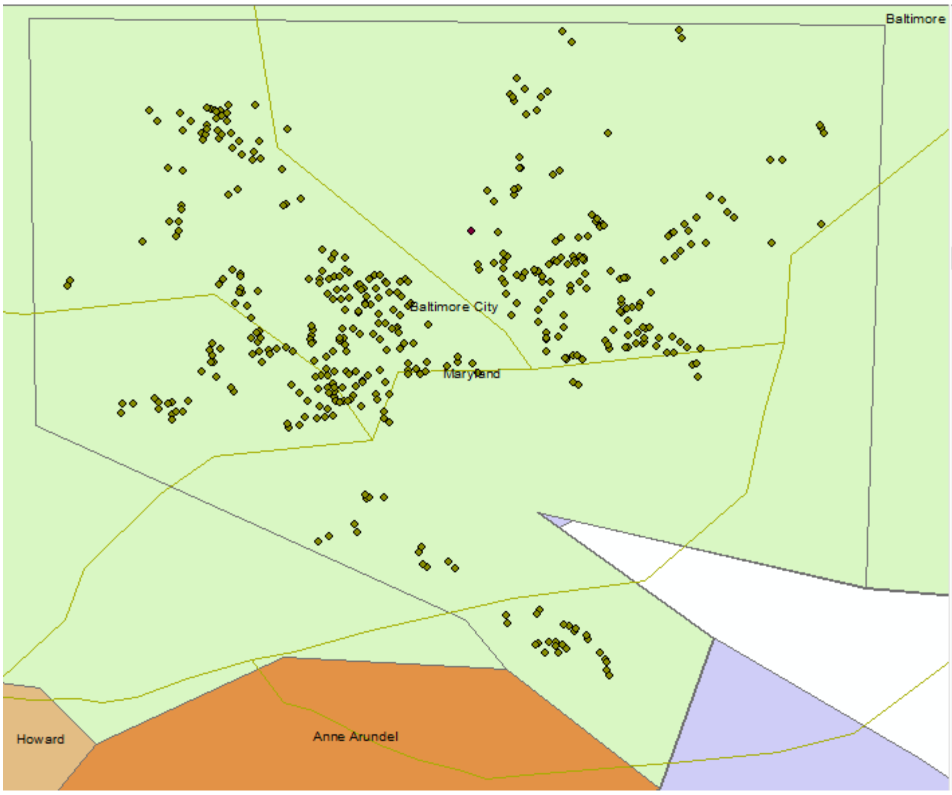
length after 369

Completeness: 0.141
Estimated number of clusters: 69
Estimated number of noise points: 75
Homogeneity: 1.000
V-measure: 0.246
Adjusted Rand Index: 0.060
Adjusted Mutual Information: 0.205
Silhouette Coefficient: 0.201
precision_score: 1.000

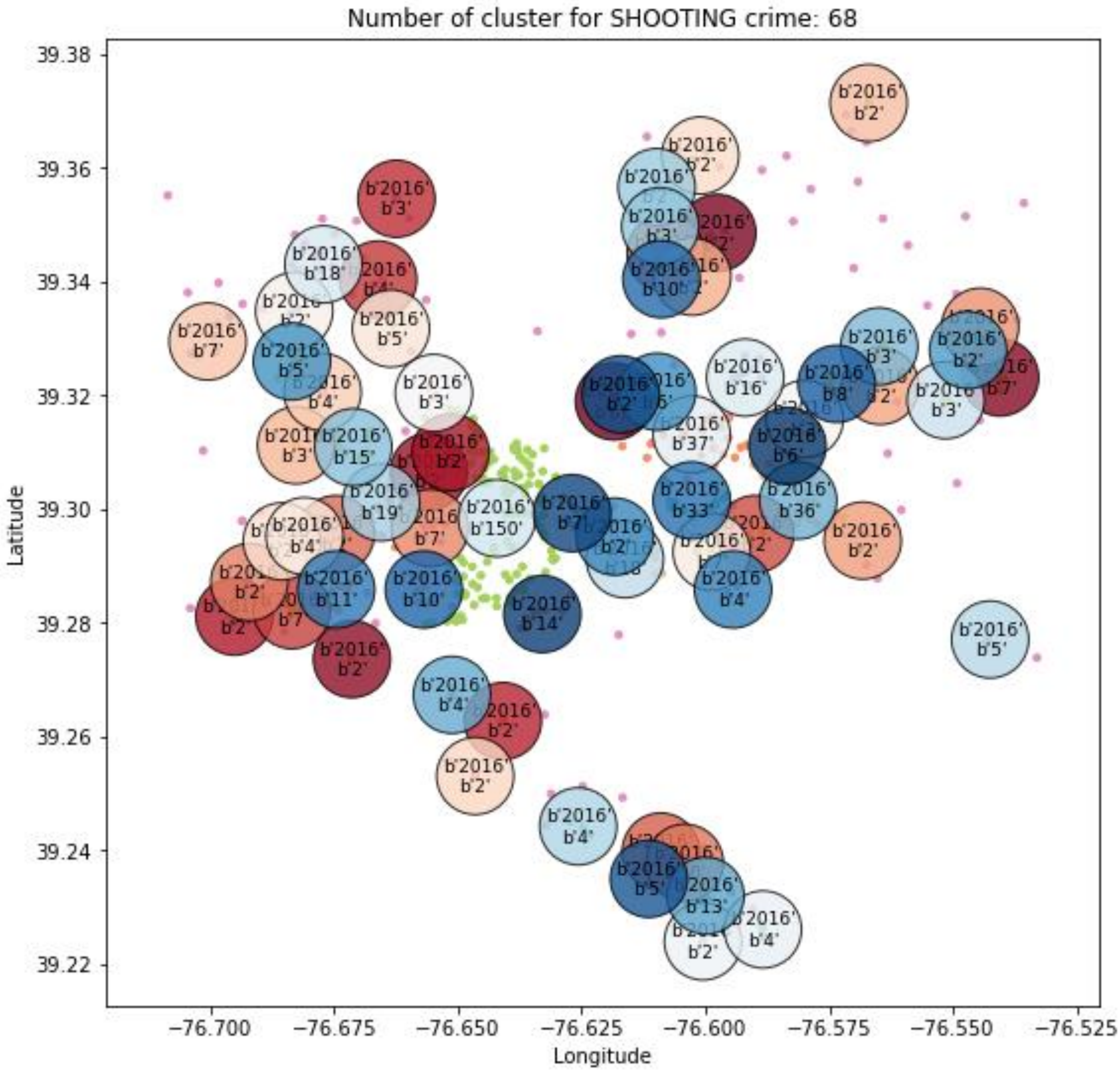

Baltimore
Baltimore City
Maryland
Howard
Anne Arundel



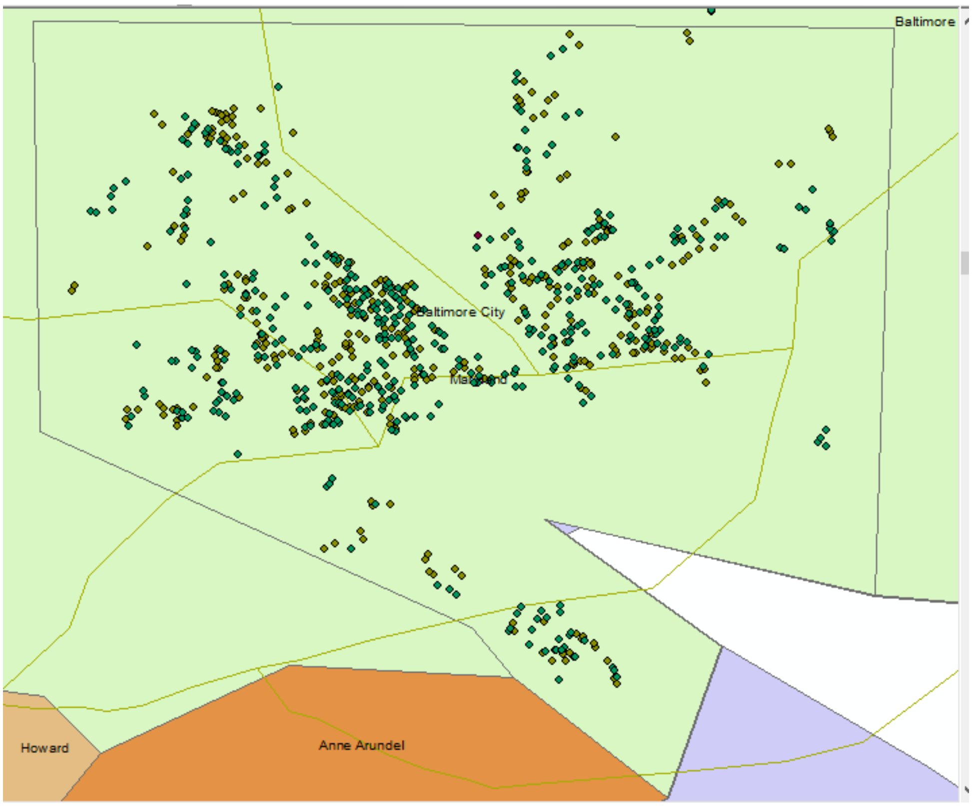
length after 634
Completeness: 0.111
Estimated number of clusters: 60
Estimated number of noise points: 68
Homogeneity: 1.000
V-measure: 0.199
Adjusted Rand Index: 0.051
Adjusted Mutual Information: 0.177
Silhouette Coefficient: 0.125
precision_score: 1.000

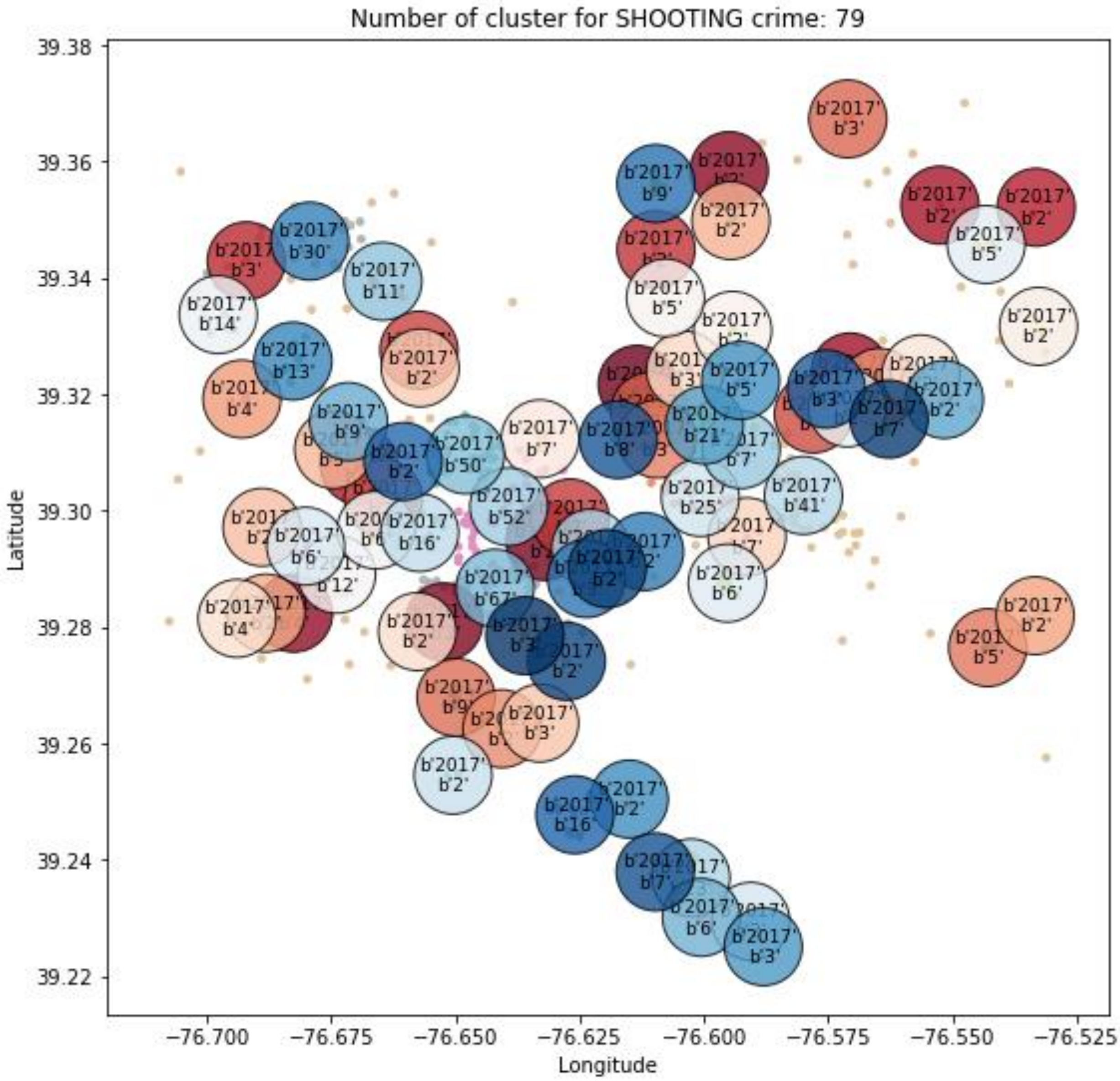


Baltimore
Baltimore City
Maryland
Howard
Anne Arundel

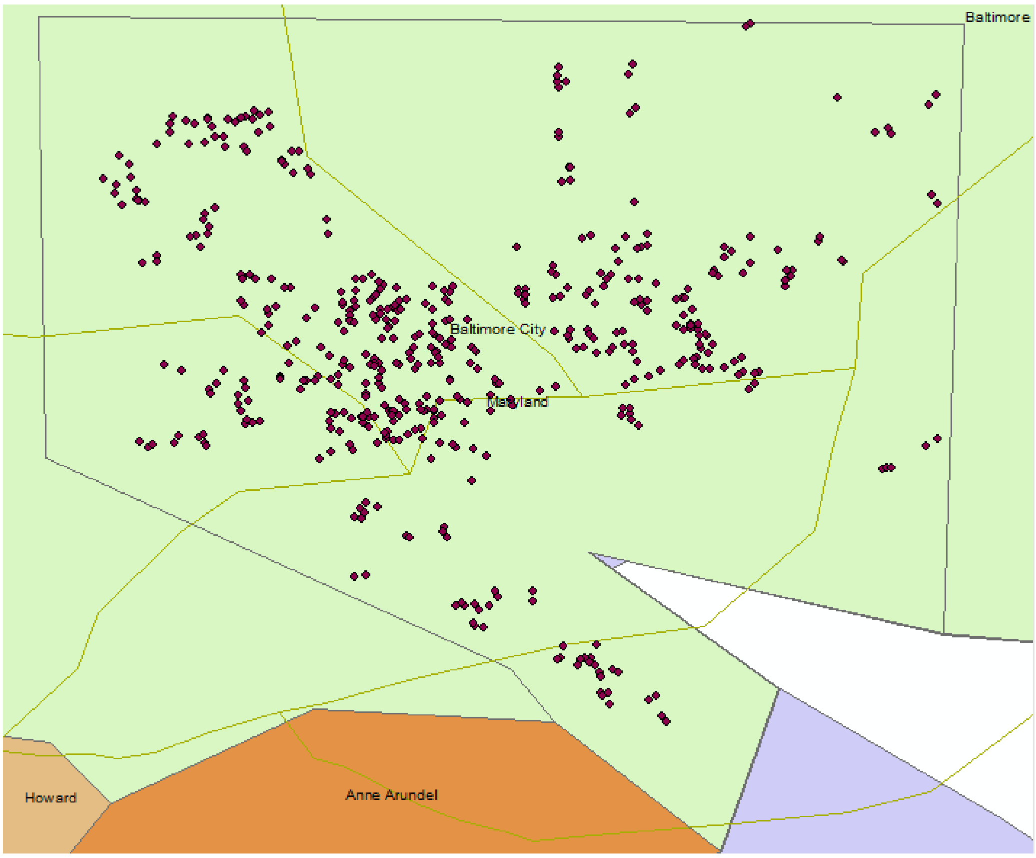


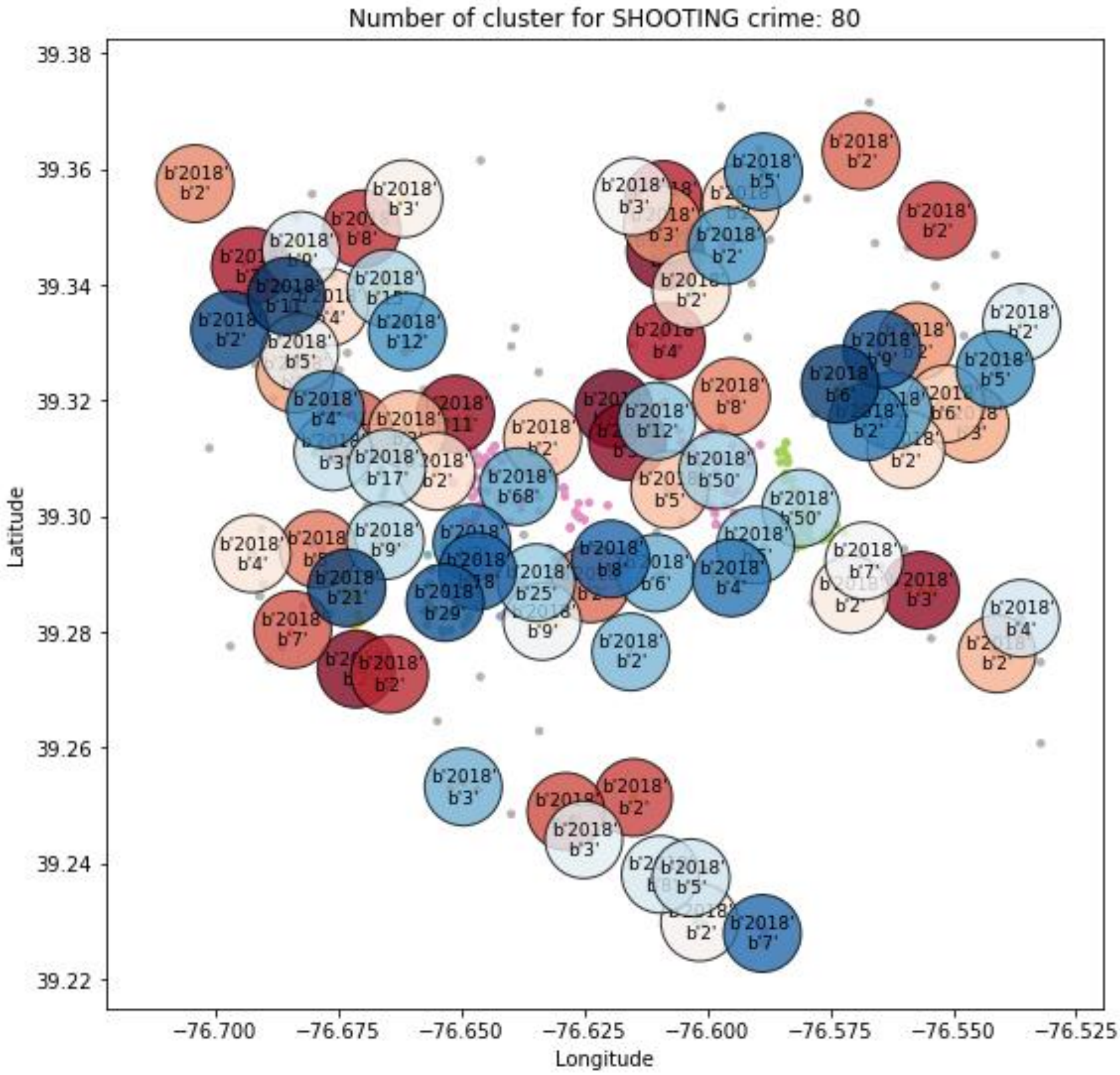
Completeness: 0.099
Estimated number of clusters: 68
Estimated number of noise points: 69
Homogeneity: 1.000
V-measure: 0.180
Adjusted Rand Index: 0.037
Adjusted Mutual Information: 0.157
Silhouette Coefficient: 0.157
precision_score: 1.000


Baltimore
Baltimore City
Howard
Anne Arundel

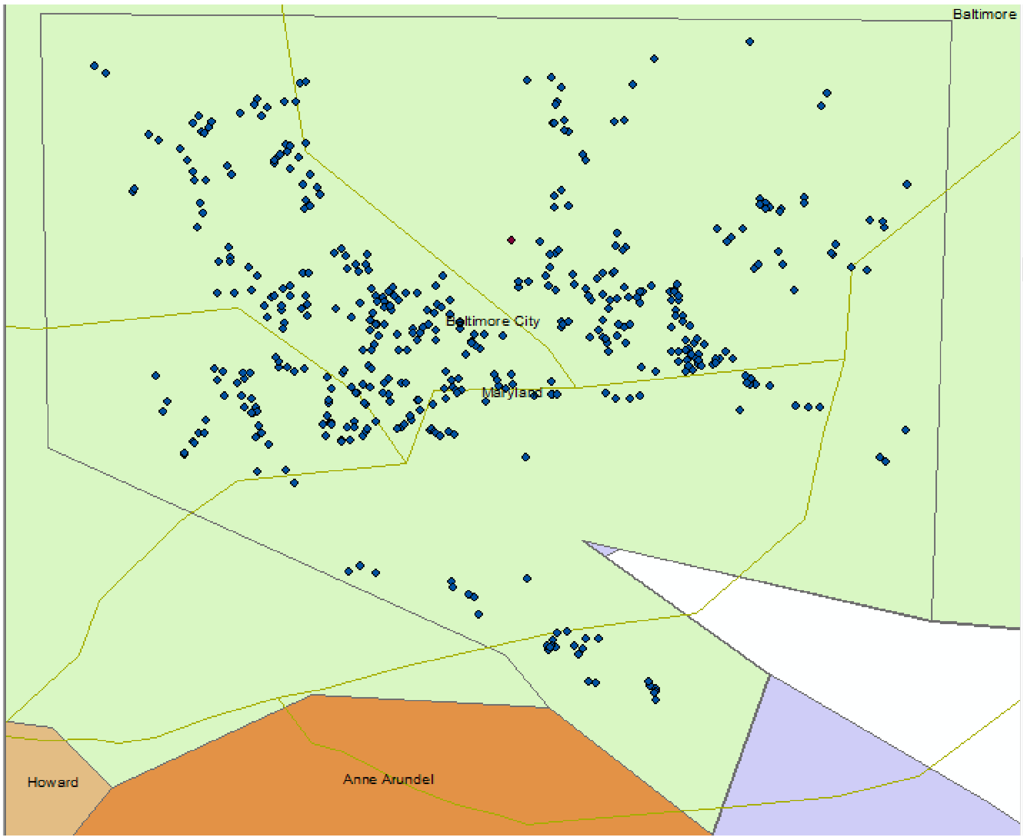


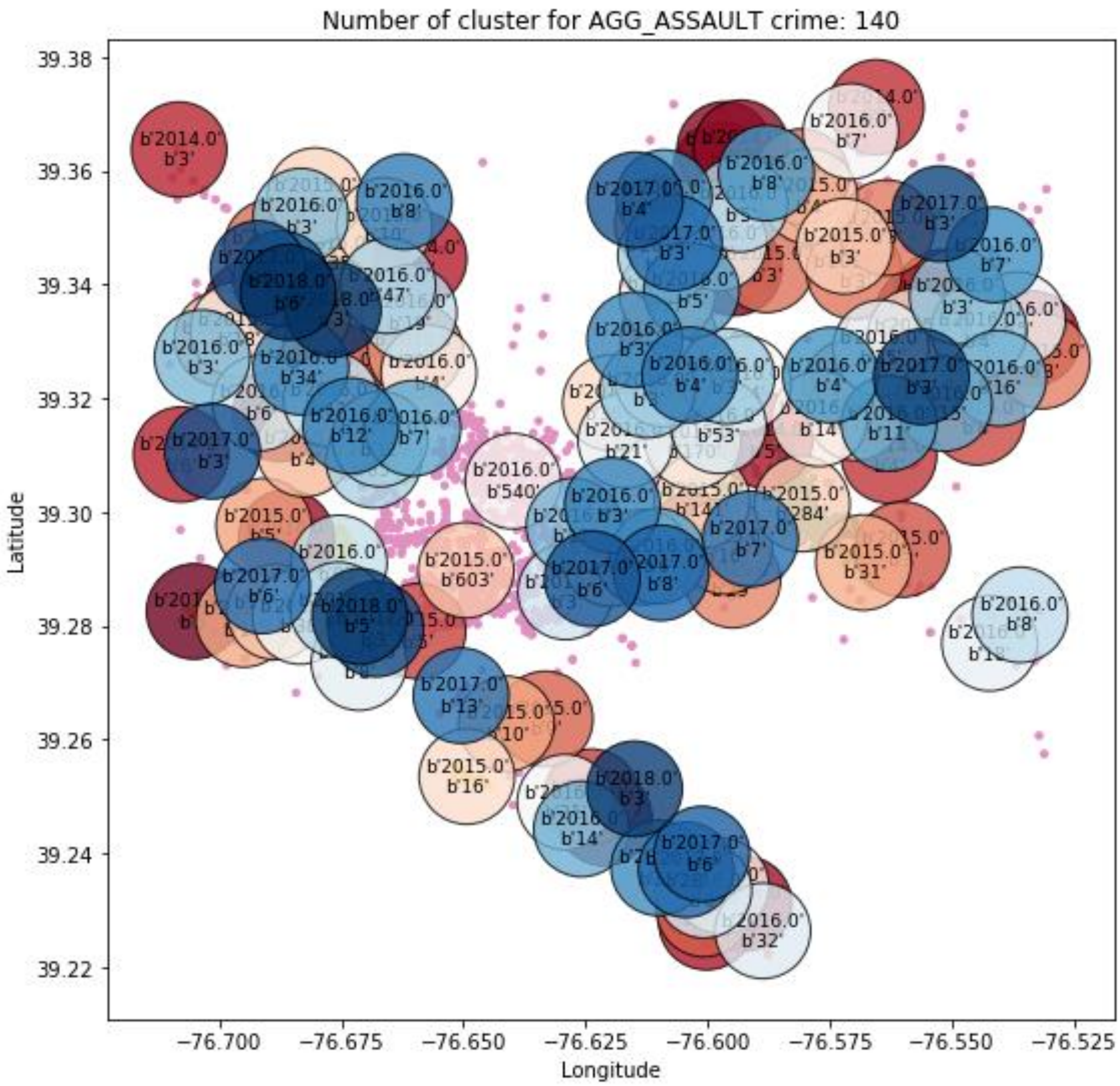
length after 703
Completeness: 0.079
Estimated number of clusters: 79
Estimated number of noise points: 61
Homogeneity: 1.000
V-measure: 0.147
Adjusted Rand Index: 0.015
Adjusted Mutual Information: 0.124
Silhouette Coefficient: 0.296
precision_score: 1.000


Baltimore
Baltimore City
Maryland
Howard
Anne Arundel

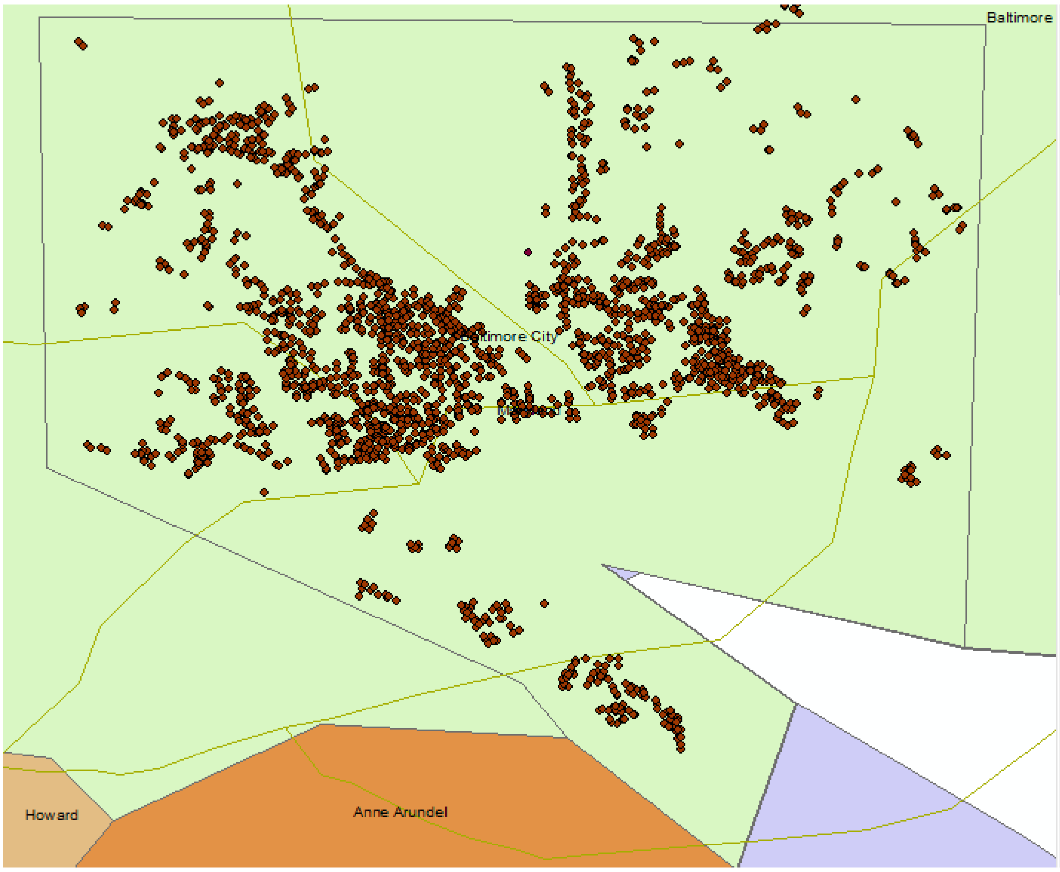


Completeness: 0.082
Estimated number of clusters: 80
Estimated number of noise points: 61
Homogeneity: 1.000
V-measure: 0.151
Adjusted Rand Index: 0.016
Adjusted Mutual Information: 0.126
Silhouette Coefficient: 0.299
precision_score: 1.000


Baltimore
Baltimore City
Maryland
Howard
Anne Arundel



Completeness: 0.083
Estimated number of clusters: 140
Estimated number of noise points: 369
Homogeneity: 0.811
V-measure: 0.150
Adjusted Rand Index: 0.034
Adjusted Mutual Information: 0.141
Silhouette Coefficient: 0.084
precision_score: 0.959



Crime from 2012-2018

Conclusion and evaluation

Crime detecting is the complex and evolving field of research in the real world environment with various Baltimore city datasets including population distributions, universities, banks, metro, library, topology, growth distributions, aimed at preventing the rates of crime from increasing. Data Mining plays a major role in the study of criminal pattern recognition and crime detection and prevention in governmental agencies for public security. This research introduces the process of identifying crime trends and predicting crimes in Baltimore cities. Clustering methods are used to detect crime and the spatial statistics procedure is used for hot spot crime prediction and most hazardous spots. Used to anticipate crimes based on the quest for similarities. The clustering of DBSCAN is incorporated and their quality is assessed on the basis of precision. The DBSCAN clustering results with high accuracy and effectiveness when comparing their performance. This system, therefore, assists government bodies in improving and accurately analyzing crime. The efficiency of the DBSCAN clustering algorithms is assessed with precision using silhouette measures and uniformity is an efficient way of mapping high density of crime-prone areas. The use of GIS / Clustering methods in Spatial-temporal analysis confirmed as an efficient means to understand the underlying correlation between incident.

This work can be extended in the future to enhance clustering and applying certain algorithms and statistical techniques algorithms to more efficiently and effectively classify criminals. It can also enhance privacy to protect the crime dataset and criminal record that are provided by police stations using a Wearhouse protected with security Measures.

Finally, it is also recommended that there should combine the Artificial intelligent and GIS to help the decision making fast and accurate to public safety.

Pseudocode

Algorithm DBSCAN (set of points D, Eps, MinPts)

```
ClusterId = label of a first cluster;
for each point p in set D do
      if (p.ClusterId = UNCLASSIFIED) then
          if ExpandCluster(D, p, ClusterId, Eps, MinPts) then
               ClusterId = NextId(ClusterId)
          endif
      endif
endfor
```

Function ExpandCluster(D, point p, clustedr label ClId, Eps, MinPts)

```
seeds = Neighborhood(D, p, Eps);
if |seeds| < MinPts then
    p.ClusterId = NOISE;
    return FALSE
else
     for each point q in seeds do                  // including point p
          q.ClusterId = ClId;
     endfor
     delete p from seeds;
      while |seeds| > 0 do
          curPoint = first point in seeds;
          curSeeds = Neighborhood(D, curPoint , Eps);
          if |curSeeds| >= MinPts then
               for each point q in curSeeds do
                    if q.ClusterId = UNCLASSIFIED then
                    /* N_Eps (q) has not been evaluated yet, so q is added to seeds */
                         q.ClusterId = ClId;
                         append q to seeds;
```

```
elseif q.ClusterId = NOISE then
/* N_Eps (q) has been evaluated already, so q is not added to seeds */
    q.ClusterId = ClId;
endif
endfor
endif
delete curPoint from seeds;
endwhile
return TRUE
endif
```

References

1. Le, K.G., P. Liu, and L.-T. Lin, *Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam.* Geo-spatial Information Science, 2019: p. 1-12.
2. Valente, R., *Spatial and temporal patterns of violent crime in a Brazilian state capital: A quantitative analysis focusing on micro places and small units of time.* Applied geography, 2019. **103**: p. 90-97.
3. Mumtaz, K. and K. Duraiswamy, *An analysis on density based clustering of multi dimensional spatial data.* Indian Journal of Computer Science and Engineering, 2010. **1**(1): p. 8-12.
4. Yi, F., et al. *An Integrated Model for Crime Prediction Using Temporal and Spatial Factors*. in *2018 IEEE International Conference on Data Mining (ICDM)*. 2018. IEEE.
5. Vural, M.S., M. Gök, and Z. Yetgin, *Analysis of incident-level crime data using clustering with hybrid metrics.* GAU J Appl Soc Sci, 2014. **6**: p. 8-20.
6. Farsi, M., et al., *Crime data mining, threat analysis and prediction*, in *Cyber Criminology*. 2018, Springer. p. 183-202.
7. Isafiade, O., A. Bagula, and S. Berman, *A revised frequent pattern model for crime situation recognition based on floor-ceil quartile function.* Procedia Computer Science, 2015. **55**: p. 251-260.
8. Wang, T., et al., *NS-DBSCAN: A Density-Based Clustering Algorithm in Network Space.* ISPRS International Journal of Geo-Information, 2019. **8**(5): p. 218.
9. McClendon, L. and N. Meghanathan, *Using machine learning algorithms to analyze crime data.* Machine Learning and Applications: An International Journal (MLAIJ), 2015. **2**(1): p. 1-12.
10. Adepoju, M., et al. *Geo-spatial technologies for Nigerian urban security and crime management—A study of Abuja crime hotspot mapping and analysis*. in *Proceedings of the ASPRS 2014 Annual Conference, Louisville, Kentucky, KY, USA*. 2014.
11. Yar, P. and J. Nasir, *GIS Based Spatial and Temporal Analysis of Crimes, a Case Study of Mardan City, Pakistan.* international Journal of Geosciences, 2016. **7**(19): p. 325-334.