

Geospatial Clustering of Psychotropic Substances Crime Locations

Kim YS^{1*}, Crump E²

DOI:10.5281/zenodo.16964389

^{1*} Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, United States of America.

² Erin Crump, Department Head, Data Analytics Department, Bridgerland Technical College, United States of America.

The worldwide prevalence of drug overdose and the misconception on psychotropic substances lead to the increased incidents of drug use disorders, drug offences and environmental harms along with financial burden on local and federal government for drug control and prevention. As a small step to reduce drug-related offences, we analyze the data sets consisting of drug- or alcohol-related crime incidents to discover temporal and seasonal patterns of such crimes. More importantly, we employ a density-based clustering algorithm to find a natural grouping of the geographic locations of crime incidents based on their longitude and latitude information. By visualizing such clusters with major crime types for each cluster, we allow residents and public safety officers to easily identify hot spots of drug-related crimes and hence develop new prevention plans to cope with drug-related crimes.

Keywords: Clustering, DBSCAN, Drug Crimes, Geospatial Analysis, Psychotropic Substances

Corresponding Author	How to Cite this Article	To Browse
Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, United States of America. Email: yong.kim@usu.edu	Kim YS, Crump E, Geospatial Clustering of Psychotropic Substances Crime Locations. Int J Engg Mgmt Res. 2025;15(4):45-55. Available From https://ijemr.vandanapublications.com/index.php/j/article/view/1785	

Manuscript Received 2025-07-05	Review Round 1 2025-07-24	Review Round 2	Review Round 3	Accepted 2025-08-10
Conflict of Interest None	Funding Nil	Ethical Approval Yes	Plagiarism X-checker 2.42	Note
 © 2025 by Kim YS, Crump E and Published by Vandana Publications. This is an Open Access article licensed under a Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/ unported [CC BY 4.0]. 				

1. Introduction

According to the 2024 World Drug Report released by the United Nations Office on Drugs and Crime (UNODC), more and more people use drugs for medical and non-medical purpose, reaching 292 million people worldwide in 2022, a 20 per cent increase over 10 years. Another online source by the National Center for Drug Abuse Statistics (NCDAS) reported that, in USA alone, over a half of people 12 and older have used illicit drugs at least once, 700,000 people died from drug overdose since 2000, and the federal government spent \$35 billion in 2020 alone for drug control and prevention.

To exacerbate the current drug-related problems that many countries face, the emergence of new synthetic drugs with mass manufacturing facility, a record high level of drug trafficking, and liberal legalization of drugs are negatively impacting people's health and well-being. For example, cannabis has been the most commonly used drug (228 million users) and its dominance is likely to continue due to the recent legalization of its production and sale for non-medical use in Canada and over a half of jurisdictions in the US. As expected in advance, more people (especially among young adults) in these two countries suffered from over (or even regular) use of cannabis, leading to psychiatric disorders and attempted suicide since its legalization (NCDAS, 2020).

It is also noteworthy that more than 16 million people in US misuse prescriptions for painkillers, opioids and sedatives in a year, which leads to a high rate of drug abuse, addiction and overdose (NCDAS, 2020). At the same time, a growing number of people are exploring the medical use of psychotropic substances to treat some mental health disorders without clinical and scientific guidelines. Note that psychotropic substances are chemical substances that can affect people's cognitive ability (consciousness and cognition), emotion (perception and mood), or behavior.

The afore-mentioned prevalence of drug overdose and the misconception on psychotropic substances naturally lead to drug use disorders, drug offences and environmental harms. For example, about 64 million people out of 292 million who used drugs worldwide are estimated to suffer from drug use disorders, only less than 10 per cent of them is currently in treatment (UNODC, 2024).

Further, in 2022, about 7 million people were arrested for drug use or possession for use globally, and over 1.6 million were convicted for drug offences (UNODC, 2024). At the same time, these pervasive illicit activities exacerbate environmental degradation due to the increased dumping of toxic and chemical wastes.

Therefore, there exists a strong need to develop actionable prevention and intervention policies towards drug-related crimes with better understanding on drug usage patterns, while concurrently providing much needed treatment and support to people affected by drug use. Personal communications with police officers in a local community also revealed that top 3 most frequent reasons for adult arrest include the possession of illegal drugs such as Marijuana. In particular, they were greatly concerned about a sharp increase (24.3%) of Marijuana possession from the past year and hence liked to find ways to curve drug-related crimes.

Strongly motivated from communications with local law enforcement officers, the authors are particularly interested in identifying seasonal patterns and geospatial characteristics of hot spots of drug-related crimes. Identifying seasonal and geospatial characteristics of drug-related crimes can be very critical for law enforcement agencies and state/federal governments. In particular, the authors pay close attention to identifying geospatial characteristics of locations associated with frequent drug-related crimes to provide complementary information to findings from seasonal and temporal patterns of drug-related crimes.

We believe that seasonal and temporal patterns of drug-related crimes provide a macro-level understanding of crime incidents in terms of when (i.e., which seasons, which day of the week, which year, and so on) drug-related crimes occur frequently. However, identified geospatial characteristics of locations that appear to be hot spots of crimes can provide useful information at a micro-level (i.e., which crime is most likely occur at a specific location) so that administrators in local law enforcement office develop efficient and effective prevention plans.

It is very reasonable to assume that while all drug-related crimes may occur more likely in a specific season, different kinds of drug-related crimes are more likely to occur at different locations.

That is, it is very possible that geographical locations associated with frequent illicit drugs use or possession of equipment to consume illicit drugs can be different from locations of frequent traffic violations under influence of drugs or alcohol beverages. Once such locations are tagged for each kind of drug crimes, local law enforcement office may reconfigure daily patrol paths with officers with more experiences and appropriate equipment for specific kinds of drug crimes. Such information will be also very valuable for public use to avoid dangerous areas for safety reasons.

In this paper, we intend to collect, analyze, and visualize crime data sets collected through incident-based reporting (IBR) by local police and sheriff's offices from 2011 through 2019 in a local community. In our definition of drug-related crimes, three types of drug-related crimes are included. The first type is Drug category that contains all crime incidents related to (suspicious) drug use and possession of equipment to consume drugs. The second type included is Liquor category such as alcohol offense and intoxication. The reason that we include alcohol offense and intoxication as drug-related crimes is that very high levels of alcohol in the body will significantly impair areas of human brain related to cognition, emotion, and behavior, which is very similar to symptoms of drug overdose. The last type considered is Traffic violation associated with driving under influence (Traffic-DUI) of drugs and alcohol.

Once we finalize the data sets of drug-related crimes, we like to review and visualize temporal and seasonal patterns of crime incidents. For advanced analysis, we intend to find a natural grouping of the geographic locations of crime incidents using clustering analysis. To this end, we employ a density-based clustering algorithm (DBSCAN), which will find geospatial clusters based on longitude and latitude information of crime locations. Finally, we will profile each cluster based on prevalent crime types within the clustered location. The visualization of such clusters on a map will make it easy for residents and public safety officers to recognize hot spots of drug-related crimes.

2. Literature Review

In this section, we like to briefly review what clustering analysis is and introduce several well-known clustering algorithms.

Clustering analysis (or a type of unsupervised learning in machine learning field) is an exploratory data analysis to group data points into clusters so that data points in the same cluster are similar while data points in different clusters are dissimilar. While grouping data points into clusters itself is an important task, clustering outcomes are often used as an input for further analysis. For example, clustering analysis has been widely adopted by marketing specialists to form the clusters of millions of customer records into a manageable set of clusters and identify the characteristics of each cluster (Schaffer & Green, 1998). Then tailored promotion campaigns can be developed and implemented for each cluster of customers. Similarly, different marketing strategies can be employed for customer clusters with unique behavior characteristics in terms of brand loyalty and price sensitivity.

Out of numerous distance-based clustering algorithms proposed in the past several decades, one of the most popular clustering algorithms is K-Means algorithm (Hartigan, 1975). The K-means algorithm starts with a random initial partition of data sets and iteratively recalculates the centroids of each partition after assigning each data point to the nearest partition measured by distance metrics between a data point and the centroid values of the partition until a convergence criterion is met (Kim et al., 2002). Note that K-Means algorithm requires the number of clusters to be specified in advance. While K-Means algorithm is very fast and intuitive to understand, its performance can be severely impacted by the initial random partitions (Babu & Murty, 1993). In addition, it may find the local optimum when data sets do not meet the assumption that data points in each partition follow spherical Gaussian distributions (Krishna & Murty, 1999).

The expectation maximization (EM) algorithm is a representative clustering algorithm based on statistical density theory. It assumes that all data points in the cluster are drawn from one of several given distributions and hence the goal is to identify the parameters of each distribution from the given data (Dempster et al., 1977). Note that the initial parameter estimates are iteratively updated based on the likelihood that data points in the cluster is drawn from a specific density function.

The EM algorithm is known to outperform other distance-based clustering algorithm (e.g., K-means) because of its applicability on both continuous and categorical data sets (Meila & Heckerman, 1998). In addition, the EM algorithm is very unique in a sense that it assigns the likelihood of the data point being in those clusters and hence each data point may belong to multiple clusters (i.e., soft clustering) while most other clustering algorithms (e.g., K-Means) assign each data point to only one cluster (i.e., hard clustering).

Another clustering algorithm group is hierarchical clustering algorithm that returns a hierarchical series of nested clusters called Dendrogram. In hierarchical clustering algorithm, there are two algorithms: Agglomerative clustering and Divisive clustering. In Agglomerative clustering, each data point forms its own cluster and, at every iteration, two nearest clusters are merged until one cluster is formed (Griffiths et al., 1984). In contrast, Divisive clustering initially assumes that all data points belong to a single cluster and recursively divide dissimilar data points into smaller clusters until each data point becomes its own cluster. While hierarchical clustering shows the hierarchical structure of the data sets, the computational cost in terms of CPU and memory requirement is relatively high and the resulting outcomes is often dependent on initial conditions for implementations (Murtagh & Contrera, 2012).

The last category of clustering algorithms is density-based clustering algorithm. In this category, the density-based spatial clustering of applications with noise (DBSCAN) is one of the most well-known and cited algorithms. The DBSCAN uses density measure as clustering criterion, and hence identifies data points with many nearby points (= a high dense region) as clusters and points in low density regions as outliers.

The key idea of the DBSCAN is that for each point to belong to a cluster, the neighborhood of a given radius (controlled by *eps* parameter) has to contain at least a minimum number of points (controlled by *minPts*). Therefore, the DBSCAN identifies each data point as a core point if it has more *minPts* within *eps*, a border point if it has fewer than *minPts* within *eps* but it is in the neighborhood of a core point, or a noise point if it is not a core point or a border point (Ester et al., 1996).

Thus, the DBSCAN starts with a randomly chosen data point and if it is a core point, all points within *eps* from it (= directly reachable points) belong to the same cluster of the core point. Then this cluster is expanded by recursively including points that are within *eps* from the closest core point in the current cluster (= reachable points). However, reachable points, which by definition is not a core point, cannot be used to reach more points. All points not reachable from any other point are outliers or noise points.

The DBSCAN has many advantages over other clustering algorithms. For example, unlike K-Means and EM algorithm, the DBSCAN can easily capture clusters of complex shapes based on density and non-linearly separable clusters. In addition, the DBSCAN can perform well on noise data sets and can be useful as outlier detector. However, it may perform poorly on data sets with large differences in densities and the quality of its performance is dependent on two parameters values, *eps* and *minPts*, which can be difficult to determine their values in advance.

In this study, we will apply the DBSCAN to group data points representing geospatial locations of crime incidents into clusters to find hot spots of various crimes along with geographical characteristics of such clusters. In particular, we will execute DBSCAN analysis with longitude and latitude values of crime incident locations in Tableau software through tabpy, an interface that allows us to exploit clustering libraries in Python from Tableau.

3. Data Sets and Local Community Background

The initial data set for this study, Cache County Sheriff Police Crime Data, was downloaded from Utah Open Data (<https://www.utah.gov/government/data.html>), which provides open access to publicly available datasets. According to meta data source, this data set contains the 15,000 rows of public safety related crime records in a local community between the year 2011 and 2018. We first present the full list of variables in Table 1.

Table 1: List of Variables

Variable	Description
case_number	Text type record ID
address	Street address of crime incident
city	City of crime incident
state	State of crime incident
latitude	Latitude value of crime incident
longitude	Longitude value of crime incident
time_created	Time the incident record created
time_updated	Time the incident record updated
time_incident	Time stamp of crime incident
day_of_week	Day of the week of crime incident
hour_of_day	Hour (0 - 23) of crime incident time
incident_description	Detailed subcategory of crimes. Same as incident_type_primary
incident_id	Numerical type record ID
parent_incident_type	Main category of crimes
incident_type_primary	Detailed subcategory of crimes

During our preliminary inspection on the data set, we found few erratic records (e.g., nine crime records whose incident year of crime was earlier than 2011), which we removed from our analysis. While we found it weird to observe only two records in the year 2013, we decided to keep them because they had complete information and it is not our main goal to study yearly trends of crimes.

The brief introduction to the local community, Cache County in Utah, USA, is necessary to understand and complete geospatial specific analysis. The authors acknowledge that the following information about Cache County is borrowed from two Web sites: Cache County official site (<https://www.cachecounty.gov/>) and Wikipedia (https://en.wikipedia.org/wiki/Cache_County,_Utah) solely for the purpose of self-containment of this paper.

The Cache County is located in the northern region of Utah state bordering Idaho state. Historically, it was first formed in 1856 and currently has 13 cities such as Logan (county seat and largest city), Hyrum, North Logan, Smithfield, Nibley, Wellsville, Amalga, and so on. It also hosts two public universities, Utah State University (USU) and Bridgerland Technical College, in Logan.

The population in Cache County in 2023 was estimated to be 142,393 and male and female groups almost evenly represent the population. However, the largest population came from the age group of [18, 64] (60.6%) followed by young group (<18, 29.3%) and senior group (>=65, 10.2%).

In terms of racial makeup, the dominant race group was White (> 80%) followed by Hispanic (11%) and many other races. With the median income of \$60,530 for a household, about 24% of the population and 9.3% of families were below the poverty line. In addition, about 63.4% and 36.6% of households were owner-occupied and renter-occupied, respectively.

4. Findings and Discussion

4.1 Analysis on All Crime Incidents in Cache County

We first plotted all crime incident records in Cache County on a map using Tableau software in Figure 1. According to Figure 1, crime incident records mostly centered around its county seat, Logan, while some portions of incident records were scattered in neighboring cities such as Hyrum, Nibley, Providence, North Logan and Smithfield.

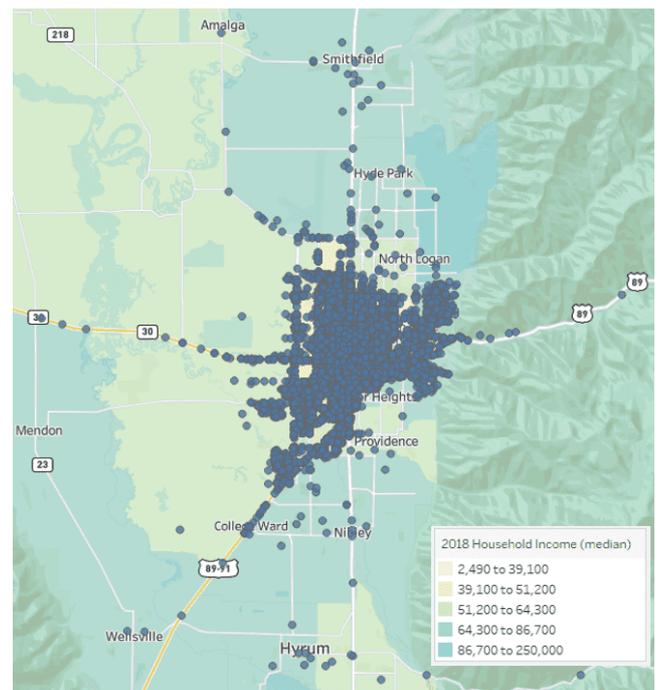


Figure 1: All Crime Incidents in Cache County

We also carried out a simple descriptive analysis to find most frequently occurred crime types in this local community. Note that this data set divides all records into 24 parent incident types (= parent_incident_type variable in Table 1), each of which contains further detailed subcategories (= incident_type_primary variable) of crimes.

Our analysis revealed that in terms of parent crime types, top 10 most frequently occurred parent crime types were Community Policing (43.4%) followed by Traffic (23.2%), Theft (5.6), Disorder (4.9%), Alarm (4.7%), Property Crime (4.0%), Drugs (2.5%), Pedestrian Stop (1.9%), Liquor (1.9%), and Assault (1.6%). We noted that the most frequent crime type, Community Policing, was mainly related to animal problem, citizen assistance, and welfare check, while the second most frequent crime type was related to traffic accident, traffic offense, and vehicle identification number (VIN) inspection.

4.2 Seasonal Patterns of Psychotropic Substances Crimes

In this subsection, we like to focus our analysis only on psychotropic substances crime (PSC) mainly because this research was in part initiated with the purpose of collaborating with local law enforcement offices in the city of Logan that want to gain insights of geospatial relationships and devise an actionable set of recommendations to control the increasing trend of PSC. In addition, Logan city is the county seat of Cache County and hence public safety issues in Logan garner the attentions of city council members and Cache County administrators. Therefore, we analyze PSC incidents only in Logan from now on.

To this end, we selected all records in two crime types, Drugs and Liquor. In addition, we also selected records in driving under the influence (DUI) incidence type within Traffic parent crime type. Finally, crime incidents only in Logan were filtered. The finalized data set contains a total of 767 PSC incidents in the city of Logan.

We graphically presented the distributions of PSC incidents in Logan by crime subtypes within their parent crime types in Figure 2. According to Figure 2, the largest portion (47.9%) of PSC incidents is attributed to Drugs, while remaining PSC incidents are attributed to Liquor (35.7%) and Traffic-DUI (16.4%) type.

Next, we intend to see if there are any seasonal patterns in PSC incidents and if so, we can find any insights from such findings. To this end, we plotted all PSC incidents across observed years and presented the outcome at the top panel of Figure 3.

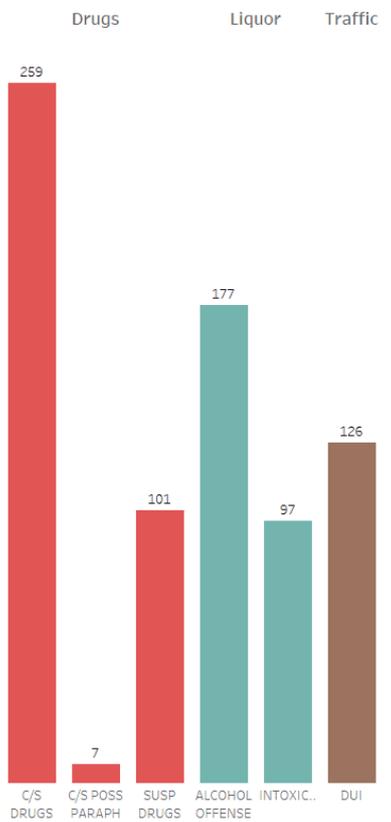


Figure 2: PSC Statistics in Logan City

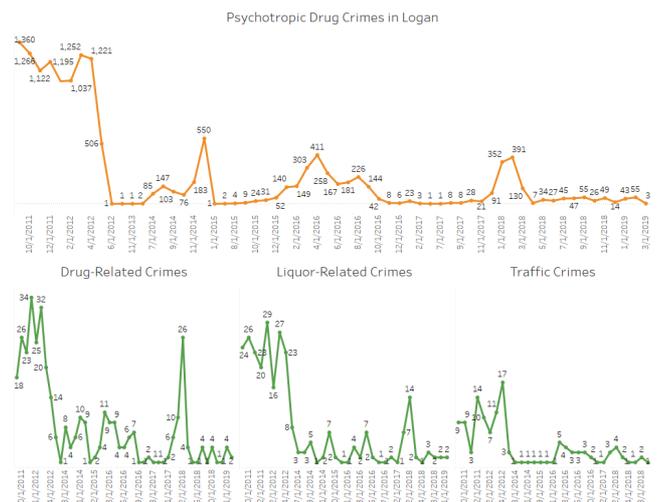


Figure 3: Seasonal Patterns of PSC in Logan City

Note that, in Figure 3, we did not pay attention to the increase or decrease of PSC over years mainly because our preliminary analysis on the initial data set revealed the erratic imbalanced distribution of records in the year 2013 due to largely missing records. We also found that there were relatively too many records for unknown reasons in the year 2011 and 2012. Therefore, we focused on incidents of PSC in each month or season for the chosen academic year (typically from late August of a year to early May of the following year).

According to Figure 3, PSC maintained a very high level of incidents in the period of September (1,360 incidents), 2011 through April (1,221), 2012. In the academic year of 2014, incidents started to increase August, peaked in November (183) and December (550), but downed to almost zero in January, 2015. During the academic year of 2015, however, incidents steadily increased from August, sharply increased from January (140) in 2016, and peaked in April (411) in 2016. Since then, it sharply decreased but still showed relatively high incidents level in May-July (over 150 incidents). In the academic year of 2017, it mainly remained stable and low until November (21), but sharply increased and maintained a very high-level during January (352) and February (391) until May (130) in 2018. The PSC incidents during the academic year of 2018 did not show a particular seasonal pattern like in previous years. Instead, it maintained a steady level of incidents between 3 and 55 over the entire year.

We also separated the seasonal patterns of Drug, Liquor and Traffic incident type and presented them at the bottom panel of Figure 3. Overall, each of Drug, Liquor and Traffic incident type presented a similar seasonal pattern we observed from the aggregated PSC category at the top panel: peaked incidents in January through March.

From these observations, we concluded that seasonal patterns of PSC incidents were closely related to academic calendars. We partially attribute this finding to the fact that the city of Logan hosts two public universities and hence it shows typical patterns observed in many other college campus towns. That is, when a new academic year starts in late August, first-time and returning students are likely to have various social meetings and hence they are most likely to be exposed to opportunity of having drinks or even psychotropic substances, and driving back to home or dormitories under the influences of alcohol.

However, this single factor alone does not fully explain why PSC incidents are the highest from January to March. After careful thought, we attribute cold winters weather in Logan city as the second factor to the high PSC incidents in winter seasons. Geographically, the Logan city is located in the highly elevated northern region (4,534 ft or 1,382 m) and has very warm and dry summers that attract many temporary residents from southern states such as Arizona and Florida.

However, it also has very long and cold winters typically starting November until March with moderate to heavy snowfall. Note that the Utah state hosted an international Winter Olympic and Paralympic Games in 2002 and was elected as the host of the 2034 Winter Olympics again. Therefore, many senior city residents seem to feel exhausted during cold winter months with heavy snow. In addition, students, faculty and staff members have to embrace for a new Spring semester starting early January after taking a short winter break and holiday seasons. So, many (college) residents with their family members and relatives are most likely to enjoy holiday seasons and greet a new year in a fun and relaxed environment with alcohol beverages. Further, long and cold winter seasons enforce most of city residents to stay inside with a very minimum physical activities, which will escalate their exhaustion and stress level even further. Ultimately, city residents who feel exhausted and stressful under unfavorable climate environments are more likely to seek for psychotropic substances including alcohol beverages, leading to higher incidents of PSC in winter seasons.

4.3 Motivation for DBSCAN Clustering Analysis

In this subsection and following subsections, we discuss why we like to run a clustering analysis with DBSCAN on PSC incidents data. Then, we present the graphical and numerical outcomes of DBSCAN clusters and share any managerial insights from public health perspectives. To this end, we first plotted all PSC incident records in Logan on a map using Tableau software in Figure 4, which indicates geographical locations expressed in longitude and latitude information on the detailed map of Logan. Overall, this map presents that PSC incidents were recorded in the most area of the city.

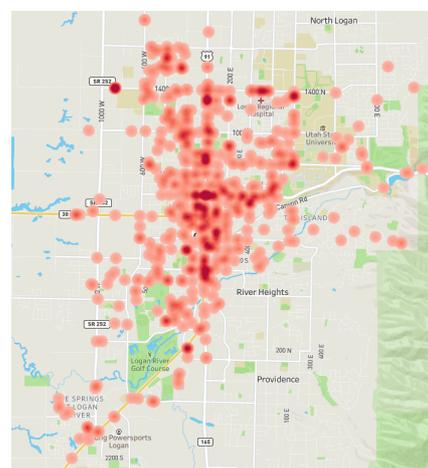


Figure 4: Geospatial Plot of PSC Incidents in Logan

Figure 4 provides one additional information of which locations of the city are hot spots of PSC incidents in darker red colors. That is, locations with darker red colors indicate multiple incidents of PSC. We immediately noted that many hot spots of PSC incidents were aligned with US highway 91 (shown along the vertical center of the map in Figure 4. Many local residents refer it as the main road) that runs through from Brigham City, Utah to Idaho Falls, Idaho. This perfectly makes sense because it is a well-known fact that the success of restaurants, bars, theaters, inns and hotels, coffee shops and many other types of business is heavily dependent on influx of more residents or travelers. Therefore, they are mostly located along the highway that passes through the center of the city because it grants residents and travelers easy access with vehicles.

However, the hottest spot of PSC incidents with most locations in darker red colors was observed at the intersection of US highway 91 and Utah State Route 30 (SR-30) (shown along the horizontal center of the map in yellow in Figure 4. Many local residents refer it as the center street). With easy access through both SR-30 and US highway 91, the intersected area of two highways indeed has served as the central location of business entities and administrative offices in the city of Logan. Therefore, it is not surprising to pinpoint this area as the hottest spot of PSC incidents.

One major limitation of findings from Figure 4 is that such findings provide only macro-level understanding of PSC incidents in the city. Therefore, they are not particularly useful for administrators in the local law enforcement office to develop efficient and effective PSC prevention plans including optimal officer allocation and patrol path configuration suited for each subcategory of PSC. For example, possibly best patrol locations for police officers to stop and arrest drivers who are suspected to commit Traffic-DUI crime may not be consistent with promising locations to spot and arrest the violators of C/S Drugs or C/S Possession of Paraphernalia (Denoted as C/S Poss Para in chart labels). Note that paraphernalia refers to any equipment to produce, conceal and consume illicit drugs.

4.4 Validation and Discussion of DBSCAN Clustering

For administrators in police and sheriff office to revise their current PSC prevention plans for each

subcategory of PSC, they would need the exact location of frequent occurrence for each subcategory of PSC. To this end, we executed DBSCAN clustering algorithm on the longitude and latitude of PSC incidents and summarized the characteristics of discovered clusters in Figure 5.

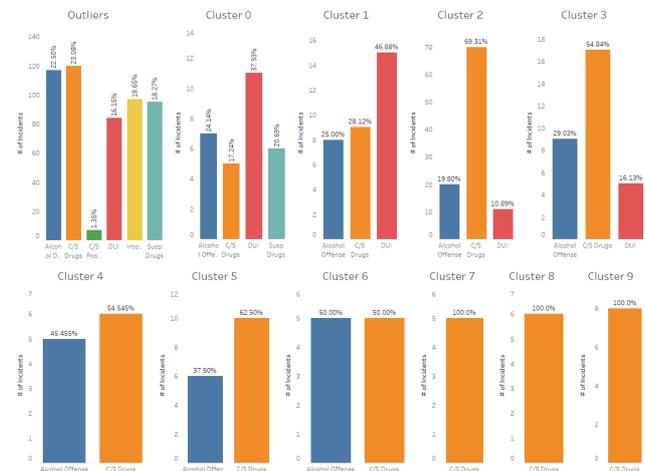


Figure 5: DBSCAN Clusters of PSC Incidents

Note that all clustering algorithms will try to put similar data points into same clusters while different data points into different clusters, minimizing differences of data points within same clusters while maximizing differences of data points across different clusters. Therefore, the first step to validate the clustering output is to inspect the characteristics of clusters.

We first noted that DBSCAN algorithm with chosen parameter values (*eps*=0.15 & *minPts*=5) discovered a total of 10 clusters and one outlier group. Unlike other clustering algorithms, the DBSCAN algorithm is known to automatically identify outliers from the data set, and, in our case, outliers are city locations with less than 5 PSC records within 0.15 distance metric value. Since data points in this group do not belong to any other clusters, this group is likely to display distinguished characteristics from other clusters.

The most outstanding characteristic of this outlier group from Figure 5 is that all PSC subcategories in the data set occupies a certain proportion ($\geq 15\%$) of records in this group except C/S Possession of Paraphernalia category (1.35%). We also noted that the composition of PSC subcategories of outlier group is somewhat similar to that of entire PSC data shown in Figure 2.

Cluster 0 in Figure 5 displays somewhat balanced distributions of four subcategories, DUI (38.0%) followed by Alcohol Offense (24.1%), Susp(ected) Drugs (20.7%) and C/S Drugs (17.2%). Cluster 0 can be also characterized with three main categories, Traffic-DUI (38.0%), Drug (37.9%) and Liquor (24.1%). Cluster 1 display a very similar distribution found in Cluster 0. Single dominant

PSC category in Cluster 1 is Traffic-DUI (46.9%) followed by Liquor-Alcohol Offense (25%) and Drugs (28.1%). In short, both Clusters 0 and 1 are commonly characterized as Traffic-DUI lead clusters while the second and third dominant categories in their clusters are different.

To inspect geospatial hot spots of Cluster 0 and Cluster 1, we plotted all data points in each cluster on a map and showed them in Figure 6. Two main observations were noted from Figure 6. First, since DBSCAN algorithm separated these two clusters, their geospatial locations were not overlapped. In particular, two hot spots in the north-west area (i.e., one near Bridgerland College and another residential area around 1800 North and 500 West) were marked for Cluster 0 while two hot spots in the north (i.e., commercial area on 1400 North) and east area (i.e., inside of USU campus near Dee Glen Smith Spectrum) were identified for Cluster 1. However, some of their hot spots were located close around the central commercial locations of the city mainly because Traffic-DUI incidents category is a dominant lead in both clusters.



Figure 6: PSC Hot Spots of Cluster 0 & 1 in Logan

Cluster 2 and Cluster 3 in Figure 4 shared a common characteristic in a sense that Drug category is a dominant category (69% and 55%, respectively) followed by Alcohol and Traffic-DUI categories with slightly different values. We also presented geospatial hot spots of Cluster 2 and Cluster 3 in Figure 7.



Figure 7: PSC Hot Spots of Cluster 2 & 3 in Logan

In Figure 7, geospatial hot spots of Cluster 2 and 3 were well separated. Hot spots of Cluster 2 were all clustered along the US Highway 91 (the main road) between 1400 North and 600 South except the segment of 600 North and 1000 North. Interestingly, one hot spot of Cluster 3 just filled in the missing US Highway 91 segment of Cluster 2. This observation makes sense considering the fact that both clusters commonly represent geospatial locations of Drug related crimes.

At first, one hot spot of Cluster 3 in the east region around 500 North and 600 East was unexpected mainly because this region is far away from central commercial regions along US highway 91. However, multiple personal visits to this region revealed that it is within a walking distance from USU dormitories and buildings, privately owned rental houses and apartments, and street parking lots, indicating noticeable heavy traffic of students and nearby local residents.

Next, we considered Clusters 4, 5 and 6 together in Figure 5 mainly because these clusters commonly contain two PSC categories, Drug ($\geq 50\%$) and Alcohol Offense (between 37.5% and 50%). Remaining three clusters, Clusters 7, 8, and 9 were again grouped together because all three clusters are represented purely by Drug category crimes (100%). We presented hot spots of these two cluster groups in Figure 8.

Figure 8 presented well separated geospatial hot spots of Cluster 4-5-6 and Cluster 7-8-9. One hot spot of Cluster 4-5-6 was located along the US Highway 91 (around 400 North). This location hosts various business entities including popular fast-food restaurants (In-N-Out Burger and Kentucky Fried Chicken) and UPS store.

In addition, this location is within a walking distance from Days Inn and Japanese restaurant, thus constantly attracting many people. Three hot spots of Cluster 4-5-6 along 1400 North also attract many people due to easy access for various business entities including Walmart (one of the largest shopping mall chains in USA), numerous restaurants and Intermountain Health Hospital.



Figure 8: PSC Hot Spots of Cluster 4-6 & 7-9

One hot spot (located at 200 East and 900 North) of Cluster 7-8-9 in Figure 8 was particularly intriguing to the authors mainly because this is where a middle school is located. While it is unknown whether violators on this hot spot were students in this school, school administrators and police and sheriff department are strongly encouraged to educate students in the school and neighborhood district about the danger of using psychotropic substances because its impact on young children tends to be more detrimental and lasts longer.

5. Conclusion

This study intends to improve the public safety in a local community by locating multiple hot spots of various crimes related to the usage of psychotropic substances. In particular, this study employs DBSCAN clustering algorithm to cluster geographical locations of such crimes and profile clusters based on dominant crime types.

To this end, we validated all 10 clusters in terms of non-overlapped geospatial locations and provided insights on what geospatial characteristics and business environmental features make each identified cluster become the hot spots of psychotropic substances crimes. Note that aforementioned findings and discussion are based on data sets extracted only from a specific local community.

However, our analysis frame work and methodology can be easily applied to other communities if necessary.

In immediate future, we like to share our findings with administrators of local police and sheriff offices so that they may use our findings to develop new patrol routines for police officers with augmented information of where and what kind of crimes types are prevalent in their patrol districts.

References

- [1] United Nations Office on Drugs and Crime. (2024). <https://www.unodc.org/unodc/en/data-and-analysis/world-drug-report-2024.html>.
- [2] National Center for Drug Abuse Statistics. (2020). <https://drugabusestatistics.org/>.
- [3] Hartigan, J.A. (1975). *Clustering algorithms*. New York: Wiley.
- [4] Kim, Y., Street, W.N., & Menczer, F. (2002). Evolutionary model selection in unsupervised learning. *Intelligent Data Analysis*, 6(6), 531-556.
- [5] Krishna, K., & Murty, M.N. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 29(3), 433-439.
- [6] Babu, G.P., & Murty, M.N. (1993). A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm, *Pattern Recognition Letters*, 14(10), 763-769.
- [7] Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1-38.
- [8] Meila, M., & Heckerman, D. (1998). An experimental comparison of several clustering and initialization methods. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 386-395.
- [9] Schaffer, C.M., & Green, P.E. (1998). Cluster-based market segmentation: Some further comparisons of alternative approaches, *Journal of the Market Research Society*, 40(2), 155-163.
- [10] Griffiths, A., Robinson, L.A., & Willett, P. (1984). Hierarchic agglomerative clustering methods for automatic document classification. *Journal of Documentation*, 40(3), 175-205.

[11] Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *WIREs Data Mining and Knowledge Discovery*, 2(1), 86-97.

[12] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). Simoudis, E., Han, J., & Fayyad, U.M. (eds.). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp.226–231.

Disclaimer / Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Journals and/or the editor(s). Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.