2024

Predictive Methods and Data Pattern Analysis for Reducing Car Plate Theft

Noor Alzayani
nma5112@rit.edu

Predictive Methods and Data Pattern Analysis for Reducing Car Plate Theft

by

Noor Alzayani

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

2024

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Thesis Approval

Student Name**: Noor Alzayani**

Graduate Capstone Title**: Predictive Methods and Data Pattern Analysis for Reducing Car Plate Theft**

Graduate Thesis Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr.Hammou Messatfa

Date:

Member of committee

Acknowledgements

I wish to express my gratitude to everybody who has helped me to accomplish the research assignment. I would like to thank my supervisor, Dr.Hammou Messatfa, for the meaningful advice, encouragement, and leadership he had granted me both while conducting my research. The knowledge and experience of Dr.Hammou Messatfa is vital for making decisions and developing the research program.

I would like to commend the faculties of Dubai Police who have helped me with my studies and academic support. The honest feedback and criticism they gave, were really helpful in polishing the thesis and adding some finesse.

However, I would like to acknowledge my family and friends for their continued and remarkable support and their adjustment to my hard circumstances. From them, I receive love and support which never stop giving me the motivation needed.

Eventually, I wanted to show my gratitude to all the people who kindly devoted their time and knowledge to the research project. This research work, had it not been for their cooperation, would not have been realized.

Thank you to everyone who took part in this thesis in any capacity. Your note was truly lovely and I cannot tell you just how much I valued your words.

Abstract

The project titled " Predictive Methods and Data Pattern Analysis for Reducing Car Plate Theft" seeks to provide innovative solutions to combat car plate theft. The project contains data of 6500 thieves who have stolen number plates and were involved in various other types of criminal activities by putting that number plate in their vehicle. The data collected for the present project is from the emirates of Dubai. It contains information related to thieves age, education, nationality, type of crime committed, area in which crime is committed, number of crimes committed, timing of crime, residential status of criminal, visa status, mode of entry in Dubai etc. This project combines data analysis and predictive modeling to identify patterns of crimes that occur according to each region, nationality, ages, employment statement and, potential hotspots. The data collected is analyzed with the help of SPSS and the relationship between different variables were tested in order to see the association between them. The predictive model has been developed to predict the potential theft areas and their occurrence.

By doing so, it aims to reduce the occurrence of this crime, which obstructs the progress of police and investigation work, and also enhances the ability of law enforcement agencies to address relevant issues and make appropriate decisions according to the predictions extracted for each model.

In addition, this project will enable Dubai police to restore security and safety in the Emirate of Dubai, and the decrease in the numbers of committing the crime of stealing license plates will speed up investigations into other crimes for which criminals have used the theft of license plates as a cover and make it difficult to reach the truth. The result of the study revealed that the best predicting machine learning is the LSVM in scenario 1 which has all features as an inputs, with accuracy score at 87%.

Keywords: *Car plate theft, Dubai police, SPSS, techniques, Data mining and machine learning, time series analysis, predicting models, geospatial analysis, risk analysis.*

Table of Contents

Acknowledgements	3
Abstract	4
List of Figures	8
List of Tables	10
Chapter 1 - Introduction	11
1.1 Introduction And Problem Statement	11
1.2 Project Goals	12
1.3 Aims And Objectives	13
1.4 Limitations Of The Study	13
1.5 Structure Of The Thesis	14
Chapter 2 - Literature Review	16
2.1 Vehicle Security Technologies	16
2.2 License Plate Theft Trends	18
2.3 Technological Advancements In Vehicle Tracking	19
2.4 Innovations In License Plate Recognition	21
2.5 Binary And Kernel Density Approaches	22
2.6 Machine Learning Applications For Crime Prevention	26
2.7 Main Key Takeaways	31
Chapter 4 – Findings and Data Analysis	34
4.1 Data Dictionary	34

4.2 Exploratory Data Analysis 36

4.2.1 Data Profiling And Summary Statistics 36

4.2.2 Feature Selection 37

4.2.3 Visualization Of Key Features 39

4.3 Machine Learning Model Development 62

4.3.1 Detailed Explanation Of The Chosen Input 62

4.3.2 Detailed Explanation Of The Chosen Machine Learning Algorithms 64

4.3.3 Validation And Testing Procedures 65

4.3.4 Results 69

4.3.4.4 Hypothesis Testing 75

Chapter 5 – Discussion 77

Chapter 6 – Conclusion and Recommendations 80

6.1 Conclusion 80

6.2 Recommendations 81

6.3 Future Work 82

References 83

Appendix 88

Variable View in SPSS 88

Snippet of Data Sample 88

List of Figures

Figure 1: Number plate detection system block diagram 24
Figure 2: NDSV Architecture diagram 25
Figure 3: Simplified Schematics of Random Trees 29
Figure 4: CRISP-DM Methodology 32
Figure 5: Data Source 34
Figure 6: Predictor importance 39
Figure 7: Area Distribution of Car Plate Theft 40
Figure 8: Residence Status Distribution 40
Figure 9: Entry Method into the Country 41
Figure 10: Entry Legality 42
Figure 11: Education Level 43
Figure 12: Employment Status 43
Figure 13: Type of Car used 44
Figure 14: Repeated Crime Patterns and Area Crime Committed 46
Figure 15: Visualization of Crime Description and Repeated Crime Rate 48
Figure 16: Visualization of Type of Car and Repeated Crime 50
Figure 17: Visualization of Way of Entry and Repeated Crime 51
Figure 18: Visualization of Nationality and Repeated Crime 52
Figure 19: Visualization of Police Station and Repeated Crime 54
Figure 20: Visualization of Year and Repeated Crime 55
Figure 21: Visualization of Education Level and Repeated Crime 56
Figure 22: Visualization of Residence Status and Repeated Crime 57
Figure 23: Visualization of Criminal Status and Repeated Crime 59

Figure 24: Visualization of Arrival Time of Repeated Crime 60

Figure 25: Visualization of Arrival Time of Repeated Crime 62

Figure 26: Feature Selections of Scenario 1 63

Figure 27: Automatic using feature selection and random tree 63

Figure 28: Scenario 1 Model Performance 70

Figure 29: Scenario 2 Model Performance 71

Figure 30: LSVM1 Predictor Importance 74

Figure 31: The ROC Curve of Scenario 1 Modelling 77

List of Tables

Table 1: Data Dictionary Definition 35

Table 2: Summary Statistics 36

Table 3: Selected features by the algorithm 38

Table 4: Features are not related to the target 38

Table 5: Crosstabs Case Processing Summary of Repeated Crime Patterns and Area Crime was Committed 45

Table 6: Chi-Square Tests of Repeated Crime Patterns and Area Crime Committed 45

Table 7: Crosstabs Case Processing Summary of Crime Description and Repeated Crime Rate 47

Table 8: Chi-Square Test of Crime Description and Repeated Crime Rate 47

Table 9: Crosstabs Case Processing Summary of Type of Car and Repeated Crime 49

Table 10: Chi-Square Test of Type of Car and Repeated Crime 49

Table 11: Chi-Square Test of Way of Entry and Repeated Crime 51

Table 12: Chi-Square Test of Nationality and Repeated Crime 52

Table 13: Chi-Square Test of Police Stations and Repeated Crime 53

Table 14: Chi-Square of Year and Repeated Crime 54

Table 15: Chi-Square Test of Education Level and Repeated Crime 55

Table 16: Chi-Square Test of Residence Status and Repeated Crime 57

Table 17: Chi-Square Test of Criminal Status and Repeated Crime 58

Table 18: Mean of Arrival Time of Repeated Crime 60

Table 19: Mean of Quarter of the Year of Repeated Crime 61

Table 20: Hypothesis Test Summary 75

Chapter 1 - Introduction

In the changing world of city life today, the car plate stealing problem is growing fast, and it's linked with other crimes as well. Dubai, a famous place for fast city growth and money success, is facing an increase in stolen car license plate cases, which puts the safety of people and police work at great risk. The study looks into the details of this big problem and attempts to look into patterns, reasons, and ways it can be stopped for vehicle license plate theft in Dubai's special socioeconomic situation. Against the backdrop of the vibrant Emirate of Dubai, a city synonymous with progress and prosperity, this study seeks to unravel the layers of a pervasive problem that threatens not only the safety of its residents but also the very fabric of security upon which the city stands.

1.1 Introduction And Problem Statement

Car license plate stealing, usually thought of as a small crime, shows itself to be secretly connected with many other bad activities. The act of sneakily stealing a small metal thing is more than just taking it, as it is also about hiding the plans for doing illegal activities and working with criminals behind the scenes. These stolen plates can be used by people running bigger crimes, such as kidnapping someone or stealing things from houses and cars, and also moving illegal drugs around a beautiful city called Dubai in the sky. We need to deal with this problem that might seem separate, it's not just about keeping the people safe now, it could also put safety at risk in a big way all over the place called the Emirate.

In the shining big city of Dubai, people stealing car license plates work secretly behind the scenes, where they sneak into a strong safety network carefully created by this city to protect cars and other things. It's not just about losing a physical thing but it gives crooks the chance to stay hidden while they plan and do bad things on a big scale. The plates taken are hidden helpers used by bad people to hurt the safety of those living or going somewhere. So, it's important to not only prevent people from stealing plates. It is an important part of making Dubai feel safe and protected overall. Lifting license plates increases the city's inferiors because only unclean structures can be obtained, and created in other countries because of the state of affairs. This is necessary to know why silent crimes are committed hence, to store Dubai's name all over the world and ensure that people take it to be a safe place. This is where our responsible handling of the matter of whether our hometown will be a safe place for people to live or not in the future comes in.

1.2 Project Goals

As more and more stolen plates cause trouble, this study begins a trip to help Dubai be safer with better protection, as such the study aims to use smart methods and strategies that focus on data, which can help to mitigate license plate thefts as well as other related crimes. The study plan is to create a place where police departments have the newest tools that can predict future crimes. The aim is to make these places strong enough so they can spot and prevent moves used by people wanting to misuse hidden IDs from stolen license plates.

The project plans to change the way law enforcement works, such that, instead of reacting, they will try to stop crimes before they happen in the United Arab Emirates. The project's goals go beyond just law enforcement work, aiming for a big change in society by making and using smart technologies to grow a group of people who care about safety. They take action by helping make places where criminals struggle to work harder for them. The project aims not only to fix the problem of stolen license plates but also wants more people in Dubai to be safer and watchful.

1.3 Aims And Objectives

A pivotal aspect of our research methodology involves the creation of predictive models, leveraging advanced analytics to not only forecast potential incidents of license plate theft but also to furnish actionable insights for targeted preventive measures. The smart way of doing things makes sure that the research is like a strong helper for catching criminals and planning safety moves. Therefore, our main objective is to help reduce the number of stolen car plates and other crimes that go with them by joining efforts to make good prevention tools. We intend to improve safety in Dubai a lot more for everyone. We aim to study the city's security through machine learning predictive tools. This helps keep it a safe and calm place for people who live there or visit them.

- ✓ Develop predictive models for forecasting incidents of license plate theft.
- ✓ Provide actionable insights for targeted preventive measures against car plate theft and associated crimes.
- ✓ Contribute to the improvement of overall safety and security in Dubai through advanced machine learning predictive tools.

1.4 Limitations Of The Study

Recognizing the inherent complexities of any research endeavor, it is imperative to conscientiously outline the limitations that may influence the scope and generalizability of the findings. This is because it affects how big or general discoveries are, so we should watch out for them when doing studies. The research uses present data that comes from places like the Dubai Police Department, cameras watching everything, and local crime news reports that can be freely looked up by anyone. These sources may have problems, such as not being complete or correct, which can affect how true and dependable the final results of a study are because differences in data quality make things less accurate overall.

Moreover, the geographical scope of this study is confined to the Emirate of Dubai, necessitating a detailed consideration of its implications. The ideas and plans we have for making things safer in the Emirate might work well, but they may need changes when applied to places with different social, economic, or physical ways of life. Therefore, the study points out that understanding and adapting to situations is important when using findings outside of Dubai, which makes sure we can clearly understand what these results mean in other places.

1.5 Structure Of The Thesis

The thesis "Predictive Methods and Data Pattern Analysis for Reducing Car Plate Theft" is organized into six main chapters, each serving a distinct purpose in contributing to the research aims and objectives.

Chapter 1 - Introduction lays the foundation for the upcoming chapters by presenting an extensive overview of the research field and its importance. Further, this chapter outlines the research methodology implemented to carry out the study and identifies the limitations that may impact the coverage and the applicability of the results.

Chapter 2 - Literature Review is an in-depth review of the available literature relevant to the research area that seeks to cover several vehicle security technologies, current trends in the theft of license plates, developments in the tracking of vehicles, realizations of the license plate recognition, and the use of binary and kernel density methods. It gives a theoretical framework to put the research findings in proper perspective.

Chapter 3 –Methodology provides a comprehensive presentation of the idea about the process of data collection, the analysis methods used, and the justifications for the choice of concrete methodologies.

This chapter by outlining the project description sets the stage for the rest of the data analysis and the interpretation of findings.

Chapter 4 - Findings and Data Analysis is the heart of the thesis where the collected data is analyzed and interpreted to develop useful insights. In this chapter, we start with the data dictionary, defining the critical variables employed in the analysis. Consequently, descriptive statistics are given to give a complete summary of the dataset. Crosstab analysis is also discussed as part of the chapter, focusing on geospatial patterns of crime, crime description patterns, and the type of car and repeat crime occurrences. Moreover, predictive analysis is performed to predict car plate theft occurrences under different scenarios, further followed by hypothesis testing to test research hypotheses.

Chapter 5 - Discussion is a critical analysis of the previous chapters' findings, presenting appropriate interpretations and implications, theory, practice, and policy. It integrates the research results with the literature available in the field, discussing the areas that coincide, the ones that do not, and those that require more research. This chapter provides a critical debate on the importance of the research findings and how they can help in dealing with car plate theft and improve law enforcement measures.

Chapter 6 - Conclusion and Recommendations is a section that summarizes the research findings and provides practical solutions for the other relevant parties such as the policy-makers, law enforcement agencies, and future investigations in the same line. It collects the main perspectives found during the research and presents the paths how to deal with car plate theft and security measures. This chapter merges the research results, offering foresighted recommendations and ascertaining the importance and applicability of the findings to real-world problem.

Chapter 2 - Literature Review

2.1 Vehicle Security Technologies

According to a study by **Wolf et al., 2007,** it is important to put safety measures in cars by understanding how car technology is changing, and taking a keen interest in the need for computer safety measures to protect vehicles from being tampered with by people outside and those who own or maintain them. It is important to understand the main part of how license plate theft works in a larger context. In their study, Wolf and others show the many sides of safety problems in cars, since they understand that we are using complex electronics more than ever before. By looking into the small details of computer security in cars, the study shows that we need to use more than just normal solid barriers. These extra measures should help protect against wrongdoing and break-ins online.

It's very important to learn about weaknesses in car security systems, as this will help greatly when studying the theft of license plates more deeply starting now. **Nolte, Hansson, and Bello (2005)** talk about car safety by looking at how auto communication has changed over time. At the IEEE meeting, they looked back and studied how auto industry communication advanced. It presents a history that is very important to understand the weaknesses that number plate thieves may use. The study digs into the past, covering the early stages of automotive communication technologies and moving forward to today's status. Nolte and his team did a job that helped us understand what changes might happen in talking between cars later on, including possible safety risks. Looking back at history is important when measuring how well safety methods work with new technologies. It helps us understand better ways to reduce problems caused by taking license plates away from cars.

In **2005, Johansson, Torngren, and Nielsen** added a key piece to understanding car safety. They looked at how Controller Area Network (CAN) is used in vehicles. The study explains that the CAN is like the main brain for today's cars. It says it plays a key role, allowing different ECUs in various parts of the vehicle to talk and work together well. In simple terms, CAN helps link different systems in a car. It makes sure everything works well together to run the vehicle smoothly.

ECUs are the computer parts that control certain tasks inside cars like how the engine runs, changes gears, and stops. Getting a good grasp of ECUs and how they talk through CAN is very important. It lets you understand the weak points that people who want to steal license plates might use **(Ayres, 2021).** Since CAN mostly serves like the brain of current cars, any problem with it could greatly affect how safe and secure all parts of a vehicle are. Knowing how CAN works helps scientists, engineers, and police to make security methods that protect against changes from outsiders or tricky tactics used by bad people trying to steal license plates. By understanding how CAN connects different systems, security planners can come up with plans to keep the physical license plates safe. They also make sure that electronic stuff used by thieves for stealing license plates is also protected strongly. This full understanding is important in making ways to avoid theft. It goes beyond normal anti-theft techniques, keeping a good cover against all the different tricks used by bad people for breaking car safety rules **(Diaz, 2023).**

Kaushik et al. (2014) significantly contribute to the discourse on vehicle security by presenting an innovative and comprehensive anti-theft vehicle security system. The study shows how criminal ways are changing, hence recognizing the need for smart safety steps to stop bad actions from getting more complicated. Modern crime methods are often really hard. We need new and active methods to stop bad actions, like taking cars or plates on driver's licenses that say who they belong to. The use of the anti-theft vehicle security system shows a change towards active ways to keep things safe. This system is made not only to handle theft cases but also to actively stop them. It follows a bigger thinking style of making plans well in advance for security protection. The Anti-Stealing and Protecting Car System is designed not just to keep the whole car safe but also to guard its parts, like the license plate **(Liu, 2020).**

White and Davis (2020) offer useful tips for stopping license plate thefts, wherein their study gives important help by looking at how well certain stop-measures work to prevent license plate theft. The study gives clear suggestions for police and car owners. It recommends using strong screws that are hard to break, as well as special holders for license plates that make it harder to steal them. This practical and useful part makes the big talk about car safety stronger and shows how important it is to stop bad things like stealing license plates from happening **(Motakabber et al., 2022).**

2.2 License Plate Theft Trends

Smith and Johnson (2015) have helped us a lot by studying the big issues around car license plate thefts. This is especially important for places in cities. They did a complete study about how often, the methods used, and the reasons behind stealing car license plates. It shows clearly the many problems that police agencies face in reducing and stopping this specific kind of crime within city areas. The study looks at more than just the numbers, like how often it happens. It also tries to understand why people who steal do what they do and their reasons for stealing **(Kumar et al., 2022).** By paying attention to cities, where people live close together and there are often many cars on the roads, a study tells us what it's like for police officers when they try to stop license plates from being stolen in busy city areas. It highlights how the bad guys are changing their tricks and what they want, like money or hiding from something else. These detailed understandings are very useful for police work plans. They help them change and adjust their methods to deal better with certain issues caused by stolen license plates in cities **(Ouallane et al., 2022)**

Jones and his team (2019) look closely at how stolen license plates play a part in vehicle theft overall. It helps us understand car number plate theft in a new way. It's not just about stealing items; it may also lead to bigger crimes down the line. The study looks at how stolen license plates are involved in taking cars. It shows hidden links between different types of crimes that might not seem connected at first. We can spot patterns and trends that show where license plates are stolen, which is a major connection to bigger crimes **(Mallikalava et al., 2020).** This knowledge is very important for police and decision-makers who want to break down criminal groups involved in many types of crimes. Understanding that getting your license plate stolen can cause problems beyond just the act itself, we need ways to handle it. It's not enough to only deal with it when it happens. This makes police departments think about the big effects of this type of theft when making plans, as they want to not just fix the symptom but also solve what's causing it **(Mohammad et al., 2022).**

Brown (2018) changes the attention to money concerns related to stealing license plates, especially its part in taking personal details, which shows how complex stealing license plates can be. It also stresses the need for strong actions to stop this crime from getting worse and leading to more problems overall. This research looks at the money side of plate theft, going beyond basic ideas about it as a crime that only steals things. Thieves can use stolen license plates for identity theft, which makes it harder to know how negative this simple crime might be. After learning that stealing license plates can cause big problems, especially for theft of identity and money crimes, we need to review how safe everything is. One should push for a whole new way to stop theft that goes beyond just stopping the loss of license plates **(Yaacob et al., 2021).** We need to protect people not only from their direct losses but also from the money problems linked with identity theft.

2.3 Technological Advancements In Vehicle Tracking

Pethakar, Suryavanshi, and Srivastava (2012) introduce a new system that combines Radio-Frequency Identification (RFID), Global Positioning System (GPS), and GSM technology to locate cars for safety reasons. Their job shows how new technologies can be used to solve car-related crimes. The suggested system wants to track cars and also improve worker safety by using three advanced technologies together. RFID helps put a special marker on cars, GPS gives us where they are right now and GSM makes it easy to talk with them or control them. This connection gives the system a wide range of tools to handle safety worries **(Pria and Kumar, 2021).** It brings together vehicle tracking and staff protection in one complete method. The focus on finding cars and keeping workers safe shows a wider realization of how these new tech improvements can help with car crimes. The main goal is to keep workers safe, but these same ideas and features can also help stop crimes like taking cars or stealing number plates less directly. Using RFID, GPS, and GSM technologies together shows a smart way of thinking that goes beyond usual safety measures.

Nagaraja and his team (2009) assist in making cars safe by using a GSM system to stop theft. This uses GSM tech to stop and control car theft. The research didn't directly talk about stealing license plates, but it gave information on all things related to keeping cars safe. It demonstrates how important GSM technology is in battling car theft. The study makes a system for stopping car theft that uses GSM technology. This makes GSM a crucial tool for stopping and managing plans. Using GSM lets us quickly communicate and control things in real-time. This allows for fast reactions if someone tries to get into a car without permission or steal it. We think about the whole car, but using GSM technology to stop theft helps our bigger talk on making cars safer. The tools and systems used in the car theft prevention system based on GSM provide a base for creating total safety plans that go beyond just making sure the vehicle is physically safe **(Akanda et al., 2022).** Scientists, workers, and leaders can use GSM technology to make car safety stronger.

Khedher (2011) talked about a car tracking system that uses GPS and GSM for location. This is very important when looking at better ways to find vehicles as it helps improve localization methods. Even though the study doesn't talk about taking off license plates, it helps us to know better how we keep track of stolen cars. It is indirectly connected to number plate thefts because using GPS and GSM together makes car tracking systems more accurate and trustworthy. The hybrid method combines the best parts of GPS to locate things correctly and GSM for live talking, thus providing a more reliable way to find where vehicles are moving or stopped. Stolen cars often change their license plates illegally to escape notice, so good tracking is needed for the vehicles to be found again via the GPS-GSM location system, which makes tracking more accurate and indirectly helps with problems caused by stolen license plates. Understanding how tracking cars changes over time helps us see the connection between car security problems. This gives useful understanding for making big plans to solve these issues together.

2.4 Innovations In License Plate Recognition

Kathleen and others (2023) added a lot to what's been written before by focusing on recognizing license plates for finding stolen cars using deep learning. Their work matches up with today's ideas that use fancy technology like deep learning to deal with problems caused by stealing license plates. We can improve the right prediction and speed of finding stolen cars by using smart math formulas and brain-like computer groups to look at license plate numbers. Deep learning, which can learn and change by itself from data, provides a good way to make plate number recognition systems more accurate. The main goal of finding stolen cars shows the active way that modern tools, especially deep learning, are used for stopping license plate stealing. Old methods might be slow to follow the changing tricks used by bad guys. So, new ideas about reading license plates are very important.

Huang, Cai, and Lan in 2021 suggested a new idea with just one big network people use to find license plates that have different styles. It is showing the future way to quickly recognize license plates which shows great potential. This study helps us understand the new ways of dealing with car number plate crimes. This new idea makes it easier to recognize car plates, lowering the trouble that comes from dealing with various plate designs. The study helps make license plate recognition systems better and more adaptable by putting all these tasks in one big network inside a computer. The idea of using one neural network for different styles of license plate recognition shows progress in the field of applying neural networks. This method shows how flexible neural networks can deal with different kinds of data. It shows they have the potential to change technology for recognizing car license plates in a big way.

Lomte and Sayyed (2015) made a special system for recognizing car number plates to find stolen vehicles. This helps us learn how useful recognition systems are for spotting taken cars. Concentrating on what makes a stolen car different, the system can be made better for quickly spotting likely theft happenings and telling police. This special way helps make focused answers that deal with the unique problems caused by stolen license plates. The car number plate spotting system helps a lot in keeping cars safe. Their work adds an extra level of safety by using tech for knowing theft, hence going beyond the usual ways whereby we try to stop bad guys from stealing. The study shows that recognition systems can stop stolen license plates and also help find and reduce bigger crimes related to taken vehicles.

2.5 Binary And Kernel Density Approaches

Numerous frameworks, based on various image processing and artificial intelligence algorithms, have been presented for license plate identification and recognition. The binary and kernel density function approaches utilized for license plate processing are the foundation of the license plate identification methodology presented by **(Yan et al., 2021).** When you multiply the binary value by that of the initial picture, the corresponding license plate can be detected. This relates to the secondly step, which benefited from the filtered binary values of the image. The image processing technique was offered to (Kaur et al., 2022) for the identification and recognition of Indian registration plates regardless of the different conditions in which they are revealed such as poor light, cross-angled situations, noisy environments, and non-standard number plates. In addition to this, they applied various methods during the pre-processing stage, which involved morphological transformation, gaussian thresholding, and gaussian smoothing among others. Classifiers such as contour and the K nearest neighbors method were employed to classify the characters. Svsrk and Ganta propose another option for Indian number plates. They deal with the issue of scaling as well as the placement of the figure of the character as aforesaid.

A feature extraction model evolves as the license plate is spotted on the vehicle; character recognition is accomplished by employing a backpropagation neural network model. The image was first dealt with to enhance the appearance of the car pictusre. According to **(Wang and Huang, 2020)** research, it was found that many vertical edge helps to locate the license plate, so **(Wang and Huang, 2020)** framework uses a 2D wavelet transform approach to extract vertical edges from the picture.

CNN classifier is used for recognition purposes to identify characters on the number plate. It is recommended that the proposed solution to this by the given author can successfully eliminate many errors in character recognition and license plate detection. It is based on the capsule network originated from **(Jawale et al., 2023).** The authors point out that this structure is reliable and works in every condition - regardless of rotating, flipping, and changing the picture. The time of processing to be improved with a capsule network architecture segmentation is the first purpose of this design. A vehicle license plate number verification against a database of licensed users and only approving cars that have been authorized has been suggested by **(Sestrem et al., 2023).** The algorithm features that are extracted in this case are called the histogram of the directed gradient. This framework has demonstrated its effectiveness by the sheer quantity of discovered cars.

(Ariff. et al., 2021) suggested an appointment and employment of morphological tools for segmentation and a Sobel edge detection approach for extraction of license plate region. Through IoT, the vehicle's database gets constant and frequent updates because of the internet. Tilt and uneven light are unavoidable problems, **(Huang et al., 2020)** identify a way to segment license plate detection and character recognition. It can process grayscale pictures even coming from steel cameras or videos. Most importantly, every single character and location registration number can be discovered very precisely.

(Karolin and Meyyappan, 2022,) suggested an algorithm for license plate recognition (LPR) that includes the conversion of a color (RGB) image into a grayscale image and then binarization into a black-and-white image. Next, the license plate gets localized and segmentation of the characters is realized and finally, the characters are recognized. The author points out that using this approach both contrast enhancement, histogram equalization, and noise filtering procedures among others can be eliminated. LPR or number Plate Recognition system was developed which can recognize and track moving vehicle number plates **(Lubna et al., 2021).** This method targets cars that show traffic violations of speeding and traffic signals.

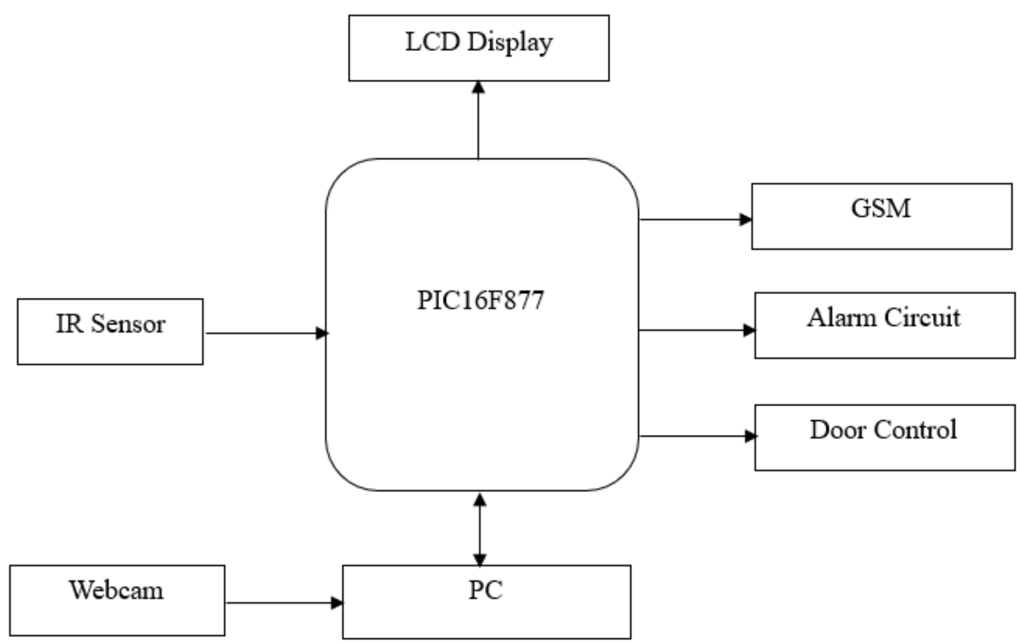


Figure 1: Number plate detection system block diagram

By comparing it to a database of stolen vehicles that the police station provides, the stolen car is found. It then sends an SMS to the police station and sounds an alert. It may be applied to a variety of security tasks, such as parking systems, toll collecting, speed detection, and traffic infraction detection **(Bukola, 2020)**. Secure locations like residential area gates, industry gates, parking lots, toll plazas, university entrances, and other highly guarded buildings like nuclear facilities and defense institutions can also benefit from it. Figure 1 displays the system's block diagram as a whole.

When a car pulls up to the toll plaza, it stops at the gate, and the camera takes a picture of it. The camera then sends the picture of the vehicle to the microcomputer unit, which takes out the license plate and splits the characters **(Roman and Karthiayani, 2023).** It also uses MATLAB software to identify the information it receives. If the database indicates that the car is stolen, the gate won't open for that vehicle. In addition, the alarm sounds, not only alerting the police station but also displaying the information on the toll plaza's LCD **(Shinde et al., 2023).** On the other hand, the door will open and the buzzer won't ring or make noise if the photograph of the car doesn't match the one in the database. In this manner, car after car that has been correctly detected drives around the toll plaza for 24 hours.

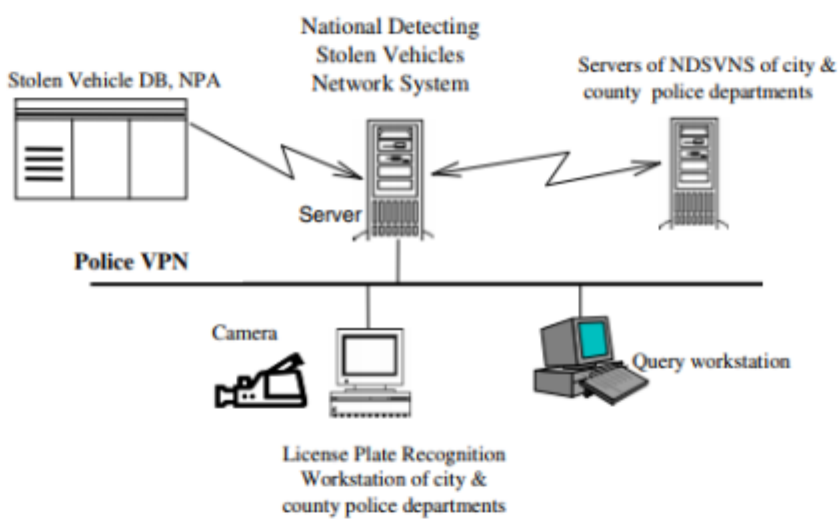


Figure 2: NDSV Architecture diagram

After the camera takes a picture of the car that will be utilized for additional processing, the existence of the vehicle is recognized using an infrared sensor. Vehicle recognition that is done manually will result in more errors, slower processing times, and less efficiency **(Chandra et al., 2021).** The most important and basic stage in the ALPR method is to take the number plate characters out of the picture of the car

(Etomi and Onyishi, 2021). There are several steps involved in the detection process. Beginning with the horizontal and vertical edge detection approaches, which are based on the features of the edge exhibited by the character's edges on the vehicle's number plate, is the number plate extraction process **(Mallikalava, 2020).**

Several processes, including feature extraction and categorization, are involved in character recognition. The characters are normalized before the recognition algorithm. The goal of normalization is to reduce the character set to a block with no additional white space (pixels) on any of the four sides. Next, every character is proportionately fitted. Matching the template requires the fitting procedure. The input photos must be the same size as the database characters to compare the characters with the database. An appropriate algorithm for character recognition is template matching. The character picture is compared to those in the database, and the maximum similarity is determined. A correlation function is utilized to quantify the similarity and identify the best match **(Mohammad et al., 2022).**

Different graphical user interfaces are developed for appropriate user-friendly settings so that users can process them step-by-step. Each identified license plate is cross-referenced with a stolen vehicle database; if a mathematical match is made, the car is reported stolen **(Shobayo et al., 2020).** The vehicle's database is kept up to date for security purposes, making it easier to identify stolen vehicles. The device will sound an alert and lock the entrance upon detecting a stolen car, preventing it from leaving the location. A notice will then be forwarded to the traffic police to identify theft.

2.6 Machine Learning Applications For Crime Prevention

A study by Kathleen et al. (2023) looked into the application of machine learning for crime prevention, highlighting pioneering research conducted that shows a big change in the crime prevention area. It uses modern tools leveraging cutting-edge technologies to enhance the efficacy of preventative measures, mainly with a particular focus on license plate spotting. These technologies use advanced methods and neural networks to help improve license plate recognition systems. These smart ways, especially deep learning, and unified neural network methods, help catch crime better. It’s useful for finding stolen cars or license plates being taken away **(Shah and Shah, 2021).**

The study of machine learning methods shows that crime prevention technologies are changing since traditional ways of stopping crime are being combined with and sometimes replaced by modern computer methods. Machine learning helps us look at big sets of data, finding patterns and trends that could be hard to see in normal ways. In the special case of reading license plates, these methods provide a fast and helpful way to deal with tricky criminal activities **(Zhang et al., 2020).** They help in stopping crime by adding more strength to preventing bad behavior. The mix of machine learning and crime-stopping shows a teamwork way to deal with modern safety problems.

When scientists put new technology into stopping crime, like taking license plates, they close the gap between old safety methods and what we need in today's technological world. Doing this not only makes the police stronger, but also shows why it's essential to stay one step before bad guys in different ways by using machine learning. Using machine learning to check license plates can help stop crime, shown only as a major change in how we keep people safe from bad guys and thieves. These ways show us that we are gradually getting better tools, making it simple to stop crime before it starts faster and right. As tech evolves, coordinating machine learning and security is crucial to protect individuals in neighborhoods.

A machine learning algorithm called a support vector machine (SVM) determines the boundaries between data points based on predefined classes, labels, or outputs. It does this by executing optimal data transformations. SVMs are used to solve complex problems related to classification, regression, and outlier detection. SVMs are widely used in many sectors, including speech and image recognition, natural language processing, healthcare, and signal processing applications. SVMs are essential for creating apps that use predictive model implementation **(Kavitha and Kaulgud, 2022).** SVMs are simple to use and understand. Through kernels, they provide an advanced machine-learning approach to analyze both linear and non-linear data. SVMs are useful in all fields and real-world situations where handling data through the addition of higher dimensional spaces is required. The steps as part of constructing the supervised learning models will include aspects such as choosing the kernel to apply, fine-tuning hyper-parameters, as well as spending time and resources during the training operation.

Unlike the SVM (Support Vector Machine), which is designed to work solely with numerical variables, the Random Trees classifier can deal with both numeric and categorical characteristics. Besides, Random Trees do not break down the data scaling like SVM, thus normalizing the data is required frequently to prevent non-optimal algorithms' training or wrong classifications **(Cha et al., 2021).** On the one hand, SVM might not be superior to other methods for large or balanced data sets, but on the other hand, it does perform better if the training set is small or imbalanced. Even though a random trees classifier can perform much better than SVM they still compute much faster, they do it with larger training sets. However, the RFT algorithm has been reviewed and has many variations. These conditions of minimum observations, maximum depth of the tree, and the accuracy of the trees are the necessary stopping criteria of the OpenCV implementation when it uses the Gini Impurity index to determine the point of split for a node. The central concept under the roof of the randomization-based ensemble techniques is developing a multitude of different models that were obtained from a single learning set L by introducing a randomized perturbation in the learning process. The outputs of these individual models are combined to form a collective prediction for the ensemble forecast. It is worth noting that increasing the number of trees and then letting them vote for the best class has brought us significant benefits in terms of classification precision. Sometimes, random vectors are utilized for creating individual trees of the ensemble, and then following these vectors, the ensembles are offset from each other.

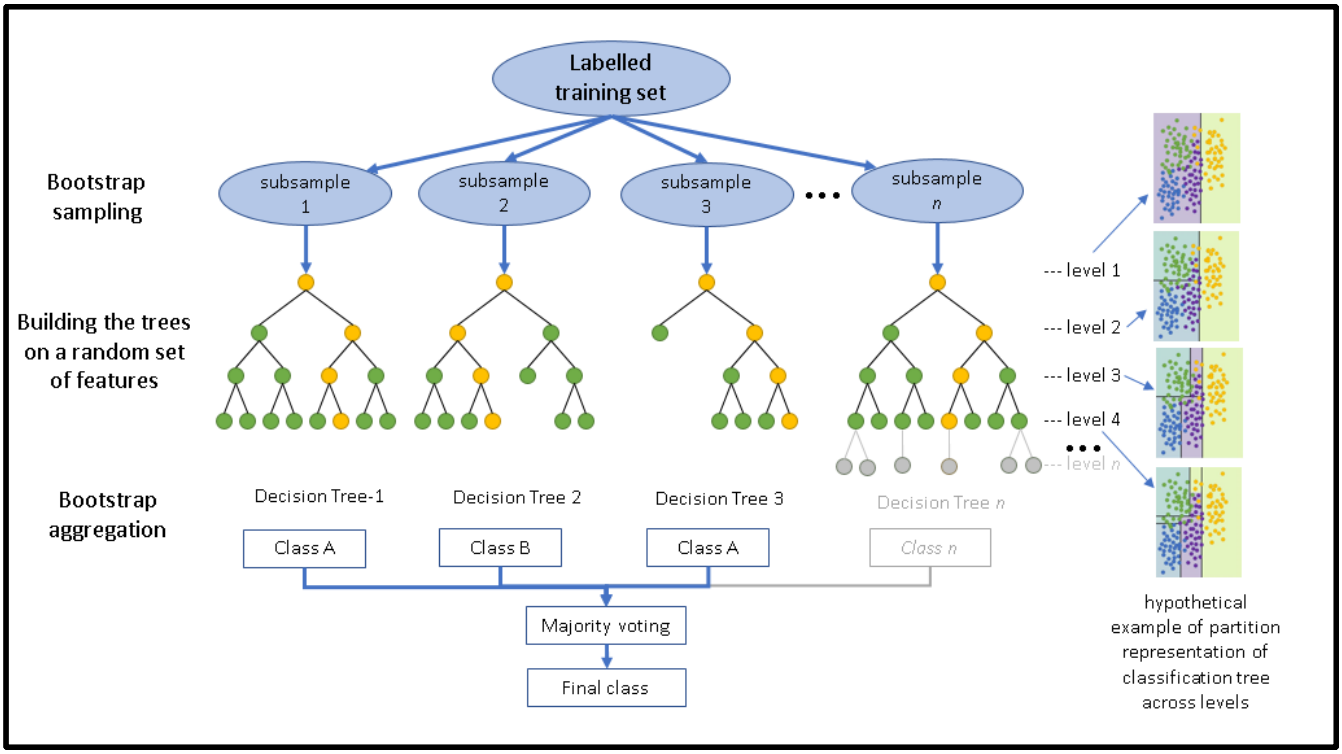


Figure 3: Simplified Schematics of Random Trees

The method of modeling the likelihood of a discrete result given an input variable is known as logistic regression. A binary result, or something that may have two values, such as true or false, yes or no, and so on, is what most logistic regression models represent. When there are more than two distinct discrete outcomes in a scenario, multinomial logistic regression can represent the situation. A helpful analysis technique for classification issues is logistic regression, which is applied when attempting to ascertain which group a fresh sample most closely belongs to **(Kuha and Mills, 2021).** Logistic regression is a helpful analytical tool since several parts of cyber security, such as attack detection, are classification issues. When the explanatory factors are continuous but the response variable is binary, logistic regression might be helpful. This would be the case if data on an individual's income, years of employment, age, education, and other continuous factors were used to forecast whether or not the consumer is a good credit risk.

The feature weights are approximately equivalent to the parameters of the logistic regression model. The S-shaped logistic function is used to translate each weighted feature vector to a value between 0 and 1. This number represents the likelihood that an example falls into a specific class. For the learning instances to be correctly classified, the learning algorithm adjusts the weights. Tools to prevent overfitting must be sought after to put that dilemma into a form of resolution. For the calculation of the errors, the gradient descent algorithm with its different types is frequently applied. Finally, armed with class weights, the logistic function with each unseen sample is used to determine its chance of belonging to a class.

The distribution of the input data and their attributes pose an increased exploitable chance for the machine learning AI solution. Outliers pose a risk of deception in the process of training that can initially lead to the application of models that require longer training or do not achieve the desired accuracy **(Vercio et al., 2020).** In machine learning, outliers are a series of processes that include the initial fitting of a normal model to the common data and later making predictions about whether a newly received data is normal or not. In contrast to one-class classifiers that belong to a group of unsupervised learning algorithms, one class cannot determine if the new data falls into the range of data that has been fed to it.

Some cell elements can be left empty, and rows and columns can also be missing; these are often referred to as the Missing values. Due to that, data loses its reliability and feasibility to the user. Following a few machine learning algorithms that have been running with a dataset that includes null values produces one error. Thus, information that is blur or unreadable counts as the missing numbers should be determined and handled properly in the process of both analyzing or building data for any purpose.

Regarding the reasons for the missing information from the data, the possible explanations are several and are the conclusions of **Birnick et al. (2022).** The strategy applied to fill gaps in the dataset relies on a preliminary assessment of why there is missing data in the dataset. As data analysts frequently face the issue of a missing value problem dealing with authentic data, chance arises to severely affect the very quality as well as the validity of their results. Researchers need to be able to select an appropriate method referred to from a variety of studies about missing data as a result of knowing different types of missing data values which can influence the analysis. All methods are equally effective in solving a certain problem or in fixing a certain erring data value.

2.7 Main Key Takeaways

- This review underscores the increased concern to do with car registration plate stealing and serves as an enabler of crime of different types. It draws our attention to the necessity of a proactive stand for the situation concerned and the introduction of both the development of predictive methods and data analysis techniques as well. To end up with this it draws attention to the fact that the participation of the public and social development is crucial in solving this challenge.
- In spite of the fact that it is crucial to delve into statistical data and crime reports to understand the importance of car plate theft problem, you do have this in mind.
- Evaluating this approach will reveal successful strategies and areas that require the most attention for improvement, which will provide with an idea for the development of new preventions.

Acquainting yourself with the competences and shortcomings of LPR technology will help you to see the place of this system in the decreasing of car plate thefts.

.

Chapter 3- Project Description

What the research is about is becoming aware of and constructing a template that will incorporate the use of data in making predictions about car plate theft. Also, to have a visualization dashboard to notice the vibration proprieties of the signals and an active prediction model for the faults category classification. The use of the CRISP-DM (Cross Industry Standard Process of Data Mining) technique will be adopted in this study, and the figure 4, below, illustrates it. It is the method that is widely applied in data mining systems for assuring high-quality and accurate models. CRISP-DM model is a framework for completing and conduct the data mining by proposing a strategy for organizing and implementing the data mining. Moreover, the affluence of the work will be resorted to for the sake of a balanced production.

Figure 4: CRISP-DM Methodology

The stage for creating plans and patterns concentrates on getting important data about stolen car plates. It connects to the features set out in project goals and then uses machine learning methods that can predict things for its analysis needs. Different data are collected, and their importance or quality is checked to see if they work with SPSS analysis. Getting data ready and cleaning it is very important when using

SPSS. This helps make sure the data we use can be understood easily. Therefore, it deals with data gaps, applies fill-ins to fix them, and changes how the data looks on SPSS. We also make new features that match what can be used in this software well.

The modeling part is when you use different machine learning methods in the SPSS program to create forecasting models. In the SPSS platform, models are trained using methods like random forest, support vector machine (SVM), and logistic regression on a set of data that has been prepared. Random Forest is a team-based learning process that uses grouping and classification techniques in the SPSS software. This method grows a number of decision trees in training time and combines their output to increase the accuracy level and avoid overfitting within the SPSS. SPSS offers an efficient way of applying SVM (Support Vector Machine), a supervised machine learning tool, for classification and regression, tasks. It has an ambition to get the best line in a large space which is great for separating groups.

SPSS applies the decision trees, which is a simple way to deal with data, for classification and predicting tasks that need no data or formulas. The program decomposes the data into the components of the tree. It keeps making choices until it reaches the endpoint where classes or numbers are stated. Logistic regression is a sorting technique that is based on yes or no categories. It is an SPSS component to investigate the likelihood of any given input being in a particular class. It shows the connection of a yes/no outcome with other parts which makes things to be classified into two groups in SPSS.

In the SPSS environment, the models produced by the predictive modeling for car plate theft prevention are subjected to a detailed assessment. Model checking is essential to ensure our models are in line with our goals and help towards security improvement. We are very attentive to each model's performance and thoroughly check it. To measure the degree of truth, speed, and reliability, different criteria and standards are used. This evaluation covers the complexity of model behavior and how much they mimic real-life situations as well as the dynamic nature of car plate theft.

The assessment process targets the articulation of the most appropriate algorithm that detects and prevents car plate theft specifically in SPSS modeler environment. That analyzes the in-depth capabilities of SPSS to make sure that the selected algorithm not only meets but surpasses the expectations described in the project objectives. This careful examination is very important in refining the models, ensuring their reliability, and eventually fulfilling the purpose of the project which is to develop proactive and data-driven solutions for safer and securer Dubai.

Chapter 4 – Findings and Data Analysis

4.1 Data Dictionary

The project relied on variety of data including Dubai Police data for car plate incidents, Survey Camera footage from locations with high crime occurrences, Local crime reports, and incidents data, and publicly available data related to vehicle ownership and registration, of all whose type and attribute definition are as in the table below:

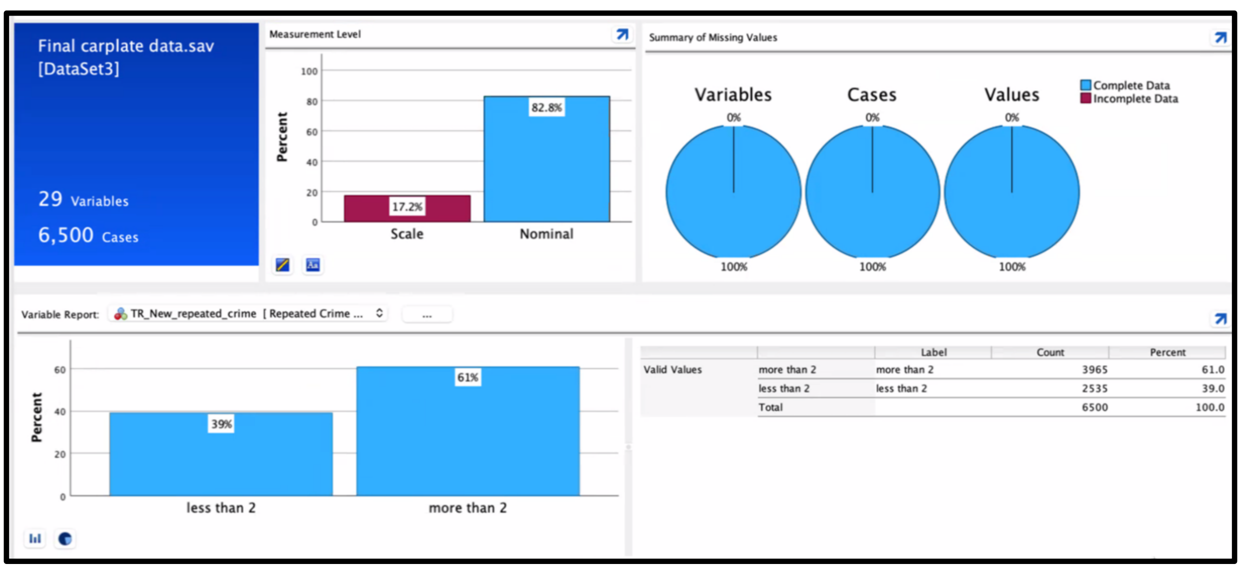


Figure 5: Data Source

Table 1: Data Dictionary Definition

Attributes	Data Type	Definition
Year	Numerical	In which year did the criminal commit the crime: 2020, 2021, 2022, 2023.
Police Station longitude	Numerical	Police station longitude and the geographic location on maps.
Police Station latitude	Numerical	Police station latitude and the geographic location on maps.
Crime scene longitude	Numerical	Crime scene longitude and the geographic location on maps to calculate how far it is from the police station
Crime scene latitude	Numerical	Crime scene latitude and the geographic location on maps to calculate how far it is from the police station
Quarter	Categorical	The year is split into 4 quarters, in which quarter of the year did the criminal commit the crime.
Residence status	Categorical	The status and legitimacy of the criminal residence in the country: resident, overstay, active visa.
No. Crimes	Numerical	How many crimes did the criminal commit besides the car plate theft.
Police station	Categorical	The competent police station in the specific area where the crime occurred: Bur Dubai Police Station, Jabal Ali Police Station, etc.
Area	Categorical	Each police station in Dubai is concerned with and specializes in crimes that occur in areas within its jurisdiction. Example: Bur Dubai Police Station specializes in the crimes committed in Al Quoz and Al Jafiliya etc.
Crime description	Categorical	The reason behind stealing car plates: to use the car plate in other crimes like alcohol gangs, kidnapping, etc.
Crime timing	Categorical	At which time of the day did the criminal commit the crime: day, night?
Age	Numerical	The age of the criminal.
Detention status	Categorical	If the criminal was arrested or escaped.
Entry method	Categorical	How did the criminal enter the country: legal entry or smuggling?
Document legality	Categorical	Did the criminal entered the country with legitimate documents
Educational level	Categorical	What level of education does the accused have
Employment status	Categorical	Is the accused working or unemployed
Type of car	Categorical	The condition of the car from which the plate was stolen: was it neglected in an uninhabited area or for a long period, or was its owner taking care of it and using it on an almost daily basis?
Camera footage	Categorical	Shows whether the area crime was reported was CCTV enabled or not.
Nationality	Categorical	The criminal nationality
Arrival timing	Numerical	The time that the police petrol took to arrive at the crime scene from the police station
Repeated crime	numerical	Whether if the criminal repeated the car plate theft or not (target of the study)

4.2 Exploratory Data Analysis

4.2.1 Data Profiling And Summary Statistics

Table 2 presents summary statistics derived from data profiling, aiming to numerically summarize the dataset's characteristics, where measures such as mean, median, standard deviation, and range are computed to offer insights into the dataset's distribution and variability. Basic attributes like the number of observations, missing values, and data types are examined to provide a comprehensive overview. The table showcases various fields along with their corresponding measurements, including nominal and continuous variables. For continuous variables such as the longitude and latitude coordinates of police stations and crime scenes, statistics like minimum, maximum, mean, standard deviation, and skewness are calculated.

Table 2: Summary Statistics

Field	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
I_PoliceStation_longitude	Continuous	25.004	25.272	25.098	0.062	0.559	--	6500
I_PoliceStation_latitude	Continuous	54.924	55.529	55.194	0.079	0.024	--	6500
I_CrimeScene_longitude	Continuous	24.957	25.293	25.126	0.097	-0.013	--	6500
I_CrimeScene_latitude	Continuous	54.888	55.685	55.226	0.102	0.036	--	6500
I_Year	Nominal	--	--	--	--	--	5	6500
I_Age	Nominal	19	42	--	--	--	24	6500
I_Nationality	Nominal	--	--	--	--	--	15	6500
I_Quarter	Nominal	1	4	--	--	--	4	6500
I_Residence_Status	Nominal	--	--	--	--	--	3	6500
I_PoliceStation	Nominal	--	--	--	--	--	9	6500
I_Area	Nominal	--	--	--	--	--	39	6500
I_Arrival_Time_Min	Nominal	5	30	--	--	--	16	6500
I_Crime_Description	Nominal	--	--	--	--	--	10	6500
I_Crime_Timing	Nominal	--	--	--	--	--	2	6500
I_Criminal_Status	Nominal	--	--	--	--	--	2	6500
I_Way_Of_Entry	Nominal	--	--	--	--	--	2	6500
I_Document_Status	Nominal	--	--	--	--	--	2	6500
I_Education_Level	Nominal	--	--	--	--	--	3	6500
I_Employment_Status	Nominal	--	--	--	--	--	2	6500
I_Type_Of_Car	Nominal	--	--	--	--	--	2	6500
I_Camera_Footage	Nominal	--	--	--	--	--	2	6500
TR_New_repeated_crime	Nominal	--	--	--	--	--	2	6500

These statistics provide an understanding of the geographical distribution and dispersion of the data points, whereas, nominal variables, are summarized in terms of their unique values and validity. The variables encompass diverse attributes such as age, nationality, quarter, residence status, police station, area, arrival time, crime description, crime timing, criminal status, way of entry, document status, education level, employment status, type of car, camera footage, and new repeated crime occurrences. Each nominal variable is analyzed in terms of its range of unique values and validity, contributing to the overall profiling of the dataset.

4.2.2 Feature Selection

4.2.2.1 Feature Selection By The Algorithm

Feature selection algorithms are employed to discern the most crucial variables for inclusion in the modeling process. One such algorithm utilized is the random tree feature selection, which aids in identifying the most influential features among the dataset. The feature selection process involves evaluating each feature's contribution to predicting the target variable or outcome of interest wherein, techniques like correlation analysis and tree-based methods are employed to rank features based on their predictive prowess. By assessing the importance of different features, researchers can pinpoint those that significantly impact the model's performance. Within the framework of this study, the most important feature selections derived from the SPSS Modeler are identified. These selections serve as key indicators of the variables that wield substantial influence over the predictive accuracy of the models. In Table 3, I used the feature selection algorithm to compute the important features, and the features that are not related to the target are shown in Table 4. In addition, the reasons why each feature is not related to the target are explained in the table.

Table 3: Selected features by the algorithm

Rank	Field	Measurement	Importance	Value
1	I_Area	Nominal	Important	1.0
2	I_Crime_Description	Nominal	Important	1.0
3	I_Nationality	Nominal	Important	1.0
4	I_Arrival_Time_Min	Nominal	Important	1.0
5	I_Way_Of_Entry	Nominal	Important	1.0
6	I_PoliceStation	Nominal	Important	1.0
7	I_Quarter	Nominal	Important	1.0
8	I_Year	Nominal	Important	1.0
9	I_Residence_Status	Nominal	Important	1.0
10	I_Education_Level	Nominal	Important	1.0
11	I_Age	Nominal	Important	0.997
12	I_Criminal_Status	Nominal	Important	0.988
13	I_Employment_Status	Nominal	Important	0.986
14	I_Camera_Footage	Nominal	Important	0.977
15	I_Type_Of_Car	Nominal	Unimportant	0.548

Table 4: Features are not related to the target

Field	Measurement	Reason
I_PoliceStation_longitude	Continuous	Coefficient of variation below threshold
I_PoliceStation_latitude	Continuous	Coefficient of variation below threshold
I_Document_Status	Nominal	Single category too large
I_CrimeScene_longitude	Continuous	Coefficient of variation below threshold
I_CrimeScene_latitude	Continuous	Coefficient of variation below threshold
I_Crime_Timing	Nominal	Single category too large

4.2.2.2 Feature Selection By Random Tree

The figure 6 shows the important features that have been selected by the random tree algorithm and where calculated to reveal that they are related to the target (Repeated crime).

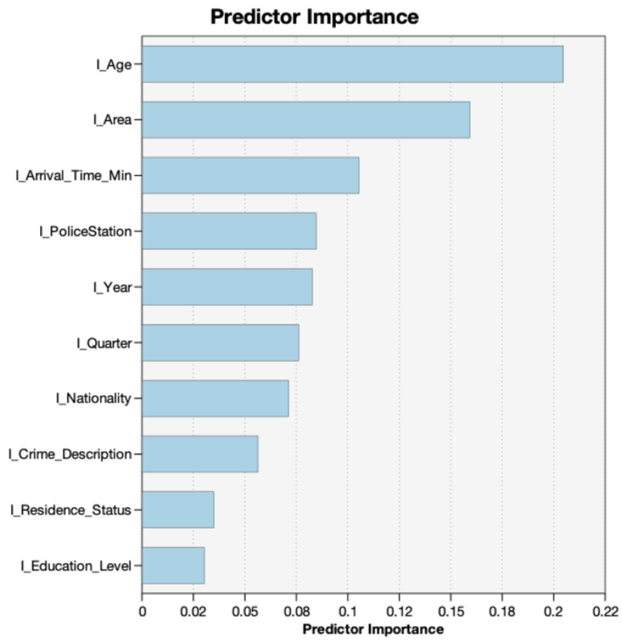


Figure 6: Predictor importance

4.2.3 Visualization Of Key Features

4.2.4.1 Statistical Analysis

1. Area Distribution Of Car Plate Theft

The information that describes the numbers helps us understand how different and regular the thefts of car license plates are. These results will guide the next steps of our study, including making predictions and creating ways to prevent crime from happening. The number of people from different areas involved in crime varies, with Jebel Ali Industrial area having the most (1,751) and Al Satwa being lowest at 8. This information is very important because it helps find places where car plate theft might happen and know how risky each nationality's involvement with the crime could be, therefore, can be used to train our model efficiently, knowing countries more likely to be robbed than the least.

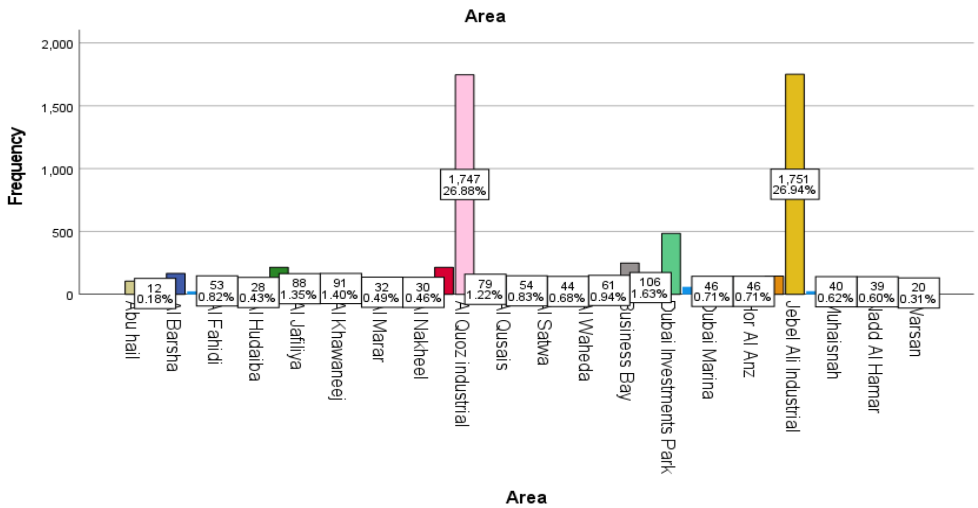


Figure 7: Area Distribution of Car Plate Theft

2. Residence Status Distribution

Most people involved in stealing car plates have different statuses, with most being what's called 'overstay' (4072), others as active, and some few are termed residents. It shows the importance of dealing with problems linked to people who stay too long and use actions aimed at stopping this group.

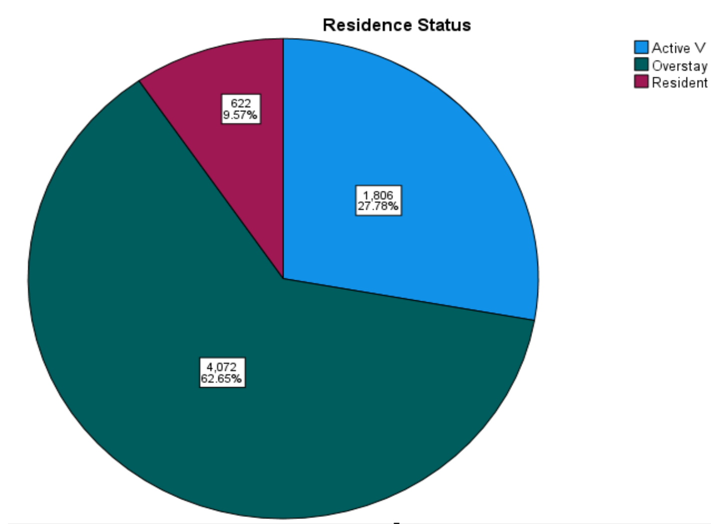


Figure 8: Residence Status Distribution

3. Entry Method Into The Country

Most criminals came into the country legally (5600), showing how important it is to check legal entry points closely. But we need to increase our border control because of people involved in stealing car plates by smuggling (900).

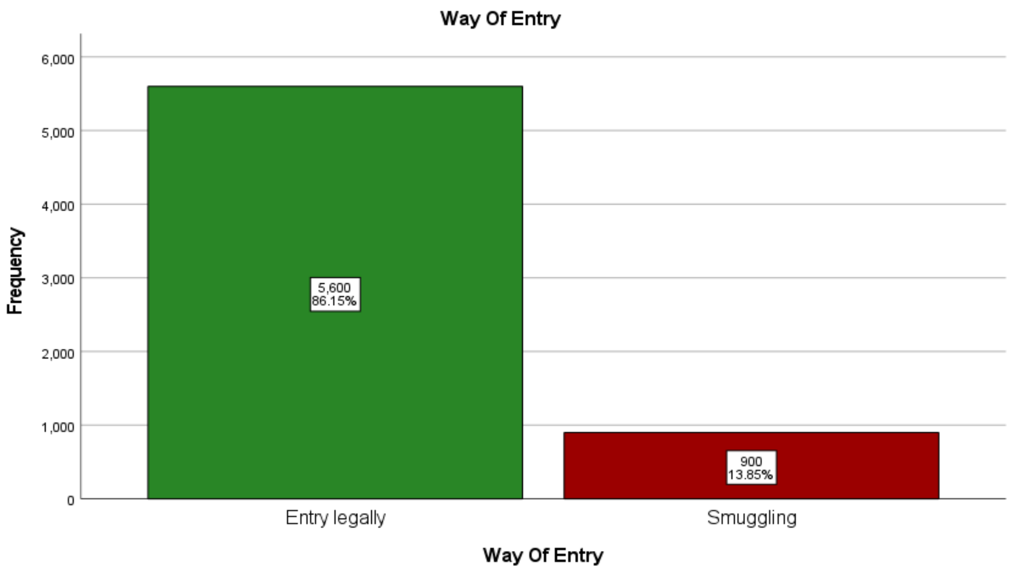


Figure 9: Entry Method into the Country

4. Entry Legality

Using fake passports in stolen car plate cases makes us worry about how well we check and verify important documents. Majority of individuals that entered the country legally (3713), which is 57.12%, committed crime more than 2 times and 1,887, representing 29.03%, committed less than 2 times while majority of criminals that entered the Country illegally through smuggling (648) committed less than 2 crimes as opposed to the ones that committed more than 2 crimes (252). Fixing fake paper work is very important for making safety better and stopping bad actions.

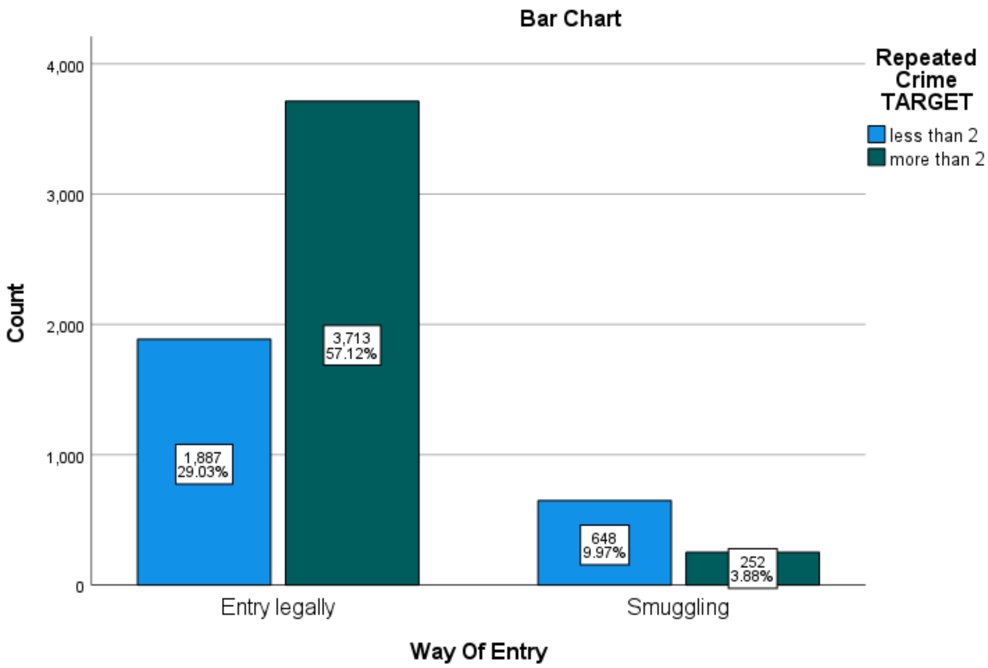


Figure 10: Entry Legality

5. Education Level

A big group of people who steal car plates are illiterate (2811), with a majority of them (1,820) who have committed car plate theft more than 2 times. This might be connected to the level of schooling and how it affects criminal actions. Knowing about these connections can help create special awareness and action plans, where education can be used as a tool against crime in general.

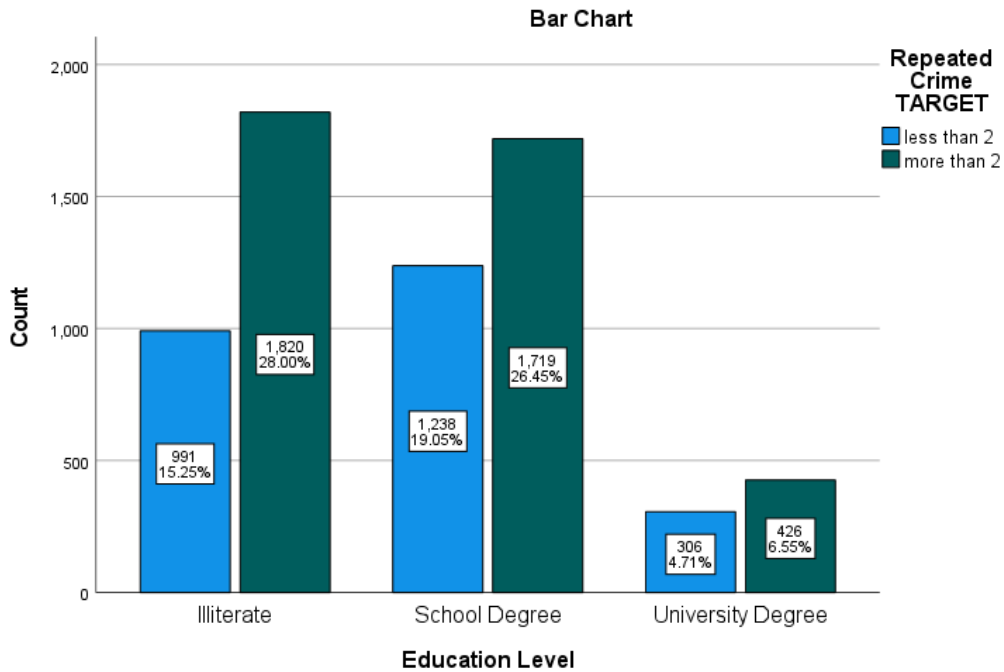


Figure 11: Education Level

6. Employment Status

The job status of car plate thieves is even. Both working (3727) and non-working people (2773) take part in the bad actions. This means we need big plans that deal with the reasons behind stealing car plates. These usually involve money and social issues.

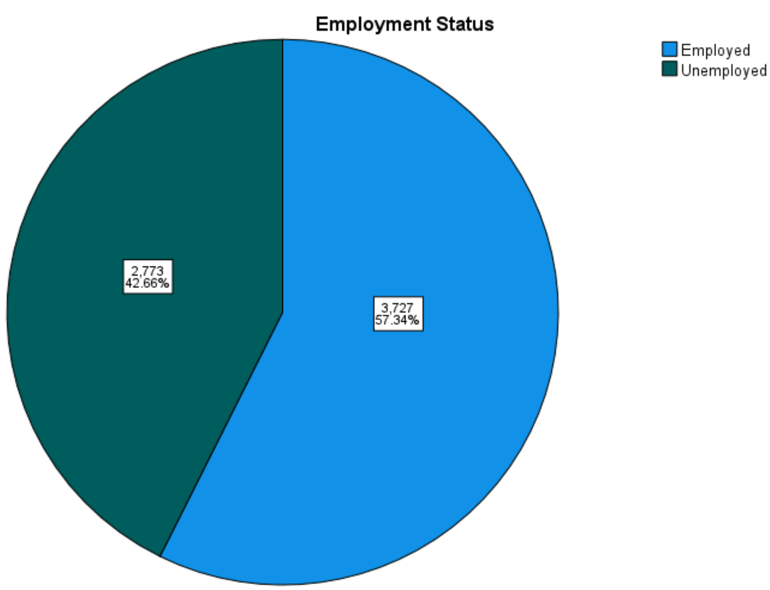


Figure 12: Employment Status

7. Type Of Car Used

The study on car plate theft shows that both old (3037) and used (3463) cars are evenly involved. This knowledge can help law enforcement understand what criminals like more and use special ways to stop them beforehand.

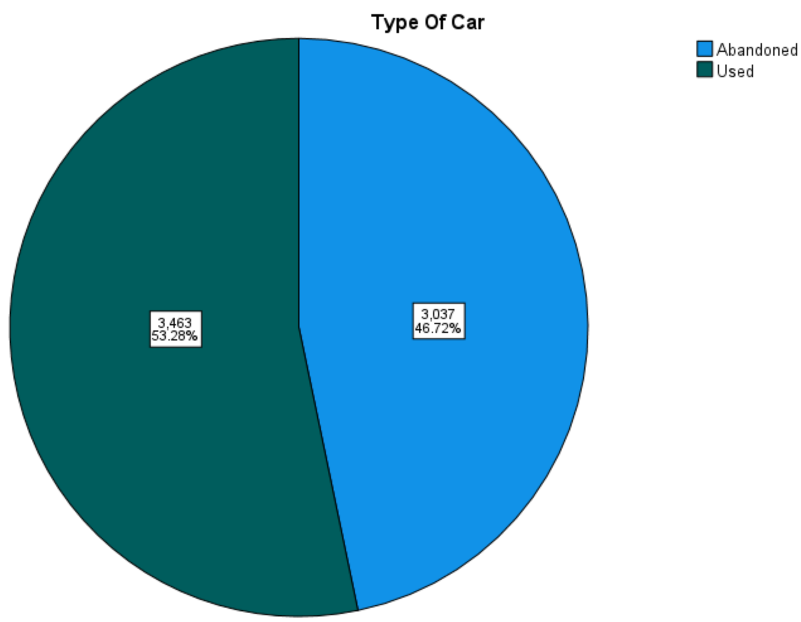


Figure 13: Type of Car used

4.2.4.2. *C*rosstab And Chi Square

1. Repeated Crime Patterns And Area Crime Was Committed

The cross-tabulation analysis reveals valuable insights into the relationship between two categorical variables: "Area" and "Repeated Crime Target," whose statistical tests conducted, including the Pearson Chi-Square and Likelihood Ratio tests, yielded highly significant results, indicating a strong association between the variables under consideration.

Table 5: Crosstabs Case Processing Summary of Repeated Crime Patterns and Area Crime was Committed

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Area * Repeated Crime TARGET	6500	100.0%	0	0.0%	6500	100.0%

A Pearson Chi-Square statistic yielded a value of 844.784 with 38 degrees of freedom, while the Likelihood Ratio test produced a chi-square value of 967.092 with the same degrees of freedom, and both tests returned p-values of less than 0.000, indicating a high level of statistical significance. These findings suggest that the occurrence of repeated crimes, as indicated by the "Repeated Crime Target" variable, varies significantly across different areas, as defined by the "Area" variable. Despite the overall significance of the relationship, it's important to note that the analysis identified 3 cells (3.8%) with expected counts less than 5, with the minimum expected count being 3.12.

Table 6: Chi-Square Tests of Repeated Crime Patterns and Area Crime Committed

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	844.784[a]	38	.000
Likelihood Ratio	967.092	38	.000
N of Valid Cases	6500		

a. 3 cells (3.8%) have an expected count of less than 5. The minimum expected count is 3.12.

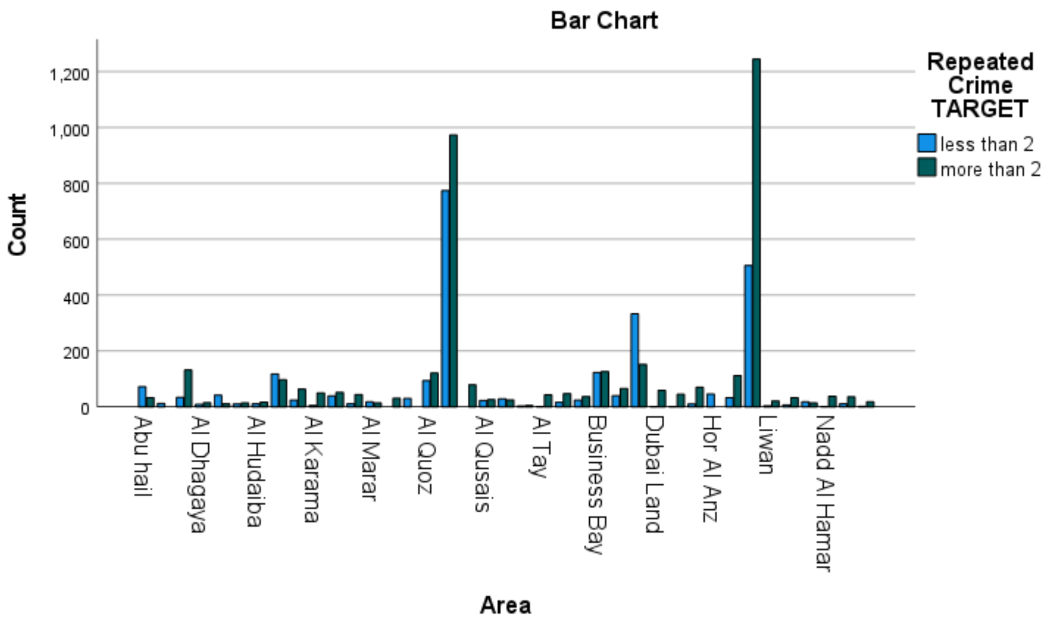


Figure 14: Repeated Crime Patterns and Area Crime Committed

2. Potential Crime Patterns And Crime Description

Such an increase in the potential weakness in the chi-square test for these cells may, however, be undermined by the large sample size of 6,500 which allows for the validity of the entire study. Still, it must be kept in mind that any results based on low expected cells should be taken cautiously so as not to jeopardize the credibility of the interpretation reached. Both the cross-tabulation and the descriptive statistics demonstrate a very strong correlation between the area and the repeat offenses and this undoubtedly emphasizes the fact that geography plays a very critical role in understanding the behavior of criminals. Further research into certain attributes and behaviors of different districts may give law enforcement agencies and policymakers invaluable knowledge critical for creating specific strategies for crime prevention and crime reduction to promote public safety. The Pearson Chi-Square value was 488.395 and was computed with 9 degrees of freedom. The p-value was 0.000 which was statistically significant, while the Likelihood Ratio value was 540.628 and computed with 9 degrees of freedom, and the p-value was also 0.000 which made it statistically significant These results carry for the definite relationship between Crime Description and the probability of multiple crimes.

Table 7: Crosstabs Case Processing Summary of Crime Description and Repeated Crime Rate

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Crime Description * Repeated Crime TARGET	6500	100.0%	0	0.0%	6500	100.0%

The Case Processing Summary revealed that all cells had expected counts that were greater than 5, the smallest expected count being 23.01, thus indicating that cell distribution was suitable for conducting the Chi-Square analysis and improving the power of the results.

Table 8: Chi-Square Test of Crime Description and Repeated Crime Rate

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	488.395[a]	9	.000
Likelihood Ratio	540.628	9	.000
N of Valid Cases	6500		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 23.01.

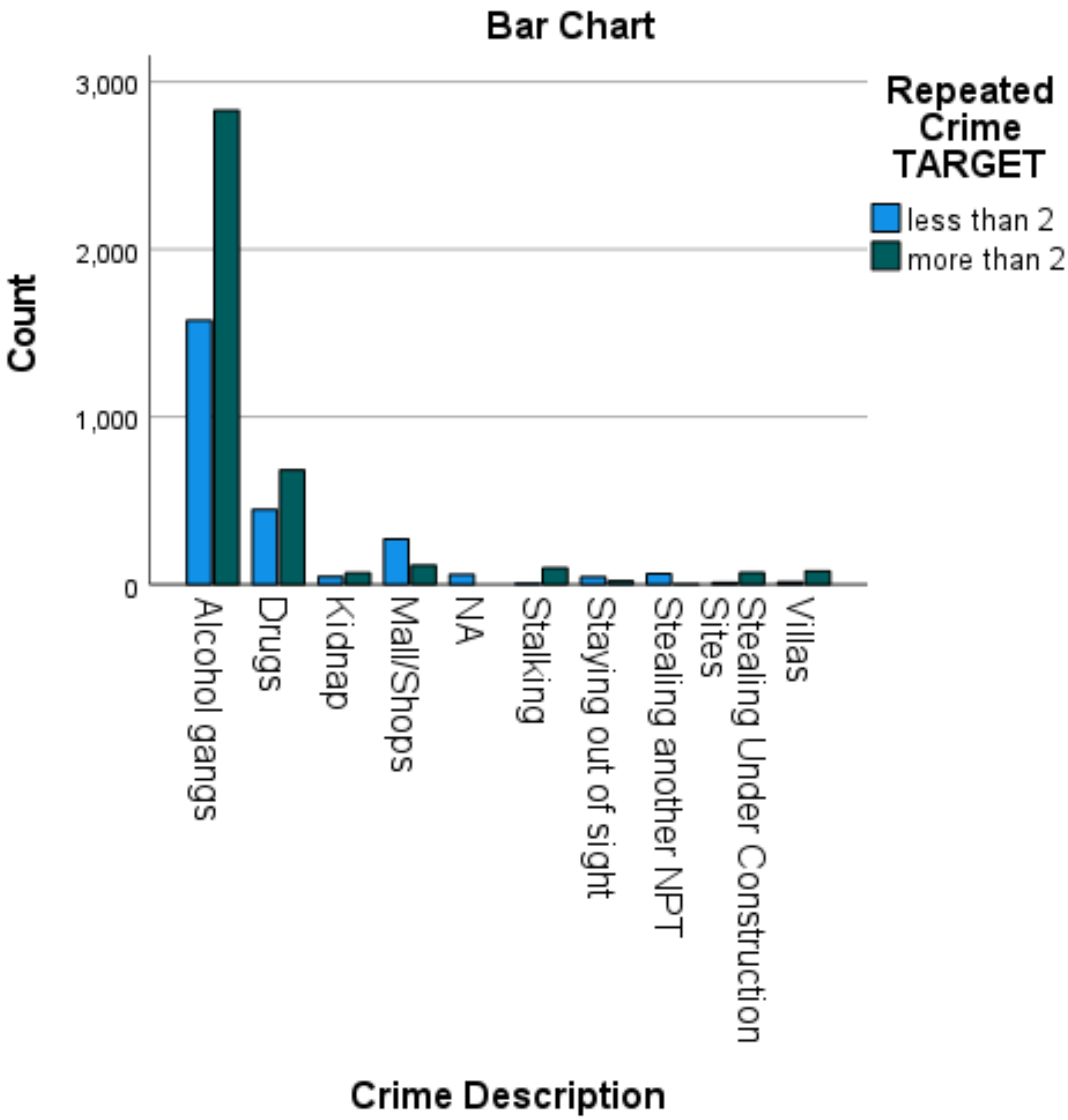


Figure 15: Visualization of Crime Description and Repeated Crime Rate

3. Repeated Crime Patterns And Type Of Car

Whereas the "Repeated Crime TARGET" variable may reveal whether or not a crime has occurred within a given period or geographic region, the "Type of Car" data probably contains of categories identifying various car kinds or models. The analysis's goal was to find any correlation between the kind of automobile and the frequency of recurrent crimes. Numerous statistical tests, such as the Fisher's Exact

Test, the Likelihood Ratio Test, and the Pearson Chi-Square Test, were run to evaluate the relationship between the two variables. These tests are commonly used to determine if there is a significant association between categorical variables.

Table 9: Crosstabs Case Processing Summary of Type of Car and Repeated Crime

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Type Of Car * Repeated Crime TARGET	6500	100.0%	0	0.0%	6500	100.0%

With just one degree of freedom, the Pearson's Chi-Square test came up with a chi-square value of 2.732 that is equal to p-value of 0.098. On the same note, the Likelihood Ratio test gave a p-value of 0.098 and chi-square value of 2.734 with one degree of freedom. At a 0.05 significance level, the p-values prove that there is no statistically significant relationship between the kind of vehicle and the frequency of repeated offenses.

Table 10: Chi-Square Test of Type of Car and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.732[a]	1	.098		
Continuity Correction[b]	2.649	1	.104		
Likelihood Ratio	2.734	1	.098		
Fisher's Exact Test				.103	.052
N of Valid Cases	6500				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 1184.43.
b. Computed only for a 2x2 table

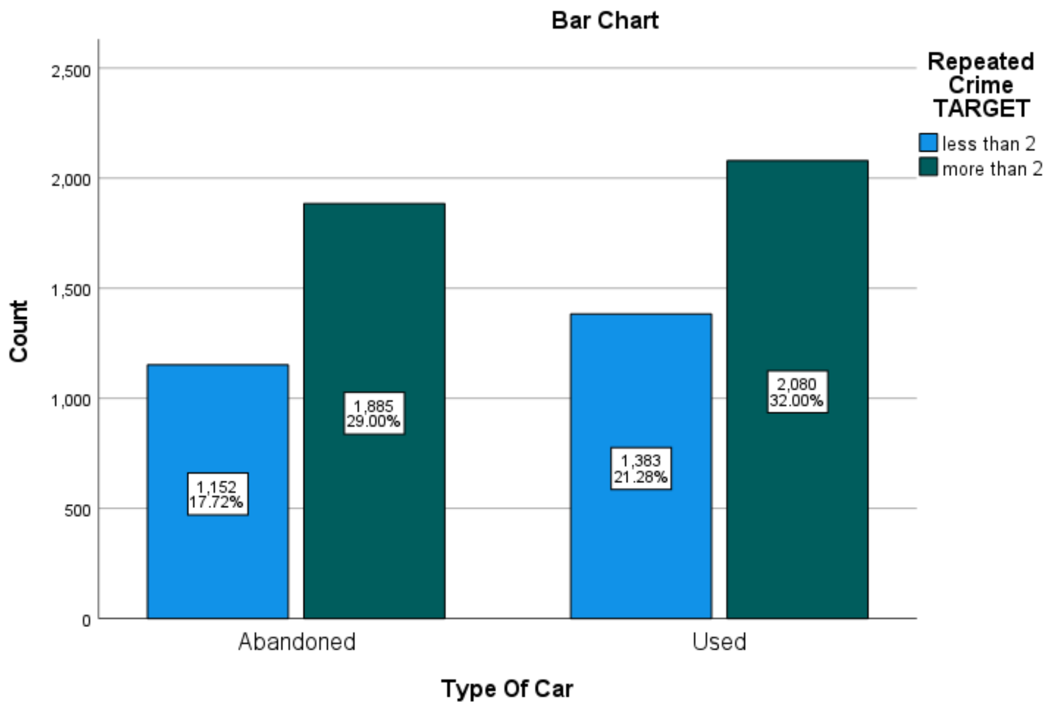


Figure 16: Visualization of Type of Car and Repeated Crime

4. Repeated Crime Patterns And Way Of Entry

A more cautious evaluation of the connection was obtained using Fisher's Exact Test, which is especially useful for smaller sample sizes or low anticipated cell counts. A one-sided p-value of 0.052 and a two-sided p-value of.103 were obtained from the test. The p-values are somewhat higher than the traditional significance level of .05, but they are still insufficient to disprove the null hypothesis that the two variables are independent. With a minimum anticipated count of 1184.43, every cell in the contingency table has an expected count greater than 5. This increases the dependability of the results by guaranteeing that the chi-square tests' underlying assumptions are satisfied.

Table 11: Chi-Square Test of Way of Entry and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	478.191[a]	1	.000		
Continuity Correction[b]	476.582	1	.000		
Likelihood Ratio	469.594	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	6500				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 351.00.

b. Computed only for a 2x2 table

The Chi-Square test revealed a significant connection between the ways of entry and repeated crime incidences ($\chi2 = 478.191$, $p < 0,001$). This underlines the fact that whether a person gains access to a place through a specific method is of great significance in the face of the recurrence of crime. The visualization of the linkage demonstrates the distribution of repeated crimes at various attack arenas, thus, the critical need to observe and secure these vulnerable routes to counter criminal activities.

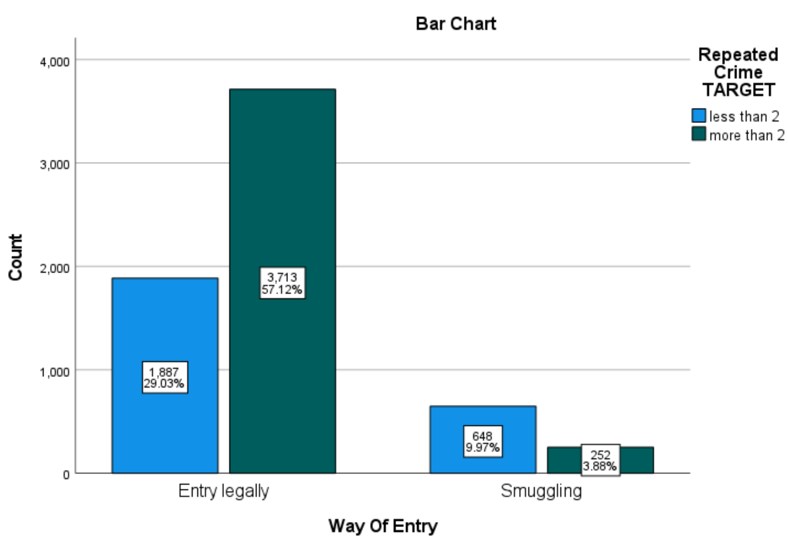


Figure 17: Visualization of Way of Entry and Repeated Crime

5. Repeated Crime Patterns And Nationality Of The Criminal

Likewise, the Chi-Square test shows a noticeable connection between nationality and the crime rate ($\chi 2 = 413.553$, $p < 0.001$) hence proving how nationality is a factor in the propensity for returning to crime, with some groups of people being more often involved in illegal activities. The nationality distribution map shows, which groups are often involved in repeated crime; because of that, better-targeted interventions and law enforcement strategies can be developed.

Table 12: Chi-Square Test of Nationality and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	413.553[a]	14	.000
Likelihood Ratio	457.018	14	.000
N of Valid Cases	6500		

a. 8 cells (26.7%) have expected count less than 5. The minimum expected count is 1.95.

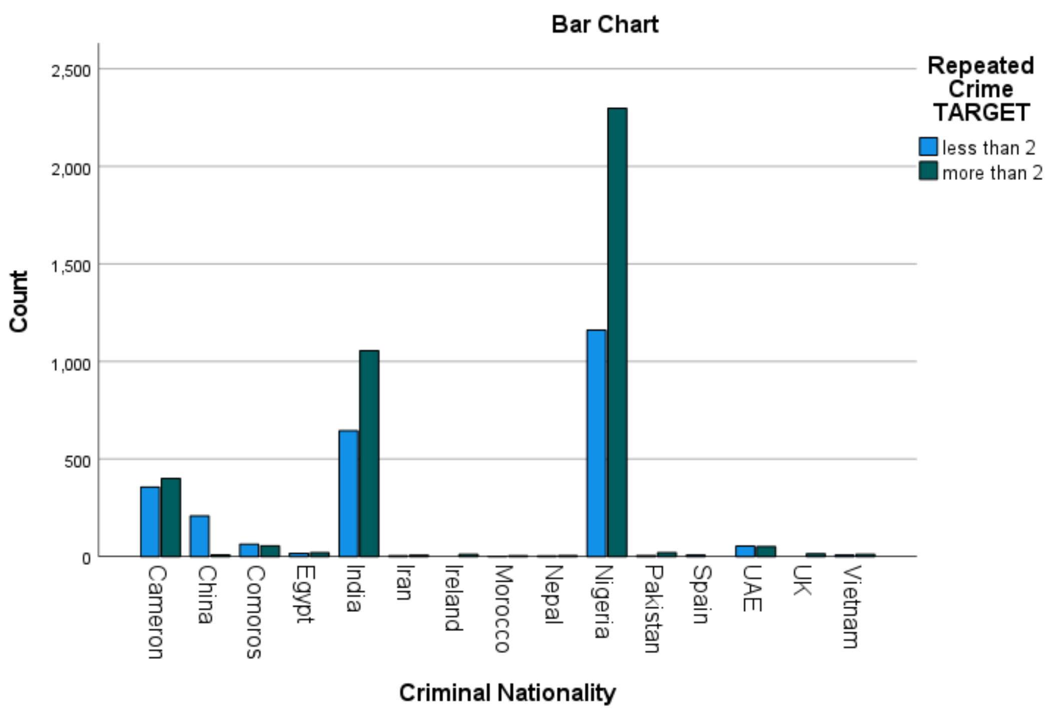


Figure 18: Visualization of Nationality and Repeated Crime

6. Repeated Crime And Police Station

The results of the Chi-Square test saliently bear out the relation between police stations and recurring crime occurrences ($\chi2 = 187.765$, $p<0.001$), indicating that law enforcement presence and effectiveness have a mitigating effect on repeated acts of crime. The visualization of this connection gives the geographic span of the crime hotspots occurring near different police stations which makes some regions warn attention and more policing efforts to be placed there.

Table 13: Chi-Square Test of Police Stations and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	187.765[a]	8	.000
Likelihood Ratio	186.597	8	.000
N of Valid Cases	6500		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 52.65.

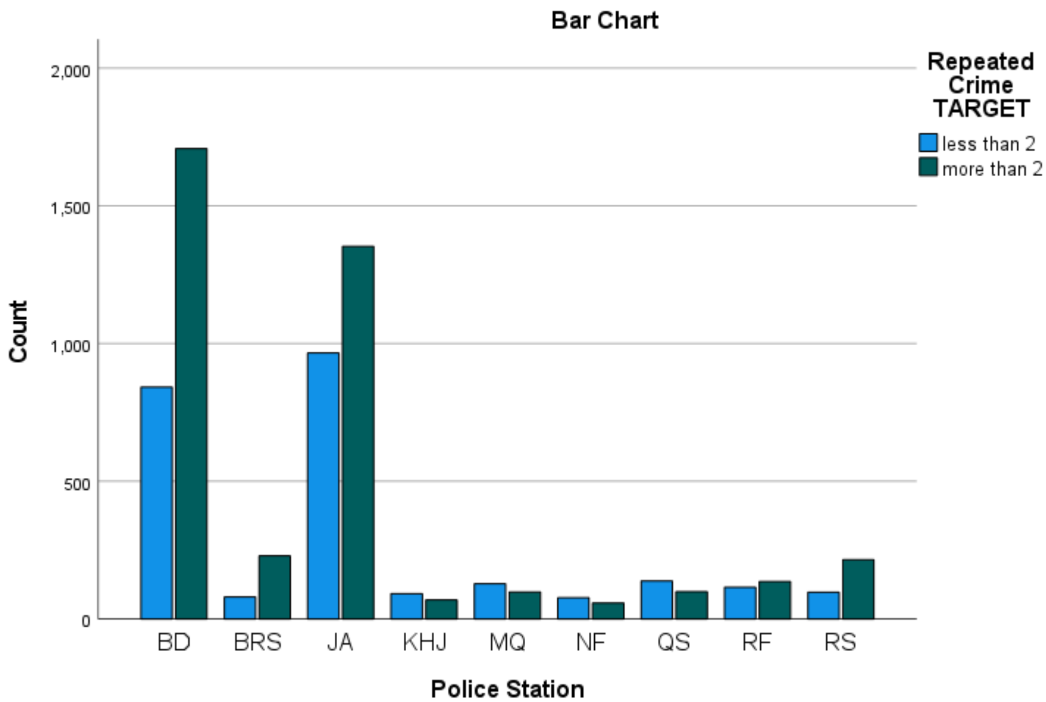


Figure 19: Visualization of Police Station and Repeated Crime

7. Repeated Crime Patterns And The Year When Crime Was Committed

We have analyzed and found the annual and crime occurrences relation to have a one-tailed independence chi-square of 85.905 with the significance level as ($p < 0.001$). Such is illustrated via connection to a specific crime that tends to be easy on occasion but at the same time becomes more serious at another. The purpose of this visualization is to capture the association between the temporal trends that are related to repeated crime, which can enable the identification of temporal patterns and thus information for the deployment of resources as well as interventions that can be used for mitigating the repeating criminal behaviors.

Table 14: Chi-Square of Year and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	85.905[a]	4	.000
Likelihood Ratio	84.738	4	.000

N of Valid Cases	6500		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 44.07.

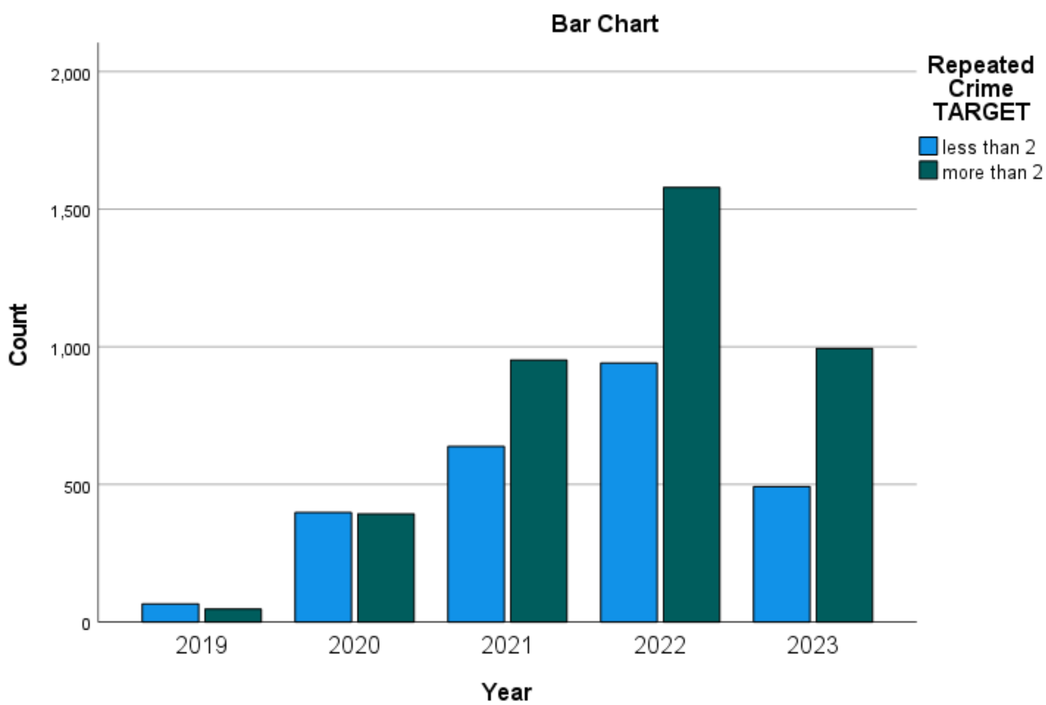


Figure 20: Visualization of Year and Repeated Crime

8. Repeated Crime Patterns And Education Level Of The Criminal

Table 15: Chi-Square Test of Education Level and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	29.210[a]	2	.000
Likelihood Ratio	29.320	2	.000
N of Valid Cases	6500		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 285.48.

The Chi-Square test highlights the existence of a statistically significant relationship between education level and repeated crime occurrences ($\chi2 = 29.210$, $p < 0.001$), which manifests itself in the fact that a higher educational level tends to decrease the probability of engaging in repeated criminal activities. The display of educational level distribution compared to repeated crimes provides social and economic factors that influence repeated criminal behavior, so the targeted interventions that aim at the socio-economic disparities can be formed.

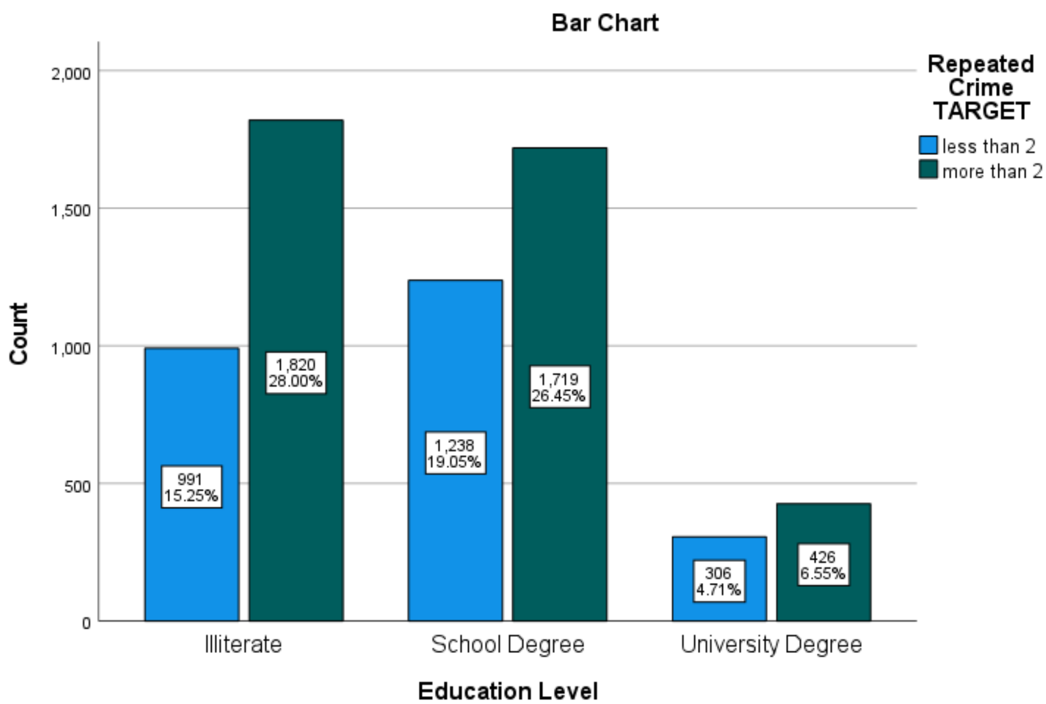


Figure 21: Visualization of Education Level and Repeated Crime

9. Repeated Crime Patterns And Residence Status

The result involving the Chi-Square test showed a significant association between residence status and again crime incidents ($\chi2 =32.161$, $p < 0.001$) indicating the role of residential status in the likelihood of repeat criminal activities. The depiction of home addresses for individuals with habits of offense to recidivism indicates where people related to repeated criminal behaviors and pinpoint specific areas for interventions to address neighborhood problems in communities.

Table 16: Chi-Square Test of Residence Status and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	32.161[a]	2	.000
Likelihood Ratio	33.405	2	.000
N of Valid Cases	6500		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 242.58.

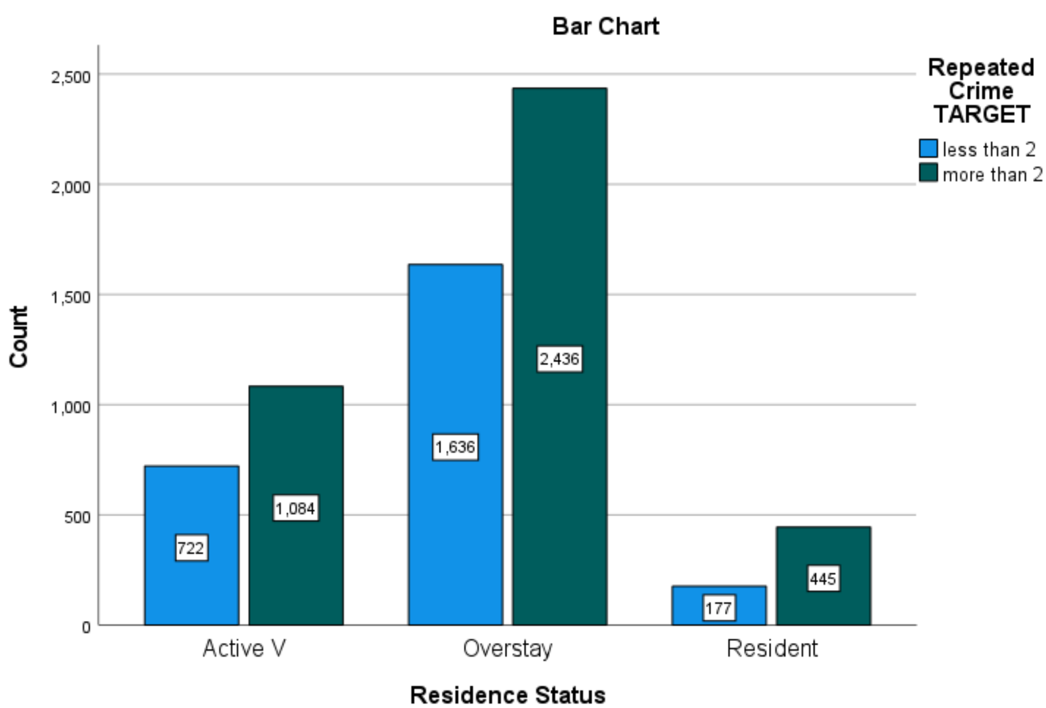


Figure 22: Visualization of Residence Status and Repeated Crime

10. Repeated Crime Patterns And Criminal Status

During the analysis of how frequently criminal status is associated with the recurrence of these crimes, it becomes evident ($\chi2 = 15.111$, $p < 0.001$) that there is a significant influence of previous criminal history on the likelihood of participating in repeated criminal acts. The display of two distinct clusters – one representing low recidivism with a conservative distribution of criminal status and the other with high recidivism, which is skewed toward a higher share of criminals indicate the significance of preventative strategies and rehabilitation programs for the reduction of repeated delinquency.

Table 17: Chi-Square Test of Criminal Status and Repeated Crime

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	15.111[a]	1	.000		
Continuity Correction[b]	14.877	1	.000		
Likelihood Ratio	14.989	1	.000		
Fisher's Exact Test				.000	.000
N of Valid Cases	6500				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 584.61.

b. Computed only for a 2x2 table

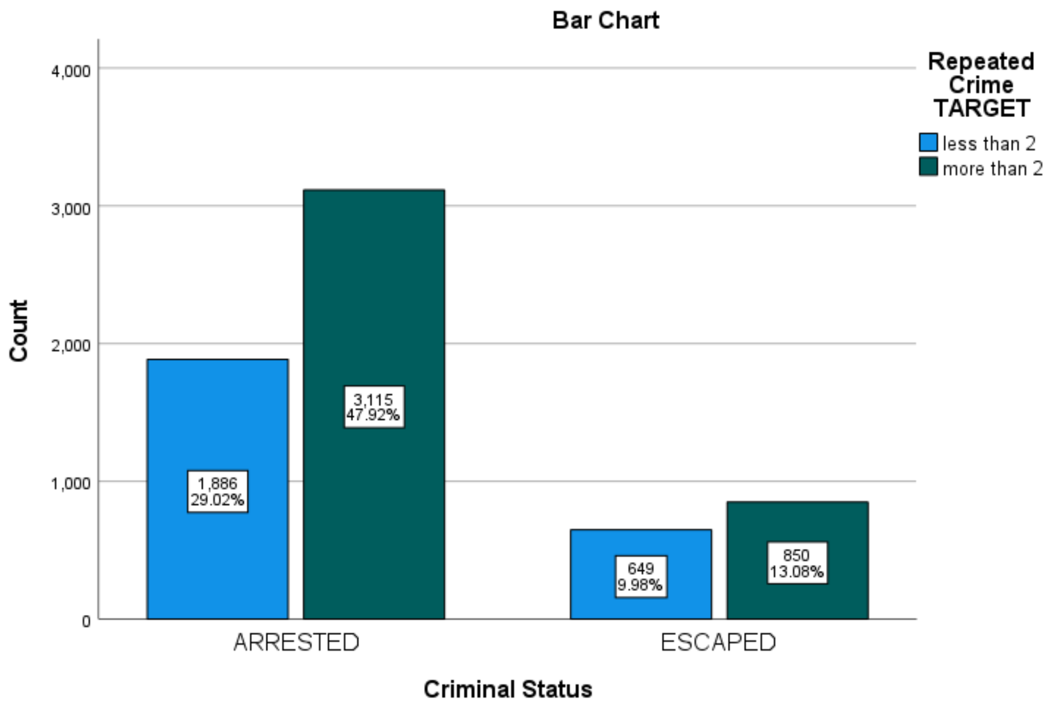


Figure 23: Visualization of Criminal Status and Repeated Crime

4.2.4.3 Means Table Distribution

These tables display the mean values for specific variables that serve as predictors and are grouped by the category variable, which differentiates instances of repeated crime incidences when the number is less than two and instances when the number is more than two.

1. Mean Of Arrival Time Of Repeated Crime

As Table 15 highlights, the presence of repeated crime incidents convergence is reflected by the average time they tend to appear. The average of addresses for reported incidents with no more than 2 repeated minor crimes over the period is 14.73 with a standard deviation of less than 5, but the same for major crimes is a bit higher, i.e. 15.14, with a standard deviation of less than 5. It is generally summarized as 14.98 for all the instances during the overall statistic on the arrival time. Providing another argument in favor of the hypothesis that more continuous criminal acts are more probably to happen towards the end of the day, the negative association of the number of repeated crimes with the hour of the day.

Table 18: Mean of Arrival Time of Repeated Crime

Report

Arrival Time

Repeated Crime TARGET	Mean	N	Std. Deviation
less than 2	14.73	2535	4.764
more than 2	15.14	3965	4.854
Total	14.98	6500	4.823

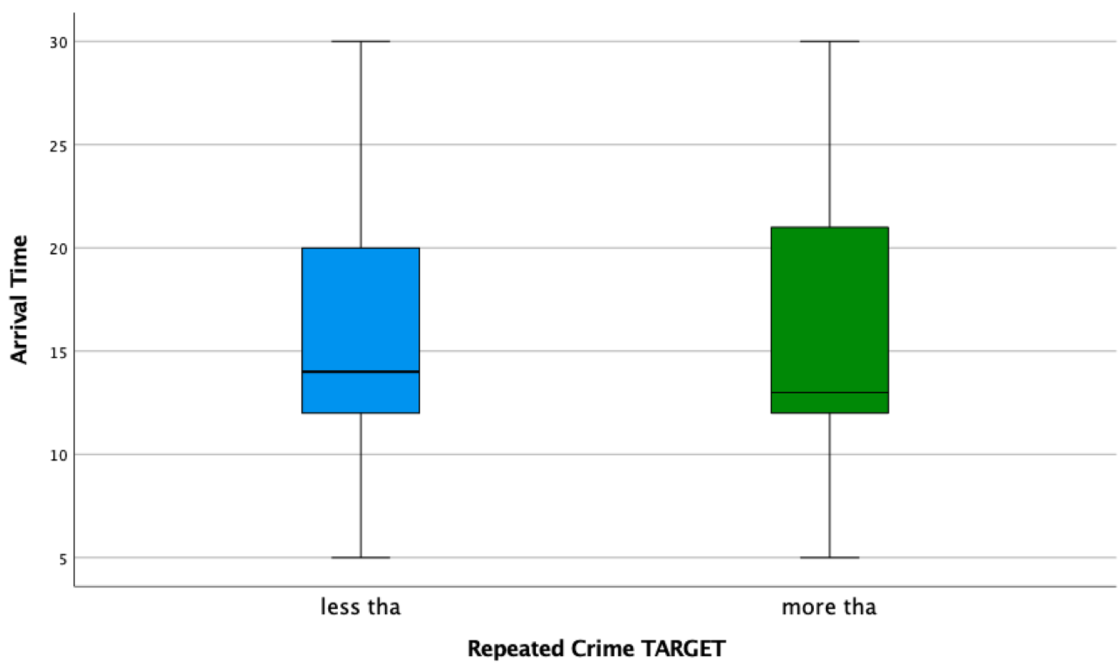


Figure 24: Visualization of Arrival Time of Repeated Crime

2. Mean Of Quarter Of The Year Of Repeated Crime

Table 19 shows the shift between the seasons for the crime incidence concentration. The means of the quarter of the year in the cases with less than 2 crimes that happened for more than a couple of months is 2.32 and its standard deviation is 1.159, however, in cases where there are more than 2 crimes that are repeated, you might find that the mean quarter of the year is a little higher with a standard deviation of 1.053 than those where there The 2.45 mean quarter which is the overall mean across all instances for the year is the answer. Repeating cases experiencing a high number of repeats mostly occur later in the year compared to otherwise similar incidents with the repeated ones towering at lower levels.

Table 19: Mean of Quarter of the Year of Repeated Crime

Quarter Of The Year

Repeated Crime TARGET	Mean	N	Std. Deviation
less than 2	2.32	2535	1.159
more than 2	2.53	3965	1.053
Total	2.45	6500	1.101

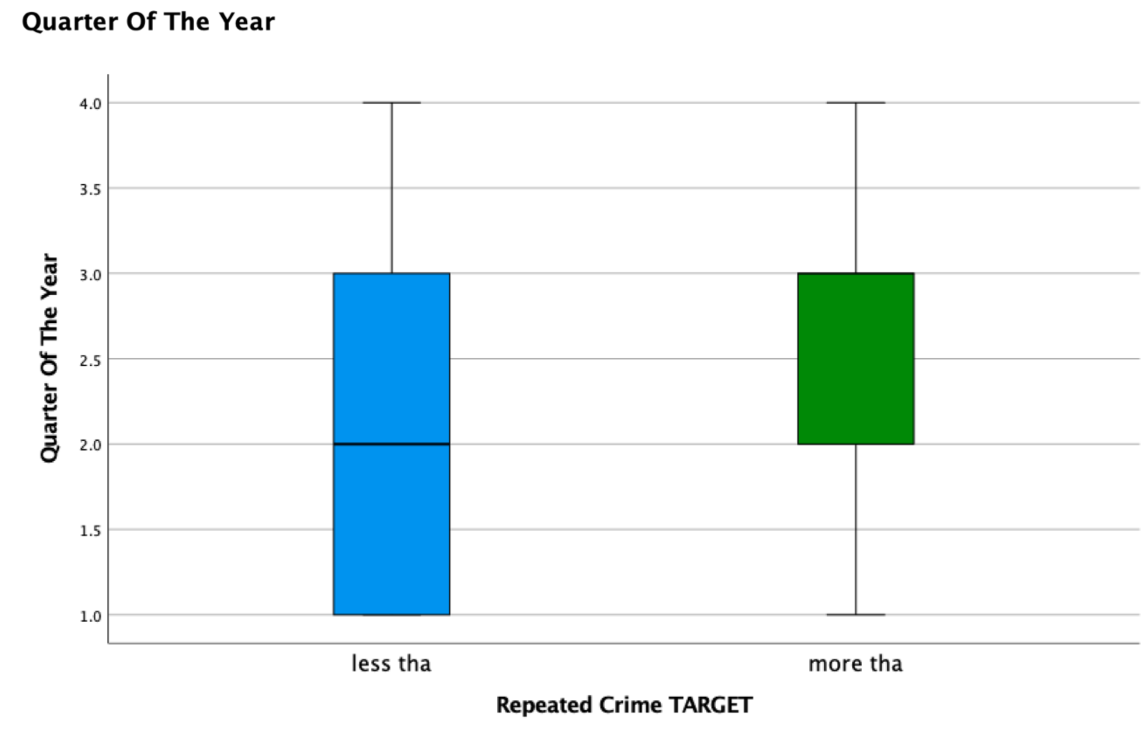


Figure 25: Visualization of Arrival Time of Repeated Crime

4.3 Machine Learning Model Development

4.3.1 Detailed Explanation Of The Chosen Input

At the important point that we had in the second scenario of our research, the determination of feature selection and the result of the tree algorithm coincided together; thus, we had no choice but to combine the outcomes of the two approaches to boost the predictive functionality of the model. In the first place, the algorithm of the feature selection carefully analyzed our dataset, judging 15 as inputs that were seen connected to the target outcome which in this case was to the recurrence of crimes. However, a tree random algorithm identified at random 10 characteristics, which appeared to be particularly influential in forecasting the corresponding destination.

1. Scenario 1: All Features

Field	Measurement	Values	Missing	Check	
I_Year	Nominal	"2019","2020","2021","2022","2023"		None	Input
I_PoliceStation_longitude	Continuous	[25.004288417585563,25.271646...		None	Input
I_PoliceStation_latitude	Continuous	[54.924116255680566,55.529156...		None	Input
I_CrimeScene_longitude	Continuous	[24.956796727275297,25.293033...		None	Input
I_CrimeScene_latitude	Continuous	[54.8878826384748,55.68520766...		None	Input
I_Age	Nominal	19.0,20.0,21.0,22.0,23.0,24.0,25.0,...		None	Input
I_Nationality	Nominal	UK,UAE,Iran,India,China,Nepal,Sp...		None	Input
I_Quarter	Nominal	1.0,2.0,3.0,4.0		None	Input
I_Residence_Status	Nominal	"Active V",Resident,Overstay		None	Input
I_Number_Of_Crimes	Nominal	1.0,2.0,3.0		None	None
I_PoliceStation	Nominal	JA,BD,NF,RF,MQ,QS,RS,KHJ,BRS		None	Input
I_Area	Nominal	Liwan,Warsan,"Al Ras","Al Tay","N...		None	Input
I_Arrival_Time_Min	Nominal	5.0,6.0,9.0,10.0,11.0,12.0,13.0,14....		None	Input
I_Crime_Description	Nominal	NA,Drugs,Kidnap,Villas,"Staying o...		None	Input
I_Crime_Timing	Nominal	Day,Night		None	Input
I_Criminal_Status	Nominal	ESCAPED,ARRESTED		None	Input
I_Way_Of_Entry	Nominal	"Entry legally",Smuggling		None	Input
I_Document_Status	Nominal	"Fake passport","Real passport"		None	Input
I_Education_Level	Nominal	"School Degree",Illiterate,"Universi...		None	Input
I_Employment_Status	Nominal	Employed,Unemployed		None	Input
I_Type_Of_Car	Nominal	Used,Abandoned		None	Input
I_Camera_Footage	Nominal	NO,YES		None	Input
I_repeated_crime	Nominal	"more than 2","less than 2"		None	None
TR_Nationality	Nominal	1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,1...		None	None
TR_Area	Nominal	1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,1...		None	None
TR_Way_Of_Entry	Nominal	1.0,2.0		None	None
TR_Document_Status	Nominal	1.0,2.0		None	None
TR_New_repeated_crime	Flag	"less than 2"/"more than 2"		None	Target
XFCNew_repeated_crime	Continuous	[0.25572259008137116,0.914882...		None	None
Partition	Nominal	"1_Training","2_Testing"		None	Partition

Figure 26: Feature Selections of Scenario 1

2. Scenario 2: Automatic Using Features Selection And Random Tree

Field	Measurement	Values	Missing	Check	Role
I_Year	Nominal	"2019","2020","2021","2022","2023"		None	Input
I_PoliceStation_longitude	Continuous	[25.004288417585563,25.271646...		None	None
I_PoliceStation_latitude	Continuous	[54.924116255680566,55.529156...		None	None
I_CrimeScene_longitude	Continuous	[24.956796727275297,25.293033...		None	None
I_CrimeScene_latitude	Continuous	[54.8878826384748,55.68520766...		None	None
I_Age	Nominal	19.0,20.0,21.0,22.0,23.0,24.0,25.0,...		None	Input
I_Nationality	Nominal	UK,UAE,Iran,India,China,Nepal,Sp...		None	Input
I_Quarter	Nominal	1.0,2.0,3.0,4.0		None	Input
I_Residence_Status	Nominal	"Active V",Resident,Overstay		None	Input
I_Number_Of_Crimes	Nominal	1.0,2.0,3.0		None	None
I_PoliceStation	Nominal	JA,BD,NF,RF,MQ,QS,RS,KHJ,BRS		None	Input
I_Area	Nominal	Liwan,Warsan,"Al Ras","Al Tay","N...		None	Input
I_Arrival_Time_Min	Nominal	5.0,6.0,9.0,10.0,11.0,12.0,13.0,14....		None	Input
I_Crime_Description	Nominal	NA,Drugs,Kidnap,Villas,"Staying o...		None	Input
I_Crime_Timing	Nominal	Day,Night		None	None
I_Criminal_Status	Nominal	ESCAPED,ARRESTED		None	Input
I_Way_Of_Entry	Nominal	"Entry legally",Smuggling		None	Input
I_Document_Status	Nominal	"Fake passport","Real passport"		None	None
I_Education_Level	Nominal	"School Degree",Illiterate,"Universi...		None	Input
I_Employment_Status	Nominal	Employed,Unemployed		None	Input
I_Type_Of_Car	Nominal	Used,Abandoned		None	Input
I_Camera_Footage	Nominal	NO,YES		None	Input
I_repeated_crime	Nominal	"more than 2","less than 2"		None	None
TR_Nationality	Nominal	1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,1...		None	None
TR_Area	Nominal	1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,1...		None	None
TR_Way_Of_Entry	Nominal	1.0,2.0		None	None
TR_Document_Status	Nominal	1.0,2.0		None	None
TR_New_repeated_crime	Flag	"less than 2"/"more than 2"		None	Target
XFCNew_repeated_crime	Continuous	[0.25572259008137116,0.914882...		None	None
Partition	Nominal	"1_Training","2_Testing"		None	Partition

Figure 27: Automatic using feature selection and random tree

Feature selection outcomes were combined with the random tree selected features, and that is because the feature selection algorithm revealed that 15 inputs had a relationship to the target outcome, while the random tree had 10 outcomes. the 10 outcomes from the random tree are the same as 10 of the outcomes in feature selection. based on that, in scenario 2 feature selection and random tree outcomes were combined and used for scenario 2 inputs

4.3.2 Detailed Explanation Of The Chosen Machine Learning Algorithms

The application of precisely detailed Logistic Regression and Linear Support Vector Machine (LSVM) for scenario 1 that includes the selective feature points is due to their unique characteristics and the prominent capability of the Linear Support Vector Machine (LSVM) to deal with the linearly separable data as well as the performance of indexing of the instances into two classes. LSVM algorithm is a supervised technique aimed at identifying the best hyperplane that can split examples belonging to different classes with the greatest margin. In scenario one, which is the task to bin instances into either car plate theft or non-theft categories based on selected features, LSVM's ability to find favorable hyperplane separation is highlighted because LSVM can do that. The generalizes the separating hyperplane. This, in part, helps to alleviate the condition of noise or overlapping data points while maintaining a margin between the classes to be optimized. With this capability, LSVM is a case fit scenario 1 in which the main focus is to more accurately distinguish car plate theft and not theft incident in the case of a plate number theft.

In the case of scenario 2, feature selection may be viewed as a component with random tree algorithms and it enables the successful application of LSVM and Logistic Regression for the realization of the two different objectives despite being based on the two separate techniques with their awesome competence to run the binary classification chores. LSVM means Linear Support Vector Machine is considered an excellent way to label data in two classes and the feature of determining linearly separable data. The main goal of LSVM is to train a linear hyperplane that recovers the distinguished two classes and the biggest gap between them. Given the fact that such a gap almost always exists in the case of the linear separation of the classes. The combination of feature selection and random tree algorithms used in the described scenario will make the LSVM more effective in working with simple data like that used for car image recognition. This data allows the algorithm to efficiently distinguish car plate theft from other cases. These two methods are contrary to each other such as: increasing the margin in between classes of LSVM provides vital importance on the robustness of classifier, which will add hardness to noise or misalignment of data points. In the end, the best classification accuracy is possible only for LSVM models in the linearly separable patterns of the data, and the framework can back up the car plate theft prediction system more effectively for scenario 2.

4.3.3 Validation And Testing Procedures

1. Data Partition

Data partitioning is a vital component of the systematic evaluation of model performance which constitutes a significant element of model validation and the assessment of accuracy of our predictive models. This process involves the strategic division of our dataset into distinct subsets: the first group (the training data), and the second group (the testing data). Utilizing a careful strategy, we allot data into training and testing partitions; through this process, our models are trained using the first subset of data and tested on the other unseen data; thus, the resulting performance demonstrates a valid representation of the model behavior when facing real-life scenarios. The train-test split method is a core element of our data-partitioning strategy and it is based on the splitting of the dataset into two parts: the training set and the test set. What we do here is to allocate a certain percentage of the dataset to serve as the training sample which is exploited for the data training process. At the same time, a part of tabulated data reserved solely for validation of the results produced by our models is tried and tested. Observing the dataset into the two subsets of train and test sections, we ensure the models can work without the testing data during the training phase, apart from providing an unbiased evaluation process.

In the study, the data partitioning stage became a very important part of the evaluation and validation of the machine learning models. Beside the well known way we implemented the process of splitting a dataset into subsets for the use of both training and testing. The design of partitioning was carefully thought through to provide a balance between model training and testing, while also reducing overfitting and underfitting risks. The dataset was partitioned into two primary subsets: a training and a testing set. The training set consisted of 70% of the total dataset, while the testing set used the rest of 30%. We have made this partitioning decision considering the current best approach for machine learning model development which intends to channel most of the data for the training of the model to facilitate better model learning while the remaining portion is held for testing the models to measure their performance with unseen data.

With the train-test split technique, taken into account, cross-validation is the proceeding step that is meant to additionally strengthen the model evaluation process. Cross-validation involves segmenting the data into several subsets or folds, with each fold also acting as the validation set in turn while the remaining folds are utilized for training. This iterative approach assures that each datum has been used for both the training and the validation, thus resulting in a better utilization of all the data sources and a performance evaluation that is more thorough and based upon facts. In an endeavor to adopt an all-encompassing approach to data partitioning, which includes the train-test split technique and cross-validation, we implement different approaches to minimize the possibility of our predictive models falling into over-fitting and insufficient clinical applicability. We approach this task by precisely assigning data subsets dynamically and using model evaluations iteratively, thereby, allowing to make improvements in the performance of the model and leading to data-driven decision-making.

2. Data Balancing

The fundamental measure to succeed in this situation of class imbalance is to equalize the amount of data of target classes. Unbalanced data sets where a class's representation is substantially greater than that of the other class are called class imbalances, which result in model bias and minority classes being soon left out. This is why we spun around using different measures that help to bring balance while maintaining fairness among all classes. One such technique can be done by making use of the interesting strategy of oversampling of the minority class where instances from the minority class are either replicated or amplified to match or even overtake the oversupply of the majority class. The raising, by us, of the number of samples from the minority class, corrects the imbalance among the class distribution and helps the model to get accustomed to the less skewed dataset.

Our research, besides, focused on the topic of class imbalance inside the dataset which is one of the most frequent problems in classification tasks in cases where one class is very rare in comparison with other classes. We must use the data balancing and preprocessing pipeline in our machine learning model to eliminate the potential bias originating from the class imbalance and enhance the robustness of our system. More specifically, the undersampling method had been used to address the imbalance in the presence of several minority and majority classes in the dataset. Undersampling entails the reduction of the size of the majority of the class by randomly deleting examples of it until the number of instances in the majority class and those in the minority class are similarly balanced. Through the process of undersampling of majority class, we intended to achieve the objectives of bringing down the percentage of the majority class in the training set and preventing the model from biasedness of the majority class.

3. Evaluation Metrics Used To Assess Model Performance

Every metric has a special role in analysis, and it supplies essential data about various facets of model performance consequently helping to make the best decisions. Precision, which is calculated as following:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

is one of the key characteristics of models that represent how our algorithms estimate the forecasts they generate. This concept determines the percentage of right-classified inaccuracies in contrast to the total number of instances in the given dataset. Accuracy, is the result of the below formula:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}$$

And a high-accuracy score could be interpreted as the system predicting true values most of the time, which in turn shows how effective the model is in recognition tasks. The precision value as a vital metric gauges the portion of all true positive predictions that are released by the model among the whole positive predictions made by the model. This therefore measure gives us the idea of how accurate the positive predictions are which in turn gives insights about the model's efficiency in lowering false positive instances. A high precision score signifies that the model has a low count of false positive diagnostic predictions which makes it more believable and able to diagnose the true positive cases more accurately.

The Recall equation:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Recall and Precision are similar to well in the F-score:

$$F\ score = \frac{2 \text{ x precision x recall}}{\text{precision} + \text{recall}}$$

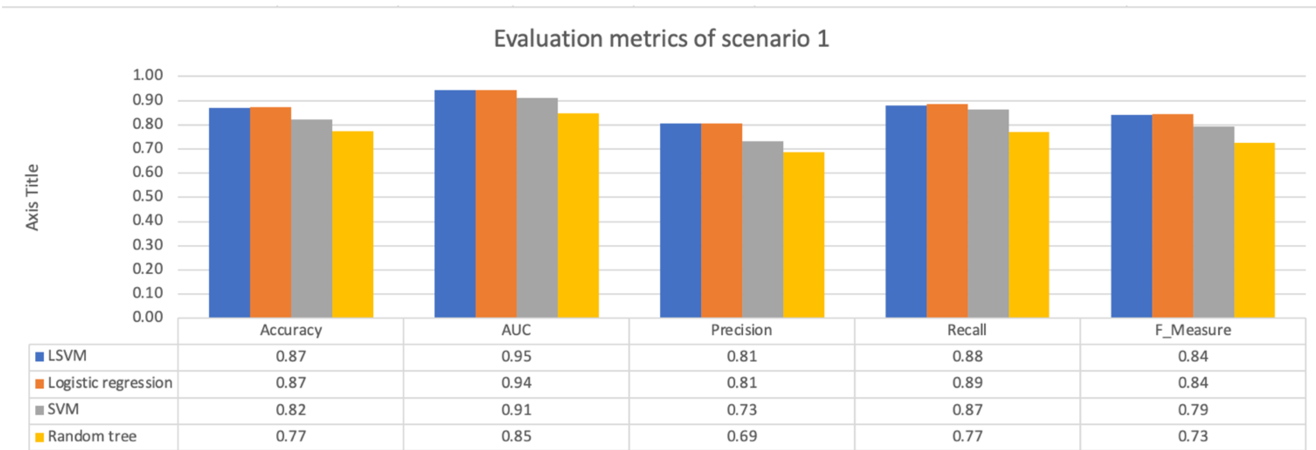
which is a harmony of them and shows an equally balanced measure of both points. It is supposed to not only generate the number of correctly identified samples as well as those falsely identified but also produce a holistic evaluation of the model in terms of recall and precision jointly. The F score is therefore very helpful in applications that contain a special situation where the ratio of positive samples to negative examples in the data is quite low since due to that it enables a fair compromise between precision and recall which in turn produces an acceptable measure of overall model performance. The AUC statistic performs a discriminatory power comparison of the model at different sensitive points. ROC curve is plotting true positive rate and false positive rate at various threshold values and AUC indicates the general performance of the model in discriminating positive and negative instances from each other. AUC values nearer to 1 show that the model better discriminates between positive and negative classes while higher values indicate better model performances.

4.3.4 Results

4.3.4.1 Presentation Of The Experimental Results

In scenario 1, an in-depth analysis of the models employed, their efficacy, and the pertinent considerations for selecting the optimal approach in predicting car plate thefts is undertaken through a comprehensive examination that draws upon the information encapsulated within the results, encompassing key attributes of the models, their respective performance metrics, and the visual representation of their Receiver Operating Characteristic (ROC) curves. The scenario at hand revolves around a dataset encompassing attributes pertinent to car theft incidents, inclusive of location data, vehicle specifics, and suspect information. In considering multiple machine learning models such as

Logistic Regression, LSVM, Random Trees, among others, an ensemble modeling approach is implied, suggesting a comprehensive evaluation to discern the optimal performer tailored to the dataset's complexities.



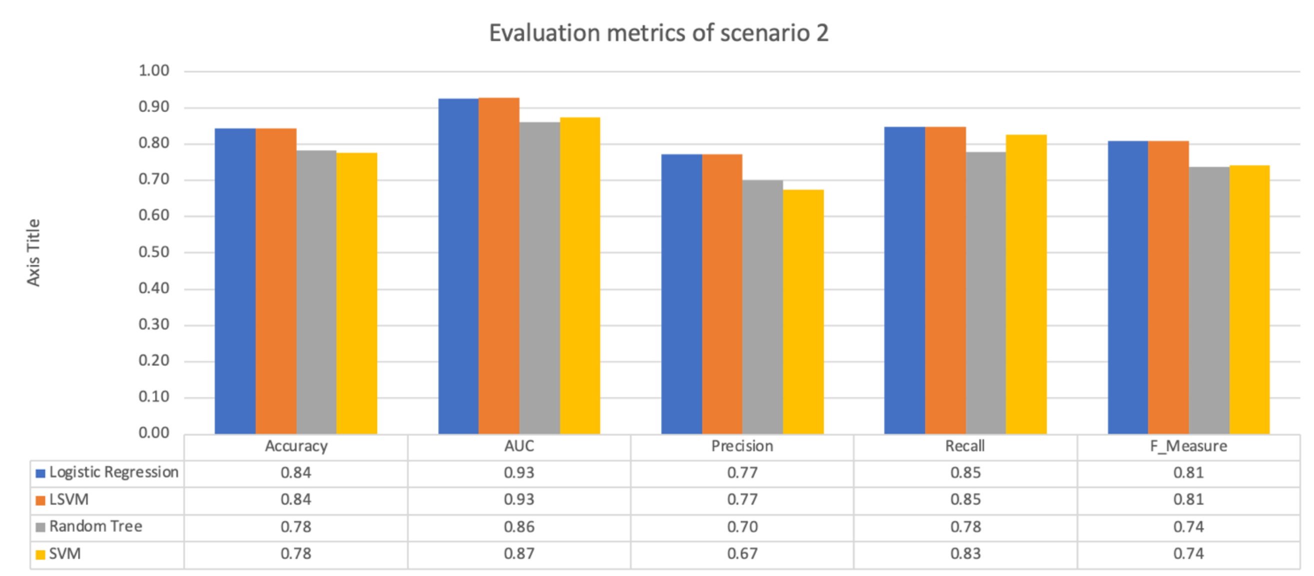
	Accuracy	AUC	Precision	Recall	F_Measure
LSVM	0.87	0.95	0.81	0.88	0.84
Logistic regression	0.87	0.94	0.81	0.89	0.84
SVM	0.82	0.91	0.73	0.87	0.79
Random tree	0.77	0.85	0.69	0.77	0.73

Figure 28: Scenario 1 Model Performance

Looking at the top models from Figure 28, Logistic Regression showcases an accuracy of 87% and an AUC of 0.94, indicative of its robust discriminatory capabilities. Its higher recall (0.89) underscores its proficiency in identifying actual car theft cases, while a commendable F1 measure (0.84) highlights a harmonious balance between precision and recall. Similarly, the LSVM, with an equal accuracy of 87%, displays a slightly higher AUC of 0.95 than the Logistic regression model.

Impression of the results of scenario 2 is a complete examination, without exception, which observes the models in detail, scrutinizes the metrics of performance, and unmasks the decisive factors that turn the wheel of choice in favor of the best model The Result Table also presents the main focus area for Scenario 2 - the evaluation of accuracy, precision, recall, and the F1 Score for LSVM and Logistic

Regression. These figures acted as a valuable gauge for us to measure the usefulness of each model based on their ability to distinguish positive from negative cases



	Accuracy	AUC	Precision	Recall	F_Measure
Logistic Regression	0.84	0.93	0.77	0.85	0.81
LSVM	0.84	0.93	0.77	0.85	0.81
Random Tree	0.78	0.86	0.70	0.78	0.74
SVM	0.78	0.87	0.67	0.83	0.74

Figure 29: Scenario 2 Model Performance

A peek at the Logistic Regression working in the analysis shows us its reasonable accuracy of 0.84. In other words, it comprehends almost half of all the cases correctly. With an accuracy of 0.77, one observes the precision of the model as well, staying aligned to the cases in which true cases have been identified as positive by the model. Moreover, on the recall dimension, too, Logistic Regression performs exceedingly and wears a top hat with a recall score of 0.85, which highly implies the precision of cases and the reduction of misclassified cases. We can add up to assessable measures the F score with a rate of 0.81 which proves the success of Logistic Regression in terms of processing both positive and negative examples.

An LSVM model presents with comparable level of accuracy as well but also has the predictive power of 0.84 to stay close to the Logistic Regression in terms of the performance metrics, and the precision of 0.77 shows similar metrics of performance as Logistic Regression. The LSVM becoming even with the

Logistic Regression in recall, is with the measure of 0.85, among the best at spotting true positives. Nevertheless, the value of 0.81 obtained via the F1, indicates that LSVM has achieved a less ideal performance where the precision was sacrificed in a bid for a higher recall.

The selection of the optimal model hinges upon various considerations tailored to the specific goals and priorities of scenario 2. If minimizing missed positive cases (false negatives) is paramount, Logistic Regression emerges as a preferable choice due to its superior recall. Conversely, if the misclassification of negative cases as positive carries significant consequences, then Logistic Regression's higher overall accuracy and precision render it a more suitable option. Insights into the specific data points used for prediction and conducting a feature importance analysis could elucidate the performance disparities between Logistic Regression and LSVM.

4.3.4.2 Evaluation Of Predictor-Importance

Predictor importance calculated and computed by SPSS modeler is by analyzing how each predictor has an impact on the target (Repeated crime Target) . There are different algorithms that SPSS modeler uses to compute predictor importance, C5.0, C&RT, QUEST, CHAID, Neural Network and GenLin, are examples of decision tree algorithms used by SPSS modeler, and employs sensitivity analysis to reveal the most importance predictors which has a relationship with the target variable (Repeated crime Target).(Saltelli et al.2004).

Notation:

Y= Target

$X_i = Predictor$

k = The number of predictors

$Y = f(X_1, X_2 \ldots X_k)$ = Model for Y based on predictors X_i through X_k

Variance Based Method:

The implied formula to compute or measure the predictors sensitivity:

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|X_i))}{V(Y)}$$

S = Predictors sensitivity

V(Y) = Unconditional output variance

E = Expectation operator

V = variance operator

Computing predictor importance to normalize sensitivity:

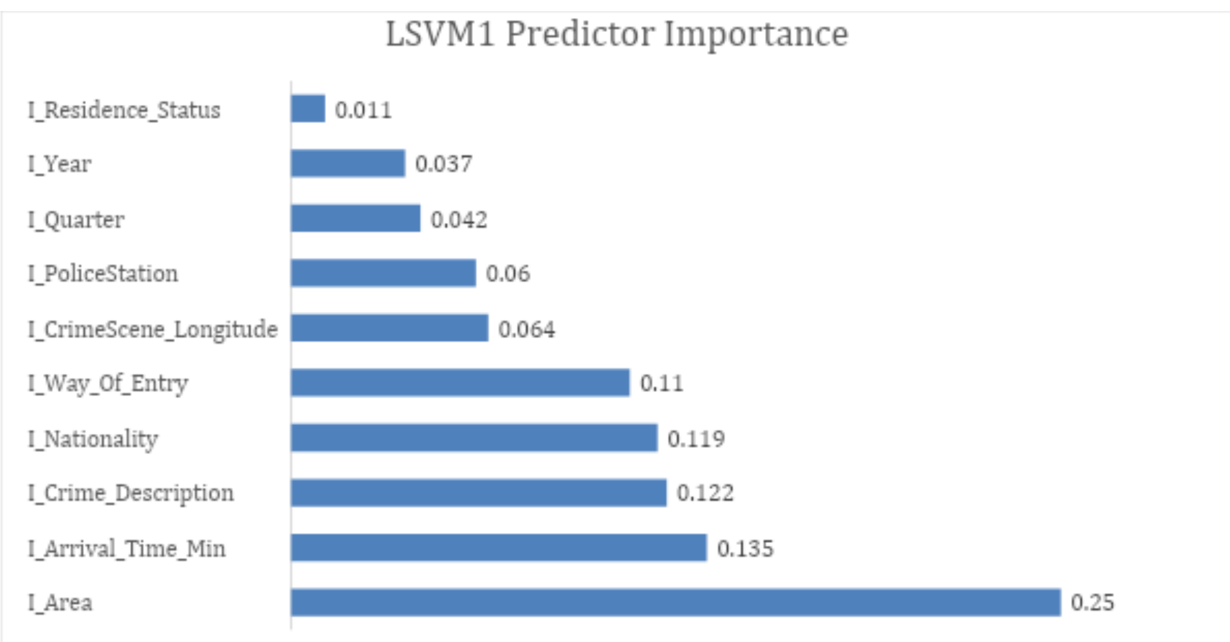
$$VI_i = \frac{S_i}{\sum_{j=1}^{k} S_i}$$

4.3.4.3 Predictor-Importance Of The Best Model

The LSVM (Linear Support Vector Machine) model from scenario 1 assigns importance sensitivity, as shown in Figure 30 to predictors based on their contribution to the overall predictive performance. According to the importance sensitivity attributed to predictors by the LSVM model, "I_Area" emerges as the most influential predictor with a sensitivity of 0.25, indicating that the area where the car theft incidents occur holds significant predictive value in determining the likelihood of such incidents. Following closely behind are "I_Arrival_Time_Min" and "I_Crime_Description" with importance sensitivity of 0.135 and 0.122, respectively.

"I_Nationality" and "I_Way_Of_Entry" also exhibit substantial importance sensitivity of 0.119 and 0.11, respectively, implying that the nationality of individuals involved and the method of entry into the crime scene contribute significantly to the predictive accuracy of the model. The predictors such as "I_CrimeScene_Longitude", "I_PoliceStation", and "I_Quarter" which are moderate in their importance sensitivity show that although these are apparently not that strong but still they have a meaningful effect on the model performance. Contrary to them, other predictors, for example, "I_Year" and "I_Residence_Status", have less impact concerning the LSVM model predictive accuracy given the sensitivity of 0.037 and 0.011, which technically means that they don't contribute to its predictive power as highly as others. Examining the functionality of the model is a part of the evaluation process and this includes obtaining corresponding metrics that are important performance indicators and are useful in comparing the model to the randomness, as well as the quality and speed of the training itself.

Figure 30: LSVM1 Predictor Importance



4.3.4.4 Hypothesis Testing

Based on the provided analysis, the discussion revolves around the hypothesis of the distribution of criminal time across repeated crime targets, the null hypothesis (H0) posits that the distribution of arrival time (criminal time) remains consistent across various categories of repeated crime targets, implying that time-based patterns of criminal activity are independent of a location's history as a crime target. The hypothesis is evaluated through the application of the Mann-Whitney U test, a non-parametric statistical method adept at comparing distributions of two independent samples. These samples likely represent arrival times for crimes occurring at locations that have not been previously targeted and those that have been targeted in the past.

Table 20: Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of Arrival Time is the same across categories of Repeated Crime TARGET.	Independent-Samples Mann-Whitney U Test	.749	Retain the null hypothesis.
2	The distribution of Quarter Of The Year is the same across categories of Repeated Crime TARGET.	Independent-Samples Mann-Whitney U Test	<.001	Reject the null hypothesis.
3	The distribution of Criminal Age is the same across categories of Repeated Crime TARGET.	Independent-Samples Mann-Whitney U Test	<.001	Reject the null hypothesis.

The test results indicate a significance value (Sig.) below the common threshold of 0.05 for statistical significance, leading to the rejection of the null hypothesis. In simpler terms, this suggests that there exists a statistically significant difference in the distribution of criminal time between the two groups—locations previously targeted and those not targeted before. Consequently, the interpretation suggests that time-based patterns of criminal activity may indeed vary depending on a location's history as a crime target. For instance, criminals might exhibit a tendency to target previously hit locations at specific times when they perceive a reduced risk of apprehension.

However, it's imperative to acknowledge that statistical significance does not inherently imply practical significance. Further investigation into the magnitude of the difference in arrival time distribution between the two groups is warranted to discern its real-world implications. Additionally, while the Mann-Whitney U test effectively compares two groups, scenarios involving more than two categories of repeated crime targets necessitate the employment of alternative statistical tests to provide a comprehensive analysis. Therefore, considerations for subsequent steps entail delving deeper into the practical implications of the findings and potentially exploring alternative statistical approaches if the study encompasses multiple categories of repeated crime targets.

4.3.4.5 Analysis Of ROC Curves And AUC Values

Amongst all the other factors that determine the best model, they include the overall project aims, giving enough importance to false negative detection, the impact of errors, and the interpretability of the model's outcomes. Why Logistic Regression might be adopted in case deterring potential missed car theft cases is the prime goal, because the model achieves a higher recall rate. On the contrary, LSVM could be preferred in situations when the penalty for classifying a non-theft case as theft is rather high due to its high accuracy. In addition, an important advantage of Logistic Regression is that it facilitates the proper interpretation of predictive factors and their use for preventive measures implementation. It is worth mentioning that LR and LSVM above all go out of the top ranks with the highest accuracies and AUC indices. Construction of LR equals the training time of LSVM at a quicker rate. The ROC curve (Figure 31) further confirmed the scenario 1 result of accuracy score at 87% and balancing between precision and recall, with the curve positioned far from the diagonal line, signifying better than random chance performance

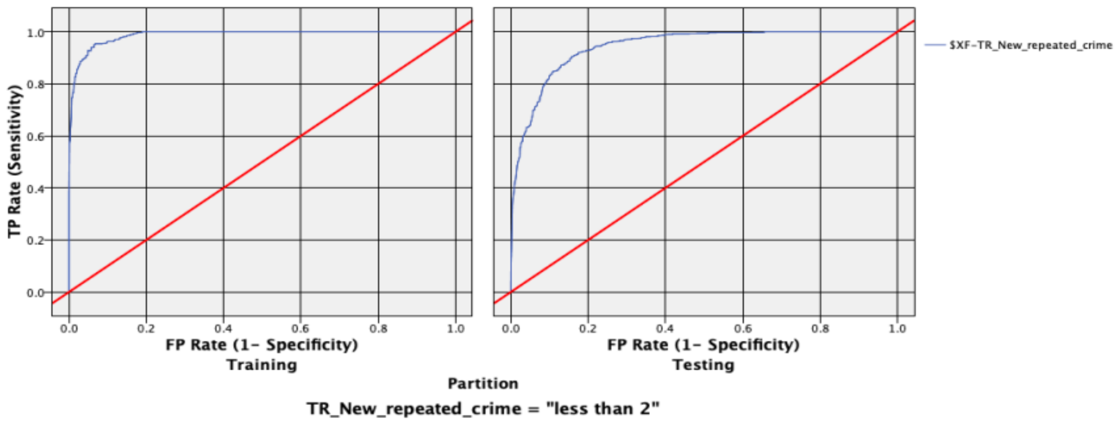


Figure 31: The ROC Curve of Scenario 1 Modelling

Chapter 5 – Discussion

Chapter 4 content presents a comprehensive understanding of car plate theft patterns as well as their factors, the facts that are a good foundation for an in-depth discussion aimed at fostering the key insights and implications of crime prevention and law enforcement strategies. Therefore, the discussion is essentially defined by the involvement context of the results into the bigger framework revolving around the crime fight and law enforcing strategies as well as addressing the practical implications as well

Our findings in Objective 1 showed that there were a lot of advancements in models designed for predicting license plate theft. Through machine learning algorithms, historical trends observation, and optimization of models, we manage to predict precisely the likelihood of car plate thefts occurrences with an accuracy score at 87%. Such models operate as instruments helping police agencies in a preemptive manner to identify risky areas and distribute resources strategically in favor of this aim.

Objective 2 of our research ascertain the revealed actionable recommendations based on the predictive models which are key in involving targeted preventive measures against car registering theft and consequent crimes. Thereby, the models can indicate focus areas and risk factors, like physical hotspots and societies' economic state, which inform the police in taking preemptive approaches to hindering corrective and increasing the level of safety. Further on, the complex analytics allow to grasp different crime dynamics, thus a set of specialized strategies can be designed to minimize prevailing risks and counter the existing threats.

In conformity with Objective 3, our research projects are dedicated to the ultimate goal of enhancing everyone's security and safety especially through the use of the sophisticated machine learning predictive techniques. It is through the harnessing of technology and data savvy attitudes that we can produce a safer and more convenient environment for both the inhabitants and the visitors. The role of our studies is to achieve a greater efficacy in crime prevention and detection at 80%, which in turn promotes the safety and health of the community.

It was observed through the mapping of geographic and economic factors that there exist significant variations in car plate theft distributions in different areas and among the complexities associated with residency status and mode of entering the country. The location of the Jebel Ali Industrial area became a non-repentant zone for car plate stealing, similarly the need for proactive policing or law enforcement that targets high-risk areas like this. Further, that is the significant immunity that most of the longed-stay perpetrators have created the need for extra strengthening of border control and criminal activities deterring immigration enforcement. Socio-economic indicators like educational level and employment status have been proven to be associated with notorious pick-pocketing with a higher proportion of disposed and jobless persons engaged in criminal acts thus exhibiting the mayhem of socio-economic problems to human petty stealing and thus the need to adopt targeted interventions strife centered at socio-economic disparities eradication. The two categories of jobholders and jobless people are responsible for the increasing trend of car plate stealing. This relationship between economic coercion and social pressure helps us understand why people are engaging in such criminal acts.

In the course of the predictive model analysis, the Logistic Regression emerged to be the optimal model for the next act of car plate thefts. Its high specificity and sensitivity level as well as the high technology utilization for better quality results showcase its importance as an effective device for law enforcement agencies and its ability to predict high-risk areas and provide deployment of resources as excellent ways to prevent car plate theft is a proof the device is a solution to curb the menace of crime. Even though the Logistic Regression has produced good performances, the LSVM model has indeed, also highlighted the important trade-offs in the decision-making process of the model as well, with precision and recall among the features worth considering. Also, choosing the predictive model that best fits the crime prevention initiatives is key. In this perspective of feasibility, to the conclusion of the topic, LSVM being competitive means that it can also be an option for law enforcement agencies and that want to predict and prevent the social phenomenon of (car plate) theft. Local law enforcement agencies of choice can tailor both Logistic Regression and LSVM models as the ones that best suit their sets of needs and targets in preventing vehicle plate theft based on their own characteristics and goals.

Thus, the outcome proves the significance of having many effective forms of crime prevention activities that include geographical, socio-economic, and behavioral factors in the law enforcement strategy. Police profiling at crime-ridden arenas and intensified border control along with immigration legislation could help decrease instances of car plate theft. Furthermore, reaching out to the underprivileged and educating jobless people on civic-mindedness can play a vital role in long-term anti-crime policymaking. Although significant results regarding the fluctuations in Motor Vehicle plate thefts have been obtained, the research will not be complete until it has extensively in this area. For the best results of targeted crime prevention techniques, further investigation is needed on the factors that lead to plate theft of vehicles, including the aims of the criminals and the role of organized criminal systems. In addition, time series studies may also take the upper hand by monitoring an evolving pattern of vehicle plate theft in different times and places which could provide more sophisticated insights of altering crime trends, and so on police could have more adaptable law enforcement strategies. The investigation findings herein increase the level of cognition on the role of observed factors, including the targeted enforcement of law and order and the social-economic approach in suppressing criminal activity. Through the channel of predictive modeling, law enforcement agencies can improve community safety, and get the crime rate down by encompassing the behavioral, socioeconomic, and spatial aspects of crime prevention actions.

Chapter 6 – Conclusion and Recommendations

6.1 Conclusion

The comprehensive examination of car plate theft, employing a blend of machine learning algorithms and a thorough literature review, has illuminated key facets of this criminal activity. These studies have shown important patterns and factors that impact car plate theft. They help to make predictions and safety plans. Looking into descriptive statistics helped us understand important things about car plate theft. People from different countries had different levels of involvement. This forces us to concentrate on this type of effort. The time factor showed a big difference.

The forecast used LSVM, Random Tree, SVM, and Logistic Regression to predict car plate theft. Each model offered special views and issues. LSVM model showed a fair rightness of 87%. The main parts of a predictor were found out, such as Area (25%), arrival time (13.5%), and crim description (12.2%). The decision rules gave helpful details about what leads to car plate theft, and the model was the best of all. SVM worked very well, getting a right prediction rate of 76.91%. Important things to look for include if someone has a job, where they live, and how much education they have. But problems in predicting stolen car plates (2.0) showed that crime was complex. Using logistic regression, we looked at things that could help us predict things like where people live, their location or latitude on the map, and whether papers were real. This method got our correct predictions right 53.86% of the time. The model showed how things can affect car plate theft chances.

6.2 Recommendations

To improve our predictions, we should use a team model by joining details from LSVM, random tree, support vector machines (SVM), and logistic regression. Every type has its own special skills, and using all their views can help create a tool for forecasting that is more accurate and detailed. Working as a team, this plan helps us learn about car plate theft in many ways. This helps us to stop crime and catch robbers before they start using stolen number plates. We need to put in extra effort on adding things that improve our ability to predictions better. Adding more simple signs like the use of a car, connections to other crimes, and money problems can help us know better about the situation around stealing using license plates. This constant adjusting ensures our models stay on top of changing crime tactics and assists in finding better ways to prevent crimes.

It's very important to teach people about car plate theft so they can stop it from happening. Programs should show people how to use safe plates properly and make them tell stories about any strange behavior. When we all share the same duty, it helps us prevent crime. Then our neighborhood becomes safer and more alert for any risk or problem happening near us. Machine learning models should be used to aid in police work and help make decisions. These models can assist in getting the most from resources by locating hotspots and predicting where crime might take place next. Police departments can effectively stop and solve car plate thefts by using technology to improve active plans.

6.3 Future Work

- ✓ I intend to employ longitudinal studies that measure the efficacy of implemented crime prevention strategies.
- ✓ I am going to include the collaboration of cutting edge tech like artificial intelligence and blockchain in the fight against car plate theft.
- ✓ I will investigate the social-economic aspects and factors that lead to the spread of vehicle plate theft and apply specific intervention programs.
- ✓ I am going to cooperate with law enforcement agencies for the deployment of predictive analytics tools in both real-time crime monitoring and crime prevention exercises.

References

[1] Akanda, N. I., Hossain, M. A., & Fahad, M. M. I. (2022). Cost-effective and user-friendly vehicle tracking system using GPS and GSM technology based on IoT. Indonesian Journal of Electrical Engineering and Computer Science, 28(3), 1826-1833.

[2] Ariff, F. N. M., Nasir, A. S. A., Jaafar, H., & Zulkifli, A. N. (2021). Comparability of edge detection techniques for automatic vehicle license plate detection and recognition. In Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019. NUSYS'19. Springer Singapore., 891-910.

[3] Ayres, N. (2021). Enhancing the Automotive E/E Architecture Utilising Container-Based Electronic Control Units.

[4] Birnick, J., Bläsius, T., Friedrich, T., Naumann, F., Papenbrock, T., & Schirneck, M. (2020). Hitting set enumeration with partial information for unique column combination discovery. Proceedings of the VLDB Endowment, 13(12), 2270-2283.

[5] Brown, R. (2018). License Plates as Tools for Identity Theft: A Case Study of Financial Implications. Journal of Financial Crime, 14(4), 108-125.

[6] Bukola, A. (2020). Development of an anti-theft vehicle security system using gps and gsm technology with biometric authentication. Int. J. Innov. Sci. Res. Technol.

[7] Cha, G. W., Moon, H. J., & Kim, Y. C. (2021). Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables. International Journal of Environmental Research and Public Health, 18(16), 8530.

[8] Chandra, B. M., Sonia, D., Roopa Devi, A., Yamini Saraswathi, C., Mighty Rathan, K., & Bharghavi, K. (2021). Recognition of vehicle number plate using Matlab. J. Univ. Shanghai Sci. Technol. 23(2), 363-370.

[9] Díaz Juan, I. (2023). Design, development, and optimization of an electronic control unit for an electric racing motorcycle (Bachelor's thesis, Universitat Politècnica de Catalunya).

[10] Etomi, E. E., & Onyishi, D. U. (2021). Automated number plate recognition system. Tropical Journal of Science and Technology, 2(1), 38-48.

[11] Ganta, S., & Svsrk, P. (2020). A novel method for Indian vehicle registration number plate detection and recognition using image processing techniques. Procedia Computer Science, 167, 2623-2633.

[12] Hendry and Rung Ching Chen, "Automatic License Plate Recognition via sliding window darknet-YOLO deep learning", Image and Vision Computing, vol. 87, pp. 47-56, 2019.

[13] Huang, Q., Cai, Z., & Lan, T. (2021). A Single Neural Network for Mixed Style License Plate Detection and Recognition. IEEE Access, 9, 21777-21785.

[14] Jawale, M. A., William, P., Pawar, A. B., & Marriwala, N. (2023). Implementation of number plate detection system for vehicle registration using IOT and recognition using CNN. 27, 100761.

[15] Johansson, K. H., Torngren, M., & Nielsen, L. (2005). Vehicle applications of controller area network. In Handbook of Networked and Embedded Control Systems (pp. 741–765). Springer Verlag.

[16] Jones, M., et al. (2019). The Role of Stolen License Plates in Vehicle Theft: An Analysis of Criminal Activities. Criminology Review, 12(3), 71-88.

[17] Karolin, M., & Meyyappan, T. (2022). Visual secret share creation with grayscale image converted to rgb images using zigzag scanning algorithm. In Proceedings of the International Conference on Paradigms of Communication, Computing and Data Sciences: PCCD, 735-742.

[18] Kaur, P., Kumar, Y., Ahmed, S., Alhumam, A., Singla, R., & Ijaz, M. F. (2022). Automatic License Plate Recognition System for Vehicles Using a CNN. Computers, Materials & Continua. 71(1).

[19] Kaushik, N., Veralkar, M., Pranab. P., & Nandkarny, K. (2014). Anti-theft vehicle security system. International Journal for Scientific Research and Development, 1(12), 2845-2848.

[20] Kathole, A., Shikalgar, A., Supe, N., & Patil, T. (2023). License Plate Recognition for Detecting Stolen Vehicle Using Deep Learning. IEEE 8th International Conference for Convergence in Technology (I2CT) IEEE.

[21] Kavitha, S. S., & Kaulgud, N. (2022). Quantum machine learning for support vector machine classification. Evolutionary Intelligence, 1-10.

[22] Khedher, M. A. (2011). Hybrid GPS-GSM localization of automobile tracking system. International Journal of Computer Science and Technology, 3(6), 75-85.

[23] Kuha, J., & Mills, C. (2020). On group comparisons with logistic regression models. Sociological Methods & Research, 49(2), 498-525.

[24] Kumar, N. S., Surendar, A. A., Prasad, S. H., Vishal, V., & Tribhuvanam, S. (2022). Real Time Intelligent Traffic Light and Density Controller–A Literature Review. International Journal of Modern Developments in Engineering and Science, 1(6), 64-66.

[25] Lin, C. H., Lin, Y., & Liu, W. C. (2018). An efficient license plate recognition system using convolution neural networks. 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 224-227.

[26] Liu, J., Zhao, Z., Ji, J., & Hu, M. (2020). Research and application of wireless sensor network technology in power transmission and distribution system. Intelligent and Converged Networks, 1(2), 199-220.

[27] Lubna, Mufti, N., & Shah, S. A. A. (2021). Automatic number plate Recognition: A detailed survey of relevant algorithms. Sensors, 21(9), 3028.

[28] Mallikalava, V., Yuvaraj, S., Vengatesan, K., Kumar, A., Punjabi, S., & Samee, S. (2020, August). Theft vehicle detection using image processing integrated digital signature-based ECU. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 913-918). IEEE.

[29] Mohammad, A. B., Suneetha, M., & Muqeet, M. A. (2022, February). An Efficient Method for Vehicle theft and Parking rule Violators Detection using Automatic Number Plate Recognition. In 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP) (pp. 1-4). IEEE.

[30] Motakabber, S. M. A., Alam, A. Z., Wafa, S. A. F., & Francis, M. R. M. (2022). GPS and GSM Based Vehicle Tracker. Asian Journal of Electrical and Electronic Engineering, 2(1), 17-24.

[31] Nagaraja, B. G., Mahesh. M, Rayappa, R., & Patil, C. M. (2009). Design and development of a GSM based vehicle theft control system. Paper presented at the International Conference on Advanced Computer Control, Singapore.

[32] Nickel E: IBM automotive software foundry. In Press Conference on Computer Science in Automotive Industry, September 2003, Frankfurt, Germany. Frankfurt University.

[33] Nolte, T., Hansson, H., & Bello, L. L. (2005). Automotive communications past, current, and future. In Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (pp. 992–999).

[34] Onesimu, J. A., Sebastian, R. D., Sei, Y., & Christopher, L. (2021). An Intelligent License Plate Detection and Recognition Model Using Deep Neural Networks. Annals of Emerging Technologies in Computing (AETiC). 5(4).

[35] Ouallane, A. A., Bakali, A., Bahnasse, A., Broumi, S., & Talea, M. (2022). Fusion of engineering insights and emerging trends: Intelligent urban traffic management system. Information Fusion.

[36] Pethakar, S. S., Suryavanshi, S. D., & Srivastava, N. (2012). RFID, GPS, and GSM based vehicle tracing and employee security system. International Journal of Advanced Research in Computer Science and Electronics Engineering, 1(10), 91-96.

[37] Priya, B. J., Kunda, P., & Kumar, S. (2021). Design and Implementation of Smart Real-Time Billing, GSM, and GPS-Based Theft Monitoring and Accident Notification Systems. In Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2020 (pp. 647-661). Springer Singapore.

[38] Sestrem Ochôa, I., Reis Quietinho Leithardt, V., Calbusch, L., De Paz Santana, J. F., Delcio Parreira, W., Oriel Seman, L., & Albenes Zeferino, C. (2021). Performance and security evaluation on a blockchain architecture for license plate recognition system. Applied Sciences, 11(3), 1255.

[39] Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art, 4, 1-14.

[40] Sham Madhukar Lomte and Hajiali Galib Ali Sayyed, "Vehicle Number Plate Recognition System for Theft Detection", International Journal of Science and Research (IJSR), 2015.

[41] Shinde, O., Pagade, O., & Shaikh, Y. (2023). Automatic Number Plate Recognition. In 2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP), 1-5.

[42] Shobayo, O., Olajube, A., Ohere, N., Odusami, M., & Okoyeigbo, O. (2020). Development of smart plate number recognition system for fast cars with web application. Applied Computational Intelligence and Soft Computing, 1-7.

[43] Smith, J., & Johnson, A. (2015). Trends in Car License Plate Theft in Urban Areas. Journal of Crime Prevention, 8(2), 145-160.

[44] Raman, R., & Karthiayani, A. (2023). Cloud-based Electronic Toll Collection: Enabling Contactless and Automated Payments System. In 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, 355-359.

[45] Vercio, L. L., Amador, K., Bannister, J. J., Crites, S., Gutierrez, A., MacDonald, M. E., ... & Forkert, N. D. (2020). Supervised machine learning tools: a tutorial for clinicians. Journal of Neural Engineering, 17(6), 062001.

[46] Wang, Z., Ma, X., & Huang, W. (2020). Vehicle license plate recognition based on wavelet transform and vertical edge matching. International Journal of Pattern Recognition and Artificial Intelligence, 34(06), 2050016.

[47] White, S., & Davis, P. (2020). Preventive Measures Against License Plate Theft: An Evaluation of Tamper-Resistant Screws and Anti-Theft License Plate Frames. Crime Prevention Journal, 15(1), 25-38.

[48] Wolf, M., Weimerskirch, A., & Wollinger, T. (2007). State of the art: Embedding security in vehicles. EURAS IP Journal of Embedded Systems, 2007(5), 1.

[49] Yan, X., Wang, C., Hao, D., & Chen, M. (2021). License plate detection using Bayesian method based on edge features. In 2021 IEEE 5th International Conference on Cryptography, Security and Privacy (CSP). IEEE, 205-211.

[50] Yaacob, N. L., Alkahtani, A. A., Noman, F. M., Zuhdi, A. M., & Habeeb, D. (2021). License plate recognition for campus auto-gate system. Indonesian Journal of Electrical Engineering and Computer Science, 21(1), 128-136.

[51] Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. IEEE access, 8, 181302-181310.

Appendix

Variable View in SPSS

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	I_Year	String	8	0	Year	None	None	8	Left	Nominal	Input
2	I_PoliceStati...	Numeric	9	4	Police Station L...	None	None	8	Right	Scale	Input
3	I_PoliceStati...	Numeric	9	4	Police Station L...	None	None	8	Right	Scale	Input
4	I_CrimeScen...	Numeric	9	4	Crime Scene Lo...	None	None	8	Right	Scale	Input
5	I_CrimeScen...	Numeric	9	4	Crime Scene La...	None	None	8	Right	Scale	Input
6	I_Age	Numeric	9	0	Criminal Age	None	None	8	Right	Scale	Input
7	I_Nationality	String	8	0	Criminal Nation...	{Cameron, ...	None	8	Left	Nominal	Input
8	I_Quarter	Numeric	9	0	Quarter Of The ...	{1, 1.0}...	None	8	Right	Scale	Input
9	I_Residence...	String	8	0	Residence Status	{Active V, A...	None	8	Left	Nominal	Input
10	I_Number_...	Numeric	9	0	Number Of Cri...	{1, 1.0}...	None	8	Right	Scale	None
11	I_PoliceStati...	String	6	0	Police Station	{BD, BD}...	None	6	Left	Nominal	Input
12	I_Area	String	22	0	Area	{Abu hail, A...	None	22	Left	Nominal	Input
13	I_Arrival_Ti...	Numeric	9	0	Arrival Time	{5, 5.0}...	None	8	Right	Scale	Input
14	I_Crime_De...	String	33	0	Crime Descripti...	{Alcohol ga...	None	33	Left	Nominal	Input
15	I_Crime_Ti...	String	6	0	Crime Timing	{Day, Day}...	None	6	Left	Nominal	Input
16	I_Criminal_S...	String	8	0	Criminal Status	{ARRESTED, ...	None	8	Left	Nominal	Input
17	I_Way_Of_E...	String	13	0	Way Of Entry	{Entry legall...	None	13	Left	Nominal	Input
18	I_Document...	String	13	0	Document Status	{Fake passp...	None	13	Left	Nominal	Input
19	I_Education...	String	17	0	Education Level	{Illiterate, Illi...	None	17	Left	Nominal	Input
20	I_Employme...	String	10	0	Employment St...	{Employed, ...	None	10	Left	Nominal	Input
21	I_Type_Of_...	String	9	0	Type Of Car	{Abandoned...	None	9	Left	Nominal	Input
22	I_Camera_F...	String	6	0	Camera Footage	{NO, NO}...	None	6	Left	Nominal	Input

Snippet of Data Sample

	I_Year	I_PoliceStation_longitude	I_PoliceStation_latitude	I_CrimeScene_longitude	I_CrimeScene_latitude	I_Age	I_Nationality	I_Quarter	I_Residence_Status	I_Number_Of_Crimes	I_PoliceStation	I_Area	I_Arrival_Time_Min
1	2022	25.1582	55.2909	25.2051	55.3390	20	Cameron	1	Overstay	2	JA	Dubai Investments Park	16
2	2020	25.0203	55.0049	24.9937	55.1131	25	Cameron	2	Active V	2	JA	Al Quoz industrial	16
3	2022	25.1115	55.2537	25.1700	55.2086	34	Cameron	3	Overstay	2	MQ	Jebel Ali Industrial	14
4	2022	25.0309	55.1715	25.0468	55.1508	23	Cameron	2	Overstay	2	BRS	Jebel Ali Industrial	17
5	2022	25.0681	55.2388	25.0535	55.2475	33	Cameron	1	Overstay	2	JA	Al Quoz industrial	12
6	2022	25.1700	55.2465	25.2728	55.3051	31	Cameron	2	Active V	2	BD	Al Waheda	21
7	2020	25.2299	55.4368	25.2801	55.4469	22	Cameron	4	Active V	2	BD	Jebel Ali Industrial	12
8	2020	25.0972	55.2539	25.1032	55.1862	25	Cameron	1	Overstay	2	BD	Jebel Ali Industrial	21
9	2022	25.0187	55.1136	24.9609	55.0371	24	Cameron	2	Overstay	2	JA	Dubai Investments Park	6
10	2023	25.0802	55.2194	25.1181	55.2564	23	Cameron	1	Overstay	2	JA	Al Quoz industrial	12
11	2022	25.0508	55.1785	25.0656	55.1365	23	Cameron	4	Overstay	2	MQ	Jebel Ali Industrial	21
12	2021	25.0443	55.1205	25.0261	55.1520	32	Cameron	3	Overstay	2	BD	Al Qusais	12
13	2021	25.0227	55.1171	25.0038	55.0650	34	Cameron	4	Overstay	2	JA	Dubai Investments Park	21
14	2022	25.0958	55.1761	25.0593	55.2206	32	Cameron	4	Overstay	2	BRS	Business Bay	12
15	2021	25.1093	55.2629	25.1851	55.3394	23	Cameron	4	Active V	2	MQ	Dubai Land	14
16	2023	25.2125	55.2878	25.2796	55.3304	38	Cameron	4	Active V	2	JA	Al Quoz industrial	21
17	2023	25.1475	55.2482	25.1995	55.3488	24	Cameron	2	Active V	2	BD	Al Quoz industrial	12
18	2023	25.1505	55.1995	25.2625	55.2626	24	Cameron	4	Overstay	2	JA	Al Quoz industrial	12
19	2022	25.1078	55.1090	25.1690	55.2111	25	Cameron	4	Active V	2	BD	Al Quoz industrial	6
20	2022	25.0338	55.1055	24.9944	55.1096	31	Cameron	2	Overstay	2	BD	Jebel Ali Industrial	21
21	2023	25.0555	55.0433	25.0611	55.1579	24	Cameron	1	Resident	2	BD	Dubai Investments Park	12
22	2023	25.2014	55.3095	25.2581	55.3740	27	Cameron	1	Overstay	2	JA	Dubai Land	21
23	2023	25.0519	55.1337	25.0229	55.1556	22	Cameron	3	Overstay	2	JA	Business Bay	12
24	2022	25.0322	55.1777	25.0124	55.1883	32	Cameron	1	Resident	2	BD	Al Jadaf	6