DURBAN UNIVERSITY OF TECHNOLOGY

Data Mining to Analyse Recurrent Crime in South Africa

By

Phirime Monyeki

Student Number: 21413700

A dissertation submitted in fulfilment of the requirement for the degree of Master of Information and Communications Technology

Faculty of Accounting and Informatics, Department of Information Technology, Postgraduate Studies

Supervisor: Dr N. Naicker

Co-Supervisor: Dr T. Adeliyi

2021

DECLARATION

I, Phirime Monyeki, declare that:

(i) The research reported in this dissertation, except where otherwise indicated, is my original research.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

- Their words have been re-written but the general information attributed to them has been referenced.
- Where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the Reference Section of this dissertation.

__________________________ 18-01-2021

Signature: P Monyeki Date

Approved: Supervisor
02 November 2021

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of the Moses Kotane Institute (MKI), Durban University of Technology (DUT), my supervisor and co-supervisor who supported me throughout this journey. I wish to thank the following people for their much-valued support and contribution:

- My supervisor, Dr N. Naicker: Thank you so much for your commitment and coaching during this study. Your guidance inspired me and made working with you a pleasure.
- My co-supervisor, Dr T. Adeliyi: Thank you very much for your support and encouragement. Working with you and sharing ideas made the work easily understandable. I appreciate your effort towards my development.
- Durban University of Technology (DUT): The support from DUT is highly appreciated. I want to thank you for the opportunity to pursue this qualification.
- Moses Kotane Institute (MKI): Thanks to MKI for advice and financial support towards this project and for supporting me to achieve this milestone.
- I would like to say thank you to all other individuals not listed here but who directly or indirectly contributed to this study.

ABSTRACT

When South Africa is compared to other countries, it has a notably high rate of crime. The country has seen a concomitantly high occurrence of murder, residential burglary, drug-related crime and carjacking (hijacking) crime. The government is desperately seeking solutions that can be implemented to reduce recurrent crime. Several reasons to explicate high crime trends in different areas include alcohol or drug abuse, low standards of education, poor parenting skills and a lack of social and vocational skills. This study aimed to gain better insight into crime trends in South Africa using data mining techniques. Decision-making linked to the data could help the government implement a coherent crime strategy to mitigate crime. The crime dataset chosen for this study was publicly available at kaggle.com. The dataset was prepared using Python programming code. The research design was utilised as an overall strategy to compile all different components of this study with an intention of answering the research questions and attaining the research objectives. To identify the significant changes, Change-Point Analysis (CPA) was performed to pinpoint the abrupt change in the South African crime dataset. Two methods called Cumulative Sum (CUSUM) and Bootstrap were implemented in this study of CPA. To analyse the trend of data, CUSUM and Bootstrap were performed to measure the occurrence of change points based on the confidence levels. The CPA outcome depicted multiple significant changes and abrupt shifts in several provinces of South Africa. Linear regression (LR) was utilised to predict the future trends of crime in South Africa from 2016 – 2022 based on the erstwhile 2005 – 2015 crime statistics. The results showed that crime has been on the increase in South Africa with certain provinces such as Western Cape, Gauteng and KwaZulu-Natal being identified as crime hotspots. Future studies on crime should focus only on one province to gain insight into the dominating crimes and hotspots within that particular province, with a view to developing highly specific crime-reduction interventions.

TABLE OF CONTENTS

DECLARATION ... ii
ACKNOWLEDGEMENTS ... iii
ABSTRACT ... iv
TABLE OF CONTENTS ... v
LIST OF TABLES ... viii
LIST OF FIGURES ... ix
LIST OF ACRONYMS AND ABBREVIATIONS ... xi
LIST OF TERMINOLOGY ... xii
CHAPTER ONE: INTRODUCTION TO STUDY ... 13
1.1 Introduction ... 13
1.2 Aim and Objectives ... 15
1.3 Problem Statement ... 16
1.4 Significance of the Study ... 16
1.5 Scope and Delimitations of Study ... 17
1.6 Research Output ... 17
1.7 Structure of the Thesis ... 17
1.8 Chapter Summary ... 18
CHAPTER TWO: LITERATURE REVIEW ... 19
2.1 Introduction ... 19
2.2 Types of Crime ... 20
2.2.1 Murder Crime ... 20
2.2.2 Burglary Crime ... 23
2.2.3 Drug-Related Crime ... 25
2.2.4 Carjacking Crime ... 28
2.2.5 Additional Crimes ... 29
2.3 Machine Learning to Solve Crime ... 31
2.4 Chapter Summary ... 40
CHAPTER THREE: RESEARCH METHODOLOGY ... 41
3.1 Introduction ... 41
3.2 Population Size ... 41

3.3 Research Design 41
3.3.1 Dataset..........42
3.3.2 Pre-processing..........42
3.3.3 Crime Pattern Theory (CPT)..........42
3.3.4 Change-Point Detection..........43
3.3.5 Crime Rate Prediction..........46
3.4 Relationship Between Change Point Analysis and Linear Regression..........47
3.5 Chapter Summary..........48
CHAPTER FOUR: RESULTS, ANALYSIS AND DISCUSSION..........49
4.1 Introduction..........49
4.2 Crime Pattern Theory Results..........49
4.2.1 Using CPT to Explain Burglary Presented in the Geographical Spatial Representation of Crime Pattern..........53
4.2.2 Using CPT to Explain Murder Presented in Geographical Spatial Representation of Crime Patterns..........58
4.2.3 Crime Pattern Theory Analysis for the Seven Remaining Provinces..........59
4.3 Murder Crime Change-Point Analysis Results..........59
4.4 Murder Crime Results Summary..........71
4.5 Burglary Crime Results..........73
4.5.1 Burglary Crime Results Summary..........84
4.6 Synopsis of Drug-related Crime Results..........86
4.6.1 Synthesis of Control Charts and CUSUM..........86
4.6.2 Drug-related Crime Summary..........87
4.7 Synopsis of Carjacking Crime Results..........87
4.7.1 Synthesis of Control Chart and CUSUM..........88
4.7.2 Carjacking Crime Summary..........88
4.8 Chapter Summary..........89
CHAPTER FIVE: PREDICTION USING LINEAR REGRESSION..........90
5.1 Introduction..........90
5.2 Murder Prediction..........90
5.2.1 Comparison of the actual murder trends and predicted values..........92
5.3 Burglary Prediction..........93

5.3.1 Comparison of the actual residential burglary trends and predicted values 94
5.4 Carjacking Prediction 95
5.4.1 Comparison of the actual carjacking trends and predicted values 97
5.5 Drug-related Crime Prediction 97
5.5.1 Comparison of the actual drug-related crimes to crime trends and predicted values 99
5.6 Chapter Summary 99
CHAPTER SIX: SUMMARY, CONCLUSIONS AND IMPLICATIONS OF STUDY 100
6.1 Introduction..100
6.2 Summary of Conclusions...100
6.2.1 To Provide Insights into Crime Trends in South Africa Using Statistics from the Extant Literature [RO 1] 100
6.2.2 To Apply the Change-Point Analysis Algorithm as an Effective Data Mining Tool in Detecting Abrupt Changes in Recurrent Crime in South Africa [RO 2] 100
6.2.3 To Predict Future Trends in Habitual Crime in South Africa Using Machine Learning [RO 3] 101
6.3 Implications of Study...101
6.3.1 Opportunities...101
6.3.2 Recommendations...102
6.3.3 Future study 102
6.4 Chapter summary 103
REFERENCES...104
Annexure A: Cover of Turn it in Report 115
Annexure B: Language Editing Certificate 116

LIST OF TABLES

Table 2.1 South African crime statistics per province (2005 – 2015) 19
Table 4.1 South African top 10 hotspots for burglary in each province (2005 – 2015) 50
Table 4.2 South African top 10 hotspots for murder in each province (2005 – 2015) 55
Table 4.3 Significant changes for murder in KwaZulu-Natal...61
Table 4.4 Significant changes for murder in Gauteng 63
Table 4.5 Significant changes for murder in Western Cape 64
Table 4.6 Significant changes for murder in Eastern Cape 66
Table 4.7 Significant changes for murder in Free State...67
Table 4.8 Significant changes for murder in Mpumalanga...69
Table 4.9 Occurrence of potential abrupt shifts for murder per province (2005 – 2015)........72
Table 4.10 Significant changes for burglary in Western Cape 74
Table 4.11 Significant changes for burglary in KwaZulu-Natal..76
Table 4.12 Significant changes for burglary in Eastern Cape 77
Table 4.13 Significant changes for burglary in Free State..79
Table 4.14 Significant changes for burglary in North West 80
Table 4.15 Significant changes for burglary in Limpopo 81
Table 4.16 Occurrence of abrupt shifts in burglary per province (2005 – 2015) 85
Table 4.17 Occurrence of abrupt shifts in drug-related crime per province (2005 – 2015) 87
Table 4.18 Occurrence of abrupt shifts in carjacking per province (2005 – 2015) 88
Table 5.1 Predicted murder trends (2016 – 2022) 91
Table 5.2 Murder prediction accuracy for South Africa (2016 – 2020) 93
Table 5.3 Predicted Residential burglary trends for South Africa (2016 – 2022) 94
Table 5.4 Residential burglary prediction accuracy for South Africa (2016 – 2019).............95
Table 5.5 Predicted carjacking trends for South Africa (2016 – 2022) 96
Table 5.6 Carjacking prediction accuracy for South Africa (2016 – 2020).............................97
Table 5.7 Predicted drug-related crime trends for South Africa (2016 – 2022) 98
Table 5.8 Drug-related crime prediction accuracy for South Africa (2016 – 2019)................99

LIST OF FIGURES

Figure 2.1 South African murder statistics per province (2005 – 2015) 22
Figure 2.2 South African Burglary statistics per province (2005 – 2015)..............................25
Figure 2.3 South African drug-related crime statistics per province (2005 – 2015) 27
Figure 2.4 South African carjacking statistics per province (2005 – 2015) 29
Figure 2.5 (a – r) Murder, burglary, carjacking and drug-related incidents per province (2005 – 2015) 39
Figure 3.1 Crime Analysis Framework...41
Figure 4.1 Burglary top 10 hotspots per province (2005 – 2015)...53
Figure 4.2 Murder top 10 hotspots per province (2005 – 2015)...58
Figure 4.3 KwaZulu-Natal murder data: background changes (2005 – 2015) 60
Figure 4.4 KwaZulu-Natal murder data: CUSUM plot (2005 – 2015)...................................61
Figure 4.5 Gauteng murder data: background changes (2005 – 2015)...................................62
Figure 4.6 Gauteng murder data: CUSUM plot (2005 – 2015) 62
Figure 4.7 Western Cape murder data: background changes (2005 – 2015)...........................63
Figure 4.8 Western Cape murder data: CUSUM plot (2005 – 2015) 64
Figure 4.9 Eastern Cape murder data: background changes (2005 – 2015) 65
Figure 4.10 Eastern Cape murder data: CUSUM plot (2005 – 2015) 65
Figure 4.11 Free State murder data: background changes (2005 – 2015) 66
Figure 4.12 Free State murder data: CUSUM plot (2005 – 2015)..67
Figure 4.13 Mpumalanga murder data: background changes (2005 – 2015) 68
Figure 4.14 Mpumalanga murder data: CUSUM plot (2005 – 2015).....................................68
Figure 4.15 Northern Cape murder data: background changes (2005 – 2015)........................69
Figure 4.16 Northern Cape murder data: CUSUM plot (2005 – 2015) 69
Figure 4.17 Limpopo murder data: background changes (2005 – 2015).................................70
Figure 4.18 Limpopo murder data: CUSUM plot (2005 – 2015) 70
Figure 4.19 North West murder data: background changes (2005 – 2015).............................71
Figure 4.20 North West murder data: CUSUM plot (2005 – 2015) 71
Figure 4.21 Western Cape burglary data: background changes (2005 – 2015).......................73
Figure 4.22 Western Cape burglary data: CUSUM plot (2005 – 2015) 74
Figure 4.23 KwaZulu-Natal burglary data: background changes (2005 – 2015) 75
Figure 4.24 KwaZulu-Natal burglary data: CUSUM plot (2005 – 2015)................................75
Figure 4.25 Eastern Cape burglary data: background changes (2005 – 2015) 76

Figure 4.26 Eastern Cape burglary data: CUSUM plot (2005 – 2015) 77
Figure 4.27 Free State burglary data: background changes (2005 – 2015) 78
Figure 4.28 Free State burglary data: CUSUM plot (2005 – 2015)..78
Figure 4.29 North West burglary data: background changes (2005 – 2015)............................79
Figure 4.30 North West burglary data: CUSUM plot (2005 – 2015) 79
Figure 4.31 Limpopo burglary data: background changes (2005 – 2015)................................80
Figure 4.32 Limpopo burglary data: CUSUM plot (2005 – 2015) 81
Figure 4.33 Gauteng burglary data: background changes (2005 – 2015)................................82
Figure 4.34 Gauteng burglary data: CUSUM plot (2005 – 2015) 82
Figure 4.35 Northern Cape burglary data: background changes 83
Figure 4.36 Northern Cape burglary data: CUSUM plot (2005 – 2015) 83
Figure 4.37 Mpumalanga burglary data: background changes (2005 – 2015) 84
Figure 4.38 Mpumalanga burglary data: CUSUM plot (2005 – 2015)....................................84
Figure 5.1 Murder trends in South Africa (2005 – 2015) 91
Figure 5.2 Murder Prediction vs Actual Murder trends (2016 – 2020) 92
Figure 5.3 Residential Burglary trends for South Africa (2005 – 2015) 93
Figure 5.4 Burglary Prediction vs Actual Residential Burglary trends (2016 – 2020)...........94
Figure 5.5 Carjacking trends for South Africa (2005 – 2015) 95
Figure 5.6 Carjacking prediction vs Actual carjacking trends (2016 – 2020) 96
Figure 5.7 Drug-related crime trends for South Africa (2005 – 2015)....................................98

LIST OF ACRONYMS AND ABBREVIATIONS

The following are the most frequently used acronyms in this dissertation:

AI	– Artificial Intelligence
ANC	– African National Congress
ARDL	– Autoregressive Distributed Lag Error Correction Model
CIAC	– Crime Information Analysis Centre
CPA	– Change Point Analysis
CPT	– Crime Pattern Theory
CUSUM	– Cumulative Sum
DUDIT	– Drug Use Disorder Identification Test
LR	– Linear Regression
MI	– Multi-Instance
ML	– Machine Learning
MSE	– Mean Square Error
SVM	– Support Vector Machine
SWOR	– Sampling Without Replacement
WEKA	– Waikato Environment for Knowledge Analysis
4IR	– Fourth Industrial Revolution

LIST OF TERMINOLOGY

The following terminology as related to the study are explained:

Artificial Intelligence: The simulation of human capabilities of doing something, for instance, digital cashier machines.

Bootstrap: Normally used to measure the accuracy, depending on the case. This statistical approach and it is usually used to estimate the sampling of the distributed data.

Data Mining: The process of manipulating and analysing a large dataset in order to gain new insight.

Change-Point Analysis: An algorithm for detecting the changes in time series data.

CUSUM: A statistical approach normally used to examine and monitor the changes in time series data.

Linear Regression: The linear approach that is utilised to indicate the connection between the variables that are dependent and independent.

Machine Learning: An application that is usually used in Artificial Intelligence, that enables machines to work automatically or digitally without being typically instructed.

Python: An Object-Oriented Programming (OOP) language designed to interact and manipulate data.

Research design: The main overall strategy that is used to carry out the whole study. It comprises different components that are needed to carry out a logical plan.

CHAPTER ONE: INTRODUCTION TO STUDY

1.1 Introduction

The study identifies the changes in crime levels in South Africa's nine provinces. Crime has escalated dramatically both in incidence and scope, such that even South African youth are deeply involved in drugs and crime, with cases being reported in schools and communities daily (Mathews *et al.* 2019).

Of all the crimes committed in Africa, violent crimes such as gang violence, alcohol and drug-related violence has seen the greatest increase, compared to all other crime categories (Africa Times Editor 2019). Murder has continued to increase dramatically over time, with 65% of murders being the result of violence (Prinsloo *et al.* 2016; Kleck and Patterson, 1993). Prinsloo *et al.* (2016) reported that serious crimes, namely serious assault, rape and attempted murder increased during the early 2000s.

Crime such as such burglary can be described as gaining illegal access to a residence or business to commit theft or further misdemeanour (Van Niekerk and Greeff 2020; Cronje and Spocter 2017). Crime trends across the developed world have been diminishing for decades, often referred to as a global crime drop (Asongu and Nwachukwu 2017b). However, not all countries have experienced a decline in the rate of crime. One of the developing countries facing a high residential burglary crime rate is South Africa. Numerous studies done in South Africa on residential burglary clearly stated that two factors have led to an increase in the commission of this particular crime; the high levels of unemployment and inequality (Musah *et al.* 2020; De Juan and Wegner 2017).

Unemployment has been shown to result in increased alcohol abuse and violence, with violent crime being recognised as the fastest growing type of crime in South Africa (Van Zyl, Wilson and Pretorius 2003; Gould, Burger and Newham 2016). Burglary target selection in South Africa does not end at the choice of a specific suburb or a city, but certain qualities of individual properties within similar suburbs are indicative of a burglary hazard (Zinn 2017; Breetzke and Cohn 2013. Inequality is considered a threat which drives this process (Chatterjee 2019; De Juanand Wegner 2017). The South African Police Service (2015) stated that between 2013 and 2014, South Africa experienced the worst violent crime levels (assault, murder, and robbery). During that period, attempted murder and robbery increased at a rate of 3.5% and 11.2% per year, respectively. Recent statistics reveal that in South Africa, the murder rate is five times greater than that of 2013 and 2014 (Van Niekerk and Greeff 2020; African Check, 2019). The statistics further depicted an increase of 5% in all types of crimes every year.

In South Africa, murder is also considered a common crime. Otieno *et al.* (2015) mentioned that murder occurs as a result of conflicts between two people or groups engaging in some activities, which include illegal mining, politics and taxi and Uber-related conflict. Recently the country has experienced a huge surge in murder committed by intimate partners. Cases of gender-based violence have increased with Abraham *et al.* (2017) linking the murder of women to gender inequality.

Their study further mentioned that sexual child murder is also an issue in South Africa, with numerous cases of young girls who are reported missing and are found deceased after a long period of searching, often with evidence of sexual abuse. Other forms of crimes are also on the increase. Davis (2014) reported that 35 000 attempted and accomplished carjackings occurred annually. Writter (2020) also mentioned that cars around 50 are stolen every day in South Africa. The South African Analysis Forum stated that 52% of attempted carjackings were successful. Nine out of 10 accomplished carjackings were reported to the police, compared to 6 in 10 attempts (Davis, 2014). The study further mentioned that cities like Johannesburg, Cape Town and Durban were known for carjacking. Johannesburg was the leading city with more crimes reported and greater numbers of people dwelling in a limited space, leading to high population and poverty related crime.

Davis (2014) further mentioned that 54% of all accomplished and attempted carjackings were committed by groups of two or more criminals. In addition, both attempted and accomplished carjackings were most likely to occur in the evening or at night and away from the victims' home. There were, however, a few instances where the offenders were working alone when committing the crimes.

Drugs are one major cause of crime; they are related to crime in numerous ways. Drug use often results in abuse and the illegal distribution of drugs. Cocaine, heroin, and marijuana are the most common examples of drugs categorised to have abuse potential (Ncontsa and Shumba 2013). The effect of drugs on people's behaviour is highly negative as it can lead to violence and other illegal activity related to drug trafficking. Violence and murder have been recognised as a global public issue (Prinsloo *et al.* 2016). Otieno *et al.* (2015) report shows a high rate of violence-related to deaths crime in South Africa.

Schwinn *et al.* (2019) stated that most people suffering from drug abuse are adult males and the youth both male and female. The study further mentioned that environments with no job opportunities and unemployment contribute to the behaviour of the youth. Hence, if drugs are readily available through drug dealers, it is more likely that the youth will take drugs and consume alcohol.

Safety and security are two major elements of a good environment. The South African crime rate has been increasing and the public authority is urgently seeking any approaches that could halt the rise of crime and support investors' trust to improve the economy (Business Insider South Africa, 2019). The study further mentioned that South Africa is one of the countries that has been labelled as one of the countries with the worst crime rates in the world (Business Insider South Africa, 2019). Numerous reasons for the high crime rate in South Africa have been put forward; these include an absence of social and professional aptitudes, the generally poor standard of education, consumption of alcohol and drug abuse (Gulma *et al.* 2013). This research study will analyse changes in murder, burglary, carjacking and drug-related crimes over time and develop a model to predict and prevent further increases. These specific crimes were selected because they are common and appear to be rising (Wessels, 2020). A review of the extant literature shows that there has been no previous research conducted that identified the changes in crime in South Africa over the last decade, using Change-Point Analysis (CPA).

1.2 Aim and Objectives

The aim of this research was to use data mining techniques to analyse crime trends in South Africa in order to support decision-making and implement a coherent crime strategy.

The objectives of the research study were as follows:

[RO 1]: To provide insights into crime trends in South Africa using statistics from the extant literature.

[RO 2]: To apply the Change-Point Analysis algorithm as an effective data mining tool in detecting abrupt changes in recurrent crime in South Africa.

[RO 3]: To-predict future trends in habitual crime in South Africa using Machine Learning (ML).

1.3 Problem Statement

The increase in South Africa's crime rate affects the image of the country, and retards economic growth as investors refuse to engage with or invest in what they perceive as dangerous markets (Business Insider South Africa, 2019). Gould, Burger and Newham (2016) stated that common reasons such as unemployment, inequality, a low standard of education, a high number of immigrants, poor housing, alcohol abuse, and poor parenting skills were identified as the common reasons for the high crime rate in South Africa. Crime affects both rich and poor societies, impacting negatively on people's lives (financially, physically and psychologically). There have been no studies conducted in South Africa to detect the changes in crime over time or models developed to predict future crime trends, hence the scope for this study.

This study will apply the CPA algorithm to provide insight into the hidden changes of crime that have occurred and further use a Linear Regression (LR) algorithm to predict future crime trends. It is essential to use these techniques in this study because they provide a clear access to the huge dataset of crime and enables the data to be managed in a way that will help to achieve the research objectives. The CPA was utilised to detect when the change or changes took place in the series data, and the confidence level at which the change occurred. The confidence level designates the probability that one or more changes took place and interval confidence designates when exactly the change took place. Prediction of crime will assist in providing a clear understanding of crimes that are likely to increase, and to pre-organise strategies to overcome or prevent their occurrence. This study will also check how Machine Learning (ML) has been used in the previous studies to solve crime related issues.

1.4 Significance of the Study

This study analyses crime trends of the most commonly committed crimes in all nine provinces of South Africa. Valuable findings obtained from this research of South African crime trends can be shared with those individuals or groups concerned with creating new crime prevention strategies. Additionally, it will provide a positive contribution to building the knowledge base of the South African Police Service (SAPS), as well as local and national government departments. Furthermore, analysis of crime trends, past and future (predicted), using data science tools and techniques will assist authorities to combat crime and this will undoubtedly provide long-term benefits to the citizens of South Africa.

1.5 Scope and Delimitations of Study

This study does not deal with all forms of crimes. It exclusively covers murder, burglary, carjacking (known locally as hijacking) and drug-related crime. These crimes were selected because they were recognised as the leading crimes in South Africa and they have shown rapid growth compared to other types of crime (Mathews *et al.* 2019). This research utilised a dataset of South African crime statistics that spans eleven years, from 2005 to 2015 (Wessels, 2020).

1.6 Research Output

The following Department of Higher Education and Training (DHET) accredited journal publication was produced towards this research study:

Monyeki, P., Naicker, N. and Obagbuwa, I. C. 2020. Change-Point Analysis: an effective technique for detecting abrupt change in the homicide trends in a democratic South Africa. *The Scientific World Journal, 2020.* Doi:10.115/2020/4158472.

1.7 Structure of the Thesis

This study is presented in five chapters and is arranged in the manner described below. A summary of each chapter is presented, as follows.

Chapter One: Introduction to Study

This chapter clearly states the problem, aim, and objectives. This chapter explains that murder, burglary, carjacking and drug-related crimes are analysed in this study. Research outputs are also highlighted in this chapter.

Chapter Two: Literature review

Chapter Two reviews previous and recent studies that are related to the aim, research questions and objectives of the study. The literature review portrays the crime landscape in the country and then reveals the approaches and methods that have been implemented to solve the crime problem in the past.

Chapter Three: Research Methodology

This chapter of the study depicts the methodology implemented in the study. The chapter explains how data mining methods are conducted to obtain results. The chapter also provides a detailed description of the datasets and tools used in this study.

Chapter Four: Results, Analysis and Discussion

This chapter contains the discussions of the results obtained from executing the experiment. Firstly, the CPA results are discussed before proceeding to analyse the predictions. Secondly, the prediction results are analysed and compared to existing crime trends.

Chapter Five: Prediction Using Linear Regression

This chapter of the study performed the Linear Regression (LR) algorithm to predict the future trends of crime from 2016–2022. The results are presented in tables and graphs. For each type of crime selected, the comparison of the actual crime trend and predicted crime trend is done and the accuracy for the prediction is calculated and presented.

Chapter Six: Summary, Conclusions and Implications of Study

This chapter examines the background, objectives and methodology of this research and then draws the conclusion based on the study findings.

1.8 Chapter Summary

This chapter has elucidated the background, introduction, significance of the study, aim and objectives. The next chapter will present an in-depth literature review, exclusively on the crimes that are specified in the problem statement and how ML has been used to overcome crime-related issues in the past.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

South Africa is one of the countries with a high crime rate (Asongu and Nwachukwu 2017a). Based on numerous previous studies, there have been two constant factors identified as responsible for the rise in crime, namely unemployment and socio-economic inequality (Musah *et al.* 2020; Gould, Burger and Newham 2016). Crimes such as burglary, violence, murder, drug abuse and carjacking are assumed to be the result of these factors, and this has had a negative impact on people's lives. Unemployment can ultimately lead people to commit crime or misuse alcohol because of stress (Marie 2017). Burglary can be described as gaining illegal access to a residence or business to commit theft or further misdemeanour. Crime rates across the modern world have been decreasing for decades, and this is often referred to as the international crime drop.

Conducting this study is essential because it enables one to gain insight into crime trends which authorities may be able to use to address crime in a way that will benefit society at large. Table 2.1 presents the crime statistics per province for the 11-year period from 2005 to 2015 for the four selected crimes (Wessels 2020).

Table 2.1 South African crime statistics per province (2005 – 2015)

Year	Western Cape	Eastern Cape	Northern Cape	Free State	KwaZulu-Natal	North West	Gauteng	Mpuma-langa	Limpopo
2005	381 825	238 977	56 515	137 987	345 784	118 840	654 817	134 829	106 983
2006	396 712	228 884	52 689	128 227	343 798	112 471	639 635	131 444	104 857
2007	395 281	220 813	48 954	127 955	328 368	112 340	615 618	125 954	97 166
2008	398 240	216 658	50 277	132 335	340 095	117 118	638 186	131 410	97 565
2009	417 619	217 230	49 746	127 512	349 103	115 680	640 074	128 814	99 610
2010	426 850	215 012	45 618	121 997	343 767	109 922	609 305	122 402	96 475
2011	447 238	214 462	45 257	126 389	348 411	111 028	577 959	122 186	113 630
2012	465 994	209 124	47 697	131 785	362 677	115 319	584 315	121 169	112 952
2013	479 022	210 248	48 947	126 290	355 729	113 935	636 195	115 996	117 638
2014	492 963	202 582	49 897	118 879	348 394	114 270	637 332	117 203	124 986
2015	490 383	196 089	50 665	117 688	342 772	114 335	622 218	119 526	129 323
Total	**4 792 127**	**2 370 079**	**546 262**	**1 397 044**	**3 808 898**	**1 255 258**	**6 855 654**	**1 370 933**	**1 201 185**

Crime has been a historical and persistent issue in most countries; few countries have managed to control this issue that affects whole countries as much as it affects individuals. Cronje and Spocter (2017) stated that South Africa has an enormous crime rate, with high occurrences of murder, residential burglary, driving under the influence of alcohol, and violent crimes. Crime statistics of South Africa have increased dramatically since early 2000s to 2014 and the government is desperately seeking solutions to overcome this issue (Gould, Burger and Newham, 2016).

According to Musah *et al.* (2020), the high rate of unemployment and inequality has driven many people to commit crime in South Africa. Almost every study conducted in South Africa on crime has identified these two factors as common causes of crime (Gulma *et al.* 2013; Langton and Steenbeek 2017). Van Zyl, Wilson and Pretorius (2003) mentioned that high levels of socio-economic inequality increase the threat of crime, because culprits sometimes identify certain characteristics of an individual's property and then target specific individuals based on perceived high property value. Gould, Burger and Newham (2016) study stated that unemployment and inequality are the major factors that contribute more to crime. Matzopoulos *et al.* (2018) mentioned that both rich and poor societies are affected equally by crime in South Africa, especially when considering the impact it has on people's lives at both individual and community levels.

2.2 Types of Crime

Crime occurs in different ways, sometimes it occurs as a result of jealousy, conflict between two or more people, through hostel violence and the use of drugs and alcohol or other reasons. However, other crimes such as theft, attack and robbery arise not because there is a conflict between the culprits and the victims but because the criminals seek to acquire assets unlawfully (Asongu and Acha-anyi 2019; Zinn 2017). The following crimes are discussed in the sub-sections below: murder; residential burglary; carjacking; and drug-related crime.

2.2.1 Murder Crime

Murder is an unlawful killing of one human by another, sometimes it occurs as a result of conflict between two or more people, either engaging in the same activities or offenders forcefully and unlawfully taking the property of a victim (Otieno *et al.* 2015). Sometimes murder can be the result of violence. South Africa is still experiencing an extremely high murder rate (Khoele *et al*. 2016), although there would have been an expectation of murder reduction due to democracy and the introduction of new laws and more political stability.

Khoele *et al*. (2016) mentioned that gender inequality is also a challenge in South Africa. Gender inequality exists in romantic partnerships, where one abuses the other it might be physical or emotional, and sometimes it leads to killing the other or committing the suicide (Khoele *et al*. 2016). The study further notes that murder against women in South Africa is on the increase and trends relating to sexual murder against minors show that young girls have a greater risk compared to young boys (Khoele *et al*. 2016). According to Abraham *et al.* (2017), crimes such as sexual assault or murder are usually committed by intimate partners, a phenomenon which is referred to as gender-based violence.

Lindegaard (2017) noted that violence in South Africa is a serious issue and is the primary cause of death. It is estimated that the occurrence of murder in South Africa is double the worldwide average. There has been a huge increase in youth conviction; the identification of young men as both perpetrators and victims is worrisome (Skelton, 2002). McCafferty (2003) suggested that crime after apartheid had increased numerically. Crime statistics depict that violent crime has the highest increase followed by murder. From 1994 to 2002, violent crimes such as assault, rape, and attempted murder recorded an increase, whilst murder reportedly declined. Business Insider South Africa (2019) noted that murder is a subcategory of violent crime which is also on the rise, and South Africa has been labelled as one of the murderous countries worldwide.

The Crime Information Analysis Centre (CIAC) agrees with McCafferty (2003) in that soon after the abolishment of apartheid, violent crimes in South Africa increased. South Africa had the highest rates of violence and murder in Cape Town in early 2003s (Van Zyl, Wilson and Pretorius 2003; Reporter 2013). The incidence of crime increased significantly after apartheid with serious crimes such as assault, robbery and murder being experienced more in big cities where people go to seek employment opportunities (Asongu and Acha-anyi 2019)

Numerous studies have identified factors such as poverty and inequality as the major contributors to the commission of different crimes. Cheteni, Mah and Yohane (2018) identified the relationship between drug-related crime and poverty in South Africa. The study utilised the Autoregressive-Distributed Lag Error Correction Model (ARDL ECM) to analyse the data. The ARDL ECM model identified a strong association between the two. The study suggested that socio-economic factors had driven many drug-related crimes in the country.

The crime statistics of South Africa show that there is a notably high rate of murder in the country compared to other African countries, with 40% of all reported murders being recorded during residential burglaries (Wessels 2020). Therefore, preventing burglaries becomes paramount as this will reduce the murder rates as a result. Statistics further show that Johannesburg is the capital city with the highest crime levels in South Africa (Van Zyl, Wilson and Pretorius 2003; Cronje and Specter 2017; McCafferty 2003).

Police brutality has also led to murder cases. For example, reports suggest that members of North West Province SAPS killed 34 miners and injured another 78 miners who were striking in Marikana, in an act of demanding a decent living wage. However, in response, the police claimed that they acted in self-defence (Marinovich 2012). Figure 2.1 presents the South African statistics of murder per province from 2005 to 2015 (Wessels 2020). The y-axis presents the total number of murder cases, and the x-axis presents the provinces of South Africa.

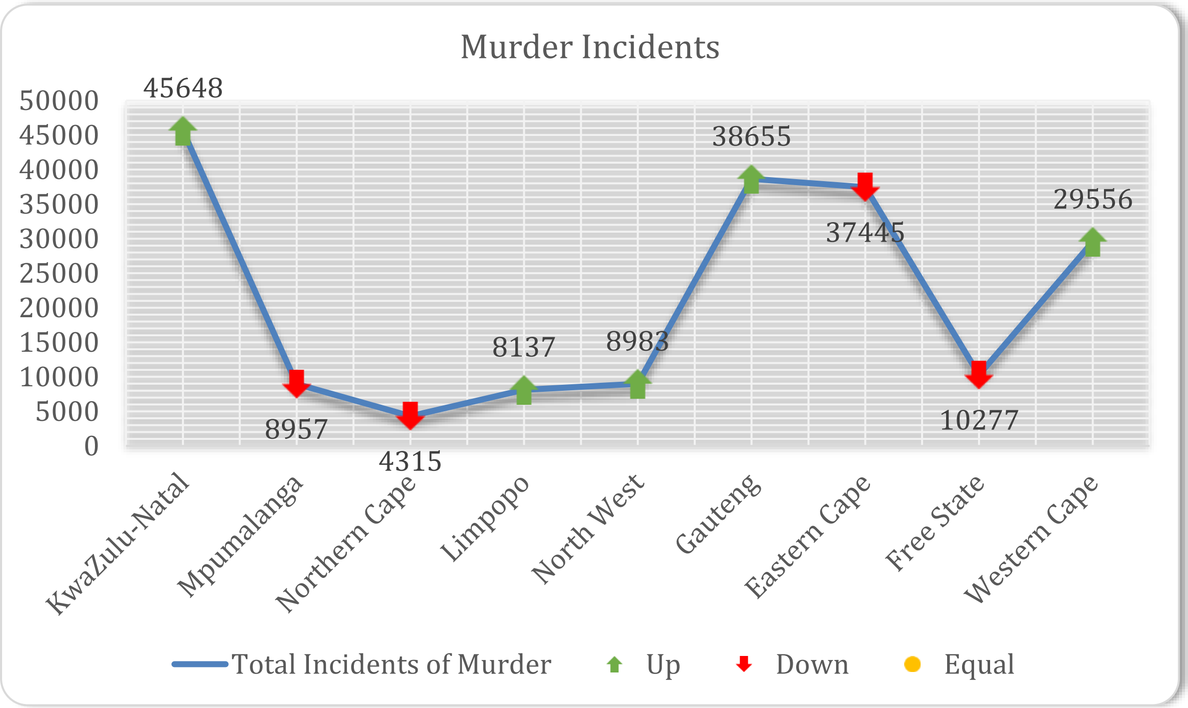


Figure 2.1 South African murder statistics per province (2005 – 2015)

The contents of Figure 2.1 are in alignment with the findings of the studies of Van Zyl, Wilson and Pretorius 2003; Asongu and Acha-anyi (2019); and Wessel (2020), that murder is a dominant crime in South Africa. As shown in Figure 2.1, murder crime in provinces such as KwaZulu-Natal, Eastern Cape and Gauteng has been on an extreme rise and several reasons have been put forward to explain this trend, including inequality, unemployment, poor environment, and the generally low standard of education (Khoele *et al*. 2016; Skelton, 2002).

2.2.2 Burglary Crime

Residential burglary remains a serious issue in South African societies. This issue is likely to be fuelled by the experience of unemployment faced by South Africans (Khoele *et al.* 2016). According to Gulma *et al.* (2013), there is an average rate of more than 590 residential burglaries per 100 000 houses per year. Therefore, residential burglary is also regarded as the most common crime and people from all spheres of life, both the rich and the poor, are vulnerable to it. According to Amiri *et al.* (2019), in 2018 and 2019 an average of 605 houses were burgled per day in South Africa; perceived high-value technological gadgets such as laptops, smart televisions, decoders, and cameras, were among the most commonly stolen items, followed by cars. Van Zyl, Wilson and Pretorius (2003) mentioned that not only had housebreaking increased in South Africa but also that school burglaries and vandalism were on the rise. Schools have long been an easy target for thieves and vandals with the issue persisting. In 2018, four schools in Mpumalanga Province of South Africa were set alight by the culprits after stealing the school asset (Cronje and Spocter 2017).

Burglary crime is committed when someone unlawfully enters residential or business premises with the intention to steal. Residential burglary is a property crime that has the highest occurrence of all law-breaking in South Africa (Hodgkinson and Andresen 2019). It remains a serious issue in South African society, and this challenge may continue as long as corruption and unemployment still exist (Gulma *et al.* 2013). An analysis by Cronje and Spocter (2017) stated that in 2015 there was an average rate of more than 590 residential burglaries per 100 000 of the houses per year.

Breetzke and Cohn (2013) conducted a study to examine the impact that gated communities had on burglary rates. Over 26 000 communities in the city of Tshwane in South Africa are registered as being gated. The findings of the study show that gated communities are more likely to be the victims of burglary even though house protection is very high, when compared to other dwellings. Thus, residing in a gated community does not necessarily reduce the risks of being targeted for burglary. The study further noted that burglars target gated communities both during the day and night. Cronje and Spocter (2017) mentioned that not all victims report crime to the police; some victims feel that police officers do not put in sufficient effort, especially when dealing with common crimes such as carjacking and residential burglary. According to Amiri *et al.* (2019), the lack of evidence sometimes relates to poor detective work by investigating officers, confirming the notion during 1997 that most investigators did not have specialised training, with only a quarter having attended a detective course and a mere

3% being fully trained. Sperino (2020) stated that the lack of evidence results to case dismissal because evidence should meet the burden proof.

Musah *et al.* (2020) agreed with a study by Amiri Kerry, Brooks Bryan and Daratha (2019) that residential burglary and property crimes are difficult to solve, with between 80 and 90% of burglaries remaining unresolved. Most burglary cases do not even reach the court and of those that do, almost 70% of the burglary cases are withdrawn during proceedings owing to insufficient evidence. The incompetence exhibited by police officers results in the continued intrusion of privacy, loss of valuable property, which fosters a climate of fear and anger (Amiri *et al.* 2019; Musah *et al.* 2020).

Langton and Steenbeek (2017) mentioned that poor housing, high number of immigrants, alcohol abuse, and lack of parenting skills are to blame as they were identified as the factors that have an enormous impact on crimes being committed. The lack of security attracts the perpetrators (Langton and Steenbeek 2017).

Van Zyl, Wilson and Pretorius (2003) mentioned that violence in schools depicts unethical behaviour within the education sector. Schools play an essential role in ensuring the development of children (Van Zyl, Wilson and Pretorius 2003). Hence, it is essential to ensure that schools provide the safest possible environment for teaching and learning. Figure 2.2 presents a total statistics of burglary per province in South Africa (Wessels 2020). The y-axis presents the total number of burglary cases, and the x-axis presents the provinces of South Africa.

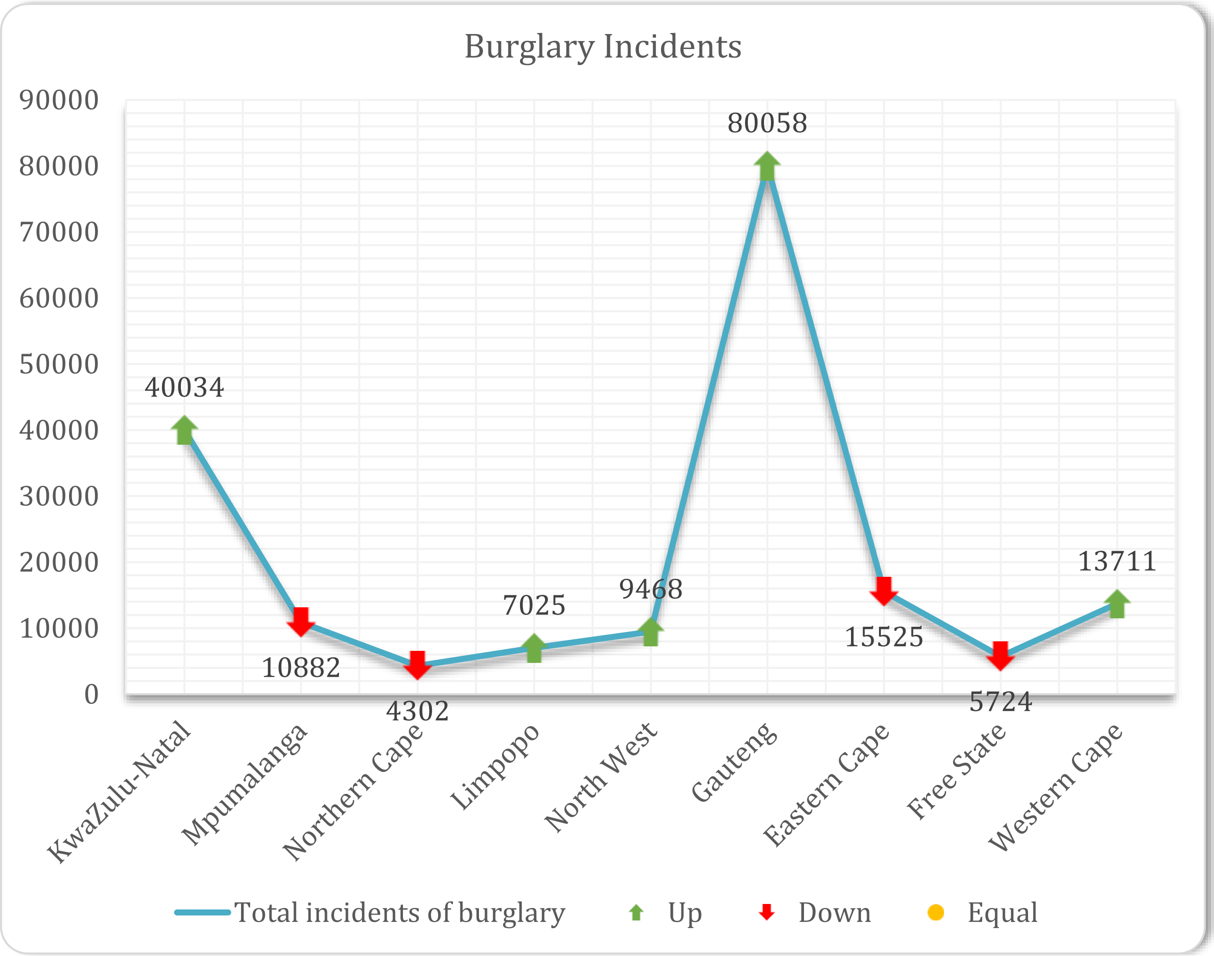


Figure 2.2 South African Burglary statistics per province (2005 – 2015)

Figure 2.2 shows the residential burglary total and the average burglaries of each province over the period under review (Wessels 2020). Figure 2.2 shows Gauteng with 80 058 incidents of burglary as the province with the highest number of cases, followed by KwaZulu-Natal with 40 034 and Eastern Cape with 15 525. Figure 2.5 depict that 2015 had the highest number of burglary cases.

2.2.3 Drug-Related Crime

Drug use has become fashionable in today's world, with different drugs available such as Kratom, cocaine, crystal meth, heroin and methadone (Brook *et al.* 2016). Some say the effect of these drugs are similar to cocaine, but they are more likely to cause serious health issues such as violent behaviour, paranoia, heart palpitations, high blood pressure and chest pain (Brook *et al.* 2016). Schwinn *et al.* (2019) conducted a study on drug abuse and prevention programs for adolescent girls. The study was conducted after it has identified that the use of drugs by youth has increased drastically, especially by young girls (Schwinn *et al.* 2019). The study suggested web-based drug abuse prevention program to reduce adolescent girls' drug use rates and associated risk factors. This program was recognised after conducting an analysis on social media platforms such as Facebook.

A study conducted by Van Zyl, Wilson and Pretorius (2003) noted that school violence endangers the lives of youth who are the future leaders of the world. The following factors were identified as the major contributors that spread school violence: access to firearms and drugs, lack of family stability, lack of community support and ineffective parenting (Van Zyl, Wilson and Pretorius 2003).

The South African Institute of Race Relations (SAIRR) (2015) reported that only 23% of learners felt safe at school. School violence in South Africa has escalated to the extent that learners are shooting, stabbing and being both physically and emotional violent to one another. Following up these instances, drugs are commonly found in these cases, 80% of fights in schools are drug-related issues. However, violence in school is not something new in schools, some of these crimes can be traced back to June 1976 when thousands of learners took to the streets of Soweto to protest about the imposition of Afrikaans as the language of teaching and learning in schools (Ncontsa and Shumba 2013). Zulu *et al.* (2004) however, disagreed with Ncontsa and Shumba (2013) and stated that the school violence and learner activism of 1976 was a protest against apartheid not crime against one other (not to be confused with the violence widely experienced in South African schools).

Some South Africans believe that immigrants are responsible for most crime waves in the country involving drugs, violence, murder, and burglaries (McCafferty 2003). A national survey was conducted to verify if immigrants are really the cause of high crime in South Africa. According to the findings, almost half the sampled individuals (48%) felt that immigrants were a crime threat, compared to 37% who thought they were a threat to jobs and the economy, while and 29% thought they were a health threat. The survey of South Africa concluded that immigrants were indeed perceived as a threat in South Africa, especially those who enter the country illegally (McCafferty 2003).

There are countless number of drugs that are utilised these days (Cheteni, Mah and Yohane 2018). A drug is a substance that cause changes in an organism’s psychology when consumed (Cheteni, Mah and Yohane 2018). Drug-related crime is the possession, or distribution of drugs classified as having the potential to be abused. Drugs are also related to numerous crimes such as drug trafficking, illegal selling, violence and carrying illegal weapons (Brook *et al.* 2016).

Berman *et al.* (2005), evaluated the drug-related crimes using the Drugs Use Disorders Identification Test (DUDIT) in Criminal Justice. This method was developed with an aim to identify individuals with drug-related problems. Hence, it was applied in a sample of heavy drug users from a prison in Sweden. DUDIT predicted drug dependence with a sensitivity of 90%. Results revealed that 78% of the convicts in prison were using drugs even though they were not arrested for drug-related crimes. In prison, drugs were taken to avoid stress and overthinking, hence the study suggested medical health assistance and therapy for inmates. Brook *et al.* (2016) mentioned that few studies have been conducted to examine the risk of using drugs, especially by adolescents. The use of drugs by the youth is not just a challenge of South Africa, it is an international health problem (Schwinn *et al.* 2019).

Figure 2.3 presents a total statistics of drug-related crimes per province in South Africa (Wessels 2020). The figure will disclose the number of drug-related reported case over the 10 years' period (2005 – 2015). The y-axis presents the total number of drug-related cases, and the x-axis presents the provinces of South Africa.

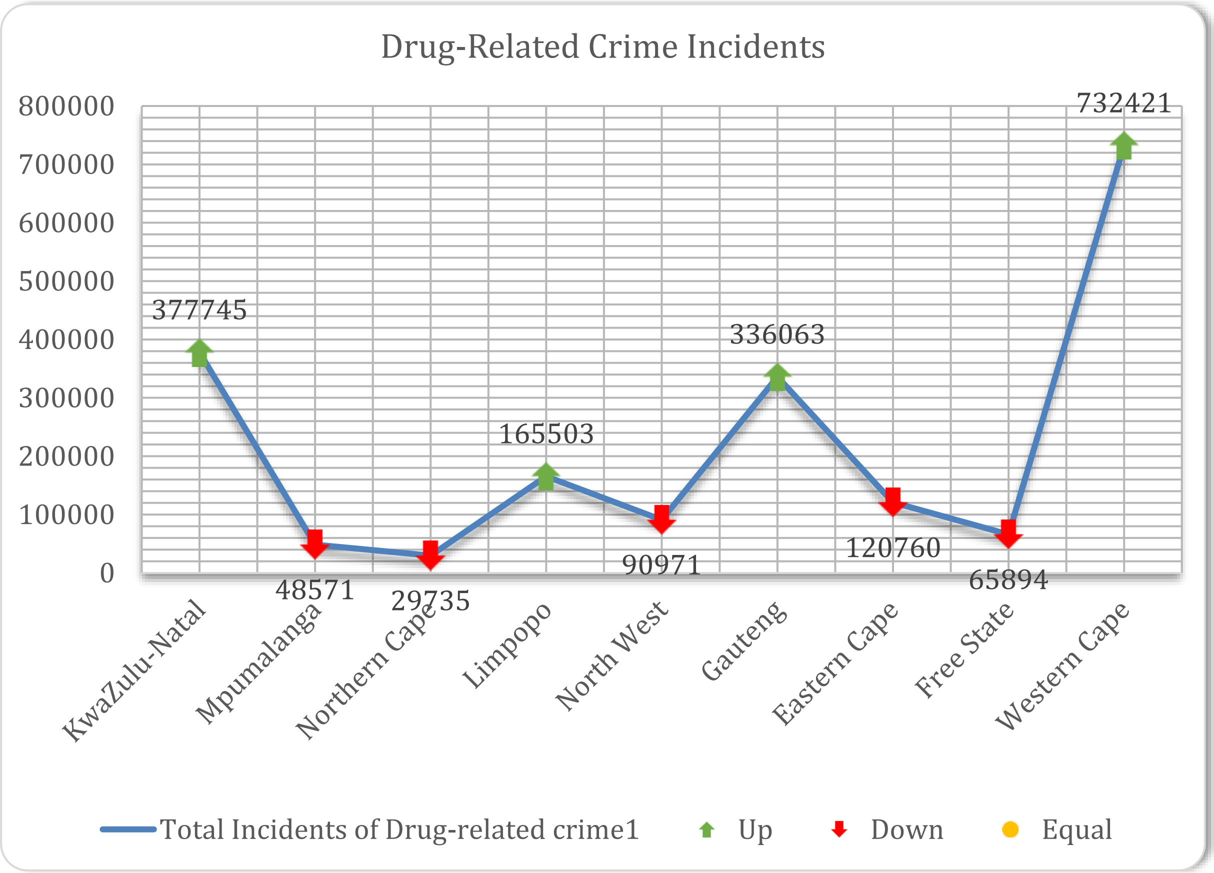


Figure 2.3 South African drug-related crime statistics per province (2005 – 2015)

There are different types of drug-related crimes, such as drug trafficking, drug production and possess. Hence the numbers are higher than some other types of crimes. Figure 2.3 shows the drug-related crimes total of each province nationally recorded during the period 2005 to 2015. Figure 2.3 shows that the province with the highest drug-related crime numbers in South Africa was Western Cape with the total of 73 2421, followed by KwaZulu-Natal with a total of 37 7745 and then Gauteng with a total of 33 6063. As shown in Figure 2.5, 2014 had the highest number of drug-related incidences.

2.2.4 Carjacking Crime

Carjacking is a worldwide phenomenon, it has increased to an extent that it is regarded as one of the leading crimes worldwide (Writter 2020). It is estimated that in South Africa, a motor vehicle is hijacked every 40 to 54 minutes. Put simply, more than 25 drivers every day become carjacking victims. Despite these statistics, there has been a few studies conducted to identify the factors contributing to the increase of carjacking. A study conducted by Davis (2014) evaluated the rising carjacking crime in South Africa and revealed that carjackings, burglaries are not the coincidences, but in fact the study found out that most perpetrators target certain individuals with certain characteristics. Factors such as unemployment and inequality were identified as major contributors to carjackings. From the findings it became evident that hijacking does not take place erratically: hijackers are selective in the choice of targets and target selection mostly takes place based on the vehicle driven by the motorist (Davis 2014). According to James (2017) carjacking remains rampant in South Africa. The police reported that after the car had been unlawfully taken from the owner it is difficult to find and locate it. Figure 2.4 presents a total statistics of carjacking per province (Wessels 2020). The y-axis presents the total number of carjacking cases, and the x-axis presents the provinces of South Africa.

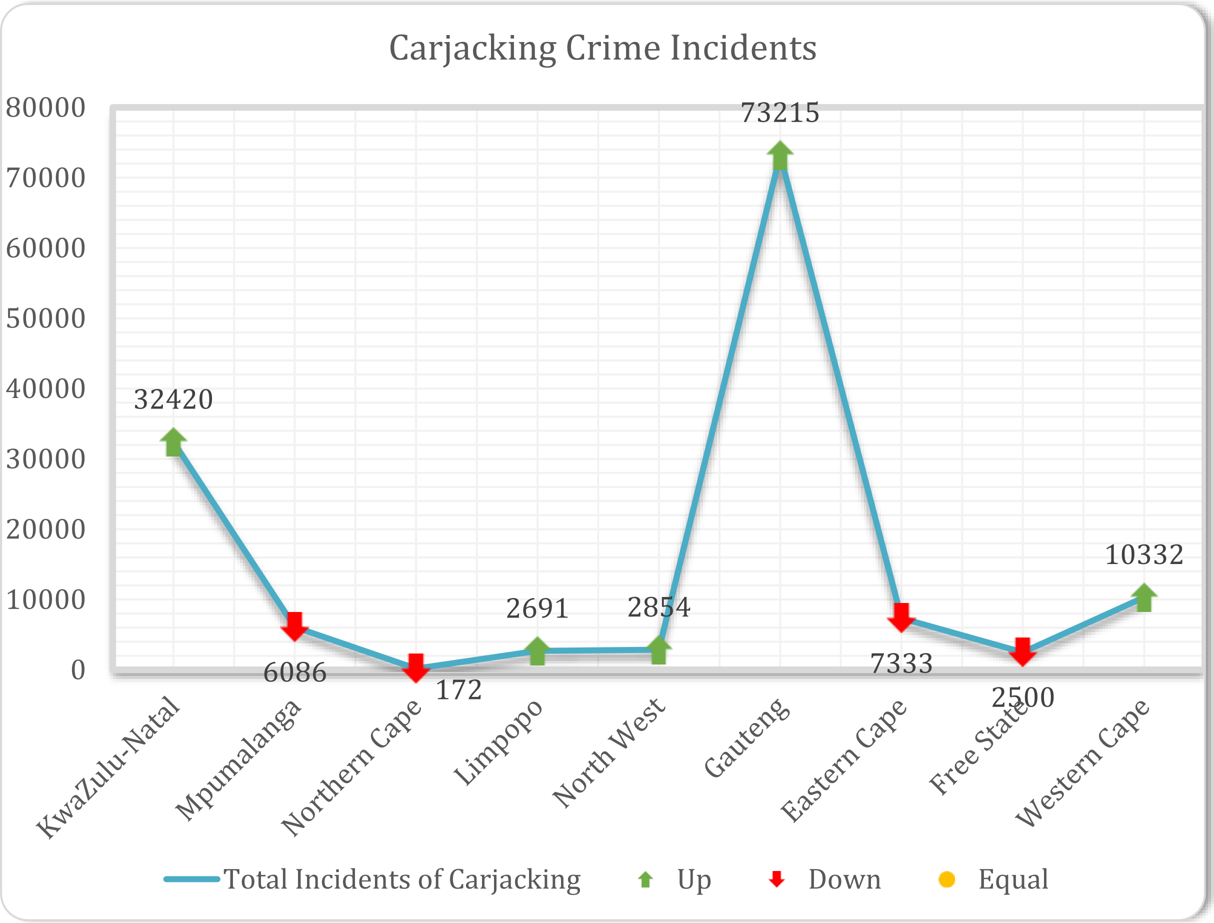


Figure 2.4 South African carjacking statistics per province (2005 – 2015)

Carjacking is commonly violent (James 2017). Most victims are threatened with a life-endangering weapon. Eighty-one percent of victims who suffered an injury as a result of physical violence required medical treatment at a hospital (Writter 2020). Other victims suffered from strong emotional stress, and 60% of victims did not get their cars back (Writter 2020). This study makes an important contribution to the field by investigating the human consequences of carjacking. It has also demonstrated that one can complete direct research with victims, despite impediments to doing so that may limit or deter researchers (Writter 2020).

2.2.5 Additional Crimes

In the past few years, numerous studies in South Africa as well as other countries have revealed that carrying of weapons and drugs in both primary and secondary schools is a growing trend (Zulu *et al.* 2004). According to Prinsloo *et al.* (2016), common violent acts committed in schools include learner to learner assaults, learner to teacher assault, teacher to learner assaults, parent to teacher assaults, vandalism and improper behaviour. Cronje and Spocter (2017) concurred with Zulu *et al.* (2004) when they confirmed that significant violence was experienced and perpetrated in SA schools. They noted that between 2011 and 2012, one in five learners reported experiencing violence at schools in the 12 months preceding data collection.

According to the study conducted by Prinsloo *et al.* (2016) during the 30-day period prior to data collection, 16% of South African high school learners reported being involved in a physical fight at school during school hours, while 11.8% of learners carried weapons on school premises. All the incidents occurred in the South African province of Western Cape. The unauthorised carrying of weapons is also experienced more at tertiary education level, where students are known to carry dangerous weapons on campus. The study further stated that in 2003, violence in schools increased to an extent that 16.8% of learners were afraid to attend school, with 20.9% of learners having been threatened or beaten by someone at school, while 32.8% reported being verbally abused at school by other learners and sometimes educators (Prinsloo *et al.* 2016).

Cheteni, Mah and Yohane (2018) found that poverty and poor services delivery by the police impacted crime levels, while disparities between the rich and poor were also blamed. Statistics indicate that crime affects mainly poorer South Africans. Gould, Burger and Newham (2016) agree with the study by Cronje and Spocter (2017), that people are starting to doubt the competence of police and further mentioned that only 52.8% of victims' report crime to the police. Other victims feel like police are incompetent or they put insufficient effort in solving reported cases. Amiri *et al.* (2019) revealed that the lack of evidence is sometimes a reflection of poor detective work by the investigating officers.

According to the SAPS (2015) during the period 2013 – 2014, crime against South African citizens and tourists was the leading crime and it was the biggest contributor to the documented crimes reported at police stations (34%). Phori's (2017) study further identified crimes related to property such as housebreaking, theft of motor vehicles, theft out of or from motor vehicles, fraud and shoplifting. The most commonly committed crime in Durban was carjacking, reported by 43.75% of study respondents, followed by car break-ins, reported by 32% of Durban respondents.

A study by Buiten and Naidoo (2016) identified the problems and factors that lead to rape in South Africa. Notwithstanding the gains made in improving women's rights and the political advances that support women in South Africa, the study revealed a huge gap between the reality of expectations and women's lived experiences. Two issues that indicates a huge difference between the public rights and lived experiences is the prevalence of sexual murder and rape committed against women. Africa Check (2020) stated that there has been an increase of sexual offence to 53 293 in 2019/20 from 52 420 in 2018/19. The sexual offence comprises of rape, sexual assault, attempted sexual offence and contact sexual offence. Buiten and Naidoo (2016) noted that rape of women in South Africa is a serious issue. The study predicted that based on the available statistics one in three women in South Africa has been or will be raped. This demonstrates the impact of gender inequality and of imbalanced gendered power relations in reference to women.

Swemmer (2019) agreed with Buiten and Naidoo (2016), and further suggested that inequality and gendered power is an issue that has been ignored. His study focused on the rape trial of the former South African president, Jacob Gedleyihlekisa Zuma. Throughout his research, the major issue identified was the silencing of rape victims as an instance of ethical loneliness. In some cases, with regards to women abuse and gender inequality, media sometimes can have negative influence (Swemmer 2019). For instance, in this case, the victim was criticised by the media and also blamed as she appeared to be a member of an opposing faction of the African National Congress (ANC), publicly attempting to tarnish the reputation of the accused. Research shows that inequality is a world-wide problem, Hence, the study concluded that victims or women suffer from the ethical loneliness (Swemmer 2019).

2.3 Machine Learning to Solve Crime

A Machine Learning (ML) technique called Change-Point Analysis (CPA) has proven to be the most accurate and reliable analytic tool for analysing series data. According to Matzopoulos *et al.* (2018), numerous studies have shown that CPA is capable of revealing the hidden existence of changes that are sometimes difficult to recognise or detect when using other techniques. Change-point analysis has been implemented in many fields of study, including health control, finance, and medicine.

Monyeki, Naicker and Obagbuwa (2020) used CPA as an effective tool to identify the abrupt changes in the murder trends in a democratic South Africa. This study was essential as it provided some insights into murder perpetrated in the country. South African crime statistics were used and grouped by province, with the data preparation done using Python. Two CPA techniques for ML were utilised, namely Cumulative Sum (CUSUM) and bootstrap. This technique was performed with the aim to offer a good interpretation of murder. Cumulative Sum (CUSUM) analysis was used to detect and analyse the crime data, while bootstrap analysis was used to measure the exact time the change points occurred. The findings indicated that during the period of 2008 to 2012 most provinces experienced the highest rate of murder and the CPA depicted all the abrupt changes (Monyeki, Naicker and Obagbuwa 2020).

Taylor (2018) described CPA as an essential technique performed on time series data to identify or detect whether one or more changes has occurred. It estimates the time of each change and determines the number of changes that took place.

Nath (2006) used Clustering to detect the crime patterns and to speed up the process of identifying and solving them. This technique works closely with detectives as it focuses on the geographical groups of crime. The study applied the k-means Clustering technique to the real time data. The contribution of the study was that it formulated crime pattern detection as a ML task and used data mining to support detection and patterns of solving crimes. Using Clustering techniques, they were able to predict the crime hotspots which will help in the deployment of police to the most likely places of crime and allow for the most effective use of police resources.

Sathyadevan, Devan and Gandadharan (2014) used data mining to predict and analyse crime. Crime analysis and prevention is considered a systematic approach for identifying and analysing patterns and trends in crime. The technique called classification was used in this study. The main reason for using this method was because the study dealt only with existing or known crimes. Therefore, this technique only relies on the existing crimes. The study used Naïve Bayes algorithm for classification. This technique is preferred when compared with Support Vector Machines (SVMs), because it does not require a lot of memory for implementation.

Using the Naïve Bayes classifier, they were able to create a model by analysing data from crime related to vandalism, murder, robbery and burglary. Patterns were discovered, and the results showed that the Naïve Bayes classifier had more than 90% accuracy. Abutabenjeh and Jaradat (2018) described classification as a supervised technique in data mining.

The use of both ML and data mining has become a vital part of crime detection and analysis. McClendon and Meghanathan (2015) conducted a study on how to use ML techniques to detect and analyse crime trends. The study utilised WEKA, which is open source data mining software that was used for comparing violent crime patterns from the communities and crime unnormalized Dataset. In this study, WEKA was used to make comparisons between the violent crime patterns and crime that is not normalised in the dataset. This data analytic tool was selected because it provides numerous ML algorithms that will enable users to implement and conduct the analysis on datasets. The study further mentioned that algorithms such as Bayes, Functions, Lazy, Meta, Multi-Instance (MI), Miscellaneous, Rules, and Trees were utilised.

The study performed three algorithms on the unnormalized dataset namely LR, Additive regression and Decision Stump algorithm. Evaluation metrix for a LR to prove efficiency and effectiveness in predicting crime based on the training set of input turned out a success. The study further suggested that the LR algorithm should be used to handle prediction because it provides small prediction errors than other techniques.

Prabakaran and Mitra (2018) agreed with McClendon and Meghanathan (2015) in proposing that LR be utilised for predictions. Additionally, many ML algorithms such Decision Tree, CPA and K-means were also recommended for data analysis to improve crime analytics.

As indicated by Ahishakiye *et al.* (2017), crime is a common problem that affects the way in which people live and the growth of communities. In this research, the Decision Tree (J48) ML algorithm was applied to assist in the context of enforcement and intelligence analysis. This approach was used to discover hidden relationships among data, and model them in a manner that would eventually assist in the decision-making process. From the experimental results, the ML used proved to be both effective and efficient, as it provided level of accuracy for the system to rely on.

Machine learning techniques are not only used for prediction but can also be used to identify factors that encourage crime in societies (McClendon and Meghanathan 2015). Mittal *et al.* (2019) identified certain factors and monitored the impact of crime in India using ML. Crime trends in India keep on changing as the population is growing rapidly. There has been an enormous rise in crimes; particularly crimes against women, children and the vulnerable sectors of society. Police spend a considerable amount of time going through the crime data thoroughly to determine the dominant factors. For that reason, this study implemented two ML techniques, namely Clustering and the Association rule mining algorithm.

The Clustering algorithm is an unsupervised learning technique that is used to partition the dataset into small groups based on the similarity measures. Since the data comes in many forms, it is essential to group it in a manner that will help discover hidden information. Therefore, after Clustering the data the Association rule mining algorithm was applied to find the correlation between the different attributes of the dataset that was created. The findings depicted that a unidirectional causation exists between the unemployment rate of level and the incidence of robbery (Mittal *et al.* 2019). Therefore, the government of India was encouraged to create more employment and education opportunities to control the crime rate (Mittal *et al.* 2019).

Lung Lin, Yang Chen and Chi Yu (2017) conducted a study on how to use ML to assist crime prevention officials in Taiwan to counter the increase in drug-related criminal activity. This study proposed a data driven method called “broken window” and spatial analysis to analyse crime data using ML algorithms. The theory implies that the lack of response from the small criminal activity could lead to more serious crimes.

Based on this theory of “broken windows”, a model is then designed that predicts the occurrence of drug-related crime in the next month based on the incidence of drug-related crime, fraud, assault, intimidation, auto theft, and burglary in the current month. As a result, that model is then extended with its spatial-temporal characteristics. Allowing each grid which splitting from the map regard as a sample accumulate. Each matrix represents different spatial-temporal status, thus this matrix will be used by algorithms such as Decision Tree, Classification algorithm and Additive Regression.

Carjacking is recognised as taking someone’s vehicle unlawfully. This type of crime is different from the car theft because it occurs in the presence of the victim and usually the offender uses weapons such as guns to threaten the lives of victims. Most cases of carjacking include other offenses such as kidnapping, murder and rape (David and Suruliandi 2017).

James (2017) conducted a study on carjacking in South Africa, in which he concludes that carjacking is still a serious issue in the country, showing an upward trend. This research used a snowball sampling method to investigate the reasons for the high rate of carjackings. However, there has been no study conducted using ML to detect the trends of carjacking in South Africa. The results from James’ (2017) research showed that carjacking is regularly violent, and most victims were threatened with weapons. Of these victims, 81% who suffered an injury as a result of the physical violence during the carjacking required medical treatment.

A study conducted by Waduge, Ranathunga and Lanka (2017) identified suitable ML approaches for detecting crime patterns in South Africa. The traditional crime solving techniques are now incapable of meeting the requirements demanded by the existing crime scenario. The most challenging aspect of this is identifying the set of crimes committed by the same individuals or same group of culprits. Machine learning was used to predict the patterns of crimes, and five steps in crime prediction were performed. These were data collection, classification of the data, identification of patterns, prediction of events and visualisation. K-means was then used to cluster the data. The study was able to identify the criminals who have committed multiple crimes and the patterns of those crimes were grouped together using clustering.

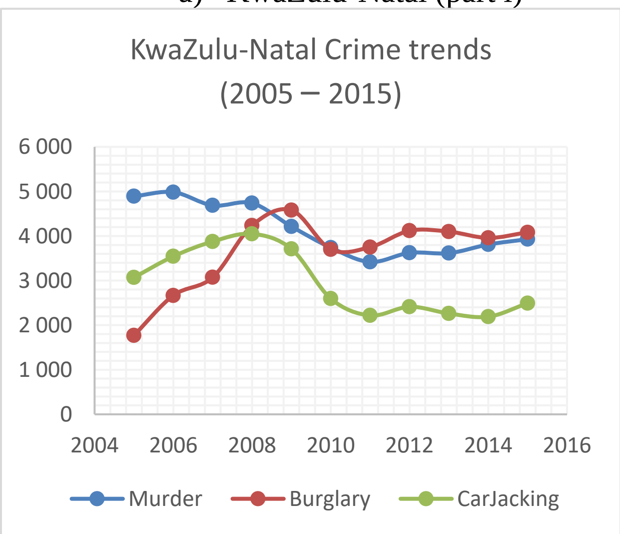
Alves, Ribeiro and Rodrigues (2018) used the ML technique called regressor to predict crime. This study was influenced by the rapidly increasing crime observed all over the world. This study primarily focused on murder, drugs, violence and carjacking (hijacking). On crime prediction, this approach achieved 97% accuracy. Regressor was able to detect the relationship of the mentioned crimes and the result revealed that most crimes were usually the result of drug abuse and alcohol consumption. The result further highlighted that in the previous five years, there had been a huge increase in drug and alcohol use. Almost 68% of crimes reported were caused by alcohol and drug abuse. This study recommended that in future, more studies should be conducted on alcohol and drug abuse as these two crimes were recognised as the major roots of crimes.

Machine learning can be used as a systematic approach for identifying and analysing patterns of crime, not only in identifying crimes but also in identifying the places that have a high probability of becoming crime hotspots. This study used a ML technique called K-means algorithm. The study mentioned few ML techniques that can be used to solve the very same crimes. ML such as classification, CPA and LR could be used to detect the patterns of crime and predict crime hotspots. Using Regressor, the study was able to identify the hotspots where crime is commonly perpetrated and a data mining framework was also developed using the geospatial plot function in Python to graphically visualize the hotspots. These results or the methods used in this study were shared with several crime departments in India in an attempt to reduce crime and solving crimes more efficiently (Jain *et al.* 2017).

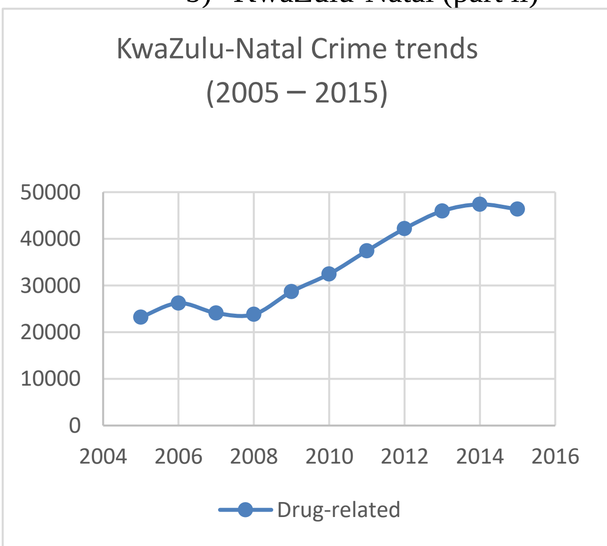
Dealing with ML usually requires a combination of methods to be performed in order to achieve the research objective. Most studies use only one method and suggest that in future, a second method should be performed following the results obtained by the first method. Almost all approaches used dealt with real time data and few studies used CPA as the data mining tool to detect crime changes. Change-point analysis has not been performed in relation to crime for any study in South Africa. For this research, after performing CPA, LR was conducted to predict the future trends based on the result obtained from the CPA. These two combinations will provide a clear understanding of crime trends based on the historical data and allow for future predictions.

The following graphs present the murder, burglary, carjacking and drug-related incidents per province. These graphs are derived from Figure 2.1 to Figure 2.4. For each province, the first column presents murder, carjacking and burglary incidents, while the second column present the drug-related incidents of the same province. These graphs are made to depict the flow of each province over the 11-year period. These graphs are grouped together because most provinces show the same trend, which is an ascending trend. However, in Gauteng (f) and Limpopo (p), drug-related crimes showed a decline between the period 2014 – 2015. Figure 2.5 (a – r) presents the four selected crimes per provinces.

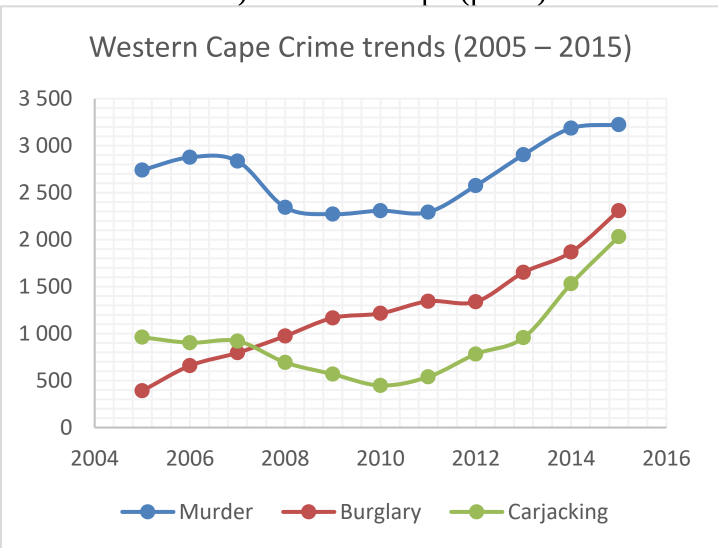
a) KwaZulu-Natal (part i)



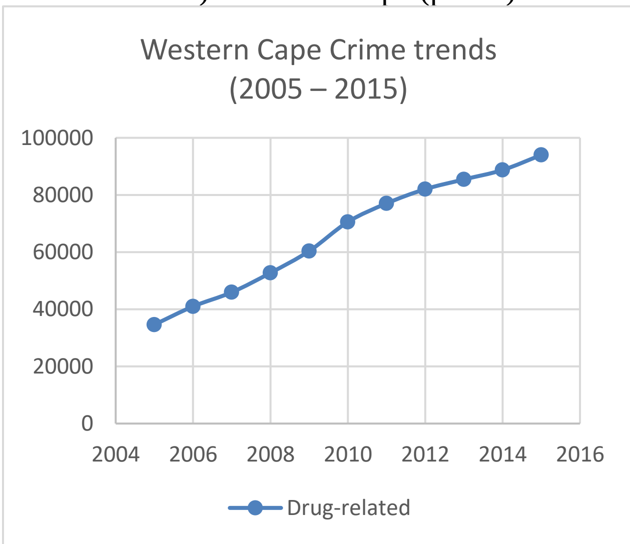
b) KwaZulu-Natal (part ii)



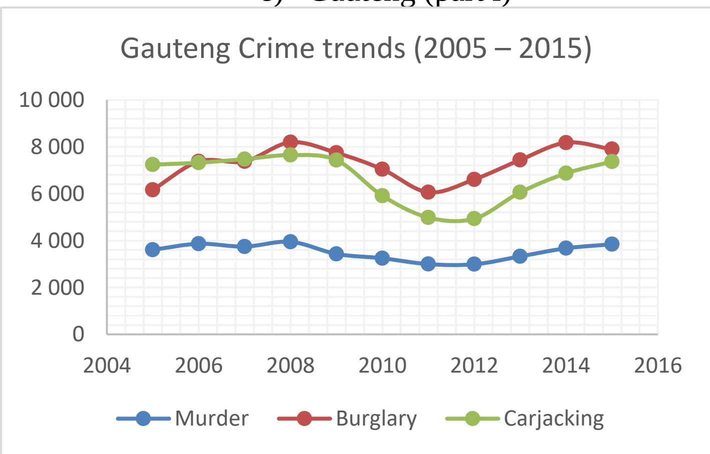
c) Western Cape (part i)



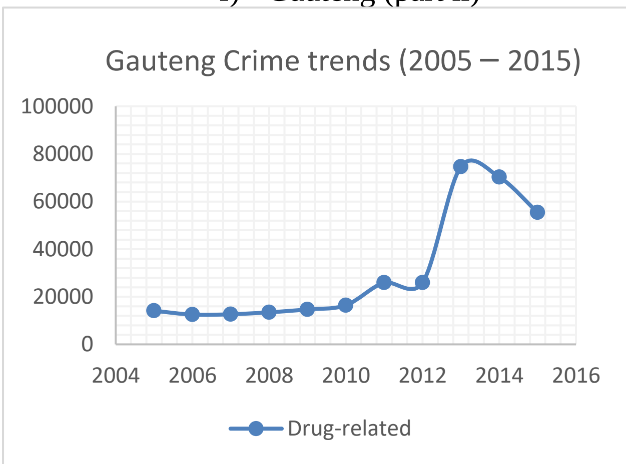
d) Western Cape (part ii)



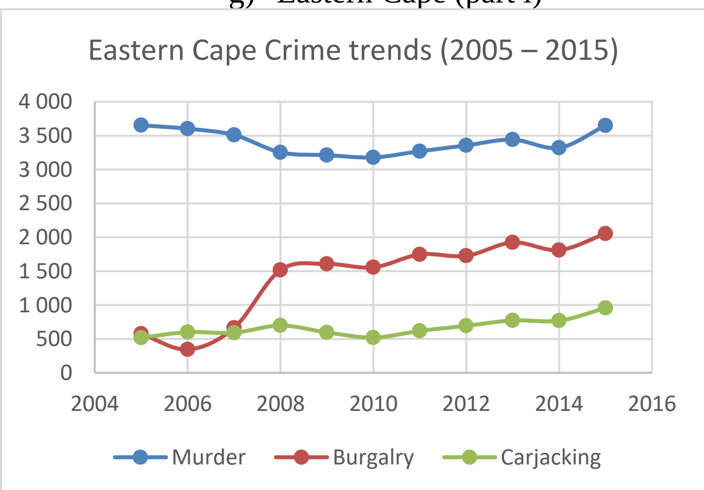
e) Gauteng (part i)



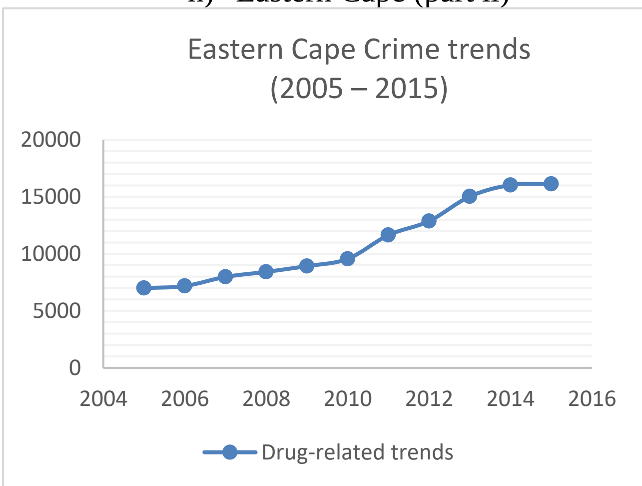
f) Gauteng (part ii)



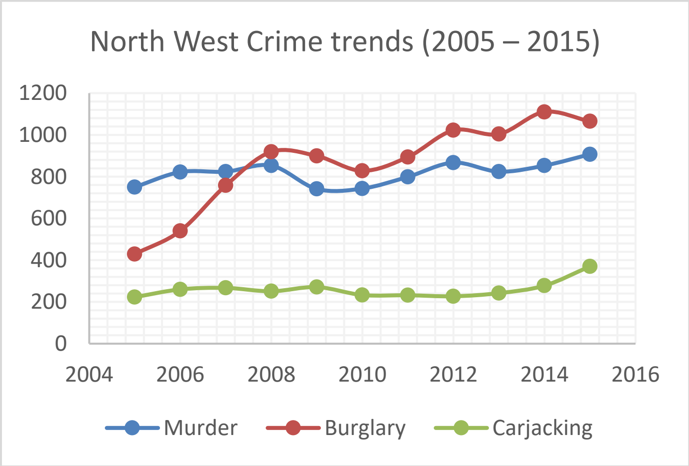
g) Eastern Cape (part i)



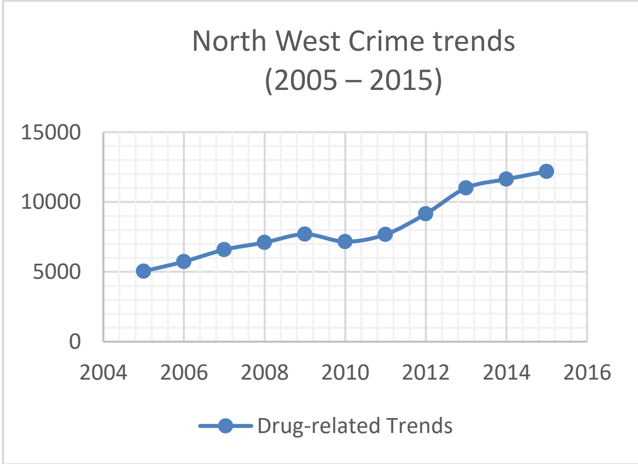
h) Eastern Cape (part ii)



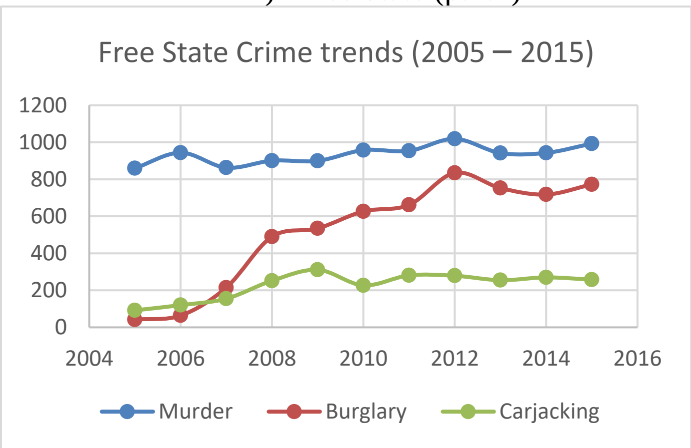
i) North West (part i)



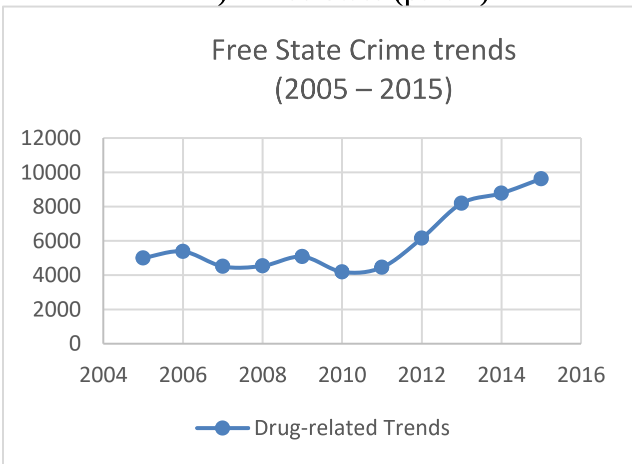
j) North West (part ii)



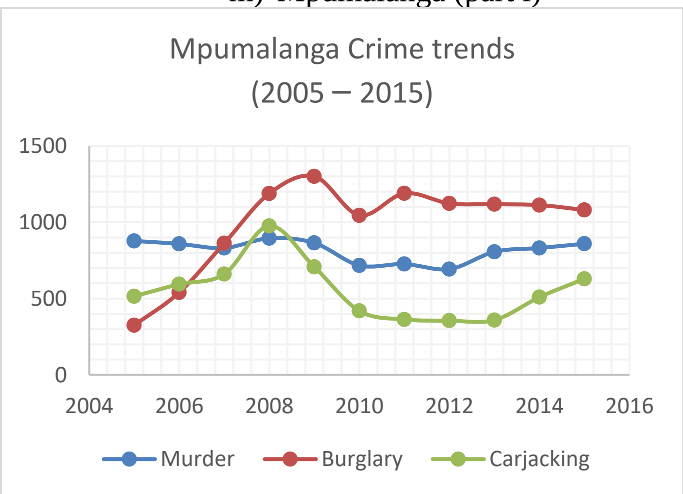
k) Free State (part i)



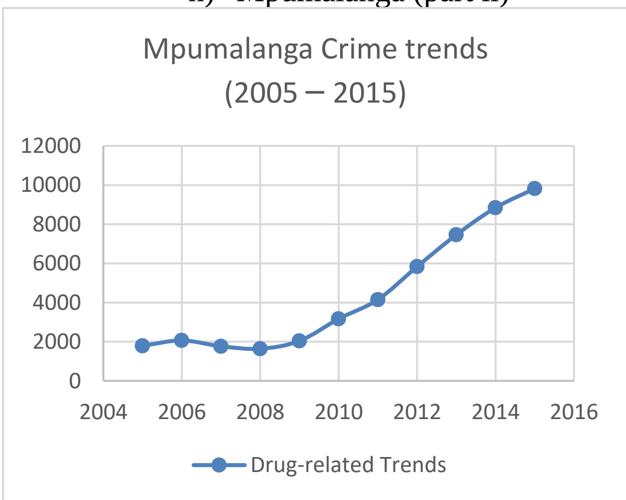
l) Free State (part ii)



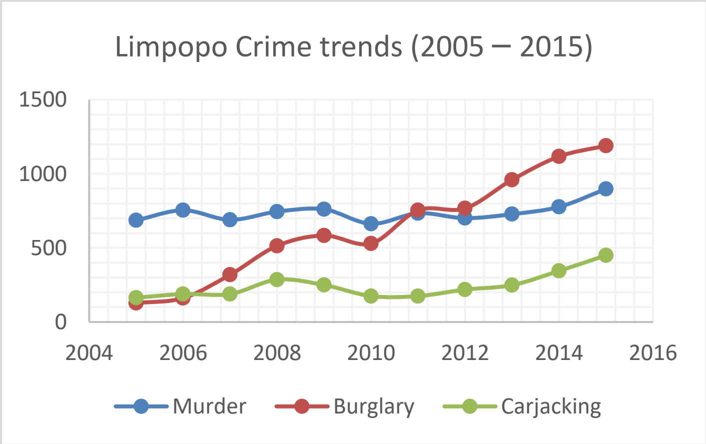
m) Mpumalanga (part i)



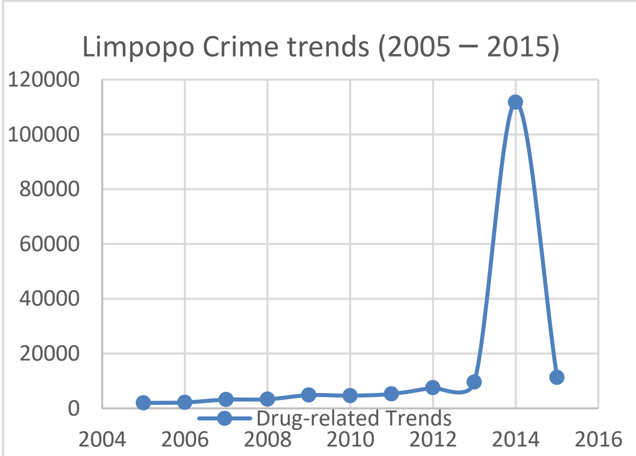
n) Mpumalanga (part ii)



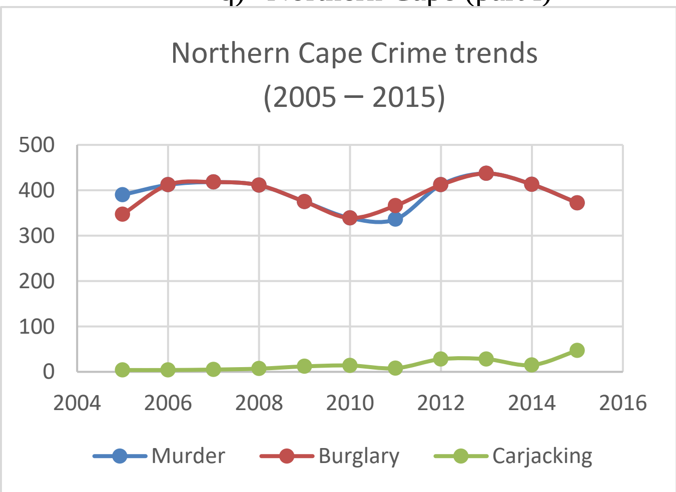
o) Limpopo (part i)



p) Limpopo (part ii)



q) Northern Cape (part i)



r) Northern Cape (part ii)

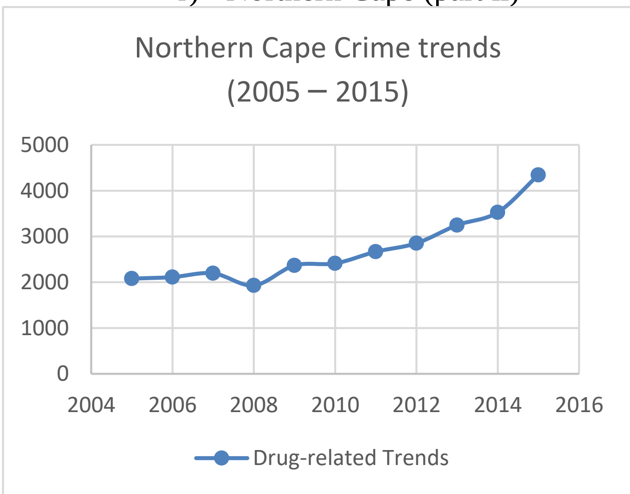


Figure 2.5 (a – r) Murder, burglary, carjacking and drug-related incidents per province (2005 – 2015)

This approach shows the hotspots of crime. Crime generators are graphically examined to explain the concentrations of crime (Hiropoulos and Porter 2014). For this study, we are concerned with burglary, carjacking, drug-related crimes and murder crime trends. These types of crimes usually occur at people's homes, and this analysis aimed to explain whether people dwelling around malls and towns are most likely to be the target of these crime. There are three different levels of crime analyses, namely, Macro, Meso and Micro (Hiropoulos and Porter 2014). Macro level refers to the level of analysis used in research, Meso level research examines on the study of teams, groups, or organisations and Micro level research examines the administrative environment including cultures and national systems (Hiropoulos and Porter 2014).

Applying Geographic Information Systems (GIS) to the CPT is well suited for data visualisation (Hiropoulos and Porter 2014). This is important as the objective with CPT is to visualise the relationship between the burglary occurrence and the hotspots. Hence, this framework is utilised to gain a better view of burglary. According to Song, Spicer and Brantingham (2013), there are three most important concepts in CPT, namely paths, edges and nodes. Nodes are referred to as the starting point and the ending point of people, where they travel to and from. Paths refer to the actual paths that people follow almost every day, such as schools, offices, and malls. Edges refer to the restriction. In this study we examined the major nodes or physical areas most likely to experience a burglary, and also to understand whether burglary occurs when people are at work or home. The CPT will help to explain how offenders find an opportunity to commit a crime in the course of their daily lives. According to Song, Spicer and Brantingham. (2013), it is the interaction of offenders with the physical and social environment that play a role in determining the targets. While engaging in routine activities, offenders may take note of a place without protection. According to CPT, crime is likely to be concentrated near hotspots (Hiropoulos and Porter, 2014).

Prabakaran and Miltra (2018) carried out a study to examine the detection techniques used in data mining. Numerous data mining techniques were analysed and combined to come up with a new combination of techniques that will detect and analyse crime.

2.4 Chapter Summary

This chapter has provided the findings of the literature review on crime in general and the different types of crimes, namely murder, burglary, carjacking, drug-related crime. This was done looking at all nine provinces of South Africa. The study further examined the major factors that contribute to crime. Unemployment and inequality were the common factors of crime mentioned throughout this chapter. Numerous studies were reviewed to determine the approaches used to solve crime across the world and in South Africa. Different types of ML were conducted in most studies to assist in reducing crime but few studies used CPA as the data mining technique to detect changes.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

In this section the selection of the appropriate research methods necessary to fulfil the research objectives and to answer the significant research questions for this study is discussed. This chapter provides insight into how to solve the aforementioned crime issues in South Africa using ML techniques

3.2 Population Size

Population size is sometimes referred to as the set of components on which the research focuses (De Vos *et al.* 2011). The study used South African crime statistics from 2005 to 2015 (an 11-year time period). The crime statistics were collected from all nine provinces of South Africa. The statistics include all documented crimes from all different areas of South Africa, from rural areas to townships, city centres and suburbs.

3.3 Research Design

This section of the study explains the techniques conducted throughout the study. The study utilised the research design formulated by the researcher to meet the objectives of the study. The research design is used as the overall strategy that integrates all the components of the study, starting by discussing the problem and analysing the data used in this study. Hence, the study further implemented the CPA. The CUSUM and bootstrap analyses were performed to detect the significant changes in crime and state when exactly the changes took place. Change-point analysis is essential in determining hidden changes (Aminikhanghahi and Cook, 2017). Figure 3.1 presents the architecture of the study.

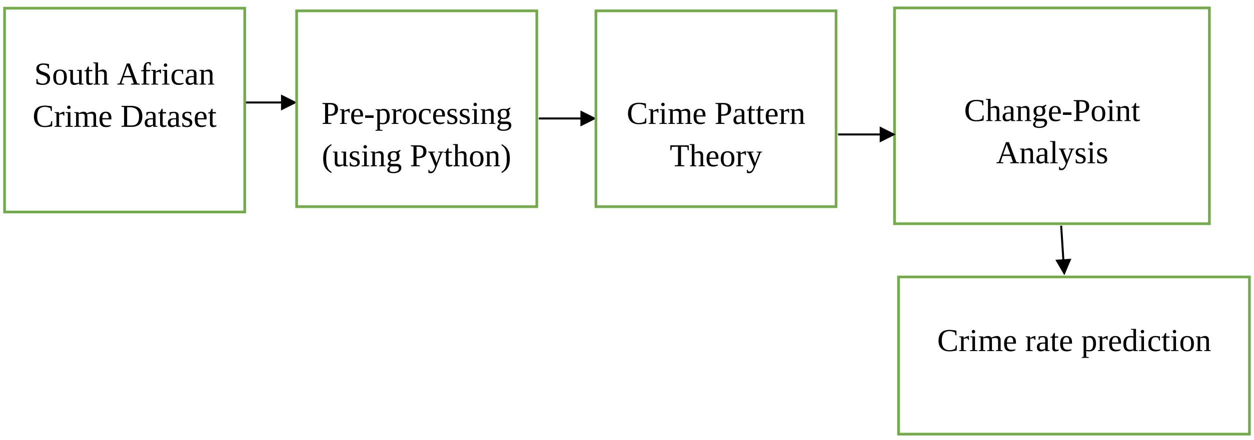


Figure 3.1 Crime Analysis Framework

3.3.1 Dataset

For this study, the South African crime statistics were used. The dataset comprises all nine provinces of South Africa and all types of crimes that have been classified and categorised. For this study, the focus is on murder, burglary, carjacking and drug-related crimes. This dataset documents crime incidence for the period 2005 to 2015, an 11-year period. The dataset of crime in South Africa is available online at https://www.kaggle.com/slwessels/crime-statistics-for-south-africa

3.3.2 Pre-processing

The pre-processing was done using Python Juypter Notebook (Anaconda 3). Actions like data cleaning, transforming data for modelling tools and breaking down a huge dataset into smaller pieces was done using a programming language called Python. It is important to do data preparation several times in no particular order (Holmostrom, Hameri and Ketokivi 2014). Pre-processing was done to prepare the data for analysis such as detecting the changes, analysing data, prediction and modelling. Python pandas allows importing and analysing of data, usually used in real-time data (Dierbach 2014). The dataset must be of time series, so CPA and Python can interact with it. Numerous stages are conducted after the data has been prepared for use.

3.3.3 Crime Pattern Theory (CPT)

To support the results obtained, the study utilised the framework of crime pattern theory. According to Song, Spicer and Brantingham (2013), the crime pattern theory was officially formalised in 1993. The CPT is an approach that examines and offers a theoretical framework to explain the dynamics of crimes in neighbourhoods (Song, Spicer and Brantingham 2013).

Eventually, the level of analysis is identified by the question being asked. This guides and determines the usefulness of the selected level of analysis. The study aims to understand the causes of burglary in certain areas of provinces where this crime is experienced most often. The CPT analyses crime in this study at the Meso level of analysis because it is mostly used for neighbourhood theories of crime. Neighbourhood theories deal with large areas such as cities, town and blocks which is most necessary for this study because the aim is to examine and gain insight into burglary in all provinces as well as identifying the hotspots. This approach discovers whether the specific provinces of South Africa have a serious problem with burglary crime.

There are different types of crime in the world, however this research decided to focus on the crimes that are mostly experienced daily in South Africa. Hence, murder, burglary, carjacking and drug-related crimes were selected for this study. The Genetic algorithm is commonly suitable for fraud detection. Crimes such as phishing and identify theft can be prevented and better understood and identified using the Genetic algorithm. Genetic algorithm can be used to assist in carjacking, drug-related crimes and burglary. This technique give insight into online theft. To find the pattern and detect the correct relationship, it is essential to have a good classification mode. Naïve Bayes was recognised as a good technique for classifying data according to their relationship. K-means Clustering was also suggested for data Clustering. After Clustering, it is important to predict and make decisions based on the prediction performed. Linear regression is one of the good leading prediction algorithms for data mining. After predicting, LR is suggested for making decisions that will best fit the model and the objectives of the study (Prabakaran and Mitra 2018).

3.3.4 Change-Point Detection

This phase implements the technique called CPA for analysing and modelling. This technique comprises of two methods called Cumulative Sum (CUSUM) and bootstrap analysis. In order to fully produce the final product of the modelling, these methods need to be executed. Then, LR is performed to predict the future trends of crime based on the existing data. This phase involved a step-by-step process to check and verify if the outcome generated is accurate, and also to check if the outcome meets the objective of the study. The final decision regarding the result obtained throughout the study is reached in this phase, after carefully considering it accuracy (Shabaz, Masood and Khan 2011).

3.3.4.1 Change-Point Analysis Detection

Change-point analysis utilises two tools to perform the complete application of change-point detection. These tools determine the significant change and the time it occurred. The CPA is usually performed on series data (Arif *et al.* 2017). The control chart consists of two lines that indicate the capacity of the data trends that are normal. These two lines are called the upper limit and lower limit. Points that appear below the lower limit or above the upper limits are recognised as outsiders, which means the crime at that time was rising or was too low. Control charts are reliable when detecting the change, but the analysis is not accurate enough (James *et al.* 2019).

Abrupt change takes place instantly, hence it was necessary for this study to perform the application of change detection to provide early advanced notice of the upcoming crime. Traditionally, changes were detected using a control chart. The difference between the two is that CPA is performed after gathering all data, while the control chart updates as soon as new data is captured (Tartakovsky *et al.* 2006). Change-point analysis detects abrupt changes that could be missed by the control chart (Taylor 2018). The CPA is recognised as the most powerful tool for detecting changes (James *et al.* 2019).

The CPA was performed to detect substantial hidden changes in the South Africa Crime statistics dataset. This study aims to expose crime and support the South African authorities of a detailed analysis of crime that could help in recognising the causes of the major changes in crime in certain provinces and years.

This study employed an experimental design with quantitative analysis. The CPA was utilised to make a solid interpretation of murder, burglary, carjacking and drug-related crime trends. Hence, CUSUM and bootstrap analyses were performed. Throughout this study, the bootstrap analysis indicates that at a certain time, the change did occur in several provinces. The minimum number that is recommended for bootstrap analysis in the experiment is 1 000 (Taylor 2018 James *et al.* 2019; Arif *et al.* 2017). To detect multiple changes, CPA uses a recursive technique, meaning it keeps on repeating the analysis until it confirms the same changes multiple times.

3.3.4.2 Change-Point Analysis Algorithm

This study applied the CPA to identify any changes in the murder, burglary, carjacking and drug-related crimes dataset. In implementing this approach, the following questions are answered:

- Did any change(s) occur?
- When did it/they occur?
- How certain is the researcher that the detected changes are accurate?

(Taylor 2018).

Three steps need to be followed and applied initially before calculating the CPA. **Do =** $\{x_1,\ldots, x_n\}$ **of size n** ($n0$ = |Do|) **The mean of** $x_1, x_2,\ldots, x_n$ **is expressed in equation 3.1**

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots x_n}{n} \tag{3.1}$$

The CUSUM always starts at zero, 0. Therefore, let S_0 be equal to zero, $So = 0$. Then, S_i is calculated as follows: S_i=S=S_{i-1}+(x_i-$\vec{x}$), i=1, 2, 3, $\ldots n$.

Before the computation of bootstrap analysis, it is essential to have restrictions for the chart, the equation below (3.2) is for calculating the magnitude of change:

$$\frac{S^i_{diff} = \max S_i - \min S_i = S_{max} - S_{min}}{i=o,\ldots,n, i=o,\ldots,n} \qquad (3.2)$$

After the action of calculating the mathematical size change, bootstrap analysis is performed multiple times N on D_0 (Arif *et al.* 2017). Each bootstrap analysis runs in the following manner:

(i) The D1 dataset of bootstrap is presented as $xj(j = 1,2,3,\ldots,n)$. The n values generate the dataset, these values are randomly stored and they are widely known as Sampling Without Replacement (SWOR).

(ii) The very same method is applied to compute the CUSUM bootstrap analysis and it is described as Sj.

(iii) The magnitude of change of CUSUM bootstrap analysis is calculated as expressed in equation 3.3:

$$\frac{S^j_{diff} = \max S_j - \min S_j = S_{max} - S_{min}}{j=o,\ldots,n, j=o,\ldots,n} \qquad (3.3)$$

(iv) When the original magnitude of change is above the magnitude of change of bootstrap, the following method is then applied.

CUSUM, $S^i_{diff} > S^j_{diff}$, the addition of bootstrap is calculated. Let N represent the number of executed bootstrap sample, and let K represent the number of bootstraps for which $S^i_{diff} > S^j_{diff}$, where the confidence level that a change has occurred as a percentage is defined in equation 3.4:

$$CL = \frac{\sum_{i=1}^{N} I(S^i_{diff} > S^j_{diff}) * 100}{N} \qquad (3.4)$$

The bootstrap ends up in an independent error structure (Albrecher *et al.* 2019), which is a distribution-free method with only one assumption. Independent error structure, as depicted below, are errors that are distributed from bootstrap, as shown in equation 3.5:

$$X_i = S_i + e_i \tag{3.5}$$

where m_i denotes the mean at the time I, e_i is a minor fault correlated with the i-th value, and the independent e_i is assumed to have a zero (0) mean value, to be distributed. Usually, $m_i = m_i - 1$, except for a small number of values of i which are called change points.

As soon the change is discovered, an estimate of the exact time of the change is calculated. CUSUM estimator formula is defined in equation 3.6:

$$\frac{|S_m| = max\ |S_i|}{i=0,\ldots,n} \tag{3.6}$$

S_m is the point in the CUSUM that is far from the zero value. Point m represents the last estimated point before the change occur while $m+1$ represents the first estimated point after the change. The second estimator, namely Mean Square Error (MSE) is used after the change occurs.

MSE(m) described as $MSE(m) = \frac{min}{i=1\ldots|D|}$ MSE(i) for a given sub-dataset D as follows (Albrecher *et al.* 2019). The bootstrap analysis is performed to detect the changes, and it is recursive (Albrecher *et al.* 2019). It detects multiple changes that could be found in the same dataset.

3.3.5 Crime Rate Prediction

After changes have been detected, LR, a supervised ML algorithm is adopted to predict future trends based on the data that already exists. The future prediction will range from 2016 to 2022 because the dataset used contains crime data until 2015. Consequently, comparison of these results will be done against the actual statistics of crime to measure the accuracy of the prediction model. The results are discussed and interpreted in the next chapter. Linear regression equations are very easy to understand and interpret; it makes the estimation procedure user-friendly when compared to other ML algorithms (Nimon and Oswald 2013). Hence, the study performed LR to obtain quick and accurate results that are easy to interpret.

3.3.5.1 Linear Regression

This phase of the study discusses how the LR was conducted and why it is important in this study. The aim of using LR is to analyse the current data in order to make an accurate prediction. Jumin *et al.* (2020) suggested that using LR is very significant for most scientific studies because it provides complete and accurate results.

Prediction of LR has the coefficients of β0 and β1 in the standard formula. This is done by studying the red straight line that indicates the relationship between the two variables (Zou, Tuncali and Silverman 2003; Jumin *et al.* 2020). The following is the standard formula of LR, as expressed in equation 3.7:

$$y = a + bx \tag{3.7}$$

In the formula, X represents the independent variable, Y represents the dependent variable, while a is the intercept, b indicates the slope of the line.

In order to produce accurate results, the confidence intervals and hypothesis tests should be calculated. Given the values of N observations on X and Y, the method of least square estimates β0 and β1. Since the red estimated line fits exactly the estimated data, a term for the discrepancy between the actual and fitted data must be included. This is represented in equation 3.8:

$$\begin{aligned} Yj &= bo + b1Xj + ej \\ &= yj + ej \end{aligned} \tag{3.8}$$

where j presents the observation number, bo estimates bo, b1 estimates b1, and ej is the discrepancy between the actual data value yj and the fitted value given by the regression equation. This discrepancy is namely residual.

Another aspect to notice here is the importance of X and Y. The equation predicts Y from X, hence Y is a dependent variable because it depends on the value of X. The influence of all variables of the value of Y is grouped into the residual.

3.4 Relationship Between Change Point Analysis and Linear Regression

In this study there is a strong link between the three concepts adopted that is, Python, CPA and LR. Firstly, Algorithms and methods were used to clean and prepare the data for the CPA to be performed. This is necessary especially for avoiding unnecessary and incorrect feedback. Secondly, CPA software was used to identify the sudden changes that occurred. Change-point analysis only uses existing data to detect minor changes that are difficult to recognise using the human brain alone. Lastly, LR is performed to predict future trends based on existing data. Linear regression makes the procedure of prediction very simple and understandable. Its interpretation of data is unique and accurate. In this case, we used South African crime statistics from 2005 to 2015 to predict the possible trends for the period 2016 to 2022. Before predicting, the data needed to be cleaned and prepared for the prediction. Linear regression was then used to predict future crime trends from 2016 to 2022. Each type of crime was analysed separately. The study further used a proportional area chart to show the relative sizes of hotspots in all nine provinces.

3.5 Chapter Summary

In this chapter, the researcher explained the dataset used, population size, techniques and designs. This chapter further described in detail how the dataset and tools were used and also how the research design was performed in order to achieve the research objectives. The chapter explained how all three different tools namely, Python, CPA and LR were utilised in analysing the dataset.

CHAPTER FOUR: RESULTS, ANALYSIS AND DISCUSSION

4.1 Introduction

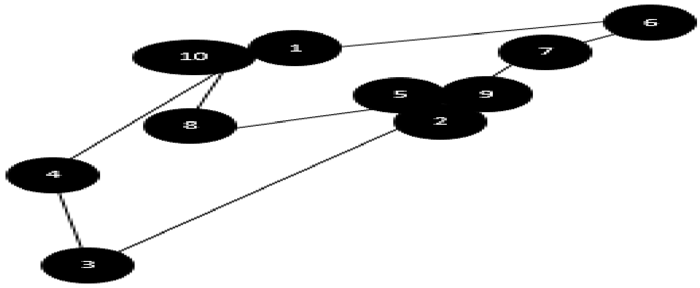
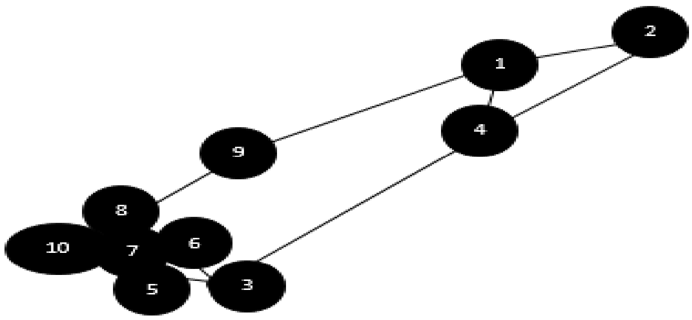
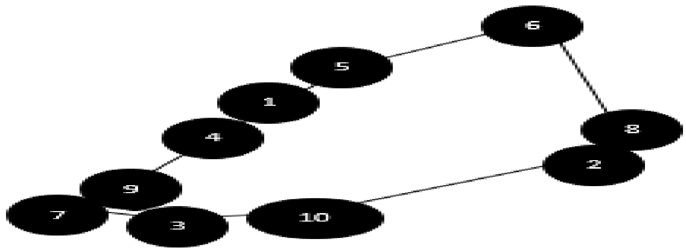
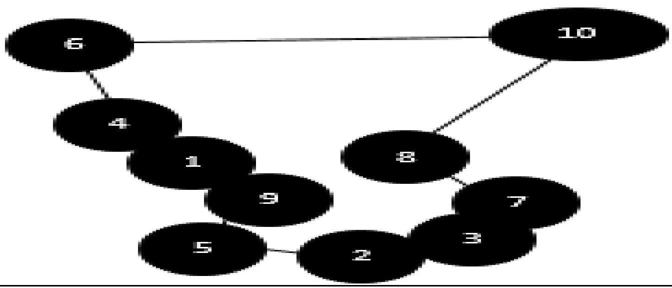
In this section, the results, discussions and interpretation of the findings are presented. The CPA was performed to determine when changes occurred and with what confidence. Liu *et al.* (2013) mentioned that CPA can be utilised in any sort of data. Control limits are horizontal lines that are used to restrict the points from exceeding the maximum number of change (Fearnhead and Rigaill 2019). All four selected types of crime in this study are analysed and discussed separately and all findings and change points are revealed and interpreted. The results of CPA are presented after the interpretation of data is performed.

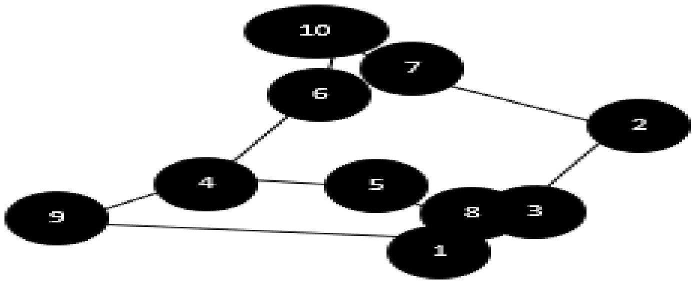
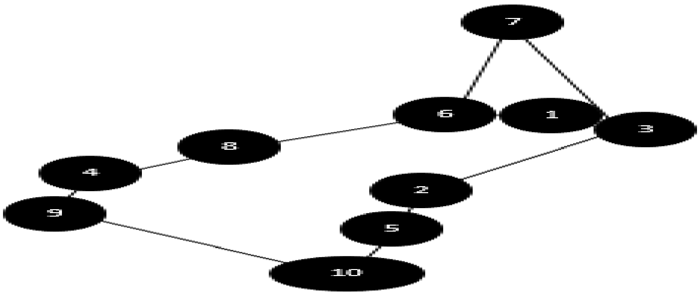
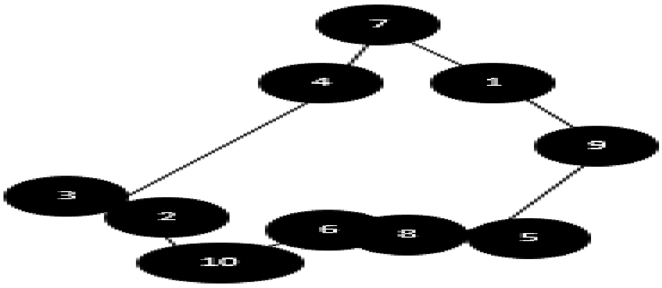
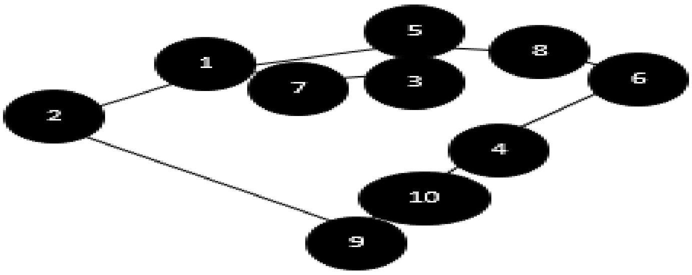
4.2 Crime Pattern Theory Results

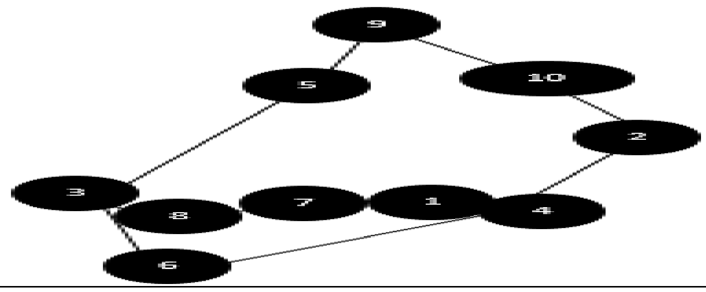
To support the results of the CPA, the study utilised the framework of the Crime Pattern Theory (CPT). Utilising this approach is essential because it provides a detailed explanation of why people commit a crime in certain areas. This approach was carried out by identifying the hotspots with high number of crime and providing the graphical spatial map of all nine provinces. The CPA detected the changes that occurred in previous years, and the CPT was used to precisely describe what attracted the offenders to target those places and why there was a sizeable change in certain years. The spatial maps presented below were derived from the official map of South Africa, and they were created to depict the hotspots of burglary crime for each province.

Table 4.1 presents ten burglary hotspots identified in the dataset used in this study. These hotspots are put in order and they are ranked in terms of the number of crimes committed. The spatial representation of these hotspots indicates the shapes and the distance between one hotspot and another. The numbers in brackets represents the incidents of residential burglary over the 11-year period.

Table 4.1 South African top 10 hotspots for burglary in each province (2005 – 2015)

KwaZulu-Natal top ten hotspots of burglary ranked	**KZN spatial representation of crime pattern**
1. Inanda (1 859)	
2. Ntuzuma (1 763)	
3. Empangeni (1 508)	
4. Esikhaleni (1 320)	
5. Pinetown (1 319)	
6. KwaDukuza (1 105)	
7. Umlazi (1 084)	
8. Hillcrest (1 032)	
9. Greenwood Park (912)	
10. Westville (836)	
Eastern Cape top ten hotspots of burglary ranked	**Eastern Cape spatial representation of crime pattern**
1. Kwazakele (1 050)	
2. New Brighton (725)	
3. Walmer (487)	
4. Mthatha (574)	
5. Ngangelizwe (374)	
6. KwaNobuhle (361)	
7. Lusikisiki (360)	
8. Duncan Village (346)	
9. Bathelsdrop (341)	
10. Butterworth (309)	
Free State top ten hotspots of burglary ranked	**Free State spatial representation of crime pattern**
1. Thabong (644)	
2. Park Road (473)	
3. Bloemspruit (439)	
4. Welkom (281)	
5. Sasolburg (252)	
6. Kopanong (242)	
7. Kagisanong (226)	
8. Phuthaditjhaba (187)	
9. Mangaung (178)	
10. Botshabelo (177)	
Western Cape top ten hotspots of burglary ranked	**Western Cape spatial representation of crime pattern**
1. Nyanga (1 420)	
2. Harare (983)	
3. Khayelitsha (842)	
4. Gugulethu (587)	
5. Mitchell's Plain (453)	
6. Milnerton (451)	
7. Mfuleni (449)	
8. Delft (439)	
9. Philippi (419)	
10. Kraaifontein (392)	

Northern Cape top ten hotspots of burglary ranked	Northern Cape spatial representation of crime pattern
1. Kimberley (79)	
2. Mothibistad (54)	
3. Kuruman (50)	
4. Galeshewe (49)	
5. Kathu (45)	
6. Bathlaros (26)	
7. Hartwater (25)	
8. Postmasburg (23)	
9. Upington (23)	
10. Kagisho (22)	
Gauteng top ten hotspots of burglary ranked	**Gauteng spatial representation of crime pattern**
1. Honeydew (3 546)	
2. Sandton (3 131)	
3. Midrand (1 929)	
4. Ivory Park (1 904)	
5. Wierdabrug (1 842)	
6. Tembisa (1 624)	
7. Douglasdale (1 568)	
8. Roodepoort (1 542)	
9. Boysens (1 467)	
10. Mondeor (1 353)	
Mpumalanga top ten hotspots of burglary ranked	**Mpumalanga spatial representation of crime pattern**
1. Witbank (840)	
2. Kanyamazane (783)	
3. Tonga (740)	
4. Kabokweni (619)	
5. Nelspruit (520)	
6. Calcutta (477)	
7. Masoyi (446)	
8. Vosman (401)	
9. Pienaar (323)	
10. Hazyview (312)	
North West top ten hotspots of burglary ranked	**North West spatial representation of crime pattern**
1. Rustenburg (1 247)	
2. Mmabatho (768)	
3. Brits (721)	
4. Boitekong (393)	
5. Mooinooi (373)	
6. Phokeng (337)	
7. Jouberton (318)	
8. Bethanie (301)	
9. Tihabane (300)	
10. Klerksdorp (274)	

Limpopo top ten hotspots of burglary ranked	Limpopo spatial representation of crime pattern
1. Thohoyandou (600)	
2. Seshego (379)	
3. Mankweng (361)	
4. Polokwane (322)	
5. Tzaneen (269)	
6. Maake (228)	
7. Musina (227)	
8. Giyani (208)	
9. Lebowakgomo (178)	
10. Makhado (170)	

The hotspots are listed on the left side. Starting from 1 to 10 are the highest burglary hotspots and the spatial representation of the crime pattern is shown on the right side for each province. The spatial representation of the crime pattern of the hotspots comprises shapes, nodes and edges. These spatial representations of crime pattern are extracted from the official map of South Africa, demarcated into nine provinces to easily interact with each province as each is treated separately. The 10 selected hotspots are the highest-ranking hotspots of burglary in all provinces. The spatial representation of crime pattern aims to provide insight us by providing the direction and the relationship between hotspot A and hotspot B. Not surprisingly, it is visible that 80% of hotspots are close to each other, and the spatial representation of crime patterns surround the cities, towns and townships where there are large numbers of people.

The CPT was performed to give insight and explain why people commit burglary crime in the hotspot identified. For this study, nodes are the person activity space, for instance, shopping centres, school, home and work. Paths represent the routes designed for movement, for instance sidewalks and streets. Edges are barriers between two places, such as rivers and seas (Hiropoulos and Porter 2014; Liu *et al.* 2013). For this study, edges and paths are not graphically represented but they are both elucidated deeply to the hotspots identified.

Figure 4.1 shows the proportional area chart of the burglary hotspots crime incidences. The essential of using the proportional area chart is to show the relative sizes of hotspots. According to Wilkinson (2011), proportional area chart is mostly used for comparing proportions (size and quantity), and it provide a clear overview of data without using the scales. Figure 4.1 presents the data that is in Table 4.1 by using the proportional area chart to provide a better view of burglary hotspots incidences.

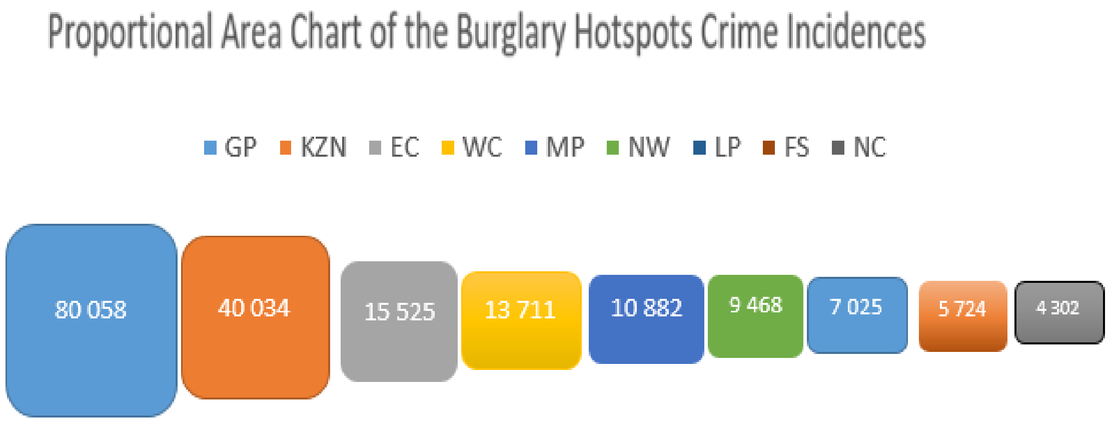


Figure 4.1 Burglary top 10 hotspots per province (2005 – 2015)

Figure 4.1 presents the top 10 hotspots of burglary crime incidents that have occurred in all nine provinces of South Africa from 2005–2015. Each square presents the province and the size present the number of cases that have been reported. The number within each square is the sum of the top 10 hotspots of burglary that has been identified in Table 4.1. In Figure 4.1, the province of Gauteng the showing the high number of 80 056 incidences of burglary, while the Northern Cape is showing the least number of 4 302 incidences of burglary.

4.2.1 Using CPT to Explain Burglary Presented in the Geographical Spatial Representation of Crime Pattern

Table 4.1 shows that Inanda is the leading hotspot in KwaZulu-Natal. Inanda is the highest burglary hotspot in KZN with 1 859 reported cases. This township is situated roughly 24km north-west of the Durban Central Business District (CBD) (South Africa 2020). Inanda is subdivided into smaller sections or villages called Inanda Newtown A, B, C; Emachobeni; Inanda Glebe; and Amaoti (South Africa 2020). When the number of crimes committed in a certain place increases, it is usually because the offenders have seen a weakness and an enticing number of targets (Hiropoulos and Porter 2014). People might increase in an area if there are job opportunities. For instance, Inanda is a township that have a big shopping centre near it namely Bridge City. Most people during the day are either at work or out shopping (Anon. 2020a). According to the CPT, this gives offenders ample opportunity to commit burglary crime. The spatial representation of the crime patterns shows the relationship or a link between the hotspots identified. The hotspots identified in KZN are not very far apart, as depicted in Table 4.1 above. This could be seen to mean that offenders know those in their communities and they are fully aware of the route people usually take every day. Many people live in houses with little to no security such as burglar guards and alarms; this is one reason offenders find it easy to commit burglaries (South Africa 2020). People migrate to Inanda in search of job opportunities, that according to Anon. (2020a) is why there are large numbers of people living in houses in very close proximity close to one another. This increases crime as offenders take advantage of the new opportunities that arise.

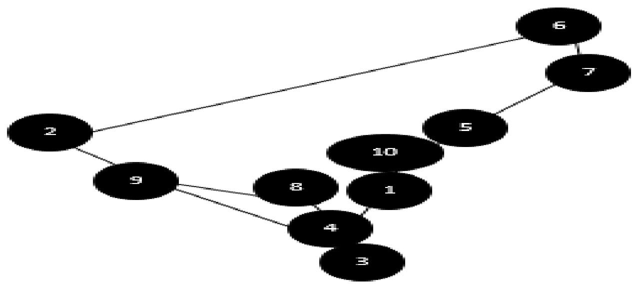
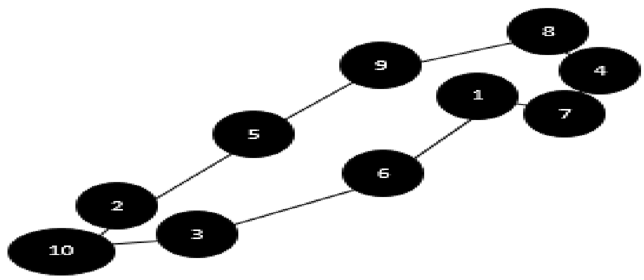
Eastern Cape (EC) is the second largest province in South Africa and its population is the third largest of all nine provinces (Busakwe and Jongikaya, 2007). Kwazakele is a township in EC with a generally high number of burglary cases reported (Wessels, 2020). A study by AmiriKerry, BrooksBryan and Daratha (2019) noted that most people do not report burglary crime to the police because of the perceived incompetence of and corruption by SAPS. Eastern Cape hotspots are very close to each other, as indicated in Table 4.1. The close hotspots identified are around cities and townships, where there are plenty of malls, shopping centres and businesses. People dwelling around town are most likely to be targeted because 60% of hotspots are in towns (Song, Spicer and Brantingham 2013). Busakwe and Jongikaya (2007) mentioned that 56% of crimes committed in Ngangelizwe and Kwazakele are burglaries. It is the most problematic crime in the area, and this crime frequently involves the use of illegal firearms.

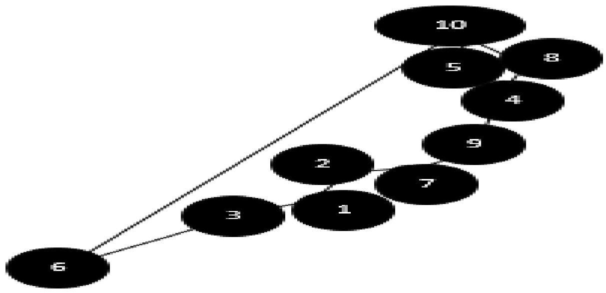
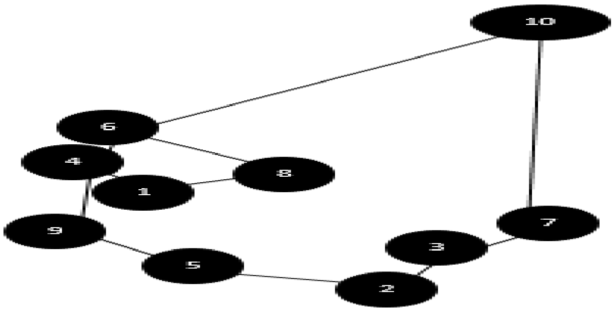
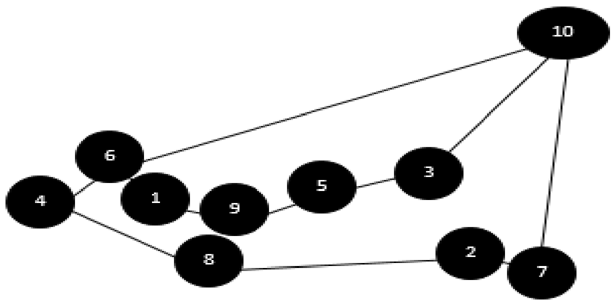
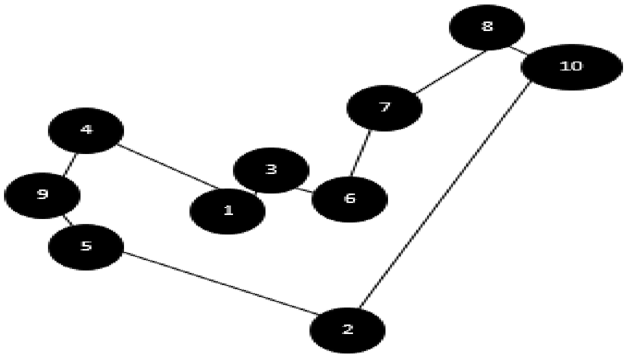
Thabong, a township in Free State (FS), had the highest number of burglary cases (644) reported to the police (Wessels, 2020). Thabong is the second largest township, after Botshabelo, in the Free State Province (Anon. 2020b). Botshabelo is also a burglary hotspot mentioned in Table 4.1. According to the Research Channel Africa (2019), the reason for the high crime rate in FS is because 73.4% of SAPS stations are understaffed. Hence, offenders take advantage of the situation. Research Channel Africa (2019) further mentioned that sometimes people in the community do not get the help that they need from the SAPS. The nodes of FS are next to each other, typically every city has a township around it. This is relevant because people dwelling in a township can travel quickly and easily to work or to the shops. Unfortunately, the lack of protection is a major problem in FS. Crime will occur if an area provides the opportunity for crime and it exists within an offender's space of awareness. Consequently, an area that provides shopping, recreation and restaurants such as at shopping malls, attract offenders to target people’s houses for burglary (Hiropoulos and Porter 2014).

For the other six provinces, the crime pattern appears to be the same. Most hotspots are located in CBDs where there are several malls and working environments. The results are similar to other hotspots represented above. Factors such as housing insecurity and the lack of support from the police have been common. Shopping centres and malls are recognised as crime attractors because people go to the mall or work every day and no one remains behind at home. That is why Amiri *et al.* (2019) mentioned that most incidents of burglary occur during the day while residents are at work. An example of how daring and brazen burglars can be, a couple from Northern Cape in South Africa whose house was protected by alarms, secured with burglar guards and a metal security door experienced a horrific robbery in their home in the early hours of one morning in 2018. Items such as laptop computers, watches, a camera and cellular phones were stolen (Beangstrom 2018). Few cases of burglary are reported on weekends. The reason for this may be because most adults do not work on weekends, neither do children usually attend school. Given the justification for burglary crime, the CPA results for the crime of murder is discussed in section 4.3.

Table 4.2 presents ten hotspots of murder identified in each province of South Africa. These hotspots are put in order and they are ranked in terms of the number of crimes committed. The spatial representation of these hotspots indicates the shapes and the distance between one hotspot and another.

Table 4.2 South African top 10 hotspots for murder in each province (2005 – 2015)

KZN top ten hotspots of Murder ranked	**KZN spatial representation of crime pattern**
1. Inanda (2 043)	
2. Umlazi (1 888)	
3. Ntuzuma (1 740)	
4. Plessislaer (1 247)	
5. Empangeni (899)	
6. Esikhaleni (766)	
7. KwaDukuza (740)	
8. Mpumalanga KZN (737)	
9. Marianhill (719)	
10. KwaMashu (675)	
Eastern Cape top ten hotspots of Murder ranked	**Eastern Cape spatial representation of crime pattern**
1. Mthatha (1 480)	
2. Kwazakele (1 042)	
3. New Brighton (1 038)	
4. Lusikisiki (1 006)	
5. Bethelsdorp (823)	
6. Ngcobo (806)	
7. Duncan Village (776)	
8. Mbizana (692)	
9. Mount Frere (666)	
10. Bethelsdorp (659)	
Free State top ten hotspots of Murder ranked	**Free State spatial representation of crime pattern**
1. Bloemspruit (793)	

2. Thabong (751)	
3. Welkom (448)	
4. Kopanong (392)	
5. Kagisanong (367)	
6. Boithuso (334)	
7. Odendaalsrus (294)	
8. Botshabelo (288)	
9. Mangaung (281)	
10. Bronville (269)	
Western Cape top ten hotspots of Murder ranked	**Western Cape spatial representation of crime pattern**
1. Nyanga (2 845)	
2. Khayelitsha (1 738)	
3. Harare (1 620)	
4. Gugulethu (1 602)	
5. Kraaifontein (1 219)	
6. Delft (1 187)	
7. Mfuleni (1 089)	
8. Mitchell's Plain (917)	
9. Philippi East (749)	
10. Bishop Lavis (598)	
Northern Cape top ten hotspots of Murder ranked	**Northern Cape spatial representation of crime pattern**
1. Galeshewe (410)	
2. Kimberley (218)	
3. Upington (174)	
4. Rosedale (167)	
5. Prieska (145)	
6. Sunrise (143)	
7. Kagisho (126)	
8. Douglas (119)	
9. Groblershoop (119)	
10. Kakamas (117)	
Gauteng top ten hotspots of Murder ranked	**Gauteng spatial representation of crime pattern**
1. Tembisa (1 175)	
2. Katlehong (860)	
3. Alexandra (852)	
4. Hillbrow (843)	
5. Ivory Park (843)	
6. Roodepoort (817)	
7. Jeppe (754)	
8. Moroko (725)	
9. Johannesburg Central (718)	
10. Dobsonville (686)	

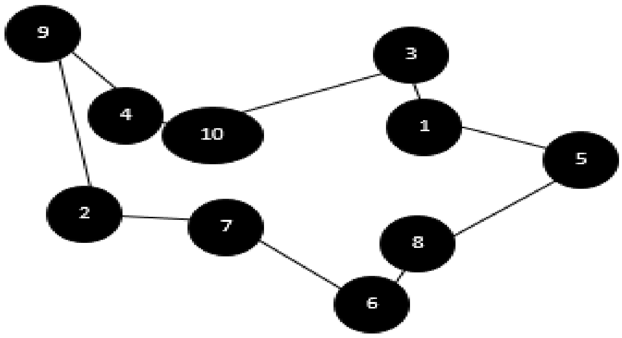
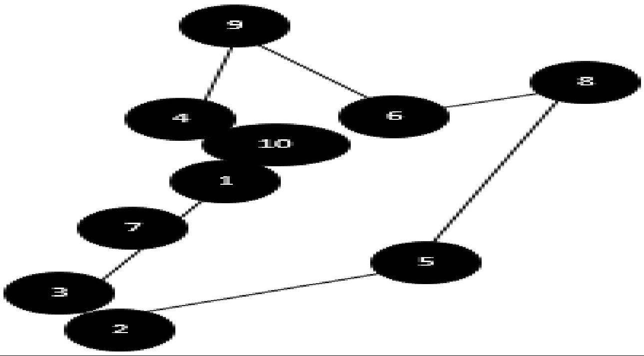
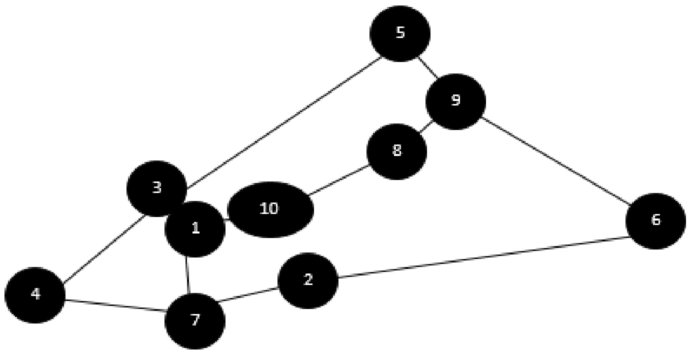
Mpumalanga top ten hotspots of Murder ranked	**Mpumalanga spatial representation of crime pattern**
1. Embalenhle (519)	
2. Vosman (401)	
3. Tonga (378)	
4. Witbank (358)	
5. Ermelo (353)	
6. KwaNyamazane (343)	
7. Calcutta (312)	
8. Piet Retief (310)	
9. Kwamhlanga (293)	
10. Masoyi (292)	
North West top ten hotspots of Murder ranked	**North West spatial representation of crime pattern**
1. Rustenburg (684)	
2. Jouberton (542)	
3. Phokeng (443)	
4. Boitekong (439)	
5. Marikana (394)	
6. Kanana (375)	
7. Brits (316)	
8. Ikageng (314)	
9. Tihabane (233)	
10. Mogwase (227)	
Limpopo top ten hotspots of Murder ranked	**Limpopo spatial representation of crime pattern**
1. Seshego (505)	
2. Thohoyandou (404	
3. Mankweng (368)	
4. Mahwelereng (300)	
5. Maake (288)	
6. Polokwane (276)	
7. Giyani (244)	
8. Bolobedu (222)	
9. Lebowakgomo (211)	
10. Ritavi (199)	

Figure 4.2 shows the proportional area chart of the murder hotspots crime incidences. The essential of using the proportional area chart is to show the relative sizes of hotspots. Figure 4.2 presents the data that is in Table 4.2 by using the proportional area chart to provide a better view of murder hotspots incidences.

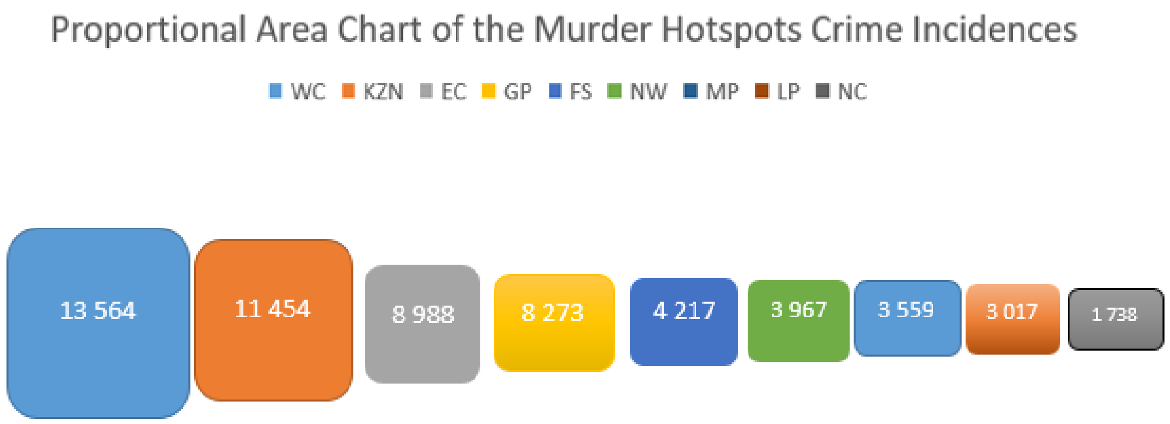


Figure 4.2 Murder top 10 hotspots per province (2005 – 2015)

Figure 4.2 presents each province and it top 10 hotspots of murder with the sum of incidents that has occurred. Each square presents the province and the size present the number of cases that has been reported. The number within each square is the sum of the top 10 hotspots of murder that has been identified in Table 4.2. The province of Western Cape shows the highest number of murder incidences while Northern Cape shows the least number of murders.

4.2.2 Using CPT to Explain Murder Presented in Geographical Spatial Representation of Crime Patterns

According to Wessels (2020), murder is one of the leading crimes in South Africa with KwaZulu-Natal (KZN) province having the highest number of murder crimes recorded in 2015. Inanda is leading with the 2 043 cases reported to the police. Waterworth, Guy and Chemaly (2020) further buttresses Wessels claim stating that out of all nine provinces of South Africa, KZN is the most dangerous province, with the highest levels of violence and the highest number of murder crimes. Murder crime in South Africa is very high. The spatial representation of crime patterns represents the relationship or link between the hotspots identified. The hotspots identified in KZN are not very far from each other, as depicted in Table 4.2. Using CPT, it was discovered that in KZN, more than 50% of murder hotspots are found in townships. According to Cele (2020), 370 people in KZN were arrested in a month of December 2020 for committing various crimes including murder. The use of drugs and alcohol has been identified as a big influence in crime, especially in townships (Cele 2020).

Gauteng remains the second-worst province in terms of murder and violent crimes. According to Phagane (2020), Gauteng had more than 4 500 murder committed between April 2019 and March 2020. Hotspots identified were located around townships, just as in the case of KZN hotspots. Most murder are committed in places of entertainment (Retorius, Cleak and Morgan 2010). Johannesburg central has been well known as a place of drug use, where alcohol and drugs are abused (Ngqakamba 2019). As depicted in Table 4.2, murder occurs mostly in townships and the hotspots are near each other.

4.2.3 Crime Pattern Theory Analysis for the Seven Remaining Provinces

The crime pattern for the other seven provinces was observed to be similar to the two provinces discussed above. The issues of violence are the same in all provinces and the number of murder and the use of drugs is increasing. Most hotspots are located in townships and a few are located in CBDs where there are a large number of people. The results are similar to other hotspots represented above. The reason for this high number of hotspots in townships is due to the high level of unemployment. According to Hiropoulos and Porter (2014), CPT is based on the claim that crime occurs if an area provides the opportunity for crime. The use of drugs and alcohol abuse is very high; the effect is often violence. For instance, in the Eastern Cape, police found that young children are used as drug mules in certain areas of townships (Poti 2019). The most common issue that leads people to commit crimes is the lack of opportunities (Matzopoulos *et al.* 2018). Hence, crime is very high especially in townships because most people do not have jobs or any activity with which to keep occupied. For this study, CPT was used to explain the factors that attract or generate crime in certain areas of the hotspots identified.

4.3 Murder Crime Change-Point Analysis Results

The crime statistics for murder, burglary, carjacking and drug-related crime were collected from all nine provinces of South Africa. This was done by collecting all the documented cases from the police stations (Wessels 2020). The breaking down of this huge dataset was conducted utilising in Python Jupyter Notebook (Anaconda 3) using methods and algorithms. The content presented in Table 2.1 presents all four selected crimes starting from 2005 to 2015. This data was utilised to perform the CPA, and each province was approached and analysed separately for each of the four types of crime selected.

Figure 2.2 shows murder trends, KwaZulu-Natal shows the highest total (45 648) of murder over the period under review. KwaZulu-Natal is followed by the province of Gauteng and

Eastern Cape with 38 655 and 37 445 murders respectively. Figure 2.5 shows that most murders were committed in 2006, within this dataset.

To fully understand the results obtained in this study, the experiment of CPA presented the results using two graphs and a table. The first graph depicts the background changes and the control limits. The y-axis of the CPA graph is the total number of crime incidents. The CPA uses ranks to represent the y-axis. According to Arif *et al.* (2017), the importance of ranks in the CPA is to sort a given data into order and replacing each value with its corresponding position in the order. The x-axis presents the years. The second graph depicts the CUSUM chart, that was responsible for indicating when exactly the change occurred. The graphical representation of murder for all nine provinces is depicted in Figures 4.3 to 4.20. Figure 4.3 presents the CPA results which comprise control limits and the background changes. Figure 4.4 presents the results of the CUSUM chart.

Figure 4.3 depicts the graphical representation of the CPA, with control limits and the background change, as related to KZN. Change-point analysis offers multiple benefits, including detecting smaller sustained shifts; it is simpler to use if numerous changes have occurred. Using this approach is essential because it detects hidden and smaller shifts that could be difficult to detect when using other methods. In Figure 4.3, changes can be observed and the existence of the blue region indicates the change or changes that has took place. The points that appear above and below the control limits are called outliers.

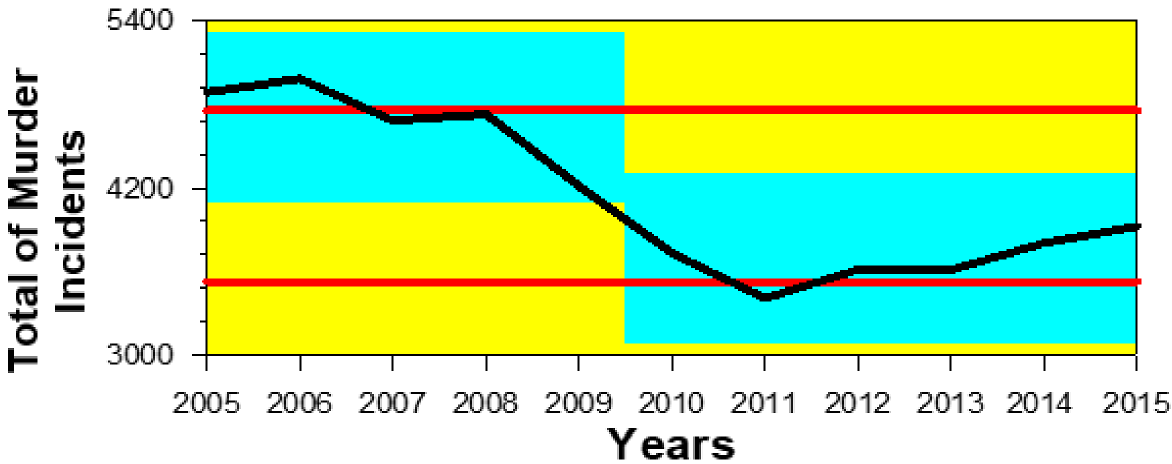


Figure 4.3 KwaZulu-Natal murder data: background changes (2005 – 2015)

Figure 4.4 presents the CUSUM chart of analysis for KZN. This chart is constructed firstly by adding the CUSUM and the plot is based on the data given (Kuncheva 2013; Fearnhead and Rigaill 2019). A significant change was observed in 2009. The slope of the chart changed once. This could be seen to mean that from 2010, incidences of murder crime in KZN started to change, as shown in Figure 2.5 (a).

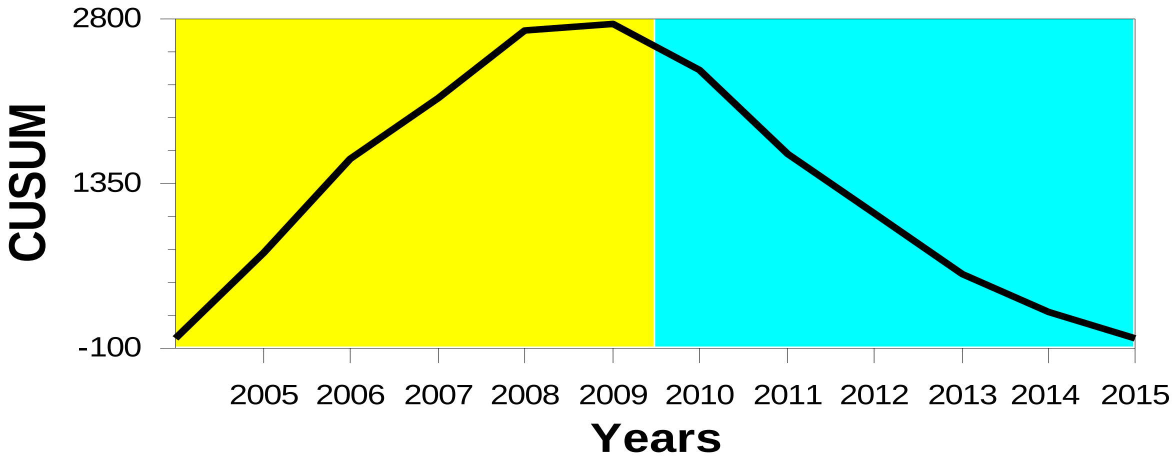


Figure 4.4 KwaZulu-Natal murder data: CUSUM plot (2005 – 2015)

Table 4.3 presents the bootstrap results of KZN murders. The confidence level of the change that occurred is 98%. The table shows level 1 change. According to Taylor (2018) the level 1 of change is the first change that is clearly visible and get detected on the first pass through the data. The level 2 changes are detected on the second pass through data.

Table 4.3 Significant changes for murder in KwaZulu-Natal

Years	Confidence Interval	Conf. Level	From	To	Level
2011	(2011, 2011)	98%	3324	3626	1

Figure 4.5 depicts the graphical representation of the CPA relating to the Gauteng data. All points appear to be within the control lines, and that indicates that there has been no drastic change in murder cases over a timed period. However, the background shift shows a hidden change that occurred.

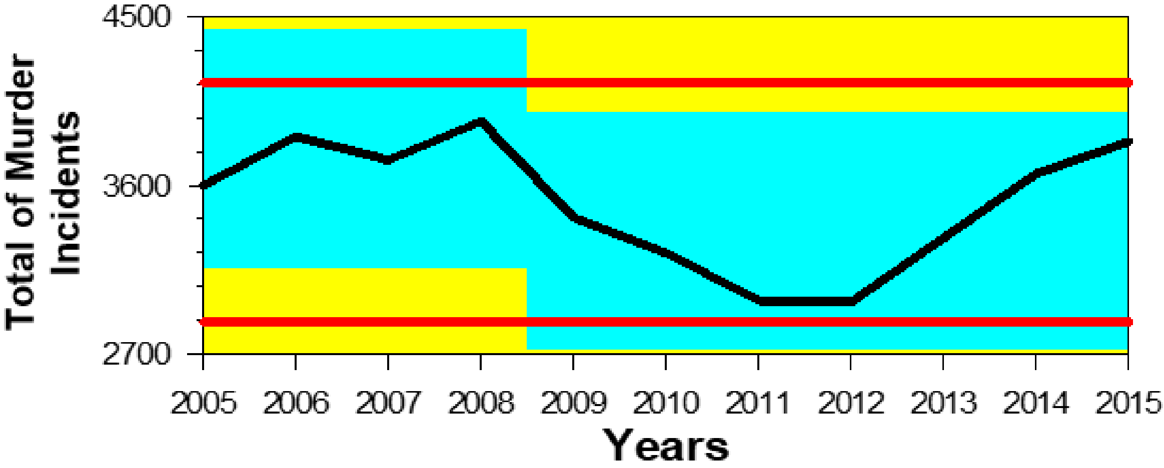


Figure 4.5 Gauteng murder data: background changes (2005 – 2015)

Figure 4.6 presents a CUSUM chart of analysis for the Gauteng data. A significant change was observed in 2011 and the background shift was observed in 2009. These two changes took place in different years and the shift indicated that the change was coming, hence CPA is very significant in providing early notice of a change. The chart was descending after the abrupt shift occurred but after the actual change took place the chart changed and started to ascend.

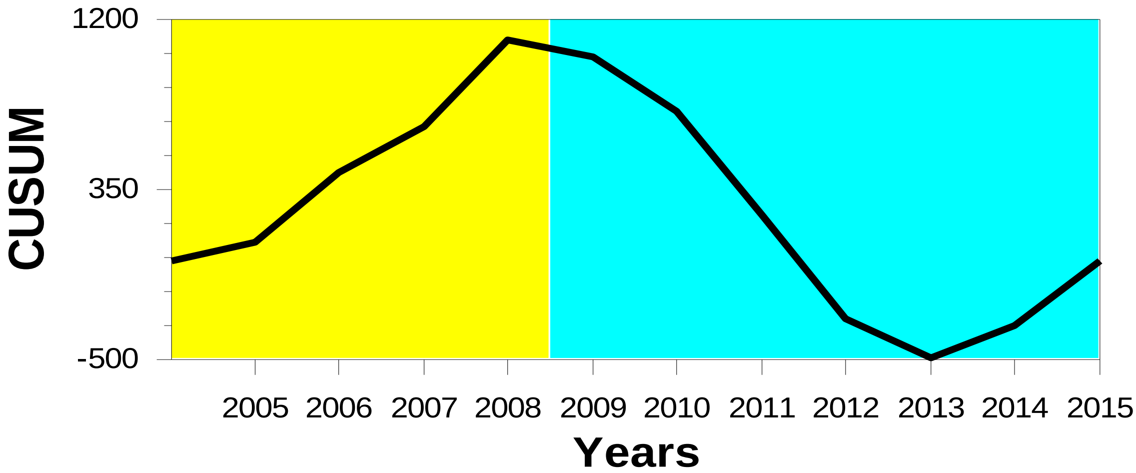


Figure 4.6 Gauteng murder data: CUSUM plot (2005 – 2015)

To gain insight into the number and timing of changes that occurred in Gauteng, bootstrap was utilised and Table 4.4 was generated. The analysis detected a level 1 change in 2011. The confidence level of the change that occurred was 97%. The table shows a level 1 change. Any number can occur depending on the number of changes that took place in the CUSUM chart analysis.

Table 4.4 Significant changes for murder in Gauteng

Years	Confidence Interval	Conf. level	From	To	Level
2011	(2009, 2012)	97%	3401.2	3012.31	1

Figure 4.7 presents the CPA results relating to the Western Cape. All points seem to be within the control limits but there is an abrupt shift of background from 2012 to 2013, this indicates the minor change that occurred.

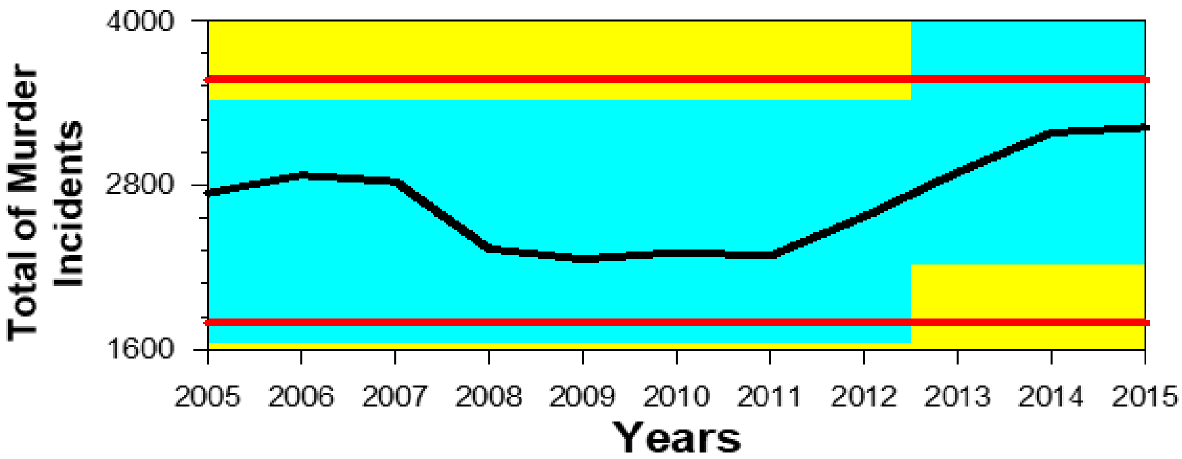


Figure 4.7 Western Cape murder data: background changes (2005 – 2015)

Figure 4.8 presents the results of the CUSUM chart of analysis for the Western Cape. Only one major significant change was detected, which occurred in 2012. The chances of the confidence level being above 95% were very high because this sudden change was recently recognised by the shift and the change occurred in the same year.

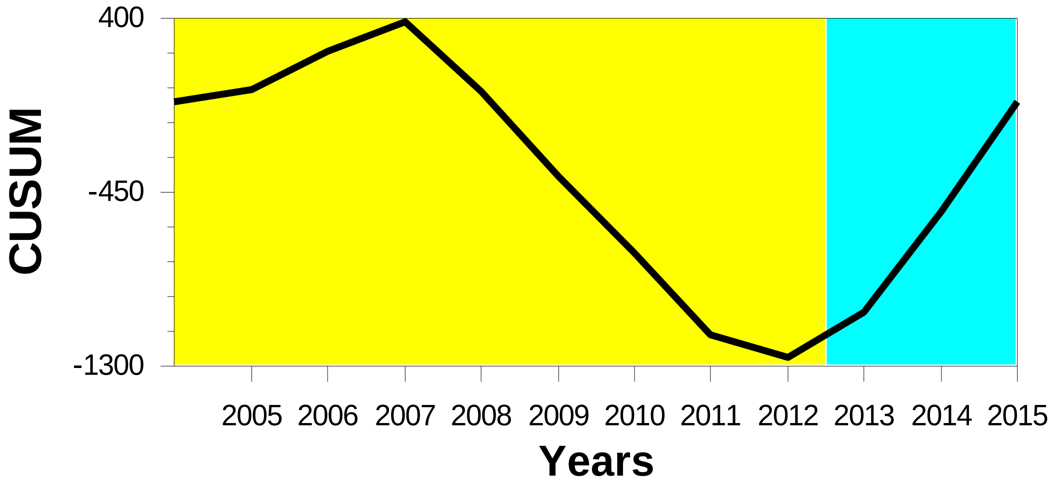


Figure 4.8 Western Cape murder data: CUSUM plot (2005 – 2015)

To get insight into the number and timing of changes that occurred in the Western Cape, bootstrap was utilised, as shown in Table 4.5. The analysis detected a level 1 change in 2012. The confidence level of the change that occurred was 97%. The table shows the level 1 change. The level 1 change is indicative of a strong change identified. As previously mentioned, any number can occur depending on the number of changes that took place in the CUSUM chart analysis. The murder trend in this province had a sudden increase over the years, which was detected picked up by the CPA algorithm.

Table 4.5 Significant changes for murder in Western Cape

Years	Confidence Interval	Conf. Level	From	To	Level
2012	(2011, 2013)	97%	2070	3106.2	1

Figure 4.9 depicts a graphical representation of CPA with control limits and background changes for the Eastern Cape. It further shows that all points appear within the control lines, but there a huge shift appeared in 2008. The chart shows that around 2005 and 2006 the points appear outside the control lines, and the huge change occurred in 2014.

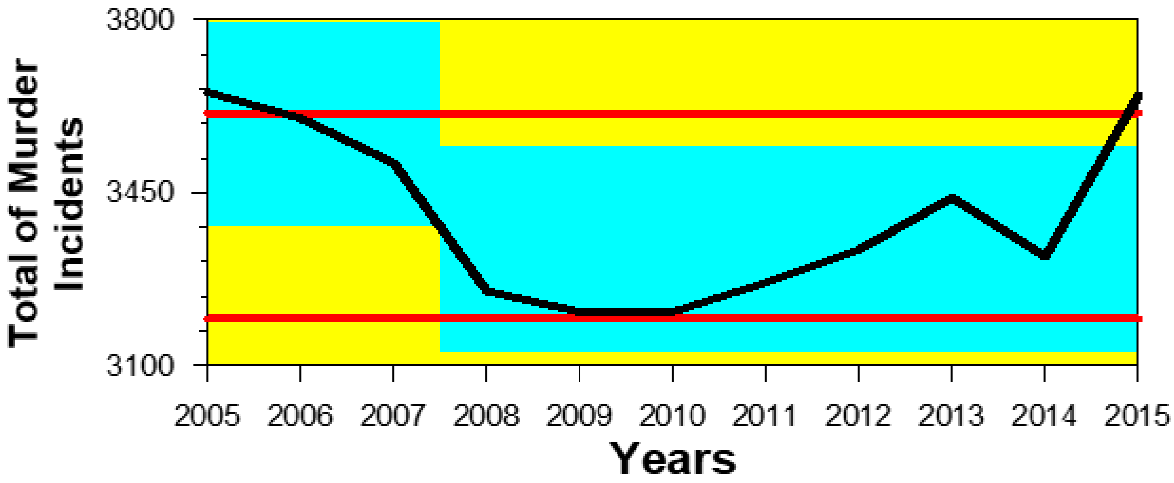


Figure 4.9 Eastern Cape murder data: background changes (2005 – 2015)

Figure 4.10 depicts the CUSUM chart of the Eastern Cape. The chart indicates that from 2007 to 2008 there was a visible change, the graph was ascending until it reached that point where it changed direction and started to descend. Then in 2014 it started to increase again.

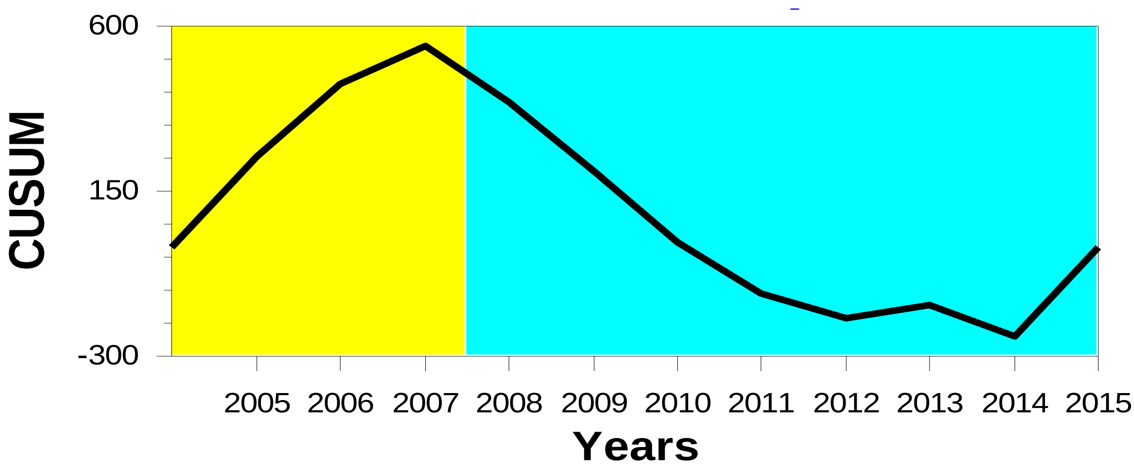


Figure 4.10 Eastern Cape murder data: CUSUM plot (2005 – 2015)

Table 4.6 depicts the bootstrap results of the Eastern Cape. The confidence level of the change that occurred was 94%. The table shows a level 1 change. The level 1 change indicates a strong change that is identified on the first pass through the data. A significant change was detected in 2008.

Table 4.6 Significant changes for murder in Eastern Cape

Years	Confidence Interval	Conf. Level	From	To	Level
2008	(2008, 2012)	94%	3284	3206	1

Figure 4.11 presents the results of the CPA for the Free State. All points appear within the restriction lines but between 2009 and 2010 there is an abrupt shift. This shift is difficult to identify when using other methods such as a control chart. After 2007 the number of murders in Free State started to increase. However, during the period of 2012-2013 the numbers dropped.

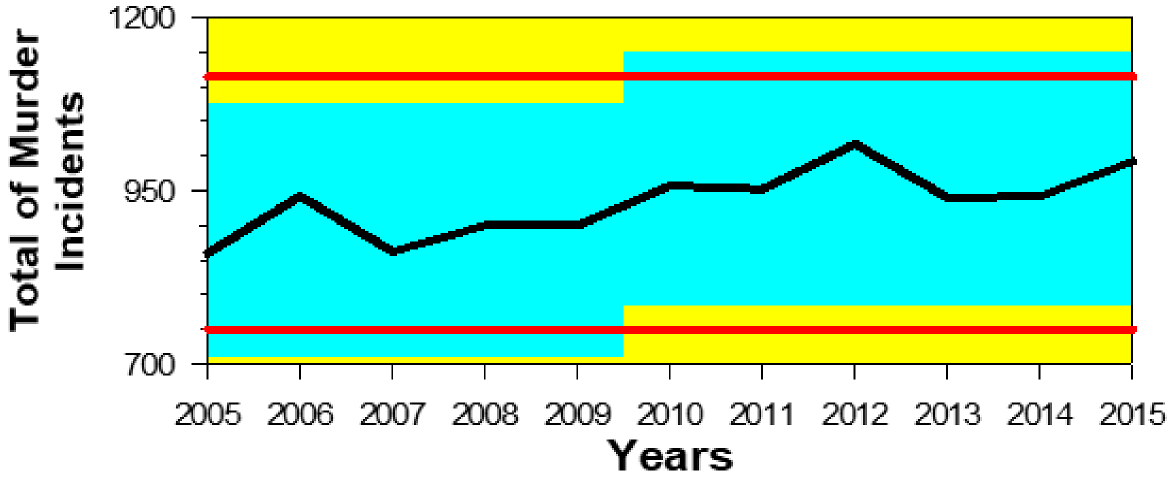


Figure 4.11 Free State murder data: background changes (2005 – 2015)

Figure 4.12 presents the results of a CUSUM chart of analysis for the Free State. Only one major significant change was detected. The chart depicts an ascending trend, however, before 2009 the graph was descending, also from 2012-2013, the graph was descending.

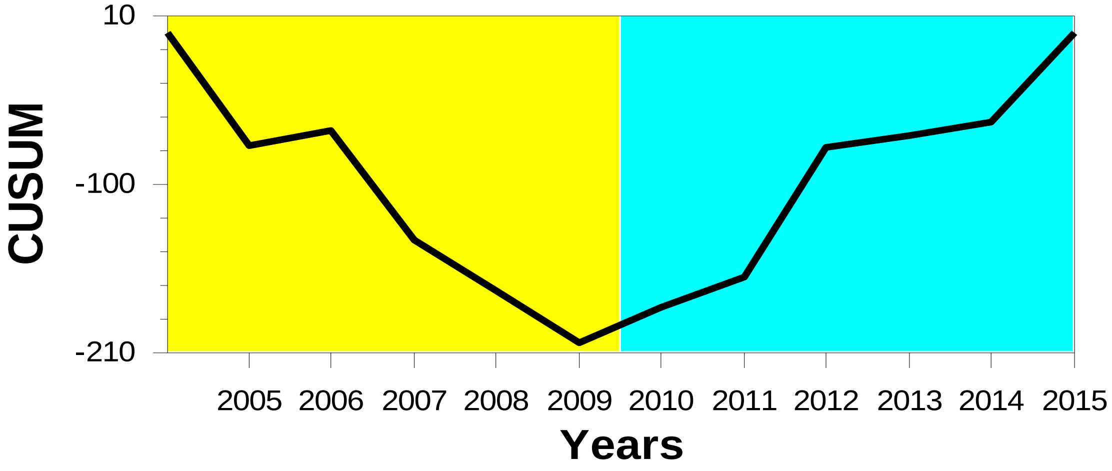


Figure 4.12 Free State murder data: CUSUM plot (2005 – 2015)

Table 4.7 presents the results of CPA on murder for the Free State. A level 1 change was discovered in 2009. The change occurred in 2009 with 92% confidence that the change did occur. The murder trend increased from 891.3 to 961.17. This indicates the confidence of change that occurred.

Table 4.7 Significant changes for murder in Free State

Years	Confidence Interval	Conf. Level	From	To	Level
2009	(2009, 2010)	92%	891.3	961.17	1

Figure 4.13 shows a graphical presentation of the results of the CPA with background changes and control limits relating to Mpumalanga. The CPA depicts the change in the blue region that occurred between 2009 and 2010, then one point below the lower limit which is seen in 2012.

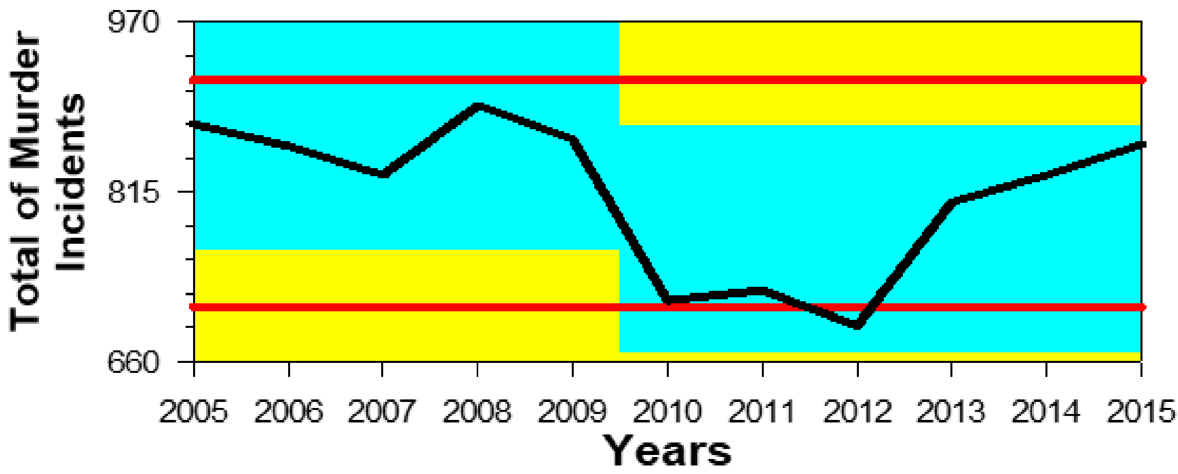


Figure 4.13 Mpumalanga murder data: background changes (2005 – 2015)

In Figure 4.14 The presence of the blue region confirms the change that occurred in the CUSUM chart. The graph depicts a downward trend of murder in Mpumalanga in 2009, this figure provides confirmation, along with Figure 4.13 that in 2009 there was an occurrence of shift.

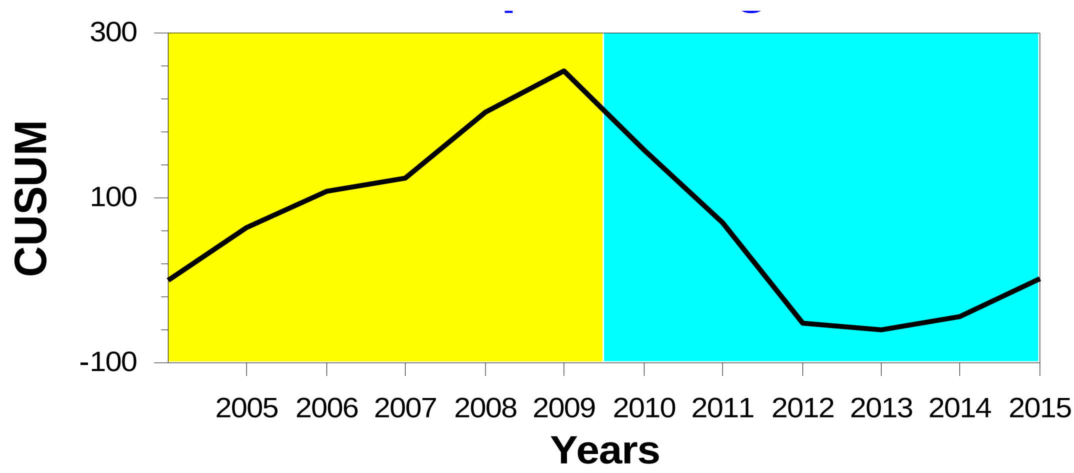


Figure 4.14 Mpumalanga murder data: CUSUM plot (2005 – 2015)

Table 4.8 depict the bootstrap results for Mpumalanga. The confidence level of the change that occurred was 92%. The table shows a level 1 change. A significant change was detected between 2011 and 2012.

Table 4.8 Significant changes for murder in Mpumalanga

Years	Confidence Interval	Conf. Level	From	To	Level
2011	(2011, 2012)	92%	711	686	1

Figure 4.15 depicts the CPA results. All crime trends are within the restriction points, which indicates the non-existence of drastic change of the incidence of murder in the Northern Cape.

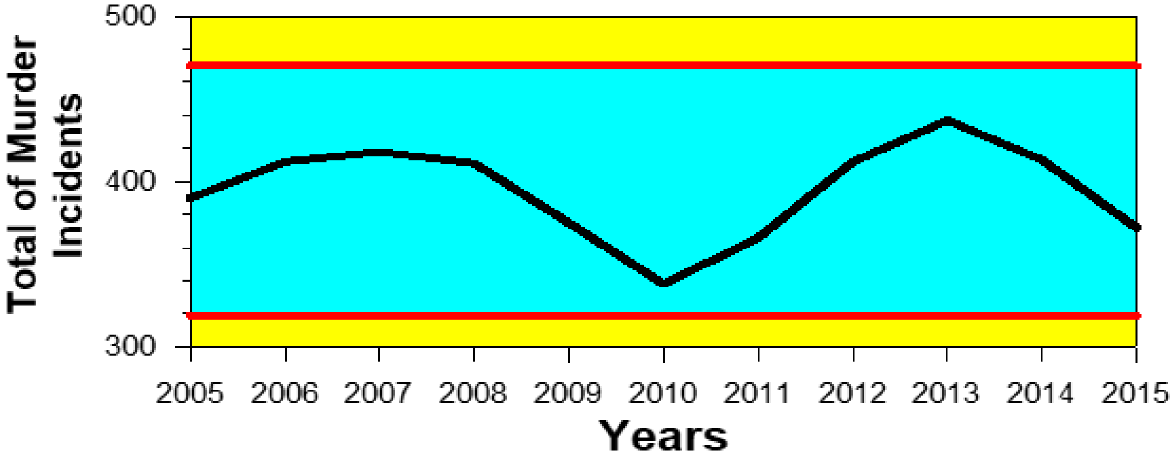


Figure 4.15 Northern Cape murder data: background changes (2005 – 2015)

Figure 4.16 presents the results of a CUSUM chart of analysis for Northern Cape. Not even one significant change was detected. The chart depicts the absence of the blue region.

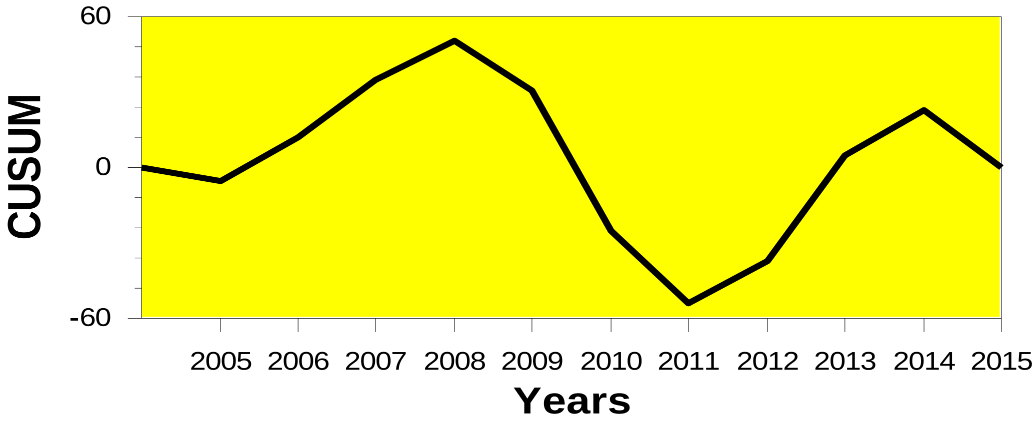


Figure 4.16 Northern Cape murder data: CUSUM plot (2005 – 2015)

Figure 4.17 depicts the CPA results for Limpopo. The absence of the blue region indicates the lack of change. All crime trends appear within the restriction points.

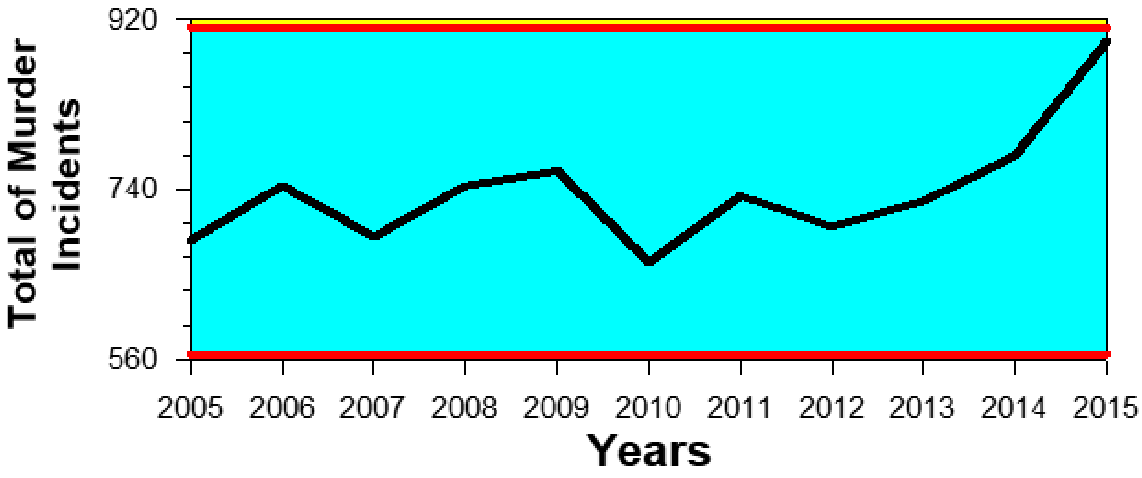


Figure 4.17 Limpopo murder data: background changes (2005 – 2015)

The bootstrap results provide a similar outcome to that of Limpopo Province, which also did not have any changes and all points appeared within the control limits.

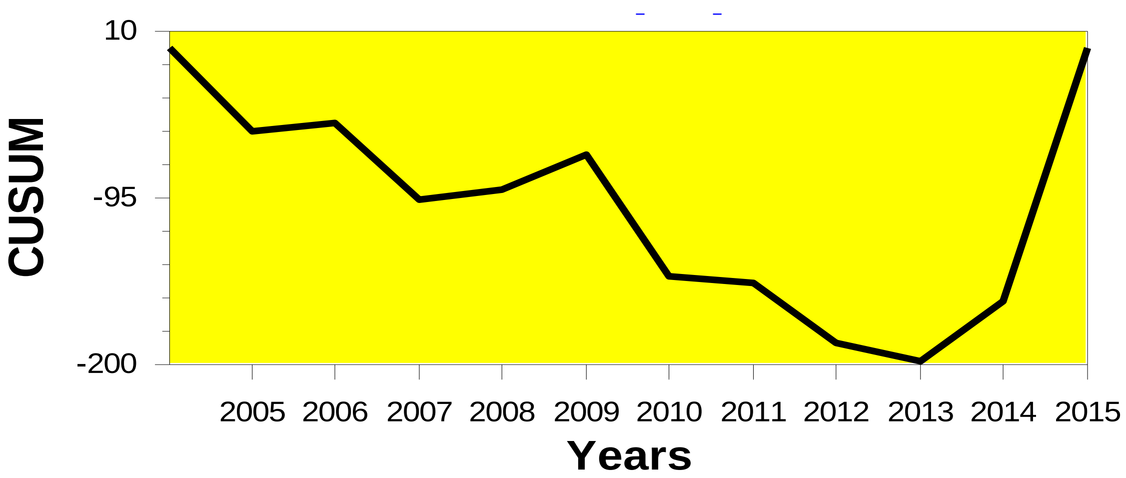


Figure 4.18 Limpopo murder data: CUSUM plot (2005 – 2015)

Figure 4.19 presents the CPA result for North West Province All points in Figure 4.19 appear within the control lines and there is no shift within the process of the analysis.

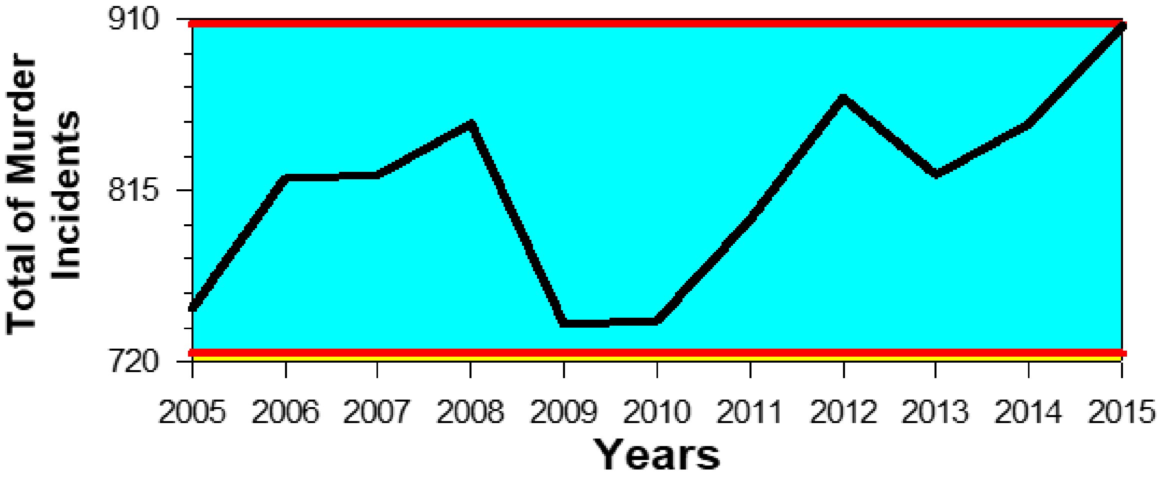


Figure 4.19 North West murder data: background changes (2005 – 2015)

Figure 4.20 presents the results of the CUSUM analysis of the data for North West. Not even one significant change was detected in this province. The chart depicts the absence of the blue region and Figure 4.20 showed no background changes.

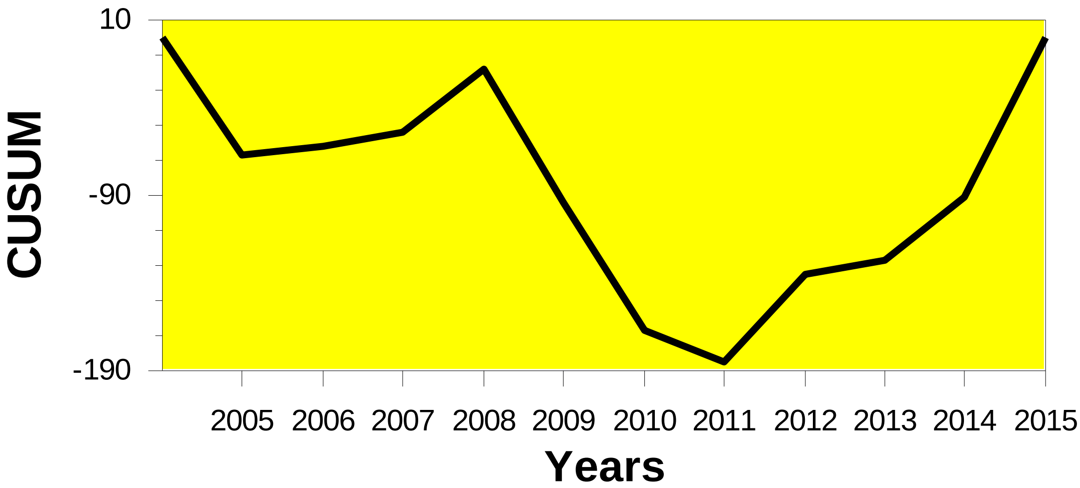


Figure 4.20 North West murder data: CUSUM plot (2005 – 2015)

4.4 Murder Crime Results Summary

The results obtained in this study agree with the discussion in Chapter Two, that murder is rising. This can be seen in the charts above where there is a change of the blue region, this indicates the change in the number of murders. The CPA results depicted that murder in most provinces is still an issue and Figure 2.5 shows that all murder trends of provinces increased as time went by. The bootstrap analysis tables depicted the level 1 change, which indicates the existence of one major significant change that caused an increase in murder. The bootstrap analysis also emphasised the persistence of the changes that occurred in 2011 and 2012 in most provinces. A Mail and Guardian staff reporter (2013) mentioned that murder crimes perpetrated between April 2011 and March 2012 in Eastern Cape were higher than the murders committed in Gauteng, even though Gauteng is known for crime and violence. Figure 2.5 also shows that between 2011 and 2012, murder cases in the Eastern Cape were higher than in Gauteng Province. These findings could contribute to future studies, by using these results to determine why in 2011 and 2012 the Eastern Cape experienced an extreme increase in murder.

Table 4.9 below shows the occurrences of abrupt shifts in murder crimes for the nine provinces of South Africa.

Table 4.9 Occurrence of potential abrupt shifts for murder per province (2005 – 2015)

Province	**Year Range**	**Shift**
KwaZulu-Natal	2011	Upward
Mpumalanga	2011–2012	Downward
Northern Cape	No significant changes	None
Limpopo	No significant changes	None
North West	No significant changes	None
Gauteng	2009–2012	Downward
Eastern Cape	2008–2012	Downward
Free State	2009–2010	Upward
Western Cape	2011–2013	Upward

Using the CPA is essential when dealing with the huge volumes of data, in this case, a murder dataset which spanned an 11-year period. The table above depicts CPA has identified that out of nine provinces, three provinces have experienced a high number of murder crimes namely KwaZulu-Natal, Western Cape and Free State. However, provinces such as Gauteng and Eastern Cape were also starting to gain momentum of increasing the number of murder crimes. For instance, Eastern Cape murder crime started to increase in 2014, and Gauteng murder crimes started to increase in 2013 just after the end of the dataset.

4.5 Burglary Crime Results

The study used a South African crime statistics dataset to perform the CPA on data relating to burglary. The data pre-processing was done using Python Jupyter (Notebook 3). Figure 2.2 presents burglary trends per provinces that are derived from the official dataset of South African crime.

Figure 2.5 shows that 2015 had the highest count of burglary with Gauteng having the highest number of burglaries, followed by KwaZulu-Natal. The provinces of Eastern Cape and Western Cape also showed a high number of burglaries during the same reporting period. The study experimented with detecting the significant change(s) and identifying the time at which the change(s) occurred.

Figure 4.21 presents the change point analysis results for Western Cape, and not all points appear within the control lines. It is observed that the graph appears below and above the restriction lines in 2005 and 2013. The presence of the background change indicates the significant change that occurred during the timed period.

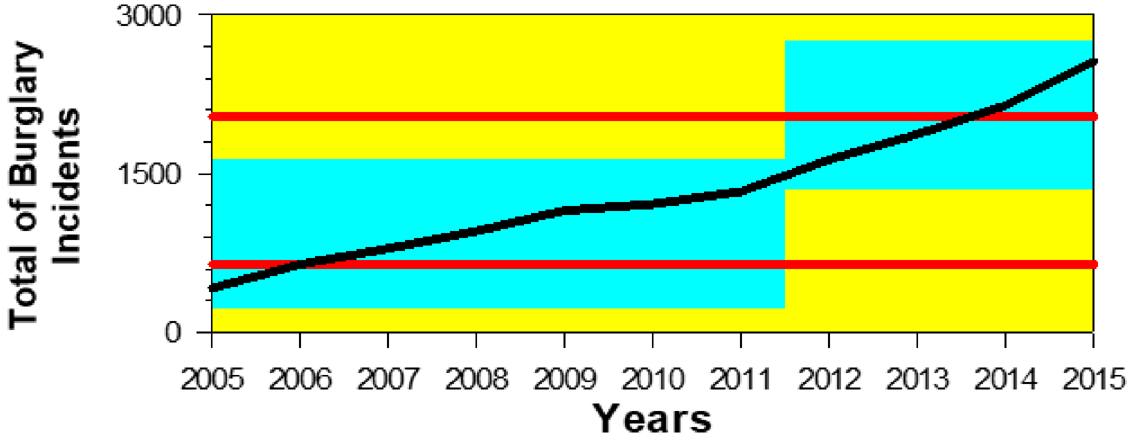


Figure 4.21 Western Cape burglary data: background changes (2005 – 2015)

Figure 4.22 depicts a graphical presentation of the CUSUM analysis relating to the Western Cape data. The CUSUM chart was used to illustrate the cumulative sum of the data points, where blue regions highlight the existence of change. The graph depicts an ascending trend. The significant change occurred from 2011 to 2013.

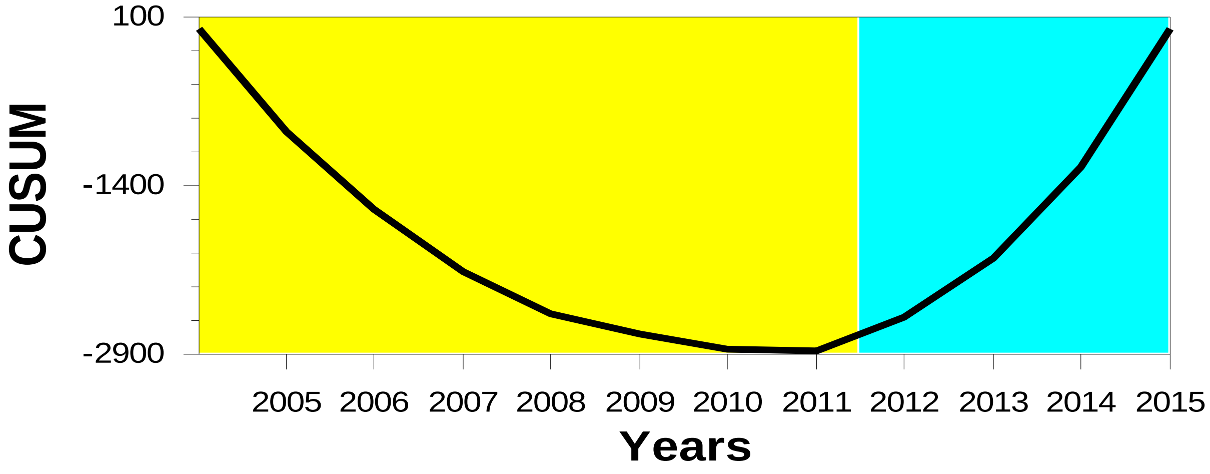


Figure 4.22 Western Cape burglary data: CUSUM plot (2005 – 2015)

Table 4.10 depicts the bootstrap results of the Western Cape. The confidence level of the change that occurred is 97%. The table shows a level 1 change. The level 1 change indicates the strong change identified on the first pass of the analysis. A significant change was detected in 2012.

Table 4.10 Significant changes for burglary in Western Cape

Years	Confidence Interval	Conf. Level	From	To	Level
2012	(2013,2013)	97%	2064.2	3000	1

Figure 4.23 presents the change point analysis results for KZN. The abrupt shift indicates the hidden sudden change that occurred. It is possible that this shift would have been missed by the control chart since it all happened within the control limits, and the control chart updates when the next point is collected. Unlike the control chart, the CPA updates after all points have been collected, hence detecting an abrupt shift is not difficult because the detection is done based on the existing data that have been collected.

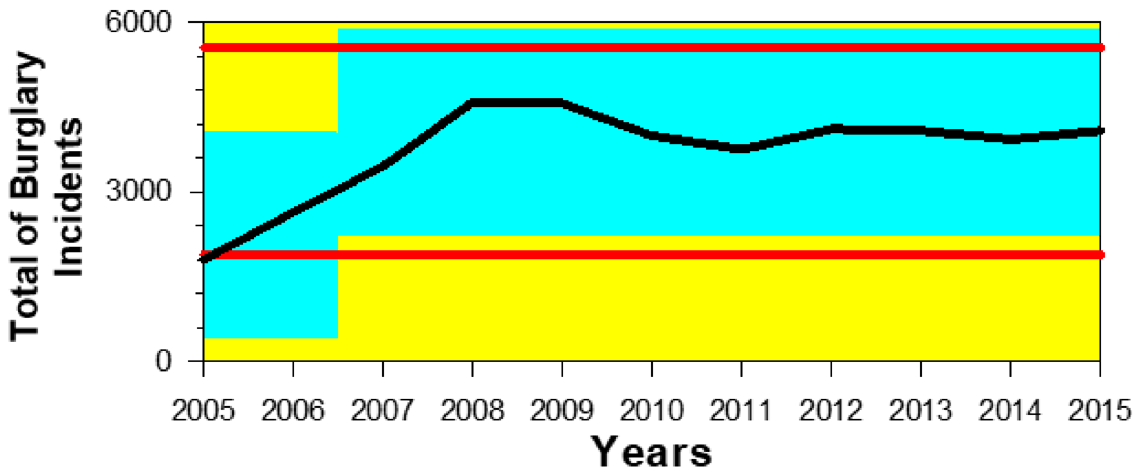


Figure 4.23 KwaZulu-Natal burglary data: background changes (2005 – 2015)

Figure 4.24 presents the CUSUM analysis relating to KZN's burglary data. The chart is showing an ascending trend from 2007, and the existence of the blue region indicates the significant occurrence of the change. To have an insight into the changes pinpointed in this case, bootstrap was performed to give a clear understanding. Table 4.11 depicts the results of the bootstrap analysis on burglary trends of KwaZulu-Natal.

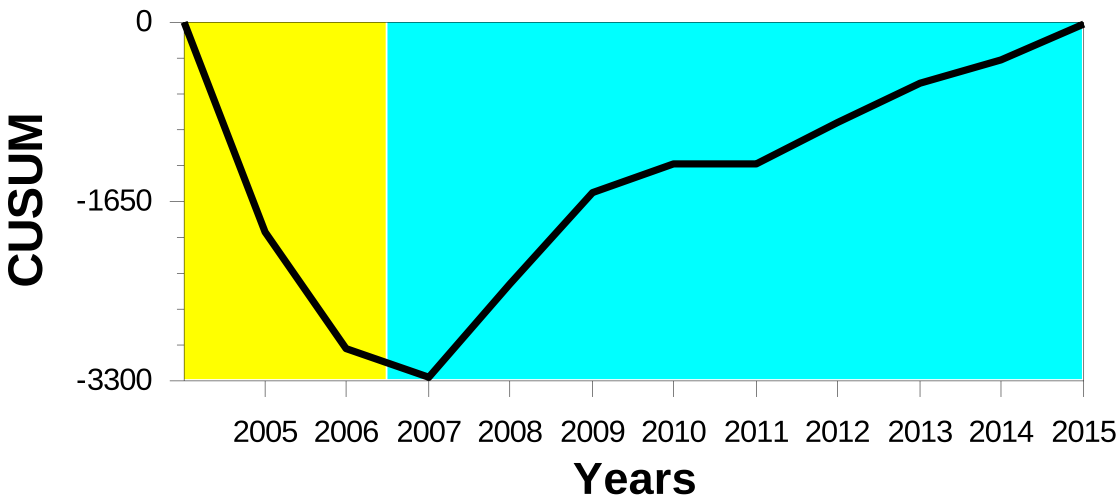


Figure 4.24 KwaZulu-Natal burglary data: CUSUM plot (2005 – 2015)

Table 4.11 depicts the results of the CPA on burglary in KZN. The analysis was 92% confident that the change occurred during the period. The bootstrap analysis shows that the abrupt shift that occurred in 2007 was detected. The abrupt shift of the background started from 2 244 to

4 074.1. The number of levels depends on the changes taking place; in this case only one change occurred.

Table 4.11 Significant changes for burglary in KwaZulu-Natal

Years	Confidence Interval	Conf. Level	From	To	Level
2007	(2007,2007)	92%	2244	4074.1	1

Figure 4.25 depicts the results of the province of Eastern Cape with a total sum of 15 568 burglaries over the 11-year period. These results of the CPA with the control limits and background region indicated abrupt shifts. All points were not within the control limits. From 2005 to 2007 the points appear below the control limit line and from 2013 to 2013 the points appear above the control limit line. This means there has been a constant increase of burglary crime in Eastern Cape.

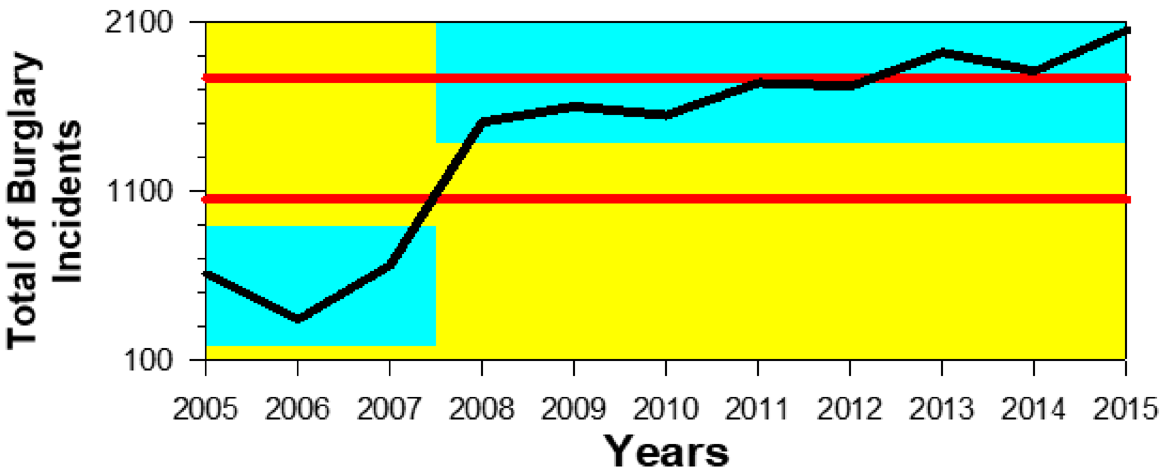


Figure 4.25 Eastern Cape burglary data: background changes (2005 – 2015)

Figure 4.26 presents a graphical representation of CUSUM analysis. This chart illustrates that there was a change in 2007 and the chart is ascending from that point. The change is also visibly depicted by the change in the background.

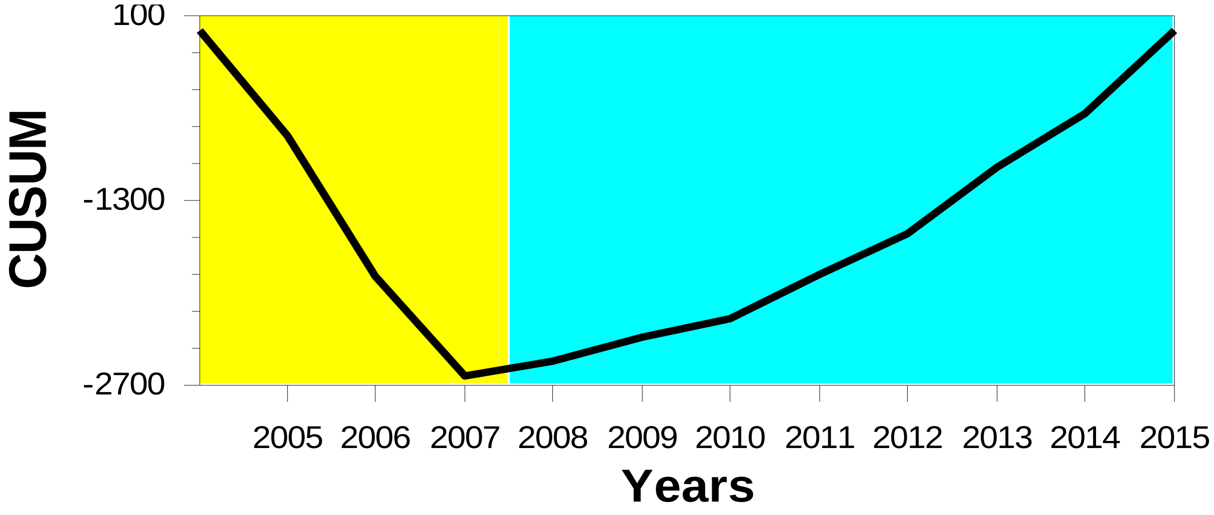


Figure 4.26 Eastern Cape burglary data: CUSUM plot (2005 – 2015)

Table 4.12 depicts the results of the CPA on burglary data for Eastern Cape. A level 1 change was detected in 2012. This is indicative of a strong change identified during the process of detecting the changes. The confidence level that the change occurred was 98%.

Table 4.12 Significant changes for burglary in Eastern Cape

Years	Confidence Interval	Conf. Level	From	To	Level
2012	(2012,2012)	98%	1625.2	192.0	1

Figure 4.27 depicts the graphical presentation of results of Free State Province with a high sum (5 724) of reported burglaries. The points appear below the control limit between 2005 and 2007. One background change occurred between 2007 and 2008.

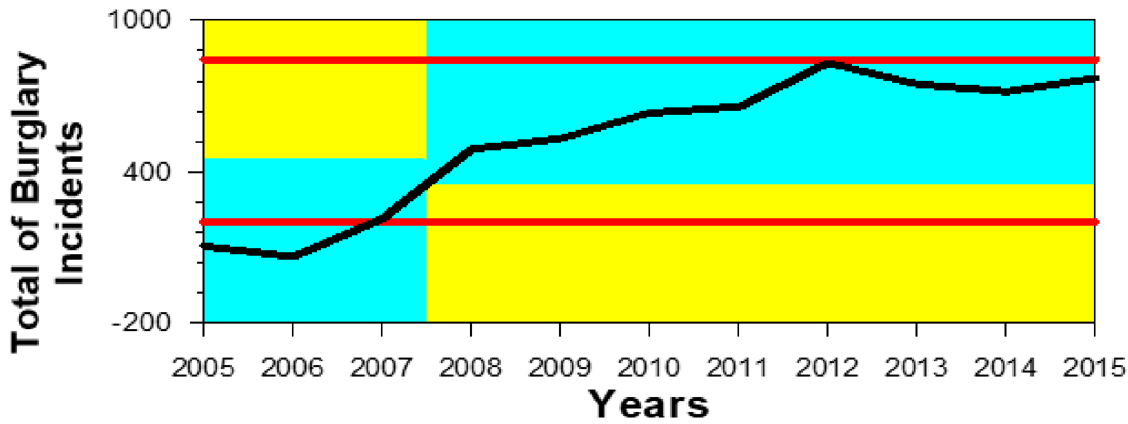


Figure 4.27 Free State burglary data: background changes (2005 – 2015)

Figure 4.28 presents the results of the CUSUM analysis relating to the Free State data, with one significant change detected between 2007 and 2008. The existence of the blue region emphasises the significant occurrence of the change.

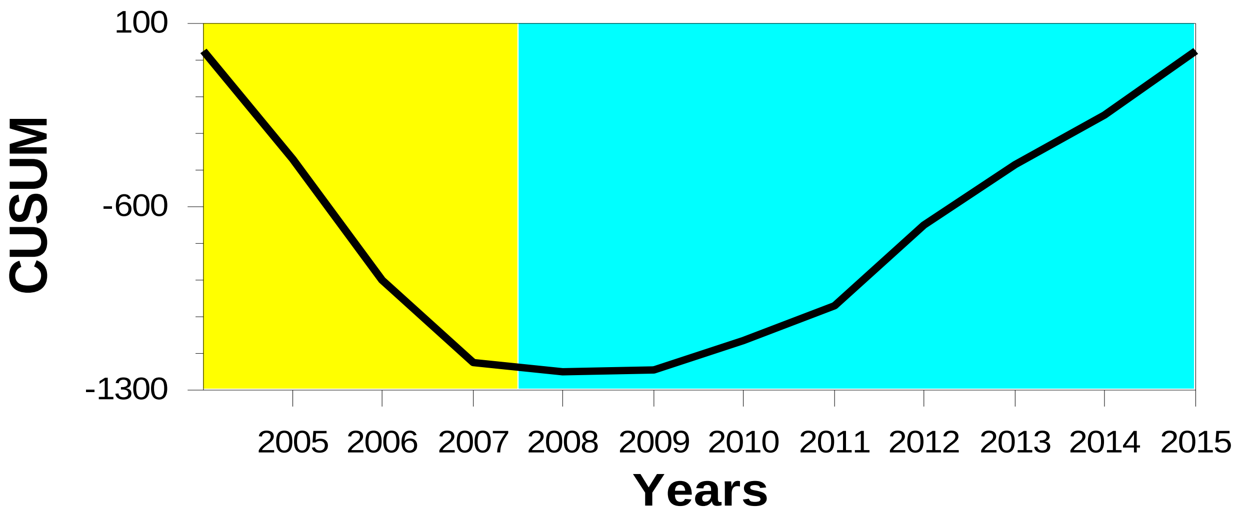


Figure 4.28 Free State burglary data: CUSUM plot (2005 – 2015)

Table 4.13 depicts the results of the CPA of the burglary data for Free State Province. A level 1 change was discovered in 2009. The change occurred in 2009, with 97% confidence that the change occurred. The burglary trend increased from 420.1 to 442.3 incidents. Figure 4.27 depicts the graph is ascending, meaning burglary crime in Free State has been on the rise; the table confirms that burglary crime has increased.

Table 4.13 Significant changes for burglary in Free State

Years	Confidence Interval	Conf. Level	From	To	Level
2009	(2009,2009)	97%	420.1	442.3	1

Figure 4.29 presents the results of the North West Province burglary data. The blue region shows an abrupt shift in 2006 and 2007. This shift is also supported by the blue region that shows a shift.

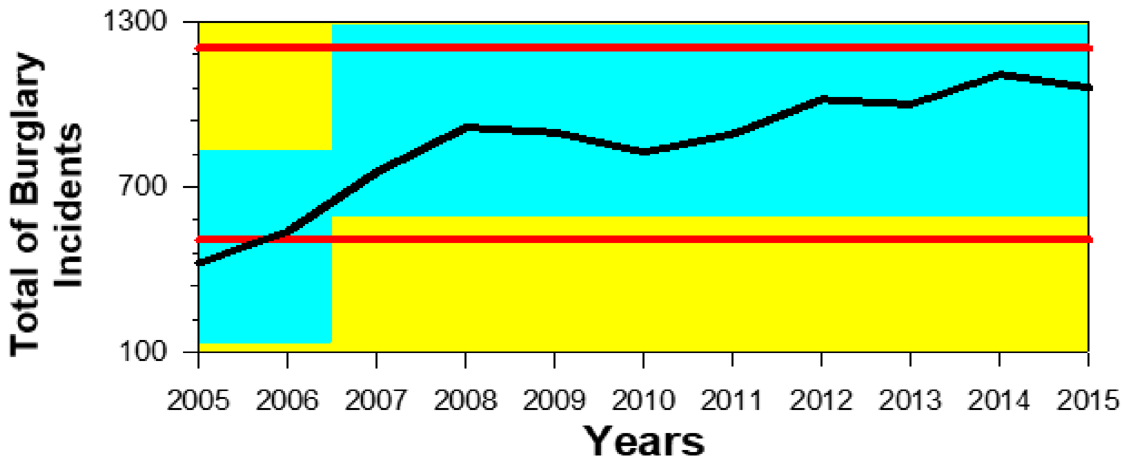


Figure 4.29 North West burglary data: background changes (2005 – 2015)

Figure 4.30, the CUSUM chart, depicts only one change that took place in 2007. The graph is in the descending order until 2007, after 2007 the number of burglaries increased and it is shown by the appearance of the blue region and the graph is in an ascending order.

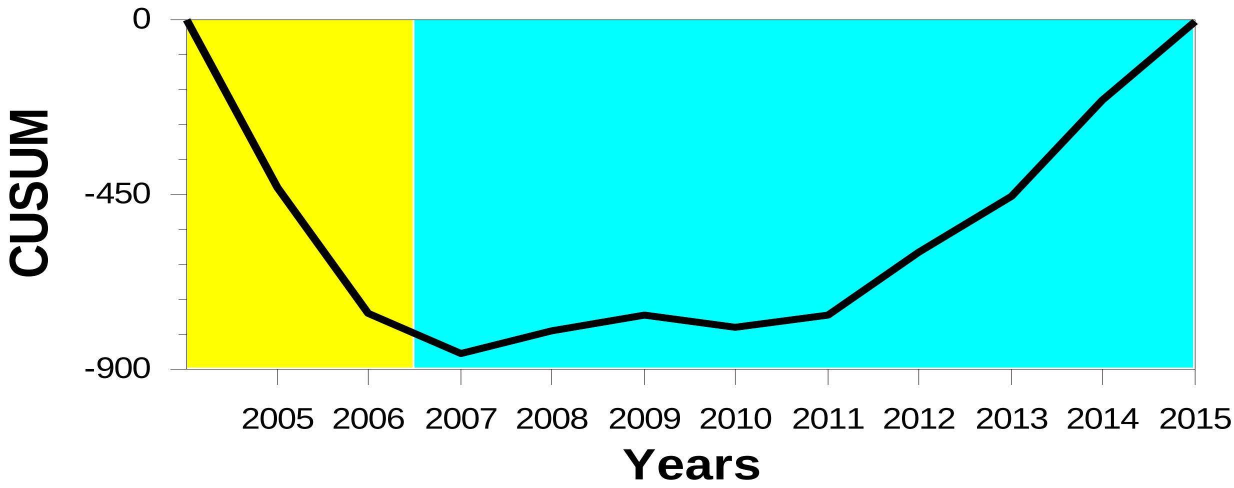


Figure 4.30 North West burglary data: CUSUM plot (2005 – 2015)

Table 4.14 depicts the results of the CPA on burglary data for the North West Province. A level 1 change was noted in 2007. The confidence level relating to the change that took place in 2007 was 90%. Figure 4.30 depicted that the graph is ascending, meaning burglary crime in North West has been on the rise and the table confirms that burglary crime increased over some time.

Table 4.14 Significant changes for burglary in North West

Years	Confidence Interval	Conf. Level	From	To	Level
2007	(2006,2007)	90%	650.2	900.1	1

Figure 4.31 depicts a graphical representation of CPA with control limits and background changes for Limpopo Province. The graph shows a constant increase in burglary over a timed period, and the line appears both below and above the control limits. There background change is in existence.

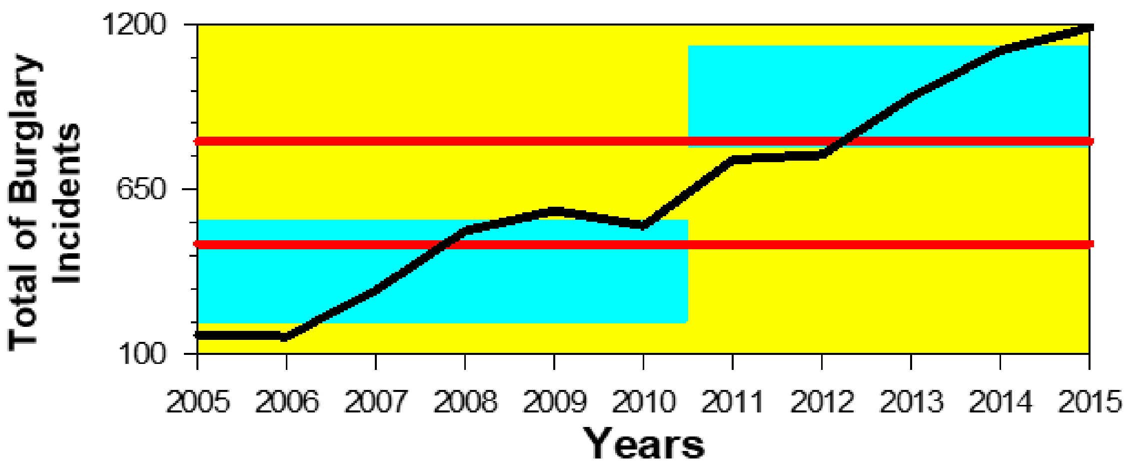


Figure 4.31 Limpopo burglary data: background changes (2005 – 2015)

The CUSUM chart in Figure 4.32 shows an ascending trend. The graph shows a curve facing up, meaning there was a turning point in Limpopo and burglary started to increase suddenly. The blue region emphasises the significant occurrence of the change that occurred between 2010 to 2012.

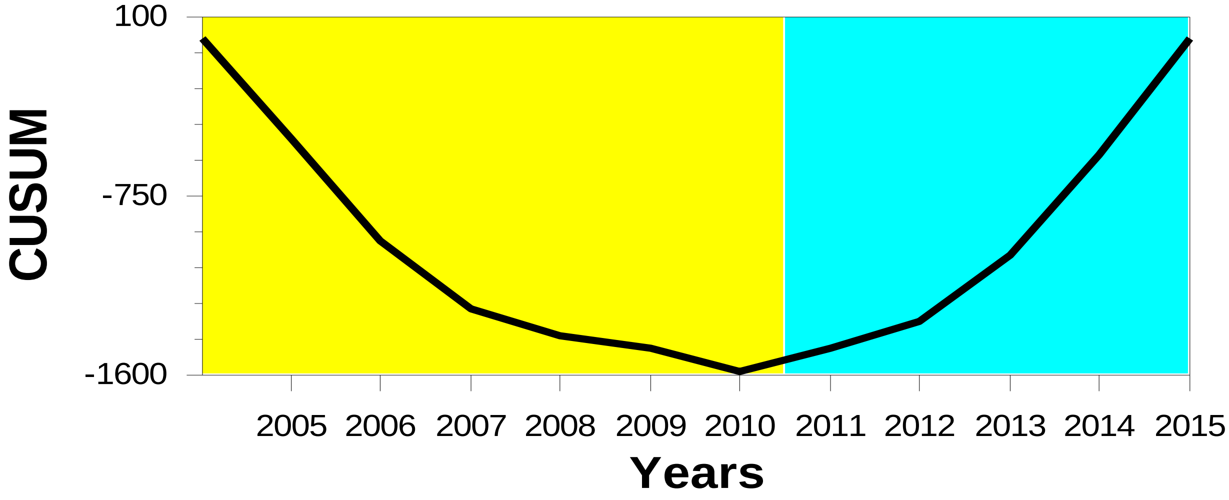


Figure 4.32 Limpopo burglary data: CUSUM plot (2005 – 2015)

Table 4.15 depicts the results of the CPA on burglary data for Limpopo. A level 1 change was identified in 2010. The change occurred between 2010 and 2012, with 98% confidence that the change occurred. Burglary incidences increased from 467 to 957.6. Figure 4.32 depicts that the graph is ascending from 2010, meaning burglary crime in Limpopo is on the rise; the table confirms that burglary crime also increased.

Table 4.15 Significant changes for burglary in Limpopo

Years	Confidence Interval	Conf. Level	From	To	Level
2010	(2010,2012)	98%	467	957.6	1

Figure 4.33 presents the results of the CPA for Gauteng Province. The blue region indicates the change points that have occurred throughout the years monitored while the red line represents the upper and lower limits. All points are within the control limits but from 2009 and 2012 there was a huge change within the control limits chart.

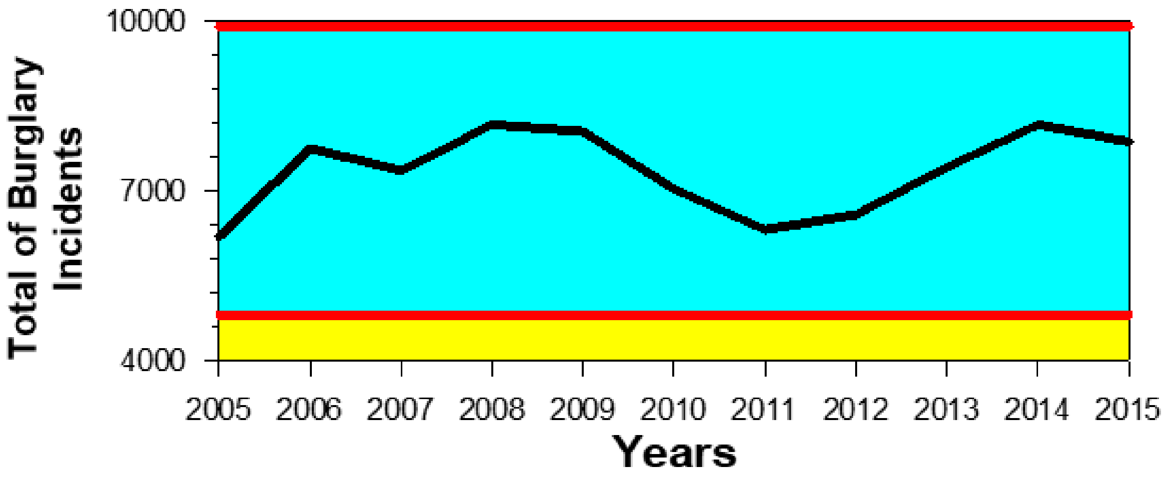


Figure 4.33 Gauteng burglary data: background changes (2005 – 2015)

In Figure 4.34, the CUSUM chart was used to illustrate the cumulative sum of the data points, where blue regions in the CUSUM chart represent the existence of a change point. All points are within the control limits.

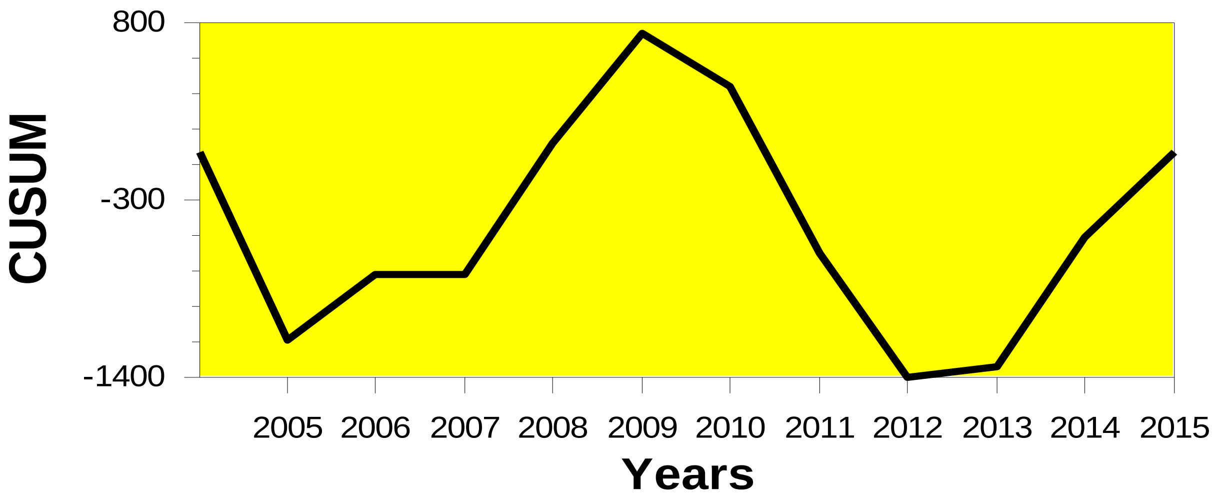


Figure 4.34 Gauteng burglary data: CUSUM plot (2005 – 2015)

Figure 4.35 depicts the CPA results with control limits and background region. All points appear within the control lines, and there was no shift seen in the background.

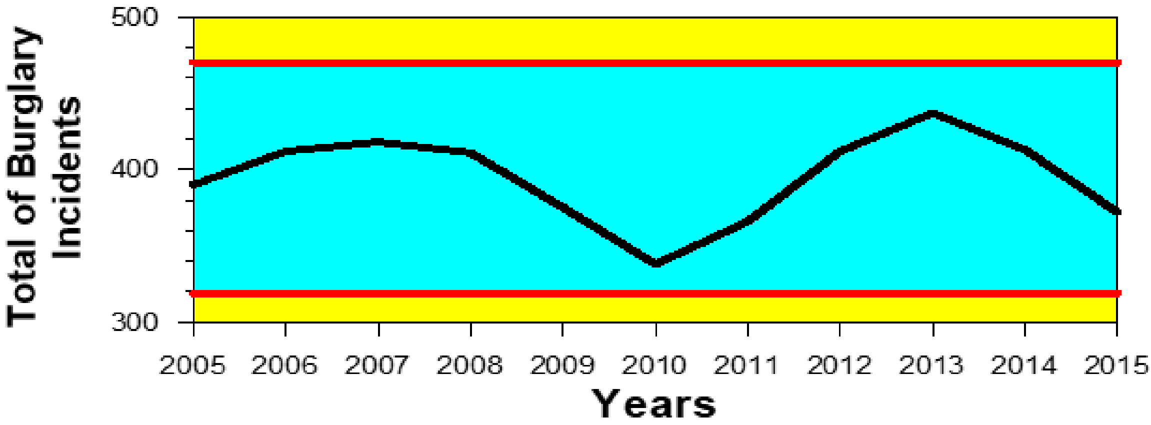


Figure 4.35 Northern Cape burglary data: background changes

In Figure 4.36 the CUSUM chart depicts the non-existence of the blue region and significant changes. Burglary trends have been persistent, and no major change or shift took place in the province of Northern Cape. Figure 2.2 depicts that Northern Cape had the lowest burglary crime over the reporting period.

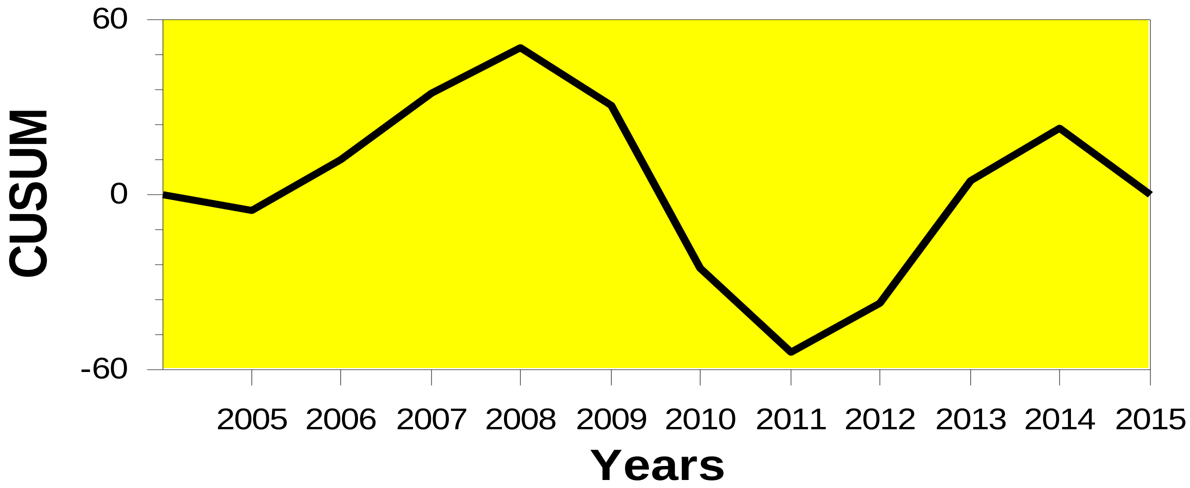


Figure 4.36 Northern Cape burglary data: CUSUM plot (2005 – 2015)

Figure 4.37 shows a graphical presentation of the results of the CPA with background changes and control limits. All points were within the restriction lines, but there was a shift of the background, meaning some change did occur.

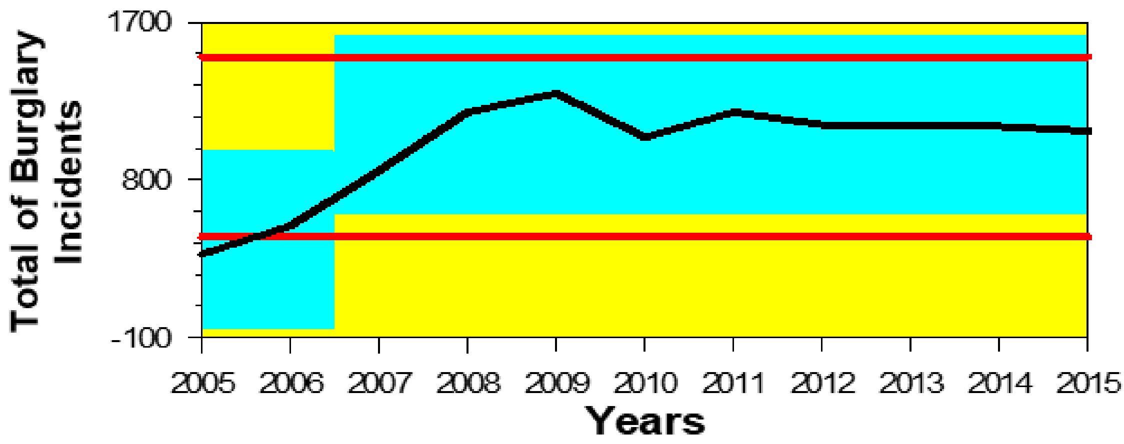


Figure 4.37 Mpumalanga burglary data: background changes (2005 – 2015)

The CUSUM chart in Figure 4.38 shows an ascending trend. The CUSUM chart shows an ascending shift that occurred from 2007 onwards. The CUSUM chart thus shows that in 2007 there was a turning point and the change occurred. This is indicated by the presence of the blue region.

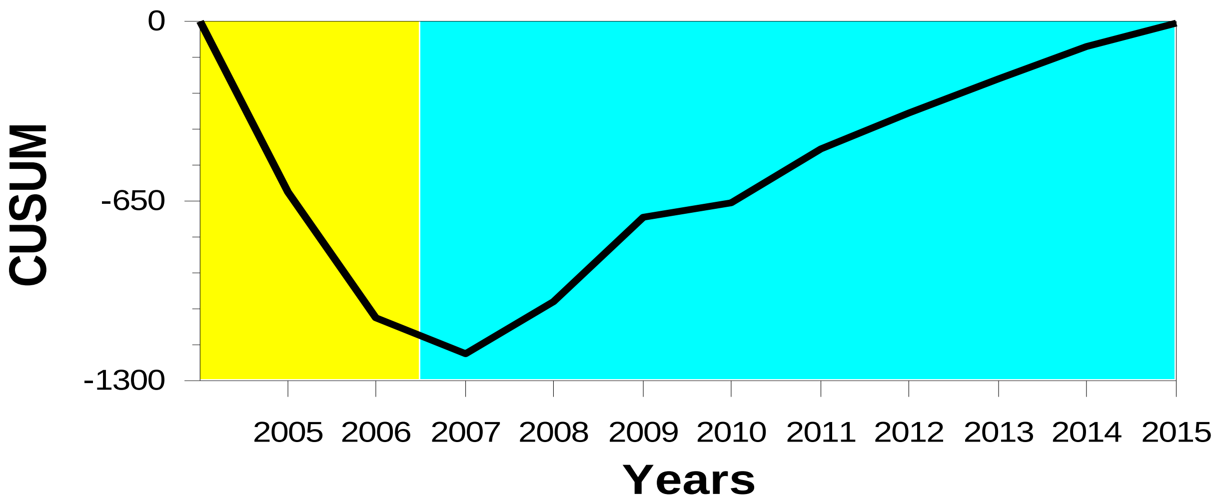


Figure 4.38 Mpumalanga burglary data: CUSUM plot (2005 – 2015)

4.5.1 Burglary Crime Results Summary

The findings obtained in this experiment agree with the study mentioned in Chapter Two about burglary statistics. The results of the CPA depicted that burglary in most provinces was still an issue except in Northern Cape, and Figure 2.5 showed that Northern Cape burglary did not extremely rise as it did in other provinces. The bootstrap analysis tables depicted a level 1 change, which indicates the existence of one major significant change that caused an increase in murder. The bootstrap analysis emphasised the persistence of the changes that occurred in 2007, 2011 and 2012 in most provinces. These findings could contribute to future studies, by using these results to determine what factors caused the changes in 2007, 2011, and 2012 in most provinces. Furthermore, it may be important to understand why the Northern Cape Province was not affected by these factors that led to the surge in burglary crime.

Table 4.16 Occurrence of abrupt shifts in burglary per province (2005 – 2015)

Province	Year Range	Shift
KwaZulu-Natal	2007	Upward
Mpumalanga	2011–2012	Downward
Northern Cape	No significant changes	None
Limpopo	2012	Upward
North West	2007–2011	Upward
Gauteng	No significant changes	None
Eastern Cape	2007–2012	Upward
Free State	2009	Upward
Western Cape	2011	Upward

Using the CPA is essential when dealing with a huge volume of data. This study used a burglary dataset that was derived from the official crime dataset. The table depicts that six out of nine provinces experienced a high level of crime. This means over 50% of crime is still experienced in South Africa, and provinces with greater numbers of people are experiencing it more than others.

Change-point analysis is always preferred over the control chart because it ensures that every point is checked and verified. With regards to the control chart, in this case it would have missed a number of changes and shifts that occurred within the control limit. Using CPA will help to gain an insight into this issue and may thus prevent and reduce the number of burglaries that are taking place in South Africa. This study depicted that burglaries in almost all provinces were increasing, and nothing has been done so far to prevent or measure the effect this has on the society.

4.6 Synopsis of Drug-related Crime Results

Drug-related crime is a common crime in South Africa. As mentioned in the literature review, this type of crime is hard to solve because people in South Africa do not report the culprits to the police. The study of James (2017) mentioned that other crimes are difficult to solve, such as carjacking, theft and drug-related crimes because the South Africans are scared to report the perpetrators to the police.

4.6.1 Synthesis of Control Charts and CUSUM

There are different types of drug-related crimes, hence the numbers are higher than some of the other crimes. The below information refers to Figure 2.3 which shows that the province with the highest drug-related crime levels in South Africa is Western Cape with a total of 732 421, followed by KwaZulu-Natal with a total of 377 745 and Gauteng with a total of 336 063. Figure 2.5 depicts that 2014 had the highest number of drug-related crimes cases with a total of 366 902, followed by 2013 and then 2015 with a total of 260 596 and 259 165 respectively.

The control chart for KwaZulu-Natal appears within the control limits but from 2009 the trends appears above and below the control lines. The CUSUM for KwaZulu-Natal shows one shift-point after 2007. The existence of the blue region indicates the significant change that occurred, and the chart shows an ascending trend. The control chart for Western Cape appears below and above the control limits, below it appears before 2010 and above it appears in 2012. The CUSUM chart shows an ascending trend starting from 2011. The control chart for Gauteng appears within the control limits but from 2009 the trends appears below the lower line. The CUSUM chart shows an ascending trend started from 2010. The blue region for Limpopo depicts no changes and the points appear within the restriction lines. The CUSUM chart depicts no changes and the results of bootstrap depict no significant changes.

The control chart for Northern Cape appears below the control limits in 2010 and above the control limits in 2012. The CUSUM chart depicts an ascending chart from 2012. For Mpumalanga, the control chart shows no changes as the blue region is absent. The CUSUM analysis depicts no changes. The CPA results show that all points of Eastern Cape appeared within the restriction lines. The table of bootstrap analysis concurs with the CPA graphs and suggests that no significant change occurred. The CUSUM analysis of North West presents that no change took place, and there is no turning point on the graph. The chart shows a descending trend. The control chart for Free State appears above the control limits since 2011, and the CUSUM chart shows one shift from 2011, with the graph ascending.

4.6.2 Drug-related Crime Summary

Table 4.17 presents the summarised results of the CPA for drug-related crime. It is observed that the numbers of drug-related crimes are very high. Dolley (2021) reported that a drug dealer in Durban was gunned down in his home by two men. Although most of the community was aware of this man's activities, but none reported him to the police. Drug-related cases can be hard to detect and solve because people protect one other, often out of fear, unlike in crimes such as murder and burglary.

Table 4.17 Occurrence of abrupt shifts in drug-related crime per province (2005 – 2015)

Province	Year Range	Shift
KwaZulu-Natal	2007–2009	Upward
Mpumalanga	No significant changes	None
Northern Cape	2010–2011	Upward
Limpopo	No significant changes	None
North West	No significant changes	None
Gauteng	2009	Upward
Eastern Cape	2012	Upward
Free State	2011	Upward
Western Cape	2010–2012	Upward

Table 4.17 above depicts that only one province in South Africa has shown a decrease in drug-related crime. Three provinces, namely Mpumalanga, Limpopo, and North West depict no significant changes. Six provinces indicate that some provinces are still experiencing an increase in drug-related crimes. This type of crime has been increasing but provinces such as North West are not affected as severely by this crime.

4.7 Synopsis of Carjacking Crime Results

Carjacking is also another serious in crime South Africa, lately carjacking increased dramatically and numerous cases have been reported. Once a vehicle is illegally taken from the owner it is difficult to locate it whereabouts (James, 2017). Thus, most victims of carjacking who report this crime do not have their cars recovered by the SAPS.

4.7.1 Synthesis of Control Chart and CUSUM

Figure 2.4 depicts that the province with the highest carjacking crime levels in South Africa is Gauteng with a total of 73 215, followed by KwaZulu-Natal with a total of 32 420 and then Western Cape with a total of 10 332. Figure 2.5 shows years with the top three counts for carjacking being 2008 with 14 856, 2007 with and 14 134 and 2009 with 13 852.

The control chart for Gauteng appears below and above the control limits. Above it appears before 2007 and below it appears after 2010. The CUSUM chart depicts two shifts in 2007 and 2010, and it shows an ascending trend. The control chart for KwaZulu-Natal appears within the control limits but from 2009 the trend appears above the upper control line and in 2013 it appears below the lower control line. The CUSUM for KwaZulu-Natal shows one shift-point after 2009. The CUSUM chart appears with one change point, and it shows an ascending trend. The control chart for Western Cape appears above the control line in 2014. The CUSUM chart shows an ascending trend starting from 2013. The province of North West presented no significant change, and the absence of the background change which is the blue region confirmed the non-existence of change. The CUSUM analysis showed no change. The bootstrap analysis presented no change of carjacking crime over a timed period.

The control chart for Eastern Cape appears above the control lines in 2014. The CUSUM chart depicts one shift in 2013, and it shows an ascending trend. The control chart for Mpumalanga appears with no changes but there is a shift of the blue region in 2009 within the process. The CUSUM chart depicts one shift in 2010. The control chart for Northern Cape appears below and above the control limits. Above it appears before 2012 and below it appears after 2007. The CUSUM chart depicts one shift in 2012, it moves in an ascending direction. The control chart for Limpopo appears with no changes and no shift of the blue region. The absence of the blue region indicates that the CUSUM chart has no significant changes. The bootstrap analysis for carjacking depicts no significant changes. The control chart for Free State appears with no changes and no shift of the blue region. The CUSUM chart shows no significant changes.

4.7.2 Carjacking Crime Summary

Table 4.18 presents the summarised results of the CPA for carjacking. The table indicates the changes that occurred over a time period, and also indicates if there was no significant change in certain provinces.

Table 4.18 Occurrence of abrupt shifts in carjacking per province (2005 – 2015)

Province	Year Range	Shift
KwaZulu-Natal	2010	Upward
Mpumalanga	2010	Downward
Northern Cape	2012	Upward
Limpopo	No significant changes	None
North West	No significant changes	None
Gauteng	2010–2011	Upward
Eastern Cape	2012–2013	Upward
Free State	No significant changes	None
Western Cape	2013	Upward

The table shows that carjacking has been rising in provinces like Gauteng and KwaZulu-Natal. Carjacking in these provinces is a serious issue and intervention is required to help minimise this crime. The mean changing point is 2010, almost all provinces had a changing point in 2010. That basically indicates that 2010 was the changing point of carjacking crimes in South Africa, that is when the trends changed dramatically.

4.8 Chapter Summary

This chapter has provided insight into past crime data. Throughout this chapter, it has been observed that provinces namely Gauteng, Western Cape, Eastern Cape and KwaZulu-Natal are crime hotspots. These provinces are very large and they consist of rural areas, townships, suburbs and city centres. Hence, the incidence of crime is higher in these provinces than in the remaining six provinces. This chapter has revealed that these provinces need serious intervention in order to minimise the level of crime. Most change points detected were with a confidence level of 95% and above. In the next chapter, the prediction using LR is presented.

CHAPTER FIVE: PREDICTION USING LINEAR REGRESSION

5.1 Introduction

In this section, predicted values are presented and compared to official crime data. The aim of this is to check the relationship between the two and the validity. The predicted values are not the actual values. However, they give an idea or a range of future trends. Presenting this section is essential in order to better understand the results obtained and to gain insight into how the prediction methods work and how to interpret the prediction results. The comparison is done between the predicted results and the actual crime trends to find the accuracy of the results obtained, equation 5.1 is utilised below to calculate the accuracy.

5.2 Murder Prediction

For prediction, the study performed LR to predict the future trends of murder crime in South Africa. The dataset utilised in this study from 2005 to 2015, hence it was necessary to start the prediction from 2016 to 2022. The prediction of seven years is performed for each type of crime selected, and the results are compared to the actual crime statistics provided by the SAPS. The prediction of crime trends is depicted using graphs and tables. Figure 2.1 shows murder trends for all nine provinces, Figure 5.1 graphically shows the murder trends, the graph is showing a huge change of murder between 2011 and 2012. Figure 5.1 shows the murder trends starting from 2005 to 2015 with no prediction trends and analysis. The aim of presenting this graph is to view the crime pattern and to gain insight into the flow of crime in the eleven-year period under study. Figure 5.1 shows that from 2005–2011 murder has dropped and increased from 2012–2015.

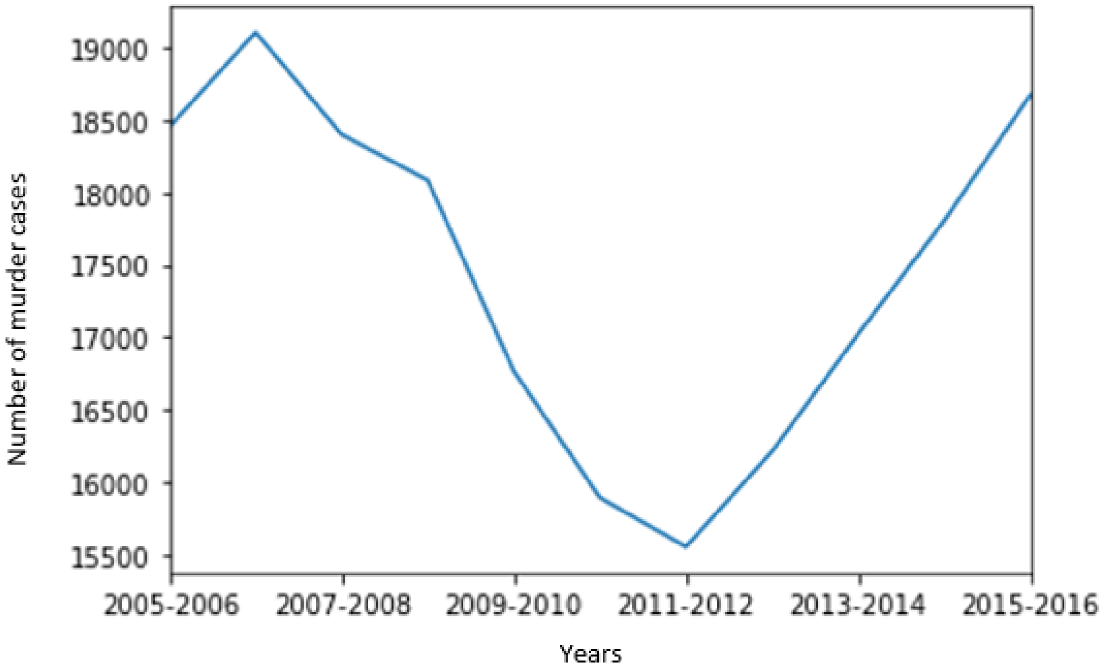


Figure 5.1 Murder trends in South Africa (2005 – 2015)

Table 5.1 depicts the predicted trends of murder from 2016 to 2022 LR is performed to obtain these results. The table displays the murder prediction descending during the seven-year period.

Table 5.1 Predicted murder trends (2016 – 2022)

Year	**Murder Prediction**
2016	18 250
2017	18 370
2018	18 690
2019	19 210
2020	19 280
2021	19 430
2022	19 710

Figure 5.2 depicts the predicted murder trends from 2016 to 2020. The red dotted line represents the linear trend line, which shows the relationship between the dependent and independent variables in the graph. The black points on the Figure:5.2 represent the predicted murder trends. The LR indicate that the graph is on the rise and the predicted trends are ascending.

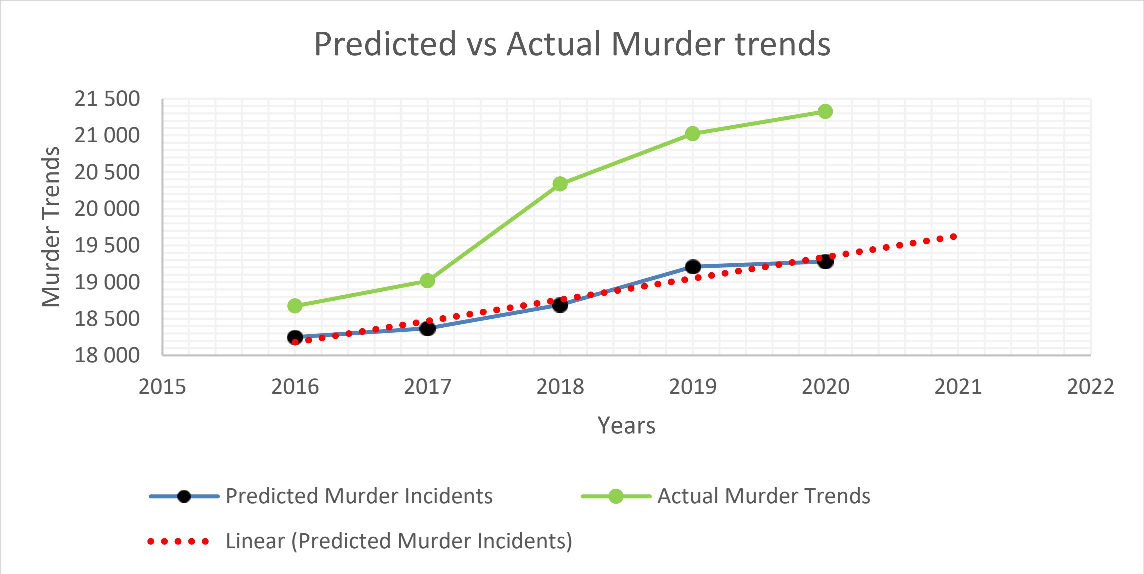


Figure 5.2 Murder Prediction vs Actual Murder trends (2016 – 2020)

5.2.1 Comparison of the actual murder trends and predicted values

South African statistics for 2019 revealed that 21 022 murders were reported during this period (African Check 2020). The study predicted that in 2019 the sum of 19 210 murders would occur. South African statistics for 2020 revealed that 21 325 murders were reported during this period (African Check 2020). The study predicted that in 2020 the sum of 19 280 murders would occurs. To confirm the accuracy of the prediction values, the percentage difference formula is utilised to calculate the accuracy of the results. Table 5.2 shows the accuracy of the prediction values for murder. An example of 2018 murder crime is shown below using equation 5.1:

$$Acc = 100 - \left(\frac{A_c - p_c}{A_c} * 100 \right) \tag{5.1}$$

Where accuracy is Acc, parameter A_c is the actual crime and P_c is the predicted crime.

= 100 – ((21 022 – 18 370)/21 022) * 100

= 100 – (2 652/21 022) * 100

= 100 – (0.120* 100)

∴ Accuracy = 100 – 12%

Accuracy = 88%

The model is showing 88% accuracy using equation (5.1). Similarly, the accuracy can be measured for the other years where the actual annual data is available.

Table 5.2 Murder prediction accuracy for South Africa (2016 – 2020)

Year	2016	2017	2018	2019	2020
Percentage	92%	97%	92%	92%	90%

5.3 Burglary Prediction

Burglary is one of the most commonly committed crimes in South Africa. According to Nkosi's (2017) study, burglary increased significantly in 2016/17. Numerous other crimes are associated with this crime, such as violence, theft, assault and murder, which are sometimes the results of burglary. Figure 5.3 illustrated that burglary is on the rise starting from 2005 – 2015. Figure 5.3 shows an increasing graph of residential burglary observed during the period 2005 – 2015.

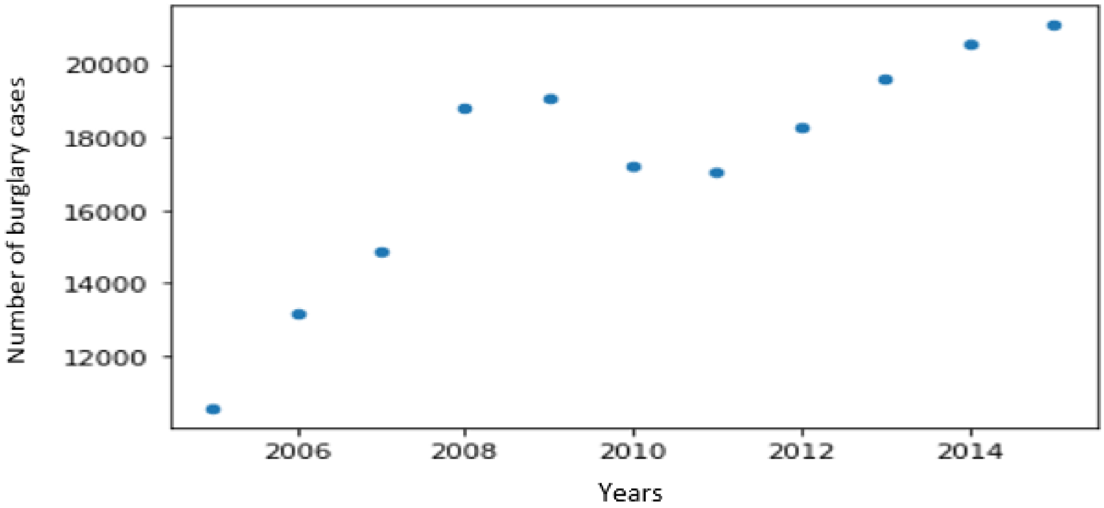


Figure 5.3 Residential Burglary trends for South Africa (2005 – 2015)

Table 5.3 depicts the predicted burglary trends for the period 2016 to 2022. Linear regression was performed to obtain these results. The table displays the burglary prediction ascending in the next 7 years. Nkosi (2017) confirmed that burglary in 2016 increased especially when compared to previous years, and it increased more than any other crime in South Africa in 2016/17. The community of Cape Town suburbs responded by taking the law into their own hands as crime increased, particularly burglary (Etheridge 2019).

Table 5.3 Predicted Residential burglary trends for South Africa (2016 – 2022)

Year	Burglary Prediction
2016	223 853
2017	232 330
2018	220 822
2019	212 310
2020	200 780
2021	200 100
2022	201 202

The Figure 5.4 shows the predicted burglary trends from 2016 to 2022. The black points in the graph below represent the predicted residential burglary trends. The LR line indicates whether the graph is ascending or descending. In this case, the graph is descending and the LR line confirms that the graph is going up.

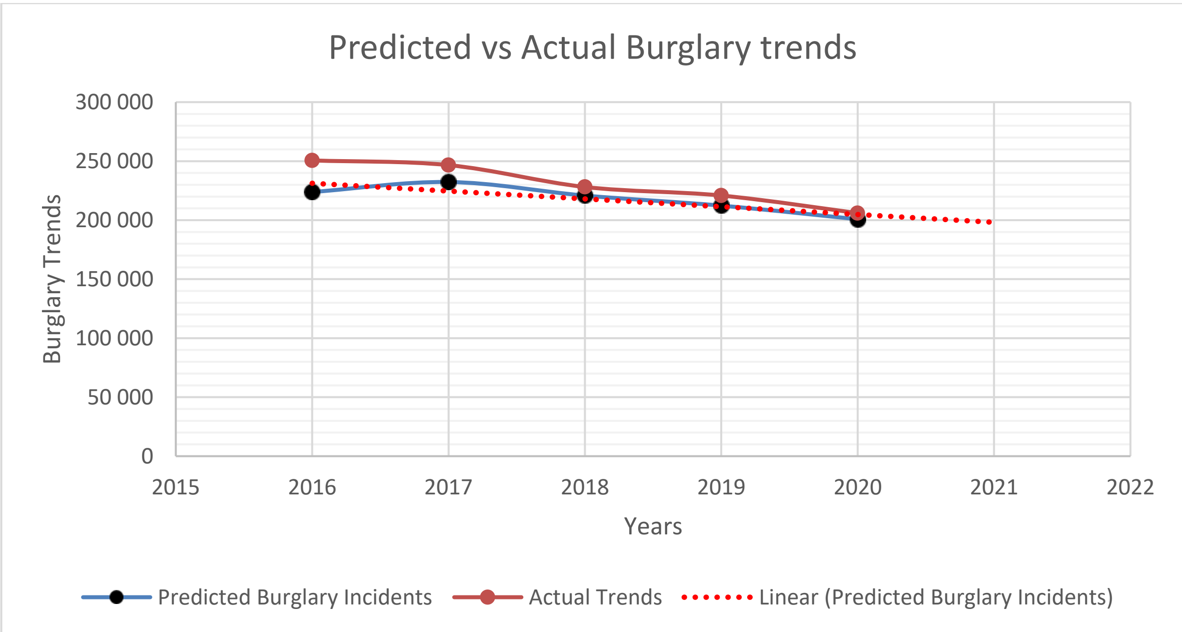


Figure 5.4 Burglary Prediction vs Actual Residential Burglary trends (2016 – 2020)

5.3.1 Comparison of the actual residential burglary trends and predicted values

South African statistics for 2018 revealed that 228 094 burglaries were reported during this period (Stats SA, 2020). The study predicted that in 2018, 220 822 burglaries would occur.

South African statistics for 2016 reveal that 223 853 burglaries were reported during this period (Writter, 2016). The study predicted that in 2016, 223 853 burglaries would be committed. To confirm the accuracy of the prediction values, the percentage difference formula is utilised to calculate the accuracy of the results. Table 5.4 shows the accuracy of the prediction values for residential burglary. Equation 5.1 was used to confirm the accuracy of the prediction model. The model for 2018 is showing 93% accuracy using equation (5.1). Similarly, the accuracy can be measured for the other years where the actual data is available.

Table 5.4 Residential burglary prediction accuracy for South Africa (2016 – 2019)

Year	2016	2017	2018	2019
Percentage	90%	95%	93%	96%

5.4 Carjacking Prediction

Carjacking is one crime that has become increasingly serious in the last decade. James (2017) mentioned that after the vehicles have been illegally taken from their owners, it is extremely difficult to find and locate those cars. During the period of 2016/17 there was 16 717 incidents of carjacking reported to the police. The study by Nkosi (2017) revealed that most victims were threatened with a life-endangering weapons, hence they did not fight for their vehicles. Figure 5.5 shows the carjacking trends from 2005 to 2015, and the graph illustrate that in 2011 carjacking started to increase again.

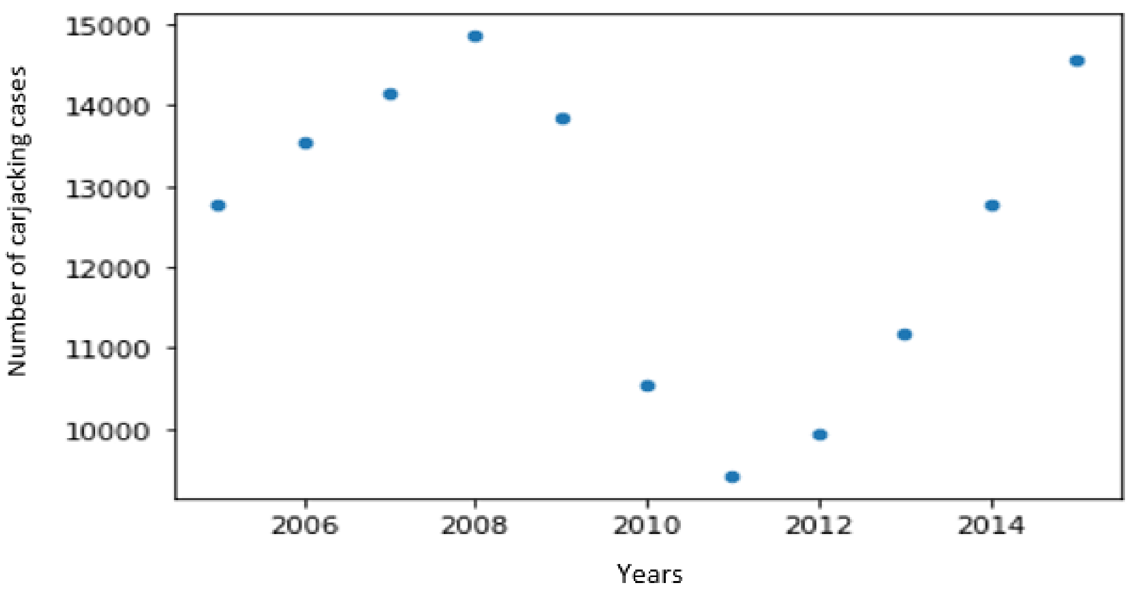


Figure 5.5 Carjacking trends for South Africa (2005 – 2015)

Table 5.5 depicts the predicted trends of carjacking from 2016 to 2022. Linear regression was performed to obtain these results. Nkosi (2017) confirmed that carjacking in 2016 had increased especially when compared to previous years, and this crime increased more than any other crime in South Africa in 2016/17.

Table 5.5 Predicted carjacking trends for South Africa (2016 – 2022)

Year	Carjacking Prediction
2016	11 563
2017	11 406
2018	11 248
2019	11 091
2020	11 933
2021	12 619
2022	12 776

Figure 5.6 depicts the predicted carjacking trends from 2016 – 2022. The black points in the graph below represent the predicted carjacking trends. The LR dotted line indicates whether the graph is ascending or descending. In this case, the graph is ascending and the LR line confirms that the graph is going up.

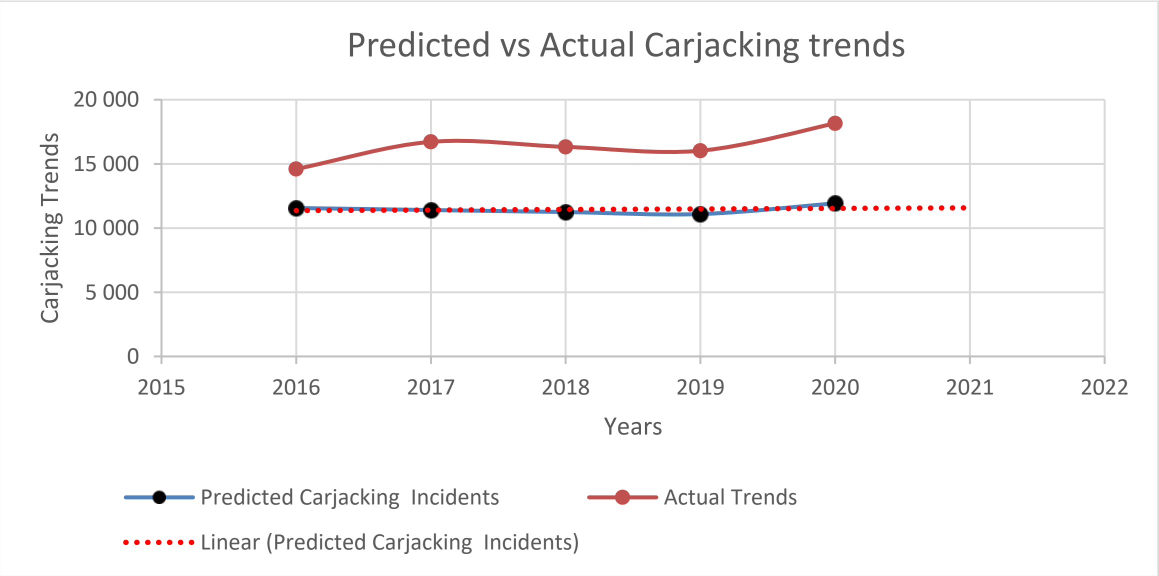


Figure 5.6 Carjacking prediction vs Actual carjacking trends (2016 – 2020)

5.4.1 Comparison of the actual carjacking trends and predicted values

South African statistics for 2016 revealed 14 602 carjacking cases reported to the police (Writter, 2016). The study predicted that in 2016, 11 563 cars would be hijacked. To confirm the accuracy of the prediction values, the percentage difference formula was utilised to calculate the accuracy of the results. Table 5.6 shows the accuracy of the prediction values for carjacking. The model for 2016 is showing 80% accuracy using equation (5.1). Similarly, the accuracy can be measured for the other years where the actual data is available.

Table 5.6 Carjacking prediction accuracy for South Africa (2016 – 2020)

Year	2016	2017	2018	2019	2020
Percentage	80%	69%	69%	70%	66%

5.5 Drug-related Crime Prediction

Most drug-related crimes are not reported to the police, especially in South Africa. The SAPS search for people who use drugs and commit these crimes. Incidents of drugs involve numerous crimes such as burglary, murder, attempted murder, violence, assault and rape. A study by Schwinn *et al.* (2019) stated that drug-related crimes are hard to solve because no one comes forward with information, even though the community knows that certain people in their environment are using drugs, they do not report them. Hence, these types of crimes are difficult to solve. According to the crime statistics for the period 2005–2015 drug-related crime was increasing. Schwinn *et al.* (2019) further mentioned that some incidents of drug-related crimes are also seen in both primary and high schools of South Africa. Figure 5.7 depicts the drugs trends for the period 2005 to 2015.

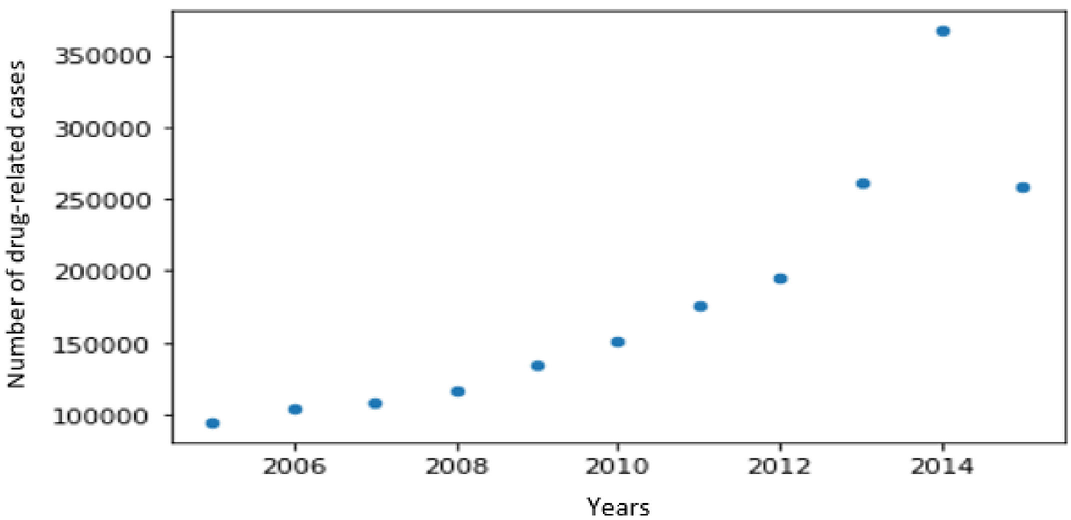


Figure 5.7 Drug-related crime trends for South Africa (2005 – 2015)

Table 5.7 presents the predicted trends of drugs-related crime for the period 2016 to 2022. Linear regression was performed to obtain these results. The table shows that the crime counts have been gradually increasing after 2015. James (2017) mentioned that drug-related crimes cases are very difficult to solve and most cases end up being dropped and are left unsolved, while others are not reported to the police. Table 5.7 shows that drug-related crime trends were ascending for the period 2016 – 2022.

Table 5.7 Predicted drug-related crime trends for South Africa (2016 – 2022)

Year	Drug-related crime Prediction
2016	246 802
2017	229 477
2018	246 534
2019	225 362
2020	208 305
2021	231 247
2022	254 190

Figure 5.8 shows the drug-related predicted crime trends for the period 2016 to 2022. The LR line indicates that the trend is ascending. Black points in the line graph present the drug-related predicted trends.

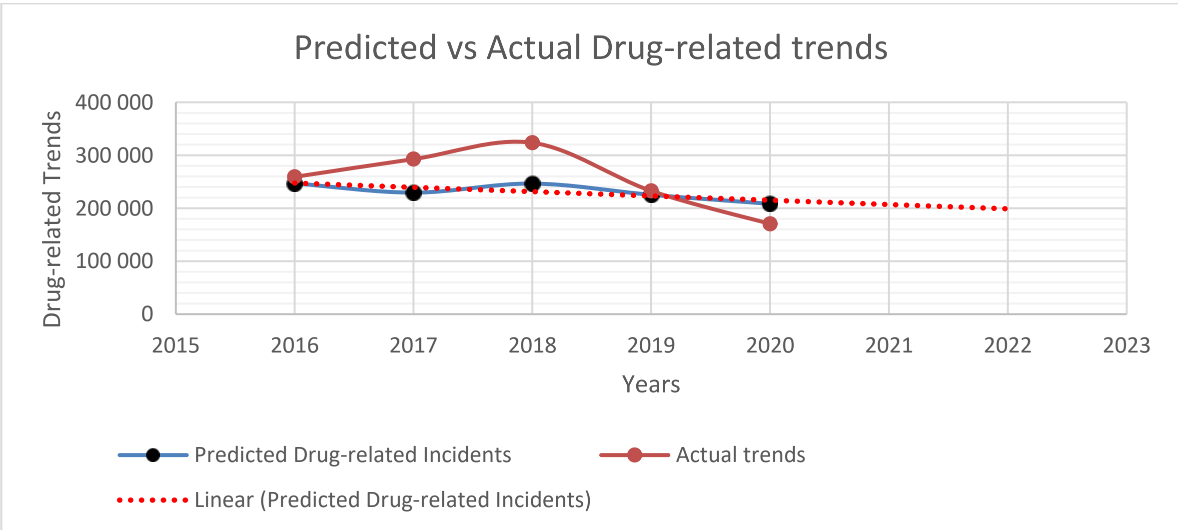


Figure 5.8 Drug-related prediction vs Actual Drug-related trends (2016 – 2020)

5.5.1 Comparison of the actual drug-related crimes to crime trends and predicted values

South African statistics for 2017 reveals that 292 689 drug-related crimes were reported during this period (Writter, 2018). The study predicted that in 2017 the sum of 339 477 drug-related crimes would occur. South African statistics for 2016 reveals that 259 165 drug-related crimes were reported during this period (Writter, 2016). The study predicted that in 2016 the sum of 316 534 drug-related crimes would occur. To confirm the accuracy of the prediction values, the percentage difference formula is utilized to calculate the accuracy of the results. Table 5.8 shows the accuracy of the prediction values for drug-related crimes. The model for 2017 is showing 79% accuracy using equation (5.1). Similarly, the accuracy can be measured for the other years where should be change to when the actual data is available.

Table 5.8 Drug-related crime prediction accuracy for South Africa (2016 – 2019)

Year	2016	2017	2018	2019
Percentage	96%	79%	77%	97%

5.6 Chapter Summary

In this chapter, the researcher conducted the application of LR to predict future trends. The researcher compared the predicted results against the official crime data in South Africa by calculating the accuracy of the results. Throughout this chapter, it is observed that the difference between the two variables (predicted and actual trends) is not above 35%. Meaning the results are at least 65% accurate. In the next chapter, the summary, conclusion and implications of the study are presented.

CHAPTER SIX: SUMMARY, CONCLUSIONS AND IMPLICATIONS OF STUDY

6.1 Introduction

This chapter presents the summary, conclusions and implications of the study. This study utilised methods in Python to do the pre-processing of dataset and implemented CPA to detect the changes that occurred during the specified period. Linear regression was performed to predict future trends. The objective of utilising these methods and techniques was to enlighten and give tangible, science-based information for consideration and use by the relevant authorities responsible for seeking crime reduction strategies in South Africa.

6.2 Summary of Conclusions

This section of the study summarises the conclusions and how each objective was achieved. Each chapter either met or contributed to the achievement of an objective.

6.2.1 To Provide Insights into Crime Trends in South Africa Using Statistics from the Extant Literature [RO 1]

The first objective of this dissertation was met in Chapter Two by reviewing literature of the crime trends in South Africa. Crimes such as murder, burglary, carjacking, drug-related crime and crime in general were covered. While providing an insight into crime trends, the study further looked at the major factors that contribute to crime. Unemployment and inequality were identified as the most common factors leading to crime, as highlighted in this chapter.

6.2.2 To Apply the Change-Point Analysis Algorithm as an Effective Data Mining Tool in Detecting Abrupt Changes in Recurrent Crime in South Africa [RO 2]

Objective two was achieved in Chapter Three and Chapter Four. The researcher carried out a study guided by a research design which combined all the significant parts of the study. Chapter Three described how CPA works and explained step-by-step how the two methods namely, CUSUM and Bootstrap were conducted to obtain accurate results. The explanation included formulas and their descriptions. This objective was accomplished in Chapter Four through the actual application of CPA. The results, analysis and discussions were provided in this chapter. All four selected crimes in this study were analysed separately. Throughout the analysis, it was observed that the Western Cape, Gauteng and KwaZulu-Natal provinces, had the highest crime rates in all four of the specified crimes (murder, burglary, carjacking and drug-related crimes).

Hence, the study suggested that these three provinces require serious attention and necessary interventions because they contribute significantly to crime in South Africa.

6.2.3 To Predict Future Trends in Habitual Crime in South Africa Using Machine Learning [RO 3]

This objective was met in Chapter Three and Chapter Five. In Chapter Three, the study described how and why LR was performed in this study. The explanations included LR formulas and descriptions. Chapter Three also provided the importance of utilising LR in this study. This objective was achieved in Chapter Five by applying LR to predict further crime trends in South Africa. The results obtained in this chapter were compared against the official crime data to examine the accuracy of the results. Throughout this chapter, it was observed that at some point from 2012 crime had increased considerably. Additionally, the difference between the two variables (predicted and actual trends) was not above 20% meaning the results of the predictions were at least 80% accurate.

6.3 Implications of Study

The results have clearly indicated that crime in South Africa is increasing rapidly especially when compared to the early 2000s. The objectives of this study were to provide an insight into crime, to guide future interventions. This was achieved by analysing the dataset, detecting changes and predicting the future trends. Change-point analysis revealed unseen occurrences of changes and shifts that would have been challenging to discover when using other methods. Linear regression gave an insight into the direction and prediction of the crime in the future. Prediction values for future trends provide advance notification of what can be expected in future (Iabal *et al.* 2013). The findings of this study were compared to the actual crime trends to determine the difference and the accuracy of the results obtained. The results of the LR clearly indicate the rise of crime in South Africa. Murder, burglary, carjacking and drug-related crimes have been increasing and are expected to increase over the next two years, as the study offered prediction until the end of 2022.

6.3.1 Opportunities

The results of this study indicate that there is a serious need to address crime in South Africa, specifically in nucleated provinces that are experiencing a high number of crime incidences, as mentioned in Chapter Four. The results of this study imply that different provinces need different kind of crime-reduction interventions. The lack of opportunities and socio-economic inequality have been singled out as the root cause of crime in South Africa. Chapter Three highlighted that poverty is the main issue that leads people to commit crime. Most people are moving away from rural areas to urban settlements in search of opportunities. That is why most crimes are reported in towns and cities, where nucleated settlements are normally found. Therefore, the creation of job opportunities by the authorities in both rural and urban settings could result in a decline in crime.

6.3.2 Recommendations

Boateng (2016), mentioned that people do not report crimes and these account for a large number of crimes committed. The results of this study suggest that different provinces experiences different crimes, provinces such as KwaZulu-Natal, Gauteng and Eastern Cape experienced higher numbers of murder than the other provinces. The study suggests that each province should be treated individually. For instance, provinces such as Gauteng, Western Cape and KwaZulu-Natal have a high population density where many people dwell in crowded areas (Kamer 2020). The study implies that more police should be deployed in the hotspots of provinces that are experiencing high crime.

The study further suggested that authorities should use technology to improve their understanding and ability to solve crime. For instance, there are places that are experiencing a high number of carjackings. The victims of carjacking do not report this crime sometimes because the police find this type of crime very difficult to solve (Boateng, 2016). The study implies that authorities should make use of technology to identify and analyse the hotspots of specific crimes in each province. This will allow for information that will guide the process of deploying SAPS officers and installing cameras in the most appropriate areas to ensure that culprits able to be identified and apprehended. However, the results also suggest that communities should also be urged to work with the SAPS to reduce crime. This is particularly relevant in the case of drug-related crime, where the available statistics do not appear to reflect the situation on the ground. Low recorded numbers do not necessarily equate to no drug-related crimes in most townships and cities of South Africa. It has been shown that many murder crimes are the direct or indirect result of drug use (Lindegaard 2017).

6.3.3 Future study

Future studies should focus on one province and to develop an insight of the dominating crimes, and hotspots of crimes. This study suggested that each province is unique and should be approached differently. Doing this for all nine provinces will assist the authorities to address crime challenges in each province.

Future studies should assess different technological approaches to analyse and solve crime. These different approaches could help in identifying the combination of techniques that work most powerfully together, in order to provide the most accurate results. Each province should ideally have a different technique approach to solve crime. By doing so, a comparison of these approaches would be done to check the most flexible and suitable approach for solving crime.

6.4 Chapter summary

In this chapter, the researcher explained how each objective of this study was achieved. How CPA as a ML algorithm has proven to be more effective for detecting abrupt and hidden changes compared to the control chart. The study illustrated the need of intervention since the analysis has confirmed the existence and the rise of crime in South Africa. There have been few studies conducted in South Africa that used ML to solve crime. Hence, conducting this study is paramount because it targets to benefit the people of South Africa. This chapter further suggested alternate plans should be put in action and explore the use of technology to solve crime. South Africa is experiencing a high rate of crime and that affects the normal living conditions of people. It is evident that sharing notions and having studies like this could help reduce crime. This study aims to ensure and encourage the safety of all South Africans with a view to a better future for all.

REFERENCES

Abrahams, N., Mathews, S., Lombard, C., Martin, L. J. and Jewkes, R. 2017. Sexual homicides in South Africa: a national cross-sectional epidemiological study of adult women and children. *PLoS One,* 12(10): 1-13.

Abutabenjeh, S. and Jaradat, R. 2018. Clarification of research design, research methods, and research methodology: a guide for public administration researchers and practitioners. *Teaching Public Administration*, 36(3): 237-258.

African Check. 2019. *Factsheet: South Africans crime statistics for 2018/19*. Available:

https://citizen.co.za/news/south-africa/crime/2178462/factsheet-south-africas-crime-statistics-for-2018-19/ (Accessed 4 May 2021).

African Check. 2020. *Factsheet: South Africans crime statistics for 2019/20*. Available: https://africacheck.org/fact-checks/factsheets/factsheet-south-africas-crime-statistics-201920 (Accessed 10 January 2021).

Africa Times Editor. 2019. *African homicide rate far higher than global average.* Available: https://africatimes.com/2019/07/08/report-african-homicide-rate-far-higher-than-global-average/ (Accessed 06 June 2021).

Ahishakiye, E., Taremwa, D., Omulo, E. O. and Niyonzima, I. 2017. Crime prediction using tree (J48) classification algorithm. *International Journal of Computer and Information Technology*, 6(3): 188-195.

Albrecher, H., Bladt, M., Kortschak, D., Prettenthaler, F. and Swierczynski, T. 2019. Flood occurrence change-point analysis in the paleo flood record from Lake Mondsee. *Global and Planetary Change*, 178: 65-76.

Alves, L. G. A., Ribeiro, H. V. and Rodrigues, F. A. 2018. Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Application*, 505: 435-443.

Aminikhanghahi, S. and Cook, D. 2017. A survey of methods for time series change-point detection. *Knowledge and Information Systems*, 51: 339-367.

Amiri, S., Brooks, K. R., Vile, B. J. and Daratha, K. B. 2019. Natural surveillance characteristics of building openings and relationship to residential burglary. *Applied Geography,* 102: 99-108.

Anon. 2020a. *Inanda, KwaZulu-Natal*. Available: https://en.wikipedia.org/wiki/Inanda,_KwaZulu-Natal (10 January 2021).

Anon. 2020b. *Thabong*. Available: https://en.wikipedia.org/wiki/Thabong (Accessed 3 December 2020).

Asongu, S. A. and Acha-anyi, P. N. 2019. The murder epidemic: a global comparative study. *International Criminal Justice Review*, 29(2): 105-120.

Asongu, S. A. and Nwachukwu, J. C. 2017a. Mobile phones in the diffusion of knowledge and persistence in inclusive human development in Sub-Saharan Africa. *Information Development*, 33(3): 289-302.

Asongu, S. A. and Nwachukwu, J. C. 2017b. The comparative inclusive human development of globalisation in Africa. *Social Indicators Research*, 134: 1027–1050.

Arif, S. N. A. M., Mohsin, M. F. M., Bakar A. A., Hamdan, A. R. and Abdullah, B. 2017. Change-point analysis: a statistical approach to detect a potential abrupt change. *Jurnal Teknologi*, 79(5): 147-159.

Beangstrom, P. 2018. *Kimberley couple tell of horrific robbery ordeal inside 'secure home'*. Available: https://www.iol.co.za/news/south-africa/northern-cape/kimberley-couple-tell-of-horrific-robbery-ordeal-inside-secure-home-17253922 (Accessed 1 December 2020).

Berman, A. H., Bergman, H., Palmstierna, T. and Schlyter, F. 2005. Evaluation of the drug use disorders identification test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *Europe Addiction Research*, 11(1): 22-31.

Boateng, F. D. 2016. Crime reporting behaviour: do attitudes toward the police matter? *Journal of Interpersonal Violence*, 33(18): 1-26.

Breetzke, D. G. and Cohn, E. G. 2013. Burglary in gated communities: an empirical analysis using routine activities theory. *International Criminal Justice Review*, 23(1): 56-74.

Brook, J. S., Morojele, N. K., Pahl, K. and Brook, D. W. 2016. Predictors of drug use among South African adolescents. *Journal of Adolescent Health*, 38(1): 26-34.

Buiten and Naidoo. 2016. Framing the problem of rape in South Africa: Gender, race, class and state histories. *Current Sociology*, 64(4): 1-16.

Busakwe, G. and Jongikaya, M. 2007. *Mthatha: Ngangelizwe*. Available: http://www.justiceforum.co.za/community%20prosecutors/6.pdf (Accessed 29 November 2020).

Business insider South Africa. 2019. *South Africa climbed the ranks to become one of the most murderous countries in the world*. Available: https://www.businessinsider.co.za/south-africa-crime-2019-11 (Accessed 10 December 2020).

Cele, B. 2020. *About 370 people arrested in KZN for various crimes, including murder since Thursday*. Available: https://www.sabcnews.com/sabcnews/about-370-people-arrested-in-kzn-for-various-crimes-including-murder-since-thursday/ (Accessed 11 December 2020).

Chatterjee, A. 2019. Measuring wealth inequality in South Africa: an agenda. *Development Southern Africa*, 36(6): 839-859.

Cheteni, P., Mah, G. and Yohane, K. 2018. Drug-related crime and poverty in South Africa. *General & Applied Economics*, 6(1): 1-16.

Crime Stats. 2020. *South African crime stats simplified*. Available: https://www.crimestatssa.com/national.php (Accessed 05 January 2021).

Cronje, C. J. and Spocter, M. 2017. Home fortification in Soneike, Cape Town, South Africa. *Journal of Housing and the Built Environment*, 33(4): 713-732.

David, H. B. F. and Suruliandi, A. 2017. Survey on crime analysis and prediction using data mining techniques. *ICTACT Journal of Soft Computing*, 7(3): 1459-1466.

Davis, L. 2014. Carjacking- Insights from South Africa to a new crime problem. *Australian & New Zealand Journal of Criminology*, 36(2): 173-191.

De Juan, A. and Wegner. E. 2017. Social Inequality, State-centered Grievances, and Protest: Evidence from South Africa. *Journal of Conflict Resolution*, 63(1): 31-58.

De Vos, A. S., Delport, C. S. L., Fouche, C. B. and Strydom, H. 2011. *Research at grass roots for the social sciences and human services professions: the social sciences and human services professions*. 4th ed. Pretoria: Van Schaik.

Dierbach, S. 2014. Python as a first programming language. *Journal of Computing Sciences in Colleges*, 29(3): 153-154.

Dolley, C. 2021. *Triple killing: two men set alight and decapitated after suspected drug kingpin shot dead in Durban*. Available: https://www.dailymaverick.co.za/article/2021-01-04-triple-killing-two-men-set-alight-and-decapitated-after-suspected-drug-kingpin-shot-dead-in-durban/ (Accessed 10 January 2021).

Etheridge, J. 2019. *Dramatic scenes in Cape Town suburb as tactical response team surround alleged burglars*. Available: https://citizen.co.za/news/south-africa/crime/2188009/dramatic-scenes-in-cape-town-suburb-as-tactical-response-team-surrounds-alleged-burglars/ (Accessed 27 December 2020).

Fearnhead, P and Rigaill, G. 2019. Change-point Detection in the Presence of Outliers. *Journal of the American Statistical Association*, 114(525): 169-183.

Gaya, H. J. and Smith, E. E. 2016. Developing a qualitative single case study in the strategic management realm: an appropriate research design? *International Journal of Business Management and Economic Research*, 7(2): 529-538.

Gould, C., Burger, J. and Newham, G. 2016. The Saps Crime Statistics: what they tell us -and what they don't. *South African Crime Quarterly*, 42: 1-12. Gould

Gulma, U. L., Evans, A., Heppenstall, A. and Malleson, N. 2013. Income inequality and crime: a review and explanation of the time series evidence. *Sociology and Criminology-Open Access*, 1(1): 1-9.

Hiropoulos, A. and Porter, J. 2014. Visualising property crime in Gauteng: applying GIS to crime pattern theory. *South African Crime Quarterly*, 47(1): 17-28.

Hodgkinson, T. and Andresen, M., A. (2019). Changing spatial patterns of residential burglary and the crime drop: the need for spatial data signatures. *Journal of Criminal Justice,* 61:90-100.

Holmostrom, J., Hameri, A. P. and Ketokivi, M .2014. Bridging practice and theory: a design science approach. *Decision Science*, 40(1): 65-87.

Iabal, R., Azrifah, A. M., Mustapha, A., Panahy, P. H. S. and Khanahmadliravi, N. 2013. An experimental study of classification algorithms for crime prediction. *India Journal of Science and Technology*, 6(3):4220-4225.

Jain, V., Bhatia, A., Sharma, Y. and Arora, V. 2017. Crime prediction using K-means algorithm. *Global Research and Development Journal for Engineering,* 2(5): 2455-5703.

James, C. 2017. Carjacking in South Africa: exploring its consequences for victims. *African Journal of Criminology & Victimology,* 30(2): 147-161.

James, N., Menzies, M., Azizi, L. and Chan J. 2019. Novel semi metrics for multivariate change-point analysis and anomaly detection. *Physica D: Nonlinear Phenomena,* 3(1): 1-33.

Jumin, E., Zaini, N., Ahmed, A. N., Abdukkah, S., Ismail, M., Sherif, M., Sefelnasr, A. and EI-Shafie, A. 2020. Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction. *Engineering Applications of Computational Fluid Mechanics*, 14(1): 713-725.

Kamer, K. 2020. *Total population of South Africa in 2018, per province*. Available:

https://www.statista.com/statistics/1112169/total-population-of-south-africa-by-province/ (Accessed 29 February 2021).

Khoele, K. B., De Wet, P. H., Pretorius, H. W. and Sommerville, J. 2016. Case series of females charged with murder or attempted murder of minors and referred to Weskoppies Hospital in terms of the Criminal Procedure Act over a period of 21 years. *South African Journal of Psychiatry*, 22(1): 1-7.

Kleck, G. and Patterson, E. B. 1993. The impact of gun control and gun ownership levels on violence rates. *Journal of Quantitative Criminology,* 9(3): 249-287.

Kuncheva, L. 2013. Change detection in streaming multivariate data using likelihood detectors. *IEEE Transactions on Knowledge and Data Engineering*, 25(5): 1175–1180.

Lewis, S. 2015. Qualitative inquiry and research design: choosing among five approaches. *Society for Public Health Education*, 16(4): 473-475.

Lindegaard, M. R. 2017. *Homicide in South Africa. The handbook of homicide*. United States: John Wiley & Sons.

Liu, S, Yamada, M, Collier, N. and Sugiyama, M. 2013. Change-point detection in time-series data by relative density-ratio estimation. *Neural Network*, 43(1): 72-83.

Lung Lin, Y., Yang Chen, T. and Chi Yu, L. 2017. Using machine learning to assist crime prevention. *IIAI International Congress on Advanced Applied Information*, 1: 1029-1030.

Mail and Guardian Staff Reporter. 2013. *Crime stats: Where murder happens in South Africa*. Available: https://mg.co.za/article/2013-09-19-where-murder-happens-in-south-africa/ (Accessed 2 January 2021).

Marinovich, G. 2012. *The murder fields of Marikana. The cold murder fields of Marikana*. Available: https://www.dailymaverick.co.za/article/2012-09-08-the-murder-fields-of-marikana-the-cold-murder-fields-of-marikana/ (Accessed 1 January 2021).

Marie, R. L. 2017. Homicide in South Africa, The Handbook of Homicide, First Edition. Edited by Fiona Brookman, Edward R. Maguire, and Mike Maguire. © 2017 John Wiley & Sons, Inc. Published 2017 by John Wiley & Sons, Inc.

Mathews, S., Abrahams, N., Martin, L. J. and Lombard, C. 2019. Homicide patterns among adolescents. *PLoS One*, 14(8): 1-10.

Matzopoulos, R., Simonetti, J., Prinsloo, M., Neethling, I., Groenewald, P., Dempers, J., Martin, L. J., Rahbar, R. A., Myers, J. E. and Thompson, L. M. 2018. A retrospective time trend study of firearm and non-firearm homicide in Cape Town from 1994 to 2013. *South African Medical Journal*, 108(3): 197-204.

McCafferty, R. ed. 2003. *Murder in South Africa: A comparison of past and present*. Available: http://www.frontline.org.za/Files/PDF/murder_southafrica%20%285%29.pdf (Accessed 24 November 2020).

McClendon, L. and Meghanathan, N. 2015. Using machine learning algorithm to analyze crime data. *Machine Learning and Applications*: *An International Journal*, 2(1):1-12.

Mittal, M., Goyal, L. M., Sethi, J. K. and Hemanth, D. J. 2019. Monitoring the impact of economic crisis on crime in India using machine learning. *Computational Economics*, 53(1): 1-19.

Monyeki, P., Naicker, N. and Obagbuwa, I. C. 2020. Change-point analysis: an effective technique for detecting abrupt change in the homicide trends in a democratic South Africa. *The Scientific World Journal*, 2020: 1-10.

Musah, A., Umar, F., Yakubu, K. N., Ahmad, M., Babagana, A., Ahmad, A., Thieme, T. A. and Cheshire, J. A. 2020. Assessing the impacts of various street-level characteristics on the burden of urban burglary in Kaduna. *Applied Geography*, 114: 1-9.

Nath, S. V. 2006. Crime pattern detection using data mining. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops*. 18-22 December. Hong Kong, China, 41-44. Available: doi: 10.1109/WI-IATW.2006.55. (Accessed 07 January 2021).

Ncontsa, V. N. and Shumba, A. 2013. The nature, causes and effects of school violence in South African high schools. *South African Journal of Education*, 33(3): 1-15.

Ngqakamba, S. 2019. *5 arrested for murder, robbery at Johannesburg tavern*. Available: https://www.news24.com/news24/southafrica/news/5-arrested-for-murder-robbery-at-joburg-tavern-20190104 (Accessed 9 December 2020).

Nimon, K. F. and Oswald, L. F. 2013. Understanding the results of multiple linear regression: Beyond standardized regression coefficient. *Organizational Research Methods*, 16(4): 650-674.

Nkosi, B. 2017. *Home is where the robbery is - over half of all South African crime is housebreaking*. Available: https://www.timeslive.co.za/news/south-africa/2017-05-31-home-is-where-the-robbery-is-over-half-of-all-sa-crime-is-housebreaking/ (Accessed 11 December 2020).

Otieno, G. E., Marinda, T., Arnighausen, B. and Tanser, F. 2015. High rates of homicide in a rural South African population. *Population Health Metrics*, 13(20): 1-9.

Phagane, T. 2020. *Gauteng most violent province in South Africa: SAPS*. Available: https://www.sabcnews.com/sabcnews/gauteng-most-violent-province-in-sa-saps/ (Accessed 8 December 2020).

Phori, M.G. 2017. Analysis of the perceptions of crime against tourists in Florida road, Durban. M.Ed., Durban University of Technology.

Poti, I. 2019. *Children reportedly being used as drug mules in Eastern Cape*. Available: https://www.sabcnews.com/sabcnews/children-reportedly-being-used-as-drug-mules-in-eastern-cape/ (Accessed 25 May 2021).

Prabakaran, S. and Mitra, S. 2018. Survey of analysis of crime detection techniques using data mining and machine learning. *Journal of Physics*, 2(1): 1-11.

Prinsloo, M., Matzopoulos, R., Laubscher, R., Myers, J. and Bradshaw, D. 2016. Validating homicide rates in the Western Cape Province, South Africa: Findings from the 2009 injury mortality survey. *South African Medical Journal*, 106(2):193-195.

Research Channel Africa. 2019. *73.4% of SAPS stations in the Free State understaffed*. Available: https://www.polity.org.za/article/734-of-saps-stations-in-the-free-state-understaffed-2019-09-20 (Accessed 8 December 2020).

Retorius, G., Cleak, J. H. and Morgan, B. 2010. The lived experience of losing a sibling through murder. *Indo-Pacific Journal of Phenomenology*, 10(1): 1-13.

South African Institute of Race Relations. 2015. *An analysis of the national crime statistics*. Available: https://www.saps.gov.za/about/stratframework/annual_report/2013_2014/crime_statreport_2014_part1.pdf (Accessed 12 December 2020).

Langton, S. and Steenbeek, W. 2017. Residential burglary target selection: an analysis at the property-level using google street view. *Applied Geography*, 86: 292-299.

Sathyadevan, S., Devan, M. S. and Gandadharan, S. S. 2014. Crime analysis and prediction using data mining. In: Negi, A. ed. *First International Conference on Networks & Soft Computing (ICNSC2014)*: Guntur India, 19-20 August 2014. Guntur India: IEEE, 405-412. Available: https://ieeexplore.ieee.org/document/6906719 (Accessed 20 January 2021).

Schwinn, T. M., Schinke, S. P., Keller, B. and Hopkins, J. 2019. Two and Three year follow up a gender specific, web-based drugs abuse prevention program for adolescent girls. *Addictive Behaviors*, 93: 86-92.

Shabaz, M., Masood, S. A. and Khan, A. 2011. Data mining methodology in perspective of manufacturing databases. *Journal of American Science*, 6(11): 999-1012.

Skelton, A. 2002. Restorative justice as a framework for juvenile justice reform: a South African perspective. *The British Journal of Criminology*, 42(3): 496-513.

Song, J., Spicer, V. and Brantingham, P. 2013. The edge effect: exploring high crime zones near residential neighbourhoods. *IEEE International Conference on Intelligence and Security Informatics.* Seattle, WA, USA, 4-7 June 2013: 245-250. Available: https://ieeexplore.ieee.org/document/6578828/metrics#metrics (22 January 2021).

South Africa. 2020. *About Inanda township*. Available: https://www.sa-venues.com/attractionskzn/inanda-township.php (Accessed 22 January 2021).

South African Police Service. 2015. *An analysis of the national crime statistics*. Available: https://www.saps.gov.za/services/crimestats.php (Accessed 2 January 2021).

Sperino, S., F. 2020. Co-Worker Evidence in Court. *Saint Louis University Law Journal*, 65(1): 1-25.

Statistics South Africa. 2020. *Housebreaking is the number one crime in South Africa*. Available: http://www.statssa.gov.za/?p=12614 (Accessed 2 January 2021).

Statista. 2020. *Number of carjackings in South Africa from the fiscal year of 2009 to the fiscal year of 2020*. Available: https://www.statista.com/statistics/1006796/number-carjackings-south-africa/ (Accessed 20 January 2021).

Stop crime. 2011. *Release of crime stats 2010/2011*. Available: https://dnaproject.co.za/blog/release-of-crime-stats-20102011 (Accessed 20 January 2021).

Swemmer, S. 2019. Ethical loneliness and the development of a victim-focused approach to rape cases in South Africa. *Aggression and Violent Behavior*, 47: 298-306.

Tartakovsky, G. A., Rozovskii, L., Blazek, R. B. and Kim, H. 2006. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, 54(9): 3372-3382.

Taylor, W.A. 2018. Change-point analysis: a powerful new tool for detecting changes. *Taylor Enterprises*, 1-19.

Van Niekerk, S. I. and Greeff, A. 2020.Factors associated with resilience in families after a house robbery incident: a mixed-methods study. *African Journal of Criminology & Victimology*, 33(1): 1-15.

Van Zyl, G. S., Wilson, G. D. H. and Pretorius, R. 2003. Residential burglary in South Africa: why individual households adopt reactive strategies. *ActaCriminologica*, 16(3): 107-123.

Waduge, D., Ranathunga, D. L. and Lanka, S. 2017. Machine Learning approaches for detect crime patterns. M.Ed., University of Moratuwa. (Accessed 10 February 2021).

Waterworth, T., Guy, D. and Chemaly, F. 2020. *KZN is a bloody battle ground and SA's most violent province - Crime Stats*. Available: https://www.iol.co.za/ios/news/kzn-is-a-bloody-battle-ground-and-sas-most-violent-province-crime-stats--8f450fc5-9020-489f-bb46-8787797e6a47 (Accessed 1 December 2020).

Wessels, S. 2020. *Crime statistics for South Africa*. Available: https://www.kaggle.com/slwessels/crime-statistics-for-south-africa (Accessed 18 December 2020).

Wilkinson, L. 2011. Exact and Approximate Area-Proportional Circular Venn and Euler Diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 18(2): 321-31.

Writter, S. 2016. *South Africa crime stats 2016: everything you need to know*. Available: https://businesstech.co.za/news/trending/135321/south-africa-crime-stats-2016-everything-you-need-to-know/ (Accessed 28 May 2021).

Writter, S. 2017. *South Africa crime stats 2017: everything you need to know*. Available: https://businesstech.co.za/news/lifestyle/207087/south-africa-crime-stats-2017-everything-you-need-to-know/ (Accessed 10 January 2021).

Writter, S. 2017. *South Africa crime stats 2018: everything you need to know*. Available: https://businesstech.co.za/news/government/270689/south-africa-crime-stats-2018-everything-you-need-to-know/ (Accessed 25 May 2021).

Writter, S. 2020. *Carjackings in South Africa – here are the major trends and hotspot areas*. Available: https://businesstech.co.za/news/motoring/421542/carjackings-in-south-africa-here-are-the-major-trends-and-hotspot-areas/ (Accessed 20 May 2021).

Zou, K. Z., Tuncali, K. and Silverman. 2003. Correlation and simple linear regression. Radiology, 227(3): 617-22.

Zulu, B. M., Urbani, G. and van der Merwe, A. 2004. Violence as an impediment to a culture of teaching and learning in some South African schools. *South African Journal of Education,* 24(2): 170-175.

Annexure A: Cover of Turn it in Report

Dissertation 02 June 2021

ORIGINALITY REPORT

12%	9%	5%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.hindawi.com Internet Source	2%
2	P. Monyeki, N. Naicker, I. C. Obagbuwa. "Change-Point Analysis: An Effective Technique for Detecting Abrupt Change in the Homicide Trends in a Democratic South Africa", The Scientific World Journal, 2020 Publication	1%
3	ir.dut.ac.za Internet Source	1%
4	Submitted to University of KwaZulu-Natal Student Paper	1%
5	researchspace.ukzn.ac.za Internet Source	1%
6	journals.co.za Internet Source	<1%
7	Submitted to Universiti Teknologi Malaysia Student Paper	<1%
8	uir.unisa.ac.za Internet Source	

Annexure B: Language Editing Certificate

Geraldine Coertze
Independent Communications Consultant
Language Practitioner –
English Language Academic Editor

Certificate of English Language Academic Editing

Author: Phirime Monyeki **Date:** 10 June 2021
Document type: Masters dissertation
Discipline: Information and Communication Technology
Academic supervisor: Dr N. Naicker
Higher Education Institution: Durban University of Technology

Dissertation title: Data Mining to Analyse Recurrent Crime in South Africa

This serves to certify that the above thesis was language edited and that assistance was provided with checking the citations and the format of the reference list.

The document was returned to the author with tracked changes, comments and a comprehensive academic editing report to facilitate the multi-level correction process. It was the responsibility of the author to accept or reject changes and to attend to all issues raised in the comments, email communication and in the report. The final, corrected version of the document was not proofread, although assistance was provided with generating the Table of Contents, List of Figures, List of Tables, and with certain aspects of document layout (cross-referencing excluded).

Ms Geraldine Coertze
Communications Consultant and Language Practitioner

Address: Pelham, Pietermaritzburg **Email:** gericoe@gmail.com **Cell:** +2782 620 2266

Individual member: South African Translators' Association
Associate member: Professional Editors' Guild (SA)
Entry-level member: Chartered Institute of Editing and Proofreading (UK)
Full member: South African Communication Association
Highest qualification: MSocSc (CCMS), UKZN