

Assignment

by Aliu Djalo

Submission date: 19-Oct-2025 11:23PM (UTC+0100)

Submission ID: 265703401

File name: 34332_Aliu_Djalo_Assignment_695044_1997657083.docx (86.72K)

Word count: 1764

Character count: 10662



10/19/2025

Crime Prediction and Visualization using Open-Source Machine Learning



Aliu Djalo (2117368)

UNIVERSITY OF WALES TRINITY SAINT DAVID

Table of Contents

1. Section One:	2
1.1. Introduction:	2
1.2. Background and Literature Review	2
1.3. Problem Statement.....	3
1.4. Aims	3
1.5. Objectives.....	3
1.6. Expected Outcomes	3
2. Section Two: Project Design.....	3
2.1. Project Overview	3
This project will focus on analysing and visualising crime data to identify patterns, hotspots, and possible future risks. It will be a development-based project that combines database design, data analysis and visualisation.....	3
2.2. Data Collection Methods.	4
2.3. Development Methods and Techniques.	4
2.4. Risk, Limitations and Ethical Issues.....	4
2.5. Data Analysis and Evaluation.	4
3. Section Three: Resources & Planning.	5
3.1.1. Resources Required.....	5
3.2. Project Timeline and Milestones.....	5
3.3. Risk, Limitations & Ethical Considerations.....	6
4. Conclusion	7

1. Section One:

1.1. Introduction:

Crime analysis is an important area of computing that helps police and communities understand where and when crimes happen. Most police forces collect large amounts of data, but it's often stored in separate systems or just shown on simple maps. This makes it hard to see deeper patterns or predict where future crimes might occur. [1]

In recent years, law enforcement and public safety agencies have begun adopting data-driven techniques, using tools such as GIS (Geographic Information Systems) dashboards, heat maps, and spatial clustering to plan patrols and allocate resources. Platforms like ESRI's ArcGIS Crime Analyst provide advanced mapping and hotspot analysis, but they are typically expensive and closed source. [2]

The UK's Police data portal offers public access to crime incident records, including spatial and temporal attributes. Yet these portals usually offer only retrospective views; maps, counts over time, filtering without predictive modelling or advanced clustering. [3]

This project will build an open-source, data-driven dashboard that analyses crime data using machine learning and Power BI visualisations to detect hotspots and forecast possible high-risk areas. It will also explore the impact of external factors such as weather or population density on crime trends, providing a richer analytical view. The system will be low-cost, transparent, and designed for smaller agencies and researchers who want to explore crime data more effectively.

1.2. Background and Literature Review

Recent research has shown that crime is not random; it happens in patterns linked to time, place, and social factors.

Studies show how clustering algorithms like DBSCAN can find "hotspots" of criminal activity, while models like Random Forest and XGBoost can predict where future incidents might occur. [4], [5]

Many current tools focus only on mapping past crimes, not predicting future ones, or require expensive licenses and technical expertise. Ethical concerns are also highlighted, as predictive algorithms may unfairly target certain communities due to biased historical data. [6]

Building on this research, this project aims to combine predictive analytics, hotspot detection, ethical awareness, and interactive visualisation in a single, open-source platform; something not yet widely available in existing crime analysis tools.

1.3. Problem Statement

Current public crime mapping tools are descriptive, not predictive. Smaller agencies and researchers need an affordable, open-source system that can detect hidden crime clusters and forecast high-risk areas using transparent machine learning methods.

1.4. Aims

To design, build and evaluate a user-friendly dashboard that detects and predicts crime patterns using SQL databases, Power BI visualisation and Python-based machine learning scripts.

1.5. Objectives

1. Design and implement a structured database (SQL or MySQL) to store and manage crime data effectively.
2. Develop and compare clustering algorithms (e.g., DBSCAN, K-Means) to detect crime hotspots.
3. Train and test a predictive model (e.g., Random Forest, XGBoost) to forecast crime risks.
4. Build and evaluate an interactive Power BI dashboard that visualises hotspots, trends, and prediction results in a user-friendly way.
5. Integrate and analyse external datasets (e.g., weather or demographic data) to explore their correlation with crime rates.

1.6. Expected Outcomes

- ✓ A working Power BI dashboard that maps, analyses, and predicts crime patterns.
- ✓ Integration of external data (weather, demographics) for deeper insights
- ✓ A clean, structured crime dataset stored in SQL.
- ✓ Evaluation results comparing models and visual performance.
- ✓ A final report discussing results, ethics, and usability.

2. Section Two: Project Design.

2.1. Project Overview

This project will focus on analysing and visualising crime data to identify patterns, hotspots, and possible future risks. It will be a development-based project that combines database design, data analysis and visualisation.

The system will be built around two main components:

1. Data System (MySQL or SQL Server) – to store and organise crime data including type, location, and date/time.

2. Power BI Dashboard – to create interactive reports and visualisations that show patterns, trends, and predictions.

To make the project more advanced, Python scripts will be integrated into Power BI to run predictive analysis, such as forecasting future crime trends or detecting patterns that are not visible in basic charts.

2.2. Data Collection Methods.

The project will use open-source datasets from the following sources:

- **UK Police Data Portal**: for crime records including location, type and time.
- **Office for National Statistics (ONS)**: for demographic data (population, income and area size).
- **OpenWeatherMap API**: for local weather data that might correlate with crime rates.

These datasets will be cleaned, combined and loaded into the SQL database. The inclusion of demographic and weather data will make predictions more realistic and add new analytical depth.

2.3. Development Methods and Techniques.

- Database Design: Structured tables will be created in MySQL (or SQL Server) to store crime, demographic and weather data.
- ETL Process: Python or SQL scripts will be used to clean and prepare data before loading it into Power BI.
- Power BI Integration: Interactive dashboards will be built with maps, charts, and trend lines.
- Predictive Analysis: Python scripts will be used within Power BI to run algorithms like Random Forest or time-series forecasting to estimate future crime trends.
- Comparative Analysis: Results will be compared between different cities (e.g., Cardiff vs Swansea) to see regional differences.

2.4. Risk, Limitations and Ethical Issues.

- Data Accuracy: Public datasets might have missed or outdated records. This will be managed through validation and cleaning.
- Bias: Predictive models could reflect biased policing patterns. Results will be clearly explained and interpreted responsibly.
- Privacy: All data will be publicly available and anonymised; no personal or identifiable information will be used.
- Technical Constraints: Power BI's Python integration has limits; only lightweight models will be used for efficiency.

2.5. Data Analysis and Evaluation.

Evaluation will include both technical and user-based assessment:

- Prediction Accuracy: Using metrics such as precision, recall and F1-score.
- Visual Effectiveness: Reviewing how clearly Power BI dashboards present information.

- User Feedback: Gathering informal feedback from classmates or mentors on usability and clarity.

3. Section Three: Resources & Planning.

3.1.1. Resources Required

Category	Details
Hardware	A personal laptop or desktop with at least 8GB RAM and stable internet connection (for database hosting, Power BI processing, and Python model training).
Software	Power BI Desktop: to design and visualise crime data dashboards and hotspot maps. MySQL / SQL Server: for data storage, cleaning, and integration. Python (with libraries like scikit-learn, pandas, and matplotlib): for clustering and predictive modelling
Data Source	UK Police Data Portal: for official UK crime records. Kaggle Datasets: for testing and comparing model performance. OpenWeather API or ONS Demographic Data: for external variables (e.g., weather, population).
Academic Resources	Access to library databases such as IEEE Xplore, ScienceDirect, and SpringerLink for research on crime analytics, clustering, and predictive policing.

3.2. Project Timeline and Milestones

The project will be completed over approximately 18 weeks, following a structured timeline aligned with key deliverables and testing phases.

Weeks	Phase	Deliverables
Weeks 1 to 3	Project Planning & Literature Review	Finalise proposal, gather references, and complete literature review on crime

		prediction models and visualisation tools.
Weeks 4–6	Data Collection & Database Design	Collect open-source crime data, clean and import into SQL database (MySQL), and test connection with Power BI.
Weeks 7–9	Exploratory Data Analysis (EDA)	Use Power BI and Python (pandas, matplotlib) to identify initial patterns and correlations.
Weeks 10–12	Machine Learning Model Development	Implement clustering (DBSCAN/K-Means) and predictive model (Random Forest/XGBoost). Test model accuracy and visualize results.
Weeks 13–14	Dashboard Development	Build an interactive Power BI dashboard showing hotspots, trends, and prediction outputs. Integrate external data (weather/demographics).
Weeks 15–16	Evaluation & Testing	Conduct user testing (usability feedback) and evaluate model performance using metrics like F1-score and Silhouette Score.
Weeks 17–18	Report Writing & Final Submission	Complete final report, including ethical considerations, discussion, and reflection.

3.3. Risk, Limitations & Ethical Considerations

- Data Limitations: Some police data may have missing fields or inconsistencies. The author will clean and validate data before analysis.
- Ethical concerns: Only anonymised public data will be used. The author will include a reflection on possible bias in predictive policing models.
- Technical risk: Potential compatibility issues between SQL and Power BI. The author will maintain data backups in CSV format to prevent data loss.

4. Conclusion

This project will build a practical, low-cost system to help identify and predict crime patterns. By combining SQL, Python, and Power BI, it aims to turn raw data into clear, useful insights. The final dashboard will not only show past crime trends but also forecast future risks, helping improve safety and resource planning.

References:

- [1] An-Chan Phung. (24th February). *Effective Crime Prevention: Data & Analytics in UK Policing*. Available: <https://www.civica.com/en-us/insights/effective-crime-prevention-connecting-data-analytics-in-uk-policing/>.
- [2] Scott Oppmann. Scott Oppmann. ArcGIS solutions 2024 product roadmap. ArcGIS solutions 2024 product roadmap. 2024. Available: <https://www.esri.com/arcgis-blog/products/arcgis-solutions/announcements/arcgis-solutions-2024-product-roadmap>.
- [3] Anonymous . *Welcome to data.police.uk*. Available: <https://data.police.uk/>.
- [4] Rashid Ahmad 1 , Asif Nawaz 1 , Ghulam Mustafa 1 , Tariq Ali 1 , Mehdi Tlija 2 , Mohammed A. El-Meligy 3 4 , Zohair Ahmed 5. Rashid Ahmad 1 , Asif Nawaz 1 , Ghulam Mustafa 1 , Tariq Ali 1 , Mehdi Tlija 2 , Mohammed A. El-Meligy 3 4 , Zohair Ahmed 5. CHART: Intelligent crime hotspot detection and real-time tracking using machine learning. *Computers, Materials and Continua* 8181(3), pp. 4171–4194. pp. 4171–4194. 2024. Available: <https://www.sciencedirect.com/org/science/article/pii/S154622182400852X>. DOI: 10.32604/cmc.2024.056971.
- [5] E. Cesario, P. Lindia and A. Vinci. E. Cesario, P. Lindia and A. Vinci. Multi-density crime predictor: An approach to forecast criminal activities in multi-density crime hotspots. Multi-density crime predictor: An approach to forecast criminal activities in multi-density crime hotspots. *Journal of Big Data Journal of Big Data* 1111(1), pp. 75. pp. 75. 2024. Available: <https://doi.org/10.1186/s40537-024-00935-4>. DOI: 10.1186/s40537-024-00935-4.
- [6] Kiana Alikhademi · Emma Drobina · Diandra Prioleau · Brianna Richardson · Duncan Purves · Juan E. Gilbert. Kiana Alikhademi · Emma Drobina · Diandra Prioleau · Brianna Richardson · Duncan Purves · Juan E. Gilbert. Untitled. Untitled. 2019. Available: <https://par.nsf.gov/servlets/purl/10276841>.

Assignment

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

62/100

PROJECT PROPOSAL – ACCA6020 INDEPENDENT PROJECT

Student: Aliu
Djalo (2117368)
Mark _____

62

Intro / A&O (400 words)		
Project title; Identification of a specific problem to be investigated; Well-defined aims and SMART objectives; Explaining why this project is worth doing; Discussion of what the project needs to do in order to be a success; Description any deliverables that will be produced at the end of the project	20	13
Specific well identified title. Define measurable success criteria for the final system and also connect it to real world impact like improve public		

safety, resource allocation etc		
Background to the Project / Literature Review (400 words)		
Description of how the work to be done by in the project relates to the work identified in the literature review; Discussion of any ethical issues raised by the project Identifying what has already been done within the field; Topic familiarity; Brief literature review using appropriate sources – 6-8 journal papers and book chapters rather than just websites; Selection of relevant models & concepts You must cite 6-8 high-quality academic sources and need to reach ~400 words of critical synthesis. Need to state how your project relates to and builds upon this existing work, leading directly to your problem statement.	20	10
Project Design (500 words)		
Clear statement of which development methods are to be used and why they are appropriate; Clear statement of which tools are to be used and why they are appropriate; Clear description of how data collected will be analysed	30	18

Mentioning the integration of Python scripts within Power BI shows an understanding of advanced analytics. Should try to integrate the evaluation plan into the "Project Design" section to make it a comprehensive methodology.		
Resources / Planning (200 words)		
List of any deliverables; Clear diagram showing both timing and resource use; Completed ethics form; Limited and relevant appendices Your table is a good start, however Gantt Chart is recommended.	15	11
Document Presentation		
Correct spelling, grammar and punctuation; Flow of narrative/fluency; Coherence; Word processed ; Correct	15	10

use of referencing with bibliography; Strict word limit		
The proposal is professionally presented. The duplication of the "Risks" section breaks the flow. Consolidating it will improve coherence.		
TOTAL	100	62

PAGE 1
PAGE 2
PAGE 3
PAGE 4
PAGE 5
PAGE 6
PAGE 7
PAGE 8