

Crime Analyzer

User Guide for Tourist Safety Application

Ferdinando Muraca, Carlo Vincenzo Stanzione

September 4, 2025

Contents

1	Executive Summary	3
1.1	System Overview	3
1.2	Key Capabilities	3
1.3	Target Users	3
2	Quick Start Guide	3
2.1	Getting Started in 3 Steps	3
2.2	Interpreting Your Result	4
2.3	Example Usage Scenario	4
3	How the System Works	5
3.1	System Overview	5
3.1.1	Core Components	5
3.2	Input Specification	5
3.2.1	Required Inputs	5
3.2.2	Derived Information	5
3.3	Output Specification	5
4	Data and Analysis	6
4.1	Data Sources	6
4.1.1	Primary Data Sources	6
4.2	Data Processing	6
5	Model Performance	6
5.1	Performance Highlights	6
5.2	Model Validation	6
5.3	Illustrative Outputs	7
6	Understanding the Predictions	8
6.1	Explainable Predictions	9
6.2	Pattern Analysis	9
6.3	Glossary	9

7	Responsible Use and Ethics	9
8	System Limitations	9
8.1	Known Constraints	10
8.2	Risk Mitigation	10
9	FAQ	10
10	Future Enhancements	10
11	Conclusion	11

1 Executive Summary

1.1 System Overview

The Crime Analyzer is an intelligent system designed to assess criminal risk for tourists in New York City. By analyzing historical crime data from the NYPD with advanced techniques, the system provides a clear risk assessment (High Risk or Low Risk), a confidence score for its prediction, and an explanation of the factors contributing to the assessment.

1.2 Key Capabilities

- Real-time safety assessment based on location, time, and user profile
- Clear explanations for each risk assessment, showing which factors are most important
- High overall accuracy with specialized handling of high-risk scenarios
- Analysis of crime patterns to reveal trends and associations

1.3 Target Users

- **Tourists:** To assess personal safety and increase awareness of local risks.
- **Individuals interested in buying or renting a home:** To check whether the address of a property is located in a safe area.
- **Other users:** Anyone who wants to know the risk level of a specific area for personal or planning reasons.

2 Quick Start Guide

2.1 Getting Started in 3 Steps

1. Provide Your Information:

- Allow the application to access your location (GPS coordinates).
- Enter basic information about yourself (age group, sex, race).
- The system automatically uses the current time.

2. Receive Your Assessment:

- The system will immediately provide a **HIGH RISK** or **LOW RISK** classification.
- You will also see a confidence score, indicating how certain the system is about its prediction.
- The key factors that influenced the result will be explained.

3. Make Informed Decisions:

- Use the result as one of several factors in your safety planning.

- Pay special attention to predictions where the confidence is not very high (e.g., between 40% and 60%).
- Consider the local crime patterns provided with your result to better understand the context.

2.2 Interpreting Your Result

- **LOW RISK:** Historical data suggests favorable conditions. However, it is always important to follow common safety practices.
- **HIGH RISK:** The system has identified indicators of increased risk. It is advisable to prefer well-lit, populated areas and consider adjusting your travel plans.
- **Confidence Score:** This score reflects the model's certainty. A result with low confidence should be interpreted with caution.
- **Context Matters:** Review the trends provided to understand which factors (time, location, surroundings) are most influencing the risk.

Recommended Actions by Confidence

Classification	Confidence band	Suggested action
LOW RISK	60–100%	Proceed, remain aware of surroundings.
LOW RISK	40–60%	Proceed with caution; check contextual trends.
HIGH RISK	60–100%	Prefer well-lit, populated routes; consider alternatives.
HIGH RISK	40–60%	Increase caution; avoid isolated areas; re-check later.

2.3 Example Usage Scenario

Scenario: A tourist is planning an evening out and wants to assess the risk of visiting Times Square around 9 PM.

- **Input:** They provide their location (Times Square), the time (9 PM), and their personal information.
- **Assessment:** The system returns a **HIGH RISK** classification with a confidence score of 71%.
- **Explanation:** The key contributing factors are the late hour and the high concentration of bars in the area.
- **Context:** The system also highlights relevant crime trends for Manhattan, such as a strong association between certain age groups and suspect profiles.
- **Decision:** Aware of the specific risk factors, the tourist decides to stay in more crowded and well-lit areas.

3 How the System Works

3.1 System Overview

The Crime Analyzer uses a sophisticated pipeline to deliver real-time crime risk assessments. The architecture is built to enterprise-grade standards, ensuring reliability and scalability.

3.1.1 Core Components

- **Data Engine:** Integrates historical NYPD complaint data and enriches it with geographic and temporal features.
- **Prediction Model:** A machine learning model that calculates the risk level based on the input data.
- **Inference Engine:** Provides the final risk assessment and confidence score.
- **Explainability Module:** Identifies and presents the key factors influencing the prediction.
- **Pattern Analysis:** Discovers underlying crime trends and patterns from the data.

3.2 Input Specification

The system requires the following information to generate an assessment:

3.2.1 Required Inputs

- **Geographic Coordinates:** Latitude and Longitude.
- **User Profile:** Age Group, Race, and Sex, based on standardized categories.

3.2.2 Derived Information

From the user's coordinates and the current time, the system automatically generates dozens of additional data points for a more accurate analysis. These include:

- **Temporal Information (e.g., Hour, Day of the Week, Season):** To understand time-based patterns.
- **Geographic Information (e.g., Borough):** To consider location-specific trends.
- **Points of Interest (POI):** Information about nearby bars, metro stations, schools, etc., to assess the environmental context.

3.3 Output Specification

The system provides a clear and concise output for each assessment:

- **Primary Output:** A binary classification of HIGH RISK or LOW RISK.
- **Confidence Score:** A percentage indicating the model's confidence in its prediction for the HIGH RISK class.

- **Contextual Trends:** The system enriches the response with relevant crime patterns for the specific time and location, helping the user understand the broader context.

4 Data and Analysis

4.1 Data Sources

The system is built on a foundation of reliable and comprehensive data to ensure the quality of its assessments.

4.1.1 Primary Data Sources

- **NYPD Complaint Data:** Historical and current crime reports from the New York City Police Department.
- **Geographic Data:** Information on Points of Interest, such as transit stations, commercial establishments, and schools.

4.2 Data Processing

To ensure accuracy, the raw data undergoes a rigorous preparation process. This includes cleaning the data, handling missing values, and integrating information from different sources. This preparation is crucial for the model to learn effectively and make reliable predictions.

5 Model Performance

The Crime Analyzer's prediction model was chosen after extensive testing and comparison with over 15 other algorithms. The final model demonstrates a strong balance of performance, reliability, and interpretability.

5.1 Performance Highlights

On a hold-out test set, the model achieved:

- **96.5% Accuracy:** The model is highly accurate in its overall predictions.
- **Strong Discriminative Ability:** The model is very effective at distinguishing between high-risk and low-risk situations.
- **Balanced Performance:** While no model is perfect, this system is tuned to provide a practical balance between identifying potential risks and avoiding excessive false alarms.

5.2 Model Validation

The model's performance has been rigorously tested to ensure it is both stable and reliable. It has been validated to perform consistently across different data samples, which indicates that it can generalize well to new, unseen situations. This robustness is critical for a system designed for real-world use.

5.3 Illustrative Outputs

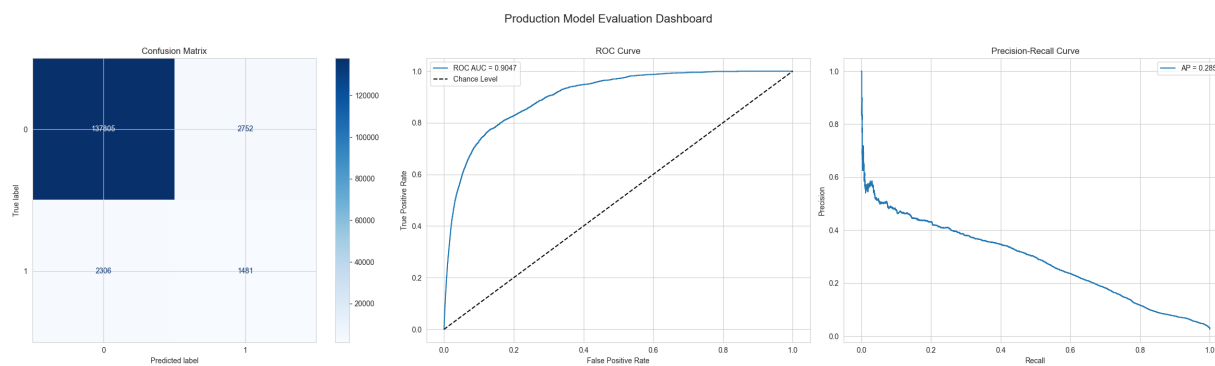


Figure 1: Model evaluation overview on the hold-out set.

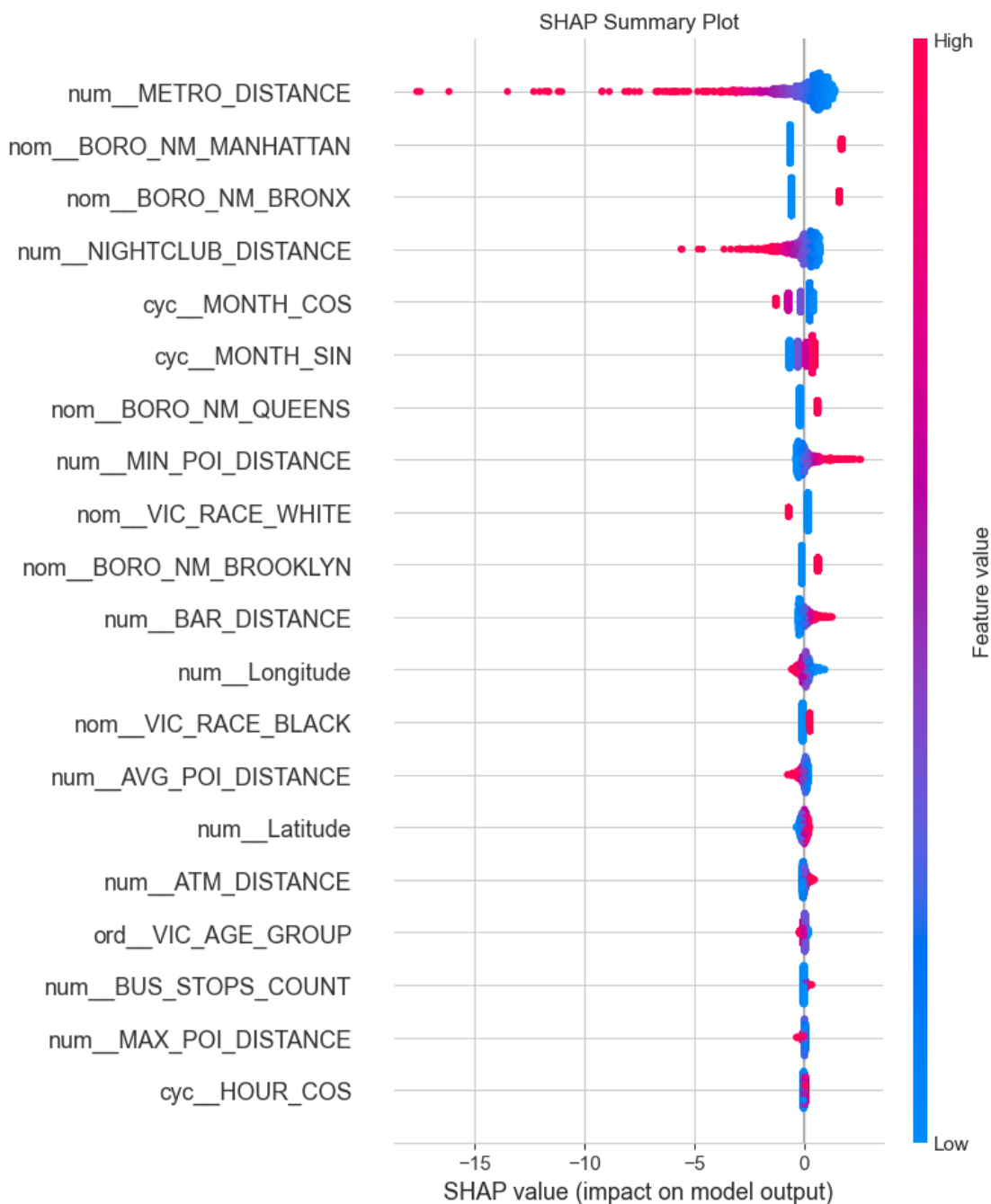


Figure 2: Explainability summary highlighting feature contributions (SHAP).

6 Understanding the Predictions

Explainability is a core principle of the Crime Analyzer. The system is designed not only to provide an assessment but also to explain the "why" behind it.

6.1 Explainable Predictions

For every prediction, the system identifies the key factors that influenced the outcome. For example, it might highlight that the time of day or the proximity to a certain type of location was a major contributor to a "High Risk" assessment. This helps users understand the context of their situation and make more informed decisions.

6.2 Pattern Analysis

In addition to individual predictions, the system analyzes the data to uncover broader crime patterns and trends. For instance, it might identify that certain types of crime are more common in specific boroughs or at particular times of the day. These insights are provided alongside the risk assessment to give users a more complete picture of the safety landscape.

Example insight (Evening time bucket):

--- Top-K Rules for TimeBucket = EVENING (172,881 records) ---

IF (SUSP_AGE=25-44) THEN (SUSP_SEX=M)

- Kulc: 0.624, Conf: 78.95%, Lift: 1.43, Support: 0.2538, Score: 0.314

IF (VIC_SEX=F) THEN (HAS_POI=NO)

- Kulc: 0.563, Conf: 74.27%, Lift: 1.14, Support: 0.2495, Score: 0.281

IF (LAW_CAT=FELONY) THEN (HAS_POI=NO)

- Kulc: 0.505, Conf: 66.66%, Lift: 1.03, Support: 0.2227, Score: 0.238

6.3 Glossary

- **Confidence score:** Model-estimated probability for the HIGH RISK class.
- **POI:** Points of Interest (e.g., bars, schools, transit stops) near a location.
- **SHAP:** A method to attribute each feature's contribution to a specific prediction.

7 Responsible Use and Ethics

- **Fairness awareness:** Inputs like age group, sex, and race mirror categories in source data and are used to reflect observed patterns, not to ascribe individual traits. Historical data can encode societal biases; interpret outputs accordingly.
- **Appropriate use:** The tool supports personal safety awareness for civilians and planners. It is not intended for law enforcement decision-making about individuals.
- **Transparency:** Explanations are surfaced to help users understand drivers of the assessment.
- **Human judgment:** Always complement outputs with local knowledge and situational awareness.

8 System Limitations

It is important to understand the limitations of the Crime Analyzer to use it effectively and responsibly.

8.1 Known Constraints

- **Based on Historical Data:** The model's predictions are based on past crime data and may not reflect sudden, real-time events.
- **Data Gaps:** The system may be less effective in areas with limited historical crime data.
- **Not a Crystal Ball:** The predictions are probabilistic estimates, not guarantees of future outcomes.

8.2 Risk Mitigation

To ensure responsible use, the system is designed with several safeguards:

- **Regular Updates:** The model is periodically retrained with the latest data to keep its knowledge current.
- **Clear Communication:** The system's limitations are clearly communicated to users.
- **User Discretion Advised:** Users are encouraged to use the assessment as one of many tools in their decision-making process and to always remain aware of their surroundings.

9 FAQ

1. **Does a LOW RISK result mean I'm safe?** No. It indicates favorable conditions historically, not a guarantee.
2. **Why is the confidence sometimes around 50%?** Inputs may be ambiguous or the situation similar to both classes; treat with extra caution.
3. **Do you track my identity?** The system uses the inputs you provide to compute the assessment. Data handling follows the deployment's privacy policy.
4. **Can I use this for cities other than NYC?** This version is trained on NYC data; broader coverage is in the roadmap.
5. **What if the result seems off?** Re-check location accuracy, time, and inputs; review contextual trends and act conservatively.

10 Future Enhancements

The Crime Analyzer is committed to continuous improvement. Future plans include:

- **Enhanced Features:** Incorporating more sophisticated data, such as real-time event information.

- **Interactive Visualizations:** Developing web-based maps to explore risk areas interactively.
- **Expansion:** Adapting the system for use in other major cities.

11 Conclusion

The Crime Analyzer is a powerful tool for enhancing personal safety. By providing accurate, explainable, and context-aware risk assessments, it empowers users to make smarter, safer decisions. Its applications extend beyond tourism to urban planning, law enforcement, and other areas where data-driven insights can contribute to a safer society.