Predictive Analysis for Campus Safety: Modeling Burglary Risk at Atlanta's Major Universities

Data Science Team (Alphabetically):
Gunn Madan, Harini Mohan, Joshua Piña, Yuntian Wu

Abstract: Routine Activity Theory teaches that "Crime requires a motivated offender, a suitable target, and the opportunity". Per the 2024 U.S Census, the Atlanta Metropolitan area is the 8th largest Metropolitan area in the United States. Students deciding to attend university in large metro areas must consider the local crime rate near the university. This project focuses on analyzing burglaries and burglary related crime in areas around major college campuses in Atlanta, with the goal of forecasting risk levels and hotspots that are safety concerns for students. This study will center around 4 major campuses, Georgia State University, Georgia Tech, Clark Atlanta University, and Spelman College. We will utilize the Atlanta Police Department's (APD) Open Data Portal to access offense type and the time and coordinates of each offense. To ensure specificity and relevance, we will use spatial filters to only include incidents that occur within a 1-mile radius of each campus, where we predict that students will be disproportionately impacted.

Drawing from our experiences on Atlanta's urban campuses, we see firsthand how student safety concerns continue to rise alongside crime rates. Burglary is becoming one of the biggest safety worries for college students as crime rates in cities like Atlanta continue to grow. In addition to causing material damage, burglaries raise students' feelings of insecurity which can influence academic concentration, enrollment choices, and general trust in their schools. By focusing on Atlanta's largest campuses, we hope to shed light on both statistical patterns and the practical effects of the safety issues that thousands of students must deal with while living and studying there. Crime patterns also fluctuate by time of day, neighborhood, and the academic calendar, indicating that our study must include spatiotemporal dynamics. We aim to apply our technical data science skills to build an accessible, student-focused application. We will construct a data pipeline to extract, clean, and transform our APD-generated data to create spatiotemporal features. The modeling phase will utilize predictive time series regression to forecast areas of higher risk and assist students in making safer decisions. Geographic Information Systems (GIS) modeling will allow us to dynamically map and forecast high-risk hotspots. We will use standard regression metrics such as RMSE to ensure model reliability. Additionally, the models' ability to detect actual burglary risk hotspots will be assessed using classification metrics including accuracy, recall, and F1-score. This dual strategy gives users assurance that our method is reliable for forecasting and practical in real-world decision making. For the project management aspect of this project, we will be following the CRISP-DM framework in an agile structure that utilizes four concise sprints that align with our assigned project checkpoints.

Upon successful completion of this project, we will be able to deliver three key outcomes. First, an ETL pipeline that allows for the automatic retrieval of APD data and ingestion into our database and model to ensure relevant and up-to-date data. Second, a validated model capable of forecasting burglary risk level to identify high-risk hotspots within a 1-mile radius of our four target campuses, backed by metrics for both classification and regression to guarantee dependability. Third, an interactive dashboard that utilizes GIS to highlight higher areas of risk in an effort of crime reduction. Together, these results will allow institutions to carefully spend safety resources while also empowering students to make safer decisions in their everyday lives. Parents and families will have more information about safety circumstances outside of the campus community, and local officials may use our results to guide more comprehensive crime prevention initiatives. By combining theoretical precision with practical implementation, our initiative guarantees to make a significant contribution to data science practice and public safety.