



Predictive Analysis for Campus Safety: Modeling Burglary Risk at Atlanta's Major Universities : **Checkpoint 1**

Alphabetically: Gunn Madan, Harini Mohan, Joshua Piña, Yuntian Wu



What are we trying to solve? - Part 1 Project

Introduction/Scoping

- Burglary related crimes frequently occur near college campuses, posing safety risks to students.
- However, **universities currently lack predictive tools** to anticipate high-risk periods and emerging crime hotspots.
- This project addresses that gap by building a data driven solution focused on Atlanta's major campuses.




What exactly is our goal? - Part 1 Project

Introduction/Scoping

To design and implement a predictive modeling and visualization system that:

- Forecasts **weekly burglary risk** around major Atlanta campuses.
- Identifies **spatial crime hotspots** within a one-mile radius of each campus.
- Delivers actionable insights through an **interactive dashboard** to support proactive safety strategies.



How will we know we have reached our goal? - Part 1

Project Introduction/Scoping

- Capture $\geq 70\%$ of actual incidents within predicted hotspot zones.
- Achieve $\geq 70\%$ accuracy in predicting high-risk weeks.
- Keep MAPE (Mean absolute percentage errors) below 15% for weekly burglary forecasts.



Tools? - Part 2 Infrastructure & Tool Setup

| Component | Tool | Why? |
|---------------------|---|---|
| Communication | Slack | Slack allows for easy integration of external apps. |
| Project Management | Notion /GitHub Projects | Notion - overall organization, GHP - to quickly raise issues |
| Version Control | Git/ GitHub | Repository hosted on GitHub |
| Documentation | Google Drive | Easy to utilize online |
| Experiment Tracking | W&B | Professor Rec. |
| Database | PostgreSQL w/ PostGIS | Best dbms for our data and GIS |
| Frontend | Streamlit | Easy application setup |
| Backend | FastAPI w/Pydantic | Postgres & Streamlit integration |



Tools? - Part 2 Infrastructure & Tool Setup

Additionally,

- Architecture:
 - LAMBDA: Batching for training with a streaming layer to maintain an evergreen state.
- Deployment:
 - Render to host the application; Render Cron + Prefect for scheduling
 - GitHub Pages serving as a static site
- Python Libraries Utilized:
 - [Std] Pandas, Numpy, Matplotlib, Seaborn
 - [ML] Scikit-Learn, Pytorch, XGBoost, StatsModels, Imblearn
 - [Spec] GeoPandas, Pandera, SQLAlchemy, GeoAlchemy2



Business Understanding - Part 3 Business & Data Understanding

Objective : Predict Burglary related crime risk around Atlanta's major college campuses (Georgia State, Georgia Tech, Spelman, and Clark Atlanta) by forecasting weekly incident counts and identifying spatial hotspots.

DS Goals

- Ingest, clean, and transform APD crime data.
- Forecast weekly burglary counts per campus using ML models
- Detect and visualize hotspot areas using spatial clustering
- Classify “high-risk” weeks based on incident history.

Business Understanding - Part 3 Business & Data Understanding

Sprint 1 : Business Understanding & EDA

- Define objectives and success criteria.
- Collect APD crime data and perform exploratory analysis time period.
- Compare DOE data
- Deliverables: EDA report + modeling plan.
- Business & Data Understanding

Sprint 2 : Data Prep & Feature Engineering

- Clean and preprocess data (offense categories, dates, missing values).
- Filter to 1-mile campus zones and engineer temporal/spatial features.
- Deliverable: Final dataset and feature set for modeling.
- Data Prep

Sprint 3 : Modeling & Evaluation

- Forecast weekly crime counts near each campus using past data and trends.
- Classify “high-risk weeks” by predicting whether burglary levels will exceed a set threshold.
- Identify hotspot areas on a map using clustering and heatmaps.
- Deliverable: Predictive models, hotspot maps, and evaluation results.
- Modeling & Evaluation

Sprint 4 : Deployment & Dashboards

- Develop dashboard with forecasts, hotspots, and risk probabilities.
- Document workflow and prepare final presentation.
- Deliverable: Dashboard, GitHub repo, final report.
- Deployment



Data Acquisition - Part 3 Business & Data Understanding

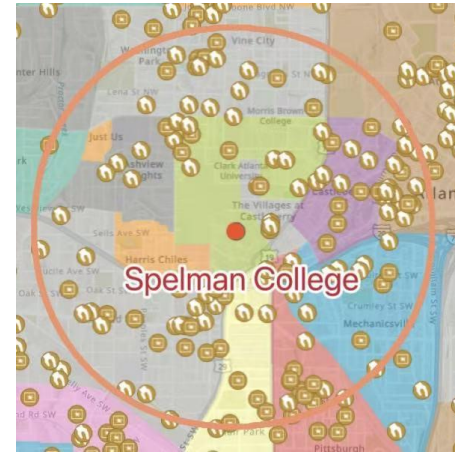
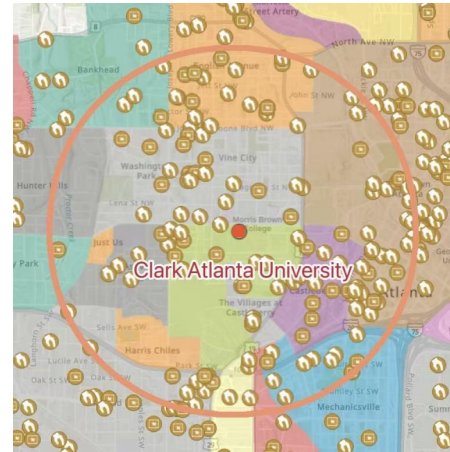
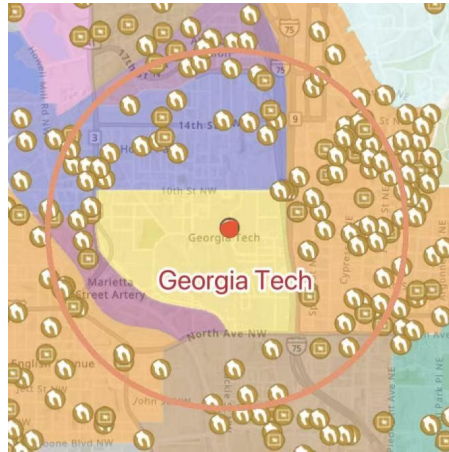
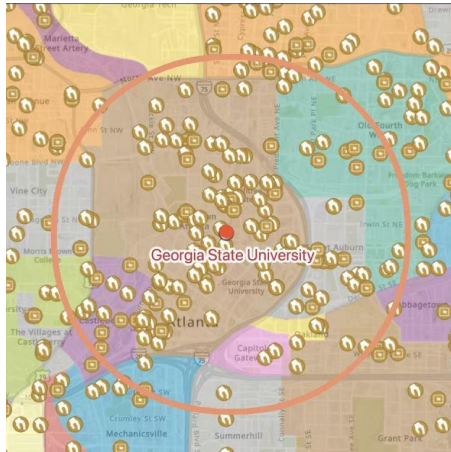
Our Primary Data Sources

- Atlanta Police Department
- US Department of Education
- University Annual Reports

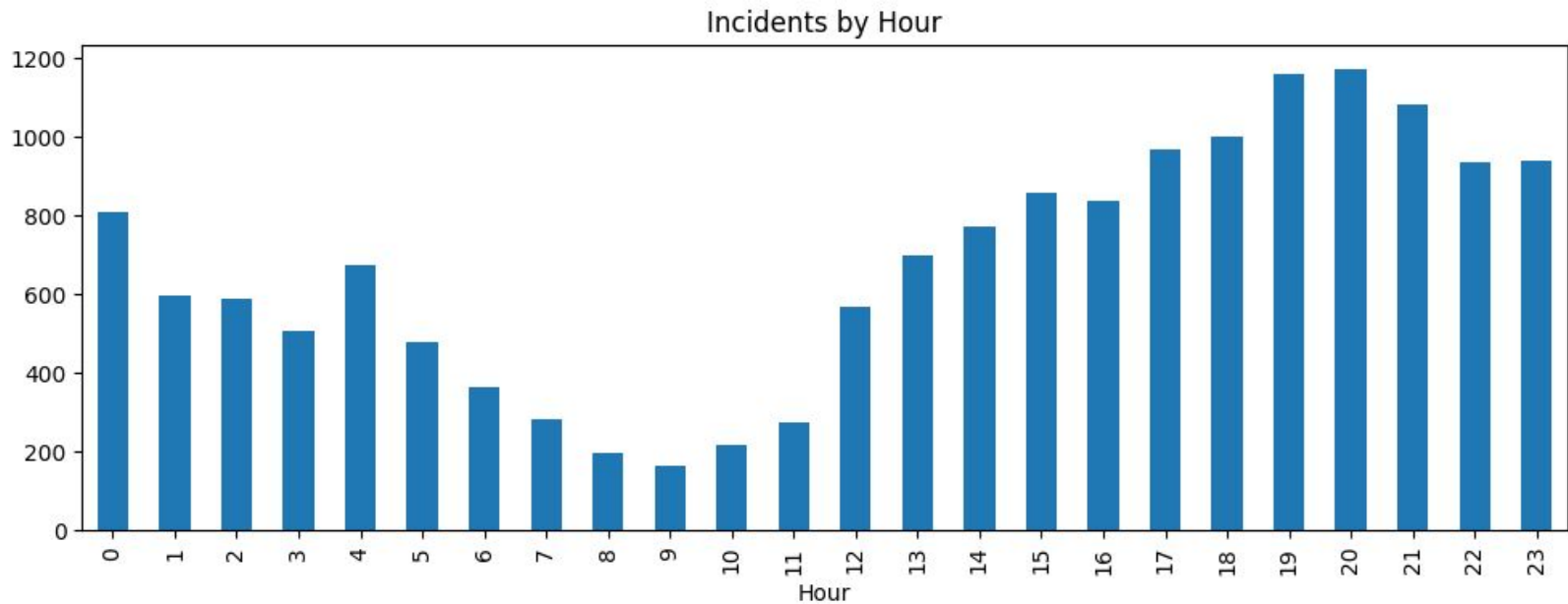
Data Links

- [Atlanta Police Department \(APD\) - Crime Maps](#)
- [Atlanta Police Department - Open Data Portal \(APD-ODP\)](#)
- [GT Police Department - Crime Logs](#)
- [US Department of Education - Campus Safety & Security \(CSS\)](#)
- [US Department of Education - CSS \(Comparison Tool\)](#)

One-Mile Radius of Each Campuses.- Part 3 Business & Data Understanding

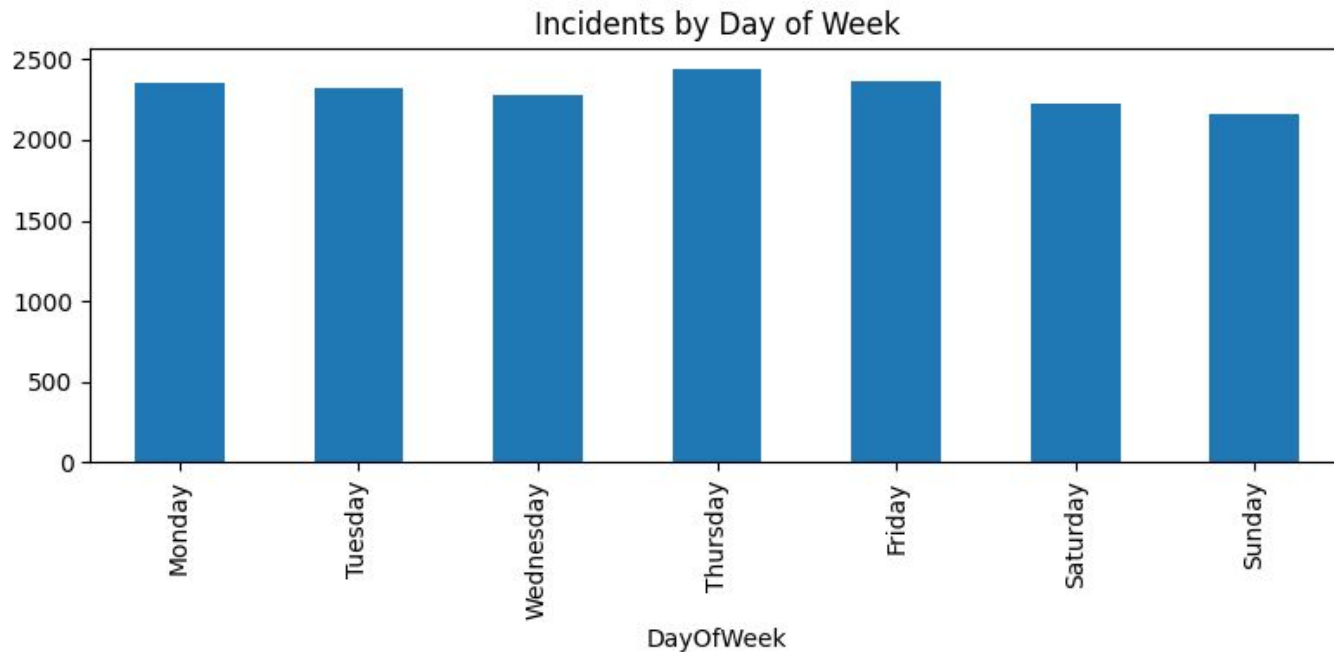


Exploratory Data Analysis - Part 3 Business & Data Understanding



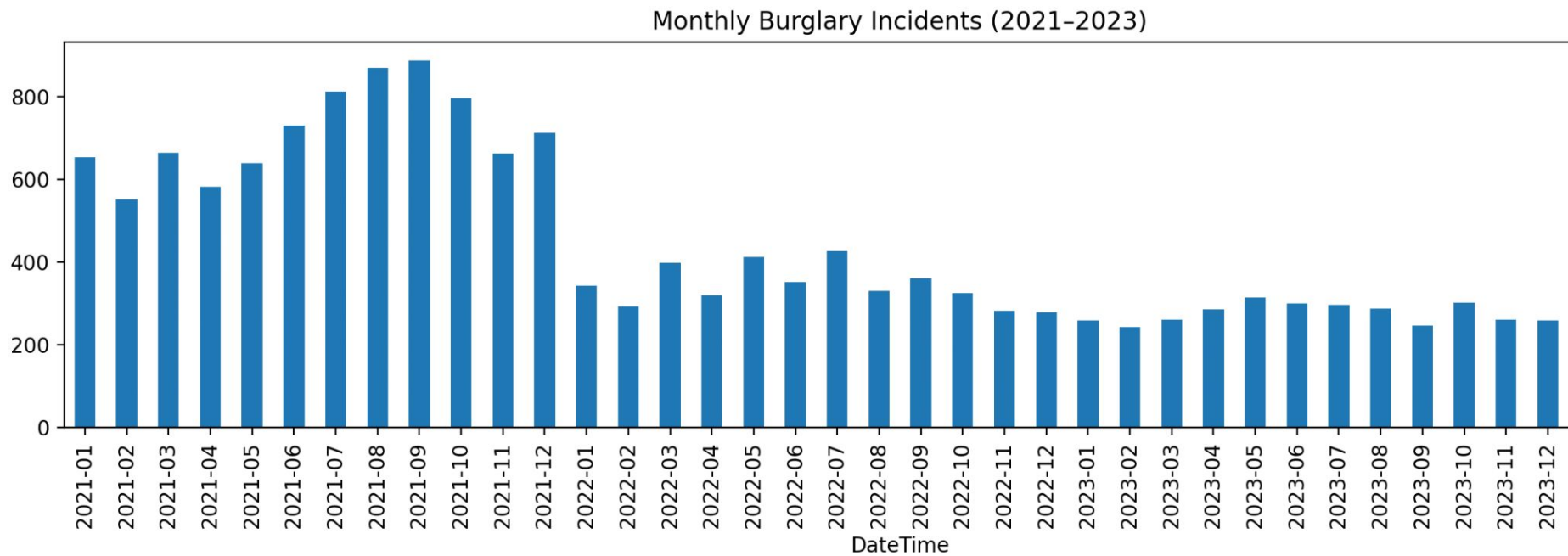
Based on APD data

Exploratory Data Analysis - Part 3 Business & Data Understanding



Based on APD data

Exploratory Data Analysis - Part 3 Business & Data Understanding



Based on APD data

Data Quality Report

DOE data differs from APD data, further emphasizing the need for the distinction between the two.

| Check | Issue Found | Action Taken |
|------------------------|---|--|
| Missing value | Some rows missing coordinates, dates, or campus fields | Dropped incomplete rows; verified remaining entries align with APD and DOE records |
| Irrelevant Fields | DOE dataset contained non-crime categories (disciplinary, hate, fire, VAWA) | Removed all unrelated fields and kept only Criminal Offenses and Arrests |
| Duplicates | Repeated APD incidents with same ID and timestamp | Deduplicated based on report_number, offense_date, and location. |
| Formatting Differences | Columns, date formats, and field names inconsistent between APD and DOE | Renamed columns and standardized date format |
| Out-of-Bounds Data | Some APD coordinates outside Atlanta or invalid (0,0) | Filtered to points within Atlanta city boundary |
| Temporal Range | DOE data annual; APD data daily | Aggregated APD to yearly totals for alignment |



Initial Findings - Part 3 Business & Data Understanding

- **Crime Concentration-** Georgia State University and Georgia Tech show the highest volume of burglary and arrest incidents within a 1-mile radius, while Spelman and Clark Atlanta remain consistently lower.
- **Temporal Trend-** Burglary incidents dipped in 2022 but rose again in 2023, aligning with DOE's annual reports- likely linked to post-COVID campus repopulation.
- **Time + Day Patterns-** Most incidents occur (Tue-Thu) and between 10 AM- 4 PM, suggesting daytime, opportunity-based offenses near student housing and campus centers.
- **Spatial Hotspots-** APD data clusters heavily in downtown Atlanta, overlapping the GSU and Georgia Tech zones, validating the need for localized predictive modeling.