

Jingjing Ye, Guangyu Min, Ziheng Xiao

jingjing@gatech.edu, gmin8@gatech.edu, zxiao76@gatech.edu

Introduction

- In our project, we gathered crime data from the Atlanta police department website and used machine learning model to generate the prediction and present visualization on Google Maps. We made use of temporal component to generate predictions for different periods in a day. The main method is spatio-temporal kernel density estimation (STKDE).
- Our analysis on the crime data could provide an immediate result and the prediction model are applied to get the result. The prediction and visualization may be needed to help policymakers and relevant departments understand the challenges and problems in the field of crime control system. It could also potentially be helpful for route planning and road safety.

Data Collection and Preparation

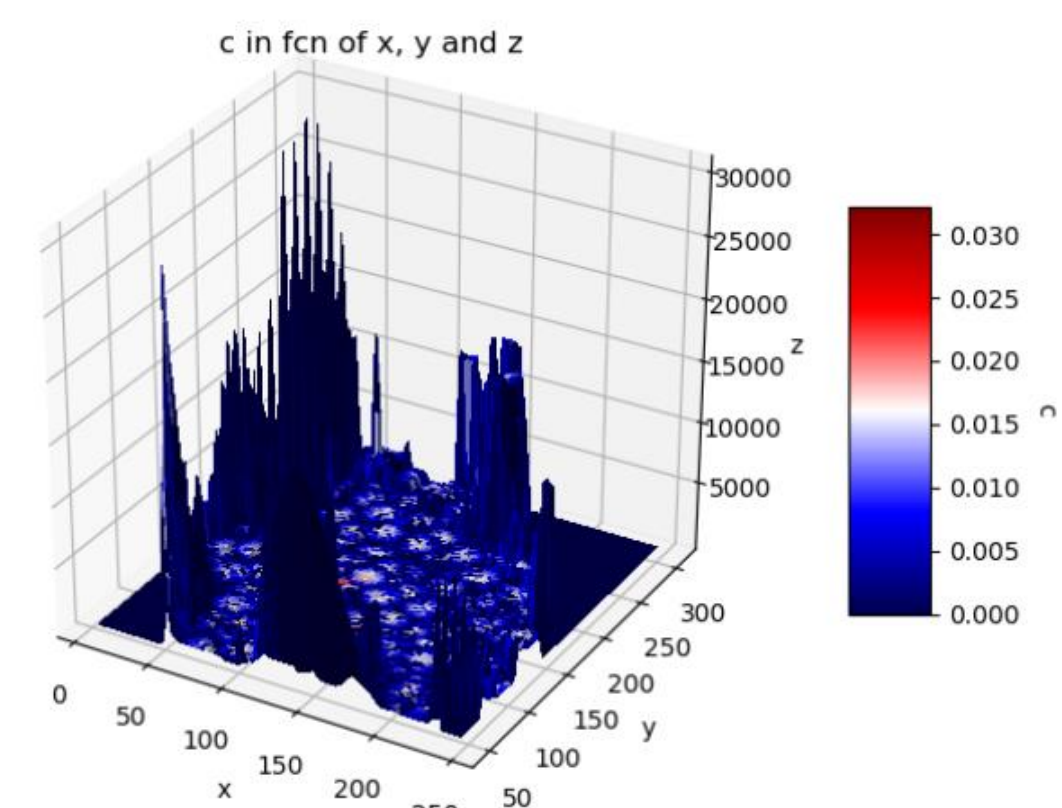
- Crime Data:**
 - Raw crime data is retrieved from Atlanta Police Department website (Crime Data) and Geographical map-ping data will be provided by Atlanta Department of City Planning (Geographical Mapping Data). The data have a size of 65MB around 380,000 records after data cleaning, and time range from 2009 to 2021.
- City Map:**
 - Geographical mapping data will be used for interactive visualization of crime prediction as well. Such data are fetched from GIS Resource provided by Atlanta Department of City Planning (Atlanta Geographical Mapping Data) in the form of GeoJSON files.
- Google Maps:**
 - We built the map background from [Google Maps JavaScript API](#). The map provides location by precise latitude and longitude to match with locations in the crime data.

Methods/Approach

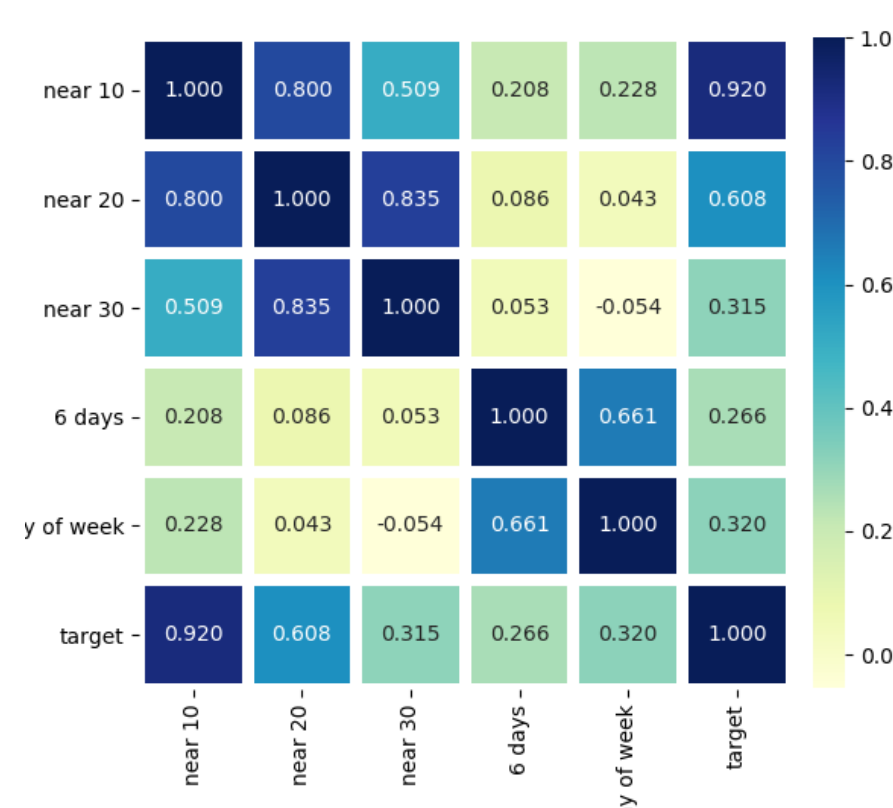
- Data Processing:**
 - We applied STKDE method to calculate the normalized density (probability) for each 3D grid combining location and timestamp.
- Prediction:**
 - In this project, Multilayer Perceptron (MLP), which consists of an input layer, three hidden layer and an output layer.
- Visualization:**
 - We imported the map using Google Maps API and draw the heatmap using latitude-longitude coordinates. For different data, the radius may need to be adjusted for a proper layout. We provided history aggregated information, as well as the predicted probabilities to show in the heatmap.
 - We also provide some options for interaction, such as the slider to choose a date, the dropdown to choose a hour range, and the click event to show a brief summary for each grid.

Experiments/Results

- Visualization:**
 - From the visualization, we could successfully observe a clear comparison between the real crime data and our prediction result for the same day.
 - The click option will show a brief summary for the chosen grid. The grid size is much smaller than a neighborhood such that more accurate prediction range can be given. This may also increase the processing and model training work. We defined a risk level table to inform people the risk for that region at a specific day. We assumed all crime accidents have a similar level of detriment.
- Prediction:**
 - In actual experiment, we use 2018-2020 three years data to predict the 2021 crime probability. Five features are extracted from kernel density in space and time dimension.
 - Input 5 features go through the network and use Adam as the optimizer.
 - Finally the output is the predicted density for that grid at a specific timestamp.



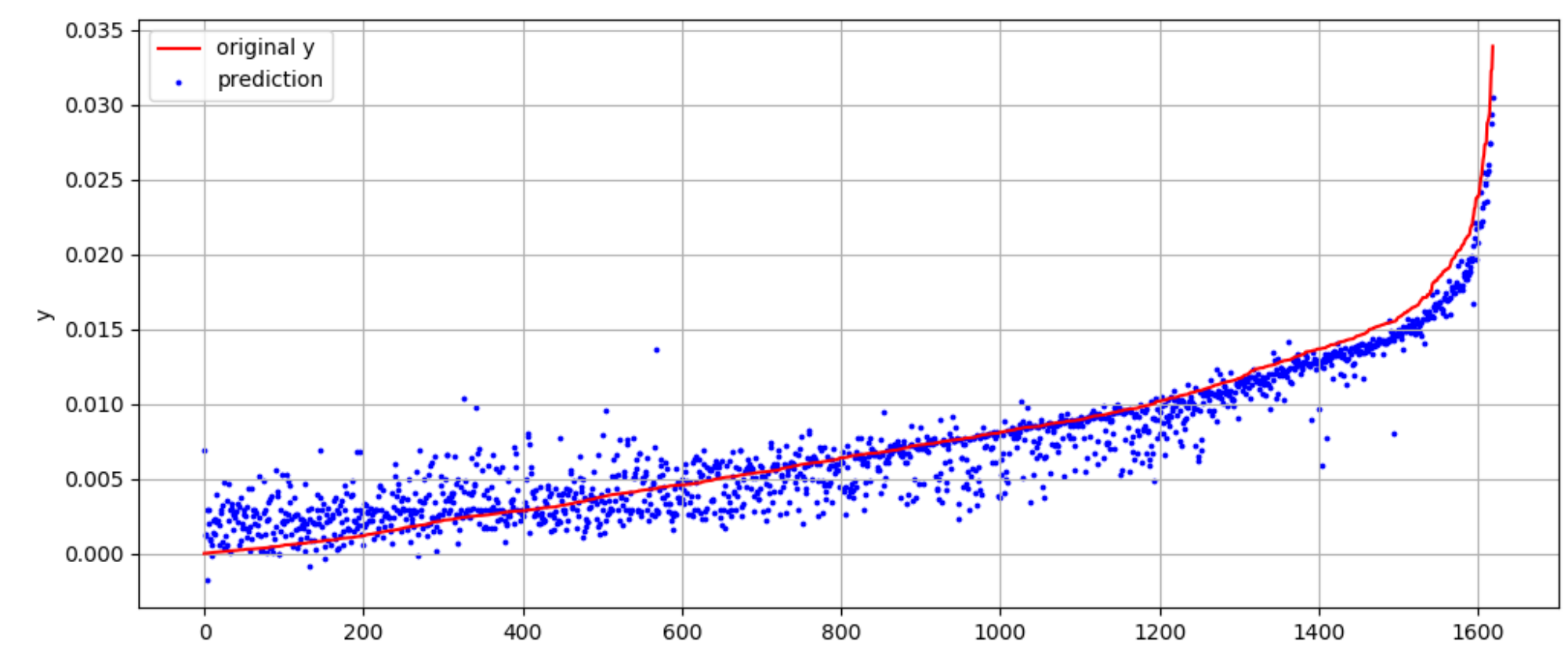
Probability distribution of crime occurrence over time



Correlation heatmap of the selected 5 features and the target

Experiments/Results (cont'd)

- Advantages:**
 - Adequate data collection
 - Low cost
 - Flexible model: change layers, learning rate, optimizer...
- Disadvantages:**
 - After SKDET, one year has about 1000 time grid, 3000 location grid so the data size is very big.

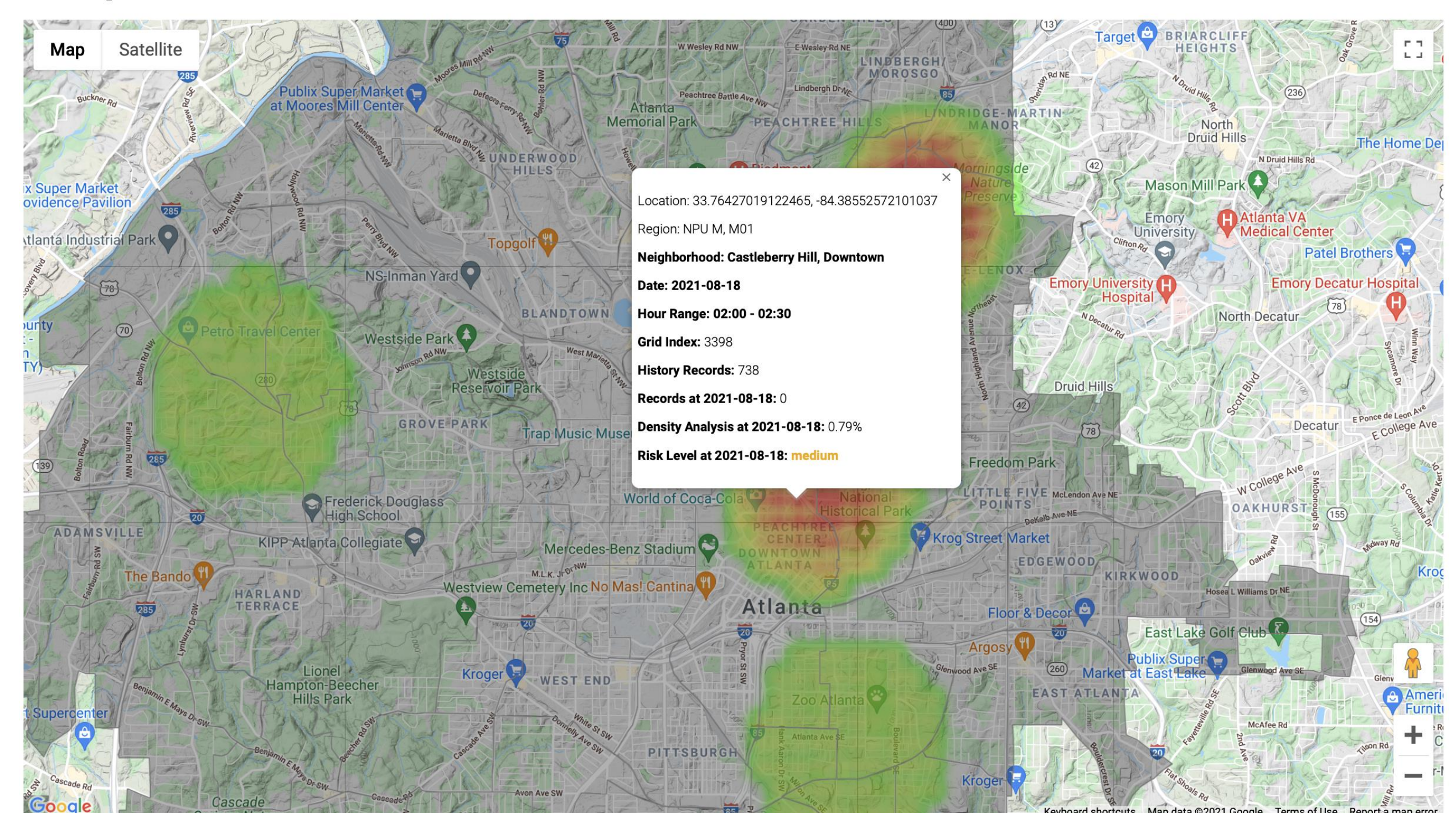


Predict result for 1300 timestamps

Visualization

- User Interface:**
 - slider option to choose a day, dropdown option to choose an hour, and click option to show information for a local grid. The crime analysis results are presented in a heatmap.

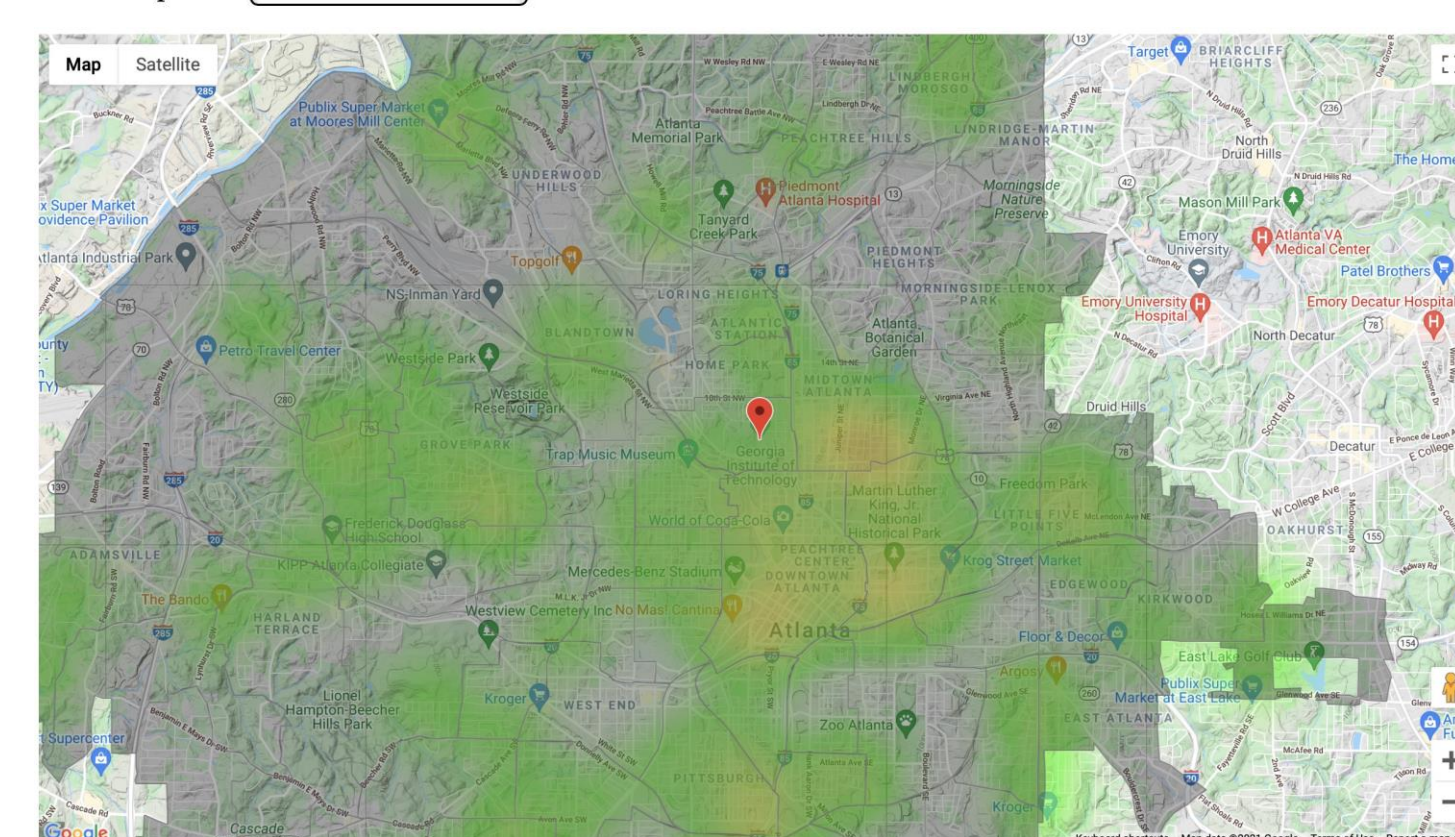
Crime Map Demo



Slide to select a previous/future date:



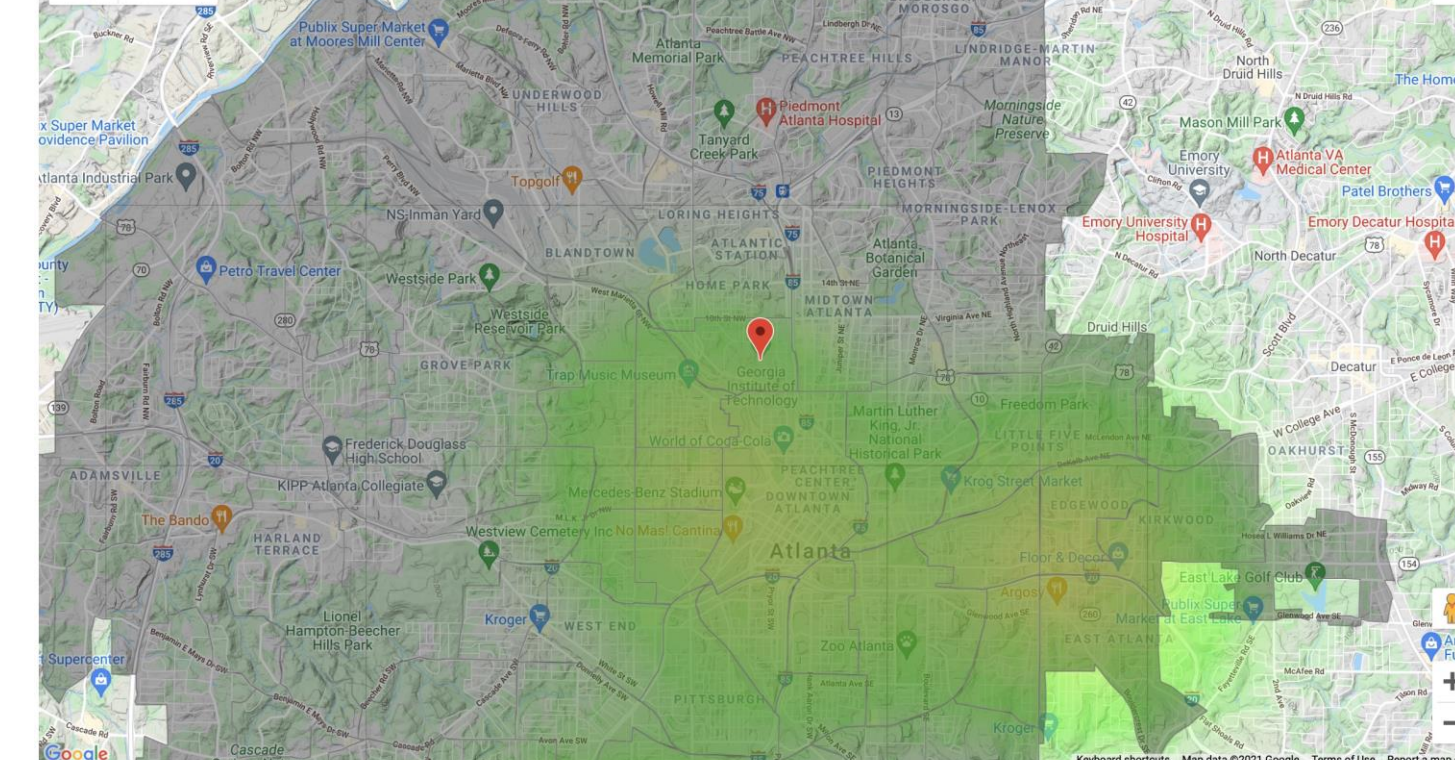
Use hour period: (02:00 - 02:30)



Slide to select a previous/future date:



Use hour period: (02:00 - 02:30)



Slide to select a previous/future date:



- The above figure shows real data presented by the heatmap for Jan 04, 2021.
- The bottom figure shows the predicted probability presented by the heatmap for Jan 04, 2021.
- As a result, the prediction matches main crime regions, but may ignore some border factors and produces a whole cluster of crime accidents in one region.