Predicting Crime in the City of Atlanta Using Event and Crime Data

Trevor Garner, Georgia Tech
Alex Wang, Georgia Tech
Morgan McGuinn, Georgia Tech
Sam Lefar, Georgia Tech
Thomas Collas, Georgia Tech
Ziad Moukadem Georgia Tech

Summary

Is there a correlation between events and crime in the City of Atlanta? We preformed the **first academic analysis** on the relation between crime and events in the City of Atlanta. Using those results we try to **predict** the number of crimes that might be committed in the City of Atlanta for the upcoming month. This information is then visualized to help inform travelers when visiting Atlanta to increase their **awareness** and improve their **safety**.

Crime Dataset

The Atlanta Police Department has crime data from **2009** - **Present**. We analyzed all that data which was **over 350,000 unique** crimes. We had access to several features for each crime, the most important features were:

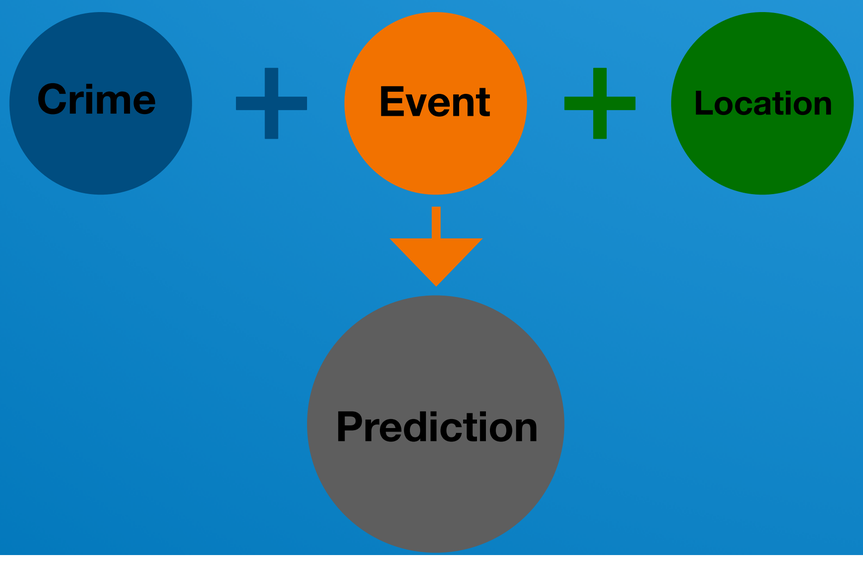
Date:	Timestamp
Beat:	Integer
Location:	String
Neighborhood:	String

Event Dataset

Google Maps and Google Events were **scrapped** using a **custom built Python script** which was able to gather event information. In addition, we used the event **API** from PredictionHQ to gather even more event data. Once grouped we had over **2,000 unique events** that occurred from **3/1/2021 - 3/31/2021**. The data was formatted as:

Date:	Timestamp
Title:	String
Category:	String
Longitude:	Float
Latitude:	Float
Beat:	Integer

Due to time complexity we were only able to scrape as far back as 3/1/2021. In the future we would like to continue gathering information from as far back as possible.



Prediction Model

The prediction model we used was a **Poisson regression** model and the metrics evaluated were Standard Error and Root Mean Squared Error (rMSE).

```
glm(formula = true_crime ~ white + income +
events_counts + pop,
   family = poisson(), data = data)
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
Income    -1.164e-05   2.34e-06    -4.8 1.7e-06
Events       1.9e-01   9.74e-02     1.9   0.046
```

We tried several other regression models such as linear and negative regression; however, Poisson **resulted** in the **lowest rMSE**. We also tested the model with and without **event data** and saw that with event data the **rMSE** value **decreased**. Additional tests were conducted for the other features and only the **statistically significant** features were **included**.

Visualizing Data

Our data was **visualized** using **d3.js**. We were able to create an **interactive website** that allows the user to explore the data based off the following topics:

‘Predicted Total Crime’
‘Previous Total Crime’
‘Total Crime Per Neighborhood’
‘Total Events Per Neighborhood’

The map is **segmented** by *beat* and **color coded** based on our prediction of future crimes. The user can **interact** with the map by **clicking** on a beat. This action will **show** the user **additional information** regarding the crime associated with that beat.

Crime Prediction Map