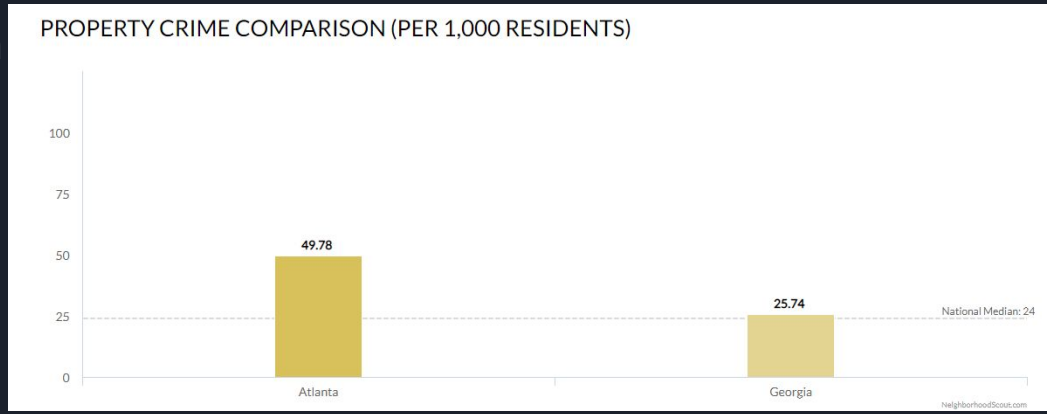
# Battle of the Neighborhoods: Atlanta Crime

By Rohan Chanani

# Background

- Atlanta is less safe than 98% of the United States
- Violent Crime rate is double the national average
- Different neighborhoods have different amounts of crime
- Neighborhoodscout.com



# Problem

- Plan is to cluster neighborhoods and classify the clusters using data about the venues in a neighborhood
- The goal is to be able to figure out what the crime will be like in a neighborhood based off of the venues in that neighborhood
- The target audience is residents and potential residents of Atlanta looking to understand and possibly avoid crime

# Data

- I will use  location data from foursquare.com and data about Atlanta Crimes from the Atlanta Police Department
- Atlanta Police Department Data can be found at https://www.atlantapd.org/i-want-to/crime-data-downloads.

Atlanta Police Data

| ur ate | Occur Time | Possible Date | Possible Time | Beat | Apartment Office Prefix | Apartment Number | Location | Shift Occurence | Location Type | UCR Literal | UCR # | IBR Code | Neighborhood | NPU | Latitude | Long |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 09-01 | 1145 | 2009-01-01 | 1148.0 | 411.0 | NaN | NaN | 2841 GREENBRIAR PKWY | Day Watch | 8 | LARCENY-NON VEHICLE | 630 | 2303 | Greenbriar | R | 33.68845 | -84. |
| 09-01 | 1330 | 2009-01-01 | 1330.0 | 511.0 | NaN | NaN | 12 BROAD ST SW | Day Watch | 9 | LARCENY-NON VEHICLE | 630 | 2303 | Downtown | M | 33.75320 | -84. |
| 09-01 | 1500 | 2009-01-01 | 1520.0 | 407.0 | NaN | NaN | 3500 MARTIN L KING JR DR SW | Unknown | 8 | LARCENY-NON VEHICLE | 630 | 2303 | Adamsville | H | 33.75735 | -84. |
| 09-01 | 1450 | 2009-01-01 | 1510.0 | 210.0 | NaN | NaN | 3393 PEACHTREE RD NE | Evening Watch | 8 | LARCENY-NON VEHICLE | 630 | 2303 | Lenox | B | 33.84676 | -84. |
| 09-01 | 1600 | 2009-01-01 | 1700.0 | 411.0 | NaN | NaN | 2841 GREENBRIAR PKWY SW | Unknown | 8 | LARCENY-NON VEHICLE | 630 | 2303 | Greenbriar | R | 33.68677 | -84.4 |

Foursquare Location Data

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | The Masquerade | Music Venue | 33.751720 | -84.389739 |
| 1 | GSU Sports Arena | College Basketball Court | 33.751735 | -84.386328 |
| 2 | Georgia Railroad Freight Depot | Event Space | 33.751479 | -84.388224 |
| 3 | Willy's Mexicana Grill #22 | Mexican Restaurant | 33.751293 | -84.385337 |
| 4 | Jamrock Restaurant | Caribbean Restaurant | 33.751554 | -84.391356 |

# Data Preparation

- Atlanta Crime Data was grouped into dataframe with each row being a neighborhood and the columns being total number of crimes, number of each type of crime, latitude, and longitude.
- Dataframe for clustering was made by dropping latitude, longitude, and neighborhood columns and normalizing the values for number of each type of crime
- Foursquare data was made into dataframe with each row being a neighborhood and each column being an average value between 0 and 1 for each venue category

Police Department Clustering Dataframe

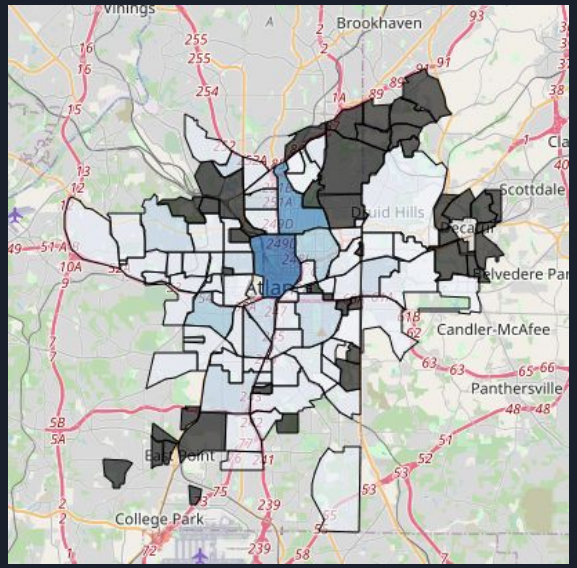
| | Count | LARCENY_NON_VEHICLE | LARCENY_FROM_VEHICLE | ROBBERY_PEDESTRIAN | ROBBERY_RESIDENCE | AUTO_THEFT | AGG_ASSAULT | BURGLARY_R |
|---|---|---|---|---|---|---|---|---|
| 0 | 2012 | 269.759732 | 297.479403 | 114.935224 | 9.465254 | 231.898717 | 133.865732 | |
| 1 | 1504 | 280.379372 | 311.130658 | 83.209362 | 13.566744 | 182.698817 | 103.107253 | |
| 2 | 2798 | 293.158036 | 339.829962 | 100.150175 | 10.695650 | 203.703511 | 147.794433 | |
| 3 | 850 | 196.842280 | 147.231624 | 88.018906 | 25.605500 | 243.252249 | 198.442624 | |
| 4 | 372 | 157.238075 | 197.461768 | 51.193792 | 3.656699 | 274.252456 | 164.551474 | |

Foursquare classification Dataframe

| Neighborhood | ATM | Accessories Store | Adult Boutique | African Restaurant | Airport Terminal | American Restaurant | Animal Shelter | Antique Shop | Aquarium | ... | Waste Facility | Weight Loss Center | Whisky Bar | Wine Bar | Wine Shop | Wings Joint | W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Adair Park | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Adams Park | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.250000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Adamsville | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Almond Park | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Amal Heights | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| Wildwood (NPU-H) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.250000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | |
| Wisteria Gardens | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Woodfield | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Woodland Hills | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.060606 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Wyngate | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |

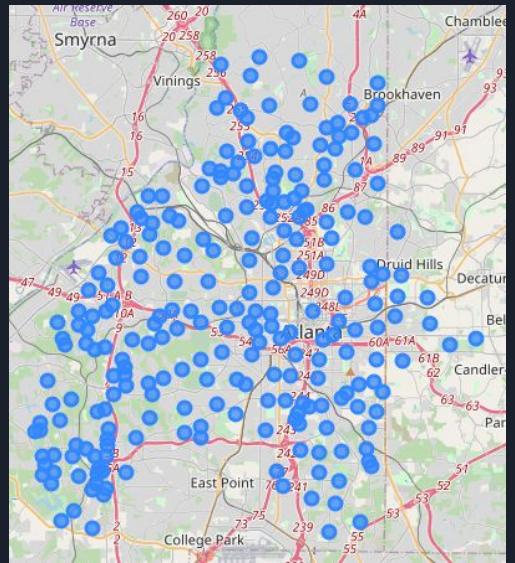
# Data Visualization

- I made two maps to understand distribution of the neighborhoods and the amount of crime in each neighborhood a bit better




← Choropleth with darker areas meaning more crime and black areas are NaN values

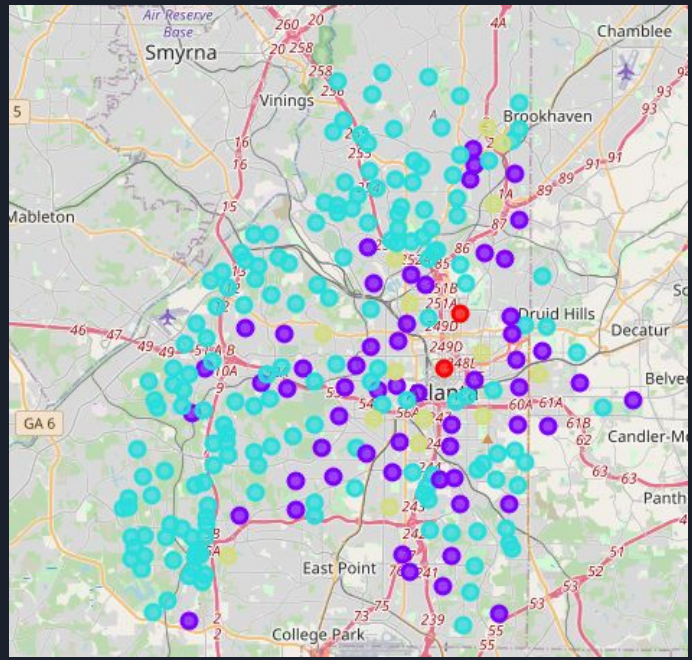
Scatter plot with each point Being a neighborhood →



# Clustering

- Used KMeans to make simple clusters so that classification would be easier
- Made four clusters: Low, Medium, High, and Very High
- The types of crime ended up having similar distribution across all neighborhoods, so no significant effect on the clusters

## Cluster Map



# Classification

- Used K-Nearest Neighbors, Support Vector Machine, and Decision Tree to classify the models based off of neighborhoods
- Used cross validation with f1 weighted score, jaccard weighted score, and accuracy score to tune and evaluate models

# Results, Discussion, and Conclusion

- None of the classifier models were accurate enough to be used to predict the cluster of a neighborhood
- Possible Reasons:
- Many of the neighborhoods had little or no venues within 800m of their average, while some had over 100, which could create inconsistencies.
- There were over 300 different categories of venues, which may have made the model too specific and caused it to over fit. A
- Also possible that the types of venues in a neighborhood don't have an effect on that neighborhood's crime,
- Grouping the categories into larger, broader categories so there's a smaller number of categories could improve the model
- Also would  be interesting to make a regression model that could predict the number of crimes in a neighborhood as opposed to the cluster that the neighborhood is in.