K-means and DBSCAN Clustering For a Desirable Living Locale

Objective

- Cluster Ottawa neighbourhoods based on:
  - Crime levels
  - Location
  - Venue popularity

- Who might this help?
  - New residents
  - Students new to Ottawa

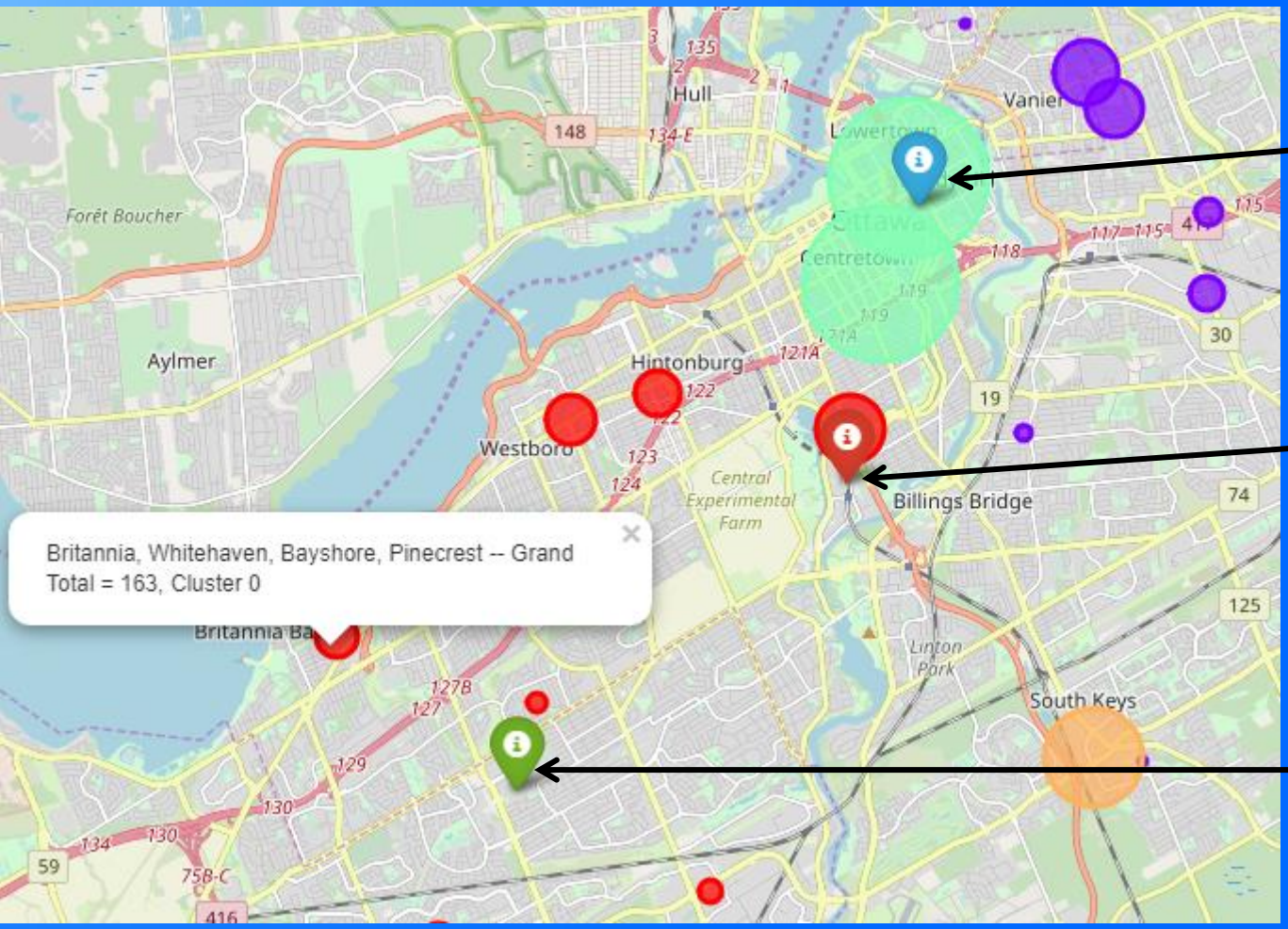
Data Used

- Ottawa Police Service [1]
  - Common Postal Code Locations
- Carleton University [2]
  - Crime data per neighbourhood, 2015
- Foursquare API
  - Popular venues throughout each neighbourhood

Methodology

1. Load postal code and crime data into a pandas dataframe
2. Cluster neighbourhoods with DBSCAN algorithm based on:
   - Latitude, longitude, and grand total of incidents
3. Cluster neighbourhoods with k-means algorthim based on:
   - Top 5 venues returned from Foursquare
4. Analyze results

DBSCAN Results

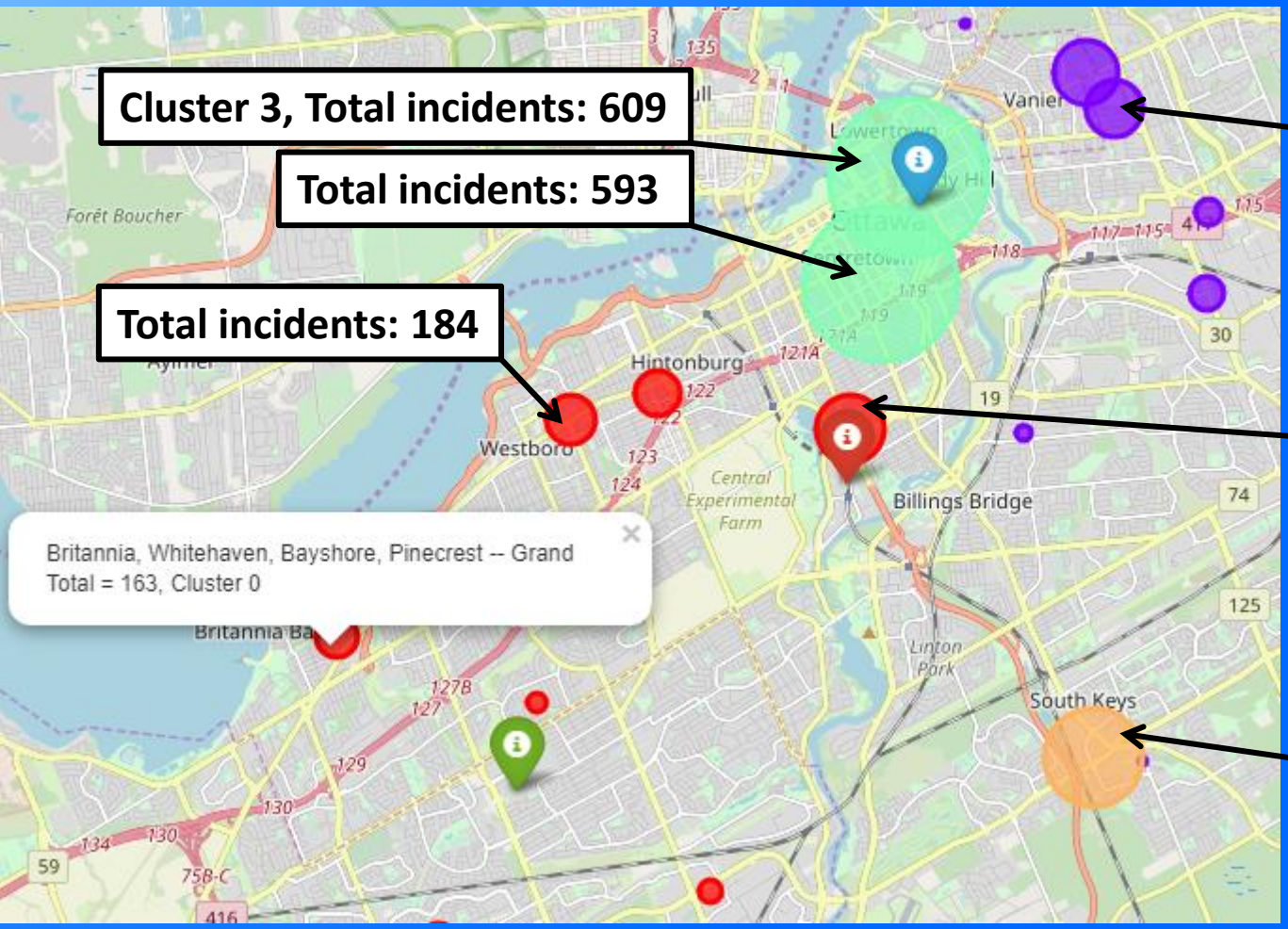


University of Ottawa

Carleton University

Algonquin College

DBSCAN Results

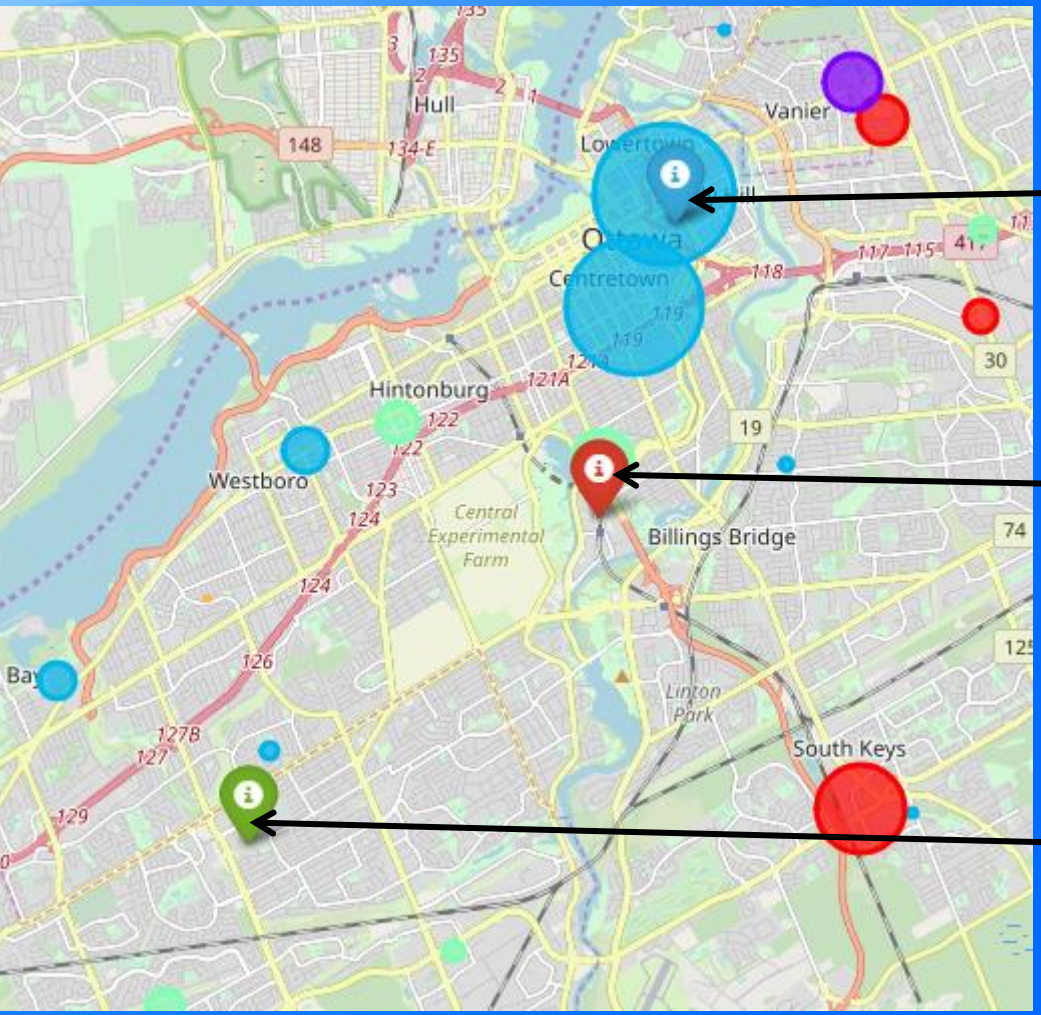


Cluster 1,
Total incidents: 209

Cluster 0,
Total incidents: 255

Outlier,
Total Incidents: 379

K-means Results

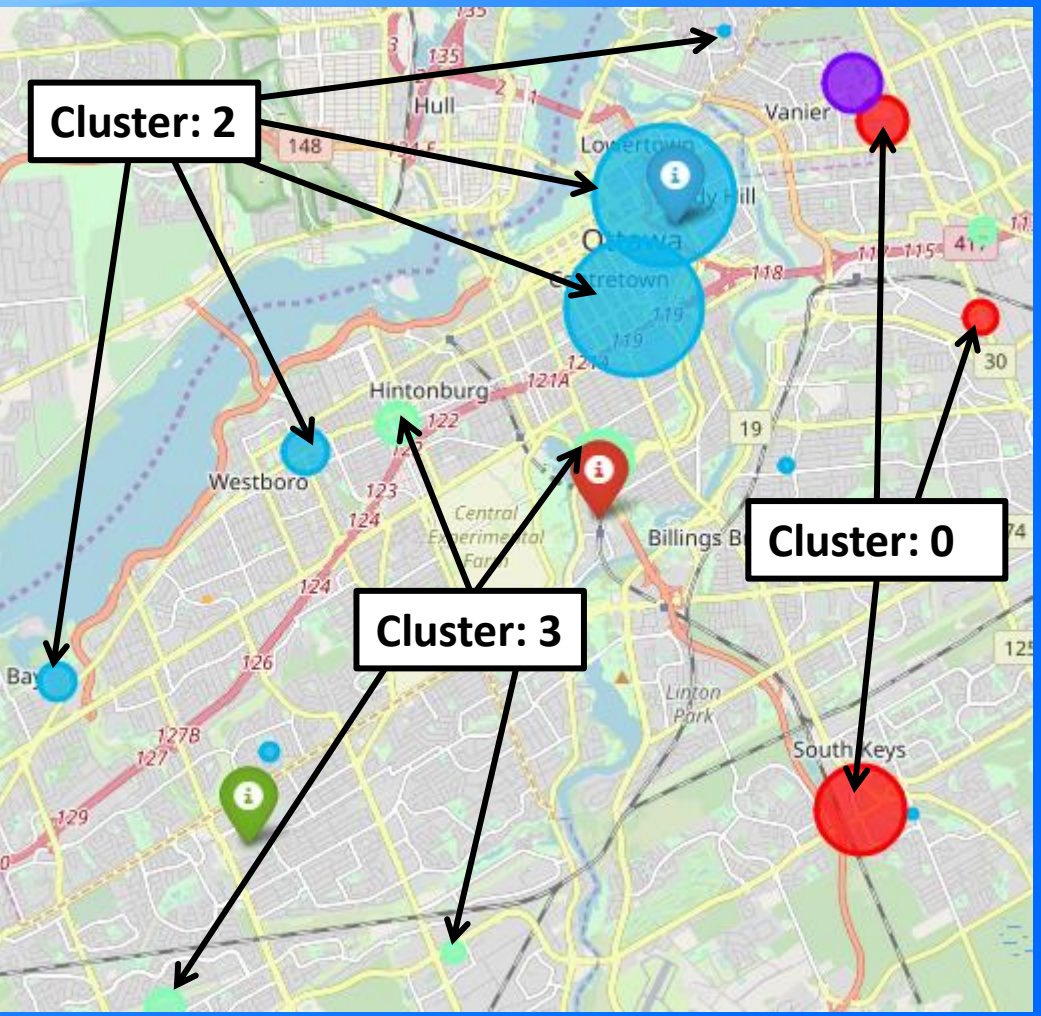


University of Ottawa

Carleton University

Algonquin College

K-means Results



Cluster 2: University of Ottawa

- Coffee shops rank as the most common venue
- large social spaces dominate (concert halls, stadiums, parks, etc.)

Cluster 3: Carleton University

- Coffee shops also rank as most common here
- Bakeries are popular within this cluster

Cluster 0: Eastern Ottawa

- Grocery and clothing stores are the most popular within this cluster

Conclusion

- Students and new residents would be exposed to higher crime levels in the downtown core
  - Particularly students commuting to (or living on) the University of Ottawa campus
  - Approximate 50% reduction in reported crimes between the two university locales
- Carleton University and Eastern Ottawa share similar crime levels, but factoring in the popular venues for each (such as proximity to grocery stores) will help individuals confirm a new home

Future Analysis

- Preprocess venue data to achieve a roughly spherical distribution [3]
  - This will help identify the most efficient number of clusters for k-means
- Incorporate coordinates for each incident
  - Ideally, current statistics will be available from the Ottawa Police Service
- Incorporate rent and utility prices
  - This will achieve a more realistic model in terms of client needs

References

[1] Ottawa Police Service. Ottawa Postal Codes. [Online]. https://www.ottawapolice.ca/en/contact-us/resources/ottawa-postal-codes.pdf

[2] Carleton University. (2016) Ottawa Police Service Crime Data. [Online]. https://library.carleton.ca/find/gis/geospatial-data/ottawa-police-service-crime-data

[3] Yellowbrick. (2016) scikit-yellowbrick. [Online]. http://www.scikit-yb.org/en/latest/api/cluster/elbow.html