



The Battle of the Neighborhoods Coursera Capstone

Daniel Miller
July 2, 2020

Overview of presentation

- Introduction
- Data
- Methodology
 - Data sources
 - Data modelling
- Results and Discussion
- Conclusion

Introduction:

A client is evaluating two former college cities as a potential location for starting a new business venture:

- When choosing a new location in a neighboring college town for a new business venture – coffee shop / music house, it is important to evaluate safety, competition, and general neighborhood demographics.
- This micro-project will use demographics data, crime data, and social venue data for two similar college towns to help provide factual content to support which location is optimal.
- The target presentation is for an audience who initially brought the question: one who was contemplating retirement in the near future and where, given the two prior home cities, would lend itself as a better opportunity for a business venture?

Data:

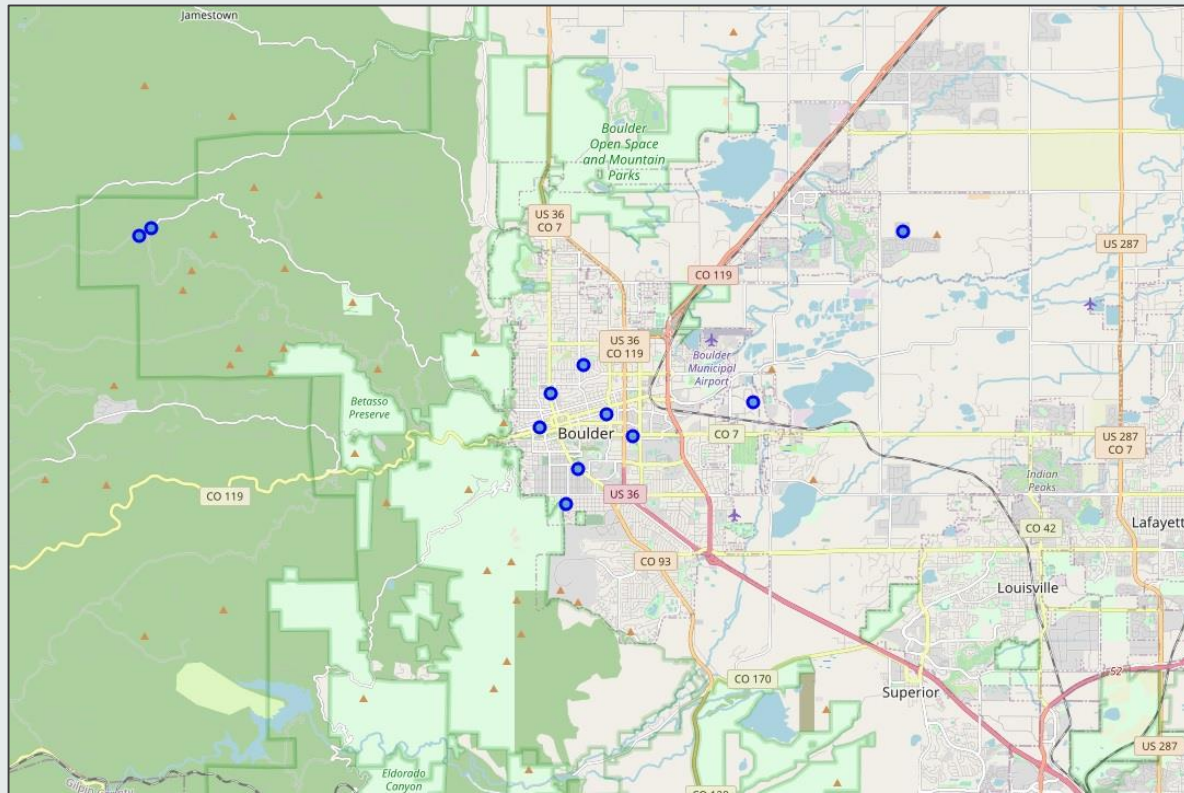
- Data Acquisition:
 - Both cities demographics, crime, and neighborhood data were gathered from open data sites. All relevant content was freely available through multiple disparate data sources.
- Data Scraping:
 - Raw data was initially acquired, and then through supplemental functions, complimentary content was appended:
 - Standardizing neighborhood names
 - Confirming geo spatial locations for crime, neighborhoods, venues.
- Data Cleansing:
 - Once data was gathered, it was cleansed through multiple iterations:
 - Outliers were removed
 - Gaps in data were either provided a default value or removed based on relevance.
 - Data was filtered to reflect the current years activities.

Methodology:

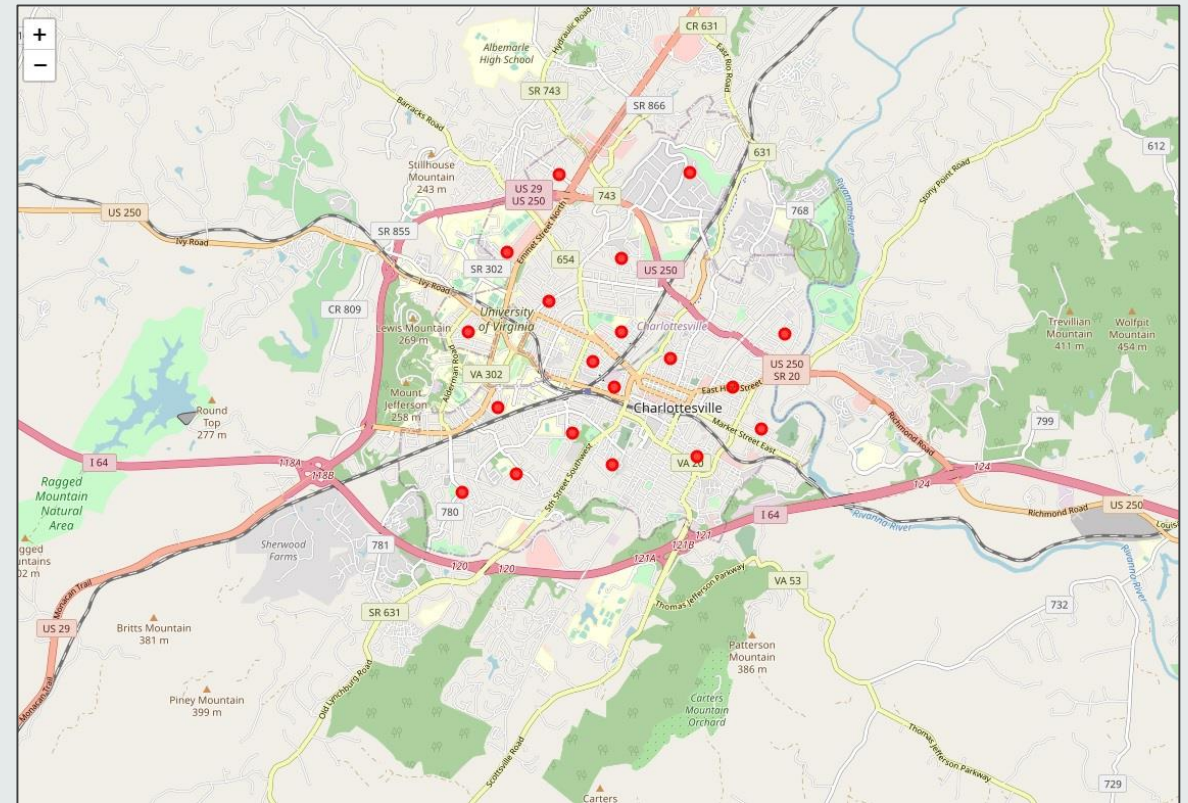
- Data Scraping from multiple open data sources
- Scrub and cleanse all current crime data
- Acquiring geographical coordinates for all data points of interest(crime, neighborhoods, venues)
- Utilizing FourSquare API to integrate common venue data
- Group neighborhood data
- Filter data based on current dates, commonalities, and other elements provided by client.
- Apply clustering algorithms against the data
- Visualize content both via maps and clustering details

City Neighborhood Disbursement

Boulder CO
11 Neighborhoods Evaluated

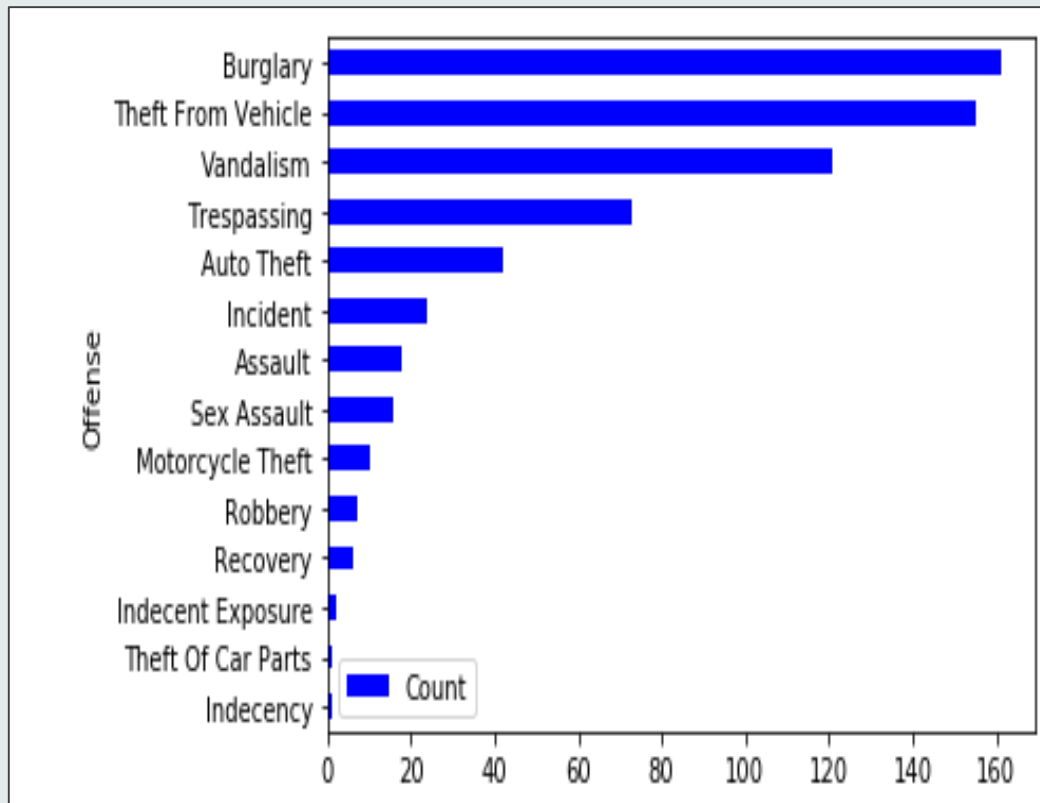


Charlottesville VA
19 Neighborhoods Evaluated

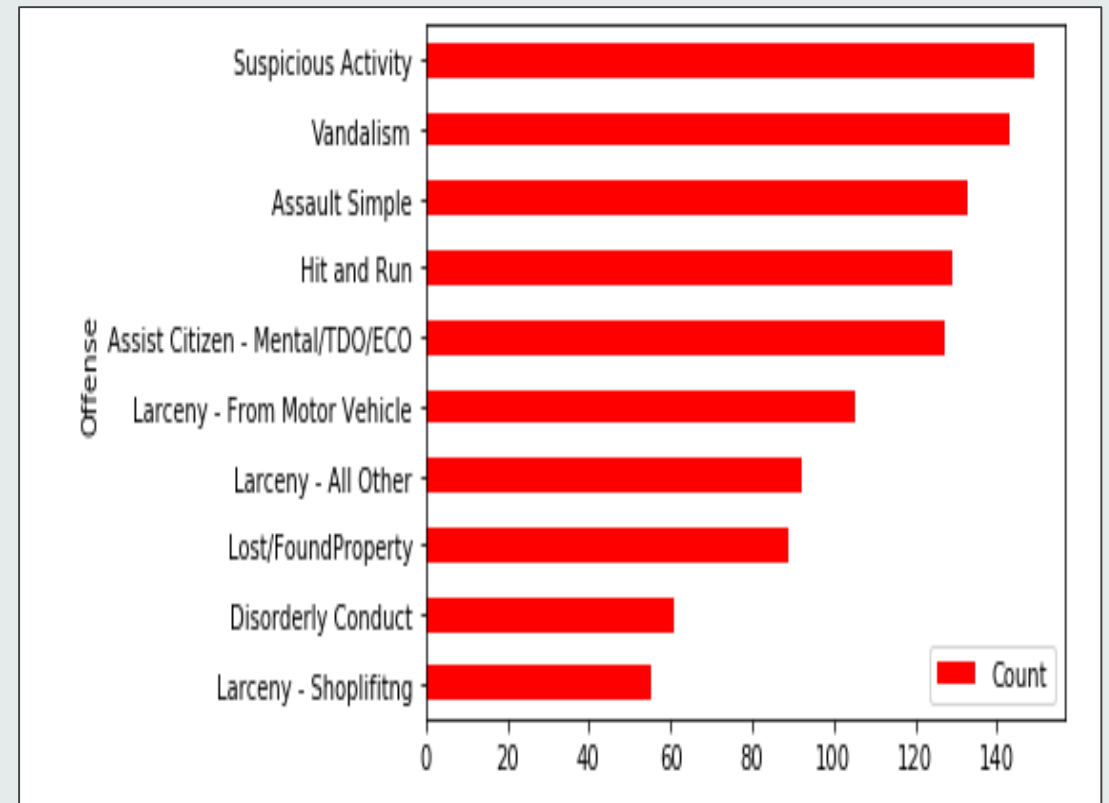


2020 City Crime Comparison

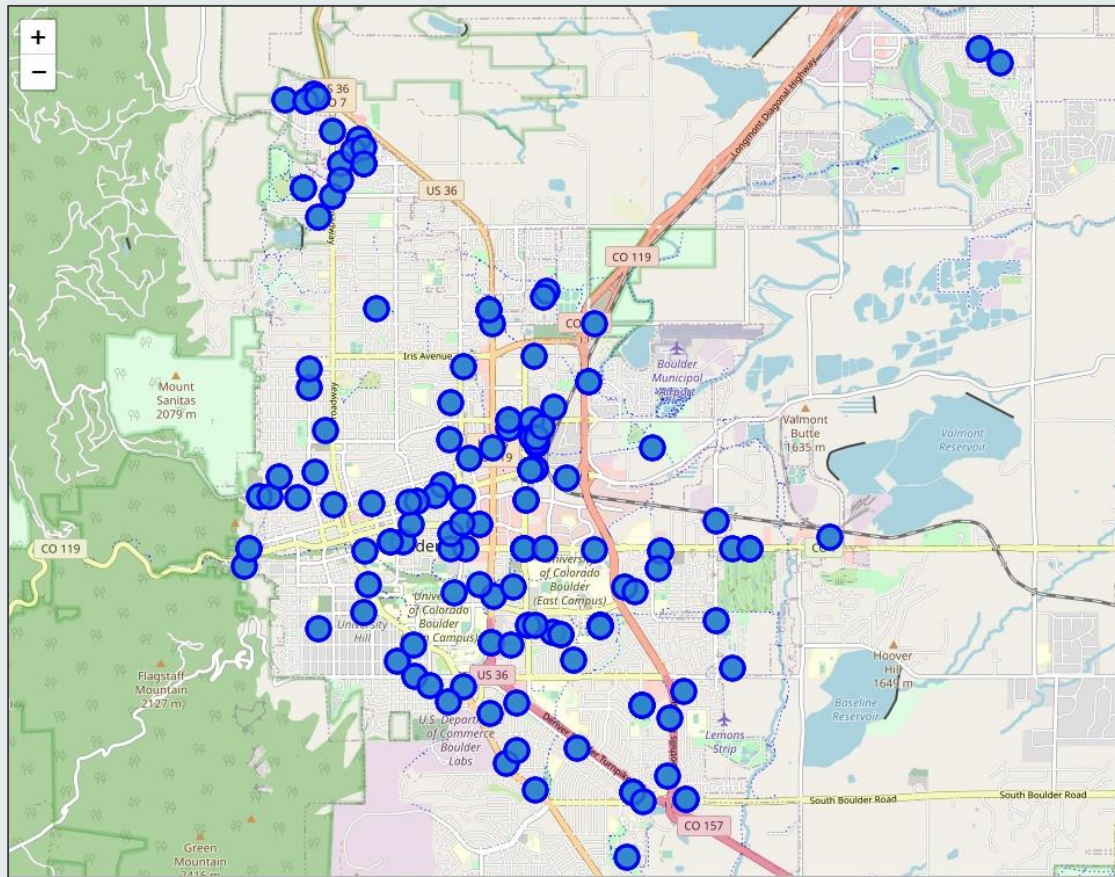
Boulder CO



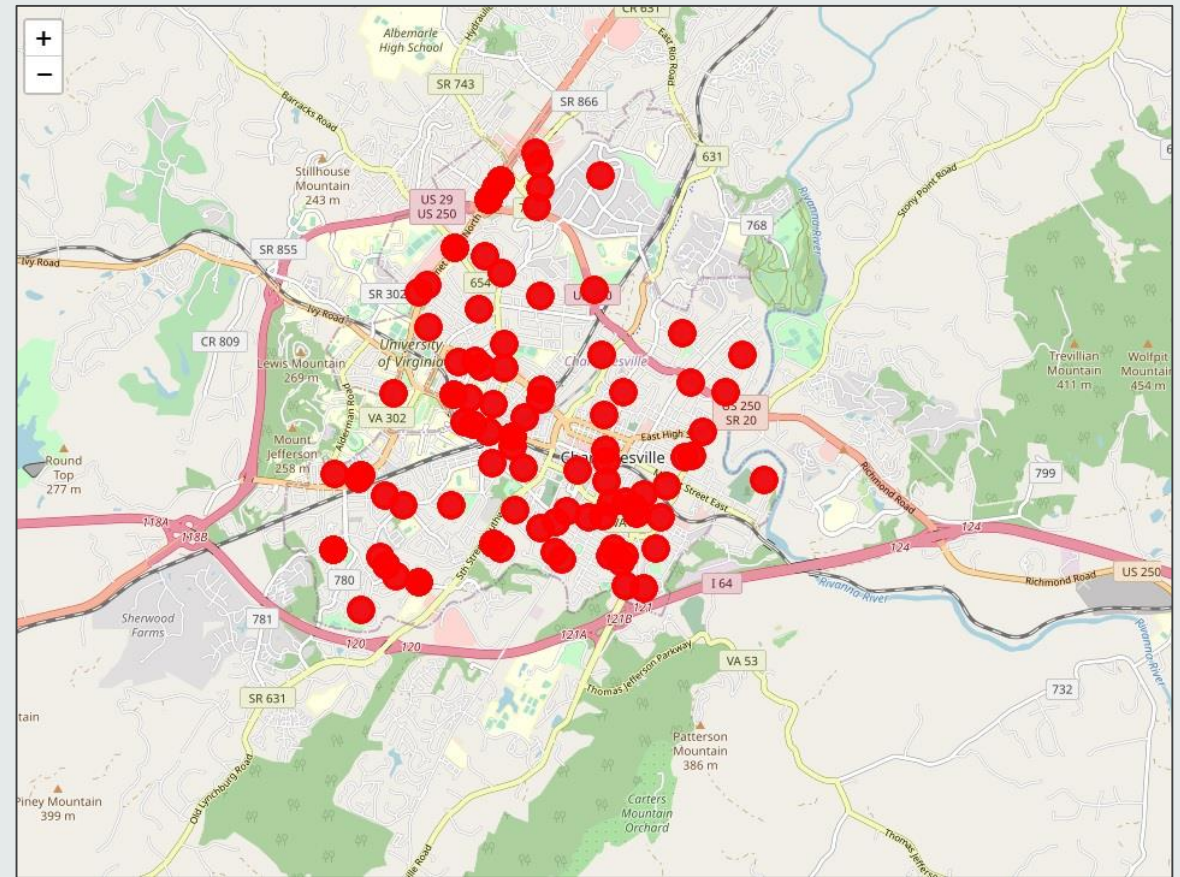
Charlottesville VA



Boulder CO



Charlottesville VA

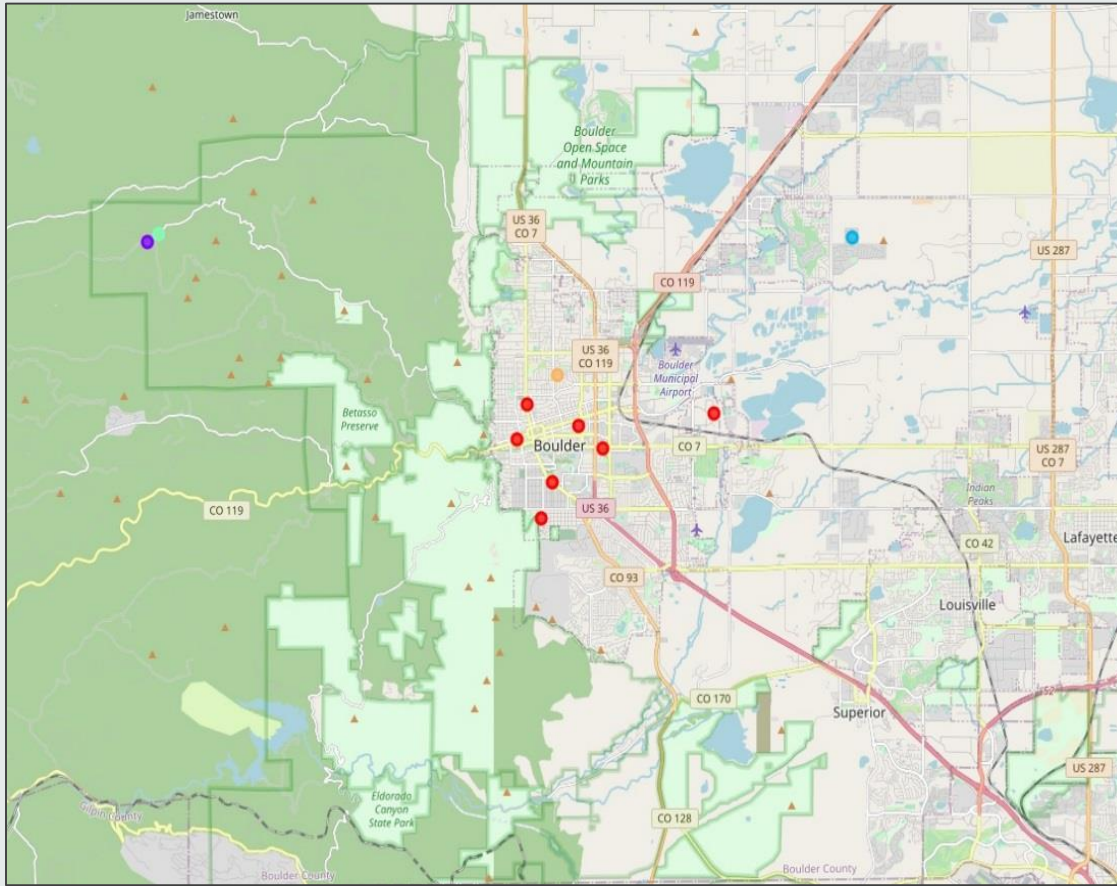


Clustering Neighborhood Focal Points

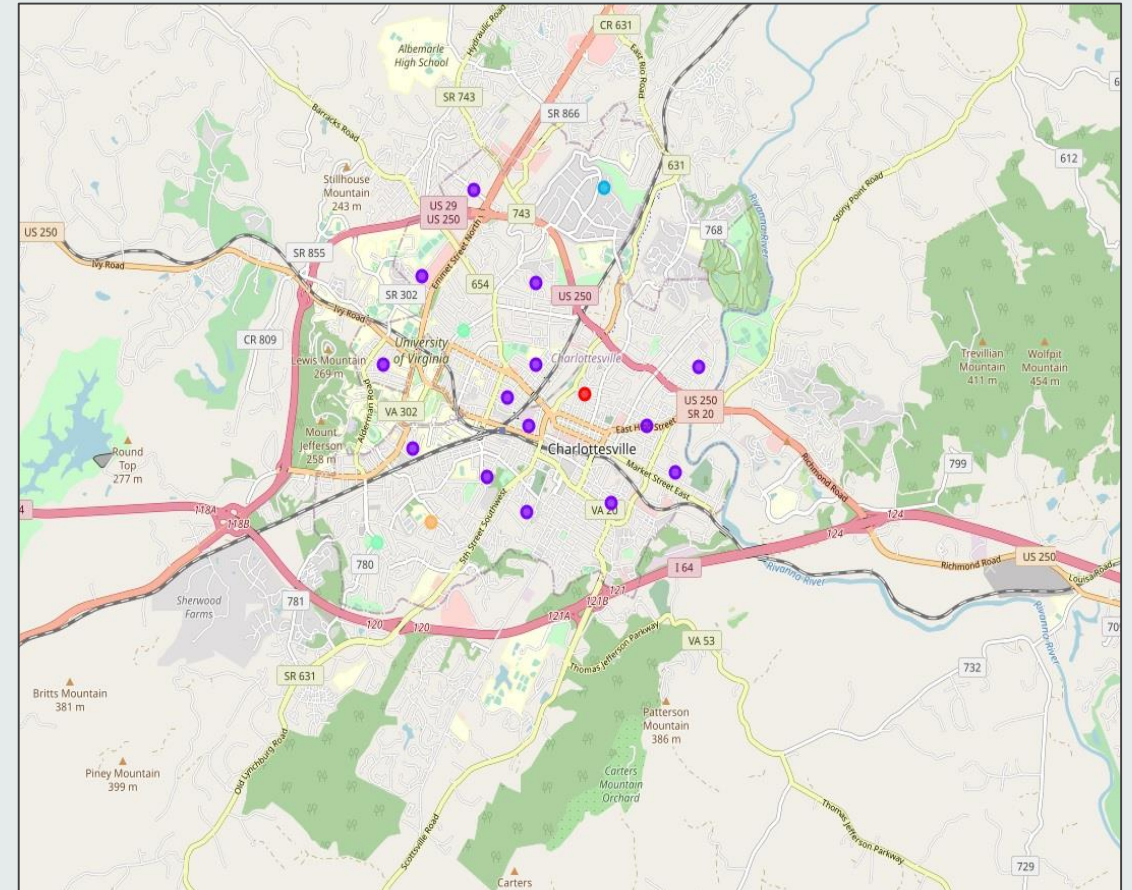
- Through using the K-Means clustering algorithms, for both cities, the neighborhoods were clustered into five common groupings.
- Basing our analysis on the typical types of venues that frequently appear within each of these clusters, we are able to easily visualize those targeted neighborhoods of interest that offer favorable opportunity.

City Neighborhood Clustering

Boulder CO
5 Neighborhood cluster depiction



Charlottesville VA
5 Neighborhood cluster depiction



Results, Discussion, Focal Points

- Both city neighborhood data, grouped in clusters, was evaluated further to determine areas where there would be a competitive advantage, as well as supporting complimentary business services nearby.
- With clustering and top venues easily available, client can quickly shortlist those neighborhoods of interest, removing initial unbeknownst bias.
- Combining the general crime figures, specific crime occurrences in the past year, and overlaying the neighborhoods gives us the ability to quickly remove those that fall out of consideration.

Neighborhood Top 3 proposal

- Minimal direct competition
- Supporting complimentary venues
- Lowest crime statistics
- Steady traffic / customer flow

Charlottesville	Boulder
Rose Hill	South Boulder
Johnson Village	Chautauqua
Lewis Mountain	Gunbarrel

Conclusion

- We can share that within this data analysis for both college cities:
 - We have identified the safest neighborhoods based on crime rates and extending focus in greater detail to vehicle type crimes.
 - Clustered neighborhoods based on venue findings and business offerings
 - Provided both visualizations and notes related to findings of each city.
- Please note that this evaluation is relatively simple in nature. It does not take into account topics such as tax rates (property, sales, business), nor does it account for cost of housing, employment index, .
- If target audience would agree, the project could be extended to look into additional factors such as the above or other accompanying