BUSINESS LOCATION OPTIMIZATION

Adam Davis

June 1, 2020

"THERE ARE THREE THINGS THAT MATTER IN PROPERTY: *LOCATION, LOCATION, LOCATION.*"

--Lord Harold Samuel, British real estate tycoon.

# Businesses succeed or fail based on their location.

- This project will attempt to find an optimum business location based on the presence of beneficial current businesses (traffic drivers) and existing competitors.

- This type of analysis would be of interest to any entrepreneur interested in starting a business or real estate owner trying to establish the value of their property.

- Specifically we will search for a good neighborhood for a nighttime live music venue. To be successful, the location should have a sufficiently robust nightlife to support another nighttime business, but not too many nighttime entertainment competitors.

- We will use the city of Baltimore in the state of Maryland, USA as our city.

Data Sources

- We use the official border of the city of Baltimore, published by the city government.

- For businesses driving nighttime traffic to a neighborhood, we will use Foursquare venues in the Food category, which includes numerous types of restaurants.

- We consider possible nighttime competitors to be businesses in the following Foursquare categories:
  - Music Venue
  - Amphitheater
  - Circus
  - Comedy Club
  - Concert Hall
  - Country Dance Club
  - Music Venue (jazz/piano bar/rock club),
  - Performing Arts Venue
  - Nightclub
 - Other Nightlife
  - Strip Club

Methodology: Geography

• To limit searches to the city boundaries, we divide Baltimore city into 901 squares approximately 500m x 500m. The corners form bounding boxes that we use as inputs to the API Foursquare.

• Use Foursquare's intent=browse and appropriate category filters to return all venues within the bounding box.

• Save 1539 traffic drivers and 234 competitors in separate datasets (panda.dataframes)

• Calculate any distances needed with latitudes and longitudes using Haversine formula.
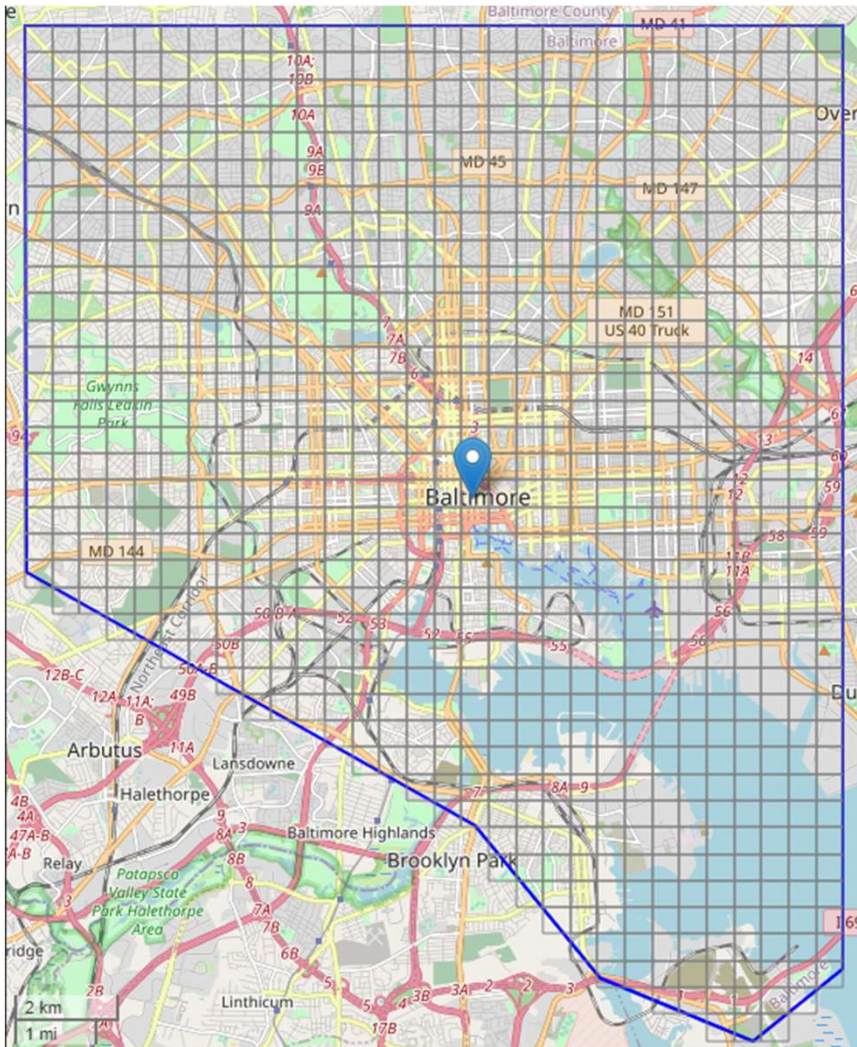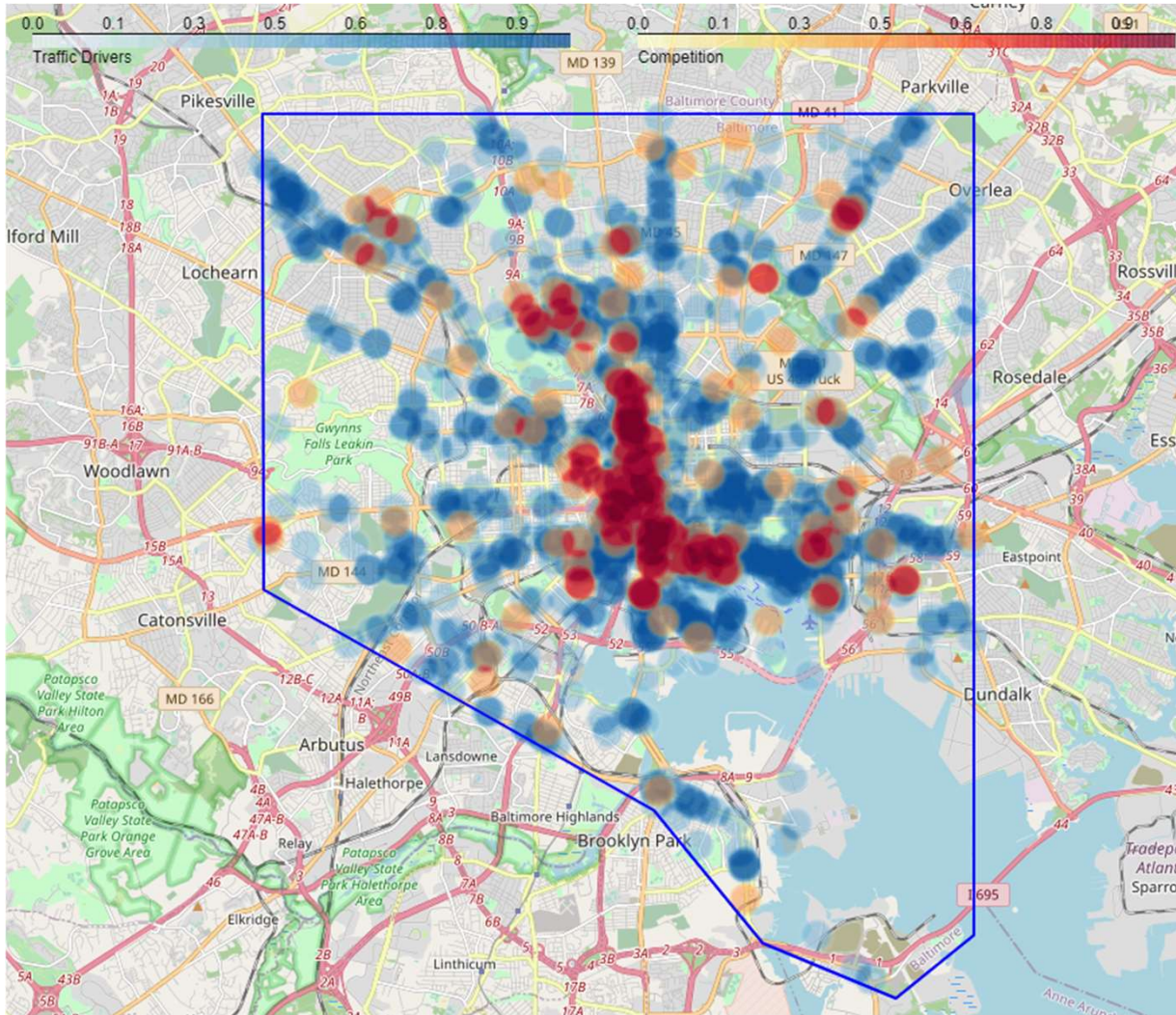
HEATMAP OF TRAFFIC DRIVERS AND COMPETITORS

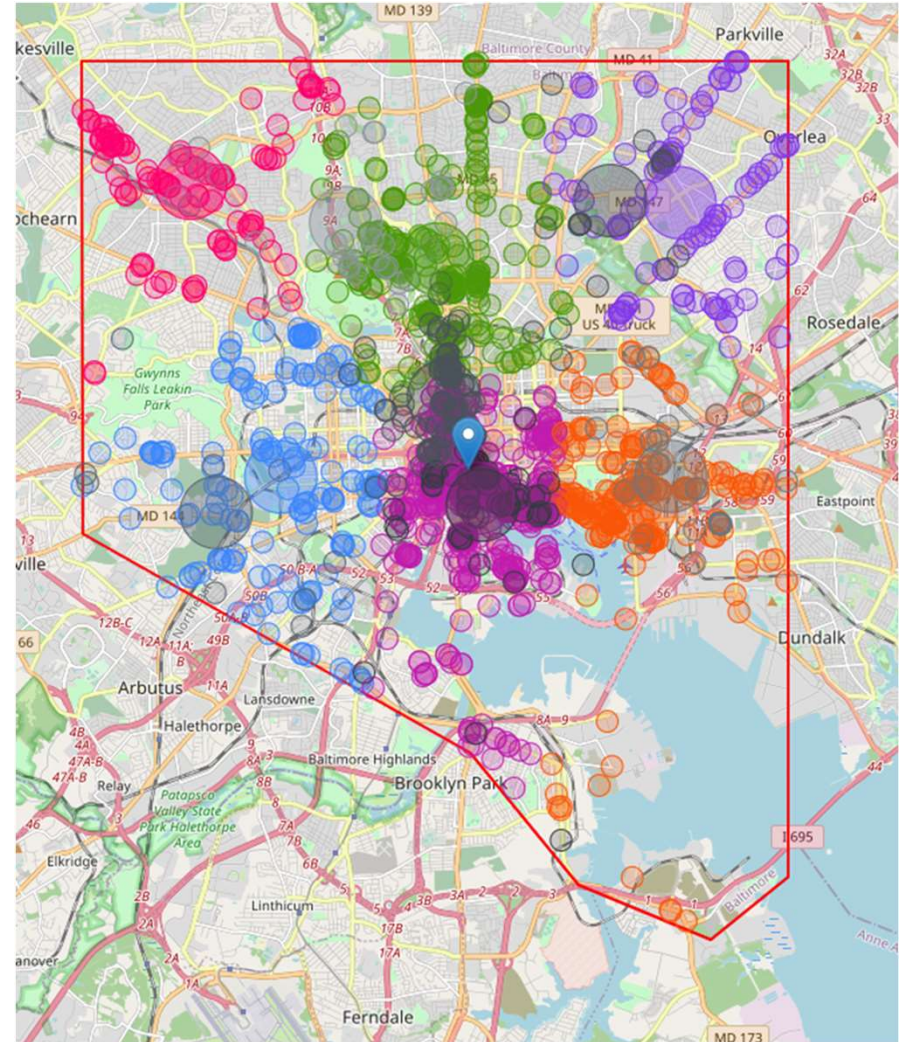# Find number of clusters

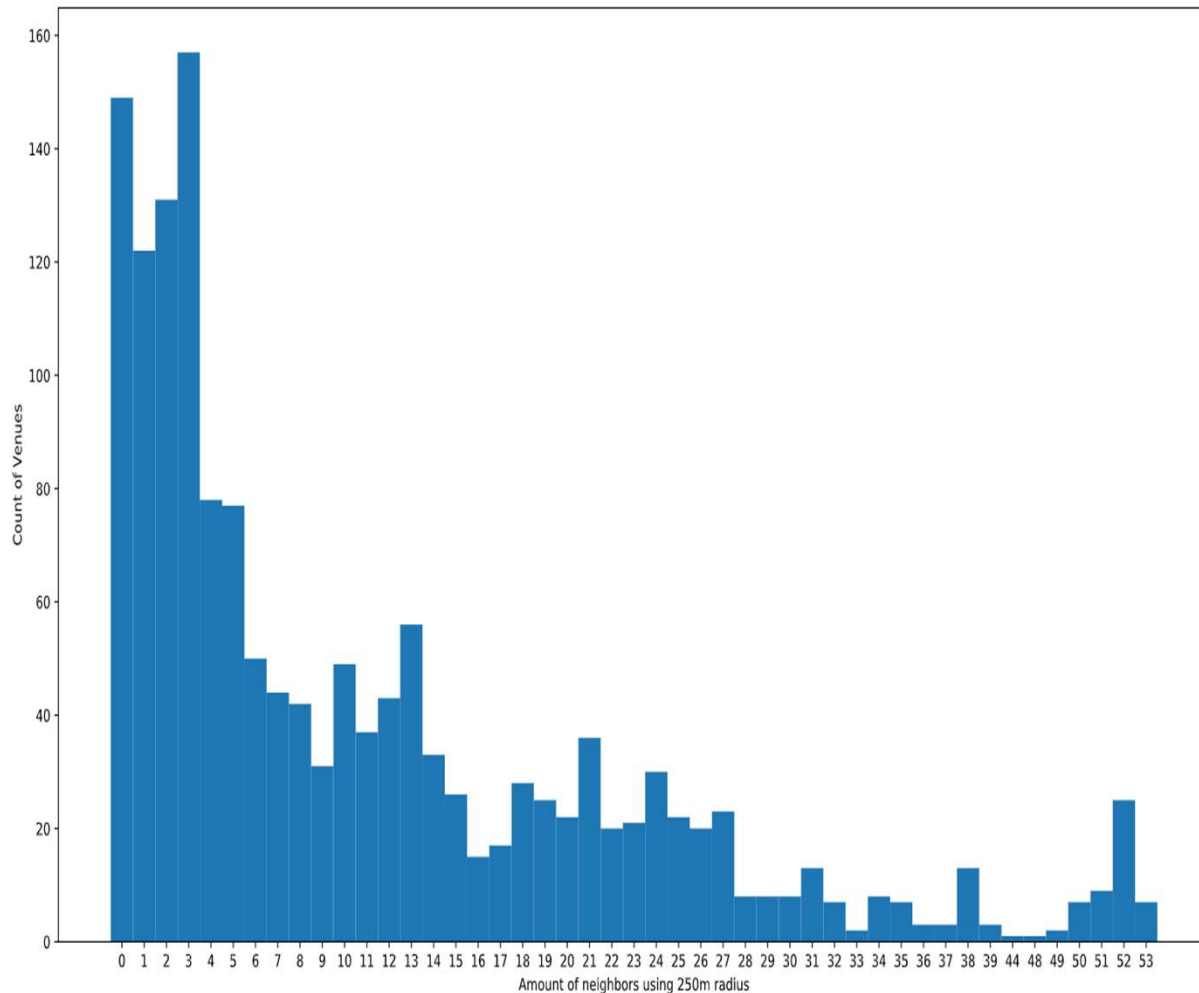The Elbow Method used to determine number of clusters
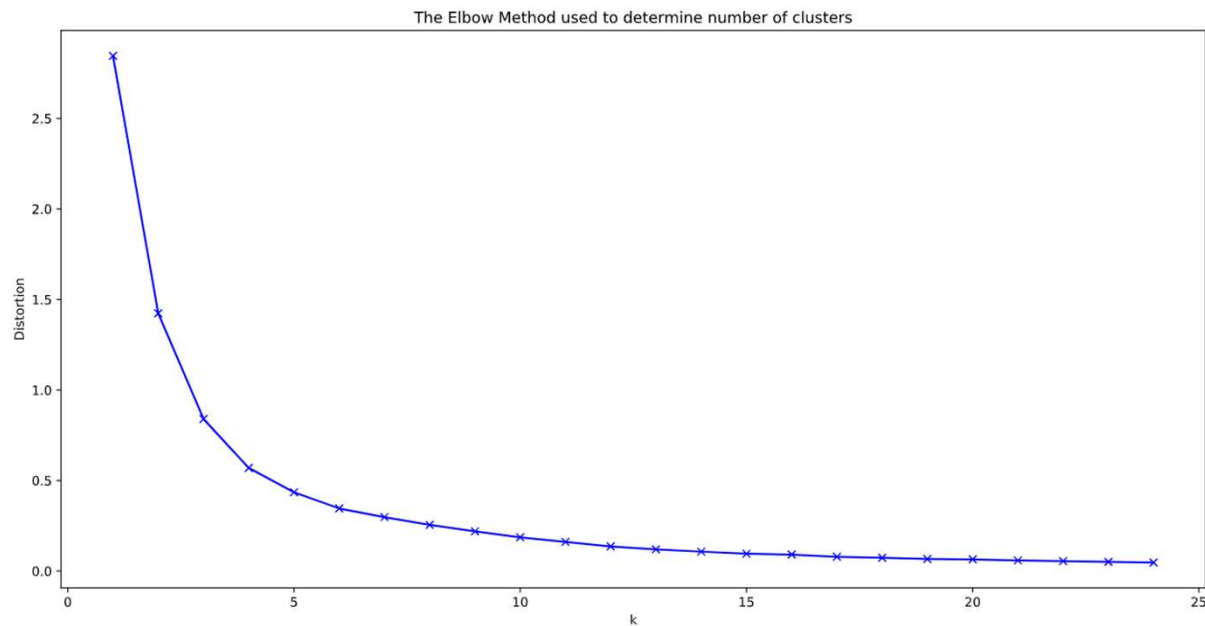
- Using the elbow method, we find 6 clusters of traffic driving venues.

Methodology: Clustering

- Traffic driving clusters are color coded.

- Competitors are drawn in grey.

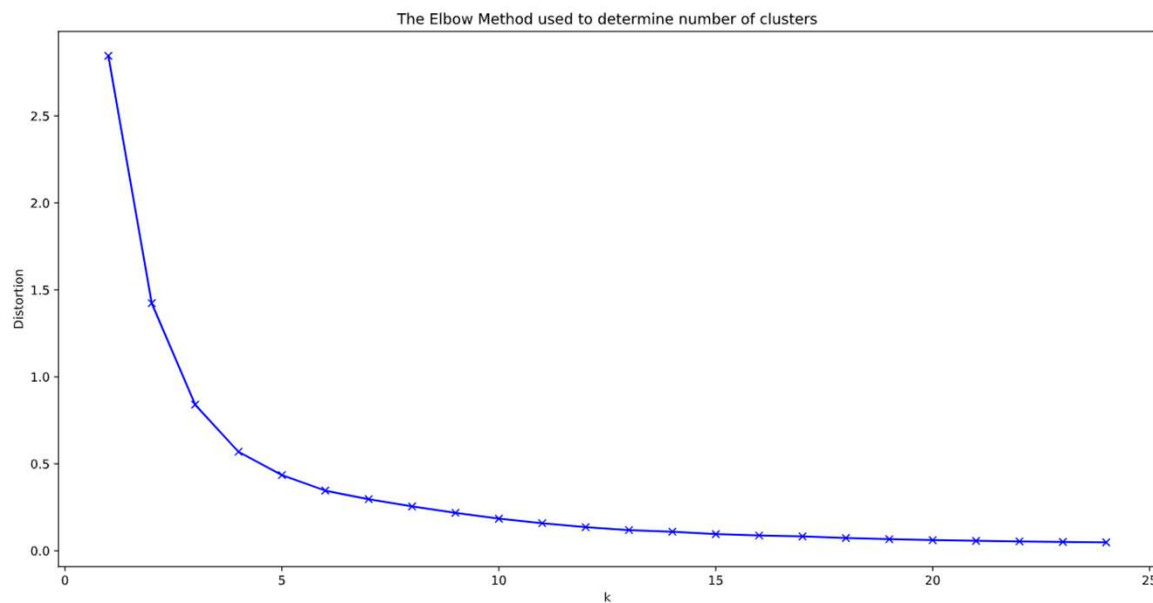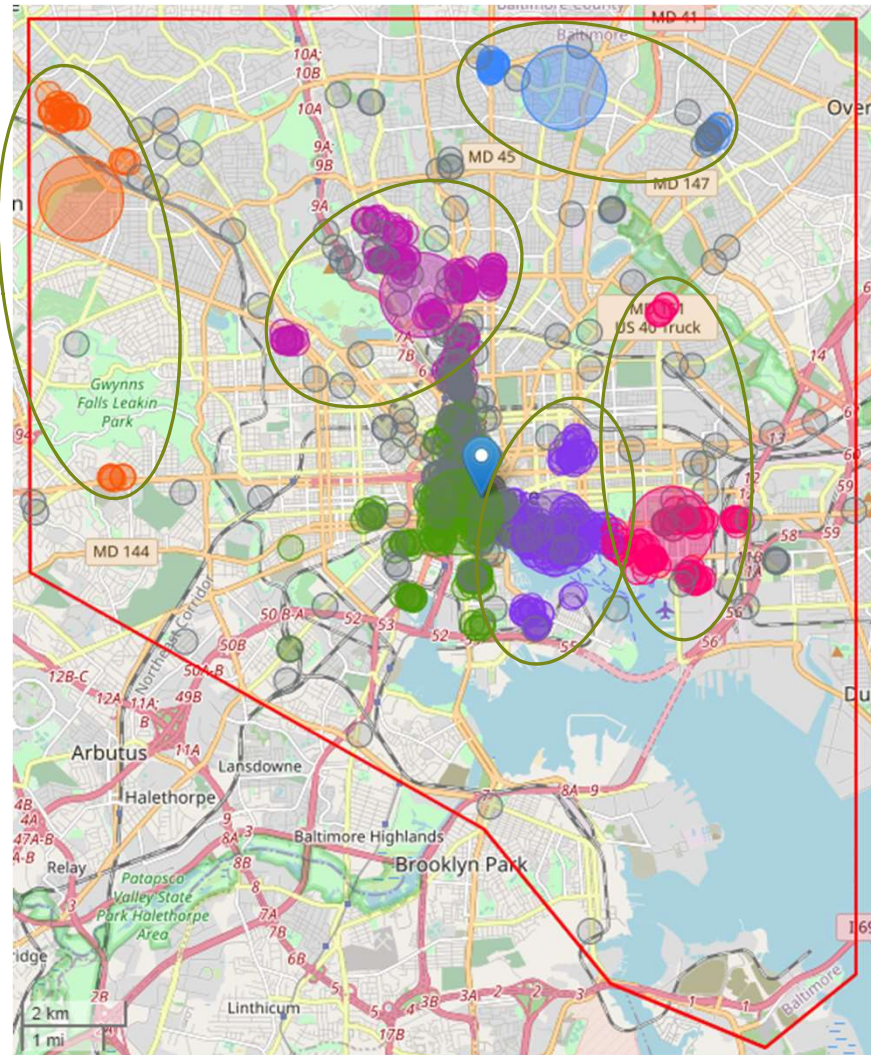- Centroids are larger circles.

- Unsatisfying!

Methodology: Data Cleansing

- Venues with few neighbors in a 250m radius are not going to be part of nightlife centers and are only adding noise

- Remove venues with only 0-5 neighbors, 46%.

- Of the original 1,539 venues, 714 remain.

The Elbow Method used to determine number of clusters

Find number of clusters (2)

- With reduced number of venues, calculate number of k-means clusters again.

- Again we find 6 clusters of traffic driving venues.

k-means Clusters
With Reduced Noise

- Highlighted anomalies are disturbing

- Several clusters seem like they have small subclusters.

- Several centroid centers (larger circles) don't seem to fall in any kind of nightlife center.

- Ex: Orange and blue centers are completely divorced from colored venue centers, and really only indicate the average location.

k-means vs. HDBSCAN

- k-means is requires unweighted data, but our lat/long data is nonlinear at 49° latitude.

- HDBSCAN doesn't require a cluster number, only minimum cluster membership

- HDBSCAN identifies core members of a cluster

- HDBSCAN identifies noise—points that aren't member of any cluster

- HDBSCAN supports Haversine formula for distance calculation
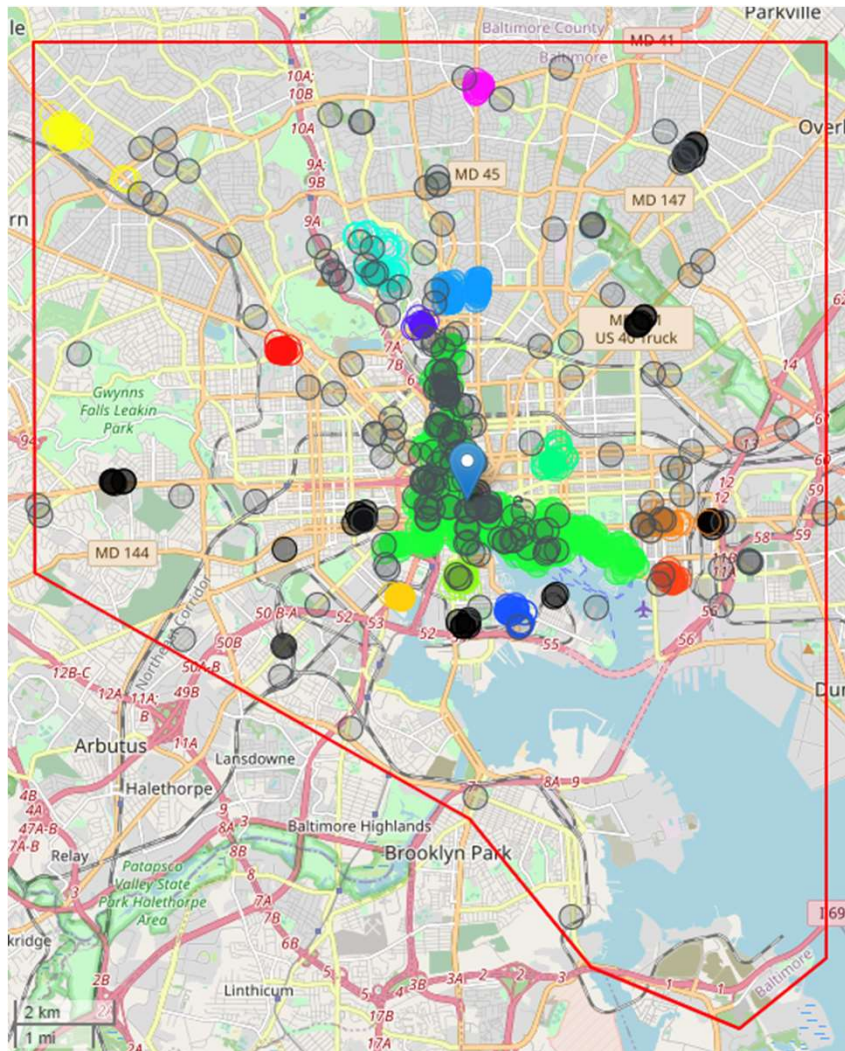
HDBSCAN is a better solution

- HDBMEANS clusters with minimum cluster membership of 10 venues.

- HDBMEANS finds 12 clusters tightly centered.

- HDBMEANS identifies noise (black circles)

- Much more satisfying!
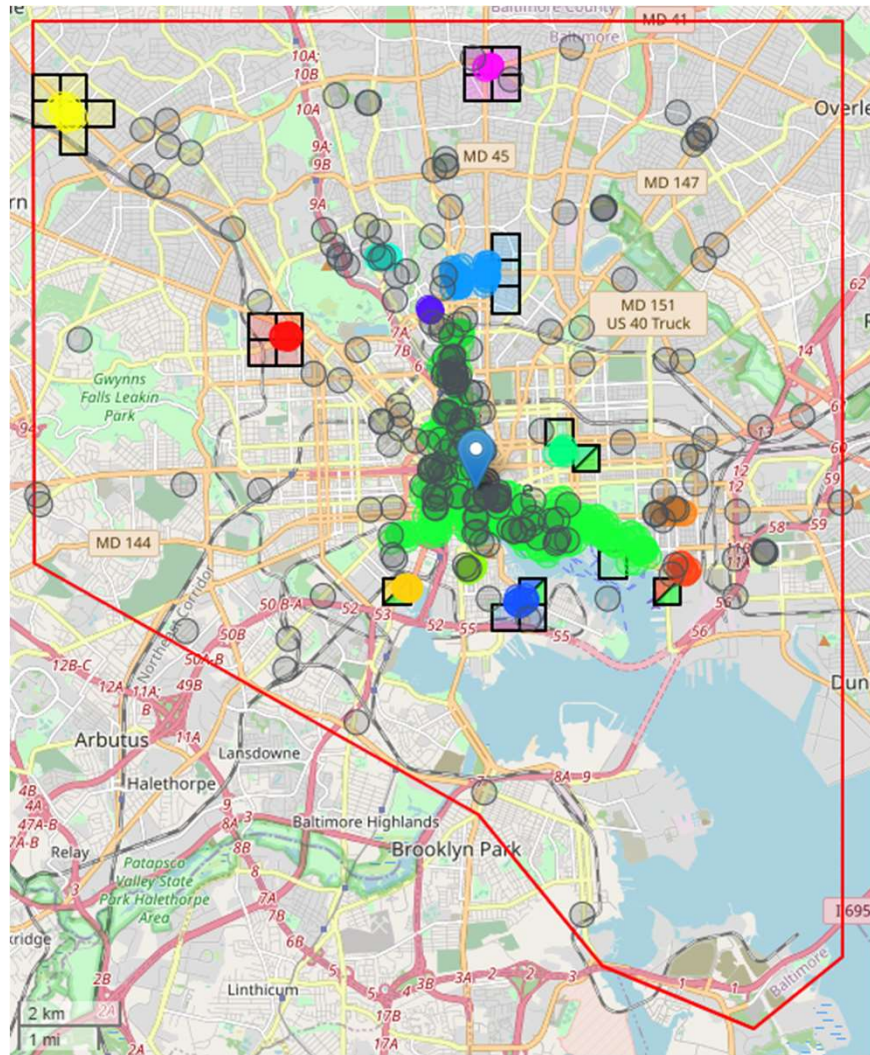
Find squares with low competition

- Identify 500m x 500m squares < .5km to traffic driving colored clusters, and with less than 2 competitors in 1km (grey circles).

- 25 locations (black squares) identified as highly viable and would benefit from further investigation.

- 4 of the identified locations straddle two nightlife clusters and are thus potentially more valuable

Conclusion

- We were able to find promising real estate near traffic driving clusters with reduced competition.

- Data inspection reveals there is likely additional noise in both traffic drivers and competitors:
  - Traffic driver data flaws included private, daytime only, or miscategorized venues that will not build nighttime traffic numbers.
  - Competitor data included daytime and private businesses that might not be valid competition for our client.
  - This additional cleaning could be done in later phases.

- Further research could be done to establish maximum music venues that might flourish in a neighborhood based on city demographics such as income, age, and population.

- HDBSCAN was the best clustering technique for data using latitude and longitude locales, required only a single parameter for use, and assisted in finding noisy points.