The Battle of Neighborhoods

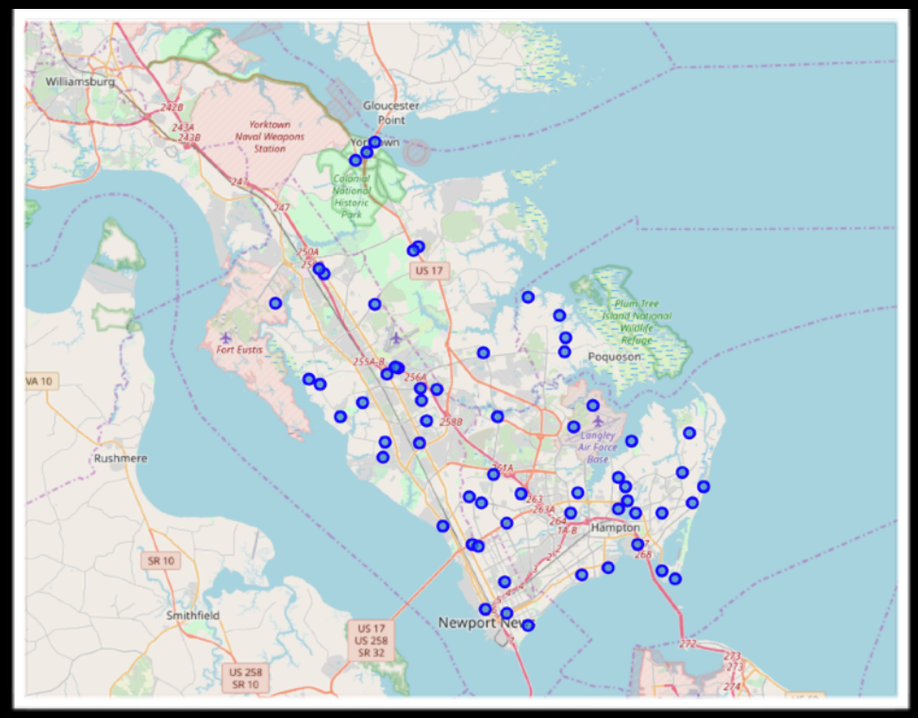
By: Jason Merten

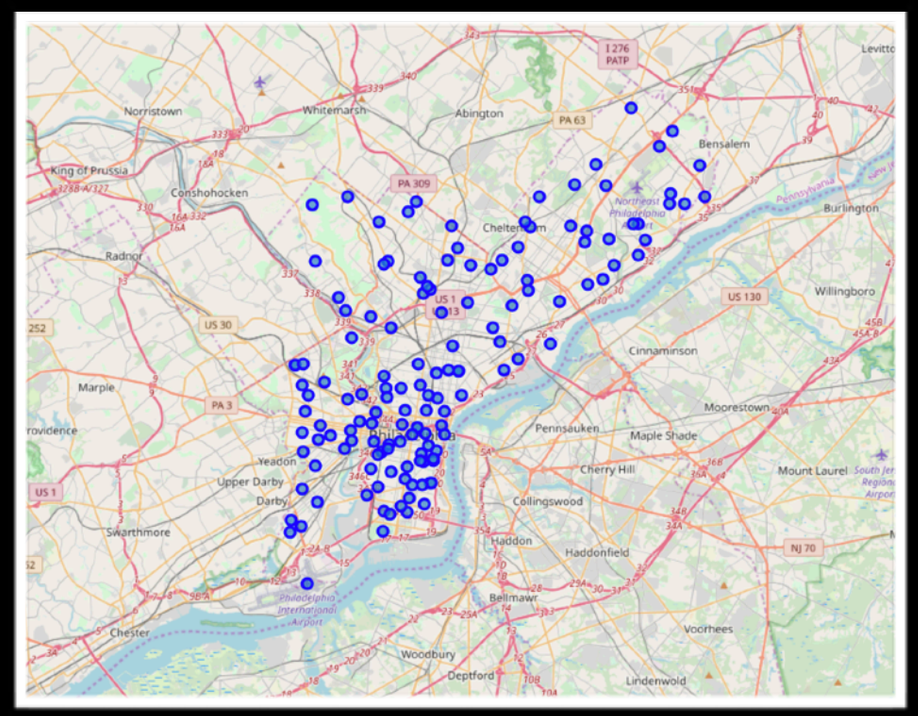
Introduction

- Conduct an analysis of what venue would be most likely to open in an area
- Base analysis on popular venues in two geographically dispersed cities
- Ensure both cities are in the same region (i.e. not New York and California)

Data

- Scrape the web for neighborhood data for each location
- Use Nominatim to get the Latitude/Longitude of each neighborhood
- Run neighborhood data into Foursquare to gather popular venue data





Methodologies

- Group venue data by venue category
- Perform one-hot encoding to find frequency of each category
- Determine the most popular categories in each city
- Run the data through k-means clustering to find similar groups/ neighborhoods

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	36th St	Convenience Store	Seafood Restaurant	Fried Chicken Joint	Discount Store	Park
1	48th St	American Restaurant	Fried Chicken Joint	Convenience Store	Park	Sandwich Place
2	Acree Acres	Pizza Place	Donut Shop	Coffee Shop	Mexican Restaurant	American Restaurant
3	Battle Park	Historic Site	History Museum	Sandwich Place	National Park	Seafood Restaurant
4	Big Bethel	Coffee Shop	Thai Restaurant	Movie Theater	Brewery	Gym

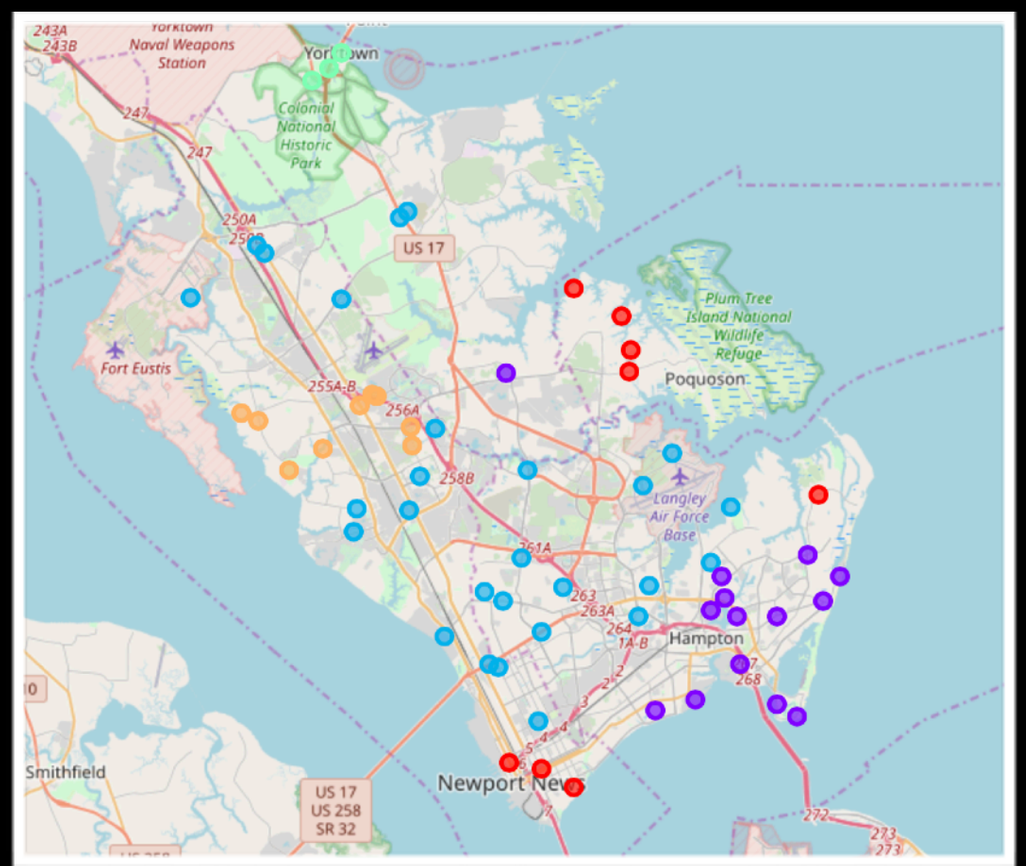
Hampton Roads Top Venues

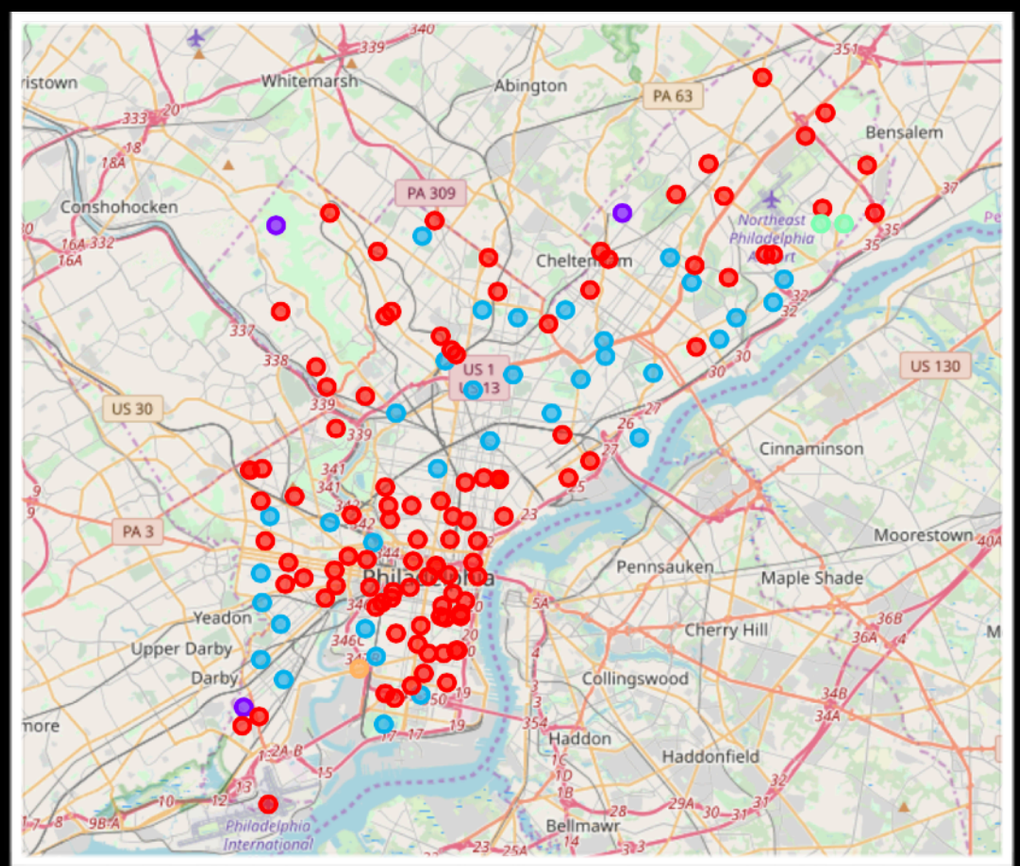
Philadelphia Top Venues

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Academy Gardens	Donut Shop	Garden	Farm	Zoo Exhibit	Farmers Market
1	Allegheny West	Intersection	Fast Food Restaurant	Sandwich Place	Grocery Store	Gym / Fitness Center
2	Andorra	Tennis Court	Playground	Zoo Exhibit	Dry Cleaner	Eastern European Restaurant
3	Angora	Park	Chinese Restaurant	Discount Store	Breakfast Spot	Light Rail Station
4	Ashton-Woodenbridge	Gym	Garden	Farmers Market	Dutch Restaurant	Eastern European Restaurant

Results

- There were a lot more data points for Philadelphia
- Modify the Foursquare search criteria before analysis
  - Philadelphia - 500m radius (more dense)
  - Hampton Roads - 5km radius (more dispersed)





Discussion

- A specific set of venues likely to open in Hampton Roads was attainable due to time constraints and limitations
- Instead, we can determine which venue category is most likely to grow based on the cluster sizes
- Hampton Roads is most likely to have an increase in parks and sports venues based on the similarities between the two cities and the large military population in the area

Hampton Roads	Cluster Name	Size
1	Convenience Store / Fast Food	Medium
2	Restaurant / Beach / Brewery	Medium
3	Miscellaneous	Large
4	Historic Site / Museum	Small
5	Pizza Restaurant	Medium

Philadelphia	Cluster Name	Size
1	Miscellaneous	Large
2	Playground	Small
3	Parks / Sports	Large
4	Cafe / Fast Food	Small
5	Art Gallery	Small

Conclusion

- Future ideas to improve the project:
  - Analysis of the brands present in each city
  - Weight the venues based on number of reviews/visits
  - Perform text analysis of the reviews to find key words

Thank you!