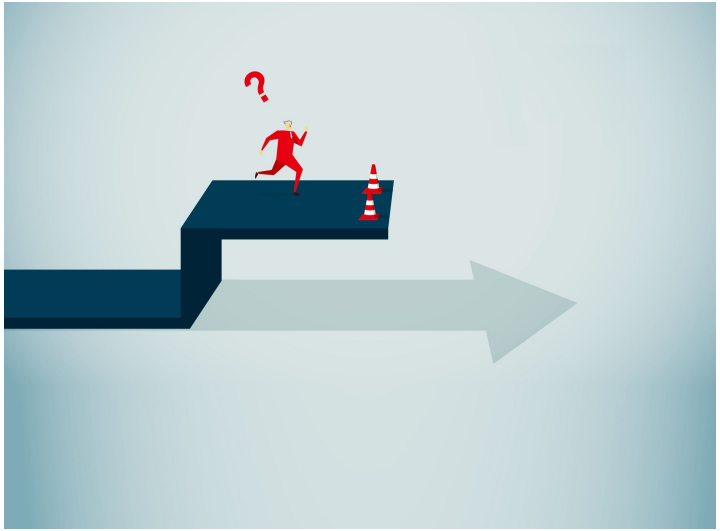
Identifying High Risk Neighborhoods in Boston

Problem

- COVID-19 has ravaged American society at an unprecedented rate
- Businesses forced to close or operate at limited capacity
- Obvious health risks limiting hospitals
- Social activity forced to close
- **Which areas to avoid COVID-19 exposure?**

Problem Limitations

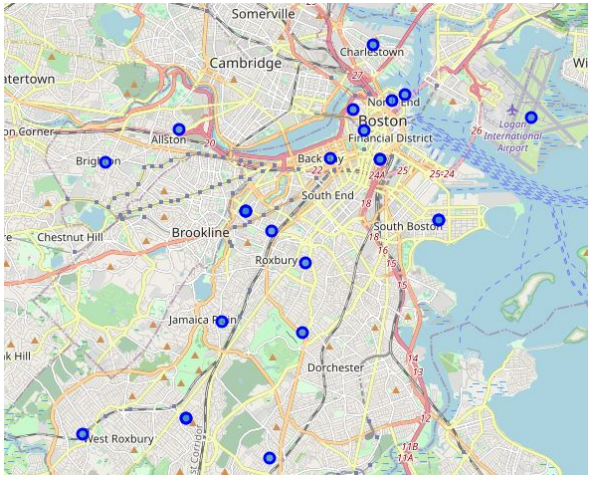
- Forced isolation difficult to enforce
- Masks offer limited protection
- People going outside regardless of risks
- Scale limited to one city: Boston



Data

- Location based data (Boston Neighborhoods by Zip Code)
- FourSquare API calls based on LAT/LNG values
  - Returning venues near each LAT/LNG value
- Folium for map display

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Beacon Hill	2	Coffee Shop	American Restaurant	Pizza Place	Italian Restaurant	French Restaurant	Sushi Restaurant	Museum	Restaurant	New American Restaurant	Plaza
1	Harbor Islands	0	Italian Restaurant	Pizza Place	Harbor / Marina	Park	Seafood Restaurant	Café	Pastry Shop	Outdoor Sculpture	Church	Diner
2	Leather District	2	Asian Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Sandwich Place	Theater	Café	Gym / Fitness Center	Restaurant
3	Chinatown	2	Asian Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Sandwich Place	Theater	Café	Gym / Fitness Center	Restaurant
4	North End	0	Italian Restaurant	Pizza Place	Seafood Restaurant	Park	Coffee Shop	Café	Sports Bar	Bakery	Pub	Sandwich Place



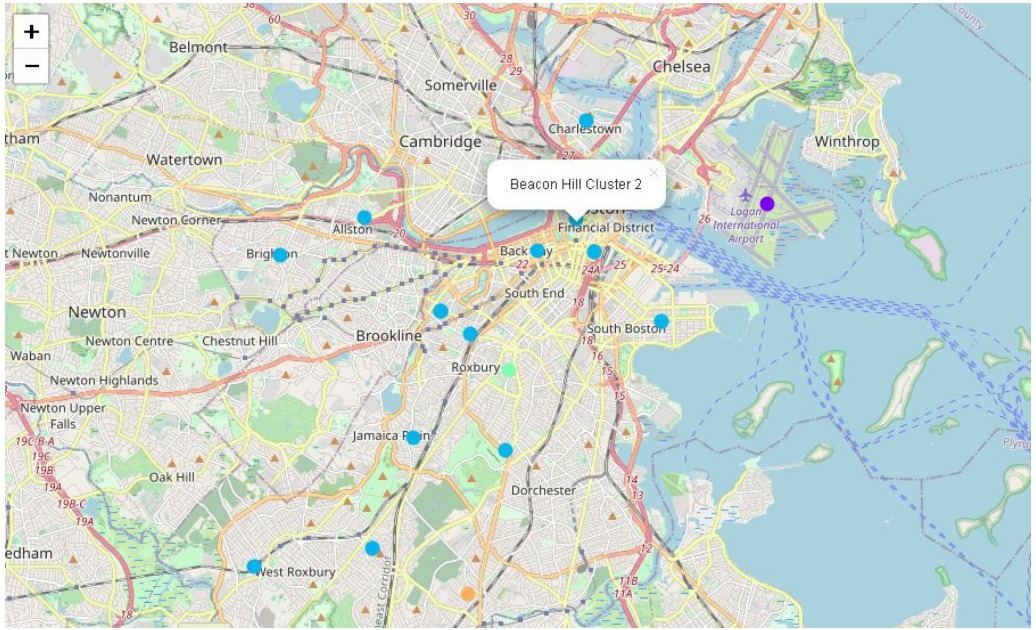
Solution

- Find areas of high COVID-19 exposure in Boston using location data
- Cluster areas together to perform in depth analysis

Methodology

- Dataframe with Latitude, Longitude, and neighborhood data
- Foursquare API call on LAT/LNG to get venues
- Score venues on 1-10 scale
- K-means clustering on venue data
- Analyze results

Results




```
Cluster 0 average score: 59.5
Cluster 1 average score: 54.0
Cluster 2 average score: 62.85
Cluster 3 average score: 48.0
Cluster 4 average score: 49.0
```

Solution Limitations

- API Call returned many unique categories instead of primary, overarching category
  - Forced to manually categorize each venue under a broader umbrella term
- Limitations of k-means application based on data
- Visual depiction of low risk, medium, high risk neighborhoods based on cluster group

Conclusion

- Problem statement answered at a base level
- Neighborhoods in cluster 2 most susceptible to COVID-19
- Neighborhoods with more bars/restaurants = higher risk