

# Correlation between Venues in Toronto Neighborhoods and Covid-19 Transmission

# 2020 – A special year

- World wide Covid-19 has infected over 84.8 millions of people and caused 1.84 million deaths
- Till December 30, 2020 there are totally 572982 Covid-19 cases and costed 15472 lives in Canada
- Ontario has the second highest total cases (178831) among all provinces and Toronto has 60000 cases since the beginning of pandemic.
- Is there any correlation between venues in Toronto neighborhoods and Covid-19 transmission?

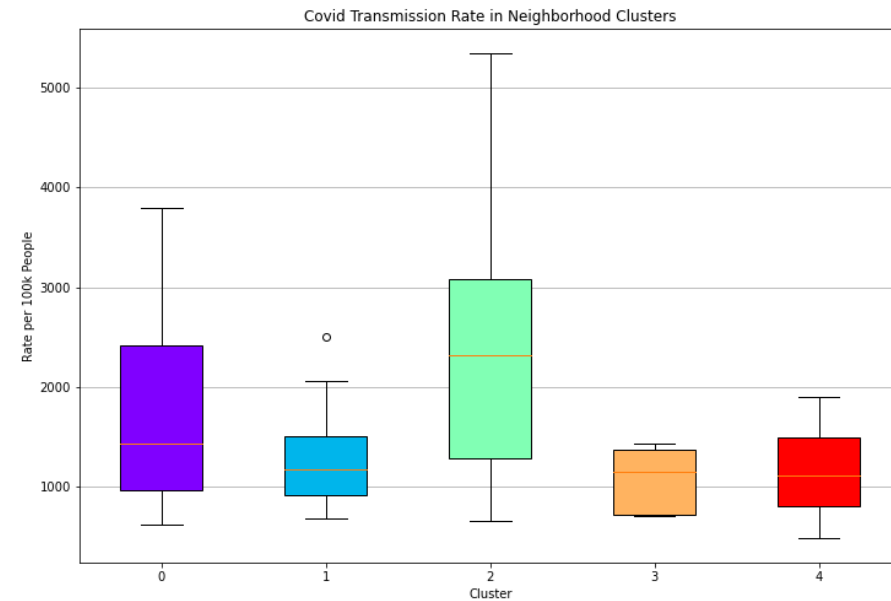
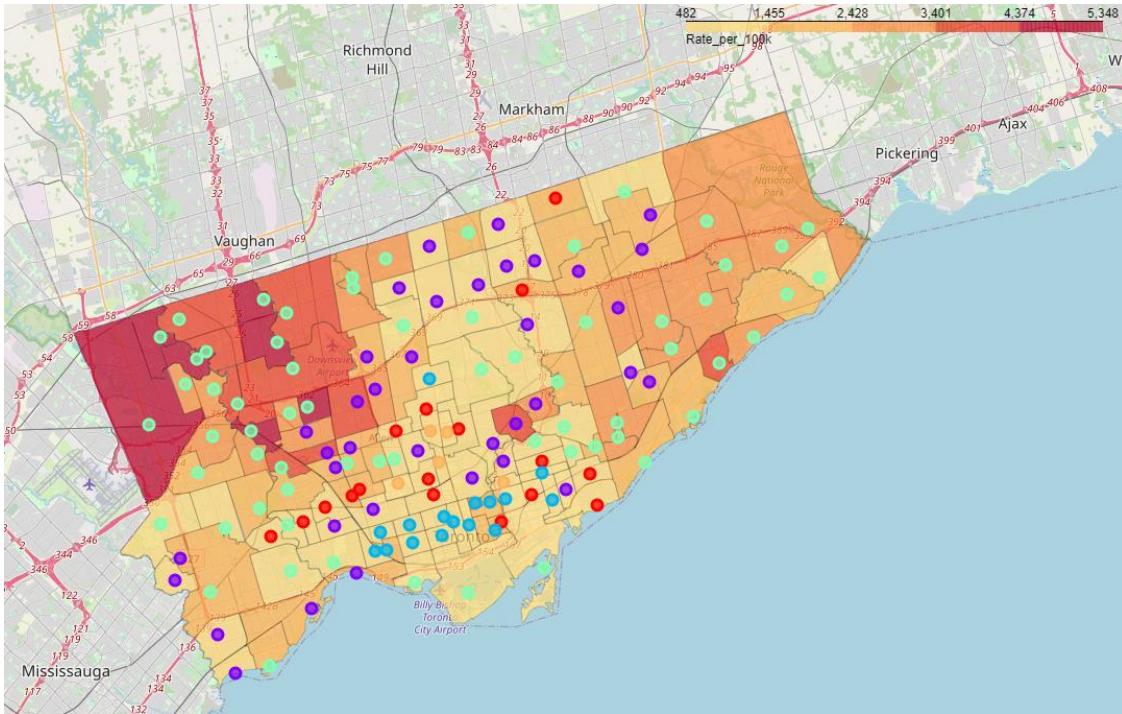
# Data Source and Cleaning

- Statistics of Covid-19 cases in Toronto as of Dec 22 from [www.toronto.ca](http://www.toronto.ca)
  - Assign preferable feature names
  - Calculate population of each neighborhood
- GEOJSON file for neighborhoods in Toronto from <https://open.toronto.ca/dataset/neighbourhoods/>
  - Extract coordinates and area of each neighborhood
  - Calculate coordinate and area of each neighborhood
- Combine above to datasets into 'toronto\_data'
  - Calculate radius of circle with equivalent area of neighborhood to pass to Foursquare

# Explore Neighborhoods in Toronto

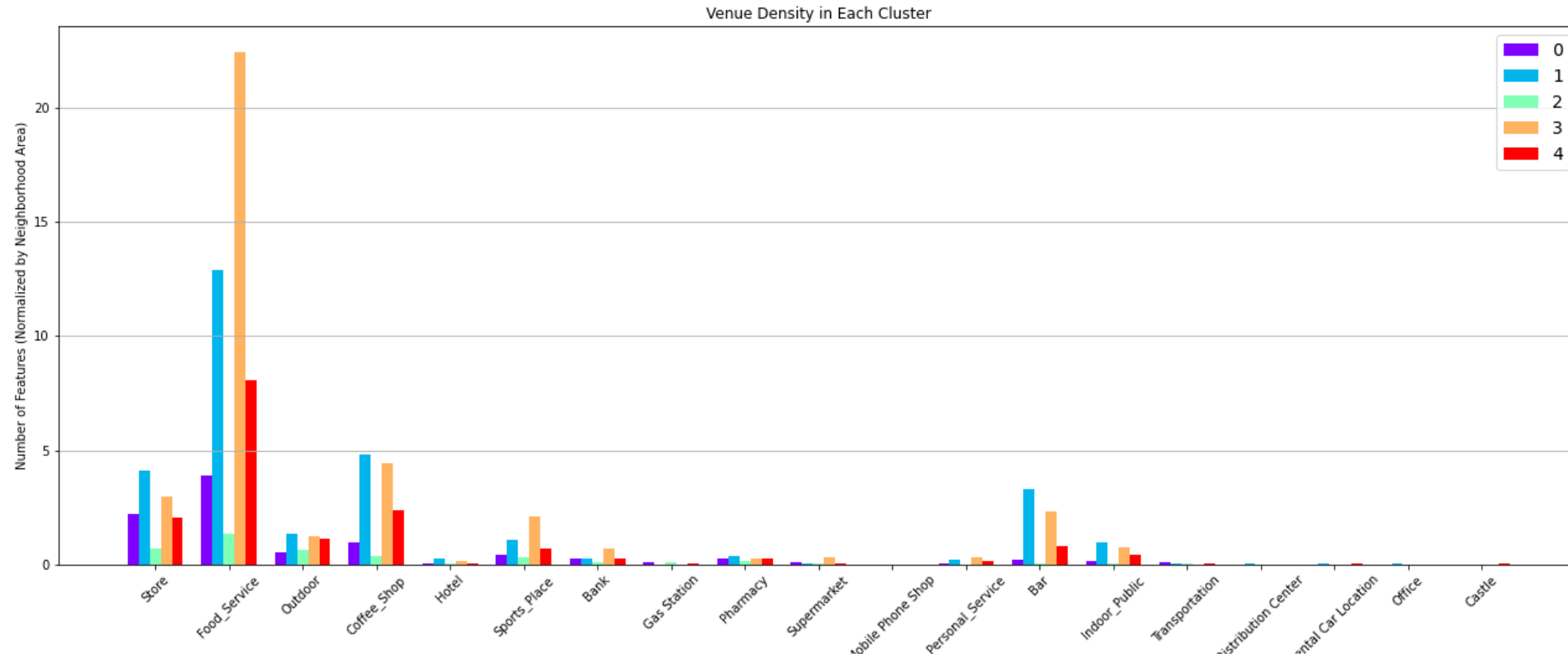
- Obtain venues in each neighborhood from Foursquare
- Perform EDA on venues
  - There are 328 different venue categories and quite a few has quantity of 1 or 2
  - Since neighborhoods have various of sizes, it makes sense to normalize number of venues with the area of neighborhood
- Group venue categories into bigger categories
- Normalize number of venues with the area of neighborhood

# Neighborhood Clustering



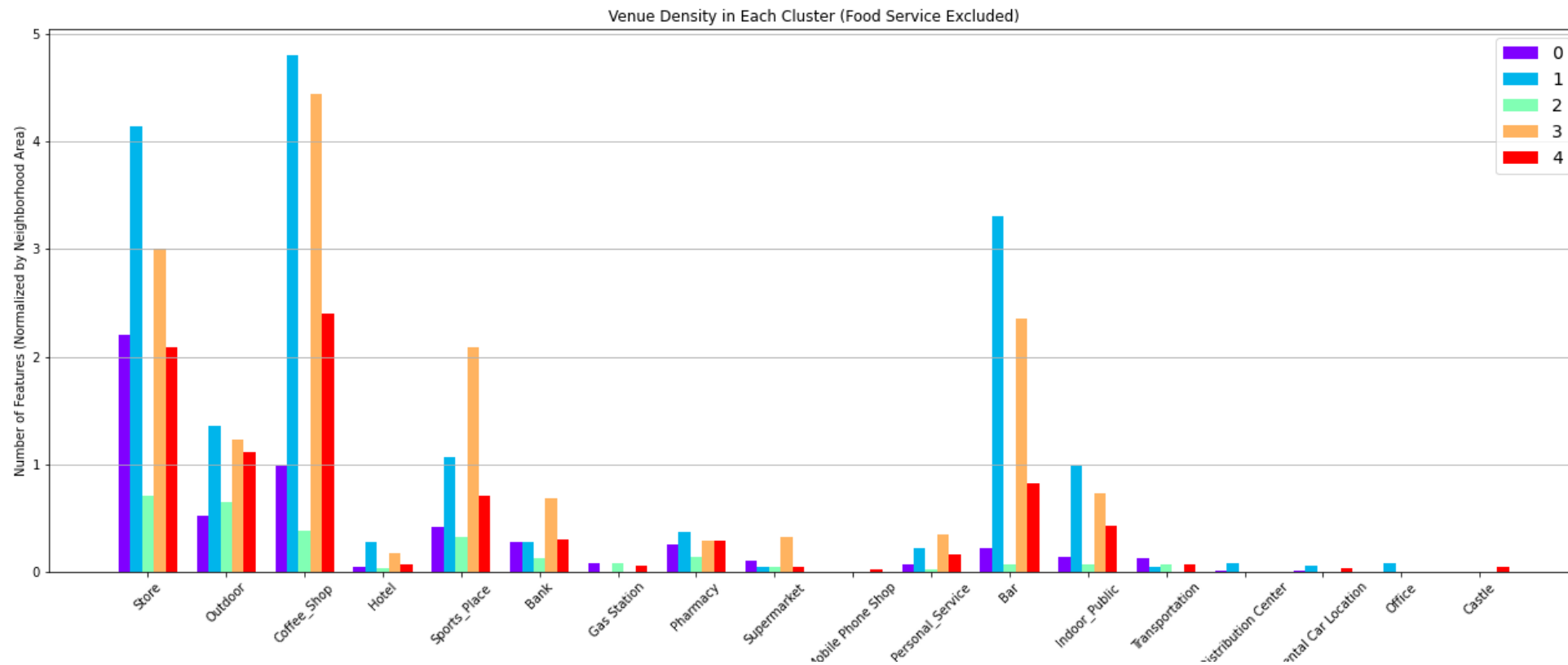
Green and purple clusters happen to have higher transmission rate. They are located outside the core area of Toronto.

# Venues in Each Cluster



Interestingly cluster green or purple does not have any venue that has significant higher density than other three clusters. In fact, cluster blue and yellow have very high venue density in food service, stores, coffee shops and bars because they are at the core of city.

# Venues in Each Cluster (Food Service Excluded)



# Conclusion

- We did not find strong correlation between venues in neighborhood and Covid-19 transmission rate
- However we have to keep in mind that:
  1. We used circle of equivalent area as boundary of neighborhood instead of the real boundary.
  2. Foursquare limited number of venues per API call to 50. Therefore for each neighborhood we can only obtain 50 venues.