



Students Need Their Coffee: Post Pandemic

DYLAN LINTHORNE

Universities Affecting New Cafes

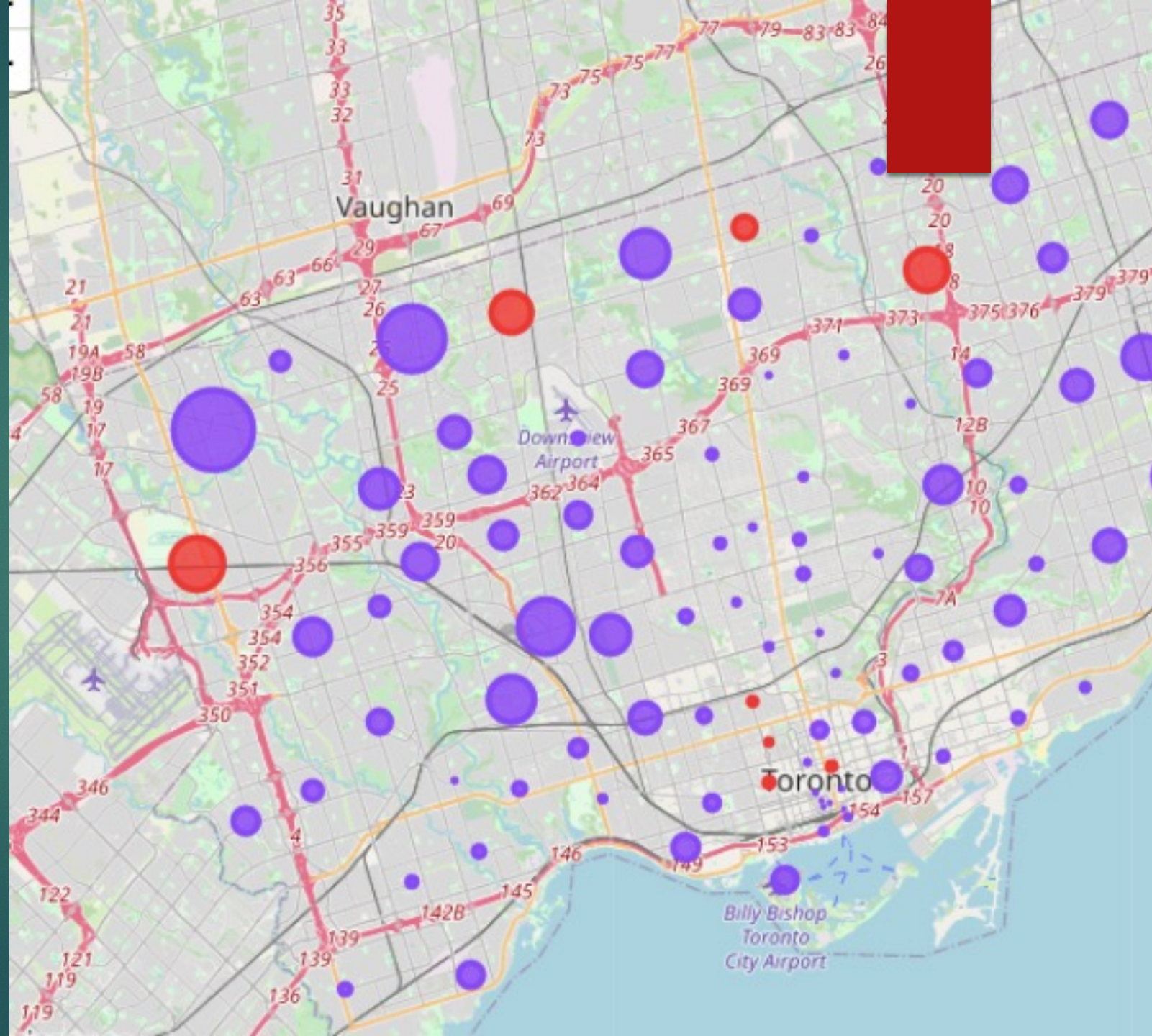
- ▶ Potential coffee shop owners might want to take advantage of the university student's consumer habits.
- ▶ Large chained coffee shops, such as, Starbucks and Second Cup might already have large share of the local markets near universities.
- ▶ On top of that, the pandemic has affecting smaller businesses more than larger chains.
- ▶ A survey was conducted to give future shop owners and stackholders the best possible option to open up shop

Data Acquisition

- ▶ Toronto neighborhoods were scrapped from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and geographical coordinates from geocoder
- ▶ University by postal code and enrolment numbers were individual taken from each schools official website
- ▶ Toronto covid case data was scrapped from the Toronto Open Data Catalogue, scrapped using their own API.
- ▶ Covid cases were normalized by the total amount of cases declared as “community spread”

Mapping it out

- ▶ GTA neighborhoods represented with each marker
- ▶ Red markers host a local university
- ▶ The size of the marker is proportional to the cumulative local covid cases.

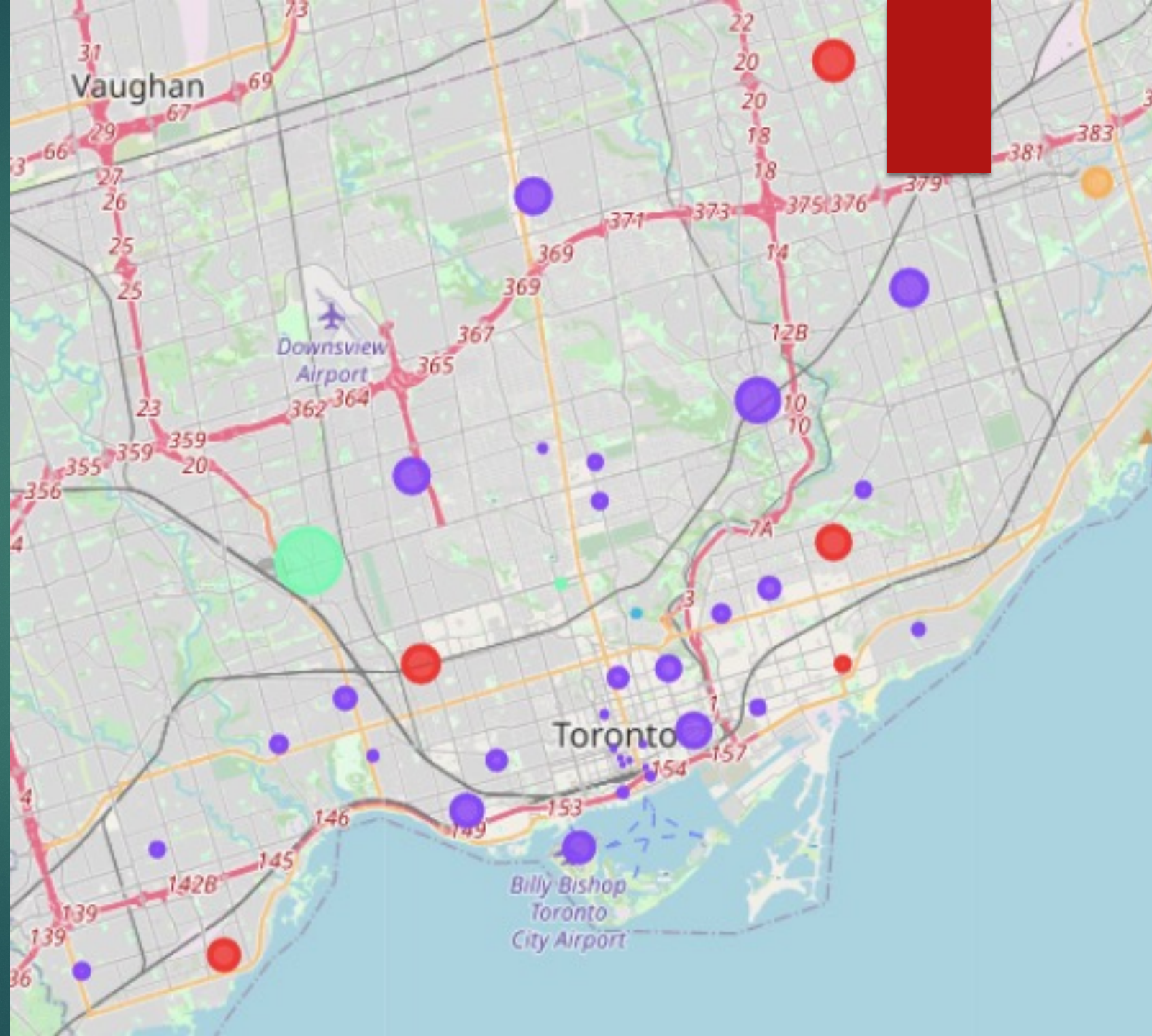


Clustering Analysis

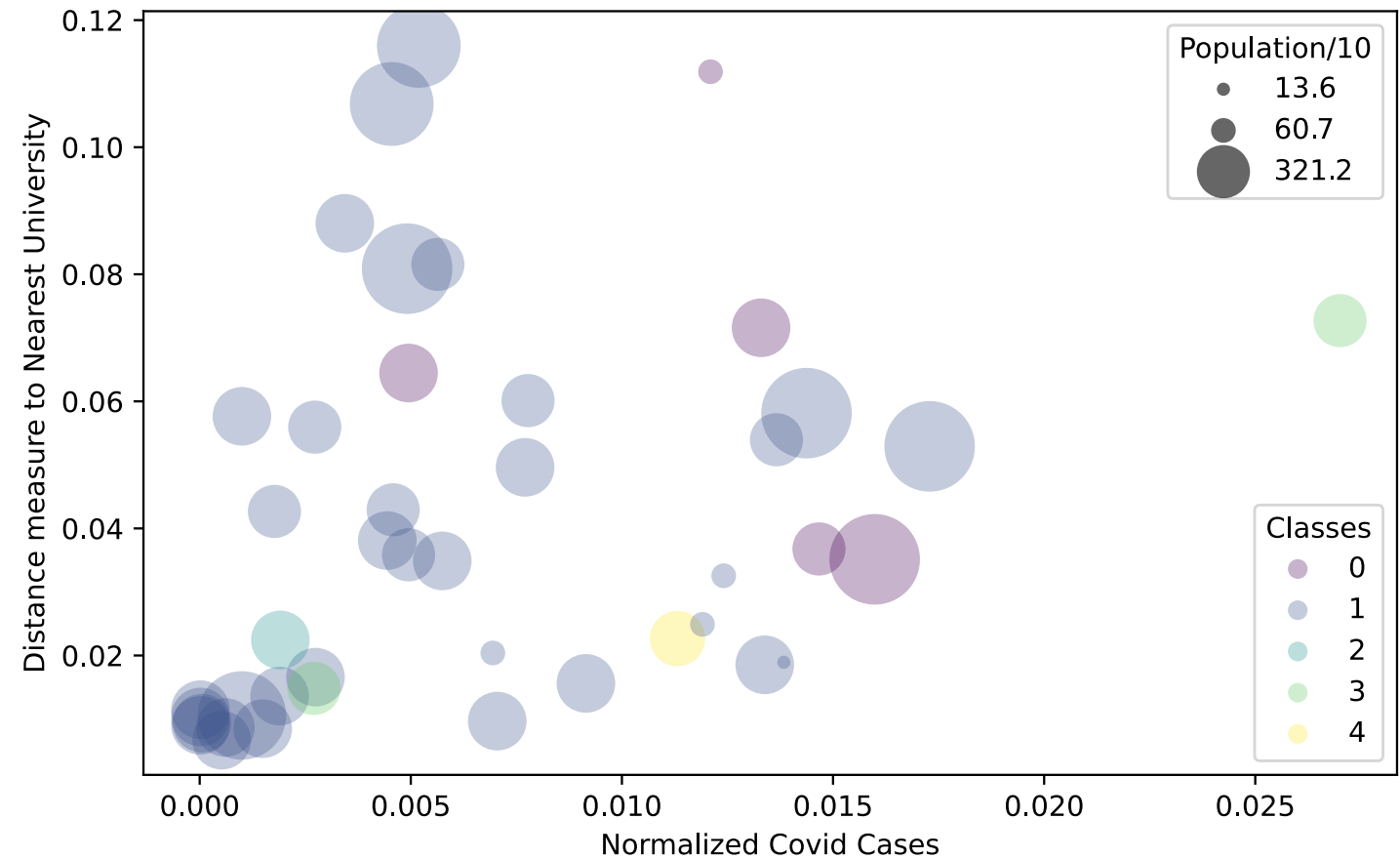
- ▶ Foursquare data was queried in each neighborhood to determine the coffee shops with the top market shares within a 500m radius.
- ▶ A k-means clustering algorithm was employed to segment the areas using the top coffee shop occurrences.
- ▶ $k = 5$ was taken, resulting in 5 distinct (non-overlapping) neighborhood segments.
- ▶ Neighborhoods were subsequently rated by how close they were to a university institution.
- ▶ The Euclidean distance was calculated for each university and neighborhood to find the closest university for each postal code.

Clusters

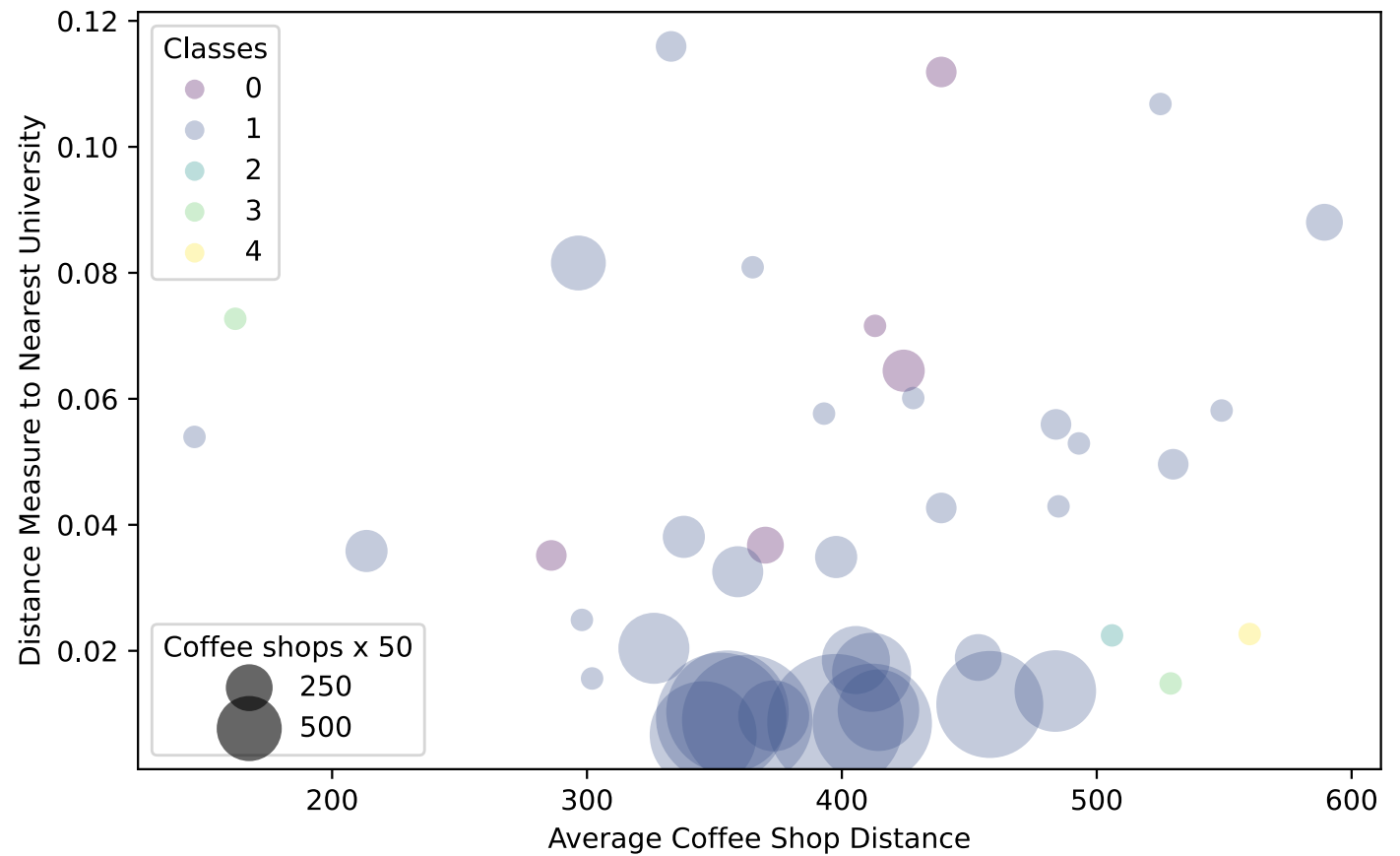
- ▶ Cluster 0 (red): large chains, close to DT core.
- ▶ Cluster 1 (purple): large chains, and highly dense
- ▶ Cluster 2 (blue): no chains, close to DT,
- ▶ Cluster 3 (green): only large chains.
- ▶ Cluster 4 (orange): no chains, further from Dt core.



- ▶ Radii of markers proportional to student population of nearest university
- ▶ Euclidean distance to calculate the distance to nearest school
- ▶ Cluster 4 shows the best compromise between covid cases and distance to schools for a given population.

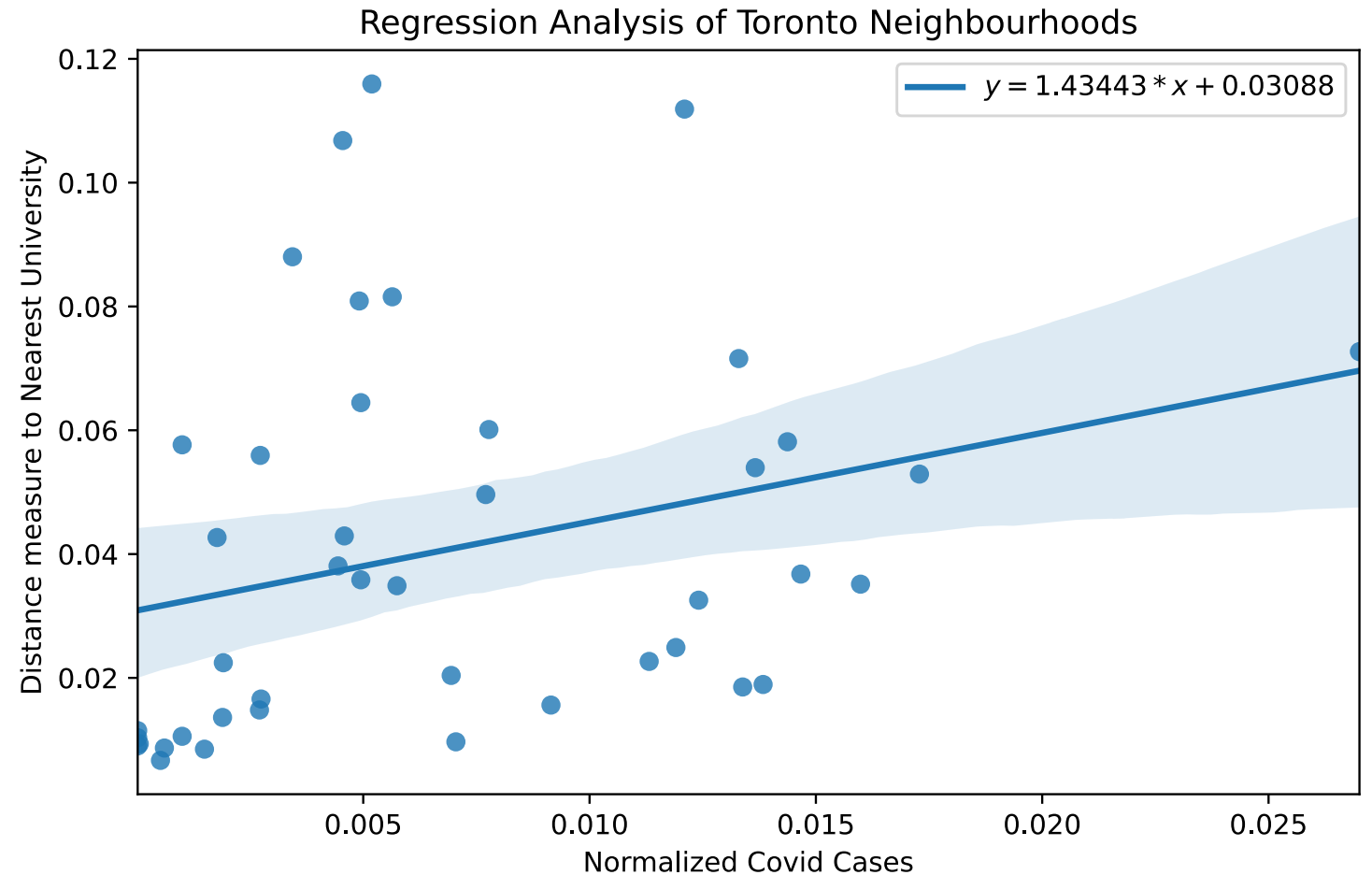


- ▶ Radii of markers are proportional to the local coffee shop density
- ▶ Cluster 1 gives the worst case scenario for a new shop owner with a highly dense area of large chains.
- ▶ Clusters 2 & 4 are near a university with only few shops within their 500m radius



Expecting a Trend

- ▶ Regression analysis employed on suspected correlations.
- ▶ Model : $y = ax + b$
- ▶ CC: 0.028
- ▶ P-value: 0.06
- ▶ Not statistically significant enough to trust as a realistic trend.



Summary

- ▶ K-means clustering of Toronto café shares in each postal code
- ▶ Given the number of cumulative covid cases and university proximity clusters 2 and 4 depicted the ideal locations for new shop owners.
- ▶ Insufficient statistical significance for any trends in covid rates and distance to their local university.
- ▶ Future work would be incorporate data from similar urban populations to validate clusters inference.