

Exploring Toronto Neighbourhood COVID-19 Rates and Venue Types



Andrew Ritchie

<https://www.linkedin.com/in/andrewmritchie/>

[Github.com/AndrewRitchie05](https://github.com/AndrewRitchie05)

[Kaggle.com/AndrewRitchie](https://kaggle.com/AndrewRitchie)

Table of Contents:

1. Introduction
2. Data
3. Methodology
4. Results
5. Discussion
6. Conclusion

Introduction

Over the course of the past year, COVID-19 has upended the lives and livelihoods of billions. In an effort to contain the spread and limit the strain placed on healthcare systems, many governments have engaged in “lockdowns” of varied methodology and severity. One such methodology has been lockdowns based on the forced closure (or limiting capacity) of businesses/venues based on the category the business/venue falls under, and the perceived risk of COVID-19 transmission for that business/venue category. This typically has significant implications for business owners in the form of lost revenues, employees in the form of lost income, taxpayers in the form of increased welfare burden, and the public in the form of lost freedom.

In many cases, for varying reasons beyond the scope of this report, it has been difficult for decision makers to develop and offer a data-based methodology to explain the rationality behind lockdowns affecting businesses and venues. This report attempts to develop a framework through which a methodology can be created to solve this problem, and allow for a more data-based approach to the closure (or limiting of capacity) for businesses and venues.

Data

We will use data sourced from:

- Scraping of Toronto Postal Code data from Wikipedia
Accessed at: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Contains data related to postal codes beginning with “M”, which corresponds to Toronto, Ontario. The dataset includes the Postal Code, Borough, and Neighbourhood.

- Foursquare venue data for Toronto, accessed through their API
Accessed at: <https://api.foursquare.com>

Dataset contains Venue Category, Postal Code, Neighbourhood, Borough Latitude, Borough Longitude, Venue Name, Venue Latitude, Venue Longitude, and Venue Category

- COVID-19 rates for neighbourhoods in Toronto, sourced from ICES
Accessed at: <https://www.ices.on.ca/~media/Files/COVID-19/ICES-COVID19-Testing-Data-FSA-percent-positivity.ashx?la=en-CA>

This dataset contains positivity rates for Toronto postal codes on a weekly basis, spanning the period August 2020 to January 2021. It contains positivity rates for both the overall population, and non-LTC (Long Term Care) individuals. For non-LTC individuals, the postal code associated with their Ontario Health Care card closest to the date of testing was used. For our purposes, we have elected to use the positivity rates for the non-LTC population as spread of the virus within LTC populations is a separate issue beyond the scope of this report.

The positivity rates for the non-LTC population were grouped by postal code, and then the median value of weekly positivity rates was calculated for each neighbourhood.

- Geospatial coordinate data, provided by Coursera
Accessed at: https://cocl.us/Geospatial_data

This dataset contains Postal Code, and the corresponding Longitude and Latitude.

Datasets were combined as necessary to provide the data frames we required for our analysis.

Methodology

Once we had retrieved the venue data using the Foursquare API, we set about grouping the venue types. The data collected from Foursquare contained many different venue types of similar characteristics, so we developed a methodology to group venues of similar characteristics. This grouping methodology attempted to create groups that are representative of groups as they have been targeted by COVID-19 “lockdowns”. Please note that there are cases where a venue does not fit perfectly in its category but was deemed *more* similar to venues in that category than to venues in any other category. The resulting venue group types were Hospital, Church, Restaurants, Fast Food, Retail, Indoor Rec, Outdoor Rec, Licensed Establishment, Personal Services, Travel, Indoor Venues, Commercial, and Essential Retail. Churches and Hospitals could potentially have been mapped to indoor venues, however due to Churches (places of worship in general) often being treated as their own category during lockdowns and the essential nature of Hospitals in the COVID-19 pandemic, we have elected to have these venues be their own distinct venue categories. For the remainder of the venue categories, they were mapped according to the following methodology:

Restaurants:

Venues whose primary business is food service, with an emphasis on sit-down dining over take-out.

['Greek Restaurant', 'Italian Restaurant', 'Restaurant', 'Tibetan Restaurant', 'Caribbean Restaurant', 'American Restaurant', 'Sushi Restaurant', 'Steakhouse', 'Seafood Restaurant', 'Middle Eastern Restaurant', 'Thai Restaurant', 'Latin American Restaurant', 'Mexican Restaurant', 'Chinese Restaurant', 'Vietnamese Restaurant', 'Japanese Restaurant', 'Taiwanese Restaurant', 'Theme Restaurant', 'Ramen Restaurant', 'Ethiopian Restaurant', 'Mediterranean Restaurant', 'Korean Restaurant', 'French Restaurant', 'Asian Restaurant', 'Modern European Restaurant', 'New American Restaurant', 'Vegetarian / Vegan Restaurant', 'Moroccan Restaurant', 'Belgian Restaurant', 'Portuguese Restaurant', 'Colombian Restaurant', 'Gluten-free Restaurant', 'Brazilian Restaurant', 'Filipino Restaurant', 'Molecular Gastronomy Restaurant', 'Cuban Restaurant', 'Indian Restaurant', 'Comfort Food Restaurant', 'Breakfast Spot', 'German Restaurant', 'Eastern European Restaurant', 'Noodle House', 'Dumpling Restaurant', 'Cajun / Creole Restaurant', 'Diner', 'Bistro']

Fast Food:

Venues whose primary business is food service, with an emphasis on take-out over sit-down dining.

['Ice Cream Shop', 'Pizza Place', 'Juice Bar', 'Dessert Shop', 'Bubble Tea Shop', 'Coffee Shop', 'Café', 'Frozen Yogurt Shop', 'Fish & Chips Shop', 'Fast Food Restaurant', 'Sandwich Place', 'Fried Chicken Joint', 'Bagel Shop', 'Snack Place', 'Burger Joint', 'Burrito Place', 'Tea Room', 'Poutine Place', 'Food Truck', 'BBQ Joint', 'Poke Place', 'Salad Place', 'Donut Shop', 'Soup Place', 'Taco Place', 'Doner Restaurant', 'Smoothie Shop', 'Airport Food Court', 'College Cafeteria', 'Creperie', 'Falafel Restaurant', 'Food Court']

Retail:

Store venues, not of an essential nature.

['Board Shop', 'Cosmetics Shop', 'Bookstore', 'Furniture / Home Store', 'Pet Store', 'Stationery Store', 'Clothing Store', 'Department Store', 'Jewelry Store', 'Gift Shop', 'Adult Boutique', 'Men's Store', 'Electronics Store', 'Antique Shop', 'Comic Shop', 'Shopping Mall', 'Miscellaneous Shop', 'Video Game Store', 'Other Great Outdoors', 'Lingerie Store', 'Tailor Shop', 'Flower Shop', 'Arts & Crafts Store', 'Record Shop', 'Baby Store', 'Athletics & Sports', 'Flea Market', 'Garden Center', 'Boutique', 'Sporting Goods Shop', 'Toy / Game Store', 'Discount Store', 'Shoe Store']

Indoor Rec:

Venues where recreational activities are primarily conducted indoors.

['Yoga Studio', 'Gym', 'Gym / Fitness Center', 'Swim School', 'Dance Studio', 'Martial Arts School', 'College Rec Center', 'Skating Rink', 'College Gym', 'Pool', 'Climbing Gym']

Outdoor Rec:

Venues where recreational activities are primarily conducted outdoors.

['Summer Camp', 'Tennis Court', 'Bike Rental / Bike Share', 'Park', 'Playground', 'Dog Run', 'Sculpture Garden', 'Plaza', 'Lake', 'Fountain', 'Monument / Landmark', 'Scenic Lookout', 'Garden', 'Skate Park', 'Trail', 'Neighborhood', 'Historic Site', 'Beach']

Licensed Establishment:

Venues where the primary service offered is the sale and consumption of liquor on-site.

['Pub', 'Brewery', 'Lounge', 'Gay Bar', 'Gastropub', 'Bar', 'Beer Bar', 'Sake Bar', 'Strip Club', 'Wine Bar', 'Hookah Bar', 'Cocktail Bar', 'Jazz Club', 'Irish Pub', 'Nightclub', 'Speakeasy', 'Sports Bar', 'Hotel Bar', 'Airport Lounge', 'Roof Deck']

Personal Services:

Venues where services offered are of a personal grooming nature.

['Spa', 'Salon / Barbershop', 'Health & Beauty Service', 'Tanning Salon', 'Optical Shop', 'Home Service']

Travel: Venues whose primary purpose is the facilitation of travel.

['Cable Car', 'Bus Line', 'Light Rail Station', 'Train Station', 'Airport', 'Harbor / Marina', 'Airport Terminal', 'Airport Gate', 'Plane', 'Airport Service', 'Rental Car Location', 'Boat or Ferry', 'Hotel', 'Distribution Center', 'General Travel', 'Intersection', 'Bed & Breakfast']

Indoor Venues:

['Movie Theater', 'General Entertainment', 'Theater', 'Event Space', 'Music Venue', 'Concert Hall', 'Basketball Stadium', 'Baseball Stadium', 'Indie Movie Theater', 'Gaming Cafe', 'Opera House', 'Stadium', 'College Auditorium', 'Performing Arts Venue', 'Art Gallery', 'Museum', 'Art Museum', 'Aquarium', 'History Museum', 'College Arts Building']

Commercial: Venues related to commerce.

['Office', 'Building', 'IT Services', 'Coworking Space', 'Business Service', 'Recording Studio']

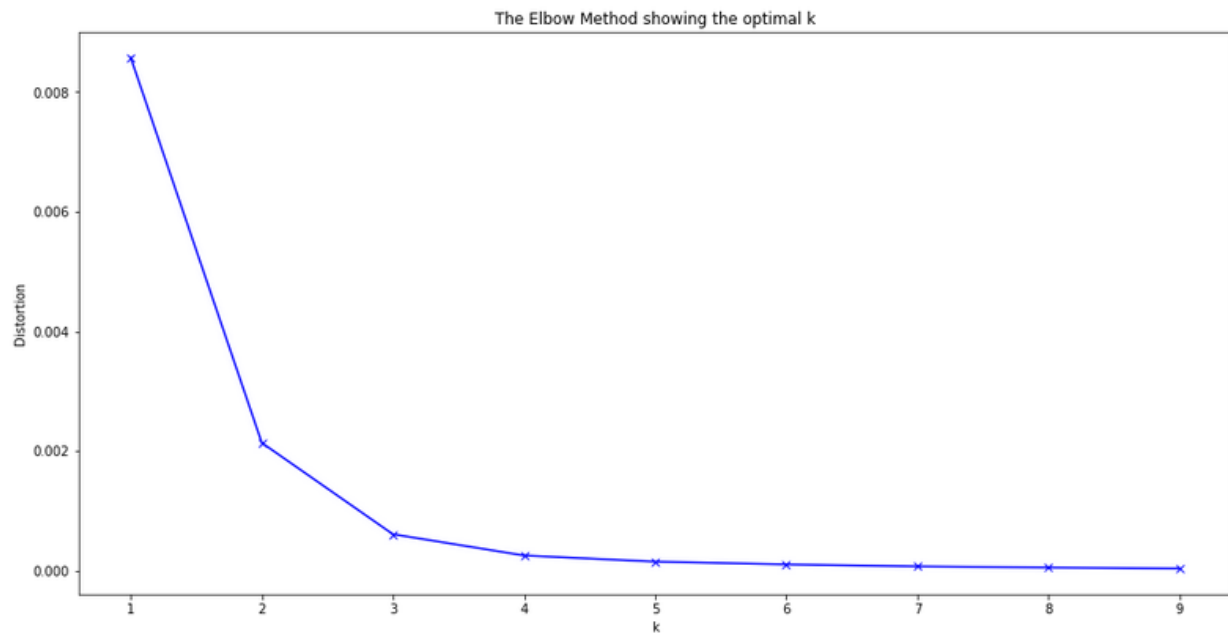
Essential Retail:

Stores or venues of an (immediately) essential nature. We included stores which sell foods which may be considered by some to be “luxury”, as we believe luxury is subjective and very difficult to universally define. We believe these “luxury” stores are closer characteristically to other venues in this category than any other category. We also included stores related to vices (liquor and tobacco), as well as Banks and Auto Workshops.

['Health Food Store', 'Fruit & Vegetable Store', 'Bakery', 'Grocery Store', 'Food & Drink Shop', 'Fish Market', 'Cheese Shop', 'Convenience Store', 'Pharmacy', 'Food', 'Gas Station', 'Farmers Market', 'Gourmet Shop', 'Supermarket', 'Butcher', 'Deli / Bodega', 'Market', 'Beer Store', 'Smoke Shop', 'Chocolate Shop', 'Wine Shop',

'Cupcake Shop', 'Organic Grocery', 'Candy Store', 'Liquor Store', 'Bank', 'Auto Workshop']

Once we had grouped venues according to the above methodology, we set about clustering postal codes by their positivity rate. This was achieved using the K-Means Clustering algorithm. We iterated over many values of K, which resulted in the following chart.



Using the elbow method, we arrived at an optimal value of $K = 4$, which we then used to implement the K-Means Clustering algorithm.

From there, a WordCloud was created for each positivity rate cluster, in an attempt to visualize the effect venues have on COVID-19 transmission. The word cloud contains the venue groups ranked by their relative presence in the cluster.

Finally, a correlation matrix was produced for the entire city so that we could explore the correlation between positivity rates and venue types.

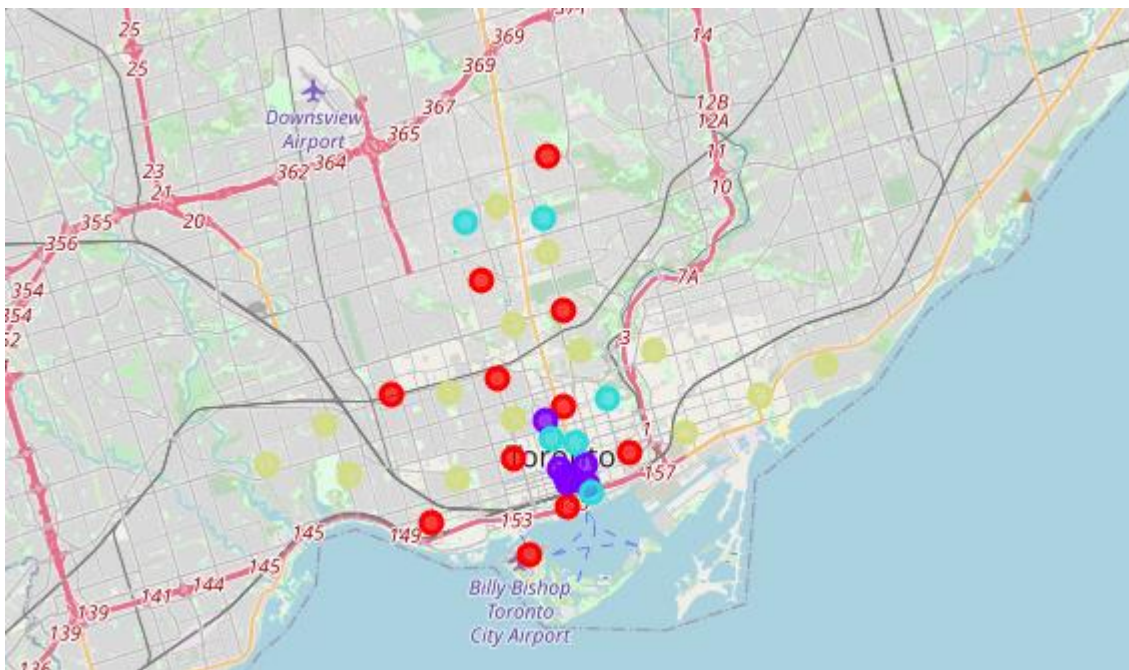
Results

K-Means Clustering for postal codes resulted in clusters with the following (median) positivity rate ranges:

- Cluster 1 3.12% - 4.05%
- Cluster 2 0%
- Cluster 3 4.42% - 5.37%
- Cluster 4 1.92% - 2.985%

The postal codes and their cluster value can be seen on the map below.

Toronto Cluster Map



Red: Cluster 1, Purple: Cluster 2, Blue: Cluster 3, Yellow: Cluster 4

WordClouds

The WordClouds for each cluster, showing the relative prominence of venue types in each cluster is shown below:



While we cannot make any statistically sound inferences from the WordClouds for each cluster, it is nonetheless informative for us to see that the relative prominence of venues in an area do appear to be related to the spread of COVID-19 in that area.

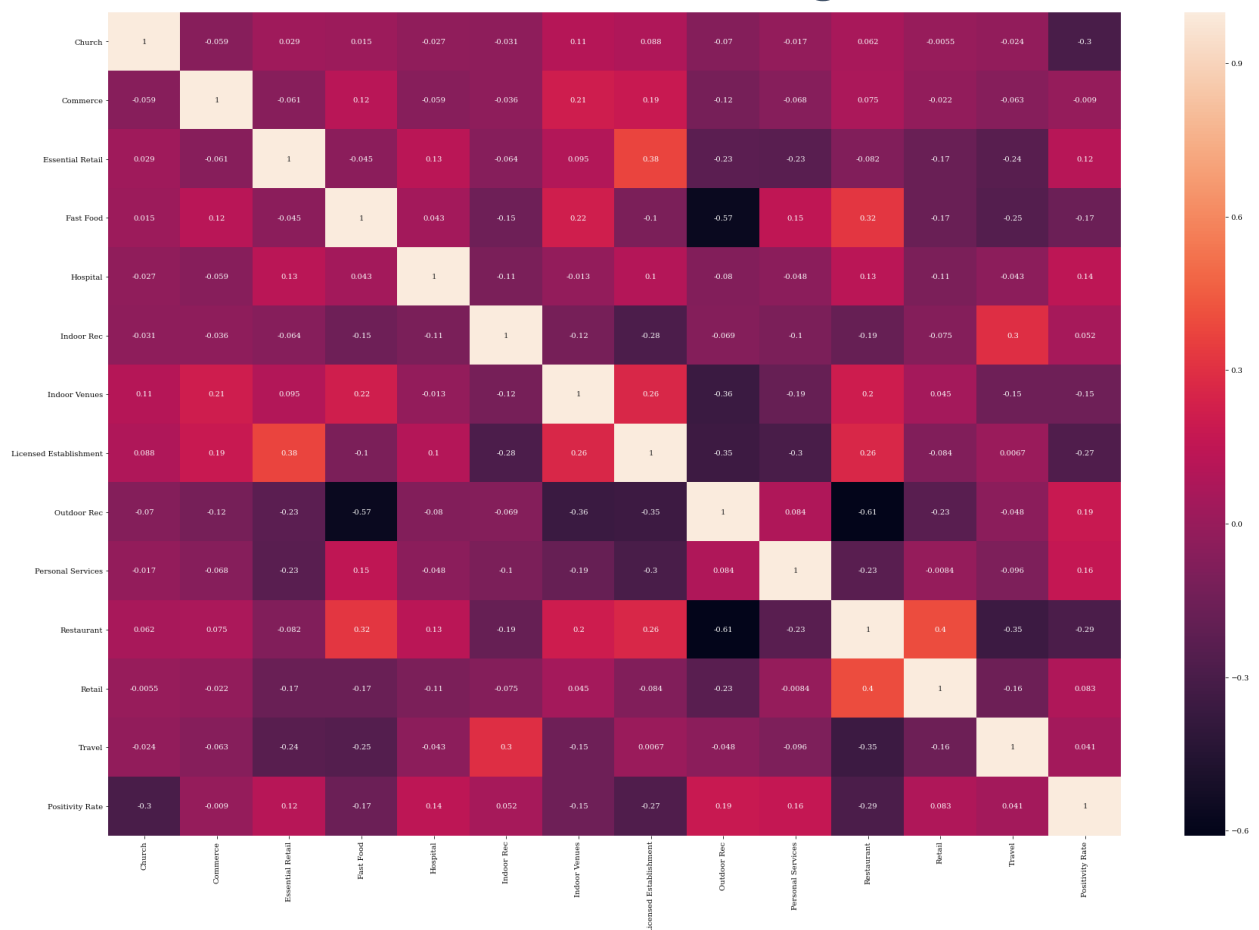
Correlation matrices

We began examining the correlation of venue prominence with COVID-19 positivity rates by calculating a correlation matrix for all postal code clusters in our dataset.

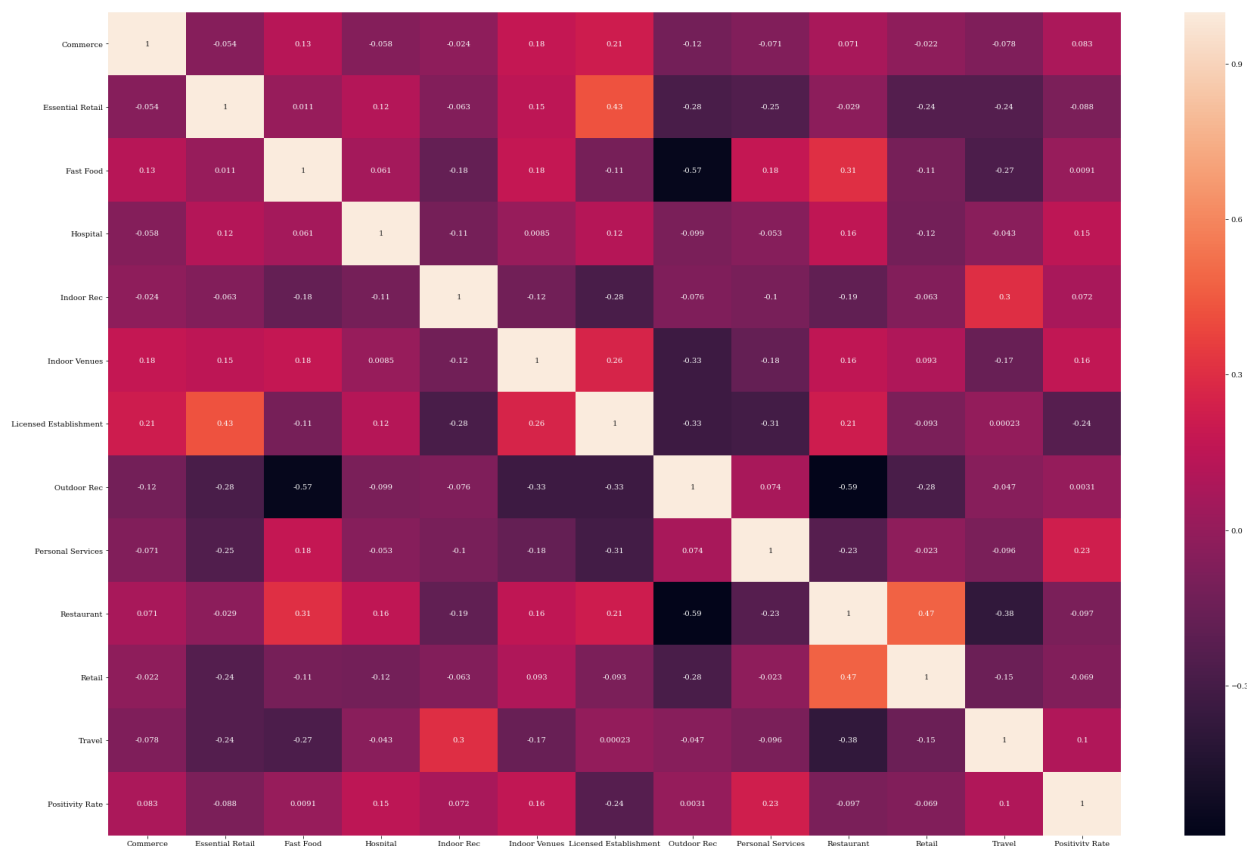
We then calculated the same correlation matrix, without the cluster containing zero positivity rate postal codes. Once the zero-rate cluster had been removed, the remaining dataset contained no presence of Churches, so Churches were also removed prior to calculating the correlation matrix.

Both Correlation Matrices are presented below.

Correlation Matrix Including All Clusters

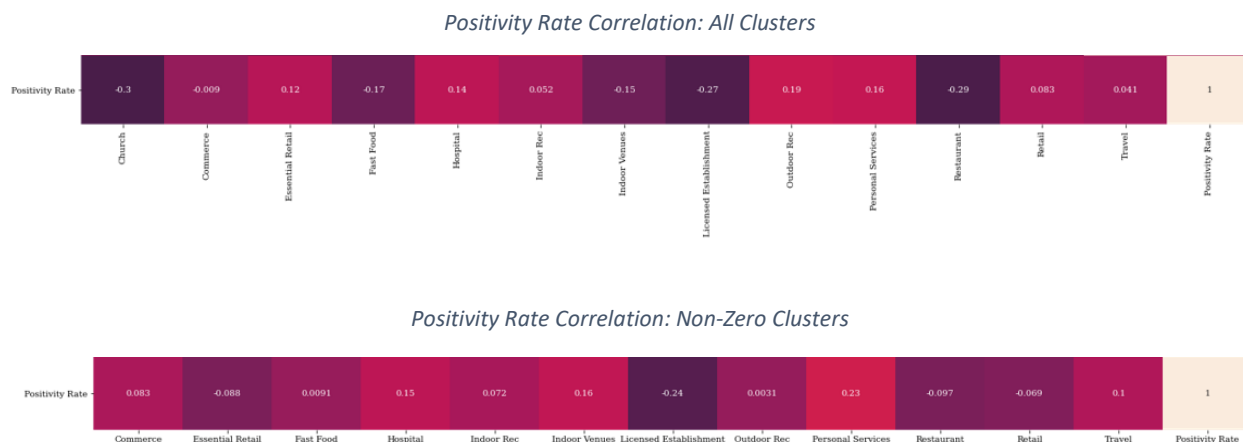


Correlation Matrix Without Zero-Rate Cluster



Discussion

Note: *Some of our results appear highly contradictory of current beliefs, and as a result we urge caution to the reader in interpreting them. The goal of this report is to attempt creation of a framework, and further adjustments and additions to the framework will likely be required.*



The correlation matrices can be interpreted as follows; a positive value indicates that a higher presence of that venue type in an area contributes to the spread of COVID-19 in that area; a negative value indicates the opposite, that increased presence of the venue reduces the spread of COVID-19 in that area. The magnitude of the value indicates the magnitude of the effect.

Looking at the results of the correlation matrices, we find results that are unsurprising as well as ones which appear very surprising.

When including all clusters, the results indicate that the prominence of churches in an area appear to have not only a negative, but the strongest negative correlation with COVID-19 positivity rates. In addition, all neighbourhoods with median positivity rates of zero include churches, and all those with non-zero positivity rates do not include churches. As churches (places of worship in general) typically engage in mass gatherings, we would expect them to have a positive correlation with positivity rates. Our results indicate the opposite. There are a couple of possible explanations we offer to the reader. The first is that churchgoers are following guidelines to restrict the spread of COVID-19 more vigorously than those who do not attend church. The second is that churches could be negatively correlated with the presence of one or more venues through which the virus spreads easily, or that churchgoers are less likely to visit those high spread venues. Further analysis is required develop more insight.

Because of the currently unexplained impact of churches, we will focus on the non-zero clusters which do not include churches for the remainder of this report.

The relative presence of commerce, fast food, indoor rec, outdoor rec, personal services, and travel venues as well as indoor venues and hospitals appear to have positive correlation with COVID-19 positivity rates in that area.

The relative presence of Essential Retail, Licensed Establishments, Restaurants, and Retail venues appear to have a negative correlation with COVID-19 positivity rates in that area.

The effects of these venues may be the result of substitution to or from other venue types, clientele differences between venues as the result of demographics, or socioeconomic factors.

Conclusion

In this report we have created a framework which can be built upon and modified to determine the effect of different venue types on the spread of COVID-19.

Some our results are expected, and others are rather unexpected. Further modeling and analysis is required in order to challenge and improve these results.

Further analysis could include data and relevant modeling for the effect of socioeconomic and demographic factors, as well as time series data regarding restrictions placed on venue types.