COVID-19 Activity Risk Calculator

The Padron Peppers: Jeev Prayaga, Gyan Prayaga, Ram Prayaga, Rena Brar Prayaga

Project Goal

To determine the risk of exposure to COVID-19 in locations in and around the City of Los Angeles by building a model and application to uncover one or more of the following:

- Features that may increase or decrease COVID-19 exposure risks
- Assist with the transition to re-open by predicting location-based risk scores
- Proposed methodology to implement risk score assessment
- Actionable steps for risk mitigation and to improve risk score

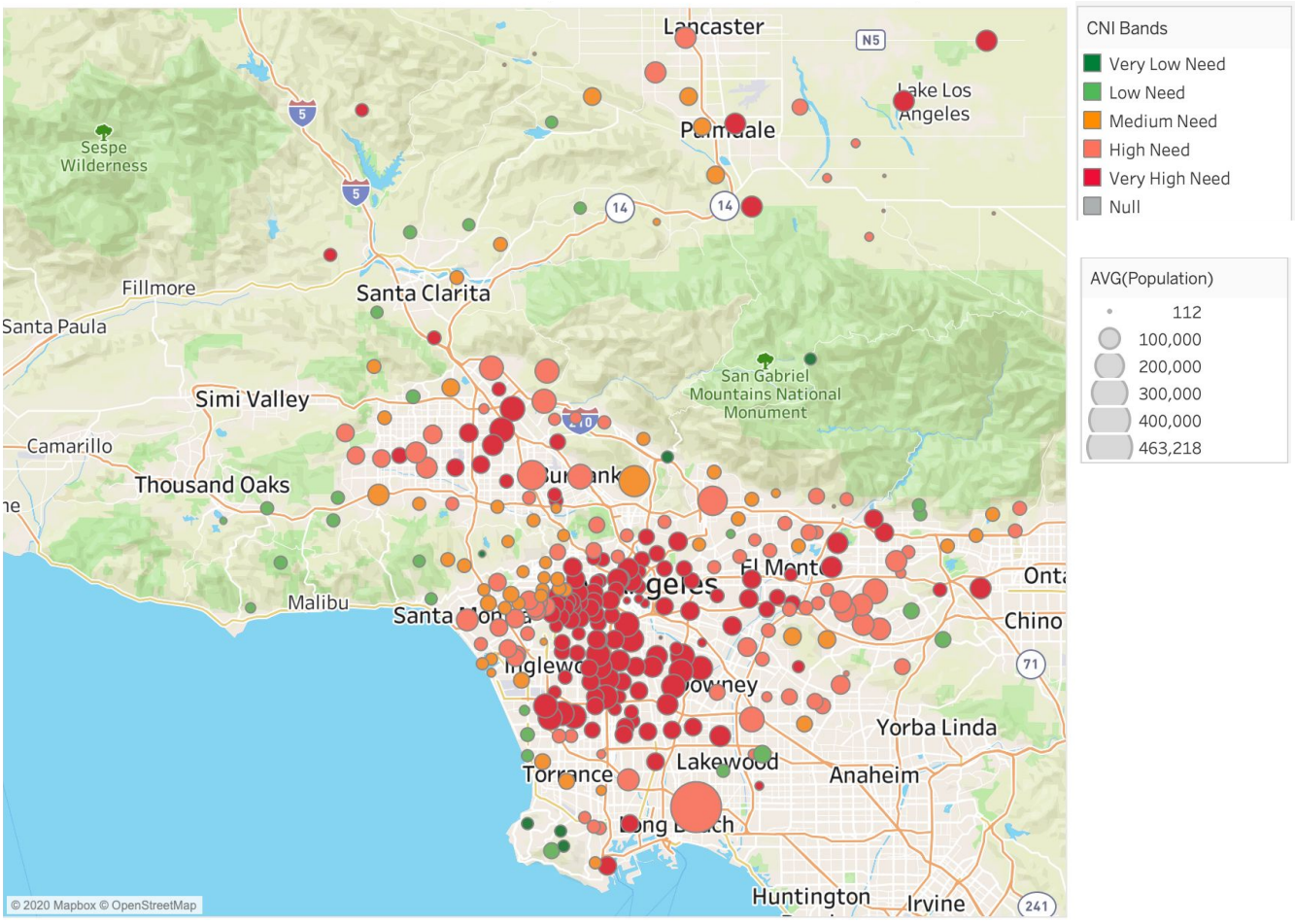
Introduction

The purpose of this project was (1) to evaluate disparities in socioeconomic factors across LA County neighborhoods using the community need index (CNI), (2) to study the role these disparities might have played in the spread of COVID-19 based on count of positive cases, and (3) to develop an application to help users evaluate the risk of leaving their home based on a personal risk profile, a neighborhood based risk profile, and an activity based risk profile (i.e. risks associated with type of activity, getting to activity, and location of activity).

Methodology

Using CNI Scores for LA County Zips

First, we created a table and visualization of population in LA County based on zip code and community need levels. To do this we used the Community Need Index (CNI), which was jointly developed by Dignity Health and Truven Health in 2004. Based on a wide array of demographic and economic statistics, the CNI provides a score for every populated ZIP code in the United States on a scale of 1.0 to 5.0. A score of 1.0 indicates a ZIP code with the least need, while a score of 5.0 represents a ZIP code with the most need. The CNI is closely linked to variations in community healthcare needs and is a strong indicator of a community's socioeconomic status and its demand for various healthcare services. The map below reflects the distribution of LA County residents within the CNI bands, ranging from very low need (dark green) to very high need (dark red).

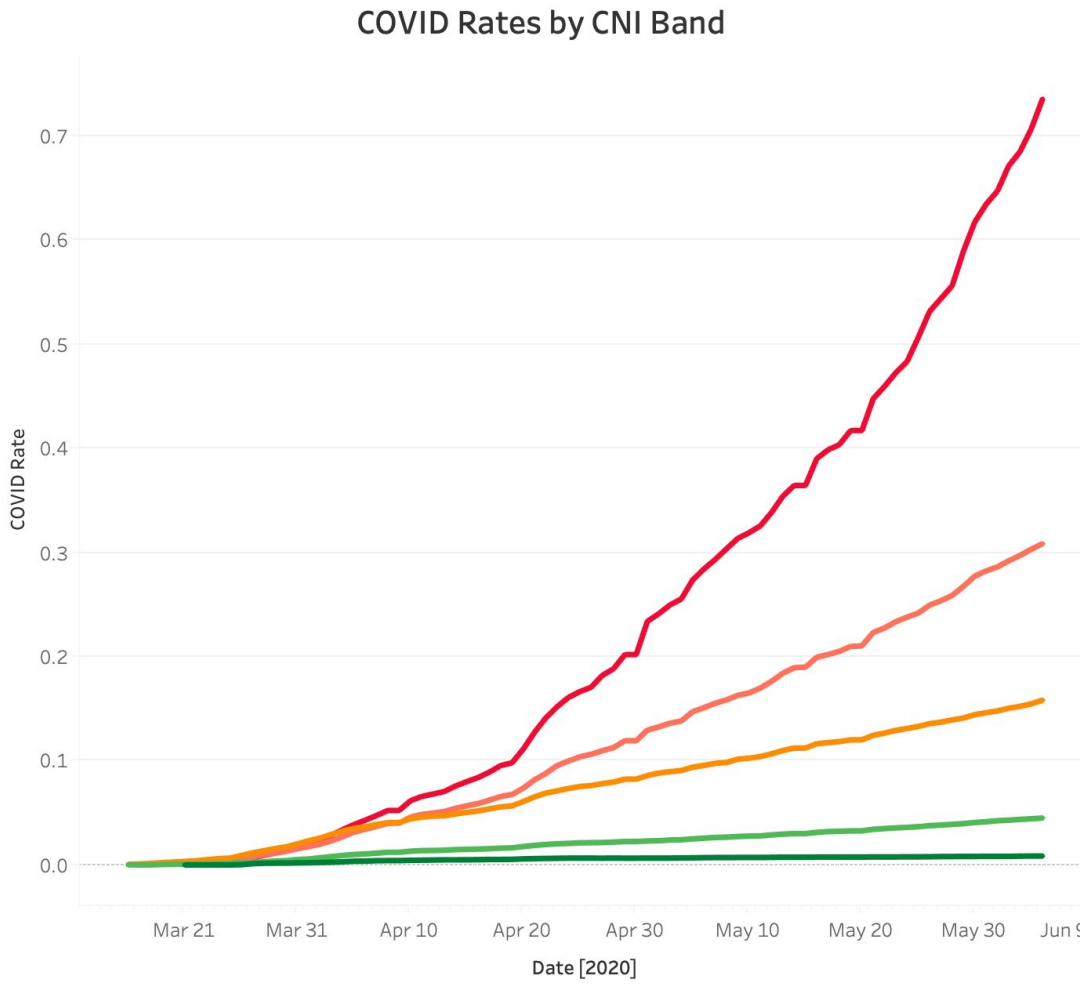


Aggregating COVID cases by zip code and neighborhoods

We then used the LA County COVID Data file to extract COVID counts and associated location (lat/long) for the "places" or neighborhoods within the county. We also computed the zip codes associated with these places, the population count for these places, and the associated CNI for these places. Since CNI is calculated on a zip code level and the places in the COVID data file often spanned multiple zip codes we averaged the CNI scores across zip codes within a place (e.g., Alhambra).

Calculating COVID rates by place and changes in COVID rate over time and by place

We analyzed COVID rates as well as the rate of change in COVID rates across places or neighborhoods within LA County. We found that areas with more rapid increases in COVID rates tended to also be areas with higher CNI values (i.e., more socioeconomically challenged neighborhoods). This strong correlation between CNI and increase in COVID rates is shown below.



Data

Calculating the neighborhood risk score

Based on the correlation between the CNI and the Covid Rates, we developed a model using a decision tree to predict the COVID exposure risk for a neighborhood. We used the following process to develop the model:

a) Prepared the data from the LA County Cases [1], within Microsoft Excel to get a LACovidData.xlsx. This file contains the following tabs:
   i) "LACCovidData" - this sheet is the file from the LA County case by date, county, city, fips, place, confirmed_cases, note, Longitude, Latitude. We added the following columns to this file:
      1) CNI - from GeoCodeZips sheet, using the city or place
      2) Population - from GeoCodeZips sheet, using the city or place
      3) Zip Code - from GeoCodeZips sheet, using the city or place
      4) Covid Rate - confirmed_cases / Population
      5) Covid Rate Change - Covid Rate (d2) - Covid Rate (d1) where d2 is a given date and d1 is the previous day's date
   ii) "LAC CNI" - this sheet contains the CNI for each LA Country Zip code, the population, City, County and State
   iii) "LAC Zips" - this sheet contains all LA County zipcodes, the classification, the City, Population, Timezone and Area codes. This file is used to generate the neighborhood risk by zip code for LA County zip codes.
   iv) "GeoCodeZips" sheet contains the a file returned from GeoCode.IO service for all places that had Covid instances. The service was used to map Lat/Long data provided in the original Covid Confirmed case file provided by the LA County to obtain zip codes to be used to find the location's CNI and population. Some values were hand coded to ensure data consistency
   v) "LACZipsPopCNI" sheet contains Zip Code, City, Population (all three values were originally from a cleaned version of LAC Zips) and CNI (which was looked up using LAC CNI).

This file was used to generate two more values:

1) COVID Change Rate Multiplied (CCRM) - this metric is a scaled value of the average change rate for a place multiplied by the most recent change rate. This value measures the amount of change in COVID cases a place/city is experiencing and the per capita significance of that average rate.
2) COVID Risk Bands (CRB) - Using a simple step function with values determined by the distribution of the CCRM, we generated the following neighborhood risk bands of "Very high risk", "High risk", "Medium risk" and "Risk"

The decision tree model was trained on the following

Independent variables:

[1] https://raw.githubusercontent.com/datadesk/california-coronavirus-data/master/latimes-place-totals.csv

- 'population',
- 'cni',
- 'covid_change_rate_multiplied'

Outcome variable:

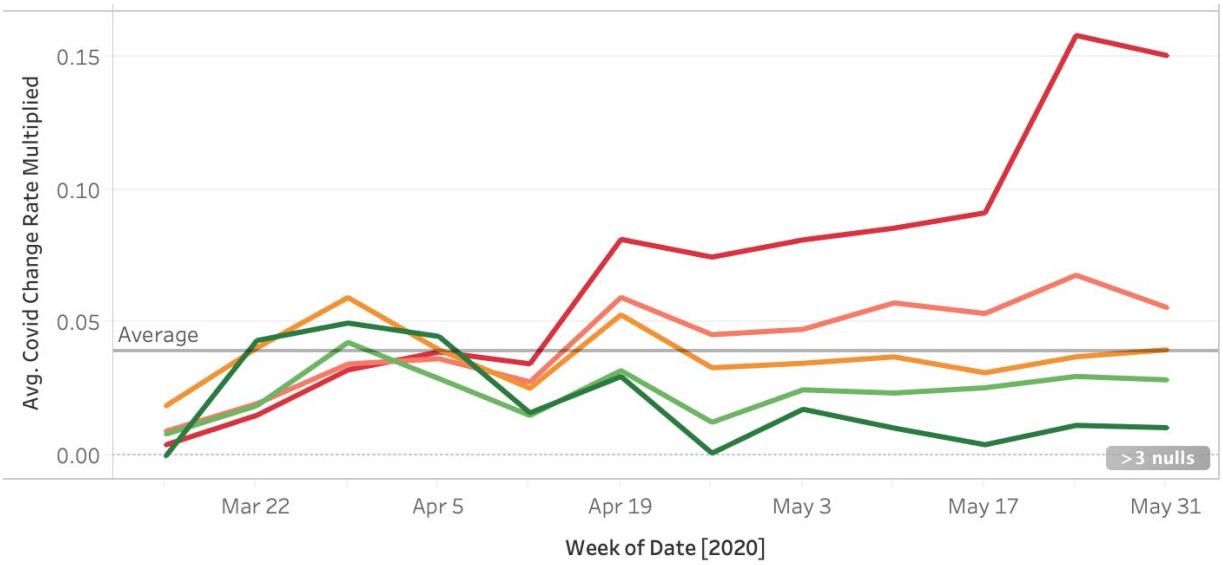
- 'covid_risk_bands'

While we are aware that the CCRM factor is highly correlated by definition with the outcome variable, we chose to include it as it helped for predictive value for instance where there is no data (zip codes that do not yet have covid cases).

We trained the model using a GINI-based decision tree model using Python's pandas and scikit-learn. The model used default parameters and obtained a Python

Accuracy : 97.94%

	precision	**recall**	**f1-score**
High Risk	0.96	1.00	0.98
Low Risk	0.96	1.00	0.98
Medium Risk	1.00	0.95	0.98
Very High Risk	1.00	0.95	0.98



Graph of CCRM broken down by CNI Bands which illustrates the impact of CNI as a highly correlatory factor in growth in COVID cases in a certain neighborhood.

Calculating risk scores & relative weights

The risk scores and relative weights assigned to personal risk profiles, activity types, transit choices, and visit locations have been guided by epidemiology guidelines and research findings. For example, the personal risk profile (increment risk based on age and health status) relies on studies showing that people over 65 and those with diabetes, hypertension and heart/lung disease are at greater risk than the general population. Similarly, the risk scores associated with activities (e.g., higher risk for basketball than tennis) rely heavily on recommendations from a group of infectious disease and infection prevention specialists. It is expected that the risk scores will need to be updated every few weeks based on new insights and improved understanding of the impact and spread of COVID. However, the present model is robust and is intended to be an early version of an activity risk scoring application that is easy to use and intuitive. It also reinforces current guidelines and best practices to stop the spread of the virus and aims to improve health literacy levels by providing tips and reminders (washing hands after a bus ride, etc.) within the application.

Middleware development

The conceptualization and design phase of the application included developing the middleware system in Python and collecting all user generated personal risk, activity risk and transport risk data. The Python API then made a request to the Google Directions API and the Geocode API, which were used to find the trip transport time, the place type(s), and the zip codes of the origin and destination. Trip length was used to assess and update risk profile if mode of transportation was public transit. The model output with neighborhood risk levels was also added to the calculation.

An individual's risk profile was calculated based on data from four separate categories: personal profile (40%), neighborhood profile (20%), activity profile (30%) and transport profile (10%) for a total risk score ranging from 0 to 100. The neighborhood risk profile is calculated based on the model output and the remaining categories are computed based on the user feedback to questions within the application.

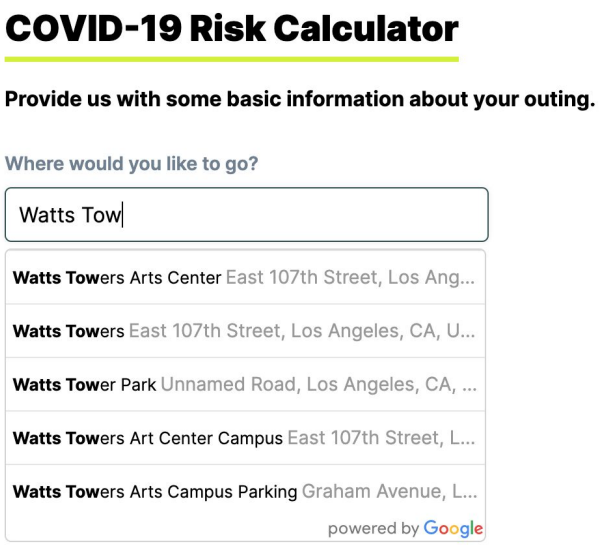
Web application development

We used the above analysis to build a simple web interface to present the LA County residents to understand their risk of exposure to the COVID virus. The user enters the requested factors (discussed above) for their requested activity within a simple web based user interface. The user location is collected using browser-based geo-location (and is done by consent by the user).

This data is sent to the middleware application which returns the risk levels, highlights of what are increasing their risks and a personalized tip. No data is stored about the user (e.g cookies or client local storage) to protect for user privacy.

The Website URL is: https://covid-risk-la.now.sh

A sample user generated risk score is shown in the visuals below. The risk levels presented to the user include the following: *very low risk*, *low risk, medium risk, high risk* and *very high risk.* The purpose of this risk is to help the user understand the risk posed by the activity and then take steps to reduce the risk by choosing a different activity. The following are screenshots of the web application:



COVID-19 Risk Calculator

Provide us with some basic information about your outing.

Where would you like to go?

Watts Towers Arts Center, East 107th Street, Lo

How will you get there?

Driving | Walking | Biking | Public Transit

What do you plan to do there?

Indoor Activity | Outdoor Activity

Going to a museum

Next →

COVID-19 Risk Calculator

To calculate the most accurate risk index, we need some more information about you.

Destination:	Transport:	Activity:	Edit
Watts Towers Arts Center	Walking	Going to a museum	

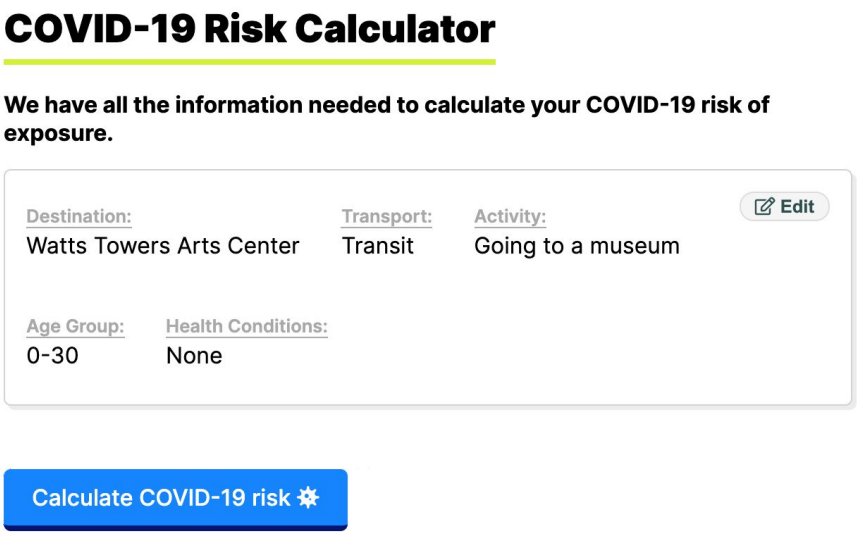
What is your age group?

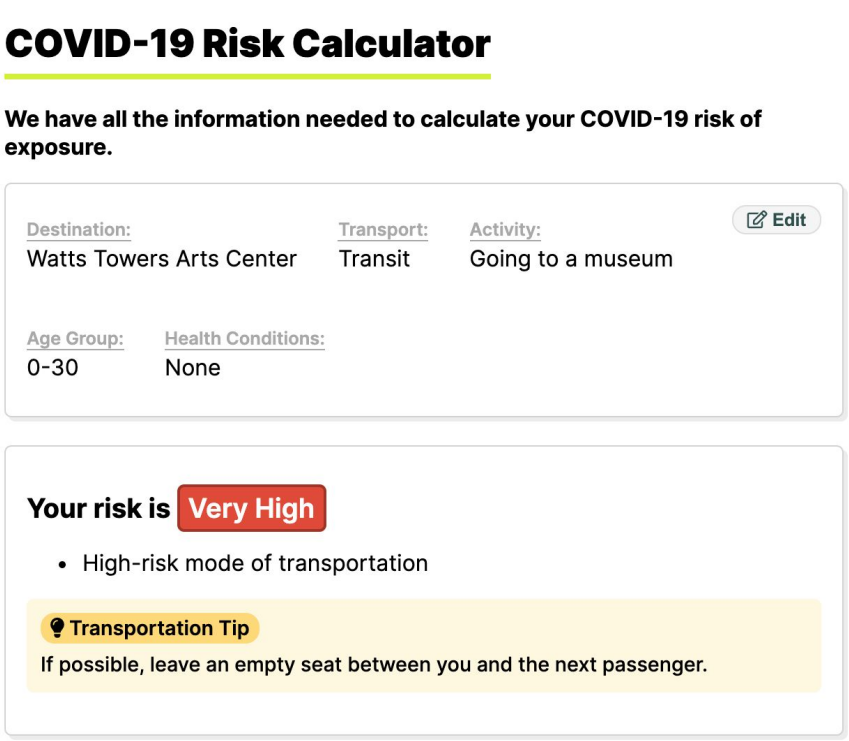
31-50

Do you have any health conditions?

- [x] Diabetes
- [ ] Hypertension
- [ ] Heart/Lung Disease

Next →





Results

Application testing with the COVID-19 Risk Calculator was performed using different user profiles (e.g., (1) 25 year old, no health issues, wants to visit Watts Tower Arts Center, plans to use bus to visit museum (see screenshots above), (2) 75 years old, diabetic, wants to go for a walk in the park, and (3) 32 years old, no health issues, wants to play basketball). In all cases the COVID-19 Risk Calculator yields activity risk scores

that are easy to view and provide users with useful information that can help guide decisions on more appropriate (and lower risk) activities.

Implementation Proposal

- Roll out application as a pilot with a smaller group to test, gather feedback and validate the risk calculations
- Increase the number of activities supported
- Update model with ongoing feed of COVID cases from LA County database

Risk Mitigation

There is a possibility that the projected risk levels associated with various activities will change over time as the threat of COVID-19 lessens and more facilities and places reopen for business. However, the application can still be a useful and effective resource if these risk levels are updated to reflect these changes.

The model is based on current COVID cases (as of June 5, 2020) and is a snapshot in time with an expected risk trajectory based on COVID rates in different neighborhoods. This too is expected to change over the upcoming months and the model would need to be updated with new case loads.

References

1. Improving Public Health & Preventing Chronic Disease: Dignity Health's Community Need Index
https://www.dignityhealth.org/-/media/Service%20Areas/arizona/PDFs/dignity-health-community-need-index-brochure3213448.ashx?la=en

2. Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS. *J Virology. 2020;*
(published online Jan 29.)DOI:10.1128/JVI.00127-20

3. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet. 2020;*
(published online Mar 11, corrected version published online on May 18)
https://www.thelancet.com/pdfs/journals/lanres/PIIS2213-2600(20)30116-8.pdf

4. CDC: People who are at higher risk for severe illness
https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html

5. COVID-19 and diabetes: Knowledge in progress. *Diabetes Res Clin Pract* (published online Apr 9) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7144611/

6. From hair salons to gyms, experts rank 36 activities by coronavirus risk level. (posted online Jun 02, 2020, updated Jun 08, 2020)
https://www.mlive.com/public-interest/2020/06/from-hair-salons-to-gyms-experts-rank-36-activities-by-coronavirus-risk-level.html