Data analysis of policies impact on COVID-19 infection increasing rate (TEAM: PiazzzzaHut) 

Introduction: This data analysis aims to demonstrate the impact of each policy on the COVID-19 infection rate in each state of the US. A panel data set was built to gather each state's cumulated infection amount and policies implementation from March 14, 2020, to June 04, 2020. The changing rate of COVID-19 infection is considered as the dependent variable and is a numeric value calculated by divide one day's cumulated infection amount by the cumulated infection amount of the previous day in each state. Independent variables are factors showing the days that policies were implemented in a state, included: stay-at-home order, mask mandates, travel restriction, and restriction on restaurants. In this study, we look for policies that effectively restrict the spreading of COVID-19 (shown as smaller infection changing rate in each day) by analyzed panel data from models and did decision tree learning. Our result ranks the impact of policies and provides suggestions for governments to make further decisions during this pandemic as well as other related public health issues.

Research Hypothesis: The implementation of four policies in states will significantly decrease the infection increasing rate.

Panel data: With panel data, we can include variables at different levels of analysis (i.e. States) which is suitable for multilevel or hierarchical modeling. Since the process of policies implementation (i.e. begin/end time) in each state through the COVID-19 outbreak are not identical, we conducted panel data analysis to measure whether those policies are effective on show down the spread of the virus or not.

Method 1: OLS/Fixed effect/Random effect model for panel data

We first do a pooled OLS estimation which is simply an OLS technique run on Panel data without considering all individually specific effects. Except applying the restaurant limit will significantly intensify the infection rate, other three policies have significant positive effects on controlling the infection. The model is adequate since the F-test provide a significant test result (p-value < 2.2e-16).

Furthermore, we do the fixed effects and random effect estimation since regular OLS regression omits heterogeneity. A fixed-effect is used when there is heterogeneity across states or time and unobserved variables can have associations with an observed variable. And Random effects are assumed to be uncorrelated with all observed variables. The Fixed effect model is insufficient for variables estimation since its p-value (0.4365>0.05) of the F-test is insignificant. Different with fixed effect model, the random effect estimation gives adequate and effective predictions with all parameters significant under 95% significance level.

Moreover, the LM test is used for comparing OLS and random-effects model with null hypothesis: OLS is better than random effect. We reject the null (p-value= 2.2e-16<0.05), thus the random effect estimation is a better method for our research topic. Consider model with random effect as the proper model for our analysis, the stay at home order, mask mandates, and travel restrictions significantly reduce the infection rate, which is consistent with our initial prediction. As implement "stay at home order" policy, each day the changing rate of infections will decrease by 0.0357 on average. The same interpretation will apply to mask mandates (-0.0352) and travel restrictions (-0.0314). However, the prediction also indicates that the restaurant limit policy will increase the infection rate by 0. 0166 on average each day, which violates our initial thought. The result is shown as figure in page 2 of our slides.

Method 2: Decision Tree Learning for Panel Data

We use a decision tree to go from observations about an item to conclusions about the item's target value. We use greedy algorithm to train the tree – a tree is built by splitting the source set, constituting the root node of the tree, into subsets, which constitute the successor children. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. Combining the model choosing analysis in method 1, we will also consider state as a random effect. We also prune the tree to prevent overfitting. The result is shown as figure in page 3 of our slides.

We found that the policies of stay-at-home and travel restrictions have more significant impact in the model. If the government do not take any restrictions, the infection rate will get higher. But if the government forces people to stay at home, then the infection rate will be controlled. An interesting thing is that if government only restricts travel but allows people to go outside, the infection rate is lower than the case where government only prohibits people from going outside but allows travel, indicating that travel restriction is more effective.

Conclusion: Combining the results of two models, this study indicates that stay-at-home and travel restrictions are two effective policies that reduce the growth of COVID-19 infection in each state. Consisting with our hypothesis, mask mandates also have a significant impact on the model considered each state's random effect. Meanwhile, limitations on restaurants unexpectedly increase the infection rate. Since COVID-19 spread through people's interaction with each other, takeout and delivery of food can only reduce the interaction within the restaurants but indirectly increase the interaction, as well as infection rate of delivery servers, and then spread as deliveries get served. Thus, we suggest governments use more effective policies such as stay-at-home order and travel restriction and pay more attention to the isolation effect of other existing policies.

References

Nytimes. (2020, June 4). Nytimes/covid-19-data. Retrieved June 5, 2020, from https://github.com/nytimes/covid-19-data

M. (n.d.). Stay at Home Orders – What is an "Essential Business"? Retrieved June 4, 2020, from https://docs.google.com/document/d/e/2PACX-1vSXZCFCbIRiRDRC-SWyc36T0S0hjXxT9wZAGM4V01_xtbywLBEn0o_kgmfs0dMJ4VbpPh30j2ZFZ3TH/pub

M. (n.d.). State and Local Covid-19 Response Information. Retrieved June 4, 2020, from https://docs.google.com/spreadsheets/d/e/2PACX-1vRlJWZJ7OkGUW57_rdA2n3xBJ3qjW6u4Z9N6K9Y5L4bM_6H7-S308qdKmJfpVstYWf300nyujvZPFSy/pubhtml?urp=gmail_link

Travel restrictions issued by states in response to the coronavirus (COVID-19) pandemic, 2020. (n.d.). Retrieved June 4, 2020, from https://ballotpedia.org/Travel_restrictions_issued_by_states_in_response_to_the_coronavirus_(COVID-19)_pandemic,_2020

Kludt, A., & Houck, B. (2020, March 15). Restaurants and Bars Shuttered Across the U.S. in Light of Coronavirus Pandemic. Retrieved June 4, 2020, from https://www.eater.com/2020/3/15/21180761/coronavirus-restaurants-bars-closed-new-york-la-chicago

Sontag, E. (2020, May 19). Where Restaurants Have Reopened Across the U.S. Retrieved June 4, 2020, from https://www.eater.com/21264229/where-restaurants-reopened-across-the-u-s

State Data and Policy Actions to Address Coronavirus. (2020, June 4). Retrieved June 5, 2020, from https://www.kff.org/coronavirus-covid-19/issue-brief/state-data-and-policy-actions-to-address-coronavirus/

WHO Coronavirus Disease (COVID-19) Dashboard. (2020, June 4). Retrieved June 5, 2020, from https://covid19.who.int/