Exploring the Impacts of Government Responses on Pandemic Control and Local Tourism at Honolulu County

DATA 512 Human Centered Data Science, Class Project Report

Andy Wang, 12/10/2022

1.Introduction

Hawaii, located in the Pacific Ocean, is one of the most visited in the US thanks to its welcoming temperature, tropical scenery, beaches, and culture. Over 9.4 million people visited the Hawaiian Islands in 2017 alone, spending more than $16 billion there (Annual Visitor Research Report, 2017). Honolulu, as the largest county and the main gateway, is the center of Hawaii's tourism. However, there was a decreasing trend in the daily visitor statistics during the pandemic. Based on the daily passenger count from Hawaii.gov between 7,000 and 11,000 tourists arrive in Honolulu every day as of May 2021. This is below the daily passenger arrival target for 2019 of 10,000 - 15,000 (Daily Passenger Count, 2019). In fact, to control the spread of COVID-19, between 2020 and 2022, the Hawaii government issued a series of public health policies such as the State of Hawaii's Safe Travels restrictions, the federal mask mandate, and Hawaii's indoor mask mandate.

Focusing on Honolulu County, the objective of this work is to evaluate the effectiveness and impacts of government responses to pandemic control and local tourism. The design of this project followed a human-centered approach and considered the social context of the data. The analysis is aimed to solve real problems related to local government, residents and tourists, and data scientists who wish to conduct COVID-related research. For local government, this project would assist decision-makers in combating COVD- 19 spread and protecting tourism development. For residents and potential tourists, this project would help them quantify and understand the impacts of government responses. For other higher-resolution data analysis related to COVID-19 modeling, this project could be used as annotated and open-source reference.

2. Background/Related Work

Since 2020, studies have been focused on the effectiveness of government responses to pandemic control and the sociological impacts of these measures. However, most of these studies are at the global or country level; Given that the enforcement of federal policies varies between states (Sun, 2020), the methodologies and conclusions from country-level studies might not be appropriate for this project.

Focused on Hawaii State, in the paper *Exploring the impacts of travel-implied policy factors on COVID-19 spread within communities based on multi-source data interpretations [1]*, the author proposed a Community Activity Score to capture the travel-related activity level and used regression methods to study its relationship with COVID-19 cases. Some inspirational methods and considerations from the paper were followed in this project. For example, using count data modeling methods instead of multiple regression, and setting different response variables to account for exposure-to-confirm temporal delays.

The Social Distancing Index (SDI), created by the Maryland Transportation Institute, quantifies daily social distance measures level. It is derived as the weighted sum of six mobility measures with the output as an index between 0 and 100 that measures how much social distance is being restricted for residents; a value of 0 means that no social distance is being used, and a value of 100 means that no residents are leaving their homes and no visitors are entering the county (China Data Lab, 2020). The output Social Distance Index describes and reflects the daily social distancing level and was applied in my regression analysis as a predictor variable to access the strength of its relationship with daily passenger arrivals and daily confirmed cases.

The database proposed in the Oxford Covid-19 Government Response Tracker project (OxCGRT) was also utilized in this project. The project has quantified the county-level government responses in 19 variables spanning containment and closure policies, economic policies, and health system policies. The OxCGRT data of Honolulu County was introduced as predictor variables in this project to quantify the level of government measures.

In summary, following the count modeling methods from the paper [1], utilizing the model outputs from SDI and OxCGRT as explanatory variables, this project focused on two study questions: 1. How do government responses affect the pandemic control (daily confirmed cases) in Honolulu County? 2. How do government responses affect the local tourism (daily passenger count) in Honolulu County?

3. Methodology

3.1 Data pre-processing

Following the human-centered analysis design, before model fitting, the collected data was processed. The variables with more than 10% of observation missing would be dropped to meet the model assumptions. For variables with less than 10% missing values, the forward fitting method was used since the data are time series. Also, variables with identical values are removed since the unchanged independent variables was not statistically significant in regression analysis.

3.2 Multicollinearity tests

Multicollinearity is one of the most common problems in regression models which might lead to skewed or misleading outcomes. Therefore, to avoid biased results, I conducted a series of multicollinearity tests before model fitting. Given Pearson correlation coefficient only measures the linear correlations between explanatory variables, I also calculated the Variance Inflation Factor (VIF) to measure the amount of multicollinearity. Combining the tests together, the predictor was dropped if variable pairs have a high (0.5 or more) absolute value of Pearson correlation coefficient and the VIF score for the variable is larger than 5.

3.3 Model selection

Given the response variable in this study are the daily confirmed cases and daily passenger count, they are all count variables. In this case, to fit OLS regression, log transformation was needed; however, it might raise issues of loss of data and lacking the capacity to model dispersion (Bruin, 2006). Instead, Poisson Regression and Negative Binomial Regression were considered. For modeling count data, Poisson regression is frequently utilized since there are several extensions to Poisson regression that are helpful for count models. The Poisson regression has the variance mean equal assumption, while the negative binomial models assume the conditional variance and mean are not equal and estimate a separate dispersion parameter. Thus, the likelihood ratio test was used to compare the assumptions of these two models.

In addition, based on the preliminary analysis, there exist zero values in daily confirmed cases data in Honolulu County. Given the possible existence of excess zeros, e.g., no COVID-19 test was conducted on some days, zero-inflated regression models are also considered, which estimate separate equations for “normal zero” and “excessive zero”. It should be noted that the Vuong test was widely used as the test for zero inflation, and some related studies have adopted this method for model selection. However, using the Vuong test for zero inflation proved to be a misuse because of the misunderstanding of the term “non-nested model”. Thus, in this project, the Vuong test was discarded and whether the zero-inflated models should be used was based on the comparison of the Akaike information criterion (AIC) and loglikelihood ratio, which is common statistical estimators accesses the goodness of model fit.

3.4 Model exposure-to-confirm temporal delays

In class, we have discussed the possible exposure-to-confirm delay for COVID, which means people may not show symptoms right away after infection and it may take a few days for the testing results to become available. Based on studies, COVID-19 symptoms may appear 2–14 days after people's exposure (CDC, 2020). Therefore, to account for this delay, I decide to fit the regression model five times. In specific, the explanatory variables remained unchanged, and the response variable would be the daily confirmed cases data, with 0, 2-, 4-, 7-, and 14-day delays respectively. Combining all these model outputs together would provide a more comprehensive understanding of how the government responses affect pandemic control both in the short-term and after exposure-to-confirm temporal delays.

4. Findings

4.1 Data pre-processing

Following the data pre-processing steps stated in section 3.1, 3 government response variables were dropped because more than 10% of observations are missing. The other 10 variables were excluded because the level of these government response variables remained unchanged, which is not statistically significant in the regression study.

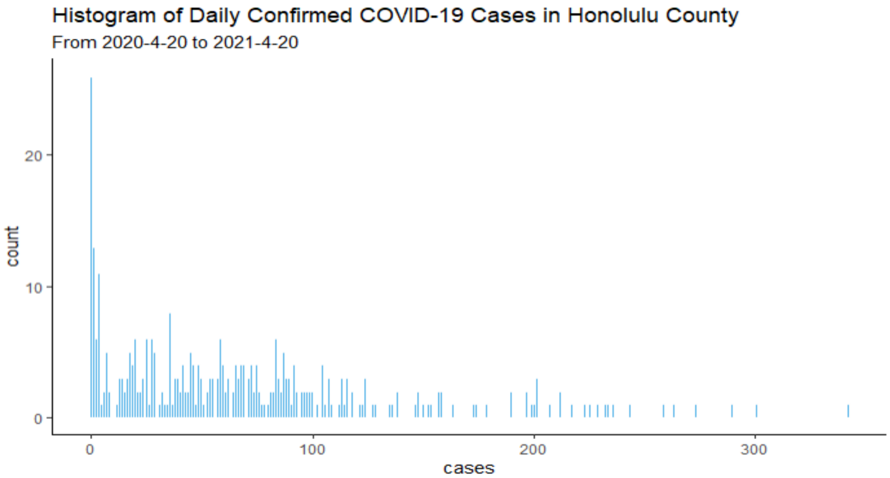

Figure 1

Additionally, the distribution of the response variables, daily passenger count, and daily confirmed cases, were also studied. One significant observation was that zero values were observed in the distribution of daily confirmed cases data (Figure 1). The plot confirmed the hypothesis that there might be excess zeros, which can be explained by either "no residents get exposure to COVID" or other reasons such as "COVID-19 test was not conducted on that given day"; therefore, zero-inflation models must be considered in the model selection.

4.2 Multicollinearity tests

The Pearson correlation coefficient was calculated for each pair of explanatory variables to detect high linear correlations, and the result can be visualized by the heatmap in Figure 2.

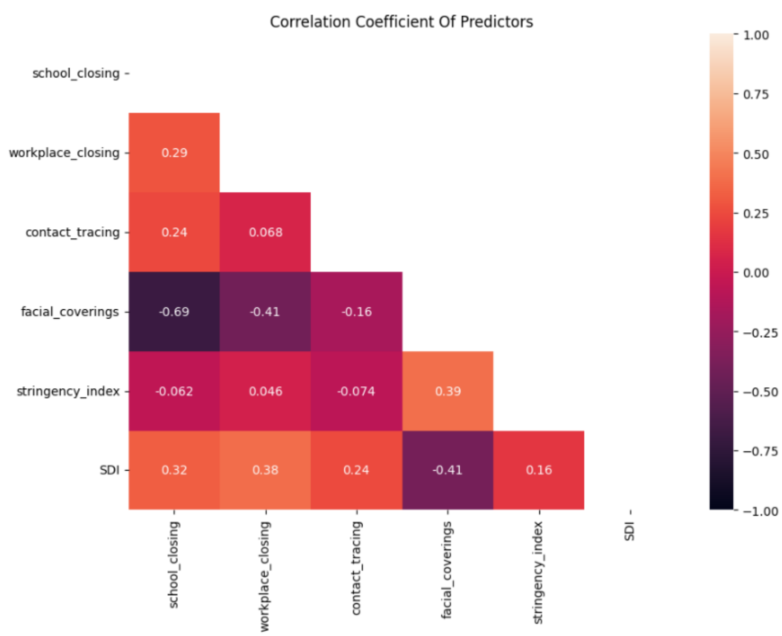

Figure 2

There was a linear correlation between some strict measures such as canceling the public event and public transportation closing. Additionally, the Variance Inflation Factor (VIF) was calculated. Combining the outputs of two methods and following the decision rule defined in section 3.1, three additional predictor variables were removed.

After data pre-processing and multicollinearity test, 6 explanatory variables were selected for regression analysis: school closing (integer 0-3), workplace closing (integer 0-3), contact tracing (integer 0-2), facial coverings (integer 0-4), stringency index (integer 0-100), and social distancing index (integer 0-100).

4.3 Model selection

For regression analysis on daily confirmed case data, following the methodologies stated in section 3.3, the Poisson regression model and Negative Binomial regression models are fitted. Then to check the model assumption (whether the conditional variance equals conditional mean or not) and the goodness of model fits are evaluated by the loglikelihood and AIC. Based on the statistics in Table 3, given the lower AIC score and significant loglikelihood ratio, it can be concluded that the Negative Binomial regression fitted data better which might be because of the overdispersion in our data. In addition, the zero-inflated Negative Binomial regression was also considered; however, based on the results, the Negative Binomial still had slightly better AIC and loglikelihood score, and a lower degree of freedom, which means the interpretability of the Negative Binomial regression model is better.

	DF	AIC	Log-Likelihood
Poisson	7	9498.251	-4742.1
Negative Binomial	8	3381.698	-1682.8
Zero-Inflated Negative Binomial	14	3394.753	-1684.4

Table 3

Similar procedures were repeated for regression analysis on daily passenger arrivals data. Based on the result of AIC and Log-Likelihood (Table 4), the Negative Binomial model was also selected for the regression analysis on daily passenger arrivals.

	DF	AIC	Log-Likelihood
Poisson	7	15498.251	-6282.029
Negative Binomial	8	5422.635	-2198.2

Table 4

4.4 Model outputs

The Negative Binomial regression model result was displayed in table 5, where the D0 – D14 represented the exposure-to-confirm temporal delays in the response variable, daily confirmed cases. Almost all the predictor variables are significant at 0.05 or more precise levels.

It can be observed that the containment and closure policies government responses such as school closing and workplace closing has negative correlations with the daily death case both immediately and after exposure-to-confirm delays. To interpret the result more specifically, for one unit increase in workplace closing level, the difference in the logs of expected counts of the daily confirmed cases after 14 days is expected to decrease by 2.288.

Based on the model outputs, the health system policies such as facial covering and contact tracing have positive correlations with the daily death case. To interpret the result more specifical, for one unit increase in Contact Tracing, the difference in the logs of expected counts of the daily confirmed cases after 14 days is expected to increase by 1.006.

Another interesting output is that the Social Distance Index is not statistically in the model where the exposure-to-delay in the model where the exposure-to-delay was not considered. On the other hand, in models which consider delays, especially with more than 7 days delays, the Social Distance Index is significant.

	Constant	School Closing	Workplace Closing	Contact Tracing	Facial Covering	Stringency Index	SDI
D0	-6.895	-0.303**	-1.855***	0.988***	2.183***	0.095***	-
D2	-6.731	-0.177*	-1.762***	0.944***	2.113***	0.091***	-0.005*
D4	-5.999	-0.008*	-2.015***	0.983***	2.185***	0.082***	-0.004*
D7	-5.955	-0.033*	-2.151***	1.099***	1.979***	0.092***	-0.009***
D14	-4.712	-0.299**	-2.288***	1.006***	1.981***	0.073***	-0.012***

** Significant at 0.05, ** significant at 0.01, *** significant at 0.001, - not significant*

Table 5

The regression model for daily passenger arrivals gives different results. Based on the output (Table 6), the workplace closing is not significant in this model. There is statistically significant evidence that the increase of School Closing and Contact tracking level were negatively correlated to the growth of daily passenger arrivals. Additionally, both Stringency Index and SDI are significant predictors.

	Constant	School Closing	Workplace Closing	Contact Tracing	Facial Covering	Stringency Index	SDI
	14.041	-0.971**	-	-0.168***	0.571***	-0.035***	-0.041***

** Significant at 0.05, ** significant at 0.01, *** significant at 0.001, - not significant*

Table 6

5. Discussion/Implications

Based on the model outputs, the containment and closure policies government responses such as school closing has a negative correlation with the daily confirmed cases and the daily passenger arrivals, which means with increasing the level of such measures, can be effective in pandemic control while at the same time, negatively affected the number of incoming tourists. The output is important for policymakers to identify the possible trade-off when issuing containment and closure policies. However, the correlation does not necessarily lead to a causal relationship between issuing these kinds of policies and the decrease in daily confirmed cases. Further study can use other statistical methods to explore whether there was a causal relationship, which can assist in better decision makings.

On the other hand, only positive correlations were observed with the level of the health system policies such as facial covering and the daily confirmed cases which indicates these measures were not as effective as the containment and closure policies in short term. However, they might still have some long-term effects on pandemic controls. However, this project was not sufficient to explore the long-term correlation between the policy and daily confirmed cases. Further studies with more observations and additional models are required to consolidate the hypothesis that health system policies have some long-term effects on pandemic control.

In addition, for other researchers and data scientists, the model output in this project demonstrated that both the Social Distance Index and Stringency Index have statistically significant correlations with the daily passenger arrivals and the daily confirmed cases in Honolulu County (with exposure-to-confirm delays), which consolidated these variables are effective predictors and indicators in COVID-19 related studies.

Besides these human-centered implications of the model results, there are several human-centered data science principles are applied in the project design. First, the purpose of this project followed scientific thinking, by identifying research questions related to the goodness of policymakers, residents, tourists, and the data science community. The assumptions for data modeling were carefully checked, and the limitations are identified. In addition, this project avoids biased or misuse of methodologies. For example, the Vuong test was discarded for the zero-inflation test to avoid potential misuse or misunderstandings. Last but not least, this project is transparent, annotated, and open to interdisciplinary communications and collaborations.

6. Limitations

It must be admitted that there are several limitations to this project. First, the model accuracy in this project relied on input data accuracy. However, some input data was not peer-reviewed. For example, the weight assignment of the Social Distance Index from the University of Maryland was not published and the raw data for calculating this index was not included. Also, the Oxford Covid-19 Government Response Tracker database might contain errors. In this project, the international travel controls variable was removed because value 4 was observed which contradicts the data document which defined the variable as an integer in the range of 0-3.

Another important limitation was that, unlike other counties which changed the social distance policies multiple times during the pandemic, Honolulu County has several government measures that remained unchanged during the pandemic, and thus these data were removed in the preliminary analysis. Some measures such as public transportation shutdown and vaccine index are important in studies of the relationship between government responses and pandemic control. Future studies on other counties might be necessary to explore how these measures affect tourism/pandemic control in general.

In addition, for the study of the relationship between local tourism and government responses in this project, only the restrictions and measures issued by the local government were studied. However, the traveling restrictions on the potential tourists from their origin country were affecting the number of arrival counts. For example, a large proportion of the tourists in Honolulu come from China and Japan, while both countries have issued traveling restrictions during the pandemic including cancellations of international airlines and temporary closures of the embassy. The impact of these policies was not considered in this project due to workload issues.

7. Conclusion

In conclusion, this project focused on Honolulu County, and conduct an analysis to evaluate the effectiveness and impacts of government responses to pandemic control and local tourism. In specific, there are two main study questions: 1. How do government responses affect the pandemic control (daily confirmed cases) in Honolulu County? 2. How do government responses affect the local tourism (daily passenger count) in Honolulu County? To answer the two study questions, this project takes SDI and OxCGRT data as explanatory variables and fitted two Negative Binomial Regression models selected by AIC and loglikelihood score.

Based on the model outputs, increasing the level of containment and closure policies government responses such as school closing is effective in disease control but largely harms the faith of incoming travelers at the same time. Health system policies such as facial covering are not effective in short-term pandemic control. The Social Distance Index and Stringency

Index are proved to be statistically significant on both daily confirmed cases and daily passenger arrivals.

The results of this project could assist the decision-makers to quantify and analyze various kinds of government measures and understand the trade-off between pandemic control and tourism development. For other higher-resolution data analysis related to COVID-19 modeling, this project revealed that the SDI and Stringency Index can be used as an indicator for future analysis on pandemic control and tourism development under COVID-19. Also, the data preprocessing and count variable modeling part of this project can be used as annotated and open-source reference.

8. References

Bruin, J. 2006. Negative binomial regression. UCLA: Statistical Consulting Group. https://stats.oarc.ucla.edu/stata/ado/analysis

China Data Lab, 2020, "Social Distancing Index from University of Maryland (US)", https://doi.org/10.7910/DVN/ZAKKCE, Harvard Dataverse.

Guo, Y., Yu, H., Zhang, G., & Ma, D. (2021). Exploring the impacts of travel-implied policy factors on COVID-19 spread within communities based on multi-source data interpretations. https://doi.org/10.1016/j.healthplace.2021.102538

Hale, T., Angrist, N., Goldszmidt, R. et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 5, 529–538 (2021). https://doi.org/10.1038/s41562-021-01079-8

Sun Z., Di L., Sprigg W., Tong D., Casal M. Community venue exposure risk estimator for the COVID-19 pandemic. Health Place. 2020:102450.

9. Data Sources

John Hopkins University COVID-19 data (CC BY 4.0)

- Confirmed cases and deaths by US county
- Link

Oxford COVID-19 government response track (CC BY 4.0)

- 19 Government responses, integer, on a scale of 1 -5, E.g.: School Closing, International Travel Controls
- Link

Social Distancing Index from University of Maryland (CC0 1.0)

- SDI: Overall Social Gathering level, float from 0 - 100
- Link

Daily Passenger Counts Hawaii.gov (DCAT-US Schema v1.1)

- Daily passenger count excludes flights from Canada
- Link

The New York Times mask compliance survey data (CC BY-NC-SA 3.0)

- Link

The CDC dataset of masking mandates by county (License)

- Link