

## Data 512 A Project Part 2 - An Extension Plan

Andy Wang 11/10/2022

### **I. Motivation/problem statement**

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. Data from the state government shows that over 9.4 million people visited the Hawaiian Islands in 2017 alone, spending more than \$16 billion there. A significant portion of Hawaii's top industries is centered around the state's 21% tourism-based economy (Wilson, 2018). Honolulu as the largest county and the main gateway of the state of Hawaii is the center of Hawaii's development of tourism.

However, there is a decreasing trend in the daily visitor statistics in recent years. 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, making up the majority of visitors to Hawaii. This is below the daily passenger arrival target for 2019 of 10,000–15,000 (Daily Passenger Count, 2022). 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. To evaluate the overall impact of COVID-19 on Honolulu tourism, and the impact of public health policies on tourism development in Honolulu, in this project, I am going to conduct a human-centered data science analysis of some available COVID-19 data on the Hawaii government website.

## **II. Research questions and/or hypotheses**

The main research question here is **How was Honolulu's local tourism influenced by the pandemic and COVID-related public health policies?** To address this question, I want to study whether there is a statistically significant difference between the domestic and international monthly passengers in Honolulu before and after the State of Hawaii's Safe Travels restrictions. I also made several hypotheses to be tested to help me understand the influence of pandemic on the local tourism:

- Every 10% change in confirmed COVID cases resulted in a 5% decrease in the daily domestic visitors count in Honolulu County.
- The average monthly international visitors decrease by 20% after the State of Hawaii's Safe Travels restrictions were issued.
- The number of visitors who arrived by air decreased by 10% after Hawaii's indoor mask mandate.

## **III. Data to be used**

In addition to the US confirm cases data from the Kaggle repository of John Hopkins University COVID-19, and the CDC [dataset of masking mandates by county](#) which is being used in the Common Analysis. The data used for the extended study relies on the Honolulu County monthly economic indicators dataset from the State of Hawaii government Department of Business, Economic Development & Tourism ([Link to the dataset](#)). The dataset that followed the Federal Open Data Policy provides a license following the [Project Open Data Metadata Schema](#) ([Link to the Policy](#)).

The dataset contains Monthly Economic Indicators (MEI) from January 1990 to September 2022, including indicators such as unemployment rate, total wage, and salary jobs,

state general funds revenues, median selling price, and the number of visitors. This project will be specifically focused on the tourism data and study the features including monthly domestic/international visitors, monthly arrivals by air, and visitors' expenditures by air. These features in the dataset, indicating the monthly tourism situations in Honolulu County, will allow me to conduct statistical tests to verify my hypothesis and will expand my common analysis to how COVID-19 and COVID-related policies affect the local tourism development.

#### **IV. Unknowns and dependencies**

It must be admitted that there are aspects of this problem that are very hard to model, and it has to be based on simplifying assumptions. One of the major issues is that the monthly international passenger count in Honolulu County not only depends on the local public health policy but is also largely affected by international traveling restrictions. For example, Japan and China, as two major sources of travelers, both published some traveling restrictions during the pandemic. Japan has established some rules for traveling during the Olympic Games in Tokyo 2020, and China has been shutting down a great number of international airlines in the previous two years. It is impossible to take into consideration of traveling restrictions from all over the world, so we have to ignore the public health policies outside Hawaii state.

Another important concern is that based on the current information, the vaccination rate is only available for citizens in Honolulu County, and it is difficult to track the vaccination status for all international travelers. Therefore, to simplify the problem, we might have to assume that the vaccination rate for international travelers is the same as the Honolulu County citizens.

## **V. Methodology**

To account for the effect of government measures on tourism in Honolulu County, I plan to build a Logistic regression model. The Logistic regression is a classification algorithm used to find the probability of success and failure, which is perfect for binary classification on whether each government measurement is affecting the tourism conditions in Honolulu County. In the model, the daily traveler count will be used as the response variable, and each government measure is an independent variable. I also plan to conduct 10-fold cross-validation on the hyperparameter to find the optimal C value for the model.

In addition, to verify our assumptions about the monthly international visitors, and visitors who arrived by air before and after the specific government measure, I plan to conduct a student's t-test, since it is designed to find whether there is a statistically significant difference between the mean values of the two observations. With this method, we can use the p-values and the type I error rate to verify whether our hypothesis is statistically significant.

## **VI. Timeline to completion**

1. Data Collection, Processing, and Exploratory data analysis ---- November 13
2. Build a Logistic regression model and conduct statistical testing --- November 20
3. Analyze and Optimize model (cross-validation), create visualization --- November 27
4. Summarize findings and presentation preparation --- December 1
5. Prepare documentation and draft final report --- December 12

## VII. Reference List

Hawaii tourism: Wilson, Reid (2013-09-27). "Hawaii's \$14 billion tourism industry back to pre-recession levels". The Washington Post. Retrieved 2018-11-23.

Honolulu tourism: "Daily Passenger Counts". dbedt.hawaii.gov. Retrieved December 14, 2020.

Logistic regression: Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 215–232.

Student t-test: Student. (1908). The probable error of a mean. *Biometrika*, 1–25.