

Capstone Project - Visit Hawaii

Applied Data Science Capstone by IBM/Coursera

The purpose of this Capstone assignment is to showcase students skills and the tools using location data to explore a geographical location. Students will have the opportunity to be as *creative* as we want and come up with an idea to leverage the *Foursquare* location data to explore or compare neighborhoods or cities of our choice or to come up with a problem that we can use the *Foursquare* location data to solve.

Section 1: Introduction & Business Problem

Background

After COVID-19 shelter-in-place policy is ended, a Seattle-based travel agency, **Capstone Travel Inc.** (aka **Capstone Travel** in short), wants to run a travel promotion called **Visit Hawaii** to bring tourists to Hawaii for a short-term vacationing or stay, in which the length of stay can be anywhere from 14 days to several months. Since there are 6 major islands in Hawaii, so one of first few questions **Capstone Travel** needs to help customers to decide is - *Which island customers should visit and stay based on their preferences or needs?*

Problem Statement

Already, there are plenty of information about Hawaii available online, so **Capstone Travel** needs to stay ahead and provide the *useful* and *decisive* information to customers on visiting Hawaii by aggregating information on island places of interest like parks & beaches, local cultural art museums, restaurants, housing, crime incident, and etc.

To help getting started with the exciting **Visit Hawaii** project, **Capstone Travel** will hire an outside data science consultant to help putting together a proposal for this new initiative to explore **the most popular venues of all 6 major islands in Hawaii**, so the final report should include venues on these beautiful islands: Kauai, Maui, Oahu, Molokai, Lanai, and Big Island.

So the proposal should include **the most popular venues of the 6 major islands in Hawaii** based the popular feedbacks received from Foursquare location data. With venue information available on each island, it can greatly help **Capstone Travel's** customers to make decision on which island they plan to spend their time on.

"Note: Big Island is also called the "Island of Hawaii"."

Target Audience

Around the world as we have already learned that public places, like schools, workplaces, and etc. have been shut down to keep people safe. All the previously well-planned and non-essential travel were either canceled or postponed. In some areas, people have been coping up with this *lock down* at least for 2-3 months. So **Capstone Travel** wants to get ready when governments have lifted the *shelter-in-place* policy when pandemic is under control. Eventually, people will need to plan an *escape* and to recover from the physiological and emotional impact of this extended *lock down*.

This **Visit Hawaii** project data science report will be targeted to customers interested in visiting Hawaii for a *short-term stay* from 14 days to several months.

Scope of the Project

The scope of the project will be focused on these areas:

- The scope of data gathering should be focusing on these 6 major Hawaii islands: Kauai, Oahu, Molokai, Lanai, Maui, and Big Island. (*Note: Kalawao County will be excluded from this project.*)
- Data acquisition and cleansing should be discussed.
- Use **Foursquare API** to retrieve location data.
- Methodology will be detailed in the report.
- Use visual tools for illustration, presentation, or any decision made.
- Due to resource constraints, economic or cost aspects of project will not be covered or discussed in this project.

Assumption

There is one assumption needs to be made here just to set the expectation:

- Pricing or cost related data will not be discussed, so it will be excluded.

Success Criteria

For this project to be successful, it needs to meet several important criteria:

- It needs to help **Capstone Travel** to provide useful information to customers before traveling to Hawaii.
- The project should provide good recommendation of island choice based on customers preference.

Below are the *data science* aspect of the project:

- Data Acquisition
- Data Cleansing
- Data Analysis
- Machine Learning
- Prediction

Section 2: Data

Data Description

There are 6 major islands to visit in Hawaii: Kauai, Oahu, Molokai, Lanai, Maui, and the island of Hawaii. Each island has its own distinct personality, characteristics, adventures, activities and sights. To briefly understand each of these Hawaiian islands it will definitely help visitors to experience the *Hawaii*, and it will also help the data science consultant to better prepare for the project.

As mentioned previously, due to resource constraints, only the following datasets will be explored and presented in the report:

1. Create a *custom* Hawaiian islands dataset. Wikipedia.org and State of Hawaii websites will be the primary source of the data.
2. Query each Hawaiian island's 20 top sites using the **Foursquare Explore API**.
3. Use the **Foursquare Search API** to retrieve the following information about each site:
 - Arts & entertainment
 - Restaurants
 - Events
4. Finally, use the open source dataset to check on the COVID-19 statistics in Hawaii. This critical information will help visitors to time and plan for the exciting trip.

Hawaiian Islands Dataset

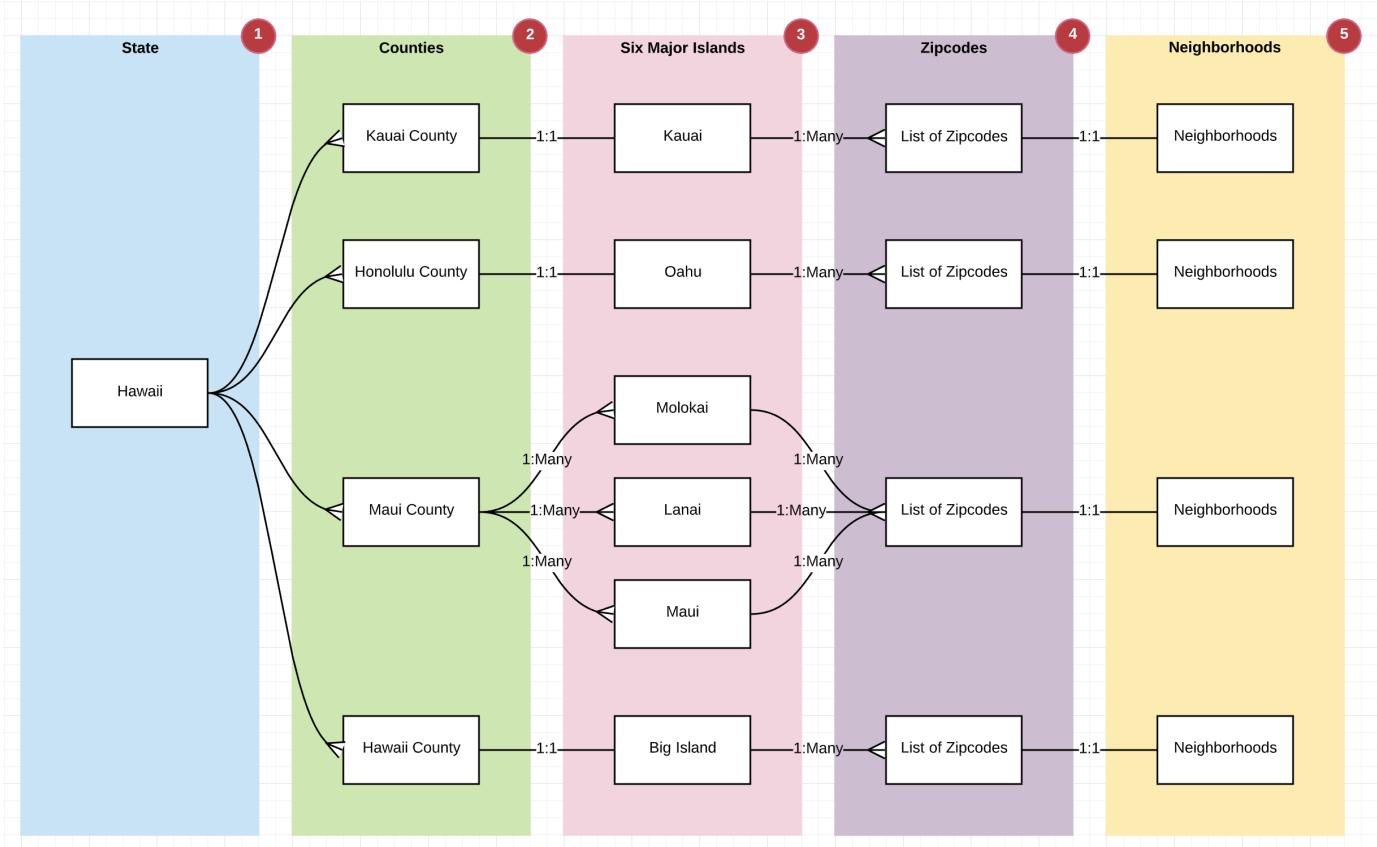
Obviously, we need to start collecting data about these 6 major islands in Hawaii, here is the list of islands in Hawaii, starting from the North to the South:

- Kauai
- Oahu
- Molokai
- Lanai
- Maui
- Big Island



While researching the availability of data, the consultant realized that the granular neighborhood data can be obtained by collecting the zip code information. Due to the size of the islands and local transportation, most of the neighborhoods have their own local post offices, and each post office has its own unique 5-digit zip code assigned. Being able to collect the neighborhood zip codes is crucial for the first step of this data science project. So data cleansing and wrangling is required before the data science process.

With a strong background in database architecture, the data science consultant comes up with a visual diagram below which help to illustrates the relationship of data entities to business, and at the same time the diagram will help him to focus on data gathering process.



Let's follow the numbers to review each data entity.

1. State of Hawaii: That's right, here is the starting point of what consultant is going to focus on - Hawaii.
2. Counties: From the high-level, there are 4 counties in Hawaii, so consultant will use Wikipedia to collect information on these counties: Kauai County, Honolulu County, Maui County, and Hawaii County.
3. 6 Major Hawaii Islands: Just to be clear, most of the counties above have one or more islands. For example, if you take a look at the map of Kauai County, you will notice that there is an island called Niihau to the south west of Kauai island. But the focus of this project is only for the 6 major Hawaiian islands. There is a "one-to-many" relationship between county and islands. It is very important to understand this. For example, Maui County consists of three major islands: Maui, Molokai, and Lanai.
4. Zip Codes: Now we get to the fun part of the data - zip codes. We can say that there is a "one-to-many" between island and zip codes. In other words, an island can have more than one zip codes associated to it. In Hawaii, usually there is a post office building for each zip code assigned.
5. Neighborhoods: Finally, there are the neighborhoods we need to apply the data science studies on. We can say that there is a "one-to-one" relationship between a zip code and a neighborhood. Understanding of this relationship will help us to make a call to Foursquare API to retrieve the venues information or help us to explore the neighborhood.

"Note: Kalawao County on Molokai island maybe listed as the fifth county in Hawaii, however this is very small, like less than 100 population using 2017 reported number. The county is under the sole jurisdiction and control of the state health department, owing to the county's history as a treatment colony for individuals suffering from Hansen's disease."*

"Note: "Hawaii County" or "Island of Hawaii" is for the Big Island. If you read "Hawaii", then it is for the whole State of Hawaii."

Data Acquisition and Cleaning

Understanding of the **Hawaiian Islands Dataset** section will definitely help the consultant to focus on the tasks and narrow his research to break down a bigger task into much smaller activities and achievable tasks.

You may have noticed that references were added to the bottom of above diagram. Let's go through each one of them how the consultant plans to acquire the data from the varied sources.

(1) State of Hawaii & (2) Counties

The information for the list of counties in Hawaii is available in Wikipedia. The consultant plans to use the web page scraping tool to retrieve these data points:

- Counties (Note: Kalawao County will be excluded from this project.)
- Capital cities
- Population based US Census report in 2007
- Area in square miles

(3) Islands

Here is the list of the data points will be screen-scraped and retrieved from the Wikipedia and GoHawaii websites. Below is the first datasets, the consultant tries to build:

Island Dataset	Data Point Description
island	Name of the Hawaiian Islands.
nickname	Each island has its own nick name, it may be helpful to learn this as a tourist.
county	Name of county associated to the island.
capital _city	Capital city of the county.
population	Island population.
density	Density of population.
area	Area size of an island.
latitude	Island latitude.
longitude	Island longitude.

"Note: To keep data consistency, the okina character (') will be removed from the data."

(4) Zip Codes & (5) Neighborhoods

Since each neighborhood is assigned a given zip code, so zip code will be used to query neighborhood information using Foursquare API. So to retrieve all the zip codes in Hawaii, the consultant will use ZipcodeToGo.com to gather the data.

Here is the second dataset the consultant will need to create:

Neighborhood Dataset	Data Point Description
island	Name of the Hawaiian Islands.
county	Name of county associated to the island.
zip_code	Name of the Hawaiian Islands.
neighborhood	Each island has its own nick name, it may be helpful to learn this as a tourist.
latitude	Neighborhood latitude.
longitude	Neighborhood longitude.
crime_incidents	Public records for crime committed using CrimeMapping data source (captured data offline and saved in CSV).
cost_of_living	Cost of living for a given zip code or neighborhood. The US standard is 100 (the base), the state of Hawaii is 176 for the cost of living.
jan_avg_temperature	Average temperature in January.
feb_avg_temperature	Average temperature in February.
mar_avg_temperature	Average temperature in March.
apr_avg_temperature	Average temperature in April.

may_avg_temperature	Average temperature in May.
jun_avg_temperature	Average temperature in June.
jul_avg_temperature	Average temperature in July.
aug_avg_temperature	Average temperature in August.
sep_avg_temperature	Average temperature in September.
oct_avg_temperature	Average temperature in October.
nov_avg_temperature	Average temperature in November.
dec_avg_temperature	Average temperature in December.
housing_rent	The cheaper rent for a smaller unit in a given neighborhood. Rent can go up as high as several thousands for a larger unit.
housing_real_estate	Average home price in a neighborhood.

Finally, once the neighborhood dataset is collected, the consultant can proceed with venue research using Foursquare API.

Island Exploration using the Foursquare API

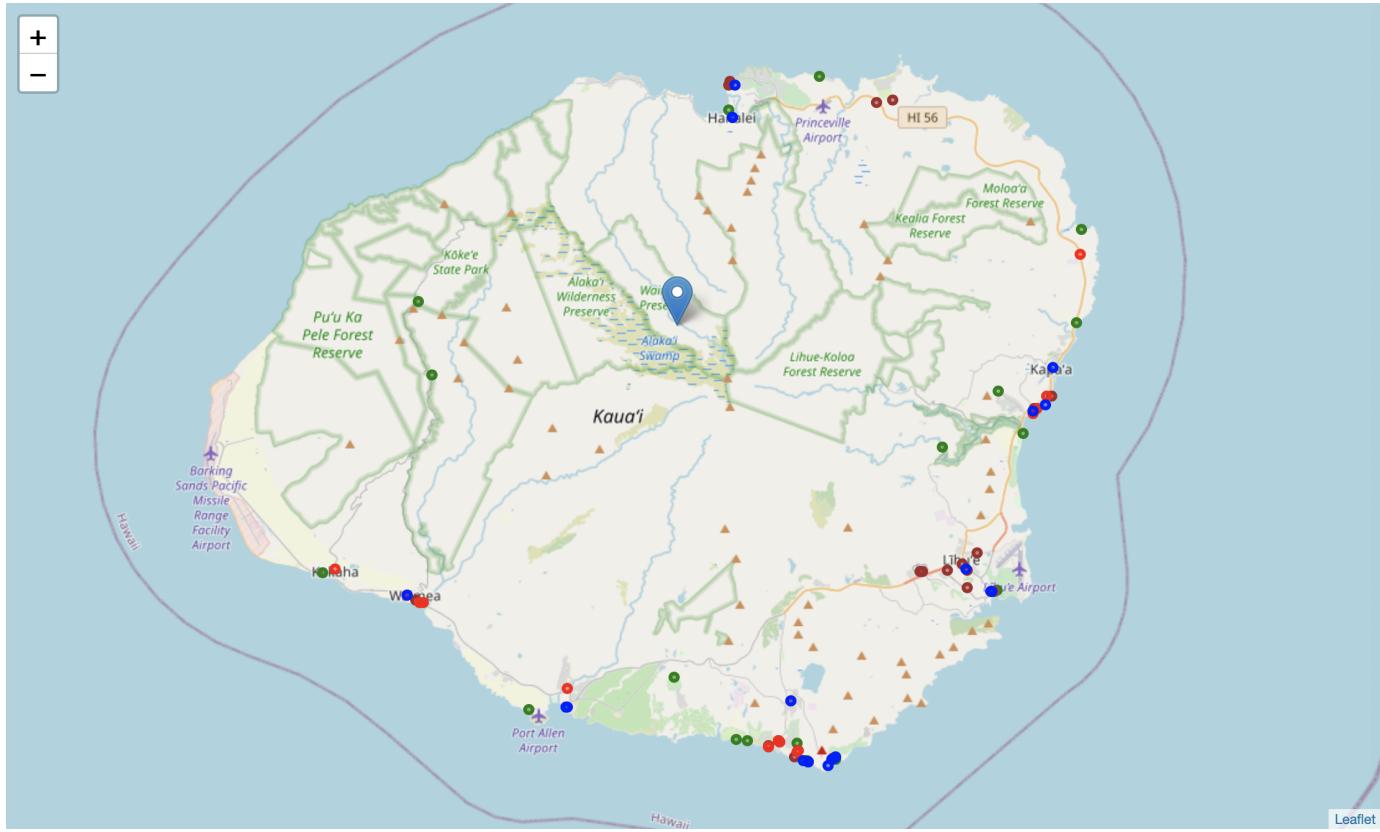
At this point, we should have the basic neighborhood information in each 6 major Hawaiian islands. This island and neighborhood information should help visitors to decide on which neighborhood they plan to the extended stay. Now we should use Foursquare API to explore each island.

Looking at the categories documented in Foursquare.com, we will try to explore venues in these categories:

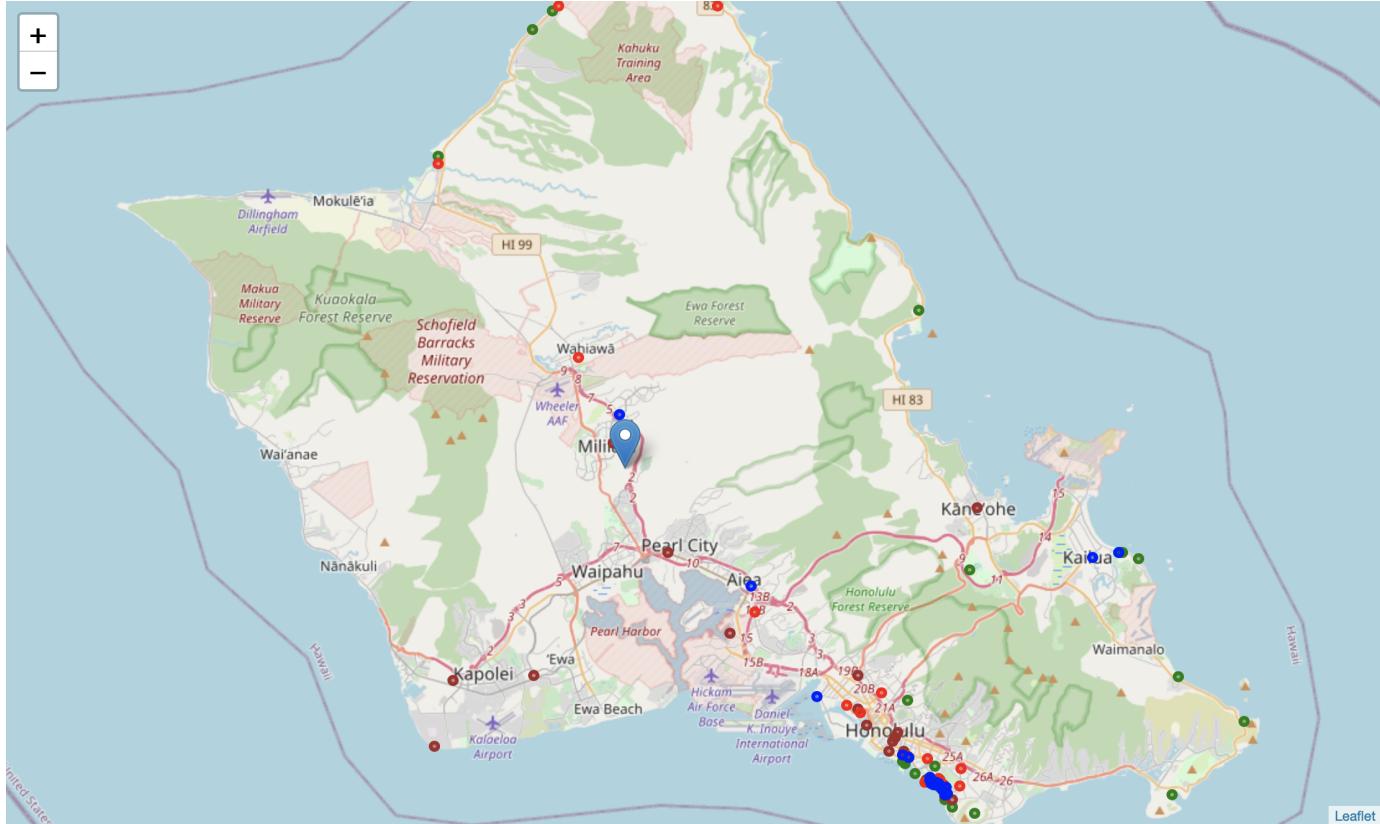
- Outdoors & Recreation ([4d4b7105d754a06377d81259](#)): This is the most important exploration category for what Hawaii has to offer. I believe most visitors coming to islands, they want to experience the beauty of the nature. This global category should include places like beaches, parks, mountains, botanical gardens, etc.
- Art & Entertainment ([4d4b7104d754a06370d81259](#)): It should include places like museum, music venue, art performance, zoo, theme parks, etc.
- Food & Restaurants ([4d4b7105d754a06374d81259](#)): This is a global category for all kind of food, I think it is fine.
- Nightlife & Bars ([4d4b7105d754a06376d81259](#)): OK, this query may be very important for visitors who may decide to stay the Honolulu Downtown area.

Due to the size of the data, the the API should just retrieve the top 20 most popular venues in each of above categories.

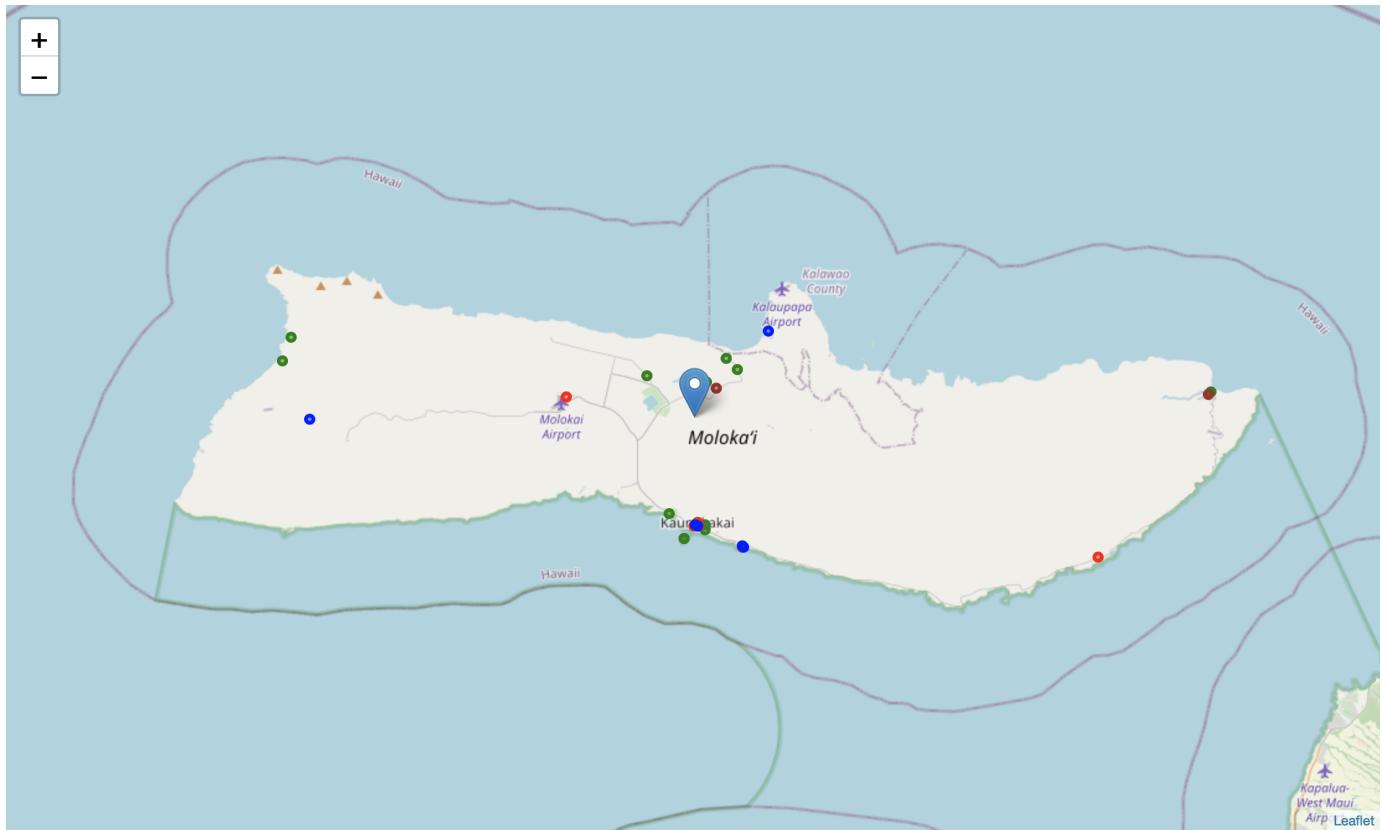
[Venues on the island of Kauai using Foursquare API](#)



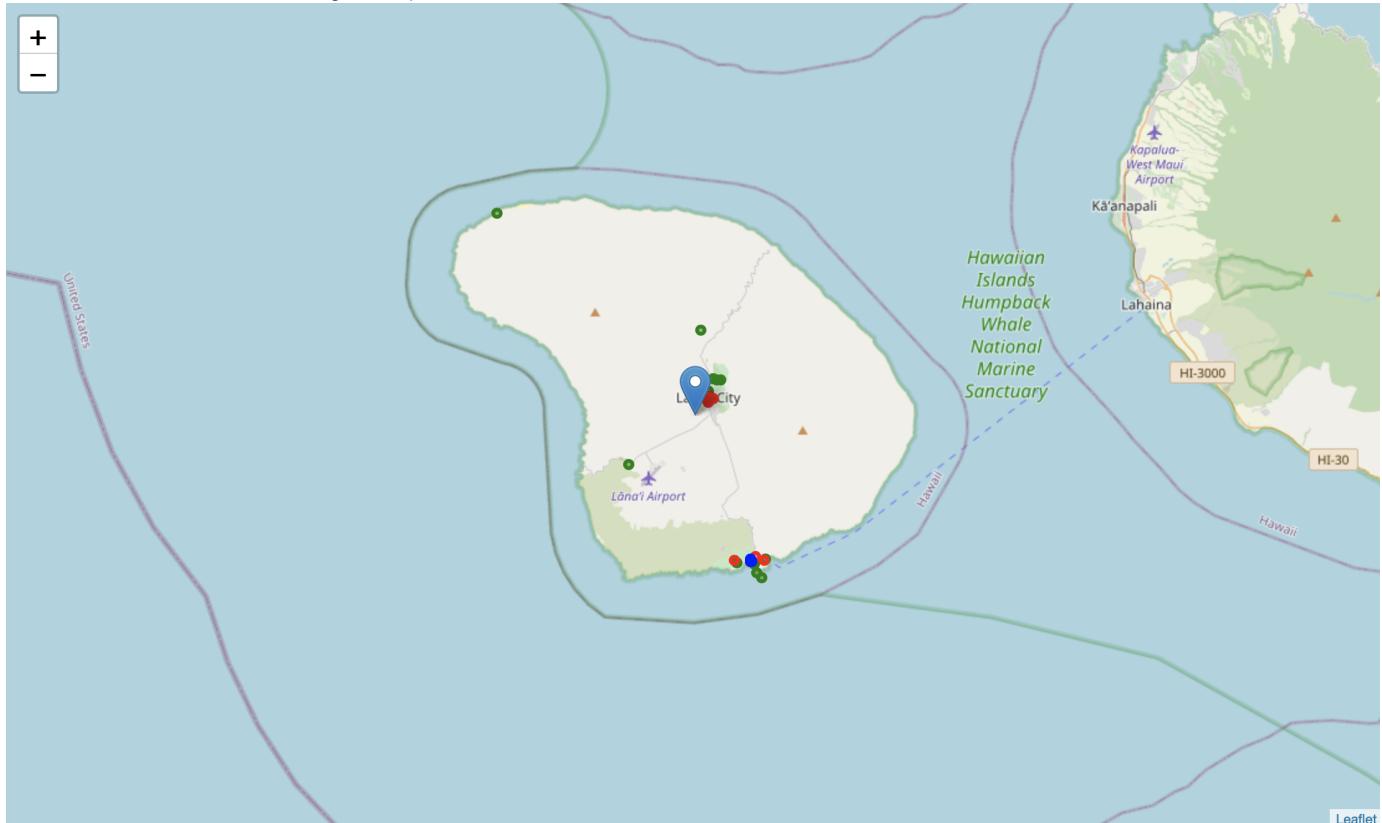
Venues on the island of Oahu using Foursquare API



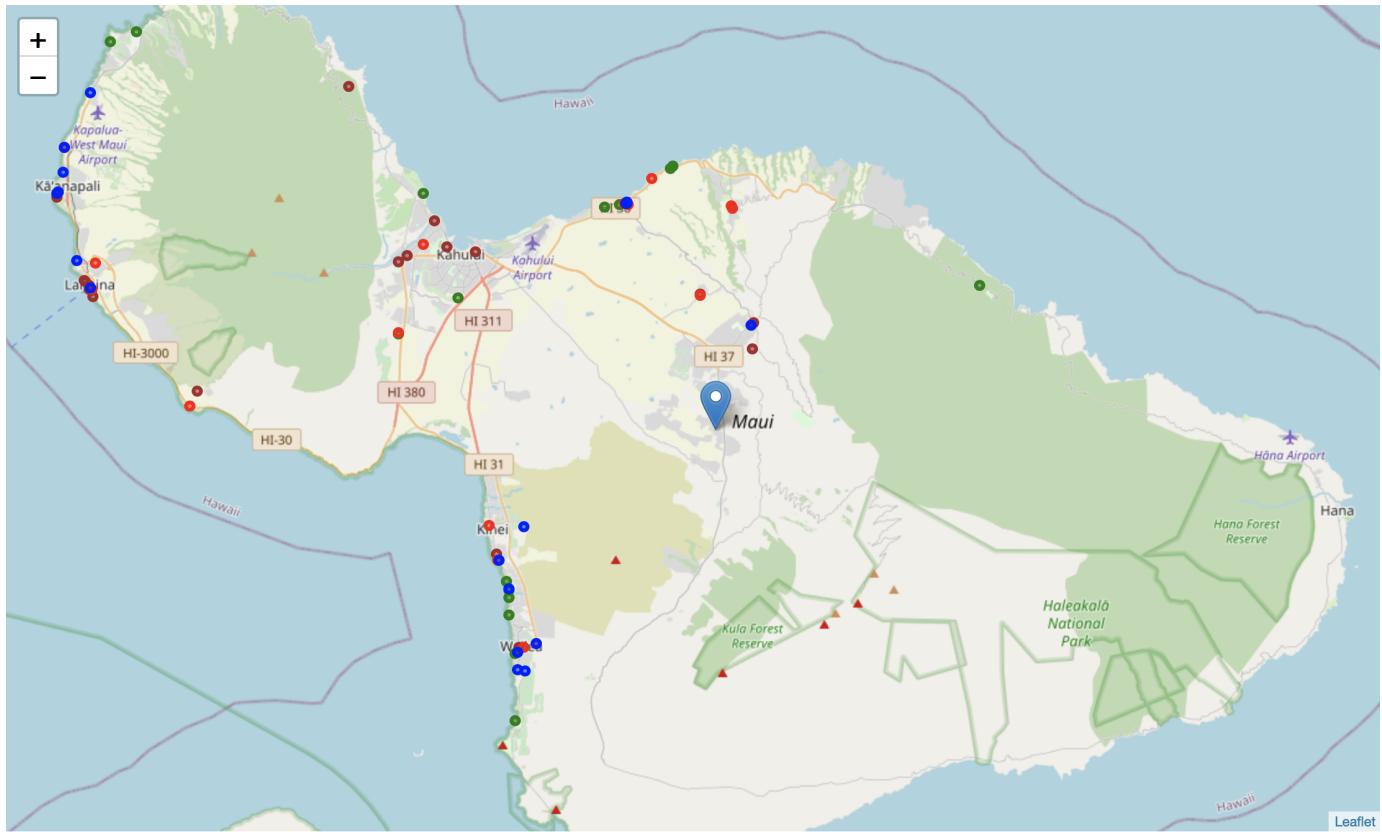
Venues on the island of Molokai using Foursquare API



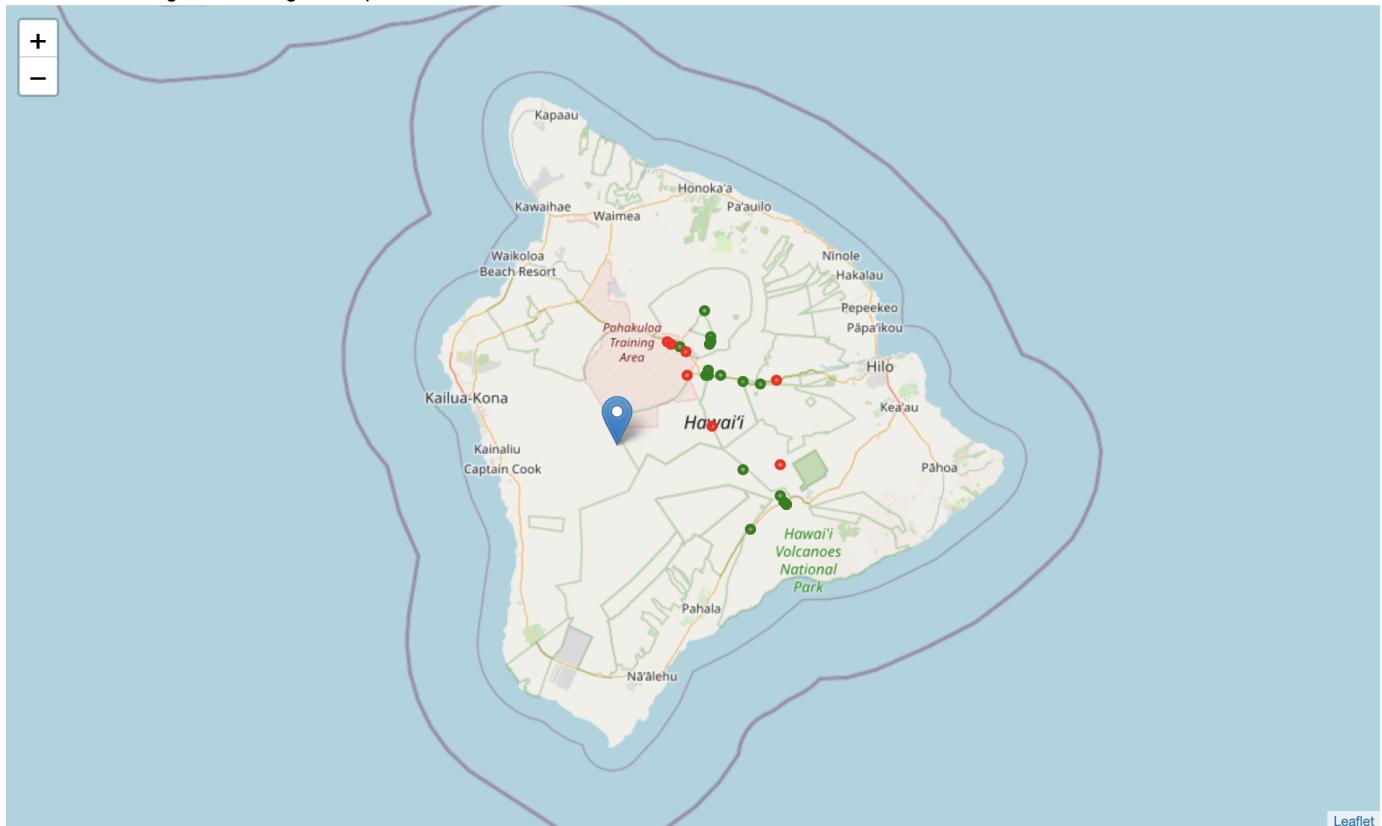
[Venues on the island of Lanai using Foursquare API](#)



[Venues on the island of Maui using Foursquare API](#)



Venues on the Big Island using Foursquare API



COVID-19 in Hawaii

Finally we cannot talk about tourism without bringing up this topic, this is very real. **Capstone Travel** wants to keep all its employees and customers safe while traveling to anywhere. Generally speaking, State of Hawaii has done an excellent job of keeping the COVID-19 cases in control.

So let data speak for itself. The most up-to-date information is available on Hawaii government Department of Health website here, <https://health.hawaii.gov/coronavirusdisease2019/>.

```
Total cases: 814 (11 newly reported)
Hawai'i County: 85
Honolulu County: 566
Kaua'i County: 29
Maui County: 122†
Pending: 0
Residents diagnosed outside of Hawai'i: 12
```

At this point, we have mapped out all venues for each island, as well as the neighborhood information documented in the first part of this section. So visitors should have the good information to have make decision on which island they plan to visit. In addition, we also review the COVID-19 data available on Hawaii government Department of Health website.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on how to visit the island.

Section 3: Methodology

Making sure that visitors will experience this spirit of *Aloha* while staying in Hawaii, the **Capstone Travel**'s business goal is to provide useful information to its customers to make a decision on which Hawaiian island they plan to spend their time on.

In the **Section 2: Data**, the data science consultant has gathered a lot of useful information for each Hawaiian island, for instance, Foursquare's *top-picks*, restaurants, parks & beaches, crime incident data, cost of living index, average weather temperature in each month, and real estate information.

In the following section, the consultant will organize the information by each Hawaiian island. This is important to group the data analysis and results in this way, and it will greatly help to present the information for each island at a time. So for each Hawaiian island, the analysis section should include the following areas of studies:

- **Crime Incidents:** The consultant will start looking into how to apply the *density data analysis of the crime incident data in each neighborhood or zip code*. The consultant plans to add the *heatmap* overlaid the map of each island to identify the low crime areas or safer neighborhoods using *k-means clustering* technique to define the *clusters of locations with low crimes*.
- **Accommodation:** Once the above neighborhoods have been identified, an additional data gathering is required using the publicly available Airbnb housing data. Instead of grouping the information by zip code like above, the Airbnb housing data will be grouped by *zones*. After research, the consultant realized that it is very common practice in Hawaii housing market to slide the housing data in this way using this *zones* due to localized weather and traffic.
- **Foursquare Top-Picks:** Additional details will be included for each Foursquare's *top-picks* venues for each island. After all, this is most critical selling points to present the beauty of each Hawaiian island.
- **COVID-19:** The employees and customers safety is the top priority of **Capstone Travel**, the consultant will provide a high-level COVID-19 data analysis for the State of Hawaii.

Finally, after present above details of each Hawaiian island to **Capstone Travel**, the consultant should prepare an executive summary of the report by including a comparison chart of each Hawaiian island. This executive report should also include overall discussion and conclusion.

Section 4: Analysis

As mentioned in the previous section, the analysis will be done by each Hawaiian island. This is important to group the information in this way, and it will greatly help to present the information for each island at a time. So for each Hawaiian island, the analysis section should include the following areas of studies:

- Crime Incidents
- Accommodation
- Foursquare Top-Picks
- COVID-19