

Opening a Coffee Shop

Jamie Fraser

July 27, 2021

Abstract

This document contains my final report for the Applied Data Science Capstone project, in fulfilment of the requirements for the IBM Data Science Professional Certificate on Coursera. I completed the course during June and July of 2021. The purpose of the project is to describe and analyze a business problem using Foursquare geospatial data in addition to inferential statistics and machine learning techniques that were learned during the course. I am writing from the point of view as a data scientist who was hired as a consultant to help determine the best area to open a new coffee shop in my hometown of Honolulu, Hawaii.

1 Background

The clients have hired me to assess the best location in Honolulu, Hawaii to open a new coffee shop. They are entrepreneurs who want to open a full service cafe and take away coffee shop that caters to visitors and local people. Their aim is to create a warm and inviting place where customers can grab a coffee, sandwiches and other small food items, and even bring their computers and get some work done.

Coffee shops are popular in Honolulu. Most neighborhoods have at least one coffee shop already, and many neighborhoods have multiple coffee shops. Due to the competition, it is important to select the right neighborhood to open this new coffee shop. They want to select an area that has enough traffic to support the coffee shop and produce a good profit, but not be in an area where there are already too many coffee shops and thus the market risks becoming oversaturated.

Finally, they want to distinguish their coffee shop to make sure that it is a comfortable and unique space for customers to enjoy their coffee and work or study. They further want to develop a menu that is unique from other coffee shops in the area and has healthy options for their customers to enjoy.

Given these business requirements, the entrepreneurs have asked me to develop a document that can guide them in making these decisions. I hope to provide enough information to guide them in data-based decision making.

2 Data

The data used for this project was obtained using the Foursquare API. Specifically, the report requires data on:

1. Number and location of coffee shops in Honolulu, Hawaii
2. ZIP codes with the highest frequency of coffee shops
3. Details on existing coffee shops

The number and location of coffee shops by ZIP code will be used to evaluate which areas are most popular for coffee shops. This can help to choose a location for the shop, as well as identify neighborhoods which may be over-saturated or have too many competing shops. Data on the venue make up of each ZIP code helps to identify areas that are most popular for customers looking for coffee shops, or where there are other businesses that can attract customers for the new coffee shop. Finally, it is important that the coffee shop stands out and is unique among competitors in the area. Thus, Foursquare data on the details on existing coffee shops, if any, provides information on competitors and help to design the menu and look of the coffee shop.

Foursquare data is supplemented by data on zip code centroid coordinates in Hawaii, from the open source data project Open Data Delaware on GitHub¹.

¹<https://github.com/OpenDataDE>

3 Methodology

This report uses advanced querying and machine learning techniques to produce usable results that can inform data-driven decision making.

3.1 Exploratory Data Analysis

First, an exploratory data analysis is performed using data obtained from the Foursquare API and mapped out using Python’s Folium package. ZIP code data is used to visualize different neighborhoods in Honolulu, and data on the types of venues in each ZIP code is obtained from Foursquare.

3.2 K-means Clustering

Venue data is compiled to find the frequency of different types of venues, including coffee shops, in each ZIP code. Cafés and coffee shops are listed as two separate categories in the Foursquare data, but for the purposes of this study these two categories are combined. After obtaining the total average frequency of each type of venue per ZIP code, a ranking is constructed that demonstrates the top 10 venue types for each ZIP code.

A k-means clustering model is initialized to cluster Honolulu ZIP codes by most popular venue types. Using the top 5 ZIP codes with the highest frequency of venues in the café/coffee shop category, this study identifies the areas where a coffee shop is most likely to be successful.

4 Results

4.1 Exploratory Data Analysis

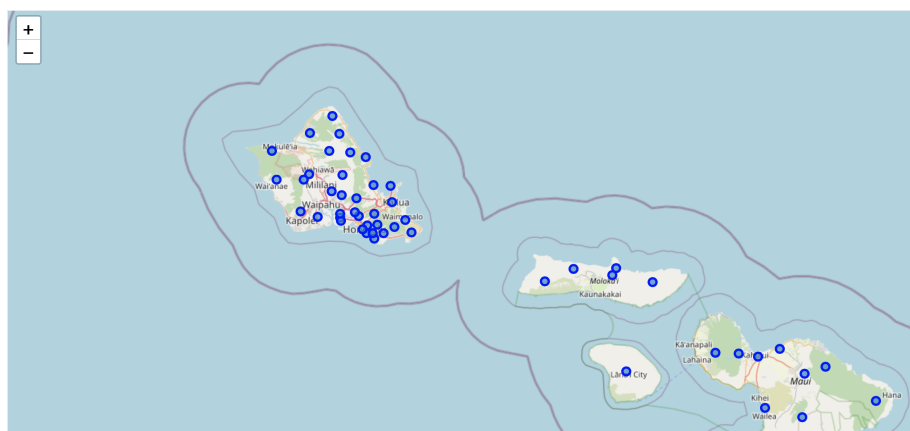


Figure 1: ZIP Codes in Honolulu, Hawaii

A visualization of ZIP codes in Honolulu shows a clear clustering in the downtown areas of Honolulu, where the most densely populated areas of the island are. ZIP codes are more spread out on the rest of the island, indicating less densely populated areas.

4.2 K-means Clustering

A k-means clustering model fitted to data on venue types in Honolulu also shows that the predominant cluster of areas with similar types of venues is centered over downtown Honolulu, in the south center part of the island. This cluster consists of areas where restaurants, fast food venues, and other retail establishments are the most popular.

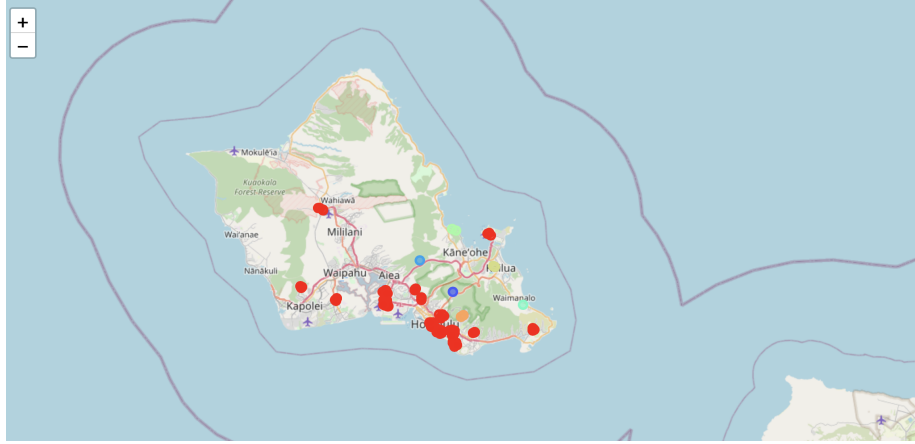


Figure 2: K-means Clustering of ZIP Codes in Honolulu

This supports results from the previous section in showing that a coffee shop would likely be most popular in downtown Honolulu compared to other areas of the island.

5 Discussion

Results of this analysis show that there are two areas that have the most potential for a successful coffee shop in Honolulu. From Figure 1, we see that ZIP codes are clustered in the south center part of the island, indicating where the most densely populated areas of Honolulu are. It follows that these more densely populated areas provide a strong customer base and likely more economic activity.

Figure 3 is visualization of the results of the ranking of venue type frequency by ZIP code. The top 5 ZIP codes with the highest frequency of coffee shops are shown, with the location of individual coffee shops in these 5 ZIP codes shown on the map. It clearly shows that there are two distinct clusters that

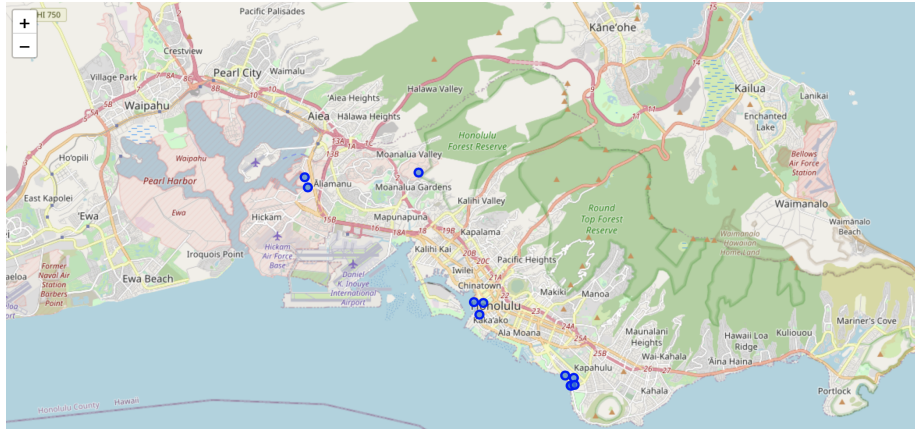


Figure 3: Clusters of Coffee Shops

are the most popular areas for coffee shops. One is on the north side of the downtown Honolulu area, and the other is on the south end of the city. This indicates that there is a pre-existing customer base for coffee shops in these two areas. Further, given that there are other successful coffee shops in these two neighborhoods, there are likely other venues and economic activity happening in these areas that could provide a platform for new coffee shops.

6 Conclusion

In conclusion, this report identifies two locations that would be ideal locations for a new coffee shop. Both are in highly-trafficked downtown areas and located in areas where there are clusters of other dinning and retail venues. Both of these factors contribute to a good environment for high volumes of customers and a vibrant neighborhood.

A new coffee shop would be successful if it has unique offerings and a welcoming atmosphere that allows it to compete with the other successful coffee shops in the area. This analysis shows that choosing one of these two identified locations for the coffee shop would provide an opportunity to stand out as well as a community that could attract and grow the customer base.