Finding locations in Kauai for a State Farm Dealership

Final Report for Coursera IBM Applied Data Science Capstone

Charles Vuono

Introduction / Business Problem

State Farm Insurance is a national chain of independently owned insurance dealerships who contract with State Farm for the right to establish a dealership which will exclusively sell State Farm insurance products from a retail location. Currently, the island of Kauai, Hawaii has no State Farm dealerships. Supposing that we are interested in opening a State Farm dealership on Kauai, the purpose of this investigation is to understand the characteristics of locations that currently support State Farm Dealerships elsewhere in Hawaii and compare them with various areas of Kauai to determine suitable candidates to establish a State Farm dealership. We will study existing State Farm Dealership locations through a variety of descriptive and predictive analytic techniques on neighborhood data, using these results to find similar Kauai neighborhoods.

Data

We will use two primary data sources to conduct our investigation:

1. Location data on every existing State Farm dealership in the United States, including longitude, latitude and full address (approximately 19,000 records)
2. Neighborhood data from the FourSquare Developer API regarding the types of commercial establishments in proximity to these dealerships.

To prepare the data set, we will load relevant Python libraries and then acquire the State Farm and the FourSquare data. Finally, we will load data on potential Kauai agency locations.

```
The State Farm data consists of 18872 rows and 18 columns.
Restricting the data to locational data looks as follows:
```

Out[2]:

	latitude	longitude	city	state	zip
0	42.885943	-77.280766	Canandaigua	NY	14424-1506
1	45.863906	-95.373885	Alexandria	MN	56308
2	41.051858	-76.220358	Nescopeck	PA	18635-0502
3	41.140139	-112.037307	Clinton	UT	84015-6608

	latitude	longitude	city	state	zip
4	38.475771	-81.079581	Clay	WV	25043-7001

The next step is to restrict the State Farm agency location data to Hawaii and turn our attention to acquiring the FourSquare data on venues in the areas surrounding these agencies.

Store	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	Kaneohe1	21.418904	-157.802846	A Place of Perfection	21.419359	-157.804072	Spa
1	Kaneohe1	21.418904	-157.802846	Fresh Catch	21.417109	-157.801815	Seafood Restaurant
2	Kaneohe1	21.418904	-157.802846	Gyu-Kaku Japanese BBQ	21.419692	-157.805073	Japanese Restaurant
3	Kaneohe1	21.418904	-157.802846	Macy's	21.420000	-157.805000	Department Store
4	Kaneohe1	21.418904	-157.802846	El Mariachi	21.418171	-157.801857	Mexican Restaurant

Finally, we load location data on the various areas of Kauai and acquire the foursquare data on those neighborhoods as well.

Store	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	Haena	22.221399	-159.546387	Wainiha Bay	22.218898	-159.543927	Beach
1	Haena	22.221399	-159.546387	Hanalei Colony Resort	22.221453	-159.545230	Resort

Store	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
2	Haena	22.221399	-159.546387	Na Pali Art Gallery And Coffee House	22.221889	-159.545516	Coffee Shop
3	Haena	22.221399	-159.546387	Kalalau B&B	22.221296	-159.546609	Bed & Breakfast
4	Haena	22.221399	-159.546387	Opakapaka Grill and Bar	22.221963	-159.545376	Restaurant

Methodology

We shall approach our problem from several angles:

1. Descriptive analysis of neighborhoods of existing State Farm agencies on other islands in Hawaii.
2. K-means clustering on existing State Farm locations and their nearby venues which will be used to measure potential Kauai locations based on the size of the overall cluster they adhere to as well as their distance to the centroid of that cluster.

We shall find a list of potential locations in Kauai using these two approaches, and develop a final list of potential locations based on those locations which are present on both lists.

Preparing the Data for analysis

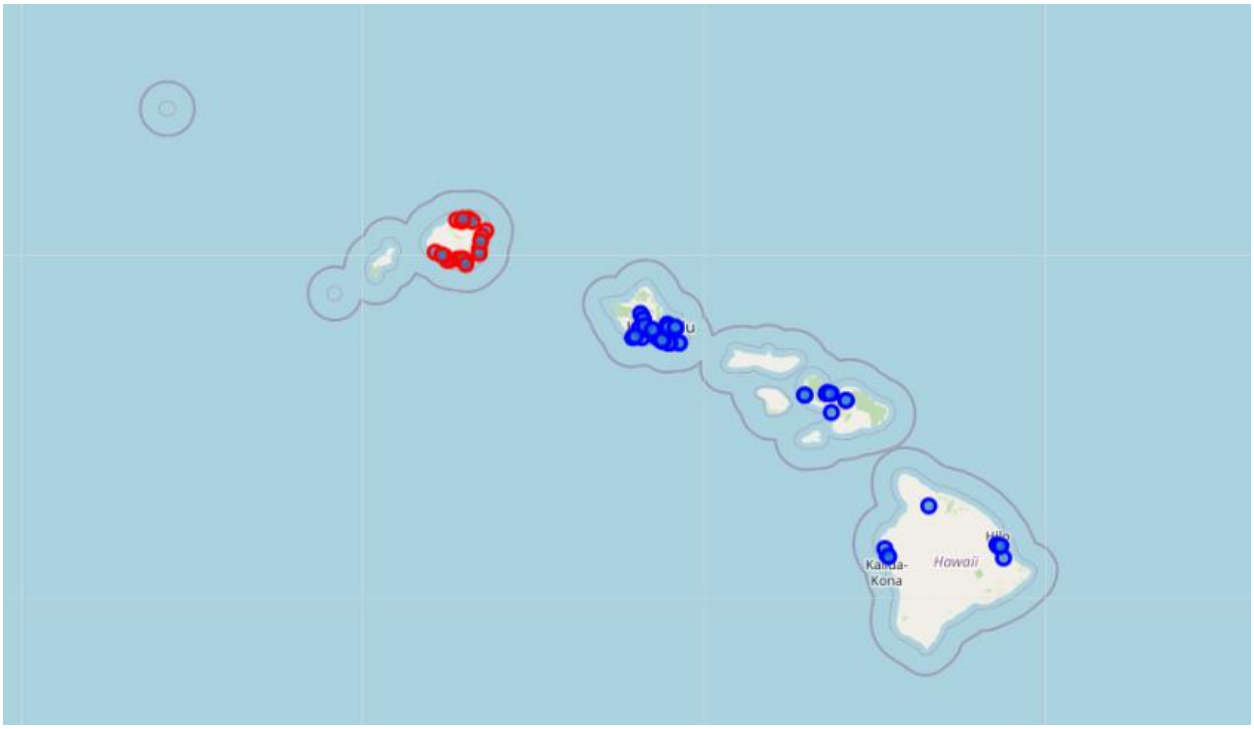
First we will organize the data in both a normalized data frame (showing a weighted presence of venues near each agency or potential agency location) and a data frame listing the most common venue near each agency or potential agency location. First we prepare the data frames on the existing locations.

Next we perform the same set of operations on the potential agency locations in Kauai.

Exploratory Data Analysis

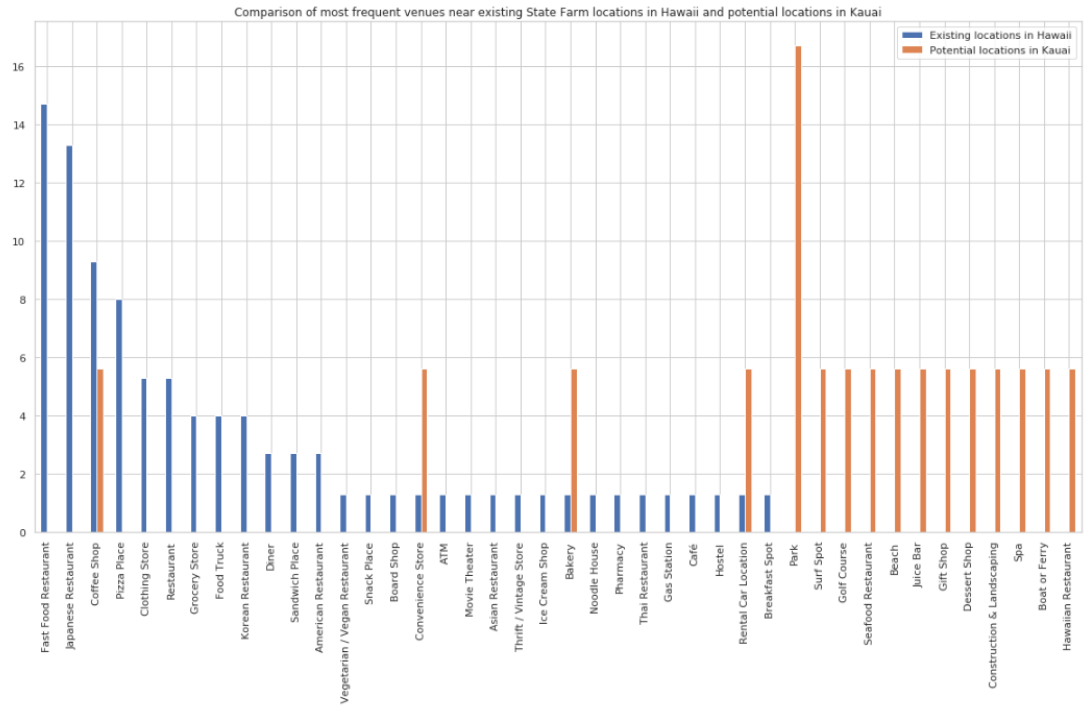
We are now in a position to provide basic descriptive analytics regarding the venues surrounding existing agencies elsewhere in Hawaii and compare them to the venues surrounding the potential locations in Kauai.

Let us begin by looking at a map of existing locations (indicated in *blue*) and a map of potential locations (indicated in *red*).



Next we shall consider the prevelance of the most common venue of existing locations and comparing this to the prevalence of the most common venues of our potential agency locations in Kauai.

To begin, we shall create a data frame with the percent frequency of the most common venue near each existing and potential location and look at the level of overlap.



Immediately, we see some potential candidates for a Kauai State Farm agency: namely, those potential locations which share a primary venue with an existing location. These locations are given as follows:

```
7     Kaumakani
10        Koloa
11        Lawai
12        Lihue
Name: Store, dtype: object
```

K-means clustering analysis

We next turn our attention to a *k-means clustering* analysis whereby we will form clusters of existing and potential locations of State Farm Agencies. This will allow us to find the suitability of potential Kauai locations by their adherence to clusters with large adherents of existing locations.

```
Best potential location(s) from K-means Clustering Analysis:
 1      Hanalei
3     Hanapepe
9      Kilauea
10       Koloa
17      Waimea
```

Results

We performed two distinct analyses: one rooted in *Exploratroy Data Analysis* showed several potential locations (indicated with orange circles in the map below) and one rooted in *k-means clustering* (indicated by blue circles in the map below). Looking at the intersection of these potential candidates leads to a single candidate for a State Farm agency in Kauai: Koloa. This locaion is indicated on the following map with the blue marker.

Discussion

The analyses provided were successful in establishing a single location for the State Farm agency in Kauai (Koloa). Although the analysis used two distinct *methods*, it was fundamentally predicated on the notion that a good location for an insurance agency can be determined by looking at nearby venues of existing agencies. While this seems like a reasonable consideration, we should introduce other indicators of an agency location viability, such as real estate values, population demographcis, accessiblility and likey foot traffic.

Conlcusion

Using rudimentary *Exploraroty Data Analysis* and *k-means clustering* can indicate viability of an inusrance location based on the narrow view of finding similarities with successful insurance agency locations regarding other nearby venues. In our analysis of comparing potential locations in Kauai with existing locations on other islands in Hawaii, we were able to suggest a candidate location for our agency. However, we recommend further study of other characteristics before settling on a location.