**Predicting New Market Viability for In-N-Out**

Brock Imel

22 May 2019

## 1. Introduction

In-N-Out Burger is an extremely popular chain of fast food restaurants that opened its first location in 1948, in Baldwin Park, CA. With its extremely small menu, consisting of only burgers, fries, shakes, and soft drinks, In-N-Out is able to concentrate on maximizing the quality of the ingredients that are used in their restaurants. This is partly because of the ownership's insistence that none of their products ever be frozen or microwaved. In order to ensure this, the company will not open any stores that are further than 600 miles from one of their three distribution centers.[1] For this reason, among others, In-N-Out has been extremely slow to expand: between 1948 and 1988, In-N-Out opened an additional 49 restaurants. For perspective, McDonald's opened a total of 10,000 restaurants between 1940 and 1988.[2]

Since 1992, however, In-N-Out has been exploring new markets outside of California, opening stores in Nevada, Arizona, Utah, Texas, and Oregon, and now operates over 300 stores.[3] With the opening of stores just south of the Oregon-Washington border, the question of whether In-N-Out would cross the border and open its first store in Washington seems to be on many consumers' minds.[4] Although In-N-Out representatives have clearly stated that the company has no interest in opening new stores in Washington at this time, the question of where In-N-Out might place stores in Washington remains an intriguing one.


In-N-Out is famous for its careful selection of new locations for expansion, and, although we do not know the exact procedure that the company uses to select new locations, we wondered if it might be possible to predict new locations for the company to target for further exploration based on the locations of a subset of current restaurants. The target area for this study will be the greater Seattle area, defined generally by the Seattle-Tacoma-Everett Metropolitan

[1] https://www.cbsnews.com/news/in-n-out-president-lynsi-snyder-keeping-burger-chain-a-family-business/
[2] https://www.britannica.com/topic/McDonalds
[3] http://www.in-n-out.com/history.aspx
[4] https://www.seattlepi.com/lifestyle/food/article/Seattle-No-In-N-Out-Washington-expansion-plans-13189594.php

Statistical Area, as seen on this map. In the following section, we define how we acquired and cleaned the data for use in our analysis.

2. Data Acquisition and Cleaning

We viewed this analysis as a pilot study of sorts for a later expanded study, and therefore aimed to reduce the dataset to a manageable test size. This allowed us to determine whether the code that we used was going to work without sacrificing too much time to lengthy database queries. The training group for this analysis was defined as all the In-N-Out restaurants located in one of three regions of California: Greater Los Angeles, the San Francisco Bay Area, and the Central Valley. These three areas were chosen because they span a continuum from dense urban population centers to more suburban and finally rural areas. The addresses for these stores were collected manually and placed in a .csv file for further processing, such as determining the latitude and longitude of each location. There are other likely data points that the company would be likely to consider, such as the sales numbers of each location and certain pieces of demographic information, such as the median income of the zip code in which the restaurant is located and the population density. Naturally, we did not have access to any of the company's sales data, so yearly sales data were randomly generated with an average value of $4.5 million, which represents the average yearly sales of a given In-N-Out location.[5] Median income and population density were found by using the Python library uszipcode. These data were supplemented by the results of queries to the Foursquare API, using a radius of 1000 meters from each restaurant. This gives us an idea of the type of businesses that are in the neighborhoods that In-N-Out selects for its stores. These queries also returned In-N-Out locations, so these had to be removed from the results.

The test data for the present study are comprised of all of the zip codes for the three county Seattle-Tacoma-Everett, WA Metropolitan Statistical Area. These were located online and entered into a .csv file manually for further processing. Using Python's geocoder library, we were able to determine the latitude and longitude of the center of each zip code. We also attempted to find the median income and population density from the uszipcode Python library. This was successful for 116 out of 165 zip codes in this MSA. The remainder were filled in with manually fetched data from the United States Census Bureau and ArcGIS' 2018 United States population map. We then ran a Foursquare query on the latitudes and longitudes returned by the geocoder library.

3. Methodology

Because we envision this as the sort of preliminary study that the company might undertake at the beginning of its exploration of a new region for expansion, we chose to avoid using predictive tools that might give us a more fine grained understanding of how an In-N-Out location might perform in each individual zip code, instead choosing to cluster the locations by using k-means

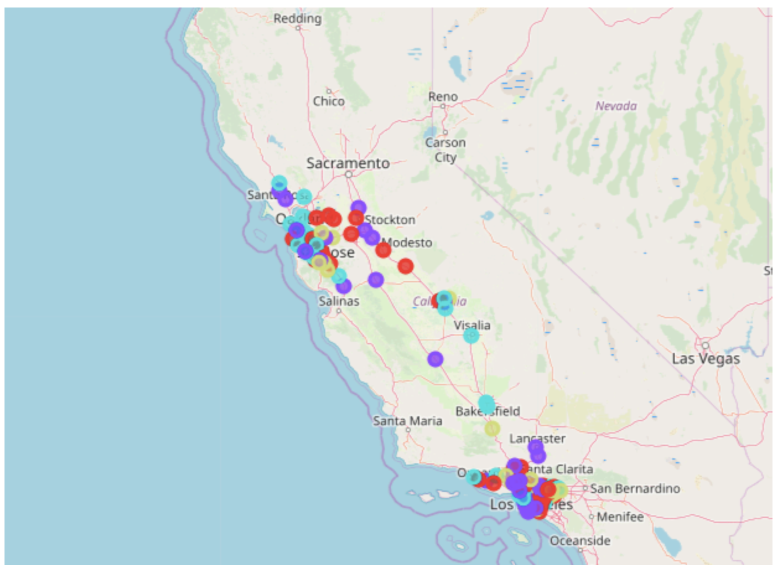
[5] https://www.forbes.com/sites/chloesorvino/2018/10/10/exclusive-in-n-out-billionaire-lynsi-snyder-opens-up-about-her-troubled-past-and-the-burger-chains-future/#4447ab334b9c

clustering, an unsupervised Machine Learning technique. Once the data returned from the Foursquare data had been normalized, we passed the training data for the existing California In-N-Out locations to an optimization function and used the elbow test to determine that the optimum number of clusters for our data was 4. We then fit the model and used the randomly generated sales data to figure out which cluster had the highest average sales. We then moved on to our test set, the Seattle zip code data. After the foursquare data was normalized, we were required by the limitations of the k-means algorithm to refit the model using only the inputs (venue types) that were common to both the training and test sets. The fit model was then used to predict how the Seattle zip codes would fall into the same four clusters. We then examined the distribution of stores/zip codes across each of the four clusters to determine whether further exploration of the Seattle area should be undertaken by the company.

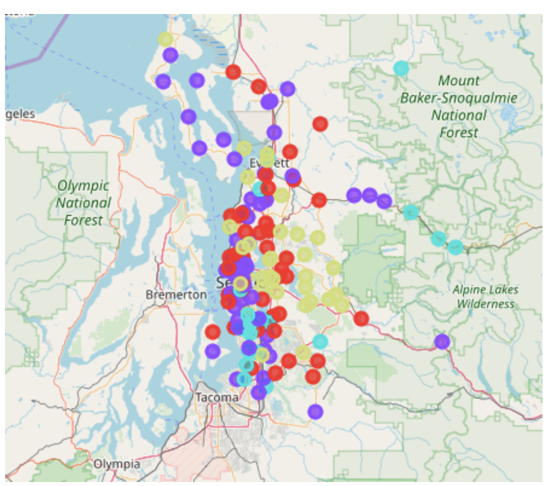
4. Results

Clustering Results: Existing California In-N-Out Locations

Legend:
- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3




Predicted Clusters: Seattle MSA Zip Codes



The preceding two maps show the results given by the clustering algorithm when applied to the existing In-N-Out data in the three regions of California and the predictions returned by the model for potential exploration by In-N-Out. The clusters were ranked with respect to their estimated profitability, based on the average of their randomly generated sales. These sales were computed by summing all of the randomly generated sales data for each cluster and dividing it by the number of members in the cluster. These average values and the resulting rank are displayed in the following table:

| Cluster Labels | # of Restaurants in Cluster | Rand_Sales | Average Yearly Sales of Restaurant in Cluster |
|---|---|---|---|
| 3 | 20 | 90638618 | 4.531931e+06 |
| 0 | 34 | 147291286 | 4.332097e+06 |
| 1 | 37 | 152682914 | 4.126565e+06 |
| 2 | 25 | 99691353 | 3.987654e+06 |

The result of the clustering is nicely shown here in the differentiation in the average yearly sales of each group, with Cluster 3 being the most profitable, followed by Cluster 0. Clusters 1 and 2 bring up the rear, though it should be noted that Cluster 1 accounts for the greatest amount of yearly sales. This should be taken as an indication of the importance of even these third-tier stores to the overall success of the company.

The following table compares the distribution of locations across each of the four clusters.

| Cluster | Actual In-N-Out Grouping | Seattle Predicted Grouping |
|---|---|---|
| Most Profitable | 20 (17.2%) | 36 (21.6%) |
| 2nd Most Profitable | 34 (29.3%) | 50 (30.1%) |
| 3rd Most Profitable | 37 (31.9%) | 57 (34.3%) |
| 4th Most Profitable | 25 (21.6%) | 23 ( 13.9%) |

## 5. Discussion

Several of the figures in this table are fairly promising. Per the predictions of the model, Seattle would have a higher number of locations which fall into the most profitable tier of stores, Cluster 3, when compared to California, where we already know these stores are extremely profitable. Seattle also edges out the current In-N-Out stores in our dataset for the next most profitable stores, if only by 0.8%. Seattle has more 3rd tier locations in the model than do the California data, but considerably fewer 4th tier locations. In light of this information, I would say that the risk is likely to be worth taking, and that the company should start scouting locations in the 1st tier zip codes. Interestingly, a majority of the top tier locations seem to be located in suburban or more rural areas, rather than being concentrated in the center of dense urban areas. This would seem to be consistent with the placement of many of the stores in California. There appear to be a greater concentration of Tier 1 in the San Francisco Bay Area region, in moderately wealthy suburbs, such as Livermore, San Ramon, and San José. Many of the more poorly performing restaurants can be found in the southern central valley and in poorer areas of the region immediately surrounding the San Francisco Bay.

## 6. Conclusion

The high percentage of predicted well performing locations in the Seattle area illustrates that further exploration may be merited in these neighborhoods. Naturally, the applicability of any of these results is hindered by the absence of real sales data for inclusion in our own study. Nevertheless, we believe clustering to be a powerful tool for preliminary sorting of locations where In-N-Out might expand, despite the lack of a distribution center within 600 miles of this location. This method could be used to examine a variety of urban and rural areas within the Mountain West and Plains regions of the US for the company’s potential expansion.