

Predicting New Market Viability for In-N-Out

In-N-Out: Background

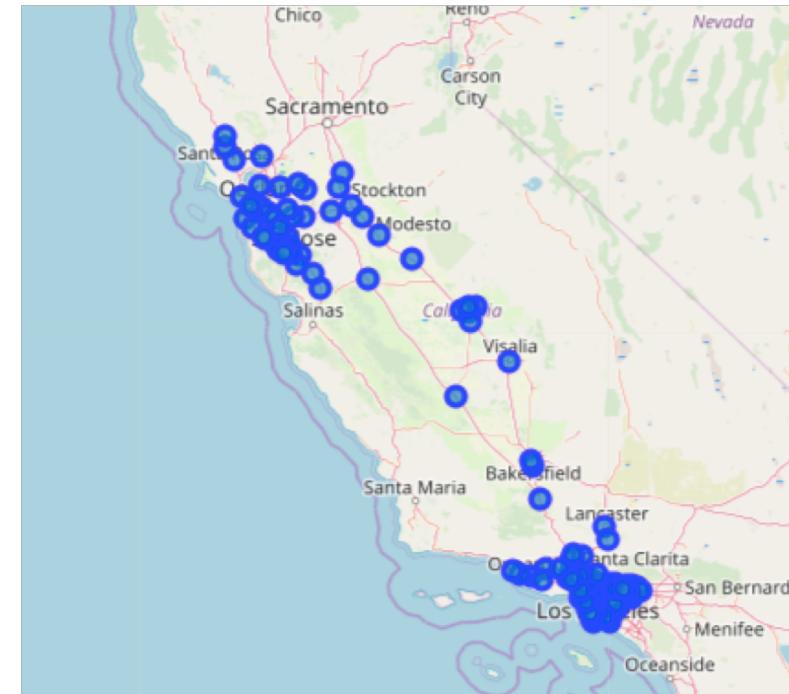
- ▶ Founded 1948 in Baldwin Park, CA
- ▶ Cautious growth
 - ▶ Opened 300th store in 2015
- ▶ Minimalist menu with extremely fresh ingredients
 - ▶ All locations within 600 miles of distribution center
- ▶ Began expansion out of California in 1992, now present in 5 other states
 - ▶ Nevada, Arizona, Texas, Utah, Oregon
- ▶ Washington State borders their current territory- What metrics for their potential success in that market could we propose?

K-means clustering

- ▶ Can we predict what zip codes would be likely to have high performing stores if In-N-Out expanded to Seattle?
 - ▶ Yes, based on similarity to previous location choices.
- ▶ K-means clustering is an ideal way to conduct preliminary research into a new region.
 - ▶ Sorts datasets into k number of groups/clusters based on similarities between datapoints -> develops a model of the data
 - ▶ Allows us to extrapolate onto other datasets with the same inputs and sort them into categories using the criteria defined by the model.

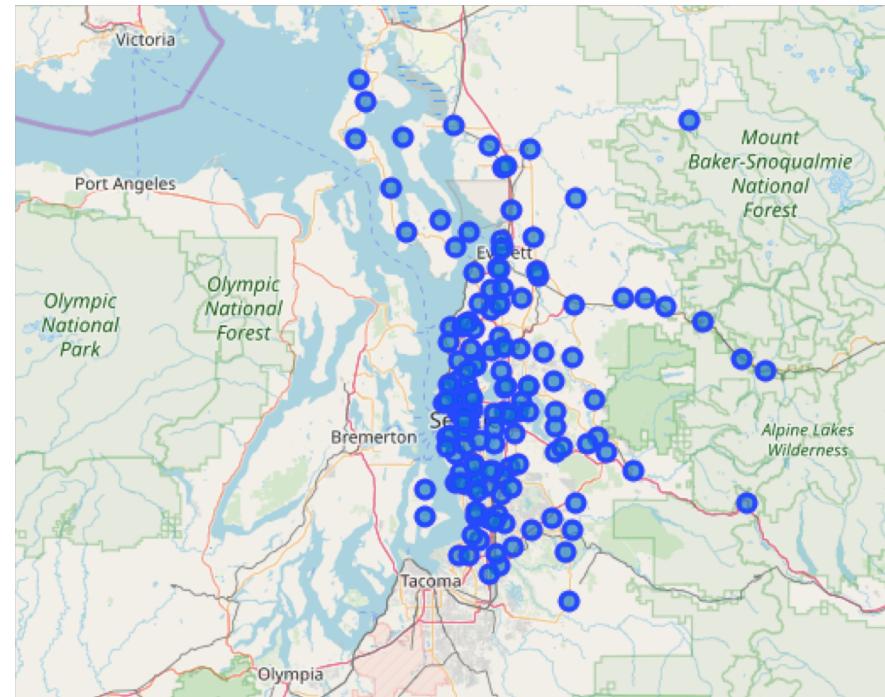
Data Collection and Cleaning- In-N-Out

- ▶ In-N-Outs from 3 Regions
 - ▶ Los Angeles, SF Bay Area, Central Valley
- ▶ Features included in dataset:
 - ▶ Latitude, Longitude
 - ▶ Randomly Generated Sales Numbers between \$2M and \$7M (Avg. \$4.5M).
 - ▶ Median Income and Population Density
 - ▶ Foursquare: All Venues within 1000m of location.



Data Collection and Cleaning- Seattle Zip Codes

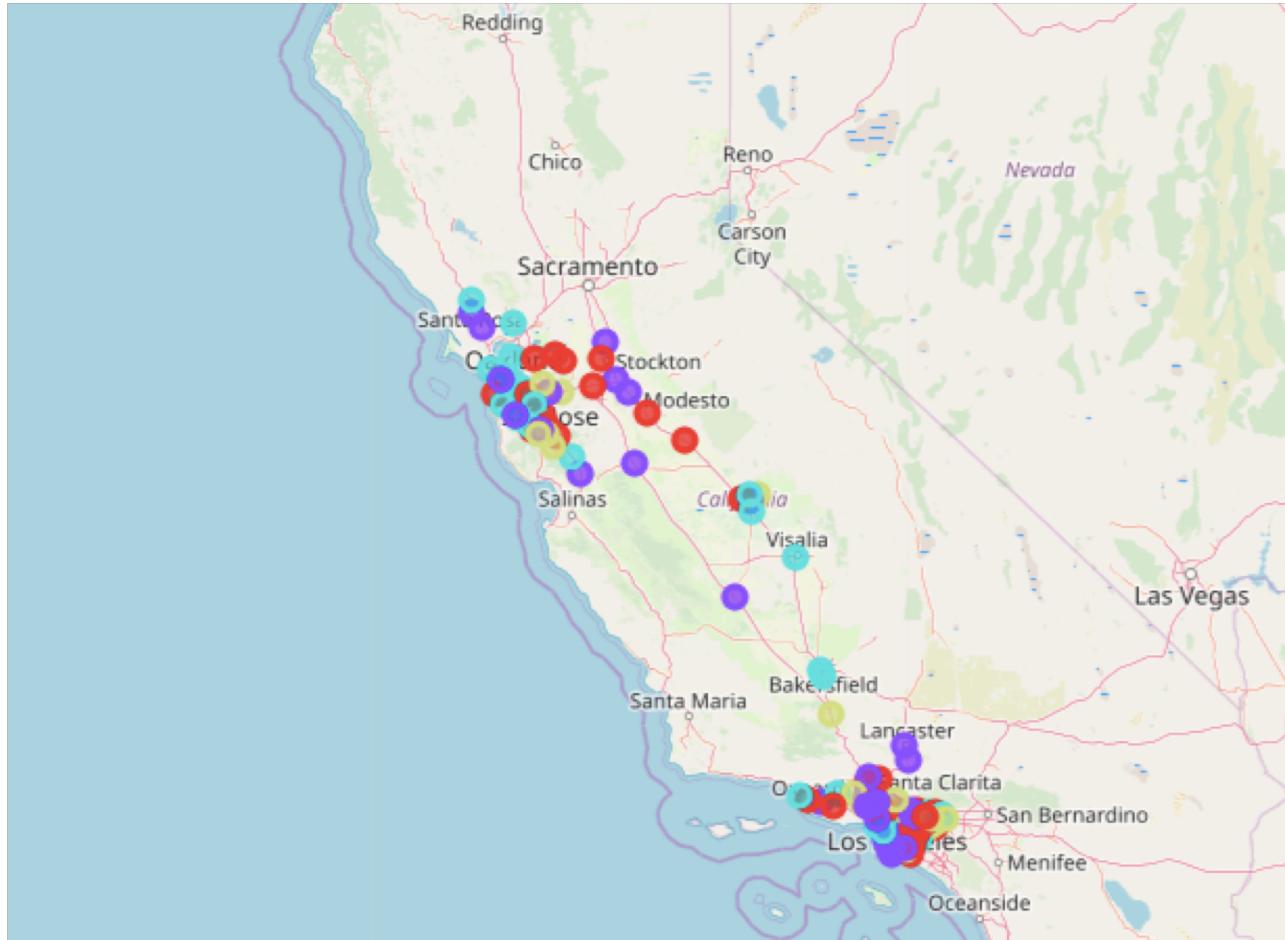
- ▶ 166 Zip Codes from Seattle-Tacoma-Everett MSA
- ▶ Features included in dataset:
 - ▶ Latitude, Longitude
 - ▶ Median Income and Population Density
 - ▶ Foursquare: All Venues within 3000m of center of Zip Code.



In-N-Out Clusters: Training Set Results

Legend:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3



In-N-Out Clusters: Training Set Results

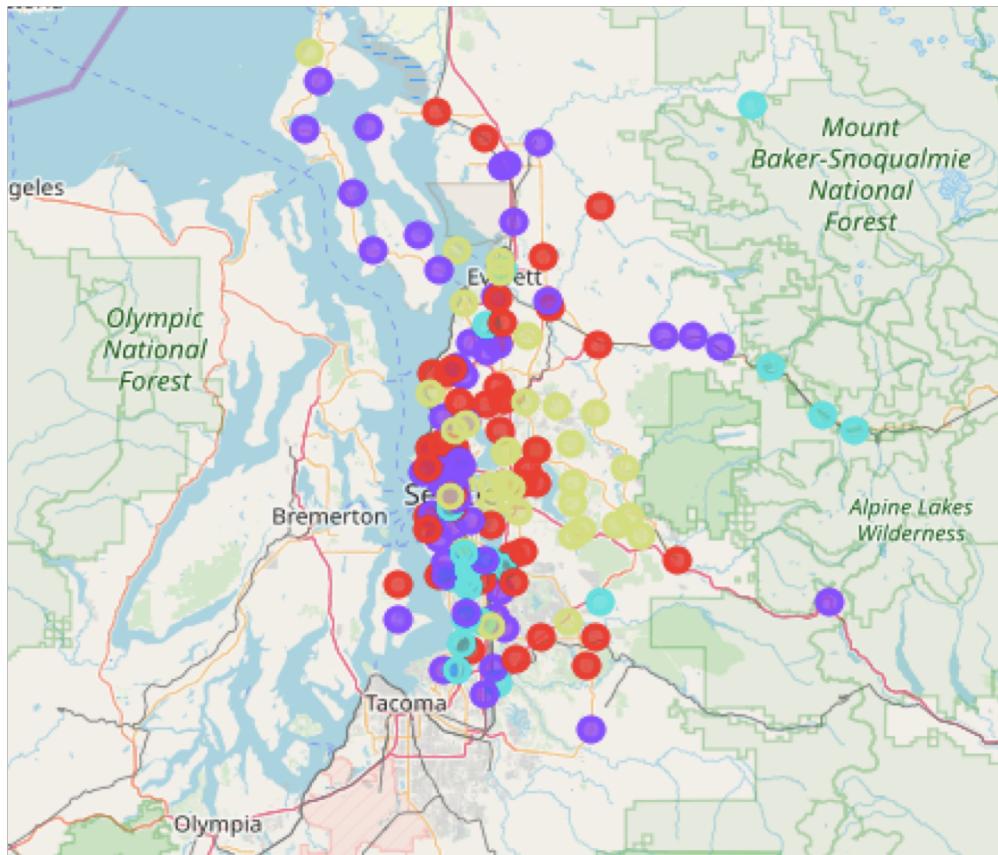
Cluster Labels	# of Restaurants in Cluster	Rand_Sales	Average Yearly Sales of Restaurant in Cluster
3	20	90638618	4.531931e+06
0	34	147291286	4.332097e+06
1	37	152682914	4.126565e+06
2	25	99691353	3.987654e+06

- ▶ Clustering successful based on clear separation in average sales between groups.
 - ▶ Shows that the features chosen for inclusion in the model are relevant to sales.
- ▶ Cluster 3 boasts the highest average yearly sales, followed by Cluster 0, Cluster 1 and Cluster 2. Cluster 1 brings in the highest revenue.
- ▶ Where are these top-tier stores located?
 - ▶ Moderately wealthy suburbs of SF (e.g. San Ramon, Livermore, San José)
 - ▶ Northwestern parts of Los Angeles (e.g. Hollywood)
- ▶ Where are the lowest performing stores located (Cluster 2)?
 - ▶ Poorer suburbs of San Francisco (e.g. Pinole, Union City)
 - ▶ Locations in the rural Central Valley (Visalia, Bakersfield)

Seattle Zip Code Cluster Predictions: Test Set Results

Legend:

- Tier 1
- Tier 2
- Tier 3
- Tier 4



Seattle Zip Code Cluster Predictions: Test Set Results

- ▶ The Northern suburbs of Seattle are predicted to perform worse than the Southern suburbs.
- ▶ Areas directly east of the city center are predicted to be the best performing.
- ▶ Further to the east, the further away from the urban core you go, the worse the areas are predicted to perform.
- ▶ I.e. the model makes very clear predictions about where the best performing regions will be.

Is Seattle a good potential market?

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%)

- ▶ Seattle has a predicted higher percentage of 1st tier and 2nd tier areas and a predicted lower percentage of 4th tier areas.
- ▶ Tier 3 brought in the most revenue of all the individual tiers in the training set. Would this remain the case in Seattle?
 - ▶ In-N-Out is well established and has a loyal following in CA
 - ▶ It can thus take more risks in opening stores that might underperform
 - ▶ Tier 3's earning potential is a long-term gamble

Conclusions and Future Directions

- ▶ Cluster model makes clear predictions as-is about which areas would be advantageous to explore further.
 - ▶ E.g. prediction of sales numbers using regression.
- ▶ Further modeling should make use of further demographic information beyond median income and population density.
- ▶ Further modeling should take into account all current In-N-Out locations.
- ▶ Advantage: this model can be used to predict viability in any area In-N-Out might wish to expand into