

Battle for Veg restaurants for Indian Settlements in Milpitas

2020

Final Report as a part of Capstone Project

VIJAY VITTAL



Battle for Vegfood

Synopsis

In and around Milpitas California, there are very few Indian Restaurant which serves pure vegetarian food. The problem often faced by the residents and customers is that, the restaurants are very crowded and operates limited time. the other problem faced by customers are delayed home delivery. the food often comes cold and many times the delivery is cancelled due to delay or cold food. in my capstone project i would like to seek answers for the below questions, I have used Anand Adiyar Bhavan (AAB) for capstone project.

1. How many Indian restaurants are there in and around Milpitas
2. What is the peak and off peak crowd
3. What is the rating of the hotel
4. Typical Delivery time
5. Best alternative place to open branch of AAB

Objective : To find feasible place to open Branch for Anand Aidyar Bhavan food outlet to increase sales and reduce the stress on main unit

Contents

1. Introduction where you discuss the business problem and who would be interested in this project.
2. Data where you describe the data that will be used to solve the problem and the source of the data.
3. Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
4. Results section where you discuss the results.
5. Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
6. Conclusion section where you conclude the report.

Introduction

There are around 500+K Indian population in California as per 2010 census. Most of the Indians are either working professionals or businesspeople. These people are located mostly in Milpitas, Sunnyvale and sacramental. Among those people, there are few people who are non-meat eaters and they follow strict vegetarian diet. There are only few Indian Restaurants which serves pure vegetarian and they don't serve any of the Meat items.

Of late, due to Covid-19, may of the customers reduced in dining facilities and started ordering online. *Due to increased order and limited staff, there has been increased waiting time and food is getting cold by the time it is delivered.* This is the main problem faced by the restaurants.

We have considered poplar food chain Anand Adiyar Bhavan (AAB) for data analysis. The set of problems discussed above applies to AAB

Population census

List of U.S. states by the population of Asian Indians

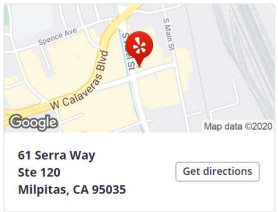
State	Asian Indian population (2010 Census)	% of state's population (2010 Census)
California	528,120	1.42%
New York	313,620	1.62%
New Jersey	292,256	3.32%
Texas	245,981	0.98%

18 more rows

[en.wikipedia.org › wiki › Indian_Americans](https://en.wikipedia.org/wiki/Indian_Americans)
Indian Americans - Wikipedia

AAB Miltpitas

Location & Hours



61 Serra Way
Ste 120
Milpitas, CA 95035

Get directions

Mon	11:30 am - 2:30 pm 5:30 pm - 11:00 pm
Tue	8:30 am - 2:30 pm 5:30 pm - 11:00 pm
Wed	8:30 am - 2:30 pm 5:30 pm - 11:00 pm
Thu	8:30 am - 2:30 pm 5:30 pm - 11:00 pm
Fri	8:30 am - 2:30 pm 5:30 pm - 11:00 pm
Sat	8:30 am - 2:30 pm 5:30 pm - 11:00 pm
Sun	8:30 am - 2:30 pm 5:30 pm - 11:00 pm

[Open now](#)

[Edit business info](#)

Ananda Bhavan–Milpitas

✓ Claimed

★ ★ ★ ☆ ☆ 183 reviews [Details](#)

\$\$ • Indian [Edit](#)

[Hours updated a few days ago](#)

[★ Write a Review](#)[Add Photo](#)[Share](#)[Save](#)

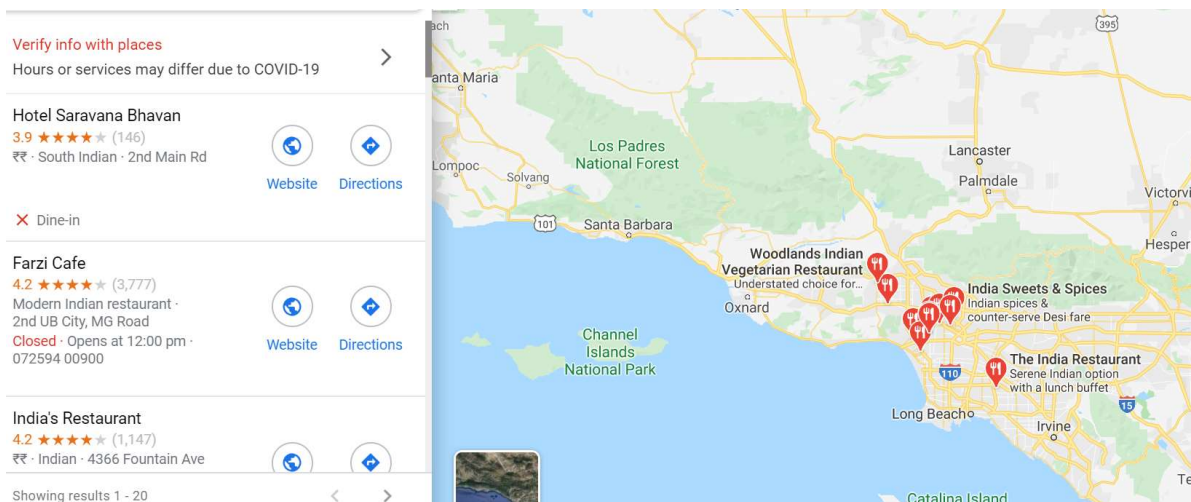
The outcome of this analysis will be helpful to AAB if they need to open a branch and if so where they need to open it.

DATA

There is no one single source of data available to perform analysis, so we have obtained data from different sources

The following sources of data obtained to perform the analysis

- Geospatial data : from www.Foursquare.com
- California postal information from <https://www.zipcodestogo.com/California/>
- Customer rating control: We collected manually since there were no data sets available and foursquare do not allow to get this for free, we created data set for K class analysis



Data Cleaning

The first data source in the described link is in json format. It initially consisted of many different classes of data. Upon examining them, the data that we are interested in was

found under 'features' category. Further formatting of the json data finally resulted in a dataframe that consists of 4 columns, namely: Milpitas, Neighborhood, Latitude and Longitude.

The second data source is a Wikipedia page that contains Postcode of the state California in a wikitable. Another problem was there were a few rows in the 'Neighborhood' column that too had no values assigned to it. As a solution, the value from the 'Indian' column of the respective row was copied into the 'Neighborhood' column

Now that we have obtained the different neighborhoods and their respective geometric coordinates for the city Milpitas, it is time to come up with different venues that the different venues have to offer. Foursquare API provides with an access to an enormous database consisting of venues from all around the world including rich variety of information such as addresses, tips, photos and comments. Having signed up for a Foursquare developer, using the Client ID and Client Secret, it is possible to make API requests in order to retrieve venue information. By feeding a function with Neighborhood name and its geometric coordinates, using Foursquare API different venues (Restaurants, Coffee shops, etc) were extracted.

Methodology section

Determining Optimal Cluster Number

K-means is a simple unsupervised machine learning algorithm that groups a dataset into a userspecified number (k) of clusters. The algorithm is somewhat naive--it clusters the data into k clusters, even if k is not the right number of clusters to use. Therefore, when using k -means clustering, users need some way to determine whether they are using the right number of clusters. One method to validate the number of clusters is the elbow method. The idea of the elbow method is to run k -means clustering on the dataset for a range of values of k (say, k from 1 to 20), and for each value of k calculate the sum of squared errors (SSE). Then, plot a line chart of the SSE

for each value of k . If the line chart looks like an arm, then the "elbow" on the arm is the value of k to be used.

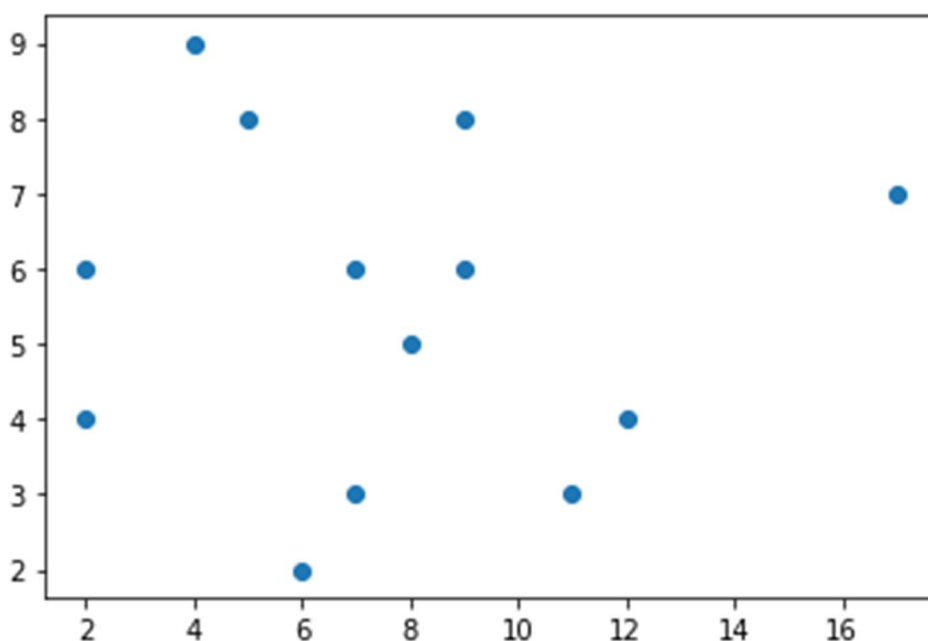
Random Initialization (for customer ratings)

Since, K-means incorporates a heuristic approach, it does not ensure converging at global optima at each iteration. Depending upon how the initial position of the clusters were set, it maybe converge into different local optima. In order to overcome this issue, numerous iterations on different random initializations were performed in order find the best set of convergence.

Results

There are not many veg hotels near Miltpitas, one is on the Sunnyvale side and there are not hotels near Sacramental side. The food is delivered to sacramental side. There is opportunity to open branch

Rating also provide similar results since rating coming from sacramental are less when compared to Sunnyvale



Conclusion: Need to open AAB branch. This will ease out the pressure for AAB on Sunnyvale side and increase sales and quality and reduced time of delivery