Coursera Capstone Report

for the

IBM Data Science Professional Certificate

Determining the optimum location of a New Vegan Restaurant in Houston, TX

By

Sharat Valluri

Date: 31-May-2020

Table of Contents

1.   Executive Summary

This report documents the evaluation procedure to determine the most optimal location to open a new vegan restaurant in one of the super Neighborhoods of the City of Houston, TX.

In this report we evaluate the various demographic, social and personal lifestyle influences that could lead to a greater adoption of plant-based diets and for the success of a purely vegan restaurant.

Neighborhoods are clustered based on the analysis of demographics and the popularity of various venue categories in each neighborhood.

The evaluation resulted in the identification of the top one or two suitable neighborhoods in each cluster and also the top three neighborhoods all over Houston which would be most favorable to a successful vegan restaurant.

2.   Introduction

Veganism is rapidly catching on in the US and the world as people become more concerned about the adverse impacts of purely animal-based diets on personal health, the heath of the planet and for personal ethical reasons such a the humane treatment of animals that has become very challenging due to the vast scope of animal factory farming.

Description of Project: The goal of this Data Science capstone project is to use various tools and techniques learned during the course of this IBM data Science program to identify the most suitable location among the Houston City Super Neighborhoods to open a new Vegan Restaurant.

I have become vegan since the last 4 years and am always on the lookout for new Vegan restaurants. However, these a few and far between even in the great gourmet city of Houston.

My wife and I have developed several amazing vegan recipes over the last few years, and we think some of these recipes are just too sumptuously, deliciously healthy for us to not share and promote them.
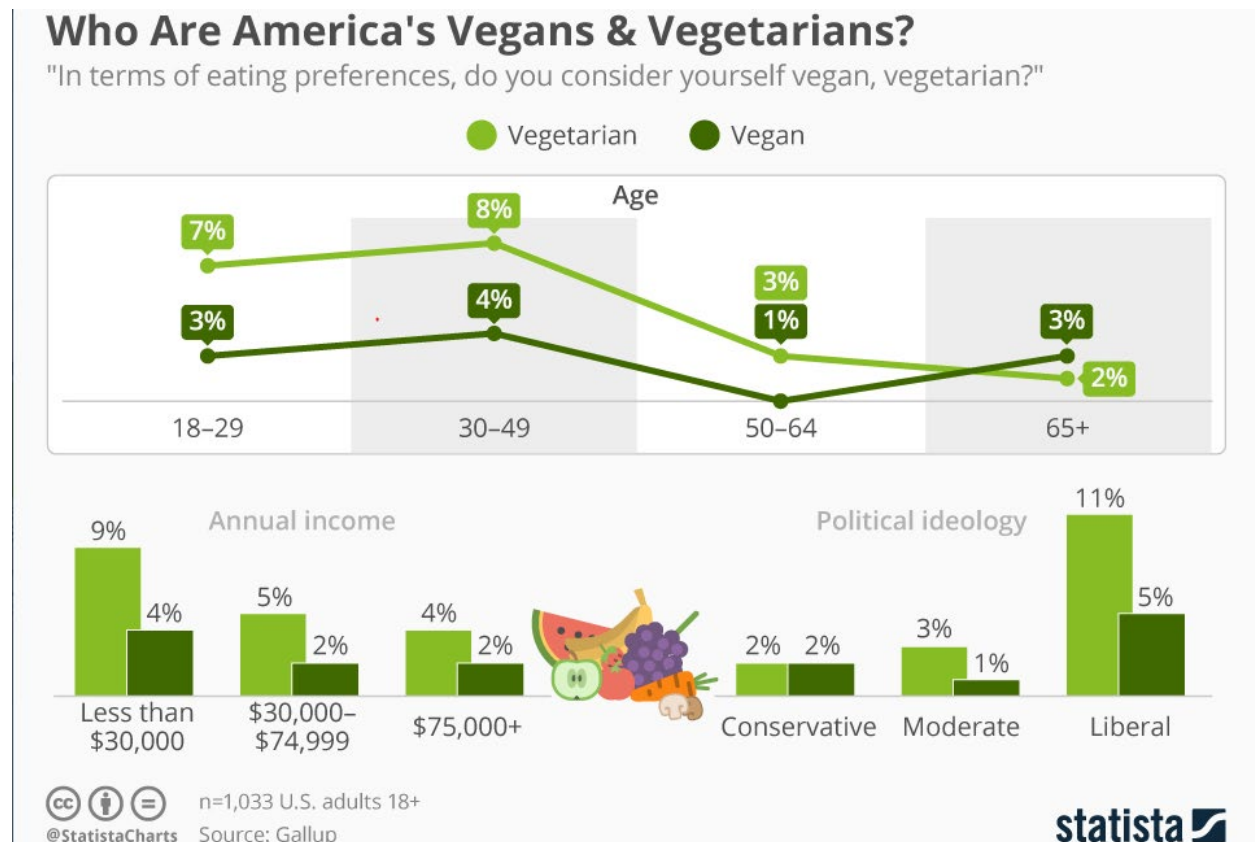
So began our quest to open our own Vegan restaurant. I believe this analysis will also be useful for other entrepreneurs who wish to start either a Vegan/ Vegetarian Restaurant in Houston. The analysis methodology can easily be applied to the evaluation of ideal locations for any other specialized cuisine as well.

3.   Literature Review

Based on results of several well publicized national and local surveys about veganism in the US, we decided to take a two-pronged approach for our location hunt.

Analyze Demographic data by each neighborhood to determine the suitability for a vegan eatery and analyze popular venue category data to evaluate and get clues on peoples tastes, preferences and pastimes.

Demographically, the factors that influence vegan diet preferences most are 1) A liberal Political Affiliation, 2) Gen Y age group, 3) Higher Education and 4) Mid-level Income (Low income levels see highest prevalence of veganism, which I believe is due to the adoption by the upwardly mobile younger professionals, who have not yet moved into the high income brackets).

Who Are America's Vegans & Vegetarians?
"In terms of eating preferences, do you consider yourself vegan, vegetarian?"

● Vegetarian   ● Vegan

Age

	18–29	30–49	50–64	65+
Vegetarian	7%	8%	3%	2%
Vegan	3%	4%	1%	3%

Annual income

	Less than $30,000	$30,000–$74,999	$75,000+
Vegetarian	9%	5%	4%
Vegan	4%	2%	2%

Political ideology

	Conservative	Moderate	Liberal
Vegetarian	2%	3%	11%
Vegan	2%	1%	5%

n=1,033 U.S. adults 18+
@StatistaCharts   Source: Gallup

statista

- From 2018 Gallup Survey, Courtesy Statista.

On the Venues front, I have determined that holistic well-being and fitness consciousness are the primary drivers of adopting a vegan lifestyle. Therefore, neighborhoods where fitness venues, nature trails and parks are popular will be prime candidate locations. Also a liberal political affiliation is often associated with acceptance of diversity in the community and and an openness to try new foods/ cuisines will be an asset, which is in turn indicated by the popularity of restaurants with cuisines from all parts of the world.

	HOLISTIC WELLBEING SEEKERS	FITNESS & LIFESTYLE COMMUNITY	ETHICAL FOODIES
Volume of posts	US: 320k | UK: 89k	US: 239k |s UK: 83k	US: 125k | UK: 43k
What drives them	self-love, self-care, wellness, mindfulness, curiosity	self-optimization, weight loss, health	altruism, compassion animal welfare, sustainability
Popular hashtags	#plantbased #vegan #cleaneating #organic #healthyfood #wellness #yummy	#vegan #healthy #fitfam #motivation #healthychoices #weightloss #plantpower #fitfood	#crueltyfree #nongmo #organic #realfood #natural #animalrights #natural #sustainable

Source: Linkfluence social data research

Based on these preliminary assumptions, data from the public domain and the tools form the course, below is my methodology and results.

4.   Methodology and Data Sources

The primary listing and description for the Houston City Super Neighborhoods is available on Wikipedia and can be scraped from that directly.

url = https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods

The Demographic Data for Houston by Super Neighborhoods is available from the Houston City Government websites.

https://www.houstontx.gov/planning/Demographics/

https://www.houstontx.gov/planning/Demographics/demograph_docs/demo_index.htm

 Unfortunately these websites provide the output in pdf format, not readily readable or scrapable. Therefore, I have compiled these demographic data from the various pdfs into a consolidated Excel Spreadsheet. This spreadsheet will then be imported into a data frame and merged with the Wiki Super Neighborhood listings dataframe.

Data regarding Political Affiliation and Racial and Ethnic diversity both of which are important determinants of Vegan preferences is not readily available on the government website. I tried to scrape a few private websites for this information, but they were blocked with a 403 error. So, I manually compiled Diversity and Political Affiliation data from the websites into a spreadsheet and merged it with my Demographic data.

Url = https://www.niche.com/places-to-live/search/most-diverse-neighborhoods/m/houston-metro-area

Then I used the Google API for getting the latitude and longitude data for the Super Neighborhoods and finally used Foursquare API venue category data to analyze the dietary and lifestyle preferences and pastimes.

All this data needed to be cleaned, wrangled and made compatible for analysis and the data clean up at each stage is documented in the Python notebook.

Finally, I used all this compiled data to cluster neighborhoods to determine which neighborhood would best support a vegan restaurant, based on the various demographic and lifestyle preferences of the members of that neighborhood.

5.   Results and Discussion:

Map courtesy – Google.com
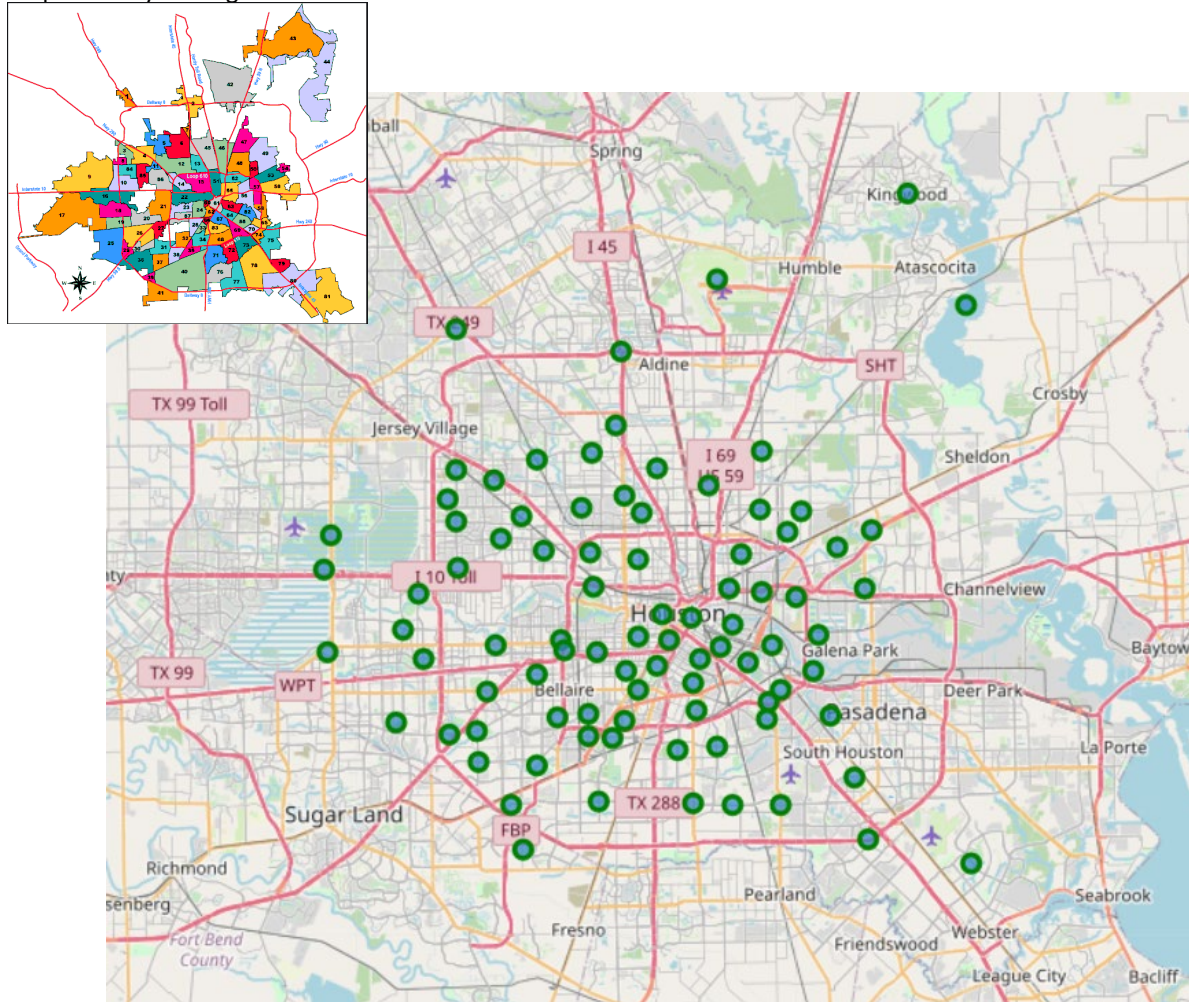
The map above shows the centers of each of the superneighborhoods which were evaluated. Once the venue data was retrieved from the FourSquare API, I used it to find the venue categories and which

were the more popular ones for each neighborhood. I used K-menas clustering to cluster neighbor hoods based on the popularity of the the venue categorises to provide insights on whether the residents preferred an active or sedentary life style, whether they were adventurous in their food habits evidenced by a profusion of international cuisnes restaurants versus traditional US foods, etc..

The demographic data I retreived was categorical so I could not combine that with the Venue data for the clustering sicne since K-means clustering does not accommodate the analysis of categorical values, and I have several Categorical values in the data set being analyzed, applying some other clustering mechanisms such as K-Modes analyses may further enhance this evaluation. This could be a futre enhancement to the analysis.

We found the Houston Neighborhood clustered by venue popularity into six gorups (I tried 5 & 7 clusters as well, but 6 gave the most efficient and dileneated clustering. Considering that Personal health and fitness were found to be the most powerful motivators for vegan diet adoption, I looked at clusteres that had a lot more physical fitness venues as being the most favorable. Then I consdered other factors in the order of :

Fitness Venues → Exotic Food interests → Specialty Foods (Gluten Free, Organic) → Nature Lovers → Art Lovers → Traditional Food Interests.

To pick specific neighborhoods among a cluster, I used the demographic data. Vegans have shown strong correlation in surveys to Liberal political affiliation. So I used the foollowing order to pick the top one or two neighbor hoods in each cluster.

Liberal → Gen Y → Racially & Ethnically Diverse → Higher Education → Mid-Income

Below are the final results form the analysis.

The top 3 Neighborhoods to open a new Vegan Restaurant in Houston based on above analysis are:

<mark>1st Choice</mark>**: Cluster 2, Neighborhood 8: Adicks Park Ten:**

Venues: Nature, Fitness, Specialty Foods, Art, Regular, Exotic

Demog: Liberal, Gen X, Very Diverse, Basic_Edn, Mid_Income, Low_Pop_Density,

<mark>2nd Choice</mark>**: Cluster 2, Neighborhood 15: Memorial Area:**

Venues: Nature, Fitness, Art, Regular Restaurants, Exotic Restaurants, Specialty Foods

Demog: Liberal, Diverse, High_Edn, High_Income, Boomers, Med_Pop_Density

<mark>3rd Choice</mark>**: Cluster 6,Candidate for Opening a Vegan Restaurant in this Cluster - 14 - Greater Heights:**

Demog: Liberal-Very Diverse-High_Edn, High_Income, Gen X High_Pop_Density

Venues - Top : Regular Restaurants, Nature, Exotic, Fitness, Art, Specialty_Foods

6.   Conclusions and Further Analysis:

It is possible to use this analysis not only to determine the optimum location of a Vegan Restaurant in Houston, but the analysis can easily be tweaked to apply to any other specialized cuisine restaurant.

I have applied the K-means clustering to only the Venue data in this analysis. However, since K-means clustering does not accommodate the analysis of categorical values, and I have several Categorical values in the data set being analyzed, applying some other clustering mechanisms such as K-Modes analyses may further enhance this evaluation.

7.   References:

https://www.houstontx.gov/planning/Demographics/demograph_docs/demo_index.htm
https://www.houstontx.gov/planning/Demographics/
https://plantproteins.co/vegan-plant-based-diet-statistics/
https://cdn.knightlab.com/libs/timeline3/latest/embed/index.html?source=1zG9HnTaxLUxgRiTS2PIuLt
WRKXCy0Si_xiwNHfC1IT4&font=Default&lang=en&timenav_position=top&initial_zoom=1&height=900