Vegan and Vegetarian Restaurants:
The Best Business Sustainable Idea

Juan Trujillo Fernández

February 28, 2021

Contents

1 Introduction

Nowadays, the world is in constant change. In a fast-growing society like ours, with the consequent new necessities, our lives must follow suit. Food habits is one main way, and a really important one. Meat consumption is a really strong pillar of our food habits, but we don't know how it impacts on the environment (or in ourselves).

Recent studies show that the relationship between meat consumption and contamination flows on CO_2 and CH_4 emissions (see [1], [3]), the use of water in agriculture for livestock, the deforestation (in this case, in Brazil) aimed to livestock (see [2]), etc. accelerating climate change. Health is another focus point on veganism and vegetarianism: the meta-analysis exposed in [4] is a summary of the benefits of this kind of diets in, for example, preventing mortality from cardio-cerebrovascular diseases.

Not just being eco-friendly and healthy; it's also sustainable. In [5], some examples in Europe highlight the improvement of this industries regarding meat industry. This obviously is extensible for new companies in the USA, as well as all over the world. In [6] there is much more information about these points.

The upcoming impact on the daily life of the mean citizen is, almost, a fact. So that, some people come up with the idea of a business vegetarian and/or vegan oriented. Here we analyse where is the better place to create your own vegan and/or vegetarian restaurant.

2 Data exploration

2.1 Data Sources

I've obtained the information of the restaurants ([7]) from Kaggle. Also, the state-level income dataset is from there. Since all the data is filled, there is no need to look for another data. We'll take too the data from Foursquare of the city of New York.

2.2 Data cleaning and organizing

In order to relate the data from all the sources, I've taken a dataframe from wikipedia relating USA states with their abbreviations. I've chosen some of the columns from both of the dataframes, specially the ones from I could extract information, such that the category of the business or the statistics relative to every state. In the vegan and vegetarian restaurants' dataframe there were many classifications of the businesses included. I simplified that information taking the first classifier.

After that, I extracted the information from the Foursquare API of the vegetarian and vegan businesses of New York in order to complete the Kaggle dataframe. With all that information, I modelled the state-level income dataframe to the information in the veg dataframe, connecting the abbreviations of the states. With that, I ended the process dropping out the repeated businesses looking them by their name.

In the dataframe 25 state make appearence, more precisely: Arizona, California, Florida, Georgia, Hawaii, Illinois, Kansas, Maine, Massachusetts, Michigan, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, Oregon, Pennsylvania, Tennessee, Texas, Utah, Virginia, Washington and Wisconsin.

The features that I dropped were no analysable, just like the menu section of the veg restaurants, or, in my personal opinion, just not related to the study (like the median age or the black/white and Latin/white comparison).

3 Exploration of the Data

The complete amount of businesses is in figure 1. It represents the distribution of this kind of business all around the 25 states presented in the study.

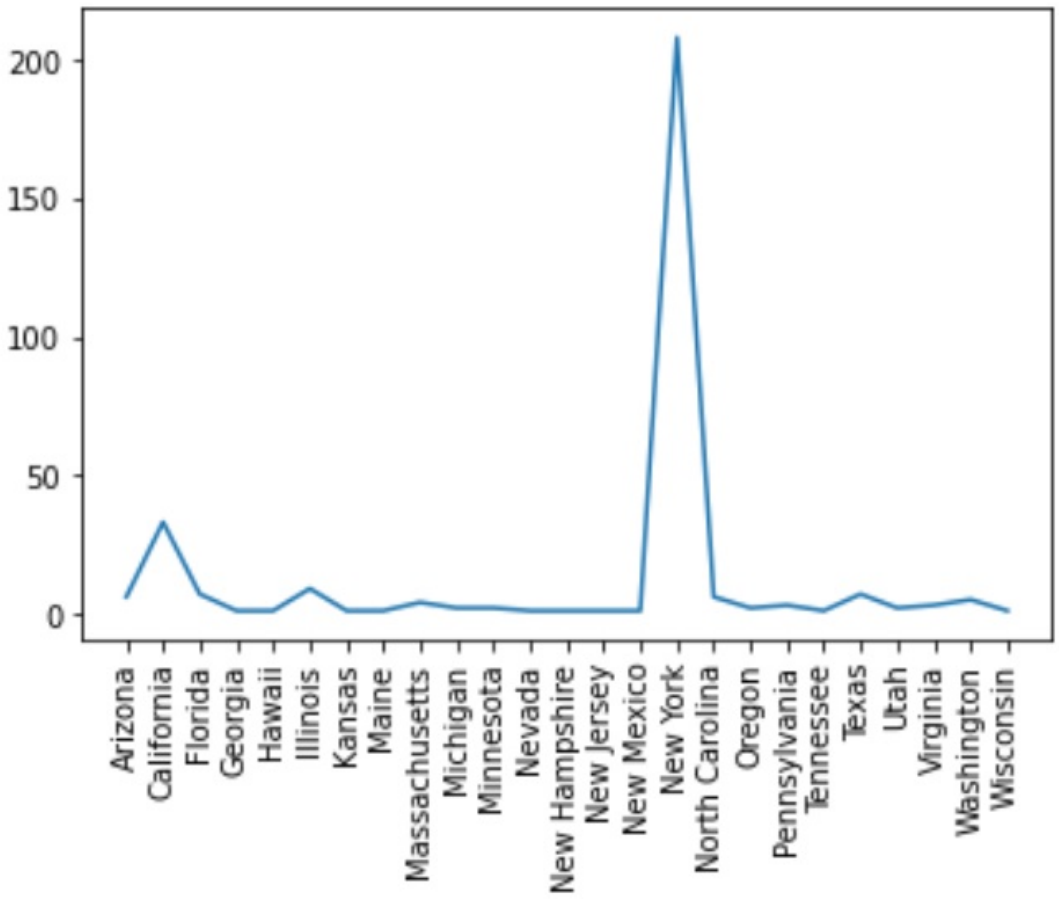

Figure 1: Total amount of businesses by state.

Having these data, it's trivial to think that the major amount of places will be on New York, but we'll try to do a generalised analysis.

This analysis will consist, first of all, in picking the k-means algorithm into the total amount of data grouped by state, trying to overcome the difficulties of the unequal distribution of the businesses. I'll pick 5 different clusters, and then extracting the conclusions over the results of the algorithm. After that I'll show the difference between the results with five different dataframes that contain the top 3 categories of every state along with the cluster level. With that I present a summary of the data obtained from [8] of the states that are involved in every level that will help with the analysis and the

conclusion because of its implications on the analysis.

3.1 Cluster level 0

The cluster is compounded of this states:

	State	Cluster Labels	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	Count
1	California	0	Restaurant	Indian Restaurant	Natural food restaurants	33
5	Illinois	0	Restaurant	Mexican Restaurant	Asian Restaurants	9
10	Minnesota	0	Beauty Salons	Restaurant	--	2
13	New Jersey	0	Natural food restaurants	--	-	1
22	Virginia	0	Asian Restaurants	Restaurant	--	3
23	Washington	0	Restaurant	Bars Clubs	--	5

I've included the Count column because there are states with less than 3 results, and it's something to be aware of. This will happen in every cluster level.

Along with this data:

	State	Cluster Labels	Labor force participation	Unemployment rate	Poverty rate	Gini coefficient	GDP per capita	People with less than 9 years of education/people with college degree or above
1	California	0	64.172727	8.281818	11.190909	0.476300	57225.127065	0.341333
5	Illinois	0	66.463636	7.536364	9.918182	0.470091	55323.663187	0.190934
10	Minnesota	0	71.400000	5.300000	6.909091	0.441000	54911.237655	0.104328
13	New Jersey	0	65.945455	6.981818	7.418182	0.467736	59402.801344	0.154662
22	Virginia	0	67.590909	5.045455	7.736364	0.461218	54667.508723	0.149070
23	Washington	0	66.081818	6.909091	8.554545	0.446655	57899.217871	0.127759

You can see that the major amount of businesses are restaurants with a vegetarian or vegan option, which is in every state. There are some other international cuisine restaurants that have vegetarian and/or vegan options in their usual menu (for example, some plates in Indian cuisine). But the difference between the amount of restaurants between one state or another, or the type of businesses, doesn't show the reasoning on putting them together. But if we watch the data we can see that, with the exception of Minnesota, they all have similar labor force participation, and, dropping California, the relation between people's education differs by 0,1. But the most revealing fact of their relation is in the GDP per capita, and I think that this is the main reason of the classification.

3.2 Cluster level 1

The cluster is compounded of this states:

	State	Cluster Labels	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	Count
3	Georgia	1	Restaurant	--	-	1
6	Kansas	1	Indian Restaurant	-	--	1
16	North Carolina	1	Restaurant	New American Restaurant	Grocery Store	6
17	Oregon	1	Restaurant	--	-	2
21	Utah	1	Restaurant	--	-	2
24	Wisconsin	1	Restaurant	--	-	1

Along with this data:

	State	Cluster Labels	Labor force participation	Unemployment rate	Poverty rate	Gini coefficient	GDP per capita	People with less than 9 years of education/people with college degree or above
3	Georgia	1	65.118182	7.436364	12.927273	0.471709	46127.402215	0.209254
6	Kansas	1	69.572727	5.309091	8.781818	0.447818	48176.654972	0.135461
16	North Carolina	1	63.536364	7.436364	12.127273	0.466782	46612.698606	0.217748
17	Oregon	1	64.136364	7.609091	10.436364	0.451882	45165.109819	0.142109
21	Utah	1	69.472727	4.772727	8.445455	0.418427	45501.514735	0.104444
24	Wisconsin	1	69.490909	6.181818	8.109091	0.432945	47745.438506	0.130980

As before, the main reason behind the classification is the low GDP per capita. Also, the accumulation of businesses isn't high, and there is one store between the results, besides the restaurants' hegemony. Also, it's worth to see the relation between people's education (the highest state has the maximum of businesses) and the relatively high labor force participation.

3.3 Cluster level 2

The cluster is compounded of this states:

	State	Cluster Labels	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	Count
8	Massachusetts	2	Restaurant	--	-	4
15	New York	2	Vegetarian / Vegan Restaurant	Restaurant	Indian Restaurant	208

Along with this data:

	State	Cluster Labels	Labor force participation	Unemployment rate	Poverty rate	Gini coefficient	GDP per capita	People with less than 9 years of education/people with college degree or above
8	Massachusetts	2	65.981818	6.127273	7.800000	0.473655	64915.326216	0.125295
15	New York	2	62.163636	6.590909	11.390909	0.502536	64815.733018	0.209493

Its huge GDP per capita is far away the highest between all other states, and determines the arising of the level. It's also interesting to see the relation between the amount of restaurants and the comparison of the people's education. Although I've taken more data from the Foursquare page, New York City is one of the bigger cities all around the world, and there was a high probability of finding a big amount of businesses in the city. Also, we can see a high Gini coefficient, especially in the state of New York.

3.4 Cluster level 3

The cluster is compounded of this states:

	State	Cluster Labels	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	Count
0	Arizona	3	Restaurant	Asian Restaurants	Vegetarian / Vegan Restaurant	6
2	Florida	3	Restaurant	Bars Clubs	New American Restaurant	7
7	Maine	3	Health Food Restaurants	--	-	1
9	Michigan	3	Pizza	Mediterranean Restaurant	--	2
14	New Mexico	3	Restaurant	--	-	1
19	Tennessee	3	Restaurant	--	-	1

Along with this data:

	State	Cluster Labels	Labor force participation	Unemployment rate	Poverty rate	Gini coefficient	GDP per capita	People with less than 9 years of education/people with college degree or above
0	Arizona	3	62.245455	7.063636	12.054545	0.458882	41854.130475	0.256026
2	Florida	3	61.709091	6.927273	10.927273	0.475800	42306.031727	0.212094
7	Maine	3	65.218182	6.172727	8.963636	0.440882	40482.568930	0.129075
9	Michigan	3	62.336364	8.727273	11.245455	0.454655	42591.754641	0.135057
14	New Mexico	3	60.554545	6.172727	15.018182	0.466709	43001.618282	0.293175
19	Tennessee	3	62.263636	7.190909	12.881818	0.472464	43883.257320	0.271175

This situation is similar to cluster level 1, with the exception on the people's education correlation: besides having to states with a high percentage of college graduates compare with other states (New Mexico and Tennessee vs Arizona and Florida), there are more businesses in the second group, and the first ones have lower percentage. Also, the poverty rate is higher compared with the other levels.

3.5 Cluster level 4

The cluster is compounded of this states:

	State	Cluster Labels	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	Count
4	Hawaii	4	Southeast Asian	--	-	1
11	Nevada	4	Japanese Restaurant	--	-	1
12	New Hampshire	4	Mexican Restaurant	--	-	1
18	Pennsylvania	4	Restaurant	Coffee Shop	--	3
20	Texas	4	Restaurant	Italian Restaurant	Gluten-free Restaurant	7

Along with this data:

	State	Cluster Labels	Labor force participation	Unemployment rate	Poverty rate	Gini coefficient	GDP per capita	People with less than 9 years of education/people with college degree or above
4	Hawaii	4	63.490909	4.763636	7.281818	0.432991	53741.171865	0.154596
11	Nevada	4	66.109091	8.418182	10.000000	0.442727	50431.417945	0.289371
12	New Hampshire	4	69.909091	4.554545	5.363636	0.427718	50816.245219	0.083327
18	Pennsylvania	4	63.818182	6.372727	8.927273	0.461364	49886.436044	0.139191
20	Texas	4	65.645455	5.945455	13.409091	0.476227	53371.744154	0.373565

In this level we find the lowest relation about people education, located in New Hampshire (which has only one result, and it's a Mexican restaurant). About the cluster relation, excluding Nevada (a rare case because of its low GDP per capita and the high education ratio) and Texas, there is a small amount of college degrees in relation to people with less than 9 years of education, and a relatively low GDP per capita. The other factors maintain the mean levels.

4 Conclusions

I've stressed the relation between higher education and the amount of veg businesses, so I think that the main reason for choosing a place to begin your business is one of the top places though. Too, the GDP per capita is important, but is not so relevant because of the partition made by the model. Also, the geography is not important such as population density, as you can see in this map (every colour represents the color of every cluster, and it's an interactive map in which you can see the name of the state):

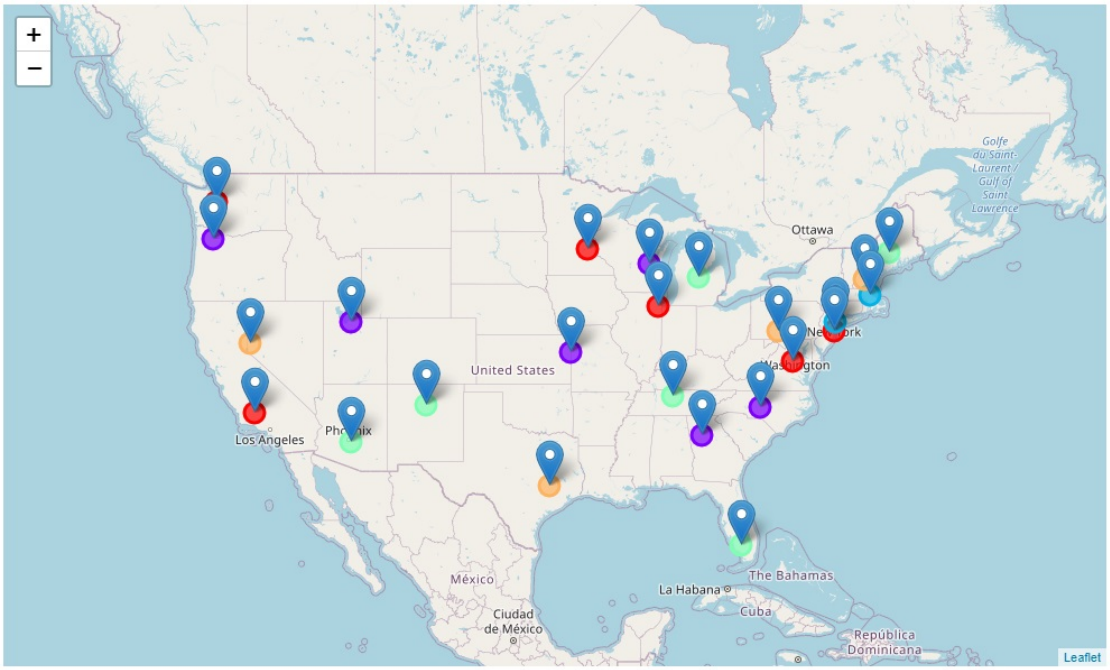

Figure 2: Map of the distribution of the clusters.

I recommend you to put a business in a rising power, such as Texas, Arizona and Massachusetts.

References

[1] Ivanova, D; Barrett, J; Wiedenhofer, D; Macura, B; Callaghan, M; Creutzig, F, 2020, 'Quantifying the potential for climate change mitigation of consumption options', *Environmental Research Letters*, vol. 15, no. 9 (link).

[2] Godfray, H. C. J; Aveyard, P; Garnett, T; Hall, J. W; Key, T. J; Lorimer, J; Pierrehumbert, R. T; Scarborough, P; Springmann, M; Jebb, S. A, 2018, 'Meat consumption, health, and the environment', *Science*, vol. 361, issue 6399 (link).

[3] Herrero, M; Gerber, P; Vellinga, T; Garnett, T; Leip, A; Opio, C; Westhoek, H. J; Thornton, P.K; Olesen, J; Hutchings, N; Montgomery, H; Soussana, J-F; Steinfeld, H; McAllister, T.A, 2011, 'Livestock and greenhouse gas emissions: The importance of getting the numbers right', *Animal Feed Science and Technology*, vol. 166-167, pp. 779-782 (link).

[4] Dinu, M; Abbate, R; Gensini, G. F; Casini, A; Sofi, F, 2017, 'Vegetarian, vegan diets and multiple health outcomes: A systematic review with meta-analysis of obser-

vational studies' *Critical Reviews in Food Science and Nutrition*, vol. 57, no. 17, pp. 3640-3649 (link).

[5] Saari, U. A; Herstatt, C; Tiwari, R; Dedehayir, O; Mäkinen, O. J, 2021, 'The vegan trend and the microfoundations of institutional change: A commentary on food producers' sustainable innovation journeys in Europe', *Trends in Food Science & Technology*, vol. 107, pp. 161-167 (link).

[6] Compend of articles on vegetarianism and veganism in https://www.vrg.org (link).

[7] 'Vegetarian and Vegan Restaurants' dataset from https://www.kaggle.com (link).

[8] 'Annual state-level income of USA' dataset from https://www.kaggle.com (link).