

Vegan and Vegetarian Restaurants: The Best Business Sustainable Idea

Juan Trujillo Fernández

February 28, 2021

Index

- 1 Introduction
- 2 Data Cleaning and Organization
- 3 Data Exploration
- 4 Conclusions

Index

- 1 Introduction
- 2 Data Cleaning and Organization
- 3 Data Exploration
- 4 Conclusions

Introduction

- Meat consumption increases pollution and climate change.
- Vegan and vegetarian diets are healthy :) .
- Also, vegan and vegetarian industry is increasing.
- Therefore it is a good opportunity for a business model!

Index

- 1 Introduction
- 2 Data Cleaning and Organization
- 3 Data Exploration
- 4 Conclusions

Sources

- Two dataframes from Kaggle (www.kaggle.com).

kaggle

- Information via Foursquare (www.foursquare.com).



- To link the datasets, one table from Wikipedia.



Data Cleaning

Steps:

- 1 Drop useless columns in both dataframes.

Data Cleaning

Steps:

- 1 Drop useless columns in both dataframes.
- 2 Take the information with Foursquare API and join it to the restaurants dataframes.

Data Cleaning

Steps:

- 1 Drop useless columns in both dataframes.
- 2 Take the information with Foursquare API and join it to the restaurants dataframes.
- 3 Reduce the categories for not overfitting the model.

Data Cleaning

Steps:

- 1 Drop useless columns in both dataframes.
- 2 Take the information with Foursquare API and join it to the restaurants dataframes.
- 3 Reduce the categories for not overfitting the model.
- 4 Summary the restaurants dataframe by unique restaurants.

Data Cleaning

Steps:

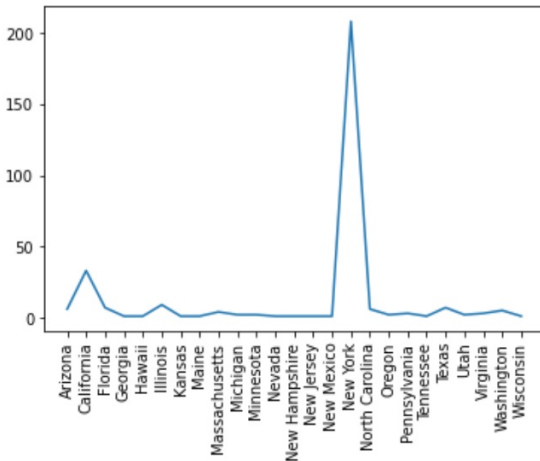
- 1 Drop useless columns in both dataframes.
- 2 Take the information with Foursquare API and join it to the restaurants dataframes.
- 3 Reduce the categories for not overfitting the model.
- 4 Summary the restaurants dataframe by unique restaurants.
- 5 Join the two main dataframes.

Index

- 1 Introduction
- 2 Data Cleaning and Organization
- 3 Data Exploration
- 4 Conclusions

Data Exploration - First Steps

Sight of the distribution of the restaurants:



Data Exploration - Methodology

- 1 Use `get_dummies` function to turn into number the categories.

Data Exploration - Methodology

- ① Use `get_dummies` function to turn into number the categories.
- ② Create a top 3 table to compare results (because some states only have 1 or 2 results).

Data Exploration - Methodology

- ① Use `get_dummies` function to turn into number the categories.
- ② Create a top 3 table to compare results (because some states only have 1 or 2 results).
- ③ Create a model of k-means algorithm to analyse the data.

Data Exploration - Methodology

- ① Use `get_dummies` function to turn into number the categories.
- ② Create a top 3 table to compare results (because some states only have 1 or 2 results).
- ③ Create a model of k-means algorithm to analyse the data.
- ④ See and interpret the final data classification.

Data Exploration - Example

	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

Figure: Data from the cluster level 2.

	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

Figure: States data from the cluster level 2.

Index

- 1 Introduction
- 2 Data Cleaning and Organization
- 3 Data Exploration
- 4 Conclusions

Conclusions - Map

Map containing all the presented states and their cluster level.

