



A model to identify new neighborhoods that could support a vegan / vegetarian restaurant

By Marina Santiago

1/2/ 2021



Introduction

Vegetarian and vegan restaurant are becoming more popular, particularly in large cities. However, opening these restaurants in smaller cities can still be risky because of the smaller population of vegetarians.

What if we could decrease this risk by predicting which neighborhoods would be most likely to support a vegan / vegetarian restaurant?



Who would be interested in this model?

- Small business owners looking to open a new restaurant
- Restaurant investors
- Developers
- Everyday people looking to move to a new neighborhood



Data acquisition

- Venue category data from Foursquare for all cities
- Longitude and Latitude data from geopy
- Neighborhood demographic data for New York City from <http://app.coredata.nyc>
- Neighborhood demographic data for Chicago from <https://datahub.cmap.illinois.gov>
- Neighborhood demographic data for Los Angeles from <http://la.myneighborhooddata.org/data>
- Neighborhood demographic data for Cleveland from <https://www.communitysolutions.com/resources/community-fact-sheets/cleveland-neighborhoods-and-wards/>

Methodology

Process training data

The data from NYC, Chicago, and LA was processed to create a simplified dataset that is generalizable to many cities.

Build model

A decision tree classification model was built using the training data from NYC, Chicago, and LA.

This model finds neighborhoods with a vegan / vegetarian restaurant with > 80% accuracy

Test model on new dataset

The model was tested using data from Cleveland, OH to find new neighborhoods that could support a vegan / vegetarian restaurant.

The model identified the only neighborhood in Cleveland with a vegan / vegetarian restaurant and two other neighborhoods



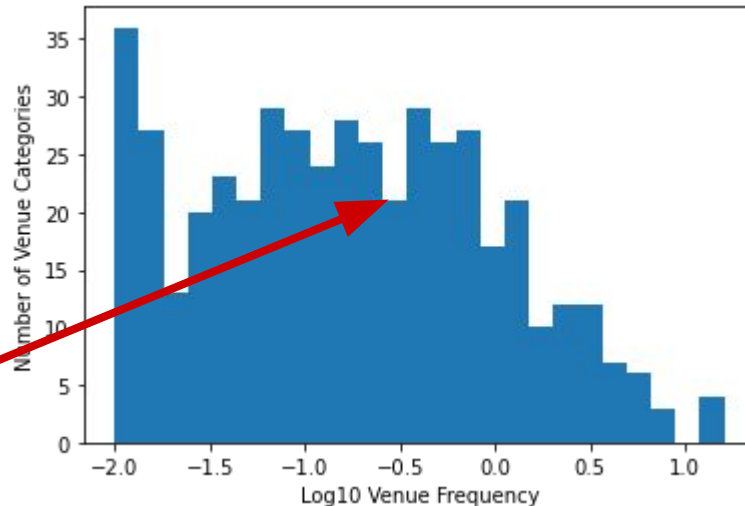
Creating a simplified and generalizable training dataset

1. Demographic data was collected for each neighborhood
 - a. Median Income
 - b. Percent white
 - c. Percent aged 65+
2. The location (latitude and longitude) of each neighborhood were identified
3. Venue categories were identified using Foursquare

Creating a simplified and generalizable training dataset

4. Venue categories were decreased by identifying the most abundant venue categories across the full dataset. Number of venue categories decreased to 186 from 469.

A cutoff of -0.5 (\log_{10} x venue frequency) was chosen because it naturally divides the data into approximately half





Creating a simplified and generalizable training dataset

5. To make the data more generalizable, the 186 venue categories were combined into 24 new categories by summing up all members of the new category. . These new categories are.

- Stores
- Offices
- Indoor Entertainment
- Outdoor Entertainment
- Health
- Transportation
- Sports related venues
- Restaurants
- Bars / Brewery
- Automotive
- Cafe / Bakery / Dessert Shop
- Food / Groceries
- High School
- ATM
- Currency Exchange
- Flea Market
- Hotel
- Intersection
- Rental Car Location
- Salon / Barbershop
- Shopping Mall
- Storage Facility
- Waste Facility
- And finally our target variable: Vegetarian / Vegan Restaurant



Building the decision tree classifier

The decision tree classifier was built using the following parameters

- The training data was split into 50% training set and 50% test set
- Entropy was used as the criterion

The model outputs a prediction accuracy of 81 - 82%.



The five most important features for the model

Feature	Importance
Bars / Brewery	0.290647
Indoor Entertainment	0.170038
Outdoor Entertainment	0.138880
Percent_white	0.095530
Transportation	0.081698

These features make sense because a restaurant is more likely to do well in areas with nearby entertainment and transportation infrastructure. Nearby bars also means that people will likely need to eat, so a restaurant is likely to be successful there. The fact that the model outputs features that you would expect to be important for a restaurant's success supports the model.



Validating the model with a new city: Cleveland, OH

The model predicts that three neighborhoods (of the 29 in Cleveland) will support a new vegan / vegetarian restaurant:

1. Buckeye-Woodhill
2. Detroit Shoreway
3. Euclid Green

Euclid-Green is actually the only neighborhood in Cleveland that already has a vegan / vegetarian restaurant

This validates the model!

It also suggests that the other two neighborhoods could support a new vegan / vegetarian restaurant

Tommy's restaurant in Euclid-Green has been very successful



- Founded in 1972
- Trip Advisor rates it #43 out of 1720 restaurants and has 4.5 stars
- They have continued to stay open during the covid pandemic suggesting that they are doing well financially and that their community supports them.

This model identifies communities that can support new vegan / vegetarian restaurants

The other communities identified in Cleveland are known for being vibrant and diverse

From <http://livecleveland.org/>





Conclusion

- I was able to create a model that predicts which neighborhoods can support a new vegetarian / vegan restaurant with more than 80% accuracy.
- This model could be improved by increasing the amount of training data, but even in its current state, it correctly identified the only neighborhood in Cleveland that currently has a vegan / vegetarian restaurant, validating the model.
- It identified two other neighborhoods that could potentially have a vegan / vegetarian restaurant.
- Further analysis is needed to confirm this model's recommendations, but based on the descriptions of these neighborhoods, they sound very promising.