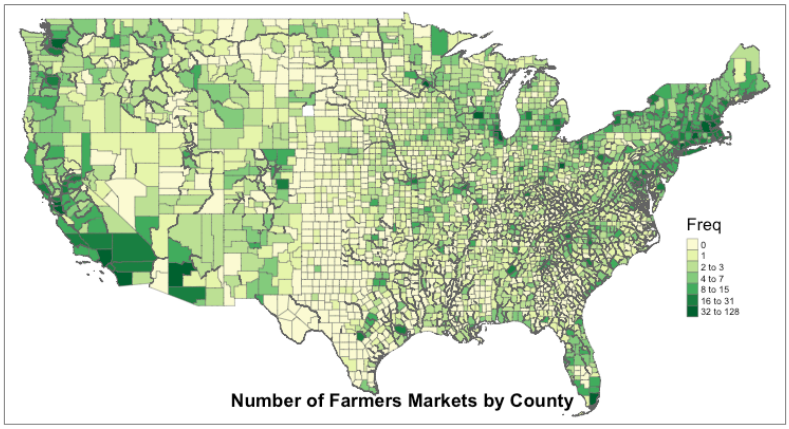
Executive Summary

Negative Binomial Regression of Farmers Market Occurrences
WGU Master's Capstone Project
Prepared by Michiel Besseling

Problem & Hypothesis

The number of farmers markets in the United States has risen steadily since the beginning of the 21st century. Farmers markets benefit the communities they serve. Their "direct to consumer" business model allows farmers to sell their produce directly to the consumer. Farmers markets across the country offer not only a place for local farmers to sell their produce, but a venue for members of the community to socialize. According to McCarthy (2010), farmers markets benefit the farmers who participate by increasing profits and establishing relationships with the local community. Consumers benefit by having access to fresh and healthy foods and by strengthening ties with the local community.



Currently in 2020, there are 8,804 farmers markets registered with the USDA. An initial examination of their geographic distribution reveals that there are likely regions that are underserved by farmers markets, providing an opportunity to open new farmers markets in those locations. This report is intended for entrepreneurs who wish to open new farmers markets in potentially promising markets. Should an entrepreneur seek to open a new farmers market with or without a location already in mind, useful insights would be gained by creating and consulting a profile of locations that includes observed and predicted number of farmers markets and that ranks locations in terms of which locations deviate most from what is projected. Thus, this study researches the question: "Given a multivariate profile of a county, how many farmers markets would one expect in that county?" In other words, where might a new farmers market prosper?

The Data Analysis Process

A table of all 8,804 farmers markets registered with the USDA was obtained and joined with publicly available county level data covering a multitude of variables relevant to the study. The final data set included 3,114 rows (counties) and 121 variables, covering topics such as frequency of farmers markets, votes in recent presidential elections by party, income, jobs, weather, housing units, age distribution, ethnicity, and others. Fewer than 10% of rows contained observations that were imputed due to missingness.

A backwards eliminated zero inflated negative binomial model was built and fitted to the data. The final model was selected over other candidate models on the grounds that the AIC was

lowest, contained the fewest number of significant variables, and conformed to model assumptions. The chi squared component for each observation was calculated and counties were ranked in with respect to the counties that deviated most from what is expected under the assumptions of the model.

Findings

The results of this study are included in the attached spreadsheet "FarmersMarketPredictions.csv," which contains each county's observed and predicted number of farmers markets, along with the values of each variable collected for analysis. The counties are ranked in order of most underserved to most saturated. This data set gives the entrepreneur valuable insight into new potential markets and sheds light on existing markets.

Limitations

Since this is an observational study, there is no evidence of causation. While this model serves as a great baseline to identify underserved counties, the model may overestimate or underestimate some counties due to confounding variables. Candidate counties should be properly investigated. This model is based on the current trends. Thus, if there are any biases currently present that inhibit or accelerate farmers market growth, those biases will be captured by the model.

Proposed Actions

The "FarmersMarketPredictions.csv" table is meant as a starting point of research should an entrepreneur seek to start a farmers market without a location in mind. It is suggested s/he look through the spreadsheet, which is ranked in order of most underserved, to find a county that may present itself. S/he can go National Farmers Market Directory hosted by the USDA to find existing farmers markets in the area to determine a city/location. If an entrepreneur should already have a location in mind, s/he can find the county on the aforementioned spreadsheet to compare the observed and expected number of farmers markets in that county. Also, s/he can find other pertinent county level statistics that may assist in determining the viability of that county.

Expected Benefits

This study has identified numerous counties that are underserved by farmers markets under the current conditions. These communities, farmers and consumers alike, will likely benefit from more farmers markets. Furthermore, with the vast amount of county level variables collected, new models can easily be created that analyze different factors. For example, one can easily estimate the effect of political affiliation on farmers markets. Lastly, this methodology is flexible and can be used to model any other count variable at the county level. For example, one can model the number of restaurants by county of a large restaurant chain and identify potentially fruitful markets.

Sources

McCarthy, R. (2010). *Evaluating the Social, Financial and Human Capital Impacts of Farmers Markets:* Market Umbrella. 1:10