The Impact of Food Environment on Obesity Rates Across the United States

Tamara Frances, Beza Getahun, Dajuan Young, Marcos Garcia, Marcos Diaz

Highlights

We have identified the top 10 actionable variables that affect a county's obesity rate in the U.S., including:

- Average soda and milk prices
- Number of full service restaurants
- Number of grocery stores

Background

Many Americans do not meet dietary recommendations for healthy eating and suffer from chronic health issues related to obesity.

Research suggests that the ubiquity of fast food and junk food has detrimental impacts on community health and well-being, particularly as it relates to obesity rates.

By studying our national food environment we can find insights on how a community's obesity rate is impacted by environmental factors.

Data

Dataset

We used data from the USDA Economic Research Service's Food Environment Atlas in our analyses. This data set includes statistics on various food environment indicators, broken down by county using Federal Information Processing Standards (FIPS) codes.

Data Cleaning

Compiled 2010 Data	Identified Missing Data	Handled Null Values
Extracted and compiled variables with 2010 data in a new table. 2010 had the largest amount of available data.	Used isnull() to determine the data missing from our dataset.	Dropped and replaced null values in our 2010 dataset.

Obesity Rate x Region	Top Actionable Variable x Region
• South : 32.13%	• South: % households no car & low access to store
• Midwest: 31.02%	• Midwest: % households no car & low access to store
• Northeast: 27.62%	• Northeast: Recreation & fitness facilities/1,000 pop
• West: 25.62%	• West: Full-service restaurants

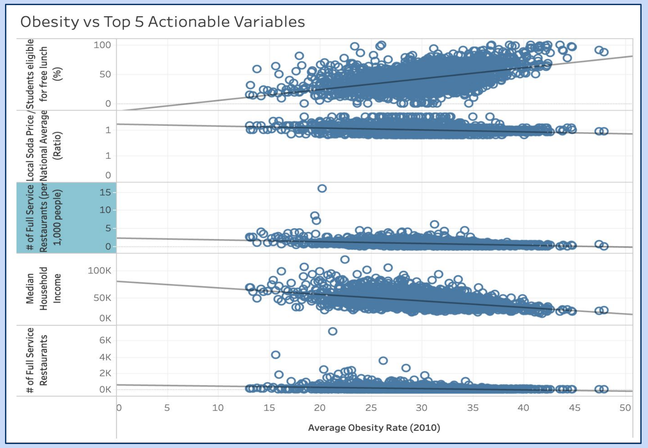
Model: Random Forest

After evaluating a variety of different machine learning models, we settled on a random forest model, which gave us the strongest accuracy at 85% in terms of predictive capabilities.

Another consideration that led to this decision was the feature selection capabilities within the same sklearn package. This allowed us to only keep the 95% most important contributing factors to reduce our variable input from 70 to 42 without reducing the accuracy of the model.

Model: Linear Regression

Using the subset of features we narrowed down, we ran a linear regression model to come up with the potential policy recommendations, based on the top variables' coefficients.



Results and Recommendations

Greatest Contributing Factors

Diabetes, race and age are the biggest factors correlated with a county's obesity rate.

The biggest *actionable* drivers are average milk and soda prices, the number of full service restaurants and number of grocery stores.

Recommendations

Policy intervention around subsidizing low-fat milk prices, taxing soda drinks, and prioritizing grocery stores and full service restaurants over fast-food and convenience store options would have the largest impact towards lowering a county's obesity rate and overall health.