CITADEL

Summer Invitational Datathon 2024

Tipping the Scales: Unraveling and Targeting Obesity Drivers in Low-Income Communities

Team 20 | Aditya Kamat, Austin Feng, Elijah Liu, Tanay Tandon

Introduction

Tanya grew up in Carroll County, Georgia, a low-income, low-food access county where fresh fruits and vegetables were a luxury she could not often afford. Her mother, a teen mom, constantly had to make tough decisions on the food she brought to the table. Fast food chains and convenience stores dominated the landscape so processed and high-fat foods were often the only affordable options. These eating habits inevitably led to health issues for her and her family, and a lifetime of worries for Tanya. [1]

Tanya's struggle is not just about food; it's about the systemic barriers that make it difficult for low-income families to maintain a healthy lifestyle. Economic insecurity, limited food access, and the lack of proper education all play significant roles in this ongoing battle. And its consequences are not trivial: in 2016, the total cost of obesity to the American healthcare system was over $260 billion, which will balloon to an average of over $400 billion from 2024-2033, in addition to an average of $260 billion in lost GDP per year in that timespan. On an individual level, people with obesity spend almost 3 times as much on health expenses than non-obese people, and for low-income individuals, this price can be devastating.

Unfortunately, Tanya's story is not unique. It reflects the harsh reality faced by 12.6 million Americans living at the intersection of obesity and poverty. [2, 3] This humbling reality was the catalyst for our study, driving us to investigate the systemic issues contributing to obesity among low-income populations.

Key Findings

Knowledge and Attitude Are Important

Our study revealed that individual factors like attitudes towards obesity are correlated with the dietary choices and health outcomes of low-income individuals. Lower-income groups are often less conscious of their own weight and have less accurate perceptions of their body size. This lack of awareness and understanding contributes to the persistence of unhealthy eating habits and obesity. Additionally, overweight and obese individuals in low-income groups are less likely to attempt weight loss and less successful when they do try, often due to unrealistic or unsustainable dieting methods.

Low-Income Leads to Low Accessibility

Accessibility to healthy food options is a critical factor in managing obesity among low-income populations. Our analysis showed that lower-income neighborhoods have greater access to fast food and sugary drinks, while healthy food options like low-fat milk are more expensive than in higher income areas. Furthermore, the proximity to fitness facilities also plays a role, with higher-income neighborhoods having better access to such amenities, thereby supporting healthier lifestyles.

Relative Importance of Factors

When weighing various factors against each other, our study found that accessibility variables have a much greater impact on obesity than knowledge and attitudes. Proximity to grocery stores, the availability of SNAP-approved retailers, and the cost of healthy foods are the most significant predictors of BMI among low-income individuals. While nutritional knowledge is important, practical barriers related to food access are far more impactful in determining obesity rates.

Policy Recommendations

Transform Food Deserts & Food Access

Invest in developing grocery stores and farmers' markets in food deserts, provide grants and loans to local entrepreneurs, expand SNAP benefits for farmers' markets and online grocery shopping, and implement zoning regulations to limit fast food restaurant density in low-income neighborhoods.

Empower Through Education

Implement comprehensive nutritional education programs in schools and community workshops, launch public health campaigns to raise awareness about healthy eating, and provide personalized support through counseling and coaching for weight management tailored to low-income individuals.

Literature Review

The current literature on the connection between obesity and socioeconomic status has predominantly focused on the relationship between low income and high obesity rates, identifying contributing factors such as the cost of healthy foods, educational attainment, and food accessibility. Our study aims to build upon this foundation by addressing gaps in the existing research.

Existing research has overwhelmingly clustered around the role of educational attainment in influencing dietary choices and obesity prevention. [6, 7] However, we believe it is just as important to analyze the attitudes, behaviors, and health related thought processes of people with obesity. Our study seeks to establish preliminary empirical examination into this field by investigating how dietary attitudes correlate with obesity rates.

Furthermore, although the concept of food deserts and their impact on obesity is well-known, the integration of food accessibility data with economic and educational factors is rare. [7. 8] We seek to use this comprehensive approach in our report to describe the various aspects of obesity in low-income communities. By analyzing the combined effect of economic constraints, educational barriers, and food accessibility, our study provides a better understanding of the systemic issues contributing to obesity.

In conclusion, our study distinguishes itself from existing work by offering a detailed analysis of an ensemble of factors contributing to obesity in low-income families. By weighing the relative importance of food costs, education, and accessibility, we aim to provide insights that can inform targeted interventions and policies.

Exploratory Data Analysis

Obesity is a multifaceted problem to tackle. Among the contributing factors, there are one's diet, exercise, income, education, and genetic factors. All of them interact with each other to complicate the overall picture. Therefore, we first verified the basics before diving deeper.

To get a macroscopic view of current dietary and physical activity trends, we used the Nutrient, Physical Activity, and Obesity dataset, and created a script to automatically plot and compare sample means to the queries stratified by demographic (in Income, Age, Race/Ethnicity, and Education, Gender). Across the board, we observed that adults in lower income groups exhibit more unhealthy behaviors across the board compared to high income groups, including lower consumption of fruits and vegetables and lower physical activity for all metrics. These results are almost perfectly monotonic across the 6 provided income groups and the vast majority of these results are statistically significant the provided confidence intervals.* As expected, lower income groups had higher obesity rates. Interestingly, higher income groups have higher proportions of overweight individuals.

Next, we hypothesized that increase in the obesity rate is mostly attributed to those of lower income. With an increasing income inequality, one would expect that a growing proportion of individuals would be exposed to socioenvironmental factors conducive to obesity, such as limited access to health-promoting resources and reduced educational opportunities. However, when plotting obesity rates over time, we observed that higher income groups had a higher increase in obesity than lower income groups from the study range of 2011-2022, with the highest income group increasing at a rate of double that of the low-income group. We conjecture that increasing screen time could have led to more disrupted and sedentary lifestyles, increasing obesity rates for white-collar groups, though investigating high income groups is beyond the scope of our report.

*We used answers from the most recent, non-pandemic year (2022 and 2019). The only non-monotonic result is fruit consumption.

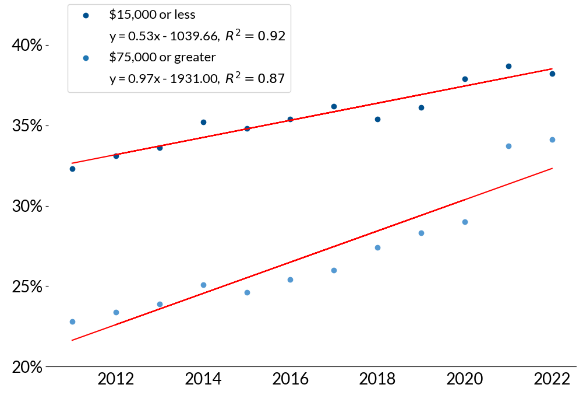

Figure 1: Obesity Rate Increase Over Time

Hoping to see if accessibility/convenience influences obesity, we investigated fast food chain in low-income areas. We chose McDonald's as our representative fast-food chain due to its status as a top 5 chain and its ubiquitous presence across the country. By plotting McDonald's locations on a map of the USA and overlaying neighborhood income data, we sought to understand which types of neighborhoods McDonald's is most prevalent in.

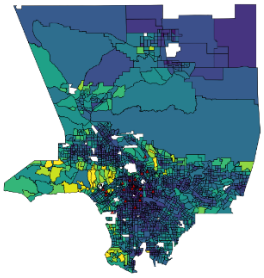

Figure 2: McDonalds Locations Across L.A, (marked in red) and Median Neighborhood Income (lighter means higher income)

From this, and more zoomed in maps (such as with L.A. above), we observe that McDonald's locations are disproportionately prevalent in lower-income neighborhoods. This pattern suggests that fast food options like McDonald's might be more accessible to individuals in these areas, potentially contributing to higher rates of fast food consumption and, consequently, higher obesity rates.

Motivation

To establish effective policy, we believe that it is imperative to better understand low-income overweight and obese individuals. This understanding is crucial for identifying barriers to healthy weight management among low-income groups and subsequently developing targeted interventions to break down these barriers.

We approach this through four angles:

1. Weight-consciousness: Are participants informed on their own weight?
2. Knowledge: How much do participants know about nutrition and health?
3. Attitude: How do participants generally view their health and weight? Do overweight participants view themselves as overweight?
4. Behavior: For those who are trying to lose weight, what methods are they using and are they successful?

Datasets and Cleaning

We to use the National Health and Nutrition Examination Survey (NHANES) dataset as it provided health, dietary, and demographic data on around 15,000 individuals. In particular, we used the 2017-March 2020 Pre-pandemic data, to establish recency while avoiding data biased by the pandemic/lockdowns. We removed data from children, as their dietary needs can be different from those of adults and could skew results. We also removed pregnant women as they are often classified as obese by BMI simply because of pregnancy and not because of any demographic or behavioral reasons.

To analyze difference in incomes, NHANES provides data via the metric Income of Family/Poverty Guidelines. We will call this the poverty ratio for simplicity. This translates to around $15,000 for a single person and $30,000 for a family of four. We classified low income as between 0 to 2 for this metric to encompass both low-income survey participants as well (Pew Research classifies low income as $30,000/single person and $60,000/family of four). We will classify high income as over 2 for this metric. This classification is mainly used to succinctly define low and high-income groups so as to avoid cluttered statistics, tables, and graphs.

Using BMI data provided, we classified participants as underweight, healthy, overweight, and obese. For analyses, we mainly look at participants that are overweight and obese as they are the target population for our proposed policies.

General Statistical Methodology

We use both questionnaire and laboratory data to analyze the above angles. Questionnaire data provides us with crucial insights on a participant's perspectives but may contain biases. Therefore, we also use laboratory data, when available, to validate questionnaire responses and provide concrete results.

To analyze differential behavior between low- and high-income participants, we calculated sample means of survey responses using one-sided z-tests, as we have a large sample size, with $\alpha = 10^{-4}$. As NHANES oversamples datasets, all our sample means are weighted by provided sample weights (to get results representative to the U.S. population). We also use effective sample size to remove additional variance from weighting and achieve a similar level of precision to a simple random sample.

Results: Weight-Consciousness

We use two main metrics to assess weight-consciousness. First, participants in NHANES were asked to classify their weight into three classes (e.g. overweight). By comparing a participant's actual weight class to their own reported class, we were able to find the proportion of participants who misclassified their weight. On average, 30% of low-income participants misclassified their weights compared to 23.26% of high-income participants, which is significantly lower as verified by our z-test (from now on we will omit references to our z-tests for succinctness).

Second, we compared a participant's self-reported weight (via questionnaire) and their measured weight. We use percent error so as to not penalize heavier participants.* One's estimation of their own weight may have internal biases, chiefly the bias to weigh less, so we consider both the absolute percent error and raw percent error. For the former, we get that on average, we have a percent error of 4.14% and 2.62% for low-income and high-income participants respectively. For the latter, we get -0.75% and -0.84% for low-income and high-income participants respectively. So, while some of the weight difference can likely be internal bias, there is still much error that can be attributed to a lack of consciousness on one's weight.

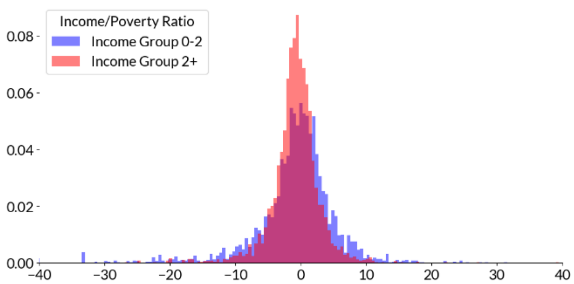

Figure 3: Raw Percent Error of Reported Weight On Questionnaire

Overall, this section suggests that lower income groups are less cognizant of their own weight, whether it be due to limited access to regular healthcare check-ups or sociocultural norms that influence body image perception. Promoting awareness can be effective in combatting obesity as it enables individuals to accurately assess their weight, potentially leading to appropriate lifestyle modifications.

Results: Knowledge

To assess the general knowledge of participants, we use two approaches. First, one of the NHANES questions quizzes participants on nutritional knowledge.* High-income participants answered correctly 70.23%, compared to a distinctly lower 49.92% for lower income groups.

Secondly, we use the Foods at Home dataset to observe participation of low-income families (defined as income of 30k per adult) in activities that can increase nutrition knowledge. Only 19.06% percent of low-income families searched the internet for nutrition information in the previous two months compared to 28.66% for high income families (95% confidence). This pattern is also consistent when considering government programs like MyPlate and MyPyramid. We also observed that low-income families are less aware of government nutrition programs. For example, only 17.59% of low-income families had previously heard of MyPlate compared to 24.98% for high-income families.

These results suggest that lower-income participants are generally less knowledgeable about nutrition and are less exposed to resources that can enhance knowledge. Therefore, education programs and effective promotion of these programs may help increase general nutrition knowledge, allowing low-income groups to establish a healthy diet.

Results: Attitude

Regarding attitude, we consider both attitude towards weight and healthy eating. For this section and the following one, we limit our scope to just overweight and obese participants as attitude and behavior of these groups are more relevant than considering all groups.

	Income/Poverty Ratio		
Query	0-2	2+	Difference
Consider Themselves Overweight	67.59%	76.28%	8.69%
Believe They are at the Right Weight	31.28%	23.05%	-8.23%
Would Like to Weigh Less	77.88%	85.59%	7.71%
Tried to Lose Weight in Past Year	46.59%	56.60%	10.01%
Currently on a Weight Loss Diet	230	381	151
Have Never Intentionally Lost 10+ lbs	28.85%	45.87%	17.02%

One-sided z-tests, $\alpha = 10^{-4}$

Figure 4: Attitudes of Overweight/Obese Participants Towards Weight Loss by Income

Using individual data from NHANES, we first observe that low-income overweight participants tend to be more satisfied with their own weight; they are likelier to view themselves as at the "right weight." Crucially, these participants are more averse to weight loss. The increasing differences between the 3rd, 4th, and 6th rows in Figure 4 paint a clear picture. Relative to high-income individuals, fewer overweight low-income individuals have motivations to undergo weight loss (3rd). Even fewer will act on these motivations (4th) and even fewer will be successful in losing weight (5th).

*A 10-pound difference in weight is much more significant if the participant is 100 pounds than if the participant is 400 pounds. Admittedly, this approach is imperfect as there are likelier more nuances in numerical estimation, but it still serves as a good proxy.

*The question is "What does 10% Daily Value for Vitamin D mean to you?" and participants were offered three answer choices

Looking at Figure 4 again, there is an interesting discrepancy between rows 4 and 5 as the ratio of high-income to low-income participants currently on a weight loss ($381/230 = 1.66$) is much higher than the ratio of participants who attempted to lose weight in the past year ($56.60/46.59 = 1.21$). Such a difference suggests either low-income overweight participants have internal biases (inflate row 4) or tend to have shorter/less frequent dieting periods. We believe this is the latter and we will unpack this in the Behavior section.

When analyzing household data from Foods at Home, we again observe that lower-income families have attitudes that are counter to healthy eating. From Figure 5, overweight families believe are likelier that healthy foods taste bad (with 95% confidence). By sample mean, lower-income obese families are likelier to believe that they eat healthy food (though this does not past a benchmark of 95%).

Income	0-30k	30k+
Healthy Foods Don't Taste Good	16.67%	12.24%
Family Already Eats Healthy Food	37.72%	33.20%

Figure 5: Attitudes of Overweight/Obese Families Towards Healthy Foods by Income

In sum, we see that overweight lower income participants tend to have attitudes that are detrimental to combatting obesity and establishing a healthy diet. Adjusting these attitudes will be instrumental in fostering healthier lifestyles and reducing obesity rates in this. This opens the door to more questions on the origins of these attitudes. Why are low-income participants mores satisfied with their weight? Why do those who want to weigh less take less action? Why do low-income participants seem to be less effective at weight loss? Answering these questions may be fundamental to effective policies as it allows us to tackle root causes. For this last question, we examine participant behavior.

Results: Behavior

To understand the efficacy of weight loss, we must look at the methods and habits employed. As all relevant NHANES data are in counts, we normalize by the number of people who are trying to lose weight so we can cross compare a low-income and a high-income individual trying to lose weight (the proportion/number of people trying to lose weight no longer matters). Of the listed methods, we remove all methods with lower sample size, $n < 1000$ (these are mainly niche methods).

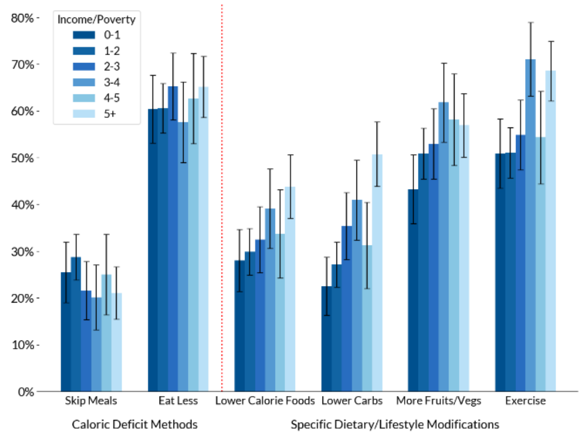

Figure 6: Percent of Participants Using Select Weight Loss Methods by Income

In Figure 6, we can see the differences in methods by income level. We classify methods into two groups: caloric deficit methods and specific dietary/lifestyle modifications. The former are more direct approaches that optimize for caloric deficit which is fundamental to weight loss. The latter are more specific methods that may not directly lead to weight loss but are auxiliary methods that also boost health. Using 95% confidence intervals, there is no statistical difference in use of caloric deficit methods. However, we see that increasing income generally leads to increased use of specific dietary/lifestyle modifications.

These methods may not directly account for the difference in efficacy but certainly can contribute. For example, eating lower calorie foods (in density) generally makes dieting easier as it allows participants to consume higher volumes of food, which lead to increased feelings of fullness. Lower carbohydrate diets can lead to reduced insulin levels, which may promote fat oxidation and decrease fat storage (Hall et al., 2015). Additionally, ketones produced during carbohydrate restriction may have appetite-suppressing effects (Gibson et al., 2015), among other benefits.

To concretely observe participant behavior, we use dietary data from NHANES. The dietary data covers all nutrient consumption in a day for two separate days. While there are few data points per individuals, with the large sample size, this provide a representative estimate of daily dietary data and is likely more accurate than questionnaire data.

Income/Poverty Ratio	0-2		2+	
Gender	Male	Female	Male	Female
Avg Caloric Deficit (kcal)	1065	1174	732	796
% with Unhealthy Diet	53.30%	52.68%	34.89%	40.69%
Avg Deficit (Unhealthy Diet)	1651	1893	1552	1583

Figure 7: Caloric Intake for Dieting Overweight Participants by Income and Gender

As obesity is most related to calories consumed, we calculate the caloric deficit. This is defined as calories consumed subtracted by the number of calories needed to maintain the participant's current body weight. This latter number can be estimated by multiplying the weight of the participant by 15. Interestingly, from Figure 7, we see that on average, lower-income participants seem to take larger caloric deficits than higher-income participants, with a difference of around 330-370 kcal.

From this, one could expect lower income participants to be more effective at a dieting. However, we see that much of this difference, is actually detrimental to weight loss as it comes from unhealthy and unsustainable deficits. While the numbers vary slightly between sources, we define an unhealthy diet as consuming <800 kcal and undergoing a deficit of over 1000 kcal. For both male and female low-income participants, 53% had an unhealthy diet compared to 34.89% and 40.69% respectively for male and female high-income participants. To compound this, the average caloric deficit for those undertaking an unhealthy diet among low-income participants was significantly higher than that of high-income participants; for female participants, this number is 300 kcal more!

While unsustainable methods can seem effective, they often lead to quick relapses due to their grueling nature and are extremely damaging to a dieter's health. So, it is no surprise that lower-income participants have more difficulties losing weight. Also, the shorter diets corroborate with the discrepancy we saw in the Attitude section.

While analyzing the dieting methods utilized by low-income groups, it's clear that a significant barrier to weight loss is a lack of knowledge on dieting methods and how to effectively execute them. The lack of specificity in these methods also suggest that low accessibility can also block low-income groups from more nuanced/targeted methods that are potentially more effective. Thus, we will include in our policy, approaches to tackle these dual barriers.

Key Findings

From our analyses we have three main conclusions.

Low Income, Low Health Cognizance

We found that low-income individuals are **less aware of their weight status** as they are likelier to misclassify their weight and have larger errors in self-reported weight. Furthermore, low-income individuals **demonstrate less nutritional knowledge and are less likely to engage with nutrition information sources**. This knowledge gap suggests could likely come from disparities in nutrition education in low-income communities as well as overall decreased priority on health.

Discrepancies in Attitude Hinder Health

Overweight and obese low-income individuals tend to be **more satisfied with their current weight** compared to their high-income counterparts. They are also **less motivated to lose weight** and less likely to act on weight loss intentions. Additionally, low-income families are more likely to believe that healthy foods taste bad.

We believe that these differing underlying perceptions comes from a decreased social stigma on obesity. With a higher proportion of overweight/obese individuals in low-income groups, a heavier weight can be considered more of the normal. Conversely, less people try to lose weight and with lower exposure to peers undergoing weight loss, low-income individuals don't receive as much positive social influence regarding nutrition and health.

Difference in Diets Limit Weight Loss

While low-income individuals attempting weight loss use similar caloric deficit methods as high-income groups, they are less likely to employ **specific dietary or lifestyle modifications**. Moreover, low-income dieters tend to adopt more **extreme and potentially unhealthy caloric deficits,** which may lead to unsustainable weight loss attempts and frequent relapses.

We attribute difference in dieting methods to a lack in knowledge on effective dieting practices as well as limited resources (such as cost barriers with gym, healthy eating).

Motivation

The average American lives just over three miles away from a McDonalds. Although McDonalds is a very prominent fast-food chain that dots America's landscape, there are plenty more, with advertisements for each one vying for the attention and spare change of the hardworking American. Beyond the previously investigated individual-level factors such as knowledge, attitudes, and behaviors, environmental elements can play a significant role in obesity.

With strong correlations demonstrated between income and obesity levels as our prior analysis showed, we propose that there exists easier access to fast food and sugary drinks in low-income neighborhoods which can sufficiently influence the obesity rate of a county.

As such, we seek to build off descriptive evidence from the US Department of Agriculture's Economic Research Service in providing the first empirical evidence in support of how this accessibility disparity contributes to higher obesity rates and poorer health outcomes in low-income areas.

Dataset & Statistical Methodology

As described anecdotally, processed foods, particularly fast food and sugary drinks, seem to be associated with increased obesity and related health risks. This should suggest that lower-income neighborhoods, where access to these unhealthy options is more prevalent and cheaper, face a heightened risk of obesity and its associated health complications.

To substantiate our hypothesis regarding the accessibility of fats and sugars in relation to income levels, we first constructed a detailed dataset integrating information on county-level median household income, relative price of soda and low-fat milk to national average prices, and the number of fast-food restaurants, grocery stores and fitness facilities.

To analyze economic disparities, we categorized counties into income bins starting from $25,000: lower-middle-class ($25,000 to $50,000), middle-class ($50,000 to $75,000), upper-middle-class ($75,000 to $90,000), affluent ($90,000 to $105,000), and very affluent (over $105,000). This classification aids in effectively comparing economic differences across regions.

In this analysis, we used the ANOVA test to evaluate differences in the means of multiple groups simultaneously. This approach is scientifically advantageous because it allows us to assess variations between different income bins without inflating the risk of Type I errors, which can occur with multiple pairwise comparisons. By using ANOVA, we effectively determine whether there are significant differences in variables like fast food restaurants per capita, fitness facilities per capita, and prices of soda and milk across these income categories.

Results: ANOVA

Using ANOVA with $\alpha = 0.05$, we observe higher frequencies of fast-food stores in low-income areas and lower frequencies of fitness centers, environmental conditions that aren't conducive to health. Interestingly, our analysis revealed that you are 15% more likely to find fast food in a lower-income county compared to other counties.

Using ANOVA with $\alpha = 0.05$, from Figure 8 on the next page, milk price, and milk-to-soda ratio show a negative correlation with median household income, while soda prices exhibit a positive correlation with median income. These price discrepancies present barriers to healthy dietary options, while pushing low-income groups to consume cheaper, unhealthy options.

The unexpected finding is that, contrary to the assumption that both soda and milk would be cheaper in low-income regions (to match affordability), soda is indeed less expensive, while milk is notably more costly. This counterintuitive result highlights a significant disparity in the affordability of these items across income levels.

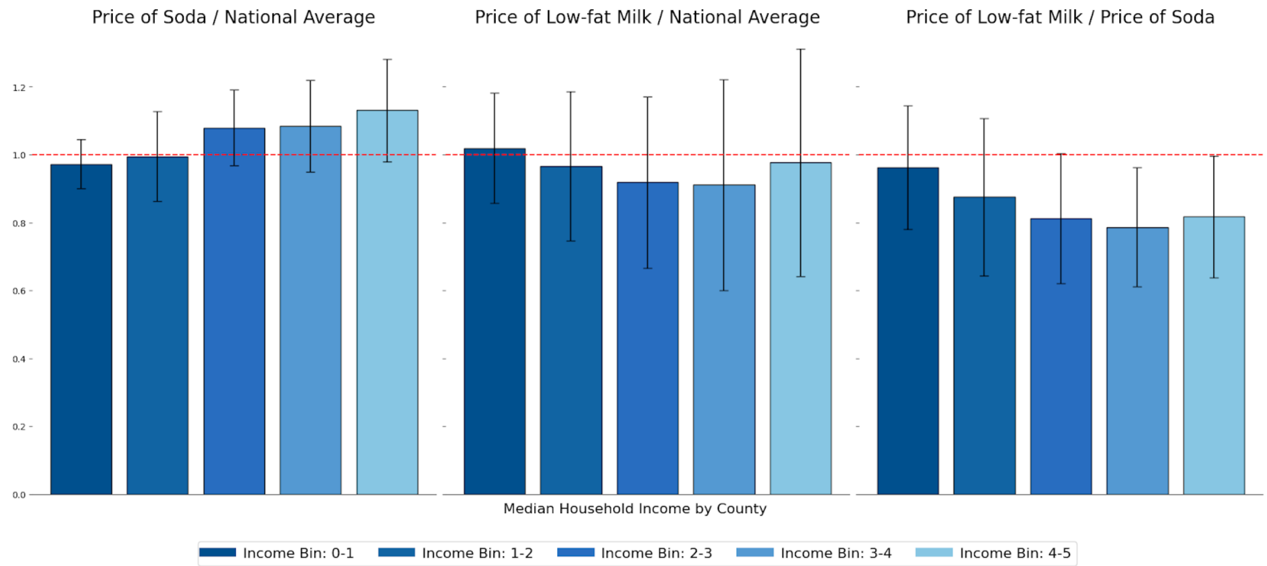

Figure 8: Ratios Between Price of Soda, Low-Fat Milk, and National Averages by Income

Statistical Methodology

Next, we use a Random Forest model to assess obesity rates in counties. In our data cleaning process, we excluded Alaska due to its unique geographic and economic conditions. Alaska's low population density and high transportation costs lead to atypical data that could skew our analysis. By removing this outlier, we ensured a more representative dataset of the contiguous United States, which better reflects the general patterns and relationships we aimed to study.

For feature engineering, we normalized fast food availability, fitness facilities, and grocery stores by population and scaled the data for uniformity. We then used grid search to optimize model parameters, improving predictive accuracy through systematic hyperparameter tuning.

To evaluate the performance of our model, we conducted cross-fold validation with five folds. This method involves partitioning the data into five subsets, training the model on four subsets, and testing it on the remaining subset.

Results: ANOVA

Our model achieved an R-squared value of 0.91 on the test data, indicating a high level of accuracy in predicting obesity rates based on the accessibility factors.

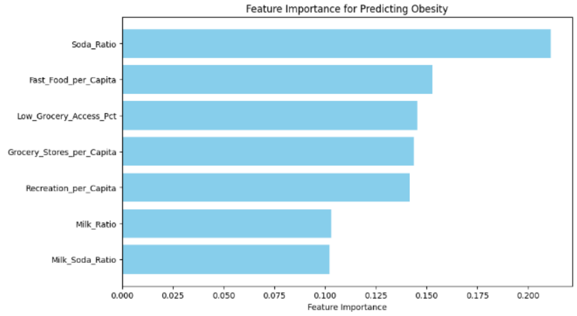

Figure 9: Random Forest Feature Regression

By analyzing the top features in our regression, we see that ratio between of a county's price of soda to the national average is the top feature, with other accessibility features ranking slightly below. This highlights the power of availability and convenience in obesity rates.

Note: Here, Soda_Ratio represents the price of soda relative to the national average, Fast_Food_per_Capita indicates the number of fast food outlets per capita, Low_Grocery reflects the percentage of the county population with low access to grocery stores, Recreation_per_Capita denotes the number of fitness facilities per capita, Milk_Ratio compares the price of low-fat milk to the national average, and Milk_Soda_Ratio compares the price of low-fat milk to that of soda.

Key Findings

From dual analysis conducted, we were able to derive two main insights.

Disparity In Access of a Healthy Lifestyle

Eating healthy is often more challenging in these lower-income areas due to the higher cost of nutritious food options like milk, coupled with the easier availability and lower cost of unhealthy alternatives like soda and fast food.

Predictive Power of Accessibility Factors

Based on our random forest regression analysis, we found that the price **of soda / national average, fast food per capita and percentage of population with low access to grocery stores are the three most important factors in predicting obesity of a county.**

Interestingly, other factors such as the number of fitness facilities per capita and the comparative prices of low-fat milk did not show as much influence in our model. This underscores **the critical role of the immediate food environment and economic accessibility** in shaping obesity trends across different regions.

Motivation

We had now seen that individual factors like knowledge and macroscopic factors were both vital in understanding the intersection of obesity and poverty so our natural next step was to synthesize and contextualize these results.

Statistical Methodology

To investigate the relative impact of accessibility and knowledge on obesity, we employed an analytical framework combining regression and relative importance analysis. We used a synthetic dataset based on the USDA FoodAPS National Household Food Acquisition and Purchase Survey due to its nationally representative household level granularity. This granularity then enabled us to merge individual BMI and nutritional knowledgeability with household specific food access information, as opposed to county specific food access like previous research.

In particular, we aimed to check, and quantify, the relative predictive power of various food access metrics such as nearest grocery store, compared to knowledge metrics such as knowledge of the food pyramid.

The variables used in the regression were split into three main groups incorporating variables we looked at in previous analyses: health knowledge/ attitude, accessibility, general food security and health habits.

The methodological steps were as follows:

1. Data Preprocessing: For knowledge and general variables, we assumed non-awareness for skipped questions. For accessibility variables, the dataset had 0 missing values. Households where BMI was not reported were dropped. Like the NHANES dataset, data from pregnant women and children was dropped. Finally, the data was filtered to include only low-income households so as to make the findings applicable.

2. Standardization: Applied standard scaling to ensure comparability across variables.

3. Regression Analysis: Conducted regression to explore the relationship between the variables and average BMI, with standardized coefficients for clear interpretation.

4. Permutation Importance: Used permutation importance with our regression model to quantify each variable's impact on BMI.

Results: Regression & Importance

Factor	Importance
Knowledge/Attitude	5.98%
Accessibility	82.96%
Food Security & Habits	13.87%

Figure 10: Relative Importance of Variables

Our analysis reveals that accessibility variables have a markedly greater influence on BMI compared to other variables, with an importance score nearly 6 times higher for food security and general variables, and 13 times higher for knowledge variables. Some of this can be due to the fact that we have more features associated with accessibility (number of feature is equal to the number of features for the remaining two groups).

In summary we have:

1. Proximity to Resources: The availability and proximity of stores, as well as participation in food assistance programs like SNAP, play a crucial role in BMI management among low-income populations. For example, the distance to grocery stores and the number of food retailers within varying distances were notable factors.
2. Impact of Food Security: Variables related to food security, such as perceptions of the cost of and time required for healthy eating, are also significant though to a lesser extent.
3. Knowledge vs. Practical Barriers: While nutrition knowledge (e.g., awareness of MyPlate and MyPyramid, reading nutrition labels) is important, our study

shows that practical barriers related to accessibility are far more impactful.

Development of longitudinal studies tracking the long-term effects of improved accessibility on BMI and overall health can help improve upon this analysis.

Transform Food Deserts & Food Access

Our study underscores the critical role of food accessibility in managing obesity rates among low-income populations, particularly in food deserts. To address this, we recommend investing in the development and maintenance of grocery stores and farmers' markets in these underserved areas. Providing grants and low-interest loans to local entrepreneurs can help establish these businesses. Additionally, expanding SNAP benefits to be usable at farmers' markets and for online grocery shopping will further enhance access to nutritious foods. Implementing zoning regulations to limit the density of fast food restaurants in low-income neighborhoods can also encourage healthier food environments and reduce obesity rates.

Empower Through Education

The findings indicate that lower-income groups often lack nutritional knowledge and have unhealthy attitudes towards weight and diet. Implementing comprehensive nutritional education programs in schools and community workshops can raise awareness about healthy eating. Public health campaigns through social media and local media outlets can further disseminate information on the benefits of a balanced diet. Personalized support through counseling and coaching for weight management, tailored to the needs of low-income individuals, can foster sustainable behavioral changes and help reduce obesity rates.

CITADEL

1. National Institute of Diabetes and Digestive and Kidney Diseases. (2023, May 1). Health risks of overweight & obesity. https://www.niddk.nih.gov/health-information/weight-management/adult-overweight-obesity/health-risks
2. In Your Own Words Reflections on Living with Obesity. (n.d.). Obesity Matters. https://obesity-matters.com/wp-content/uploads/2023/01/In-Your-Own-Words-Reflections-on-Living-with-Obesity_EN-Digital_FINAL-2.pdf
3. Cawley, J., Biener, A., Meyerhoefer, C., Ding, Y., Zvenyach, T., Smolarz, B. G., & Ramasamy, A. (2021). Direct medical costs of obesity in the United States and the most populous states. Journal of managed care & specialty pharmacy, 27(3), 354–366. https://doi.org/10.18553/jmcp.2021.20410
4. Schweikert, D. (2023, June 27). Chapter 3, the Social Cost of Obesity of the 2023 Report of the Joint Economic Committee Congress of the United States on the 2023 Economic Report of the President. United States Joint Economic Committee. https://www.jec.senate.gov/public/_cache/files/d1774be7-8a29-4ecc-996d-3bda421310f3/2023-erp-ch3-republican-response.pdf
5. Telesford, I., Schwartz, H., Claxton, G., & Cox, C. (2023, July 6). How have costs associated with obesity changed over time? Peterson-KFF Health System Tracker. https://www.healthsystemtracker.org/chart-collection/how-have-costs-associated-with-obesity-changed-over-time/
6. US Census Bureau. (2023, September 12). Poverty in the United States: 2022. Census.gov. https://www.census.gov/library/publications/2023/demo/p60-280.html
7. Rehkemper, G., & Ridmann, C. (2019, May 3). *Gavinr/usa-McDonalds-locations: USA McDonalds locations*. GitHub. https://github.com/gavinr/usa-mcdonalds-locations
8. Office of the Assistant Secretary for Planning and Evaluation. (2024). Poverty guidelines. ASPE. Retrieved August 3, 2024, from https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines
9. US Department of Agriculture. (2023, May 24). FoodAPS national household food acquisition and purchase survey. USDA ERS. https://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey.aspx
10. Edmisten, M. (2024, March 29). Discover The U.S. Town Furthest Away From Its Closest McDonalds. 247 Wall St. https://247wallst.com/food/2024/03/29/discover-the-u-s-town-furthest-away-from-its-closest-mcdonalds
11. Gallup. (2022). A Global Analysis of Cooking Around the World. https://www.gallup.com/analytics/512897/global-cooking-research.aspx