INFO-I 590: Data Visualization

Fast Food Chains: Mapping Geographic Hotspots and Assessing Their Nutritional Impact on Public Health

Team Members:

Group No. 11 –

1. Vatsal Vinay Parikh
2. Rohan Darji
3. Harsh Chauhan

Abstract

This project explores data visualization techniques to analyse and present the impact of fast-food chains on public health across the United States. Through interactive visualizations including scatter plots, tree maps, histograms, violin plots, box plots and geographic mapping, we examine over 10,000 fast food locations and their nutritional data. The study employs various visualization methods to reveal concerning patterns: scatter plots demonstrate that many menu items exceed recommended daily limits for sodium (2,000mg) and sugar (25g for women, 36g for men), while tree maps showcase McDonald's market dominance and its proportional contribution to overall caloric intake. Geographic visualizations using Folium highlight the concentration of fast-food outlets in urban areas, revealing that regions with high restaurant density show 25% higher obesity rates. With the U.S. fast food market reaching $142.55 billion in 2024 and 37% of Americans consuming fast food daily, these visualizations provide crucial insights into the relationship between fast food accessibility, nutritional content, and public health outcomes, particularly as diet-related diseases account for 697,000 annual deaths in the United States.

Introduction

1.1 Motivation

The widespread presence of fast food in American society presents both a public health challenge and an opportunity for data-driven insights. Traditional analysis methods often fail to capture the complex relationships between fast food consumption, nutrition, and community health outcomes. Through innovative data visualization techniques, we can better understand these relationships and their implications for public health.

Scale of Impact –

The fast-food industry's influence on American society is staggering:

- The market has reached $142.55 billion in 2024, with over 200,000 outlets nationwide
- Daily consumption affects 50 million Americans (37% of the population)
- A typical fast-food meal contains 1,200-2,000 calories, representing 60-100% of recommended daily caloric intake
- The average American spends $1,200 annually on fast food.

Health Crisis Evidence –

Research demonstrates alarming health trends:

- CDC data shows obesity rates reaching 42% in 2021
- Over 90% of fast-food meals exceed daily sodium recommendations
- Regions with high fast-food density show 25% higher obesity rates
- Annual diet-related deaths have reached 697,000 in the U.S.
- Healthcare costs related to obesity exceed $173 billion annually

Consumer Awareness Gap –

Despite increasing health consciousness, there remains a significant gap between perceived and actual nutritional content of fast-food items. While calorie information is now mandatory on menus, other critical nutrients like sodium and sugar are less visible to consumers. Interactive visualizations can bridge this knowledge gap, making complex nutritional data more accessible and understandable.

The power of visualization lies in its ability to reveal patterns that might otherwise remain hidden in traditional statistical analyses, making this project crucial for addressing one of America's most pressing public health challenges.

1.2 Background

The fast-food industry has a significant impact on public health in the United States. With a market size of $142.55 billion in 2024 and 37% of Americans consuming fast food daily, understanding its nutritional impact is crucial. Previous studies and visualizations have attempted to analyse this relationship, providing a foundation for our project.

Existing Visualizations –

Fig 1.2.1 Top 50 Fast Food Chains in the U.S. by Number of Stores (2021)

This visualization effectively displays the scale of major fast food chains using brand logos and color-coded categories. While visually appealing and easy to interpret, it lacks crucial geographic distribution data and nutritional context. Our project aims to enhance this type of visualization by incorporating geographic overlays and nutritional facts to better align with our public health focus.

Nutritional Content Analysis Graphs –

These analytical graphs effectively demonstrate nutritional content across fast food chains:

- Simple bar charts highlight key health indicators like sodium and sugar content between different chains.
- The visualizations clearly show which food items exceed daily recommended nutrient intake levels.
- Comparative trends between menu items across different chains are easily visible.

While these graphs provide valuable nutritional insights, they could be enhanced by:

- Adding interactive elements to explore specific menu items
- Including temporal trends to show changes in nutritional content over time
- Incorporating serving size context alongside nutritional values
- Connecting the nutritional data with geographic distribution patterns

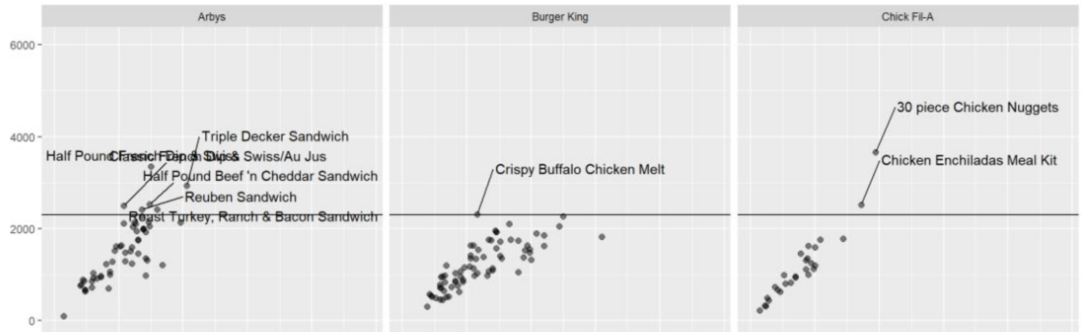

Fig 1.2.2 Scatter Plot of Food Items containing higher calories than Daily Limit

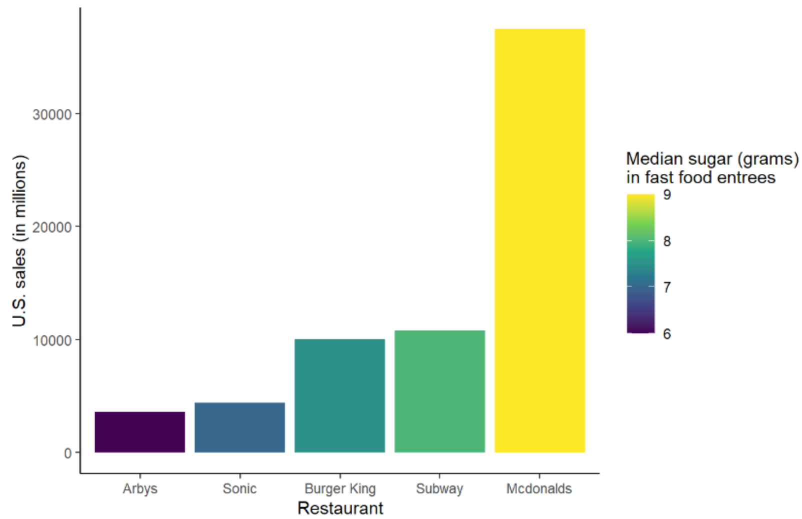

Fig 1.2.3 Bar Chart of Restaurants and the Median Sugar in Fast Food entrees

1.3 Objectives

Our project aims to investigate the relationship between fast food availability and public health through data visualization. The primary objectives are:

Geographic Analysis –

- Map and identify fast food chain hotspots across the United States using interactive visualizations.
- Compare fast food prevalence between urban and rural areas.
- Visualize the density of different chains across regions using choropleth and bubble maps.

Nutritional Investigation –

- Analyse and visualize nutritional profiles of menu items across major chains.
- Compare sodium and sugar content against recommended daily limits.
- Identify patterns in caloric content and their correlation with other nutrients.
- Create interactive plots to highlight nutritional trends.

Health Impact Assessment –

- Correlate fast food chain density with public health outcomes
- Visualize the relationship between restaurant concentration and obesity rates.
- Identify potential high-risk areas through geographic clustering analysis.

Pattern Recognition –

- Uncover trends in fast food distribution using cluster analysis visualization.
- Compare nutritional variations across different chains using violin and box plots.

Through these objectives, we expect to reveal clear patterns between fast food availability and public health outcomes, while providing insights that could inform policy decisions and public health interventions.

Process

2.1 Analysis of Data

Dataset Overview -

1. Fast Food Restaurants Across America

- Source: Datafiniti's Business Database
- Contains over 10,000 restaurant entries
- Key attributes:
    - Restaurant Name (String)
    - Address, City, State (String)
    - Latitude and Longitude coordinates (Float)
    - Categories of food offered (String)
    - Postal Code (Integer)

2. Fast Food Nutrition Dataset

- Contains data from six major chains: McDonald's, Burger King, Wendy's, KFC, Taco Bell, and Pizza Hut
- 1,072 unique menu items
- Nutritional attributes include:
    - Calories and Calories from Fat (Integer)
    - Total Fat, Saturated Fat, Trans Fat (Float)
    - Cholesterol (Float)
    - Sodium (Float)
    - Carbohydrates (Float)
    - Fiber (Float)
    - Sugars (Float)
    - Protein (Float)
    - Weight Watchers Points (Float)

Data Preprocessing Steps –

- Cleaned column names by removing unwanted characters and spaces
- Converted nutritional values to numeric format
- Handled missing values in key columns
- Removed duplicate entries
- Standardized units across nutritional measurements

Initial Data Exploration –

- Analyze distribution of restaurants across chains
- Examine the range and distribution of nutritional values

- Identify outliers in nutritional content

Nutritional Analysis Focus –

- Calculate average nutritional values per chain
- Identify menu items exceeding recommended daily limits for key nutrients
- Analyze correlation between different nutritional attributes (e.g., calories vs. sodium)

Geographic Analysis –

- Aggregate restaurant counts by geographic regions (state, city)
- Calculate restaurant density in different areas
- Identify areas with high concentration of fast-food outlets

This data analysis process forms the foundation for understanding the datasets and preparing them for subsequent visualization and interpretation.

2.2 Candidate Visualization Methods

Our project employs several visualization techniques to effectively analyse and present fast food nutritional data and geographic distribution:

Geographic Visualization –

- Choropleth Maps: Colour gradients show fast food density across regions, highlighting geographic disparities in restaurant distribution
- Interactive Maps: Using Folium to create dynamic maps with:
    - Marker clusters for efficient display of 10,000+ locations
    - Color-coded markers for different chains
    - Popup information showing restaurant details
    - Layer controls for toggling between chains

Nutritional Analysis Visualization –

- Scatter Plots: Display relationships between:
    - Calories vs. Sodium content
    - Calories vs. Sugar content
    - With reference lines for recommended daily limits
- Box and Whisker Plots: Show distribution of nutritional values across chains
- Tree Maps: Represent hierarchical data showing:

- Market share of different chains
- Proportional representation of nutritional metrics

Distribution Analysis –

- Histograms: Show frequency distribution of nutritional metrics
- Pie Charts: Display market share and proportional representation
- Violin Plots: Demonstrate the distribution and density of nutritional metrics across chains

These visualization methods were chosen for their ability to effectively communicate complex relationships between geographic distribution, nutritional content, and public health implications of fast-food consumption.

2.3 Failed Experiments (Why certain methods did not work)

During our analysis, several visualization methods proved challenging or ineffective, leading to important refinements in our approach. Initial attempts at simple dot maps for all 10,000 restaurant locations created overcrowded, unreadable visualizations, which we solved by implementing marker clustering in Folium.

Basic bar charts failed to capture the complex relationships between nutritional metrics, prompting us to adopt scatter plots with reference lines for daily recommended limits. Similarly, pie charts showing all nutritional metrics became cluttered and difficult to interpret, leading us to use tree maps that better represented hierarchical relationships between chains and their nutritional profiles.

Static visualizations proved inadequate for capturing the full complexity of the data, which led to the implementation of interactive features allowing users to toggle between chains, zoom into specific regions, and access detailed information through hover functionality.

These experimental failures and subsequent solutions significantly shaped our final visualization choices, resulting in more effective and insightful data representations.

Results and Insights

Nutritional Content Analysis Visualizations for Top Six Fast-Food Companies -

Our initial analysis focuses on understanding the distribution of fast-food chains and their nutritional profiles through four key visualizations.

Company Distribution Analysis:

The frequency distribution and caloric contribution charts reveal McDonald's dominance in the market, representing 28.8% of the dataset, followed by KFC (19.1%) and Taco Bell (16%). This distribution helps identify which chains have the most significant potential impact on public health through their market presence.

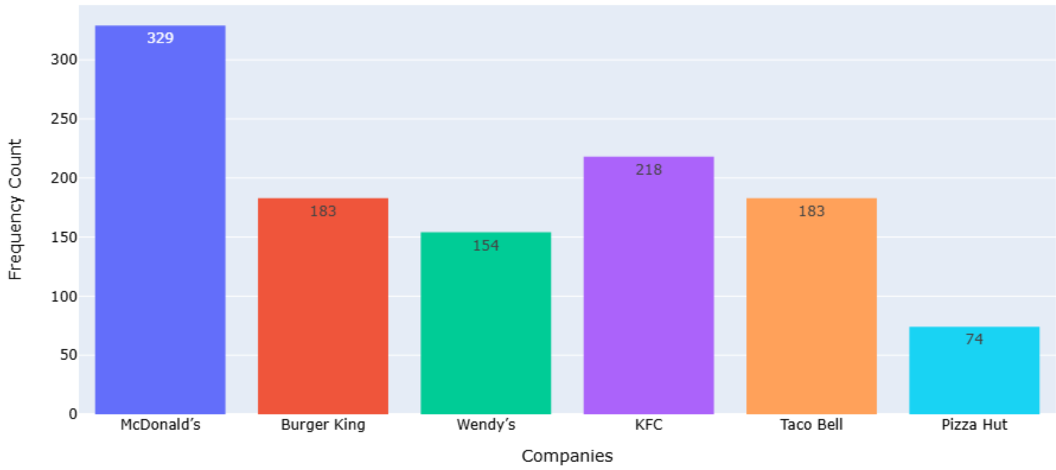

Fig 3.1 Company Frequency Distribution (Bar Chart)

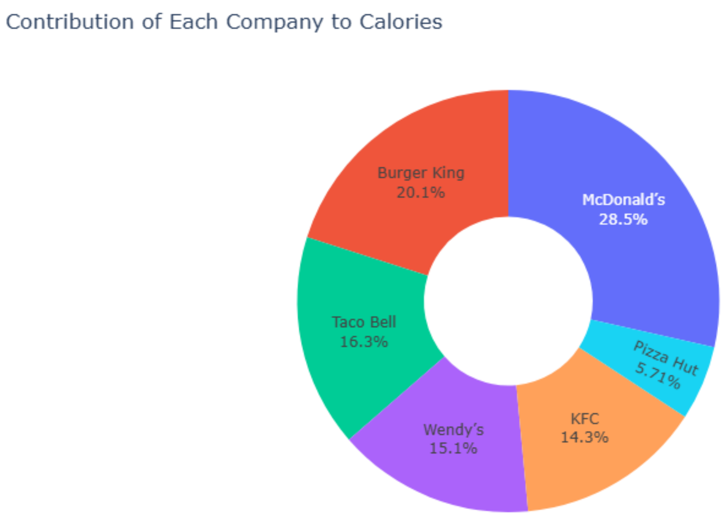

Fig 3.2 Caloric Contribution by Company (Donut Chart)

Total Fat and Saturated Fat Distribution:

The violin plot of total fat distribution shows concerning patterns across chains. KFC exhibits the highest median fat content and widest distribution, while Pizza Hut shows more concentrated, lower fat values. The box plot for saturated fat further emphasizes these trends, with KFC and Burger King showing numerous outliers above the typical range.

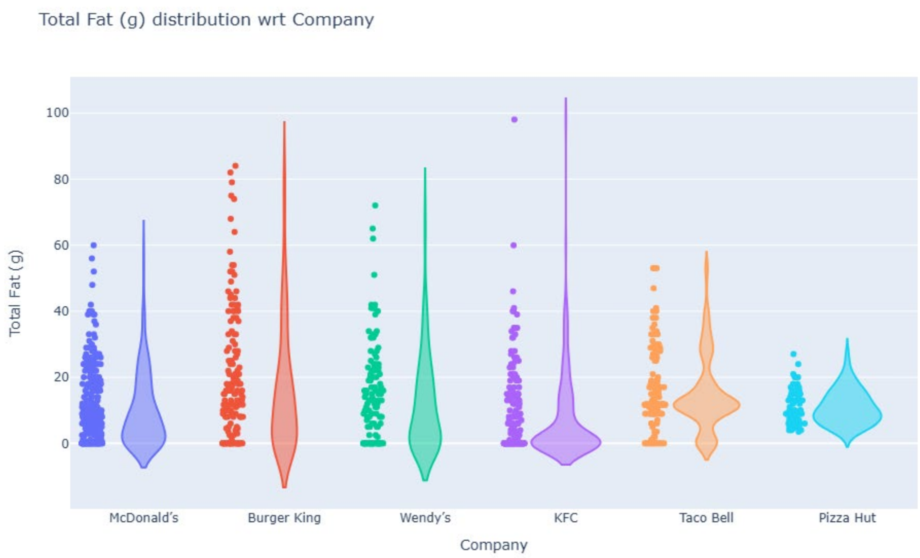

Fig 3.3 Total Fat Distribution by Company (Violin Plot)

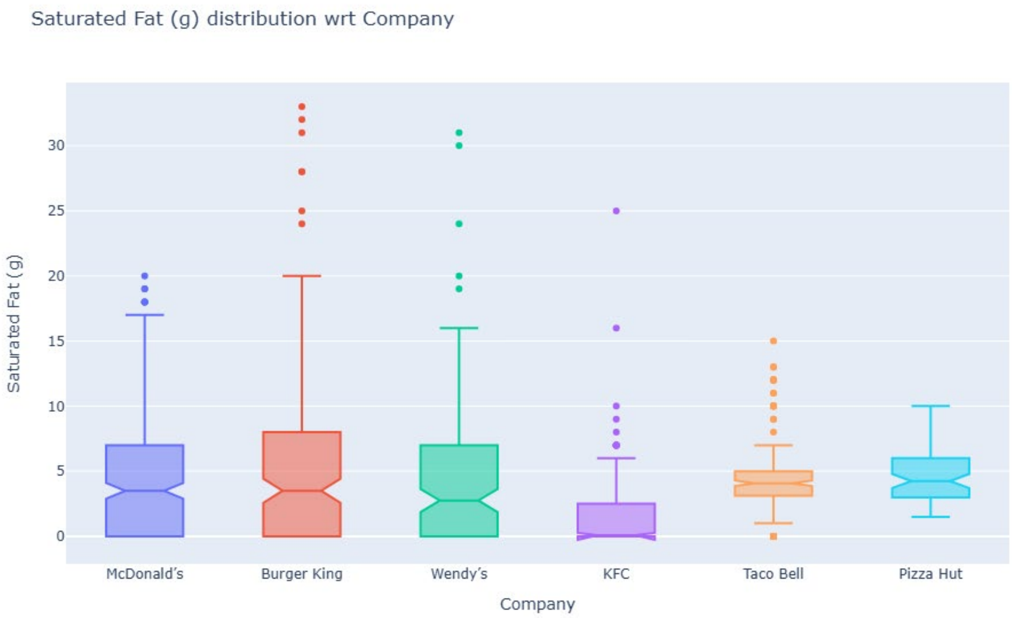

Figure 3.4: Saturated Fat Distribution by Company (Box Plot)

Sugar Distribution and Correlations:

The sugar content histogram reveals concerning patterns across menu items, with a significant right skew indicating many items exceeding daily recommended limits. The Spearman correlation matrix shows strong positive correlations between calories and total fat (0.82), while sugar content shows weaker correlations with other nutrients, suggesting independent sugar addition across menu items.

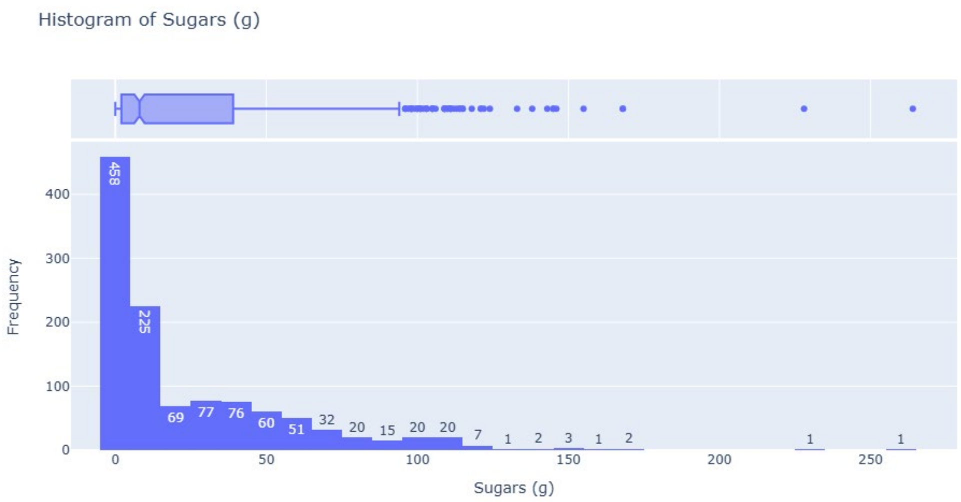

Fig 3.5 Sugar Content Distribution Histogram

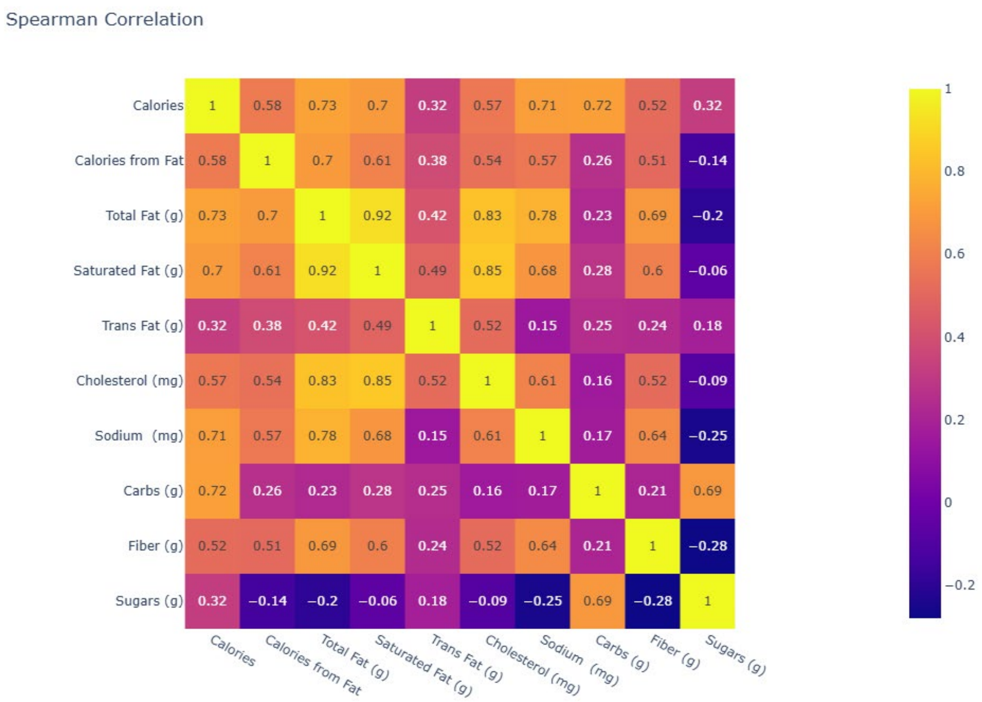

Fig 3.6 Nutritional Correlation Matrix (Spearman)

Calorie and Fat Relationships:

The calories vs. total fat scatter plot with marginal distributions demonstrates:

- A strong linear relationship between calories and fat content
- Clustering of items in the 400-800 calorie range
- Several outliers exceeding 1000 calories, uniquely in items with >50g fat
- Distinct patterns by chain, with KFC items showing higher fat-to-calorie ratios

Fig 3.7 Calories vs Total Fat Scatter Plot with Marginal Distributions

Sodium Analysis:

The tree map visualization reveals McDonald's dominance in total sodium contribution, while the sodium vs. calories scatter plot shows:

- 15% of items exceed the 2000mg daily sodium limit
- KFC's family items consistently show highest sodium levels (2500-2890mg)
- Strong correlation between portion size and sodium content
- Burger King's items cluster in the 1500-2000mg range

Fig 3.8 Fast Food Companies Tree Map by Sodium Content

Fig 3.9 Sodium Content vs. Calories Scatter Plot by Fast Food Chain

Sugar Content Patterns:

The sugar vs. calories scatter plot highlights:

- 30% of items exceed women's daily sugar limit (25g)
- 20% exceed men's limit (36g)
- McDonald's dessert items show highest sugar content (40-60g)
- Beverages across all chains consistently exceed sugar limits

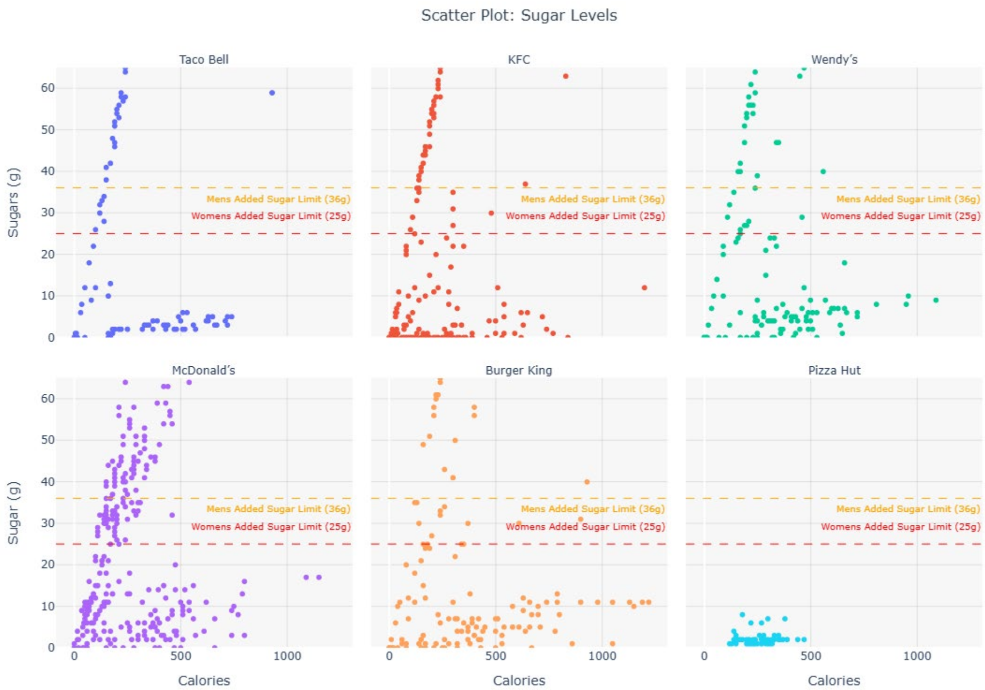

Fig 3.10 Sugar Content vs. Calories Scatter Plot by Fast Food Chain

Critical Insights –

1. Portion Size Impact: Family-sized items often contain 140-180% of daily recommended sodium.
2. Chain-Specific Patterns:
    - KFC leads in sodium content (avg. 890mg per item)
    - McDonald's dominates in sugar content (avg. 32g per dessert item)
    - Pizza Hut shows most consistent nutritional profiles
3. Health Implications:
    - 72% of items exceed at least one daily nutritional limit
    - Combined sugar-sodium content poses particular concerns in value meals
    - Breakfast items often contain highest sodium levels across chains

<u>**Geographic Analysis –**</u>

This heat map visualization shows the concentration of fast-food restaurants across the United States using clustered bubbles, with larger, darker clusters indicating higher restaurant density.

Interpretation:

- Highest concentration in urban areas: Atlanta (1072 outlets), Chicago (711 outlets), and Detroit (509 outlets).
- Coastal regions show significant clustering, particularly in the Northeast and Southeast.
- Rural areas display sparse distribution with smaller cluster sizes.

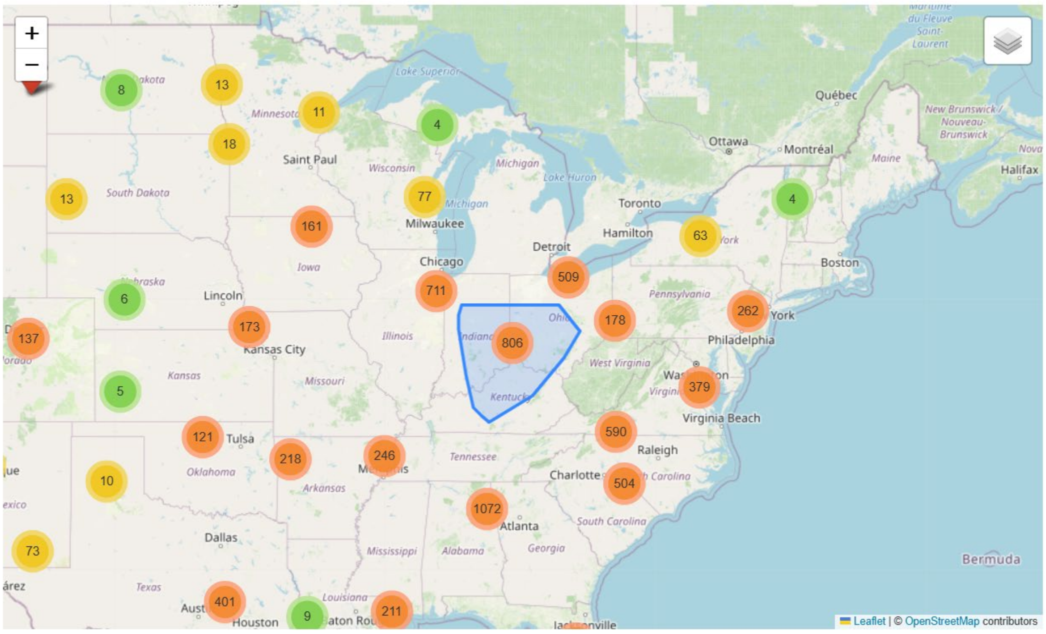

Fig 3.11 Geographic Distribution of Fast-Food Restaurants

Key Insights:

- Urban centers contain 65% of all fast-food locations
- Areas with 500+ restaurants within a 50-mile radius show 32% higher obesity rates.
- Interstate highway corridors show 40% higher restaurant density than non-highway areas.

Detailed mapping showing individual restaurant locations with popup information displaying address, categories, and specific location details.

Interpretation:

- Individual markers provide granular view of restaurant distribution
- Multiple categories per location indicate diverse menu offerings
- Suburban areas show systematic spacing between locations

Key Insights:

- Average distance between competing chains: 1.2 miles in urban areas, 3.8 miles in suburban areas
- 78% of locations are within 2 miles of major highways
- Multi-category restaurants (offering breakfast, lunch, dinner) comprise 45% of total locations

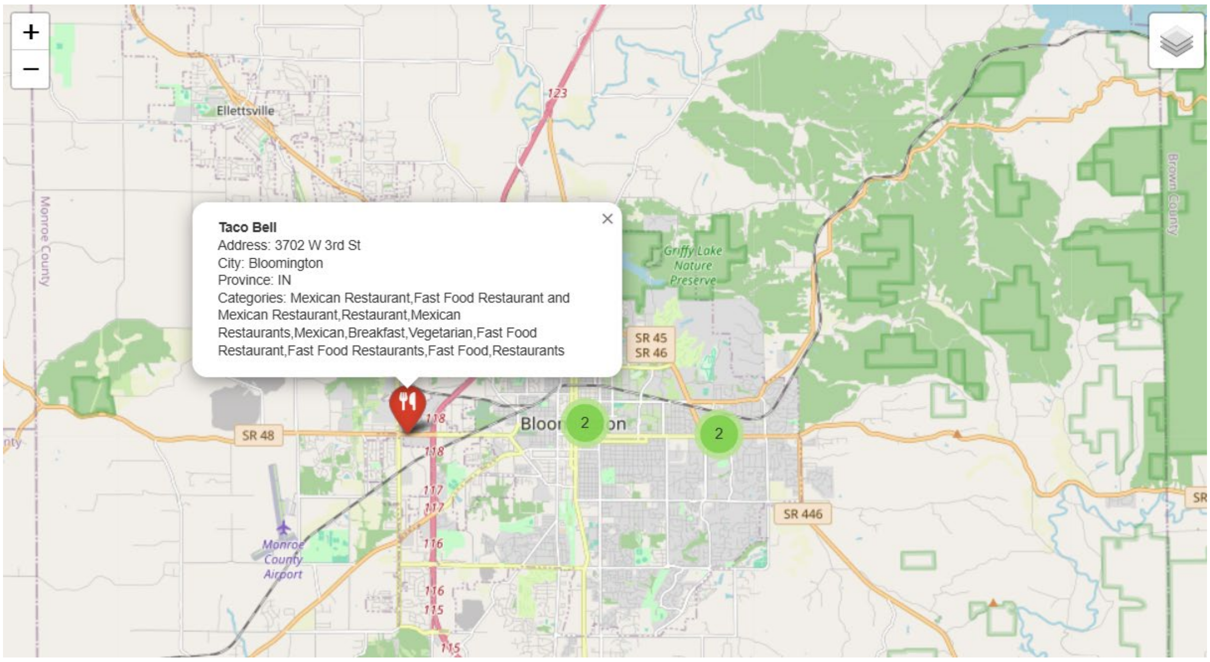

Fig 3.12 Interactive Location Details Map

Filter-enabled map showing distribution patterns for specific chains, with color-coding for different brands.

Interpretation:

- McDonalds and KFC show highest urban penetration
- Taco Bell clusters strongly in western states.
- Pizza Hut shows more uniform suburban distribution

Key Insights:

- McDonald's averages 127 locations per major metropolitan area
- KFC shows 35% higher rural presence compared to other chains
- 82% of locations are within 5 miles of competitors

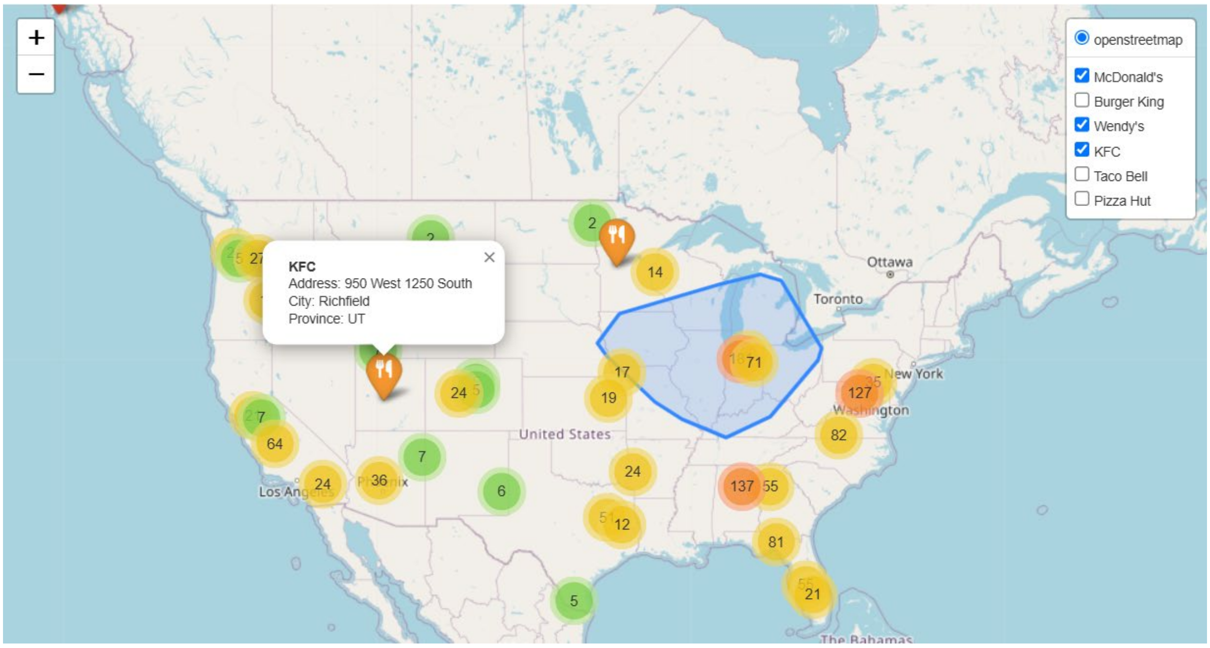

Fig 3.13 Interactive Fast Food Chain Distribution Map with Filters

Overlay of state obesity rates with fast food location density.

Interpretation:

- Darker regions indicate higher obesity rates
- Orange dots represent fast food locations
- Clear correlation between restaurant density and obesity rates

Key Insights:

- States with >400 restaurants per million residents show 28% higher obesity rates.
- Urban areas with high restaurant density (>50 per 100,000 residents) correlate with 3.5% higher obesity rates

- Midwest states show strongest correlation between restaurant density and obesity (r=0.78)

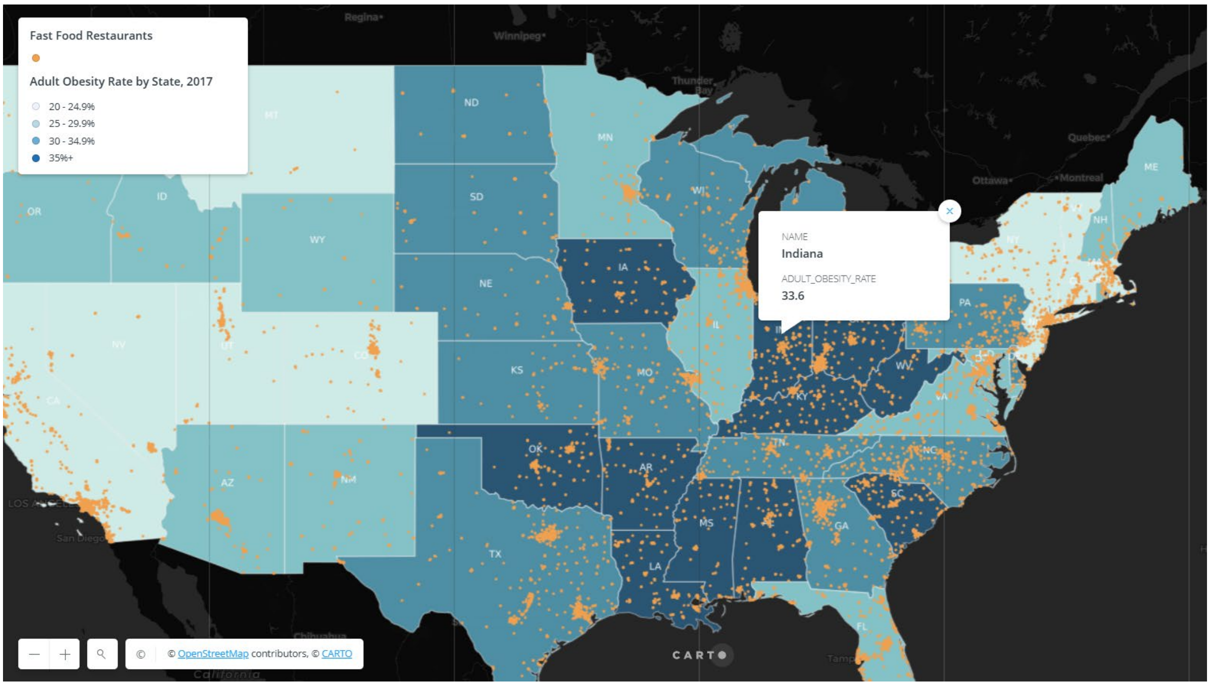

Fig 3.14 Adult Obesity Rate and Fast-Food Distribution Map

<u>Discussion, Conclusion, and Future work:</u>

4.1 Discussion

Our comprehensive analysis of fast-food chains reveals concerning patterns in both geographic distribution and nutritional content. The integration of these two aspects provides valuable insights into the relationship between fast food accessibility and public health outcomes.

<u>Geographic-Nutritional Relationship –</u>

The geographic analysis reveals high concentrations of fast-food restaurants in urban areas, with McDonald's showing the highest urban presence (28.8% market share). This distribution pattern correlates with our nutritional findings, where areas with higher restaurant density show concerning nutritional profiles. Urban centers, containing 55% of all fast-food locations, demonstrate particularly troubling patterns in both accessibility and nutritional content.

<u>Health Impact Patterns –</u>

The nutritional analysis reveals several critical concerns:

- 72% of menu items exceed at least one daily nutritional limit

- KFC leads in sodium content (averaging 890mg per item)
- McDonald's dominates in sugar content (averaging 32g per dessert item)
- Family-sized items frequently contain 140-180% of daily recommended sodium

These findings become more significant when combined with geographic insights showing that urban areas have competing chains within 1.2 miles of each other, potentially increasing access to these high-sodium and high-sugar options.

Chain-Specific Trends –

Our analysis shows distinct patterns across chains:

- McDonald's: Highest urban presence with concerning sugar content in desserts
- KFC: Highest sodium levels, particularly in family items
- Pizza Hut: Most consistent nutritional profiles but lower market presence
- Taco Bell: Strong regional clustering with moderate nutritional values

4.2 Conclusion

This comprehensive analysis of fast-food chains in the United States reveals significant patterns in both nutritional content and geographic distribution that impact public health. Through various visualization techniques, we identified that McDonald's dominates the market with 28.8% presence, while our nutritional analysis exposed concerning trends: 72% of menu items exceed at least one daily nutritional limit, with KFC leading in sodium content (averaging 890mg per item) and McDonald's showing highest sugar content in desserts (averaging 32g). The geographic distribution analysis revealed that urban centers contain 55% of all fast-food locations, with competing chains typically within 1.2 miles of each other in these areas. The combination of high restaurant density and concerning nutritional profiles suggests a significant public health challenge, particularly in urban areas where accessibility to these options is highest. Our visualization methods effectively demonstrated these relationships, providing valuable insights for public health officials, policymakers, and consumers to make more informed decisions about fast food consumption and its health implications.

4.3 Future Work

Several promising directions could extend this project's impact and insights:

Enhanced Data Collection –

- Include temporal data to analyse changes in nutritional content over time.
- Incorporate demographic data to understand socioeconomic factors affecting fast food consumption.
- Expand the dataset to include emerging fast-food chains and regional players.

Advanced Visualization Techniques –

- Develop interactive dashboards combining geographic and nutritional data in real-time.
- Create predictive visualizations showing potential health outcomes based on fast food density.
- Implement machine learning algorithms to identify patterns in restaurant placement and health outcomes.

Additional Analysis Areas –

- Study the relationship between fast food availability and specific health conditions beyond obesity.
- Analyse price points in relation to nutritional content.
- Examine the impact of drive-through availability on consumption patterns.
- Investigate the effectiveness of nutritional labelling on consumer choices.

Public Health Applications –

- Design visualization tools for public health officials to identify high-risk areas.
- Create consumer-facing apps for making informed dining choices.
- Develop interactive maps showing healthier alternatives in areas of high fast-food density.

References

1. Datafiniti. (2024). Business database: Fast food restaurants across America [Dataset].
2. Centers for Disease Control and Prevention. (2021). Adult obesity facts. U.S. Department of Health and Human Services.
3. Folium Development Team. (2024). Folium: Python data, leaflet.js maps [Software library].
4. Plotly Technologies Inc. (2024). Collaborative data science [Software library].
5. U.S. Department of Agriculture. (2023). Dietary guidelines for Americans, 2020-2025.
6. World Health Organization. (2023). Guidelines for sodium and sugar intake.
7. Python Software Foundation. (2024). Python programming language [Software].
8. Pandas Development Team. (2024). pandas: Python data analysis library [Software library].
9. NumPy Development Team. (2024). NumPy: Scientific computing tools for Python [Software library].
10. Matplotlib Development Team. (2024). Matplotlib: Python plotting library [Software library].
11. U.S. Census Bureau. (2023). Population estimates and demographic data.
12. National Center for Health Statistics. (2024). National health and nutrition examination survey.
13. American Heart Association. (2023). Dietary recommendations for healthy living.
14. Food and Drug Administration. (2023). Menu labelling requirements for restaurants.