

Proposal

2024-11-08

Group Members

Lahari Koganti – lk2776
Vidit Tripathi – vt2415
Tai Yue - ty2554

Project Title:

Analysis of Milk Production and Consumption Patterns in the United States

Motivation:

Understanding state-wise milk production and the consumption of whole, low-fat, and skimmed milk is crucial for optimizing supply chains, ensuring food security, and supporting sustainable agriculture. Using USDA data, our project aims to analyze these trends and provide insights for policymakers, dairy farmers, and businesses. Additionally, our project seeks to study consumer behaviour in choosing the type of milk they prefer to drink.

Anticipated Data Sources

- Data for milk sales quantities by product (<https://www.ers.usda.gov/data-products/dairy-data> - Fluid beverage milk sales quantities by product (Annual))
- Data for region subtotals for milk cows, milk per cow, and milk production. (<https://www.ers.usda.gov/data-products/dairy-data> - Milk cows and production by State and region (Annual))

Final Products

Build a website detailing the project goals and motivations, data cleaning process, exploratory analysis, and visualizations of milk production and consumption trends over the years and across regions in the USA. Create a two-minute screencast summarizing the project report.

Analysis, Visualizations and Coding Challenges

- Analysis and Visualizations:
 - Analysis of Whole vs. Other-than-Whole Milk Consumption Across the Years in the United States: Show the results using line plots and quantify the relationship between year and milk consumption for both whole and other-than-whole milk types using linear regression models.

- Analysis of Consumption Patterns for Flavored Whole vs. Flavored Other-than-Whole Milk Across the Years in the United States: Show the results using line plots.
- Analysis of the Distribution of Milk Cows Across Regions and Years in the United States: Present the results using data distribution plots and a geographic distribution map. Additionally, examine how the distribution of milk cows is influenced by geographical factors over time.
- Analysis of the Distribution of Milk Production per Cow Across Regions and Years in the United States: Present the results using plots (eg., violin plots) and a geographic distribution map. Apply regression models to study how milk production per cow varies across regions and years.
- Coding Challenges:
 - Data Tidying: The datasets are large, so we need to tidy them and use relevant data.
 - Year Selection: Select the years for analysis, as the data spans from 1970 to 2023.
 - Visualization: Use the appropriate types of plots for visualizing the data.
 - Regional and State-wise Data: Ensure that comparisons are made correctly across regions and states.
 - Interpretability: Ensure that the analysis and plots are clear, easy to read, and understandable.

Planned Timeline

Weeks	Tasks
November 11-15	Project Review Meeting
November 18-22	Data Tidying
November 25-29	Data Analysis
December 2-6	Website Build
December 7	Project Submission