## Olivia Yuengling DATA 205 CRN 33334 Professor Perine May 14th 2025

## DATA 205 Final Report

## Introduction

Over the past decade, U.S. grocery prices have risen steadily, with sharp spikes during 2021–2022 that outpaced wage growth, disproportionately impacting low-income households. This project aims to analyze national data to explore food price trends, expenditure patterns across income levels, and the relationship between inflation and consumer behavior.

## Project Objectives

This project examines the relationship between core inflation, food prices, and consumer food spending patterns in the United States from 2013 to 2023. Through analyzing food price and expenditure data, the project seeks to: (1) identify food categories most susceptible to price volatility; (2) Analyze inflation and food price trends from 2013 to 2023; (3) compare expenditure trends across different income groups; (4) detect seasonal fluctuations in food prices; and (5) develop econometric models to link household grocery expenditures with related food spending behaviors, such as dining out and school meals. By integrating price data with consumer expenditure survey responses, this project aims to provide a comprehensive understanding of how inflationary pressures and other food prices influence food-related decision-making at the household level.

## Tools & Methods

The initial data extraction, ZIP‐file handling, and merging were implemented in Python using the zipfile, os, re, and pandas libraries for the CES dataset. Subsequent data cleaning, visualization, and statistical modeling were performed in R with the tidyverse, dplyr, ggplot2, stringr, readr, tidyr, and scales packages.

## Data Overview

This project draws on multiple datasets to examine food prices, core inflation, and consumer spending behavior in the U.S. from 2013 to 2023. The primary source is the Consumer Expenditure Interview Survey (CES) from the Bureau of Labor Statistics (BLS), which provides detailed information on household expenditures and demographics. Data from various zip files were extracted, merged, and cleaned using Python, with key variables including year, income group, average spending on food (total, groceries, dining out, school meals), and the number of consumer units per group.

The average prices dataset, also from the BLS, contains monthly food price data by item and geographic area. Key fields include series ID, year, period, item name, area name, and price. This dataset was merged with core inflation data from the Federal Reserve Bank of St. Louis, which includes the monthly median core CPI, enabling analysis of how food price trends align with broader inflationary patterns.

## Data Cleaning

This project and cleans three primary datasets: the Consumer Expenditure Interview Survey (2013–2023), average food prices from the Bureau of Labor Statistics (BLS), and core inflation data. All raw quarterly data files from the CES data from 2013 to 2023 were extracted from their respective ZIP folders and merged into single dataframes by using Python.

The Consumer Expenditure Survey dataset (cex\_raw) originally contained 930,344 rows and 40 columns. It was filtered and manipulated to create relevant variables: year, income group, mean expenditures across categories (total food, food at home, food away from home, and school meals), and the number of households per income group. Redundant or unused variables were dropped.

The food\_prices dataset initially contained 193,079 rows and 6 columns, including series\_id, year, period, value, item\_name, and area\_name. Items were categorized into 14 specific food groups (e.g., “Beef,” “Dairy,” “Fruits,” “Grains & Baked Goods”) through the creation of a new variable: “food\_category”. Entries under the categories “Other” and “Utilities” were excluded. The dataset was then filtered to include only entries from 2013 to 2023, resulting in 26,018 rows with 7 columns.

Next, the core\_inflation dataset. It contains 132 rows with the variables year, period, and MEDCPIM158SFRBCLE (the median core consumer price index) was renamed for clarity (med\_core\_price\_index). It was joined with the filtered food\_prices dataset using a combination of year and period variables via an inner join, increasing the number of columns in food\_prices to eight.

This cleaning process ensured that all datasets were aligned by year and period, allowed for meaningful category-level comparisons, and enabled subsequent modeling and analysis of inflation-adjusted food spending trends across demographic groups. Full cleaning scripts are available in the GitHub repository.

## Basic Statistics

The average\_prices dataset covers years 2013 to 2017 with 26,018 data entries. It records monthly food prices for 12 distinct categories, such as "Dairy," "Poultry," and "Fruits." The price values range from $0.30 to $19.13. The dataset additionally includes core inflation data (from the core inflation dataset), with the median core CPI averaging approximately 3.03 during the observed period, peaking at 8.19. Out of all entries, only two missing values were identified and were subsequently excluded during the cleaning process.

The CES dataset spans from 2013 to 2023 and includes a total of 67 data entires representing various household income groups. The dataset includes variables related to food expenditures across seven income groups, with grocery spending ranging from $53.40 to $1,574.33. The average total food expenditure within this dataset reached up to $2,487. On average, each income group consists of roughly 13,886 individuals, with null entries removed during data cleaning to maintain consistency in calculations.

## Final Data Products

A graph of growth in food expenses

AI-generated content may be incorrect.This heatmap (Figure 2) visualizes the year-over-year percentage change in total food spending across different income groups from 2014 to 2017. Each row represents an income bracket, ranging from "Less than $5,000" to "$70,000 and over," while each column represents a year. The color intensity of each cell indicates the growth rate in food expenditure for that specific income group and year. Lighter shades of pink represent smaller increases, while darker shades signify larger increases. Notably, grey cells would indicate a decline in food expenditure.

Figure 1

Looking at the data, we can observe varying patterns of food expenditure growth across income levels and time. For instance, lower-income groups appear to have experienced more significant percentage increases in their food spending in 2015 compared to higher-income groups. The patterns shift in subsequent years, suggesting that economic factors and inflation may have differentially impacted food budgets across different income segments. Analyzing the specific color intensities for each income group and year allows for a detailed understanding of how food spending evolved for various segments of the population during this period.

A screenshot of a computer

AI-generated content may be incorrect.To investigate the potential association between food expenditure growth and income group, an Analysis of Variance (ANOVA) was conducted (Figure 2), with GrowthRate as the dependent variable and IncomeGroup as the independent variable. The resultant F-statistic was 0.824, yielding a corresponding p-value of 0.589.

Figure 2

Given that the observed p-value (0.589) exceeds the conventional alpha level of 0.05, the null hypothesis, which posits no statistically significant differences in mean food expenditure growth rates across income groups, fails to be rejected. This outcome suggests that, within the parameters of the current dataset, income group does not exert a statistically significant effect on food expenditure growth.

However, it is imperative to interpret this finding with circumspection. The analysis is predicated on a temporal window of four years, representing a limited sample that may not adequately capture long-term trends or inherent variability within the data. Consequently, although the statistical test did not identify a significant relationship, the limited temporal scope of this analysis may undermine the robustness and generalizability of these findings.

## References & Resources

U.S. Bureau of Labor Statistics. (n.d.). Consumer Expenditure Surveys. Retrieved May 13, 2025, from https://www.bls.gov/cpi/data.htm

U.S. Bureau of Labor Statistics. (n.d.). Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average [CPILFESL]. FRED, Federal Reserve Bank of St. Louis. Retrieved May 13, 2025, from https://fred.stlouisfed.org/series/CPILFESL

Ritz, R. (n.d.). Average Price Data (BLS) [Dataset]. Kaggle. Retrieved May 13, 2025, from <https://www.kaggle.com/datasets/robertritz/average-price-data-bls>

OpenAI. (2023). *ChatGPT* (GPT‑4) [Large language model]. Retrieved May 14, 2025, from <https://chat.openai.com/>

Anthropic. (2023). *Claude* [Large language model]. Retrieved May 14, 2025, from <https://claude.ai/>