Forecast Accuracy Analysis and Visualization System for CPI and PPI

Dataset → https://www.ers.usda.gov/data-products/food-price-outlook

This dataset is curated by the United States Department of Agriculture Economic Research Service (USDA ERS) and it provides data on food prices and forecasts annual food price changes up to 18 months in the future. This dataset contains eight different tables, broken up into four categories: Consumer Price Index (CPI), Producer Price Index (PPI), Historical Data, and Archived Data. Going on to the individual data level, we have the following:

- CPI forecast - Includes recent point estimates and various percentage changes (monthly, year-over-year, year-to-date, etc.) using the latest methods. These forecasts reflect the most recent data and are updated monthly
- PPI forecast - Current forecasts of annual PPI for food-related items. Similar to the CPI forecast, this contains recent point forecasts for prices paid to domestic producers (e.g., farm-level and wholesale food)
- CPI forecast archived - Historical CPI forecasts as originally published. These reflect the actual point estimates and ranges that were released monthly, based on the old forecasting methodology used before July 2023
- PPI forecast archived - Historical PPI forecasts as originally published. Like the CPI archive, this contains original forecasts and ranges released before the switch to the new methodology
- CPI historical forecast - These show what the CPI forecasts would have been in the past (starting at 2003) if the current methodology had been used, including point estimates and prediction intervals
- PPI historical forecast - Same as the CPI historical forecast file, but for producer-level price data. Includes midpoints and confidence intervals.

- Historical CPI - This shows the real historical percent change in the CPI for food, as released by the Bureau of Labor Statistics. Useful for comparing forecasts to actuals.
- Historical PPI - Contains the real historical PPI percent changes, also used for evaluating forecast performance.

The full dataset description and files can be found at the link above.

Project Goals

- Design and implement a relationship database relating all PPI and CPI data, ranging from current to historical to centralize food price outlook queries
- Track forecast accuracy over time by comparing forecasts with the actual values
- Compare and analyze differences in old methodology and new methodology when it comes to forecasting PPI and CPI
- Analyze trends across categories to provide insights on food price
- Provide forecast confidence evaluations to showcase to users the uncertainty of the predictions
- Build a simple GUI that combines all of these goals in order to make the data accessible and digestible in the form of tables and graphs

Project Details

The data will most likely need to be normalized– already there are discrepancies between the different data tables that will need to be resolved. This data is updated regularly and from the USDA, meaning there is no generated or falsified data. Using the data tables described above, this project will be able to answer queries such as “Show all forecasts for Meat in 2020” or “Compare archived vs historical forecasts for 2007.” There’ll be visualizations using graphs and charts to convey this information, as well as a table displaying the raw data on the side. There will be filtering options such as a year range, item type, forecast method, consumer

or producer only, etc. The database will be done using MySQL (using connection pooling) and the frontend and backend will be done using Java and Java libraries. The preferred IDE will be IntelliJ for its comprehensive package managing and compilation using libraries such as Maven or Gradle. Version control will be done using GitHub.