## Hook Document

*Supermarket Swings: Forecasting Grocery Price Volatility*

**Scenario:**

You've just joined the Virginia Cooperative Extension as a data consultant. Rising inflation is putting pressure on food budgets across the Mid-Atlantic. Your job? Use data from the Bureau of Labor Statistics to uncover which grocery items are most susceptible to price swings—and which are predictably stable.

**Context/Motivation:**

Families, food banks, and policymakers rely on food price forecasts to make strategic decisions. But which items can be trusted to remain steady? Which spike during crises like COVID-19? Using time-series forecasting, you'll uncover trends in eggs, dairy, produce, and more.

**Mission:**

You'll analyze 10 years of BLS grocery pricing data, build SARIMA (or alternative) forecasting models, and identify the most volatile products. Then, you'll make a clear, communicative notebook or report that helps non-technical readers understand the stakes.

**Outcome:**

Create an accessible and reproducible analysis that answers:
*Which grocery items have shown the greatest price volatility in the past decade, and what might explain it?*

## Rubric

Case Study Rubric – Supermarket Swings

Target Audience: 2nd-Year UVA Data Science Students

Assignment Type: Case Study (Individual)

Deliverable Format: GitHub repo + Notebook/Report + Visualizations

## **Overview**:

You will use historical data from the Bureau of Labor Statistics to explore price volatility in grocery items over the past 10 years in the Mid-Atlantic region. This includes time-series modeling (SARIMA or justified alternatives), data cleaning, and interpretation of trends. Your deliverable should be a clear analysis notebook or mini-report.

## **Required Components:**

| ***Component*** | ***Description*** | ***Points*** |
|---|---|---|
| **Notebook or Report** | Includes problem statement, analysis steps, SARIMA model, volatility comparison, and conclusions | 40 |
| **Data Cleaning Pipeline** | Clear documentation of how missing data was handled, how items were filtered, and data was prepped for modeling | 20 |
| **Visualizations** | Appropriate line plots, forecasts, and boxplots that effectively communicate trends | 15 |
| **Interpretation & Insights** | Reasoned explanation of volatility patterns and contextual factors (e.g., perishability, inflation, seasonality) | 15 |
| **README & Reproducibility** | GitHub repo includes README, code runs without errors, all data and outputs are accessible | 10 |

| Criteria | Meets Expectations |
|---|---|
| **Time-Series Modeling** | Implements SARIMA or alternative; justifies parameter selection and model performance |
| **Clarity of Communication** | All findings clearly presented in notebook or report format |
| **Thoughtful Decision-Making** | Explains tricky tradeoffs (e.g., why exclude some items, why certain years are prioritized) |
| **Relevance** | Connects volatility to external factors (e.g., supply chain, COVID, seasonality) |
| **Reproducibility** | Code runs without issues; GitHub repo is self-contained and well-documented |

**3. Supporting Materials**

**Suggested Explainer Blog Post:**

- Neptune.ai SARIMA Blog
  *Good entry point for understanding time-series forecasting techniques.*

**Suggested Article for Technical Depth:**

- *BLS Average Retail Food and Energy Prices, Mid-Atlantic Dataset:*

  https://www.bls.gov/regions/mid-atlantic/data/averageretailfoodandenergyprices_usandwest_table.htm
  *Includes methodology explanation and public dataset context.*

*Why this matters:*

- *https://www.sciencedirect.com/science/article/pii/S0306919213001188*
- *https://libertystreeteconomics.newyorkfed.org/2024/07/what-was-up-with-grocery-prices/*