In this project I built a linear regression model that aims to understand the influence of various economic indicators on food prices'.

Our **dataset comprises observations from multiple economic sectors**, including energy, food, gasoline, housing, and mortgage interest rates. We collected data over several years to capture both short-term fluctuations and long-term trends.

Initial Variable Selection

"We chose our independent variables based on theoretical considerations and empirical evidence suggesting their relevance in economic modeling. These variables include:
- **All Items**: A general index of prices.
- **Energy and Gasoline**: Key components of cost in many industries.
- **Food**: An essential sector affecting and affected by economic conditions.
- **Mortgage and Interest Rates**: Indicators of housing market dynamics and financial policy.
- **Shelter**: Reflecting real estate market conditions.
- **Time Variables (Year and Month)**: To account for temporal trends and seasonal effects."

Model Building Process

"Our model building involved several key steps:
1. **Preprocessing**: We cleaned the data, handled missing values, and verified the integrity of our data.
2. **Encoding Categorical Data**: For categorical variables like geographic regions, we applied one-hot encoding to transform them into a numerical format suitable for regression analysis.
3. **Model Estimation**: We fitted the model to our preprocessed and encoded data.
4. **Diagnostics**: We conducted diagnostic tests to check for issues such as multicollinearity, autocorrelation, and non-normality in residuals."

Model 1

The first linear regression model to understand how various economic factors impact the 'value' was based on oil seeds other members did other commod. Our model indicates a strong predictive capability and explains approximately 81% (adj r = 0.808) of the variation in 'value

Significantly, factors like 'mortgage and interest' and 'energy' show substantial negative impacts on 'value', while 'shelter', 'food', and 'gasoline' have positive effects.

Our analysis shows that while our model is robust in explaining a significant portion of the variability in economic values, there are areas such as autocorrelation and multicollinearity that we need to address (two or more predictor variables in a regression model are highly correlated.)

Heat Map

The correlations were investigated via a heat map.

Our heatmap analysis revealed notable negative correlations between mortgage rates and other economic indicators such as housing prices and construction spending. This suggests an inverse relationship where increases in mortgage rates may lead to decreases in these sectors. These insights are crucial for understanding market dynamics and can influence policy formulation, economic forecasting, and investment strategies.

We can use this to see which variables are highly correlated.

Updated Regression 2
Impact of Removing Variables:

Removing 'all_items', 'year and month, reduced multicollinearity, which initially might have masked the true relationships between variables like 'shelter' and the dependent variable.

Updated Regression 3
Impact of Removing Variables:

The removal of 'food' might have eliminated a confounding effect, where 'food' was capturing some of the effects.

Summary of Regression Models and Coefficients:

1. **First Model**:

 - **Shelter**: Initially positive.
   - **Food**: Consistently positive until it was removed.
   - **Energy**: Consistently positive.
   - **Gasoline**: Initially positive.
   - **Mortgage and Interest**: Initially significantly negative.

2. **Second Model** (after removing 'all_items', 'year', and other variables):

- **Shelter**: Turned negative.
   - **Energy**, **Gasoline**, and **Mortgage and Interest**: Continued being significant with varied influence.

3. **Third Model** (after further removing 'food'):

 - **Shelter**: Turned positive again.
   - **Energy**: Increased positive influence.
   - **Gasoline**: Became negative.
   - **Mortgage and Interest**: Became insignificant.

Analysis of Coefficient Changes:

Energy: The increasing significance and magnitude of the 'energy' coefficient across models suggest that as other confounding or interacting variables were removed, the more direct and robust relationship of 'energy' with the dependent variable became apparent.

Gasoline: The change from positive to negative could indicate that its initial positive association was due to interactions with variables like 'food', which when removed, revealed a more typical expected inverse relationship.

Economic and Industrial Connections Between Oilseeds and Gasoline:

Biofuel Production:

: A significant portion of oilseeds is used in the production of biofuels. As biofuel production increases, the demand for oilseeds increases, potentially raising their prices.

- Impact on Gasoline: Increased biofuel production can lead to a substitution effect where biofuels partially replace traditional fossil fuels like gasoline. This might initially increase the price of gasoline due to blend requirements (adding biofuels to gasoline), which could explain an initial positive correlation. However, as biofuel production becomes more efficient and widespread, it might reduce the overall demand for gasoline, potentially leading to lower gasoline prices, hence a shift to a negative correlation.

Conclusion

In our sequential model refinement process, we observed significant changes in the coefficients of our predictors, particularly 'shelter', 'energy', and 'gasoline'. These changes were influenced by the removal of variables like 'all_items', 'year', and 'food', which altered the dynamics within our model. These modifications reveal the intricate balance in capturing the true economic relationships and underscore the necessity for careful model construction and validation. Each step in our analysis brought new insights and highlighted the critical role of thorough diagnostic testing and theoretical alignment in econometric modeling.

This approach not only helps in improving the model but also in communicating the complexity and iterative nature of model-building in econometrics to stakeholders or academic peers.

The observed shift from a positive to a negative relationship between gasoline prices and oilseed values underscores the complex interplay between energy markets and agricultural commodities. The initial positive correlation may have reflected an increase in biofuel

production driving up both gasoline and oilseed prices. Over time, however, as biofuels potentially displace some gasoline demand, we see a reversal in this trend. This dynamic prompts us to consider broader economic forces and policy impacts when analyzing commodity markets, ensuring our model reflects underlying real-world interactions.

Understanding these connections and their implications can significantly enhance the robustness and relevance of your econometric modeling, providing critical insights for stakeholders in both the energy and agricultural sectors.