De La Salle University

Final Paper for DATA100

Exploring Trends in Wholesale Goods Prices Per Region as Economic Indicators

Group 1
ARDALES, DARIUS VINCENT CIRUJALES
PE, STEVEN DANIEL TINGA
SIABABA, AUBREI GOULD REMUDARO
YAO, ZEYAO
YU, JOHN XAVIER TECSON

I. Details of Dataset

The dataset under consideration includes the Wholesale Selling Prices of Agricultural Commodities in the Philippines. Its key features provide a comprehensive view of the pricing dynamics within the agricultural sector. The geographical location of the data ranges from the national level down to specific regions. The time period of the dataset ranges from 2010 to 2023 including a diverse array of agricultural goods like cereals, root crops, beans and legumes, condiments, fruit vegetables, leafy vegetables, fruits, commercial crops, poultry, and fish. Time intervals are also captured both monthly and annually, allowing for a nuanced temporal perspective on the fluctuations and patterns within the data.

The data was collected from the PSA OpenStat Prices, specifically the Wholesale Selling Prices (New Series) tables. The dataset is extensive, covering a wide range of agricultural commodities and geographical locations within the Philippines. To access the data, users can navigate through the PSA OpenStat Prices interface, selecting the desired commodity, geolocation, year, and period.

The dataset is a dynamic and evolving resource, regularly updated to reflect the latest market information. As of the latest update on 11/22/2023, the dataset is current for certain commodities, providing a timely and reliable foundation for the analysis of wholesale prices in the agricultural sector.

II. Overview

The wholesale and retail industry involves buying and selling goods and services at various stages of the supply chain, often acting as intermediaries between the agrarian and retail sectors (Manuel, 2022). Agricultural products, such as fresh produce and raw materials, are typically sold in bulk to wholesalers who then distribute them to retailers. Retailers, in turn, sell these products in smaller quantities to end consumers (Abrons, 2017). The supply chain linkage between wholesalers and retailers facilitates the coordinated delivery of agricultural products, acting as a crucial intermediary between the agrarian sector and end consumers.

The wholesale price index in the Philippines is an indicator designed to measure the changes in the price levels of commodities that flow from the wholesaler to the retailer and the wholesaler to the producer (end-user). It covers the wholesale prices of commodities from two levels of bulk distribution and is a vital guide in economic analysis and policy formulation, as well as a basis for price adjustments in business contracts and projects. The average domestic wholesale price of well-milled rice in the Philippines was approximately 38.36 Philippine pesos per kilogram in 2022, while the average domestic wholesale price of premium rice was 42.03 Philippine pesos per kilogram in the same year (Moody's Analytics, n.d.). This information demonstrates the significance of understanding the dynamics of wholesale goods prices in agriculture, especially in a country where the agricultural sector plays a critical role in the economy. Aside from this, price trends and fluctuations in agricultural commodities have significant implications for the economy, inflation, and the well-being of the farming sector. The dataset for this paper is sourced from PSA OpenStat Prices, offering insights that can cover a diverse range of commodities. The dataset captures the wholesale prices of a diverse range of agricultural commodities. Understanding these prices is integral to comprehending the economic dynamics of the supply chain, from producers (farmers) to the wholesale market. In addition, wholesale prices are a fundamental component in determining the cost structure of goods within the supply chain. These costs influence retail prices, thereby impacting consumer behavior and market trends. Analyzing wholesale prices provides insights into the broader economic implications for both businesses and consumers within the Retail and Wholesale Industry (Cho, 2020). The dataset incorporates economic indicators such as the Gross Regional Domestic Product (GRDP) and GRDP per capita. This linkage between wholesale prices and economic metrics establishes a crucial connection, as the Retail and Wholesale Industry significantly contributes to regional economic performance. This means that policymakers can benefit from insights derived from the dataset to formulate targeted interventions and policies that support the Retail and Wholesale Industry. The dataset's coverage of various commodities and regions enables a nuanced approach to policy development, addressing specific challenges faced by different segments of the industry. Lastly, this data can be used by agricultural supply chain businesses to make informed decisions about pricing strategies, supply chain management, and market expansion. Consumers benefit indirectly as well, as a transparent understanding of wholesale prices contributes to a more stable and affordable food supply.

This paper aims to delve into the dynamics of wholesale goods prices over time, considering the impact of inflation, regional variations, and economic indicators. Understanding the trends and patterns in wholesale goods prices is crucial for various stakeholders, including policymakers, businesses, and consumers.

III. Research Questions

This paper aims to examine the relationship between trends in wholesale good prices with GRDP and GRDP per capita:

i. Which regions experience the most/least expensive wholesale prices for agricultural commodities?
ii. Have wholesale prices remained constant over time, or have there been significant changes?
iii. Is there a correlation between wholesale prices and the Gross Regional Domestic Product (GRDP) or GRDP per capita?
iv. What types of agricultural goods are consistently the most/least expensive, and how has this changed over time?

IV. Objectives

The paper aims to investigate the dynamics of wholesale goods pricing over time, taking into account the effects of inflation, geographical variances, and economic indicators. The paper includes a comprehensive analysis that includes adjusting prices for inflation, assessing trends in average prices, identifying places with the most/least costly prices, and investigating relationships with GRDP and GRDP per capita. The primary objectives of this paper are to:

1. Analyze wholesale goods prices over time, considering inflation, regional variations, and economic indicators.

2. Investigate trends in average prices and identify regions with the most/least expensive prices.
3. Assess the consistency of price changes over time and their reflection in GRDP or GRDP per capita data.
4. Explore the pricing dynamics of different agricultural commodities.

V. Methodology (the tools, techniques, code fragments)

This paper employs various statistical techniques such as exploratory data analysis (EDA) and regression analysis, with the potential for additional insights from seasonal trends in monthly wholesale goods data. The dataset will be adjusted for inflation and the implementation will utilize Python programming for data manipulation, statistical analysis, and visualization.

a. CPI and Inflation

As we all know, money depreciates due to inflation. This means using the same amount of money from 2023, we get fewer products than we would if we were in 2013. Therefore, it is important to eliminate the effect of inflation on the price of the product before we can compare them.

The consumer price index (CPI) is a popular measure of the inflation rate by getting the change in the money spent by consumers over time. It uses the price of the market basket at a specific time as the basis and compares the prices of the market basket at other times (Eq.1). It can be applied to study the general inflation nationally and annually or the specific inflation rate of an industrial over an amount of time (Fernando, 2023). The CPI was used in this study to find the specific inflation in the gross regional domestic product (GRDP) per region per year by setting 2018 as the base year.

$$CPI = \frac{\text{price in the market basket in the current time}}{\text{price in the market basket in the base time}} \times 100 \qquad \text{(Eq.1)}$$

In order to adjust the prices of the products, Equation 2 is used. The original price of the product is to be multiplied by the ratio between the CPI of the base year and the CPI of the target year, which is to essentially divide the price by the inflation rate of the target year. The price after adjustment represents the amount of money needed to purchase the product in the target year using the money with purchasing power from the base year (*How to adjust for inflation in monetary data sets*, 2021).

$$Price_{after\ adjustment} = \frac{Price_{before\ adjustment} \times CPI_{base\ year}}{CPI_{target\ year}} \qquad \text{(Eq.2)}$$

b. Regression Analysis

Regression analysis covers a set of statistical techniques that are used to assess the correlation between multiple variables. In general, the different techniques employed in regression analysis work by generating and then tuning a mathematical model derived from data points. The accuracy of the model is determined by the deviation of the data points within the data set from the mathematical model. Regression analysis is a near ubiquitous tool in different fields of study including the natural sciences, social sciences, business analytics, and many more.

i. Linear Regression

Linear regression is a technique used in order to determine the degree to which a set of independent and dependent variables linearly correlate. In its most basic form, linear regression mathematically reduces to an expression in which a single dependent variable is related to an independent variable through a correlation coefficient and an error term (*Regression Analysis - Encyclopedia of Mathematics*, n.d.). For the purposes of this study, a multiple regression is performed in which multiple independent variables (wholesale goods prices) are correlated with a single dependent variable (GRDP per capita) with a model with the following mathematical model.

$$\mathbf{y} = a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n + \epsilon_i \qquad \text{(Eq. 3)}$$

Where:

$\mathbf{y} = \text{dependent variable}$

$a = \text{correlation coefficient}$

$\mathbf{x} = \text{independent variable}$

$\epsilon_i = \text{error term}$

***NOTE:** for succeeding equations, these variables are retained

ii. Regularization Models (Ridge Regression)

Ridge regression, also called Tikhonov regularization, is a method for calculating regression coefficients for a multiple regression. In particular, it is useful in situations where there are reasons to believe that the independent variables may be correlated with one another (Igual & Seguí, 2017). Mathematically, Ridge regression seeks to minimize the following expression:

$$\min\left(\sum_{i=0}^{n}(a_0 + \sum_{j=1}^{d} a_j x_{ij} - y_i)^2 + \alpha \sum_{j=1}^{d} {a_j}^2\right) \qquad \text{(Eq. 4)}$$

iii. Sparse Models (Lasso Regression)

The last of the regression models to be used in this paper, Lasso regression is yet another technique used to model multiple regression systems. Lasso regression seeks to hone in on the independent variables which are the best predictors of the independent variables while reducing all less relevant variables to a coefficient of zero (Igual & Seguí, 2017). It is mathematically formulated as:

$$\min\left(\sum_{i=0}^{n}(a_0+\sum_{j=1}^{d}a_j x_{ij}-y_i)^2+\alpha\sum_{j=1}^{d}|a_j|\right) \qquad \text{(Eq. 5)}$$

c. python implementation (code fragments with descriptions)

i. Adjustment for Inflation

In Python, the GRDP in each region in different years is divided by the GRDP of that region in 2018 and then multiplied by 100 to get the CPI of that region and that year using a for loop. Then it was stored in this data frame called CPI (Fig.1).

```
#calculate for the consumer price index (CPI) based on 2018 price
for col in CPI.columns:
    CPI[col] = (GRDP['At Constant 2018 Prices ' + str(col)]/GRDP['At Constant 2018 Prices ' + str(2018)])*100
```

Figure 1. Code fragment for the CPI computation

Then, the wholesale datasets were treated with this function named adjust(). It first converts the data types of the prices from objects to numeric values and rewrites the missing prices as NaN. Then, the prices of goods from different categories had their price adjusted accordingly with the CPI data obtained using Equation 2 for them to all be based on the year 2018 (Fig.2).

1.2 Adjust the price

```
#A function that adjusts all the wholesale prices based on the computed CPI based on 2018
def adjust(csv):
    #import the file
    data = pd.read_csv(csv)

    #rename the columns
    for col in data.columns:
        if col not in ['Geolocation', 'Commodity']:
            new_col = col.replace(' Annual', '')
            data.rename(columns={col: new_col}, inplace=True)
            #adjust the data type for the price columns to be numeric, and the missing values will be replaced with NaN
            data[new_col] = pd.to_numeric(data[new_col], errors='coerce')

    #create a loop that will divide each price with the appropirate CPI based on the year and the region it is in
    for index, row in data.iterrows():
        reg = row['Geolocation']
        for col in CPI.columns:
            data.loc[index, str(col)] /= CPI[str(col)][reg]/100

    #the output will be the adjusted dataset
    return data
```

Figure 2. Code fragment for the function adjust() that eliminates the effect of inflation in the prices of the wholesale goods

ii. GRDP per Capita Feature Generation

	Geographic Location	2010	2015	2020
0	..NATIONAL CAPITAL REGION (NCR)	11855975	12877253	13484462
1	..CORDILLERA ADMINISTRATIVE REGION (CAR)	1616867	1722006	1797660
2	..REGION I (ILOCOS REGION)	4748372	5026128	5301139
3	..REGION II (CAGAYAN VALLEY)	3229163	3451410	3685744
4	..REGION III (CENTRAL LUZON)	10137737	11218177	12422172

Figure 3. Initial observations in the census dataset

To effectively analyze trends in wholesale goods prices, there is a need for the Gross Regional Domestic Product (GRDP) per capita. This metric serves as a reliable measure of the average wealth within a region, offering insights into the income available for expenditures, especially on essential commodities such as wholesale goods. However, at present, the GRDP per capita features are yet to be generated within our dataset.

To address this gap, we employed an approach utilizing both the census and GRDP datasets. As illustrated in Figure 3, the initial observations in the census dataset provide population data for the years 2010, 2015, and 2020. However, a complete analysis requires data spanning the years 2010 through 2022. To achieve this, we undertook the creation of new columns corresponding to the absent years (2011-2014, 2016-2019, and 2021-2022). Subsequently, the values for these years were set as NaN, to temporarily fill in the missing values as seen in Figure 4.

	Geographic Location	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0	..NATIONAL CAPITAL REGION (NCR)	11855975	NaN	NaN	NaN	NaN	12877253	NaN	NaN	NaN	NaN	13484462	NaN	NaN
1	..CORDILLERA ADMINISTRATIVE REGION (CAR)	1616867	NaN	NaN	NaN	NaN	1722006	NaN	NaN	NaN	NaN	1797660	NaN	NaN
2	..REGION I (ILOCOS REGION)	4748372	NaN	NaN	NaN	NaN	5026128	NaN	NaN	NaN	NaN	5301139	NaN	NaN
3	..REGION II (CAGAYAN VALLEY)	3229163	NaN	NaN	NaN	NaN	3451410	NaN	NaN	NaN	NaN	3685744	NaN	NaN
4	..REGION III (CENTRAL LUZON)	10137737	NaN	NaN	NaN	NaN	11218177	NaN	NaN	NaN	NaN	12422172	NaN	NaN

Figure 4. Census dataset with complete year columns (2010 - 2022)

To further visualize the patterns within the current census dataset, it should be reshaped into a long table format with each year column, along with its corresponding geographical location and population, represented in a single row. This transformation can be done using the melt method, as shown in Figure 5. After that, the population trends over time in different regions for years 2010, 2015, and 2020 can be visualized via line plot as can be seen in Figure 6.

```python
years = [str(year) for year in range(2010, 2023)]
# Melt the census to transform year columns into observations based on the Geographic Location
melted_census = new_census.melt(id_vars = ['Geographic Location'], value_vars = years, var_name='Year', value_name='Population')
```

Figure 5. Melting the census dataset

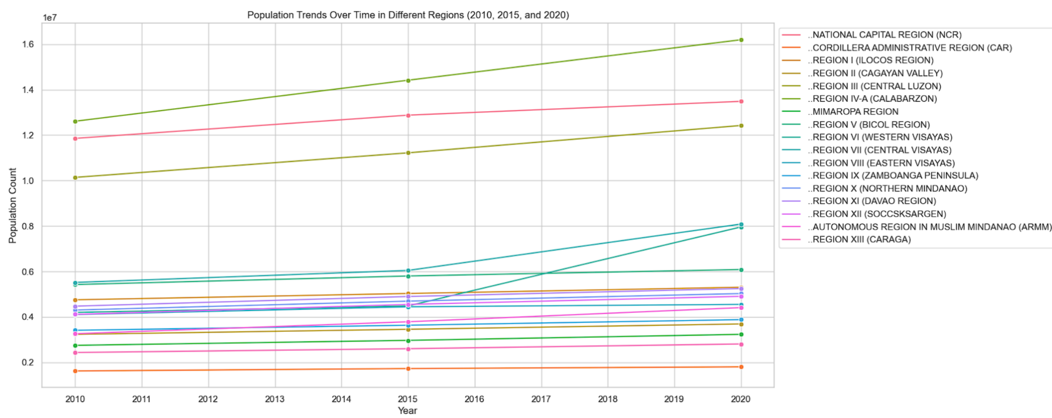


Figure 6. Population Trends Over Time in Different Regions for the years 2010, 2015, and 2020

The line plot illustrates a consistent linear increase in population trends across various regions for the years 2010, 2015, and 2020. This apparent linearity in the data points indicates a viable approach for imputation which is imputation via linear interpolation. This method utilizes non-missing values from neighboring data points to compute accurate estimates for missing data points as seen in Figure 7.

```python
# Convert year column from type object to type int
melted_census['Year'] = melted_census['Year'].astype(int)

# Ensure that the melted_census is sorted by Geographic Location and Year for better interpolation (Will reset this after interpolation)
melted_census = melted_census.sort_values(['Geographic Location', 'Year'])

# Group by 'Geographic Location' and apply linear interpolation to the 'Population' column within each group
melted_census['Population'] = melted_census.groupby('Geographic Location', group_keys=False)['Population'].apply(lambda region: region.interpolate(method='linear'))

# Reset the sorting order
melted_census = melted_census.sort_index()
```

Figure 7. Imputation via Linear interpolation on the melted census dataset

After the imputation of all the missing values from the melted census dataset, the inverse of the melt method, which is the pivot method, will be used to reshape the melted census dataset back to its original form by transforming the years from rows into columns as shown in Figure 8.

```python
# Revert the melted census back to it's original format where years are columns by using pivot
reverted_census = melted_census.pivot(index='Geographic Location', columns='Year', values='Population').reset_index()
# DONE PREPARING THE CENSUS DATASET
```

Figure 8. Reverting the melted census dataset back to its original format via pivoting

Finally, GRDP per capita computation will span the years 2010 to 2022. For each year in this range, the formula employed involves dividing the Gross Regional Domestic Product (GRDP) of the current year by the corresponding population in the respective region and year as demonstrated in Figure 9.

```
# Initialize the years
years = [str(year) for year in range(2010,2023)]

# Loop through the years to calculate GRDP per capita for each region by year
for year in years:
  GRDP_census['GRDP per Capita '+ year] = GRDP_census['At Current Prices ' + year] / GRDP_census['Population_' + year]

GRDP_census.info()
```

Figure 9. GRDP per capita computation for years 2010 - 2022

iii. EDA

1. Average Price of Goods by Region

For the exploratory data analysis, A custom function calculate_yearly_average() which grouped regions and calculated the mean for each year, is applied to all cleaned datasets. Then, they are combined into a single dataset combined_averages using pd.concat() giving a dataset with 120 rows. Regions are grouped again and the mean of each year is calculated as overall_average. A line graph is then plotted to show the yearly average price of all goods by each region.

2. Wholesale Prices vs GRDP and Wholesale Prices vs GRDP per Capita

a list is created for each GRDP column using the code fragment below:

```
# We will use the columns 'At Constant 2018 Prices' from 2010 to 2022
grdp_YearColumns = ['At Constant 2018 Prices ' + str(year) for year in range(2010, 2023)]
```

Figure 10. Generation of list containing annual GRDP values

The list is used in summing the ‘At Constant 2018 Prices’ for all regions by year and stored in aggregated_grdp. Another dataframe is created for a single average price per year for all regions. This is done by taking the mean of overall_average and storing it to aggregated_wholesale_prices. Then, the aggregated_grdp and aggregated_wholesale_prices are combined into a single dataframe, combined_data. From there, the combined data is plotted on a scatterplot. The correlation coefficient is also calculated using .corr().

The same method is also used for Wholesale Prices vs. GRDP per Capita but substituting datasets specific to GRDP per Capita, e.g. aggregated_grdp = aggregated_grdpPerCapita

3. Changes in Average Price of Goods

A graph is also created to show the change in the average price of goods. Each average dataset (average_BeanAndLegumes, etc.) is plotted on a line graph with plt.plot(); x being the year and y being the mean.

iv. Regression Analysis

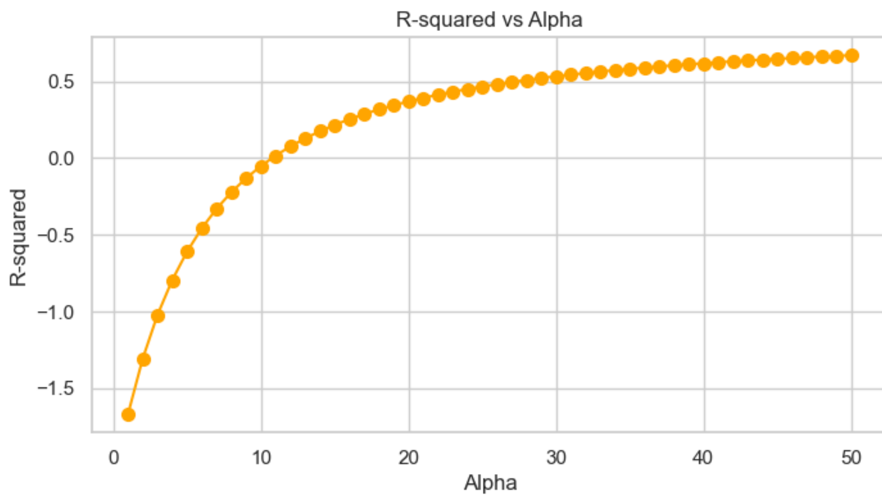
In order to better understand how to most effectively implement the linear, Ridge, and Lasso regression models on the data set, the models were first applied on a smaller selection of the data set. In particular, these models were used on three different good types each paired with a region of the Philippines. These pairings were: NCR with Beans and Legumes, Region VII (Central Visayas) with Seafood, and Region XI (Davao Region) with Grains

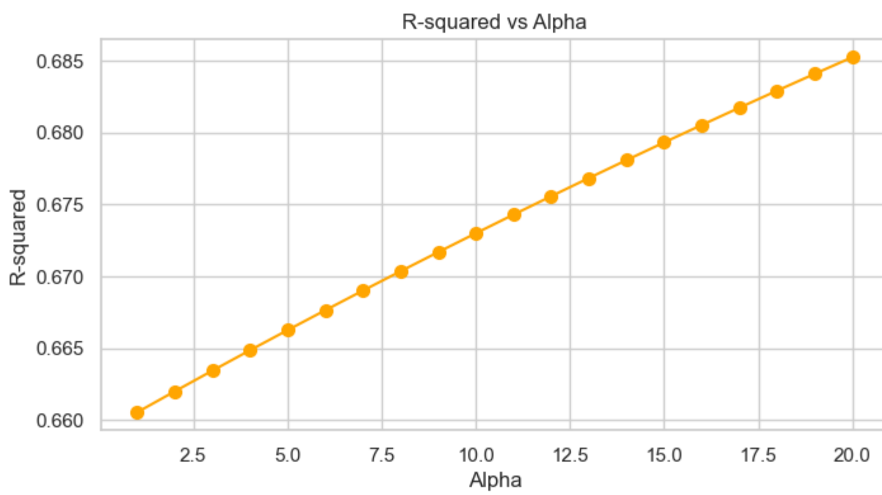
Through their implementation in scikit-learn, both the Ridge and Lasso regression models include an argument "alpha" that may be specified. This is essentially a tuning parameter which dictates how much the model takes into account variables with smaller correlation coefficients. A higher alpha results in a simpler model more focused on the best indicators. Once the data was formatted for the regression models, a loop was set up to run through values of alpha between 1-50 in order to understand the behavior of the models and how their R^2 and MSE scores change with alpha.

```
# Initialize alpha values from 1 to 50
alphas = np.arange(1, 51)
coefficients = []
r2_values = []
mse_values = []

# Fit Ridge Regression for each alpha value
for alpha in alphas:
    ridge_model = Ridge(alpha=alpha)
    ridge_model.fit(features_train, target_train)

    # Coefficients
    coefficients.append(ridge_model.coef_)

    # R-squared
    predictions = ridge_model.predict(features_test)
    r2_values.append(r2_score(target_test, predictions))

    # Mean Squared Error
    mse_values.append(mean_squared_error(target_test, predictions))
```

Figure 11. Loop implementing regression models while iterating through values of alpha

Plots of these would also be generated for ease of interpretation. These plots would reveal several key factors. Firstly, the manner in which the R^2 scores would change in response to the varying alpha would be different from model to model. On some models, these two variables would have a logarithmic relationship while in others, a linear relationship would be observed. Yet sometimes, the relationship would be some other non-elementary function with different local minima and maxima. Despite this, it was found that the most accurate models generally occurred for values of alpha between 10-20. As such for the total regional regression, an alpha of 15 was chosen as a generally good middle ground.





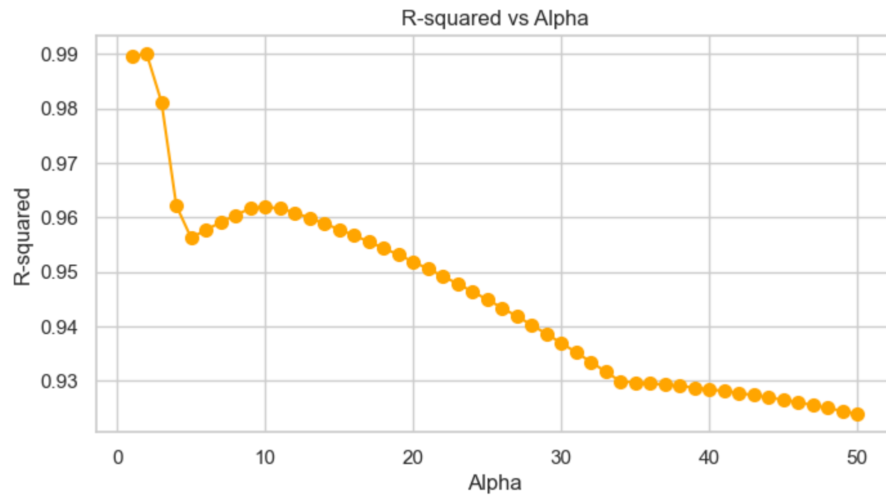
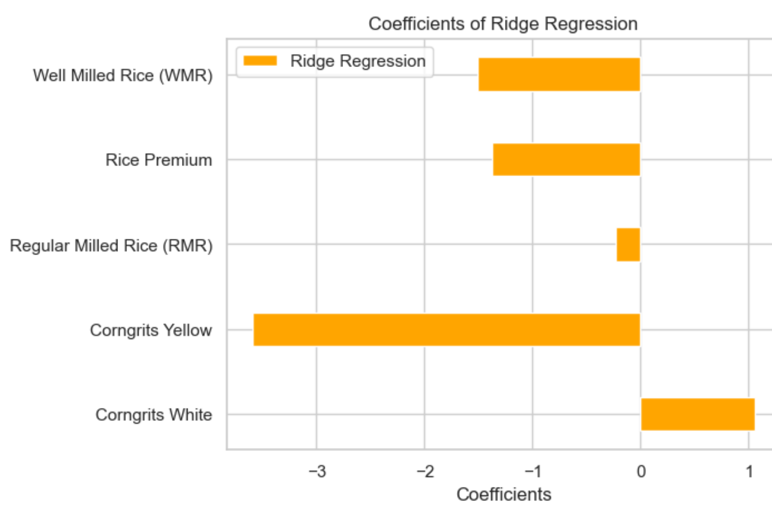


Figure 12-14. Various behaviors of R^2 with changing alpha values

Once the ideal value of alpha per test case model was found, the regression was done for that value with the correlation coefficients of each good being plotted as a bar plot. This gave a rough estimate of how the total model would behave along with some insights on possible variables that may be expected to have a larger influence on the GRDP per capita of a region.



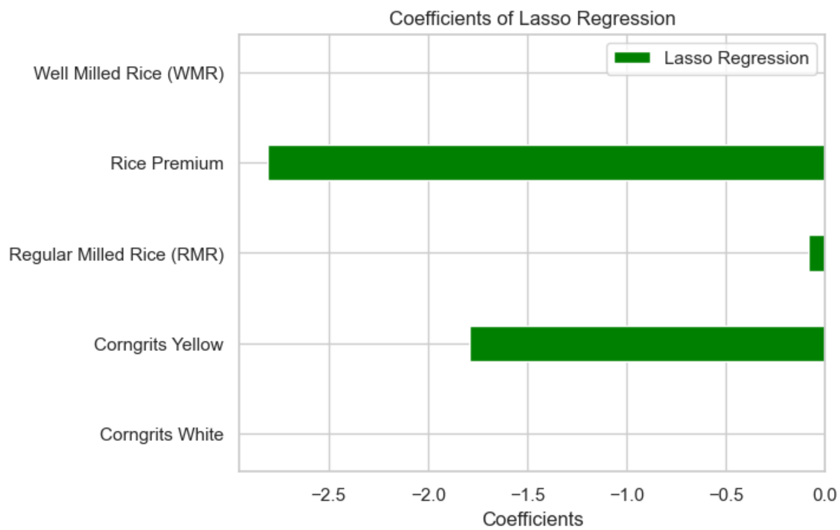


Figure 15-16. Bar plots of coefficient values for Ridge and Lasso regressions

After the tuning process, certain variables and functions were prespecified and predefined as they would be used in the general regression by region. In particular, a custom bar plot function was created which would plot the absolute value of coefficients. Additionally, if the correlation coefficient were positive, its color would be blue, however if the value were negative, its color would be red along with adding a "--" to the good's label.

```
# Function to plot horizontal barplot with sign change indication and color differentiation
def plot_horizontal_barplot(coefficients, title):
    # Sort coefficients by absolute value in descending order
    sorted_coeffs = coefficients.abs().sort_values(ascending=False)
    top_coeffs = sorted_coeffs.head(top)

    # Check if the sign of the coefficient changes
    changed_sign = ['--' + label if coefficients[label] < 0 else label for label in top_coeffs.index]

    # Set colors based on the sign of the coefficient
    colors = ['red' if coefficients[label] < 0 else 'blue' for label in top_coeffs.index]

    # Plotting
    plt.barh(changed_sign, top_coeffs, color=colors)
    plt.xlabel('Absolute Coefficient Value')
    plt.title(title)
    plt.show()
```

Figure 17. Custom function for plotting coefficients as colored barplot

The next section of code would create a large dataframe containing all wholesale goods data within a region along with the GRDP per capita within the region. The three regression models were then run on the data with variable tracking results like the values of the coefficients and R^2 values per model. The plotting of results was also done with the aforementioned horizontal bar plot function. This process was looped throughout every region contained within the dataset in order to run the regressions per region. A summary of the results after all the regressions had completed.

VI. Results and Conclusions

a. Trends by Region
b. Trends by Wholesale Good
c. Suitable Models

Amongst all the regions, every model was the best at least once. Summing these up based on their R^2 scores, linear regression was best 1 time, Ridge was best 7 times, and Lasso was best the most times at 9 instances. Firstly, this suggests that a simple linear regression is generally unsuitable for this task as there are too many variables to consider and as such, the more complicated model that arises from linear regression is not generalizable to most of the data.

Ridge regression was shown to be more prevalent in Mindanao and in less populated regions of the Visayas. This suggests that there are moderate numbers of influential variables in the regions though not enough to require Lasso regression and as such, Ridge regression provides the most appropriate regression model.

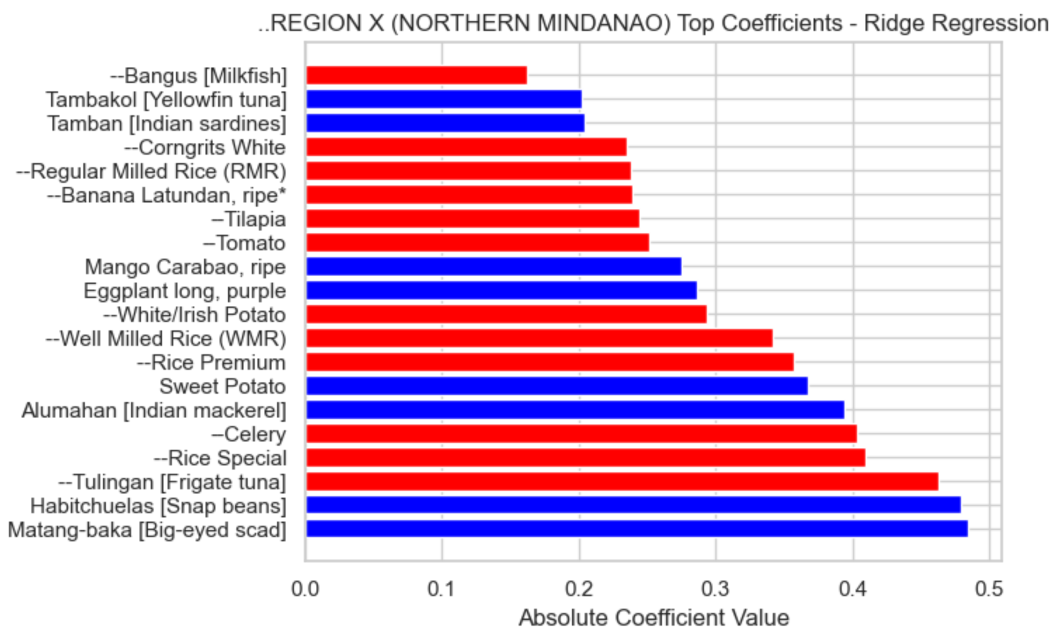


Figure 18. Ridge regression coefficients of Region X

Finally, Lasso regression is overwhelmingly prevalent in Luzon and the more populated regions of the Visayas. This points to the fact that Luzon has access to a larger pool of wholesale goods that other regions do not have. As such, the Lasso model is ideal for this situation as it is able to simplify the more complicated nature of these regions into their most crucial correlators.

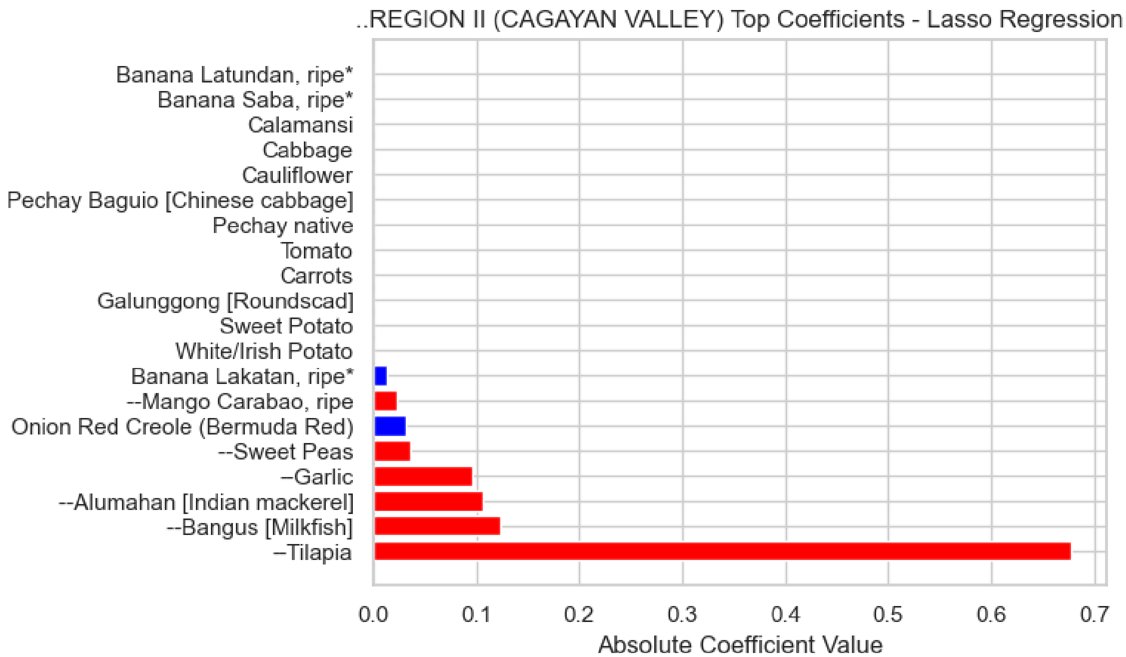


Figure 19. Lasso regression coefficients of Region II

d. Best Indicators

The general analysis reveals significant regional variations in GRDP indicators, highlighting the diverse economic landscapes within the country. The need for employing all three regression models—linear, lasso, and elastic net—emerges as crucial for accurately describing each region due to these inherent differences.

The exploration of the best indicators reveals key guidelines for subsequent regressions. Notably, basic goods such as grains and vegetables consistently prove to be effective indicators of GRDP per capita. Additionally, more expensive seafood contributes significantly to describing economic indicators. The identified alpha values, ranging between 10 and 20, shed light on the appropriate degree of simplification for regression models, indicating a balance between complexity and accuracy.

In examining regional trends, the National Capital Region (NCR) stood out, showcasing a wealth-driven economy where various goods serve as robust indicators of GRDP per capita. In contrast, the Cordillera Administrative Region (CAR) demonstrated the significance of vegetable prices in reflecting economic trends. Region I exhibited positive correlations with fruit prices but negative correlations with vegetables, illustrating the unique economic dynamics in each locale. Interestingly, the correlation analyses revealed that, in rural regions, particularly for grains and cereals like rice, lower prices were consistently associated with higher GRDP per capita. This suggests the economic importance of these staple commodities in rural households, influencing both consumption patterns and economic activities.

The best indicators for GRDP per capita were identified, emphasizing the importance of selecting an appropriate alpha value in regression models. Basic goods, such as grains and vegetables, consistently proved effective indicators, while more expensive seafood also contributed significantly. The regional variations in correlators highlighted the complexity of economic influences, reinforcing the need for region-specific analyses.

SUMMARY OF FINDINGS TABLE:

Region	Model	R^2	Indicators
NCR	Lasso	0.81	Seafood, Fruit, Beans, Vegetables
CAR	Ridge	0.77	Vegetables, Root Crops, Beans
REGION I	Lasso	0.97	Fruit, Root Crops, Seafood, Beans, Grains
REGION II	Lasso	0.98	Seafood, Beans, Vegetables, Fruit
REGION III	Lasso	0.64	Seafood
REGION IV-A	Lasso	0.71	Seafood, Fruit, Condiments
REGION IV-B	Lasso	0.83	Seafood
REGION V	Lasso	0.71	Root Crops, Fruit, Vegetables, Seafood
REGION VI	Ridge	-2.07	Inaccurate models
REGION VII	Ridge	0.97	Beans, Seafood, Fruit, Root Crops, Condiments
REGION VIII	Linear	0.85	Seafood, Grains
REGION IX	Lasso	0.71	w3zeSeafood
REGION X	Ridge	0.90	Seafood, Beans, Grains, Vegetables, Root Crops, Fruit
REGION XI	Ridge	0.93	Grains, Seafood, Root Crops, Vegetables
REGION XII	Ridge	0.91	Rice and fruit based correlations
REGION XIII	Ridge	0.95	Vegetables, Grains, Fruit
ARMM	Lasso	0.90	Grains

However, it is crucial to acknowledge the limitations and areas for improvement. The absence of data for certain regions, notably Regions VII and XII, poses a challenge to the comprehensiveness of the study. Future analyses should prioritize completing the dataset, especially for regions in Visayas and Mindanao. Moreover, the relatively low R^2 values for Region VI suggest potential inaccuracies in correlations drawn from the model, indicating the need for caution in interpretation.

The study suggests avenues for future research. The unique combinations of correlators in each region underscore the need for alternative approaches, such as multi-output regression or clustering methods, to better capture regional variations. The variations in economic activities, environmental conditions, and human development factors across regions call for a robustness check in different parts of the Philippines before generalizing the findings. Moreover, the study highlights the potential for further regional-level analysis, exploring the effectiveness of economic development policies and identifying sectors with growth potential in each region.

VII. References

Abrons, R. (2017, November 21). Farm Marketing Wholesale Direct vs. wholesale distribution. Small Business - Chron.com. https://smallbusiness.chron.com/farm-marketing-wholesale-direct-vs-wholesale-distribution-25030.html

Cho, S.-H. (2020, April). Retailer buyer power does influence wholesale market price, new study finds. Tepper School of Business - Carnegie Mellon University.

Fernando, J. (2023, August 11). *Consumer Price Index (CPI) Explained: What It Is and How It's Used*. Investopedia. https://www.investopedia.com/terms/c/consumerpriceindex.asp

How to adjust for inflation in monetary data sets. (2021, September 29). Time Series Analysis, Regression, and Forecasting. https://timeseriesreasoning.com/contents/inflation-adjustment/#:~:text=The%20formula%20for%20inflation%20adjustment,multiplying%20the%20result%20by%20100.&text=This%20is%20an%20important%20formula.

Igual, L., & Seguí, S. (2017). *Introduction to data Science: A Python Approach to Concepts, Techniques and Applications*. Springer.

Manuel, L. (2022). 9.3 Role of Wholesale Intermediaries . Global Marketing In a Digital World.

Moody's Analytics. (n.d.). Philippines - wholesale price index. Philippines Wholesale Price Index. https://www.economy.com/philippines/wholesale-price-index

Regression analysis - Encyclopedia of Mathematics. (n.d.). https://encyclopediaofmath.org/index.php?title=Regression_analysis