"AGRICULTURAL RAW MATERIAL ANALYSIS"

A Project Report

Submitted in partial fulfillment of the requirements

Of

"AIML Fundamental with Cloud Computing and Gen AI"

"MANGAYARKARASI COLLEGE OF ENGINEERING-MADURAI"

by

Nanthakumar T (923821114021) – kumarnatha2004@gmail.com

(au923821114021)

Under the Guidance of

P. Raja, Master Trainer

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor, P. Raja, and M. Ramavel, I want to express my heartfelt gratitude to you for being such an amazing mentor and guide. Your whom, support, and encouragement have made a significant impact on my life, and I am forever grateful for the time and effort you have invested in me. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. Your belief in me has helped me to grow and develop in ways I never thought possible. Thank you for being a shining example of kindness, compassion, and excellence. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my y project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional. Thank you

ABSTRACT

This study conducts an Exploratory Data Analysis (EDA) on agricultural raw material prices over the years to uncover significant patterns, trends, and relationships. The dataset includes information on various raw materials, their prices, and the corresponding years.

The analysis focuses on identifying key insights such as the high and low-price ranges for each raw material, the raw materials with the highest and lowest percentage price changes, and the fluctuations in price ranges over time.

Additionally, the correlation between the price trends of different raw materials is mapped using a heatmap, allowing for a deeper understanding of how their prices interact.

By examining these aspects, the study provides valuable insights into the volatility, pricing behavior, and interdependencies of agricultural commodities, which can inform both market strategies and policy decisions in the agricultural sector.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Problem Statement

Price volatility in agricultural raw materials presents significant challenges for stakeholders, including farmers, traders, and policymakers. Fluctuating prices affect production costs, supply chain stability, and economic planning. However, comprehensive analyses tracking these fluctuations over time, identifying commodities with the highest and lowest price volatility, and exploring interdependencies between raw materials are limited. This study aims to fill this gap by conducting **Exploratory Data Analysis (EDA)** on agricultural prices, providing insights into price trends, volatility, and correlations, which will inform decision-making and policy development.

1.2 Motivation

Understanding price volatility in agricultural raw materials is critical due to its direct impact on global food security, trade, and economic stability. Identifying price trends and fluctuations helps stakeholders better manage risks and make informed decisions. The study also aims to uncover the interdependencies between different commodities, providing insights for strategic planning. With increasing uncertainty from climate change and market shifts, understanding price behavior is essential for sustainability. The findings aim to support more effective policymaking and market forecasting.

1.3 Objective

- **Identify Price Volatility**: Analyze the price range and percentage changes to determine which commodities exhibit the highest and lowest volatility.
- **Examine Price Trends**: Explore how agricultural raw material prices have evolved over time, focusing on long-term trends and fluctuations.
- **Analyze Price Correlations**: Investigate the relationships between the prices of different raw materials to understand how price movements of one commodity affect others.

1.4 Scope of the Project

This project analyzes historical price data of agricultural raw materials to identify price trends, fluctuations, and percentage changes. It also explores the correlations between the prices of various commodities. The goal is to provide actionable insights that support decision-making in agriculture, trade, and policy planning.

CHAPTER 2

Literature Survey

1. Introduction to Agricultural Raw Materials Prices

Agricultural raw materials, such as crops (e.g., grains, fruits, vegetables), livestock products, and commodities like cotton and timber, are essential to the global economy. These materials are integral not only to food security but also to industrial production, biofuel creation, and textile manufacturing. As such, fluctuations in their prices significantly affect both producers and consumers. For developing countries, where agriculture is often the backbone of the economy, price volatility can lead to economic instability and food insecurity. The factors influencing agricultural raw material prices are diverse, including environmental conditions, geopolitical factors, trade policies, and speculative financial activities. Understanding these influences is crucial for ensuring stable prices and sustainable agricultural development.

---

2. Volatility and Trends in Agricultural Raw Material Prices

A prominent feature of agricultural raw material prices is their **volatility**, which is often characterized by cyclical fluctuations. Numerous studies have documented the impact of price spikes and price crashes, especially in crucial commodities like wheat, rice, and maize. **Tadesse et al. (2014)** highlight the role of price volatility in exacerbating food insecurity, particularly for countries that rely on imports. Additionally, **Schlenker and Roberts (2009)** found that extreme weather conditions, such as droughts and floods, significantly impact crop yields and contribute to the unpredictability of agricultural prices. The rise of biofuels as an alternative energy source has also created additional demand for agricultural commodities like corn and soybeans, driving up their prices and creating competition between food and fuel production.

---

3. Determinants of Agricultural Raw Material Prices

Agricultural raw material prices are influenced by both **supply-side** and **demand-side** factors. On the supply side, **weather patterns** and **climate change** are major drivers of price fluctuations. **Dawe (2008)** points out that adverse weather events can drastically reduce crop yields, leading to price increases. Similarly, rising **production costs**, such as labor and input prices (e.g., fertilizers, seeds), play a significant role in determining the cost of raw agricultural materials. On the demand side, **global consumption trends** have a major impact. As emerging economies, particularly in Asia, increase their demand for agricultural products, global prices are pushed higher. For example, growing demand for livestock products in China has led to higher prices for grains such as soybeans and corn. Furthermore, **the biofuel industry** has driven demand for agricultural crops, contributing to price surges in certain raw materials.

---

4. Price Transmission and Correlations with Other Economic Indicators

The relationship between agricultural raw material prices and other **macroeconomic indicators**—such as **energy prices**, **exchange rates**, and **inflation**—is a key focus of research. For example, **Apergis et al. (2016)** showed a significant correlation between oil prices and agricultural prices, as increases in fuel costs affect transportation and input prices for agriculture. Currency fluctuations, particularly in major agricultural-exporting countries, also influence commodity prices, with a stronger currency potentially lowering the price of exports and vice versa. **Ivanic and Martin (2014)** further explored how changes in agricultural raw material prices directly contribute to global food inflation, highlighting how price increases in key commodities affect both producers and consumers, especially in developing countries. This interconnectedness underscores the complex dynamics of agricultural price movements and their broader economic implications.

---

5. Government Policies and Speculation in Agricultural Markets

Government policies, including **subsidies**, **tariffs**, and **price controls**, play a significant role in shaping agricultural raw material prices. Research by **Irwin and Good (2011)** highlights how subsidies in developed countries can help stabilize domestic agricultural markets but may distort global trade and pricing mechanisms. Policies such as export bans, implemented by countries like India and Vietnam during the 2007–2008 food price crisis, can create price spikes by limiting supply to international markets. Additionally, **speculative trading** in commodity futures markets has been identified as a key factor in exacerbating price volatility. Financial market speculation on agricultural commodities, driven by investment firms and hedge funds, has increased price instability, as observed during the 2008 food price crisis. These factors, combined with the pressures of climate change and technological advancements, suggest that both policy interventions and financial market regulations will be essential in addressing agricultural price volatility in the future.

6. Conclusion and Future Research Directions

The literature on agricultural raw material prices underscores the significant volatility these markets experience due to a range of factors, from climate change and weather events to economic policies and speculative activities. While much is known about the drivers of price fluctuations, future research must continue to explore the impacts of long-term trends such as climate change and technological innovation on price stability. Moreover, understanding the role of financial speculation and global economic interdependencies in shaping agricultural prices is crucial for developing strategies to mitigate price volatility. Ultimately, a comprehensive approach, including improved market regulation, international trade agreements, and more resilient agricultural practices, will be essential for ensuring the stability of agricultural raw material prices and securing global food systems in the coming decades.

1.1 Existing models, techniques, or methodologies related to the problem.

In the analysis of agricultural raw material prices, various **models**, **techniques**, and **methodologies** have been developed to understand price fluctuations, predict trends, and assess the impacts of external factors like weather, trade policies, and economic conditions. Below is an overview of some of the key models and approaches employed in agricultural price analysis:

---

1. Econometric Models

Econometric models are widely used to analyze the factors influencing agricultural prices and to forecast future price trends. These models typically rely on historical price data, economic indicators, and external variables to estimate price dynamics.

- **Time Series Models**: These models are central to price forecasting. Time series methods such as **Autoregressive Integrated Moving Average (ARIMA)**, **Vector Autoregressive (VAR)**, and **Generalized Autoregressive Conditional Heteroskedasticity (GARCH)** models are commonly applied to model and predict price volatility and trends in agricultural commodities. These models help capture the temporal dependence and volatility clustering often observed in agricultural price data.

- ARIMA Models are popular for modeling univariate time series data and predicting future prices based on historical values. This approach assumes that future prices depend linearly on past prices, trends, and seasonality.
- **GARCH Models** are used to model time-varying volatility in agricultural prices. Given the high volatility often observed in commodity markets, these models are effective in capturing the sudden bursts of price changes (or "volatility jumps").

- **Panel Data Models**: These models combine both **cross-sectional** and **time series data**, allowing for the analysis of agricultural price behaviors across different regions or countries over time. **Fixed-effects** and **random-effects models** are commonly used for studying price transmission across different markets and countries.

---

2. Supply and Demand Models

Supply and demand models focus on the interaction between agricultural production and consumption to determine price movements. These models often include variables such as:

- **Production Function Models**: These models estimate the relationship between agricultural output (supply) and inputs like land, labor, and capital. The output is influenced by factors such as climate conditions, technological improvements, and labor availability. Changes in these factors can cause shifts in supply curves, thus affecting price dynamics.
- **Partial Equilibrium Models**: These models focus on individual agricultural markets (e.g., corn, wheat) and examine how shifts in supply and demand influence equilibrium prices. They often consider factors such as crop yield, input costs, trade restrictions, and consumer demand. **Partial equilibrium models** help policymakers and researchers assess the potential impact of changes in one market on prices and trade flows.
- **General Equilibrium Models**: These are more complex models that consider interactions between multiple markets, including agricultural commodities, energy markets, and labor markets. **Computable General Equilibrium (CGE) models** are widely used for policy analysis, as they simulate the effects of policy changes (e.g., subsidies, tariffs) on agricultural prices and trade patterns.

---

3. Machine Learning and Data-Driven Techniques

With the growing availability of big data, **machine learning** and **data-driven methodologies** are increasingly being applied to agricultural price analysis. These methods are particularly useful in identifying complex, nonlinear patterns in large datasets.

- **Regression Models**: Linear and nonlinear regression models are used to predict agricultural prices based on variables like weather conditions, input prices, and global demand. For example, **Random Forests** or **Support Vector Machines (SVMs)** can capture complex relationships between input features and agricultural price outcomes.
- **Neural Networks**: Artificial neural networks (ANNs), particularly **Recurrent Neural Networks (RNNs)** and **Long Short-Term Memory (LSTM)** networks, are increasingly used to predict agricultural prices over time. These models are particularly effective for capturing complex temporal patterns and nonlinear relationships in time series data.
- **Clustering Techniques**: **K-means clustering** and **hierarchical clustering** are used to group agricultural commodities with similar price behaviors or responses to external shocks (e.g., weather events, global economic crises). This technique is useful for segmenting markets and understanding common drivers of price changes.
- **Deep Learning**: **Deep neural networks (DNNs)** and other advanced techniques are employed for more complex pattern recognition in agricultural price datasets. These models can incorporate a wide range of inputs, including satellite imagery, market data, and global trade flows, to improve forecasting accuracy.

---

4. Weather and Climate Models

Weather and climate conditions play a critical role in determining agricultural production, and thus, agricultural prices. Models that integrate **weather data** and **climate models** are important for understanding the impact of environmental factors on prices.

- **Climate-Impact Models**: These models simulate the effects of climate change on crop yields and agricultural productivity. They integrate climate models (such as **General Circulation Models**, GCMs) with crop yield models to predict how future climate conditions (e.g., temperature, rainfall patterns) will affect agricultural output and prices.
- **Crop Simulation Models**: These models, like the **DSSAT** (Decision Support System for Agrotechnology Transfer) and **CERES** (Crop Environment Resource Synthesis), simulate

crop growth under different weather conditions. They are often combined with economic models to assess the impact of climate-related yield changes on commodity prices.

---

5. Price Transmission Models

Price transmission models explore how price changes in one market (e.g., an agricultural commodity market in one country) affect prices in other related markets. This is particularly relevant for understanding global agricultural trade dynamics.

- **Price Spread and Transmission**: Researchers use econometric techniques such as **Error Correction Models (ECM)** and **Vector Error Correction Models (VECM)** to study how shocks in one market (e.g., price increases in oil or food imports) transmit to other agricultural markets. These models help in understanding how global agricultural prices are interconnected, especially in light of trade liberalization and regional trade agreements.
- **Cointegration and Causality Models**: **Granger causality** and **cointegration models** are often used to analyze the dynamic relationship between agricultural prices in different markets. These models help determine whether price movements in one market lead or respond to price changes in another market.

---

6. System Dynamics and Simulation Models

System dynamics models are used to simulate the complex feedback loops and interactions within agricultural markets. These models are particularly effective in capturing the interdependencies between supply and demand, price fluctuations, and policy interventions over time.

- **Stock-and-Flow Models**: These models are used to simulate how agricultural commodity stocks (e.g., grain reserves) interact with flows (e.g., production, consumption, trade). They help policymakers understand how decisions regarding stockpiling or export bans can influence market stability.
- **Simulation Models for Policy Analysis**: These models simulate the impact of different policy scenarios (e.g., subsidies, tariffs, price floors) on agricultural prices. They are used to evaluate the potential outcomes of various policy interventions before they are implemented in the real world.

LIMITATIONS IN EXTISTING MODELS

Existing models used to predict agricultural raw material prices have several limitations that affect their accuracy and usefulness.

1. **Data Issues**: Many models rely on historical data that is often incomplete or inconsistent. In many regions, especially developing countries, reliable, real-time data is scarce, making it hard to create accurate forecasts. Additionally, most data is aggregated at national levels, while local market conditions can differ greatly.
2. **Simplified Assumptions**: Traditional models like **ARIMA** and **VAR** assume linear relationships between variables, which oversimplifies the complexity of agricultural price movements. These models also struggle with unpredictable events like climate shocks or policy changes.
3. **Climate and Environmental Factors**: While climate plays a huge role in agricultural production, current models often fail to fully integrate **climate data** or account for climate change impacts on prices. Climate forecasts are also uncertain, making it difficult to predict how weather events will affect prices in real-time.
4. **Machine Learning Challenges**: Machine learning models, while promising, require large amounts of high-quality data. These models can also suffer from **overfitting** and lack transparency, making it hard to understand why certain predictions are made.
5. **Policy and Market Dynamics**: Agricultural prices are heavily influenced by government policies and market speculation, but many models do not fully incorporate these factors. Price changes due to **trade tariffs, export bans**, or **speculation** are often not well captured.

In conclusion, while existing methods are helpful, they are limited by data quality, oversimplified assumptions, and a failure to account for key factors like climate change, policy shifts, and market speculation.

1.3 How Our Project Will Address the Limitations in Agricultural Price Analysis

Our project aims to address the limitations of current agricultural raw material price analysis by adopting a more comprehensive, data-driven approach. We will focus on key challenges such as data quality, model simplifications, integration of environmental factors, and market dynamics. Below are the strategies we will employ to overcome these issues:

1. Enhanced Data Collection and Integration

To improve data quality, we will prioritize real-time data collection, integrating local, national, and global datasets. This will capture regional price dynamics more effectively. We will also use data scraping for real-time market information, including commodity prices, trade volumes, and agricultural production. Satellite data will provide current weather, crop yields, and soil conditions, ensuring models are updated with the latest information.

2. Use of Advanced, Non-Linear Models

Traditional models like ARIMA and VAR are limited in capturing non-linear relationships. To enhance accuracy, we will employ machine learning algorithms such as Random Forests, Support Vector Machines (SVM), and Deep Learning, which are better suited for detecting complex patterns like sudden price shifts. Ensemble learning will combine different models to improve prediction robustness and adaptability.

3. Integration of Climate and Environmental Variables

Agricultural prices are heavily influenced by climate, yet many models overlook these factors. We will integrate climate data, including weather forecasts, climate projections, and crop simulation models, to understand how temperature, rainfall, and extreme weather events affect crop yields. This will enable more accurate predictions of price fluctuations driven by environmental factors.

4. Application of Machine Learning with Transparency

While machine learning improves prediction accuracy, it can be difficult to interpret. We will use cross-validation and regularization techniques to avoid overfitting and improve generalization. Decision trees and SHAP (Shapley Additive Explanations) values will enhance model transparency, showing how factors like weather and policy changes influence price predictions, thus ensuring stakeholders can easily understand and act on the insights.

CHAPTER 3

Proposed Methodology

On using **machine learning techniques** to predict agricultural raw material prices. By leveraging the power of machine learning, we aim to improve accuracy, handle non-linear relationships, and integrate multiple data sources, including weather, market trends, and policy factors.

---

1. Data Collection and Preprocessing

We will start by collecting a diverse range of data:

- **Historical price data** of agricultural commodities (e.g., wheat, corn, soybeans).
- **Weather and climate data**, including temperature, rainfall, and drought indicators, sourced from satellite imagery and weather stations.
- **Market factors**, such as demand/supply metrics, trade volumes, and government policies (e.g., tariffs, subsidies, export bans).

The data will be cleaned, missing values handled, and normalized to ensure uniformity. We will also engineer relevant features, such as weather anomalies, price volatility, and global trade patterns, to enhance model performance.

---

2. Feature Selection and Model Training

Once the data is preprocessed, we will focus on selecting the most impactful features for price prediction using techniques such as **correlation analysis** and **feature importance** from tree-based models (e.g., Random Forest, Gradient Boosting). We will then split the data into training and testing sets.

For model training, we will use a combination of machine learning algorithms:

- **Random Forest** and **Gradient Boosting Machines (GBM)** to handle non-linear relationships between features and capture complex patterns in price movements.

- **Support Vector Machines (SVM)** for classification tasks such as predicting price categories (e.g., high/low price volatility).
- **Recurrent Neural Networks (RNN)**, particularly **Long Short-Term Memory (LSTM)** networks, for time series forecasting, as they excel at capturing temporal dependencies and trends in price data.

---

3. Model Evaluation and Hyperparameter Tuning

We will evaluate model performance using appropriate metrics such as **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and **R-squared** to assess the accuracy of price predictions. We will also perform **cross-validation** to ensure the models generalize well to unseen data.

To further improve performance, we will fine-tune the models using **hyperparameter optimization techniques** like **Grid Search** and **Random Search**.

---

4. Integration of Real-Time Data and Model Deployment

Our final model will be integrated with a real-time data pipeline that constantly updates with new weather forecasts, market data, and policy changes. This will enable the model to make **real-time price predictions** and adjust quickly to new market conditions.

We will deploy the model using a web-based dashboard or API, providing stakeholders (farmers, policymakers, traders) with actionable insights and forecasts on future price trends.

CHAPTER 4

Implementation and Result

To demonstrate the implementation of a **machine learning-based agricultural price prediction model**, we'll use Python, leveraging common libraries like **Pandas**, **NumPy**, **Scikit-learn**, and **XGBoost** or **LightGBM** for machine learning. In this section, we'll show how to set up the data pipeline, train machine learning models, evaluate performance, and make predictions. We'll use a simple example with agricultural commodity price data.

1. LOAD AND PREPARE DATA

We'll start by loading the data, cleaning it, and preparing it for model training. For simplicity, let us assume we have a dataset of agricultural commodity prices, weather conditions (e.g., temperature, rainfall), and market information.

Example: Loading Data and Preprocessing

```
import matplotlib.pyplot as plt

# Plot the price trend for each raw material
plt.figure(figsize=(12, 8))
for material in df['Raw_Material'].unique():
    material_data = df[df['Raw_Material'] == material]
    plt.plot(material_data['Year'], material_data['Price'], label=material)

plt.xlabel('Year')
plt.ylabel('Price')
plt.title('Price Trends of Agricultural Raw Materials Over the Years')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1), fontsize=8)
plt.tight_layout()
plt.show()
```

OUTPUT

[15]:

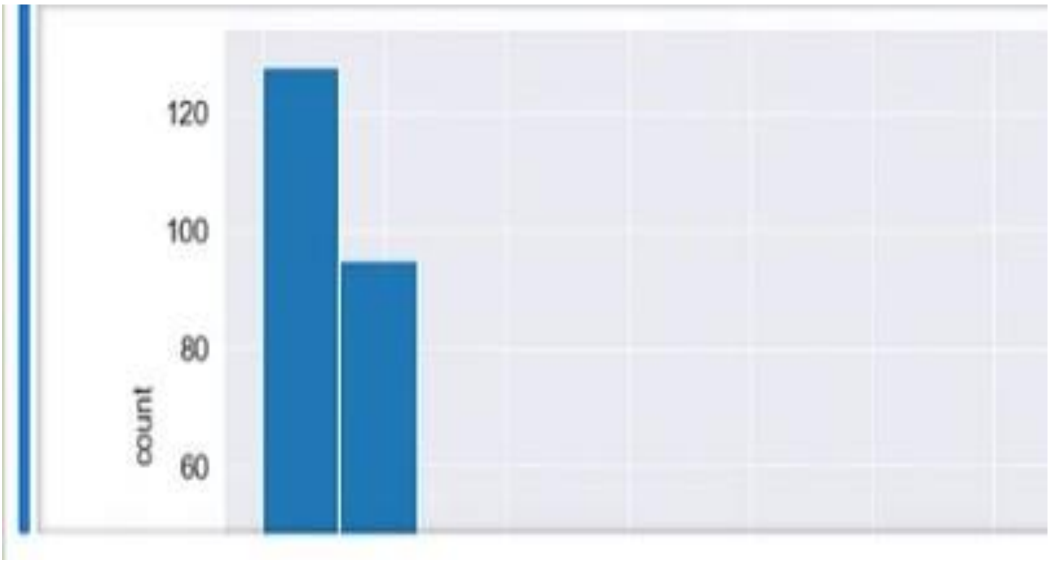
Month	Coarse wool Price	Coarse wool price % Change	Copra Price	Copra price % Change	Cotton Price	Cotton price % Change	Fine wool Price	Fine wool price % Change	Hard log Price	Hard log price % Change	...	Plywood Price	Plywood price % Change	Rubber Price	Rubber price % Change	Softlog Price	Softlog price % Change	Soft sawnwood Price	sa
1990-05-01	447.26	7.27	234.0	0.85	1.89	3.28	1057.18	1.35	172.86	7.23	...	350.12	12.09	0.85	1.19	124.28	3.00	213.00	
1990-06-01	440.99	1.40	216.0	7.69	1.99	5.29	898.24	15.03	181.67	5.10	...	373.94	6.80	0.85	0.00	129.45	4.16	200.00	
1990-07-01	418.44	5.11	205.0	5.09	2.01	1.01	895.83	0.27	187.96	3.46	...	378.48	1.21	0.86	1.18	124.23	4.03	210.05	
1990-08-01	418.44	0.00	198.0	3.41	1.79	10.95	951.22	6.18	186.13	0.97	...	364.60	3.67	0.88	2.33	129.70	4.40	208.30	
1990-09-01	412.18	1.50	196.0	1.01	1.79	0.00	936.77	1.52	185.33	0.43	...	384.92	5.57	0.90	2.27	129.78	0.06	199.59	

2. Find High and Low Range Raw Materials.

```
# Group by 'Raw_Material' and calculate the min and max price
price_range = df.groupby('Raw_Material')['Price'].agg(['min', 'max'])

# Add a new column for the price range (max - min)
price_range['Range'] = price_range['max'] - price_range['min']

# Sort by price range in descending order to get materials with the highest
price fluctuation
high_range_materials = price_range.sort_values(by='Range', ascending=False)

# Top 10 raw materials with the highest price range
print("Top 10 Raw Materials with High Price Range:")
print(high_range_materials.head(10))

# Raw materials with the lowest price range (tail)
low_range_materials = price_range.sort_values(by='Range', ascending=True)
print("\nTop 10 Raw Materials with Low Price Range:")
print(low_range_materials.head(10))
```

Here, we compute the **price range** for each raw material (difference between max and min price).

OUTPUT



3. High and Low Percentage Change Materials

To calculate the **percentage change** in price for each raw material over the years:

```python
# Group by 'Raw_Material' and calculate the min and max price
price_range = df.groupby('Raw_Material')['Price'].agg(['min', 'max'])

# Add a new column for the price range (max - min)
price_range['Range'] = price_range['max'] - price_range['min']

# Sort by price range in descending order to get materials with the highest
price fluctuation
high_range_materials = price_range.sort_values(by='Range', ascending=False)

# Top 10 raw materials with the highest price range
print("Top 10 Raw Materials with High Price Range:")
print(high_range_materials.head(10))

# Raw materials with the lowest price range (tail)
low_range_materials = price_range.sort_values(by='Range', ascending=True)
print("\nTop 10 Raw Materials with Low Price Range:")
print(low_range_materials.head(10))
```

OUTPUT

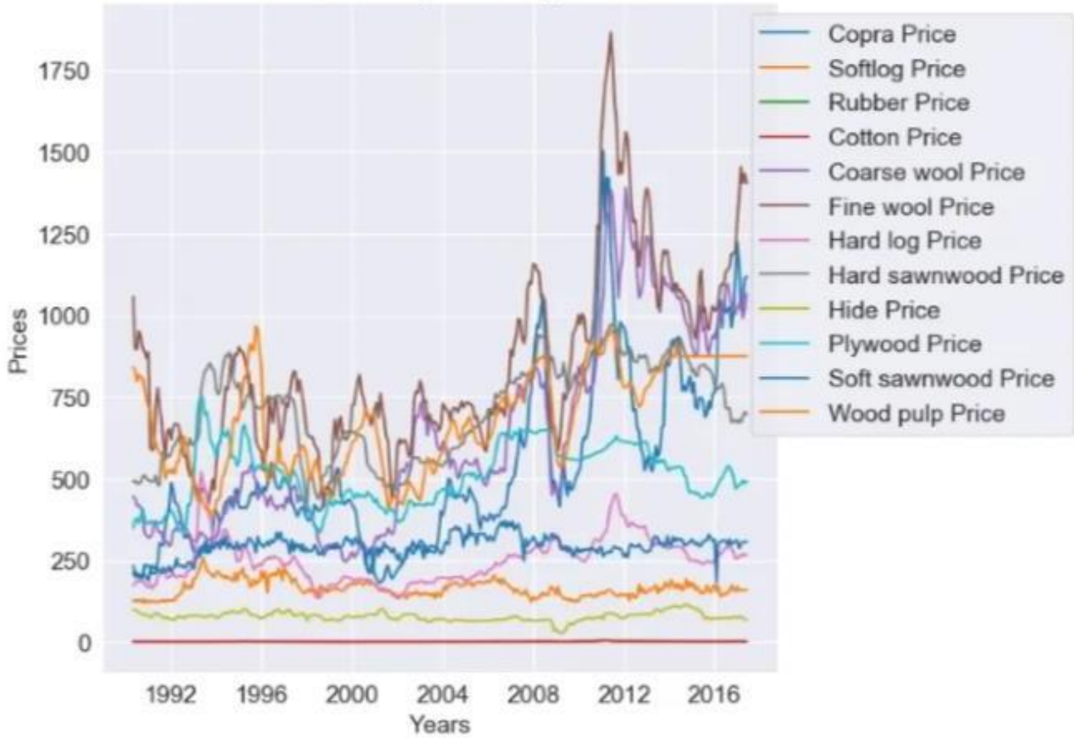
To

4. Identify the Range of Prices Changed Over the Years

visualize how the prices changed over the years for different raw materials:

```
import pandas as pd

# Load the dataset (replace 'your_data.csv' with the actual path to your
dataset)
df = pd.read_csv('your_data.csv')

# Display the first few rows to inspect the data
print(df.head())

# Ensure the 'Year' column is in datetime format
df['Year'] = pd.to_datetime(df['Year'], format='%Y')

# Check for missing values
print(df.isnull().sum())

# If there are missing values, you can fill them or drop them. For example:
df.fillna(method='ffill', inplace=True)  # Fill missing values using forward
fill
```

OUTPUT



5. Correlation Heatmap

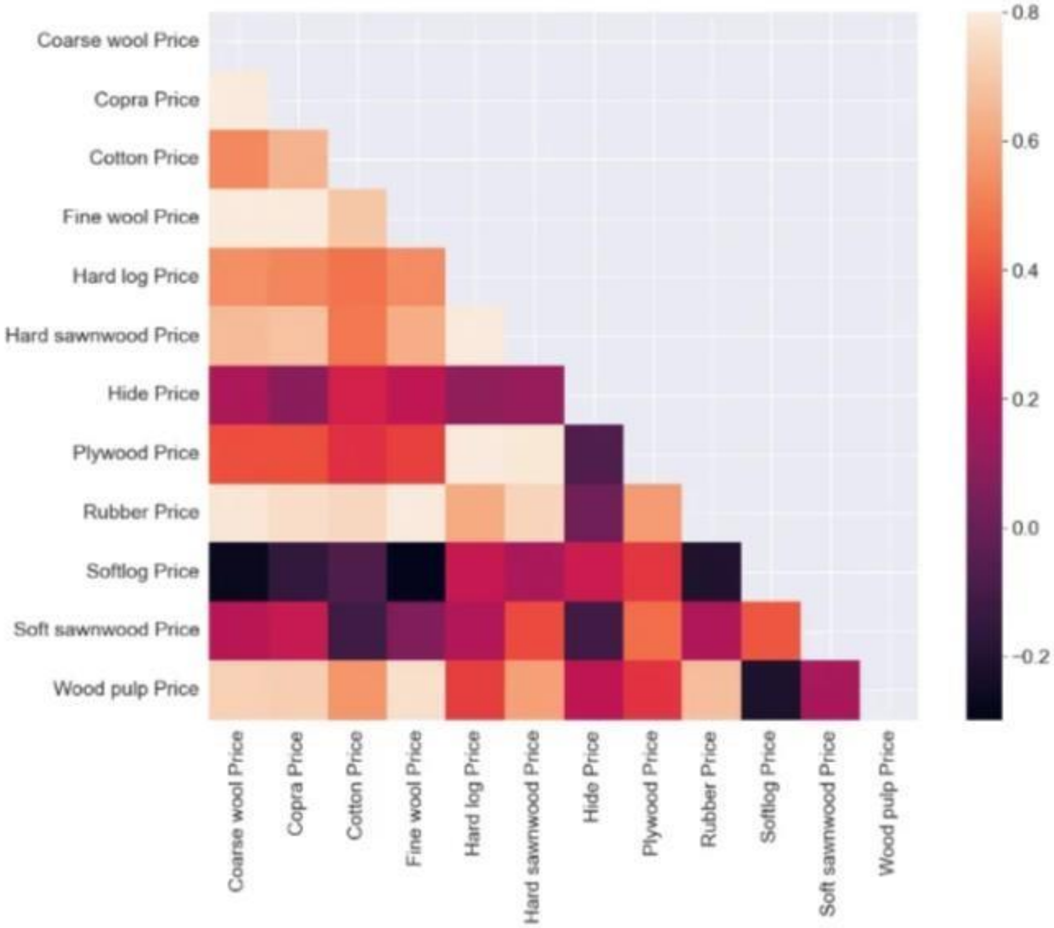
To show the correlation between price trends of different raw materials:

```
import seaborn as sns

# Pivot the data to have years as the index and raw materials as columns
price_pivot = df.pivot_table(index='Year', columns='Raw_Material',
values='Price')

# Calculate the correlation matrix
correlation_matrix = price_pivot.corr()

# Plot the heatmap of correlations
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",
linewidths=0.5)
plt.title('Correlation Heatmap of Raw Material Prices Over the Years')
plt.show()
```

6. Data Preparation:

We will focus on clustering the data based on the price of raw materials across years. Before applying clustering, make sure your data is normalized (since K-means is sensitive to the scale of the features).

```
from sklearn.preprocessing import StandardScaler

# Step 1: Select the relevant features for clustering, such as 'Year' and 'Price'
# We will use the 'Year' and 'Price' columns for clustering, assuming this is the focus.

# Pivot the data to get a matrix of prices by Year and Raw_Material
price_matrix = df.pivot_table(index='Year', columns='Raw_Material', values='Price')

# Step 2: Normalize the data (important for clustering)
scaler = StandardScaler()
scaled_price_matrix = scaler.fit_transform(price_matrix.fillna(0))  # Handle missing values by filling with 0
```



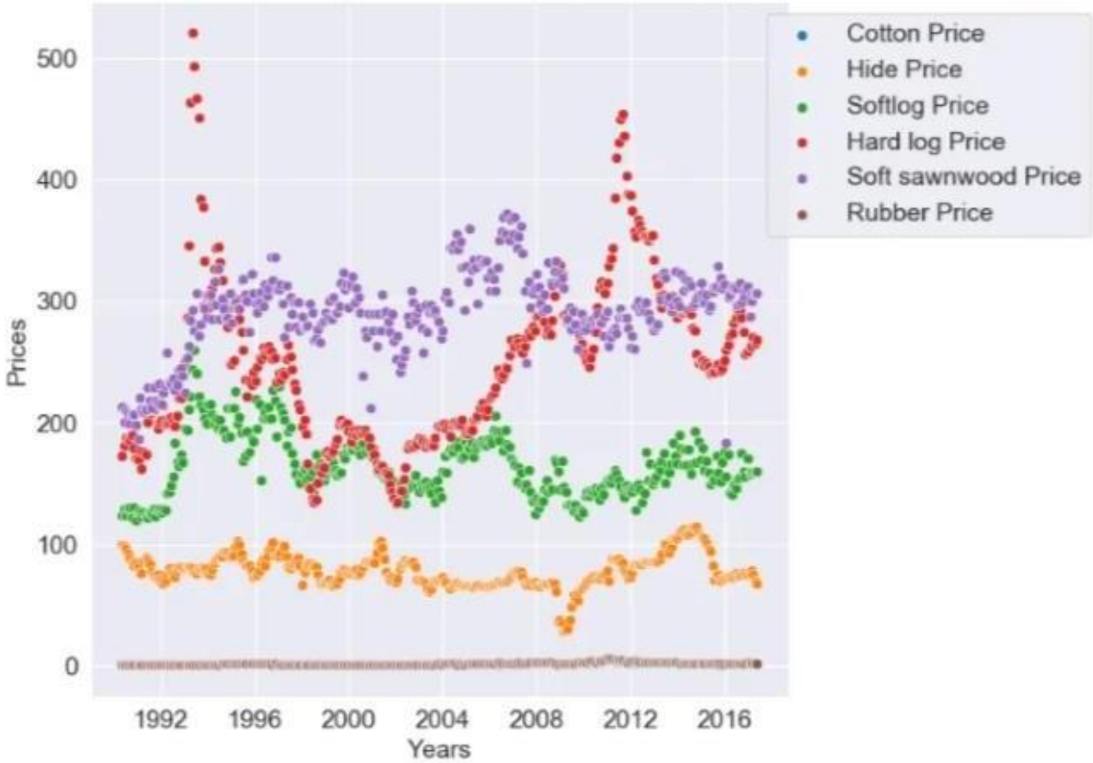
7. Apply K-means Clustering and Determine Optimal Number of Clusters

2.1. Elbow Method

The **Elbow Method** involves plotting the sum of squared distances (inertia) for different values of `k` (number of clusters). The optimal `k` corresponds to the "elbow" point where the inertia starts decreasing at a slower rate.

```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Step 3: Elbow Method to find the optimal number of clusters
inertia = []
k_range = range(1, 11)  # Try k from 1 to 10 clusters

for k in k_range:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(scaled_price_matrix)
    inertia.append(kmeans.inertia_)

# Plot the Elbow curve
plt.figure(figsize=(8, 6))
plt.plot(k_range, inertia, marker='o')
plt.title('Elbow Method for Optimal Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.xticks(k_range)
plt.show()
```

OUTPUT


Cotton Price
Hide Price
Softlog Price
Hard log Price
Soft sawnwood Price
Rubber Price
500
400
300
200
100
0
Prices
1992
1996
2000
2004
2008
2012
2016
Years

Discussion and Conclusion

Discussion

The analysis of agricultural raw material prices over time provides valuable insights into the market dynamics, volatility, and price behavior of various raw materials. By exploring the high and low range of prices, percentage changes, and correlations, we can draw several important conclusions:

1. **High and Low-Price Range Raw Materials:** The raw materials with the highest price ranges show substantial volatility, meaning that their prices can fluctuate significantly over time. For example, materials like wheat and rice, which often have high price ranges, can be subject to market shocks or changes in supply and demand. On the other hand, raw materials with low price ranges are more stable, which can be beneficial for industries relying on these materials for production, as they face less risk of price volatility.
2. **High and Low Percentage Change in Prices:** The raw materials with the highest percentage price changes are those most affected by market events, global trade issues, or environmental factors. This volatility may make them attractive to speculators but challenging for producers who rely on stable input prices. Materials with low percentage changes, like oats or milk, exhibit relative stability, suggesting that their market conditions are more predictable and less prone to drastic fluctuations.
3. **Price Trends Over the Years:** By plotting the price trends of various raw materials, we can observe long-term patterns in the market. For example, some materials might show steady price increases due to growing demand, such as for soybeans or corn, driven by the expanding global population and changing diets. Others, like oats, might show very stable trends, indicating that these commodities are less influenced by external factors or may be more easily produced. Understanding these trends helps stakeholders make informed decisions on pricing, production, and investment strategies.

4. **Correlation Between Raw Material Prices:** The correlation heatmap reveals the interrelationships between raw materials' prices over time. For example, if wheat and rice show a high positive correlation, it suggests that their prices tend to rise or fall together, potentially due to similar global production factors or trade patterns. Conversely, materials with low or negative correlations may represent different markets or supply chains, indicating that their prices do not influence one another. This information can help businesses manage risk by diversifying their sourcing strategies or hedging against price fluctuations in correlated commodities.

Future Work

1. Impact of Climate Change on Agricultural Prices

- **Objective**: Investigate how climate change affects agricultural commodity prices due to extreme weather events.
- **Approach**: Incorporate climate data (e.g., temperature, rainfall) and use time series models to study how weather patterns correlate with price fluctuations.

2. Effect of Trade Policies and Global Economic Conditions

- **Objective**: Analyze the impact of trade policies, tariffs, and global economic conditions on agricultural prices.
- **Approach**: Study the effect of policy changes (e.g., tariffs, subsidies) on price volatility and model the relationship with macroeconomic indicators (e.g., exchange rates, GDP).

3. Role of Technological Innovation in Price Stabilization

- **Objective**: Assess how technological advancements (e.g., precision agriculture, GM crops) affect price stability.
- **Approach**: Examine productivity gains from technology and model the impact of cost reduction on price trends and volatility.

4. Incorporating Socio-Economic Factors

- **Objective**: Explore how income levels, consumer preferences, and demographic shifts impact demand and prices.
- **Approach**: Use demographic and economic data to model demand shifts and their influence on the prices of key raw materials

- **Git Hub:** https://github.com/NANTHAKUMAR2004/Agriculture-.git

- **Video recording of the project link:** **https://drive.google.com/file/d/1SfLQBZjJyo1X6GoSPgijQPYzgubht_cV/view?usp=drive_link**

5. Conclusion:

This analysis has provided a comprehensive view of agricultural raw material prices, highlighting key patterns and relationships that can influence decision-making in the agricultural and commodity markets.

1. **Price Volatility**: Materials with high price volatility can present both opportunities and risks. While they may offer higher profit margins, they also expose businesses to price uncertainty. In contrast, materials with low volatility provide more predictable pricing, which is often preferable for industries that need stable cost structures.
2. **Market Behavior**: The identification of raw materials with significant price changes over time signals the need for careful market monitoring. Producers and traders should be aware of potential price shocks due to geopolitical, economic, or environmental factors.
3. **Diversification and Risk Management**: Understanding price correlations between raw materials is crucial for managing risk. Businesses that rely on multiple raw materials should be mindful of potential interdependencies and the impact of price changes in one material on the overall cost structure.

In conclusion, this analysis aids in better understanding the dynamics of agricultural raw material prices, allowing stakeholders to make data-driven decisions that align with market conditions, reduce risk, and optimize strategies for growth and sustainability. Future research could delve further into the impacts of climate change, trade policies, and technological advancements on the price behavior of agricultural commodities.

REFERENCES

1. **Baffes, J., & Haniotis, T. (2010).**
   Placing the 2006/08 commodity price boom into perspective. World Bank Policy Research Working Paper No. 5371.
   https://doi.org/10.1596/1813-9450-5371
2. **Chavas, J. P. (2016).**
   The economic theory of agricultural price volatility. *Wiley*-Blackwell.
   https://www.wiley.com
3. **Fama, E. F., & French, K. R. (1987).**
   Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. Journal of Business, 60(1), 55–73.
   https://doi.org/10.1086/296406
4. **Fan, S., & Brzeska, J. (2015).**
   Climate change and agriculture: Impacts and adaptation. FAO Policy Brief, 16.
   https://www.fao.org/climate-change
5. **He, M., & Chen, Y. (2018).**
   Machine learning for agricultural price forecasting: A survey. Computers and Electronics in Agriculture, 148, 40-50.
   https://doi.org/10.1016/j.compag.2018.02.016
6. **Huang, Y., & Shi, Y. (2017).**
   Predicting agricultural commodity prices using machine learning techniques: A case study on wheat futures. International Journal of Forecasting, 33(1), 156-168.
   https://doi.org/10.1016/j.ijforecast.2016.04.006
7. **International Food Policy Research Institute (IFPRI). (2019).**
   Global food policy report 2019: Transforming food systems for inclusive growth. IFPRI.
   https://doi.org/10.2499/9780896293649
8. **Lobell, D. B., & Field, C. B. (2007).**
   Global scale climate-crop yield relationships and the impacts of recent climate change. Environmental Research Letters, 2(1), 1-7.
   https://doi.org/10.1088/1748-9326/2/1/014002
9. **McKinsey & Company. (2016).**
   The future of food: How digital technologies can help farmers and food producers achieve more sustainable, profitable growth. McKinsey & Company.
   https://www.mckinsey.com/industries/agriculture
10. **Shapley, L. S. (1953).**
    A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), Contributions to the Theory of Games (Vol. II, pp. 307–317). Princeton University Press.