****

**University of Sulaymaniyah**

**College of Science**

**Computer Department**

**Stage 4**

**Data analysis on wheat yearly prices analyzing and representation**

**Prepared:**

**Meer Shwan**

**Zaniar Xalid Dana Hussien**

**Lectured:**

**DR. Miran Taha Abdulah**

**29/1/2024**

**Contents**

[**Introduction 3**](#_Toc157522216)

[**Problem Statement 4**](#_Toc157522217)

[**Solution Method 4**](#_Toc157522218)

[**Pseudo code 4**](#_Toc157522219)

[**Implementation: 6**](#_Toc157522220)

[**Graphs 7**](#_Toc157522221)

[**Results Discussion: 8**](#_Toc157522223)

[**Project Conclusion: 8**](#_Toc157522224)

# 

# **Introduction**

Welcome to our Wheat Price Analysis project, a comprehensive exploration aimed at understanding the dynamics of wheat prices and their broader implications. In choosing this project, we recognize the pivotal role that wheat plays in the global economy and the lives of individuals worldwide. The project was born out of a desire to shed light on the intricate factors influencing wheat prices, with the ultimate goal of contributing valuable insights to society.

Wheat, a staple food for a significant portion of the global population, holds immense economic and social importance. The fluctuations in wheat prices impact not only farmers and stakeholders in the agricultural sector but also resonate through various facets of daily life, from food prices to economic stability. The relevance of our Wheat Price Analysis project lies in its potential to unravel patterns, trends, and correlations within the complex web of factors influencing wheat prices.

By undertaking this project, we aim to provide a different understanding of how production trends, and external factors contribute to the dynamics of wheat prices. The societal impact is substantial, as these insights can inform decision-makers in making more informed choices related to agriculture, trade, and food security.

The potential benefits for daily life are manifold. From helping farmers make strategic decisions to empowering consumers with knowledge about potential price fluctuations, our project seeks to contribute to a more resilient and informed society. By leveraging data science methodologies, we aspire to make a meaningful impact on the accessibility, affordability, and sustainability of one of the world's most vital commodities.

## **Problem Statement**

The intricate dynamics of wheat pricing in the agriculture industry are the focus of our investigation. We attempt to identify the factors driving wheat prices using a dataset that includes a variety of attributes, such as average closing prices, annual percentage changes, and output levels. Problems including missing values, duplicate data, and complex variable linkages need to be addressed. The goal of the project is to shed light on these dynamics so that decision-makers in sustainable agriculture and policymakers can make well-informed choices. Better price forecasting models and a deeper comprehension of the major factors influencing wheat pricing are among the expected results.

# **Solution Method**

Our method starts with data cleaning to remove duplicates and missing variables, then moves on to exploratory data analysis with visualizations to find patterns and connections. In order to comprehend important variables, we use statistical measures and, for categorical data, label encoding. Information about attribute dependencies is provided by the correlation analysis. Predictive modeling may optionally incorporate machine learning models. The findings provide a thorough grasp of the dynamics of wheat prices through lucid graphics and data summaries.

## **Pseudo code**

1. Step 1: Begin
2. Step 2: The user describes the goal of the use. The goal is to analyze wheat prices data and generate insights through data visualization and statistical analysis.
3. Step 3: The functional requirements for reusability are collected

- Read wheat prices data from an Excel file.

- Display the first and last few rows of the dataset.

- Plot scatter plots, bar charts, and line plots to visualize relationships.

- Calculate and display statistical measures such as mean, max, min.

- Handle missing values in the dataset.

- Encode categorical variables using LabelEncoder.

- Generate and save a correlation matrix.

1. Step 4: Generate test cases, execute, and collect functional coverage report

- Verify that the data is read correctly.

- Validate statistical measures.

1. Step 5: Select metrics for identification and choose its bounds

- Metrics: Duplication count, mean, max, min.

- Bounds: Define thresholds for acceptable values.

1. Step 6: Identify candidate components satisfying criteria from Step 4

- Components: Data reading, data visualization, statistical analysis, data preprocessing.

1. Step 7: Update the dynamic component metrics library

- Store metrics for identified components.

1. Step 8: Find out the optimal path for the identified set of components

- Optimal path: Sequence of execution for data reading, visualization, and analysis.

1. Step 9: Calculate the reuse frequency for each component

- Reuse frequency: How often each component is reused in similar analyses.

1. Step 10: The reuser makes the final decision about the reuse of components in the qualified set.

- Decide whether to use qualified components for future analyses.

Step 11: End

# **Implementation:**

The following crucial actions are included in the project implementation:

Data Loading: To load the 'wheat\_prices.xlsx' dataset into a DataFrame, use the Pandas package.

Data cleaning: Use appropriate imputation techniques to fill in the missing values in the 'produced\_amount' column.

Remove duplicate items from all pertinent columns to improve data integrity.

Correlation Analysis:

To measure the connections between various attributes, compute the correlation matrix.

Use a heatmap to visually represent correlations to get a rapid understanding of attribute interdependence.

Encoding Labels:

When converting categorical data into numerical format, use Scikit-Learn's LabelEncoder to make machine learning integration easier

To gain statistical insights into important variables, compute the mean, maximum, minimum, and standard deviation.

Machine Learning Integration :

Utilize historical data to train algorithms for wheat price forecasting.

Exploratory Data Analysis (EDA):

To examine patterns and connections in the dataset, create visualizations with Matplotlib and Seaborn.

To illustrate important characteristics like average closing prices and production numbers, create scatter plots, bar charts, and line graphs.

## **Graphs**

# 

# **Results Discussion:**

The acquired data provide insightful information about the dynamics of wheat prices. Trends and linkages are revealed through visualizations such as scatter plots and line graphs. The mean, maximum, and minimum are examples of statistical measurements that provide information on core patterns. The impact of significant factors on wheat prices is discussed, laying the groundwork for well-informed decision-making. By providing a thorough understanding of the dataset, the findings aid in the development of future applications and analysis.

# **Project Conclusion:**

In conclusion, by using a solid technique, our project has effectively handled the challenges associated with comprehending wheat pricing. By means of data cleansing, exploratory data analysis, and correlation analysis, we were able to acquire important knowledge regarding the driving forces behind wheat prices. The outcomes, which are displayed using statistical metrics and visualizations, add to a thorough comprehension of the dataset. In the agriculture sector, the cornerstone for well-informed decision-making is the identification of trends and linkages. In the long run, these results can help support sustainable agricultural practices and direct future efforts at predictive modeling. Through the research, knowledge gaps have been successfully filled and the foundation for future developments in wheat price analysis has been established.