# **Vegetable Price Analysis and Forecasting**

## **Overview**

This project focuses on analyzing vegetable price trends, comparing prices across different categories, and building predictive models for forecasting future prices. It combines exploratory data analysis (EDA), visualization, and machine learning techniques to deliver actionable insights and forecasts.

## **Objectives**

* Understand the trends and patterns in vegetable pricing.
* Identify correlations between different vegetable prices.
* Predict future prices using machine learning models.
* Provide actionable insights for stakeholders in the agriculture and retail sectors.

## **Dataset**

* **Source**: Kaggle (provided in CSV format).
* **Columns**:
  + Price Dates: Dates corresponding to the price recordings.
  + Vegetable categories: Prices of Bhindi (Ladies finger), Tomato, Onion, Potato, Brinjal, Garlic, Peas, Methi, Green Chilli, and Elephant Yam (Suran).

## **Methodology**

1. **Data Preprocessing**:
   * Handled missing values and cleaned data for analysis.
   * Converted date columns to a datetime format for temporal analysis.
2. **Exploratory Data Analysis (EDA)**:
   * Visualized price trends and distributions.
   * Analyzed correlations between vegetable prices using scatter plots and pair plots.
   * Examined proportions of average prices using pie charts.
3. **Feature Engineering**:
   * Created new features like Price Differences and lag-based features for time series forecasting.
4. **Model Building**:
   * Applied models such as:
     + Linear Regression.
     + Random Forest Regressor.
     + ARIMA for time series forecasting.
   * Evaluated models based on Mean Absolute Error (MAE) and R-squared metrics.
5. **Visualization of Results**:
   * Generated visual insights for price trends, proportions, and model predictions.

## **Key Findings**

* **Trends**: Most vegetables exhibited seasonal trends with occasional outliers caused by external factors.
* **Correlations**: Strong positive correlations between certain vegetable pairs, suggesting market dependencies.
* **Forecasting**: Random Forest Regressor provided the most accurate predictions, closely capturing price fluctuations.

## **Technologies Used**

* **Python Libraries**:
  + pandas, numpy: Data manipulation and analysis.
  + matplotlib, seaborn: Data visualization.
  + sklearn: Machine learning.
  + statsmodels: Time series forecasting.

## **How to Run**

Clone this repository:  
bash  
Copy code  
git clone https://github.com/your-repository/vegetable-price-analysis.git

Install the required Python libraries:  
bash  
Copy code  
pip install -r requirements.txt

Run the Jupyter notebook:  
bash  
Copy code  
jupyter notebook Vegetable\_Price\_Analysis.ipynb

## **Results and Insights**

* Insights from EDA and model predictions are stored in the results/ directory.
* Visualizations are available in the figures/ directory.

## **Contributors**

* **Keamogetswe Mothoa**Role: Data Analyst and Project Lead  
  Contributions: Data analysis, modeling, and reporting.

## **References**

* Data Source: [Kaggle](https://www.kaggle.com/)
* Python Documentation: pandas, seaborn, [sklearn](https://scikit-learn.org/)

## **License**

This project is licensed under the MIT License. See the LICENSE file for details.