# ***PREDICTION MODEL JOURNEL***

* Classified the dataset on the basis of crop type to obtain the following 6 csv sheets for data analysis and prediction.

1. CORN
2. COTTON
3. ONION
4. POTATO
5. RICE
6. TOMATO

* SORTED THE DATA ON THE BASIS OF DATE
* STARTED WORKING ON CLEANING AND PREPROCESSING THE DATA IN PYTHON

To simplify the visualizations and focus on the predicted prices with clear labels, I'll streamlined the code to:

1. Train the model using RandomForestRegressor.
2. Predict the future modal prices for the next 10 days.
3. Visualize the predicted modal prices clearly.

* HERE ARE THE STEPS USED IN PYTHON

## **Data Loading and Preprocessing:**

* Load the dataset containing crop price data, ensuring that the date column is parsed correctly.
* Drop any irrelevant columns such as crop ID, crop name, and arrival quantity.
* Convert the date column to the appropriate datetime format.
* Extract additional features from the date, such as year, month, and day, to facilitate analysis.

1. **Feature Engineering:**

* Create lag features by shifting the modal price data by a certain number of days. This step helps capture temporal dependencies in the data and is crucial for time series forecasting.

1. **Data Encoding:**

* Encode categorical variables, such as the name of the mandi, using techniques like label encoding. This step converts categorical data into a numerical format suitable for machine learning models.

1. **Model Training:**

* Split the dataset into training and testing sets to evaluate the model's performance.
* Choose an appropriate machine learning model for the task. In this project, a RandomForestRegressor is used due to its ability to capture nonlinear relationships and handle tabular data effectively.
* Train the model using the training data, fitting it to the features (input variables) and the target variable (modal price).

1. **Model Evaluation:**

* Make predictions on the test set using the trained model.
* Evaluate the model's performance using metrics such as Mean Absolute Error (MAE), which measures the average absolute difference between predicted and actual prices.

1. **Future Price Prediction:**

* Use the trained model to predict crop prices for the next 10 days.
* Utilize the lag features and the most recent known data point to forecast future prices iteratively.

1. **Visualization:**

* Visualize the predicted modal prices for the next 10 days using a line plot.
* Annotate each data point with its corresponding predicted price to provide clear labels and aid interpretation.

1. **Project Journal Entry:**

* Document each step of the data preprocessing, feature engineering, model training, evaluation, and prediction process in the project journal.
* Include explanations of the techniques used, rationale behind the choices made, and any challenges encountered during the project.
* Reflect on the model's performance, discussing its strengths, limitations, and potential areas for improvement.

# Outputs:

# **CORN**

A graph with blue lines

Description automatically generated

# **COTTON**

A graph with a line going up

Description automatically generated

# **ONION**

A graph with numbers and lines

Description automatically generated

# **POTATO**

A graph with numbers and lines

Description automatically generated

# **RICE**

A graph with blue lines

Description automatically generated

# **TOMATO**

A graph with blue lines

Description automatically generated