

Agricultural Market Price Data Analysis and Prediction

1.Introduction:

This project analyses agricultural commodity market prices to provide insights into market behavior and aid data-driven agricultural planning for stakeholders such as farmers, traders, and policymakers. The goal is to explore the dataset, conduct thorough analysis, build predictive models, and uncover patterns that can inform better decision making.

2.Data Overview:

The dataset, “Agricultural Market Data BDA.csv,” contains 3,236 records with 10 columns. Features include State, District, Market, Commodity, Variety, Grade, ArrivalDate, Min Price, Max Price, and Modal Price. There are multiple unique values for each categorical feature, reflecting a wide coverage across Indian states, districts, markets, and commodities. No missing or duplicate values were identified, demonstrating completeness and reliability

3. Data Cleaning and Preprocessing:

The cleaning steps included:

- Removal of duplicate records (none found).
- Checks and handling of missing values (none present).
- Data type conversions for consistency, such as converting date columns and encoding categorical variables.
- Outlier detection using the IQR method revealed some potential outliers in price columns, likely due to special market events or seasonal effects.
- The dataset emerged fully clean, containing 3,236 entries ready for analysis.

4. Exploratory Data Analysis & Visualizations:

EDA covered statistical and visual examination of price distributions, commodity frequency, and regional patterns:

- Price columns (Min, Max, Modal) display slightly right-skewed distributions, with most commodities priced moderately and a few outliers at high values.

- Pairwise correlation among price columns is strong—indicating consistent market dynamics.
- Top ten commodities (such as tomato, onion, rice) were highlighted, showing substantial variation in average modal prices.
- State-level average prices reveal significant regional differences, likely due to logistics and local market forces.
- Seasonal arrival patterns were visualized, suggesting harvest-related or festival-driven supply influxes.
- Multiple bar plots, histograms, and heatmaps were used in the notebook for these visual insights

5. Final Analysis: Identifying Highly Successful Apps:

The final analysis involved predictive modeling to estimate modal price:

- A Linear Regression model was developed using price columns and encoded categorical features (state, commodity, etc.).
- The model demonstrated strong performance, with an R^2 score of 0.987, low RMSE, and close matches between actual and predicted modal prices for test samples.
- Modal price was shown to be highly predictable based on available features, confirming the consistency of market mechanisms.
- Certain commodities and regions stood out as having regularly higher modal prices, reflecting successful market segments.
- Recommendations for next steps include trying advanced machine learning models (Random Forest, Gradient Boosting) and forecasting frameworks for future trends.

6. Conclusion:

This project covered the complete analytics and modeling pipeline using agricultural price data. Key results:

- Data was well-organized, clean, and consistent.
- Market prices exhibited clear patterns, with strong correlations across price metrics and noticeable regional and seasonal trends.
- The analysis provides actionable insights to optimize supply chains, inform market interventions, and improve farmer and trader outcomes.

