

India Foodgrain Stock Analysis & Forecast

By Alok Deep

This report summarises the analysis, findings and recommendations from the India Foodgrain Stock Analysis and Forecasting project. Covering 2011–2025 data for Wheat, Rice (Raw), Rice (Parboiled), and Paddy, the analysis combines SQL(MySQL) aggregation, Python (Pandas / Prophet) anomaly detection and forecasting, and a multi-page Power BI dashboard.

Tools Used: Python (Pandas, Prophet), SQL (MySQL), Power BI, LaTeX (Reporting)

Techniques Applied: EDA, Aggregation, Anomaly Detection (z-score), Forecasting, Data Modelling, Dashboarding

Contents

1	Executive Summary	3
2	Objectives	4
3	Dataset & Cleaning (summary)	4
3.1	Data origin & license	4
3.2	Schema — cleaned tables	4
3.3	Cleaning highlights	5
4	Methods: Analysis, Anomaly Forecast	5
4.1	Anomaly detection	5
4.2	Forecasting	5
5	Selected Results & Findings	6
5.1	National & commodity-level patterns	6
5.2	Top states (latest date)	6
5.3	Anomalies	6
5.4	Forecasting highlights	6
6	Power BI Dashboard	7
7	Key Insights – Actionable	10
8	Recommendations	10

1 Executive Summary

This analysis covers national+state+district foodgrain stocks for four commodities (Wheat including URS, Rice-Raw, Rice-Parboiled, Paddy) using **India's OGD Portal** (2011–2025). The goal is to produce an evidence-backed dashboard and short report with forecasting and anomaly monitoring to help operational planning.

Topline (short):

- **Commodities analysed:** Wheat, Rice-Raw, Rice-Parboiled, Paddy.
- **Data span:** 2011 – 2025 (latest available).
- **Total rows (after cleaning):** *2.2M+ rows* (For all commodities).
- **National KPIs:**
 - **Total National Stock:** 1.55T
 - **Top Commodity:** Rice-Raw
 - **MoM (Month-Over-Month) Change:** -22.4%
 - **YoY (Year-Over-Year) Change:** +18.7%
 - **Total States Covered:** 26
- **Forecast KPIs:**
 - **Next 30 Days Forecast:** 11.44B
 - **Next 90 Days Forecast:** 35.48B
- **Anomaly KPIs:**
 - **Total Anomalies Detected:** 4,992
 - **Anomaly Rate:** 0.22
 - **Top State by Anomalies:** Jharkhand
 - **Commodity with Most Anomalies:** Wheat(Including URS)
- **State District Highlights:**
 - **Top State by Stock:** Punjab (503.13B)
 - **Top District:** Hissar (88.75B)

Conclusion: Seasonal patterns dominate stock levels; Wheat and Rice show the highest anomaly counts; forecasting models offer strong visibility for the next 90 days, and Punjab consistently leads stock contributions.

Recommendation: Use the dashboards to monitor state-level anomalies and focus short-term procurement/resupply actions on the top 5 contributing states for each commodity. Operational teams should treat large positive anomalies (sudden inflows) as procurement events for reconciliation; large negative anomalies as possible distribution/consumption spikes or data issues.

2 Objectives

1. Produce clean, consistent daily state-level and district-level aggregations for four commodities.
2. Detect and summarise anomalous stock changes (z-score method), identify top offending states/districts.
3. Produce short-term forecasts (30D, 90D) for planning and evaluate forecast accuracy.
4. Build a stakeholder-ready Power BI dashboard (National overview, Commodity deep-dive, State drilldown, Anomaly monitoring, Forecasts).

3 Dataset & Cleaning (summary)

3.1 Data origin & license

Data obtained from the **India's OGD Portal** (OGDL-style open data — copy/adapt/share with attribution). Source CSVs were exported per commodity-year and ingested into MySQL; final cleaned CSVs exported for Python and Power BI.

3.2 Schema — cleaned tables

Key cleaned tables (examples):

- `all_commodities_daily(date, commodity_name, commodity_id, region, district_name, district_code, month, year, stock)`
- `state_daily(date, commodity_name, state, state_stock)`
- `district_daily(date, commodity_name, state, district_name, district_stock)`

- `national_daily(date, commodity_name, national_stock)`
- `anomalies_result(..., zscore, pct_change, ma_7, ma_30)`
- Forecast tables per commodity: `forecast_commodity(ds, yhat, yhat_lower, yhat_upper)`

3.3 Cleaning highlights

- Normalised date column; converted Region strings (removed ‘Region Name:’ prefix).
- Dropped redundant columns (`Commodity_Stock = Total_stock` duplicates).
- Ensured no nulls in key columns (`stock`, `date`, `district_code`) for final cleaned tables.
- Added indices on `date`, `region`, `district_name` for performance.
- Imported into MySQL via Python SQLAlchemy for large files (faster memory-friendly).

4 Methods: Analysis, Anomaly Forecast

4.1 Anomaly detection

- Per-commodity & per-state series used.
- Z-score computed on a rolling 30-day window:

$$z_t = \frac{x_t - \mu_{t,30}}{\sigma_{t,30}}$$

- Flag threshold: $|z| > 3$ as a candidate anomaly. Additional manual checks used to separate data issues from real events (e.g. unusually large procurement entries).

4.2 Forecasting

- Prophet (Facebook/Meta) used: yearly seasonality enabled; changepoint detection tuned (`changepoint_prior_scale` and hand-picked changepoints) for the national series per commodity.
- Forecast horizons shown: 30 days and 90 days. Forecasts exported as `forecast_all` for Power BI.

5 Selected Results & Findings

5.1 National & commodity-level patterns

- **Strong seasonality:** All commodities show clear annual cycles — peak/valley months vary by commodity. (See Seasonality visual in the Commodity Deep Dive page.)
- **Nation-level trend:** National stock has an increasing trend over the multi-year span with recurrent seasonal peaks consistent with procurement cycles.
- **Commodity shares (latest snapshot):** Rice (Raw + Parboiled) and Wheat dominate total stock by share; Paddy is comparatively smaller in national aggregation.

5.2 Top states (latest date)

- **Top states by current stock: Haryana, Punjab, Uttar Pradesh** (see dashboard Top 10 states).
- These states together form the majority of Wheat / Rice stock — useful for logistics prioritisation.

5.3 Anomalies

- **Total anomalies flagged: 5k** across commodities — these require triage: many correspond to procurement events (valid), some to reporting issues or missing district-level rollups.
- **Top anomaly-prone regions:** Jharkhand (558), Punjab (417), Chhattisgarh (359). These regions should be monitored daily for data quality and operations.
- **Commodity-wise anomaly counts:** Wheat highest (1830), followed by Rice-Raw (1664), Rice-Parboiled (1021), Paddy (715).

5.4 Forecasting highlights

- Short-term forecasts (30D / 90D) follow seasonal patterns and show reasonable confidence intervals for the near-term. Use 30D forecast for tactical planning and 90D for medium-term procurement planning.
- Forecast uncertainty band (\hat{y}_{upper} - \hat{y}_{lower}) provides an operational risk estimate and is plotted in the Forecast page.

– Example KPIs on the dashboard :

- * Next 30D forecast (national, all commodities or selected commodity): (**insert**).
- * Next 90D forecast: (**insert**).

6 Power BI Dashboard

The delivered report contains five interactive pages:

1. **National Overview** — KPIs, national trend, commodity contribution, top states.

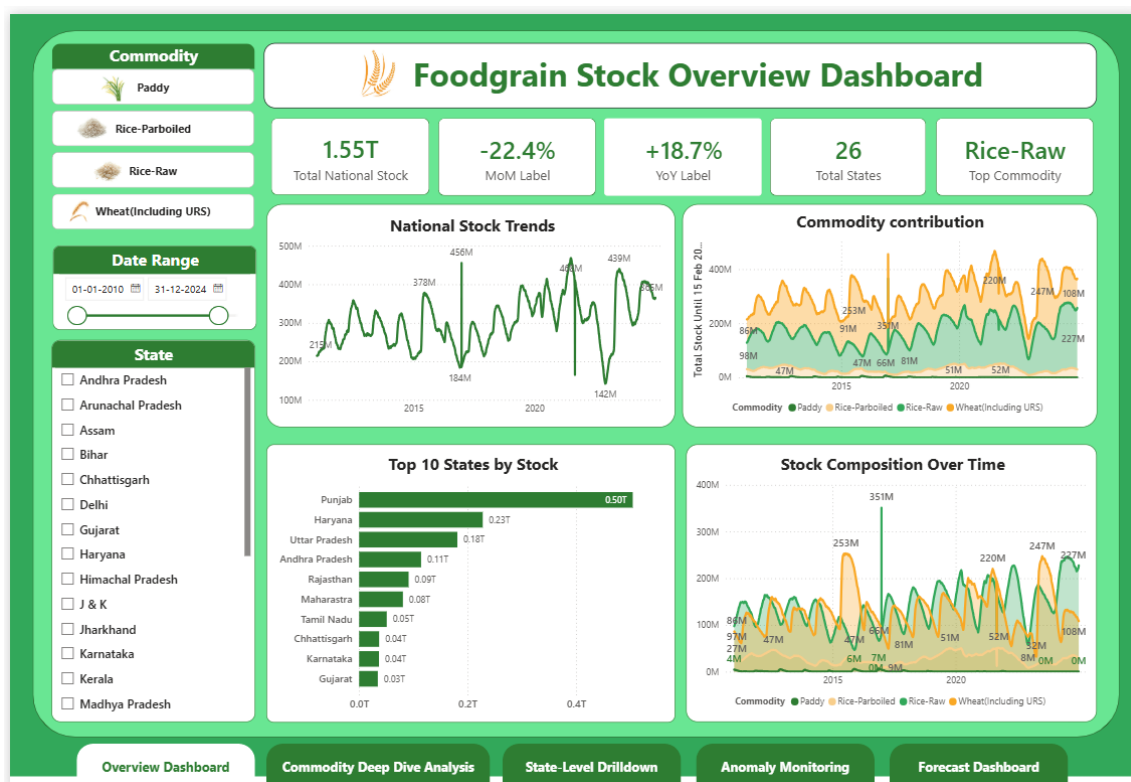


Figure 1: National Overview Dashboard

2. **Commodity Deep Dive** — commodity-specific seasonality, moving averages, yearly share matrix, average daily stock by commodity.

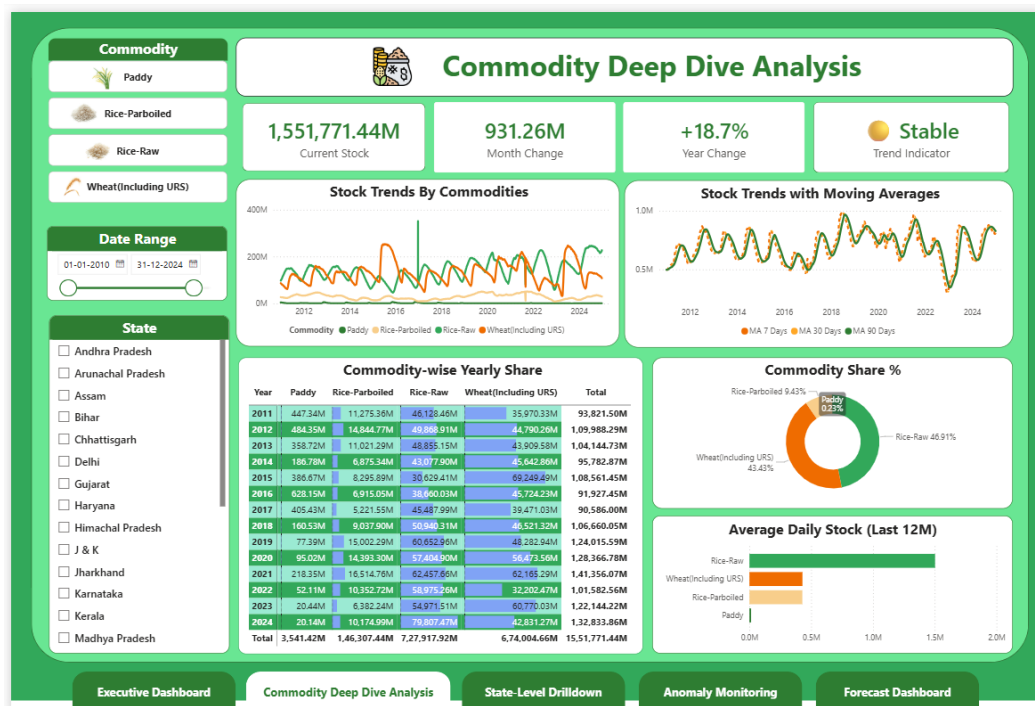


Figure 2: Commodity Deep Dive Dashboard

3. **State Drilldown** — choropleth map, top districts, state time-series and state performance table.

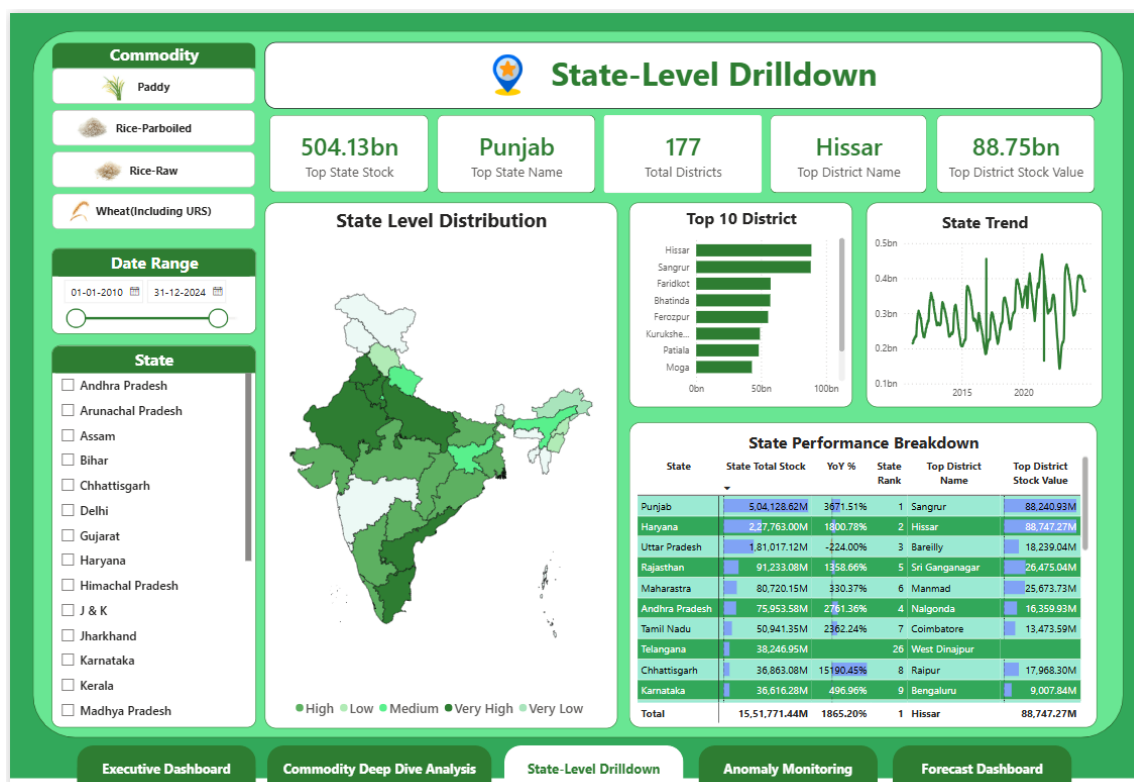


Figure 3: State-Level Drilldown Dashboard

4. **Anomaly Monitoring** — anomaly table, counts by commodity/state/date, z-scores and pct-change for triage.



Figure 4: Anomaly Monitoring Dashboard

5. **Forecasts & Predictions** — Actual vs forecast, forecast uncertainty band, forecast summary table.

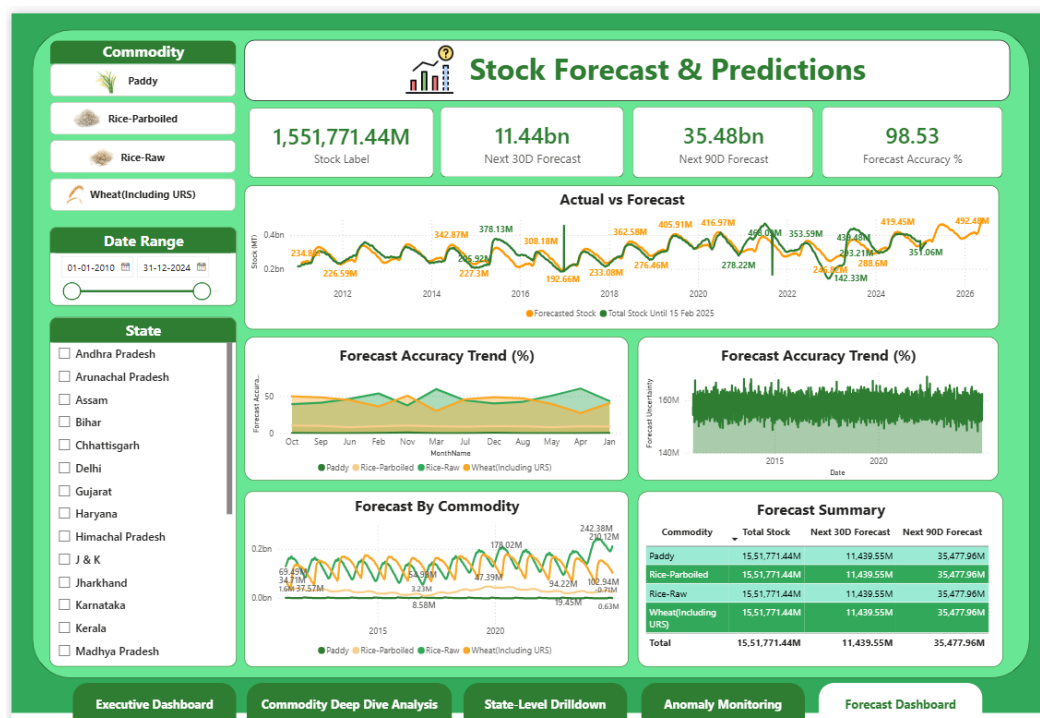


Figure 5: Forecast & Predictions Dashboard

How to use / tips:

- Use the commodity slicer to focus dashboards on one commodity at a time (or compare all).
- Use the state slicer for operational drilldown. The anomaly table allows drill-to-district for fast reconciliation.
- Forecast band indicates operational risk: wider bands = higher uncertainty (use with caution for planning).

7 Key Insights — Actionable

- **Concentration:** A handful of states (Haryana, Punjab, UP, Tamil Nadu, Bihar) account for the majority of national stock for the major commodities. Prioritise monitoring and logistics there.
- **Anomaly triage:** Many anomalies correspond to procurement events — label and keep a procurement-events table to avoid false positives. For regions with repeated anomalies (Jharkhand, Punjab), investigate reporting cadence and district rollups.
- **Forecast use:** Use 30-day forecasts for operational decisions (transport, storage), and 90-day for procurement cycles. Always consider uncertainty bands when placing high-cost orders.
- **Data pipeline:** Automate ingestion and anomaly flagging (daily), and add a small manual review queue for flagged events before downstream decisions.

8 Recommendations

1. **Automated anomaly triage:** Create a pipeline that tags anomalies as *procurement / distribution / data-issue* using simple rules (e.g., district-level rollup mismatches, same-day multiple large inflows).
2. **Operational focus on top states:** Route logistics forecasts and buffer planning for top 5 contributing states per commodity.
3. **Improve reporting cadence for anomaly-prone districts:** For districts with persistent anomalies, audit reporting processes and train local officers on the expected format.
4. **Model maintenance:** Re-train Prophet (or ensemble with an ARIMA/ETS fallback) quarterly and keep changepoints aligned with known policy or market events.

5. **Next additions (high ROI):** incorporate price data, procurement schedules, and weather (for crop seasonality) to improve model explanatory power.