Market Segmentation Report: Indian Agriculture (Geographic & Pricing)

Prepared by: Automated Analysis

1. Objective

Analyze the provided Indian agriculture dataset focusing on geographic segmentation (state, district, market) and pricing segmentation (modal_price) for final produce sales.

2. Data & Approach

The dataset contains 2,238 records. We focused on the top commodities by frequency: Potato, Tomato,
Onion, Paddy(Dhan)(Common), Brinjal, Cauliflower, Banana, Green Chilli. We computed state-commodity
price statistics, used PCA to reduce dimensionality, and applied KMeans clustering (k=3) on state
price profiles.

3. Key Findings

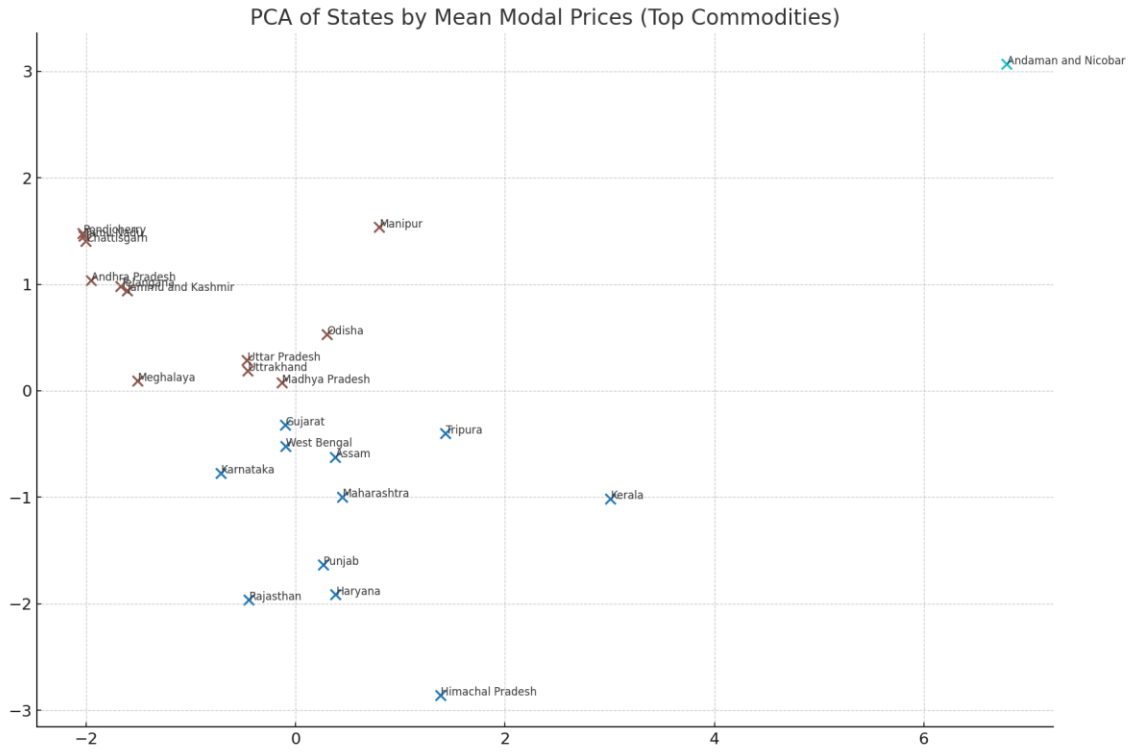
- Geographic concentration: a few states dominate market arrivals for top commodities.

- Price segmentation: distinct clusters of states with low/mid/high mean modal prices.

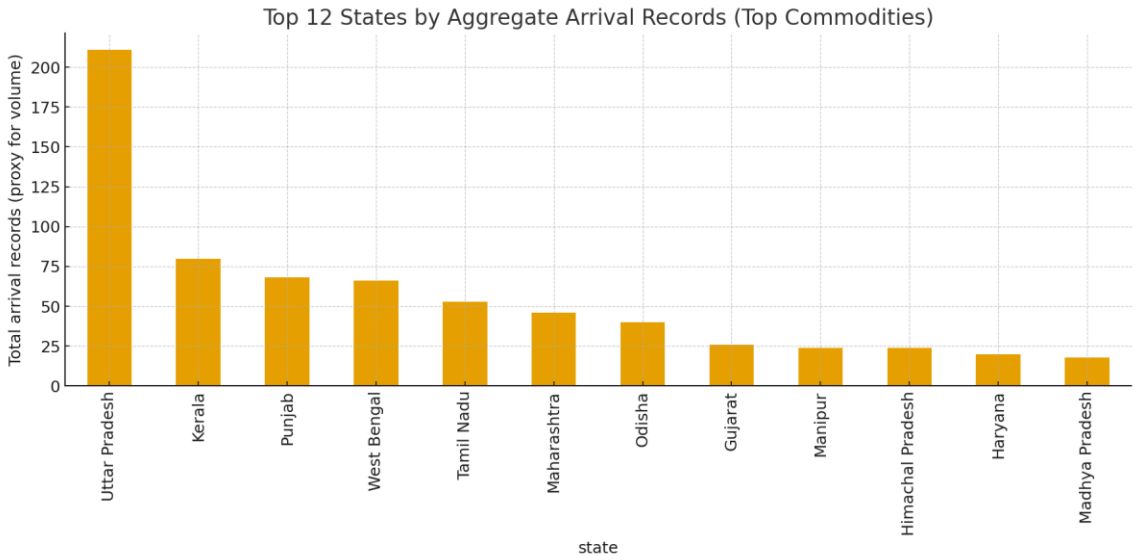
- Perishables show high price dispersion across regions, indicating seasonality and potential for distress sales.

4. Data-driven Insights

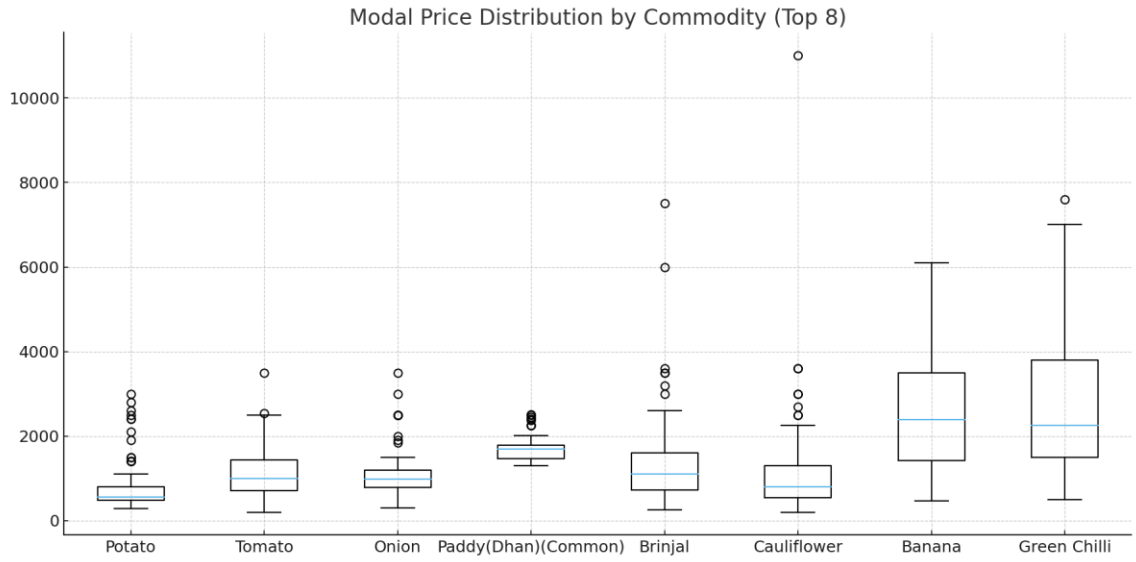
The PCA scatter below positions states by principal components of standardized mean modal prices. Nearby states in the plot have similar pricing structures for the selected commodities.



Quantity concentrations (Top 12 states) suggest that infrastructure investments (cold storage, processing) in these states would disproportionately reduce waste and stabilize prices. The bar chart below shows top states by aggregate arrival records for the top commodities.



Price distributions across the top commodities reveal large interquartile spreads and outliers, particularly in perishable crops. The boxplot below highlights this dispersion and suggests where price risk is highest.



5. Geographic & Pricing Segmentation (Details)

State clusters (0/1/2) derived from KMeans on PCA-transformed mean prices are saved alongside the summary CSV. Use these clusters to target interventions for infrastructure, market linkages, and price stabilization.

State	Cluster
Maharashtra	0
Tripura	0
Rajasthan	0
Punjab	0
Kerala	0
Karnataka	0
West Bengal	0
Haryana	0
Gujarat	0
Assam	0
Himachal Pradesh	0
Jammu and Kashmir	1
Uttar Pradesh	1
Andhra Pradesh	1
Telangana	1
Tamil Nadu	1
Pondicherry	1
Odisha	1
Meghalaya	1
Manipur	1
Uttrakhand	1
Madhya Pradesh	1
Chattisgarh	1

Andaman and Nicobar	2

6. Implications and Recommendations

1. Invest in post-harvest infrastructure in high-arrival states to reduce distress sales.

2. Enhance market linkages and transport incentives between surplus and deficit regions.

3. Promote crop diversification and staggered planting to avoid synchronized gluts.

4. Explore price-stabilization mechanisms for highly volatile crops (buffer stocks, price support).

7. Supplementary Files

- summary_by_state.csv (summary_by_state.csv)

- state_clusters.csv (state_clusters.csv)

- PCA scatter, quantity bar chart, and boxplot are embedded in this report.

Github link-

8.Conclusion

Indian agricultural markets show **strong geographic segmentation and significant price variation across states**:

1. **Production and arrival concentration**
    - A handful of states dominate arrivals for major commodities (Potato, Tomato, Onion, Paddy, Brinjal, Cauliflower, Banana, Green Chilli).
    - These states act as "surplus hubs," while many others are structurally deficit and rely on imports.
2. **Price segmentation**
    - **Surplus states** (high arrivals) form a cluster with generally **lower modal prices** — oversupply often depresses farmer incomes.
    - **Deficit or remote states** form a cluster with **higher prices**, reflecting transport costs and limited local supply.
    - A middle cluster exists with moderate arrivals and prices.
3. **Commodity volatility**

- Perishable crops (Tomato, Onion, Potato) show **wide price dispersion** and many outliers across markets, confirming high volatility and risks of distress sales.
- More stable crops (Paddy, Banana) show narrower distributions, indicating less volatility.

4. **Farmer welfare & value chain implications**
    - Farmers in surplus states suffer from **low farm-gate prices** due to glut conditions.
    - Farmers in deficit states face higher input costs and logistics challenges but can capture better margins if market linkages exist.
    - The mismatch signals inefficiencies in the **supply chain and storage infrastructure**.