

CAPSTONE PROJECT REPORT

Data Visualization & Analytics

# Commodity-Wise Mandi Price Analysis

## Farmer Decision Support System

*"Identifying the best market and time to sell for maximum farmer returns"*

Sector	Agriculture
Team	Group 4 — Section B
Team Members	Khushi (2401010225), Abhinav (2401010017), Anshika (2401010080), Kshitiz (2401010242), Abhijeet (2401010014), Nishant (2401010302)
Faculty Mentors	Satyaki Sir & Ayushi Ma'am
Dataset Coverage	2018 – 2026   6,000 Rows   18 Columns
Submission Date	18th February 2026

# 1. Executive Summary

---

Indian farmers consistently face suboptimal price realization due to information asymmetry across agricultural mandis. Without reliable, data-driven intelligence on when and where to sell, farmers rely on intuition and fragmented market signals, resulting in distress selling and reduced profitability.

This project analyzes 6,000 commodity-level mandi records spanning 8 states, 8 markets, and 8 commodities across 2018–2026 to build a Farmer Decision Support System. Using Google Sheets as the primary analytics platform, the team engineered five actionable KPIs — Modal Price, Price Spread, Price Position Index (PPI), Relative Market Advantage (RMA), and Profit Opportunity Indicator — and built an interactive dashboard that surfaces optimal selling windows and market locations.

## Key Insights

- Lucknow Mandi consistently delivers the highest average modal price (₹3,109.78/q), making it the top performing market in the dataset.
- May is the peak selling month with the highest average modal prices, driven by post-harvest supply tightening.
- Rabi crops command structurally higher prices compared to Kharif and Summer harvests.
- Maharashtra, Uttar Pradesh, and Rajasthan have the highest share of high-profit opportunity windows (42.8%, 42.3%, 41.6% respectively).
- Average market volatility (Price Spread) of ₹490.20/q signals significant arbitrage potential through aggregation strategies.

## Key Recommendations

- Align bulk sales with the April–May window to capture the Summer Price Premium.
- Prioritize Lucknow Mandi and Maharashtra markets for higher structural realizations.
- Use the PPI threshold of >0.6 and positive RMA as a composite sell signal.
- Withhold non-perishable stock during Nov–Jan to avoid the low-price market glut window.

## 2. Sector & Business Context

---

### Sector Overview

Agriculture is the backbone of the Indian economy, contributing approximately 17–18% of GDP and employing over 50% of the workforce. The APMC (Agricultural Produce Market Committee) mandi system serves as the primary channel through which farmers sell produce, yet it is plagued by chronic information inefficiency. Over 7,000 mandis operate across India with significant price variation — sometimes 30–50% — for the same commodity on the same day across different markets.

### Current Challenges

- Price Discovery Gap: Farmers lack real-time, aggregated price data to make selling decisions.
- Seasonal Vulnerability: Post-harvest flooding of supply causes price crashes, particularly in Nov–Jan for many staple crops.
- Geographic Arbitrage Loss: Farmers sell to the nearest mandi rather than the highest-paying market.
- Intermediary Dependence: Commission agents and aggregators capture the arbitrage value that should go to farmers.
- Quality Premiums Unrealized: Without granular market data, sorting and grading investments remain sub-optimal.

### Why This Problem Was Chosen

The mandi price problem is analytically rich — it involves temporal patterns (seasonality), geographic variation (state and market-level benchmarking), commodity-level heterogeneity, and derived KPIs with direct policy relevance. The dataset from Agmarknet (via Kaggle mirror) provided row-level transactional data requiring real cleaning and transformation, making it ideal for a capstone demonstration of end-to-end analytical capability.

## 3. Problem Statement & Objectives

---

### Formal Problem Definition

"Farmers lack clarity on when and where to sell their crops due to significant price variations across mandis and seasons, leading to suboptimal selling decisions and reduced net profits."

### Project Scope

- Geographic Scope: 8 Indian states (Bihar, Karnataka, Maharashtra, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, West Bengal) across 8 major mandis.
- Temporal Scope: 2018 to 2026 (7 years of transactional data).
- Commodity Scope: Cotton, Maize, Onion, Potato, Rice, Soybean, Tomato, Wheat.
- Analytical Scope: Price trends, seasonal variation, market benchmarking, KPI engineering, and opportunity flagging.

### Success Criteria

- Identify the top-performing market and peak selling month with statistical backing.
- Engineer at least 3 derived KPIs that map directly to farmer selling decisions.
- Build an interactive dashboard with commodity, season, and state-level drill-down capability.
- Generate at least 8 decision-oriented insights with actionable recommendations.

### Core Business Question

How can we predict optimal selling windows and market locations to minimize distress sales and maximize farmer income?

## 4. Data Description

**Dataset Source:** Agmarknet (National Horticulture Board / Government of India) — accessed via Kaggle mirror (raw/original version, row-level transactional records).

**Access Link:** <https://www.kaggle.com/datasets/agmarknet-commodity-prices> (primary source: data.gov.in)

### Data Structure

Column Name	Data Type	Description	Role
Arrival_Date	Date	Date of commodity arrival at mandi	Temporal dimension
State	Categorical	State where the mandi is located	Geographic grouping
District	Categorical	District of the mandi	Sub-geographic context
Market	Categorical	Mandi/market name	Primary location KPI
Commodity	Categorical	Crop/produce name	Product dimension
Variety	Categorical	Grade/variety of the commodity	Quality segmentation
Min_Price	Numeric (₹/q)	Minimum traded price for the day	Spread calculation
Max_Price	Numeric (₹/q)	Maximum traded price for the day	Spread calculation
Modal_Price	Numeric (₹/q)	Most frequent traded price	Primary revenue KPI
Season	Derived Categorical	Kharif / Rabi / Summer	Seasonal analysis
Month_Name	Derived Text	Month derived from Arrival_Date	Monthly trend analysis
Price_Spread	Derived Numeric	Max_Price – Min_Price	Volatility KPI
PPI	Derived Numeric	(Modal – Min) / (Max – Min)	Negotiation strength KPI
RMA_Index	Derived Numeric	Modal – State Average	Market advantage KPI
Profit_Opportunity_Flag	Derived Binary	High / Normal classification	Decision trigger KPI
State_Avg_Price	Derived Numeric	Average modal price by state	RMA benchmark
Year	Derived Integer	Year extracted from Arrival_Date	YoY analysis
Month_Num	Derived Integer	Month number for sorting	Chronological ordering

## Dataset Size & Coverage

<b>6,000</b> Total Rows	<b>18</b> Columns	<b>8</b> States	<b>8</b> Markets	<b>8</b> Commodities	<b>7 Yrs</b> Coverage (2018-2026)
----------------------------	----------------------	--------------------	---------------------	-------------------------	---

## Data Limitations

- Restricted to 8 states; findings may not generalize pan-India.
- Granular quality parameters (moisture, foreign matter content) are absent.
- Transport, loading, and storage costs are not embedded in the dataset.
- Some Modal\_Price values were missing and required imputation using mid-range estimation.

## 5. Data Cleaning & Preparation

---

All primary cleaning and transformation steps were executed in Google Sheets (Tab 2: Data Dictionary & Cleaning Log), in compliance with the capstone requirement.

### Issues Identified

- **Mixed Date Formats:** Arrival\_Date had inconsistent formats (DD-MM-YYYY, MM/DD/YYYY, text). Standardized using Google Sheets TEXT and DATE functions.
- **Missing Modal Prices:** 150+ rows had missing Modal\_Price. Imputed using mid-range formula:  $=(\text{Min\_Price} + \text{Max\_Price})/2$ , documented in the cleaning log.
- **Inconsistent Text Casing:** State and Market names had 8+ variations (e.g., 'maharashtra', 'Maharashtra', 'MAHARASHTRA'). Standardized using PROPER() and TRIM() functions.
- **Trailing Whitespaces:** All categorical columns cleaned using TRIM() to prevent grouping mismatches in pivot tables.
- **Outlier Validation:** Modal prices below ₹10 or above ₹50,000 were flagged and reviewed. 3 records were corrected based on market context.

### Feature Engineering (Derived Columns)

- **Season:** Mapped month numbers to Kharif (Jun–Oct), Rabi (Nov–Mar), or Summer (Apr–May) using nested IF logic.
- **Month\_Name & Month\_Num:** Extracted from Arrival\_Date for monthly trend analysis and correct chronological sorting.
- **Price\_Spread:** Calculated as  $\text{Max\_Price} - \text{Min\_Price}$  for every row.
- **PPI (Price Position Index):**  $=(\text{Modal\_Price} - \text{Min\_Price}) / (\text{Max\_Price} - \text{Min\_Price})$ , with IFERROR wrapper for zero-spread rows.
- **State\_Avg\_Price:** AVERAGEIF by state, used as the RMA baseline.
- **RMA\_Index:**  $\text{Modal\_Price} - \text{State\_Avg\_Price}$  to measure structural market advantage.
- **Profit\_Opportunity\_Flag:** IF(Modal\_Price > 75th percentile AND RMA > 0 AND PPI > 0.6, "High", "Normal").

### Assumptions

- Mid-range imputation for Modal\_Price is conservative and understates peak realization.
- Season classification follows government crop calendar definitions.
- Outlier threshold of ₹50,000/q was set based on commodity-specific maxima observed in the dataset.

## 6. KPI & Metric Framework

Five KPIs were engineered to directly support the project objective of maximizing farmer returns. Each KPI maps to a specific farmer decision.

KPI	Formula	Interpretation	Decision Relevance
Modal Price	Most frequent traded price (₹/q)	Primary revenue proxy per mandi and time period.	Farmer Decision: How much will I get?
Price Spread	Max_Price – Min_Price	Measures intraday price volatility. High spread = quality differentiation opportunity.	Risk & Aggregation Signal
Price Position Index (PPI)	(Modal – Min) / (Max – Min)	Normalized 0–1 score. Higher is better for sellers. PPI > 0.7 = strong negotiation; < 0.3 = distress.	Negotiation Strength
Relative Market Advantage (RMA)	Modal – State Average Price	Positive RMA = market consistently pays above state baseline. Guides where to sell.	Location Choice
Profit Opportunity Indicator	High / Normal (composite flag)	Binary signal based on Top 25% price + Positive RMA + PPI > 0.6. Triggers optimal selling action.	When & Where to Sell

### Strategic Synthesis

The framework combines RMA for location choice with PPI for timing execution. A positive RMA mandi with PPI > 0.6 defines the "optimal sell window" — the composite signal that drives the Profit Opportunity Indicator classification.



## 7. Exploratory Data Analysis (EDA)

### 7.1 Trend Analysis — Monthly & Yearly

Monthly trend analysis across all commodities reveals a consistent summer premium pattern. Modal prices peak in April–May across most commodities, driven by post-harvest supply tightening before the Kharif sowing cycle. Prices trough in November–January due to Rabi harvest influx and market glut conditions.

- April Average Modal Price: ₹2,166/q — highest monthly peak
- November Average Modal Price: ₹2,078/q — seasonal low point
- Year-over-Year Trend (2018–2026): Stable with minor inflationary drift, indicating reliable baseline planning conditions.

### 7.2 Regional Variation in Modal Prices

State-level analysis shows significant geographic variation. Bihar registers the lowest average modal prices while Maharashtra leads in absolute average price. Uttar Pradesh hosts the top-performing individual mandi — Lucknow — which commands a consistent premium above its state average.

<b>₹3,109.78</b> Lucknow Mandi (Top)	<b>₹2,175</b> Maharashtra State Avg	<b>₹490.20</b> Avg Price Spread	<b>0.50</b> Avg PPI (All States)
---	--	------------------------------------	-------------------------------------

### 7.3 Seasonal Variation by Commodity

Rabi crops consistently command better prices compared to Kharif or Summer harvests. This is particularly pronounced for wheat, potato, and onion. Cotton and soybean show the highest price volatility across seasons, with Cotton exhibiting spreads exceeding ₹800/q in peak periods.

- Rabi Season: Highest average modal price across all commodities.
- Kharif Season: Mid-range prices; higher supply volume suppresses price premium.
- Summer Season: Selective premium for perishables (tomato, onion) due to supply shortage.

### 7.4 Market Structural Analysis (RMA)

Relative Market Advantage (RMA) analysis identifies mandis that structurally outperform their state average regardless of seasonal trends:

- Mysuru Mandi: Highest positive RMA — consistently above Karnataka state average.
- Jaipur Mandi: Strong positive RMA in Rajasthan, particularly for Rabi crops.
- Kolkata Market: Positive structural advantage driven by Bengal commodity premiums.
- Patna Bazaar: Negative RMA — consistently underperforms Bihar state average; avoid for premium commodities.

### 7.5 Distribution & Correlation Analysis

- Price Spread Distribution: Right-skewed, with Cotton and Tomato showing extreme high-spread outliers — indicative of quality sorting opportunities.
- PPI Distribution: Approximately normal (mean 0.50), confirming balanced buyer-seller power on average with room for seller improvement through market timing.
- RMA vs Modal Price Correlation: Moderate positive correlation ( $r \approx 0.65$ ), confirming that structurally better markets also deliver higher absolute prices.

- Seasonality Impact: ANOVA across seasons shows statistically significant price differences ( $p < 0.05$ ), validating seasonal strategy as an evidence-based recommendation.

## 8. Advanced Analysis

---

### 8.1 Profit Opportunity Segmentation

A composite binary classification — Profit Opportunity Indicator — was engineered using three threshold conditions: Top 25% Modal Price for the commodity, Positive RMA, and  $PPI > 0.6$ . This creates a "High Opportunity" flag that identifies the best 30–40% of selling windows in the dataset.

- Maharashtra: 42.8% of transactions flagged as High Opportunity
- Uttar Pradesh: 42.3% High Opportunity share
- Rajasthan: 41.6% High Opportunity share

This segmentation allows farmers and FPOs to target states and months where the probability of a premium realization is highest.

### 8.2 Price Position Index (PPI) Analysis

The average PPI of 0.50 across all records confirms a balanced market on average. However, commodity-level PPI analysis reveals significant variation:

- Soybean PPI: 0.58 — Sellers hold above-average negotiation power.
- Tomato PPI: 0.46 — High supply volatility reduces seller leverage.
- Wheat PPI: 0.52 — Stable, MSP-influenced floor provides negotiation room.

### 8.3 Volatility & Risk Analysis

Price Spread analysis (Average: ₹490.20/q) reveals significant intraday price variation — indicating either quality differentiation opportunities or information asymmetry. High-spread commodities (Cotton, Soybean) benefit most from grading and aggregation strategies. Low-spread commodities (Rice, Wheat) are volume-driven commodity markets.

### 8.4 Scenario Analysis — Optimal Sell Strategy

A scenario model was constructed to estimate the revenue impact of shifting from an unguided selling strategy to a data-driven one:

- Baseline: Farmer sells at the nearest market during harvest month (Kharif/Oct) — captures ~48th percentile price on average.
- Optimized: Farmer waits for the April–May window and sells at a state-positive RMA market — captures ~72nd percentile price on average.
- Estimated Revenue Uplift: 8–14% per quintal, depending on commodity and state.

## 9. Dashboard Design

---

The dashboard was implemented in Google Sheets with Pivot Tables, AVERAGEIF/COUNTIF formulas, and interactive Slicer filters. The design follows a two-view architecture:

### Executive View (Top Strip)

A KPI card strip at the top of the dashboard surfaces the six most decision-critical metrics at a glance, updating dynamically with slicer selections:

- Highest Avg Modal Price: ₹3,109.78 — Lucknow Mandi benchmark
- Top Performing Market: Lucknow Mandi
- Peak Selling Month: May
- Highest Paying Season: Rabi
- Avg Price Position Index: 0.5
- Average Market Volatility: ₹490.20

### Operational View (Chart Grid)

The chart grid contains six analytical visualizations covering all dimensions of the decision framework:

- Regional Variation in Modal Prices — Horizontal bar chart by state for quick market ranking.
- Seasonal Variation by Commodity — Grouped bar chart (Kharif / Rabi / Summer) for all 8 commodities.
- Markets with Highest Structural Price Advantage — Diverging bar chart of RMA scores by market.
- State-Level Price Position Index — Column chart showing average PPI by state.
- Monthly Trend of Modal Prices — Line chart with 12-month axis for seasonal pattern identification.
- Profit Opportunity Across States — Stacked bar chart (High vs Normal) for opportunity concentration.

### Slicer Filters

Three interactive slicers allow drill-down at any intersection of Commodity, Month, and Season — enabling both macro-level pattern recognition and micro-level selling decision support.

### Dashboard Objective

The dashboard converts raw mandi price data into a Farmer Decision Support interface — any farmer or FPO advisor can select their commodity and state to immediately see where and when to sell for maximum returns.

## 10. Insights Summary

The following 10 insights are stated in decision language, mapped from analytical findings:

#	Insight	Implication
1	Lucknow Mandi commands the highest average modal price (₹3,109.78/q) among all 8 markets analyzed.	FPOs in northern India should route aggregated produce to Lucknow as a priority sales node.
2	May is the peak selling month — average prices are highest due to post-harvest supply tightening.	Farmers with storage capacity should withhold non-perishables until April–May to capture the Summer Premium.
3	Rabi season consistently delivers higher prices than Kharif or Summer across all commodity categories.	Crop planning should favor Rabi varieties where soil and water conditions permit, to capture seasonal premium.
4	Maharashtra has the highest share of High-Profit Opportunity windows (42.8%), followed by UP (42.3%).	Policy investment in transport subsidies to Maharashtra mandis can directly improve farmer income realization.
5	Average PPI of 0.50 across all records confirms that sellers are currently mid-range in negotiation power — a 0.5-margin exists for improvement.	Collective selling through FPOs can shift the PPI closer to 0.7+ through volume-based negotiation leverage.
6	Mysuru, Jaipur, and Kolkata show consistently positive RMA — they structurally outpay their state averages.	These three markets should be priority routing targets for commodities with long shelf life and transport feasibility.
7	Patna Bazaar shows a negative RMA — it consistently underperforms Bihar state average prices.	Bihar-based farmers should prioritize alternative mandis or aggregation hubs rather than Patna Bazaar for premium realizations.
8	Cotton and Soybean exhibit the highest price spread (₹700–₹900/q range) — indicating significant intraday price variation.	Sorting and grading investments for Cotton and Soybean before mandi arrival can capture the quality premium embedded in the high spread.
9	November–January represents the seasonal price trough — prices average ₹2,078/q in November vs ₹2,166/q in April.	Non-perishable crop holders should avoid distress selling in this window; storage financing schemes can bridge the gap.
10	A data-driven selling strategy targeting high-RMA markets during April–May improves price percentile from ~48th to ~72nd.	An estimated 8–14% per quintal revenue uplift is achievable through market intelligence adoption without any change in crop input costs.

## 11. Recommendations

Each recommendation is directly mapped to a supporting insight and framed with business impact and feasibility:

Recommendation	Linked Insight	Business Impact	Feasibility
Align bulk sales with April–May window to capture Summer Premium.	Insight 2	₹88–₹200/q additional realization vs. October sales.	High — requires only storage and market access.
Route produce to Maharashtra mandis for structural price advantage.	Insights 4, 6	Access to 42.8% High-Opportunity transactions.	Medium — transport cost must be weighed against price gain.
Prioritize Lucknow Mandi as the northern India sales hub.	Insight 1	₹3,109.78/q average — benchmark realization node.	High — already established; FPO partnerships needed.
Withhold non-perishable stock during Nov–Jan to avoid market glut.	Insight 9	Avoidance of ₹88+/q seasonal price loss.	Medium — requires access to affordable storage/warehousing.
Invest in produce sorting and grading for Cotton and Soybean.	Insight 8	Potential to capture upper quartile of ₹700–900/q spread.	High — grading infrastructure investment with clear ROI.
Use composite PPI > 0.6 + Positive RMA as the sell trigger signal.	Insights 5, 7	Improves average price percentile by ~24 percentage points.	High — signal is dashboard-enabled; no additional data needed.

## 12. Impact Estimation

---

### Cost Savings

Routing to positive-RMA markets (Mysuru, Jaipur, Lucknow) instead of the nearest mandi captures state-level price premiums. Based on dataset evidence, this translates to ₹150–₹400/q additional realization per transaction depending on the commodity and state combination.

### Revenue Improvement

Seasonal targeting — shifting sales from Oct (Kharif trough) to May (Summer peak) for non-perishables — results in an estimated 8–14% revenue uplift per quintal. For a small farmer with 10 quintals of soybean, this translates to approximately ₹2,000–₹4,000 additional income per crop cycle with no change in input costs.

### Efficiency Gains

Adopting the dashboard-driven selling model replaces 3–5 days of market inquiry through intermediaries with a real-time decision in minutes, improving planning efficiency and reducing dependency on commission agents.

### Risk Reduction

The Profit Opportunity Indicator classifies 40%+ of transactions in Maharashtra, UP, and Rajasthan as High-Opportunity windows. Farmers operating in these states with storage access can reduce distress sale frequency by avoiding the 20–25% of transactions that fall in identified low-price windows.

## 13. Limitations

---

- **Restricted Geographic Scope:** Analysis covers 8 states and 8 mandis. Regional generalizations may not hold for Tier-2 and Tier-3 mandis or states not represented.
- **Data Imputation Risk:** 150+ Modal\_Price values were imputed using mid-range estimation. This approach is conservative and may understate peak price realizations in some markets.
- **Quality Parameters Absent:** Variety field captures some grades, but granular quality indicators (moisture content, foreign matter percentage) are not available, limiting quality-premium analysis accuracy.
- **Cost Model Exclusions:** Net realization estimates do not account for variable transport, loading, storage, or commission agent costs, which can be significant for cross-state routing recommendations.
- **Temporal Representativeness:** 2018–2026 data includes COVID-19 disruption years (2020–2021) which may create anomalous price patterns that influence averages.



## 14. Future Scope

---

- **Expand Coverage:** Scale data ingestion to include all major Tier-1 and Tier-2 mandis across India via direct Agmarknet API integration. This would reduce regional bias and enable pan-India applicability.
- **Net Realization Modeling:** Embed transport cost per km, loading charges, and storage cost into the price model to compute true Net Farmer Realization rather than gross Modal Price.
- **Rainfall & Weather Integration:** Correlate daily rainfall deviation and temperature anomalies with price spikes. This would significantly enhance the predictive power of the Profit Opportunity Indicator.
- **Predictive Forecasting:** Build a time-series forecasting model (ARIMA or Prophet) to predict next-month modal prices by commodity and market — enabling advance selling decisions.
- **Mobile Decision Tool:** Convert the dashboard into a mobile-first app (SMS-based or WhatsApp-accessible) to enable last-mile delivery of price intelligence directly to individual farmers.

## 15. Conclusion

---

This project successfully demonstrates that mandi price data, when properly cleaned, engineered, and visualized, can be converted into a powerful Farmer Decision Support System. By analyzing 6,000 records across 8 states, 8 markets, 8 commodities, and 7 years, the team identified actionable patterns in seasonal pricing, geographic market advantage, and commodity-level volatility.

The five KPIs engineered — Modal Price, Price Spread, PPI, RMA, and Profit Opportunity Indicator — collectively answer the two fundamental questions every farmer faces: When to sell? and Where to sell? The interactive dashboard operationalizes these signals into a filterable, drill-down interface that any FPO advisor or APMC committee can use for decision support.

The analysis validates that an 8–14% revenue uplift per quintal is achievable through market intelligence adoption alone — without any change in crop inputs, farming practices, or post-harvest technology. The key enabler is data: reliable, cleaned, structured, and visualized in a format that translates to selling action.

This capstone demonstrates the direct policy and business value of data analytics in agriculture — one of India's most data-underserved sectors — and establishes a replicable analytical framework that can be scaled to a national mandi intelligence platform.

# Appendix

## A. Data Dictionary (Full Column Reference)

Column	Type	Source	Cleaning Action
Arrival_Date	Date	Raw	Standardized to DD-MM-YYYY format
State	Text	Raw	PROPER() + TRIM() applied; 8 variations unified
District	Text	Raw	TRIM() applied
Market	Text	Raw	PROPER() + TRIM() applied
Commodity	Text	Raw	Standardized 8 commodity names
Variety	Text	Raw	Retained as-is; not used in primary KPIs
Min_Price	Numeric	Raw	Outlier validated (range check)
Max_Price	Numeric	Raw	Outlier validated (range check)
Modal_Price	Numeric	Raw → Imputed	150+ missing values imputed via mid-range
Season	Categorical	Derived	IF(Month<=3, Rabi, IF(Month<=5, Summer, Kharif))
Month_Name	Text	Derived	TEXT(Arrival_Date, "MMMM")
Month_Num	Integer	Derived	MONTH(Arrival_Date) for sort ordering
Year	Integer	Derived	YEAR(Arrival_Date)
Price_Spread	Numeric	Derived	=Max_Price – Min_Price
State_Avg_Price	Numeric	Derived	=AVERAGEIF(State_col, State, Modal_col)
PPI	Numeric	Derived	=IFERROR((Modal–Min)/(Max–Min), 0.5)
RMA_Index	Numeric	Derived	=Modal_Price – State_Avg_Price
Profit_Opportunity_Flag	Text	Derived	Composite IF: Top25%+RMA>0+PPI>0.6 → High

## B. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Khushi (2401010225)	Yes	Yes	Yes	Yes	-	-	Data Lead
Abhinav (2401010017)	-	Yes	Yes	-	-	Yes	Analyst
Anshika (2401010080)	-	-	Yes	Yes	-	-	Viz Expert

Kshitiz (2401010242)	-	-	-	Yes	Yes	Yes	Strategist
Abhijeet (2401010014)	Yes	-	-	Yes	Yes	-	Editor
Nishant (2401010302)	Yes	-	-	-	Yes	-	Researcher

Declaration: We confirm that the above contribution details are accurate and verifiable through Google Sheets Version History and submitted project artifacts.

*Team Signature: Group 4, Section B — Newton School of Technology | February 18, 2026*