Market Data Analysis Report

1. Dataset Description

1.1 Source: Government agricultural market dataset containing daily arrivals and prices of various commodities across Indian states and districts.
Total records: **6,205** (collected from multiple regional markets).

1.2 Columns:

State – Represents the Indian state (e.g., Andhra Pradesh, Gujarat, Kerala).

District – Administrative district under the respective state.

Market – Local agricultural market name.

Commodity – Product traded (e.g., Tomato, Potato, Onion, Brinjal, Banana, etc.).

Variety – Type or sub-category of the commodity (e.g., Hybrid, Local).

Grade – Quality specification (FAQ, Standard, Non-FAQ).

Arrival_Date – Date of market arrival/recording.

Min_x0020_Price – Minimum selling price on that day (in ₹ per quintal).

Max_x0020_Price – Maximum selling price on that day.

Modal_x0020_Price – Most common (modal) market price, used as a benchmark.

1.3 Data Quality:

- Clean and structured with consistent column names.
- Numeric conversions applied for price columns.
- Minor formatting issues resolved (column renaming, data type conversions).
- No major missing values after preprocessing.

2. Operations Performed

2.1 Data Cleaning & Exploration

- Removed special characters from column names.
- Converted price columns to numeric types.
- Parsed Arrival_Date into standard date format.
- Checked unique values for categorical columns (State, Commodity, Market, etc.).
- Verified distribution of records across states and commodities.

2.2 Descriptive Analytics

Visualizations and metrics were created using PySpark DataFrames and Matplotlib/Seaborn:

- **Commodity Distribution** – Bar & pie charts showing most frequently traded commodities.
- **State-wise Market Activity** – Horizontal bar chart comparing total records per state.
- **Price Distribution** – Histogram and KDE plot showing spread of modal prices.
- **Market-wise Analysis** – Top 10 active markets visualized using bar graphs.
- **Grade-wise Variation** – Violin and box plots highlighting price variability by grade.

2.3 Relationship Analysis

- **Min vs. Max Price Correlation:** Scatter and heatmap plots show strong positive relation (~0.9).
- **Commodity vs. Modal Price:** Bar plots identify premium commodities with higher average modal prices.
- **Trend Analysis:** Line chart displays price changes over time for key commodities (e.g., Tomato, Onion).
- **State vs. Average Price:** Horizontal bar chart for regional comparison in pricing.

3. Key Insights

3.1 Commodity Insights

- Over 50 commodities recorded; Potato, Onion, Wheat, Tomato, and Brinjal dominate dataset frequency.
- **High-value commodities** such as Cardamom, Almond, Pepper, Cashewnut, and Coconut Oil command much higher modal prices.
- Perishable items (Tomato, Banana) show wide daily price variations due to supply fluctuations.

3.2 Workforce Demographics

- Dataset covers **15+ Indian states**, with **Uttar Pradesh, Kerala, Gujarat, Maharashtra, and Himachal Pradesh** leading in record count.
- **Tripura and Kerala** show higher average modal prices (>₹5,000) due to premium commodities (spices, coconut, cardamom).
- **Northern and Western states** have larger market volumes, while **Southern states** show higher per-unit prices.

3.3 Price Insights

- **Modal Price Range:** ₹5 – ₹1,87,500
- **Average Modal Price:** ~₹3,942
- **Median Price:** ₹3,000
- **Standard Deviation:** ~₹4,845 → indicates significant variability in market pricing.
- **Correlation (Min–Max–Modal):** Strong positive trend

3.4 Market & Variety Insights

- **Top Markets:** Cuddapah, Tiruvuru, Vayalapadu, and Palamaner show maximum trading activity.
- **Common Grades:** FAQ and Standard dominate, indicating regular market quality.
- **Popular Varieties:** Local, Hybrid, and Sona are most traded across commodities.

3.5 Temporal Trends

- Prices exhibit short-term fluctuations influenced by seasonality and local supply-demand.
- Daily modal price patterns confirm **spike behavior** for high-demand commodities like Tomato and Onion.

4.Recommendations

4.1 Price Stability & Forecasting

- Apply **time-series forecasting models (ARIMA, Prophet)** to predict future prices for volatile commodities.
- Use predictive analytics to alert farmers about upcoming price surges or declines.

4.2 Commodity Optimization

- Promote diversification towards high-profit commodities (e.g., Spices, Dry Fruits) in states with suitable climate conditions.
- Introduce regional specialization — e.g., Kerala for Cardamom, Gujarat for Groundnut.

4.3 Market Efficiency

- Encourage digital market reporting systems for **real-time price tracking**.
- Implement uniform grading and quality assurance to reduce regional price disparities.

4.4 Data-driven Decision Making

- Use analytics dashboards to help **policymakers** identify supply-demand imbalances.
- Identify **underperforming markets** (low price or volume) for targeted interventions.

4.5 Future Analytics Opportunities

- Build **commodity price prediction** models using PySpark MLlib.
- Perform **clustering** of markets based on price similarity.
- Correlate **climate and rainfall data** with commodity prices for improved yield and pricing strategy.
- Integrate sentiment analysis from **news or farmer feedback** to study price impact factors.