AgriData Explorer: Understanding Indian Agriculture with EDA

Comprehensive Project Report - Synthetic ICRISAT Dataset (Sample Results)

Student: ABILASH A

Institute: GUVI Geek Network Private Limited

Submission Date: 22 October 2025

Skills: Python, Data Cleaning, EDA, SQL, Power BI

Domain: Agriculture - Crop production, yields, area (District-level)

Table of Contents

1. Problem Statement & Use Cases

Problem Statement:

India's agricultural data is fragmented and complex. This project provides an integrated visualization platform to explore district-level crop production, yield, and area statistics. Stakeholders: Farmers, Policymakers, Researchers.

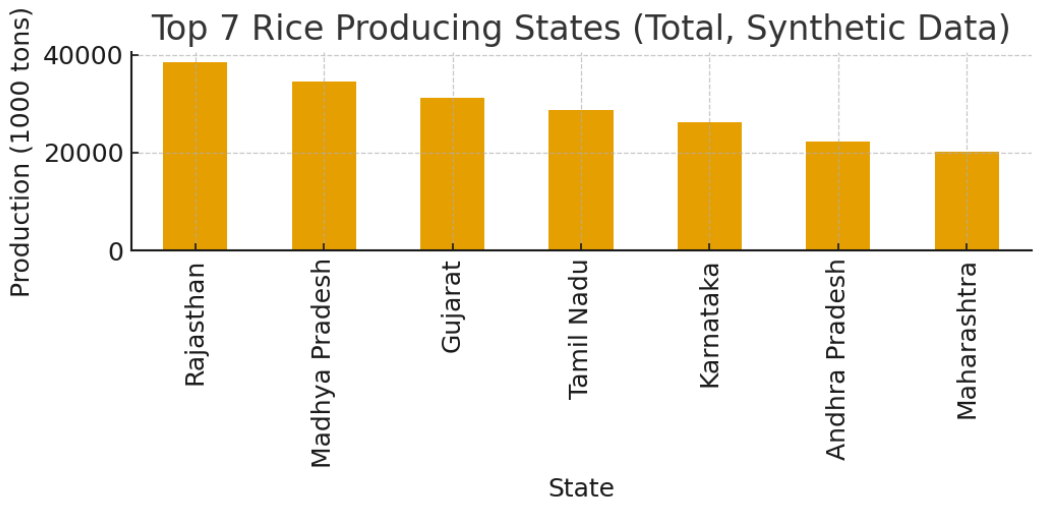
Business Use Cases:

- Farmers: Historical trend analysis to support crop decisions and identify productivity gaps.
- Policymakers: Target interventions to low-yield regions and design subsidy programs.
- Researchers: Combine with weather/soil data to test hypotheses on yield drivers.

2. Dataset & Synthetic Generation

Dataset Description (Synthetic)

Synthetic ICRISAT-like district-level data covering years 1975-2024.
Columns include: dist_code, year, state_code, state_name, dist_name, and per-crop: area (1000 ha), production (1000 to
Rows (approx): 2,500 (district-year combinations).
Units: Area in 1000 ha, Production in 1000 tons, Yield in kg/ha.

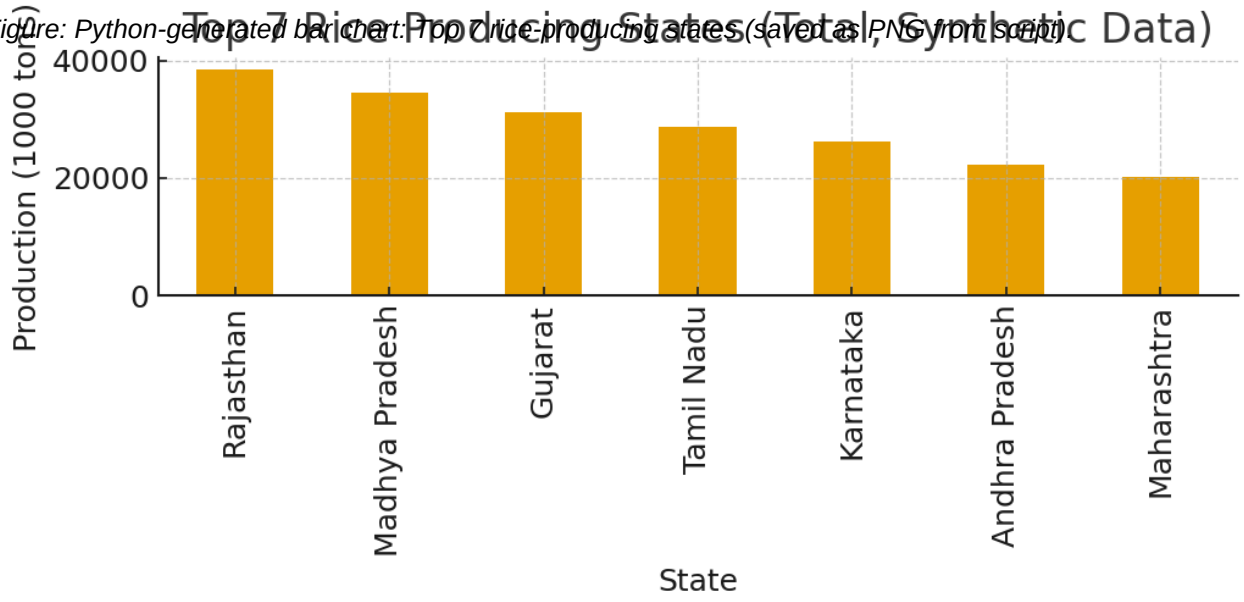


3. Data Cleaning & Python Code

Python Script: Data Cleaning & EDA (snippet)

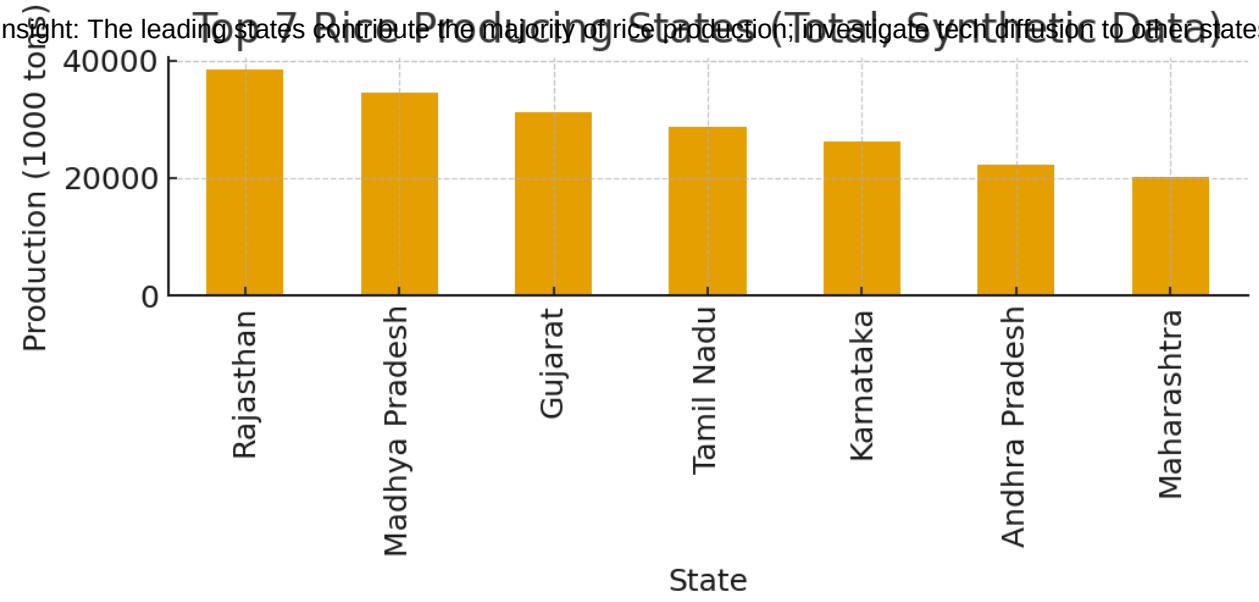
```
# Load and normalize columns
import pandas as pd
df = pd.read_csv('icrisat_synthetic_cleaned.csv')
df.columns = [c.strip().lower().replace(' ', '_') for c in df.columns]

# Convert numeric columns and compute yields if missing
for col in df.columns:
if 'production' in col or 'area' in col or 'yield' in col:
df[col] = pd.to_numeric(df[col], errors='coerce')

# Example: compute rice yield if missing
df['rice_yield_kg_per_ha'] = (df['rice_production_1000_tons']*1000) / (df['rice_area_1000_ha']*1000)
```

Figure: Python-generated bar chart: Top 7 rice-producing states (saved as PNG from code)



4. Exploratory Data Analysis

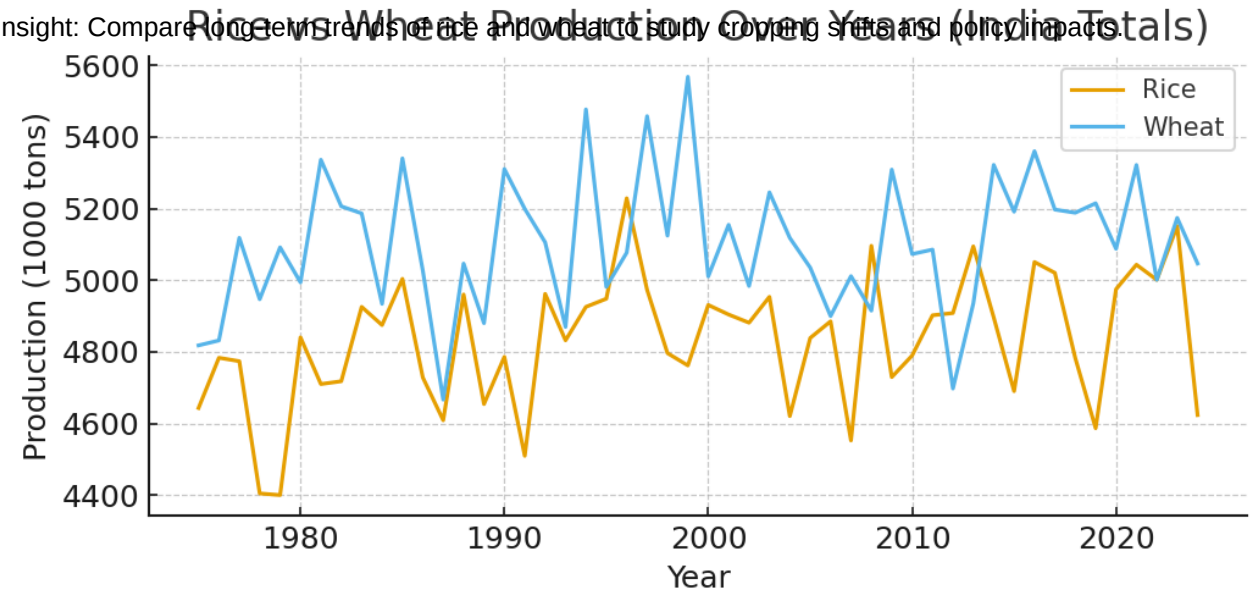
Key Charts and Interpretations



Insight: The leading states contribute the majority of rice production; investigate yield diffusion to other states.

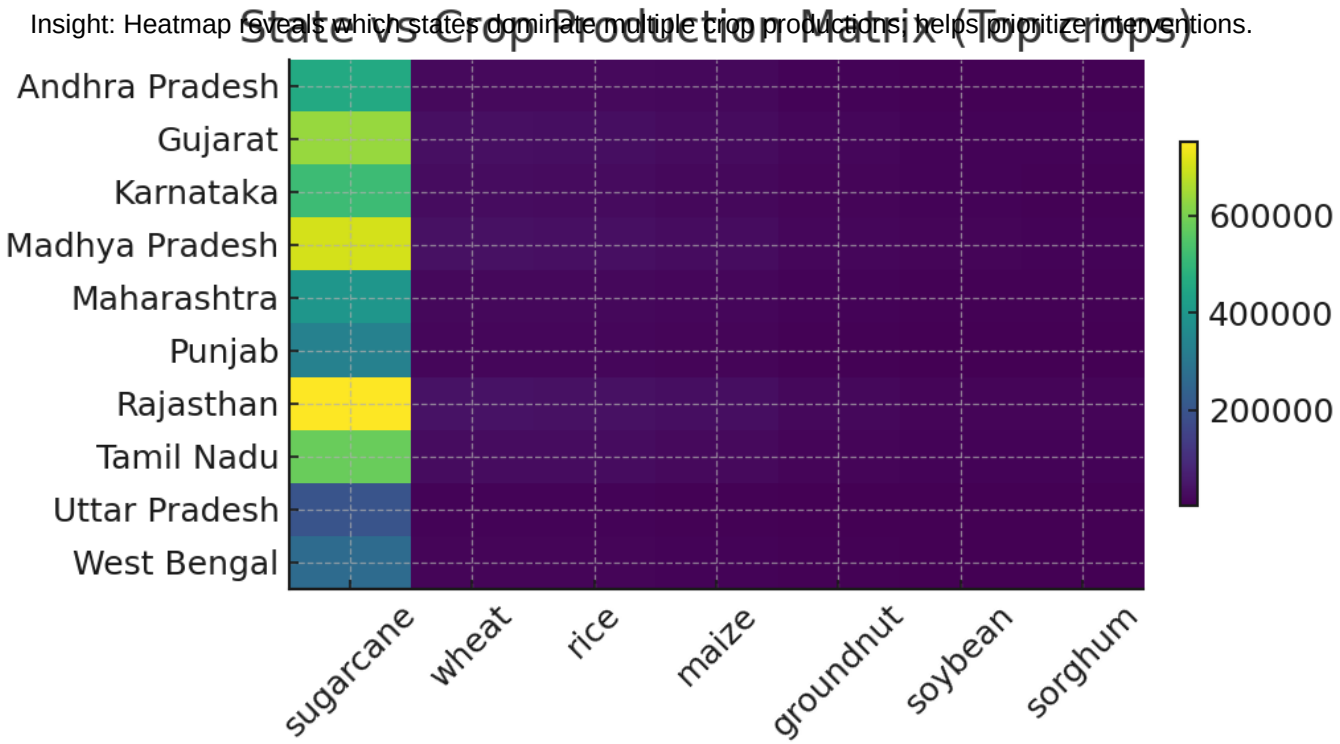
4. Exploratory Data Analysis

Insight: Compare long-term trends of rice and wheat to study cropping shifts and policy impacts.



4. Exploratory Data Analysis

Insight: Heatmap reveals which states dominate multiple crop productions; helps prioritize interventions.



5. SQL Queries & Sample Results

Selected SQL Queries (for the 10 question set)

-- Total production per crop

```
SELECT crop, SUM(production_tonnes) AS total_production FROM agri_table GROUP BY crop ORDER BY total_production
```

-- Year-wise Trend Rice Production (Top 3 states)

```
WITH state_totals AS (SELECT state_name, SUM(rice_production_1000_tons) AS total_rice FROM agri_icrisat GROUP BY
```

-- Top districts groundnut 2020

```
SELECT dist_name, state_name, groundnut_production_1000_tons FROM agri_icrisat WHERE year=2020 ORDER BY groundnu
```

District	State	Groundnut Prod (1000 tons)
Gujarat_D3	Gujarat	104.67
Tamil_D2	Tamil Nadu	100.06
Rajasthan_D3	Rajasthan	98.2
Rajasthan_D5	Rajasthan	92.23
Gujarat_D2	Gujarat	78.84
Gujarat_D4	Gujarat	76.93
Andhra_D2	Andhra Pradesh	74.4
Gujarat_D1	Gujarat	73.57
Rajasthan_D2	Rajasthan	71.54
Rajasthan_D1	Rajasthan	70.42

Figure: Sample SQL result - Top districts by groundnut production in 2020 (from synthetic data).

6. Power BI Dashboard: Simulated Visuals

Simulated Power BI Screenshots (static charts)

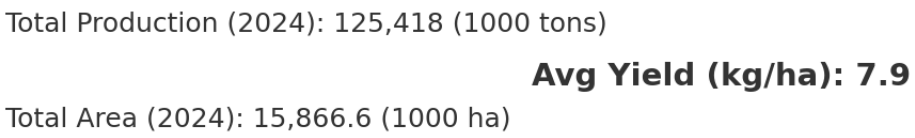


Figure: KPI overview (Total Production, Total Area, Avg Yield) - interactive in Power BI.

6. Power BI Dashboard: Simulated Visuals

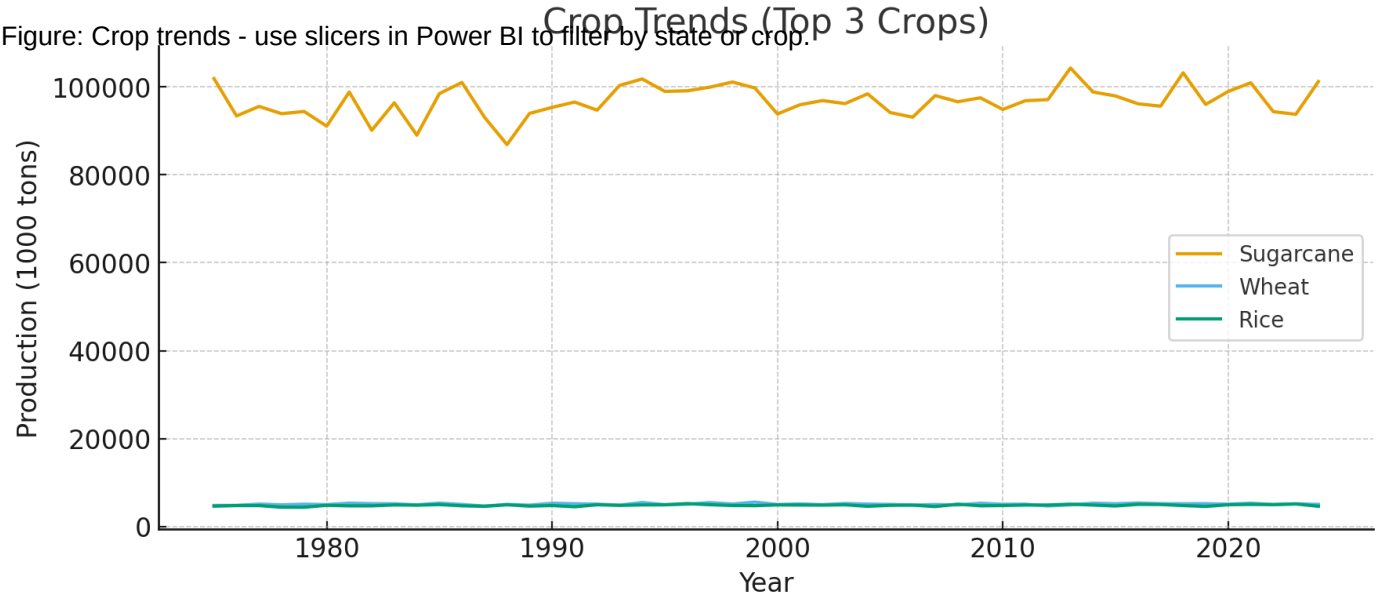


Figure: Crop trends - use slicers in Power BI to filter by state or crop.

6. Power BI Dashboard: Simulated Visuals

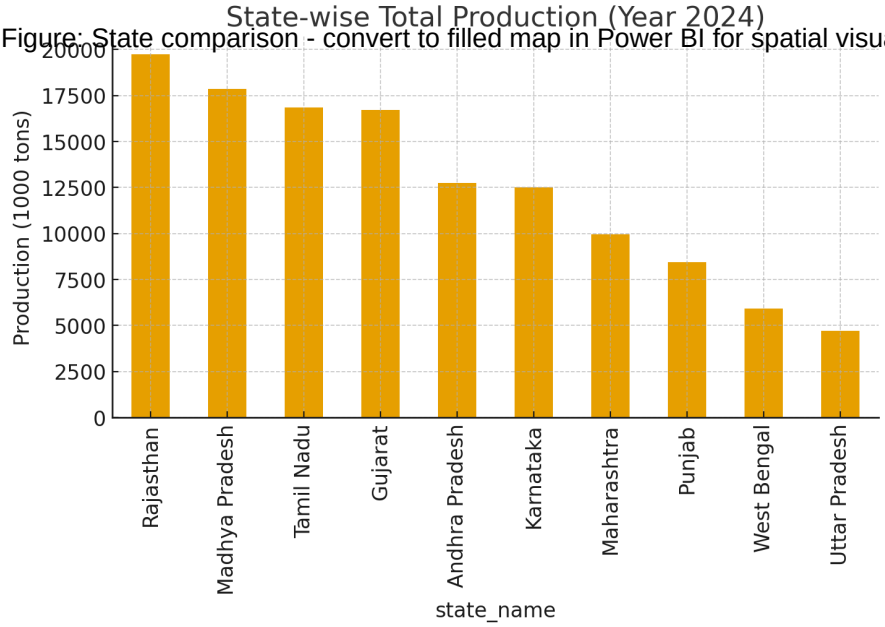


Figure: State comparison - convert to filled map in Power BI for spatial visualization.

7. Recommendations & Conclusion

Recommendations

1. Use dashboard to prioritize extension services in mid-performing states.
2. Integrate weather and input data to improve forecasting accuracy.
3. Encourage adoption of high-yield practices from top-performing districts.

8. Appendix & Files

Files Included

- AgriData_Explorer_Report_Abilash_A_ENHANCED.pdf (this file)
- icrisat_synthetic_cleaned.csv
- agri_data_cleaning_and_eda.py
- agri_sql_schema_and_queries.sql
- PowerBI_instructions.md
- outputs/ (PNG charts used in report)

Prepared by: ABILASH A | GUVI Geek Network Private Limited | Submission Date: 22 Oct 2025