

# **Horticulture Analytics Report — Horticulture (2022–2023)**

## **1. Executive summary**

This report summarizes exploratory and descriptive analytics performed on the horticulture dataset for 2022–2023 (source: Horticulture (2022-2023).ipynb). It covers data description and quality, data cleaning steps, key metrics (total production, total area harvested, average yield per hectare, revenue), crop & region performance, seasonality and trends, producer segmentation, risk signals (pests/returns/losses), and actionable recommendations for agronomy, supply chain and policy.

**Topline (example placeholders — replace with values from the notebook):**

- Total production (tons): **[TOTAL\_PRODUCTION\_TONS]**
  - Total area harvested (ha): **[TOTAL\_AREA\_HA]**
  - Average yield (t/ha): **[AVERAGE\_YIELD]**
  - Total estimated revenue (local currency): **[TOTAL\_REVENUE]**
  - Top 3 crops by production: **[CROP\_A], [CROP\_B], [CROP\_C]**
  - Peak production months: **[MONTHS]**
- 

## **2. Dataset description**

**Source:** Horticulture (2022-2023).ipynb (project data tables / CSV exports embedded in the notebook).

**Typical columns expected** (adapted from the notebook):

- record\_id — unique row id
- farm\_id / farmer\_id
- region / district / village
- crop / variety
- planting\_date / harvest\_date
- area\_ha — area harvested (hectares)
- production\_tons — production in metric tons (or kg)
- yield\_t\_per\_ha — derived (production / area)

- market\_price — price at sale time (per ton)
- revenue — derived (production \* price)
- input\_cost — cost of seeds, fertilizer, labor (if present)
- irrigation — irrigation type (rainfed / drip / flood)
- pest\_incident / loss\_percent — flagging losses

## 2.1 Data quality (summary)

- Missing values: (report which columns have NAs and counts) — **[MISSING\_SUMMARY]**
  - Duplicates: number of duplicate records removed — **[DUPLICATES\_REMOVED]**
  - Date parsing: planting/harvest dates parsed and validated.
  - Outliers: extremely high yields or negative production rows inspected and either corrected or removed.
  - Units: production units normalized to tons and area to hectares.
- 

## 3. Operations performed

### 3.1 Data cleaning & preprocessing

(Short narrative of transformations; replace placeholders with counts from notebook)

- Removed test/placeholder rows: **[N\_ROWS\_REMOVED]**
- Removed records with missing farm\_id or production\_tons: **[N\_CRITICAL\_MISS]**
- Derived fields: harvest\_month, harvest\_week, yield\_t\_per\_ha, revenue.
- Standardized crop names/varieties (mapped synonyms).
- Aggregations computed at: daily / monthly / crop-level / region-level / farmer-level.

### 3.2 Descriptive analytics & visualizations

- Time-series of monthly production and revenue.
- Top crops by production and revenue (bar charts / Pareto).
- Region-wise production maps or bar charts.
- Yield distribution per crop (boxplots / histograms).
- Price trend lines (monthly average price per crop).

### 3.3 Producer analytics (farmer-level)

- Farmer segmentation: by production volume, revenue, and yield efficiency (High/Medium/Low producers).
- Cohort analysis (e.g., farmers who started producing in 2022 and retention in 2023).
- Basic CLV-like estimate: lifetime revenue per farmer (if multiple seasons present).

### 3.4 Advanced analyses

- Seasonality & trend decomposition for major crops (monthly decomposition).
  - Yield drivers: correlation of yield with irrigation type, fertilizer usage, and input cost.
  - Risk signals: crops/regions with high loss rates or pest incidents.
  - Pairwise crop planting patterns (crop rotation / co-occurrence) if multi-crop per farm data available.
- 

## 4. Key insights (high-level)

*(Replace bracketed items with exact numbers/tables/figures from notebook)*

### 4.1 Production & revenue

- **Total production (2022–2023):** [TOTAL\_PRODUCTION\_TONS].
- **Total revenue:** [TOTAL\_REVENUE].
- **Average yield across all crops:** [AVERAGE\_YIELD] (t/ha).
- **Monthly seasonality:** Peak harvest during [MONTHS]; low months [MONTHS].

### 4.2 Crop & variety performance

- **Top revenue-generating crops:** [CROP\_1], [CROP\_2], [CROP\_3] (with revenue shares).
- **Top yield per ha:** [CROP\_HIGH\_YIELD] ([YIELD\_VALUE] t/ha).
- **Pareto observation:** top 20% of crop-varieties contribute [X%] of total production/revenue.

### 4.3 Farmer behavior & segmentation

- High-performing farmers (top 10% by production) produce [X%] of the volume.

- Median farmer farm size: **[MEDIAN\_AREA\_HA]**.
- Repeat-season producers vs single-season entrants: **[PERCENT\_REPEAT]**.

#### **4.4 Geography & irrigation**

- Regions with highest production: **[REGION\_A]**, **[REGION\_B]**.
- Irrigation impact: farms with drip irrigation show **[X%]** higher average yield than rainfed.

#### **4.5 Operational and risk signals**

- Loss concentration: Specific crops/regions showed higher pest incidence (e.g., **[CROP\_OR\_REGION]** with **[LOSS%]**).
  - Price volatility: crop prices had standard deviation **[PRICE\_SD]**, affecting farmer revenue predictability.
- 

### **5. Recommendations**

#### **5.1 Production & agronomy**

- Promote irrigation practices (drip) in regions where they significantly increase yield.
- Provide targeted agronomic advisory for low-yielding varieties / farms with highest input cost.

#### **5.2 Market & revenue**

- Encourage value-chain linkages for top revenue crops to stabilize prices in off-peak months.
- Bundle produce (cooperative aggregation) to reduce price volatility and reduce marketing costs.

#### **5.3 Farmer support & retention**

- Target high-potential but low-yield farmers with extension programs.
- Implement early-warning pest monitoring in regions flagged with high losses.

#### **5.4 Supply chain & operations**

- Prioritize cold-chain investments for perishable high-revenue crops in top producing districts.
- Improve quality-grading and packaging for crops with high return/quality issues.

### **5.5 Analytics roadmap**

- Build predictive yield models using weather, irrigation, input-cost features.
- Implement a live dashboard (PowerBI / Tableau) with region/crop filters, monthly KPIs, and farmer-level drilldowns.
- Create A/B trials for improved seed/fertilizer packages and measure yield uplift.