

REPORT: SEASONALITY OF FARMER QUESTIONS IN KENYA, UGANDA AND TANZANIA (WEFARM / PRODUCERS DIRECT)

Analysis based on the complete, stable years 2019–2020

1. INTRODUCTION

The purpose of this report is to provide a detailed analysis of the seasonality observed in farmer questions submitted to the WeFarm platform across three East African countries: Kenya (KE), Uganda (UG), and Tanzania (TZ). WeFarm operates as a digital knowledge-sharing system where farmers ask questions about crop production, livestock, pests and diseases, market prices, weather, and various other practical agricultural matters. Understanding the seasonality of these questions makes it possible to determine **when** farmers most urgently need information and **what kinds** of content are most relevant at different times of the year.

The analysis outlines how the number of incoming questions changes month by month, but more importantly, it reveals the *informational cycle* that farmers go through: from early questions about planting, to issues that arise during crop growth, to decisions related to harvesting and selling. This knowledge supports operational planning, such as designing SMS outreach, thematic campaigns, and country-specific communication schedules.

2. DATA SOURCES AND PREPARATION

2.1. Input data

The analysis is based on the aggregated pivot table `questions_aggregated_pivot_dashboard`, derived from the raw set of WeFarm farmer questions. Each row in the pivot table represents the number of questions for a specific combination of:

- year and month,
- country code,
- question language,
- dictionary-based category (`dictionary_category`),
- WeFarm topic group (`wefarm_topic_group`),
- original WeFarm topic (`wefarm_question_topic`),
- aggregated count (`count`).

This table serves as both the analytical dataset for this report and the primary data source for the Tableau dashboard.

2.2. Normalization and categorization

Before building the pivot table, the data underwent:

- cleaning and standardization of column names,
- unifying country and language codes,
- mapping WeFarm's detailed topics into simplified topic groups (`wefarm_topic_group`),
- assigning categories using the custom semantic dictionary (`dictionary_category`),
- extracting year and month from timestamps.

This structure ensures consistency across countries, years, and thematic categories.

2.3. Countries included in the analysis

The report includes only the countries present in the Tableau dashboard:

- **Kenya (KE)**
- **Uganda (UG)**
- **Tanzania (TZ)**

Other countries appearing in the raw dataset (e.g., GB) have extremely small volumes and were excluded because they are not relevant for seasonality analysis.

3. TEMPORAL STABILITY OF THE DATASET (2017–2022) AND JUSTIFICATION FOR THE CHOICE OF YEARS

Although the full dataset spans 2017–2022, the years differ substantially in completeness, stability, and overall question volume.

To conduct a reliable seasonality analysis, it is essential to use only years that meet two criteria:

1. **A complete 12-month cycle is available,**
2. **Question volume reflects actual farmer behaviour,** not platform-driven effects (growth phase, decline phase, partial shutdown, etc.).

Below is the assessment of each year.

3.1. Year 2017 – only late-year data available

- The dataset begins only in the final months of 2017.
- All earlier months are missing.
- A full annual cycle cannot be reconstructed.
- Cross-category or cross-country comparisons would be invalid.

3.2. Year 2018 – platform growth distorts the data

In 2018, the number of questions grows rapidly toward the end of the year because the platform was still scaling:

- low activity in early 2018,
- accelerating uptake later in the year,
- question volume reflects *platform growth*, not agricultural seasonality.

Including 2018 would incorrectly distort seasonal patterns.

3.3. Year 2019 – first complete and stable year

This is the first year that:

- contains all 12 months,
- shows stable activity,
- has sufficient volume in every month,
- is not affected by growth or decline effects.

It is the first truly reliable year for seasonality analysis.

3.4. Year 2020 – second complete and stable year

Similar to 2019:

- complete with all 12 months,
- high, consistent question volume,
- clear and interpretable monthly patterns.

This year enables meaningful year-to-year comparisons.

3.5. Year 2021 – sharp and unnatural decline in question volume

In 2021, platform activity drops significantly:

- volumes decrease sharply,
- the pattern appears to reflect **platform decline**, not changes in farmer behaviour,
- monthly values lose their agricultural interpretability.

3.6. Year 2022 – only the first months available

- The dataset contains only the beginning of the year, with very few questions,
- no complete cycle,
- cannot be used for seasonal analysis.

3.7. Analytical decision

Only two years — 2019 and 2020 — fulfil the criteria of completeness and stability.

These are the only reliable years for studying true agricultural seasonality.

All seasonality findings in this report are therefore based solely on **2019–2020**.

4. THEMATIC CATEGORIES AND THEIR RELATIVE IMPORTANCE

Across KE, UG and TZ (2019–2020), the distribution of dictionary-based categories is:

- **planting_growing** – ~31%
- **livestock** – ~21%
- **pests_disease** – ~17%
- **market_price** – ~10%
- **weather** – ~4%
- **timing_harvest** – ~3%
- **other** – ~14%

The platform's activity is dominated by crop-related topics (planting, growth, pest management), reflecting the typical needs of smallholder farmers.

5. OVERALL SEASONALITY (2019–2020 COMBINED)

When merging data for all three countries, a clear annual rhythm emerges:

- **January–February:** relatively low activity
- **March:** initial seasonal rise
- **April–May:** **primary high-season peak**, approx. 21% of annual questions
- **June:** drop following the early-season peak
- **July–August:** secondary rise
- **September–October:** stable, moderate level
- **November:** **late-season spike**, important across multiple categories
- **December:** moderately low activity

This shows a dual-peak annual cycle:
a strong **April–May** peak and a secondary **November** peak.

6. SEASONALITY BY THEMATIC CATEGORY

6.1. *Planting_growing (planting and crop growth)*

The most dominant category exhibits:

- strong peaks in **April and May**,
- elevated activity from **March through September**,
- an additional, unusually strong **November** rise in 2020.

Year-to-year comparison:

- **2019**: main peak in **May**,
- **2020**: main peak in **April**, plus a secondary surge in **November**.

This category defines the overall seasonal structure of the platform.

6.2. *Pests_disease (plant health issues)*

This category correlates with sensitive crop growth phases:

- primary peaks in **April–May**,
- secondary peaks in **July–August**,
- an additional late-year rise in **November 2020**.

Farmers seek advice on pests and diseases shortly after planting and during periods of rapid vegetative growth.

6.3. *Livestock*

This is the most evenly distributed category:

- relatively stable question volume across the entire year,
- subtle rise in **April**,
- minimal month-to-month volatility.

Livestock concerns do not follow the same strict crop-linked seasonality.

6.4. *Market_price*

This category shows two distinct seasonal windows:

- **May** and **July**,
- **November** — especially strong in 2020.

These patterns point to decision-making points for selling produce or engaging with local markets.

6.5. *Timing_harvest*

- most active in **July**, **August**, and **November**,
 - reflects periods when farmers reassess crop maturity and harvest windows.
-

6.6. *Weather*

- activity concentrated in **April** and **November**,
 - aligns with periods when weather conditions become critical for crop progress.
-

7. CROSS-COUNTRY DIFFERENCES

7.1. *Kenya (KE)*

- The most concentrated seasonal pattern among the three countries,
- high volumes from **January to May**,
- clear maximum in **April**,
- gradual decline after June,
- relatively low activity in the final months of the year.

7.2. *Uganda (UG)*

- The most evenly distributed question volume,
- noticeable peaks in **April** and **October**,
- no extreme spikes,
- category patterns are smoother than in KE.

7.3. *Tanzania (TZ)*

- Very distinct pattern compared to KE and UG,

- low activity in the early part of the year,
- major increases during **November–December**,
- many categories reach their highest values in late-year months.

Conclusion:

KE and UG concentrate their information needs in the first half of the year, while TZ shows a strong shift toward **late-year information demand**.

8. YEAR-TO-YEAR SHIFTS (2019 VS 2020)

Despite only two stable years, meaningful shifts are visible:

- `planting_growing` → peak shifts from **May (2019)** to **April (2020)**,
- `pests_disease` → same shift from **May** → **April**,
- `market_price` → **November** becomes particularly strong in 2020,
- several categories exhibit elevated late-year activity in 2020.

This demonstrates that seasonality is **not perfectly fixed**. Peaks can shift by one or two months, and new secondary peaks can appear.

9. OPERATIONAL IMPLICATIONS

9.1. Communication must be country-specific

- **Kenya and Uganda** → the central seasonal point is **April**; communication should ramp up from February onward.
- **Tanzania** → the dominant season is **November–December**; campaigns should focus on late-year timing.

9.2. Content should be delivered before seasonal peaks

- planting information → February–March for KE/UG, later for TZ,
- pest/disease alerts → just before April–May,
- harvest guidance → June–July,
- market price content → May, July, and November.

9.3. Seasonality should be reassessed annually

Because peak months shift between years, the communication calendar should be updated every year based on the latest seasonality signals.

10. LIMITATIONS

- Only two years (2019–2020) are suitable for seasonality analysis; earlier and later years were incomplete or unstable.
 - Variations in question volume across categories may affect monthly proportions.
 - Classification quality may vary by language.
 - Analysis is based on number of questions, not unique farmers.
-

11. SUMMARY

1. Farmer questions follow a clear annual rhythm with major peaks in **April–May** and a secondary peak in **November**.
2. Crop-related categories (planting_growing and pests_disease) dominate the platform's seasonality.
3. KE and UG share similar seasonal calendars, while TZ shows a strong late-year pattern.
4. 2020 exhibits measurable seasonal shifts, including earlier planting-related peaks and amplified November activity.
5. Operational communication should be **thematic, country-specific, and adjusted yearly** based on updated seasonality trends.