Farming Data Analysis

Ameya Ghatpande

EXPECTATION 1 - YIELD DATA

The expectation cannot be proved. The problems are -

- Most of the harvested_area columns have NaN/Null values

- Unit of harvested_area column not mentioned

- Same crop mentioned multiple times for same year, org and field ID

- Planted_area which seems like harvested_area also has NaN/Null values

EXPECTATION 2 - FIELD ACREAGE DATA

The expectation cannot be proved. The problems are -

- Each Field under a same organization has multiple field areas per year.

- If we GroupBy org_id, field_id, year and then add the field_area, we observe field areas are different for different years.

OVERALL DATA QUALITY ANALYSIS

- No 1:1 mapping between field data and yield data. There should be a primary key associated with each row
- Field_area_unit in the field table has only one value (acre). This column is not required
- Yield data (harvested_area) has Null values. This is very important data which should not be null
- Yield data (harvested_area) has no units mentioned
- Yield data (Planted_area) is very confusing and looks duplicate to harvested_area and also has Nulls.
- New_org_id in the toy_orgs data has no business value.

Thank you

Ameya Ghatpande