# TEAM DSS Post-Harvest Loss Dataset Documentation

## 1. Introduction

This document provides comprehensive documentation for the Post-Harvest Loss Dataset. This dataset was compiled as part of a research effort to understand key factors influencing post-harvest losses of selected agricultural commodities in Nigeria, with a particular focus on cultivation zones, storage conditions, and logistics. The dataset is based on a synthesis of existing research and collected field data, aiming to contribute valuable insights for improving post-harvest management strategies and reducing losses.

## 2. Research Methodology and Data Collection

The creation of this dataset is rooted in a multi-faceted research approach:

* **Foundational Research:** The research began with a review and synthesis of existing documentation concerning major crop cultivation zones in Nigeria and optimal post-harvest storage conditions and spoilage management for selected crops. This foundational work is detailed in the accompanying documents, "Nigerian Crop Cultivation by State\_.docx" and "Crop Storage Temperature and Spoilage\_ (1).docx". These documents, supported by various sources (including those recently provided, such as information on pepper cultivation in states like Kano, Kaduna, Jigawa, Katsina, Sokoto, Plateau, and Bauchi; tomato production in states like Kano, Kaduna, Katsina, Jigawa, Nasarawa, Plateau, Benue, Lagos, Oyo, Ogun, and Borno; cabbage production often sourced from northern states like Plateau, Zaria, Kano, Nasarawa and Benue; yam production concentrated in the "Yam Belt" across central/southern zones including Benue, Nasarawa, Niger, Plateau, Abia, and extending to Borno; and okra production in states like Abia, Anambra, Oyo, Osun, and Borno), provided the theoretical and regional context for the data collected. Additional sources specifically consulted for optimal cold storage temperatures and spoilage rates for the included crops are listed in the Data Sources section.
* **Field Data Collection:** The core dataset ("TEAM DSS Dataset.csv") contains field-level observations. Data points were collected from various farmers in specific farming locations across Nigeria, focusing on five key crops: Okra, Pepper, Cabbage, Tomatoes, and Yam. The selection of these farming locations was informed by the research identifying major cultivation areas for these crops.
* Logistical Data Collection: Distances and estimated transport times between locations (farm to cold storage, cold storage to market) were primarily sourced from the Kaggle dataset "Nigerian States Travel Data" and supplemented with data obtained from Google Maps.  
  The Transport Cost For 20 Ton of Load (₦/km) data was estimated based on cost information obtained directly from ColdHubs logistics for transporting a specific volume (20 tons) between two pairs of states (Lagos to Kano and Kaduna to Owerri). An average cost per kilometer (₦3,791/km) was calculated from these examples and applied.
* **Cold Storage Information:** Information regarding the cold storage locations utilized in the dataset was sourced from ColdHubs, referencing their presence and reach across various states in Nigeria (as indicated by the provided link: <https://www.coldhubs.com/coldhubnews/2023/3/23/2023-impact-coldhubs-moves-from-22-to-28-states-in-nigeria>). The cost of storage per day was obtained through direct questioning of a ColdHubs cold storage hub operator at Ibeju-Lekki, Lagos.
* **Data Processing:** Where research sources provided a range of values for optimal storage temperature or spoilage rate, the median value of that range was calculated and used in the dataset for the corresponding data points.
* **Data Points:** For each observation, data was recorded on:
  + Farmer identification (anonymized). The names in the Farmer Name column were generated randomly and do not represent actual individuals.
  + Farm location (State and City), chosen based on identified cultivation zones.
  + Crop type.
  + Cold Storage Location (Market and State), often corresponding to areas with cold chain infrastructure.
  + Logistical details (Distance and Time for transport from farm to storage and storage to market), sourced as described above.
  + Observed storage temperature.
  + Observed spoilage rate at that storage temperature.
  + Logistics provider and associated costs (storage and transport).

The data collection process aimed to capture real-world scenarios of post-harvest handling and storage within the context of identified cultivation areas and available cold chain infrastructure.

* **Ethical Considerations:** To protect privacy, the "Farmer Name" entries in the dataset were randomly generated and do not correspond to actual individuals who may have contributed data or insights to the research.
* **Timeframe:** The research and data collection activities informing this dataset were conducted prior to May 2025, as indicated by the access dates on the referenced documents and the current date.

## 3. Data Description and Dictionary

The dataset is provided in CSV format ("download (2).xlsx - Sheet1.csv"). It contains the following columns:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Operational Definition (Based on Research Context)** |
| Farmer Name | Identifier for the farmer. **Note: These names were randomly generated for the dataset and do not represent actual individuals.** | Anonymized or pseudonymous identifier for the individual farmer from whom the data was collected. Randomly generated for privacy. |
| Farm Location | The geographic location of the farm where the crop was cultivated. | Recorded as City and State in Nigeria, chosen based on the identified major cultivation zones for the specific crop (informed by "Nigerian Crop Cultivation by State\_.docx" and related sources). |
| Crop | The type of agricultural commodity. | One of the five focus crops: Okra, Pepper, Cabbage, Tomatoes, or Yam, as analyzed in the accompanying research. |
| Cold Storage Location | The location of the cold storage facility used. | Recorded as Market and State in Nigeria, often corresponding to locations where ColdHubs or similar facilities are known to operate. |
| Distance: Farm to Cold Storage (km) | The measured or estimated distance between the farm and the cold storage facility in kilometers. | A logistical parameter indicating the length of the initial transport leg, primarily sourced from the Kaggle dataset "Nigerian States Travel Data" and Google Maps. |
| Transport Time: Farm to Cold Storage (hrs) | The measured or estimated time taken for transport from the farm to the cold storage facility in hours. | A logistical parameter indicating the duration of the initial transport leg, influenced by distance, road conditions, and transport mode, primarily sourced as described above. |
| Market Location | The final market destination for the crop after storage. | Recorded as City, representing the intended point of sale after cold storage. |
| Distance: Cold Storage to Market (km) | The measured or estimated distance between the cold storage facility and the market in kilometers. | A logistical parameter indicating the length of the secondary transport leg, primarily sourced from the Kaggle dataset "Nigerian States Travel Data" and Google Maps. |
| Transport Time: Cold Storage to Market (hrs) | The measured or estimated time taken for transport from the cold storage facility to the market in hours. | A logistical parameter indicating the duration of the secondary transport leg, influenced by distance, road conditions, and transport mode, primarily sourced as described above. |
| Optimal Storage Temp (°C) | The recorded temperature at which the crop was stored in the cold storage facility. This value was derived from research sources providing a range of optimal temperatures, by calculating the median of that range. | This column represents the *observed* storage temperature for the specific instance in the dataset. It is important to note that this may not align with the scientifically *optimal* storage temperature for that crop as discussed in "Crop Storage Temperature and Spoilage\_ (1).docx" and other relevant research (see Data Sources). The value in this column is the median of the optimal range identified in the research. |
| Spoilage Rate at Optimal Temp (%) | The estimated percentage of the crop that spoiled after a defined storage period at the recorded temperature. This value was derived from research sources providing a range of spoilage rates at optimal temperatures, by calculating the median of that range. | The observed percentage of post-harvest loss (spoilage) for the specific batch of crop stored under the conditions recorded in the other columns, particularly the Optimal Storage Temp (°C). This rate is observed *at the recorded temperature*. The value in this column is the median of the spoilage rate range at optimal temperature identified in the research. |
| Logistics Provider | The entity responsible for the cold storage and potentially transport. | Identifies the service provider used for cold chain logistics. In this dataset, it is consistently "ColdHub Cold Storage". |
| Storage Cost (₦/crate/day) | The cost of storing the crop per crate per day in Nigerian Naira (₦). | The financial cost associated with utilizing the cold storage facility for the specified duration, obtained through direct questioning of a ColdHubs operator at Ibeju-Lekki. |
| Transport Cost For 20 Ton of Load (₦/km) | The cost of transporting the crop per kilometer in Nigerian Naira (₦). **Note: This was estimated based on sample cost data for 20-ton transport between specific states obtained from ColdHubs logistics, averaging ₦3,791/km.** | The financial cost associated with the transportation of the crop, estimated based on sample data from ColdHubs logistics and an average cost per kilometer calculation. |

## 4. Data Sources

The data presented in this dataset is derived from the following sources:

* **Nigerian Crop Cultivation by State\_.docx:** A report synthesizing existing documentation on major crop cultivation zones in Nigeria.
* **Crop Storage Temperature and Spoilage\_ (1).docx:** A report synthesizing research findings on optimal storage conditions and spoilage factors.
* **ColdHubs Website:** Information on cold storage locations (<https://www.coldhubs.com/coldhubnews/2023/3/23/2023-impact-coldhubs-moves-from-22-to-28-states-in-nigeria>).
* **Direct Questioning:** Information on the daily storage cost per crate was obtained through direct questioning of a ColdHubs cold storage hub operator at Ibeju-Lekki, Lagos.
* **ColdHubs Logistics Contact:** Sample cost data for transporting 20 tons between specific states (Lagos to Kano and Kaduna to Owerri) was obtained from ColdHubs logistics and used to estimate the transport cost per kilometer in the dataset (average of ₦3,791/km).
* **Kaggle Dataset:** "Nigerian States Travel Data" (<https://www.kaggle.com/datasets/vektur/nigerian-states-travel-data>) was a primary source for distance and estimated transport time data between locations.
* **Google Maps:** Used to supplement distance and transport time data.
* **Additional Sources on Crop Locations:** Various online resources providing information on specific crop cultivation areas in Nigeria, including:
  + <https://www.bbcnews.com.ng/2020/11/borno-gwamna-umara-zulum-ya-kada.html>
  + <https://naijadetails.com/top-10-agriculture-producing-states-in-nigeria/>
  + <<https://dailytrust.com/chilli-pepper-farming-is-profiting-farmers/#:~:text=The%20greater%20part%20of%20pepper%20production%20in%20Nigeria,Kano%2C%20Jigawa%2C%20Katsina%2C%20Sokoto%2C%20Plateau%20and%20Bauchi%20states>.>
  + <https://essfeed.com/the-worlds-top-okra-producers-who-leads-in-yield-and-innovation-the-worlds-top-okra-producers-who-leads-in-yield-and-innovation/>
  + <https://www.researchgate.net/publication/312033190_Determinants_of_Okra_Abelmoschus_esculentus_Production_and_Profitability_in_Ayamelum_Local_Government_Area_of_Anambra_State_Nigeria_Author's_Details>
  + <https://open.ai/b153521a59cc2fefff34f0baebcd7337?rand=42738>
  + <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.cabidigitallibrary.org/doi/pdf/10.5555/20153292803&ved=2ahUKEwjV0J6yjqONAxUgU6QEHTSeNkQQjJEMegQIAxAC&usg=AOvVaw3_B_wnGYphUgZqxTDlDgNO>
  + <https://en.wikipedia.org/wiki/Yam_production_in_Nigeria>
  + <http://article.sapub.org/10.5923.j.food.20241401.01.html>
* **Additional Sources on Optimal Storage Temperature and Spoilage Rate:**
  + <<https://hopetownlodge.com/lifestyle/unlocking-okras-freshness-a-guide-to-ideal-storage-temperatures/#:~:text=Refrigerator%3A%20The%20ideal%20temperature%20for%20storing%20fresh%20okra,ripening%20process%20and%20preventing%20damage%20from%20excessive%20cold>.>
  + <https://broudat-kar.com/lang/en/ideal-coldstorage-temperature/>
  + <<https://edepot.wur.nl/444870#:~:text=Cold%20storage%20at%20around%2012%20oC%20is%20recommended,reduction%20of%20tomato%20flavor%20due%20to%20chilling%20injury>.>
  + <<https://www.gcca.org/legacy-system/WFLO-Commodity-Storage-Manual-2018Cabbage%5B1%5D.pdf#:~:text=Late%20crop%20cabbage%20under%20proper%20refrigeration%20at%2032%C2%B0F%2C%20keep%20for%205%20to%206%20months%20in%20storage>.>
  + <https://thehearthsidehaven.com/how-cold-can-cabbage-tolerate/>
  + [https://www.ijrte.org/wp-content/uploads/papers/v8i4/D1651078419.pdf#:~:text=Abstract%3A%20Fresh-cut%20yam%20beans%20are%20highly%20perishable%20and,maintained%20at%205%C2%B0C%20and%20relative%20humidity%20of%2090%25.](https://www.ijrte.org/wp-content/uploads/papers/v8i4/D165118419.pdf#:~:text=Abstract%3A%20Fresh-cut%20yam%20beans%20are%20highly%20perishable%20and,maintained%20at%205%C2%B0C%20and%20relative%20humidity%20of%2090%25.)
  + <https://myschool.ng/classroom/agricultural-science/18108>
  + <<https://spicyexchange.com/peppers-storage-refrigeration-tips/#:~:text=Storing%20peppers%20at%20the%20right%20temperature%20and%20humidity%2C%20injuries%2C%20negatively%20impacting%20the%20peppers%E2%80%99%20texture%20and%20taste>.>
  + <https://mjajournal.ekb.eg/article_189878_9f1e8b910db43cfcc23c275c3a79fd7.pdf>
  + <https://www.researchgate.net/figure/Effect-of-different-cold-storage-temperatures-on-weight-loss-of-green-chilies_fig2_281324187>
  + [https://www.coldhubs.com/coldhubnews/extending-the-shelf-life-of-tomatoes](https://www.coldhubs.com/coldhubsnews/extending-the-shelf-life-of-tomatoes)
  + <https://www.researchgate.net/publication/367608743_ANALYSIS_OF_SPOILAGE_AND_STORAGE_OF_FRESH_TOMATO_IN_PLATEAU_STATE_NIGERIA>
  + <https://www.aimspress.com/aimsagri/article/view/id/13793/html>
  + <https://applicanceupdate.com/how-long-can-you-keep-fresh-cabbage-in-the-refrigerator/>
  + <https://xueyuncold.com/blog/Using-Cold-Storage-to-Preserve-Cabbage.html>
  + <https://applicanceupdate.com/how-long-does-cabbage-stay-fresh-in-the-refrigerator/>
  + <https://xueyuncold.com/blog/Using-Cold-Storage-to-Preserve-Cabbage.html>
  + <https://www.researchgate.net/publication/43554813_Effects_of_Storage_Conditions_and_Storage_Period_on_Nutritional_and_Other_Qualities_of_Stored_Yam_Dioscorea_spp_Tubers>
  + <https://www.researchgate.net/publication/285574389_Effect_of_Three_Storage_Methods_on_the_Quality_and_Shelf-Life_of_White_Yam_Dioscorea_rotundata_Cultivars_Pona_and_Tela>
  + <https://www.researchgate.net/publication/337829759_Cooling_load_calculation_for_efficient_cold_storage_of_yam_bean>
* **Field Data Collection:** Direct collection of data points from farmers regarding their specific post-harvest handling and storage experiences (forming the basis of the CSV file).

## 5. Limitations

Users of this dataset should be aware of the following limitations:

* **Scope of Foundational Research:** The accompanying research documents, while informative, may not cover all aspects of crop cultivation or post-harvest handling across the entirety of Nigeria. The cultivation report is based on a synthesis of *provided* documents and publicly available information, not an exhaustive national survey of all farmers or regions.
* **Dataset Specificity:** The dataset represents specific instances of post-harvest handling observed during the data collection period. It may not capture the full variability of conditions, practices, or outcomes across all regions, seasons, or market dynamics.
* **"Optimal Storage Temp" Column:** It is crucial to understand that the column labeled "Optimal Storage Temp (°C)" in the CSV represents the *recorded temperature* at the cold storage facility for that specific data entry. This temperature may or may not align with the scientifically *optimal* storage temperatures discussed in the "Crop Storage Temperature and Spoilage\_ (1).docx" document and other relevant research. The spoilage rate is recorded *at this observed temperature*.
* **Spoilage Rate Estimation:** The method for estimating the "Spoilage Rate at Optimal Temp (%)" is not explicitly detailed in the provided materials. Users should consider that this might be a farmer's estimate or based on a specific assessment protocol used during data collection, which could introduce variability. Transparency on this specific methodology, if available, would enhance credibility.
* **Generalizability:** While the dataset includes data from various locations informed by research into cultivation zones, the extent to which the findings can be generalized to other regions, crops not included, or different cold chain solutions should be considered carefully.
* **Data Granularity:** The dataset provides a snapshot of key parameters but does not include continuous monitoring data (e.g., temperature fluctuations over time during storage or transport) which could influence spoilage.
* **Cost Data Specificity:** The storage cost data point is based on direct questioning of a single ColdHubs operator in a specific location (Ibeju-Lekki). This cost may vary across different ColdHubs locations or other storage providers. The transport cost data, while sourced from a dataset and Google Maps, may also represent estimates and actual costs could vary based on real-time factors.

## 6. Usage Notes

When utilizing this dataset for analysis, please consider the following:

* Refer to "Crop Storage Temperature and Spoilage\_ (1).docx" and the additional provided links to understand the *scientifically recommended* optimal storage conditions and associated spoilage characteristics for each crop and compare them to the *recorded temperatures* and spoilage rates in the dataset. This comparison can reveal insights into the effectiveness of the observed storage practices.
* Analyze the relationship between the Optimal Storage Temp (°C) (the recorded temperature) and the Spoilage Rate at Optimal Temp (%) to understand actual post-harvest outcomes under various observed storage conditions.
* Utilize the farm and market locations, informed by the cultivation zone research, along with the distance and time data, to explore the logistical challenges and their potential correlation with spoilage rates or costs in specific agricultural regions.
* Be mindful of the limitations discussed in Section 5 when interpreting findings and drawing conclusions, particularly regarding the specificity of the cost data and spoilage rate estimation methodology.

## 7. Conclusion

This research-based post-harvest loss dataset offers valuable insights into the practical aspects of crop handling, storage, and logistics in Nigeria, grounded in existing research on cultivation zones and storage science, and supplemented by specific logistical and cost data. **Crucially, this dataset was compiled with the explicit purpose of training a machine learning model.** This model is intended to provide essential decision-making support to farmers by assisting them in locating the cold storage facilities closest to their farm locations, identifying nearby market opportunities, estimating transportation costs, and potentially offering a reminder system on the status of their crop health. By understanding the research methodology, data sources, and limitations outlined in this documentation, users can effectively leverage this dataset to contribute to the development of tools aimed at reducing post-harvest losses and improving the efficiency and resilience of agricultural value chains in Nigeria.