Report on the Nyagatare Training Session

# Background and Context

The training in Nyagatare took place over 5.5 days, from July 28 to August 2. The structure was as follows:

* **Monday to Friday (July 28 – August 1):** Sessions ran from 8:30 a.m. to 5:00 p.m. Each day consisted of approximately five hours of lectures and hands-on programming activities, followed by two hours of participant group discussions. Please refer to the detailed program for more information.
* **Saturday, August 2:** On the final day, participants from 23 groups each presented a summary of their proposed data use case. Presentations lasted about five minutes per participant, beginning at 9:00 a.m. and concluding around 12:00 noon.

The rest of this report provides insights into what transpired, focusing on the extent of content covered, participant attendance, performance on assessments, and feedback on the training.

# Curriculum Coverage

Please refer to the course website for the full curriculum. During the Nyagatare training, the plan was to cover three to four modules: Python Foundations, Data Analysis with Python, Spatial Data, and Time Series. However, because participants had not completed the mandatory course preparation—specifically the basics of Python—the pace of the training was slower than expected. As a result, we were only able to cover two modules: *Module 1 (Python Foundations) and Module 2 (Data Analysis with Python).*

# Attendance and Participation

## Overall attendance

Overall, participants were expected to engage in three main categories of activities in order to be considered as having fully participated in the course: attending lecture sessions led by instructors, actively contributing during group discussions, and completing the required assessments (including quizzes and submissions). Statistics on the completion of quizzes and assessments will be presented in the next section. The table below focuses specifically on class attendance across each day.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Day** | **Date** | **Expected Number of Participants** | **Number Present** | **Percent Present** |
| 1 | July 28 |  |  |  |
| 2 | July 29 |  |  |  |
| 3 | July 30 |  |  |  |
| 4 | July 31 |  |  |  |
| 4 | August 1 |  |  |  |

Table 1: Attendance by each day

## Participation in the Data Use Case

The data use case is an integral component of this training program. In this component, participants are expected to design and execute a data science project. During the Nyagatare training, participants were asked to develop an initial idea for their project (or data use case) and present it to the entire group. With the exception of a few selected institutions, each institution was required to propose its own project. As a result, institutions with a single participant formed one-person project teams, while those with multiple participants formed larger teams. For example, NISR, with 15 participants, was divided into three project teams. The summary below shows the number of participants who submitted and presented a project.

* **Participants involved in data use case concept development and presentation**
* **Participants involved in data use case concept development but didn’t present**

In Table 2 below, we present a list of all projects by institution. For more details on these projects, including team members and descriptions, please refer to the participant projects section on the [course website](https://dmatekenya.github.io/AIMS-DSCBI/projects.html).

|  |  |
| --- | --- |
| **Institution** | **Project title** |
| Ministry of Finance and Economic Planning | Macro-economic dashboard |
| Ministry of Trade and Industry | A Data-Driven Approach to Industrial Price Monitoring in Rwanda |
| Rwanda Basic Education Board | ICT Device Management Schools |
| Rwanda Revenue Authority | Data Discovery Chatbot |
| Ministry of Defence | Optimized Defense Resource Allocation |
| Ministry of Youth and ICT | Predictive Modeling of Youth Unemployment in Rwanda |
| Ministry of Justice | Predictive Crime Mapping Using Machine Learning to Identify High-Risk Crime Locations |
| Ministry of Environment,  Rwanda Environment Management Authority | Predicting urban Air Quality in Kigali City using machine learning |
| National Institute of Statistics of Rwanda | Land cover classification using high resolution satellite images |
| Gender Monitoring Office | Interactive Gender Based Violence Monitoring Dashboard |
| National Institute of Statistics of Rwanda (NISR) | Exploring the Geospatial Relationship Between Altitude and Child Stunting in Rwanda |
| Ministry of ICT and Innovation | Data Sharing & Privacy Bot |
| Ministry of Infrastructure (MININFRA) | Leveraging Data for Evidence-Based Infrastructure Needs Assessment & Planning |
| National Institute of Statistics of Rwanda (NISR) | Prioritizing Environmental and Climate Change Statistics in National Strategic Frameworks |
| National Institute of Statistics of Rwanda (NISR) | Job Transitions and Employment Stability in Rwanda: A Panel Data Analysis Using Labour Force Surveys |
| Rwanda Energy Group (REG) | Forecasting electricity demand and optimizing distribution in national grid. |
| Rwanda Biomedical Center | Natural Language Processing for Public Health Intelligence |
| Rwanda Agriculture Board | Predicting crop yield using ML models: case of maize in Rwanda |
| Rwanda Information Society Authority (RISA) | AI-Policy Chatbot |
| Ministry of Education | Real-Time Education Data Monitoring and Quality Feedback System |
| National Institute of Statistics of Rwanda (NISR) | Modernizing Rwanda’s Trade Statistics System: Enhancing Data Quality and Efficiency |
| WASAC Utility | Analyzing the Impact of Customer Complaints Patterns on Operational Efficiency at WASAC |
| Rwanda Convention Bureau (RCB) | Quantifying the Economic Impact of MICE Events Using Data Science |
| Directorate general of immigration and emigration (DGIE) | High level passenger travel analysis |
| Meteo Rwanda | TBA |

**Table 2: List of data use cases**

# Performance on Assessments

## Evaluation and Assessment Framework

To track participant performance and ensure completion of all course activities, the training included three types of graded assessments:

* **Quizzes:** Short assessments, typically lasting no more than 25 minutes, designed to test participants’ understanding of key concepts.
* **Programming Assignments:** Given that this is a hands-on, skills-based training program, programming assignments are essential for allowing participants to practice their skills while also enabling instructors to evaluate how well the concepts have been understood.
* **Data Use Case:** Serving as the capstone project of the training, this component contributes significantly to course completion and applies the skills learned in a practical context.

For the Nyagatare session, the graded assessments included:

* Two quizzes
* One programming assignment

## Performance Outcomes and Insights

All three assessments described above were graded out of 100. The final score was initially calculated as a simple average, without applying any weighting. Please note that a weighting scheme will be introduced later, with assignments carrying more weight than quizzes.

With regard to participation in these assessments, the majority (90% and above) of participants took part. Therefore, the figures below focus on the number of participants who missed the assessments.

* Quizz-1. 5 participants missed this assessment
* Quiz-2: 5 participants missed this assessment
* Assignment-1: 3 participants missed this assessment

### Aggregate Performance on All Assessments

In terms of performance, the majority of participants did well. For the overall score—calculated as an unweighted average across the three assessments—the median was 70%. This indicates that more than half of the participants scored above this level.

The bar chart below illustrates the distribution of participants across performance groups. As shown, the largest share falls within the **mid-performing range (60–79%)**, highlighting that most participants were able to grasp the key concepts and demonstrate them in assessments. A smaller proportion achieved **high-performing scores (≥80%)**, reflecting strong mastery of the material, while a notable minority remained in the **low-performing group (≤60%)**, indicating the need for additional support and targeted follow-up.

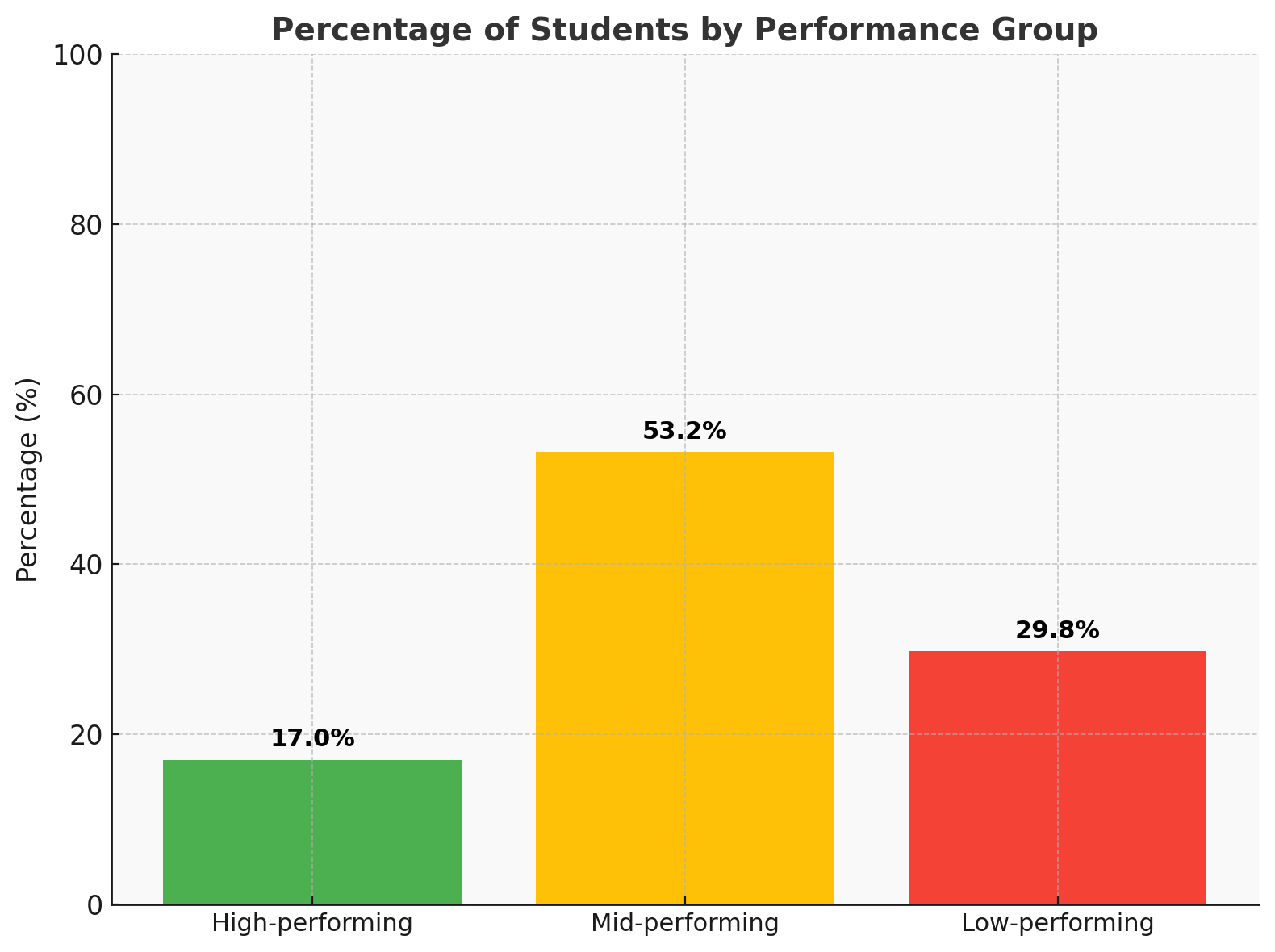


Figure-1: Aggregate Performance across all assessments

### Detailed Performance per Assessment

The distribution of performance across the three assessments highlights important differences in how participants engaged with quizzes versus the assignment as shown the bar plot below in Figure 2.

* **Quiz 1:** Out of 44 participants who attempted this quiz, scores ranged from **37.5% to 100%**, with a **median of 68.8%** and an **average of 66.9%**. The results show a balanced distribution across performance groups, though a notable portion remained in the low-performing range.
* **Quiz 2:** Similarly, 44 participants completed Quiz 2. Performance improved slightly compared to Quiz 1, with scores ranging from **45% to 95%**, a **median of 70%**, and an **average of 71.8%**. The bar plot confirms that more participants moved into the mid- and high-performing groups relative to Quiz 1.
* **Assignment 1:** A total of 47 participants submitted the first programming assignment. Scores ranged from **53% to 97%**, with a **median of 74%** and an **average of 74.5%**. This assessment recorded the strongest results overall, with the majority of participants falling into the mid- and high-performing categories, and very few in the low-performing group.

**Overall,** the results suggest that participants performed better on the assignment than on the quizzes, possibly reflecting the value of more time and hands-on engagement compared to timed quizzes. The bar plot visually reinforces this trend, showing a larger share of high- and mid-performing students in the assignment.

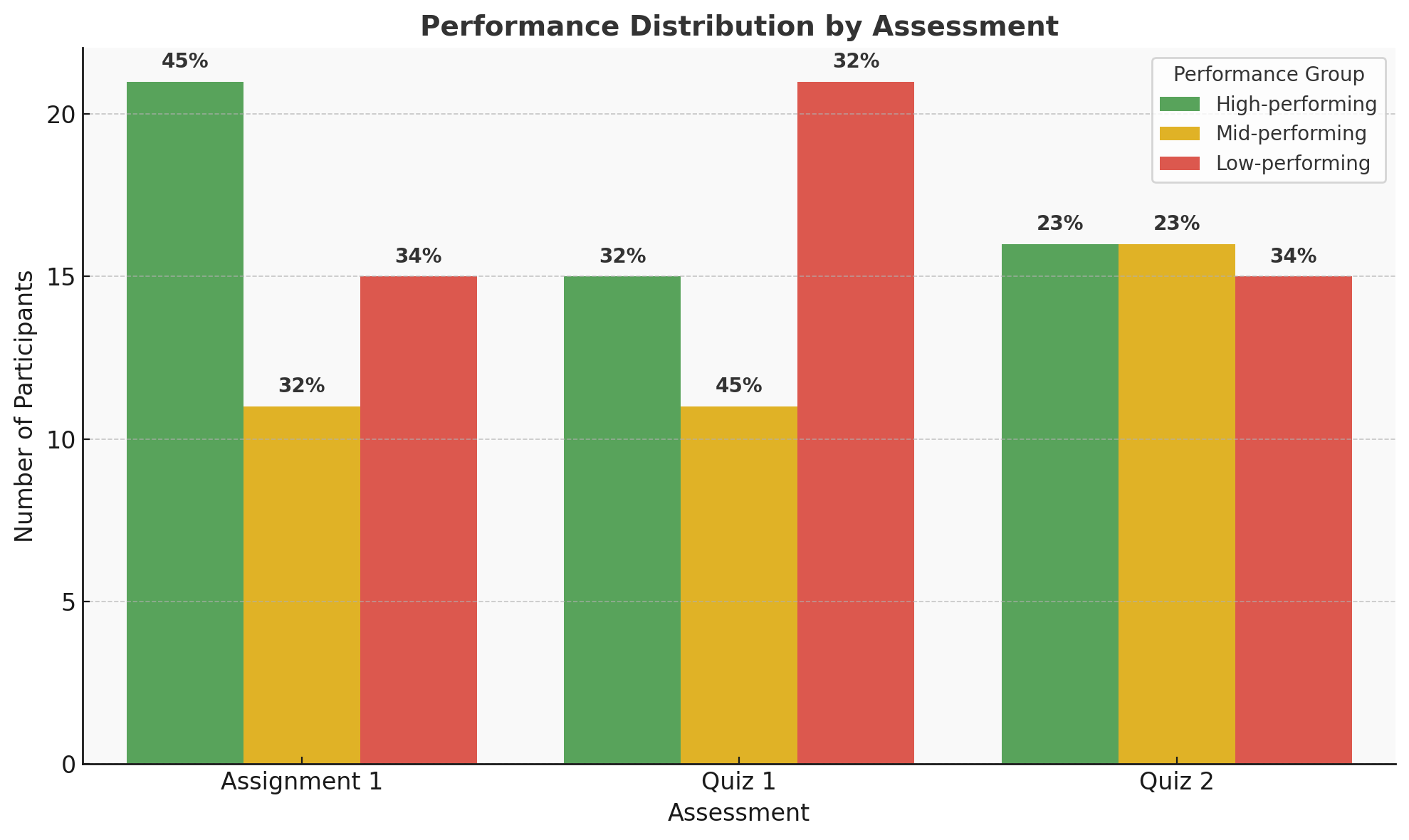


Figure-2: Performance by each assessment

Participant Feedback and Suggested Improvements

## Participants Feedback

### Key Takeaways from Participant Feedback

Throughout the course, instructors collected daily feedback from participants. At the end of the training, a comprehensive survey was also conducted to capture overall participant reflections. The key takeaways are summarized below.

* **Improved Confidence:** Most participants reported stronger confidence in Python programming and data analysis skills after the training.
* **Valuable Content:** Python for Data Science (NumPy, Pandas) was rated as the most valuable component, followed by Python basics, EDA, and statistical analysis.
* **Pace Adjustment Needed:** The majority found the course pace too fast, indicating the need for slower delivery or additional time for practice.
* **Challenges Identified:** Statistical concepts, hands-on exercises, and core programming concepts were the most difficult areas for participants.
* **Call for More Practice:** Feedback strongly emphasized increasing practice sessions and providing additional support during exercises.

### Participant Feedback in Detail

The detailed feedback from participants provides deeper insights into their learning experience. It highlights overall satisfaction with the training, changes in confidence levels, and the topics they found most valuable. It also captures areas where participants struggled, perceptions about the course pace, and their suggestions for improvement. Together, these insights point to what worked well and where adjustments may be needed in future iterations of the program.

Overall Satisfaction  
Overall satisfaction with the course content was positive, with:

* **21% rating it as excellent**,
* **63% rating it as good**,
* Only a small minority (16%) giving it average or below average ratings .

Confidence Gains  
Participants reported clear improvements in their confidence levels:

* In **Python programming**, most participants moved from mid-range confidence (level 3) to higher levels (4–5), with **85% rating themselves at 3 or above after the course**.
* In **data analysis with Python**, a similar trend was observed, with over **88% rating their confidence at level 3 or above after the course** .

Valuable Topics  
When asked about the most valuable topics, participants highlighted:

* **Python for Data Science (NumPy, Pandas, etc.)** – 44%
* **Python Basics** – 15%
* **Exploratory Data Analysis (EDA)** – 15%
* **Statistical Analysis in Python** – 15%
* **Data Processing with Pandas** – 12%

This indicates strong appreciation for both foundational Python skills and applied data science techniques.

Pace of the Course  
A majority (78%) felt the course was **too fast**, while 22% thought the pace was “just right.” None rated it as too slow. This suggests that future iterations should consider a slightly slower pace or more time for practice and reflection .

Challenging Areas  
Participants identified the most challenging areas as:

* **Statistical concepts (32%)**
* **Hands-on exercises (32%)**
* **Python programming concepts (16%)**
* Smaller shares noted challenges with visualization syntax (8%), pandas data manipulation (5%), and interpreting results (5%) .

Suggestions for Improvement  
From the open-ended feedback, participants emphasized the need for:

* **More practice sessions and hands-on exercises**
* **Additional time for intensive learning**
* **Stronger support during exercises**
* Opportunities to **refresh and revisit concepts** as the course progresses

## Feedback from Instructors

In addition to participant evaluations, the instructors also provided feedback and reflections on various aspects of the training. Their observations cover not only the performance and engagement of participants, but also interactions with key stakeholders such as CENFRI and NISR, as well as logistical considerations including internet availability at the venue and other operational issues. This feedback provides a valuable complement to participant perspectives, offering a more complete view of the training experience.

### Participant Engagement and Preparation

Overall, attendance and participation levels were strong throughout the training. However, a key bottleneck—and an important area for improvement—is ensuring that participants complete the pre-session preparatory course material when it is provided. During the sessions themselves, participants are encouraged to carefully work through the programming notebooks step by step, paying close attention to the instructions in both the markdown text and code comments.

## Training Venue Facilities and Connectivity

The facilities at the Nyagatare conference center were generally adequate. However, the first day was affected by a 1–2 hour delay due to the lack of a functional projector. In addition, participants faced internet connectivity challenges, especially when downloading large datasets and installing Python packages.

## Stakeholder Coordination

[ADD COORDINATION CHALLENEGS IF YOU FEEL THEY ARE NEEDED]

# Activities Leading Up to the Next In-Person Session

This section outlines the key activities that trainees are expected to undertake before the next in-person session. These activities fall into five main categories:

1. **Assigned Programming Exercises** – These may be graded or ungraded and are designed to help trainees practice and reinforce essential skills. So far, two assignments have been given.
2. **Project Meetings** – Each project team will hold weekly meetings (40–60 minutes) with the trainers. Some weeks will be **update meetings** (mandatory for all teams), while other weeks will be structured as **office hours**, where team members can join as needed to ask questions and seek guidance.
3. **Project Work** – Teams will continue to develop and advance their projects, supported by advisors and targeted skill-building to match the requirements of their chosen projects.
4. **Additional Sessions** – Trainers may organize supplementary sessions as needed (e.g., on spatial data or chatbot development). For example, a dedicated session on Building Chatbots with LLMs was held on August 23, 2025, for teams working on chatbot development or LLM-related projects.
5. **Course Website and GitHub Repository** – Instructors have created a [course website](https://dmatekenya.github.io/AIMS-DSCBI/), hosted on GitHub, to provide materials and resources that support ongoing learning. The website will act as

The purpose of these activities is to ensure that participants remain actively engaged, continue practicing core Python skills, and sustain momentum on their projects. We will also use this period to **catch up on modules that were not covered in the previous session** and to provide a **solid foundation for the next in-person training**. In addition, trainees will receive the specific skills and advisory support needed to advance the projects they are undertaking.

# **Conclusion and Next Steps**

This training session in Nyagatare successfully built participants’ skills in Python programming and data analysis, while also providing valuable lessons for instructors and stakeholders. The evaluations show clear gains in confidence and appreciation for the course content, alongside constructive feedback on areas that require adjustment—particularly the pace of delivery, the need for more practice, and better pre-session preparation.

Looking ahead, the next session presents an opportunity to build on this foundation. Incorporating the feedback received will be key to ensuring an even more impactful learning experience. Specific adjustments—such as allocating additional time for practice, providing stronger support during exercises, slowing down the pace, and reinforcing preparatory work—will help participants achieve greater mastery of the material. Continued collaboration with stakeholders such as CENFRI and NISR will also strengthen the relevance and sustainability of the program.

In addition, the upcoming session will adjust the sequencing of modules to better align with the projects participants are pursuing. While all modules were already planned, certain topics—such as Large Language Models (LLMs) and chatbots—will be introduced earlier than initially scheduled to support the teams working on these areas. Similarly, modules on machine learning will be prioritized for groups whose projects require these methods. This flexible approach ensures that the training remains directly relevant and responsive to participants’ ongoing project needs.