Data Analytics
Project Report

Report on the Data Analytics Training Course

Programme:
BRICS Astronomy Societal Impact Program

Prepared by:
BRICS Astronomy

Date:
August 2025

1. Introduction

The rapid advancement of large-scale astronomical projects such as the SKA and the Rubin Observatory LSST is ushering in an era of mega-data production in astronomy. These projects are generating unprecedented volumes of data on astrophysical transients, survey science, and time-domain astronomy, demanding new scientific methods and advanced data/compute infrastructures. Current research practices, collaboration models, and data analysis tools are insufficient to handle this scale and complexity. While BRICS countries are already contributing significantly to these global initiatives, there remains a critical need for a coordinated effort that integrates forefront observational facilities with cutting-edge big data solutions. Without such an initiative, the ability of BRICS researchers to fully exploit these transformative opportunities and lead in the development of innovative methodologies for the study of the dynamic Universe will be severely constrained.

As part of the Societal Impact Program, BRICS Astronomy, in collaboration with the Inter-University Institute for Data Intensive Astronomy (IDIA), has already demonstrated the value of capacity building through the successful launch and completion of a Data Analytics Training Course. This initiative equipped participants with both fundamental and advanced data analytics skills, particularly in astronomy-related applications. The strong response and wide pool of applicants highlight not only the demand for such training across BRICS countries and beyond, but also the necessity of scaling up efforts to prepare the next generation of researchers for the data-intensive future of astronomy.

2. Applications and Participation

- **Applications received:** Over 900 (with more than 800 from BRICS countries).
- **Accepted:** All applicants, facilitated by the scalability of online platforms such as Google Colab.
- **Active weekly participation:** Between 400 and 500 participants.
- **Session attendance:** ~200 participants in the morning sessions and ~150 in evening sessions.
- **YouTube views (for those unable to attend live):** Approximately 300 per session.
- **Capstone projects submitted:** 170.
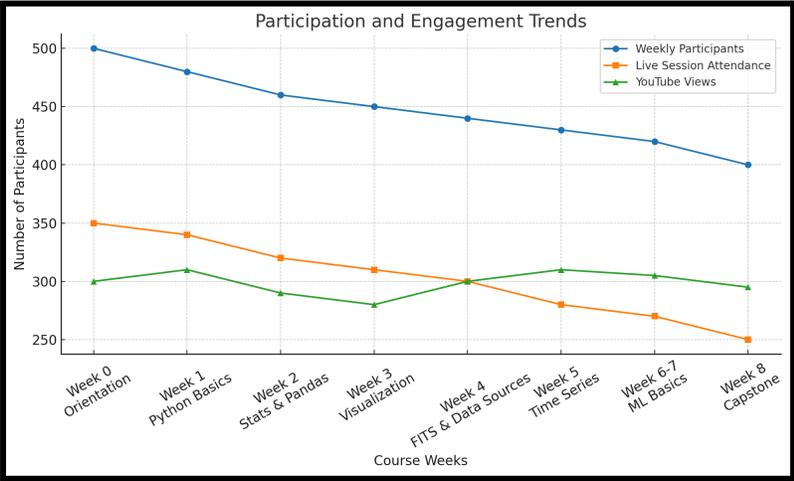- **Countries that Participated:** 39 (See Annexure A)

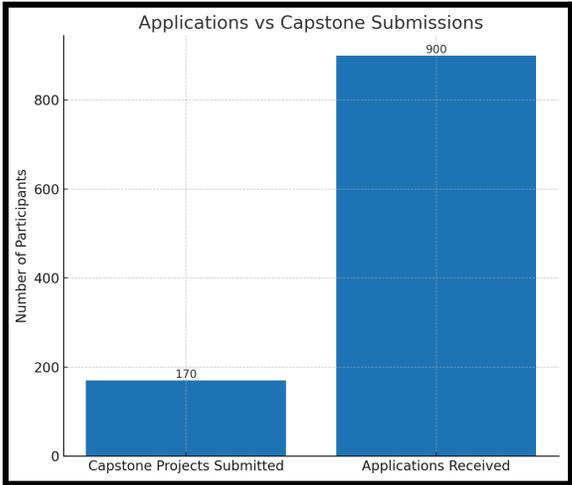*Figure 1: Participation and Engagement Trends*

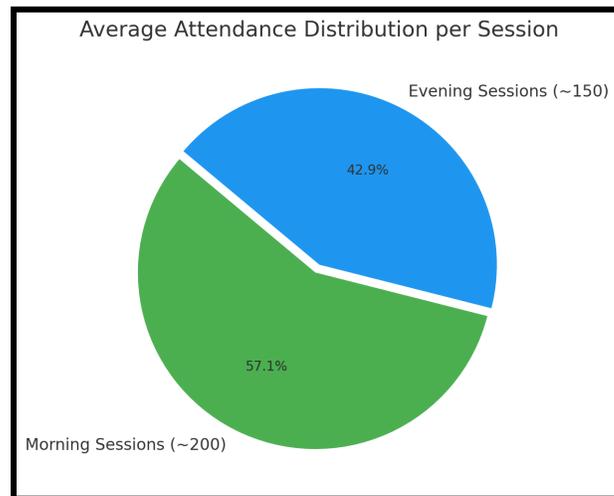*Figure 2: Applications vs Capstone Submissions*

Figure 3: Average Attendance Distribution per Session

3. Methodology

The course was structured over eight weeks plus an orientation week, combining synchronous and asynchronous learning.

Key features included:

- **Instructors:** 9 contributors who developed and delivered learning material.
- **Learning repository:** A GitHub repository housing compulsory and supplementary resources.
- **Sessions:** Two per week, accommodating different time zones.
- **Delivery:** Sessions were livestreamed and recorded, with replays available on YouTube.

4. Project Timeline

- **Week 0:** Orientation.
- **Week 1:** Introduction to Python (installation, variables, control flow, functions, NumPy).
- **Week 2:** Data analysis with Pandas, statistical methods, and hypothesis testing.
- **Week 3:** Data visualization with Matplotlib & Seaborn; colour theory and astronomical imaging.
- **Week 4:** Astronomical data sources, handling FITS files (Astropy, Astroquery).
- **Week 5:** Time series analysis for astronomical data (Lightkurve).
- **Weeks 6–7:** Machine learning basics for astronomy (clustering, classification, regression using Scikit-learn).
- **Week 8:** Capstone project submission and presentations.

5. Resources

Participants had access to:

- **Software:** Python (>3.6), Jupyter Notebooks, Anaconda.
- **Libraries:** NumPy, Pandas, Matplotlib, Seaborn, SciPy, Astropy, Astroquery, Scikit-learn.
- **Datasets:** Gaia, Sloan Digital Sky Survey (SDSS), TESS.
- **Instructional support:** Dedicated instructors providing lectures and resources.
- **Software developer**: Designed and developed the platform to submit the capstone projects.
- **Github**: Dedicated GitHub repo with the necessary resources.

6. Outcomes and Impact

- Trained hundreds of participants in essential data science and astronomical data analysis skills.
- Received 170 capstone projects, showcasing applied learning and problem-solving with real datasets.
- Fostered collaboration and knowledge sharing across BRICS countries.
- Built a foundation for scaling similar societal benefit programs in the future.

7. Team Members

Name	Institution	Country
Prof. Najam Hasan	Maulana Azad National Urdu University	India
Dr. Priya Hasan	Maulana Azad National Urdu University	India
Mr. Daniel Egbo	South African Astronomical Observatory	South Africa
Prof. Sally Macfarlane	Inter-University Institute for Data Intensive Astronomy	South Africa
Mrs. Narusha Isaacs-Klein	Inter-University Institute for Data Intensive Astronomy	South Africa

Dr. Aleksey Dianchenko	Ural Federal University	Russia
Dr. Nikita Rawat	South African Astronomical Observatory	South Africa
Dr. Jyoti Prakash	Inter-University Centre for Astronomy and Astrophysics	India
Mr. Moleboge Lekoloane	South African Astronomical Observatory	South Africa
Ms. Duduzile Kubheka	South African Astronomical Observatory	South Africa

8. Lessons Learned

The Data Analytics Training Course provided valuable insights into both the opportunities and challenges of delivering large-scale training across BRICS countries. Several key lessons emerged:

1. **Accessibility Challenges**
   Participants from countries such as China, Brazil, and Russia faced significant difficulties accessing course materials and platforms due to restrictions and technical barriers. This limited their ability to fully participate in the training. In addition, language barriers further hindered engagement for participants from China, Russia, and Brazil, pointing to the need for localized resources.

2. **Time Zone Differences**
   Coordinating across multiple time zones proved to be a persistent challenge. While offering two sessions per day helped mitigate the issue, participants' varying availability still made it difficult to ensure equal access to live sessions.

3. **Language Barriers**
   Language emerged as one of the most pressing challenges. To improve inclusivity and learning outcomes, it is recommended that future iterations of the course involve lecturers who can prepare and deliver materials in their native languages, or that multilingual support is provided.

4. **Platform Accessibility**
   The reliance on Google Colab highlighted disparities in platform accessibility. Some

countries could not access the same tools, creating unequal learning experiences.

5. **Need for Increased Support**
One-on-one support was identified as a critical factor for participant success. The course would benefit from additional support staff who can provide individualized guidance, helping participants navigate technical difficulties and deepen their understanding of the material.

9. Spin-offs

A significant spin-off of the program was the development of a dedicated management system, initially designed to support the coordination and monitoring of capstone projects. Following a detailed evaluation of the pilot phase, it was determined that a more comprehensive solution was required. This led to the conceptualization and development of a fully integrated Learning Management System (LMS) tailored to the course's requirements.

The LMS is designed to address several limitations identified during the pilot, including the lack of centralized access to instructions, assessments, and support services. Key functionalities include:

- **Instruction Delivery**: Hosting all course instructions and structured learning materials in a centralized environment.

- **Assessment Management**: Automated handling of weekly assessments to track progress and ensure participants are adequately prepared for the capstone project.

- **Support Scheduling**: Built-in functionality for participants to schedule one-on-one sessions with support staff to resolve technical and conceptual challenges.

In addition, the LMS establishes a uniform platform accessible across all BRICS countries, enabling standardized access to course materials, integrated translation features for multilingual delivery, and equitable access to assessments. This uniformity ensures consistent learning experiences and mitigates the accessibility issues observed in the pilot.

The system also demonstrates potential scalability. Beyond the immediate application to the Data Analytics Training Course, the LMS could be extended as a shared digital infrastructure for other BRICS working groups, providing a standardized environment for delivering free, high-quality educational content and training programs across member countries.

9. Conclusion

The Data Analytics Training Course was a resounding success, with high participation, strong engagement, and impactful outcomes. By leveraging online platforms and open-source tools,

the program was able to provide inclusive access to a global audience, empowering participants with skills relevant not only to astronomy but to data-driven industries more broadly. This initiative demonstrates the value of collaborative efforts in advancing science education and capacity building within the BRICS framework.

Annex A

Countries that Participated
1. India
2. South Africa
3. Ethiopia
4. Nigeria
5. Ghana
6. Russia
7. Kenya
8. Brazil
9. Egypt
10. Zambia
11. Madagascar
12. China
13. Burkina Faso
14. Philippines
15. Nepal
16. Bangladesh
17. Botswana
18. Pakistan
19. Algeria
20. Germany
21. Zimbabwe
22. United States
23. Tanzania

24. Swaziland
25. Mozambique
26. Uganda
27. United Arab Emirates
28. Italy
29. Angola
30. Cameroon
31. United Kingdom
32. Democratic Republic of the Congo
33. Saudi Arabia
34. France
35. Sri Lanka
36. Taiwan
37. Morocco
38. Chile
39. Belgium