CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
1. Kuncham Varun Teja – CMR INSTITUTIONS OF TECHNOLOGY – COMPUTER SCIENCE AND ENGINEERING – DATA SCIENCE

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

- Clean drinking water is essential for health, dignity, and daily life, yet many communities in India—especially in rural and underdeveloped regions—still lack reliable access. While urban areas often benefit from piped water and infrastructure, rural households may rely on handpumps, wells, or unsafe sources. This inequality affects health outcomes, increases time spent on water collection (especially by women), and limits overall well-being. Despite national progress and efforts under Sustainable Development Goal 6, significant disparities remain across regions and social groups. Some states have made advances, while others lag behind due to poor infrastructure or limited data-driven planning. This project analyzes data from the 78th Round of the Multiple Indicator Survey to identify where these gaps exist. By exploring links between water access, sanitation, clean fuel usage, and migration, the goal is to highlight vulnerable regions and help inform equitable government policies that ensure safe drinking water reaches every household.

PROPOSED SOLUTION

- To analyze disparities in access to improved drinking water across India using the 78th Round of the Multiple Indicator Survey (MIS).To identify underperforming regions and socio-economic patterns affecting water accessibility. The solution will consist of the following components
- Data Collection:
  - Use the publicly available dataset from AI Kosh (MIS – 78th Round).
  - Include Indicators: Improved Water Access, Clean Cooking Fuel, Sanitation Access, Migration Rate, Area Type and State.
- Data Preprocessing:
  - Clean missing and inconsistent data.
  - Standardize state names and numeric columns.
  - Feature engineering for new indicators like Water Access Gap = 100 - %Access.
- Exploratory Data Analysis (EDA):
  - Compare rural vs urban access levels.
  - Identify state-wise and regional disparities.
  - Correlation analysis between water access and other indicators (fuel, sanitation, migration).
- Deployment:
  - Deploy an interactive dashboard using IBM Watson Studio.
  - Host the solution on IBM Cloud Lite, ensuring scalability, easy access for stakeholders.
- Data Visualization:
  - Bar Charts, Pie Charts, Heatmaps, Scatter Plots.
  - Visually present disparities in drinking water access across states, rural-urban divisions, and their correlation with sanitation, fuel usage trends.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the data analysis solution to assess disparities in access to improved drinking water across India.

- System requirements
  - Technology Stack:
    - IBM Cloud Lite
    - IBM Watson Studio
- Library required to build the model
  - Python Libraries
    - Pandas
    - Sea born, ,Matplotlib for Visualization
- Work Flow Strategy
  - Data Cleaning and Preprocessing
  - Feature Engineering
  - Correlation and Pattern Detection

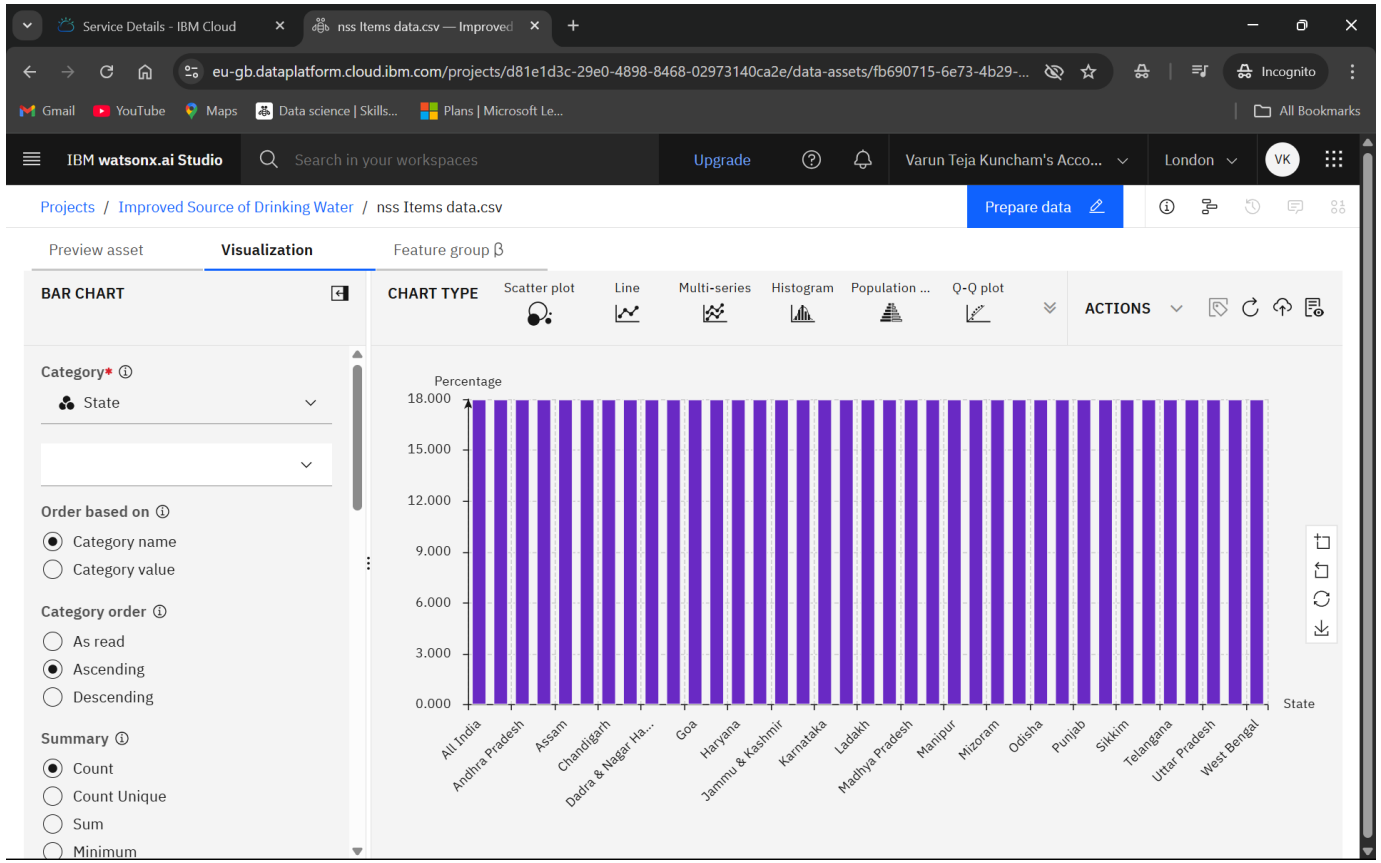
ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - No predictive model is used here; instead, the project focuses on descriptive analytics and correlation analysis to identify disparities and patterns in access to drinking water.
  - **The approach includes:**
    - Group-wise aggregation
    - Statistical comparison
- **Data Input:**
  - **State** - Name of the state or union territory
  - **Age Group** – (e.g., "15 years and above")
  - **Sector** - Area type: **Rural**, **Urban**, or **All**
  - **Gender** - Male, Female, or Person
  - **Indicator** - The specific metric being measured
  - **Value** - The numeric percentage or value corresponding to the indicator.
- **Training Process:**
  - Load and clean the dataset using **Pandas**. Perform descriptive statistics to get summary insights.
  - Conduct **correlation analysis** to identify strong relationships. Visualize insights using **Matplotlib** and **Seaborn**
- **Deployment :**
  - Develop an interactive analysis notebook using IBM Watson Studio that presents water access trends, comparisons, and correlation insights through graphs and summaries.

RESULT

1. Bar Chart:

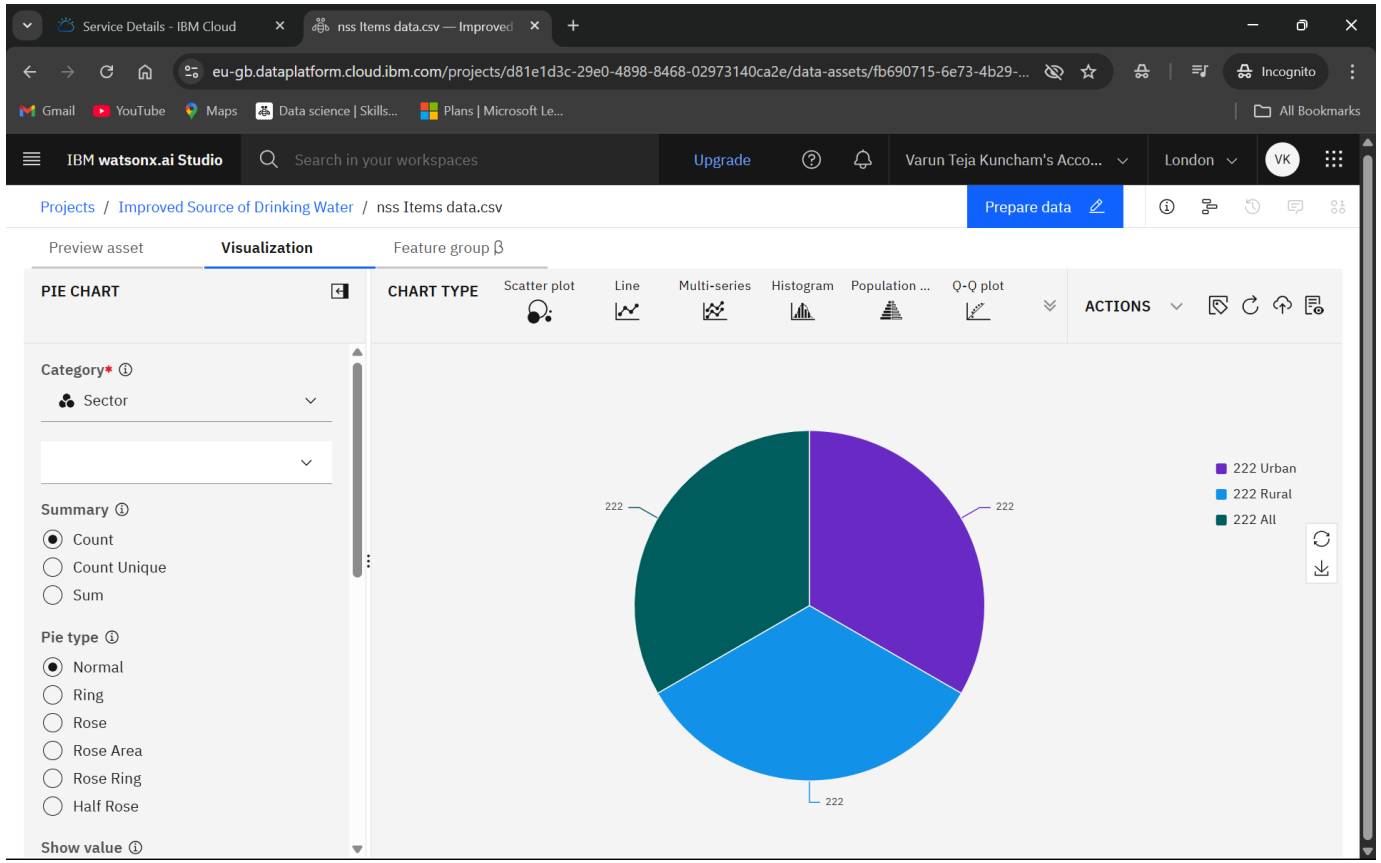
State-wise Count of Improved Drinking Water Access Records



RESULT

2. Pie Chart:

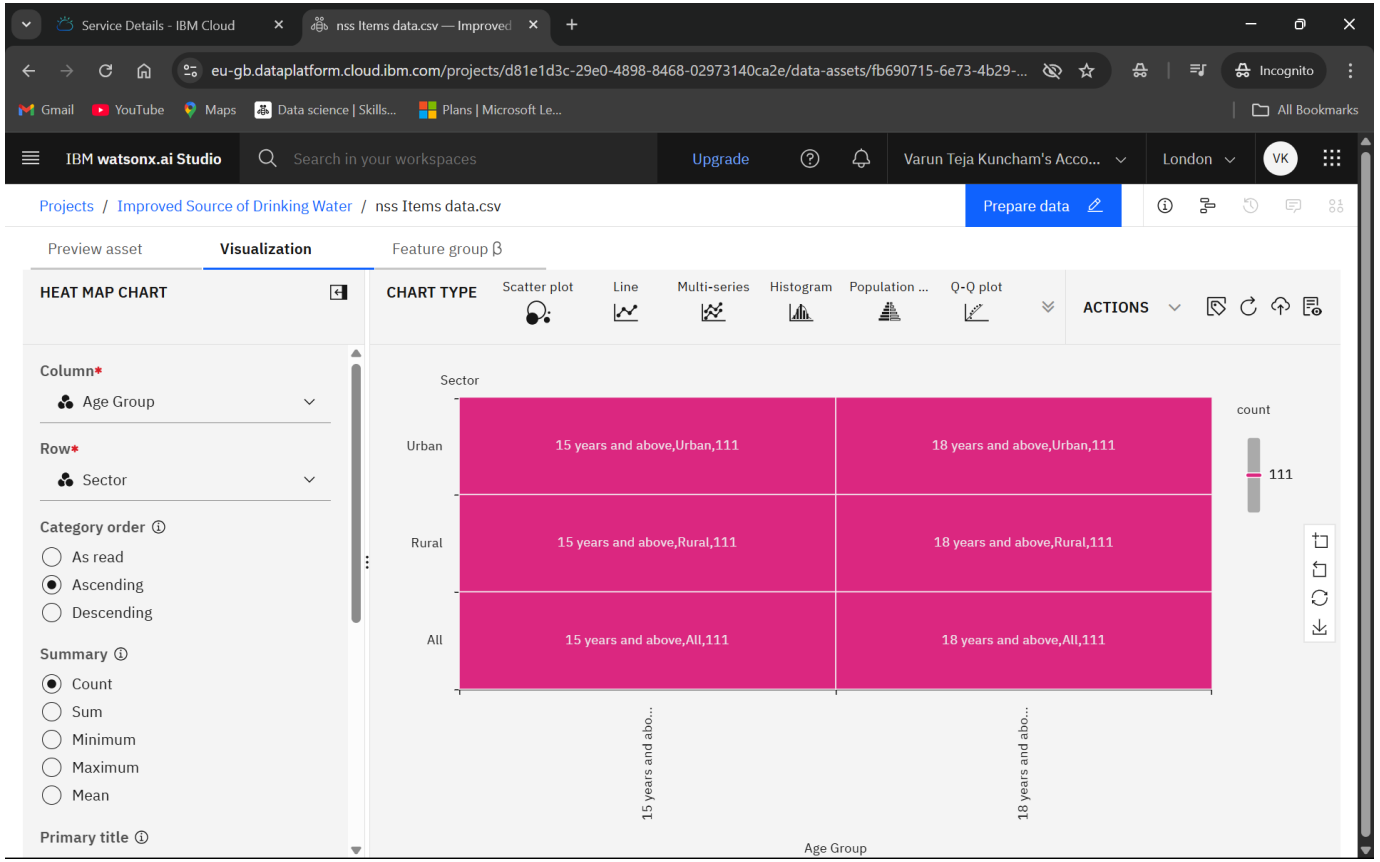
Sector-wise Distribution of Survey Records (Urban, Rural, All)



RESULT

3. Heat Map Chart:

Sector vs Age Group – Survey Response Intensity

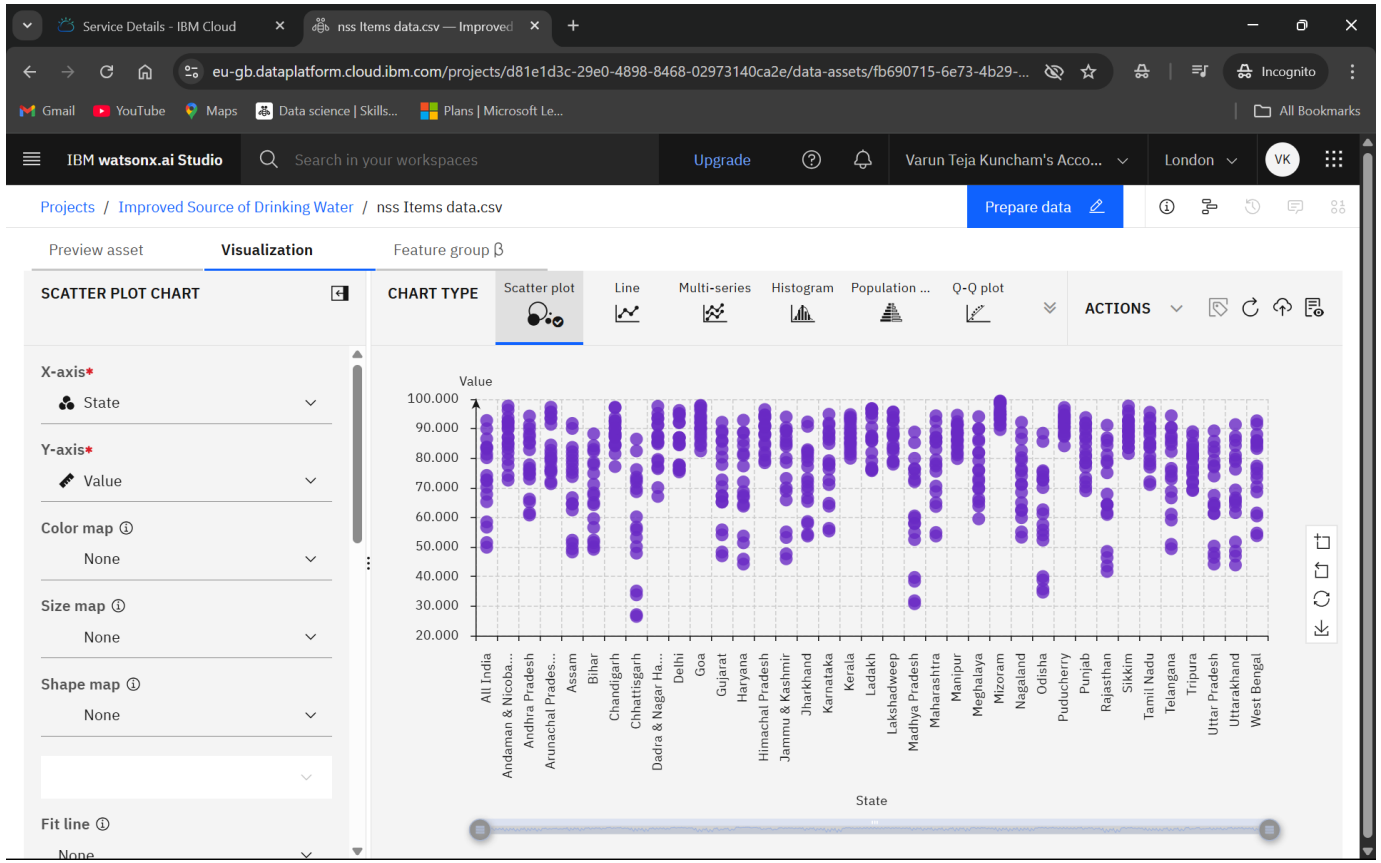


edunet foundation

RESULT

4. Scatter Plot Chart:

Analysis of Value Distribution Across States and Overall Frequency

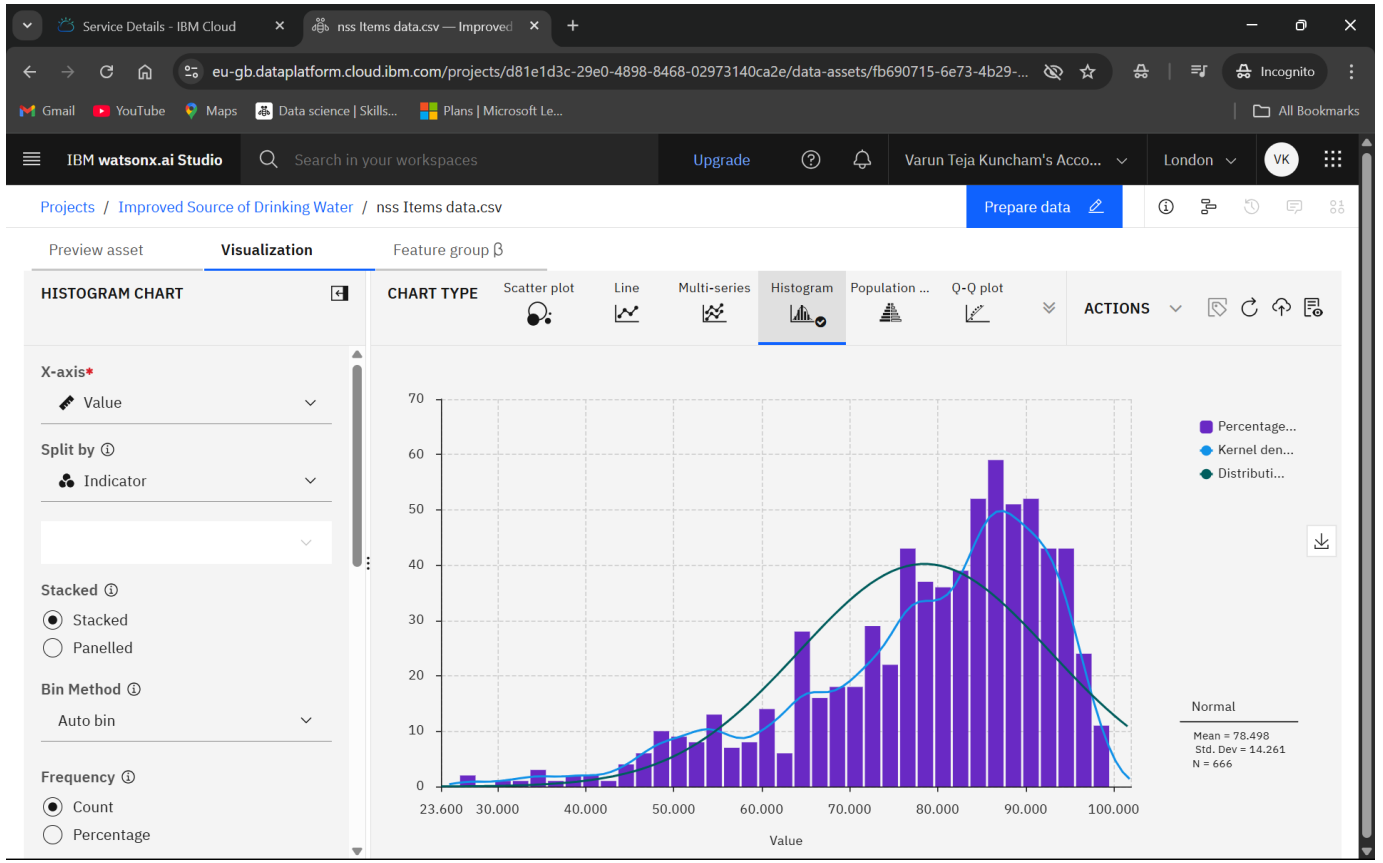


edunet foundation

RESULT

5. Histogram Chart:

Distribution of Value with Mean and Standard Deviation



CONCLUSION

- The analysis of the 78th Round of the Multiple Indicator Survey highlights regional and sectoral disparities in access to improved drinking water across India.
- Data visualization using bar charts, pie charts, and heatmaps revealed that while survey representation is well-balanced across states, sectors, and age groups, the real-world access to safe drinking water still varies significantly—especially between rural and urban populations.
- The correlation between water access, sanitation, and clean fuel usage suggests that multi-dimensional development policies are needed, focusing on infrastructure, awareness, and accessibility.
- This project provides data-driven insights to help policymakers prioritize investments, design targeted interventions, and ultimately contribute to achieving Sustainable Development Goal 6: Clean Water and Sanitation for All.

FUTURE SCOPE

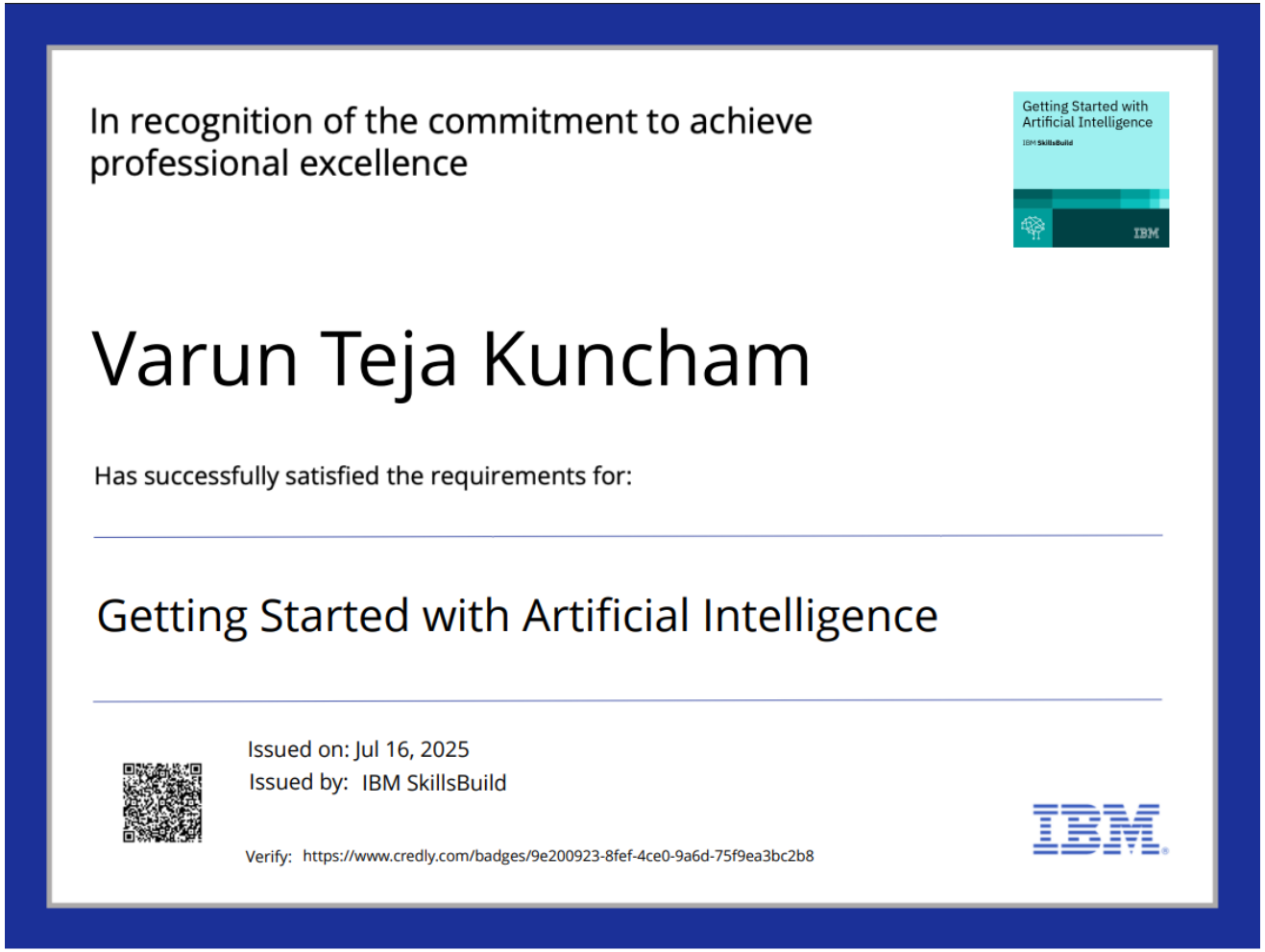
- **Granular-Level Analysis:**
  - Extend the study to the **district or village level** for more precise insights into water accessibility gaps.
- **Time-Series Comparison:**
  - Analyze data from multiple survey rounds to observe **trends and progress over time** in water access.
- **Predictive Modeling:**
  - Incorporate **machine learning models** to predict regions at risk of poor water access based on socio-economic and geographic indicators.
- **Public Dashboard Deployment:**
  - Build a **web-based interactive dashboard** for stakeholders, enabling dynamic exploration of visual insights by region, sector, or indicator.
- **Multidimensional Correlation Studies:**
  - Expand analysis to include variables like **education, income, climate data, and health outcomes** for a holistic development model.
- **Real-time Data Integration:**
  - Integrate with **IoT-based water quality sensors** or public infrastructure databases to monitor real-time availability and safety of drinking water.
- **Policy Simulation Models:**
  - Develop simulation models to **predict policy outcomes**, testing the impact of different interventions before implementation.

REFERENCES

- **National Sample Survey (78th Round) – Multiple Indicator Survey**
  Government of India, Ministry of Statistics and Programme Implementation (MoSPI)
  - https://mospi.gov.in
- **AI Kosh – Open Government Data Platform**
  Ministry of Electronics and Information Technology (MeitY), Government of India.
  - https://data.gov.in
- **Sustainable Development Goal 6 – Clean Water and Sanitation**
  United Nations Department of Economic and Social Affairs
  - https://sdgs.un.org/goals/goal6
- **IBM Watson Studio Documentation**
  IBM Cloud Services for Data Science and AI
  - https://www.ibm.com/cloud/Watson-studio
- **Python Libraries Used:**
  - **Pandas:** Data manipulation – https://pandas.pydata.org
  - **Matplotlib & Seaborn:** Visualization – https://matplotlib.org , https://seaborn.pydata.org

IBM CERTIFICATIONS

- Screenshot/ credly certificate(getting started with AI)



IBM CERTIFICATIONS

- Screenshot/ credly certificate(Journey to Cloud)

IBM CERTIFICATIONS

- Screenshot/ credly certificate(RAG Lab)

IBM **SkillsBuild**

Completion Certificate

This certificate is presented to

Varun Teja Kuncham

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU

edunet foundation