Concept Note

Goal :

To analyze DRINKING WATER and SANITATION using correlation.

To identify

- [ ] Key challenges and
- [ ] Propose ACTIONABLE SOLUTIONS aligned with
- [ ] SDG goal 6 :"CLEAN WATER AND SANITATION"

Identify TRENDS, KEY ISSUES and propose DATA-DRIVEN SOLUTIONS

Objectives of the project

- [ ] To collect and analyze data on "water and sanitation access"
- [ ] To identify the "primary factors" contributing to 'Lack of access to clean water' and 'sanitation'
- [ ] To understand TEMPORAL(changes or patterns that occur over time) and SPATIAL(changes or patterns that occur across different locations or 'regions') eg- Regional patterns such as differences between URBAN and RURAL areas / between different regions or COUNTRIES
- [ ] To develop PREDICTIVE MODELS for future access based on "current data"
- [ ] To propose actionable solutions and policy recommendations to enhance 'access to clean water and sanitation'
  - This is happening due to this, so we should do this(CAUSE-SOLUTION)
- [ ] To asses the potential impact of these "solutions" on achieving SGD goal 6
  - How its gonna impact, increase rate of success (by our solutions)
  - How its aligned/ connected with SDG 6

Data sources used

1. Unicef access to drinking water

- Query data : Build own dataset according to indicators, or
- Download one of the 6 given datasets
- Read some reports

https://data.unicef.org/topic/water-and-sanitation/drinking-water/

2. JMP Data

Data | JMP

The WHO/UNICEF Joint Monitoring Programme (JMP) is the custodian of global data on Water Supply, Sanitation and Hygiene (WASH).

https://washdata.org/data

3. WHO GLAAS water

GLAAS data portal

GLAAS provides policy- and decision-makers at all levels with reliable, easily accessible, comprehensive data on water, sanitation and hygiene (WASH) systems, including on governance, monitoring, human resources and finance. GLAAS monitors elements of

https://glaas.who.int/

4. Data.gov (USA website)
5. Data.gov.in (India website)
6. AODN open access to ocean data

https://portal.aodn.org.au/

7. Kaggle.com

- WASH dataset (Access to drinking water and sanitation)

Features of dataset

1. **Location :** Countries/ regions (For SPATIAL)
2. **Access levels :** Proportion of population with access to “Safely managed drinking water” and “ sanitation”
3. **Time :** Temporal data including year of recordings (Year)
4. **Demographics :** (Population) density, (Urban) vs (Rural) population, and socioeconomic status (poor, rich, richest)*

(written inside these) : Columns of dataset

Resources

1. Python for Data analysis by Wes Mckinney

https://wesmckinney.com/book/

2. Pandas official documentation

https://pandas.pydata.org/docs/

3. Numpy (for mathematical analysis) official documentation

> https://numpy.org/doc/

4. Seaborn (Matplotlib based visualization library) official documentation

> https://seaborn.pydata.org/

5. Scikit-learn (for Machine Learning) official documentation

> https://scikit-learn.org/stable/index.html

6. Matplotlib (easy to learn visualization tool) official documentation

> https://matplotlib.org/stable/index.html

7. DataCamp blog on data analysis
8. Youtube “IBM” self playlist https://youtube.com/playlist?list=PL7JM7wyIxWcRhd5It-2RZPKK_B3nG0qQf&si=6nH_RNwccj8PwxjF
   - Alex the analyst (excel, python)
   - freecodecamp
   - Kenji explains (Excel)
9. UCI machine learning repository UCI Machine Learning Repository
10. Python Data Science Handbook" by Jake VanderPlas https://jakevdp.github.io/PythonDataScienceHandbook/

11. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron https://github.com/Akramz/Hands-on-Machine-Learning-with-Scikit-Learn-Keras-and-TensorFlow

https://github.com/ageron/handson-ml2

Pdf downloaded

12. Chatgpt docx

Python_Libraries_Commands.docx

13. "Data Visualization with Python and Seaborn" tutorials
    a. https://thenextweb.com/news/a-beginners-guide-to-data-visualization-with-python-and-seaborn
14.