GreenData
Solutions

Introduction: At GreenData Solutions, we are fully committed to catalysing green environmental actions and healthy living across Africa based on sustainable utilisation of reliable data.

Vision Statement: Empowering sustainable futures through data-driven environmental and health insights.

Mission Statement: Providing accurate, actionable, and accessible environmental and health data to inform decision-making and drive positive change for sustainable living.

Core Values: Our core values are encapsulated in the acronym ASICI - Achieving Sustainability In Collaborative Innovations, as follows:

Accuracy: Delivering high-quality, reliable data to support informed decision-making.

Sustainability: Promoting environmentally responsible practices and minimizing our own ecological footprint.

Innovation: Embracing cutting-edge technologies and methodologies to stay at the forefront of environmental data services.

Collaboration: Fostering partnerships and open communication to drive environmental and health stewardship.

Integrity: Upholding transparency, accountability, and ethics in all aspects of our work.

GreenData Solutions

DATA CLEANING CHALLENGE: HEALTHCARE DATASET

DATA DESCRIPTION

1. **ID:** A unique identifier assigned to each individual or record.

2. **Name:** The full name of the individual.

3. **Age:** The age of the individual (in years).

4. **Gender:** The gender of the individual (e.g., Male, Female, Other).

5. **City:** The city where the individual resides.

6. **Blood Type:** The blood group of the individual (e.g., A+, B-, O+).

7. **Education:** The highest level of education attained by the individual (e.g., High School, Bachelor's, Master's).

8. **Employment Status:** The individual's current employment status (e.g., Employed, Unemployed, Student).

9. **Salary:** The individual's annual or monthly salary (in monetary units).

10. **Health Condition:** A description of the individual's overall health condition (e.g., Excellent, Good, Poor).

11. **Credit Score:** A numerical representation of the individual's creditworthiness (e.g., 720, N/A).

12. **Date of Admission:** The date the individual was admitted to a hospital or healthcare facility.

.

INTRODUCTION

You are hired as a Data Analytics Consultant in a healthcare system. Your primary responsibility is to analyze and clean messy data related to patients' health records. This involves identifying and addressing inconsistencies, missing values, and formatting errors, all of which are commonly encountered in real-world datasets.

This Data Cleaning Challenge is designed to help you enhance your critical thinking, attention to detail, and problem-solving skills while working with real-world messy data.

TASK DETAILS

This dataset is deliberately messy and includes various errors and inconsistencies. Your task is to:

1. Download the dataset attached to this email.

2. Perform exploratory data analysis (EDA) to understand the dataset's structure, identify anomalies, and gain preliminary insights.

3. Clean and prepare the data for further analysis, ensuring it is consistent, accurate, and ready for use.

4. Document your cleaning process, detailing:

*   The steps taken.
*   The logic behind each decision.
*   Any assumptions made during the cleaning process.

DELIVERABLES

1. Clean Dataset: Submit a cleaned version of the dataset that is ready for analysis.

2. Cleaning Process Report: Provide a detailed report that includes:

- A step-by-step description of the cleaning process.
- The decisions made and the reasoning behind them.
- Any limitations of the cleaned data.

2a. Identify cities or demographics with higher healthcare needs based on health conditions and admission dates.

3. Document Your Findings:

Document your findings and cleaning process using any of the following platforms:

- LinkedIn

- Medium

- GitHub

- Hashnode

- Any other platform you are comfortable with.

Share the link to your published report for review. We will feature selected reports on our page for more visibility.

HOW TO SUBMIT

- A link for submission will be provided before the end of the challenge.

- Use the hashtag #DataCleaningChallenge to share your journey.

BENEFITS

- Improve your data cleaning and analysis skills.

- Build a robust portfolio showcasing your ability to handle real-world data challenges.

- Gain visibility by having your work published on our LinkedIn page.

We are excited to see your creativity and expertise in action. Let's get started!.