Neodocs Healthcare – Data Analyst Assignment

Dataset Provided

uACR Testing Data (attached)
Result: Urine Albumin-to-Creatinine Ratio (uACR) value

Objective

You will use the provided dataset and/or external public health data to:

1. Explore and analyze patterns and trends in uACR test results.
2. Validate and contextualize your findings using publicly available health data (for example, state-level diabetes, hypertension, kidney-disease prevalence, census, or lifestyle data).
3. Provide code and a PDF report that explain your thought process and findings.

Question 1: Data Exploration & Preprocessing

● Handle missing or inconsistent data.
● Create meaningful descriptive statistics and visualizations:
● Age and gender distribution of tests.
● State/city-wise distribution of tests and abnormal results.
● Identify possible data-quality issues.

   Deliverables

   ● Cleaned dataset
   ● Summary tables and key visualizations
   ● Short discussion of data issues and how you addressed them

Question 2: Pattern & Trend Analysis

● Define at least three features of interest (e.g., age group, gender ratio, urban vs. rural clusters, seasonal trends if applicable).
● Classify and visualize states/cities by risk profile (e.g., high, medium, low prevalence of abnormal results).
● Discuss how your findings could guide healthcare or diagnostic strategy.

   Deliverables

   ● Feature definitions and rationale
   ● Analytical plots (heatmaps, bar charts, maps if desired)
   ● Key observations

Question 3: External Data Validation

- Identify and acquire at least one external dataset (examples: National Family Health Survey (NFHS)
- ICMR or WHO data on kidney disease, diabetes, hypertension State-level census or lifestyle statistics)
- Show how external data supports, explains, or challenges your patterns.
- Discuss possible causal or correlational factors.

   Deliverables

   - External dataset(s) source and brief description
   - Code and analysis combining internal and external data
   - Written discussion of validation and insights

Submission Instructions

1. Code:
   - One Python notebook or .py script with well-structured functions and comments.
   - Include data-cleaning, analysis, visualization
2. PDF (4–8 pages) explaining:
   - Approach and thought process
   - Key findings and validation Plots and tables for clarity
   - Keep it clear and business-relevant.
3. Reproducibility
   - Ensure code runs end-to-end with no hard-coded paths.
   - Document any external dataset sources.