This dataset contains information that is focused on health-related parameters of patients, potentially used for diagnosing chronic kidney disease (CKD). Here is a breakdown of the data and its relevance from my perspective:

|  |  |  |
| --- | --- | --- |
| Column name | datatype | About |
| id | int | A unique identification number for each patient |
| age | int | Age of the patient; useful for demographic insights |
| bp | int | Blood pressure readings; critical for understanding cardiovascular health |
| sg | float | Specific gravity of urine; indicative of kidney functionality |
| al | int | Albumin levels; an important biomarker for kidney health |
| su | int | Sugar levels in urine; connected to diabetes management |
| rbc | str | Red blood cell appearance; shows if there’s any abnormality related to anemia or other conditions |
| pc | str | Pus cell appearance; can indicate infections or inflammations |
| pcc | str | Presence of pus cell clumps; another infection marker |
| ba | str | Presence of bacteria in urine; highly indicative of infections |
| bgr | int | Blood glucose random levels; crucial for identifying diabetes |
| bu | int | Blood urea; a measure of kidney efficiency |
| sc | float | Serum creatinine; another essential parameter for kidney health |
| sod | float | Sodium levels; linked to electrolyte balance |
| pot | float | Potassium levels; critical for nerve and muscle function |
| hemo | float | Hemoglobin levels; used to assess anemia |
| pcv | int | Packed cell volume; associated with oxygen-carrying capacity |
| wc | int | White cell count; indicative of immune response |
| rc | float | Red cell count; tied to overall blood health |
| htn | str | Hypertension status (yes/no); a major CKD risk factor |
| dm | str | Diabetes mellitus status (yes/no); linked to kidney damage |
| cad | str | Coronary artery disease status (yes/no); connected to overall health risks |
| appet | str | Appetite (good/poor); general health indicator |
| pe | str | Presence of pedal edema (yes/no); a symptom of kidney issues |
| ane | str | Anemia status (yes/no); a result of decreased kidney function |
| classification | str | Target label indicating CKD status (ckd/notckd) |

This dataset is highly versatile for exploring data cleaning, visualization, and supervised learning. It allows for both regression (e.g., predicting blood pressure levels) and classification (e.g., identifying CKD). From my perspective, this dataset provides a rich opportunity for data preprocessing, feature engineering, and modeling to uncover insights about CKD.

col={'age': 'age',

'bp': 'blood\_pressure',

'sg': 'specific\_gravity',

'al': 'albumin',

'su': 'sugar',

'rbc': 'red\_blood\_cells',

'pc': 'pus\_cell',

'pcc': 'pus\_cell\_clumps',

'ba': 'bacteria',

'bgr': 'blood\_glucose\_random',

'bu': 'blood\_urea',

'sc': 'serum\_creatinine',

'sod': 'sodium',

'pot': 'potassium',

'hemo': 'hemoglobin',

'pcv': 'packed\_cell\_volume',

'wc': 'white\_blood\_cell\_count',

'rc': 'red\_blood\_cell\_count',

'htn': 'hypertension',

'dm': 'diabetes\_mellitus',

'cad': 'coronary\_artery\_disease',

'appet': 'appetite',

'pe': 'pedal\_edema',

'ane': 'anemia',

'classification': 'class'}

df.rename(columns=col, inplace=True)

df