

Data Preprocessing:

- **Created `pth_div_phosphorus`:** Calculated as `pth / phosphorus`, with 0 used when either value is missing or zero.
- **Created `alp_value`:** Derived from the average of `alp` and `alp_final`, using the non-missing value if only one is available, and defaulting to 0 if both are missing.
- **Created `corrected_calcium_by_albumin_div_phosphorus`:** Calculated as `BX_Corrected_Calcium_By_Albumin / phosphorus`, with 0 used when either value is missing or zero.
- **Created `bmi_value`:** Computed as `weight / (height * height)`, with 0 used when height or weight is missing or zero.
- **Created `alp_combined_with_cl_po4_newValue`:** Calculated as `alp_value * (chlorine / phosphorus)`, applied only when all values are non-zero and `alp_value` is within the reference range (44–147 IU/L); otherwise set to 0.
- **Dropped 31 columns:** Removed due to having more than 90% missing values.

```

columns_to_drop = [
    "radius_t_score",
    "radius_z_score",
    "99mtc_sestamibi",
    "4d_mri",
    "l1_l4_z_score",
    "femur_total_z_score",
    "99mtc_mibi",
    "femur_total_t_score",
    "l4_t_score",
    "alp_combined_with_cl_po4",
    "l3_t_score",
    "l2_t_score",
    "l1_l4_t_score",
    "l1_t_score",
    "total_protein",
    "first_surgery_pathology_report",
    "4d_ct",
    "l2_z_score",
    "l3_z_score",
    "l4_z_score",
    "l1_z_score",
    "ionized_calcium",
    "urinary_creatinine_24h",
    "femoral_neck_z_score",
    "femoral_neck_t_score",
    "parathyroid_scintigraphy",
    "bmi",
    "alp",
    "alp_final",
    "parathyroid_spect",
    "neck_ultrasound"
]

```

Resultant Data set:

Total number of columns without missing values: 25

Columns without missing values: ['aphp', 'nhph', 'secondary_hp', 'gender', 'tobacco', 'fracture_present', 'kidney_stones_present', 'abdominal_pain', 'fatigue', 'myalgia', 'constipation', 'insomnia', 'polydipsia', 'polyuria', 'muscle_weakness', 'headache', 'nausea', 'amnesia', 'gallstones', 'nephrolithiasis', 'pth_div_phosphorus', 'alp_value', 'corrected_calcium_by_albumin_div_phosphorus', 'bmi_value', 'alp_combined_with_cl_po4_newValue']

Total number of columns with missing values: 23

Columns with missing values: ['age', 'height', 'weight', 'magnesium', 'albumin', 'serum_calcium',
'serum_creatinine', 'gfr', 'phosphorus', 'urinary_calcium_24h', 'Triglycerides', 'ldl_cholesterol',
'total_cholesterol', 'ast', 'alt', 'ggt', 'bun', 'chlorine', 'serum_25_hydroxy_vitamin_d', 'pth',
'corrected_calcium_by_albumin', 'serum_calcium_phosphorus_ratio', 'pf_index']

Total number of columns with missing values: 23

Column Name	Missing %	Column Name	Missing %
urinary_calcium_24h	85.08	Triglycerides	53.68
chlorine	82.74	magnesium	48.53
gfr	82.03	serum_25_hydroxy_vitamin_d	40.32
serum_calcium_phosphorus_ratio	80.35	albumin	37.33
phosphorus	80.29	corrected_calcium_by_albumin	37.09
height	79.87	age	32.83
weight	79.75	ast	32.47
pf_index	75.79	pth	30.50
total_cholesterol	68.00	alt	26.90
bun	66.27	serum_creatinine	19.59
ldl_cholesterol	56.08	serum_calcium	13.84
ggt	55.06		
