**Data Cleaning Steps**

**Beneficiary Data**

1. Column Removal:
   * Deleted the DOD column due to over 99% null values.
2. Data Transformation:
   * Replaced Y in the RenalDiseaseIndicator column with 1 (Yes) and kept 0 as No. Converted its type to an integer.
3. Duplicates:
   * Removed all duplicate rows.
4. Data Type Conversion:
   * Converted financial columns (IPAnnualReimbursementAmt, IPAnnualDeductibleAmt, OPAnnualReimbursementAmt, OPAnnualDeductibleAmt) to fixed decimal numbers for consistency.
5. Disease Columns:
   * Replaced all 2 values in the chronic condition columns (e.g., ChronicCond\_Alzheimer, ChronicCond\_Heartfailure) with 0 (indicating no disease).
   * Simplified column names by removing the prefix ChronicCond\_, leaving only the disease names for clarity.
6. Gender Column:
   * Replaced 1 with "Male" and 2 with "Female" in the Gender column for better interpretability.

**Inpatient Data**

1. Duplicates:
   * Removed duplicate rows.
2. Data Type Conversion:
   * Converted InscClaimAmtReimbursed to a fixed decimal number.
   * Replaced NA in DeductibleAmtPaid with 0 (indicating no payment) and converted it to a fixed decimal number.
3. DiagnosisGroupCode Handling:
   * Removed rows where DiagnosisGroupCode contained less than 1% of data as OTH (Others), as it caused errors during type conversion.

**Outpatient Data**

1. Duplicates:
   * Removed duplicate rows.
2. Data Type Conversion:
   * Converted InscClaimAmtReimbursed and DeductibleAmtPaid to fixed decimal numbers.
3. Missing Values:
   * Replaced empty values in ClmAdmitDiagnosisCode with NA for consistency.

**Provider Data**

1. Header Adjustment:
   * Converted the first row to column headers.
2. Duplicates:
   * Removed duplicate rows.