Data Fields Overview

ENCOUNTER

- **ENCOUNTERID**
  Specifies an ID for a given encounter (used to identify a unique encounter).
- **PATID**
  Specifies a unique patient ID (used to identify a unique patient).
- **ADMIT_DATETIME, DISCHARGE_DATETIME**
  The admit date and time at which a patient was admitted for a given encounter; the discharge date and time for a given patient encounter (used to obtain the patient window to extract data).

ENC_TYPE (One Hot Encoding)

Specifies the type of patient encounter:

- **ENC_TYPE_ED**: ED = Emergency Department
- **ENC_TYPE_OS**: OS = Observation Stay
- **ENC_TYPE_AV_TH**: AV = Ambulatory Visit, TH = Telehealth
- **ENC_TYPE_IP**: IP = Inpatient Hospital Stay
- **EI**: Emergency Department Admit to Inpatient Hospital Stay (permissible substitution)
- **ENC_TYPE_CLPSED**: IS = Non-Acute Institutional Stay, IC = Institutional Professional Consult (permissible substitution), OA = Other Ambulatory Visit, NI = No information, UN = Unknown, OT = Other

DISCHARGE_STATUS (One Hot Encoding)

Specifies how the patient was discharged:

- **DISCHARGE_STATUS_HO**: HO = Home / Self Care
- **DISCHARGE_STATUS_HH**: HH = Home Health
- **DISCHARGE_STATUS_SN_RH**: RH = Rehabilitation Facility, SN = Skilled Nursing Facility
- **DISCHARGE_STATUS_AM**: AM = Against Medical Advice
- **DISCHARGE_STATUS_CLPSED**: SH = Still in Hospital, AL = Assisted Living Facility, EX = Expired, HS = Hospice, IP = Other Acute Inpatient Hospital, NH = Nursing Home (Includes ICF)

- ADMIT_DURATION
  The difference between discharge and admit datetime for a given encounter (Numerical Feature).
- days_since_prior_enc_patient_level
  Difference in days between the ADMIT_DATETIME of the current encounter and the DISCHARGE_DATETIME of the prior encounter (Numerical Feature).
- days_since_prior_IP_OS_ED_patient_level_int
  Difference in days between the ADMIT_DATETIME of the current encounter and the DISCHARGE_DATETIME of the prior IP/OS/ED encounter (Numerical Feature).
- prior_IP_OS_ED_count
  Count of the patient's IP, OS, or ED encounters within a 2-year look-back period (Numerical Feature).
- prior_other_encounter_types_count
  Cumulative count of prior non-'AV_TH' encounters for each patient up to the current encounter (Numerical Feature).
- distance
  Calculates the distance between the FACILITY_LOCATION and ZIP_CODE using query_postal_code (Numerical Feature).
- readmission (One Hot Encoding)
  Marks a second inpatient encounter as a readmission if it occurs within 30 days of the first. If a second inpatient encounter occurs between 31-33 days, it checks if the patient had any ED, OS, OA, or AV encounters in the 3 days prior. If found, it is marked as a readmission.

DEMOGRAPHICS

- **age**
  Calculated from the Patient's DOB to the date they were admitted (**ADMIT_DATETIME**) in the encounter table (Numerical Feature).

SEX (One Hot Encoding)

- **SEX_F**: Female
- **SEX_M**: Male
  (Removed patients with missing fields. Carried forward sex values for all encounters of a given patient with at least one sex identification.)

HISPANIC (One Hot Encoding)

- **HISPANIC_N**: No
- **HISPANIC_Y**: Yes
- **HISPANIC_CLPSED**: R = Refuse to answer, NI = No information, UN = Unknown, OT = Other

RACE (One Hot Encoding)

- **RACE_01_02_04**: 01 = American Indian or Alaska Native, 02 = Asian, 04 = Native Hawaiian or Other Pacific Islander
- **RACE_03**: 03 = Black or African American
- **RACE_05**: 05 = White
- **RACE_CLPSED**: 06 = Multiple race, 07 = Refuse to answer, NI = No information, UN = Unknown, OT = Other

RUCA_CODE (One Hot Encoding)

- **RUCA_CODE_1.0**: Metropolitan area core: primary flow within an urbanized area (UA)
- **RUCA_CODE_2.0**: Metropolitan area high commuting: primary flow 30% or more to a UA
- **RUCA_CODE_3+**: Other codes indicating varying levels of urbanization and commuting behavior
- **RUCA_CODE_CLPSED**: Missing value for this feature

SDI Quantile (One Hot Encoding)

SDI linked to clinical data based on patient residential ZIP codes mapped to ZCTA-level data:

- **SDI Quantile 0 (1-53)**: Least deprived
- **SDI Quantile 1 (53.3-83)**
- **SDI Quantile 2 (83.3-95)**
- **SDI Quantile 3 (95.33-98.66)**
- **SDI Quantile 4 (99-100)**: Most deprived

VITALS

TOBACCO (One Hot Encoding)

- **TOBACCO_01**: Current user
- **TOBACCO_02**: Never
- **TOBACCO_03**: Quit/former user
- **TOBACCO_CLPSED**: Passive or environmental exposure, Not asked, NI = No information, UN = Unknown, OT = Other

BMI (One Hot Encoding)

Imputed from height and weight values of a given encounter or carried forward from the previous encounter if missing:

- **BMI_<18.5**
- **BMI_18.5 – 24.9**
- **BMI_25.0 – 29.9**
- **BMI_30.0+**
- **BMI_CLPSED**: Missing BMI

DIASTOLIC (Blood Pressure Reading)

For each encounter, calculates the min, max, and mean of the diastolic values, normalized using global min/max values, stored in a semicolon-separated string format.

SYSTOLIC (Blood Pressure Reading)

Same normalization process as diastolic.

LAB RESULT CM

- **Loincs_results_modifier_unit**
  Patient's lab results for a given encounter. The result for each LOINC code is normalized using population minimum and maximum values for the specific LOINC code, becoming a vector of features representing that patient's lab results.

CKD (One-Hot Encoding)

Five stages of chronic kidney disease based on GFR lab codes: '69405-9', '33914-3', '48642-3':

- **CKD_0**: No chronic kidney disease (GFR codes not found)
- **CKD_1**: Normal or high GFR (>90 mL/min)
- **CKD_2**: Mild CKD (GFR = 60-89 mL/min)
- **CKD_3**: Moderate CKD (GFR = 30-59 mL/min)
- **CKD_4**: Severe CKD (GFR = 15-29 mL/min)
- **CKD_5**: End Stage CKD (GFR <15 mL/min)

DIAGNOSIS

- **n_dx**: Number of ICD-10 diagnostic codes for a given encounter (before being converted to CCS)
- **n_px**: Number of ICD-10 procedural codes for a given encounter

DX_CCS_codes (One Hot Encoding + SVD)

Chi-square significant diagnostic codes converted to CCS.

PX_CCS (One Hot Encoding + SVD)

Chi-square significant procedural codes converted to CCS.

ELIXHAUSER (One Hot Encoding)

Created based on diagnostic ICD-10 codes.

PRESCRIBING

Diabetes RXNORM_CUI (One Hot Encoding)

- **Insulins**
- **Glucagon-like peptide-1 (GLP-1) analogues**
- **Biguanides**
- **Sulfonylureas**
- **Alpha glucosidase inhibitors**
- **Thiazolidinediones**
- **Dipeptidyl peptidase 4 (DPP-4) inhibitors**
- **Sodium-glucose co-transporter 2 (SGLT2) inhibitors**
- **Other blood glucose lowering drugs, excl. insulins**
- **Aldose reductase inhibitors**

Non-Diabetes RXNORM_CUI (One Hot Encoding)

- **AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM**
- **ANTIHYPERTENSIVES**
- **BETA BLOCKING AGENTS**
- **CORTICOSTEROIDS FOR SYSTEMIC USE**
- **LIPID MODIFYING AGENTS**