

```
##
## ### Summary of Table: admissions ###
## Number of rows: 58976
## Number of columns: 19
## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## row_id           integer             0         0.00
## subject_id        integer             0         0.00
## hadm_id           integer             0         0.00
## admittime         POSIXct             0         0.00
## dischtime         POSIXct             0         0.00
## deathtime         POSIXct          53122        90.07
## admission_type     character           0         0.00
## admission_location character           0         0.00
## discharge_location character           0         0.00
## insurance         character           0         0.00
## language          character           0         0.00
## religion          character           0         0.00
## marital_status     character           0         0.00
## ethnicity         character           0         0.00
## edregtime         POSIXct          28099        47.64
## edouttime         POSIXct          28099        47.64
## diagnosis         character           0         0.00
## hospital_expire_flag integer           0         0.00
## has_chartevents_data integer           0         0.00
```

```
##
## ---
```

```
## ### Summary of Table: antibiotics ###
## Number of rows: 164927
## Number of columns: 16
## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## icustay_id      integer           74         0.04
## starttime       POSIXct           24         0.01
## endtime         POSIXct            7         0.00
## amount          numeric            7         0.00
## amountuom       character           0         0.00
## rate            logical        164927       100.00
## rateuom         logical        164927       100.00
## ordercategoryname character         0         0.00
## patientweight   numeric           31         0.02
## totalamount     integer          8315         5.04
## totalamountuom  character           0         0.00
## statusdescription character         0         0.00
## label           character           0         0.00
## abbreviation    character           0         0.00
## antibiotic      integer            0         0.00
## dbsource        character           0         0.00
```

```
##
## ---
##
```

```
## ### Summary of Table: bloodculture ###
## Number of rows: 632506
## Number of columns: 10
```

```

## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## hadm_id      integer           0         0.00
## icustay_id    integer       156798        24.79
## dy            integer       156798        24.79
## hr            integer       187726        29.68
## charttime     POSIXct         41791         6.61
## chartdate     POSIXct          0         0.00
## org_name      character        0         0.00
## positiveculture integer        0         0.00
## ab_name       character        0         0.00
## antibioticresistance character  0         0.00
##
## ---
##
## ### Summary of Table: gcs_hourly ###
## Number of rows: 1515342
## Number of columns: 7
## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## icustay_id    integer           0         0.00
## hr            integer           0         0.00
## gcs           integer           0         0.00
## gcseyes       integer       1444         0.10
## gcsmotor      integer       3768         0.25
## gcsverbal     integer       3568         0.24
## endotrachflag integer           0         0.00
##
## ---
##
## ### Summary of Table: icd9_diag ###
## Number of rows: 651047
## Number of columns: 7
## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## row_id        integer           0         0.00
## subject_id     integer           0         0.00
## hadm_id        integer           0         0.00
## seq_num        integer         47         0.01
## icd9_code      character        0         0.00
## short_title    character        0         0.00
## long_title     character        0         0.00
##
## ---
##
## ### Summary of Table: icustays ###
## Number of rows: 61532
## Number of columns: 12
## Column names and data types:
##           Data_Type Missing_Count Missing_Pct.
## row_id        integer           0         0.00
## subject_id     integer           0         0.00
## hadm_id        integer           0         0.00
## icustay_id     integer           0         0.00
## dbsource       character        0         0.00
## first_careunit character        0         0.00

```

```

## last_careunit character      0      0.00
## first_wardid integer        0      0.00
## last_wardid integer        0      0.00
## intime POSIXct             0      0.00
## outtime POSIXct            10      0.02
## los numeric                10      0.02
##
## ---
##
## ### Summary of Table: labs_hourly ###
## Number of rows: 928195
## Number of columns: 22
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## icustay_id integer              0      0.00
## hr integer                      0      0.00
## neutrophil numeric             853757    91.98
## creactiveprotein numeric       926422    99.81
## whitebloodcell numeric        576380    62.10
## partialpressureo2 numeric     473825    51.05
## bicarbonate numeric          534726    57.61
## lactate numeric              763646    82.27
## troponin numeric             886778    95.54
## bloodureanitrogen numeric     549233    59.17
## creatinine numeric           547907    59.03
## alaninetransaminase numeric    836577    90.13
## aspartatetransaminase numeric  836642    90.14
## hemoglobin numeric           514069    55.38
## intnormalisedratio numeric    675513    72.78
## platelets numeric            559952    60.33
## albumin numeric              866292    93.33
## chloride numeric             496072    53.44
## glucose numeric              401175    43.22
## sodium numeric              524290    56.48
## bilirubin numeric            816141    87.93
## hematocrit numeric           487396    52.51
##
## ---
##
## ### Summary of Table: output_hourly ###
## Number of rows: 3325543
## Number of columns: 3
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## icustay_id integer              0      0.00
## hr integer                      0      0.00
## urineoutput numeric            11920     0.36
##
## ---
##
## ### Summary of Table: patients ###
## Number of rows: 46520
## Number of columns: 8
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## row_id integer                  0      0.00

```

```

## subject_id    integer           0          0.00
## gender        character          0          0.00
## dob           POSIXct           0          0.00
## dod           POSIXct        30761        66.12
## dod_hosp      POSIXct        36546        78.56
## dod_ssn       POSIXct        33142        71.24
## expire_flag   integer           0          0.00
##
## ---
##
## #### Summary of Table: pt_icu_outcome ####
## Number of rows: 61533
## Number of columns: 17
## Column names and data types:
##
##          Data_Type Missing_Count Missing_Pct.
## row_id      integer           0          0.00
## subject_id   integer           0          0.00
## dob          POSIXct           0          0.00
## hadm_id      integer           0          0.00
## admittance   POSIXct        12348        20.07
## dischtime    POSIXct        12348        20.07
## icustay_id   integer           0          0.00
## age_years    numeric           0          0.00
## intime       POSIXct           0          0.00
## outtime      POSIXct          10          0.02
## los          numeric          10          0.02
## hosp_deathtime POSIXct        59256        96.30
## icu_expire_flag integer           0          0.00
## hospital_expire_flag integer        12348        20.07
## dod          POSIXct        37341        60.68
## expire_flag   integer           0          0.00
## ttd_days      integer        37341        60.68
##
## ---

```

```
##
## ### Summary of Table: pt_stay_hr ###
## Number of rows: 3687586
## Number of columns: 9
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## icustay_id    integer             0          0.00
## hadm_id       integer             0          0.00
## subject_id    integer             0          0.00
## intime        POSIXct             0          0.00
## outtime       POSIXct             0          0.00
## starttime     POSIXct             0          0.00
## endtime       POSIXct             0          0.00
## hr            integer             0          0.00
## dy            integer           1398          0.04
##
## ---
##
## ### Summary of Table: pt_weight ###
## Number of rows: 396241
## Number of columns: 11
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## icustay_id    integer             0          0.00
## dy            integer             0          0.00
## starttime     POSIXct             0          0.00
## endtime       POSIXct             0          0.00
## admissionweight numeric        331567         83.68
## dailyweight   numeric        241485         60.94
## previousweight numeric        312334         78.82
## echoweight    numeric        262244         66.18
## avg_weight_naive numeric        20263          5.11
## min_weight    numeric        20263          5.11
## max_weight    numeric        20263          5.11
##
## ---
```

```
##
## ### Summary of Table: pv_mechvent ###
## Number of rows: 694958
## Number of columns: 21
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## icustay_id      integer           0          0.00
## charttime       POSIXct          76          0.01
## starttime       POSIXct           0          0.00
## endtime         POSIXct           0          0.00
## duration_hours  numeric           0          0.00
## ventnum         integer           0          0.00
## minutevolume    numeric        233141        33.55
## settidalvolume  numeric        461276        66.37
## obstidalvolume  numeric        314290        45.22
## sponttidalvolume numeric        439527        63.25
## setpeep         numeric         21446         3.09
## totalpeep       numeric        669702        96.37
## pressurehighapr integer        694415        99.92
## pressurelowapr  numeric        694428        99.92
## timehighapr     numeric        694419        99.92
## timelowapr      numeric        694423        99.92
## meanairwaypressure numeric      236821        34.08
## peakinsppressure numeric      322901        46.46
## neginspforce    numeric        693945        99.85
## insptime        numeric        572792        82.42
## plateaupressure numeric        564736        81.26
```

```
##
## ---
```

```
##
## ### Summary of Table: transfers ###
## Number of rows: 261897
## Number of columns: 13
## Column names and data types:
##          Data_Type Missing_Count Missing_Pct.
## row_id      integer           0          0.00
## subject_id   integer           0          0.00
## hadm_id      integer           0          0.00
## icustay_id   integer        174176        66.51
## dbsource     character         0          0.00
## eventtype    character         0          0.00
## prev_careunit character         0          0.00
## curr_careunit character         0          0.00
## prev_wardid  integer        58933         22.50
## curr_wardid  integer        58943         22.51
## intime       POSIXct          24          0.01
## outtime      POSIXct        58976         22.52
## los          numeric        58976         22.52
```

```
##
## ---
```

```
##
## ### Summary of Table: vasopressors ###
## Number of rows: 314964
## Number of columns: 11
## Column names and data types:
```

```
##                               Data_Type Missing_Count Missing_Pct.
## icustay_id                   integer           834           0.26
## starttime                    POSIXct             0           0.00
## endtime                      POSIXct          231185          73.40
## norepinephrine_rate          numeric          231436          73.48
## norepinephrine_amount        numeric          237721          75.48
## epinephrine_rate             numeric          259713          82.46
## epinephrine_amount           numeric          278082          88.29
## dopamine_rate               numeric          181993          57.78
## dopamine_amount              numeric          227051          72.09
## dobutamine_rate              numeric          272450          86.50
## dobutamine_amount            numeric          286648          91.01
##
## ---
##
## ### Summary of Table: vitals_hourly ###
## Number of rows: 7292362
## Number of columns: 11
## Column names and data types:
##                               Data_Type Missing_Count Missing_Pct.
## icustay_id                   integer             0           0.00
## hr                           integer             0           0.00
## spo2                         numeric          1972385          27.05
## fio2                         numeric          6341965          86.97
## temperature                  numeric          5726438          78.53
## resprate                     numeric          2550173          34.97
## heartrate                    numeric           872012          11.96
## sysbp                        numeric          2647890          36.31
## diasbp                       numeric          2648668          36.32
## glucose                      numeric          6144343          84.26
## meanarterialpressure          numeric          2631818          36.09
##
## ---
```

```
##
## All CSV tables have been successfully loaded and summarized!
```

```
## Loaded tables:
## admissions, antibiotics, bloodculture, gcs_hourly, icd9_diag, icustays, labs_hourly, output
## _hourly, patients, pt_icu_outcome, pt_stay_hr, pt_weight, pv_mechvent, transfers, vasopressors,
## vitals_hourly
```

Step 2: Data Preprocessing

Objectives:

- Address missing data.
- Create a clean dataset for feature engineering.

Plan:

1. Handle Missing Values:

- Use domain knowledge to impute or discard variables:
 - For vital signs (`vitals_hourly`), calculate mean/median values for missing hourly data.