COL362/632 - DBMS - Assignment 1: SQL Queries

Deadline

All submissions should be made on Moodle by **3rd Feb 2025, 11:59 PM**.

General Instructions

Follow all instructions. Submissions not following these instructions will not be evaluated and will be given **Zero** marks.

1. Kindly ensure that this assignment is completed independently. Collaboration with external entities, including individuals, AI agents, websites, discussion forums, etc., is strictly prohibited. You can discuss and post questions on the piazza to seek clarification **until 1st Feb**.
2. You must use **PostgreSQL 15.x** for this assignment.
3. There are a total of 35 queries in this assignment.
4. **Marking scheme:** Queries carry different weights based on the perceived difficulty level of the query. These weights are not revealed to you (intentionally). The weights will be released after the evaluation.
5. We will evaluate your assignment in an automated fashion, so ensure that your folder names, file names, and directory structure strictly follow the instructions. There is a zip file provided for your convenience. Make sure to change the filename.
6. Each SQL query will be graded in a binary; there will be no partial grading.
7. Note that there will be no deadline extensions. But you can use your 72 late hours. Refer to lecture notes on course organization for late-hour rules.
8. Download the cleaned-up data from this link. It is a .sql file that can be imported directly to PostgreSQL. Refer to lecture notes on importing a .sql file into PostgreSQL.

9. A single zip has to be submitted (for you convinence we have provided an empty file that you should use).
    a. The zip is structured such that upon deflating, all submission files appear under a directory with the student's entry number. For, suppose a student's entry number is 22XXCSXX123. In that case, the zip submission should be named 22XXCSXX123.zip, and upon deflating, all contained files should be under the directory named ./22XXCSXX123 only (names should be in uppercase). Your submission might be rejected and not be evaluated if you do not adhere to these specifications.
    b. There are four sections (each of different difficulty levels) in the assignment, i.e., "Basic", "Intermediate", "Advanced", and "Graph". For this, we have created four different folders named "BASIC", "INTERMEDIATE", "ADVANCED", and "GRAPH".
    c. In each designated section for every question, write your query in the .sql file following the naming convention <folder name>_<question number>.sql. For instance, the solution to question "1" within the "Basic" section should be saved in a file named "**BASIC_1.sql**". Similarly, the response to question "2" in the "Graph" section must be saved in a file named "**GRAPH_2.sql**". Please do not write comments or other text except for an SQL query in the .sql files. **Note don't start your queries with %sql or %%sql, that is only specific to notebook.**

d. The final directory structure should look like below:

```
22XXCSXX123/
|
├── BASIC/
|   ├── BASIC_1.sql
|   ├── BASIC_2.sql
|   └── ...
|
├── INTERMEDIATE/
|   ├── INTERMEDIATE_1.sql
|   ├── INTERMEDIATE_2.sql
|   └── ...
|
├── ADVANCED/
|   ├──ADVANCED_1.sql
|   ├── ADVANCED_2.sql
|   └── ...
|
└── GRAPH/
    ├── GRAPH_1.sql
    ├── GRAPH_2.sql
    └── ...
```

Dataset

This dataset is structured into two primary modules—hospital ("hosp") and intensive care unit ("icu")—which together span tens of thousands of patient encounters and aggregate millions of individual records. We highly suggest to also examine the schema and data before writing your queries.

Hospital Module ("hosp"):

- **Scope:** Derived from a hospital-wide electronic health record, this module covers a broad spectrum of patient care data primarily recorded during hospital stays, with some information extending to outpatient settings.

- **Data Included:**
  - Patient and admission details (e.g., basic demographics, admissions, transfers)
  - Laboratory measurements and related metadata
  - Microbiology culture results
  - Provider orders and medication administration details
  - Medication prescriptions and pharmacy information
  - Billing data, including diagnostic and procedure codes, billing events, and service-related information
  - Additional service-related records

ICU Module ("icu"):

- **Scope:** Sourced from a clinical information system specific to intensive care, this module focuses on high-resolution ICU data. It has been organized into a star schema linking ICU stays to a variety of detailed event records.
- **Data Included:**
  - Patient ICU stays and associated metadata
  - Records of intravenous fluids and medications, including their ingredients
  - Patient outputs, procedures performed, and charted observations
  - Events captured as specific date/time entries
- **Structure:** Each event record is linked by stay and item identifiers for detailed traceability across tables.

Description

hosp Module

Table Name	Description
admissions	Information about each hospital admission (hadm_id), including admission and dis type, location, etc.
d_hcpcs	Dimension table describing codes used in **hcpcsevents** (mostly CPT codes).
d_icd_diagnoses	Dimension table describing ICD-9/ICD-10 diagnosis codes used in **diagnoses_icd**.
d_icd_procedures	Dimension table describing ICD-9/ICD-10 procedure codes used in **procedures_ic**
d_labitems	Dimension table describing laboratory test items. Links to **labevents**.
diagnoses_icd	ICD-9/ICD-10 diagnoses billed for each hospital admission. Links to **d_icd_diagno**
drgcodes	Diagnosis-related group (DRG) codes associated with hospital stays.
emar	Electronic Medication Administration Record (eMAR) data, documenting medicatio (barcode scanning at bedside).
emar_detail	Detailed breakdown of each eMAR event, including dose given and route.
hcpcsevents	Hospital billing events containing HCPCS codes. Links to **d_hcpcs**.
labevents	Laboratory measurements and results from hospital-wide labs. Links to **d_labitem**
microbiologyevents	Microbiology cultures and antibiotic sensitivities.
omr	Online Medical Record data, often containing outpatient-type measurements (heig
patients	Core patient-level information (gender, anchor age, date of death).
pharmacy	Detailed records of filled medications, including dosing information and status. Lin **emar** via *pharmacy_id*.
poe	Provider Order Entry (POE) records. General interface for entering patient orders (etc.).
poe_detail	Supplemental detail (in an Entity-Attribute-Value format) for entries in **poe**.
prescriptions	Medications ordered for a patient, including route, frequency, and identifying code **pharmacy** and **emar**.
procedures_icd	ICD-9/ICD-10 procedure codes billed during hospital stays. Links to **d_icd_proced**
provider	Lists de-identified provider identifiers used across *hosp* (e.g., *order_provider_id*).
services	Tracks clinical services assigned to the patient (e.g., “Surgery,” “Cardiology”).
transfers	Tracks physical movement of patients across different wards/units in the hospital.

icu Module

Table Name	Description
caregiver	De-identified caregiver identifiers for the ICU module.
chartevents	The majority of ICU charted data (vital signs, ventilator settings, etc.). Links to d_i itemid.
datetimeevents	Date/time-type measurements charted in the ICU (e.g., insertion dates).
d_items	Dimension table defining itemid for ICU measurements.
icustays	Defines individual ICU stays, with admission and discharge times.
ingredientevents	Ingredients used within the same solution for a drug (e.g., multiple additives in an
inputevents	Continuous infusions or intermittent administrations (IV fluids, drips, etc.).
outputevents	Output data (urine output, drain volumes) recorded in the ICU.
procedureevents	Procedures documented during the ICU stay (e.g., mechanical ventilation, imaging start/end times and caregiver info.

Detailed Tables of Columns

Below is a reference table for **each** dataset table. Columns appear in the following format:

```
| Column Name | Data Type | Description |
```

hosp Module

- **admissions**

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing the *patients* table (each patient
hadm_id	bigint	**Primary key** for each hospital admission (unique per ad
admittime	text	Date and time the patient was admitted.
dischtime	text	Date and time the patient was discharged.
deathtime	text	If applicable, date and time of in-hospital death (often n
admission_type	text	General classification of the admission's urgency or typ etc.).
admit_provider_id	text	De-identified ID of the provider who admitted the patien **provider** table if needed).
admission_location	text	Location of the patient before arriving (e.g., emergency
discharge_location	text	Location to which the patient was discharged (e.g., hom
insurance	text	Type of insurance covering the patient.
language	text	Recorded primary language of the patient.
marital_status	text	Patient's marital status as documented.
race	text	Race documented at admission.
edregtime	text	Date/time the patient was registered in the emergency d
edouttime	text	Date/time the patient left the emergency department.
hospital_expire_flag	bigint	Indicates if patient died during the admission (1) or surv

Keys and Links

- **subject_id** → *patients.subject_id*
- **hadm_id** is unique within this table (defining each admission).

- d_hcpcs

Column Name	Data Type	Description
code	text	HCPCS/CPT code (five characters). **Primary key** for this
category	double precision	Broad numeric category for the code.
long_description	text	Longer descriptive text for the code.
short_description	text	Short descriptive text for the code.

Keys and Links

- Joins to **hcpcsevents** on hcpcs_cd = d_hcpcs.code.

- **d_icd_diagnoses**

Column Name	Data Type	Description
icd_code	text	ICD diagnosis code. Together with icd_version, this form diagnosis.
icd_version	bigint	ICD version (9 or 10).
long_title	text	Full descriptive title for the ICD code (e.g., "Cholera due

Keys and Links

- Joins to **diagnoses_icd** on (icd_code, icd_version).

- **d_icd_procedures**

Column Name	Data Type	Description
icd_code	text	ICD procedure code. Together with **icd_version**, forms procedure.
icd_version	bigint	ICD version (9 or 10).
long_title	text	Descriptive name for the procedure code.

Keys and Links

- Joins to **procedures_icd** on (icd_code, icd_version).

- **d_labitems**

Column Name	Data Type	Description
itemid	bigint	**Primary key**. Unique identifier for a laboratory concept.
label	text	Human-readable description of the lab test (e.g., "Sodiu
fluid	text	The substance on which the measurement was taken (e
category	text	Higher-level grouping for the lab measurement (e.g., "C "ABG").

Keys and Links

- Joins to **labevents** on itemid.

- **diagnoses_icd**

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing *patients*.
hadm_id	bigint	**Foreign key** referencing *admissions*.
seq_num	bigint	Order/priority of the diagnosis for billing purposes.
icd_code	text	ICD code for the diagnosis.
icd_version	bigint	ICD version (9 or 10).

Keys and Links

- Joins to **d_icd_diagnoses** on (icd_code, icd_version).

- **drgcodes**

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing *patients*.
hadm_id	bigint	**Foreign key** referencing *admissions*.
drg_type	text	The DRG ontology used.
drg_code	bigint	DRG code associated with the stay.
description	text	Description of the DRG code.
drg_severity	double precision	Severity of illness rating linked to the DRG.
drg_mortality	double precision	Mortality risk rating linked to the DRG.

- emar

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing *patients*.
hadm_id	double precision	**Foreign key** referencing *admissions*.
emar_id	text	Unique identifier for the eMAR entry.
emar_seq	bigint	Consecutive integer that orders eMAR entries chronolog
poe_id	text	**Foreign key** referencing **poe**. Links to the provider orde
pharmacy_id	double precision	**Foreign key** referencing **pharmacy**.
enter_provider_id	text	De-identified ID of the provider entering the eMAR. (a d the provider who entered the medication administration
charttime	text	Time the medication administration was charted.
medication	text	Name of the medication administered.
event_txt	text	High-level info about administration status (e.g., "Admin
scheduletime	text	Time the administration was scheduled, if available.
storetime	text	Time the administration was documented in the system.

- emar_detail

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing *patients*.
emar_id	text	**Foreign key** referencing **emar**. Unique ID for each order
emar_seq	bigint	Consecutive integer which sorts eMAR entries.
parent_field_ordinal	double precision	Identifies multiple doses/formulary items for the same e
administration_type	text	Type of administration (e.g., “IV Bolus,” “Medication Infu
pharmacy_id	double precision	**Foreign key** referencing **pharmacy**.
barcode_type	text	Type of barcode scanned (if any).
reason_for_no_barcode	text	Reason no barcode was scanned if barcode scanning wa
complete_dose_not_given	text	Indicator if the full dose was not given.
dose_due, dose_due_unit	text	The expected dose and units for the administration.
dose_given, dose_given_unit	text	The actual dose administered and units.
will_remainder_of_dose_be_given	text	Indicates if leftover dose will be administered later.
product_amount_given	double precision	The actual amount of product administered.
product_unit	text	The unit for the product amount.
product_code, product_description	text	Code/description for the scanned product.
prior_infusion_rate	double precision	Infusion rate prior to this administration.
infusion_rate	double precision	The new infusion rate.
infusion_rate_adjustment	text	Indicates if the rate has been adjusted.
infusion_rate_adjustment_amount	double precision	Amount of the rate adjustment.
infusion_rate_unit	text	The unit for the infusion rate (e.g., mL/hr).
route	text	Route of administration (e.g., IV, Oral).

infusion_complete	text	Whether the infusion completed as planned.
completion_interval	text	Additional time or instructions for completing the infusio
new_iv_bag_hung	text	Indicates if a new IV bag was hung.
continued_infusion_in_other_location	text	If the infusion continued after patient transfer.
restart_interval	text	Interval before restarting an infusion.
side, site	text	Documented side/site of administration (e.g., left arm).
non_formulary_visual_verification	text	Documents if a non-formulary medication was visually v

---

- hcpcsevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
chartdate	text	Date associated with the coded event.
hcpcs_cd	text	The HCPCS/CPT code; join with d_hcpcs.code.
seq_num	bigint	Sequence number for the code within a single admissio
short_description	text	Short textual description for the code.

- labevents

Column Name	Data Type	Description
labevent_id	bigint	Unique identifier for each row in **labevents**.
subject_id	bigint	**Foreign key** referencing *patients*.
hadm_id	double precision	**Foreign key** referencing *admissions*.
specimen_id	bigint	Identifier grouping measurements from the same specir
itemid	bigint	Lab item tested; join with **d_labitems.itemid**.
order_provider_id	text	Provider who ordered the lab test (de-identified).
charttime	text	Time the sample was taken or measurement made.
storetime	text	Time the measurement was finalized and stored in the d
value	text	The lab test result in text form.
valuenum	double precision	Numeric value of the lab test result if applicable.
valueuom	text	Unit of measurement.
ref_range_lower	double precision	Lower bound of the normal range for this lab test, if kno
ref_range_upper	double precision	Upper bound of the normal range for this lab test, if kno
flag	text	Indicates if the result is abnormal or out of normal range
priority	text	Lab priority (routine/stat).
comments	text	Additional notes regarding the measurement. Deidentifi

---

- microbiologyevents

Column Name	Data Type	Description
microevent_id	bigint	Unique identifier for each row in **microbiologyevents**.
subject_id	bigint	**Foreign key** referencing *patients*.
hadm_id	double precision	**Foreign key** referencing *admissions*.
micro_specimen_id	bigint	Identifier grouping measurements from the same microb
order_provider_id	text	Provider ID for the microbiology test order.
chartdate	text	Date of the microbiology observation.
charttime	text	Time of the observation if available (NULL if only the da
spec_itemid	bigint	Item ID describing the type of specimen.
spec_type_desc	text	Text describing the specimen (e.g., “Blood,” “Urine”).
test_seq	bigint	Sequence number differentiating multiple tests on the s
storedate	text	Date when the result was finalized.
storetime	text	Date/time when the result was finalized.
test_itemid	bigint	ID corresponding to the test performed.
test_name	text	Name of the microbiology test (e.g., “Gram Stain,” “Cult
org_itemid	double precision	ID of the organism that grew (if any).
org_name	text	Name of the organism.
isolate_num	double precision	The isolate number within the specimen (multiple isolate
quantity	double precision	Approximate quantity of organism grown.
ab_itemid	double precision	ID of the antibiotic tested against the organism.
ab_name	text	Name of the antibiotic tested.
dilution_text	text	Free-text describing antibiotic dilution.
dilution_comparison	text	Comparison operator for the dilution (e.g. “>”, “<”).
dilution_value	double precision	Measured dilution value.
interpretation	text	Interpretation of antibiotic sensitivity (S = sensitive, R = etc.).
comments	text	Additional deidentified text about the culture or sensitiv

- omr

Column Name	Data Type	Description
subject_id	bigint	**Foreign key** referencing *patients*.
chartdate	text	Date on which the observation was recorded.
seq_num	bigint	Monotonically increasing integer to distinguish multiple same day.
result_name	text	Description of the observation (e.g., “Blood Pressure,” “
result_value	text	The value associated with the observation (e.g., “120/80

- patients

Column Name	Data Type	Description
subject_id	bigint	**Primary key**. Unique patient identifier across the datase
gender	text	Documented gender.
anchor_age	bigint	Age at the de-identified anchor year (capped at 90+ as
anchor_year	bigint	Shifted year of the patient’s anchor date.
anchor_year_group	text	Range of years during which the anchor_year occurred
dod	text	De-identified date of death for out-of-hospital mortality hospital discharge).

- pharmacy

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
pharmacy_id	bigint	Primary key within the pharmacy table, used to link to p
poe_id	text	Reference to an order in poe.
starttime	text	Start time for the prescribed medication.
stoptime	text	Stop time for the prescribed medication.
medication	text	Name of the medication.
proc_type	text	Type of order (e.g., “IV Piggyback,” “Unit Dose”).
status	text	Whether the prescription is active, discontinued, etc.
entertime	text	Time at which the prescription was entered in the pharr
verifiedtime	text	Time at which the order was verified.
route	text	Route of administration (e.g., IV, Oral).
frequency	text	Frequency of administration (e.g., “Q6H,” “BID”).
disp_sched	text	Time(s) of day for scheduled administration (e.g., “08,
infusion_type	text	Code for type of infusion (e.g., ‘B’, ‘C’, ‘N’).
sliding_scale	text	Indicates whether a sliding scale is used (Y/N).
lockout_interval	double precision	Interval for patient-controlled analgesia (time between doses).
basal_rate	double precision	Basal continuous administration rate.
one_hr_max	double precision	Maximum one-hour dose for the medication.
doses_per_24_hrs	double precision	Number of doses expected per 24 hours.
duration	double precision	Numeric duration for this prescription.
duration_interval	text	Unit associated with the duration, e.g. “Doses,” “Weeks
expiration_value	double precision	Numeric length of time until medication expires.
expiration_unit	text	Unit associated with the expiration (days, hours, etc.).
expirationdate	double precision	The de-identified date of expiry.

dispensation	text	Source of dispensation.
fill_quantity	double precision	Quantity dispensed.

- poe

Column Name	Data Type	Description
poe_id	text	Primary key within poe. Uniquely identifies the provide poe_seq).
poe_seq	bigint	Monotonically increasing sequence for the order.
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
ordertime	text	Time at which the order was placed.
order_type	text	Broad classification of order (e.g., Lab, Medications, R
order_subtype	text	More specific subtype of the order (e.g., “Holter Monit
transaction_type	text	Action performed (e.g., “New,” “Change,” “D/C”).
discontinue_of_poe_id	text	If this order discontinues a previous order, references
discontinued_by_poe_id	text	If this order was discontinued by another future order,
order_provider_id	text	Provider who placed this order (de-identified).
order_status	text	Whether the order is active or has been discontinued.

- poe_detail

Column Name	Data Type	Description
poe_id	text	Foreign key referencing poe. Unique identifier for the
poe_seq	bigint	Sequence number for the order.
subject_id	bigint	Foreign key referencing patients.
field_name	text	The label/name of the attribute for this order detail (e.
field_value	text	The value of that attribute (e.g., “Psychiatry,” “Clear Li

- prescriptions

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
pharmacy_id	bigint	Foreign key referencing pharmacy.
poe_id	text	Links to poe.
poe_seq	double precision	Monotonically increasing sequence for the prescriptio
order_provider_id	text	Provider who placed the prescription order (de-identi
starttime	text	Time the prescription was started.
stoptime	text	Time the prescription was stopped.
drug_type	text	Indicates if the drug is a main component, base, or ad
drug	text	Name of the medication.
formulary_drug_cd	text	Hospital-specific code for the medication.
gsn	text	Generic Sequence Number (internal medication codin
ndc	double precision	National Drug Code identifier.
prod_strength	text	Strength/composition description (e.g., “12.5 mg Tabl
form_rx	text	Container format for the dose (e.g., “TABLET,” “VIAL”)
dose_val_rx	text	Prescribed dose value.
dose_unit_rx	text	Prescribed dose unit.
form_val_disp	text	Amount of medication contained in a single formulary
form_unit_disp	text	Unit for the formulary dosage.
doses_per_24_hrs	double precision	Number of doses expected in 24 hours.
route	text	Route of administration (IV, Oral, etc.).

- procedures_icd

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
seq_num	bigint	The priority of the procedure's billing code.
chartdate	text	Date associated with the billed procedure.
icd_code	text	ICD code representing the procedure.
icd_version	bigint	ICD version (9 or 10).

Keys and Links

- Joins to **d_icd_procedures** on (icd_code, icd_version).

- **provider**

Column Name	Data Type	Description
provider_id	text	Primary key. Unique de-identified provider ID used t

---

- **services**

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
transfertime	text	Time the service assignment changed from prev_se
prev_service	text	The previous service (if any).
curr_service	text	The current service. E.g., "SURG," "MED," "PSYCH,"

- **transfers**

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	double precision	Foreign key referencing admissions.
transfer_id	bigint	Unique transfer event identifier (often used to build
eventtype	text	Type of transfer event (e.g., 'admit,' 'transfer,' 'disch
careunit	text	The unit or ward to which the patient was transferre
intime	text	Date/time the patient entered this unit.
outtime	text	Date/time the patient left this unit.

---

icu Module

- caregiver

Column Name	Data Type	Description
caregiver_id	bigint	Primary key in the ICU module. De-identified identi

---

- chartevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	double precision	De-identified caregiver ID (reference to caregiver).
charttime	text	Time the observation was recorded.
storetime	text	Time the observation was validated/entered.
itemid	bigint	Foreign key referencing d_items for a measuremen
value	text	Recorded value in text form.
valuenum	double precision	Numeric version of the recorded value (if applicabl
valueuom	text	Units of measurement.
warning	double precision	Indicates if a warning was triggered for this charte

- datetimeevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	bigint	Reference to caregiver.
charttime	text	Time of the event’s measurement.
storetime	text	Time data was validated/entered into the system.
itemid	bigint	Foreign key referencing d_items (for date/time-typ
value	text	The recorded date/time value (e.g., “2023-05-01”)
valueuom	text	Typically “Date” for these entries.
warning	bigint	Indicates whether a warning was documented.

- d_items

Column Name	Data Type	Description
itemid	bigint	Primary key. Unique numeric identifier for each ICU
label	text	Human-readable label (e.g., “Heart Rate,” “Respira
abbreviation	text	Short abbreviation used in the ICU.
linksto	text	The name of the events table to which this itemid “inputevents”).
category	text	Category grouping (e.g., “Vital Signs,” “IV Medicat
unitname	text	Unit of measurement if consistent (e.g., “bpm,” “m
param_type	text	The type of parameter (numeric, text, date).
lownormalvalue	double precision	Recorded lower normal bound for numeric measur
highnormalvalue	double precision	Recorded upper normal bound for numeric measu

- icustays

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Primary key identifying a single ICU stay.
first_careunit	text	The first ICU care unit (e.g., MICU, SICU) the patie
last_careunit	text	The last ICU care unit the patient stayed in (can d
intime	text	Date/time the patient was transferred into the ICU
outtime	text	Date/time the patient was transferred out of the IC
los	double precision	Length of stay in the ICU (days).

- ingredientevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	bigint	Links to caregiver.
starttime	text	Start time for the infusion/ingredient administrati
endtime	text	End time for the infusion/ingredient.
storetime	text	Time this entry was documented.
itemid	bigint	Foreign key referencing d_items.
amount	double precision	The total amount of the ingredient administered
amountuom	text	Units of the ingredient amount.
rate	double precision	The rate at which the ingredient was administere
rateuom	text	Units of the rate (e.g., mL/hr).
orderid	bigint	Identifies multiple ingredients in the same solutio
linkorderid	bigint	Links repeated changes under the same base or
statusdescription	text	Status of the infusion (e.g., “Changed,” “Finished
originalamount	bigint	The initial planned total amount in the bag at star
originalrate	double precision	The initial planned rate.

- inputevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	bigint	Reference to caregiver.
starttime	text	When the input began.
endtime	text	When the input ended.
storetime	text	Documentation time.
itemid	bigint	Foreign key referencing d_items.
amount	double precision	Amount administered (e.g., volume in mL).
amountuom	text	Units of amount.
rate	double precision	Rate of administration (e.g., mL/hr).
rateuom	text	Units of rate.
orderid	bigint	Identifies multiple items (drugs/fluids) in the sam
linkorderid	bigint	Links repeated modifications of the same order.
ordercategoryname	text	Higher-level category for the order (e.g., “IV med
secondaryordercategoryname	text	Secondary category if available.
ordercomponenttypedescription	text	Description of the order component (additive, ba
ordercategorydescription	text	More verbose description of the category.
patientweight	double precision	Patient weight (kg) at the time of the order.
totalamount	double precision	Total volume in the bag.
totalamountuom	text	Units of total volume.
isopenbag	bigint	Indicates if the bag was open.
continueinnextdept	bigint	Whether the infusion continued upon transfer ou
statusdescription	text	Reason the infusion ended (e.g., “Changed,” “Sto
originalamount	double precision	The initial planned total amount at starttime.

originalrate	double precision	The initial planned rate at starttime.

- outputevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	bigint	Reference to caregiver.
charttime	text	Time the output event was documented.
storetime	text	Time data was finalized/entered.
itemid	bigint	Foreign key referencing d_items (e.g., type of flu
value	bigint	The volume of output.
valueuom	text	Unit for the output volume (e.g., “mL”).

- procedureevents

Column Name	Data Type	Description
subject_id	bigint	Foreign key referencing patients.
hadm_id	bigint	Foreign key referencing admissions.
stay_id	bigint	Foreign key referencing icustays.
caregiver_id	double precision	Reference to caregiver.
starttime	text	Time the procedure began.
endtime	text	Time the procedure ended.
storetime	text	Time the procedure was documented in the syst
itemid	bigint	Foreign key referencing d_items (the type of pro
value	double precision	Duration or other numeric measure of the proce long they lasted).
valueuom	text	Unit for the value (e.g., “min,” “hour”).
location	text	Specific physical location on the patient’s body i Arm”).
locationcategory	text	Higher-level category of the location (e.g., “Inva
orderid	bigint	Links multiple procedure items in the same orde
linkorderid	bigint	References the original order if repeated.
ordercategoryname	text	High-level name for the procedure order (e.g., “P
ordercategorydescription	text	More descriptive text for the order category.
patientweight	double precision	Patient weight at the time.
isopenbag	bigint	Indicates if the procedure was from an open bag
continueinnextdept	bigint	If the procedure continued upon transfer.
statusdescription	text	Indicates final status (e.g., “FinishedRunning,” “S
originalamount	double precision	Present but not clinically meaningful for many pr 1).
originalrate	bigint	Also present but not clinically meaningful (often

Queries

Basic Queries

Note (for this section): In case of a tie, output all the answers following the order. Use the strings (along with the case) given in the query for string comparison.

The format **"X days, HH:MM:SS"** is a general format for interval differences. If the duration is **24 hours or more**, it is displayed as **"X days, HH:MM:SS"**. If the duration is **less than 24 hours**, it is shown as **"HH:MM:SS"**

1. List the subject_id of all female patients with anchor age more than 89.

   output columns: subject_id

   order by columns: subject_id
2. List the top 5 years in which the most number of patients were admitted to hospitals.

   output columns: count, year

   order by columns: count(descending), year
3. List the hadm_id, gender, admission duration(**HH:MM:SS or** X days, HH:MM:SS) of all admissions which have non-null dischtime. Admission duration is defined as the difference between dischtime and admittime.

   output columns: hadm_id, gender, duration

   order by columns: duration, hadm_id
4. List the number of distinct medications ordered by each valid(non-null) provider.

   output columns: enter_provider_id, count

   order by columns: count(descending)
5. Find the number of distinct admissions where the reason for the barcode not being present is ‘Barcode Damaged’ and the marital status of the patient is not ‘MARRIED’.

   output: count
6. For each patient, count the number of times they have been admitted to ICU(including 0 number of stays).

   output columns: subject_id, count

   order by columns: count, subject id

7. For every patient whose ‘dod’ is not null, find their latest ‘hadm_id’.

   output columns: subject_id, latest_hadm_id, dod

   order by columns: subject_id

8. List all the ‘pharmacy_id’ which are present in the pharmacy but not written in any prescriptions.

   output columns: pharmacy_id

   order by columns: pharmacy_id

9. List all unique ‘icd_code’, ‘icd_version’ values present in both’ diagnoses_icd’ and ‘procedures_icd’ tables.

   output columns: icd_code, icd_version

   order by columns: icd_code, icd_version

10. List the unique ‘hcpcs_cd’ and a short description of all hcpcs events related to ‘Hospital observation’. You can assume that an event is related to x if its description contains x(case-insensitive).

    output columns: hcpcs_cd, short_description

    order by columns: hcpcs_cd, short_description

Intermediate Queries

Note (for this section): In case of a tie, output all the answers following the order. Use the strings (along with the case) given in the query for string comparison.

The format **"X days, HH:MM:SS"** is a general format for interval differences. If the duration is **24 hours or more**, it is displayed as **"X days, HH:MM:SS"**. If the duration is **less than 24 hours**, it is shown as **"HH:MM:SS"**

1. For each patient who has a non-NULL value in the ‘dod’ (date of death) column, write an SQL query to find the ‘hadm_id’ corresponding to their earliest ‘admittime’.

   output columns: subject_id, ~~latest_hadm_id~~ hadm_id, dod

   order by columns: subject_id

2. Find the average duration of admissions(**HH:MM:SS or** X days, HH:MM:SS) having diagnosis icd_code='4019' and icd_version='9'. Admission duration is defined as the difference between dischtime and admittime.

   output columns: avg_duration

3. For each caregiver_id, find the number of procedure events, chart events, and date-time events, documented by that caregiver.

   output columns: caregiver_id, procedureevents_count, chartevents_count, datetimeevents_count

   order by columns: caregiver_id, procedureevents_count, chartevents_count, datetimeevents_count

4. Find out if any patient has been diagnosed with an 'infection' in multiple admissions in the same year(on the basis of admission time). You can assume that the title of such a diagnosis will contain the word 'infection' (case insensitive).

   output columns: subject_id, count_admissions, year

   order by columns: year, count_admissions(descending), subject_id

5. List all the patients who had an admission type of 'URGENT' and died in the same admission(use hospital_expire_flag). Find how many ICD procedures were ordered and how many diagnoses were recorded for that patient in that admission.

   output columns: subject_id, hadm_id, count_procedures, count_diagnoses

   order by columns: subject_id, hadm_id, count_procedures(descending), count_diagnoses(descending)

6. Find the average BMI (kg/m2) of patients who were prescribed the medications 'OxyCODONE (Immediate Release)' and 'Insulin' both (case sensitive). Submit the results with exactly **10 places after the decimal**. [Hint - use hosp.omr table]

   output columns: avg_BMI

7. Find the patients who had the same diagnoses when they were admitted for the first time and most recently(use admittime). Find the percentage distribution of these patients on the basis of their gender. Round off to 2 decimal places.

   output columns: gender, percentage

   order by columns: percentage(descending)

8. Calculate the average admission duration of hospital stays (**HH:MM:SS or** X days, HH:MM:SS) for patients across all unique admissions. Admission duration is defined as the difference between dischtime and admittime.

   output columns: subject_id, avg_duration

   order by columns: subject_id

9. List the patients who were prescribed the same medication (same pharmacy_id) multiple times. List their 'subject_id' in the order(descending) of the number of times a medication was prescribed to the same patient. If the count is same, order by subject_id, pharmacy_id.

   output columns: subject_id, pharmacy_id

10. List the patients who were diagnosed with a kidney disease in their first admission and readmitted after this diagnoses. You can assume that the long title of such a diagnosis will contain the word 'kidney'(case insensitive). List the 100 patients who were admitted most recently(based on the 'admittime' of their first admission)

    output columns: subject_id

    order by columns: subject_id

Advanced Queries

Note (for this section): In case of a tie, output all the answers following the order.