Consider the mimiciii database, further described at <https://mimic.physionet.org/>.

For this task, the relevant tables and columns are:

# **ADMISSIONS table**

| **Column** | **Postgres data type** |
| --- | --- |
| SUBJECT\_ID | INT |
| HADM\_ID | INT |
| ADMITTIME | TIMESTAMP(0) |

# The ADMISSIONS table contains general information regarding each patient admission to the hospital. Each row of the ADMISSIONS table contains a unique HADM\_ID, which represents a single patient’s admission to the hospital. It is possible for this table to have duplicate SUBJECT\_ID, indicating that a single patient had multiple admissions to the hospital. The ADMISSIONS table can be linked to the PATIENTS table using SUBJECT\_ID. ADMITTIME provides the date and time the patient was admitted to the hospital.

# **PATIENTS table**

| **Column** | **Postgres data type** |
| --- | --- |
| SUBJECT\_ID | INT |
| GENDER | VARCHAR(5) |
| DOB | TIMESTAMP(0) |

# The PATIENTS table contains basic information about each patient. SUBJECT\_ID is a unique identifier which specifies an individual patient. SUBJECT\_ID is a candidate key for the table, so is unique for each row. GENDER is the genotypical sex of the patient. DOB is the date of birth of the given patient.

# **CHARTEVENTS table**

| **Name** | **Postgres data type** |
| --- | --- |
| SUBJECT\_ID | NUMBER(7,0) |
| HADM\_ID | NUMBER(7,0) |
| ITEMID | NUMBER(7,0) |
| CHARTTIME | TIMESTAMP(0) |
| VALUE | VARCHAR2(200 BYTE) |

The CHARTEVENTS table contains all the charted data available for a patient during their hospitalization. ITEMID is the identifier for a single measurement type in the database. Each row is associated with one ITEMID (e.g. ITEMID = 212) corresponds to an instance where that ITEMID was measured (e.g. heart rate). CHARTTIME records the time at which an observation was made. VALUE contains the value measured for the concept identified by the ITEMID.

For the tasks below, the relevant medical background and necessary ITEMIDs in the CHARTEVENTS table are as follows:

Invasive mechanical ventilation is a type of lung support that some patients receive during a hospitalization.

ITEMID = 226732, which has the label = "O2 Delivery Device(s)"

ITEMID 226732 corresponds to the type of lung support a patient received during hospitalization. Lung support is documented regularly during a hospitalization because it can change over a hospitalization. Values of this ITEMID include "Aerosol-cool", "High flow neb", "Nasal cannula", "Endotracheal tube", "Bipap mask", "None". The first time that "Endotracheal tube" is documented during a hospitalization indicates the start of an episode mechanical ventilation. The last time "Endotracheal tube" is documented prior to documentation of another type of lung support, e.g. "Nasal cannula" indicates the end of an episode of invasive mechanical ventilation.

ITEMID = 224684, which has the label "Tidal Volume (set)"

This ITEMID corresponds to the tidal volumes (size of breaths) that patients received while on invasive mechanical ventilation. Tidal volumes are documented multiple times over a course of a mechanical ventilation episode.

**Tasks**

Using SQL (and if needed, any other statistical or programming language) please write code to accomplish the following 3 tasks. Please document your code.

1. Create a dataset of hospitalization information for all adults (Age > 18) who received invasive mechanical ventilation. Columns should include HADM\_ID, SUBJECT\_ID, ADMITTIME, GENDER, Age. Age should be determined based on DOB and ADMITTIME. Assume that documentation of ITEMID = 226732 in the CHARTEVENTS table with value of "Endotracheal tube" indicates that a patient received invasive mechanical ventilation during the hospitalization.
2. For hospitalizations identified in #1, generate a dataset of all documented tidal volumes and O2 device recordings during the hospitalization. Columns should include HADM\_ID, SUBJECT\_ID, CHARTTIME, O2\_device, tidal\_volume. The O2\_device column should contain all values of ITEMID = 226732, with missing values at chart times when a tidal volume was documented but not the o2 device. The tidal\_volume column should contain all values of ITEMID = 224684, with missing values at chart times when an o2 device was documented but not a tidal volume.
3. For hospitalizations identified by #1, generate a dataset of the start time, end time, and total length of the episode of mechanical ventilation (or first episode if more than one during the hospitalization). Table columns should include HADM\_ID, SUBJECT\_ID, start\_time, end\_time, ventilation\_hours.