Part 1: Data Challenge

Time: 1.5 hours

Purpose

The purpose of this task is to get a feel for your understanding of how datasets are related to one another and to gauge your technical aptitude for organizing and shaping data.

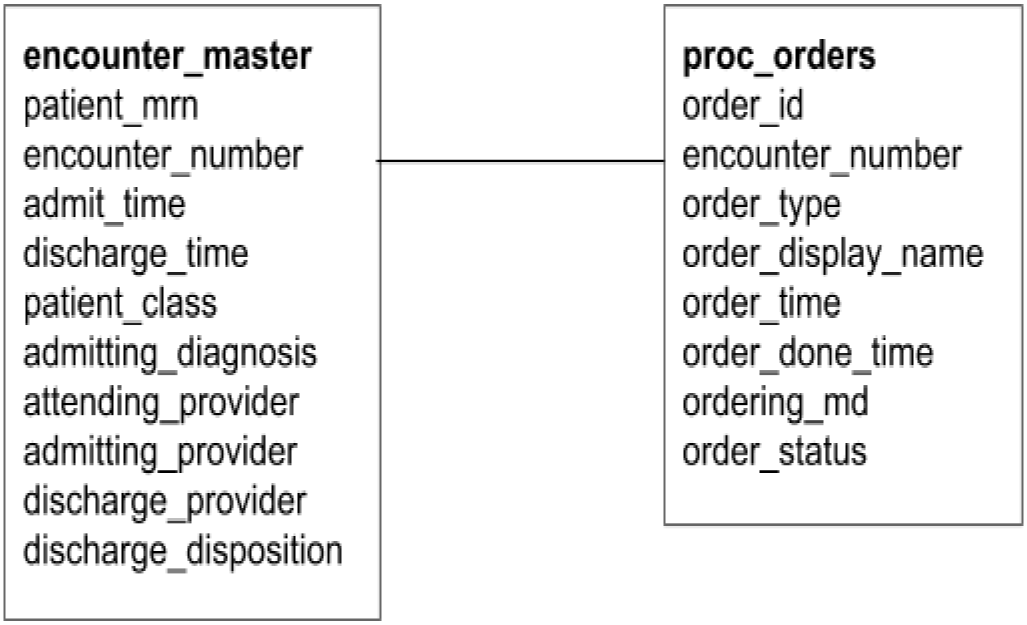
Instructions

Using the datasets and background provided below, please answer the questions below. Most people use SQL and/or Python, but feel free to use whatever's most comfortable.

Note that describing your approach and explaining any limitations or shortcomings is as important as the code.

One last note: please do not spend more than **1.5 hours** on this task. If you're not able to get to some questions, that is OK. We're more interested in seeing your approach. And you need to save some time for Part 2!

Background



encounter_master
patient_mrn
encounter_number
admit_time
discharge_time
patient_class
admitting_diagnosis
attending_provider
admitting_provider
discharge_provider
discharge_disposition

proc_orders
order_id
encounter_number
order_type
order_display_name
order_time
order_done_time
ordering_md
order_status

The data model above simulates two datasets we typically receive when ‘ramping up’ a new inpatient deployment. This type of deployment evaluate the workflows around patients who spend more than one night in the hospital. A hospital visit is usually defined as an “encounter” and a hospital visit involving an overnight stay at the hospital is designated as an inpatient encounter.

Each row on the encounter_master table represents a patient’s inpatient visit to the hospital.

- **patient_mrn -** A unique number representing the patient
- **encounter_number** - A unique number representing this specific visit
- **admit_time** - The date/time patient was “checked in” for their visit
- **discharge_time**- The date/time the patient was released from the hospital.
- **admit_class -** This is the patient’s classification when they first arrived at the hospital, as many patients arrive to be treated and not expecting to stay in the hospital overnight. See below for the standard classification of patients
    - *Inpatient - the patient will be treated in the hospital overnight*
    - *Emergency - the patient is initially treated in the Emergency Department*
    - *Observation - the patient will be placed under observation for precautionary purposes with the hope of releasing the patient in less than 24 hours*
- **admit_source -** This describes the patient’s point of origin when they are admitted as an inpatient. Examples of this value are “home”, “referral”, “transfer from another facility”
- **current_department -** This is the patient’s current location or unit in the hospital if they are still in house. If the patient is already discharged, this is their last location

- **admitting_department -** The first inpatient department that the patient is admitted to
- **admitting_diagnosis -** A code and freetext description of the reason that the patient arrived for their visit, typically described as a problem or disease, i.e. “Back Pain”, “Heart Attack”
- **attending_provider -** The most recent provider listed as responsible for treating the patient.
- **admitting_provider -** The provider responsible for approving the patient’s inpatient stay at the hospital
- **discharge_provder -** The provider provider responsible for approving the patient’s readiness to leave and ensuring a safe and appropriate departure from the hospital
- **discharge_disposition** - The physician’s determination of what should happen next to the patient. See below for explanations of the values
    - *home* - The patient was sent home
    - *IP Rehab* - The patient was sent to an acute rehabilitation facility
    - *SNF* - The patient was sent to a Skilled Nursing Facility
    - *expired* - The patient died during their visit

Each row on the proc_orders table represents an order or set of instructions directed by a provider for care or treatment of the patient by a specific department or care team.

- **order_id -** A unique identifier representing an order to treat the patient during the encounter
- **encounter_number -** A unique number representing this specific visit
- **order_type** - A category representing the type of order placed. Examples of order_type are “Lab”, “Radiology”, “Nursing”, “Physical Therapy”
- **order_display_name**
- **order_time** - the date/time that the provider placed the order
- **order_done_time** - the date/time that the order was completed
- **ordering_md** - the provider responsible for placing the order
- **order_status** - the current status of the order. See below for explanation of the values:
    - Ordered - order has been placed but no action has been taken on the order
    - In progress - order is currently being acted on
    - Completed - treatment/care instructions

Questions

Given the table schema above, please provide the logic/statement(s) that you would use to address the following questions:

1. What percentage of patient visits are still admitted or not discharged yet from the hospital?

2. Length of stay or LOS is a calculated metric that measures the total number of midnights between the admit_time and the discharge_time or current time, if the patient is still admitted. What is the Median LOS by admit_class for encounters admitted in the month of June?

3. We received a question from a client asking if LOS has gone down after Jan. 1 of this year, normalizing for things like seasonality and other factors (your choice). Can you describe how you would write a script / model to evaluate this? Please be as explicit as possible (some code would be great!).

Part 2: Storytelling Challenge

Time: 1.5 hours

Purpose

The purpose of this challenge is to explain how you would analyze whether or not a machine learning model is creating real-world impact.

Instructions

Using the background provided below, please outline an “analytics story” which would show as conclusively as possible that our model is “working” as expected (i.e., having a real-world impact on hospital operations). Please use data visualizations to help illustrate the story. Feel free to use slides, a document, hand drawings, etc.

Imagine that you’re going to present this analysis to a customer who’s pretty savvy about data, but not a machine learning engineer.

Also, please point out where you think your assumptions / analysis might be inconclusive, or where you’d have further questions.

Background

One of our models predicts when a “surge” is going to happen in an emergency department. It will “fire” if, in the next few hours, we predict that the emergency department census will exceed a critical threshold (the threshold can vary per hospital, depending on the emergency department’s size).

The model is a regression model, but it only “fires” if it predicts that census will cross the threshold (so in some sense, it acts as a classifier in the way the output is produced). Remember that the model may be wrong on occasion (either the census isn’t going to be as bad as we think, or the model doesn’t fire early enough, etc.).

Once the model “fires,” a number of people at the hospital (emergency department director, charge nurses, hospitalists, transporters, cleaning staff, etc.) will all be looped into a communication thread through our applications to help mitigate the bottleneck before it forms (think of a group chat). In its simplest form, the message from Qventus that kicks off the thread might be something like “High census expected in next 3 hours. ED, please reply with needs.”

Hospital staff might then respond to that nudge and/or take actions like expediting cleaning of rooms, working to get patients discharged faster or transported more quickly, etc.

Other pieces of context:

- We usually have a large amount of historical data that can be used to baseline.
- Think about how you'd show causality here. How can we get as close as possible to proving that our intervention is what's driving operational improvement?
- Remember that there may be multiple steps in the journey of proving impact here. For example: is the model predicting events correctly? Are users engaging with the output? Is an operational outcome being achieved? And more…
- Assume that we have all the data you'd need to tell this story (including all operational data from the hospital).