We are excited to invite you to participate in this take-home challenge that will allow you to showcase your skills and creativity in the field of healthcare analytics.

Problem Statement

A pharmaceutical company is looking to generate insights using data from routine clinical practice to better understand the treatment of major depressive disorder (MDD). MDD is typically treated with antidepressants and psychological therapy. Individuals who do not respond to an antidepressant therapy may be co-prescribed an adjunctive medication (e.g., antipsychotic, anticonvulsant, or lithium) alongside their antidepressant to help improve their symptoms.

You have access to a synthetic electronic health record (EHR) dataset (**attached in the email**) of approximately 3,700 patients who received mental healthcare for MDD in a major hospital system. The data dictionary can be found at the end of this document. Given the data at hand, suggest **one research question** which can help understand the treatment that patients with MDD receive in routine care.

Goals & Expectations

The goal of this challenge is to assess your capability in designing your own study given a problem statement and data at hand, consider the potential strengths and limitations of applying various analytical techniques, and assess your ability in "getting your hands dirty" with real-world data.

Two expected outcomes are:

1. Your code (written in Python and/or R) in a GitHub repository (recommended).
2. Your slide deck (PowerPoint recommended) summarizing your data, study design, approach, results, and key findings.

This challenge should take **up to two hours of effort**, including the time taken to develop the deliverables. As such, we expect the analysis to be limited in scope, but it is important for you to articulate what your analysis achieves and how this could be refined in further research.

You are expected to present your work at the next interview. The presentation should take **10 minutes**. Please allow at least 24 hours between submitting your code and presentation and the scheduled interview so our team can be well prepared for the interview.

m holmusk

Data Dictionary

The synthetic dataset provided contains four CSV files: (1) patient demographics, (2) clinical data, (3) billing identifiers, and (4) billing amount. All the information regarding the variables can be found below.

Table 1. Variables in the demographics table.

Field Name	Description
patient_id	Unique patient ID
gender	Patient's gender
race	Patient's race: Chinese/Indian/Malay/Others
resident_status	Patient's residential status: Citizen/Permanent resident/Foreigner
date_of_birth	Patient's date of birth

Table 2. Variables in the clinical data table.

Field Name	Description
id	Unique patient ID
date_of_admission	Date of an inpatient or emergency admission
date_of_discharge	Date of an inpatient or emergency discharge
medical_history_dia	Diagnosis of diabetes in the past 24 months (0=No, 1=Yes)
medical_history_sud	Diagnosis of substance use disorder in the past 24 months (0=No, 1=Yes)
medical_history_hbp	Diagnosis of high blood pressure in the past 24 months (0=No, 1=Yes)
medical_history_ren	Diagnosis of renal failure in the past 24 months (0=No, 1=Yes)
medical_history_tum	Diagnosis of solid tumour in the past 24 months (0=No, 1=Yes)
medical_history_anx	Diagnosis of anxiety disorder in the past 24 months (0=No, 1=Yes)
medical_history_mood	Diagnosis of other mood disorders in the past 24 months (0=No, 1=Yes)
trt_anx	Current treatment of anxiolytics (0=No, 1=Yes)
trt_con	Current treatment of anticonvulsants (0=No, 1=Yes)
trt_adt	Current treatment of antidepressants (0=No, 1=Yes)
trt_ssr	Current treatment of SSRI (0=No, 1=Yes)
trt_the	Current treatment of psychotherapy (0=No, 1=Yes)
trt_oth	Current treatment of other psychiatric medications (0=No, 1=Yes)
symptom_1	Current presence of abnormal sleep patterns (trouble falling asleep/sleeping too much) (0=No, 1=Yes)
symptom_2	Current presence of anhedonia (loss of interest/pleasure) (0=No, 1=Yes)
symptom_3	Current poor appetite (0=No, 1=Yes)
symptom_4	Currently feeling depressed or hopeless (0=No, 1=Yes)
symptom_5	Current presence of suicidal thoughts (0=No, 1=Yes)
weight	Current weight of the patient (kg)
height	Current height of the patient (cm)
cgis_adm	Clinical Global Impression Scale- Severity at the time of an inpatient or emergency admission (Score from 1-7, with 7 being most severe)
cgis_dis	Clinical Global Impression Scale- Severity at the time of an inpatient or emergency discharge (Score from 1-7, with 7 being most severe)
gaf_lv	Global Assessment of Functioning in the past 24 months (0=0, 1=1-10, 2=11-20, 3=21-30, 4=31-40, 5=41-50, 6=51-60, 7=61-70, 8=71-80, 9= 81-90 , 10=91-100)

m holmusk

Table 3. Variables in the billing identifiers table

Field Name	Description
bill_id	Billing ID
patient_id	Patient ID
date_of_admission	Date of an inpatient or emergency admission

Table 4. Variables in the billing amount table

Field Name	Description
bill_id	Billing ID
amount	Amount billed to the patient ($)

Abbreviations:

SSRI = selective serotonin reuptake inhibitor

holmusk