WhiteCoat - Data Analyst Assignment

Foreword

This assignment aims to assess a candidate's knowledge in handling normalised data and analytical skills, as well as writing clean and presentable code.

This test is confidential and not to be circulated, shared or redistributed in any format, in part or in their entirety. This includes the exercises and your answers, analysis and code.

Dataset

The dataset can be downloaded here: https://tinyurl.com/wc-download-data
The data dictionary is found below.

Assignment Questions

Question 1: SQL Queries

Based on the `consults`, `diagnoses` and `prescriptions` datasets, calculate:

a) Top 5 chronic diagnoses **overall**
b) Top 5 non-chronic (acute) diagnoses **by month**
c) Top 5 medications prescribed for each of the top 5 diagnoses in a)

Please submit the following:

1. A single Excel file, with each answer part in a separate sheet.
2. SQL queries used to extract answers
3. A short write-up detailing how you approached the problem and assumptions, if any.

Question 2: Data Wrangling & Aggregation

Use the `action_log` dataset to calculate **overall** and **by month**:

a) % dropoff at each stage of the booking funnel (see table for explanation)
b) Average time patients spent in each stage before successfully moving to the next stage

Order	Action	Description
1	Joined Queue	Patient filled up necessary details and system assigns a queue number
2	Doctor Ready	Patient reached front of queue and doctor is ready for consultation

3	Consult Start	Patient received notification that doctor is ready, and clicks to start consult
4	Consult End	Doctor completed patient assessment and clicks to end consult
5	Meds Delivered	Medication (if purchased) has been delivered to patient's location

Please submit the following:

1. A single Excel file, with each answer part in a separate sheet.
2. Source code / documentation.
3. A short write-up detailing how you approached the problem and assumptions, if any.

Question 3: Analysis and Visualisation

The Marketing team is looking for personalisation strategies to provide our existing user base with engaging EDM content. Using the datasets given, identify **3 (three)** interesting insights and / or user segments, including your recommendation(s) for potential targeting strategies. Feel free to highlight any additional data required that you think would strengthen your analysis.

Please submit the following:

1. Report of your analysis.
2. Source code / documentation, for other analysts to reproduce your work independently.
3. A visualisation tool / script for your recommendations, and explain how this tool will be maintained and updated.

Data Dictionary

Additional Information

- All data is limited to General Practitioner (GP) services in Singapore only.
- All timestamps & dates are in Singapore Standard Time (UTC +0800).
- Columns with the same names can be used as join keys.

Table: `consults`

List of completed WhiteCoat consultations in a given period of time.

Column Name	Data Type	Description
consult_date	Date	Date of consultation
consult_id	String	Unique identifier for consults
doctor_id	String	Unique identifier for doctors
member_id	String	Unique identifier for patients
is_under_insurance	Boolean	Whether consult was covered under an insurance policy
mc_days	Integer	Number of Medical Certification (MC) days prescribed by doctor If 0, assume no MC was issued for the consult
is_referral_letter_issued	Boolean	Whether a referral letter for specialist was issued
delivery_type	Integer	Medication delivery method selected by patient. • 0 = no delivery • 1 = self-collection • 2 = standard delivery (within 6 hours) • 3 = express delivery (within 2 hours)
consult_revenue	Float	Total cost of consultation including medication, standardised across entire dataset (calculation method)

Table: `diagnoses`

Doctor's diagnosis of patient for each consult.

Column Name	Data Type	Description
consult_id	String	Unique identifier for consultations
diagnosis_code	String	ICD-10 code of diagnosed illness
is_chronic	Boolean	Whether the given diagnosis is a chronic illness, broadly defined as "conditions that last 1 year or more and require ongoing medical attention or limit activities of daily living or both." (CDC)

Table: `patients`

Demographic information of patients who successfully consulted a WhiteCoat GP.

Column Name	Data Type	Description
member_id	String	Unique identifier for patients
member_dob	String	Patient date of birth
gender	String	Patient gender

Table: `prescriptions`

Medication(s) prescribed by doctor for a given consult.

Column Name	Data Type	Description
consult_id	String	Unique identifier for consults
medication_id	String	Unique identifier for medication
prescription_quantity	Integer	Amount prescribed by doctor Note: Assume units **within the same medication** are consistent, but units **across** medications might not be comparable.
is_med_bought	Boolean	Whether patient decided to purchase the prescribed medication

JSON: `action_log`

Newline-delimited JSON files for patients' behaviour in queue.

Field Name	Data Type	Description
consult_id	String	Unique identifier for consults
member_id	String	Unique identifier for patient
doctor_id	String	Unique identifier for doctors
action	String	All possible values can be found in Question 2
timestamp	Timestamp	Timestamp of when action was taken