

Analyst Assessment

Cobbs Creek Healthcare, LLC

Requirement

There are two sections in this assessment: SQL and Python. If you are stuck on a single question, feel free to skip to the next one, just try to complete as many as possible. The final submission should contain two files, one `.sql` and one `.py` or `.ipynb`, non-coding question answers can be put into comment block. Please pack them into a `.zip` archive and send it back to us.

SQL

*[SQLiteStudio](#) might be helpful if you don't have existing database environment. Please also avoid using `CREATE TABLE/VIEW` in your query.

Q1

Please describe your approach to understand and solve the problem when facing this error message.

```
001030(42010): syntax error line 3 at position 14 unexpected '('
```

Q2

Data: `sales_example.csv` `sales_mapping.csv`

We have Drug D sales data for indication A, B and C from the past two years. We want to know about the Drug's performance(performance is evaluated by total prescriptions - TRx, the higher the better) for each specialty distributions(SDs).

Please create an output with 2021 quarterly indication A performance trend for all available SDs. The output should have 3 columns: SD name, quarter(e.g. 2024Q1) and total sales number. For all records cannot be linked to SD, put them into `NO_CONTRACT_SD` group.

Q3

Data: `dx_example.csv`

When patient gets a diagnosis from physician, a diagnosis code will be provided to identify what disease, disorders, symptoms or other reasons for patient encounters. We want to define a patient cohort for certain therapeutical area based on the diagnosis they get. The 3 groups of interested diagnosis code are:

- Group A: C83.0, C83.00, C83.01, C83.02, C83.03

- Group B: C91, C91.1, C91.10, C91.11, C91.12
- Group C: C95.10, C95.90

Please write a query with window function to return the 2nd distinct diagnosis (i.e. only patient with more than 1 diagnosis code will be included) of each patient in Group A by service date.

Q4

Data: `dx_example.csv`

How many patients have diagnosis from two different groups within the same calendar month? (Hint: not a lot)

Python

- Please make sure all the necessary dependencies are imported.

Q1

Explain what's the purpose of `if __name__ == __main__` section in a python script.

Q2

Given two integer array `observed` and `predicted`, write a single python function returns the variance of `observed` and `predicted`, and R square. Please do not use any package rather than built-in functions, i.e. no import.

Q3

Data: `dx_example.csv`

Following `SQL Q3`, how many patients have diagnosis from two different groups within the same month? (Hint: not a lot)

Q4

Data: `dx_example.csv`

A patient need have at least one diagnosis from the 3 groups in `SQL Q3` to enter the cohort. Within the patient cohort, how many patients have consecutive diagnosis for more than 3 calendar months since their first diagnosis?