Exploring the relationship between age and health.

WGU | D205 Data Acquisition

Performance AssessMent

Narcisse, Laurie

2023

1. **Provide a question that can be answered using structured query language (SQL) to acquire data from both the original database and the add-on CSV file data. The question should require data from both data sources.**

* Are patients who have arthritis more likely to be over 40, and how many of them have had high blood pressure or a stroke?

**A1. Identify which data from the original data set and the add-on CSV file are needed to answer the question including all tables, columns, and data types.**

* The CSV file that I used is mservices.csv

|  |  |  |  |
| --- | --- | --- | --- |
| **Patient** | **Data Type** | **Mservices.csv** | **Data Type** |
| Patient\_id | Varchar | Patient\_id | Varchar |
| Age | Int | arthritis | Varchar |
| gender | Varchar |  |  |
| Highblood pressure | Varchar |  |  |
| stroke | Varchar |  |  |

1. **Create an entity relationship diagram (ERD) for the add-on CSV file and any other tables and columns used to answer the question from part A by evaluating the data contained in the file and identifying the m:n relationships and relational constraints.**

**A close-up of a table

Description automatically generated**

**B1. Write SQL code, in text format, that creates a table based on the ERD and specifies the columns and relevant keys.**

CREATE TABLE public.mservices

(

patient\_id character varying(50) NOT NULL,

services character varying(100),

overweight character varying(10),

"arthritis" character varying(10),

"Diabetes" character varying(10),

"Hyperlipidemia" character varying(10),

"BackPain" character varying(10),

"Anxiety" character varying(10),

"Allergic\_rhinitis" character varying(15),

"Reflux\_esophagitis" character varying(15),

"Asthma" character varying(10),

      PRIMARY KEY (patient\_id)

);

ALTER TABLE public.mservices

      ADD FOREIGN KEY (patient\_id)

      REFERENCES patient(patient\_id);

ALTER TABLE public.mservices

OWNER to postgres;

**B2. Write SQL code, in text format, that loads the data from one of the add-on CSV files into the table created in part B1.**

* COPY public.mservices (patient\_id, services, overweight, "arthritis", "Diabetes", "Hyperlipidemia", "BackPain", "Anxiety", "Allergic\_rhinitis", "Reflux\_esophagitis", "Asthma")

FROM '/path/to/your/csv/file.csv' CSV HEADER;

1. **Write a SQL statement or statements in text format for a query or queries that answer the question from part A.**

* select pat.patient\_id, pat.age, pat.stroke, pat.hignblood, ms.arthritis

from patient pat

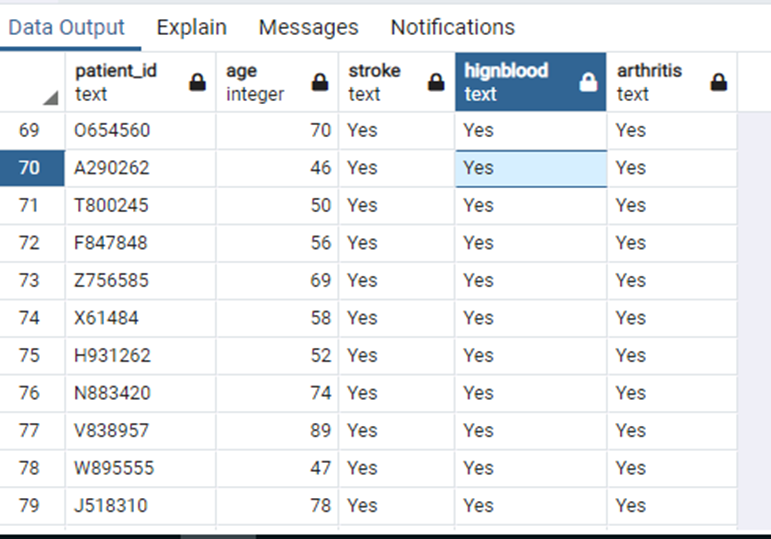
left join public.mservices ms on pat.patient\_id = ms.patient\_id

where ms.arthritis like 'Yes%'

and pat.age>40

and pat.stroke like 'Yes%'

and pat.hignblood like 'Yes%';

* ****

**C1. Provide a data file or files that capture the results from the query or queries.**

* Please see file named: C1 Results

1. **Identify the specific time period for how often the add-on file should be acquired and refreshed in the database for the data to remain relevant to the business and the question from part A.**

* I would advise that the data needs to be updated and refreshed at least once a month to keep it pertinent for the business’s needs in answering the question about patients with arthritis.

**D1. Explain why the time period identified in part D is relevant to the business needs.**

* The frequency complies with customary healthcare decision cycles, reporting standards, and operational cost management while providing up-to-date patient information. However, the exact update rate should be regularly reassessed to ensure it continues to meet evolving business needs and data dynamics.

1. **Provide a Panopto video recording that includes the presenter and a vocalized demonstration showing all code used, the code being executed, and the results of all code used in the task.**

* A link and an upload has been provided.

1. **Acknowledge web sources used to acquire data or segments of third-party code to support the application. Be sure the web sources are reliable.**

* SQL tutorial. (n.d.). <https://www.w3schools.com/sql/default.asp>
* Lucidcharts. Lucid visual collaboration suite: Log in. (n.d.). [https://lucid.app/lucidchart/5c59bf0f-0675-4846-9d94-389cc82a5db4/edit?page=0\_0#](https://lucid.app/lucidchart/5c59bf0f-0675-4846-9d94-389cc82a5db4/edit?page=0_0)