A: Research Question

For this project, my research question is, “How many patients are diabetic by age group?” This question is relevant to business because it helps identify trends across different age groups and informs the distribution of necessary resources, such as specialized care from endocrinologists for the most affected groups. By analyzing the number of diabetic patients by age group, healthcare providers can create tailored treatment plans and educational materials. This targeted approach can lead to improved patient outcomes, such as reduced diabetes-related complications and a more effective healthcare delivery system that ensures allocated resources meet the specific needs of diabetic patients in each age group.

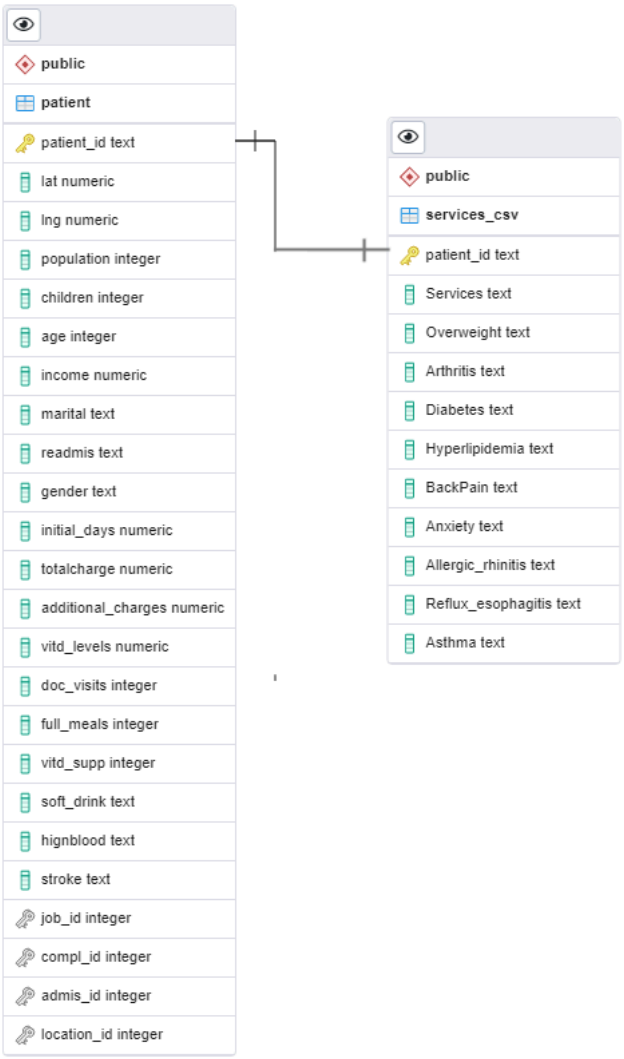
A1: Question Justification

There are two data sources required to answer the research question: the “patient” table, located in the original “medical-data” database, and the add-on CSV file “mservices.csv.” The “patient” table contains a unique identifier for each patient, “patient_id,” as well as an “age” column that would be used to group patients by age. The “mservices.csv” file also has a “patient_id” column and a “Diabetes” column that indicates whether a patient has been diagnosed with diabetes. Using these two data sources, we can determine the number of diabetic patients in an age group.

A2: Identifying Data

We need the “patient_id” and “age” columns from the “patient” table. From the table created from the CSV file, named “services_csv,” we require the “patient_id” and “Diabetes” fields. The “patient_id” field will join the patient table and the add-on CSV table. The “patient_id” and “Diabetes” fields use the text data type, and the “age” field uses the integer data type.

B1: ERD



The pgAdmin4 ERD beta creation tool generated the above ERD. The “patient_id” is the primary key in both the “patient” and “services_csv” tables, making each “patient_id” unique in both tables. Therefore, this creates a one-to-one relationship because each patient can only have one corresponding record in the “services_csv” table.

B2: SQL Code For Add-On CSV

```
CREATE TABLE public.services_csv (
        patient_id text NOT NULL,
        "Services" text NOT NULL,
        "Overweight" text NOT NULL,
        "Arthritis" text NOT NULL,
        "Diabetes" text NOT NULL,
        "Hyperlipidemia" text NOT NULL,
        "BackPain" text NOT NULL,
        "Anxiety" text NOT NULL,
        "Allergic_rhinitis" text NOT NULL,
        "Reflux_esophagitis" text NOT NULL,
        "Asthma" text NOT NULL,
        PRIMARY KEY (patient_id),
        CONSTRAINT patient_id_fkey FOREIGN KEY (patient_id)
        REFERENCES public.patient (patient_id) MATCH SIMPLE
        ON UPDATE NO ACTION
        ON DELETE NO ACTION
        NOT VALID
);


ALTER TABLE public.services_csv
        OWNER to postgres;
```

B3: SQL Code For Loading CSV File Data

```
COPY services_csv

FROM 'C:\LabFiles\Medical\mservices.csv'

DELIMITER ','

CSV HEADER;
```

C: Research Question SQL Query

```
SELECT
    CASE
        WHEN patient.age BETWEEN 0 AND 17 THEN '0-17'
        WHEN patient.age BETWEEN 18 AND 24 THEN '18-24'
        WHEN patient.age BETWEEN 25 AND 34 THEN '25-34'
        WHEN patient.age BETWEEN 35 AND 44 THEN '35-44'
        WHEN patient.age BETWEEN 45 AND 54 THEN '45-54'
        WHEN patient.age BETWEEN 55 AND 64 THEN '55-64'
        ELSE '65+'
    END AS age_groups,
    COUNT(patient.patient_id) AS diabetic_patients
FROM
    patient
INNER JOIN
    services_csv
ON
    patient.patient_id = services_csv.patient_id
WHERE
    services_csv."Diabetes" = 'Yes'
GROUP BY
    age_groups
ORDER BY
```

age_groups;

C1: Research Query Output

Data Output | Explain | Messages | Notifications

	age_groups text	diabetic_patients bigint
1	18-24	259
2	25-34	409
3	35-44	379
4	45-54	359
5	55-64	365
6	65+	967

D: Add-On CSV File Refresh Interval

To monitor and manage the number of diabetic patients by age group, the add-on CSV should be refreshed daily. This is important because it ensures the data stays current and accurately reflects the most recent patient information and diagnoses. Daily updates are essential for studying trends, allowing healthcare professionals to identify shifts in diabetes prevalence across different age groups. With this information, healthcare resources, such as endocrinologists, educational materials, and targeted treatment plans, can be allocated more efficiently to meet the needs of specific age groups experiencing an increase in diabetes cases.

E: Panopto Video

The Panopto video is in the D205 Student Creators folder under “D205 PA (011039266).”

F: Web Sources

For this project, I referred to Jain’s article on the SQL CASE statement from GeeksforGeeks, which guided in incorporating conditional logic to categorize diabetic patients by age groups (Jain, 2024). Additionally, PostgreSQL Tutorial’s article on importing CSV files into a table was crucial for understanding how to transfer data from the “mservices.csv” file to the “services_csv” table (Import CSV File Into PostgreSQL Table, n.d.).

References

Import CSV File Into PostgreSQL Table. (n.d.). PostgreSQL Tutorial. Retrieved June 19, 2024, from https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/

Jain, S. (2024, May 14). *SQL CASE Statement*. GeeksforGeeks. Retrieved June 19, 2024, from https://www.geeksforgeeks.org/sql-case-statement/