Healthcare Analytics Assignment -   
Part 1: Data Pipeline Documentation

# 1. Setup Overview

The pipeline was built using Docker, Python, and PostgreSQL. SQL transformations were performed using DBeaver.

## Approach A: Docker + PowerShell

- Cloned the repo and navigated using PowerShell  
- Installed Docker, Python, and dependencies  
- Started the PostgreSQL container using `docker-compose up -d`  
- Loaded the dataset using `python load\_data.py`

## Approach B: DBeaver (GUI)

- Connected to PostgreSQL using DBeaver GUI  
- Setup connection using Docker host settings:  
 Host: localhost  
 Port: 5432  
 Username/Password: postgres  
- Verified that tables were loaded properly  
- Performed transformations using SQL Editor

# 2. Data Quality Checks

- Handled missing ages in patients by assigning 'Not Available' group  
- Ensured appointment dates were valid and calculated time gaps between appointments  
- Standardized prescription categories by medication name  
- Verified provider IDs for consistency

# 3. Design Decision: Non-Destructive Transformations

Instead of modifying the original tables using ALTER TABLE, transformed versions of the tables were created. This modular approach supports clean rollbacks and aligns with best practices in production systems.

# 4. Data Transformations

## a. patients\_transformed

Description: Added age\_group and patient\_type for better segmentation.

SQL Code:

CREATE TABLE patients\_transformed AS  
SELECT \* ,  
CASE  
 WHEN age BETWEEN 0 AND 18 THEN '0-18'  
 WHEN age BETWEEN 19 AND 30 THEN '19-30'  
 WHEN age BETWEEN 31 AND 50 THEN '31-50'  
 WHEN age BETWEEN 51 AND 70 THEN '51-70'  
 WHEN age IS NULL THEN 'Not Available'  
 ELSE '70+'  
END AS age\_group,  
CASE  
 WHEN registration\_date::date >= NOW()::date - INTERVAL '6 months' THEN 'New'  
 WHEN registration\_date::date >= NOW()::date - INTERVAL '24 months' THEN 'Regular'  
 ELSE 'Long-term'  
END AS patient\_type  
FROM patients;

* Issue handled: Missing ages were tagged as 'Not Available'.

## b. appointments\_transformed

Description: Added day\_of\_week and time gap from last appointment.

SQL Code:

CREATE TABLE appointments\_transformed AS  
SELECT \* ,  
to\_char(appointment\_date::date,'Day') AS day\_of\_week,  
(appointment\_date::date - LAG(appointment\_date::date)  
 OVER(PARTITION BY patient\_id ORDER BY appointment\_date::date)) AS days\_since\_last\_appointment  
FROM appointments;

* Issue handled: Handled NULLs using LAG for first appointments.

## c. prescriptions\_transformed

Description: Standardized medication categories and calculated prescription frequency.

SQL Code:

CREATE TABLE prescriptions\_transformed AS  
SELECT \* ,  
CASE  
 WHEN medication\_name = 'Ibuprofen' THEN 'Pain Relief'  
 WHEN medication\_name = 'Metformin' THEN 'Diabetes'  
 WHEN medication\_name IN ('Lisinopril', 'Aspirin', 'Atorvastatin') THEN 'Heart'  
 WHEN medication\_name = 'Amoxicillin' THEN 'Antibiotic'  
 ELSE 'Others'  
END AS medication\_category,  
CASE  
 WHEN ROW\_NUMBER() OVER(PARTITION BY patient\_id ORDER BY prescription\_date::date) = 1 THEN 'First-time'  
 ELSE 'Repeat'  
END AS prescription\_frequency  
FROM prescriptions;

* Issue handled: Cleaned and mapped medication names for consistent grouping.

Analysis Results

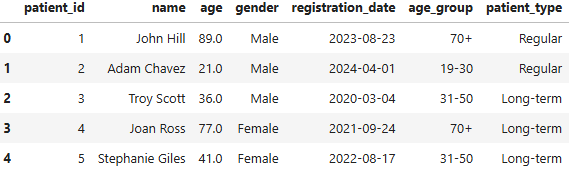
## Question 1 (a)

What is the distribution of patients across age groups?

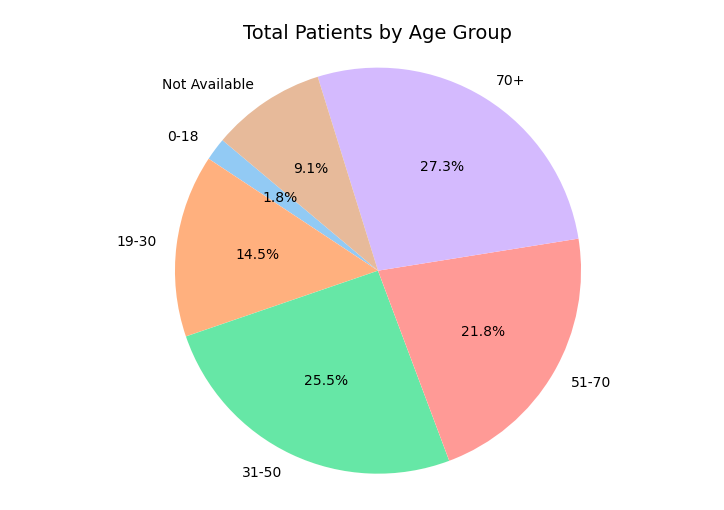
### SQL Code

SELECT \* FROM patients\_transformed;

### Result Table Screenshot (Partial)



### Chart Screenshot



### Interpretation:

The distribution helps identify which age groups are more actively engaging with the healthcare system. This insight can guide age-specific outreach or resource planning.

## Question 1 (b)

How does the appointment frequency vary by patient type?

### SQL Code

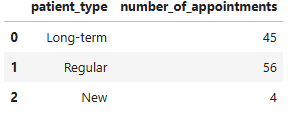
SELECT T1.patient\_type, COUNT(T2.appointment\_id) AS number\_of\_appointments

FROM patients\_transformed AS T1

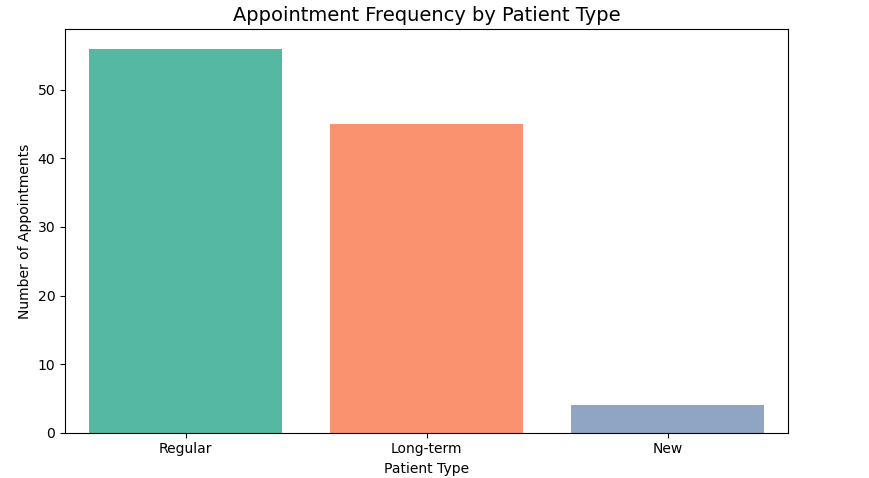
JOIN appointments\_transformed AS T2 ON T1.patient\_id = T2.patient\_id

GROUP BY 1;

### Result Table Screenshot



### Chart Screenshot



### Interpretation:

New or long-term patients may have different appointment frequencies, which can inform strategies to improve engagement or retention.

## Question 2 (a)

What are the most common appointment types by age group?

### SQL Code

SELECT T1.age\_group, T2.appointment\_type, COUNT(\*) AS appointment\_count

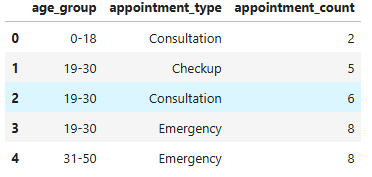
FROM patients\_transformed AS T1

JOIN appointments\_transformed AS T2 ON T1.patient\_id = T2.patient\_id

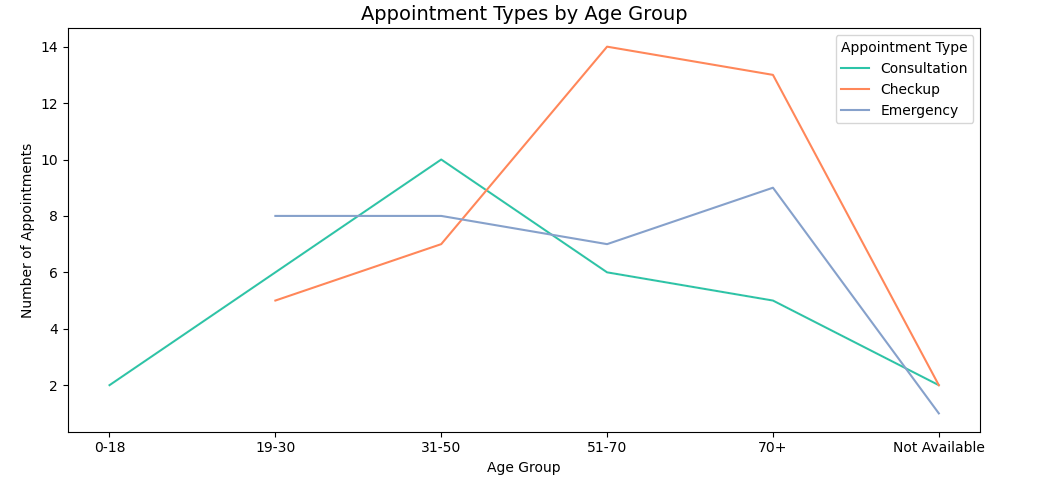
GROUP BY 1,2

ORDER BY 1 ;

### Result Table Screenshot



### Chart Screenshot



### Interpretation:

Understanding popular appointment types by age group allows providers to tailor services for specific demographics, enhancing patient satisfaction and care efficiency.

## Question 2 (b)

Are there specific days of the week with higher emergency visits?

### SQL Code

SELECT day\_of\_week , COUNT(\*) AS appointment\_count

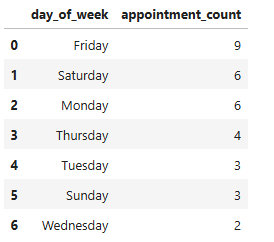
FROM appointments\_transformed

WHERE appointment\_type = 'Emergency'

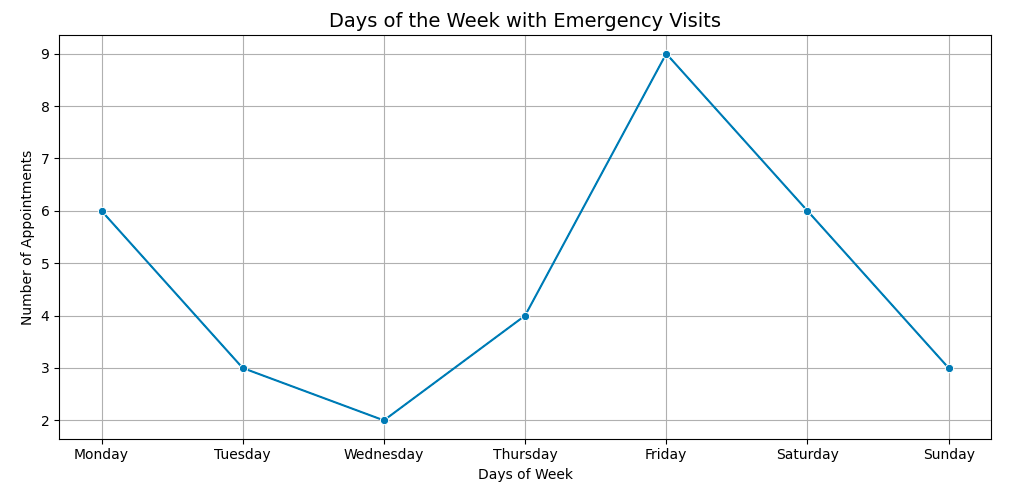
GROUP BY 1

ORDER BY 2 DESC ;

### Result Table Screenshot



### Chart Screenshot



### Interpretation:

Analyzing emergency visit patterns by weekday helps identify operational stress points, enabling better scheduling and resource allocation.

## Question 3 (a)

What are the most prescribed medication categories by age group?

### SQL Code

SELECT T2.age\_group, T1.medication\_category, COUNT(\*) AS total\_patient

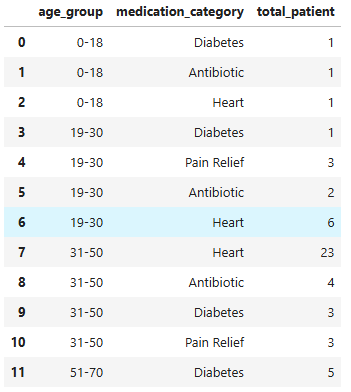
FROM prescriptions\_transformed AS T1

JOIN patients\_transformed AS T2 ON T1.patient\_id = T2.patient\_id

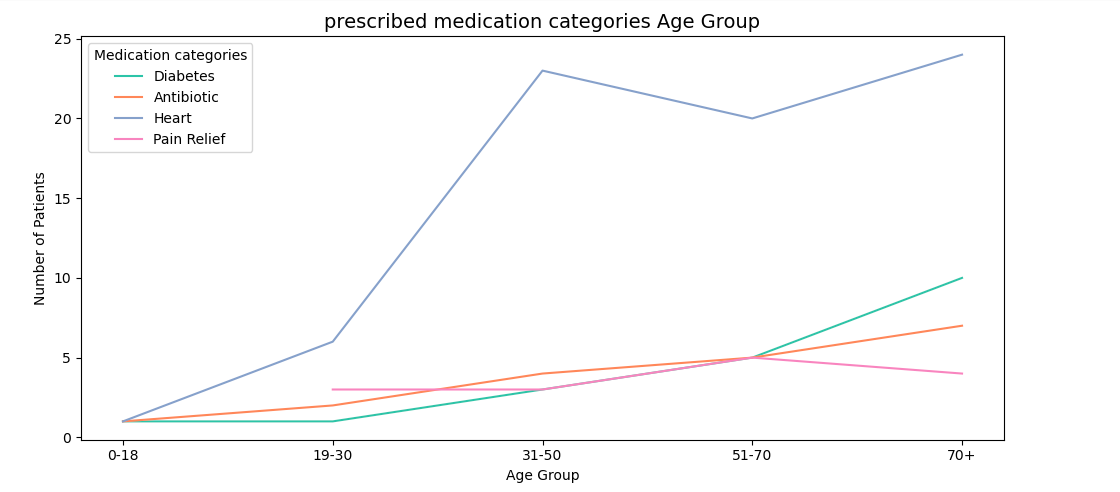
GROUP BY 1,2

ORDER BY 1 ;

### Result Table Screenshot (Partial)



### Chart Screenshot



### Interpretation:

Identifying medication categories prescribed most by age group assists in inventory planning, preventive care initiatives, and chronic disease management.

## Question 3 (b)

How does prescription frequency correlate with appointment frequency?

### SQL Code

WITH appointment\_counts AS (

SELECT

patient\_id,

COUNT(\*) AS appointment\_count

FROM appointments\_transformed

GROUP BY patient\_id

),

prescription\_counts AS (

SELECT

patient\_id,

COUNT(\*) AS prescription\_count

FROM prescriptions\_transformed

GROUP BY patient\_id

)

SELECT

a.patient\_id,

a.appointment\_count,

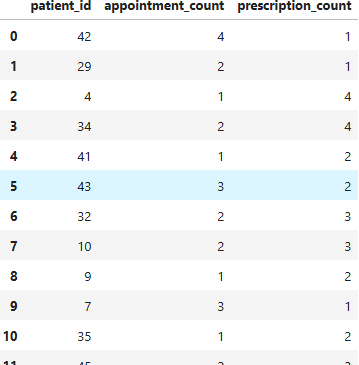
p.prescription\_count

FROM appointment\_counts a

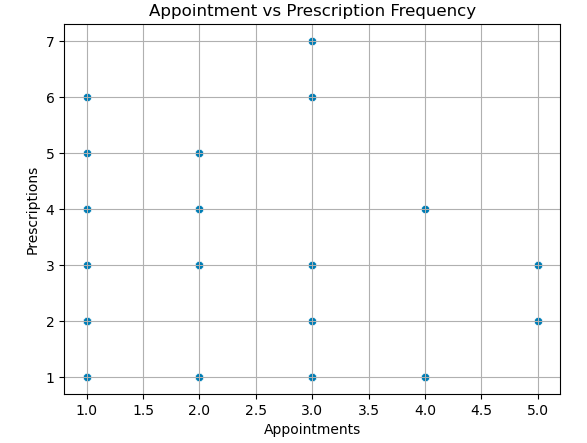
JOIN prescription\_counts p ON a.patient\_id = p.patient\_id

;

### Result Table Screenshot (Partial)



### Chart Screenshot





### Interpretation:

A correlation between appointment and prescription frequencies may indicate whether higher patient engagement results in more active treatment or monitoring.

Business Insights

## Insight 1

Finding: The majority of patients fall in the **70+** age group, indicating high engagement from senior patients.

Suggestion: Enhance preventive and wellness programs tailored for seniors (70+ age group).

## Insight 2

Finding: Emergency visits peak on **Friday** with 9 visits.

Suggestion: Bolster staffing and introduce end‑of‑week tele‑triage support to smooth Friday surges.

## Insight 3

Finding: **Regular** patients (6–24 months since registration) have the highest appointment count (56), ahead of long‑term (45) and new patients (4).

Suggestion: Investigate which touchpoints drive strong engagement in this group and replicate them for new and long‑term cohorts.

## Insight 4

Finding: **Heart** medications are the top prescription category for older age groups (24 for 70+, 20 for 51–70).

Suggestion: Review cardiovascular drug inventory and patient education to ensure adherence and supply continuity.

## Insight 5

Finding: Patients with **more appointments** also receive **more prescriptions**, suggesting comprehensive care but raising overtreatment questions.

Suggestion: Perform an outcome analysis to confirm whether higher visit‑to‑prescription ratios improve health metrics or signal unnecessary prescribing.

## Insight 6

Finding: Among the 51+ cohort, **Checkup** is the most common appointment type, reflecting a preventive‑care focus

Suggestion: Expand proactive screening and chronic‑disease management initiatives to leverage this preventive mindset.