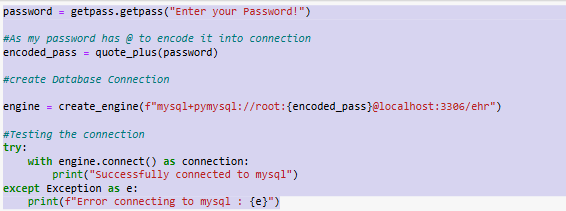
**DOCUMENTATION**

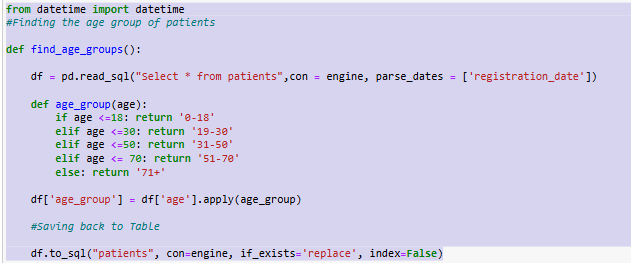
**Transformation Approach**The transformation phase was designed to enrich raw data from four core tables—**patients, appointments, prescriptions, and providers**—to enable meaningful analysis and insights. The transformations were executed using **Python (Pandas)** after ingestion into a **MySQL** database.

****

****

**Patient-Level Transformations**

1. **Age Group Classification**  
   Patients were segmented into 5 age groups for demographic analysis:
   * 0-18, 19-30, 31-50, 51-70, and 71+  
     The transformation was applied using a custom function mapping the patient’s age into these ranges.



**Patient Type**  
Based on the time since registration:

**New**: < 6 months

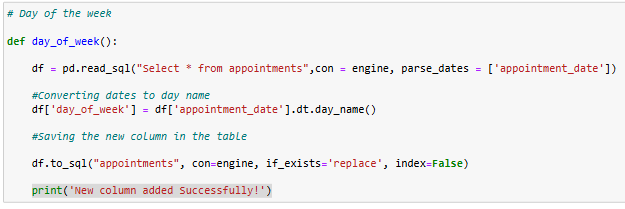
**Regular**: 6–24 months

**Long**-**term**: > 24 months  
 This was calculated by subtracting registration\_date from the current date and converting to months.



**Appointment-Level Transformations**

1. **Day of the Week**  
   Extracted from appointment\_date using dt.day\_name() to analyze visit patterns across the week.



1. **Time Since Last Appointment**  
   Calculated per patient by sorting appointments chronologically and subtracting the days from the previous appointment date to get the number of days.



**Prescription-Level Transformations**

1. **Medication Category**  
   Mapped common medication names to categories like Pain Relief, Diabetes, Heart, etc., using a dictionary.



1. **Prescription Frequency**  
   Classified as First-time or Repeat based on whether a patient had received the same medication before. Used duplicated() to identify repeat prescriptions.

**Analysis**

* **What is the distribution of patients across age groups?**

SELECT age\_group, round(count(patient\_id)\*100/(select count(\*) from patients),2) as number\_of\_patients from patients

group by age\_group

order by age\_group;

**RESULT**

| **Age Group** | **Number of Patients (%)** |
| --- | --- |

|  |  |
| --- | --- |
| 0-18 | 1.82% |

|  |  |
| --- | --- |
| 19-30 | 14.55% |

|  |  |
| --- | --- |
| 31-50 | 25.45% |

|  |  |
| --- | --- |
| 51-70 | 21.82% |

|  |  |
| --- | --- |
| 71+ | 36.36% |

**Patients above 50 make approx 50% of the patients**

* **How does the appointment frequency vary by patient type?**

SELECT p.registration\_type, COUNT(a.appointment\_id) AS appointment\_count

FROM appointments a

JOIN patients p ON a.patient\_id = p.patient\_id

GROUP BY p.registration\_type

ORDER BY p.registration\_type;

**RESULT**

| **Registration Type** | **Appointment Count** |
| --- | --- |
| Long Term | 39 |
| New | 4 |
| Regular | 62 |
|  |  |

**62 Patients were registered between 6-24 months from current date**

* **What are the most common appointment types by age group?**

with cte as (

select p.age\_group, a.appointment\_type, count(a.appointment\_type) as count,

dense\_rank() over (partition by age\_group order by count(a.appointment\_type) desc) as rnk from appointments a join patients p

on a.patient\_id = p.patient\_id

group by p.age\_group, a.appointment\_type

order by p.age\_group)

select age\_group, appointment\_type, count from cte where rnk = 1;

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

select day\_of\_week, count(patient\_id) as number\_of\_visits from appointments

where appointment\_type = 'Emergency'

group by day\_of\_week

order by count(patient\_id) desc;

**day\_of\_week number\_of\_visits**

**Friday 9  
 Saturday 6  
 Monday 6  
 Thursday 4  
 Tuesday 3  
 Sunday 3  
 Wednesday 2**

**Friday records the highest Emergrncy appointments followed by Saturday. Sunday with low availability due no doctors present.**

* **What are the most prescribed medication categories by age group?**

with cte as (

select pt.age\_group, p.med\_category, count(p.patient\_id) as count\_of\_patients, dense\_rank() over (partition by age\_group order by count(p.patient\_id) desc) as rnk

from prescriptions p join patients pt on p.patient\_id = pt.patient\_id

group by pt.age\_group, p.med\_category

order by pt.age\_group, count\_of\_patients desc)

select age\_group, med\_category, count\_of\_patients from cte where rnk=1

**RESULT**

**Age\_group Med\_category number\_of\_patients**

**0-18 Infection 1  
0-18 Heart 1  
0-18 Diabetes 1  
19-30 Pain Relief 5  
31-50 Heart 16  
51-70 Pain Relief 15  
71+ Pain Relief 15**

**People above age group 50 tend to take more Painkiller medicines and people of age group 31-50 are more likely to suffer from heart related issues.**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| **Key Findings to Improve Healthcare Operations**   1. **High Engagement from Long-Term Patients**    * Long-term patients account for a significant portion of appointment volume.    * **Implication**: These patients are more likely to follow through on care plans, making them ideal candidates for chronic condition management programs and long-term wellness initiatives. 2. **Underutilization by New Patients**    * New patients show very low appointment frequency.    * **Implication**: There may be barriers to engagement early in the care journey (e.g., poor onboarding, limited access). Targeted welcome programs and follow-up reminders could help increase their engagement and retention. 3. **Age-Driven Healthcare Demand**    * The **71+ age group** makes up the largest proportion of the patient population and likely drives demand for prescriptions and appointments.    * **Implication**: Resource planning (staffing, appointment slots, specialized care) should account for the needs of this growing demographic. |  |

**Suggestions for Further Analysis**

* Explore missed appointments and no-show rates by patient type and weekday.
* Analyze provider efficiency based on appointment load and prescription volume.
* Study cost implications of prescription patterns across different demographics.