Additional Questions

One question I was curious about was whether GPA differs significantly between American and international applicants for Fall 2025. I was curious as my wife was an international student from Japan, and a much better student than I, so I wondered how this tracked with a larger dataset. To answer this, I grouped applicants by their reported citizenship category and averaged GPA for each group.

```
WITH parsed AS (
  SELECT
    *,
    CASE
      WHEN status ~ '^\d{4}-\d{2}-\d{2}$'
      THEN to_date(status, 'YYYY-MM-DD')
      ELSE NULL
    END AS decision_date_parsed
  FROM applicants
)
SELECT
  lower(us_or_international) AS category,
  ROUND(AVG(gpa)::numeric, 3) AS avg_gpa
FROM parsed
WHERE decision_date_parsed BETWEEN DATE '2025-01-01' AND DATE '2025-08-31'
  AND us_or_international IS NOT NULL
GROUP BY category;
```

Another question I explored was which universities had the highest acceptance rates within the Fall 2025 window. The query calculates acceptance percentages by university, filters out those with very few applicants, and sorts the results so we can focus on schools with a meaningful number of submissions.

```
WITH parsed AS (
```

```
 SELECT
  *,
  CASE
    WHEN status ~ '^\d{4}-\d{2}-\d{2}$'
    THEN to_date(status, 'YYYY-MM-DD')
    ELSE NULL
  END AS decision_date_parsed
 FROM applicants
)
SELECT
 llm_generated_university AS university,
 COUNT(*) AS n,
 ROUND(100.0 * SUM(CASE WHEN lower(status) LIKE 'accepted%' THEN 1 ELSE 0 END) /
COUNT(*), 2) AS acceptance_rate_pct
FROM parsed
WHERE decision_date_parsed BETWEEN DATE '2025-01-01' AND DATE '2025-08-31'
 AND llm_generated_university IS NOT NULL
GROUP BY llm_generated_university
HAVING COUNT(*) >= 5
ORDER BY acceptance_rate_pct DESC;
```