

Grad Café Data Analysis Report

Module 3 Programming Assignment

Author: Wei Liu

Date: September 13, 2025

Introduction

This report presents the results of SQL queries performed on a PostgreSQL database containing 50,000 grad school admission records scraped from GradCafe. The database includes information about applications, admission outcomes, student demographics, and academic metrics. All queries were designed to answer specific research questions about graduate school admissions patterns.

Query Results and Analysis

Question 1

Question: How many entries do you have in your database who have applied for Fall 2025?

Result: 6,631

SQL Query:

```
SELECT COUNT(*) FROM applicant_data WHERE term ILIKE '%Fall 2025%'
```

Explanation:

This query counts all records where the term field contains 'Fall 2025'. The ILIKE operator provides case-insensitive pattern matching to capture variations in formatting.

Question 2

Question: What percentage of entries are from international students (not American or Other)?

Result: 53.53%

SQL Query:

```
SELECT ROUND((COUNT(*) FILTER (WHERE us_or_international ILIKE '%International%') * 100.0 / COUNT(*)), 2) FROM applicant_data WHERE us_or_international IS NOT NULL AND us_or_international != ''
```

Explanation:

This query calculates the percentage of international students using a filtered count. The FILTER clause counts only records where the student type is 'International', then divides by the total count of non-null entries.

Question 3

Question: What is the average GPA, GRE, GRE V, GRE AW of applicants who provide these metrics?

Result: GPA: 3.77, GRE: 176.86, GRE V: 165.44, GRE AW: 5.50

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 2), ROUND(AVG(gre)::numeric, 2),  
ROUND(AVG(gre_v)::numeric, 2), ROUND(AVG(gre_aw)::numeric, 2) FROM applicant_data  
WHERE (gpa IS NOT NULL OR gre IS NOT NULL OR gre_v IS NOT NULL OR gre_aw IS NOT  
NULL) AND (gre IS NULL OR gre <= 170) AND (gre_v IS NULL OR gre_v <= 170) AND (gre_aw  
IS NULL OR gre_aw <= 6)
```

Explanation:

This query computes averages for each academic metric separately, including only records that have at least one non-null value. Each average is rounded to 2 decimal places for readability.

Question 4

Question: What is the average GPA of American students in Fall 2025?

Result: 3.77

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 2) FROM applicant_data WHERE term ILIKE '%Fall  
2025%' AND us_or_international ILIKE '%American%' AND gpa IS NOT NULL
```

Explanation:

This query filters for Fall 2025 American students with non-null GPA values, then calculates the average. The result shows American students maintain a high average GPA.

Question 5

Question: What percent of entries for Fall 2025 are Acceptances?

Result: 35.88%

SQL Query:

```
SELECT ROUND((COUNT(*) FILTER (WHERE status ILIKE '%Accept%') * 100.0 / COUNT(*)),  
2) FROM applicant_data WHERE term ILIKE '%Fall 2025%'
```

Explanation:

This query calculates the acceptance rate for Fall 2025 applications by counting acceptances and dividing by total Fall 2025 applications. The ~36% acceptance rate reflects competitive admission standards.

Question 6

Question: What is the average GPA of applicants who applied for Fall 2025 who are Acceptances?

Result: 3.76

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 2) FROM applicant_data WHERE term ILIKE '%Fall 2025%' AND status ILIKE '%Accept%' AND gpa IS NOT NULL
```

Explanation:

This query finds the average GPA of accepted Fall 2025 applicants. Interestingly, the average GPA of accepted students (3.76) is nearly identical to the overall American student average (3.77).

Question 7

Question: How many entries are from applicants who applied to JHU for a masters degrees in Computer Science?

Result: 35

SQL Query:

```
SELECT COUNT(*) FROM applicant_data WHERE (program ILIKE '%Johns Hopkins%' OR program ILIKE '%JHU%' OR llm_generated_university ILIKE '%Johns Hopkins%') AND (program ILIKE '%Computer Science%' OR program ILIKE '%CS%' OR llm_generated_program ILIKE '%Computer Science%') AND degree ILIKE '%Masters%'
```

Explanation:

This query uses multiple pattern matching conditions to capture various ways JHU and Computer Science might be represented in the data. It searches both original and LLM-generated fields to ensure comprehensive matching.

Question 8

Question: How many entries from 2025 are acceptances from applicants who applied to Georgetown University for a PhD in Computer Science?

Result: 6

SQL Query:

```
SELECT COUNT(*) FROM applicant_data WHERE (program ILIKE '%Georgetown%' OR
llm_generated_university ILIKE '%Georgetown%') AND (program ILIKE '%Computer
Science%' OR program ILIKE '%CS%' OR llm_generated_program ILIKE '%Computer
Science%') AND degree ILIKE '%PhD%' AND status ILIKE '%Accept%' AND date_added >=
'2025-01-01'
```

Explanation:

This query combines university matching, program matching, degree filtering, status filtering, and date filtering to find specific Georgetown CS PhD acceptances in 2025.

Question 9

Question: Pennsylvania State University International Student Percentage for Fall 2025

Result: 58.33%

SQL Query:

```
SELECT ROUND((COUNT(*) FILTER (WHERE us_or_international ILIKE '%International%') *
100.0 / COUNT(*)), 2) FROM applicant_data WHERE term ILIKE '%Fall 2025%' AND
(program ILIKE '%Pennsylvania State%' OR program ILIKE '%Penn State%' OR program
ILIKE '%PSU%' OR llm_generated_university ILIKE '%Pennsylvania State%' OR
llm_generated_university ILIKE '%Penn State%') AND us_or_international IS NOT NULL AND
us_or_international != ''
```

Explanation:

This additional query examines Penn State's international student percentage for Fall 2025, showing a higher international percentage (58.33%) compared to the overall dataset (53.53%).

Question 10

Question: Pennsylvania State University Fall 2025 Acceptances

Result: 25

SQL Query:

```
SELECT COUNT(*) FROM applicant_data WHERE (program ILIKE '%Pennsylvania State%' OR
program ILIKE '%Penn State%' OR program ILIKE '%PSU%' OR llm_generated_university
ILIKE '%Pennsylvania State%' OR llm_generated_university ILIKE '%Penn State%') AND
status ILIKE '%Accept%' AND date_added >= '2025-01-01'
```

Explanation:

This additional query counts Penn State's Fall 2025 acceptances, reflecting a total of 25 records.

Technical Implementation Notes

Database: PostgreSQL with 50,000 records from GradCafe

Query Language: SQL with PostgreSQL-specific functions

Pattern Matching: ILIKE for case-insensitive searches

Data Quality: Used both original and LLM-generated fields for comprehensive matching

Null Handling: Appropriate filtering for NULL values in calculations