SQL QUERIES

QUESTION 1: How many entries do you have in your database who have applied for Fall 2026?

SQL Query:

```
SELECT COUNT(*)
FROM applicants
WHERE term = 'Fall 2026';
```

Result:

```
gradcafe=> SELECT COUNT(*)
FROM applicants
WHERE term = 'Fall 2026';
 count
-------
  4735
(1 row)
```

Explanation: The query counts all entries where the program term is Fall 2026.

QUESTION2 : What percentage of entries are from international students (not American or Other) (to two decimal places)?

SQL Query:

```
SELECT
  ROUND(
    100.0 * SUM(CASE WHEN us_or_international = 'International' THEN 1 ELSE 0 END)
    / NULLIF(COUNT(*), 0),
    2
  )
FROM applicants
WHERE us_or_international IS NOT NULL;
```

Result:

```
gradcafe=> SELECT
  ROUND(
    100.0 * SUM(CASE WHEN us_or_international = 'International' THEN 1 ELSE 0 END)
    / NULLIF(COUNT(*), 0),
    2
  )
FROM applicants
WHERE us_or_international IS NOT NULL;
 round
-------
 44.24
```

(1 row)

Explanation: The dataset contains two citizenship categories (American and International). The percentage was calculated using all entries with known citizenship.

QUESTION 3: What is the average GPA, GRE, GRE V, GRE AW of applicants who provide these metrics?

SQL Query:

```
SELECT
  ROUND(AVG(gpa)::numeric, 3)    AS avg_gpa,
  ROUND(AVG(gre)::numeric, 3)    AS avg_gre,
  ROUND(AVG(gre_v)::numeric, 3)  AS avg_gre_v,
  ROUND(AVG(gre_aw)::numeric, 3) AS avg_gre_aw
FROM applicants;
```

Result:

```
gradcafe=> SELECT
  ROUND(AVG(gpa)::numeric, 3)    AS avg_gpa,
  ROUND(AVG(gre)::numeric, 3)    AS avg_gre,
  ROUND(AVG(gre_v)::numeric, 3)  AS avg_gre_v,
  ROUND(AVG(gre_aw)::numeric, 3) AS avg_gre_aw
FROM applicants;
 avg_gpa | avg_gre | avg_gre_v | avg_gre_aw
---------+---------+-----------+------------
   3.797 | 279.682 |   161.238 |      4.202
(1 row)
```

Explanation:

This query calculates the average GPA, GRE Quantitative score, GRE Verbal score, and GRE Analytical Writing score for applicants in the dataset. The SQL `AVG()` function automatically excludes NULL values, ensuring that only applicants who provided each metric are included in the calculation.

The results show that applicants generally have strong academic performance, with an average GPA close to 3.8 and competitive GRE scores.

QUESTION 4: What is their average GPA of American students in Fall 2026?

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 3)
FROM applicants
```

```
WHERE term = 'Fall 2026'
  AND us_or_international = 'American'
  AND gpa IS NOT NULL;
```

Result:

```
gradcafe=> SELECT ROUND(AVG(gpa)::numeric, 3)
FROM applicants
WHERE term = 'Fall 2026'
  AND us_or_international = 'American'
  AND gpa IS NOT NULL;
 round
-------
 3.811
(1 row)
```

Explanation: This query calculates the average GPA of applicants who identified as American and applied for Fall 2026 programs. The query filters applicants based on the application term and citizenship status and excludes entries where GPA data is missing. The results show that the average GPA of American applicants applying for Fall 2026 is **3.811**, indicating strong academic performance among this applicant group.

QUESTION 5: What percent of entries for Fall 2025 are Acceptances (to two decimal places)?

SQL Query:

```
SELECT
  ROUND(
    100.0 * SUM(CASE WHEN status = 'Accepted' THEN 1 ELSE 0 END)
    / NULLIF(COUNT(*), 0),
    2
  )
FROM applicants
WHERE term = 'Fall 2025';
```

Result:

```
gradcafe=> SELECT
```

```
  ROUND(
    100.0 * SUM(CASE WHEN status = 'Accepted' THEN 1 ELSE 0 END)
    / NULLIF(COUNT(*), 0),
    2
  )
FROM applicants
WHERE term = 'Fall 2025';
 round
-------
 50.00
(1 row)
```

Explanation:

This query calculates the percentage of applicants who were accepted among all Fall 2025 submissions. The numerator counts accepted applicants, and the denominator counts total Fall 2025 submissions.

The acceptance rate for Fall 2025 applicants was **50.00%**.

QUESTION 6: What is the average GPA of applicants who applied for Fall 2026 who are Acceptances?

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 3)
FROM applicants
WHERE term = 'Fall 2026'
  AND status = 'Accepted'
  AND gpa IS NOT NULL;
```

Result:

```
gradcafe=> SELECT ROUND(AVG(gpa)::numeric, 3)
FROM applicants
WHERE term = 'Fall 2026'
  AND status = 'Accepted'
  AND gpa IS NOT NULL;
 round
-------
 3.780
(1 row)
```

Explanation: This query calculates the average GPA of applicants who applied for Fall 2026 programs and were accepted. Only entries with available GPA data were included.

The average GPA of accepted Fall 2026 applicants is **3.780**, indicating that admissions decisions consider multiple evaluation factors beyond GPA alone.

QUESTION 7: How many entries are from applicants who applied to JHU for a masters degrees in Computer Science?

SQL Query:

```
SELECT COUNT(*)
FROM applicants
WHERE degree ILIKE 'Master%'
  AND llm_generated_program ILIKE '%computer science%'
  AND (
        llm_generated_university ILIKE 'Johns Hopkins%'
     OR llm_generated_university ILIKE '%JHU%'
     OR llm_generated_university ILIKE '%John%Hopkins%' );
```

Result:

```
gradcafe=> SELECT COUNT(*)
FROM applicants
WHERE degree ILIKE 'Master%'
  AND llm_generated_program ILIKE '%computer science%'
  AND (
        llm_generated_university ILIKE 'Johns Hopkins%'
     OR llm_generated_university ILIKE '%JHU%'
     OR llm_generated_university ILIKE '%John%Hopkins%' );

 count
-------
     1
(1 row)
```

Explanation: The query includes both full university name matching ("Johns Hopkins") and abbreviation matching ("JHU") to ensure all relevant applicant entries were captured.

QUESTION 8: How many entries from 2026 are acceptances from applicants who applied to Georgetown University, MIT, Stanford University, or Carnegie Mellon University for a PhD in Computer Science?

SQL Query:

```
SELECT COUNT(*)
FROM applicants
WHERE term ILIKE '%2026%'
  AND status = 'Accepted'
  AND degree ILIKE 'PhD%'
  AND llm_generated_program ILIKE '%computer science%'
  AND (
        llm_generated_university ILIKE '%Georgetown%'
     OR llm_generated_university ILIKE '%MIT%'
     OR llm_generated_university ILIKE '%Massachusetts Institute of Technology%'
     OR llm_generated_university ILIKE '%Stanford%'
     OR llm_generated_university ILIKE '%Carnegie Mellon%'
     OR llm_generated_university ILIKE '%CMU%'
      );
```

Result:

```
gradcafe=> SELECT COUNT(*)
FROM applicants
WHERE term ILIKE '%2026%'
  AND status = 'Accepted'
  AND degree ILIKE 'PhD%'
  AND llm_generated_program ILIKE '%computer science%'
  AND (
        llm_generated_university ILIKE '%Georgetown%'
     OR llm_generated_university ILIKE '%MIT%'
     OR llm_generated_university ILIKE '%Massachusetts Institute of Technology%'
     OR llm_generated_university ILIKE '%Stanford%'
     OR llm_generated_university ILIKE '%Carnegie Mellon%'
     OR llm_generated_university ILIKE '%CMU%'
      );
 count
-------
     2
(1 row)
```

Explanation:

This query counts applicants from 2026 who were accepted into PhD programs in Computer Science at Georgetown University, MIT, Stanford University, or Carnegie Mellon University. The query filters applicants by term year, admission status, degree type, and program field using LLM-generated university and program names for consistent matching. The results show that **2 applicants** met these criteria.

QUESTION 9: Do your numbers for question 8 change if you use LLM Generated Fields (rather than your downloaded fields)?

Solutions: When the query was executed using raw downloaded program fields instead of LLM-generated structured fields, the number of matching applicants changed from **2 to 0**. This difference occurs because scraped program text often contains inconsistent formatting, abbreviations, or incomplete university names. The LLM-generated fields standardize university and program names, allowing more accurate and reliable matching. Therefore, LLM-generated fields improve data quality and analysis accuracy.

CUSTOM SQL QUESTIONS:

QUESTION10: What are the top 5 most frequently applied-to programs for Fall 2026?

SQL Query:

```
SELECT program, COUNT(*) AS application_count
FROM applicants
WHERE term = 'Fall 2026'
GROUP BY program
ORDER BY application_count DESC
LIMIT 5;
```

RESULT:

```
gradcafe=> SELECT program, COUNT(*) AS application_count
FROM applicants
WHERE term = 'Fall 2026'
GROUP BY program
ORDER BY application_count DESC
LIMIT 5;
         program         | application_count
-------------------------+-------------------
 Physics PhD             |               251
 Chemistry PhD           |               248
 Clinical Psychology PhD |               217
 Sociology PhD           |               181
 English PhD             |               143
(5 rows)
```

Explanation:This query identifies the five graduate programs that received the highest number of applicant submissions for Fall 2026 by counting application entries for each program.

QUESTION 11: What is the average GPA of International applicants?

SQL Query:

```
SELECT ROUND(AVG(gpa)::numeric, 3)
```

```
FROM applicants
WHERE us_or_international = 'International'
  AND gpa IS NOT NULL;
```

RESULT:

```
gradcafe=> SELECT ROUND(AVG(gpa)::numeric, 3)
FROM applicants
WHERE us_or_international = 'International'
  AND gpa IS NOT NULL;
 round
-------
 3.777
(1 row)
```

Explanation: This query calculates the average GPA of applicants who identified as international students. Only applicants who provided GPA information were included.