

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

Answer: 6990

SQL:

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

Why this query:

Counts all rows where term equals 'Fall 2026' (TRIM avoids trailing-space mismatches).

---

2) What percentage of entries are from international students (to two decimal places)?

Answer: 44.36

SQL:

```
SELECT  
  ROUND(  
    (100.0 * COUNT(*) /  
    NULLIF((SELECT COUNT(*) FROM applicants), 0)  
  )::numeric,  
    2  
  )  
  FROM applicants  
 WHERE us_or_international = 'International';
```

Why this query:

International count divided by total entries times 100, rounded to 2 decimals. NULLIF prevents divide-by-zero.

---

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

Answer: (Decimal('3.75'), Decimal('204.89'), Decimal('160.42'), Decimal('8.51'))

SQL:

```
SELECT  
  ROUND(AVG(gpa)::numeric, 2),  
  ROUND(AVG(gre)::numeric, 2),  
  ROUND(AVG(gre_v)::numeric, 2),  
  ROUND(AVG(gre_aw)::numeric, 2)  
  FROM applicants  
 WHERE gpa IS NOT NULL  
    OR gre IS NOT NULL  
    OR gre_v IS NOT NULL  
    OR gre_aw IS NOT NULL;
```

Why this query:

Averages numeric columns; restricts to rows where at least one metric exists.

---

4) What is the average GPA of American students in Fall 2026?

Answer: 3.79

SQL:

```
SELECT ROUND(AVG(gpa)::numeric, 2)
FROM applicants
WHERE TRIM(term) = 'Fall 2026'
AND us_or_international = 'American'
AND gpa IS NOT NULL;
```

Why this query:

Filters to Fall 2026 + American + non-null GPA, then averages.

---

5) What percent of entries for Fall 2026 are Acceptances (to two decimal places)?

Answer: 24.32

SQL:

```
SELECT
CASE
WHEN (SELECT COUNT(*) FROM applicants WHERE TRIM(term) = 'Fall 2026') = 0
THEN 0
ELSE ROUND(
(100.0 * COUNT(*) /
(SELECT COUNT(*) FROM applicants WHERE TRIM(term) = 'Fall 2026')
)::numeric,
2
)
END
FROM applicants
WHERE TRIM(term) = 'Fall 2026'
AND status = 'Accepted';
```

Why this query:

Accepted count divided by total Fall 2026 count times 100; CASE prevents division by zero.

---

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

Answer: 3.76

SQL:

```
SELECT ROUND(AVG(gpa)::numeric, 2)
FROM applicants
WHERE TRIM(term) = 'Fall 2026'
AND status = 'Accepted'
AND gpa IS NOT NULL;
```

Why this query:

Filters to accepted Fall 2026 applicants with GPA and averages.

---

7) How many entries are from applicants who applied to JHU for a masters degree in Computer Science?

Answer: 6

SQL:

```
SELECT COUNT(*)
  FROM applicants
 WHERE program ILIKE '%Johns Hopkins%'
    AND program ILIKE '%Computer Science%'
    AND degree = 'Masters';
```

Why this query:

Case-insensitive matching with ILIKE on program text and restricts degree to Masters.

---

8) How many Fall 2026 acceptances are from applicants who applied to Georgetown, MIT, Stanford, or CMU for a PhD in CS?

Answer: 2

SQL:

```
SELECT COUNT(*)
  FROM applicants
 WHERE TRIM(term) = 'Fall 2026'
    AND degree = 'PhD'
    AND status = 'Accepted'
    AND program ILIKE '%Computer Science%'
    AND (
      program ILIKE '%Georgetown%' OR
      program ILIKE '%MIT%' OR
      program ILIKE '%Stanford%' OR
      program ILIKE '%Carnegie Mellon%'
    );
```

Why this query:

Filters by term/degree/status and matches university keywords inside the original program text.

---

9) Do the numbers for question 8 change if you use LLM-generated fields?

Answer: 2

SQL:

```
SELECT COUNT(*)
  FROM applicants
 WHERE TRIM(term) = 'Fall 2026'
    AND degree = 'PhD'
    AND status = 'Accepted'
    AND llm_generated_program ILIKE '%Computer Science%'
    AND (
      llm_generated_university ILIKE '%Georgetown%' OR
      llm_generated_university ILIKE '%MIT%' OR
      llm_generated_university ILIKE '%Stanford%' OR
      llm_generated_university ILIKE '%Carnegie Mellon%'
    );
```

Why this query:

Same as Q8 but matches CS + universities using normalized LLM-generated fields.

---

10) Additional Question: What are the top 5 universities by number of entries (LLM university field)?

Answer: University of California: 1774; University of Michigan: 1136; Stanford University: 1125; Columbia University: 1038; Yale

SQL:

```
SELECT STRING_AGG(x, ';' )
FROM (
  SELECT llm_generated_university || ':' || COUNT(*) AS x
  FROM applicants
  WHERE llm_generated_university IS NOT NULL
  GROUP BY llm_generated_university
  ORDER BY COUNT(*) DESC
  LIMIT 5
) t;
```

Why this query:

Counts entries per LLM-normalized university and returns the top 5 as a readable summary string.

---

11) Additional Question: What is the overall acceptance rate across all terms (to two decimals)?

Answer: 35.21

SQL:

```
SELECT
  ROUND(
    (100.0 * (SELECT COUNT(*) FROM applicants WHERE status='Accepted') /
    NULLIF((SELECT COUNT(*) FROM applicants), 0)
  )::numeric,
  2
);
```

Why this query:

Computes accepted / total \* 100 and rounds to 2 decimals. NULLIF prevents divide-by-zero.

---