1. **How many entries do you have in your database who have applied for Fall 2024?**

```
SELECT 'Applicant count: ' || COUNT(*) FROM applicants WHERE term in ('Fall 2024','F24')
```

I saw in the data that term can be the full name or shorted filtered for both cases.

2. **What percentage of entries are from international students (not American or Other) (to two decimal places)?**

```
SELECT 'Percent International: ' || ROUND((SUM(CASE WHEN us_or_international =
 'International' THEN 1 ELSE 0 END)::numeric / COUNT(*) * 100), 2) || '%'
AS percent_international FROM applicants;
```

The field field that captures international students so I just filtered where it's set to International and divided it by all applications.

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

```
SELECT 'Average GPA: ' || ROUND(AVG(NULLIF(gpa,0))::numeric,2) || ',
Average GRE: ' || ROUND(AVG(NULLIF(gre,0))::numeric,2) || ', Average GRE V: ' ||
ROUND(AVG(NULLIF(gre_v,0))::numeric,2) || ', Average GRE AW: ' ||
ROUND(AVG(NULLIF(gre_aw,0))::numeric,2) FROM applicants;
```

In this case, I set the scores and GPA to be NULL in the cases where they were set to 0 and likely not entered. I then averaged each of the scores.

4. **What is their average GPA of American students in Fall 2025?**

```
SELECT 'Average GPA American: ' || ROUND(AVG(NULLIF(gpa,0))::numeric,2)
FROM applicants WHERE term = 'Fall 2025' AND us_or_international = 'American'
```

I filtered for the Fall 2025 term and American applications and then averaged the CPA. I again removed 0 from scores before averaging.

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

```
```

```
SELECT 'Acceptance percent: ' || ROUND(SUM(CASE WHEN status = 'Accepted' THEN 1 ELSE 0 END)::numeric / COUNT(*) * 100,2) || '%' FROM applicants WHERE term IN ('Fall 2025','F25')
```

I filtered for Fall 2025 applications. On the numerator I filtered for applications where the status is set to Accepted and then the denominator is all applications.

6. **What is the average GPA of applicants who applied for Fall 2025 who are Acceptances?**

```
SELECT 'Average GPA Acceptance: ' || ROUND(AVG(NULLIF(gpa,0))::numeric,2) FROM applicants WHERE term in ('Fall 2025','F25') AND status = 'Accepted'
```

I filtered for Fall 2025 applications where the status is Accepted. I then averaged the GPA, again setting 0s to NULL so they'd be removed from the average calculation.

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

```
SELECT 'JHU Computer Science Masters Applications: ' || COUNT(*) FROM applicants WHERE llm_generated_university = 'Johns Hopkins University' AND llm_generated_program = 'Computer Science' AND degree = 'Masters'
```

I used the LLM normalized university and degrees and filtered for JHU and computer science. I also filtered for degree set to masters. I then counted all the applications.

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

```
SELECT 'Georgetown University Computer Science PhD Acceptances: ' || COUNT(*) FROM applicants WHERE llm_generated_university = 'Georgetown University' AND llm_generated_program = 'Computer Science' AND term LIKE '%2025%' AND status = 'Accepted' AND degree = 'PhD'
```

I again used the LLM normalized columns to filter for the university and program. I then filtered for the PhD program and status is Accepted. I then counted all the applications.

9. **What universities with at least 10 applications for master's programs in computer science have the lowest and highest acceptance rates?**

```
SELECT 'Lowest acceptance rate: ' || (SELECT university || ' - ' || ROUND(acceptance_rate * 100, 2)::text || '%' FROM (SELECT llm_generated_university AS university, SUM(CASE WHEN status = 'Accepted' THEN 1 ELSE 0 END)::numeric / COUNT(*) AS acceptance_rate,
```

```
COUNT(*) AS total_apps FROM applicants WHERE degree = 'Masters' AND
llm_generated_program = 'Computer Science' GROUP BY 1 HAVING COUNT(*) >= 10) a
ORDER BY acceptance_rate ASC LIMIT 1) || ' | Highest acceptance rate: ' || (SELECT
university || ' - ' || ROUND(acceptance_rate * 100, 2)::text || '%' FROM (SELECT
llm_generated_university AS university, SUM(CASE WHEN status = 'Accepted' THEN 1 ELSE
0 END)::numeric / COUNT(*) AS acceptance_rate, COUNT(*) AS total_apps FROM applicants
WHERE degree = 'Masters' AND llm_generated_program = 'Computer Science' GROUP BY 1
HAVING COUNT(*) >= 10) b ORDER BY acceptance_rate DESC LIMIT 1)
```

I grouped by the normalized university name and calculated the acceptance rate, sorting both ascending and descending to get the highest and lowest acceptance rates. I had a minimum of 10 applications because I wanted to make sure there's a large enough sample to get an estimate.

10. **What are the most popular PhD programs in 2024 and 2025 by applications?**

```
SELECT 'Most popular PhD program in 2024: ' || (SELECT llm_generated_program || ' at ' ||
llm_generated_university || ' - ' || COUNT(*) FROM applicants WHERE degree = 'PhD' AND
term LIKE '%2024%' GROUP BY llm_generated_program, llm_generated_university ORDER
BY COUNT(*) DESC LIMIT 1) || ' | Most popular PhD program in 2025: ' || (SELECT
llm_generated_program || ' at ' || llm_generated_university || ' - ' || COUNT(*) FROM applicants
WHERE degree = 'PhD' AND term LIKE '%2025%' GROUP BY llm_generated_program,
llm_generated_university ORDER BY COUNT(*) DESC LIMIT 1)
```

I filtered by term and then counted the applications by normalized program and university for 2024 and 2025. I then just took the first row.