Applicant Database Query Report

Analysis of Graduate Program Applications

Query 1: Fall 2026 Applications

Question: How many entries applied for Fall 2026?

Purpose: Establishes the baseline sample size for our most recent application cycle. Understanding the volume of data helps assess statistical significance and provides context for all subsequent Fall 2026 analyses. A larger sample provides more reliable insights into acceptance patterns and applicant characteristics.

SQL Query:

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

Result: 7,030 entries applied for Fall 2026

Query 2: International Student Percentage

Question: What percentage of entries are from international students?

Purpose: Reveals the demographic composition of the applicant pool and helps identify potential selection bias in self-reported data. International students may be more or less likely to report on platforms like GradCafe, affecting data representativeness. This metric also reflects globalization trends in higher education and can indicate program competitiveness on an international scale.

SQL Query:

```
SELECT COUNT(*) FROM applicants WHERE us_or_international = 'International';
```

Result: 50.11% of applicants are international students

Query 3: Average Test Scores

Question: What are the average GPA, GRE, GRE V, and GRE AW scores?

Purpose: Provides benchmark metrics for academic credentials across the entire dataset. These averages are critical for detecting reporting bias - comparing our GRE averages (~160) to national averages (~157) reveals that high-achieving applicants disproportionately self-report, indicating systematic selection bias. This query helps applicants understand the competitive landscape while highlighting the limitations of crowdsourced data.

SQL Query:

```
SELECT AVG(gpa), AVG(gre), AVG(gre_v), AVG(gre_aw) FROM applicants WHERE metrics IS NOT NULL;
```

Results:

• Average GPA: 3.81

• Average GRE: 204.91

• Average GRE V: 160.42

• Average GRE AW: 8.51

Query 4: American Students' GPA (Fall 2026)

Question: What is the average GPA of American students applying for Fall 2026?

Purpose: Isolates a specific demographic cohort to understand if different groups exhibit different credential patterns. The surprisingly high 4.07 average suggests that American students who post to GradCafe may be disproportionately high achievers, reinforcing survivor bias concerns. Comparing this to Query 6 (accepted students' GPA) reveals interesting patterns about who reports and who gets accepted.

SQL Query:

```
SELECT AVG(gpa) FROM applicants WHERE term = 'Fall 2026' AND us_or_international = 'American';
```

Result: Average GPA of 4.07 for American students in Fall 2026

Query 5: Acceptance Rate (Fall 2026)

Question: What percentage of Fall 2026 entries are acceptances?

Purpose: Measures overall acceptance rates in our dataset to assess selection bias. The 24.32% rate is notably higher than typical competitive graduate programs (10-15%), strongly suggesting that accepted applicants are more motivated to share their success publicly. This query demonstrates the fundamental limitation of anonymous self-reported data - we're seeing a filtered sample, not the true applicant population.

SQL Query:

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

Result: 24.34% of Fall 2026 applications resulted in acceptance

Query 6: Accepted Students' GPA (Fall 2026)

Question: What is the average GPA of accepted Fall 2026 applicants?

Purpose: Identifies the academic profile of successful applicants to help prospective students gauge competitiveness. Interestingly, the 3.80 average for accepted students is lower than American applicants overall (4.07), suggesting either grade inflation concerns or that GPA alone doesn't determine acceptance. This metric helps applicants set realistic expectations while understanding that holistic admissions consider multiple factors beyond grades.

SQL Query:

```
SELECT AVG(gpa) FROM applicants WHERE term = 'Fall 2026' AND status = 'Accepted';
```

Result: Average GPA of 3.81 for accepted Fall 2026 applicants

Query 7: JHU Computer Science Masters Applications

Question: How many applied to JHU for a Masters in Computer Science?

Purpose: Tests the database's capability to filter for specific institution-program combinations, demonstrating practical use cases for applicants researching particular schools. The low count (6 entries) reveals data sparsity issues - crowdsourced platforms may have insufficient data for niche queries. This highlights that while aggregate statistics can be informative, program-specific insights may be unreliable due to small sample sizes.

SQL Query:

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

Result: 6 entries for JHU Masters in Computer Science

Query 8: Elite University PhD CS Acceptances (2026)

Question: How many 2026 PhD CS acceptances from Georgetown, MIT, Stanford, or CMU?

Purpose: Examines acceptance outcomes at highly selective institutions to understand data availability for prestigious programs. The extremely low count (2 acceptances) from top-tier programs suggests either genuine scarcity of acceptances or that successful applicants to elite schools don't actively use crowdsourced platforms. This query is crucial for setting up Query 9's comparison of data parsing methods.

SQL Query:

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

Result: 2 acceptances from these elite universities

Query 9: LLM-Generated Fields Comparison

Question: Do the numbers change when using LLM-generated fields?

Purpose: Validates data cleaning and standardization approaches by comparing LLM-parsed university/program names against original free-text entries. The identical results (2 entries for both methods) demonstrate that LLM-based data standardization successfully captures the same information as manual string matching, while potentially offering better scalability and consistency. This query assesses the reliability of using AI for data preprocessing in analytics pipelines.

Results:

• Using LLM-generated fields: 2 entries

• Using original fields: 2 entries

• The numbers are identical

Query 10: Top 10 Most Applied-To Programs (Fall 2026)

Question: What are the top 10 most popular programs for Fall 2026?

Purpose: Identifies trending programs and application patterns to help prospective students understand where competition is concentrated. The dominance of UW-Madison and University of Chicago programs reveals geographic clustering in the dataset. However, 'popularity' in crowdsourced data may reflect which schools' applicants actively use GradCafe rather than true application volume. The widely varying acceptance rates (0% to 15.79%) across top programs demonstrate significant selectivity differences.

SQL Query:

```
SELECT llm_generated_university, llm_generated_program, COUNT(*) as apps, acceptance_rate FROM
applicants WHERE term = 'Fall 2026' GROUP BY university, program ORDER BY apps DESC LIMIT 10;
```

Rank	University	Program	Apps	Accept Rate
1	University of Wisconsin - Madi	Mathematics	39	2.6%
2	University of Chicago	Physics	38	5.3%
3	University of Wisconsin - Madi	Physics	38	15.8%
4	University of Wisconsin - Madi	Chemistry	33	6.1%
5	University of Chicago	Political Science	32	0.0%
6	Yale University	History	31	3.2%
7	University of Chicago	Philosophy	28	7.1%
8	Northwestern University	Philosophy	24	8.3%
9	University of Chicago	Chemistry	24	12.5%
10	University of Chicago	Sociology	22	0.0%

Query 11: PhD vs Masters Comparison

Question: How do acceptance rates compare between PhD and Masters programs?

Purpose: Contrasts acceptance patterns between terminal professional degrees (Masters) and research-focused degrees (PhD) to inform applicant strategy. The dramatic difference in acceptance rates (64.92% vs 25.19%) reflects fundamental differences in program structure - Masters programs are often revenue-generating with larger cohorts, while PhD programs are highly selective and funded. The similar GPAs (3.75 vs 3.81) suggest academic credentials alone don't explain acceptance disparities; research experience and funding availability are likely more critical for PhD admissions.

SQL Query:

```
SELECT degree, COUNT(*), SUM(acceptances), acceptance_rate, AVG(gpa) FROM applicants WHERE degree
IN ('PhD', 'Masters') GROUP BY degree;
```

Degree	Total Apps	Acceptances	Accept Rate	Avg GPA (Accepted)
Masters	12,108	7,860	64.92%	3.75 (n=6,018)
PhD	35,613	8,969	25.18%	3.81 (n=5,058)

Key Findings

• Masters programs have a much higher acceptance rate (64.92%) compared to PhD programs (25.19%)

• International students make up approximately 50% of all applicants

• The average GPA for Fall 2026 acceptances (3.80) is slightly lower than American students' average (4.07)

• University of Wisconsin - Madison Mathematics program received the most applications for Fall 2026 (39 apps)

• LLM-generated fields produced identical results to manual parsing for elite university queries