

## MODULE 3 - DATABASE QUERIES ASSIGNMENT

### JHU Modern Software Concepts

=====

QUESTION 1: How many entries applied for Fall 2026?

-----

Answer: 10,526

SQL Query:

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

Explanation: This query counts all rows where the term field equals 'Fall 2026'.

QUESTION 2: What percentage are international students? (to 2 decimal places)

-----

Answer: 49.00%

SQL Query:

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

```
-- Divided by total count
```

Breakdown:

- Total entries: 32,000
- International: 15,681
- American: 16,303
- Other: 16
- Percentage:  $(15681 / 32000) * 100 = 49.00\%$

Explanation: We count entries where us\_or\_international is 'International' and divide by total.

QUESTION 3: What is the average GPA, GRE, GRE V, GRE AW?

-----

Answer:

- Average GPA: 3.7566
- Average GRE (Quantitative): 166.00
- Average GRE Verbal: 160.47
- Average GRE AW: 4.38

SQL Queries:

```
SELECT AVG(gpa) FROM applicants WHERE gpa IS NOT NULL
```

```
SELECT AVG(gre) FROM applicants WHERE gre IS NOT NULL
```

```
SELECT AVG(gre_v) FROM applicants WHERE gre_v IS NOT NULL
```

```
SELECT AVG(gre_aw) FROM applicants WHERE gre_aw IS NOT NULL
```

Explanation: We calculate averages only for entries that have non-NULL values for each metric.

Note: GRE scores are sparse in this dataset because many programs no longer require them post-COVID.

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

-----

Answer: 3.8017

SQL Query:

```
SELECT AVG(gpa) FROM applicants
WHERE us_or_international = 'American'
AND term = 'Fall 2026'
AND gpa IS NOT NULL
```

Explanation: Filter for American students applying Fall 2026 who reported their GPA.

QUESTION 5: What percent of Fall 2025 entries are Acceptances? (to 2 decimal places)

-----

Answer: 35.74%

SQL Query:

```
SELECT COUNT(*) FROM applicants WHERE term = 'Fall 2025' -- Total: 19,764
```

```
SELECT COUNT(*) FROM applicants WHERE term = 'Fall 2025' AND status ILIKE '%Accept%' -- Accepted: 7,063
```

Breakdown:

- Total Fall 2025: 19,764
- Accepted: 7,063
- Percentage:  $(7063 / 19764) * 100 = 35.74\%$

Explanation: Using ILIKE '%Accept%' catches variations like "Accepted", "Acceptance", etc.

QUESTION 6: What is the average GPA of Fall 2026 Acceptances?

-----

Answer: 3.7725

SQL Query:

```
SELECT AVG(gpa) FROM applicants
WHERE term = 'Fall 2026'
AND status ILIKE '%Accept%'
AND gpa IS NOT NULL
```

Explanation: Filter for accepted Fall 2026 applicants who reported their GPA.

QUESTION 7: How many applied to JHU for master's in computer science?

-----

Answer: 7

SQL Query:

```
SELECT COUNT(*) FROM applicants
```

```
WHERE (program ILIKE '%johns hopkins%' OR program ILIKE '%jhu%')
AND program ILIKE '%computer science%'
AND degree ILIKE '%master%'
```

Explanation: Search for both "Johns Hopkins" and "JHU" since users may write either.

The program field contains both program name and university combined.

QUESTION 8: 2026 PhD CS Acceptances at Georgetown, MIT, Stanford, CMU?

-----

Answer: 20

SQL Query:

```
SELECT COUNT(*) FROM applicants
WHERE term ILIKE '%2026%'
AND status ILIKE '%Accept%'
AND degree ILIKE '%PhD%'
AND program ILIKE '%computer science%'
AND (
    program ILIKE '%georgetown%'
    OR program ILIKE '%mit%'
    OR program ILIKE '%massachusetts institute%'
    OR program ILIKE '%stanford%'
    OR program ILIKE '%carnegie mellon%'
    OR program ILIKE '%cmu%'
)
```

Explanation: Using the raw 'program' field which contains both program and university.

Multiple variations are searched to catch different spellings (MIT vs Massachusetts Institute of Technology).

QUESTION 9: Do numbers change using LLM Generated Fields?

-----

Answer: YES - Changed from 20 to 10

SQL Query (using LLM fields):

```
SELECT COUNT(*) FROM applicants
WHERE term ILIKE '%2026%'
AND status ILIKE '%Accept%'
AND degree ILIKE '%PhD%'
AND llm_generated_program ILIKE '%computer science%'
AND (
    llm_generated_university ILIKE '%georgetown%'
    OR llm_generated_university ILIKE '%massachusetts institute%'
    OR llm_generated_university ILIKE '%mit%'
    OR llm_generated_university ILIKE '%stanford%'
    OR llm_generated_university ILIKE '%carnegie mellon%'
)
```

Explanation: The numbers changed because the LLM classified MIT's "Electrical Engineering

and Computer Science" (EECS) program as "Electrical Engineering" rather than "Computer Science".

This shows that LLM cleaning can affect query results based on classification decisions.

CUSTOM QUESTION 1: What are the top 10 most popular programs?

-----

Answer:

1. Psychology: 2,771 applications
2. Computer Science: 2,290 applications
3. Chemistry: 2,112 applications
4. Biology: 1,693 applications
5. Physics: 1,553 applications
6. Economics: 1,290 applications
7. Data Science: 1,198 applications
8. English: 1,092 applications

9. Mathematics: 934 applications

10. Statistics: 922 applications

SQL Query:

```
SELECT llm_generated_program, COUNT(*) as count
FROM applicants
WHERE llm_generated_program IS NOT NULL
GROUP BY llm_generated_program
ORDER BY count DESC
LIMIT 10
```

Explanation: Using LLM-generated program names for cleaner grouping (avoids duplicates

like "CS" vs "Computer Science" vs "Comp Sci"). Psychology is the most popular, followed

by Computer Science and Chemistry.

CUSTOM QUESTION 2: What is the acceptance rate by degree type?

-----  
Answer:

- PhD: 24.36% (5,861 accepted / 24,056 total)
- Masters: 66.52% (4,419 accepted / 6,643 total)
- PsyD: 15.10% (85 accepted / 563 total)
- MFA: 33.97% (161 accepted / 474 total)
- Other: 56.40% (119 accepted / 211 total)

SQL Query:

```
SELECT
    degree,
    COUNT(*) as total,
    SUM(CASE WHEN status ILIKE '%Accept%' THEN 1 ELSE 0 END) as
accepted,
    ROUND(100.0 * SUM(CASE WHEN status ILIKE '%Accept%' THEN 1
ELSE 0 END) / COUNT(*), 2) as rate
FROM applicants
```

WHERE degree IS NOT NULL

GROUP BY degree

ORDER BY total DESC

Explanation: PhD programs have much lower acceptance rates (24%) compared to Masters (67%).

This makes sense as PhD programs are typically more competitive and have fewer spots.

**Reference:**

These answers were completed with the help of AI tools: Cursor Claude 4.5, Gemini 3.0, and Copilot.