

Danielle Chan

Modern Software Concepts in Python

9/14/25

Module 3 Assignment: Querying Grad Cafe Data

1. Number of Fall 2025 applicants: 6631

Query Description: Counted the rows where term = 'Fall 2025'

Why: We filter by term because that column directly tells us which semester/year the applicant applied for.

2. Percentage of international applicants: 48.03%

Query Description: We filter by term because that column directly tells us which semester/year the applicant applied for.

Why: We only want the share of applicants who are explicitly international. Excluding "American" and "Other" ensures the percentage reflects true international applicants.

3. Averages:

- a. GPA: 3.79
- b. GRE: 178.25
- c. GRE V: 159.62
- d. GRE AW: 6.71

Query Description: Compute AVG() for gpa, gre, gre_v, gre_aw, ignoring NULLs.

Why: Some applicants don't provide test scores or GPA. Ignoring missing values ensures averages are based only on valid numbers.

4. Average GPA of American applicants in Fall 2025: 3.78

Query Description: Filter term = 'Fall 2025' and us_or_international = 'American', then average GPA.

Why: We restrict to American applicants in that specific term to measure GPA trends for that group only.

5. Percentage of Fall 2025 entries that are acceptances: 35.88%

Query Description: Filter term = 'Fall 2025', count how many have status = 'Acceptance', divide by total count.

Why: Acceptance rate is the proportion of accepted applications, so we compare accepted to total applications in that term.

6. Average GPA of accepted applicants for Fall 2025: 3.77

Query Description: Filter term = 'Fall 2025' and status = 'Acceptance', then average GPA.

Why: We're measuring GPA only among students who were admitted, not all applicants.

7. Number of applicants for JHU master's in computer science: 0

Query Description: Match llm_generated_university for variants of "Johns Hopkins", and llm_generated_program for variants of "Master's" and "Computer Science".

Why: Applicants may write "Johns Hopkins", "JHU", or variations. Similarly, "Master's" may be "MS" or "M.S.". We include these to capture all possible matches.

8. Number of 2025 Georgetown PhD Computer Science acceptances: 0

Query Description: Filter term = 'Fall 2025', status = 'Acceptance', university variants of "Georgetown", and program variants of "PhD" and "Computer Science".

Why: Like JHU, applicants may write the university/program in different ways. We combine multiple filters to make sure we only get Georgetown PhD Computer Science admits in 2025.

9. Number of Fall 2025 Data Science applicants: 60

Query Description: Filter term = 'Fall 2025' and llm_generated_program containing "Data Science".

Why: Program names vary (e.g., "MS in Data Science" vs "Data Science"), so we search for "Data Science" anywhere in the program name.

10. Number of applicants who submitted a GRE score: 3804

Query Description: Program names vary (e.g., "MS in Data Science" vs "Data Science"), so we search for "Data Science" anywhere in the program name.

Why: Not all applicants submit GRE scores, and missing ones are stored as NULL. Checking for non-null values tells us how many provided scores.