Query Descriptions
Modern Software Concepts in Python
Emily Hammer (ehammer5)

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

Answer: 7207

Query:	**Query Description:**
SELECT COUNT(p_id) FROM results WHERE term = 'Fall 2026'	Count the number of entries in the unique ID column that have a term of 'Fall 2026'
Additional Python Calculation:	
None.	

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

Answer: 50.48%

Query:	**Query Description:**
SELECT us_or_international, COUNT(p_id) FROM results GROUP BY us_or_international ORDER BY us_or_international desc;	Group us_or_international column by contents, then count the number of entries in each group
Additional Python Calculation:	
Add results of grouped counts together to find total, then find the International group as a percent of that total, apply rounding.	

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

Answer: Average GPA: 3.77, Average GRE: 220.00, Average GRE V: 161.28, Average GRE AW: 9.83

Query:	**Query Description:**
SELECT avg(gpa), avg(gre), avg(gre_v), avg(gre_aw) FROM results;	Calculate the average of each column gpa, gre, gre_v, and gre_aw
Additional Python Calculation:	
Round each average to two decimal places.	

4. What is their average GPA of American students in Fall 2026?

Answer: 3.80

Query:	**Query Description:**
SELECT avg(gpa) FROM results WHERE term = 'Fall 2026' AND us_or_international = 'American';	Calculate the average of the column gpa for all entries where the term = 'Fall 2026' and us_or_international = 'American'
Additional Python Calculation:	
Round average to two decimal places.	

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

Answer: 35.98%

Query:	**Query Description:**
SELECT status, COUNT(p_id) FROM results WHERE term = 'Fall 2025' GROUP BY status	Group status column by contents for all entries where the term = 'Fall 2025', then count the number of entries in each group

ORDER BY status asc;	
Additional Python Calculation:	
Add results of grouped counts together to find total, then find the Accepted group as a percent of that total, apply rounding.	

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

Answer: 3.77

Query:	**Query Description:**
SELECT avg(gpa) FROM results WHERE term = 'Fall 2026' AND status = 'Accepted';	Calculate the average of the column gpa for all entries where the term = 'Fall 2026' and status = 'Accepted'
Additional Python Calculation:	
Round average to two decimal places.	

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

Answer: 8

Query:	**Query Description:**
SELECT COUNT(p_id) FROM results WHERE degree = 'Masters' AND program LIKE ('Computer Science%') AND (program LIKE ('%Johns Hopkins%') OR program LIKE ('%JHU%') OR program LIKE ('%Hopkins%'));	Count the number of entries in the unique ID column that have a degree of 'Masters' and a program that starts with 'Computer Science' and a program that contains 'Johns Hopkins,' 'JHU,' or 'Hopkins'
Additional Python Calculation:	
None.	

8. How many entries from 2026 are acceptances from applicants who applied to Georgetown University, MIT, Stanford University, or Carnegie Mellon University for a PhD in Computer Science?

Answer: 124

Query:	**Query Description:**
SELECT COUNT(p_id) FROM results WHERE degree = 'PhD' AND program LIKE ('Computer Science%') AND (program LIKE ('%Georgetown%') OR program LIKE ('%MIT%') OR program LIKE ('%Stanford%') OR program LIKE ('%Carnegie%'));	Count the number of entries in the unique ID column that have a degree of 'PhD' and a program that starts with 'Computer Science' and a program that contains 'Georgetown,' 'MIT,' 'Stanford,' or 'Carnegie'
Additional Python Calculation:	
None.	

9. Do your numbers for Q8 change if you use LLM Generated Fields (rather than your downloaded fields)?

Answer: 117

Query:	**Query Description:**
SELECT COUNT(p_id) FROM results WHERE degree = 'PhD' AND llm_generated_program LIKE ('Computer Science%') AND (llm_generated_university LIKE ('%George town%') OR llm_generated_university LIKE ('%Massachusetts Institute of Technology%')	Count the number of entries in the unique ID column that have a degree of 'PhD' and a llm_generated_program that starts with 'Computer Science' and a llm_generated_university that contains 'George town,' 'Massachusetts Institute of Technology,' 'Stanford,' or 'Carnegie'

OR llm_generated_university LIKE ('%Stanford%') OR llm_generated_university LIKE ('%Carnegie%'));	
Additional Python Calculation:	
None.	

Additional Questions

I added questions 10, 11, 12, and 13 to my queries because I wanted to compare the best and the worst of each program and university. While it might have been more helpful to visualize this information as 'Top 10' tables or even the top 5 of each, it was interesting nonetheless to see that the program that accepted the most and rejected the most was Computer Science, and the university that accepted the most and rejected the most was University of California. I'm not sure whether that is due to misattribution on the LLM's part or a higher likelihood of applicants to those programs to self-report.

10. What program accepted the most applicants? How many did they accept?

Answer: Computer Science, 903

Query:	**Query Description:**
SELECT llm_generated_program, COUNT(p_id) as total_count FROM results WHERE status = 'Accepted' GROUP BY llm_generated_program ORDER BY total_count desc;	Group llm_generated_program column by contents for all entries where the status = 'Accepted,' then count the number of entries in each group and order acceptance count descending so the first entry in the results would be the program with the most acceptances
Additional Python Calculation:	
None.	

11. What program rejected the most applicants? How many did they accept?

Answer: Computer Science, 892

Query:	**Query Description:**
SELECT llm_generated_program, COUNT(p_id) as total_count FROM results WHERE status = 'Rejected' GROUP BY llm_generated_program ORDER BY total_count desc;	Group llm_generated_program column by contents for all entries where the status = 'Rejected,' then count the number of entries in each group and order rejection count descending so the first entry in the results would be the program with the most rejections
Additional Python Calculation:	
None.	

12. What university accepted the most applicants? How many did they accept?

Answer: University of California, 414

Query:	**Query Description:**
SELECT llm_generated_university, COUNT(p_id) as total_count FROM results WHERE status = 'Accepted' GROUP BY llm_generated_university ORDER BY total_count desc;	Group llm_generated_university column by contents for all entries where the status = 'Accepted,' then count the number of entries in each group and and order acceptance count descending so the first entry in the results would be the university with the most acceptances
Additional Python Calculation:	
None.	

13. What university rejected the most applicants? How many did they accept?

Answer: University of California, 628

Query:	**Query Description:**

SELECT llm_generated_university, COUNT(p_id) as total_count FROM results WHERE status = 'Rejected' GROUP BY llm_generated_university ORDER BY total_count desc;	Group llm_generated_university column by contents for all entries where the status = 'Rejected,' then count the number of entries in each group and order rejection count descending so the first entry in the results would be the university with the most rejections
Additional Python Calculation:	
None.	