Kate Borisova eboriso1@jh.edu EN605.256 Spring 2025 Module 3 Databases

Part A. Answers to questions

1. How many entries do you have in your database who have applied for Spring 2025?
68

2. What percentage of entries are from international students (not American or Other) (to two decimal places)?
57.46%

3. What is the average GPA, GRE, GRE V, GRE AW of applicants who provide these metrics?
Average GPA: 3.70

Average GRE: 165.27

Average GRE AW: 4.19

Average GRE V: 158.94

4. What is their average GPA of American students in Spring 2025?
3.71

5. What percent of entries for Spring 2025 are Acceptances (to two decimal places)?
51.47%

6. What is the average GPA of applicants who applied for Spring 2025 who are Acceptances?
3.69

7. How many entries are from applicants who applied to JHU for masters degrees in Computer Science?
9

Part B. Limitations

It was a good exercise in the light of web scrapping and loading the data into a database, but overall it is not too representative to rely on the data completely. I see the following reasons why the data is not very reliable:

1. The applicants' identities are not verified, so anyone can post anything.
2. The data input form does not seem to validate most of input values, as we can see data is not consistent. I have seen someone's GPA set to 99.99, which is hardly true.
3. Sharing application progress is absolutely voluntary, so the database represents only those applications that some students decided to share.

The analytic responses do not surprise me, since they are just values shared by some students who like sharing. Thus, we know that people who like to share have the identified scores or shares.

Overall, the selection of data can show timeframes for accepts and declines. Universities are likely to accept first and then decline after a while the rest of applications. I think I will apply the web scraping approach in the future. It is a good exercise and Python libraries are awesome.