Deciphering Big Data:

Individual Reflection (1,000 words)

BY
Shaun Bell-Gibson

ASSIGNMENT: Individual Reflection

July 2023

Introduction

The Gibbs reflective cycle will be the framework to facilitate reflection on the activities conducted during the Deciphering Big Data (DBD) module. The primary focus on the reflection will be on experience of working as part of a team, employing different techniques and the impact on professional development during the module. The Gibbs model uses six stages of reflection: Description, Feelings, Evaluation, Analysis, Conclusion and Action plan (UoE, 2020). These six individual stages presented by Gibbs create an emphasis on obtaining a deeper understanding and encouragement of learning guiding users to an improvement in future practice (Husebø et al., 2015).

Working within a team

The two other team members of Group3 have had a greater amount of experience working in the field of Data Science which put me in position to be able to learn from their collective experience, however there was also a limitation to this situation. After reviewing the draft for the first assessment I noticed that there was insufficient critical analysis and references within the document. I raised this with the group, and they agreed it needed more. A few more references were added however I still wasn't happy with the level of criticality, this time I felt uncomfortable raising the issue again due to the other members seemingly being content with the document and having more experience on the topic. This aligns with research by Brown and Dutton (1995) which states that individuals who perceive themselves as less experienced may feel pressured to conform to the opinions of more experienced group members, inhibiting the willingness to contradict them.

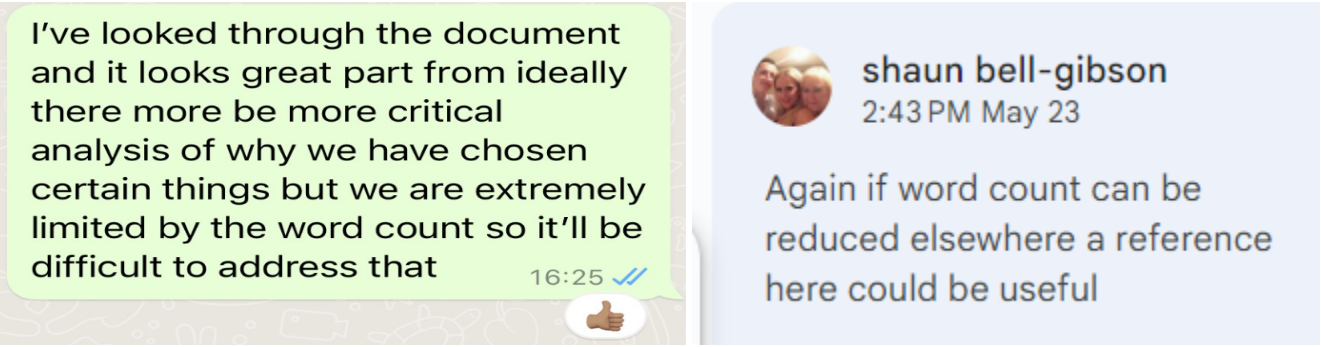

Figure 1 & 2: Review comments for Assessment 1

Feedback from the assessment reflected that more critical discussion should have been included. For the second team assessment I learned from my previous shortcomings and was much more assertive and proactive in driving to move the team forward. This positive mind set allows individuals with less confidence and experience to significantly contribute to group tasks (Dweck, 2006). In the future I will persist in addressing an identified problem even if there are other team members with more experience. I will utilise the experience from the difference I saw in my mindset from the first and second assessments to dictate how I act in future group tasks.

Feedback comments

– Knowledge and understanding of the topic/ issues under consideration An very good demonstration of knowledge and understanding, displaying some originality and in-depth understanding in all key areas of knowledge relevant to the work. The report focuses on the design which is indeed appropriate. There needs to be a good balance between theory and practice and that does not come across in the report - it is more theoretical and require better application focus. See below for further comments. Application of knowledge and understanding A satisfactory demonstration of the application of knowledge and understanding to address the learning outcomes assessed by the assignment within a real-world application. The application is a hotel management database, and the report offers much details on the design. However, there are issues with demonstrating knowledge in a practical setting. For example, the data management pipeline in section 5 is a diagram offering basic descriptions without any reference to the report's application. Please consider that you need to demonstrate how your understanding of the theory assisted with the design process. The requirements section also needs better articulation with the design section, as each is rather stand alone in the report. Criticality A reasonable demonstration of critical analysis, with some insights and reflection, but the linking between theory and practice is not great. There was some critical discussions in section 6 but really limited. More is needed for section 3

Figure 3: Assessment 1 feedback

Database Build

Over the course of the module I have grown in ability and confidence in completing tasks such as the data build task. My first attempt at creating a database was conducted only using MySQL workbench with data from the Normalisation task, this proved to be quite time consuming and required more steps than using code to create the database. I attempted to create the database using SQLmagic however this repeatedly caused errors. As it was my first time using MySQL workbench I struggled to build this simple database and come to an understanding of the software.

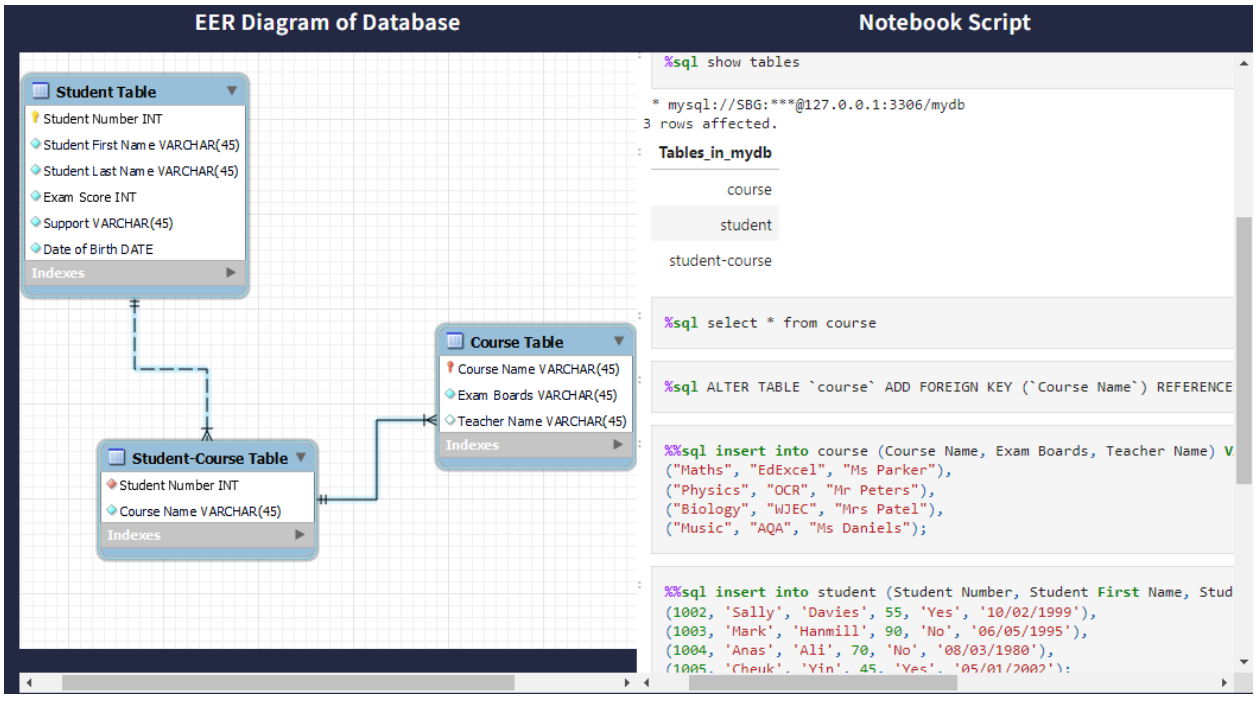

Figure 4: First database attempt

My second attempt at creating a database was conducted using Jupyter notebook, mysql.connector and MySQL workbench with data from the group 3's database proposal. This

time I found the process more straightforward and easy to use due to my previous experience in using python, even with a more complicated database.

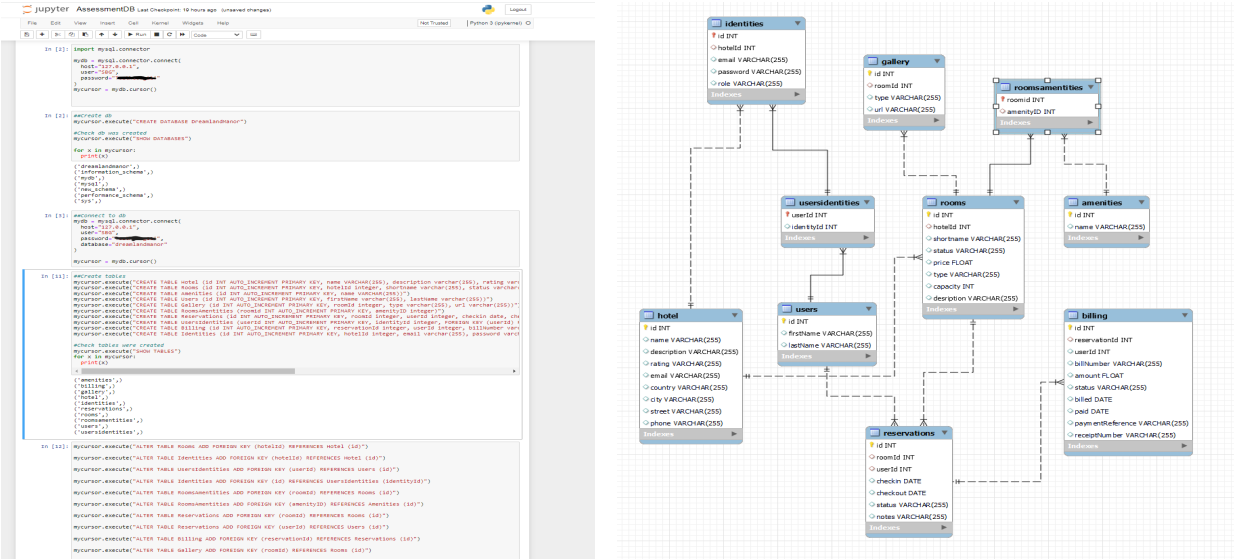

Figure 5: Second database attempt

The final attempt of building a database was a collaborative effort between team 3 to create and populate a database for the Summative assessments in Units 6 & 11 allowing for a deeper understanding and further critical analysis of the database. Within these three attempts to build a database I explored many different methods (Workbench, SQL magic and MySQL.connector) before finding the one that best suited me and my skill set, each time I built on what I had learned from the previous attempt. To improve my database building skills I will experiment using different DBMSs and become more competent in advanced techniques such as stored procedures.

Normalisation Task

The normalisation task required a table with data that is in an unnormalised form to be normalised to the 3rd Normal Form (3NF), showing each step of the process. The principles of normalisation assist in eliminating data redundancy, update anomalies, and data inconsistencies in databases (Elmasri & Navathe, 2019). Initially I was confident in completing the task as I had previously converted unnormalised data to 3NF. However I soon realised that I had not properly learned how to go through each individual step in the process and required a deeper understanding especially between 2NF and 3NF.

1NF involves removing repeating groups and ensuring atomicity within the data elements. 2NF requires identifying and addressing partial dependencies, which often involves breaking tables into separate entities to avoid data redundancy (Silberschatz, Korth, & Sudarshan, 2010). 3NF involves eliminating transitive dependencies, ensuring that non-key attributes depend solely on

the primary key. This requires careful analysis and understanding of the relationships between different attributes, which is something that I did not place as much importance on previously.

The normalisation process significantly improved data integrity and eliminated data anomalies. However, I realised that normalisation can lead to increased complexity as data is queried across multiple normalised tables. Ramakrishnan & Gehrke (2003) state that it is essential to strike the right balance between normalisation and denormalisation.

To further enhance my skill in data normalisation, I intend to continue exploring real-world datasets to use as practise. Once confident, I will research advanced normalisation techniques, such as Boyce-Codd normal form and higher normal forms, to gain a more comprehensive understanding of database design. Applying these principles in future data science projects will help me create more advanced databases that are robust and efficient.

Web Scraping

The web scraping activity required the development of a Python script to extract data containing the keyphrase ‘data science’ from a website and store it in a json file. I have technical skills in Python but lack experience in web scraping techniques. Initially I was confident that I could easily complete this task however I quickly became frustrated by errors appearing in the code.

Through trial and error and some technical improvement, I was able to complete the activity and successfully scrape data from a website. This activity assisted me in learning key lessons on how to adapt and overcome unforeseen challenges in writing python scripts. I will continue to refine my error handling mechanisms and learn about techniques to scrape websites with dynamic content.

Conclusion

The DBD module provided me with the opportunity to build on the skill from previous modules and highlight key areas of improvement that I will endeavour to address in the future. Working with my peers facilitated the opportunity to learn from their combined experience.

References

Brown, M. E., & Dutton, J. E. (1995). The limits of attributability: Responsibility for poor performance in an organizational setting. Administrative Science Quarterly, 40(3), 403-427.

Dweck, C. S. (2006). Mindset: The New Psychology of Success. Random House.

Eisenberg, J., Post, C. and DiTomaso, N. (2019) 'Team dispersion and performance: The Role of Team Communication and Transformational Leadership', Small Group Research, 50(3), pp. 348–380. doi:10.1177/1046496419827376.

Elmasri, R., & Navathe, S. B. (2019). Fundamentals of database systems (8th ed.). Pearson.

Husebø, S.E., O'Regan, S. and Nestel, D. (2015) "Reflective practice and its role in simulation," *Clinical Simulation in Nursing*, 11(8), pp. 368–375. Available at: https://doi.org/10.1016/j.ecns.2015.04.005.

Marlow, S., Lacerenza, C. and Salas, E. (2017) 'Communication in virtual teams: A conceptual framework and research agenda', Human Resource Management Review, 27(4), pp. 575-589. doi: 10.1016/j.hrmr.2016.12.005

Ramakrishnan, R., & Gehrke, J. (2003). Database management systems (3rd ed.). McGraw-Hill.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2010). Database system concepts (6th ed.). McGraw-Hill.

UoE (1970) *Gibbs' reflective cycle*, *The University of Edinburgh*. Available at: https://www.ed.ac.uk/reflection/reflectors-toolkit/reflecting-on-experience/gibbs-reflective-cycle#:~:text=One%20of%20the%20most%20famous,analysis%2C%20conclusion%20and%20action%20plan. (Accessed: April 4, 2023).