**Reflection on Learning and Individual Contributions**

**Unit 1: Introduction to Big Data**  
This unit was a great introduction to the course, covering the fundamentals of big data, including the concepts of Volume, Variety, Velocity, and Veracity, which were new to me. I also learned about data cleansing, standardization, and normalization techniques. The distinction between structured and unstructured data, along with the various data formats (human-readable like Excel and PDF, and non-human-readable like JSON and XML), broadened my understanding of how data is used in real-world applications.

**Unit 2: Data Wrangling**  
In this unit, I strengthened my ability to identify and address data-wrangling challenges. I learned to recognize security risks, limitations, and potential opportunities when working with large datasets. The critical thinking skills I developed allowed me to choose appropriate tools and techniques for preparing, cleaning, exploring, creating, optimizing, and evaluating big data. Additionally, working in a virtual team setting gave me experience in collaboration, adapting to team roles, and understanding organizational dynamics.

**Unit 3: Web Scraping**  
Unit 3 presented challenges, particularly with the task of web scraping. As this was a new skill for me, it was tough, especially considering many websites have measures to block scrapers. This made completing the task more difficult than anticipated, but it provided a valuable learning experience about the complexity of data extraction from the web.

**Unit 4: Data Cleaning and Transformation**  
I gained a deeper understanding of data cleaning and transformation techniques, essential for managing large datasets. I learned about the data management pipeline and the factors affecting data cleaning. The unit also covered the requirements for designing and automating data cleaning processes, and by the end, I was able to clean and transform data effectively using these techniques.

**Unit 5: UNICEF Data Set Activity**  
The activity in this unit focused on cleaning and converting a UNICEF dataset. I successfully converted the SAV file to CSV, though extracting and organizing the headers was difficult. Despite the challenge, I persevered and gained valuable experience in handling real-world datasets, highlighting the complexities of data preprocessing.

**Unit 6: Database Design**  
Unit 6 was more familiar, as I already had experience with databases at work. I am responsible for creating and maintaining databases using PostgreSQL and Snowflake. However, the main focus of this unit was the group project. I contributed by presenting ideas and solutions, and by overseeing the logical structure of the project, including code and visuals. I was also in charge of formatting the document and adding visuals to the appendices, ensuring the content was clear and well-presented.

**Unit 7: Data Normalization**  
This unit was particularly practical for me, as I frequently work with unorganized data at my job. I used SQL, which I am comfortable with, to create normalized tables and organize the data. I first created an Excel file for the normalization activity and then used SQL to create tables in the third normal form (3NF). This hands-on experience reinforced my ability to apply SQL for data structuring tasks.

**Unit 8: Compliance Frameworks**  
In Unit 8, I gained insights into data compliance frameworks, focusing on regulations like GDPR. The comparison between GDPR and U.S. data protection laws helped me understand how compliance obligations affect data storage and handling. The unit emphasized the importance of ensuring data privacy, security, and accountability in real-world scenarios, which is crucial for any data professional.

**Unit 9: Database Management Systems (DBMS)**  
This unit deepened my understanding of various DBMS, both relational (SQL, PostgreSQL, Oracle) and non-relational (MongoDB, SAP). I also learned about modern tools like Hadoop and Data Lakes. The unit covered database design and security concerns, helping me understand how to evaluate DBMS options and select the best system for large datasets.

**Unit 10: APIs and Security**  
Unit 10 focused on the importance of APIs in data parsing and communication. I worked with the OpenWeatherMap API, gaining experience in managing authentication risks and ensuring data integrity. I also learned about API security measures like key management, rate limiting, and secure data storage. This experience enhanced my ability to manage API security in real-world applications.

**Unit 11: Executive Summary for Individual Project**  
In Unit 11, I focused on drafting the Executive Summary for my Individual Project, which involved creating a logical database for a client organization. I summarized the project findings and recommendations for a non-technical audience. I also analyzed SQL and No-SQL DBMS options, discussed GDPR compliance, and prioritized recommendations to meet business needs. This unit helped me improve my ability to convey complex information in a clear, concise manner.

**Unit 12: Machine Learning and Data Analytics**  
Unit 12 introduced me to machine learning and its role in big data analytics. I learned about emerging trends and how compliance frameworks protect individual data. Most of my time, however, was spent finalizing my e-portfolio and reflection. Writing a reflection on my own work was challenging, and building the e-portfolio was a new experience. Despite the learning curve, I successfully organized and uploaded the content to GitHub.

Throughout the course, I developed a range of technical skills, including data extraction, cleaning, and transformation. The group project in Unit 6 was an opportunity to apply these skills in a team setting. I contributed by presenting ideas, providing logical insights, and ensuring the project’s documentation and visuals were clear and accessible. My experience in database management, particularly with SQL and Snowflake, played a crucial role in the project’s success. I also learned to work collaboratively, adapting to different team roles and supporting others with their tasks.

Working on this project made me realize the importance of clear communication and effective collaboration in team activities. I found that my ability to adapt to new tools and technologies improved as the course progressed. I also developed greater confidence in my ability to apply these skills in real-world scenarios, which has strengthened my approach to data management and database design.

This module has been instrumental in expanding my knowledge and skills in data science, particularly in areas like data wrangling, database design, and compliance frameworks. My contributions to the group project, alongside my individual learning, have prepared me to handle complex data challenges and make informed decisions in data management. The skills I’ve gained will be valuable in my current and future work, and I’m excited to continue applying them in real-world projects.