**Data Management Competency Assessment**

**1. Introduction to Data Management**

In the MS-CISBA program, I completed two key courses that deepened my knowledge of data organization, querying, and analysis:

* **CIDM 6350 – Data and Information Management (Dr. Lazrig, Summer 2022)**
* **CIDM 6355 – Data Mining Methods (Dr. Chen, Fall 2022)**
* **CIDM 6351 – Business Data ETL (Dr Hupherys, Fall 2024)**

CIDM 6350 focused on relational database design, normalization, data integrity, and SQL-based data manipulation using MySQL. CIDM 6355 emphasized analytical thinking and data mining methods using tools like RapidMiner and R. These courses laid a critical foundation for how I model, manage, and analyze business data.

**2. What I Know (Strengths)**

My major strengths from these courses include:

* **Entity Relationship Modeling**: I learned to design ER diagrams and translate business rules into relational models. CIDM 6350 taught me how to organize data efficiently and logically for scalable applications.
* **SQL Proficiency**: I built and queried databases using SQL DDL and DML, practiced complex joins, and applied filters and aggregations. I used MySQL Workbench and SQLFiddle for hands-on tasks.
* **Normalization and Data Integrity**: Understanding functional dependencies and how to normalize tables into 2NF and 3NF helped me reduce redundancy and enforce consistency.
* **Data Mining Techniques**: In CIDM 6355, I applied algorithms such as decision trees, logistic regression, clustering, and association rules using RapidMiner and R, and explored real-world data sets.
* **OLAP and Warehousing Concepts**: I became familiar with concepts like cubes, rollups, and data marts, which help structure data for analytics and reporting purposes.

**3. Where I Am Weak**

Some concepts and tasks were more challenging:

* **Advanced SQL Optimization**: While I can write functional queries, optimizing them for performance and scalability in large datasets is still a skill I am developing.
* **Text Mining**: CIDM 6355 briefly introduced text analytics, but I would like to go deeper into sentiment analysis and unstructured data modeling.
* **• Building Automated Pipelines:** Although touched on in CIDM 6351, I’m still learning how to fully automate ETL processes and integrate data into production workflows. I have been exposed to the concepts of using Apache Airflow for workflow orchestration, but have not yet built a full production pipeline with it.

**4. What I Wish I Knew**

Looking back, I would have benefited from:

* **Cloud-Based Data Architecture**: I wish I had learned more about deploying databases using cloud platforms like AWS RDS or BigQuery.
* **Real-Time Data Streaming**: Incorporating tools like Kafka or Spark Streaming into my learning would round out my understanding of dynamic systems.

**5. Supporting Evidence**

Some key examples of my work include:

* **ER Diagram and Schema Assignment** (CIDM 6350): Designed a normalized relational model from scratch using Draw.io.
* **SQL Query Projects**: Built and queried a MySQL database with DDL and DML statements, performing data transformations.
* **RapidMiner Labs** (CIDM 6355): Created workflows for classification and clustering models using historical customer data.
* **Group Data Mining Project**: Applied association rule mining to retail data, submitted both a written report and a recorded presentation.
* **Python Web Scraping Assignment** (CIDM 6351): Extracted structured web data using Python requests, response, and BeautifulSoup libraries. This knowledge later enabled me to build the data scraping portion of my capstone prototype.
* **Data Dictionary and Field Parsing** (CIDM 6355): Learned to define data fields precisely and parse composite fields into multiple structured columns. I later applied these principles when structuring the bank bonus dataset for my capstone project, including separating fields like “monthly fees” and “how to avoid fees.”

I plan to include these artifacts in my GitHub portfolio under a DataManagement directory for Capstone documentation.

**6. Capstone Readiness and Integration**

The skills I’ve gained in data modeling, querying, and mining are foundational to my Capstone:

* I will use **normalized schemas and SQL queries** to structure the back-end of my prototype.
* I will apply **data mining algorithms** to derive insights from operational data.
* My understanding of **data warehouses and marts** will inform how I structure the analytical output layer.

This area complements:

* **Data Analytics**: Well-managed data enables more accurate, useful analytics.
* **Software Systems**: A clean, reliable data model supports scalable system design.
* **Cybersecurity**: Access control, data integrity, and transactional safety are essential in securing information.

**7. Conclusion**

I am confident in my ability to design, manage, and analyze data at an intermediate to advanced level. I have demonstrated mastery of core database principles, supplemented with mining techniques for added insight. I recognize opportunities to grow in cloud, NoSQL, and pipeline automation areas, and plan to continue expanding my expertise.