Capstone Assessment: Data Management

*Masyn Grisel 2/23/25*

1. Competencies and Strengths (What I Know)

My coursework in Data Engineering (CIDM 6351) and Data Mining Methods (CIDM 6355) has provided me with a solid foundation in data extraction, transformation, database management, and data mining techniques. I have developed strong skills in working with structured and unstructured data, applying SQL, Python, and R to manage and analyze datasets.

Key Skills and Knowledge Areas:

* Data Engineering & ETL Processes
  + Extracting, cleaning, normalizing, and transforming data from multiple sources.
  + Creating ETL pipelines for structured and unstructured data.
* Database Management & SQL
  + Writing advanced SQL queries for data retrieval, joins, aggregations, and indexing.
  + Managing relational databases for business intelligence and reporting.
* Data Mining & Machine Learning Techniques
  + Implementing classification, clustering, and association rule mining techniques.
  + Applying sampling techniques for business intelligence applications.
* Online Analytical Processing (OLAP)
  + Using OLAP cubes to analyze multi-dimensional data efficiently.
  + Creating data warehouses to store and process large datasets.
* Programming in Python and R
  + Using Python libraries (Pandas, NumPy, Scikit-learn) for data processing.
  + Utilizing R for statistical analysis and visualization.

Examples of Work:

* Built an ETL pipeline to extract and clean data from external sources.
* Applied classification models (e.g., decision trees, logistic regression) to categorize data.
* Designed SQL queries for data mining and business intelligence reporting.
* Implemented sampling methods to improve the accuracy of predictive models.

2. Weaknesses and Areas for Improvement (Where I Am Weak)

While I have a strong foundation in data engineering and mining techniques, I need to improve in the following areas:

* Advanced Machine Learning for Data Mining – I need more experience with deep learning techniques, random forests, support vector machines (SVMs), and ensemble methods.
* Big Data & NoSQL Databases – My experience is primarily with relational databases, and I have not worked extensively with NoSQL databases like MongoDB, Cassandra, or Hadoop.
* Optimization of SQL Queries & Storage Efficiency – I need to improve my ability to optimize SQL queries using indexing, partitioning, and query tuning for faster execution on large datasets.

3. Gaps in Knowledge (What I Wish I Knew)

There are several key areas I wish I had learned in more depth:

1. Cloud-Based Data Storage & Processing – I want to gain experience using AWS Redshift, Google BigQuery, and Azure SQL for handling large-scale data storage and retrieval.
2. Real-Time Data Processing & Streaming – I wish I had worked with technologies like Apache Kafka, Spark Streaming, or Flink to process data in real-time.

4. Knowledge Sources & Examples of Work

To improve my skills and proficiency, I have utilized the following resources:

* Course materials from CIDM 6351 & CIDM 6355
* Python and R documentation for data processing and statistical modeling
* SQL tutorials and database optimization guides
* Hands-on data mining projects in Python (Scikit-learn, Pandas)
* DataCamp for tutorials in Python, SQL, and R
* Youtube for a full on Python beginner tutorial

5. Summary Statement & Preparedness

My coursework in Data Engineering and Data Mining has equipped me with strong skills in ETL processes, SQL database management, data mining, and business intelligence techniques. I have successfully applied Python, R, and SQL to manage and analyze data, build predictive models, and support business decision-making.

However, to fully prepare for real-world data management challenges, I need to deepen my knowledge in cloud-based data platforms, real-time data processing, and advanced machine learning techniques. Moving forward, I plan to enhance my expertise in big data technologies and database optimization to apply these skills effectively in business intelligence and analytics projects, including my Capstone prototype.