**Capstone Project 2: Data Modeling**

**Assignment Type: Individual**

**Submission Format: Report + SQL Scripts + ERD Diagram**

**Due Date: 27 April 2025 23:59**

**Objective:**

The goal of this project is to design and implement a structured data model based on the dataset collected in **Capstone Project 1**. You will create a **conceptual, logical, and physical data model**, ensuring the data is structured efficiently for storage and retrieval. This project will also introduce database design principles and SQL querying.

**Project Guidelines:**

**Step 1: Understanding the Dataset**

* Review the dataset collected in Capstone Project 1.
* Identify key data components such as entities, attributes, and relationships.
* Determine the type of database system that best fits the data (Relational vs. Non-relational).

**Deliverable:** A short summary of the dataset structure and its main attributes.

**Step 2: Conceptual Data Modeling**

* Define key **entities, attributes, and relationships** based on the dataset.
* Identify the **primary entities** (e.g., customers, transactions, products, etc.).
* Establish relationships between entities (one-to-one, one-to-many, many-to-many).
* Create a **Conceptual Entity-Relationship Diagram (ERD)**.

**Deliverable:**

* **Conceptual ERD** showing entities and relationships. At least 8 entities.
* **Explanation of why this model is suitable for the case study**.

**Step 3: Logical Data Modeling**

* Translate the conceptual model into a **logical schema**.
* Define entity attributes, data types, and constraints (e.g., primary keys, foreign keys, unique constraints).
* Apply **Normalization (3NF)** to eliminate data redundancy and improve efficiency.
* Update the **ERD** with attributes and relationships.

**Deliverable:**

* **Logical ERD** in **3NF** with attributes, keys, and relationships.

**Step 4: Physical Data Modeling & Database Schema**

* Convert the logical model into a **physical database schema**.
* Choose a **Database Management System (DBMS)** (e.g., MySQL, PostgreSQL, MongoDB, SQLite).
* Create database tables using **SQL scripts**.
* Define **indexes, constraints, and relationships** in the database schema.
* Populate tables with your data collected from Capstone project 1.

**Deliverable:**

* **Physical ERD**
* **SQL Scripts for table creation.**
* **Screenshots of Data Inserted.**

**Step 5: Querying the Database**

* Write and execute **basic SQL queries** to interact with the database:
  + Retrieving data using SELECT
  + Filtering data with WHERE
  + Sorting data using ORDER BY
  + Using joins (INNER JOIN, LEFT JOIN, etc.)
  + Aggregations (COUNT, SUM, AVG, etc.)
* Easy query:
  + **Easy Query**: Create a query that uses two tables. It should involve basic filtering, grouping, and aggregation such as SUM, COUNT, or AVG. These types of queries are useful for simple KPIs or visualizations like total sales, customer counts, or daily averages.
  + **Medium Query**: Create a query that joins 3 or more tables. Include conditions such as date filters or top-n results. This query should be able to support dashboard visuals like bar charts, top 5 lists, or trend lines filtered by category. Combine filtering by one of the tables + aggregation + sorting + string function.
  + **Difficult Query**: Create a query that uses four or more joined tables with Combination of filtering by one table + aggregation + sorting + string function + date function + nested query.

**Deliverable:**

* **At least 5 SQL queries with their results. With at least 2 Easy, 2 Medium and 1 Difficult.**
* **Brief explanation of what each query does.**

**Submission Requirements**

Your final submission should include:

1. **Dataset Summary (Step 1)**
2. **Conceptual Model (ERD + Explanation) (Step 2)**
3. **Logical Model (Normalized ERD) (Step 3)**
4. **Physical Model (SQL Schema + Data) (Step 4)**
5. **SQL Queries & Execution Results (Step 5)**

**Evaluation Criteria (Rubric)**

| **Category** | **Points** | **Description** |
| --- | --- | --- |
| Dataset Understanding | 15 | Clarity in defining dataset structure. |
| Conceptual Data Model | 15 | Proper identification of entities and relationships. At least 8 entities. |
| Logical Data Model | 20 | Correct normalization and ERD refinement. |
| Physical Data Model | 20 | Correct implementation of tables, keys, and constraints. |
| SQL Queries & Execution | 30 | Effective queries demonstrating data retrieval. At least 5 queries with at least 4 marks each (no result shown, no marks). With good reason. |

**Total:** 100 points

**Tools & Resources**

* **Database Options:** MySQL, PostgreSQL, SQLite, MongoDB (for NoSQL)
* **SQL Tools:** MySQL Workbench, pgAdmin, SQLite Browser
* **ERD Tools:** Draw.io, Lucidchart, Microsoft Visio, dbdiagram.io
* **Python Libraries (Optional for Data Handling):** Pandas, SQLAlchemy