

## Project – Stage 1

### Requirement Engineering and Database Design

#### Objective

The objective of Stage 1 is to introduce you to Requirement Engineering and Entity-Relationship Diagram (ERD) design. You are required to select a raw CSV dataset from platforms like Kaggle and proceed to identify use cases associated with the dataset. You will then convert these use cases into an Entity-Relationship Diagram. You will be using this ERD in the next stages of the project where you are required to work with database creation with PostgreSQL.

#### Stage 1 Tasks:

##### Dataset Selection:

- Pick a dataset that piques your interest from Kaggle: <https://www.kaggle.com/datasets>
- Ensure the dataset is available in CSV format for ease of importing into PostgreSQL.

##### Requirement Engineering:

- Identify the use cases associated with the selected dataset. Use cases should represent real-world scenarios or operations that the database is expected to support. How do they relate to the chosen domain? How can they be interconnected? Write down your observations.
- **A minimum of 15 Relations is to be present in the ERD.** A Use Case representing the relationship being read by both sides of the entities mentioning the cardinality and the participation of each entity in the relationship should be documented.
- **A minimum of two “at least”, two “at most”, and two “exactly one” relationships are to be present.**
- Document each use case using the following template (similar to the format of Assignment 2):
  - **Use Case Name:** [Name of the use case]
  - **Description:** [Detailed description of the use case mentioning the cardinality and the participation of the entities involved.]

##### Entity-Relationship Diagram (ERD) Design:

- Take the knowledge gained from understanding the use cases and craft an ERD Diagram using a tool of your choice.
- **Entities:** Identify the main entities in your dataset.
- **Attributes:** List the attributes associated with each entity and specify their keys.
- **Relationships:** Define relationships between entities. Explain how they connect based on the identified use cases. Specify cardinalities. Specify participations (total and partial participation).

### Write-Up:

- **Explore a Current Topic:** Find a recent computer science-based issue, development, or trend related to society in the current environment. Investigate its impact and significance within the computer science field and potentially even on a broader societal level. The issue can be anything from dataset analysis to design – it is your choice as long as it pertains to the course material.
- **Evaluate & Choose:** Once you've identified a relevant issue, delve deeper. Explore and assess the different existing approaches or solutions related to this issue. Based on your evaluation, choose one option that you believe holds the most promise and incorporate it to the tasks at hand.
- **Justify Your Choice:** Clearly explain your decision for choosing a specific option. Be sure to present a solid and well-reasoned argument explaining why you believe this option is the most effective or impactful in addressing the chosen issue.

### Submission Requirements:

- Submit a PDF document containing:
  - PART 1: Chosen dataset details and reference link.
  - PART 2: Use case descriptions for each identified use case, with emphasis on elucidating their relevance to the domain and their interconnections.
  - PART 3: ERDs for each use case, with clear explanations of the relationships based on the identified use cases.
  - PART 4: Write-Up investigating a recent computer science issue related to our course content, evaluate existing approaches and options, and justify the selection of the most promising option.

### Evaluation Criteria:

- Dataset relevance.
- Clarity and completeness of use case descriptions.
- Depth of understanding and linkage between identified use cases.
- Accuracy and completeness of the ERD design.
- Relevance of the issue identified and reasoning of the chosen option.

**Important Notes:**

- This is an **individual** assignment.
- **Academic Integrity Violation:**
  - In this open project environment, it is strictly prohibited to substantially duplicate ideas or code from other projects. If we find any, we'll have to check for possible academic misconduct.
  - Be careful not to copy code from the internet. If you do, it could mean failing the project and facing the academic integrity committee.
- Reach out to the instructor or any of the TA's for any clarification or guidance.
- Ensure **proper citation** if external sources are used.

**Deadline for Stage 1 Submission: March 22<sup>nd</sup>, 2024**