FACULTY OF ENGINEERING AND BASIC SCIENCES
ACADEMIC PROGRAM: DATA ENGINEERING AND ARTIFICIAL INTELLIGENCE

COURSE: ETL (G01)
ETL-Project: First Delivery

✅ Getting Started

In this first delivery, you will demonstrate your knowledge in data management, data architecture, and visualizations. Your task will begin with gathering project requirements and evaluating potential data sources. Based on this, you will select a suitable dataset (minimum 10,000 rows and at least 10 features) to conduct exploratory data analysis (EDA) and create insightful visualizations.

You will design a simple data architecture, choose the appropriate technological stack, and migrate the dataset into a database. Your EDA and visualizations must be based on data stored in the database (not the original CSV file).

You can start coding from scratch, and the technologies we expect to evaluate are described in the technologies section.

✅ What is Expected

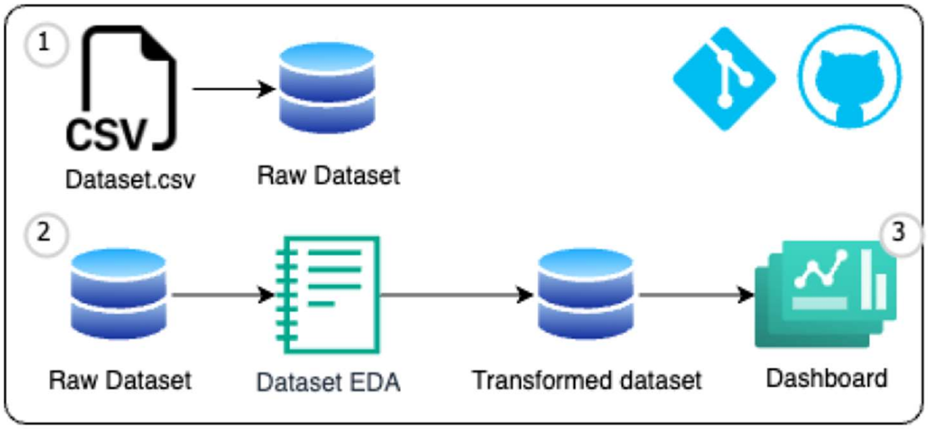
- **Requirements Gathering**: Define the objectives and needs of the data analysis task.
- **Data Source Evaluation**: Justify your choice of dataset based on availability, relevance, structure, and quality. It is expected that you get the CSV file of the dataset you choose and create an ETL application to migrate the data to a relational database.
- **Architecture Design**: Propose a basic ETL architecture (e.g., data ingestion, storage, transformation, visualization, and Data Warehouse).
- **Data Model**: Your data model design structures how your data is stored and accessed(e.g., star schema, snowflake schema).
- **Technology Stack Selection**: Justify the tools chosen for each layer of your architecture (e.g., database, visualization, transformation).
- **Data Migration**: Load the dataset into a relational database (e.g., PostgreSQL, MySQL).
- **Visualizations and reports**: Build meaningful charts, dashboards and/or reports using the cleaned dataset from the database.
- **Final Report**: Include a brief technical document that explains:
  - The dataset you selected and why
  - Requirements and evaluation criteria
  - The architecture and tools you chose
  - Summary of the EDA process and the visualizations produced

✅ Technologies

- Python
- Jupiter Notebook
- Database (you choose)

- Visualizations
- Git - GitHub

✅ Diagram



📁 Evaluation

Item	GitHub Repo	Readme	Gitignore	Migration of data to the database	Data Analysis	Extracting data from the database	Visualizations	Report	Presentation
Weight									