**Project Title:**  
**Data Cleaning and Preprocessing Using MySQL**

**Project Description:**  
This project focuses on cleaning and preprocessing raw data using MySQL to ensure it is accurate, consistent, and ready for analysis. The dataset used contains information regarding corporate layoffs during 2022. It has fields like Company, location, industry, total\_laid\_off, percentage\_laid\_off, date, stage, country, funds\_raised\_millions which initially included missing values, duplicates, inconsistent formats, and irrelevant entries.

The primary goal of the project was to perform data cleaning process using SQL queries thereby enhancing data quality and ensuring reliable downstream analytics.

**Key Objectives:**

* Import raw data into MySQL from CSV.
* Identify and handle missing or null values.
* Remove duplicate records.
* Standardize inconsistent data formats (e.g., dates).
* Detect and correct outliers or invalid entries.

**Tools & Technologies:**

* MySQL 8.x
* SQL Workbench
* CSV file imports using LOAD DATA INFILE

Data Source: https://www.kaggle.com/datasets/swaptr/layoffs-2022

**Outcome:**  
The cleaned dataset is now free from inconsistencies and errors, enabling reliable reporting, analysis, and machine learning applications. The SQL scripts developed can be reused for periodic data cleaning tasks and serve as a foundational step in the data pipeline.

**Project Steps:**

1. Import the data in CSV format
2. Convert CSV to JSON (which fixed many data import issues)
3. Import the table data to database by “Table data import wizard” in MySQL workbench
4. Check for duplicates and remove them
5. Standardize data and fix errors
6. Check null values and handle correctly
7. Remove any columns and rows that are not necessary - few ways