Data Cleaning with SQL

This repository showcases practical applications of SQL for data cleaning and exploration tasks. It focuses on a real-world dataset related to layoffs sourced from publicly available CSV files.

Project Overview

The project demonstrates skills in:

Data Acquisition

- Identify and source relevant public datasets for analysis.

Data Cleaning

- Utilize SQL statements to identify and address inconsistencies, missing values, and formatting errors within the layoff data. Techniques employed may include:
    - Handling missing values (imputation or deletion)
    - Standardizing data formats (dates, currencies, etc.)
    - Removing duplicates or outliers
    - Data validation and filtering

Data Exploration

- Perform exploratory data analysis (EDA) on the cleaned dataset to derive meaningful insights and visualizations.

Steps to Reproduce

Database Creation

- First, create a database called `World_layoffs`, where we will import the raw data from `layoffs.csv`.

Data Cleaning

- Remove duplicates
- Standardize the data
- Remove null or blank values
- Remove any columns that are not relevant

Data Exploration

- Analyze the cleaned data to uncover trends, patterns, and insights.
- Utilize various SQL queries to explore different aspects of the dataset.