DATASET: HR ANALYTICS – EMPLOYEE MANAGEMENT SYSTEM

ABOUT THE DATASET

This dataset simulates a real-world HR management system containing employee, salary, attendance, and performance data.

It is intentionally designed with messy and inconsistent data to help learners practice data cleaning, transformation, and analysis using SQL.

OBJECTIVE

The goal of this dataset is to:

- Practice data cleaning techniques
- Handle real-world messy data scenarios
- Build end-to-end SQL project skills
- Prepare for data analyst interviews

DATASET OVERVIEW

The dataset consists of multiple related tables:

- Employees
- Departments
- Salaries
- Attendance
- Performance

Each table represents a different aspect of an organization's HR system.

TABLE SCHEMA

1. Employees

Column Name	Description
emp_id	Unique employee ID
emp_name	Employee name (may contain inconsistencies)
age	Employee age
city	Employee location
dept_id	Department ID
hire_date	Date of joining

2. Departments

Column Name	Description
dept_id	Unique department ID
dept_name	Department name

3. Salaries

Column Name	Description
salary_id	Unique salary record ID
emp_id	Employee ID
salary	Salary amount
salary_date	Salary record date

4. Attendance

Column Name	Description
attendance_id	Unique attendance ID
emp_id	Employee ID
attendance_date	Date
status	Present/Absent

5. Performance

Column Name	Description
emp_id	Employee ID
rating_2022	Performance rating (2022)
rating_2023	Performance rating (2023)
rating_2024	Performance rating (2024)

DATA QUALITY ISSUES

This dataset includes multiple real-world data problems:

- **NULL and empty values**
- **Duplicate records**
- **Inconsistent text (case & spacing issues)**
- **Invalid values (negative salary, unrealistic age)**
- **Outliers**
- **Inconsistent date formats**
- **Data type issues**
- **Referential integrity issues**
- **Missing values across tables**
- **Mixed formatting**

SKILLS YOU WILL LEARN

By working with this dataset, learners will gain:

Data Cleaning (SQL)
Data Transformation
Data Validation
Business Logic Implementation
Real-world Problem Solving
Portfolio Building

USE CASES

This dataset can be used for:

- **SQL practice projects**
- **Data cleaning challenges**
- **Interview preparation**
- **Portfolio projects**

DISCLAIMER

This dataset is completely synthetic and created for educational purposes only.
Any resemblance to real individuals or organizations is purely coincidental.