



DATA MANAGEMENT Assignment 1 - 2

Students:

- Andrea Potì 2008416
- Giulio D'Erasmo 1859130

The Dataset: Human Resource

HR: Analytics

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement
JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked
Over18	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StandardHours	StockOptionLevel
TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager

HR: Data Generator

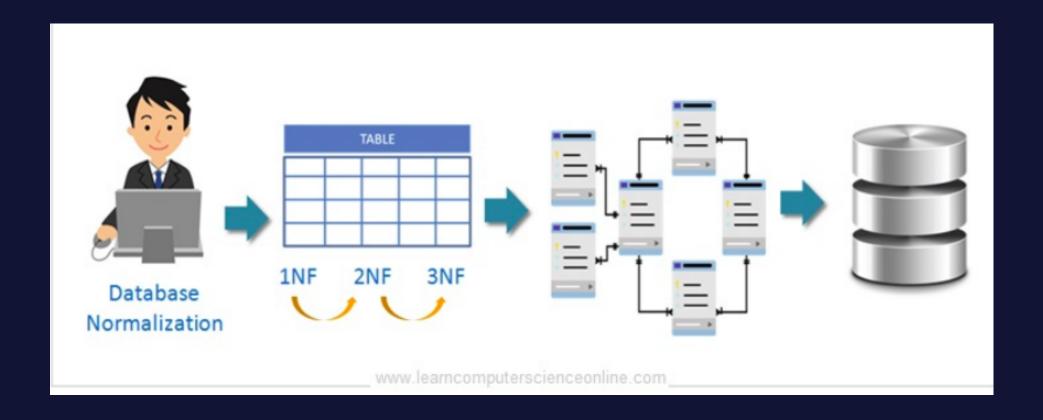
Emp ID	Name Prefix	First Name	Middle Initial	Last Name	Gender	E Mail
Father's Name	Mother's Name	Mother's Maiden Name	Date of Birth	Time of Birth	Age in Yrs.	Weight in Kgs.
Date of Joining	Quarter of Joining	Half of Joining	Year of Joining	Month of Joining	Month Name of Joining	Short Month
Day of Joining	DOW of Joining	Short DOW	Age in Company (Years)	Salary	Last % Hike	SSN
Phone No.	Place Name	County	City	State	Zip	Region
User Name	Password					

GOALS:

Avoid redundancy of data:

NORMALIZATION

- Constrain:
 - Primary key;
 - Foreign key.
- Optimized data types
- Data preprocessing



Address

id_emp int

country varchar(50)

city varchar(50)

state character(2)

zip int

ssn character(12)

region varchar(50)

Login

id_emp int
username varchar(50)
password varchar(50)

Education

id_emp int
education_no int
educational_field varchar(20)

ExtraInfoEmployee

id_emp int fathers_name varchar(50) mothers_name varchar(50) mothers_last_name varchar(50) weight int marital_status varchar(20) number_company_worked int total_working_years int work_life_balance int distance_from_home int int years_in_company

SatisfactionAndProductivity

id_emp int
hire_id int
job_satisfaction int
enviroment_satisfaction int
job_involment int
business_travel varchar(100)

performance_rating int

Employee

id int first_name varchar(20) middle_initial varchar(20) last_name varchar(20) gender varchar(5) email varchar(200) age int phone_no character(12) date_of_birth datetime

Hire

id_emp int

hire_id int

department varchar(40)

job_role varchar(40)

jobLevel int

attrition varchar(3)

hire_date date

Salary

id_emp int
hire_id int
daily_rate int
hourly_rate int
monthly_rate int
monthly_income int
percent_salary_hike int

ExtraInfoHire

hire_id int
quarter_of_joining character(2)
half_of_joining varchar(5)
year_of_joining int
month_of_joining varchar(10)
day_of_joining int
day_name varchar(10)

```
CREATE TABLE `Employee` (
    `id` int PRIMARY KEY,
    `first_name` varchar(20),
    `middle_initial` varchar(20),
    `last_name` varchar(20),
    `gender` varchar(5),
    `email` varchar(200),
    `age` int,
    `phone_no` character(12),
    `date_of_birth` datetime
);
```



HOW?

- DDL Statement
- Primary Key +
 Foreign Key
- Special syntax to import CSV file

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Data/hr_database/Employee.csv'
INTO TABLE Employee
FIELDS TERMINATED BY ','
IGNORE 1 ROWS;
```