**Data Cleaning SQL File**

use world\_layoffs;

create table layoff\_stage like layoffs;

INSERT into layoff\_stage

select \* from layoffs;

--- Data Cleaning involves three steps:

--- 1. Data Duplication

--- 2. Data Standardization

--- 3. Removing Null Values

--- 4. Removing unnecessary columns

**1.Removing Data Duplicates**

select \* from layoff\_stage where company Like 'Yahoo';

-- WITH data\_duplicate AS

-- (Select \* , ROW\_NUMBER()

-- OVER(Partition by location,industry,total\_laid\_off,percentage\_laid\_off,date,stage,country,funds\_raised\_millions) AS row\_num

-- from layoff\_stage)

CREATE TABLE `layoff\_stage2` (

`company` varchar(29) NOT NULL,

`location` varchar(16) NOT NULL,

`industry` varchar(15) DEFAULT NULL,

`total\_laid\_off` int DEFAULT NULL,

`percentage\_laid\_off` decimal(6,4) DEFAULT NULL,

`date` date DEFAULT NULL,

`stage` varchar(14) DEFAULT NULL,

`country` varchar(20) NOT NULL,

`funds\_raised\_millions` decimal(10,4) DEFAULT NULL,

`row\_num` INT

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

Insert into layoff\_stage2

Select \* , ROW\_NUMBER()

OVER(Partition by location,industry,total\_laid\_off,percentage\_laid\_off,date,stage,country,funds\_raised\_millions) AS row\_num

from layoff\_stage;

Select \* from layoff\_stage2 where row\_num>1;

delete from layoff\_stage2 where row\_num>1;

**2. Standardizing Data**

UPDATE layoff\_stage2 SET company = TRIM(company);

SELECT Industry from layoff\_stage2 where industry like 'Crypto%';

UPDATE layoff\_stage2 SET INDUSTRY = "Crypto" where industry like 'Crypto%';

SELECT DISTINCT(Industry) from layoff\_stage2;

SELECT DISTINCT(Location) from layoff\_stage2;

SELECT DISTINCT(Country) from layoff\_stage2 order by Country;

UPDATE layoff\_stage2 SET Country = "United States" where Country like 'United S%';

SELECT (Date) from layoff\_stage2 ;

**3. Removing Null Values**

Select \* from layoff\_stage2 where company is null;

Select \* from layoff\_stage2 where industry is null;

Select \* from layoff\_stage2 where total\_laid\_off is null and percentage\_laid\_off is null;

Select \* from layoff\_stage2 where stage is null or stage = 'unknown';

Select \* from layoff\_stage2 where funds\_raised\_millions is null;

Select \* from layoff\_stage2 as l1

Join layoff\_stage2 as l2

on l2.company=l1.company and l1.location=l2.location

where (l1.industry is NULL or l1.industry = "") and l2.industry is not null;

Update layoff\_stage2 as l1

Join layoff\_stage2 as l2

on l2.company=l1.company and l1.location=l2.location

SET l1.industry = l2.industry

where (l1.industry is NULL or l1.industry = "") and l2.industry is not null;

**4. Deleting unnecessary rows**

Delete from layoff\_stage2 where total\_laid\_off is null and percentage\_laid\_off is null;

ALTER Table layoff\_stage2 drop column row\_num;