**Data Cleaning Steps**

1. Creating Another table instead of using Raw table:

Create table layoffscleaning

LIKE layoffs;

1. Inserting values from original raw table to new table:

INSERT INTO layoffscleaning

Select \* From layoffs;

1. Finding Duplicates

WITH DUPLICATECTE AS

(

Select \*, Row\_Number() Over(Partition BY company, location, industry, total\_laid\_off, percentage\_laid\_off, `date`, stage, country, funds\_raised\_millions) As row\_num

From layoffscleaning

)

SELECT \* FROM DUPLICATECTE

WHERE ROW\_NUM>1

1. Creating Another Table with row\_num column so that we can delete the duplicate

CREATE TABLE `layoffscleaning2` (

`company` text,

`location` text,

`industry` text,

`total\_laid\_off` int DEFAULT NULL,

`percentage\_laid\_off` text,

`date` text,

`stage` text,

`country` text,

`funds\_raised\_millions` int DEFAULT NULL,

`row\_num` int

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

Above query will create blank table called layoffscleaning2

1. Now we want to insert the values from the table created earlier to the third table with row\_num column

Insert into layoffscleaning2

Select \*, Row\_Number() Over(Partition BY company, location, industry, total\_laid\_off, percentage\_laid\_off, `date`, stage, country, funds\_raised\_millions)

As row\_num

From layoffscleaning

1. Following query is used to find the duplicate from the new table

Select \* From layoffscleaning2

Where row\_num>1

1. Deleting duplicate

Delete From layoffscleaning2

Where row\_num>1

**Standardizing Data**

**1) Finding Extra Space**

WITH trimcte As

(

Select company, LENGTH(COMPANY), LENGTH(Trim(Company))

From layoffscleaning2

)

Select \* From trimcte

Where LENGTH(COMPANY)>LENGTH(Trim(Company))

2) Updating Column by removing space

Update layoffscleaning2

SET COMPANY= TRIM(COMPANY)

3) Finding Similar values in the column.

USE Distinct and Order By so that we can nearly similarly value at one place.

Select Distinct industry

From layoffscleaning2

Order By Industry

4) Once we find nearly similar values then following query is used to update the detail. For example crypto and Crypto Currency is same.

UPDATE layoffscleaning2

Set Industry= 'Crypto'

Where Industry like 'Crypto%'

**Changing Data Types**

1. Converting Text to Date Data Type:

UPDATE layoffscleaning2

SET `date` = STR\_TO\_DATE(`date`, '%m/%d/%Y');

ALTER TABLE layoffscleaning2

MODIFY COLUMN `date` DATE;

**Null OR Blank**

1. To find the Null or blank value following is the query used

Select \* From layoffscleaning2

Where industry IS NULL

OR industry= '';

1. We need to try to find if we can fill this value with the Relevant value. E.g let say Company is Airbnb and location is blank. We can find if we have another row for Airbnb where we have industry column filled.

Select \* From layoffscleaning2

Where company='Airbnb'

3)

Updating

Update layoffscleaning2

Set industry='Travel'

Where company='Airbnb'

1. For any column if we find Null and if we are sure that we can delete it as it does not make sense to keep it. So we will Delete such rows:

DELETE From layoffscleaning2

Where total\_laid\_off IS NULL AND

percentage\_laid\_off IS NULL

**Dropping Column**

There will be scenario where we have to drop any column. In this case we have added extra column called Row\_Num. We have to delete that.

Alter table layoffscleaning2

Drop column row\_num