-- DATA CLEANING

SELECT \*

FROM layoffs;

-- 1. check for duplicates and remove any

-- 2. standardize data and fix errors

-- 3. Look at null values and see what

-- 4. remove any columns and rows that are not necessary

-- 1. Remove Duplicates

CREATE TABLE layoffs\_staging

LIKE layoffs;

SELECT \*

FROM layoffs\_staging;

INSERT layoffs\_staging

SELECT \*

FROM layoffs;

-- CRETAING NEW TABLE TO DELETE DUPLIACTES FROM THE PREVIOUS TABLE..

CREATE TABLE `layoffs\_staging2` (

`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;

SELECT \*

FROM layoffs\_staging2;

SELECT \*,

ROW\_NUMBER() OVER(

PARTITION BY company , location , industry , total\_laid\_off , percentage\_laid\_off , stage , country , funds\_raised\_millions, `date` ) AS row\_num

FROM layoffs\_staging;

WITH duplicate\_cte AS

(

SELECT \*,

ROW\_NUMBER() OVER(

PARTITION BY company , location , industry , total\_laid\_off , percentage\_laid\_off , stage , country , funds\_raised\_millions, `date` ) AS row\_num

FROM layoffs\_staging

)

SELECT \*

FROM duplicate\_cte

WHERE row\_num>1;

INSERT INTO layoffs\_staging2

SELECT \*,

ROW\_NUMBER() OVER(

PARTITION BY company , location , industry , total\_laid\_off , percentage\_laid\_off , stage , country , funds\_raised\_millions, `date` ) AS row\_num

FROM layoffs\_staging;

DELETE

FROM layoffs\_staging2

WHERE row\_num >1;

SELECT \*

FROM layoffs\_staging2

WHERE row\_num>1;

-- STANDARDIZING DATA

SELECT company , TRIM(company)

FROM layoffs\_staging2;

UPDATE layoffs\_staging2

SET company=TRIM(company);

SELECT \*

FROM layoffs\_staging2

WHERE industry LIKE 'crypto%';

UPDATE layoffs\_staging2

SET industry='Crypto'

WHERE industry LIKE 'crypto%';

SELECT DISTINCT country, TRIM( TRAILING '.' FROM country)

FROM layoffs\_staging2

ORDER BY 1;

UPDATE layoffs\_staging2

SET country = TRIM(TRAILING '.' FROM country)

WHERE country LIKE 'United States%';

SELECT \*

FROM layoffs\_staging2;

SELECT `date`

FROM layoffs\_staging2;

UPDATE layoffs\_staging2

SET `date`= str\_to\_date(`date` , '%m/%d/%Y');

ALTER TABLE layoffs\_staging2

MODIFY COLUMN `date` DATE;

SELECT t1.industry ,t2.industry

FROM layoffs\_staging2 t1

JOIN layoffs\_staging2 t2

ON t1.company=t2.company

WHERE ( t1.industry IS NULL or t1.industry = ' ')

AND t2.industry IS NOT NULL;

UPDATE layoffs\_staging2 t1

JOIN layoffs\_staging2 t2

ON t1.company=t2.company

SET t1.industry =t2.industry

WHERE ( t1.industry IS NULL or t1.industry = ' ')

AND t2.industry IS NOT NULL;

UPDATE layoffs\_staging2

SET industry=NULL

WHERE industry =' ';

SELECT \*

FROM layoffs\_staging2

WHERE total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

DELETE

FROM layoffs\_staging2

WHERE total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

ALTER TABLE layoffs\_staging2

DROP COLUMN row\_num;

SELECT \*

FROM layoffs\_staging2

ORDER BY 1;

SELECT \*

FROM layoffs

ORDER BY 1;