SELECT \*

FROM world\_layoffs.layoffs;

CREATE TABLE world\_layoffs.layoffs\_staging

LIKE world\_layoffs.layoffs;

INSERT layoffs\_staging

SELECT \* FROM world\_layoffs.layoffs;

SELECT \*

FROM world\_layoffs.layoffs\_staging

;

SELECT company, industry, total\_laid\_off,`date`,

ROW\_NUMBER() OVER (

PARTITION BY company, industry, total\_laid\_off,`date`) AS row\_num

FROM

world\_layoffs.layoffs\_staging;

SELECT \*

FROM (

SELECT company, industry, total\_laid\_off,`date`,

ROW\_NUMBER() OVER (

PARTITION BY company, industry, total\_laid\_off,`date`

) AS row\_num

FROM

world\_layoffs.layoffs\_staging

) duplicates

WHERE

row\_num > 1;

SELECT \*

FROM world\_layoffs.layoffs\_staging

WHERE company = 'Oda'

;

SELECT \*

FROM (

SELECT company, location, industry, total\_laid\_off,percentage\_laid\_off,`date`, stage, country, funds\_raised\_millions,

ROW\_NUMBER() OVER (

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

) AS row\_num

FROM

world\_layoffs.layoffs\_staging

) duplicates

WHERE

row\_num > 1;

WITH DELETE\_CTE AS

(

SELECT \*

FROM (

SELECT company, location, industry, total\_laid\_off,percentage\_laid\_off,`date`, stage, country, funds\_raised\_millions,

ROW\_NUMBER() OVER (

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

) AS row\_num

FROM

world\_layoffs.layoffs\_staging

) duplicates

WHERE

row\_num > 1

)

DELETE

FROM DELETE\_CTE

;

WITH DELETE\_CTE AS (

SELECT company, location, industry, total\_laid\_off, percentage\_laid\_off, `date`, stage, country, funds\_raised\_millions,

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

FROM world\_layoffs.layoffs\_staging

)

DELETE FROM world\_layoffs.layoffs\_staging

WHERE (company, location, industry, total\_laid\_off, percentage\_laid\_off, `date`, stage, country, funds\_raised\_millions, row\_num) IN (

SELECT company, location, industry, total\_laid\_off, percentage\_laid\_off, `date`, stage, country, funds\_raised\_millions, row\_num

FROM DELETE\_CTE

) AND row\_num > 1;

ALTER TABLE world\_layoffs.layoffs\_staging ADD row\_num INT;

SELECT \*

FROM world\_layoffs.layoffs\_staging

;

CREATE TABLE `world\_layoffs`.`layoffs\_staging2` (

`company` text,

`location`text,

`industry`text,

`total\_laid\_off` INT,

`percentage\_laid\_off` text,

`date` text,

`stage`text,

`country` text,

`funds\_raised\_millions` int,

row\_num INT

);

INSERT INTO `world\_layoffs`.`layoffs\_staging2`

(`company`,

`location`,

`industry`,

`total\_laid\_off`,

`percentage\_laid\_off`,

`date`,

`stage`,

`country`,

`funds\_raised\_millions`,

`row\_num`)

SELECT `company`,

`location`,

`industry`,

`total\_laid\_off`,

`percentage\_laid\_off`,

`date`,

`stage`,

`country`,

`funds\_raised\_millions`,

ROW\_NUMBER() OVER (

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

) AS row\_num

FROM

world\_layoffs.layoffs\_staging;

-- now that we have this we can delete rows were row\_num is greater than 2

DELETE FROM world\_layoffs.layoffs\_staging2

WHERE row\_num >= 2;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;

-- if we look at industry it looks like we have some null and empty rows, let's take a look at these

SELECT DISTINCT industry

FROM world\_layoffs.layoffs\_staging2

ORDER BY industry;

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE industry IS NULL

OR industry = ''

ORDER BY industry;

-- let's take a look at these

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE company LIKE 'Bally%';

-- nothing wrong here

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE company LIKE 'airbnb%';

UPDATE world\_layoffs.layoffs\_staging2

SET industry = NULL

WHERE industry = '';

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE industry IS NULL

OR industry = ''

ORDER BY industry;

UPDATE layoffs\_staging2 t1

JOIN layoffs\_staging2 t2

ON t1.company = t2.company

SET t1.industry = t2.industry

WHERE t1.industry IS NULL

AND t2.industry IS NOT NULL;

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE industry IS NULL

OR industry = ''

ORDER BY industry;

SELECT DISTINCT industry

FROM world\_layoffs.layoffs\_staging2

ORDER BY industry;

UPDATE layoffs\_staging2

SET industry = 'Crypto'

WHERE industry IN ('Crypto Currency', 'CryptoCurrency');

SELECT DISTINCT industry

FROM world\_layoffs.layoffs\_staging2

ORDER BY industry;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;

SELECT DISTINCT country

FROM world\_layoffs.layoffs\_staging2

ORDER BY country;

UPDATE layoffs\_staging2

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

SELECT DISTINCT country

FROM world\_layoffs.layoffs\_staging2

ORDER BY country;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;

UPDATE layoffs\_staging2

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

ALTER TABLE layoffs\_staging2

MODIFY COLUMN `date` DATE;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE total\_laid\_off IS NULL;

SELECT \*

FROM world\_layoffs.layoffs\_staging2

WHERE total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

DELETE FROM world\_layoffs.layoffs\_staging2

WHERE total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;

ALTER TABLE layoffs\_staging2

DROP COLUMN row\_num;

SELECT \*

FROM world\_layoffs.layoffs\_staging2;