use world\_layoffs;

select \* from layoffs;

-- Removing Duplicates

create table layoffs\_staging like layoffs;

select \* from layoffs\_staging;

select \*,

ROW\_NUMBER()

OVER(PARTITION BY company, industry,total\_laid\_off,percentage\_laid\_off,`date`) as row\_num

from layoffs\_staging;

insert layoffs\_staging select \* from layoffs;

with duplicate\_cte 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 layoffs\_staging

)

select \* from duplicate\_cte

where row\_num >1;

select \* from layoffs\_staging where

company= 'Casper';

with duplicate\_cte 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 layoffs\_staging

)

delete from duplicate\_cte

where row\_num >1;

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 where row\_num >1;

insert into layoffs\_staging2

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 layoffs\_staging;

delete from layoffs\_staging2 where row\_num >1;

select \* from layoffs\_staging2 ;

-- Standarizing 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 industry from layoffs\_staging2;

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 `date`,

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

from layoffs\_staging2;

update layoffs\_staging2

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

select `date`

from layoffs\_staging2;

alter table layoffs\_staging2

modify column `date` date;

-- null and blank values

select \*

from layoffs\_staging2 where total\_laid\_off is null and percentage\_laid\_off is null;

update layoffs\_staging2 set industry = null

where industry = '';

select \*

from layoffs\_staging2 where industry is null or industry ='';

select \*

from layoffs\_staging2 where company = 'Airbnb';

select \*

from layoffs\_staging2 where company like'Bally%';

select t1.industry, t2.industry

from layoffs\_staging2 t1

join

layoffs\_staging2 t2

on t1.company= t2.company

-- and t1.location = t2.location

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;

select \*

from layoffs\_staging2

where total\_laid\_off is null

and percentage\_laid\_off is null;

-- delete nulls

delete

from layoffs\_staging2

where total\_laid\_off is null

and percentage\_laid\_off is null;

select \* from layoffs\_staging2;

alter table layoffs\_staging2

drop column row\_num;