Data cleaning of world_layoffs
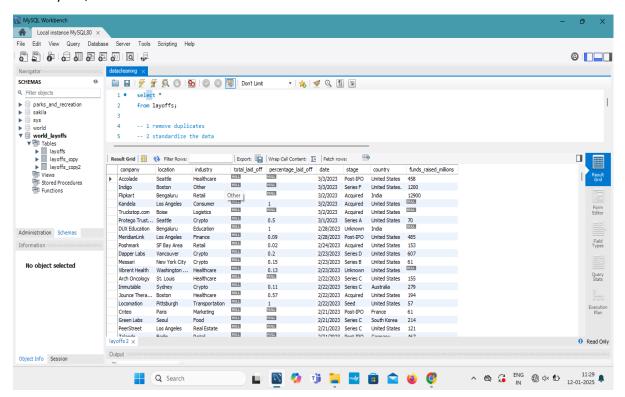
select *

from layoffs;

-- 1 remove duplicates

-- 2 standardize the data

-- 3 null values or empty values

-- 4 remove any colums or rows

create table layoffs_copy

like layoffs;

select *

from layoffs_copy;

insert layoffs_copy

```sql
select *
from layoffs;


select *,
row_number() over(
partition by company,industry,total_laid_off,percentage_laid_off,`date`)as row_num
from layoffs_copy;


with duplicatte_cte as
(
select *,
row_number() over(
partition by
company,industry,total_laid_off,percentage_laid_off,`date`,stage,country,funds_raised_millions)as
row_num
from layoffs_copy
)

select*
from duplicatte_cte
where row_num > 1;

CREATE TABLE `layoffs_copy2` (
  `company` text,
  `location` text,
  `industry` text,
  `total_laid_off` int DEFAULT NULL,
  `percentage_laid_off` text,
  `date` text,
  `stage` text,
  `country` text,
```
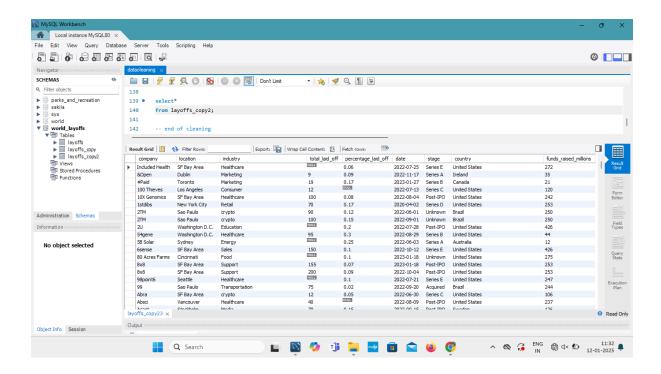
```sql
  `funds_raised_millions` int DEFAULT NULL,

  `row_num` int

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;



select*

from layoffs_copy2;


insert into layoffs_copy2

select *,

row_number() over(

partition by
company,industry,total_laid_off,percentage_laid_off,`date`,stage,country,funds_raised_millions)as
row_num

from layoffs_copy;



DELETE

from layoffs_copy2

where row_num > 1;


select*

from layoffs_copy2;


-- standardizing data


select company,trim(company)

from layoffs_copy2;


update layoffs_copy2

set company = trim(company);
```

```sql
select  industry
from layoffs_copy2
where industry like 'crypto%';


update layoffs_copy2
set industry = 'crypto'
where industry like 'crypto%';


select distinct country, trim(trailing '.' from country)
from layoffs_copy2
order by 1;


update layoffs_copy2
set country = trim(trailing '.' from country)
where country like 'united states%';


select `date`,
str_to_date(`date`,'%m/%d/%Y')
from layoffs_copy2;


update layoffs_copy2
set `date` = str_to_date(`date`,'%m/%d/%Y');


alter table layoffs_copy2
modify column `date` date;


select*
from layoffs_copy2
where total_laid_off is null
and percentage_laid_off is null;
```

```sql
update layoffs_copy2
set industry = null
where industry = ' ';


select *
from layoffs_copy2 as t1
join layoffs_copy2 as t2
    on  t1.company = t2.company
where (t1.industry is null or t1.industry = ' ')
and t2.industry is not null;


UPDATE layoffs_copy2  l1
JOIN layoffs_copy2  l2
    ON  l1.company = l2.company
SET l1.industry = l2.industry
WHERE l1.industry IS NULL
AND l2.industry IS NOT NULL;


select*
from layoffs_copy2
where total_laid_off is null
and percentage_laid_off is null;


delete
from layoffs_copy2
where total_laid_off is null
and percentage_laid_off is null;



alter table layoffs_copy2
drop column row_num;
```

select*

from layoffs_copy2;

-- end of cleaning