```sql
# Data cleaning project

-- Remove duplicates
-- Standardize the data
-- Dealing with NULL and emplty values
-- Removing unnecessary columns and rows

select * from layoffs;

# Creating a copy of raw data to perform operations

create table layoffs_copy
like layoffs;

insert into layoffs_copy
select * from layoffs;

select * from layoffs_copy;

# 1. Removing duplicates
with duplicate_cte as (
 select * ,
 row_number() over(
 partition by company, location, industry, total_laid_off, percentage_laid_off, `date`, stage, country, funds
 from layoffs_copy
)
select * from duplicate_cte
where row_num >=2;

-- checking
select * from layoffs_copy
where company = 'casper';

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,
  `funds_raised_millions` int DEFAULT NULL,
  `row_num` int
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;

insert into layoffs_copy2
select * ,
 row_number() over(
 partition by company, location, industry, total_laid_off, percentage_laid_off, `date`, stage, country, funds
 from layoffs_copy;

select * from layoffs_copy2;

-- deleting duplicate rows
delete
from layoffs_copy2
where row_num >=2;

select * from layoffs_copy2
where row_num >=2;

# Standardizing data

select company ,trim(company)
from layoffs_copy2;

update layoffs_copy2
```

```sql
set company = trim(company);

select * from layoffs_copy2;

select distinct industry
from layoffs_copy2
order by 1 ;

select *
from layoffs_copy2
where industry like 'crypto%';

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

select distinct country
from layoffs_copy2
order by 1;

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

update layoffs_copy2
set country = trim(trailing '.' from country )
where country = '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');

select * from layoffs_copy2;

alter table layoffs_copy2
modify column `date` date;

# Dealing with NULL and emplty values

select * from layoffs_copy2
where industry is null
or industry = '';

select *
from layoffs_copy2 as lc1
JOIN layoffs_copy2 as lc2
 on lc1.company = lc2.company
where (lc1.industry is null
or lc1.industry ='')
and lc2.industry is not null;

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

update layoffs_copy2 lc1
join layoffs_copy2 lc2
set lc1.industry = lc2.industry
where lc1.industry is null
and lc2.industry is not null;

# Removing unnecessary columns and rows

select * from layoffs_copy2
where total_laid_off is null
```

```sql
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;
```