```
use Data_cleaning;

Create table layoffs_stage(select * from layoffs); -- Creating duplicates of source data
select * from layoffs_stage;

-- 1. Remove duplicates
-- 2. Standardize the data
-- 3. Handle null and blank values
-- 4. Handle date format issues
-- 5. Change data types, remove columns, rename tables and column names/types

-- 4. Handling date format issues
alter table layoffs add column new_date date; -- Create a new date column
update layoffs set new_date = str_to_date(date, "%m/%d/%Y"); -- Update the new date column using STR_TO_DATE function
alter table layoffs drop column date; -- Drop the existing date column
alter table layoffs change column new_date date date; -- Change the column name from new_date to date

-- 1. Remove duplicates

with cte as (
    select *,
        row_number() over(partition by company, location, industry, total_laid_off, percentage_laid_off, stage, country, date) as Row_num
    from layoffs_stage
)
select * from cte where row_num > 1; -- To identify duplicate values

create table layoffs_stage2 (
    select *,
        row_number() over(partition by company, location, industry, total_laid_off, percentage_laid_off, stage, country, date) as Row_num
    from layoffs_stage
); -- Create another table
select * from layoffs_stage2 where Row_num > 1;
delete from layoffs_stage2 where Row_num > 1;
alter table layoffs_stage2 drop column Row_num; -- Drop the column that is not needed further
-- rename table layoffs_stage2 to layoffs_stage

-- 2. Standardize text
select company, trim(company) from layoffs_stage2;
update layoffs_stage2 set company = trim(company);

select distinct industry from layoffs_stage2 order by 1;
select * from layoffs_stage2 where industry like 'crypto%';
update layoffs_stage2 set industry = 'Crypto' where industry like 'crypto%';

select distinct country from layoffs_stage2 order by 1;
select * from layoffs_stage2 where country like 'United States%';
update layoffs_stage2 set Country = 'United States' where Country like 'United States%';

-- 3. Handle null and blank values
```

```
-- For string values
select distinct industry from layoffs_stage2 order by 1;
select * from layoffs_stage2 where industry is null;
update layoffs_stage2 set industry = 'Entertainment' where industry is null; -- Found industry type through company name using Google
select * from layoffs_stage2 where company = 'Bally''s Interactive';

select * from layoffs_stage2 where industry = '' or industry is null;

select * from layoffs_stage2 where company = 'Airbnb';
update layoffs_stage2 set industry = 'Travel' where company = 'Airbnb';

select * from layoffs_stage2 where company = 'carvana';
update layoffs_stage2 set industry = 'Transportation' where company = 'carvana' and industry = '';

select * from layoffs_stage2 where company = 'Juul';
update layoffs_stage2 set industry = 'Consumer' where company = 'Juul' and industry = '';

select * from layoffs_stage2 where total_laid_off is null and percentage_laid_off is null; -- Delete unnecessary data from the table
delete from layoffs_stage2 where total_laid_off is null and percentage_laid_off is null;

describe layoffs_stage2;
alter table layoffs_stage2 modify column percentage_laid_off int; -- Changing column type
```