***Data Cleaning***

1. ***Remove Duplicate***
2. ***Standardize The Data***
3. ***Check Null Values and Blank Values***
4. ***Remove Any Columns***

***-----------------------------------------------------------------------------------------------------------------------------***

***First, We Create A Table (name is layoffs\_staging) Like Our Current Table (name is layoffs).***

***We work on this table, because we want the table with the raw data, in case something happens. We did not miss our raw data, because the data is so important.***

**CREATE TABLE** layoffs\_staging

**LIKE** layoffs;

***(We create a table named layoff\_staging, which is like the layoffs table, which is our raw data.)***

***Second, we put all data or insert data on our create new table name layoffs\_staging from layoffs table***

**INSERT** layoffs\_staging

SELECT\*

FROM layoffs;

***(We are inserting everything in the layoffs\_staging table from the layoffs table.)***

***--------------------REMOVE DUPLICATE-------------------------***

To remove duplicates we give the unique raw id

***First Step: Check Duplicate***

***Removing duplicate : we want to make sure we don’t have any duplicate. We add extra column name UNIQUE ROW ID which made it really easy to remove duplicate. Given Row Number we all match it against all of these columns then we see if there any duplicate***

SELECT\*,

ROW\_NUMBER ()

OVER(

PARTION BY company, location, industry, total\_laid\_off, date, stage, country, funds\_raised, ) AS row\_num ---partition by all of these columns

FROM layoffs\_staging;

We can filter where the row number is greater than 2. If it has two or above that means there’s duplicate. We are going to take this all above and put them into subqueries or CTE. We are create CTE for this because it’s really easy.

WITH duplicate\_cte AS (

SELECT \*

ROW\_NUMBER () OVER (PARTION BY industry, company, location, total\_laid\_off, date, stage, country, funds\_raised)

FROM layoff\_staging)

SELECT \*

FROM duplicate\_cte

WHERE row\_number > 1;

Showing many duplicate after run the code, now check one duplicate how many they are.

SELECT \*

FROM layoff\_staging

WHERE company = ‘casper’;

***Second Step: Lets the row\_num column add in our database table, and create new table name layoff\_staging2. Without doing this we cannot delete the duplicate, because the operation cannot perform with out adding row\_num in database.***

CREATE TABLE layoff\_staging2

Let’s create the table manually

***Layoff\_staging table*-right click, then select**-copy to click board, after click- create statement:

Show format…………………………………………

……………………………………………………………

……………………………………………………………

`row\_num` INT // add the row number in the table

***Third Step: Delete the duplicate from the table***

DELETE

FROM layoff\_staging2

WHERE row\_num > 1;