SELECT \*

FROM layoffs;

--1) Stage my Raw data(duplicate) and name it **Layoffs\_staging.**

--2) Remove any duplicates

--3) Standardize data i.e check for errors and other things to make for a uniform data

--4) Check for NULLS and blank spaces

--5) Remove any unnecessary columns or rows

**--1) Stage my Raw data(duplicate) and name it Layoffs\_staging.**

CREATE TABLE [dbo].[layoffs\_staging](

[company] [nvarchar](255) NULL,

[location] [nvarchar](255) NULL,

[industry] [nvarchar](255) NULL,

[total\_laid\_off] [float] NULL,

[percentage\_laid\_off] [float] NULL,

[date] [datetime] NULL,

[stage] [nvarchar](255) NULL,

[country] [nvarchar](255) NULL,

[funds\_raised\_millions] [float] NULL

);

INSERT layoffs\_staging

SELECT \*

FROM layoffs;

**--2) Remove any duplicates**

WITH cte\_duplicate AS

(SELECT \*, ROW\_NUMBER()

OVER( PARTITION BY company, 'location', industry, total\_laid\_off,

percentage\_laid\_off, 'date', stage,country, funds\_raised\_millions ORDER BY company) AS row\_num

FROM layoffs\_staging)

SELECT \*

FROM cte\_duplicate

WHERE row\_num>1;

SELECT \*, ROW\_NUMBER()

OVER( PARTITION BY company, 'location', industry, total\_laid\_off,

percentage\_laid\_off, 'date', stage,country, funds\_raised\_millions ORDER BY company) AS row\_num

FROM layoffs\_staging;

**--3) Standardize data i.e check for errors and other things to make for a uniform data**

SELECT

company, TRIM(company)

FROM

layoffs\_staging;

UPDATE layoffs\_staging

SET

company = TRIM(company);

SELECT

\*

FROM

layoffs\_staging

WHERE

industry IN ('%Crypto%');

UPDATE layoffs\_staging

SET

industry = 'Crypto'

WHERE

industry = 'Cryptocurrency';

SELECT DISTINCT

country

FROM

layoffs\_staging

ORDER BY 1

;

UPDATE layoffs\_staging

SET

country = 'United States'

WHERE

country = 'United States.';

**--4) Check for NULLS and blank spaces**

ALTER TABLE layoffs\_staging

ALTER COLUMN "date" DATE NULL;

SELECT

\*

FROM

layoffs\_staging

WHERE

total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

**--5) Remove any unnecessary columns or rows**

DELETE

layoffs\_staging

WHERE total\_laid\_off IS NULL

AND percentage\_laid\_off IS NULL;

SELECT

\*

FROM

layoffs\_staging

WHERE

company LIKE 'Airbnb';

WITH t3 AS

(SELECT t1.industry as industry1,t2.industry as industry2

FROM layoffs\_staging t1

JOIN layoffs\_staging t2

ON t1.company = t2.company AND

t1.location= t2.location

WHERE t1.industry IS NULL AND t2.industry IS NOT NULL)

UPDATE t3

SET industry1 =industry2

WHERE industry1 IS NULL AND industry2 IS NOT NULL;

SELECT

\*

FROM

layoffs\_staging t1

JOIN

layoffs\_staging t2 ON t1.company = t2.company

AND t1.location = t2.location

WHERE

t2.company LIKE 'JUUl';

SELECT

\*

FROM

layoffs\_staging