SQL Script With Result

1.

```
3 •   select *
4     from layoffs;
5
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: ⫯A | Fetch rows:

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_rai
Atlassian	Sydney	Other	500	0.05	3/6/2023	Post-IPO	Australia	210
SiriusXM	New York City	Media	475	0.08	3/6/2023	Post-IPO	United States	525
Alerzo	Ibadan	Retail	400	NULL	3/6/2023	Series B	Nigeria	16
UpGrad	Mumbai	Education	120	NULL	3/6/2023	Unknown	India	631
Loft	Sao Paulo	Real Estate	340	0.15	3/3/2023	Unknown	Brazil	788
Embark Trucks	SF Bay Area	Transportation	230	0.7	3/3/2023	Post-IPO	United States	317
Lendi	Sydney	Real Estate	100	NULL	3/3/2023	Unknown	Australia	59
UserTesting	SF Bay Area	Marketing	63	NULL	3/3/2023	Acquired	United States	152
Airbnb	SF Bay Area		30	NULL	3/3/2023	Post-IPO	United States	6400
Accolade	Seattle	Healthcare	NULL	NULL	3/3/2023	Post-IPO	United States	458
Indigo	Boston	Other	NULL	NULL	3/3/2023	Series F	United States.	1200
Zscaler	SF Bay Area	Security	177	0.03	3/2/2023	Post-IPO	United States	148
MasterClass	SF Bay Area	Education	79	NULL	3/2/2023	Series E	United States	461
Ambev Tech	Blumenau	Food	50	NULL	3/2/2023	Acquired	Brazil	NULL
Fittr	Pune	Fitness	30	0.11	3/2/2023	Series A	India	13

ayoffs 1 ✕

2.

```
create table layoffs_staging
like layoffs;
```

3.

```
14 •   select *
15     from layoffs_staging;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: 🔠

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions

ayoffs_staging 3 ×

4.

```
insert layoffs_staging
select *
from layoffs;
```

5.

```
select *
from layoffs_staging;
```

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions
Atlassian	Sydney	Other	500	0.05	3/6/2023	Post-IPO	Australia	210
SiriusXM	New York City	Media	475	0.08	3/6/2023	Post-IPO	United States	525
Alerzo	Ibadan	Retail	400	NULL	3/6/2023	Series B	Nigeria	16
UpGrad	Mumbai	Education	120	NULL	3/6/2023	Unknown	India	631
Loft	Sao Paulo	Real Estate	340	0.15	3/3/2023	Unknown	Brazil	788
Embark Trucks	SF Bay Area	Transportation	230	0.7	3/3/2023	Post-IPO	United States	317
Lendi	Sydney	Real Estate	100	NULL	3/3/2023	Unknown	Australia	59
UserTesting	SF Bay Area	Marketing	63	NULL	3/3/2023	Acquired	United States	152
Airbnb	SF Bay Area		30	NULL	3/3/2023	Post-IPO	United States	6400
Accolade	Seattle	Healthcare	NULL	NULL	3/3/2023	Post-IPO	United States	458
Indigo	Boston	Other	NULL	NULL	3/3/2023	Series F	United States.	1200
Zscaler	SF Bay Area	Security	177	0.03	3/2/2023	Post-IPO	United States	148
MasterClass	SF Bay Area	Education	79	NULL	3/2/2023	Series E	United States	461
Ambev Tech	Blumenau	Food	50	NULL	3/2/2023	Acquired	Brazil	NULL
Fittr	Pune	Fitness	30	0.11	3/2/2023	Series A	India	13

6.

```
with duplicate_cte as
(
select *,
row_number()
over(partition by company,location, industry, total_laid_off, percentage_laid_off, `date`, stage, country, funds_raised_millions
from layoffs_staging
)
select *
from duplicate_cte
where row_num>1;
```

In duplicate_CTE create a new column name row_num in which there is 1 for all unique rows and 2 for duplicate rows.

Then from duplicate_CTE display all those columns where row_num is greater than 1. (actually these are duplicate rows)

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions	row_num
Casper	New York City	Retail	NULL	NULL	9/14/2021	Post-IPO	United States	339	2
Cazoo	London	Transportation	750	0.15	6/7/2022	Post-IPO	United Kingdom	2000	2
Hibob	Tel Aviv	HR	70	0.3	3/30/2020	Series A	Israel	45	2
Wildlife Studios	Sao Paulo	Consumer	300	0.2	11/28/2022	Unknown	Brazil	260	2
Yahoo	SF Bay Area	Consumer	1600	0.2	2/9/2023	Acquired	United States	6	2

7.

Created layoffs_staging2 to store cleaned data with row_num.

```sql
CREATE TABLE `layoffs_staging2` (
    `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;
```

8.

```sql
select *
from layoffs_staging2;
```

9.

```
insert into layoffs_staging2
select *,
row_number()
over(partition by company,location, industry, total_laid_off, percentage_laid_off, `date`, stage, country, funds_raised_mill
from layoffs_staging;
```

Insert all data from layoffs_staging table to layoffs_staging2 along with row_num column as by using the window function.

10.

i. Removing Duplicates

```
delete
from layoffs_staging2
where row_num>1;
```

ii. Standardizing Data

11.

```
select company, trim(company)
from layoffs_staging2;
```

As there are extra white spaces in the company column, trim function is used to remove these spaces.

company	trim(company)
E Inc.	E Inc.
Included Health	Included Health
&Open	&Open
#Paid	#Paid
100 Thieves	100 Thieves
100 Thieves	100 Thieves
10X Genomics	10X Genomics
1stdibs	1stdibs
2TM	2TM
2TM	2TM
2U	2U
54gene	54gene
5B Solar	5B Solar
6sense	6sense
80 Acres Farms	80 Acres Farms

12.

```
update layoffs_staging2
set company= trim(company);
```

This will upadate the company column by removing extra white spaces.

13.

Check industry column

```
select distinct(industry)
from layoffs_staging2
order by 1;
```

industry
NULL
Aerospace
Construction
Consumer
Crypto
Crypto Currency
CryptoCurrency
Data
Education
Energy
Fin-Tech
Finance
Fitness
Food

Result 8 ×

We need to standardized the Crypto or Crypto Currency to one standard name.

14.

```
update layoffs_staging2
set industry='Crypto'
where industry like 'Crypto%';
```

This will update industry column with single standard name Crypto for all where industry column is start with Crypto or any text.

15.

Check location column

```
select distinct(location)
from layoffs_staging2
order by 1;
```

No blanks or duplicates and the data is in standard form.

location
Accra
Ahmedabad
Albany
Amsterdam
Ann Arbor
Athens
Atlanta
Auckland
Austin
Baltimore
Bangkok
Barcelona
Baton Rouge
Beau Vallon
Beijing

Result 10 ×

16.

```sql
select distinct(country)
from layoffs_staging2
order by 1;
```

check country column

United States is not in standard form.

country
Singapore
South Africa
South Korea
Spain
Sweden
Switzerland
Thailand
Turkey
United Arab Emirates
United Kingdom
United States
United States.
Uruguay
Vietnam

17.

```
update layoffs_staging2
set country=trim(trailing '.' from country)
where country like 'United States%' ;
```

Here update the country column by triming the dot from united states and hence used a single standard name.

18.

```
select `date`,
str_to_date(`date`,'%m/%d/%Y')
from layoffs_staging2;
```

Convert date column to a standard date formate.

9

date	str_to_date(`date`,'%m/%d/%Y')
▶ 12/16/2022	2022-12-16
7/25/2022	2022-07-25
11/17/2022	2022-11-17
1/27/2023	2023-01-27
7/13/2022	2022-07-13
1/10/2023	2023-01-10
8/4/2022	2022-08-04
4/2/2020	2020-04-02
6/1/2022	2022-06-01
9/1/2022	2022-09-01
7/28/2022	2022-07-28
8/29/2022	2022-08-29
6/3/2022	2022-06-03
10/12/2022	2022-10-12
1/18/2023	2023-01-18

Result 16 ✕

19.

The date column have a text data type.

Convert it into date data type.

```
alter table layoffs_staging2
modify column `date` date;
```

iii.Removing NULL or Blank Values

20.

```
select *
from layoffs_staging2
where company='Airbnb';
```

	company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions	row_r
▶	Airbnb	SF Bay Area		30	NULL	2023-03-03	Post-IPO	United States	6400	1
	Airbnb	SF Bay Area	Travel	1900	0.25	2020-05-05	Private Equity	United States	5400	1

21.

```
update layoffs_staging2 as t1
join layoffs_staging2 as t2
    on t1.company=t2.company
set t1.industry=t2.industry
where (t1.industry is null or t1.industry = ' ')
and t2.industry is not null;
```

This query self join the layoffs_staging2 tables on company column and populate the null or blank industry column of t1 with the t2 industry column values but for that the company name must be the same.

22.

```
-- 4. Remove any columns or rows
select *
from layoffs_staging2
where total_laid_off is null
and
percentage_laid_off is null;
```

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions	
E Inc.	Toronto	Transportation	NULL	NULL	2022-12-16	Post-IPO	Canada	NULL	1
100 Thieves	Los Angeles	Retail	NULL	NULL	2023-01-10	Series C	United States	120	1
Accolade	Seattle	Healthcare	NULL	NULL	2023-03-03	Post-IPO	United States	458	1
Ada	Toronto	Support	NULL	NULL	2023-02-01	Series C	Canada	190	1
Adara	SF Bay Area	Travel	NULL	NULL	2020-03-31	Series C	United States	67	1
Addi	Bogota	Finance	NULL	NULL	2022-06-14	Series C	Colombia	376	1
AirMap	Los Angeles	Aerospace	NULL	NULL	2020-04-30	Unknown	United States	75	1
Airtasker	Sydney	Consumer	NULL	NULL	2022-07-04	Series C	Australia	26	1
Akerna	Denver	Logistics	NULL	NULL	2022-05-27	Unknown	United States	46	1
Akerna	Denver	Logistics	NULL	NULL	2020-09-02	Post-IPO	United States	NULL	1

layoffs_staging2 28 ×

23.

```
delete
from layoffs_staging2
where total_laid_off is null
and
percentage_laid_off is null;
```

iv. Removing Columns or Rows

24.

```
alter table layoffs_staging2
drop column row_num;
```

25.

```
select *
from layoffs_staging2;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions
BounceX	New York City	Marketing	77	0.2	2020-04-06	Series B	United States	75
Brainly	Krakow	Education	25	NULL	2022-11-06	Series D	Poland	148
Branch	New York City	Retail	3	0.27	2020-06-11	Seed	United States	2
Branch Metrics	SF Bay Area	Marketing	100	0.2	2020-04-07	Series E	United States	367
Brave Care	Portland	Healthcare	40	0.33	2022-09-06	Series B	United States	42
Breathe	New Delhi	Healthcare	50	0.33	2022-06-14	Series A	India	6
Breather	Montreal	Real Estate	120	0.8	2020-12-16	Series D	Canada	131
Brex	SF Bay Area	Finance	62	0.15	2020-05-29	Series C	United States	732
Brex	SF Bay Area	Finance	136	0.11	2022-10-11	Series D	United States	1500
Bridge Conn...	Nashville	Healthcare	154	1	2020-11-17	Series B	United States	45
Bridgit	Waterloo	Construction	13	0.13	2022-12-06	Series B	Canada	36
Bright Machines	SF Bay Area	Data	30	0.08	2022-06-28	Unknown	Israel	250
Bright Money	Bengaluru	Finance	100	0.5	2022-07-15	Series A	India	31
Brighte	Sydney	Energy	58	NULL	2022-09-09	Series C	Australia	145
Brighte	Sydney	Energy	30	0.15	2022-06-20	Series C	Australia	145