

MYSQL

DATA CLEANING

CONTENT:

- 1.Removing Duplicates.**
- 2.Standardize the Data.**
- 3.Null OR Blank Values.**
- 4.Remove Column.**

Navigator ::::::::::::::::::::::::::::::::::

DATA CLEANING IN MY SQL* x layoffs_staging2

Filter objects

- student
- sys
- world_layoffs**
 - Tables
 - layoffs
 - layoffs_staging
 - layoffs_staging2
 - Views
 - Stored Procedures
 - Functions
- zomato

Administration Schemas

Information

No object selected

Object Info

```

1      -- DATA CLEANING
2
3
4
5      |
6  ●   USE WORLD_LAYOFFS;
7  ●   SELECT *
8      FROM LAYOFFS;
9
10
11     # CREATE A TABLE LIKE "LAYOFFS"
12  ●   CREATE TABLE layoffs_staging
13      LIKE LAYOFFS;
14
15     #INSERT ALL THE DATA IN "LAYOFFS_STAGING" TABLE
16  ●   INSERT layoffs_staging
17      SELECT *
18      FROM LAYOFFS;
19
20
21
22
23
24
25

```

Output

Session

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

- student
- sys
- world_layoffs
 - Tables
 - layoffs
 - layoffs_staging
 - layoffs_staging2
 - Views
 - Stored Procedures
 - Functions
- zomato

Administration Schemas

Information

No object selected

DATA CLEANING IN MY SQL* x layoffs_staging2

Limit to 5000 rows

```
39
40
41
42 • CREATE TABLE `layoffs_staging` (
43     `company` text,
44     `location` text,
45     `industry` text,
46     `total_laid_off` int DEFAULT NULL,
47     `percentage_laid_off` text,
48     `date` text,
49     `stage` text,
50     `country` text,
51     `funds_raised_millions` int DEFAULT NULL
52 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
53
54
55
56
57
58
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
89	13:32:39	WITH STAGE_YEAR (stage , years , total_laid_off) AS (SELECT stage , YEAR(`date`) ,SUM(t...	56 row(s) returned	0.016 sec / 0.000 sec
90	13:33:07	WITH STAGE_YEAR (stage , years , total_laid_off) AS (SELECT stage , YEAR(`date`) ,SUM(t...	20 row(s) returned	0.016 sec / 0.000 sec

Object Info Session

MySQL Workbench

Local instance MySQL80

FileEditViewQueryDatabaseServerToolsScriptingHelp

SQLSQL

Limit to 5000 rows

Limit to 5000 rows

DATA CLEANING IN MY SQL*

SCHEMAS

Filter objects

student

sys

world_layoffs

Tables

layoffs

layoffs_staging

layoffs_staging2

Columns

company

location

industry

total_laid

percentage_laid_off

date

stage

country

funds_rai

row_num

Indexes

30

31 • WITH DUPLICATE_CTE AS

32 (

33 SELECT * ,

34 ROW_NUMBER() OVER(

35 PARTITION BY company, location , industry, total_laid_off, percentage_laid_off,`date`,

36 stage,country,funds_raised_millions) AS ROW_NUM

37 FROM world_layoffs.layoffs_staging

38)

39 SELECT *

40 FROM DUPLICATE_CTE

41 WHERE ROW_NUM > 1 ;

42

Administration

Schemas

Information

Column: date

Definition: date date

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

	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

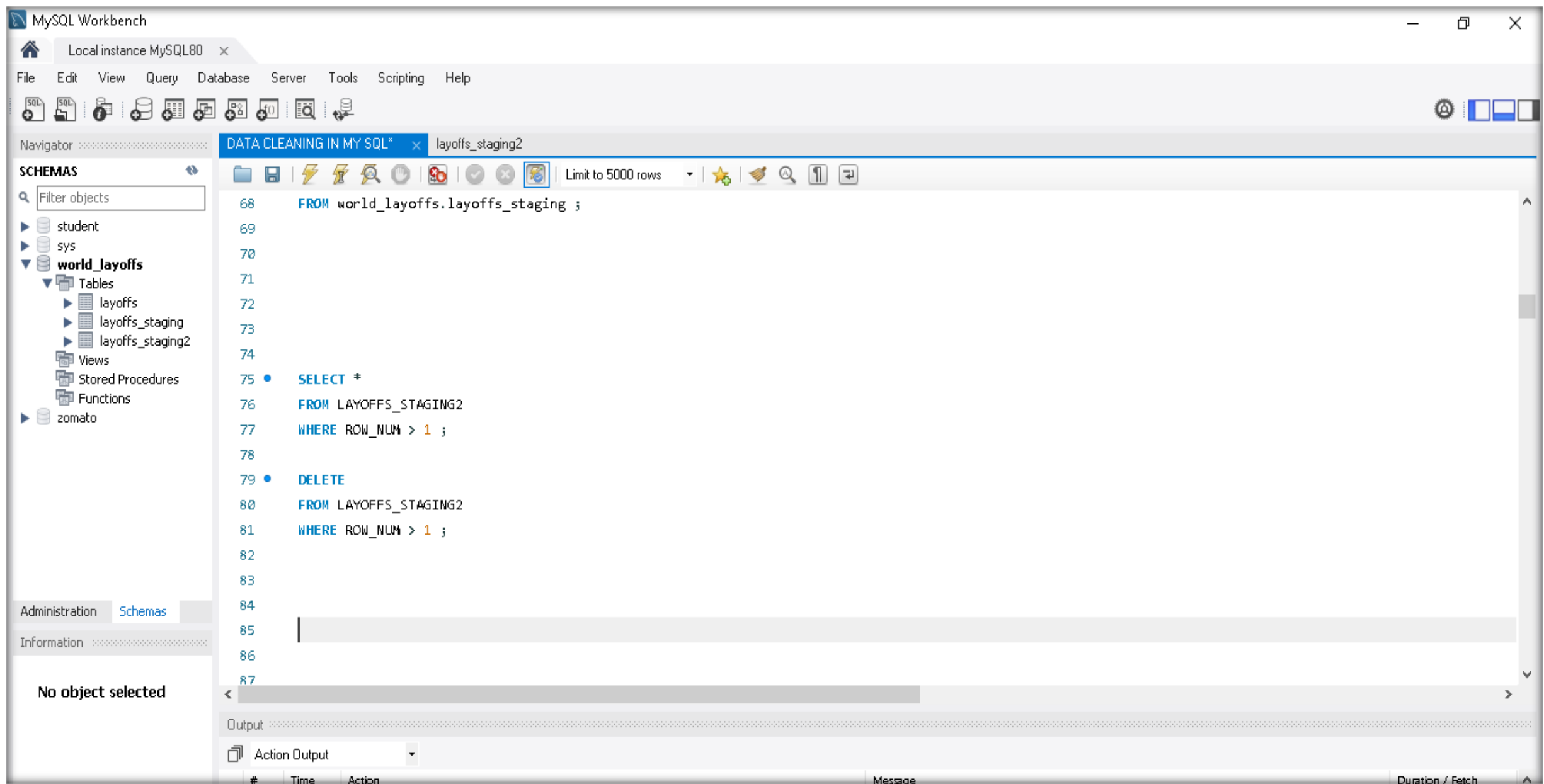
Result 46

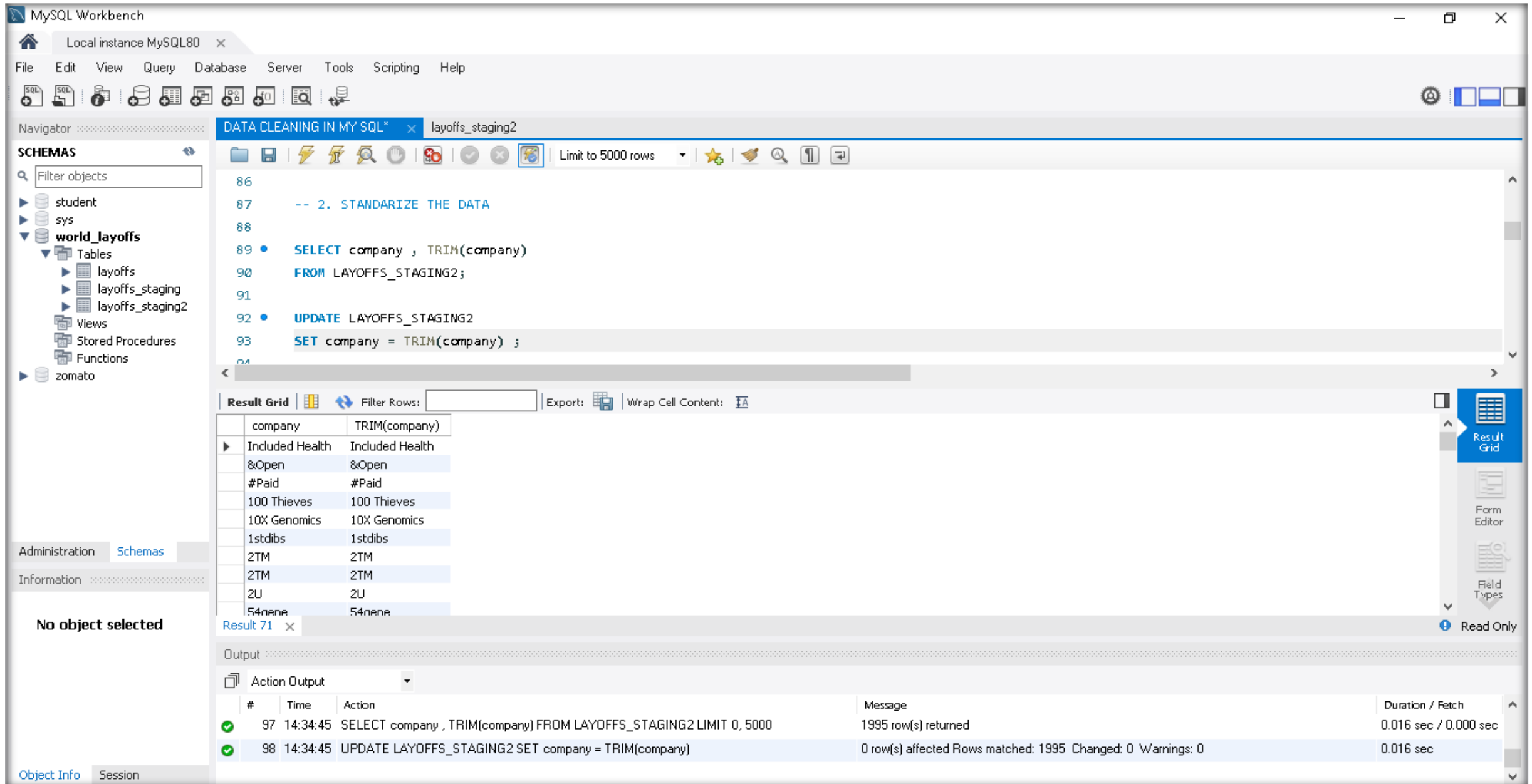
Read Only

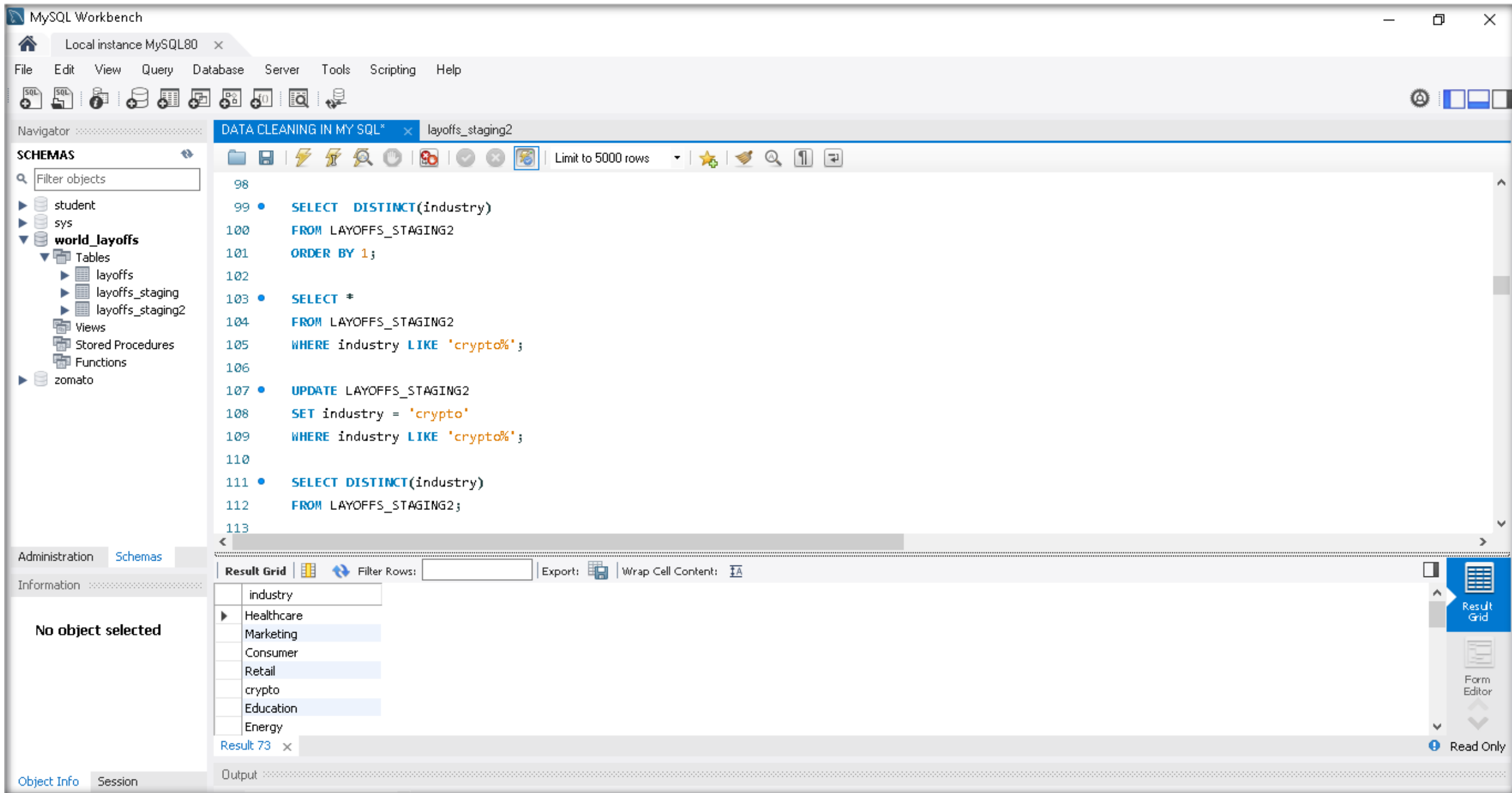
Object Info

Session

Output







MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

- student
- sys
- world_layouts
 - Tables
 - layoffs
 - layoffs_staging
 - layoffs_staging2
 - Views
 - Stored Procedures
 - Functions
- zomato

DATA CLEANING IN MY SQL* layoffs_staging2

Limit to 5000 rows

```
109 • SELECT DISTINCT location
110 FROM LAYOFFS_STAGING2 ORDER BY 1 ;
111
112 • SELECT *
113 FROM LAYOFFS_STAGING2
114 WHERE country LIKE 'united states%' ORDER BY 1;
115
116 • SELECT DISTINCT country , TRIM(TRAILING '.' FROM country)
117 FROM LAYOFFS_STAGING2 ORDER BY 1;
118
119 • UPDATE LAYOFFS_STAGING2
120 SET country = TRIM(TRAILING '.' FROM country)
121 WHERE country LIKE 'United States' ;
122 • SELECT *
123 FROM LAYOFFS_STAGING2;
124
```

Administration Schemas

Information

No object selected

Result Grid

	country	TRIM(TRAILING '.' FROM country)
▶	Argentina	Argentina
	Australia	Australia
	Austria	Austria
	Brazil	Brazil
	Bulgaria	Bulgaria
	Canada	Canada

Result 78

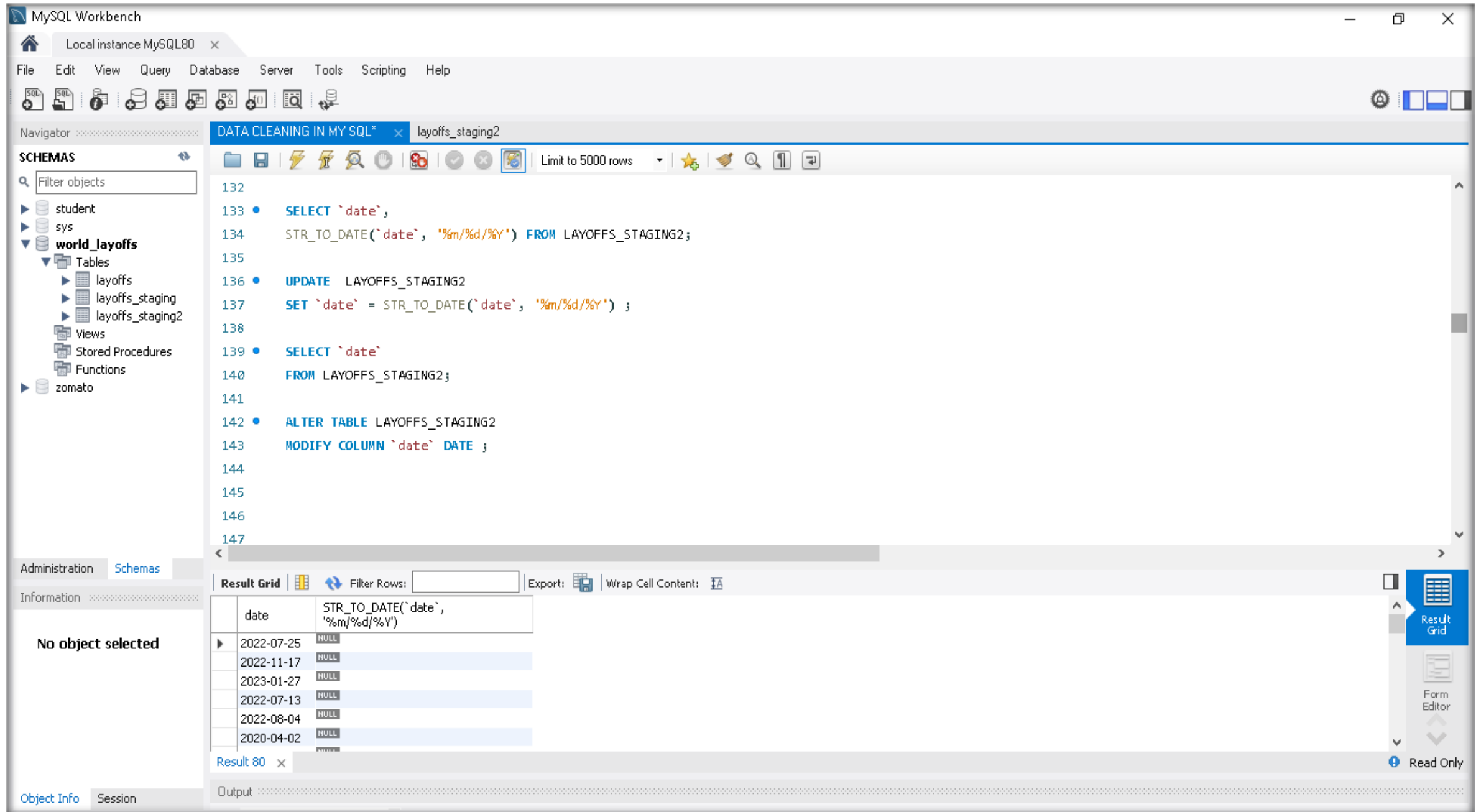
Object Info Session

Output

Result Grid

Form Editor

Read Only



MySQL Workbench

Local instance MySQL80 x

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

- student
- sys
- world_layoffs
 - Tables
 - layoffs
 - layoffs_staging
 - layoffs_staging2
 - Views
 - Stored Procedures
 - Functions
- zomato

Administration Schemas

Information

No object selected

DATA CLEANING IN MY SQL* x layoffs_staging2

Limit to 5000 rows

```
148 -- 3. NULL VALUES OR BLANK VALUES
149
150 • SELECT * FROM
151 LAYOFFS_STAGING2
152 WHERE total_laid_off IS NULL AND
153 percentage_laid_off IS NULL;
154
155 • SELECT t1.industry , t2.industry
156 FROM LAYOFFS_STAGING2 t1
157 JOIN LAYOFFS_STAGING2 t2
158     ON t1.company = t2.company
159     AND t1.location = t2.location
160 WHERE (t1.industry IS NULL OR t1.industry = '' )
161 AND t2.industry IS NOT NULL
162 ;
163
164 • UPDATE LAYOFFS_STAGING2 t1
165 JOIN LAYOFFS_STAGING2 t2
166     ON t1.company = t2.company
167 SET t1.industry = t2.industry
168 WHERE t1.industry IS NULL
169 AND t2.industry IS NOT NULL
170 ;
171
172 • UPDATE LAYOFFS_STAGING2
```

Object Info Session

Output

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

student
sys
world_layoffs
 Tables
 layoffs
 layoffs_staging
 layoffs_staging2
 Views
 Stored Procedures
 Functions
zomato

Administration Schemas

Information

No object selected

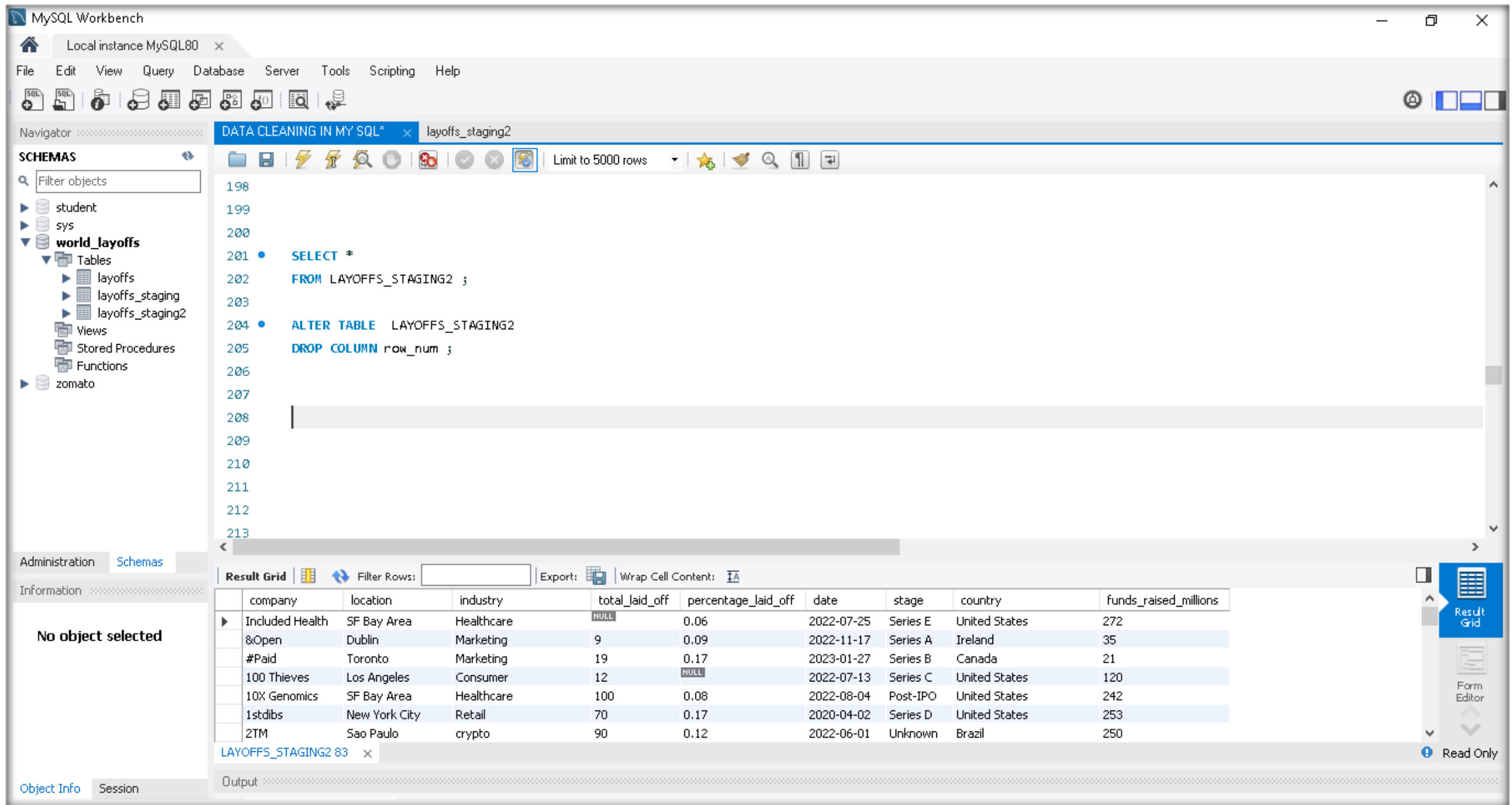
DATA CLEANING IN MY SQL* x layoffs_staging2

Limit to 5000 rows

```
175
176 • SELECT * FROM
177     LAYOFFS_STAGING2
178     WHERE company LIKE 'Bally%' ;
179
180 • SELECT *
181     FROM LAYOFFS_STAGING2
182     WHERE industry IS NULL
183     OR industry = '' ;
184
185 • SELECT *
186     FROM LAYOFFS_STAGING2 ;
187
188 • SELECT * FROM
189     LAYOFFS_STAGING2
190     WHERE total_laid_off IS NULL AND
191     percentage_laid_off IS NULL;
192
193 • DELETE
194     FROM LAYOFFS_STAGING2
195     WHERE total_laid_off IS NULL AND
196     percentage_laid_off IS NULL;
197
198 • SELECT *
199     FROM LAYOFFS_STAGING2 ;
```

Object Info Session

Output



THANK YOU