

SQLite

Michael Levine

Background: Using SQLite, let's explore a Human Resources dataset ("Core_Dataset.csv"). The dataset itself comes from Dr. Rich Huebner, director of data science and architecture at Houghton Mifflin Harcourt in Boston, MA. Please refer to the citation page for more information. Moreover, this tutorial was created while using SQLite Online, an interactive SQLite browser. For more information on it, please also refer to the citation page at the end of this document.

Scenario: Let's suppose we have a database within a Human Resources department of a company. The department is looking to study the top three sources its highest performing employees were hired from (Monster.com, a job fair, etcetera). It then wants to interview those employees to better understand how they interacted with those ads. To do this, we need to first identify what employment source most of our highest performing employees came from. Then we need to pull the employee names associated with those employment sources.

To begin, the 'Core_Dataset.csv' file was then uploaded to SQLite Online. Now let's first see what each of the fields are in the database:

```
PRAGMA table_info(Employees);
```

cid	name	type	notnull	dflt_value	pk
0	Last Name		0	null	0
1	First Name		0	null	0
2	Employee Number		0	null	0
3	State		0	null	0
4	Zip		0	null	0
5	DOB		0	null	0
6	Age		0	null	0
7	Sex		0	null	0
8	MaritalDesc		0	null	0
9	CitizenDesc		0	null	0
10	Hispanic/Latino		0	null	0
11	RaceDesc		0	null	0
12	Date of Hire		0	null	0
13	Date of Termination		0	null	0
14	Reason For Term		0	null	0
15	Employment Status		0	null	0
16	Department		0	null	0
17	Position		0	null	0
18	Pay Rate		0	null	0
19	Manager Name		0	null	0
20	Employee Source		0	null	0
21	Performance Score		0	null	0

As we can see, we have several fields in our database. Now, let's see what the values are in the 'Performance Score' field.

SELECT DISTINCT

"Performance Score"

FROM

Employees;

Performance Score

Fully Meets

N/A- too early to review

90-day meets

Exceptional

Needs Improvement

Exceeds

PIP

We have several performance scores in this field. Now, let's select all active employees who were most highly rated in their most recent performance review. To do this, Let's limit our SQL statement to only "Exceeds" and "Exceptional" (and filter on "Active" from the Employment Status column). Here's our SQL statement:

SELECT

"First Name",

"Last Name",

"Employee Number",

"Date of Hire",

"Employment Status",

"Employee Source",

"Performance Score"

FROM

"Employees"

WHERE

("Performance Score" = "Exceptional" **OR** "Performance Score" = "Exceeds"
) **AND** ("Employment Status" = "Active")

This gives us the following output:

First Name	Last Name	Employee Number	Date of Hire	Employment Status	Employee Source	Performance Score
Jennifer	Zamora	1112030816	4/10/2010	Active	Employee Referral	Exceptional
Thomas	Murray	1406068403	11/10/2014	Active	Diversity Job Fair	Exceptional
Thelma	Petrowsky	1108027853	11/10/2014	Active	Employee Referral	Exceptional
Jason	Foss	1192991000	4/15/2011	Active	Professional Society	Exceptional
Eric	Dougall	1101023754	1/5/2014	Active	Professional Society	Exceeds
Leonara	Lindsay	602000312	1/21/2011	Active	Diversity Job Fair	Exceeds
Elisa	Bramante	1006020066	1/5/2009	Active	Other	Exceeds
Webster L	Butler	1110029990	1/28/2016	Active	Pay Per Click - Google	Exceeds
Ketsia	Liebig	1103024679	9/30/2013	Active	Website Banner Ads	Exceeds
Wilson K	Adinolfi	1409070522	7/5/2011	Active	MBTA ads	Exceeds
Linda	Anderson	1304055947	1/9/2012	Active	MBTA ads	Exceptional
Helen	Billis	1308060366	7/7/2014	Active	Professional Society	Exceeds
Evelyn	Girifalco	1204032927	9/29/2014	Active	Professional Society	Exceeds
Cayo	Gonzalez	1411071212	7/11/2011	Active	Diversity Job Fair	Exceeds
Kara	Harrison	1404066622	5/12/2014	Active	Monster.com	Exceeds
Shari	Ngodup	1403066020	4/1/2013	Active	Diversity Job Fair	Exceeds
Adeel	Osturnka	1404066949	9/30/2013	Active	On-campus Recruiting	Exceeds
Haley	Rivera	1405067642	11/28/2011	Active	Search Engine - Google Bing Yahoo	Exceeds
Elias	Robinson	1011022887	7/8/2013	Active	Employee Referral	Exceptional
Tayana	Jeannite	1008020942	7/5/2011	Active	MBTA ads	Exceeds
Yen	Johnston	1306057810	7/7/2014	Active	Professional Society	Exceptional
Lisa	Lunquist	1001504432	8/19/2013	Active	Glassdoor	Exceeds
Luisa	Monterro	1001103149	5/13/2013	Active	Billboard	Exceptional
Joe	Smith	1001970770	9/29/2014	Active	Pay Per Click - Google	Exceeds
Latia	Costa	1409070567	1/10/2011	Active	Other	Exceeds
Michael	Riordan	1502072711	1/9/2006	Active	Billboard	Exceeds
Andrew	Szabo	1201031324	7/7/2014	Active	MBTA ads	Exceptional

Now let's say we want to see how many of these employees came from each individual employee source. Let's write our new SQL statement:

```
SELECT
    "Employee Source",
    COUNT("Employee Source") AS "Number of Employees"
FROM
    Employees
WHERE
    "Employment Status" = "Active" AND ("Performance Score" = "Exceptional"
    OR "Performance Score" = "Exceeds")
GROUP BY
    "Employee Source"
ORDER BY COUNT
    ("Employee Source") DESC;
```

The output from our SQL statement above is as follows:

Employee Source	Number of Employees
Professional Society	5
Diversity Job Fair	4
MBTA ads	4
Employee Referral	3
Billboard	2
Other	2
Pay Per Click - Google	2
Glassdoor	1
Monster.com	1
On-campus Recruiting	1
Search Engine - Google Bing Yahoo	1
Website Banner Ads	1

"Professional Society," "Diversity Job Fair," and "MBTA Ads" are our top three employment sources. Now let's create a SQL statement that pulls the active and high performing employee names associated with those sources. From here, our Human Resources department will be able to see the names of the employees they'd like to interview:

SELECT

"First Name",
"Last Name",
"Employee Number",
"Date of Hire",
"Employment Status",
"Employee Source",
"Performance Score"

FROM

"Employees"

WHERE

("Performance Score" = "Exceptional" OR "Performance Score" = "Exceeds"
)

AND ("Employment Status" = "Active")

AND ("Employee Source" = "Professional Society" OR "Employee Source" =
"Diversity Job Fair" OR "Employee Source" = "MBTA ads")

ORDER BY

"Employee Source";

First Name	Last Name	Employee Number	Date of Hire	Employment Status	Employee Source	Performance Score
Thomas	Murray	1406068403	11/10/2014	Active	Diversity Job Fair	Exceptional
Leonara	Lindsay	602000312	1/21/2011	Active	Diversity Job Fair	Exceeds
Cayo	Gonzalez	1411071212	7/11/2011	Active	Diversity Job Fair	Exceeds
Shari	Ngodup	1403066020	4/1/2013	Active	Diversity Job Fair	Exceeds
Wilson K	Adinolfi	1409070522	7/5/2011	Active	MBTA ads	Exceeds
Linda	Anderson	1304055947	1/9/2012	Active	MBTA ads	Exceptional
Tayana	Jeannite	1008020942	7/5/2011	Active	MBTA ads	Exceeds
Andrew	Szabo	1201031324	7/7/2014	Active	MBTA ads	Exceptional
Jason	Foss	1192991000	4/15/2011	Active	Professional Society	Exceptional
Eric	Dougall	1101023754	1/5/2014	Active	Professional Society	Exceeds
Helen	Billis	1308060366	7/7/2014	Active	Professional Society	Exceeds
Evelyn	Girifalco	1204032927	9/29/2014	Active	Professional Society	Exceeds
Yen	Johnston	1306057810	7/7/2014	Active	Professional Society	Exceptional

Sources

Huebner, R., & Patalano, C. (n.d.). Human Resources Data Set. Retrieved July 13, 2019, from

<https://www.kaggle.com/rhuebner/human-resources-data-set>

SQLite Online. (n.d.). Retrieved July 13, 2019, from <https://sqliteonline.com/>