codecademy

Funnel Project

Learn SQL from Scratch

Declan Tidd

June 15 2018

Example Table of Contents

1.1 Original Table

Here is the original table, opened with the code below. We can see the five “quiz” questions that each user is asked, and how many distinct users answered each question

```
select question, count(distinct user_id)
from survey
group by 1;
```

question	count(distinct user_id)
1. What are you looking for?	500
2. What's your fit?	475
3. Which shapes do you like?	380
4. Which colors do you like?	361
5. When was your last eye exam?	270

Percentages calculated using Excel

Using Excel, we are able to calculate the percentage of users answered each question.

It seems the biggest drop in percentages happens after the second question, and after the fourth question

question	users answering	percent answering	drop in percentage answering		total users
1. what are you looking for	500	100%	-		500
2. what's your fit	474	95%	5%		
3. which shapes do you like	380	76%	19%		
4. which colors do you like	361	72%	4%		
5. when was your last eye exam	270	54%	18%		

Individual Tables

- Before we can analyze the data we need, we need to look at all the data we have. In this case, we have three separate tables with repeated data, and data we don't really need.

```
select *
from quiz
limit 5;
select *
from home_try_on
limit 5;
select *
from purchase
limit 5;
```

user_id	style	fit	shape	color
4e8118dc-bb3d-49bf-85fc-cca8d83232ac	Women's Styles	Medium	Rectangular	Tortoise
291f1cca-e507-48be-b063-002b14906468	Women's Styles	Narrow	Round	Black
75122300-0736-4087-b6d8-c0c5373a1a04	Women's Styles	Wide	Rectangular	Two-Tone
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	Women's Styles	Narrow	Square	Two-Tone
ce965c4d-7a2b-4db6-9847-601747fa7812	Women's Styles	Wide	Rectangular	Black

user_id	number_of_pairs	address
d8addd87-3217-4429-9a01-d56d68111da7	5 pairs	145 New York 9a
f52b07c8-abe4-4f4a-9d39-ba9fc9a184cc	5 pairs	383 Madison Ave
8ba0d2d5-1a31-403e-9fa5-79540f8477f9	5 pairs	287 Pell St
4e71850e-8bbf-4e6b-accc-49a7bb46c586	3 pairs	347 Madison Square N
3bc8f97f-2336-4dab-bd86-e391609dab97	5 pairs	182 Cornelia St

user_id	product_id	style	model_name	color	price
00a9dd17-36c8-430c-9d76-df49d4197dcf	8	Women's Styles	Lucy	Jet Black	150
00e15fe0-c86f-4818-9c63-3422211baa97	7	Women's Styles	Lucy	Elderflower Crystal	150
017506f7-aba1-4b9d-8b7b-f4426e71b8ca	4	Men's Styles	Dawes	Jet Black	150
0176bfb3-9c51-4b1c-b593-87edab3c54cb	10	Women's Styles	Eugene Narrow	Rosewood Tortoise	95
01fdf106-f73c-4d3f-a036-2f3e2ab1ce06	8	Women's Styles	Lucy	Jet Black	150

HERE WE CAN COMBINE THE TABLES USING A LEFT JOIN TO ONLY GET THE INFORMATION THAT WE WANT

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc-bb3d-49bf-85fc-cca8d83232ac	1	3 pairs	0
291f1cca-e507-48be-b063-002b14906468	1	3 pairs	1
75122300-0736-4087-b6d8-c0c5373a1a04	0	Ø	0
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	1	5 pairs	0
ce965c4d-7a2b-4db6-9847-601747fa7812	1	3 pairs	1
28867d12-27a6-4e6a-a5fb-8bb5440117ae	1	5 pairs	1
5a7a7e13-fbcf-46e4-9093-79799649d6c5	0	Ø	0
0143cb8b-bb81-4916-9750-ce956c9f9bd9	0	Ø	0
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	1	5 pairs	0
b1dded76-cd60-4222-82cb-f6d464104298	1	3 pairs	0

```
SELECT DISTINCT Q.USER_ID, H.USER_ID IS NOT NULL
AS 'IS_HOME_TRY_ON', H.NUMBER_OF_PAIRS,
P.USER_ID IS NOT NULL AS 'IS_PURCHASE'

FROM QUIZ Q

LEFT JOIN HOME_TRY_ON H ON Q.USER_ID =
H.USER_ID

LEFT JOIN PURCHASE P ON P.USER_ID = Q.USER_ID

LIMIT 10;
```

It’s easiest to work with this data if we use a “with” statement. After that, we can look at the data as a whole rather than by individual users. Here we can see which percentage of users tried something on, and which percentage of users actually purchased something

num_browse	num_is_try_on	num_of_pairs	percent_trying_on	percent_purchase
10	7	27.0	0.7	0.3

```
with funnel as (select distinct q.user_id, h.user_id is not null as 'is_home_try_on', h.number_of_pairs, p.user_id is not null as 'is_purchase'
from quiz q
left join home_try_on h on q.user_id = h.user_id
left join purchase p on p.user_id = q.user_id
limit 10)
select count(*) as 'num_browse', sum(is_home_try_on) as 'num_is_try_on', sum(number_of_pairs) as 'num_of_pairs',
1.0*sum(is_home_try_on)/count(user_id) as 'percent_trying_on', 1.0* sum(is_purchase)/count(user_id) as 'percent_purchase'
from funnel;
```

SQL allows us to play with this data however we please. Using two simple lines, we can determine the average price of our products

```
select count(*), avg(price)
from purchase;
```

count(*)	avg(price)
495	112.717171717172

Or the group our products by their color

```
select color, count(color)
from purchase
group by color
order by 2 desc;
```

color	count(color)
Jet Black	86
Driftwood Fade	63
Rosewood Tortoise	62
Rose Crystal	54
Layered Tortoise Matte	52
Pearled Tortoise	50
Elderflower Crystal	44
Sea Glass Gray	43
Endangered Tortoise	41