codecademy

Usage Funnels with Warby Parker

Learn SQL from Scratch

Shirin Mohajer

March, 24th 2019

Table of Contents

1. What columns does the “survey” table have?
2. What is the number of responses for each question?
3. *Which question(s) of the quiz have a lower completion rates? What do you think is the reason?*
4. Finding out whether or not users who get more pairs to try on at home will be more likely to make a purchase

    Analysis: Calculate overall conversion rates based on aggregating across all rows.

    Compare conversion from quiz→home_try_on and home_try_on→purchase

    Calculate the difference in purchase rates between customers who had 3 vs 5 number_of_pairs
5. Analysis: most common results of the style quiz.
6. Analysis: most common types of purchase made.

1. What columns does the “survey” table have?

1. What columns does the “survey” table have?

The users' responses are stored in a table called ”survey”. The easiest way to answer this question is to select some rows from this table. Here, we are going to limit our select to 10, and based on the result we can say this table has three coeloms as follow:

- question
- user_id
- response

(Here is a sample of what going to be displayed in the output)

```
-- Quiz Funnel

SELECT *

FROM survey

LIMIT 10;
```

question	user_id	response
1. What are you looking for?	005e7f99-d48c-4fce-b605-10506c85aaf7	Women's Styles
2. What's your fit?	005e7f99-d48c-4fce-b605-10506c85aaf7	Medium
3. Which shapes do you like?	00a556ed-f13e-4c67-8704-27e3573684cd	Round
4. Which colors do you like?	00a556ed-f13e-4c67-8704-27e3573684cd	Two-Tone

2. What is the number of responses for each question?

2. What is the number of responses for each question?

Users will "give up" at different points in the survey. In this section, we are going to analyze how many users move from Question 1 to Question 2, etc.

Here, we are going to create a quiz funnel using the GROUP BY command.

This query:

- displays each “question”
- COUNTs number of DISTINCT “user_id”s
- Groups results based on the “question”s

```
SELECT question,
    COUNT(DISTINCT(user_id))
FROM survey
GROUP BY question;
```

Result

Basically, this query counts how many times a question got answered by DISTICT “user_id”s and displays each COUNT in front of each “question”.

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

3. Which question(s) of the quiz have a lower completion rates? What do you think is the reason?

3. Which question(s) of the quiz have a lower completion rates? What do you think is the reason?

To answer this question

1. From previous question, we know how many users responded to each survey question. Now, using a spreadsheet program like Excel or Google Sheets, we can calculate the percentage of users who answer each question.
2. In the second step we can analyze the results.

3.1 calculate the percentage of users who answer each question.

- From the previous question we got the table on the right.
- Next, we can calculate the percentage of completion for each question. To calculate that, using an Excel Sheet, we can divide the number of people completing each step by the number of people completing the previous step.

Question	Dividing by previous	Percentage of Completion
1. What are you looking for?	500/500	%100
2. What's your fit?	475/500	%95
3. Which shapes do you like?	380/475	%80
4. Which colors do you like?	361/380	%95
5. When was your last eye exam?	270/361	%74.79

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

3.2 Analysis: Which question(s) of the quiz have a lower completion rates? What do you think is the reason?

The last question “When was your last eye exam?” has the lowest completion rate.
This can be because this a personal and medical question and also people might not clearly remember the date.
Another low rate question is the second question and it is about “Which shapes do you like?”. This question can be hard to answer because people probably cannot decide until they try it.

4. Finding out whether or not users who get more pairs to try on at home will be more likely to make a purchase.

4.1 Examine the first five rows of each table What are the column names?

In this part, we are going to look into the “quiz”, “home_try_on”, and “purchase” tables to get a better understanding of our date. We are going to limit our result into 5 rows.

quiz table has the following columns:

- user_id
- style
- fit
- shape
- color

home_try_on table has the following columns:

- user_id
- number_of_pairs
- address

purchase table has the following columns:

- user_id
- product_id
- style
- model_name
- color
- price

```
SELECT *
FROM quiz
LIMIT 5;

SELECT *
FROM home_try_on
LIMIT 5;

SELECT *
FROM purchase
LIMIT 5;
```

quiz

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

home_try_on

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

purchase

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

4.2 Combine Three Tables to create a new layout

In this part, we are going to create a new table with the following layout:

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc	True	3	False
291f1cca	True	5	False
75122300	False	NULL	False
user_id	is_home_try_on	number_of_pairs	is_purchase

To do so, we need to Left Join all these tables with the query on the right side.

- If the user has any entries in home_try_on, then is_home_try_on will be ‘True’.
- number_of_pairs comes from home_try_on table
- If the user has any entries in is_purchase, then is_purchasewill be ‘True’.
- Select only the first 10 rows from this table

```
SELECT DISTINCT q.user_id,
  CASE
    WHEN h.user_id IS NOT NULL THEN 'True'
    ELSE 'False'
  END AS 'is_home_try_on',
   h.number_of_pairs,
  CASE
   WHEN p.user_id IS NOT NULL THEN 'True’
   ELSE 'False'
  END 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;
```

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

4.3 Analysis: Calculate overall conversion rates based on aggregating across all rows.

We should aggregate all these three tables (quiz, home_try_on, purchase) and create a funnel which shows users who tried pairs at home, number of pairs and whether not they purchased anything.
This funnel can be later used to do more analysis on different marketing strategies used by the company.

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc-bb3d-49bf-85fc-cca8d83232ac	True	3 pairs	False
291f1cca-e507-48be-b063-002b14906468	True	3 pairs	True
75122300-0736-4087-b6d8-c0c5373a1a04	False	0	False
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	True	5 pairs	False

```
-- aggregating across all rows
with funnel as (
SELECT DISTINCT q.user_id,
    CASE
WHEN h.user_id IS NOT NULL THEN 'True'
ELSE 'False'
END AS 'is_home_try_on',
h.number_of_pairs,
CASE
  WHEN p.user_id IS NOT NULL THEN 'True'
ELSE 'False'
END 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)
```

4.3.1 Calculating overall conversion rate

Here, we are trying to answer this question on what percentage of people who tried pairs at home (regardless of the number), eventually ended up buying something.

To do so, we should divide number of pairs purchased by total number of pairs tried at home.

The answer is .66 which means %66.

conversion_rate
0.66

```
-- Calculating Conversion Rate: How many people who tried on at home, ended up purchaseing
SELECT
1.0* sum(
CASE
    WHEN funnel.is_home_try_on == 'True'
            and funnel.is_purchase = 'True'
     THEN 1
ELSE 0
END
)
/sum(
CASE
WHEN funnel.is_home_try_on == 'True' then 1
ELSE 0
END ) as conversion_rate
from funnel;
```

4.3.2 Compare conversion from quiz→home_try_on and home_try_on→purchase

To calculate the conversion rate for home_try_on→purchase and quiz→home_try_on,

we should find the percentage of the people who took he quiz and request try on pairs, and then made purchase. We can use the query on the right side to find these numbers.

Results show that not everybody who took the quiz, asked to try glasses at home, and not every one who tried pairs at home made purchase. We Can also use an Excel sheet to find conversion rates.

brows	home_try	purchased	brows_to_checkout	checkedout_to_purchase
1000	750	495	0.75	0.66
1000/1000	750/1000	495/750		
%100	%75	%66		

```
-- aggregating across all rows
with funnel as (
SELECT DISTINCT q.user_id,
                    h.user_id IS NOT NULL AS 'is_home_try_on',
  p.user_id IS NOT NULL AS 'is_purchased'
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)
SELECT
count (user_id) AS 'brows',
SUM(is_home_try_on) AS 'home_try',
SUM(is_purchased) AS 'purchased',
1.0 * SUM(is_home_try_on)/ count (user_id)
AS brows_to_checkout,
1.0 * SUM(is_purchased) / SUM(is_home_try_on)
AS checkedout_to_purchase
FROM funnel;
```

4.3.3 calculate the difference in purchase rates between customers who had 3 number_of_pairs with ones who had 5.

To calculate and compare the conversion rates for the customer who had 3 pairs and who had 5 pairs we can retrieve some more columns from the database. The result shows people who received 5 pairs to try a little bit more likely to make purchase. ~ %76 compare to ~ 74. The different is not very high and we also need to consider sending more pairs to customers is more costly and there will be a higher number products in the hand customers for review while they can be presented to others instead.

brows	home_try	purchased	brows_to_checkout	checkedout_to_purchase
1000	750	495	0.75	0.66

three_pairs	five_pairs
0.7656565656566	0.74949494949495

```
with funnel as (
SELECT DISTINCT q.user_id, h.number_of_pairs,
                    h.user_id IS NOT NULL AS 'is_home_try_on',
  p.user_id IS NOT NULL AS 'is_purchased'
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)
SELECT
count (user_id) AS 'brows',
SUM(is_home_try_on) AS 'home_try',
SUM(is_purchased) AS 'purchased',
1.0 * SUM(is_home_try_on)/ count (user_id)
AS brows_to_checkout,
1.0 * SUM(is_purchased) / SUM(is_home_try_on)
AS checkedout_to_purchase,
 1.0* SUM(number_of_pairs like '%3%')/SUM (is_purchased)
                    AS 'three_pairs',
 1.0* SUM(funnel.number_of_pairs like '%5%')/SUM (is_purchased)
                    AS 'five_pairs'
FROM funnel;
```

5. Analysis: most common results of the style quiz

5.1 Most common results of the style quiz. Popular (style)s

Here we are trying to find out how is the popularity of each different styles based on the answers we collect from the style quiz.
We can use the query on the right side to find out different possible answers, number of responses and percentage of each.

style	count_num	percentage
I'm not sure. Let's skip it.	99	9.9
Men's Styles	432	43.2
Women's Styles	469	46.9

```
/*
Women's Styles
Men's Styles
I'm not sure. Let's skip it.
*/
SELECT style, COUNT(style) AS 'count_num',
 100. * COUNT(style)
/(SELECT COUNT(style) FROM quiz ) AS 'percentage'
FROM quiz
GROUP BY style;
```

5.2.1 (style)s conversion based on the answers to style quiz.

Here we are trying to find out that for each separate answers, how many of them took pairs home for a try on and how many of them ended up purchasing something.

style	count_num	is_home_try_on	is_purchased
I'm not sure. Let's skip it.	99	69	0
Men's Styles	432	320	243
Women's Styles	469	361	252

```
SELECT q.style,
    COUNT(q.style) AS 'count_num',
    COUNT(h.user_id) AS 'is_home_try_on',
    COUNT(p.user_id) AS 'is_purchased'
FROM quiz q
LEFT JOIN purchase p
ON
  q.user_id = p.user_id
LEFT JOIN home_try_on h
ON
   q.user_id = h.user_id
GROUP BY q.style;
```

5.2.2 (style)s conversion rate based on the answers to style quiz

TO get a better understanding we need to calculate the conversion rate for each step using the query on the right side.

style	count_num	is_home_try_on	is_purchased
I'm not sure. Let's skip it.	99	69	0
Men's Styles	432	320	243
Women's Styles	469	361	252

```
SELECT q.style,
    COUNT(q.style) AS 'count_num',
    COUNT(h.user_id) AS 'is_home_try_on',
    COUNT(p.user_id) AS 'is_purchased'
FROM quiz q
LEFT JOIN purchase p
ON
  q.user_id = p.user_id
LEFT JOIN home_try_on h
ON
   q.user_id = h.user_id
GROUP BY q.style;
```

5.2.2 (style)s conversion rate based on the answers to style quiz

TO get a better understanding we need to calculate the conversion rate for each step using the query on the right side.

```
SELECT q.style,
                COUNT(q.style) AS 'count_num',
                100. * COUNT(q.style)
  /(SELECT COUNT(style) FROM quiz) AS
'style_percentage',
                COUNT(h.user_id) AS
'is_home_try_on',
                100. * COUNT(h.user_id)
  /(SELECT COUNT(style) FROM quiz) AS
'home_try_percentage',
                COUNT(p.user_id) AS
'is_purchased',
                100. * COUNT(p.user_id)
  /(SELECT COUNT(style) FROM quiz) AS
'percentage'
FROM quiz q
LEFT JOIN purchase p
ON
                q.user_id = p.user_id
LEFT JOIN home_try_on h
ON
                q.user_id = h.user_id
GROUP BY q.style;
```

Result Shows that out 1000 answers, 99 people said “I'm not sure. Let's skip it.” People who skipped this question, took 6.0 percent of the home_try_on, but they did not purchase anything. People who were looking for “Men’s Style” were %43.2 of the answers and they took home %32 of the pairs and purchased %24.3. People who answered “Women’s Style” were %46.9 percent of the answers and they took home %32 of the pairs and purchased %25.2. Result shows that people who are looking for women’s style glasses are more likely to try more glasses and do more purchase. We can do the same analysis for the rest of the quiz columns.

	number_answered_quiz					
	1000					
style	count_num	style_percentage	is_home_try_on	home_try_percentage	is_purchased	Purchased_percentage
I'm not sure. Let's skip it.	99	9.9	69	6.9	0	0.0
Men's Styles	432	43.2	320	32.0	243	24.3
Women's Styles	469	46.9	361	36.1	252	25.2

5.2.3 (color)s conversion rate based on the answers to style quiz

We can do the same analysis on colors too. We just need to replace style -> color. Result shows the most popular color is black and the second most popular color is tortoise. Therefore, we can invest on more black and tortoise pairs.

number_answered_quiz
1000

color	count_num	color_percentage	is_home_try_on	home_try_percentage	is_purchased	purchased_percentage
Black	280	28.0	220	22.0	150	15.0
Crystal	210	21.0	165	16.5	104	10.4
Neutral	114	11.4	79	7.9	48	4.8
Tortoise	292	29.2	213	21.3	144	14.4
Two-Tone	104	10.4	73	7.3	49	4.9