Warby Parker Project

Jonas Persson

2018-07-08

Question 1

- **Question:** Select all columns from the first 10 rows. What columns does the table have?
- **Answer:** question, user_id, response

Query:

```
SELECT *
FROM survey
LIMIT 10;
```

Output:

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
1. What are you looking for?	00a556ed-f13e-4c67-8704-27e3573684cd	I'm not sure. Let's skip it.
2. What's your fit?	00a556ed-f13e-4c67-8704-27e3573684cd	Narrow
5. When was your last eye exam?	00a556ed-f13e-4c67-8704-27e3573684cd	<1 Year
3. Which shapes do you like?	00bf9d63-0999-43a3-9e5b-9c372e6890d2	Square
5. When was your last eye exam?	00bf9d63-0999-43a3-9e5b-9c372e6890d2	<1 Year
2. What's your fit?	00bf9d63-0999-43a3-9e5b-9c372e6890d2	Medium

Question 2

- **Question:** What is the number of responses for each question?
- **Answer:** Q1 = 500, Q2 = 475, Q3 = 380, Q4 = 361, Q5 = 270

Query:

```
SELECT question,
   COUNT(DISTINCT user_id)
FROM survey
GROUP BY 1;
```

Output:

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

Question 3

- **Question:** Which question(s) of the quiz have a lower completion rates? What do you think is the reason?
- **Answer:** Question 5 ”When was your last eye exam?” (75%) followed by question 3 ”Which shapes do you like?” (80%) had the lowest completion rates. Regarding question 5 there is a response available called ”Not Sure. Let’s Skip It” which could be utilized by customers who are unsure of when their last eye exam took place, however, as indicated by the data, many people still chose not to answer this question at all. That could indicate that people for some reason are unwilling to give up this personal information about themselves. Regarding question 3 there is an option available called ”No Preference” for customers who might not have a particular preference in regards to the glasses shape. Yet, as indicated by the data, many people still are unable to complete this question. The shape of the glasses does not appear to be as private a question as perhaps question 5 was, and maybe the likely reason for people not answering question 3 as opposed to question 5 is perhaps that they are unable to find the kind of shape that they are looking for among the provided answer alternatives or it could be that they simply have a problem making their mind up as to what shape they would like and realise that they need to think a bit more about it, rather than them withholding a response for the reason of not wanting to give away personal information.

Output:

Question Number	Percent Completing this Question	(Calculation)
1	100%	(500/500)
2	95 %	(475/500)
3	80 %	(380/475)
4	95 %	(361/380)
5	75 %	(270/361)

Question 4

- **Question:** Examine the first five rows of each table What are the column names?
- **Answer:**
  quiz: user_id, style, fit, shape, color
  home_try_on: user_id, number_of_pairs, address
  purchase: user_id, product_id, style, model_name, color, price

Queries:

```
SELECT *
FROM quiz
LIMIT 5;
```

Outputs:

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

```
SELECT *
FROM home_try_on
LIMIT 5;
```

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

```
SELECT *
FROM purchase
LIMIT 5;
```

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

Question 5

- **Question:** Use a LEFT JOIN to combine the three tables (quiz, home_try_on, purchase), starting with the top of the funnel (browse) and ending with the bottom of the funnel (purchase).

Query:

```
SELECT DISTINCT q.user_id,
    h.user_id IS NOT NULL 'is_home_try_on',
    h.number_of_pairs,
    p.user_id IS NOT NULL 'is_purchase'
 FROM quiz AS 'q'
 LEFT JOIN home_try_on AS 'h'
   ON q.user_id = h.user_id
 LEFT JOIN purchase AS 'p'
   ON p.user_id = q.user_id
 LIMIT 10;
```

Output:

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

Question 6

- **Question:**

1. We can calculate overall conversion rates by aggregating across all rows.
2. We can compare conversion from quiz→home_try_on and home_try_on→purchase.

Answer: The overall conversion rate and quiz→home_try_on and home_try_on→purchase-rates are calculated using a temporary funnels-table created by a WITH-statement:

Query:

```
 WITH funnels AS (SELECT DISTINCT q.user_id,
    h.user_id IS NOT NULL 'is_home_try_on',
    h.number_of_pairs,
    p.user_id IS NOT NULL 'is_purchase'
 FROM quiz AS 'q'
 LEFT JOIN home_try_on AS 'h'
   ON q.user_id = h.user_id
 LEFT JOIN purchase AS 'p'
   ON p.user_id = q.user_id)
SELECT COUNT(*) AS 'num_quiz',
  SUM(is_home_try_on) AS 'num_home_try_on',
  SUM(is_purchase) AS 'num_purchase',
  1.0 * SUM(is_purchase) / COUNT(user_id) AS 'overall_conv_rate',
  1.0 * SUM(is_home_try_on) / COUNT(user_id) AS 'quiz_to_home_try_on',
  1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS 'home_try_on_to_purchase'
FROM funnels;
```

Output:

num_quiz	num_home_try_on	num_purchase	overall_conv_rate	quiz_to_home_try_on	home_try_on_to_purchase
1000	750	495	0.495	0.75	0.66

Comment: About 50% (49.5) of all people who go on the website end up making a purchase (overall conversion rate). 75% of the people who go on the website end up completing the quiz and hence go on to the home-try-on stage, and of those people 66% end up actually puchasing a pair.

Question 6 (continued)

- **Question:**

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

Answer: Using the same query as in the previous question it is possible to compare the stated purchase rates by grouping by ”number_of_pairs”

Query:

```
WITH funnels AS (SELECT DISTINCT q.user_id,
                                 h.user_id IS NOT NULL 'is_home_try_on',
   h.number_of_pairs,
   p.user_id IS NOT NULL 'is_purchase'
 FROM quiz AS 'q'
 LEFT JOIN home_try_on AS 'h'
   ON q.user_id = h.user_id
 LEFT JOIN purchase AS 'p'
   ON p.user_id = q.user_id)
SELECT number_of_pairs,
  COUNT(*) AS 'num_quiz',
  SUM(is_home_try_on) AS 'num_home_try_on',
  SUM(is_purchase) AS 'num_purchase',
  1.0 * SUM(is_home_try_on) / COUNT(user_id) AS 'quiz_to_home_try_on',
  1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS 'home_try_on_to_purchase'
FROM funnels
GROUP BY number_of_pairs
ORDER BY number_of_pairs;
```

Output:

number_of_pairs	num_quiz	num_home_try_on	num_purchase	quiz_to_home_try_on	home_try_on_to_purchase
Ø	250	0	0	0.0	Ø
3 pairs	379	379	201	1.0	0.530343007915567
5 pairs	371	371	294	1.0	0.792452830188679

Comment: Of the 1000 people who went on the webiste, 250 did not complete the quiz and thus did not go on to the home-try-on stage, 379 had a 3-pair home-try-on and 371 had a 5-pair home-try-on. The people who received the 5-pair home-try-on were considerably more likely to go through with completing a purchase (about 79% ended up buying) as compare to the people who received the 3-pair home-try-on (about 53% ended up buying).

Question 6 (continued)

- **Question:**

4. The most common results of the style quiz.

Answer: By selecting ”response” and ”question” and counting the distinct user_id from the survey-table and grouping by ”response” and ordering by the COUNT(user_id) we obtain a table of the most common answers to the quiz:

Query:

```
SELECT response,
  question,
  COUNT(DISTINCT user_id)
FROM survey
GROUP BY 1
ORDER BY 3 DESC;
```

Output:

response	question	COUNT(DISTINCT user_id)
Men's Styles	1. What are you looking for?	242
Women's Styles	1. What are you looking for?	209
Narrow	2. What's your fit?	208
<1 Year	5. When was your last eye exam?	141
Rectangular	3. Which shapes do you like?	141
Medium	2. What's your fit?	132
Square	3. Which shapes do you like?	119
Tortoise	4. Which colors do you like?	117
Black	4. Which colors do you like?	112
I'm not sure. Let's skip it.	1. What are you looking for?	92
Round	3. Which shapes do you like?	91
Wide	2. What's your fit?	88
Crystal	4. Which colors do you like?	69
1-3 Years	5. When was your last eye exam?	56
3+ Years	5. When was your last eye exam?	37
Neutral	4. Which colors do you like?	36
Not Sure. Let's Skip It	5. When was your last eye exam?	36
No Preference	3. Which shapes do you like?	29
Two-Tone	4. Which colors do you like?	27

Comment: The most and next most common answers were in relation to question 1 ”What are you looking for?”. This is not a big surprise since this is the first question of the quiz and it only has two alternatives (men’s or women’s) meaning both of the alternatives will get a big number of replies as opposed to many of the other questions which have many more options in terms of possible answers. The table also tells us that the most popular shape is ”rectangular” (141), the most popular fit is ”medium” (132), the most popular color is ”Tortoise” (117) and most people had an eye exam less than a year ago (141).

Question 6 (continued)

- **Question:**

5. The most common types of purchase made.

Query:

```
SELECT product_id,
   style,
   model_name,
   color,
   price,
   COUNT(DISTINCT user_id)
       AS 'tot_num_purchases'
FROM purchase
GROUP BY 1
ORDER BY 6 DESC;
```

Output:

product_id	style	model_name	color	price	tot_num_purchases
3	Men's Styles	Dawes	Driftwood Fade	150	63
10	Women's Styles	Eugene Narrow	Rosewood Tortoise	95	62
9	Women's Styles	Eugene Narrow	Rose Crystal	95	54
1	Men's Styles	Brady	Layered Tortoise Matte	95	52
6	Women's Styles	Olive	Pearled Tortoise	95	50
4	Men's Styles	Dawes	Jet Black	150	44
7	Women's Styles	Lucy	Elderflower Crystal	150	44
2	Men's Styles	Brady	Sea Glass Gray	95	43
8	Women's Styles	Lucy	Jet Black	150	42
5	Men's Styles	Monocle	Endangered Tortoise	50	41

Comment: The most popular model configuarion is the Dawes model in Driftwood Fade color in the men's styles range with a total number of purchases of 63. The most popular women's styles model configuration is the Eugene Narrow in Rosewood Tortoise (62). The least popular model is the men's styles Monocle in Endangered Tortoise (41). Perhaps the name of this color option (endangered tortoise) is putting some people off even though they actually like the color? Time for a name change?

Question 6 (continued)

- **Question:**

Extra question: We can find the most popular style, model and color among the glasses purchased:

Query:

```
--most popular style
SELECT style,
  COUNT(DISTINCT user_id)
   AS 'tot_num_purchases'
FROM purchase
GROUP BY 1
ORDER BY 2 DESC;

--most popular model
SELECT model_name,
  style,
  COUNT(DISTINCT user_id)
   AS 'tot_num_purchases'
FROM purchase
GROUP BY 1
ORDER BY 3 DESC;

--most popular color
SELECT color,
  style,
  COUNT(DISTINCT user_id)
   AS 'tot_num_purchases'
FROM purchase
GROUP BY 1
ORDER BY 3 DESC;
```

Output:

style	tot_num_purchases
Women's Styles	252
Men's Styles	243

model_name	style	tot_num_purchases
Eugene Narrow	Women's Styles	116
Dawes	Men's Styles	107
Brady	Men's Styles	95
Lucy	Women's Styles	86
Olive	Women's Styles	50
Monocle	Men's Styles	41

color	style	tot_num_purchases
Jet Black	Women's Styles	86
Driftwood Fade	Men's Styles	63
Rosewood Tortoise	Women's Styles	62
Rose Crystal	Women's Styles	54
Layered Tortoise Matte	Men's Styles	52
Pearled Tortoise	Women's Styles	50
Elderflower Crystal	Women's Styles	44
Sea Glass Gray	Men's Styles	43
Endangered Tortoise	Men's Styles	41

Question 6 (continued)

- **Question:**

Extra question: We can calculate the average price of sold glasses as well as group by style:

Query (average price):

```
SELECT AVG(price),
  style,
  color,
  COUNT(DISTINCT user_id) AS 'tot_num_purchases'
FROM purchase
ORDER BY 1 DESC;
```

Output:

AVG(price)	tot_num_purchases
112.717171717172	495

Query (average price grouped by style):

```
--calculate average model price of sold glasses
 SELECT AVG(price),
  style,
  COUNT(DISTINCT user_id) AS 'tot_num_purchases'
FROM purchase
GROUP BY 2
ORDER BY 1 DESC;
```

Output:

AVG(price)	style	tot_num_purchases
113.769841269841	Women's Styles	252
111.625514403292	Men's Styles	243

Comment: The average sales price is 114$ for Women's styles glasses and 112$ for Men's styles glasses. Hence there is not much of a price difference if any between the glasses sold in the two ranges and they are both close to the average sales price (113$), which perhaps would indicate that both styles have a good distribution of glasses in different price categories available to the customer.