codecademy

Warby Parker Purchase Data

Learn SQL from Scratch

Jack Owens

December 2018

Table of Contents

Warby Parker Style Quiz

```
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
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

2. *Users will "give up" at different points in the survey. Let's analyze how many users move from Question 1 to Question 2, etc.*

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

column1	column2
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

Calculate the percentage of users who answer each question.:

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

column1	column2	column3
1. What are you looking for?	500	1
2. What's your fit?	475	.95
3. Which shapes do you like?	380	.76
4. Which colors do you like?	361	.72
5. When was your last eye exam?	270	.54

Question 5 has the lowest completion rate followed by 4 and 3. This is likely because they are later in the survey and suffer from those who did not complete question 2. These questions are also personal preference questions that users may either be less concerned with or unsure about.

Question 5 in particular is a more difficult question to answer than 4 and 3 because it may be difficult for some individuals to remember

4. Warby Parker Purchase funnel is: Take the Style Quiz → Home Try-On → Purchase the Perfect Pair of Glasses
During the Home Try-On stage, we will be conducting an A/B Test:
50% of the users will get **3** pairs to try on
50% of the users will get **5** pairs to try on
Let's find out whether or not users who get more pairs to try on at home will be more likely to make a purchase.
The data will be distributed across three tables:

- quiz
- home_try_on
- purchase

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

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

User_id, style, fit, shape, color

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

User_id, number_of_pairs, address

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

User_id, product_id, style, model_name, color, price

QUIZ	HOME_TRY_ON	PURCHASE
User_id	**User_id**	**User_id**
style	**Number_of_pairs**	**Product_id**
fit	**address**	**style**
shape		**Model_name**
color		**Color, price**

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

User_id, style, fit, shape, color

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**

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

User_id, number_of_pairs, address

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**

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

User_id, product_id, style, model_name, color, price

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

5. Each row will represent a single user from the browse table:
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_purchase will be 'True'.
Use a LEFT JOIN to combine the three tables, starting with the top of the funnel (browse) and ending with the bottom of the funnel (purchase).
Select only the first 10 rows

```
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;
```

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		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		0
0143cb8b-bb81-4916-9750-ce956c9f9bd9	0		0
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	1	5 pairs	0
b1dded76-cd60-4222-82cb-f6d464104298	1	3 pairs	0

Larger Sample Size of home_try_on Query

```
SELECT DISTINCT q.user_id AS
      user_id,
(SELECT COUNT(*)
FROM home_try_on ht WHERE
      ht.user_id = q.user_id) 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 h.user_id = q.user_id
LEFT JOIN purchase p
  ON p.user_id = q.user_id
WHERE h.user_id IS NOT NULL
LIMIT 30;
```

OBJECTIVES OF NEW QUERY:

Want to analyze larger data set of the population who got to try on the glasses at home. Did the USERS who got 3 PAIRS to try on actually purchase more frequently than USERS who got 5 PAIRS?

*In extending the Limit of the sample to 30, I found that 14/30 or 46.66% of the home try on group actually bought a pair of glasses (TABLE ON NEXT SLIDE *abbriged)*

Of the 14, 11 received 5 pairs (78.57%)

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
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
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	1	5 pairs	0
b1dded76-cd60-4222-82cb-f6d464104298	1	3 pairs	0
9fc1bcfe-1c3b-4b78-bb3b-af3586c2f05c	1	5 pairs	1
20b03d28-d39c-46cf-81af-9fb479e823c0	1	5 pairs	1
ffe1b116-6f09-4408-9aba-f0d268c67fbe	1	3 pairs	0

What Else Increased Purchases?

```
SELECT DISTINCT q.user_id,
    h.user_id IS NOT NULL AS
'is_home_try_on',
    h.number_of_pairs,
    p.model_name, p.color,
    p.price,   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;
```

OBJECTIVES OF NEW QUERY:

Want to analyze other factors that may have played a role in purchase decisions for the home_try_on group including PRICE, MODEL NAME, and COLOR from the PURCHASE table

user_id	is_home_try_on	number_of_pairs	model_name	color	price	is_purchase
4e8118dc-bb3d-49bf-85fc-cca8d83232ac	1	3 pairs				0
291f1cca-e507-48be-b063-002b14906468	1	3 pairs	Eugene Narrow	Rosewood Tortoise	95	1
75122300-0736-4087-b6d8-c0c5373a1a04	0					0
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	1	5 pairs				0
ce965c4d-7a2b-4db6-9847-601747fa7812	1	3 pairs	Eugene Narrow	Rose Crystal	95	1
28867d12-27a6-4e6a-a5fb-8bb5440117ae	1	5 pairs	Olive	Pearled Tortoise	95	1
5a7a7e13-fbcf-46e4-9093-79799649d6c5	0					0
0143cb8b-bb81-4916-9750-ce956c9f9bd9	0					0
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	1	5 pairs				0
b1dded76-cd60-4222-82cb-f6d464104298	1	3 pairs				0

Actionable Insights for Warby Parker *all based on 30-person sample size

- **Continue home try on method as purchase success rate is high at 46%**
- **5 pair try on responsible for 78.57% of home try on purchases – send 5**
- **Eugene Narrow appears as most popular Model (3/7 purchases or 42.86%). Sell more.**
- **$95 purchase price occurred on 4/6 purchases 66%**