code|cademy

Usage Funnels with Warby Parker
Learn SQL from Scratch – Capstone Project

Patrick Marmo
pmarmo@digiscribe.info

Course Dates: June 5, 2018 – July 31, 2018

Table of Contents

1. Warby Parker Overview
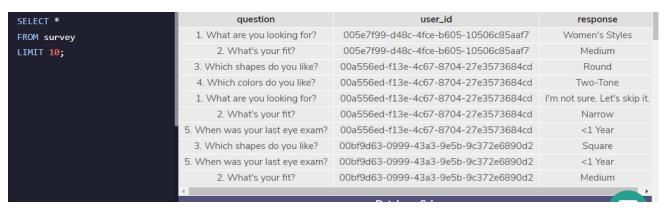2. Warby Parker's Quiz Funnel
3. Warby Parker's Home Try-On Funnel

1. Warby Parker Overview

Question 1 - Warby Parker Quiz Funnel

Warby Parker is a designer eyewear company that offers an online marketplace where potential customers can browse the selection of eyewear they offer. Warby Parker provides customers with a free 'Style Quiz' that can be used to help determine what type of eyewear might best suit their own personal style.

Warby Parker is interested in knowing how far customers move along in their quiz, and would like to analyze what type of questions might cause users to close out of the quiz without fully completing it. Using SQL, we are able to analyze their database to see what questions they offer, how many users complete each question, and what percentage of users move onto the next question.

Using a simple SELECT query, we see that the 'survey' table (where the resulting quiz data is stored) structure contains three columns storing three pieces of data: the QUESTION asked of the customer, the USER_ID the website assigned the customer, and the RESPONSE provided by the customer.

```
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

In the query results we can see the questions asked in the dataset returned, but this query makes it hard to be certain we're seeing all the questions. A separate SELECT DISTINCT query shows us that there are indeed only 5 questions, and using ORDER BY we can be sure they are sorted in the order we desire by inserting ASC or DESC after providing the sort column.

```
SELECT DISTINCT question
FROM survey
ORDER BY question ASC
LIMIT 10;
```

question
1. What are you looking for?
2. What's your fit?
3. Which shapes do you like?
4. Which colors do you like?
5. When was your last eye exam?

2. Warby Parker's Quiz Funnel

Question 2 - Warby Parker's Quiz Funnel (Cont)

The queries we've run so far have only provided us with basic data stored in the survey table. In order to provide more useful information and direction to Warby Parker regarding the quiz results, we need to see how many users moved from question 1, to question 2, and so on. This is called a funnel and is extremely useful in analyzing how users move through a given process, whether it be a 5 question quiz or a 10 step e-commerce purchase.

Using a simple select query allows us to specify that we want to query the 'question' column. Including the COUNT operation on the 'question' column allows us to also count that column, while GROUP BY allows us to group our results by each question. Finally, using 'AS' allows us to rename the columns in our query results to whatever values desired; in this case, capitalizing 'question' and naming the count column to "COUNT".

```
SELECT question AS
'QUESTION',
        COUNT(question) AS
'COUNT'
FROM survey
GROUP BY question
;
```

QUESTION	COUNT
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 - Warby Parker's Quiz Funnel (Cont)

	A	B	C
1	Survey Question	# User Completing	Completion Rate (as %)
2	1. What are you looking for?	500	100
3	2. What's your fit?	475	95
4	3. Which shapes do you like?	380	80
5	4. Which colors do you like?	361	95
6	5. When was your last eye exam?	270	74.79224377

The data returned from this funnel tells us that the very first question was answered by 500 people. This is our starting user count and will only decrease from here as users stop moving forward with the quiz. Using Excel to calculate the completion rate on each question, we see that Questions 3 and 5 have lower completion rates compared to questions 1, 2, and 5.

This would point to questions 3 and 5 being questions that users may not want answer. Question 3, asking about shapes, might limit users unnecessarily; maybe users want to be able to select more than 1 shape when taking the quiz. Question 5, relating to the date of a user's last eye exam, could turn users off by making them think they have to know the exact date of their eye exam in order to complete the quiz.

Whatever the case may be, with questions 3 and 5 having much lower completion rates (a 15% drop from question 2 to 3, and a 10% drop from question 4 to 5), Warby Parker may want to consider replacing those quiz questions with some less restrictive or less personal ones in order to encourage higher completion rates.

3. Warby Parker's Home Try On Funnel

Question 4 - Warby Parker's Home Try-On Funnel

Using SQL, we can look at data across multiple tables and combine the data in those different tables to provide a more useful set of data for analysis.
For example, using SQL to query the 3 tables below (quiz, home_try_on, and purchase) we can see that each table has a unique set of columns:
- 'quiz' contains 'user_id', 'style', 'fit', 'shape', and 'color'
- 'home_try_on' contains 'user_id', 'number_of_pairs', and 'address'
- 'purchase' contains 'user_id', 'product_id', 'style', 'model_name', 'color', and 'price
Note that the user_id column is contained in all 3 tables. Since this unique value is stored in all 3 tables, this column will become the unique key that we use to match the data in all 3 tables together and create a combined set of data. We will do this using JOIN commands, found in the next section.

```
SELECT *
FROM quiz
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

```
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. Warby Parker's Home Try-On Funnel (Cont)

Using the LEFT JOIN command in SQL, we are able to combine the data from the 3 tables noted above using the 'user_id' column that is contained in each table. Using the JOIN command, we start with the 'quiz' table and join the 'home_try_on' table to it using the 'user_id' in each table as our key. After that, we then join the 'purchase' table to our 'quiz' table, again joining on the 'user_id' column. We have now combined all 3 tables into one. Once combined, any columns that were not populated in a particular table will be null in our new joined table. In order to be able to read the new table clearly, the CASE statements shown below allow us to specify that when a particular column is populated (i.e. not blank or 'NULL'), then to output the value TRUE in the field, otherwise output 'False". This will allow us to see if a user completed a home try-on, how many pairs they tried, and if they subsequently made a purchase.

```
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 as 'q'
LEFT JOIN home_try_on as 'h'
    ON q.user_id = h.user_id
LEFT JOIN purchase as 'p'
    ON q.user_id = p.user_id
LIMIT 10;
```

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
ce965c4d-7a2b-4db6-9847-601747fa7812	True	3 pairs	True
28867d12-27a6-4e6a-a5fb-8bb5440117ae	True	5 pairs	True
5a7a7e13-fbcf-46e4-9093-79799649d6c5	False	0	False
0143cb8b-bb81-4916-9750-ce956c9f9bd9	False	0	False
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	True	5 pairs	False
b1dded76-cd60-4222-82cb-f6d464104298	True	3 pairs	False

```
WITH newtable 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 as 'q'
LEFT JOIN home_try_on as 'h'
    ON q.user_id = h.user_id
LEFT JOIN purchase as 'p'
    ON q.user_id = p.user_id)
SELECT COUNT(*) AS 'Total',
  SUM(is_home_try_on) AS 'Number Home Try-Ons',
  SUM(is_purchase) AS 'Number Purchases',
  1.0*SUM(is_home_try_on) / COUNT(user_id) AS 'Home Try-On %',
  1.0*SUM(is_purchase) / SUM(is_home_try_on) AS 'Purchase %'
FROM newtable
;
```

Once we have the results of combined table above, now we can start aggregating the data to see what the results of Warby Parker's home try-on test were. Using the query shown to the left to aggregate our combined table, we can see that we had a total of 1000 users take the style quiz. Of those 1000 users, we can see that 750 went forward with a home try on. Of that 750 home try-ons, 495 users actually made a purchase. This tells us that 75% of all survey respondents tried Warby Parker's products on at home after completing the style quiz. Of those 750 home trials, 495 of them resulted in a purchase, which is a 66% purchase rate of the trials that were completed. This is an OVERALL CONVERSION RATE since it provides the overall purchase rate for this quiz -> home try-on -> purchase funnel.

Query Results				
COUNT(*)	Number Home Try-Ons	Number Purchases	Home Try-On %	Purchase %
1000	750	495	0.75	0.66

Question 6 - Analyzing the A/B Test Results of Warby Parker

The results shown above indicate that overall, 750 home try-ons were completed and 495 purchases were made of those 750 try-ons. But what of our A/B test where some users were given 3 pairs of glasses to try on while other users were given 5? By adding some additional logic to our query to group our results based on the 'number_of_pairs' column from the 'home_try_on' table, we can now see our results based on the number of pairs provided to the user to try on. This allows us to do a side-by-side comparison.

```sql
project.sqlite

WITH newtable 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 as 'q'
LEFT JOIN home_try_on as 'h'
    ON q.user_id = h.user_id
LEFT JOIN purchase as 'p'
    ON q.user_id = p.user_id)
SELECT number_of_pairs AS 'Num_Pairs',
    SUM(is_home_try_on) AS 'Number Home Try-Ons',
    SUM(is_purchase) AS 'Number Purchases',
    1.0*SUM(is_purchase) / SUM(is_home_try_on) AS 'Purchase %'
FROM newtable
GROUP BY number_of_pairs
ORDER BY number_of_pairs ASC
;
```

Query Results

Num_Pairs	Number Home Try-Ons	Number Purchases	Purchase %
Ø	0	0	Ø
3 pairs	379	201	0.530343007915567
5 pairs	371	294	0.792452830188679

Database Schema

home_try_on		750 rows
user_id		TEXT
number_of_pairs		TEXT
address		TEXT

purchase		495 rows
user_id		TEXT
product_id		INTEGER
style		TEXT
model_name		TEXT
color		TEXT
price		INTEGER

survey		1986 rows
question		TEXT
user_id		TEXT
response		TEXT

quiz		1000 rows

By grouping our data this way, we can see that with the A (i.e. 3 pair) test group there were 379 Home Try-Ons with 201 purchases, for a 53% purchase rate. By comparison, we can see that with the B (i.e. 5 pair) test group there were 371 Home Try-Ons with 294 purchases, for a 79% purchase rate. This shows us that even though both groups started with a similar number of home try-on users, there was a 26% increase in the number of purchases for the variant group that had a higher number of pairs to try-on!

So what does this data tell us? First off, this is a huge increase and should indicate to Warby Parker that they should continue to offer a 5 pairs of sunglasses instead of 3 during home trials in order to boost their home-try-on to purchase rate. However, these test results only take a single variant group into account in changing from 3 pairs to try-on at home to 5 pairs. What if Warby Parker increased the number of home-try on pairs of glasses to 7? Or even 10? This data indicates that they should possibly try higher numbers to see if they can boost the home try-on to purchase ratio even more.