codecademy

Capstone Project:
Usage Funnels with Warby Parker

Learn SQL from Scratch

Bertie Sheridan

25th September 2018 Class

Table of Contents

1. Getting familiar with Warby Parker

1.1 Getting familiar with Warby Parker

Warby Parker is a ‘transformative lifestyle brand’ producing designer glasses and sunglasses. The company focuses on creative thinking, smart design, & doing good in the world – for every pair of glasses sold, another pair is donated to someone in need.

We’ve been provided with four sets of data relating to two aspects of Warby Parker’s marketing. The first table, `survey`, contains details of user’s answers to a set of questions relating to their preferences. This will be analysed to see how far user’s progressed through the survey.

The other three tables, `quiz`, `home_try_on`, and `purchase`, will be combined to see how far users progressed through these three stages of the Warby Parker marketing funnel, analysing how many of the customers who completed the quiz went on to the ‘Home Try On’, and how many of those customers went on purchase Warby Parker’s product.

We’ll see what other insights we can pick up on the way.

The remainder of this section will provide some more detail on the data tables listed above.

1.2 survey table details

* All fields are in TEXT format unless otherwise stated

Survey A table describing which answers customers gave to a set of survey questions.	
Column	**Description***
question	The question asked to the user
user_id	A unique identifier for each customer
response	The answer given by each user to the above question

Survey Questions

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?

```
SELECT DISTINCT question
FROM survey
ORDER BY 1;
```

Query 1.2a

Sample Data

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

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

Query 1.2b

1.3 `quiz` table details

* All fields are in TEXT format unless otherwise stated

Quiz A table containing the results of a quiz relating to customer preferences	
Column	**Description***
user_id	A unique identifier for each customer
style	The customers preferred style category
fit	The customers preferred fit
shape	The customers preferred shape
color	The customers preferred color

Sample Data

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
28867d12-27a6-4e6a-a5fb-8bb5440117ae	Women's Styles	Narrow	Rectangular	Black
5a7a7e13-fbcf-46e4-9093-79799649d6c5	Women's Styles	Wide	Rectangular	Tortoise
0143cb8b-bb81-4916-9750-ce956c9f9bd9	Women's Styles	Wide	Rectangular	Two-Tone
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	Women's Styles	I'm not sure. Let's skip it.	Square	Tortoise

```
SELECT *
FROM quiz
LIMIT 9;
```

(Selection reduced due to lack of space)

Query 1.3

1.4 home_try_on table details

* All fields are in TEXT format unless otherwise stated

home_try_on A table with details of those customers who used the 'home try on' facility	
Column	**Description***
user_id	A unique identifier for each customer
number_of_pairs	The number of pairs of glasses shipped for 'home try on'
address	The shipping address used for the customer

Sample Data

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
4c10e298-53c8-4009-adda-bbcaecb7e8b6	5 pairs	312 Frawley Cir
5a3ee321-517d-4a21-a351-d6815ab2edd5	5 pairs	301 Fred Douglass Cir
4d895ccf-4877-4f13-8183-13d7d0a20a47	3 pairs	77 Margaret Corbin Dr
39e8a811-75b9-4dc3-bdff-c92b6db0431d	3 pairs	482 Coenties Slip
9d2656a0-d066-4b42-bce1-77825f34ded9	5 pairs	364 De Peyster St

```
SELECT *
FROM home_try_on
LIMIT 10;
```

Query 1.4

1.5 purchase table details

* All fields are in TEXT format unless otherwise stated

Purchase A table containing details relating to customer orders	
Column	**Description***
user_id	An identifier for each unique customer
product_id	(INTEGER) The product code for each item sold
style	The style category for the item sold
model_name	The product name for each item sold
color	The color of the item sold
price	(INTEGER) The price record for each item sold

```
SELECT *
FROM purchases
LIMIT 8;



(Selection reduced due
to lack of space)
```

Query 1.5

Sample Data

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
021901a5-74ee-482d-9f03-0089d17ff7d6	3	Men's Styles	Dawes	Driftwood Fade	150
026e11d9-926c-4206-af7c-273e3ee6ad7f	2	Men's Styles	Brady	Sea Glass Gray	95
028b47b0-e075-4b2c-bcf3-f963d8089449	10	Women's Styles	Eugene Narrow	Rosewood Tortoise	95

2. Analysing the Quiz Funnel

2.1 Style Quiz Analysis

What is the number of responses for each question?
The number of responses to each question are shown in the table to the right.

What is the percentage of users that answer each question?
The percentage of users that answered each question are shown in the table to the right.

Which question(s) of the quiz have a lower completion rates?
The completion rates for questions 3, 4, & 5 are significantly lower than questions 1 & 2. The rates drop significantly between questions 2 & 3, and again between questions 4 & 5.

What do you think is the reason?
It's possible that the completion rates drop between questions 2 & 3 so significantly because the answers to questions 3 & 4 are likely be very subjective. Answers to 1 & 2 are absolute answers, making them somewhat easier to answer.

Whilst the answer to question 5 is also likely to be absolute, it is likely that the person completing the survey doesn't have this information to hand, or that they have never had an eye exam, especially if they are looking for sunglasses, as oppose to corrective glasses.

Completion rates between the questions never increase, suggesting that once the user finds a question that they can't answer, they tend to give up on the quiz.

question	#	%
1. What are you looking for?	500	100.0
2. What's your fit?	475	95.0
3. Which shapes do you like?	380	76.0
4. Which colors do you like?	361	72.2
5. When was your last eye exam?	270	54.0

```
SELECT question, COUNT(*) AS '#', (100.0 *
COUNT(*) / 500) AS '%'
FROM survey
GROUP BY 1
ORDER BY 1;
```

Query 2.1

3. Analysis of A/B Testing with Home Try On Funnel

3.1 Setting up the Home Try On Funnel

In the customer journey being examined in this marketing funnel, the customer follows three steps:

1. Complete a preferences quiz.
2. Opt for a ‘Home Try On’.
3. Purchase a pair of glasses.

The first objective of the analysis is to determine, of the customers that complete the preferences quiz, how many opt for a Home Try On, and how many then purchase a pair of glasses.

The second objective of the analysis is to determine how the number of glasses that a potential customer receives as part of the Home Try On influences the likelihood of the purchasing a pair of glasses.

In order to do this, we need to combine the data from three different tables (`quiz`, `home_try_on`, and `purchase`) in to a temporary table that we can query in order to conduct the analysis. This is done using the query to the top right. This query starts with all of the unique User ID’s in the quiz table (i.e. all of the users that took the quiz), then determines whether that user opted for a Home Try On (True/False), if so, how many pairs of glasses were received, and finally, whether that user purchased any glasses (True/False). See the table to the right for a sample of what this temporary table looks like.

```
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 q.user_id = p.user_id
LIMIT 10;
```

Query 3.1

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

3.2 Analysing the Home Try On Funnel

With the temporary table on the previous slide as a sub-query, we can perform an aggregate query (by counting the User ID's, and summing the 'True' values for Home-Try-On's and Purchases) to see how many users have completed each step. These figures are summarised in the table below, with a further table indicating the percentage conversions.

So, 75% of customers that take the quiz opt for a 'Home Try On', and 49.5% of customers that take the quiz purchase a pair of glasses. Overall, about a 50% quiz to purchase rate.

```
WITH purchase_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 q.user_id = p.user_id)
SELECT COUNT (*) AS 'quizzes_complete', SUM(is_home_try_on) AS
'count_home_try_on', SUM(is_purchase) AS 'count_purchases'
FROM purchase_funnel;
```

Query 3.2

quizzes_complete	count_home_try_on	count_purchases
1000	750	495

quizzes_complete	count_home_try_on	count_purchases
100%	75%	49.5%

3.3 Further Analysis of the Home Try On Funnel

Using the same temporary table, but adjusting the main query slightly as per the query on the right), we can examine how the purchase conversion rate changes depending on whether an individual received 3 pairs of glasses vs. 5 pairs of glasses as part of their Home Try On.

In order to make this analysis a little more meaningful, we'll adjust our conversion percentage to express purchases as a percentrage of Home Try On's, rather than as a percentage of quizzes complete (495 / 750 = 66.0%).

```
WITH purchase_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 q.user_id = p.user_id)
SELECT number_of_pairs, SUM(is_purchase) AS 'count_purchase',
SUM(is_home_try_on) AS 'count_home_try_on', (1.0 * SUM(is_purchase) /
SUM(is_home_try_on))
FROM purchase_funnel
WHERE number_of_pairs IS NOT NULL
GROUP BY 1
ORDER BY 1 DESC;
```

Query 3.3

number_of_pairs	count_purchases	count_home_try_on	%
5 pairs	294	371	79.2
3 pairs	201	379	53.0

The results of this query show that there is a significantly improved purchase conversion rate when users are given 5 pairs of glasses to try on vs. 3 pairs. This may be because when more glasses are received by the potential customer, there is a higher chance that the user will find a pair that they like.

The results of this A/B test would suggest that Warby Parker should issue 5 pairs of glasses for every home try on, though they may need to conduct further analysis, for example, cost analysis, in order to determine whether this is viable.

50% of customers were due to receive 3 pairs of glasses, and the other 50% 5 pairs of glasses. It's probably of little significance, but we can see from the above that a slightly higher number of customers received 3 pairs.

4. Further Insights for Warby Parker

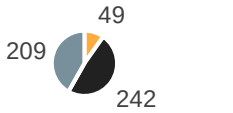
4.1 survey results distribution

Using the query to the right, we're able to produce the data that underlies the graphical analysis of the survey question answers provided below.

```
SELECT question, response, COUNT(*)
FROM survey
GROUP BY 1, 2
ORDER BY 1, 2;
```

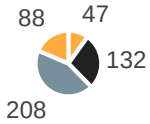
Query 4.1

Question 1: What (style) are you looking for?



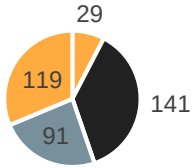
- I'm not sure. Let's skip it.
- Men's Styles
- Women's Styles

Question 2: What's your fit?



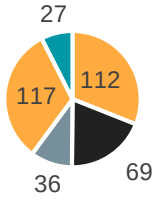
- I'm not sure. Let's skip it.
- Medium
- Narrow
- Wide

Question 3: Which shapes do you like?



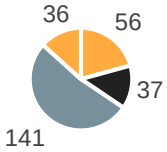
- No Preference
- Rectangular
- Round
- Square

Question 4: Which colors do you like?



- Black
- Crystal
- Neutral
- Tortoise
- Two-Tone

Question 5: When was your last eye exam?



- 1-3 Years
- 3+ Years
- <1 Year
- Not Sure. Let's Skip It

This information could provide useful insight into customer preferences, helping, for example, inventory planning and future design development.

From this data we can see that 'Narrow' is the most common frame fit, 'Rectangular' is the most common shape preference, and it's more or less a two way tie between black and tortoise for the most popular colour.

Similar data from purchase information tells a slightly different story overleaf.

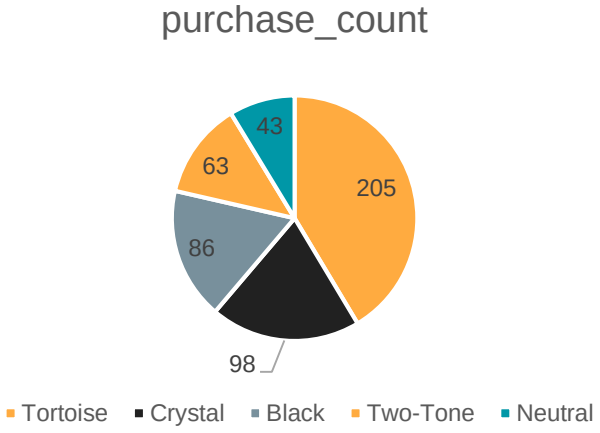
4.2 purchase preference distribution

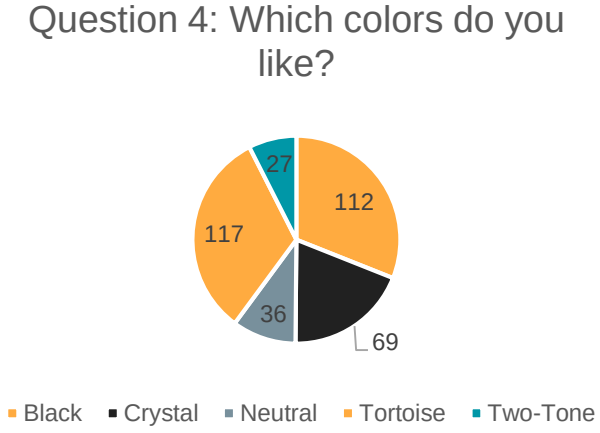
In order to ascertain the preferences of customers who purchased glasses, we need to use a slightly more complex query involving CASE and LIKE statements to categorise the purchases in a way that aligns this data with the survey data. See the new query to the right that retrieves data about the color preferences of customers that purchased glasses.

The resulting pie chart shows that ‘Tortoise’ is by far and away the most popular color category purchased, and also that there is otherwise little alignment between the two data sets.

```
SELECT CASE
    WHEN color LIKE ('%black%') THEN 'Black'
    WHEN color LIKE ('%tortoise%') THEN
'Tortoise'
    WHEN color LIKE ('%crystal%') THEN 'Crystal'
    WHEN color LIKE ('%driftwood%') THEN 'Two-
Tone'
    WHEN color LIKE ('Sea Glass Gray') THEN
'Neutral'
END AS 'color_category', COUNT (*)
FROM purchase
GROUP BY 1
ORDER BY 2 DESC;
```

Query 4.2a





Each data set has it’s own strength’s and weaknesses, for example, the purchase dataset is based on a larger sample (495 vs. 361), but purchases is retrospective information, whereas the survey could be prospective customers.

It’s hard to extract too much meaning from the comparison. I quick INNER JOIN (shown below) shows that there are no common User ID’s between the datasets.

Query 4.2b

```
SELECT COUNT (*)
FROM survey
INNER JOIN purchase ON survey.user_id = purchase.user_id
WHERE survey.question LIKE ‘4. %’
```

4.3 Other Insights

Other simple aggregate queries can be used to gain insight into Warby Parker's customer preferences, based on the data that we've been provided with.

For example, the query to the right returns the number of glasses purchased at each price point. This may help the company understand the price sensitivity of their product, and therefore to make pricing decisions. It would seem that the most popular price point is $95.00 – the middle price point.

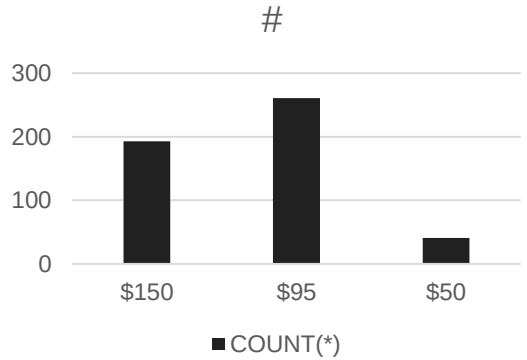
Alternatively, we could examine which model of glasses was most popular with customers that purchased glasses. See the query and graph below.

Again, this might help with promotions or inventory planning for future purposes if these trends are maintained over time.

The Eugene Narrow is the most popular product. The Monocle is the least popular product, and also the least expensive – the only product at the $50.00 price point. It may be worth Warby Parker further looking at the profitability of this product, as it may need to be cut.

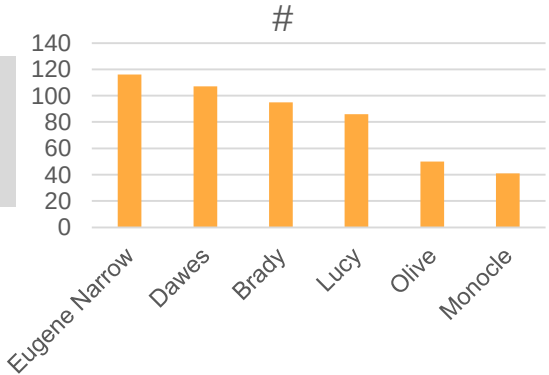
```
SELECT price, COUNT(*) AS '#'
FROM purchase
GROUP BY 1
ORDER BY 1 DESC;
```

Query 4.3a




```
SELECT model_name, COUNT(*) AS
'#'
FROM purchase
GROUP BY 1;
```

Query 4.3b



End of Presentation