codecademy

Warby Parker Capstone

Learn SQL from Scratch
Marcia Molettieri
August 7, 2018

Table of Contents

1. Get familiar with Warby Parker
2. Results of the Survey
3. A/B Testing
4. Conclusions & Interesting Notes

1. Get Familiar with Warby Parker

1.1 Get Familiar with Warby Parker, Survey

Warby Parker is an American online retailer for premier prescription eye wear. Founded in 2010, Warby Parker primarily sells prescription glasses and sunglasses through their website.

In order to help clients find the pair of glasses that is right for them, the company provides it's clients with a Style Survey, which consists of a few questions about their eye wear preferences.

The table below is a snapshot of the data that Warby Parker has collected from their surveys. This table is named "Survey." It contains the survey questions, responses to each question and a user ID (unique identifier) for each client that responded to the survey. This data will be further analyzed to help the company better understand their clients and ultimately produce better sales results moving forward.

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.

```
Query to find columns in
table named "Survey":

SELECT * FROM survey
 LIMIT 10;
```

2. Results of Survey

2.1 Results of Survey, Funnel

The Warby Parker Survey consists of 5 questions. The company has found that a decent number of clients do not complete the survey and ultimately "give up" along the way. If clients do not complete the survey, they do not purchase glasses. Warby Parker would like to know which questions clients are less likely to answer.

In order to determine the number of responses for each question, the table must be grouped by each of the 5 questions and the number of indiduals answering each question (user_id) must be counted.

```
Query to show funnel of questions answered
in survey:

SELECT question, COUNT(DISTINCT user_id)
          AS 'Answer Count'
          FROM survey
  GROUP BY question;
```

Questions	Answer 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

2.2 Results of Survey, Funnel by Percentage

The results below show each of the 5 question in the survey and how many clients responsed to each question.
- We notice that there is a significant drop in responses between Question 2 and Question 3 as well as Question 4 and Question 5

When Excel is used to calcuate the percent change between questions, we see that only 80% of clients that answered Question 2 moved forward to answer Question 3, and only 75% of clients that answered Question 4 moved forward to answer Question 5.
- The drop from Question 2 to Questions 3 may be contrubuted to the fact that there are a variety of shape options to pick from and without seeing the glasses on clients may not know which shapes they perfer.
- The drop from Question 4 to Question 5 is likely due to the fact that this is a more personal question. Clients may feel less comfortable providing information about their health/doctor's visits.

```
How to calculate percentage
change between questions:


Divide the number of clients
that completed each step by the
number of clients that
completed the previous step

For example:

Number of Answers to Question 2
--------------------------------
Number of Answers to Question 1
```

Questions	Answer 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
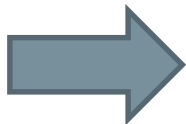
Questions	Percent Completed Question
1. What are you looking for?	100%
2. What's your fit?	95%
3. Which shapes do you like?	80%
4. Which colors do you like?	95%
5. When was your last eye exam?	75%

3. A/B Testing

3.1 A/B Testing, Tables

After clients have completed Warby Parker's Survey, they can have several pairs of glasses sent to their home to try-on to help them find their best fit. Warby Parker conducted an A/B test to determine if sending more try-on options resulted in more purchases.

The A/B Test consisted of:
- 50% of clients were given **3 pairs** to try-on at home
- 50% of clients were given **5 pairs** to try-on at home

In order to detemine whether clients that received more pairs were more likely to purchase a pair, three tables needed to be analyzed. The tables "Quiz," "Home_Try_On," and "Purchase" contain information about each client that started the quiz and potentially tried on pairs at home and potentially purchased pairs after the try-on.

```
The query results below were limited to
only 5 in order to save processing time

Query to see all columns in "Quiz"

SELECT * FROM quiz
LIMIT 5;


Query to see all columns in "Home_Try_On"

SELECT * FROM home_try_on
LIMIT 5;


Query to see all columns in "Purchase"

SELECT * FROM purchase
LIMIT 5;
```

3.1 A/B Testing, Tables

"Quiz" columns provide the user_id of clients who started the Survey as well as their responses to the first four quetsions

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" columns provide the user_id of clients who had pairs sent to their home as well as the number of pair and their home address

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" columns provide the user_id of clients who tried on pairs at home and purchased one as well as information about the pair they purchased

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

3.2 A/B Testing, Home Try-On Funnel

In order to see how many clients went through each step in the proccess, the three tables need to be joined on the user_id column.

By creating the table below named "Funnel", we can see all the clients (user_id's) that began or completed the survey, all the clients that tried on pairs at home, how many pairs they were given to try (3 or 5) and whether they purchased any of those pairs.

Because the "Is_home_try_on" and "is_purchased" columns were not preexisting in any of the original tables, they were created by identifyting whether the number_of_pairs column from "Home_try_on" and product_id column from "Purchase" were null when the three tables were left joined.

```
Query to show funnel from Quiz to Home_Try_On to Purchase

WITH funnel AS
(SELECT quiz.user_id,
            CASE
            WHEN hto.number_of_pairs IS NOT NULL THEN 'True'
    ELSE 'False'
    END AS 'Is_home_try_on',
  hto.number_of_pairs,
  CASE
            WHEN purchase.product_id IS NOT NULL THEN 'True'
    ELSE 'False'
    END AS 'Is_purchase'
FROM quiz
LEFT JOIN home_try_on AS 'hto'
            ON quiz.user_id = hto.user_id
LEFT JOIN purchase
            ON hto.user_id = purchase.user_id)
SELECT * FROM funnel
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		False
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	True	5 pairs	False

3.2 A/B Testing, Home Try-On Funnel

To make the analysis of the new table "Funnel" a little bit easier, the True/False statements can be changed to 1's and 0's. 1's represnting True and 0's represnting False.

This small change will allow for simpler aggragations, such as counting or summing the columns.

```
Query to show funnel from Quiz to Home_Try_On to Purchase

WITH funnel AS
(SELECT quiz.user_id,
        hto.number_of_pairs IS NOT NULL AS
'Is_Home_Try_On',
   hto.number_of_pairs,
   purchase.product_id IS NOT NULL AS 'Is_Purchase'
FROM quiz
LEFT JOIN home_try_on AS 'hto'
        ON quiz.user_id = hto.user_id
LEFT JOIN purchase
        ON hto.user_id = purchase.user_id)
SELECT * FROM funnel
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

3.3 A/B Testing, Funnel Calculations

When changing the True/False statements to 1's and 0's, the user_id column, Is_home_try_on column and Is_purchase column can be aggregated to see how many clients tried on pairs at home and how many of those clients purchased a pair.

The user_id column must be counted, while the Is_home_try_on and Is_purchase columns can be summed now that the columns contain 1's and 0's.

```
Query to show funnel from Quiz to Home_Try_On to Purchase

WITH funnel AS
(SELECT quiz.user_id,
         hto.number_of_pairs IS NOT NULL AS 'Is_Home_Try_On',
   hto.number_of_pairs,
   purchase.product_id IS NOT NULL AS 'Is_Purchase'
FROM quiz
LEFT JOIN home_try_on AS 'hto'
         ON quiz.user_id = hto.user_id
LEFT JOIN purchase
         ON hto.user_id = purchase.user_id)
SELECT COUNT(user_id) AS 'Quiz Count',
         SUM(Is_Home_Try_On) AS 'Try On Count',
   SUM(Is_purchase) AS 'Purchase Count'
FROM funnel;
```

Quiz Count	Try On Count	Purchase Count
1000	750	495

3.3 A/B Testing, Funnel Calculations

Once the aggregates have been calcuated, the conversion percentage of Quiz to Home Try On and Home Try On to Purchase can be calculated.

Through this query, we found that 75% of clients that started the Warby Parker Survey got pairs to try-on at home. Additioanlly, 66% of clients that tried on pairs at home purchased a pair. This means that just less that 50% of clients that started the survey went through the entire process and purchased a pair of glasses.

```
Query to show funnel from Quiz to Home_Try_On to Purchase

WITH funnel AS
(SELECT quiz.user_id,
            hto.number_of_pairs IS NOT NULL AS 'Is_Home_Try_On',
  hto.number_of_pairs,
  purchase.product_id IS NOT NULL AS 'Is_Purchase'
FROM quiz
LEFT JOIN home_try_on AS 'hto'
            ON quiz.user_id = hto.user_id
LEFT JOIN purchase
            ON hto.user_id = purchase.user_id)
SELECT
  1.0 * sum(is_home_try_on) / count(user_id) AS 'Quiz to Try-on',
  1.0 * sum(is_purchase) / sum(is_home_try_on) AS 'Try-on to
Purchase',
 ROUND( 1.0 * sum(is_purchase) / count(user_id), 2) AS 'Quiz to
Purchase'
FROM funnel;
```

Quiz to Try-on	Try-on to Purchase	Quiz to Purchase
0.75	0.66	0.49

3.3 A/B Testing, Funnel Calculations

In addition to determining overall conversion rates, we can determine the results of the A/B Test. By grouping the table "Funnel" by the number_of_pairs column and aggregating User_id, Is_Home_Try_On and Is_Purchase, we can see whether clients were more likely to purchase a pair with 3 or 5 try-on options.

We notice that significantly more clients purchased a pair of glasses if they received 5 pairs to try-on. Only 53% of clients that received 3 try-on options purchased a pair, but almost 80% of client that received 5 try-on opitions purchased a pair.

```
Query to determine A/B Test results:

WITH funnel AS
(SELECT quiz.user_id,
            hto.number_of_pairs IS NOT NULL AS 'Is_Home_Try_On',
   hto.number_of_pairs,
   purchase.product_id IS NOT NULL AS 'Is_Purchase'
FROM quiz
LEFT JOIN home_try_on AS 'hto'
            ON quiz.user_id = hto.user_id
LEFT JOIN purchase
            ON hto.user_id = purchase.user_id)
SELECT number_of_pairs,
            SUM(is_home_try_on) AS 'Number of Clients Tried on Pairs',
   SUM(is_purchase) AS 'Number of Clients Purchased Pair(s)'
FROM funnel
GROUP BY 1;
```

number_of_pairs	Number of Clients Tried on Pairs	Number of Clients Purchased Pair
NULL	0	0
3 pairs	379	201
5 pairs	371	294

4. Conclusions

4.1 Conclusions & Interesting Notes

1. Question 3 and Question 5 were less likely to be answered than other questions in the Warby Parker Survey.
    - Warby Parker may want to consider doing additional research into why this is the case and either a) reword the questions in the survey or b) remove the questions from the survey. They should consider this particualrly for Question 5.

2. 75% of clients that started the survey actually completed it and have pairs sent to their home to try-on. If Warby Parker can improve their survey to have more clients complete it, they can increase the number of clients that try-on pairs at home.

3. The A/B Test showed that clients were more likely to purchase a pair of glasses when they were given 5 options over 3 options.
    - Warby Parker should consider sending 5 pairs to all clients in the furute to increase the likelihood that client will purchase a try-on pair.

The most popular pair of glasses was purchased 63 times:
Product ID - 3, Model Name – Dawes

The prices for glasses spands from $50 to $150. The least expensive pair is surprsingly the least popular pair, while the most popular pair was at the high end of the price range.

product_id	model_name	price	Product ID Count
3	Dawes	150	63
10	Eugene Narrow	95	62
9	Eugene Narrow	95	54
1	Brady	95	52
6	Olive	95	50
4	Dawes	150	44
7	Lucy	150	44
2	Brady	95	43
8	Lucy	150	42
5	Monocle	50	41

```
SELECT product_id, model_name,
   price, COUNT(*) AS 'Product
ID Count'
  FROM purchase
  GROUP BY 1
  ORDER BY COUNT(*) DESC;
```