



# Uncovering Opportunity at Warby Parker

Learn SQL from Scratch

Matzen Shirley

01/01/2019

# Table of Contents

## Getting familiar with Warby Parker

- What columns does the table have?

## Survey Funnel Analysis & Conclusions

- What is the number of responses for each question?
- Which questions of the quiz have lower completion rates?
- What do I think is the reason?

## Warby Parker A/B Test: The merits of try-on options

- Are users who get more pairs to try on at home more likely to purchase?
- What are the column names of each of the tables?
- What are the over-all conversion rates for the funnel?
- How do conversion rates compare on a step-to-step basis?
- How do conversion rates compare between 3-pair, and 5-pair try-on tests?
- What are some actionable insights for Warby Parker?

# Getting familiar with Warby Parker

# Get familiar: Survey table

**Objective:** to identify the various columns and data elements in the 'Quiz' table

**Observations:** Each user can go through a maximum of 5 questions. The response column reveals that the data's structure mostly, if not entirely, consists of string values.

**Question #1: What columns does the [Survey] table have?**

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

## Results

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

# **Survey funnel analysis and conclusions**

# Survey table summary

**Background:** We should treat this survey data from the survey table as a funnel on its own. Therefore, our first step to zeroing in on areas for improvement is to identify the number of responses (or users) at each step.

**Observations:** As with any funnel, we see that people drop off in different amounts at each step. At first glance, it looks like there's a relatively large drop off from question #4 to question #5

**Next:** We will do some conversion rate calculations outside of this environment

question	users_responded
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 #2: What is the number of responses for each question?**

```
/* We want to GROUP BY the
questions column and use the
aggregating COUNT function
so that we can see the total
number of users that
completed each given
question in the survey */
```

```
1  SELECT question, COUNT(DISTINCT user_id)
   AS users_responded
2  FROM survey
3  GROUP BY 1;
```

# Survey table analysis

**Observations:** very small drop-offs in # of responses from question 1 to 2 and from 3 to 4. Conversely, the largest drop-offs are from **questions #2 to #3 and from #4 to #5**

**Initial Hypothesis':** Respondents that finished 4 but not 5 questions may have been experiencing voter fatigue to a greater degree

**Question #3a & #3b:** What questions have lower completion rates & what might be the reason?

Using excel, I calculated the completion rates for each question

question	users_responded	% completion
1. What are you looking for?	500	100%
2. What's your fit?	475	95%
3. Which shapes do you like?	380	80%
4. Which colors do you like?	361	95%
5. When was your last eye exam?	270	75%

```
1  SELECT question, COUNT(DISTINCT user_id)
   AS users_responded
2  FROM survey
3  GROUP BY 1;
```

# Survey table insights and conclusions

## Question fatigue setting in?

### Evidence:

- 12% of those who responded to 4 total quiz questions (call them 4's) selected 'No Preference' for question #3
- ONLY 6% of those who responded to 5 total quiz questions (call them 5's) selected 'No Preference' for question #3

**Assumption:** Let's assume that the selections of 'No Preference' or '...skip it' may be used as a proxy for quiz impatience – an indication of quiz-question fatigue

**Conclusion:** The disparity between the 'No-Preference' selection rate of the 4's and that of the 5's suggests that question fatigue may have been at least partially responsible for the relatively large dip in question 5's response rate. This association becomes even more plausible when considering the fact that for questions #1 and #2 for which 'No Preference' and '...skip it' are quiz options, the 4's and the 5's selected those options at very comparable rates, indicating that it was not until further into the quiz (question #3) that the 4's may have experienced question fatigue and gave up.

### Actions to take →

- Decrease quiz length or concatenate two questions into one while maintaining both clarity and brevity

## Question #4: What might be the reason for lower completion rates

**Step 1:** Generate a table that contains a user\_id in one column and the count of questions completed in the other column

**Step 2:** Join the new table with the survey table

**Step 3:** Filter the joined table to compare the 'No Preference' selection rates

```
WITH funnel_progress AS (SELECT user_id, COUNT(user_id) AS
number_answered
FROM survey
GROUP BY 1
ORDER BY 2 DESC)

SELECT response, COUNT(survey.user_id) AS num_response,
survey.user_id AS user_id
FROM survey
JOIN funnel_progress ON funnel_progress.user_id =
survey.user_id
WHERE question LIKE '3%' AND number_answered = 4
GROUP BY 1
ORDER BY 2 DESC;
```

response	num_response	user_id
Rectangular	35	ff8461f7-e500-458c-9087-98fa63562d99
Square	25	f78079b2-77ee-47d6-8431-880645ac145b
Round	20	fcbb78b4-cc17-48dc-9fbc-92619abf0242
No Preference	12	fceacac3-80ef-4090-bc6d-4a10f5d209db



# **Warby Parker A/B Test: The Merits of Try-On Options**

# A/B Test - Get Familiar: Quiz, Home\_Try\_On, and Purchase

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

user_id	number_of_pairs	address
d8add87-3217-4429-9a01-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

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: What are the column names (of the three tables)?**

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

**Column Names:** user\_id, style, fit, shape, color

-----

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

**Column Names:** user\_id, number\_of\_pairs, address

-----

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

**Column Names:** user\_ids, product\_id, style, model\_name, color, price

# A/B Test: Setting the table...

**Background:** In order for us to analyze the health of Warby Parker's purchase funnel from quiz to home-try on to purchase, we need to have the data from those three steps in one single table that we can then manipulate and query.

**Result:** Now we are ready to perform full-funnel analysis.

**\*\*The 'Null' fields in 'number\_of\_pairs' column indicate that the user took the quiz, but did participate in the home-try-on phase, thus they do not have any pairs assigned to them\*\***

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-a301-27dd93b12e2	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

/\*Combine the quiz, home\_try\_on, and purchase tables in order to get data ready for funnel analysis

Because home\_try\_on is a subset of quiz, and purchase is a subset of each, we will use a LEFT JOIN to combine 'home\_try\_on' with 'quiz' and another LEFT JOIN to combine 'purchase' with 'quiz'.\*/

```
153 SELECT DISTINCT q.user_id,
154 h.user_id IS NOT NULL AS 'is_home_try_on',
    h.number_of_pairs, p.user_id IS NOT NULL AS 'is_purchase'
155 FROM quiz AS 'q'
156 LEFT JOIN home_try_on AS 'h'
157 ON q.user_id = h.user_id
158 LEFT JOIN purchase AS 'p'
159 ON p.user_id = q.user_id
160 LIMIT 10;
```

# A/B Test: Single-Funnel Conversion Metrics

I used a WITH clause in order to isolate the table that now contains the data we need in order to perform single-funnel analysis across the original three tables we were given. I used the SUM aggregate function to identify the number of users that reached each step.

num_quiz	home_try_on	purchase_complete	try_on rate	purchase rate	full_conversion
1000	750	495	0.75	0.66	0.495

**Question #6:** The full-conversion rate for the funnel is 49.5%

**Question #7:** While the try\_on\_rate is a relatively healthy 75%, the purchase rate, at 66%, may contain larger weaknesses that we can address

**Question #6:** What are the over-all conversion rates for the funnel?

**Question #7:** How do conversion rates compare on a step-to-step basis?

```
162 WITH funnels AS (SELECT DISTINCT q.user_id,
163                    h.user_id IS NOT NULL AS 'is_home_try_on',
164                    h.number_of_pairs, p.user_id IS NOT NULL AS
165                    'is_purchase'
166 FROM quiz AS 'q'
167 LEFT JOIN home_try_on AS 'h'
168     ON q.user_id = h.user_id
169 LEFT JOIN purchase AS 'p'
170     ON p.user_id = q.user_id)
171
172 SELECT COUNT(user_id) AS 'num_quiz',
173        SUM(is_home_try_on) AS 'home_try_on',
174        SUM(is_purchase) AS 'purchase_complete',
175        1.0 * SUM(is_home_try_on) / COUNT(user_id)
176 AS 'try_on rate', 1.0 * SUM(is_purchase) /
177 SUM(is_home_try_on) AS 'purchase rate', 1.0
178 * SUM(is_purchase) / COUNT(user_ID) AS
179 'full_conversion'
180 FROM funnels;
```

# A/B Test: Multi-Funnel Conversion Comparison

**Background:** While conversion rate is important in determining the actions Warby Parker should consider, I thought it important to not rule out the possibility that other outcomes may differ between the two groups as well, such as the average price per pair purchased.

**Query Methodology:** I used the ROUND function to truncate the avg\_price result to two decimal places. The original result included a row with users who never reached the home\_try\_on phase. However, we're only concerned with users who reached 'home-try-on' and the resulting A/B test participation. To clean this up, I added a WHERE clause to filter for ONLY those users who had been included in either side of the A/B test.

number_of_pairs	home_try_on	purchase_complete	purchase rate	avg_price
3 pairs	379	201	0.530343007915567	113.26
5 pairs	371	294	0.792452830188679	112.35

**Question #8:** How do the 3-pair and 5-pair tests' conversion rates compare?

**Question #9:** Are users who receive more try-on pairs more likely to purchase?

```
140 WITH funnels 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',
    p.price
141 FROM quiz AS 'q'
142 LEFT JOIN home_try_on AS 'h'
143     ON q.user_id = h.user_id
144 LEFT JOIN purchase AS 'p'
145     ON p.user_id = q.user_id)
146
147 SELECT number_of_pairs,
    SUM(is_home_try_on) AS 'home_try_on',
    SUM(is_purchase) AS 'purchase_complete',
    1.0 * SUM(is_purchase) /
    SUM(is_home_try_on) AS 'purchase rate',
    ROUND(AVG(price),2) AS avg_price
148 FROM funnels
149 WHERE number_of_pairs = '3 pairs'
150     OR number_of_pairs = '5 pairs'
151 GROUP BY number_of_pairs;
```

# A/B Test: Results & Impact Projections

**Observations:** The 3-pair test group purchased at a 53% rate while the 5-pair test group did so at a 79% rate! The average price per purchase was roughly the same between both groups. Here's a realistic, yet hypothetical scenario that illustrates just how large the impact of these findings could be:

## 3 vs. 5-pair Revenue Generation Potential

### Assumptions:

- purchase rates from test remain constant as does average price
- # of try-ons is equal to the total number of users who went through the test

Group	# of try-ons	# purchased	\$ / pair	Total Rev. \$	Rev % Change
3-Pair Group	750	397	\$113	\$44,861	
5-Pair Group	750	593	\$112	\$66,416	
Revenue Delta				<b>+ \$21,555</b>	<b>+ 48%</b>

number_of_pairs	home_try_on	purchase_complete	purchase_rate	avg_price
3 pairs	379	201	0.530343007915567	113.26
5 pairs	371	294	0.792452830188679	112.35

**Question #10:** What are some actionable insights for Warby Parker?

```
140 WITH funnels 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',
p.price
141 FROM quiz AS 'q'
142 LEFT JOIN home_try_on AS 'h'
143 ON q.user_id = h.user_id
144 LEFT JOIN purchase AS 'p'
145 ON p.user_id = q.user_id)
146
147 SELECT number_of_pairs,
SUM(is_home_try_on) AS 'home_try_on',
SUM(is_purchase) AS 'purchase_complete',
1.0 * SUM(is_purchase) /
SUM(is_home_try_on) AS 'purchase_rate',
ROUND(AVG(price),2) AS avg_price
148 FROM funnels
149 WHERE number_of_pairs = '3 pairs'
150 OR number_of_pairs = '5 pairs'
151 GROUP BY number_of_pairs;
```

# A/B Test: Suggested Actions

## Suggested Actions:

- Push all users into the '5-pair' try-on funnel
- Run another test to see whether similar results can be achieved with 4 pairs – could save \$\$ on circulating inventory

Group	# of try-ons	# purchased	\$ / pair	Total Rev. \$	Rev % Change
3-Pair Group	750	397	\$113	\$44,861	
5-Pair Group	750	593	\$112	\$66,416	
Revenue Delta				<b>+ \$21,555</b>	<b>+ 48%</b>

number_of_pairs	home_try_on	purchase_complete	purchase rate	avg_price
3 pairs	379	201	0.530343007915567	113.26
5 pairs	371	294	0.792452830188679	112.35

**Question #10:** What are some actionable insights for Warby Parker?

```
140 WITH funnels 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',
p.price
141 FROM quiz AS 'q'
142 LEFT JOIN home_try_on AS 'h'
143 ON q.user_id = h.user_id
144 LEFT JOIN purchase AS 'p'
145 ON p.user_id = q.user_id)
146
147 SELECT number_of_pairs,
SUM(is_home_try_on) AS 'home_try_on',
SUM(is_purchase) AS 'purchase_complete',
1.0 * SUM(is_purchase) /
SUM(is_home_try_on) AS 'purchase rate',
ROUND(AVG(price),2) AS avg_price
148 FROM funnels
149 WHERE number_of_pairs = '3 pairs'
150 OR number_of_pairs = '5 pairs'
151 GROUP BY number_of_pairs;
```

## Extras: Other areas of potential funnel optimization

**Initial Thought:** I was curious to know whether any particular pattern could be found within the quiz responses that might be indicative of a greater propensity to purchase and/or a greater propensity to purchase a higher-priced pair of glasses.

**Query Methodology:** I went back to the quiz table containing all the questions and responses and joined it with the home\_try\_on and purchase tables.

### Interesting Findings:

- People who selected 'I'm not sure. Let's skip it.' on the 'Fit' question not only purchased at a significantly higher rate than those who selected other answers, but also spent \$6 dollars more per purchase
- Those who selected 'No Preference' on the 'Shape' question also purchased at significantly higher rate than those who selected other answers

```
185 WITH funnels AS (SELECT 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',  
    q.fit, p.price  
186 FROM quiz AS 'q'  
187 LEFT JOIN home_try_on AS 'h'  
188     ON q.user_id = h.user_id  
189 LEFT JOIN purchase AS 'p'  
190     ON p.user_id = q.user_id)  
191  
192 SELECT fit, SUM(is_home_try_on) AS  
    num_try_on, SUM(is_purchase) AS  
    num_purchased, 1.0 * SUM(is_purchase) /  
    SUM(is_home_try_on) AS purchase_rate,  
    ROUND(AVG(price),2) AS avg_price  
193 FROM funnels  
194 GROUP BY fit  
195 ORDER BY purchase_rate DESC;
```



## Extras: Other areas of potential optimization

### Interesting Findings:

- People who selected 'I'm not sure. Let's skip it.' on the 'Fit' question not only purchased at a significantly higher rate than those who selected other answers, but also spent \$6 dollars more per purchase

fit	num_try_on	num_purchased	purchase_rate	avg_price
I'm not sure. Let's skip it.	64	45	0.703125	119.67
Wide	150	105	0.7	107.95
Medium	234	152	0.64957264957265	112.93
Narrow	302	193	0.639072847682119	113.52

- Those who selected 'No Preference' on the 'Shape' question also purchased at significantly higher rate than those who selected other answers

shape	num_try_on	num_purchased	purchase_rate	avg_price
No Preference	71	53	0.746478873239437	111.32
Round	140	95	0.678571428571429	116.26
Rectangular	288	189	0.65625	114.26
Square	251	158	0.629482071713147	109.21

Fit table Query (same as previous slide)

```
190 WITH funnels AS (SELECT q.user_id,
191   h.user_id IS NOT NULL AS
192   'is_home_try_on', h.number_of_pairs,
193   p.user_id IS NOT NULL AS 'is_purchase',
194   q.fit, p.price
195 FROM quiz AS 'q'
196 LEFT JOIN home_try_on AS 'h'
197   ON q.user_id = h.user_id
198 LEFT JOIN purchase AS 'p'
199   ON p.user_id = q.user_id)
200 SELECT fit, SUM(is_home_try_on) AS
201   num_try_on, SUM(is_purchase) AS
202   num_purchased, 1.0 * SUM(is_purchase) /
203   SUM(is_home_try_on) AS purchase_rate,
204   ROUND(AVG(price),2) AS avg_price
205 FROM funnels
206 GROUP BY fit
207 ORDER BY purchase_rate DESC;
```

Shape table  
Query

```
197 WITH funnels AS (SELECT q.user_id,
198   h.user_id IS NOT NULL AS
199   'is_home_try_on', h.number_of_pairs,
200   p.user_id IS NOT NULL AS 'is_purchase',
201   q.fit, p.price, q.shape
202 FROM quiz AS 'q'
203 LEFT JOIN home_try_on AS 'h'
204   ON q.user_id = h.user_id
205 LEFT JOIN purchase AS 'p'
206   ON p.user_id = q.user_id)
207 SELECT shape, SUM(is_home_try_on) AS
208   num_try_on, SUM(is_purchase) AS
209   num_purchased, 1.0 * SUM(is_purchase) /
210   SUM(is_home_try_on) AS purchase_rate,
211   ROUND(AVG(price),2) AS avg_price
212 FROM funnels
213 GROUP BY shape
214 ORDER BY purchase_rate DESC;
```

## Extras: Suggested next steps

I acknowledge that the sample size here may be too small for the data to be considered significant, and that there may be some lurking variables causing the difference in conversion rates and dollars spent.

### Actions:

However, I would suggest that Warby Parker dig further in order to isolate any identifiable, repeating attributes among those who select 'No Preference' or '...skip it'. The funnel behavior for these users seems to indicate that Warby Parker's best customers (in terms of likelihood to purchase and dollars spent per purchase) are those who aren't clear on what they need.

If Warby Parker is able to push more users who look and behave like those identified by the most recent two queries, they could see as much as 15% lift in revenue.

# Summary of suggested action and potential impact

## Suggested Course of Action

## Potential Impact

- Push all users into the '5-pair' try-on funnel
- Run another test to see whether similar results can be achieved with 4 pairs – could save \$\$ on circulating inventory
- Decrease quiz length or concatenate two questions into one while maintaining both clarity and brevity
- Strive to isolate any identifiable, repeating attributes among those who select 'No Preference' or '...skip it'.



**Increase revenue by ~ 48%**



**Increase purchase rate**



**Increase revenue by up to  
15%**