

**code**cademy + WARBY PARKER

# Capstone Project: Funnels with Warby Parker

Learn SQL from Scratch

Laura Milanez

2018.08.26

# Table Of Contents

Section 1: Getting familiar with Warby Parker

Section 2: User Preferences

Section 3: The Quiz Funnel

Section 4: The Home Try-On Funnel

Section 5: Conclusions and Actionable Insights

Section 1

# Getting familiar with Warby Parker

# Survey Table

```
3 -- THE SURVEY TABLE
4 -- What columns does the table have?
5 SELECT *
6 FROM survey
7 LIMIT 10;
8 -- 1986 rows
9
10 -- How many possible answers are there per question?
11 -- This assumes all answers were selected at least once
12 SELECT
13 question,
14 COUNT(DISTINCT response) AS 'unique responses'
15 FROM survey
16 GROUP BY 1 ORDER BY 1;
17
18 -- Example responses (getting insight into the questions)
19 -- subbed in 1, 2, 4, and 5
20 SELECT
21 DISTINCT response
22 FROM survey
23 WHERE question LIKE '3%';
24
25 -- Number of users that took the survey
26 SELECT COUNT(DISTINCT user_id)
27 FROM survey;
28
29 -- Average number of answers per user, given 1986 rows on table
30 SELECT 1986.0/500;
```

The survey table stores users' responses to the survey quiz. It contains following columns:

- **Question** – the questions asked on the survey (5 distinct ones).
- **User\_id** – the unique identifier of each respondent (primary key).
- **Response** – the response selected by each user to the given survey question.

The table contains 1986 total responses, which were given by 500 unique users. This indicates that on average, each user answered 3.97 (approximately 4) out of 5 questions of the survey.

The table on the left shows the number of unique responses per question. The table on the right shows an example of the types of answers that were given (question #3 used as an example). The example responses coupled with low number of unique responses indicates that the survey uses multiple choice rather than open text format, limiting the user's choices, but allowing for easier analysis.

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

response
Round
Square
Rectangular
No Preference

# Quiz Table

The quiz table stores users' selected preferences, helping Warby Parker better understand that types of glasses that might be appealing to each user. It contains following columns:

- **User\_id** – the unique identifier of each respondent (primary key).
- **Style** – style of glasses, to be selected between Women's, Men's or neither.
- **Fit** – description of facial features to better understand the fit of the glasses
- **Shape** – the preferred shape of the frames
- **Color** – the preferred color of the frames

The table contains 1000 unique respondents. Of those, 99 skipped style selection, 432 (~43%) selected Men's Styles, and 469 (~47%) selected Women's Styles. Disregarding the 99 respondents that didn't state a preference, the data above suggests a slightly higher rate of female respondents.

```
1  -- THE HOME TRY-ON FUNNEL TABLES
2  -- What columns does the table have?
3  SELECT *
4  FROM quiz
5  LIMIT 5;
6  -- 1000 rows
7
8  -- Looking at the split between style preferences to infer gender distribution of
   respondents
9  SELECT
10 style,
11 COUNT (DISTINCT user_id) as 'number of unique users'
12 FROM quiz
13 GROUP BY 1 ORDER BY 1;
```

# Home Try-On Table

```
16 -- THE HOME TRY-ON FUNNEL TABLES
17 -- What columns does the table have?
18 SELECT *
19 FROM home_try_on
20 LIMIT 5;
21 -- 750 rows
22
23 --Are entries on the table unique users?
24 SELECT COUNT(DISTINCT user_id)
25 FROM home_try_on;
26
27 --Confirming the split between variants
28 SELECT
29 number_of_pairs,
30 COUNT(*)
31 FROM home_try_on
32 GROUP BY 1;
```

The **home\_try\_on** table stores user information around number of pairs tried on and location of delivery. Specifically, it contains the following columns:

- **User\_id** – the unique identifier of each respondent (primary key).
- **Number\_of\_pairs** – identifies whether the user is part of variant A (3 pairs) or variant B (5 pairs)
- **Address** – the delivery address of the user

The table contains a total of 750 entries, representing 750 unique users. Of those users, 379 (50.53%) were given 3 pairs to try at home, and 371 were given 5 pairs.

Example of the table format and entries below:

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-accb-49a7bb46c586	3 pairs	347 Madison Square N
3bc8f97f-2336-4dab-bd86-e391609dab97	5 pairs	182 Cornelia St

# Purchase Table

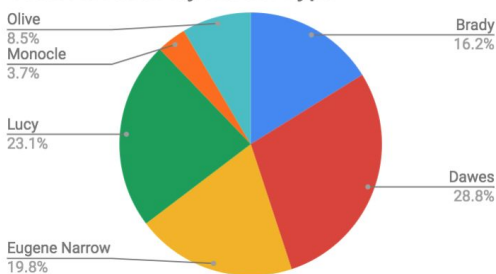
```
33 -- THE PURCHASE TABLE
34 -- What columns does the table have?
35 SELECT *
36 FROM purchase
37 LIMIT 5;
38 -- 495 rows
39
40 -- Number of products available
41 SELECT COUNT(DISTINCT product_id)
42 FROM purchase;
43
44 -- Did unique users purchase more than once?
45 SELECT COUNT(DISTINCT user_id)
46 FROM purchase;
47
48 -- Total Revenue
49 SELECT
50 1.0*SUM(price) AS 'total revenue'
51 FROM purchase;
52
53 -- Revenue Split by Gender
54 SELECT
55 style,
56 COUNT(*) AS 'total purchases',
57 1.0*SUM(price) AS 'total revenue'
58 FROM purchase
59 GROUP BY 1 ORDER BY 1;
60
61 -- Split by Model Name:
62 SELECT
63 model_name,
64 COUNT(*) AS 'total purchases',
65 1.0*SUM(price) AS 'total revenue',
66 price
67 FROM purchase
68 GROUP BY 1 ORDER BY 1;
```

The purchase table stores information on the purchase made by each user. It contains the following columns:

- **User\_id** – the unique identifier of each respondent (primary key).
- **Product\_id** – the unique identifier of each product (10 total possibilities present in the table)
- **Style** – style of glasses, to be selected between Women's or Men's.
- **Model\_name** – the name of the model purchased
- **Color** – the color of the model purchased
- **Price** – the price of the model purchased

The table contains a total of 495 purchases by unique users, accounting for a total revenue of \$55,795.00. When we look at revenue split by model, we see that:

Total Revenue by Model Type



As we can see on the chart, slightly higher revenue is derived from Women's styles. This seems to be in part due to the higher percentage of users purchasing that style, and in part, due to slightly higher average purchase amount (W: \$113.77 M: \$111.63).

Dawes and Lucy derive the most revenue, followed closely by Eugene Narrow and Brady. Of those, Dawes and Lucy are, as expected, the most expensive models, with a price of \$150, as compared to \$95 for the other two.



More insights on the purchase table are discussed in Section 2 (User Preferences).

Section 2

# User Preferences

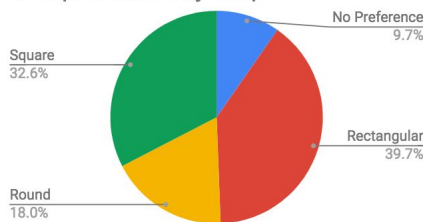


# User Preferences

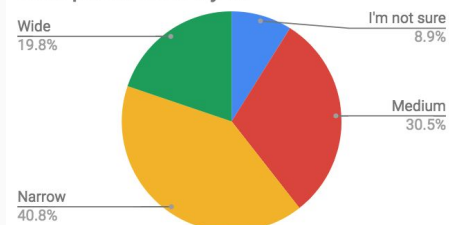
```
5 -- USER PREFERENCES
6 -- by color
7 SELECT
8 color,
9 COUNT(*) AS 'total selection'
10 FROM quiz
11 GROUP BY 1 ORDER BY 1;
12 -- by style
13 SELECT
14 style,
15 COUNT(*) AS 'total selection'
16 FROM quiz
17 GROUP BY 1 ORDER BY 1;
18 -- by fit
19 SELECT
20 fit,
21 COUNT(*) AS 'total selection'
22 FROM quiz
23 GROUP BY 1 ORDER BY 1;
24 -- by shape
25 SELECT
26 shape,
27 COUNT(*) AS 'total selection'
28 FROM quiz
29 GROUP BY 1 ORDER BY 1;
```

This section of the report is meant to be informational only. The data present here will not be used to inform the final recommendations, as the results below should already be informing which glasses are presented to each user for selection, and therefore, should already be implicitly included in the funnel.

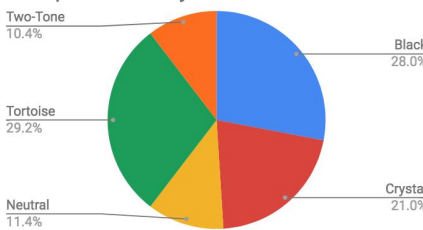
user preference by shape



user preference by fit



user preference by color



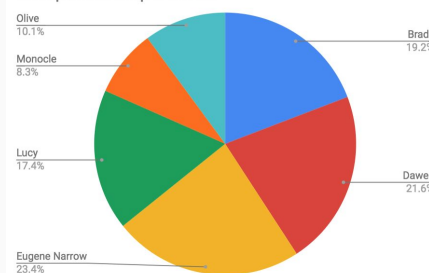
The three charts above show that:

- There's a strong preference for square and rectangular glasses (72.3%).
- Black and Tortoise account for 57.2% of color selection. However, Crystal is a close third with 21% of selection.
- Lastly, most respondents have medium or narrow faces.

Looking lower down on the funnel, we see that of the glasses purchased:

- Six different models were represented.
- Eugene Narrow and Dawes are the most popular models, followed closely by Brady and Lucy.
- The top two models represent 45% of all purchases made.

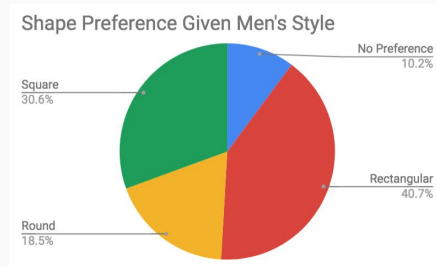
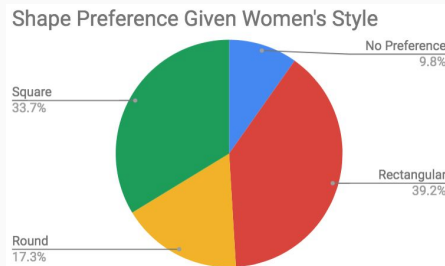
total purchases per model



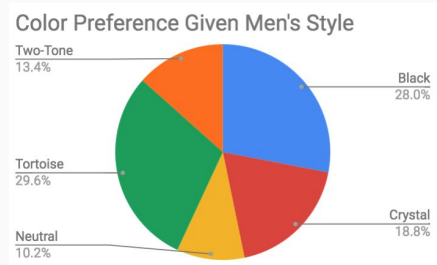
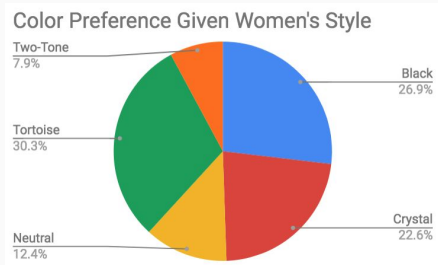
# User Preferences by Gender

```
31 -- BY GENDER THEN BY:  
32 -- by shape  
33 SELECT  
34 fit,  
35 COUNT(*) AS 'total selection'  
36 FROM quiz  
37 WHERE style LIKE 'Women%'  
38 GROUP BY 1 ORDER BY 1;  
39  
40 SELECT  
41 fit,  
42 COUNT(*) AS 'total selection'  
43 FROM quiz  
44 WHERE style LIKE 'Men%'  
45 GROUP BY 1 ORDER BY 1;  
46  
47 -- by color  
48 SELECT  
49 color,  
50 COUNT(*) AS 'total selection'  
51 FROM quiz  
52 WHERE style LIKE 'Women%'  
53 GROUP BY 1 ORDER BY 1;  
54  
55 SELECT  
56 color,  
57 COUNT(*) AS 'total selection'  
58 FROM quiz  
59 WHERE style LIKE 'Men%'  
60 GROUP BY 1 ORDER BY 1;
```

Below we'll take a look at whether preferences vary by gender\*:



Shape preferences are comparable when it comes to men and women (Rectangular > Square > Round). The notable difference is that women seem to prefer Square frames slightly more often than men.



When looking at color selection given style preference, the most notable difference is the preference for two-tone glasses in Men's Style (almost double that of women's). Women also seem to prefer Crystal and Neutral glasses more often than Men. However, the differences in Black and Tortoise are almost negligible.

\*Disclaimer: gender is inferred from the answer to the question about style preference – Women's vs. Men's style. Given that style preference is not binary, this is not meant to be a true representation of the gender split, but rather an approximation based on the information available to us on this dataset.

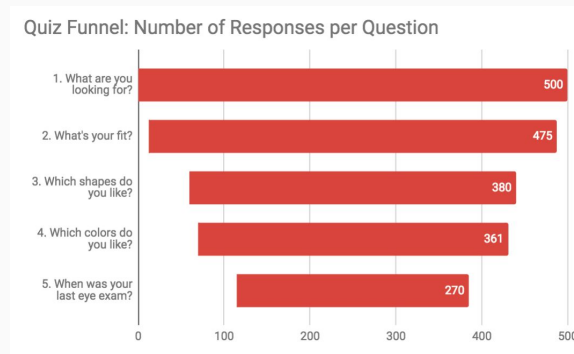
Section 3

# The Quiz Funnel

# The Quiz Funnel

```
1 -- Creating a quiz funnel
2 -- What is the number of responses for each question?
3 SELECT question,
4     COUNT(DISTINCT user_id) AS 'number of responses'
5 FROM survey
6 GROUP BY 1 ORDER BY 1;
```

Leveraging the survey table, we create a quiz funnel that gives us insight into the completion rate per question:



The table below shows the number of responses and completion rate per question. The completion rate is calculated relative to the previous survey question, meaning we are addressing the following question: *of the people who answered the previous question, how many also answered the subsequent question?*

question	number of responses	percent completion of question
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	70%

# The Quiz Funnel (cont'd)

```
1 -- Creating a quiz funnel
2 -- What is the number of responses for each question?
3 SELECT question,
4     COUNT(DISTINCT user_id) AS 'number of responses'
5 FROM survey
6 GROUP BY 1 ORDER BY 1;
```

question	number of responses	percent completion of question
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	70%

We can see on the table above that Questions 2 and 4 have high completion rates, but Questions 3 and 5 have lower rates. There are various possible explanations for the lower rates seen in Questions 3 and 5, a few of which are explored below:

The low completion rate in Question 3 might be attributed to the fact that people are unsure of their glasses shape preference without trying them on. It's possible that they are first time buyers or unfamiliar with the shapes available.

The low completion rate in Question 5 suggests that last doctor visit is a more sensitive question and people might be reluctant to address it. Alternatively, it's possible that responders are unsure when their last visit was, resulting in lower rate of completion.

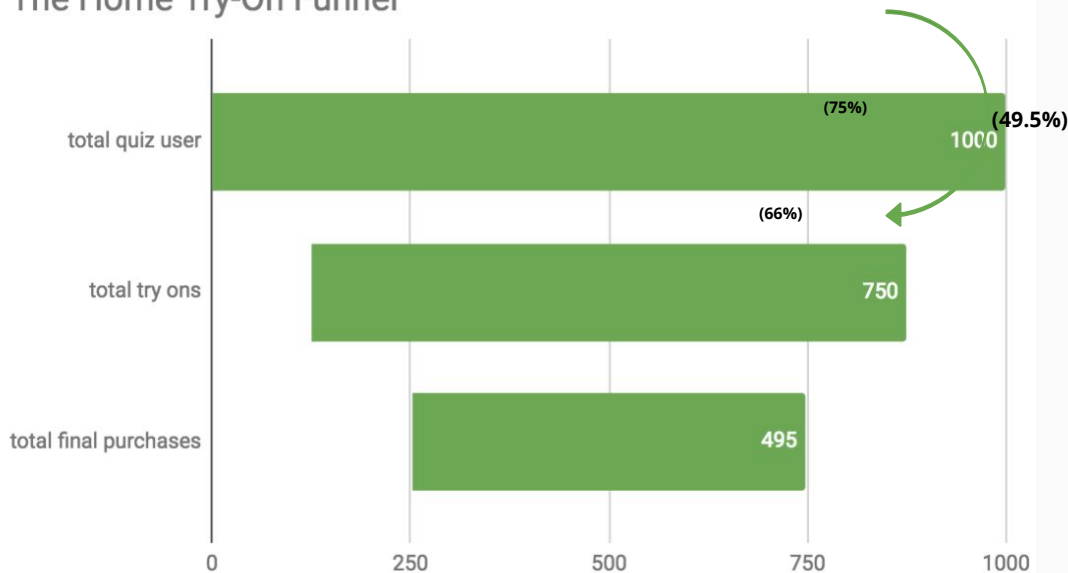
Section 4

# The Home Try-On Funnel

# Home Try-On Funnel

```
1 -- New Funnel table
2 -- Creating temporary table combining all three tables from the Home Try-On Funnel
3 WITH funnel AS
4 (
5     SELECT DISTINCT q.user_id,
6         ht.user_id IS NOT NULL AS 'is_home_try_on',
7         ht.number_of_pairs,
8         p.user_id IS NOT NULL AS 'is_purchase'
9     FROM quiz AS 'q'
10 -- Left joining the two additional tables on user_id
11     LEFT JOIN home_try_on AS 'ht'
12         ON ht.user_id = q.user_id
13     LEFT JOIN purchase AS 'p'
14         ON p.user_id = ht.user_id
15 )
16 -- Calculating funnel metrics from aggregate table above
17 SELECT
18 -- Total quiz takers
19     COUNT(*) AS 'total quiz user',
20 -- Total home try-ons
21     SUM(is_home_try_on) AS 'total try ons',
22 -- Total final purchases
23     SUM(is_purchase) AS 'total final purchases',
24 -- Conversion of quiz to home try-on
25     1.0 * SUM(is_home_try_on) / COUNT(user_id) AS '% quiz to home try ons',
26 -- Conversion of home try-on to final purchase
27     1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS '% home try ons to final purchase'
28 FROM funnel;
```

The Home Try-On Funnel



Above we can see the home try-on funnel. Of the users who take the quiz, 75% go on to have a home try-on, however, only 66% of those go on to make an overall purchase. This results in an overall conversion rate of 49.5%.

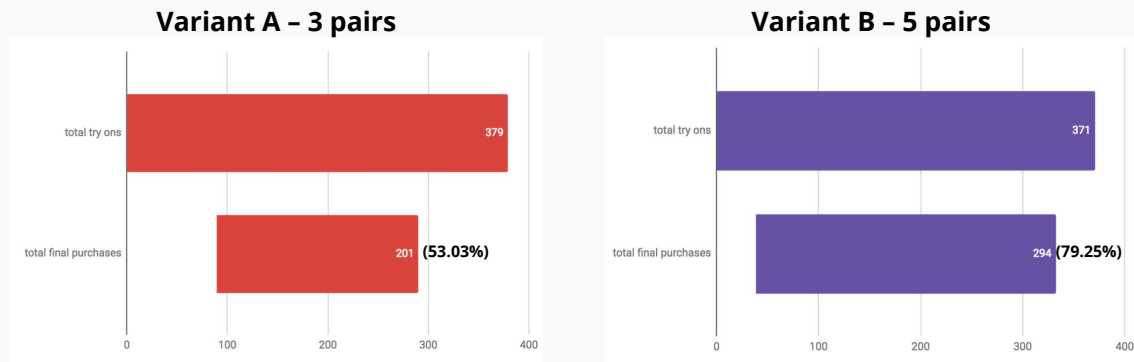
Additional insights on the funnel and final recommendations are discussed in Section 5.

# A/B Testing Home Try-On Variants

```
1 -- New Funnel table
2 -- Creating temporary table combining all three tables from the Home Try-On Funnel
3 WITH funnel AS
4 (
5     SELECT DISTINCT q.user_id,
6         ht.user_id IS NOT NULL AS 'is_home_try-on',
7         ht.number_of_pairs,
8         p.user_id IS NOT NULL AS 'is_purchase'
9     FROM quiz AS 'q'
10 -- Left joining the two additional tables on user_id
11     LEFT JOIN home_try-on AS 'ht'
12         ON ht.user_id = q.user_id
13     LEFT JOIN purchase AS 'p'
14         ON p.user_id = ht.user_id
15 )
16 -- Calculating A/B testing metrics from aggregate table above
17 SELECT
18     -- Number of Pairs
19     number_of_pairs,
20     -- Total quiz takers
21     COUNT(*) AS 'total quiz user',
22     -- Total home try-ons
23     SUM(is_home_try-on) AS 'total try ons',
24     -- Total final purchases
25     SUM(is_purchase) AS 'total final purchases',
26     -- Conversion of home try-on to final purchase
27     1.0 * SUM(is_purchase) / SUM(is_home_try-on) AS '% home try ons to final purchase'
28 FROM funnel
29 GROUP BY 1
30 ORDER BY 1;
```

As we saw when analyzing the home\_try-on table, 75% of the users who took a quiz went on to have a home try-on. For the next piece of our analysis, we'll be looking specifically at this 75%. We'll attempt to address the question: *of the people who had a home try-on, who ended up purchasing a pair?*

Of the 750 total home try-ons, 379 (50.53%) were given 3 pairs to try at home, and 371(49.47%) were given 5 pairs. Below we see funnels for the two variants for comparison:



Given that the sample sizes were comparable, and assuming the users were selected at random, we can see in brackets on the funnels above that Variant B (5-pairs to try at home) had significantly higher try-on to purchase conversion rate than Variant A (3-pairs to try at home).

This indicates that by increasing the number of options presented to a user, we increase the probability that they will purchase a pair of glasses.

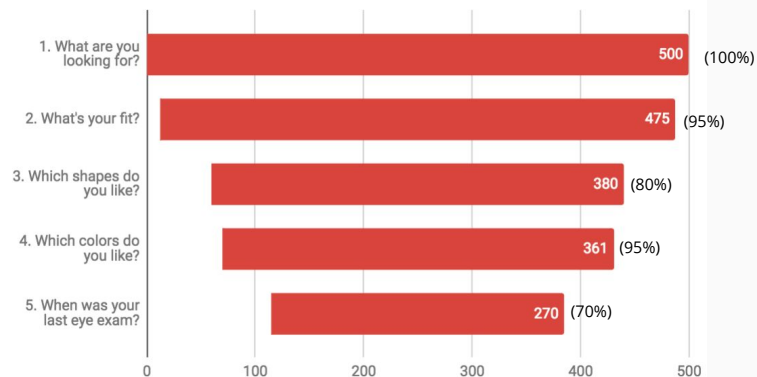


Section 5

# Conclusions and Actionable Insights

# Suggestions for the Quiz Funnel

Quiz Funnel: Number of Responses per Question



In order to increase overall purchase rates, we should start by looking at the top of the funnel (completion of the quiz). The importance of increasing quiz completion should not be understated, as driving higher conversion at the lower levels of the funnel can certainly increase conversion at the upper end.

In psychology this is known as foot-in-the-door (FITD) technique, and it aims at getting a person to agree to a larger request (in this case purchasing a pair of glasses) by having them agree to one or more modest requests at first (in this case trying on glasses at home and ultimately purchasing a pair).

Drawing insights from our quiz funnel analysis, in order to increase completion of the survey, Warby Parker should focus on reducing drop off rates\* for Questions #3 and #5, which are the highest at 20% and 30%, respectively.

\*Disclaimer: we don't have enough information to know whether we're looking at the drop off or exit rate. However, due to the larger percentages seen (20/30%), and since this has no direct impact in the conclusions drawn above, we'll refer to this number as the drop-off rate.

# Suggestions for the Quiz Funnel (cont'd)

In Section 4, we learnt that presenting users with more glasses options drove higher conversion rates. With the same reasoning, **Warby Parker should consider including additional glass shape options to Question #3.** In the case that additional shapes are not available for purchase, **user preference research can be conducted through an open ended question such as, "Other, please specify: \_\_\_\_\_."**

As for Question #5, in Section 4 we noted that the higher sensitivity of questions related to doctor appointments is a likely culprit for the lower conversion rates. More research is needed to better understand the actual cause of the lower conversion rate, however, if our hypothesis above is correct, one possible option is saving that question for later.

Given that Question #5 has the highest drop-off rate and assuming the user must submit an eye exam before purchasing a pair of glasses, **Warby Parker should omit Question #5 from the initial Quiz all together.**

# Suggestions for the Home Try-On Funnel

When we look at the home try on funnel, we see three important conversion rates:

- Quiz to try-on (at 75% conversion)
- Try-on to purchase (at 66% conversion)
- Overall conversion rate: quiz to purchase (at 49.5%)

As mentioned at the beginning of Section 5, in order to improve the overall conversion rate, we should focus on tackling issues within each step of the funnel and improving individual steps' conversion rates. This will result in higher overall conversion.

Given that we just discussed what Warby Parker should do to increase quiz completion (bottom of the funnel), we now look at the next step of the funnel: try-on to purchase conversion rate.

When A/B testing the home try-on variants (Section 4), we saw that variant B (5 pairs), had significantly higher try-on to purchase conversion rates than variant A (3 pairs). Specifically, variant B saw a conversion rate of 79.25%, while variant A saw a conversion rate of only 53.03%. This suggests that, in order to increase overall try-on to purchase conversion rate, **Warby Parker should send 5 rather than 3 pairs for their users to try-on.**

Additionally, Warby Parker should go further and **A/B test 5 pairs against 4, 6, or 7 to find the optimal number of pairs to increase conversion at the top of the funnel.**

# Summary of Recommendations

Below is a summary of the actionable insights included in the conclusion of this report:

1. Include additional glass shape options to Question #3.

OR

2. Conduct user glass shape preference research through an open ended question on the survey.
3. Omit Question #5 from the initial Quiz all together.
4. Send 5 rather than 3 pairs for their users to try-on.
5. A/B test 5 pairs against 4, 6, or 7 to find the number of pairs that maximizes home try-on to purchase conversion rate.