codecademy

Usage Funnels with Warby Parker

Learn SQL from Scratch

Sarah Schofield

Cohort 07/03/2018

codecademy

1. **Quiz Funnel**
2. Purchase Funnel
3. More analysis with Warby Parker data

1.1 *survey* table analysis

Warby Parker runs a 5 question style quiz to help customers find the perfect frame. The data generated from the style quiz are stored in a table `survey`. From the data stored in this table, you can determine the completion rate for each question, and at which questions customers tend to drop out.

Before starting funnel analysis to determine the completion rate for each question, we should have a look at the `survey` table structure and contents.

The query below produces the output on the right. It selects all columns (`question`, `user_id` and `response`) for the first 10 records from the `survey` table.

```
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

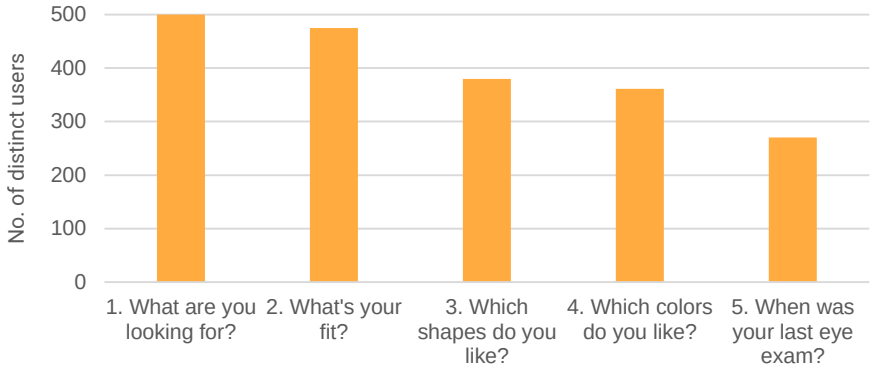
1.2 Creating the Quiz Funnel

To determine the number of responses per question, you can use the aggregate function `COUNT` with the `user_id` column as the parameter. To ensure the same customer has not answered multiple times, you can use `DISTINCT` in the command.

Finally, you can the group the count for each value in `question` by using the `GROUP BY` statement, thereby giving the count for each question in the style quiz. In this case, I have used the column reference number 1 instead of referring to `question.`

```
SELECT question AS 'Question',
    COUNT(distinct user_id) AS 'No. of distinct users'
FROM survey
GROUP BY 1;
```

In this example, I have optionally renamed the table columns for output readability by using the command `AS`.



Question	No. of distinct users
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

1.3 Analysing the Quiz Funnel

As you can see from the graph and table on the previous slide, the response rate decreases for each question in the quiz.
The percentage completion rate for each questions is lowest for question 3 and question 5 (80% and 75%), indicating customers are dropping out of the quiz at these points.

So why are customers dropping out the quiz at these points?

Question	No. of distinct users	% Completion rate
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%

Possible reasons for quiz dropout

3. Which shapes do you like?
- Customers aren't sure what glasses shapes suit their face shape
- People generally don't have favourite shapes in the same way people have favourite colors

Suggestion
Add a skip option for this question.

5. When was your last eye exam?
- Customer can't remember when they last had an eye exam
- Too personal a question
- Question acts as a prompt to get an eye test, therefore they have dropped out before they start ordering glasses to try out

Suggestion
Remove the question from the quiz. There doesn't seem to be value in asking this upfront as this is unlikely to influence the customer's recommendations, and the customer will be asked for their prescription if they purchase the glasses.

For each suggestion, Warby Parker could trial these changes using A/B testing to see whether there is a greater completion rate for the quiz.

codecademy

1. Quiz Funnel
2. **Purchase Funnel**
3. More analysis with Warby Parker data

2.1 Table analysis for Purchase Funnel

The question Warby Parker wants to answer using Purchase Funnel analysis is whether customers who are sent 5 pairs of glasses to try on are more likely to make a purchase compared to customers sent 3 pairs of glasses.
The data required to create our Purchase funnel is stored in 3 tables:

- `quiz`
- `home_try_on`
- `purchase`

```
SELECT *
FROM quiz
LIMIT 5;

SELECT *
FROM home_try_on
LIMIT 5;

SELECT *
FROM purchase
LIMIT 5;
```

Using the query on the left, 5 example rows from each table are shown on the right.
`quiz` has 5 columns, `home_try_on` has 3 columns, and `purchase` has 6 columns.

quiz

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

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

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

2.2 Table analysis for Purchase Funnel

In order to create the a Purchase Funnel for analysis, you can combine the 3 tables in to a single table using the `LEFT JOIN` function. The tables can be joined using the `user_id` column present in each table.
In order to create the funnel, you start the `LEFT JOIN` first from the `quiz` table, then to the `home_try_on` table, and then the `purchase` table.

In the query below, I have renamed columns for ease of understanding using the `AS` command.
To simplify writing the funnel, I have used alias's for the table names:

- `quiz` = `q`
- `home_try_on` = `h`
- `purchase` = `p`

Where the user is present in `home_try_on`, then the value in `is_home_try_on` will be 1, while not present will be 0.
Where the user is present in `purchase`, then the value in `is_purchase` will be 1, while not present will be 0.

```
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;
```

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	Ø	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	Ø	0
0143cb8b-bb81-4916-9750-ce956c9f9bd9	0	Ø	0
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	1	5 pairs	0
b1dded76-cd60-4222-82cb-f6d464104298	1	3 pairs	0

2.3 Conversion analysis (part 1)

Now that we have a Purchase Funnel which combines all 3 tables worth of data by `user_id`, the funnel can be analysed to show conversion rates between steps in the customer journey.
Using the `WITH` command, I can create a temporary table called `funnel` on which I can complete my analysis.
Because `is_home_try_on` and `is_purchase` are columns with 1 and 0 to represent if customer is present or not present in their respective tables, I can use the aggregate function `SUM` on these columns to indicate how many customers participated in the home try on stage, or purchased a pair of glasses.

Total quiz conversion
Using the query on the right, the total conversion rate between customer's who complete the quiz and those that proceed on to purchase glasses can be determined.
In this case, the total conversion rate is **49.5%.**
The output of this query produces the data below.

num_quiz	num_purchase	quiz to purchase
1000	495	0.495

```
WITH 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 'num_quiz',
    SUM(is_purchase) AS 'num_purchase',
    ROUND(1.0 * SUM(is_purchase) / count(user_id), 3) AS 'quiz
to purchase'
FROM funnel;
```

2.3 Conversion analysis (part 2)

Conversion by step (quiz --> try on --> purchase)
Using the query on the right, the output is the conversion rate for each step is the customer journey.
The conversion rate between quiz to home try on is **75%.**
The conversion rate between home try on to purchase is **66%** .

num_quiz	num_try_on	num_purchase	quiz to home_try_on	home_try_on to purchase
1000	750	495	0.75	0.66

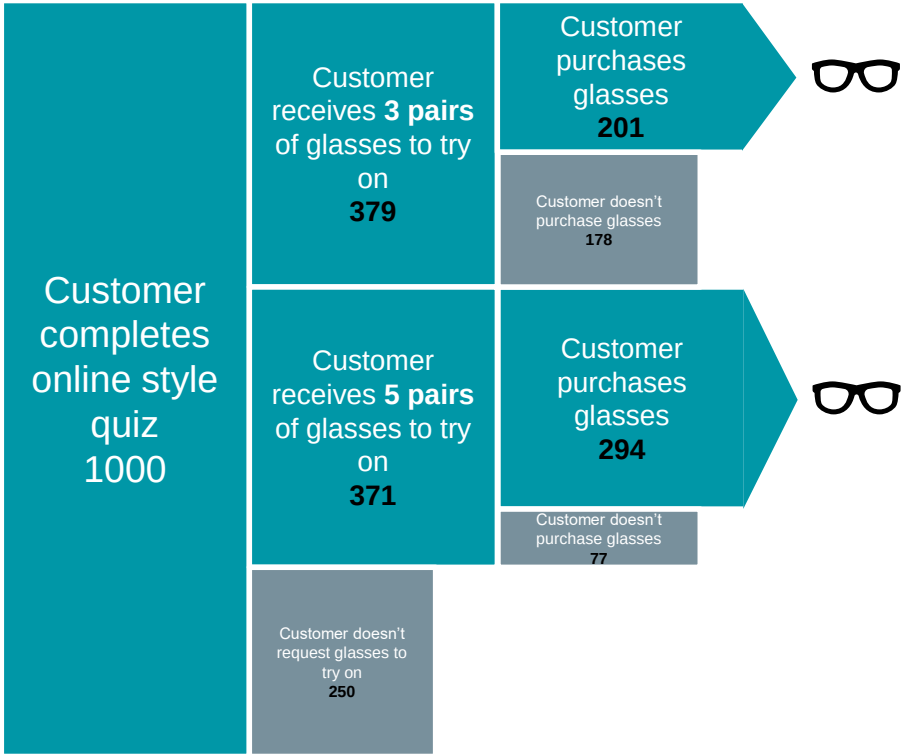
```
WITH funnel AS (
    SELECT DISTINCT q.user_id,
        h.user_id IS NOT NULL AS 'is_home_try_on',
        h.number_of_pairs,
        p.user_id IS NOT NULL AS 'is_purchase'
    FROM quiz q
    LEFT JOIN home_try_on h
        ON q.user_id = h.user_id
    LEFT JOIN purchase p
        ON q.user_id = p.user_id
    )
SELECT COUNT(*) AS 'num_quiz',
    SUM(is_home_try_on) AS 'num_try_on',
    SUM(is_purchase) AS 'num_purchase',
    ROUND(1.0 * SUM(is_home_try_on) / count(user_id),3) AS 'quiz to home_try_on',
    ROUND(1.0 * SUM(is_purchase) / sum(is_home_try_on),3) AS 'home_try_on to purchase'
FROM funnel;
```

Conversion by number of glasses (A/B test)
Using the query on the right, the output is the conversion rate between home try on and purchase for either 3 pairs or 5 pairs of glasses to try on.
The conversion rate is higher when the customer has 5 pairs of glasses to try compared to 3 pairs (**79.2%** vs **53%**).

number_of_pairs	num_try_on	num_purchase	home_try_on to purchase
3 pairs	379	201	0.53
5 pairs	371	294	0.792

```
WITH 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_home_try_on) AS 'num_try_on',
    SUM(is_purchase) AS 'num_purchase',
    ROUND(1.0 * SUM(is_purchase) / sum(is_home_try_on),3) AS 'home_try_on to purchase'
FROM funnel
WHERE number_of_pairs IS NOT NULL
GROUP BY number_of_pairs;
```

2.4 Conclusion from Purchase Funnel analysis



Of the **1000** customers who completed the style quiz, **75%** of customers continue to the home try on option.
As a test, 50% of customers were assigned either 3 pairs to try on or 5 pairs to try (**379** vs **371** respectively).
From the A/B test, 5 pairs of glasses for home try in resulted in a higher conversion rate – an additional **26.2%** conversion rate compared to customers who received 3 pairs of glasses (**79.2% vs 53%**). This suggests Warby Parker should adopt sending 5 pairs of glasses to all customers who reach the home try stage.

Next steps

- Analysis to determine why 25 % of customers who complete the survey do not proceed to the home try on stage (e.g. do those customer's come in-store?)
- Cost-benefit analysis of sending additional pair of glasses for home try on. Although the net result is increased sales with 5 pairs of glasses to try on, what are the costs associated with this? E.g. Returns costs, failure to return samples, damaged samples

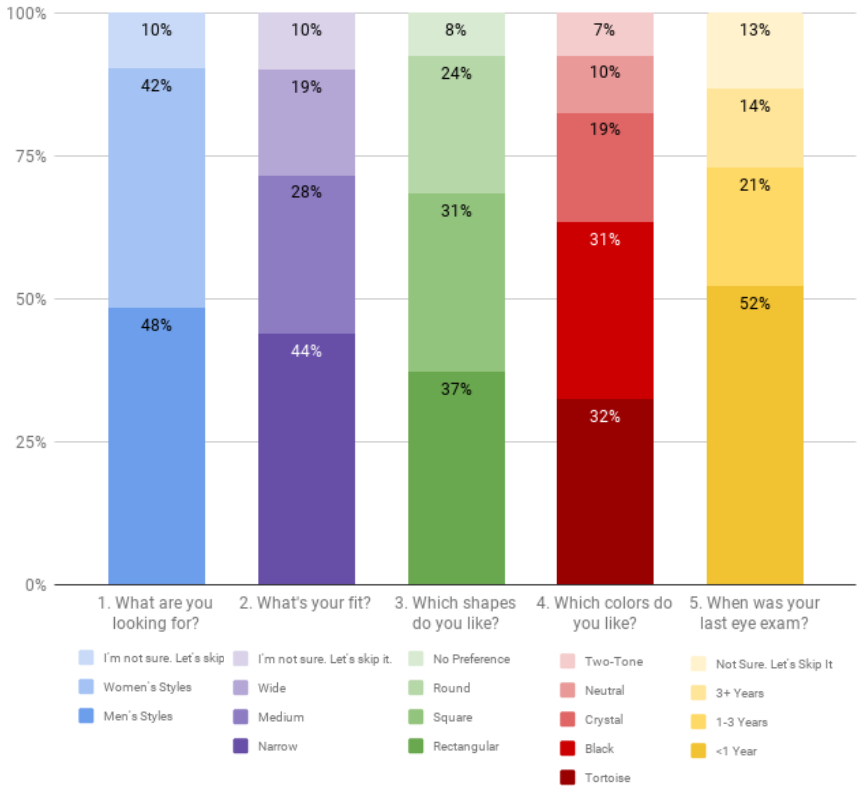
codecademy

1. Quiz Funnel
2. Purchase Funnel
3. **More analysis with Warby Parker data**

3.1 *survey* table analysis – response distribution

Firstly, I checked how many different responses I was expecting for each question in the `survey` table. The results were 3, 4, 5, 5 and 4 different options for question 1 to 5, as shown to the right.
I then ran a query to determine the number of users who chose each response for each question, the results of which have been plotted in the graph.

Question	Observation	So what?
1. What are you looking for?	Slightly more men than women looking for glasses	Suggest more men use the site than women. This could mean men are happier than women to browse glasses online, or women are less likely to require glasses (e.g. contact lenses). Could suggest an opportunity for marketing towards men e.g. men's lifestyle magazines, websites etc. Or some percentage of women choose to look at both styles and select the skip option.
2. What's your fit?	Customers prefer a narrow fit	Warby Parker (WP) should focus their efforts on designing glasses for narrow fit, as they appear to be the majority of customers
3. Which shapes do you like?	Round glasses are the least popular shape	WP should not prioritise their efforts on designing round frames, and focus on rectangular and square frames
4. Which colors do you like?	Color preference is divided mostly between tortoise and black frames	Suggests customers prefer standard glasses colors, and WP should focus their design efforts with traditional colors. **In general**, answers to 2 – 4 suggest customers prefer more traditional glasses styles, rather than bold statement items.
5. When was your last eye exam?	Customers most likely to have had an eye test recently	Suggests majority of customers are getting their eyes tested and immediately choosing WP for actual designs / price, compared to the traditional experience of being tested and purchasing in the same place. Seems reasonable that WP do not invest in opticians services if customers are happy to separate testing and purchasing glasses.

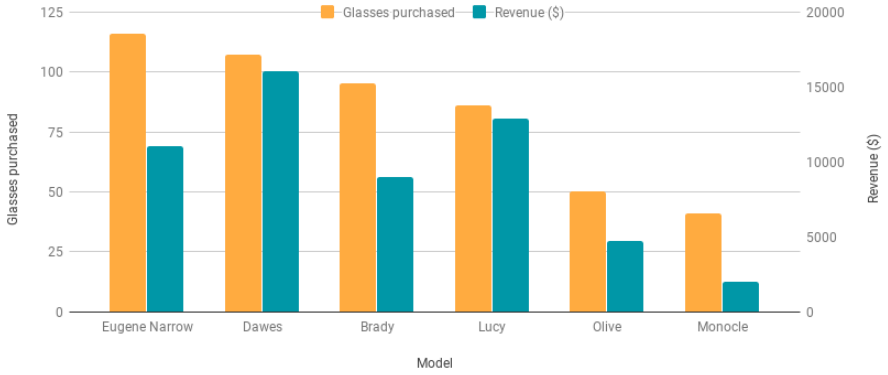


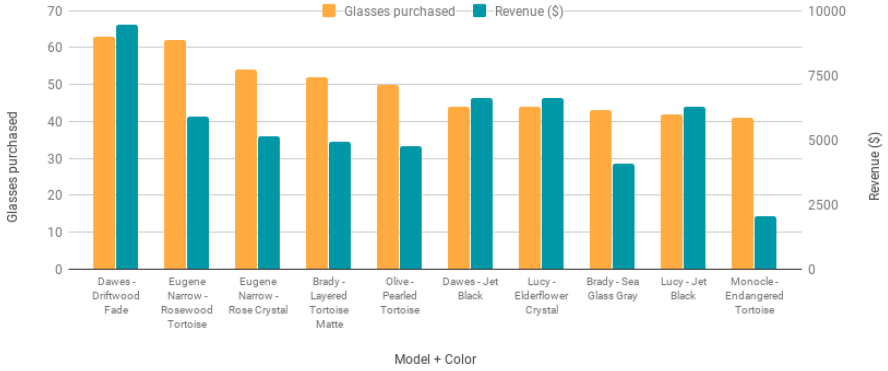
3.2 *purchase* table analysis

Based on previous analysis of the `purchase` table, you can see that there is price variation based on glasses model. I analysed the difference in sales by units sold and total revenue to determine which model was actually generated most revenue for Warby Parker. The left y-axis shows the number of glasses purchased and the right y-axis shows the revenue generated in dollars.
Total numbers analysis shows **495** pairs of glasses were purchased, generating a total of **$55,795**.

As shown in the first graph, the most popular model was Eugene Narrow (116), while Monocle was the least popular (41). However, the model that generated the most revenue was Dawes ($16,050). Eugene Narrow was placed 3rd out 6 ($11,020), while Monocle still generated the least revenue ($2,050).
In the second graph, I analysed how color influenced purchasing choice. Although Eugene Narrow is still a popular choice in two different colors, Dawes in Driftwood Fade was the most popular, but was only placed 6th in Jet Back. Based on revenue, Dawes in Driftwood Fade is still number 1, however Dawes in Jet Black and Lucy in Elderflower Crystal both generated $6600, so placed joint 2nd in terms of revenue.

Based on this analysis, Warby Parker could set up a randomised test to assess whether the prices should be changed for both cheap but popular options and pricier but less popular options. An assessment of profit in addition to revenue would also be required to fully assess the optimal price option. Lastly, as Dawes appears to generate most revenue, Warby Parker could investigate more color options (e.g. Rosewood Tortoise) that could increase popularity of the model further.





3.3 Do number of try-on pairs alter purchase choice?

Based on the A/B test analysis for home try-on, we would expect an increase in number of sales when customers are sent 5 pairs to try on. But is there any difference in the pair for try-on and purchased once normalised for this fact? For this analysis, I used a `LEFT JOIN` to combine the `purchase` table and the `home_try_on` table. I rearranged the data to produce the table on the right.

The average pairs of glasses sold for the 5 try-on group was 29.4 and 20.1 for the 3 try-on group. By normalising the 5 pair group using the two averages, you can see the new values accounting for the fact more pairs of glasses are sold in general if 5 pairs of glasses are sent for home try-on.

Example: Brady – Layered Tortoise Matte
28 / (29.4/20.1) = 19

From the normalised results, you can see which models were more likely to sell if in a 5 pair or 3 pair try-on bundle.
If the customer is part of the 5 pair try-on test, Olive in Pearled Tortoise was more likely to chosen and be purchased, while if the customer is part of the 3 pair try-on test, Eugene Narrow in Rose Crystal was more likely to be chosen and be purchased.

If Warby Parker were to decide not to proceed with sending all customers 5 pairs of glasses, this analysis could be used to 'nudge' customers towards glasses which are more likely to sell based on how many pairs are sent for try-on, such as reordering the search results or branding as a 'top pick'.

```
WITH temp AS (
    SELECT *
    FROM purchase p
    LEFT JOIN home_try_on h
        ON p.user_id = h.user_id
    )
SELECT product_id,
    style,
    model_name,
    color,
    number_of_pairs,
    COUNT(*),
    price
FROM temp
GROUP BY 3,4,5
ORDER BY 1, 6 desc;
```

	Actuals	
Model + Color	5 pairs	3 pairs
Brady - Layered Tortoise Matte	28	24
Brady - Sea Glass Gray	28	15
Dawes - Driftwood Fade	33	30
Dawes - Jet Black	25	19
Monocle - Endangered Tortoise	26	15
Olive - Pearled Tortoise	36	14
Lucy - Elderflower Crystal	27	17
Lucy - Jet Black	29	13
Eugene Narrow - Rose Crystal	28	26
Eugene Narrow - Rosewood Tortoise	34	28

	Normalised		
Model + Color	5 pairs	3 pairs	Ratio
Brady - Layered Tortoise Matte	19	24	80%
Brady - Sea Glass Gray	19	15	128%
Dawes - Driftwood Fade	23	30	75%
Dawes - Jet Black	17	19	90%
Monocle - Endangered Tortoise	18	15	119%
Olive - Pearled Tortoise	25	14	176%
Lucy - Elderflower Crystal	18	17	109%
Lucy - Jet Black	20	13	153%
Eugene Narrow - Rose Crystal	19	26	74%
Eugene Narrow - Rosewood Tortoise	23	28	83%