



WARBY PARKER  
eyewear

# Capstone Project: Funnels

Learn SQL from Scratch

Samantha Pita

January 3, 2019

## TABLE OF CONTENTS

*Part I: Warby Parker's Survey Funnel*

*Part II: The People Have Spoken...*

*Analyzing Response Data*

*Part III: A/B Testing with Home Try-On Funnel*

# Part I: Warby Parker's Survey Funnel



## 1.1 What is a Marketing Funnel?

In the context of marketing analysis, a **funnel** is a theory or diagram of how a customer travels towards the purchase of a product or service. Warby Parker, an innovative designer eyewear brand, uses a Style Quiz to steer new customers towards a purchase. The customers answer 5 questions, and all of their responses get stored into a database table called *survey*.

- To get an idea of what columns exist in the *survey* table, I wrote the following query:

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

- The results are on the following slide.

## 1.2 Query 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

By looking at 10 sample quiz responses, you can see that there are 3 columns:

- **question:** This column displays which of the 5 questions the row of data is referring to
- **user\_id:** This column shows us which user answered by displaying the user's *unique* identifying number
- **response:** This column displays the user's response

Warby Parker can learn a lot about what type of glasses people are looking for by looking at these results as a whole.

# Part II: The People Have Spoken... Analyzing Response Data



## 2.1 Quiz Response Rates

One of the very first things Warby Parker probably wants to know is, “How many responses does each question have?”.

- **In relation to SQL, this means we want to find out the total number of responses by unique users** (we use the *COUNT* and *DISTINCT* commands here) **for each question**. We also sort our results by question using the *GROUP BY* command.
- The query I used to find this information is to the right.
- The results are below:

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

question	COUNT(DISTINCT user_id)
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

You can see that the number of responses *declines* with each question. This shows that customers are giving up and falling off the funnel at different points in the quiz.

## 2.2 Sniffing Out Questions with Low Completion

### Which questions lose the most potential customers?

- This is important because Warby Parker will want to adjust or remove any questions that seem to derail a lot of quiz takers. So, we want to find out which questions have the lowest completion rates. Let's look at the table from the previous slide showing the number of responses for each question (*table, top right*).

### How do we calculate completion rates?

- Using simple arithmetic, **I divide the # of people completing each step by the # of people completing the previous step.** The # of people who responded to the *first* question will start at 100%, because we haven't lost anyone yet. The results of each question's completion rates are shown here (*table, bottom right*).

question	COUNT(DISTINCT user_id)
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	completion rate
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%



## 2.3 Low Completion Results

### Results

- Our results show that Questions 3 and 5 have the lowest completion rates.

### Analysis

- Question 3: “Which shapes do you like?” My inference as to why this question has a low completion rate is that people do not know what shapes they are looking for yet. They may feel uncomfortable making a decision until they see examples of what’s available first. Perhaps they just don’t want to pigeonhole themselves into a particular shape yet.
- Question 5: “When was your last eye exam?” I assume that people defer from answering this question because they do not remember when their last eye exam was. I can relate because as I get older, it seems that time to escape from work to go to the eye doctor becomes less and less of a priority. If people are ashamed that they don’t remember their last eye exam, they may get discouraged and leave the quiz.

question	COUNT(DISTINCT user_id)
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	completion rate
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%

Part III:  
A/B Testing with  
Home Try-On Funnel



## 3.1 Introduction to A/B Testing

**A/B testing** is a method of comparing two versions of a single variable. The control is the original element that is currently in use. The variant is the new version that will be tested out in comparison against the original.

Let's say Warby Parker is analyzing their Home Try-On Funnel:

### Home Try-On Funnel:

- quiz
- home\_try\_on
- purchase

Warby Parker wants find out whether or not users who get more pairs of glasses to try on at home will be more likely to make a purchase. The variable being tested here is ***the number of glasses you get to try on at home.***

## 3.2 Testing Our Variables

**Variable 1:** 50% of users will get 3 pairs of glasses to try on at home

**Variable 2:** 50% of users will get 5 pairs of glasses to try on at home

We currently have different tables for each step in the funnel. We are going to have to combine these into one data set to see the results of our A/B testing. Let's start by seeing what we're tables we're working with by using the following query:

```
1 SELECT * FROM quiz
2 LIMIT 5;
3
4 SELECT * FROM home_try_on
5 LIMIT 5;
6
7 SELECT * FROM purchase
8 LIMIT 5;
```

The query results are on the following slide.

## Query:

```
SELECT * FROM quiz
LIMIT 5;

SELECT * FROM home_try_on
LIMIT 5;

SELECT * FROM purchase
LIMIT 5;
```

## Results:

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		
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-acco-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

From doing this exercise, I see that every table has one column in common: **user\_id**. We're now going to use a series of **left joins** to merge the three different tables using the **user\_id** column. We use a left join specifically because it allows us to keep all information even if some values are *null*.

## 3.3 Merging Tables

Here's the query I used to merge these tables using their matching **user\_id** columns:

```
SELECT distinct quiz.user_id, home_try_on.user_id IS  
NOT NULL AS 'is_home_try_on',  
home_try_on.number_of_pairs, purchase.user_id IS NOT  
NULL AS 'is_purchase'  
FROM quiz  
LEFT JOIN home_try_on  
    ON home_try_on.user_id = quiz.user_id  
LEFT JOIN purchase  
    ON purchase.user_id = quiz.user_id  
LIMIT 10;
```

*This SELECT statement allows me to retrieve the following columns:*

- 1) The user id of everyone who took the Quiz
- 2) Whether or not that user did the Home Try-On
- 3) How many glasses he/she tried on
- 4) If that user made a purchase or not

*These are my left joins on the common user\_id columns*

*I am retrieving only 10 rows so that the results will load quickly*

And the results....

Query Results			
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

This is our merged table of the Quiz, Home Try-On, and Purchase steps of the Home Try-On Funnel which we can nickname “Funnels”. I can take this new table and use it to deduce new insights like conversion rates between the steps. We just need to use a **With** statement like so, and we can write some more queries from this table:

```
WITH funnels AS (
  ...
  ...
  ...
)
SELECT COUNT(*) AS 'num_browse'
FROM funnels;
```

## 3.4 Calculating Conversion Rates

Here's the query I used for calculating the conversation rate between funnel stages:

```
WITH funnels AS (  
  SELECT distinct quiz.user_id,  
    home_try_on.user_id IS NOT NULL AS  
    'is_home_try_on', home_try_on.number_of_pairs,  
    purchase.user_id IS NOT NULL AS 'is_purchase'  
  FROM quiz  
  LEFT JOIN home_try_on  
    ON home_try_on.user_id = quiz.user_id  
  LEFT JOIN purchase  
    ON purchase.user_id = quiz.user_id)  
SELECT COUNT(*) AS 'num_quiz',  
  SUM(is_home_try_on) AS 'is_home_try_on',  
  SUM(is_purchase) AS 'is_purchase'  
FROM funnels;
```

*This is the  
table that  
we just  
merged!*

*This SELECT  
statement will  
show us how  
many users  
completed each  
step of the  
funnel.*

And here is the result:

Query Results		
num_quiz	is_home_try_on	is_purchase
1000	750	495

750 divided by 1000

**Conversion Rate  
from Quiz to  
Home Try-On:**

**75%**

495 divided by 750

**Conversion Rate  
from Home Try-  
On to Purchase:**

**66%**



## 3.4 Calculating Conversion Rates (continued)

Now in the previous slide, I did the math to find the conversion rates manually, but I can actually use a query to do the math for me if I add in two additional columns like so:

```
WITH funnels AS (  
  SELECT distinct quiz.user_id, home_try_on.user_id IS NOT  
  NULL AS 'is_home_try_on', home_try_on.number_of_pairs,  
  purchase.user_id IS NOT NULL AS 'is_purchase'  
  FROM quiz  
  LEFT JOIN home_try_on  
    ON home_try_on.user_id = quiz.user_id  
  LEFT JOIN purchase  
    ON purchase.user_id = quiz.user_id)  
SELECT COUNT(*) AS 'num_quiz',  
  SUM(is_home_try_on) AS 'is_home_try_on',  
  SUM(is_purchase) AS 'is_purchase',  
  1.0 * SUM(is_home_try_on) / COUNT(user_id) AS  
  'quiz_to_home_try_on',  
  1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS  
  'home_try_on_to_purchase'  
FROM funnels;
```

Query Results				
num_quiz	is_home_try_on	is_purchase	quiz_to_home_try_on	home_try_on_to_purchase
1000	750	495	0.75	0.66

I added this  
part to the  
SELECT  
statement...

...and got the  
same answer  
as I did on  
the previous  
slide!

## 3.5 A/B Test Results.... Finally

It's great that I found the overall conversion rates, but I still don't know the difference between the **users who tried on 3 pairs of glasses at home** and **users who tried on 5 pairs of glasses at home**. Now that I've gotten this far, all I just need to use a GROUP BY statement so I can clearly see the conversion rates per number of glasses tried on at home.

```
WITH funnels AS (  
  SELECT distinct quiz.user_id,      home_try_on.user_id IS NOT  
    NULL AS 'is_home_try_on', home_try_on.number_of_pairs,  
    purchase.user_id IS NOT NULL AS 'is_purchase'  
  FROM quiz  
  LEFT JOIN home_try_on  
    ON home_try_on.user_id = quiz.user_id  
  LEFT JOIN purchase  
    ON purchase.user_id = quiz.user_id)  
SELECT number_of_pairs, COUNT(*) AS 'num_quiz',  
  SUM(is_home_try_on) AS 'is_home_try_on',  
  SUM(is_purchase) AS 'is_purchase',  
  1.0 * SUM(is_home_try_on) / COUNT(user_id) AS  
  'quiz_to_home_try_on',  
  1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS  
  'home_try_on_to_purchase'  
FROM funnels  
GROUP BY number_of_pairs  
ORDER BY number_of_pairs;
```



*I added **number\_of\_pairs** to the  
SELECT statement*



*Here is my **GROUP BY** statement followed by an  
**ORDER BY** statement! The latter statement just means  
the results will show in order.*

And finally.... when I ran my query, I got these results!

## Results:

Query Results					
number_of_pairs	num_quiz	is_home_try_on	is_purchase	quiz_to_home_try_on	home_try_on_to_purchase
Ø	250	0	0	0.0	Ø
3 pairs	379	379	201	1.0	0.530343007915567
5 pairs	371	371	294	1.0	0.792452830188679

### Conversion Rates:

**3 Pairs: 53%**

**5 Pairs: 79%**

## Final Analysis:

The obvious winner of our A/B Test is **5 pairs of glasses**. Users were more likely to make a purchase if they tried on 5 pairs of glasses at home as opposed to 3 pairs of glasses. In fact, they were **26% more likely to make a purchase**. This is clear evidence that Warby Parker should switch to allowing users to try on 5 pairs of glasses at home after taking the style quiz. The extra shipping cost is probably minimal, and the revenue increase will certainly be there. I personally agree with the findings, because when I go to a glasses store I have try on ALL the glasses to find the perfect pair!

This is just one great example of how SQL can be used to analyze data and help businesses stay informed of their customers' needs and wants. I feel like this course gave me a great stepping stone into the world of SQL and data analysis. Thank you to everyone who contributed to this course and giving me this opportunity to learn something valuable. My journey has just begun!