



# Warby Parker Capstone

Learn SQL from Scratch

Schuyler Berland - sberland631@gmail.com

June 22 2019

# Table of Contents

1. Warby Parker - Background & Funnel Project
2. Quiz Funnel
3. A/B Testing with Home Try-On Funnel

# **1. Warby Parker - Background & Funnel Project**

# Warby Parker - Background & Funnel Project

**Company Background:** “[Warby Parker](#) is a transformative lifestyle brand with a lofty objective: to offer designer eyewear at a revolutionary price while leading the way for socially conscious businesses. Founded in 2010 and named after two characters in an early Jack Kerouac journal, Warby Parker believes in creative thinking, smart design, and doing good in the world. For every pair of eyeglasses and sunglasses sold, a pair is distributed to someone in need.” (description source, Codecademy)

**Funnel:** For my capstone project, the objective was to analyze various marketing funnels used by Warby Parker. I was provided two funnels with the following tables:

## 1. Quiz Funnel:

- `survey`

## 2. Home Try-On Funnel:

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

## **2. Quiz Funnel**

# Quiz Funnel, Question 1

To help users find their perfect frame, Warby Parker has a [Style Quiz](#) that has the following questions:

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

The users' responses are stored in a table called `survey`.  
**Select all columns from the first 10 rows. What columns does the table have?**

**EXPLANATION:**  
"SELECT \*" SELECTS ALL COLUMNS FROM A TABLE.  
"FROM SURVEY" SELECTS ALL ROWS AND COLUMNS FROM THE SURVEY TABLE.  
"LIMIT 10" LIMITS THE SELECTION TO THE FIRST 10 ROWS OF THE SURVEY TABLE.

**THE QUIZ FUNNEL CONTAINS THREE COLUMNS: QUESTION (the survey question), USER\_ID & RESPONSE (the selected answer from the user)**

QUERY:

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

question	user_id	response
1. What are you looking for?	005e7f99-d48c-4fce-b605-10506cd85aaf7	Women's Styles
2. What's your fit?	005e7f99-d48c-4fce-b605-10506cd85aaf7	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

# Quiz Funnel, Question 2

QUESTION: Users will “give up” at different points in the survey. Let’s analyze how many users move from Question 1 to Question 2, etc.

Create a quiz funnel using the `GROUP BY` command.

*What is the number of responses for each question?*

## EXPLANATION:

“SELECT QUESTION” SELECTS THE QUESTIONS AND MAKES IT THE FIRST COLUMN OF QUERY.

“COUNT (DISTINCT user\_id)” COUNTS THE NUMBER OF USERS WHO REACHED EACH QUESTION.

“GROUP BY QUESTION” GROUPS THE COUNT BY QUESTION, THUS ALLOWING FOR THE RESULTS ON THE RIGHT.

I chose distinct in case a user took the survey twice.

## QUERY:

```
SELECT question,  
COUNT(DISTINCT user_id)  
FROM survey  
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

# Quiz Funnel, Question 3

Using a spreadsheet program like Excel or Google Sheets, calculate the percentage of users who answer each question.:

*Which question(s) of the quiz have a lower completion rates?*

*What do you think is the reason?*

Add this finding to your presentation slides!

QUESTION	USER_ID COUNT	% QUESTION TO QUESTION COMPLETION
1. What are you looking for?	500	
2. What's your fit?	475	95.00%
3. Which shapes do you like?	380	80.00%
4. Which colors do you like?	361	95.00%
5. When was your last eye exam?	270	74.79%

## FINDINGS:

**THE LOWEST COMPLETION RATES WERE FOR QUESTIONS 3 AND 5.**

## THIS LIKELY MEANS THE FOLLOWING:

1. **THOSE COMPLETING THE SURVEY ARE PERHAPS NOT ENTIRELY SURE WHICH SHAPES THEY LIKE. THIS MAY PROVIDE AN OPPORTUNITY TO REPHRASE THE QUESTION AND SUGGEST SHAPES, OR SEND A SURPRISE STYLE TO CLIENTS.**
2. **THOSE COMPLETING THE SURVEY ARE NOT FAMILIAR WITH WHEN THEY LAST RECEIVED AN EYE EXAM. AS THIS HAS THE LOWEST COMPLETION, IT WOULD BE INTERESTING TO TEST IT AS THE FIRST QUESTION (TO SEE IF PEOPLE ARE GETTING BORED BY END OF SURVEY, OR IF PEOPLE ARE FILLING THIS OUT LEAST BECAUSE THEY CANNOT RECALL THE LAST TIME THEY RECEIVED AN EYE EXAM).**



### **3. A/B Testing with Home-Try-On funnel**

# A/B Testing with Home-Try-On Funnel, Question 4

Warby Parker's purchase funnel is:

Take the Style Quiz → Home Try-On → Purchase the Perfect Pair of Glasses

During the Home Try-On stage, we will be conducting an A/B Test:

- 50% of the users will get **3** pairs to try on
- 50% of the users will get **5** pairs to try on

*Let's find out whether or not users who get more pairs to try on at home will be more likely to make a purchase.*

The data will be distributed across three tables:

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

Examine the first five rows of each table

What are the column names?

**THE EXPLANATION FOR THIS IS SIMILAR TO QUESTION ONE, "SELECT \*" SELECTS ALL COLUMNS AND ROWS FOR THE TABLE, I ADDED A LIMIT 5 AS ONLY THE FIRST FIVE ROWS ARE TO BE EXAMINED.**

**THE COLUMN NAMES FOR THE TABLES ARE THE FOLLOWING:**

**QUIZ: USER\_ID,STYLE, FIT, SHAPE, COLOR**

**HOME\_TRY\_ON: USER\_ID, NUMBER\_OF\_PAIRS, ADDRESS**

**PURCHASE: USER\_ID,PRODUCT\_ID, STYLE,\_MODEL\_NAME,COLOR\_PRICE**

**QUERY:**

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

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

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

# A/B Testing with Home-Try-On Funnel, Question 5

We'd like to create a new table with the following layout:

Each row will represent a single user from the browse table:

If the user has any entries in `home_try_on`, then `is_home_try_on` will be 'True'.

`number_of_pairs` comes from `home_try_on` table

If the user has any entries in `is_purchase`, then `is_purchase` will be 'True'.

Use a `LEFT JOIN` to combine the three tables, starting with the top of the funnel (browse) and ending with the bottom of the funnel (purchase).

Select only the first 10 rows from this table (otherwise, the query will run really slowly).

## EXPLANATION:

I first created a table with 4 columns: `user_id`, `is_home_try_on`, `number_of_pairs` and `is_purchase`. I renamed table `quiz` as `q`, table `home_try_on` as `h` and table `purchase` as `p` to simplify the query. I did a left join combining all three tables. I started my left join from the `quiz` table because it keeps all user ids from the `quiz` table in the final results. For `is_home_try_on` and `is_purchase`, I created the columns as "IS NOT NULL" because it creates values of 1 and 0, which can be summed later on for conversions.

As the question asked for the results to appear as True or False, and not as 1 or 0, I then had to wrap the query into a `WITH` statement. I could then create a table using my previous results where 1 = True and 0 = False (using a `CASE` statement).

I did `LIMIT 10` to select only the first 10 rows.

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc8dc	True	3	False
291f1cca	True	5	False
75122300300	False	NULL	False

## QUERY:

```
WITH funnel AS (SELECT q.user_id AS user_id,
h.user_id IS NOT NULL AS 'is_home_try_onn',
h.number_of_pairs AS number_of_pairs,
p.user_id IS NOT NULL AS 'is_purchasee'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
ON p.user_id = h.user_id)
SELECT user_id,
CASE
WHEN is_home_try_onn = 0 THEN 'False' ELSE 'True' END AS
'is_home_try_on',
number_of_pairs,
CASE
WHEN is_purchasee = 0 THEN 'False' ELSE 'True' END AS 'is_purchase'
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
75bcefebd-40cd-4e1d-a301-27d6d93b12e2	True	5 pairs	False
ce965c46-7a2b-4d86-9847-60174767812	True	3 pairs	True

# A/B Testing with Home-Try-On Funnel, Question 6

Once we have the data in this format, we can analyze it in several ways:

- We can calculate overall conversion rates by aggregating across all rows.
- We can compare conversion from `quiz`→`home_try_on` and `home_try_on`→`purchase`.
- We can calculate the difference in purchase rates between customers who had 3 `number_of_pairs` with ones who had 5.
- And more!

We can also use the original tables to calculate things like:

- The most common results of the style `quiz`.
- The most common types of `purchase` made.
- And more!

*What are some actionable insights for Warby Parker?*

## EXPLANATION 1: CONVERSION

**This query looks at the conversion rates from `quiz` → `home try on` → `purchase`.**

**I used the query from the last question, except I removed the case statements. The 0 and 1s allow for easy sums to calculate conversions.**

**I continue to use the `WITH` statement, and then I create a count column for quiz users and sums for the home try on and purchase users columns. I then divide the sum of home try ons over the count of quiz ids to find the conversion rate for the home try on program. I then divided the sum of purchasing users over the sum of `home_try_on` users to calculate that conversion rate.**

**The percent of those who purchase after requesting the home try on is actually very high (66%), so this is quite an impressive number! The % of those interested in the home try on is also very high (75%). I think it would be interesting to later see what percent of those who decline the home try on eventually make a purchase at the store. I think we can work on increasing the % purchase number as well.**

## FIRST QUERY:

```
WITH funnel AS (SELECT q.user_id AS user_id,
h.user_id IS NOT NULL AS 'is_home_try_on',
h.number_of_pairs AS number_of_pairs,
p.user_id IS NOT NULL AS 'is_purchase'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
ON p.user_id = h.user_id)
SELECT COUNT(*) as 'num_quiz', SUM(is_home_try_on) AS
'num_home_try_on',
SUM(is_purchase) as 'num_purchase',
1.0*SUM(is_home_try_on)/COUNT(user_id) AS '%_home_try_on',
1.0*SUM(is_purchase)/SUM(is_home_try_on) AS '%_purchase'
FROM funnel;
```

num_quiz	num_home_try_on	num_purchase	_%_home_try_on	_%_purchase
1000	750	495	0.75	0.66

# A/B Testing with Home-Try-On Funnel, Question 6

Once we have the data in this format, we can analyze it in several ways:

- We can calculate overall conversion rates by aggregating across all rows.
- We can compare conversion from `quiz`→`home_try_on` and `home_try_on`→`purchase`.
- We can calculate the difference in purchase rates between customers who had 3 `number_of_pairs` with ones who had 5.
- And more!

We can also use the original tables to calculate things like:

- The most common results of the style `quiz`.
- The most common types of `purchase` made.
- And more!

*What are some actionable insights for Warby Parker?*

**EXPLANATION 2: DIFFERENCE IN PURCHASE RATES**

This query is very similar to the query in the last slide, except that I grouped by the number of pairs from `home_try_on`. The findings show that those who try on 5 pairs are far more likely to purchase (79%) as opposed to (53%) who try on only 3 pairs.

One experiment may be to test only providing 5 pairs to try on for a month or two. We can analyze if the percentage is still as high, or at least higher than 53%.

**SECOND QUERY:**

```
WITH funnel AS (SELECT q.user_id AS user_id,
h.user_id IS NOT NULL AS 'is_home_try_on',
h.number_of_pairs AS number_of_pairs,
p.user_id IS NOT NULL AS 'is_purchase'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
ON p.user_id = h.user_id)
SELECT number_of_pairs,COUNT(number_of_pairs),
SUM (is_purchase) as 'number_purchase',
1.0 * SUM(is_purchase)/COUNT(number_of_pairs) AS '%_purchase'
FROM funnel
GROUP BY number_of_pairs
HAVING number_of_pairs IS NOT NULL;
```

number_of_pairs	COUNT(number_of_pairs)	number_purchase	%_purchase
3 pairs	379	201	0.530343007915567
5 pairs	371	294	0.792452830188679

# A/B Testing with Home-Try-On Funnel, Question 6

Once we have the data in this format, we can analyze it in several ways:

- We can calculate overall conversion rates by aggregating across all rows.
- We can compare conversion from `quiz`→`home_try_on` and `home_try_on`→`purchase`.
- We can calculate the difference in purchase rates between customers who had 3 `number_of_pairs` with ones who had 5.
- And more!

We can also use the original tables to calculate things like:

- The most common results of the style `quiz`.
- The most common types of `purchase` made.
- And more!

*What are some actionable insights for Warby Parker?*

## EXPLANATION 3: QUIZ STYLES & SHAPE

For these queries, I looked at if mens or womens styles were more favored, as well as popular shapes. It appears women's is favored, but not by a large margin. For shapes, rectangular and square were the most popular shapes.

For the query I selected styles first (in order to later GROUP BY styles). I then counted the `user_ids`, and divided by 1000, as this was the total number of those who took the quiz.

For shapes, I did a simple query with select, count and GROUP BY in order to view the count by style.

### THIRD QUERY:

```
STYLE:SELECT style,
COUNT (DISTINCT user_id) AS 'user_count',
1.0* COUNT(user_id)/1000 AS '% of users'
FROM quiz
GROUP BY style;
```

```
SHAPE: SELECT shape,
COUNT (shape)
FROM quiz
GROUP BY shape;
```

style	user_count	% of users
I'm not sure. Let's skip it.	99	0.099
Men's Styles	432	0.432
Women's Styles	469	0.469

shape	COUNT (shape)
No Preference	97
Rectangular	397
Round	180
Square	326

# A/B Testing with Home-Try-On Funnel, Question 6

Once we have the data in this format, we can analyze it in several ways:

- We can calculate overall conversion rates by aggregating across all rows.
- We can compare conversion from `quiz`→`home_try_on` and `home_try_on`→`purchase`.
- We can calculate the difference in purchase rates between customers who had 3 `number_of_pairs` with ones who had 5.
- And more!

We can also use the original tables to calculate things like:

- The most common results of the style `quiz`.
- The most common types of `purchase` made.
- And more!

*What are some actionable insights for Warby Parker?*

## EXPLANATION 4: PURCHASE PRICE

The middle price point of 95 dollars is the most popular purchase price for a pair of glasses (at 52% of purchases). The premium styles are second, with 39%. Only 8% of clients purchase the 50 dollar price point. One actionable insight may be to expand where we are marketing the \$50 dollar glasses. The current client base is not buying them, are we marketing \$50 glasses to the wrong markets? One action may also be to continue highlighting the premium selection, as it does well, and expand on that line.

### FOURTH QUERY:

```
Price: SELECT price,
COUNT (user_id),
1.0*COUNT(user_id)/495 as '%_purchase'
FROM
purchase
GROUP BY price;
```

```
Shape by Gender: WITH results AS (SELECT user_id,
REPLACE (style,""," ") as styles,
fit,
shape,
color
FROM quiz)
SELECT color, COUNT (color),
1.0* COUNT (color)/531
FROM results
WHERE styles = 'Womens Styles'
GROUP BY color
ORDER BY 3 DESC;
```

price	COUNT (user_id)	['_purchase]
50	41	0.0828282828282828
95	261	0.527272727272727
150	193	0.389898989898989