code_cademy

Capstone: Funnels - Warby Parker

Learn SQL From Scratch
Jamie Lynne Weingard
July, 2018

Table of Contents

1: Getting Familiar with Warby Parker's Data

1.1 Warby Parker – Style Quiz

Question 1:
Warby Parker's Style Quiz contains 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?

Analysis:
Columns for this table include the questions above, the user_id associated with each question and the user's response for said question. I've limited the results to 10.

```
-- All from Survey Table, Limit 10

SELECT *
FROM survey
LIMIT 10;
```

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

2: What is the Quiz Funnel?

2.1 What is the Quiz Funnel?

Question 2:
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?

Analysis:

Q1: 500
Q2: 475
Q3: 380
Q4: 361
Q5: 270

We see that as the responders progress through the survey, they respond to fewer of the questions.

```
-- Query # 1

SELECT question,
COUNT(DISTINCT user_id) AS 'User Count'
FROM survey
GROUP BY question;

-- OR --

--Query # 2
SELECT question,
COUNT(DISTINCT user_id) AS 'User Count'
FROM survey
GROUP BY 1;
```

question	User Count
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

2.2 What is the 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?

Analysis:
The number of responses reduces as the survey progresses. The total number of responders is 500 yet only 270 answered the last question.

We can see that some users don't answer all questions or may not be sure of what they are looking for.   If the user isn't sure what you are looking for then it can be difficult to answer the remaining questions.

-- No statement requested – table below completed in Excel

question	User Count	Percentage Completed
1. What are you looking for?	500	100%
2. What's your fit?	475	95%
3. Which shapes do you like?	380	76%
4. Which colors do you like?	361	72%
5. When was your last eye exam?	270	54%

3: A/B Testing with Home Try-On Funnel

3.1 A/B Testing w/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

What are the column names?

Analysis:
Form the quiz table, the columns names are: user_id, style, fit, shape, color

From the home_try_on table, the column names are: user_id, number_of_pairs, address

From the purchase table, the column names are: user_id, product_id, style, model_name, color, price

See screen shot to right that shows 5 records from each table

```
SELECT *
FROM quiz
LIMIT 5;

SELECT *
FROM home_try_on
LIMIT 5;

SELECT *
FROM purchase
LIMIT 5;
```

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-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

4: Create New Table
using Joins & Analyze Data

4.1 Create New Table Using Joins

Question 5:
We'd like to create a new table with the following layout

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc	True	3	False
291f1cca	True	5	False
75122300	False	NULL	False

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).

```
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 AS 'q'
LEFT JOIN home_try_on AS 'h'
    ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
    ON h.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

4.2 Create New Table Using Joins

Continuing from the previous slide, I created the table however the results for True were showing 1, False was showing 0 and Null was showing 0. So I then took this a step further convert 1 to True , 0 to False and 0 (pairs) to NULL. See new results below.

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc	True	3	False
291f1cca	True	5	False
75122300	False	NULL	False

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	NULL	False
75bc6ebd-40cd-4e1d-a301-27ddd93b12e2	True	5 pairs	False
ce965c4d-7a2b-4db6-9847-601747fa7812	True	3 pairs	True
28867d12-27a6-4e6a-a5fb-8bb5440117ae	True	5 pairs	True
5a7a7e13-fbcf-46e4-9093-79799649d6c5	False	NULL	False
0143cb8b-bb81-4916-9750-ce956c9f9bd9	False	NULL	False
a4ccc1b3-cbb6-449c-b7a5-03af42c97433	True	5 pairs	False
b1dded76-cd60-4222-82cb-f6d464104298	True	3 pairs	False

```
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'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
  ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
  ON h.user_id = p.user_id)
SELECT user_id,
CASE
  WHEN is_home_try_on = 1 THEN 'True'
    ELSE 'False'
  END as 'is_home_try_on',
CASE
  WHEN number_of_pairs IS NULL THEN 'NULL'
    ELSE number_of_pairs
  END AS 'number_of_pairs',
CASE
  WHEN is_purchase = 1 THEN 'True'
    ELSE 'False'
  END AS 'is_purchase'
FROM funnels
LIMIT 10;
```

4.3   Create New Table Using Joins & Analyze Data

Question: 5
Now that 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 see that there are 1000 responders with 750 home try ons, 2992 pairs & 495 purchases

We can compare conversion from quiz→home_try_on and home_try_on→purchase.
- We see that of the 1000 responders – 75% did a home try on and 66% of these responders made a purchase
- We can also see that of the 1000 responders – 49% made a purchase

--Aggregate Across All Rows

num_browses	home_try_ons	pairs	purchases
1000	750	2992.0	495

--Conversion Rate From Quiz > home_try_on & home_try_on > purchase

num_browses	home_try_ons	purchases	quiz_to_home_try_on	home_try_on_to_purchase	quiz_to_purchase
1000	750	495	0.75	0.66	0.495

--Aggregate Across All Rows
```
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'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
  ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
  ON h.user_id = p.user_id)
SELECT COUNT(*) AS 'num_browses',
 SUM(is_home_try_on) AS 'home_try_ons',
 SUM(number_of_pairs) AS 'pairs',
 SUM(is_purchase) AS 'purchases'
FROM funnels;
```

--Conversion Rate From Quiz > home_try_on & home_try_on > purchase
```
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'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
  ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
  ON h.user_id = p.user_id)
SELECT COUNT(*) AS 'num_browses',
  SUM(is_home_try_on) AS 'home_try_ons',
  SUM(is_purchase) AS 'purchases',
  1.0 * SUM(is_home_try_on) / COUNT(*) AS 'quiz_to_home_try_on',
  1.0 * SUM(is_purchase) / SUM(is_home_try_on) AS 'home_try_on_to_purchase'
  1.0 * SUM(is_purchase) / COUNT (*) AS 'quiz_to_purchase'
FROM funnels;
```

4.4 Create New Table Using Joins & Analyze Data

Continued….

We can calculate the difference in purchase rates between customers who had 3 number_of_pairs with ones who had 5.

Analysis:
We see that there are a total of 495 purchases made, below is the breakout:

• 201 (41%) purchases were made for 3 pairs
• 294 (59%) purchases were made for 5 pairs

purchases	3P	3P %	5P	5P %
495	201	0.41	294	0.59

```sql
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'
FROM quiz AS 'q'
LEFT JOIN home_try_on AS 'h'
  ON q.user_id = h.user_id
LEFT JOIN purchase AS 'p'
  ON h.user_id = p.user_id)
SELECT SUM(is_purchase) AS 'purchases',
  COUNT(DISTINCT CASE
   WHEN number_of_pairs = '3 pairs' THEN user_id
  END) AS '3P',
   ROUND(1.0 * COUNT(DISTINCT CASE
   WHEN number_of_pairs = '3 pairs' THEN user_id
   END) / SUM(is_purchase), 2) AS '3P %',
    COUNT(DISTINCT CASE
   WHEN number_of_pairs = '5 pairs' THEN user_id
   END) AS '5P',
   ROUND(1.0 * COUNT(DISTINCT CASE
   WHEN number_of_pairs = '5 pairs' THEN user_id
   END) / SUM(is_purchase), 2) AS '5P %'
FROM funnels
WHERE is_purchase <> 0;
```

5: Additional Queries & Analysis – Quiz Table

5.1 Quiz Table – Queries & Analysis

What are the most common results from the Quiz?

Analysis:
For the first query, I pulled fit and counted the number of user responses by fit and ordered by in descending order. For the 2nd query, I followed the same format but used shape. For the 3rd query I followed the same format and used color.

As you can see the most common overall responses (regardless of style), users are looking for a Narrow fit, Rectangular shape and Tortoise color, followed quickly by black (the old classic).

The least common color was Two-Tone with 104 responses. Still pretty relevant.

There are 89 that weren't sure of the fit and another 97 that weren't sure of the shape.

Warby Parker may want to increase their Narrow/Rectangular frame options

```
--Top Overall Responses
SELECT fit, COUNT(user_id) AS 'COUNT'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC;

SELECT shape, COUNT(user_id) AS 'COUNT'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC;

SELECT color, COUNT(color) AS 'COUNT'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC;
```

fit	COUNT
Narrow	408
Medium	305
Wide	198
I'm not sure. Let's skip it.	89

shape	COUNT
Rectangular	397
Square	326
Round	180
No Preference	97

color	COUNT
Tortoise	292
Black	280
Crystal	210
Neutral	114
Two-Tone	104

5.2 Quiz Table – Queries & Analysis

What else can we learn from the quiz table.

Analysis:
For each of these 3 queries, I have pulled in the fit, shape, color, then filtered the data for each style type and then counted all of the user ids for each. I have limited to the top 5 responses for each style.

For the Men's Styles, we can see that 23 users indicated narrow, rectangular, tortoise glasses.

For the Women's Styles, we can see that 20 users indicated narrow, rectangular black glasses.

For the users that just weren't sure, we can see that 8 users indicated narrow, square, black glasses.

fit	shape	color	Mens Count
Narrow	Rectangular	Tortoise	23
Medium	Rectangular	Tortoise	19
Narrow	Rectangular	Black	18
Medium	Rectangular	Black	17
Narrow	Square	Tortoise	16

fit	shape	color	Womens Count
Narrow	Rectangular	Black	20
Narrow	Rectangular	Tortoise	20
Medium	Rectangular	Tortoise	16
Medium	Square	Tortoise	16
Narrow	Square	Crystal	16

fit	shape	color	Unsure Count
Narrow	Square	Black	8
Narrow	Rectangular	Black	7
Wide	Rectangular	Crystal	6
Medium	Square	Black	5
Medium	Square	Tortoise	5

```
--TOP COMBINATIONS

--MEN'S COUNT - TOP 5
SELECT fit, shape, color,
COUNT(CASE
    WHEN style LIKE 'Men%' THEN user_id
    END) AS 'Mens Count'
FROM quiz
GROUP BY 1, 2, 3
ORDER BY 4 DESC
LIMIT 5;

--WOMEN'S COUNT - TOP 5
SELECT fit, shape, color,
COUNT(CASE
    WHEN style LIKE 'Wo%' THEN user_id
    END) AS 'Womens Count'
FROM quiz
GROUP BY 1, 2, 3
ORDER BY 4 DESC
LIMIT 5;

--UNSURE/UNISEX COUNT - TOP 5
SELECT fit, shape, color,
COUNT(CASE
    WHEN style LIKE '%not%' THEN user_id
    END) AS 'Unsure Count'
FROM quiz
GROUP BY 1, 2, 3
ORDER BY 4 DESC
LIMIT 5;
```

5.3 Quiz Table – Queries & Analysis

Let's keep looking for new ways we can look at this data.

Analysis:
For the first query, I pulled the fit and counted the responses by style. Then I grouped the responses by fit and ordered the responses in descending order from Mens to Womens to Unsure/Unisex. I followed the same method for shape & color

For Fit, we can see that 174 Men, 189 Women & 45 Unsure/Unisex are looking for Narrow glasses

For Shape, we can see that 176 Men, 184 Women & 37 Unsure/Unisex are looking for Rectangular glasses

For Color, we can see that 128 Men, 142 Women & 22 Unsure/Unisex are looking for Tortoise glasses

Notice there are more folks who are unsure that know they want black versus tortoise glasses.

fit	Mens Count	Womens Count	Unsure/Unisex Count
Narrow	174	189	45
Medium	142	131	32
Wide	79	103	16
I'm not sure. Let's skip it.	37	46	6

shape	Mens Count	Womens Count	Unsure/Unisex Count
Rectangular	176	184	37
Square	132	158	36
Round	80	81	19
No Preference	44	46	7

color	Mens Count	Womens Count	Unsure/Unisex Count
Tortoise	128	142	22
Black	121	126	33
Crystal	81	106	23
Two-Tone	58	37	9
Neutral	44	58	12

```
--TOTAL COUNT BY STYLE & FIT
SELECT fit,
COUNT(CASE
    WHEN style LIKE 'Men%' THEN user_id
    END) AS 'Mens Count',
COUNT(CASE
    WHEN style LIKE '%Wome%' THEN user_id
    END) AS 'Womens Count',
COUNT(CASE
    WHEN style LIKE '%sure%' THEN user_id
    END) AS 'Unsure/Unisex Count'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC, 3 DESC, 4 DESC;

--TOTAL COUNT BY STYLE & SHAPE
SELECT shape,
COUNT(CASE
    WHEN style LIKE 'Men%' THEN user_id
    END) AS 'Mens Count',
COUNT(CASE
    WHEN style LIKE '%Wome%' THEN user_id
    END) AS 'Womens Count',
COUNT(CASE
    WHEN style LIKE '%sure%' THEN user_id
    END) AS 'Unsure/Unisex Count'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC, 3 DESC, 4 DESC;

--TOTAL COUNT BY STYLE & COLOR
SELECT color,
COUNT(CASE
    WHEN style LIKE 'Men%' THEN user_id
    END) AS 'Mens Count',
COUNT(CASE
    WHEN style LIKE '%Wome%' THEN user_id
    END) AS 'Womens Count',
COUNT(CASE
    WHEN style LIKE '%sure%' THEN user_id
    END) AS 'Unsure/Unisex Count'
FROM quiz
GROUP BY 1
ORDER BY 2 DESC, 3 DESC, 4 DESC;
```

5.4 Quiz Table – Queries & Analysis

I was curious to see what the most common type of eyeglass frame users were looking for when responding to the quiz.

Analysis:
For Women Styles, the most popular type of glasses are Narrow/Rectangular (74 responses) while the lowest number would be for Medium/Round (14 responses)

For Men's Styles, the most popular type of glasses are Narrow/Rectangular (72 responses) while the lowest number would be for Wide/Round (18 responses)

We know based on the available responses to the quiz that users don't have to pick every option or they can respond "I'm Not Sure…". Some users may not have a preference on the style (Men's vs Women's) but they did know what type of fit/shape they were looking for. The highest number again was for Narrow/Rectangular glasses.

Warby Parker may want to include a Unisex category that provides both Men's or Women's frames suitable for both genders.

style	Narrow/Rectangular	Medium/Rectangular	Wide/Rectangular	Unsure/Rectangular	Narrow/Round	Medium/Round	Wide/Round	Unsure/Round	Narrow/Square	Medium/Square	Wide/Square	Unsure/Square
Women's Styles	74	55	40	0	40	14	17	0	58	44	37	0
Men's Styles	72	61	30	0	30	26	18	0	56	38	27	0
I'm not sure. Let's skip it.	17	11	9	0	8	3	4	0	18	15	1	0

See larger scale of chart on next slide…

```
--COUNTS BY STYLE/SHAPE COMBO
SELECT style,
COUNT(CASE
 WHEN shape = 'Rectangular'
  AND fit = 'Narrow' THEN user_id
END) AS 'Narrow/Rectangular',
COUNT(CASE
  WHEN shape = 'Rectangular'
   AND fit = 'Medium' THEN user_id
END) AS 'Medium/Rectangular',
COUNT(CASE
   WHEN shape = 'Rectangular'
    AND fit = 'Wide' THEN user_id
END) AS 'Wide/Rectangular',
COUNT(CASE
    WHEN shape = 'Rectangular'
     AND fit = '%not_sure%' THEN user_id
END) AS 'Unsure/Rectangular',
 COUNT(CASE
    WHEN shape = 'Round'
     AND fit = 'Narrow' THEN user_id
END) AS 'Narrow/Round',
COUNT(CASE
    WHEN shape = 'Round'
     AND fit = 'Medium' THEN user_id
END) AS 'Medium/Round',
COUNT(CASE
    WHEN shape = 'Round'
     AND fit = 'Wide' THEN user_id
END) AS 'Wide/Round',
COUNT(CASE
    WHEN shape = 'Round'
     AND fit = '%not_sure%' THEN user_id
END) AS 'Unsure/Round',
COUNT(CASE
    WHEN shape = 'Square'
     AND fit = 'Narrow' THEN user_id
END) AS 'Narrow/Square',
COUNT(CASE
    WHEN shape = 'Square'
     AND fit = 'Medium' THEN user_id
END) AS 'Medium/Square',
COUNT(CASE
    WHEN shape = 'Square'
     AND fit = 'Wide' THEN user_id
END) AS 'Wide/Square',
OUNT(CASE
    WHEN shape = 'Square'
     AND fit = '%not_sure%' THEN user_id
END) AS 'Unsure/Square'
FROM quiz
GROUP BY 1
ORDER BY 1 DESC;
```

5.4a Quiz Table – Queries & Analysis

style	Narrow/Rectangular	Medium/Rectangular	Wide/Rectangular	Unsure/Rectangular	Narrow/Round	Medium/Round	Wide/Round	Unsure/Round	Narrow/Square	Medium/Square	Wide/Square	Unsure/Square
Women's Styles	74	55	40	0	40	14	17	0	58	44	37	0
Men's Styles	72	61	30	0	30	26	18	0	56	38	27	0
I'm not sure. Let's skip it.	17	11	9	0	8	3	4	0	18	15	1	0

5.5 Quiz Table – Queries & Analysis

What are the least common results from the Quiz?

Analysis:
I wanted to see of the users who chose Men's or Women's, what is the least common combinations for the responses.

We can see that Round, Wide, Two-Tone glasses are not common, only 1 response.

```
--BOTTOM 20 COMBINATIONS - REMOVING UNSURE RESPONSES
SELECT fit, shape, color, COUNT(user_id)
FROM quiz
WHERE fit NOT LIKE '%sure%'
  AND style NOT LIKE '%sure%'
  AND shape NOT LIKE 'No Preference'
GROUP BY 1, 2, 3
ORDER BY 4
LIMIT 20;
```

fit	shape	color	COUNT(user_id)
Wide	Round	Two-Tone	1
Medium	Round	Two-Tone	4
Wide	Round	Neutral	4
Medium	Round	Crystal	5
Medium	Round	Neutral	5
Narrow	Round	Two-Tone	5
Wide	Rectangular	Neutral	6
Wide	Square	Two-Tone	6
Medium	Square	Two-Tone	7
Wide	Round	Black	7
Narrow	Round	Neutral	8
Wide	Rectangular	Two-Tone	8
Wide	Round	Crystal	8
Wide	Square	Neutral	8
Medium	Rectangular	Neutral	11
Wide	Rectangular	Crystal	11
Medium	Round	Black	12
Medium	Square	Neutral	12
Narrow	Square	Neutral	13
Narrow	Square	Two-Tone	13

5.6 Quiz Table – Queries & Analysis

What are the results if the users is only looking for a specific color or shape?

Analysis:
Only 9 people are looking only by color

color	COUNT
Black	5
Crystal	2
Tortoise	2

32 people are looking for Square, 28 Rectangular and 20 Round for a total of 80 people searching for a particular shape.

shape	COUNT
Square	32
Rectangular	28
Round	20

```
--Looking only by color
SELECT color, COUNT(color) AS 'COUNT'
FROM quiz
WHERE fit LIKE '%not%'
  AND shape LIKE '%no%'
GROUP BY 1;

--Looking only by shape
SELECT shape, COUNT(shape) AS 'COUNT'
FROM quiz
WHERE fit LIKE '%not%'
  AND shape NOT LIKE 'No%'
GROUP BY 1;
```

6: Additional Queries & Analysis – Purchase Table

6.1 Purchase Table – Queries & Analysis

Now let's analyze the purchase table

Analysis:
First SQL gives us all columns from the purchase table; limited to 10 records.

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
021901a5-74ee-482d-9f03-0089d17ff7d6	3	Men's Styles	Dawes	Driftwood Fade	150
026e11d9-926c-4206-af7c-273e3ee6ad7f	2	Men's Styles	Brady	Sea Glass Gray	95
028b47b0-e075-4b2c-bcf3-f963d8089449	10	Women's Styles	Eugene Narrow	Rosewood Tortoise	95
029e417e-a0b5-4ec7-9190-a852016619d9	5	Men's Styles	Monocle	Endangered Tortoise	50
02bd0f5a-3f04-4902-b005-82188b07785f	7	Women's Styles	Lucy	Elderflower Crystal	150

The 2nd SQL gives us a catalog of available items to purchase. There are a total of 10 types of frames available (half are Men's and the other half are Women's).

product_id	style	model_name	color	price
1	Men's Styles	Brady	Layered Tortoise Matte	95
2	Men's Styles	Brady	Sea Glass Gray	95
3	Men's Styles	Dawes	Driftwood Fade	150
4	Men's Styles	Dawes	Jet Black	150
5	Men's Styles	Monocle	Endangered Tortoise	50
6	Women's Styles	Olive	Pearled Tortoise	95
7	Women's Styles	Lucy	Elderflower Crystal	150
8	Women's Styles	Lucy	Jet Black	150
9	Women's Styles	Eugene Narrow	Rose Crystal	95
10	Women's Styles	Eugene Narrow	Rosewood Tortoise	95

```
--Purchase Table – Limit 10
SELECT *
FROM purchase
LIMIT 10;


--Product Listing
SELECT DISTINCT product_id, style, model_name,
color, price
FROM purchase
ORDER BY 1;
```

6.2 Purchase Table – Queries & Analysis

Analysis:

The 1st SQL gives us the Rounded Average Price from the purchase table

AVG PRICE
112.72

The 2nd SQL gives us the Maximum & Minimum price from the purchase table

MAX(price)	MIN(price)
150	50

The 3rd SQL gives us the Total Sales which is a Rounded Sum of all Prices

Total Sales	Purchases
55795.0	495

The 4th SQL gives us the Total Sales by Style

style	Total Purchased	Total Sales
Women's Styles	252	28670
Men's Styles	243	27125

Let's see what the break out would by in % (done in excel)

Total Purchases = 495
Women's Styles =   252   51%
Men's Styles =   243   49%

Total Sales = $55,795.00
Women's Styles = $28,670.00   49%
Men's Styles = $27,125.00   51%

Based on this there are more purchases made for Women's styles but the Men's styles bring in more revenue (not by much as the split is almost 50/50)

```
--Rounded Avg Price
SELECT ROUND(AVG(price), 2) AS 'AVG PRICE'
FROM purchase;

--Min & Max Prices
SELECT MAX(price),
MIN(price)
FROM purchase;

--Rounded Sum of all Prices = Total Sales
SELECT ROUND(SUM(price), 2) AS 'Total Sales', COUNT(user_id) AS 'Total
Purchases'
FROM purchase;

--Total Sales by Style
SELECT style, COUNT(*) AS 'Total Purchased', SUM(price) AS 'Total Sales'
from purchase
GROUP BY 1
ORDER BY 2 DESC;
```

6.3 Purchase Table – Queries & Analysis

Now let's find the most common products sold

Analysis:
The 1st SQL gives us a count of purchases by Color. The most common/popular color is Jet Black.

color	Color Count
Jet Black	86
Driftwood Fade	63
Rosewood Tortoise	62
Rose Crystal	54
Layered Tortoise Matte	52
Pearled Tortoise	50
Elderflower Crystal	44
Sea Glass Gray	43
Endangered Tortoise	41

The 2nd SQL gives us a count of purchases by Model Name. The most common is Eugene Narrow

model_name	Model Count
Eugene Narrow	116
Dawes	107
Brady	95
Lucy	86
Olive	50
Monocle	41

Continue to next page for more details

```
--Most common Color Purchased
SELECT color, COUNT(user_id) AS 'Color Count'
FROM purchase
GROUP BY 1
ORDER BY 2 DESC;

--Most common Model Name purchased
SELECT model_name, COUNT(user_id) AS 'Model Count'
FROM purchase
GROUP BY 1
ORDER BY 2 DESC;
```

6.4 Purchase Table – Queries & Analysis

Analysis:

The 1st SQL gives us a count of products sold for Men's Styles by product_id. The most common is product_id = 3 with 63 purchases. The least common is product_id = 5 with 41 purchases.

style	product_id	Total Purchased	price	Total
Men's Styles	3	63	150	9450
Men's Styles	1	52	95	4940
Men's Styles	4	44	150	6600
Men's Styles	2	43	95	4085
Men's Styles	5	41	50	2050

Product_ID	% of Purchases (243)	% of Total $27,125
3	26%	35%
1	21%	18%
4	18%	24%
2	18%	15%
5	17%	8%

The 2nd SQL gives us a count of products sold for Women's Styles by product_id. The most common is product_id = 10 with 62 purchases. The least common is product_id = 8 with 42 purchases.

style	product_id	Total Purchased	price	Total Sales
Women's Styles	10	62	95	5890
Women's Styles	9	54	95	5130
Women's Styles	6	50	95	4750
Women's Styles	7	44	150	6600
Women's Styles	8	42	150	6300

Product_ID	% of Purchases (252)	% of Total $28,670
3	25%	21%
1	21%	18%
4	20%	17%
2	17%	23%
5	17%	22%

```
--Most Common Product Sold – Men's
SELECT style, product_id, COUNT(*) AS 'Total Purchased',
price, SUM(price) AS 'Total'
FROM purchase
WHERE style LIKE 'Men%'
GROUP BY 2
ORDER BY 3 DESC;

--Most Common Product Sold – Women's
SELECT style, product_id, COUNT(*) AS 'Total Purchased',
price, SUM(price) AS 'Total Sales'
FROM purchase
WHERE style LIKE 'Women%'
GROUP BY 2
ORDER BY 3 DESC;
```

6.5 Purchase Table – Queries & Analysis

Analysis:
This SQL gives us a look at all products sold by Style and Product ID. This also gives us the total number purchased and the total sales.

We can see that product_id =3 is the top seller overall while product_id = 5 is the lowest seller overall.

```
--Most Common Product Sold – ALL
SELECT style, product_id, COUNT(*) AS 'Total Purchased',
price, SUM(price) AS 'Total Sales'
FROM purchase
GROUP BY 2
ORDER BY 3 DESC;
```

style	product_id	Total Purchased	price	Total Sales
Men's Styles	3	63	150	9450
Women's Styles	10	62	95	5890
Women's Styles	9	54	95	5130
Men's Styles	1	52	95	4940
Women's Styles	6	50	95	4750
Men's Styles	4	44	150	6600
Women's Styles	7	44	150	6600
Men's Styles	2	43	95	4085
Women's Styles	8	42	150	6300
Men's Styles	5	41	50	2050

style	product_id	Total Purchased	price	Total Sales	% Purchased	% Sales
Men's Styles	3	63	$150.00	$ 9,450.00	13%	17%
Women's Styles	10	62	$ 95.00	$ 5,890.00	13%	11%
Women's Styles	9	54	$ 95.00	$ 5,130.00	11%	9%
Men's Styles	1	52	$ 95.00	$ 4,940.00	11%	9%
Women's Styles	6	50	$ 95.00	$ 4,750.00	10%	9%
Men's Styles	4	44	$150.00	$ 6,600.00	9%	12%
Women's Styles	7	44	$150.00	$ 6,600.00	9%	12%
Men's Styles	2	43	$ 95.00	$ 4,085.00	9%	7%
Women's Styles	8	42	$150.00	$ 6,300.00	8%	11%
Men's Styles	5	41	$ 50.00	$ 2,050.00	8%	4%
	10	495		$55,795.00	100%	100%

6.6 Purchase Table – Queries & Analysis

Analysis:
This SQL gives us the total purchases and totals sales by price point. We see there are 3 different prices $50, $95 & $150. We can see that the $95 price point is the most common while the $50 has very low sales in comparison.

```
SELECT price, COUNT(*) 'AS Total Purchased', SUM(price) AS
'Total Sales'
FROM purchase
GROUP BY 1
ORDER BY 1;
```

price	AS Total Purchased	Total Sales
50	41	2050
95	261	24795
150	193	28950

7: Capstone Project Review

7.1 Learn SQL from Scratch-Recap

Now that we have looked at all of the data provided by Warby Parker, let's recap what we have discovered.

- 5 Questions on the survey; responses decline as the users progresses through the survey.
- 1000 users responded to the survey
- 750 users did the Home Try On – 75% of survey responders
- Of the 750 users who did the home try on, 66% made a purchase
- 495 purchases were made – 49% of survey responders
- Based on the data provided, each users only purchased one item

Quiz Table:
- we find that the most popular response for fit was Narrow, Shape was Rectangular and Color was Tortoise
- For the Men's Styles, we can see that 23 users indicated narrow, rectangular, tortoise glasses.
- For the Women's Styles, we can see that 20 users indicated narrow, rectangular black glasses.
- For the users that just weren't sure, we can see that 8 users indicated narrow, square, black glasses.
- For Fit, we can see that 174 Men, 189 Women & 45 Unsure/Unisex are looking for Narrow glasses
- For Shape, we can see that 176 Men, 184 Women & 37 Unsure/Unisex are looking for Rectangular glasses
- For Color, we can see that 128 Men, 142 Women & 22 Unsure/Unisex are looking for Tortoise glasses
- For Women Styles, the most popular type of glasses are Narrow/Rectangular (74 responses) while the lowest number would be for Medium/Round (14 responses)
- For Men's Styles, the most popular type of glasses are Narrow/Rectangular (72 responses) while the lowest number would be for Wide/Round (18 responses)

Purchase Table:
- We see the average selling price is $112.72, the Min Price is $50 and Max Price is $150.
- Total Sales is 495 items for $55,795.00 (Men's – 243 - $27,125.00 & Women's – 252 - $27,125.00)
- There were more Women's items sold but the revenue is higher for Men's purchases
- The most common/popular color is Jet Black.
- The most common is Eugene Narrow
- For Men's Styles, the most common is product_id = 3 with 63 purchases. The least common is product_id = 5 with 41 purchases.
- For Women's Styles, the most common is product_id = 10 with 62 purchases. The least common is product_id = 8 with 42 purchases.
- From the data, we can see that product_id =3 is the top seller overall while product_id = 5 is the lowest seller overall.

8: Learn SQL from Scratch-Recap

8.1 Learn SQL from Scratch-Recap

Here is a recap of what I learned in this class:
- How to create a select query
- How to limit the number of records returned in the results
- How to aggregate (Sum, Count, Average, Min, Max, Group, Having)
- How to order data (sort)
- How to do data funnels
- How to performs Joins (Inner, Left, Cross, Union)
- How to calculate churn rates – Not used in this presentation
- How to find first and last attribution – not used in this presentation
- And more….

Thank you for taking the time to review my SQL and findings. Should you have any questions, please let me know. Have a wonderful day!!

9: Appendix A: Capstone Project Questions

9.1 Appendix A – Questions

Quiz Funnel:

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?

2. 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?

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!

Home Try-On Funnel:

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?

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

user_id	is_home_try_on	number_of_pairs	is_purchase
4e8118dc	True	3	False
291f1cca	True	5	False
75122300	False	NULL	False

   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).

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?

7. Great work! Now take all your queries and findings from this data and put together your presentation.

   Make sure it is readable and visually appealing so that the reviewer will understand!