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.

In this project, we will analyze different Warby Parker's marketing funnels in order to calculate conversion rates. Here are the funnels and the tables given:

Quiz Funnel:

- `survey`

Home Try-On Funnel:

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

This project was a collaboration with Warby Parker's Data Science team (thank you!) and uses fictional data.

Let's get started!

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?

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

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?

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

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

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?

We divide the number of people completing each step by the number of people completing the previous step:

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?

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

```sql
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

5.

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

User_id || is_home_try_on||

user_id	is_home_try_on	Number of pairs	is_purchase

01fdf106-f73c-4d3f-a036-2f3e2ab1ce06	true	3	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 (quiz) 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 p.user_id = q.user_id
LIMIT 10;
```

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

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?

```
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 p.user_id = q.user_id)
SELECT COUNT(*),
SUM(is_home_try_on) AS 'num_home_try',
SUM(is_purchase) AS 'num_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_to_purchase'
 FROM funnels;
```

Query Results

COUNT(*)	num_home_try	num_purchase	quiz_to_home_try_on	home_try_to_purchase
1000	750	495	0.75	0.66

```
SELECT
 COUNT(DISTINCT CASE
  WHEN number_of_pairs = '5 pairs' THEN user_id

  END) AS '5 pair home try',
   COUNT(DISTINCT CASE
  WHEN number_of_pairs = '3 pairs' THEN user_id
  END) AS '3 pair home try'
FROM home_try_on;
```

5 pair home try	3 pair home try
371	379

The sqlite code for details of the funnels:

```
WITH q AS(
  SELECT '1-quiz' AS stage ,COUNT(DISTINCT user_id)
FROM quiz
),
h AS (
 SELECT '2-home try on' AS satge,COUNT(DISTINCT user_id)
 FROM home_try_on
),
p AS (
 SELECT '3-purchase' AS satge,COUNT(DISTINCT user_id)
 FROM purchase
)
SELECT * FROM q
```

```
UNION ALL
SELECT * FROM h
UNION ALL
SELECT * FROM p;
```

And the funnel looks like

stage	COUNT(DISTINCT user_id)
1-quiz	1000
2-home try on	750
3-purchase	495

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

Code is updated as conversion rates.sqlite in github handle.