

👍 **LiveLab** | Terracotta Survey Analysis

INTRODUCTION: Eben, the owner of a local plant shop called Terracotta, surveyed former, current, and even prospective customers who visited his shop to better understand their buying patterns and preferences. He's hired you to analyze the survey data and provide data-driven recommendations to make sure he really nails a pivot from a brick and mortar storefront to an online-only plant shop.

HOW IT WORKS: Follow the prompts in the questions below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write and **blue boxes** for text-based answers. No need to submit this document to your LiveLab Team – it's just for you!

SQL App: [Here's that link](#) to our specialized SQL app, where you'll write your SQL queries and interact with the data.

– Data Set **Description**

plant_types – Survey responses to the question “What type of plants do you typically purchase?” Values are 'other', 'terrariums', 'large or small trees', 'indoor plants', 'low-maintenance plants', 'tropical plants', 'succulents or cacti', 'aquatic plants', 'outdoor plants', and 'herbs or spices'.

reason_for_purchase_words – Survey responses to the question “What factor is most important to you when choosing a plant to buy?” This was a free-response question.

reason_for_purchase – Categories based on common themes in `reason_for_purchase_words`. Values are 'care requirements', 'benefits', 'pet-friendly or safe for children', 'size', 'rare or unique species', and 'price'.

freq – Survey responses to the question “How often do you typically purchase plants?” Values are 'weekly', 'monthly', 'every few months', 'once a year', and 'rarely, only for special occasions'

care_info – Categories based on common themes in survey responses to the question “What plant care information would you like to see provided in-store or online?” Values are 'pet-safe information', 'light requirements', 'watering frequency & amount', 'propagation techniques', and 'soil type & fertilization'

pots – Survey responses to the question “What kind of pots or planters do you prefer for your plants?” Values are 'ceramic', 'terracotta', 'plastic', 'metal', 'hanging baskets', and 'wall-mounted planters'

– Task 1: Frequent Purchases

Eben asked us to uncover which types of plants are most frequently purchased by customers. We'll query the survey data to find out the answer. Let's dive in!

- A.** To warm up, Write a query that returns all the columns in the dataset (`terracotta.survey`). Copy and paste your query into the box below.

```
SELECT
  *
FROM
  terracotta.survey;
```

- B.** While you can certainly peek at the information just above the output from your previous query, write a new query using the `COUNT` function that returns the exact number of rows in the survey table. What is the exact number of rows in the survey table?

```
SELECT
  COUNT(*)
FROM
  terracotta.survey;
```

- C. Next, write a query that groups the survey responses by the question “What type of plants do you typically purchase?” and counts the number of responses for each type of plant listed.

```
SELECT
    plant_types,
    COUNT(*)
FROM
    terracotta.survey
GROUP BY
    plant_types;
```

- D. Modify your query to order the number of survey responses for each type of plant by largest to smallest. What type of plants are most frequently purchased by customers?

```
SELECT
    plant_types,
    COUNT(*)
FROM
    terracotta.survey
GROUP BY
    plant_types
ORDER BY
    count DESC;
```

Low-maintenance plants — 51

- E. What percentage of all survey responses list low-maintenance plants as the type of plant they most frequently purchase?

NOTE: Divide what you get from part **D.** with what you found in part **B.** – you won't calculate this proportion in a query.

23.39%

– Task 2: Reasons for Purchase

Eben is really eager to understand what's driving plant moms and plant dads to decide which ready-to-nurture plants they want to take home. Lucky for us, a more senior analyst already assigned a category to each and every free response from the survey. The `reason_for_purchase` column will be key for this Task.

- A. Write a query that groups the survey responses by the `reason_for_purchase` categories and counts the number of responses in each category. Sort the number of responses in each category from biggest to smallest.

```
SELECT
    reason_for_purchase,
    COUNT(*)
FROM
    terracotta.survey
GROUP BY
    reason_for_purchase
```

```
ORDER BY
  count DESC;
```

- B.** Add another expression to your query to answer the following question: How many categories had at least **30** responses?

```
SELECT
  reason_for_purchase,
  COUNT(*)
FROM
  terracotta.survey
GROUP BY
  reason_for_purchase
HAVING
  COUNT(*) ≥ 30
ORDER BY
  count DESC;
```

3 categories

- C.** What percentage of all survey responses list care requirements (e.g. maintenance, light requirements, watering frequency, etc.) as the most important factor in choosing a plant to buy?

NOTE: Divide what you get from part **B.** with what you found in Task 1, part **B.**
– you won't calculate this proportion in a query.

26.61%

- D. What percentage of all survey responses list whether or not a plant is pet-friendly or safe for children as the most important factor in choosing what to buy?

NOTE: Same as before, just divide what you get from part **B.** with what you found in Task 1, part **B.** – you won't calculate this proportion in a query.

26.15%

– Task 3: A Recommendation for Terracotta

Let's get back to Eben's concern that motivated this analysis. How might Eben use what you found as part of a potential marketing or advertising strategy? Are there ways that you can use what you learned to drive success in online plant sales?

Together with your team, craft a data-driven recommendation to Eben.

He can use his car requirements and his plants' pet friendliness / safety for children as a way to market his plants

– LevelUp: Free Response Data

Categorizing free-response data is a common task for a data analyst. As part of their work, data analysts often deal with unstructured data, such as free-response survey data, and need to find ways to organize and analyze it effectively. But categorizing free-response data can be challenging, since language and terminology used by respondents can vary widely.

Consider the `reason_for_purchase_free_response` column. The response "I have a dog, so I always check if the plant is pet-friendly." and "I love my cat too much to take any chances, so I opt for non-toxic plants." should both be categorized as **pet-friendly**. But only one of these responses includes the subphrase "pet-friendly".

Similarly, "I want to enjoy my plants without worrying about my toddler getting into them." and "I prefer plants that are safe for children, even if they're a bit more expensive." should both be categorized as **safe for children**. But only one of these includes the subphrase "safe for children".

Can you write a SQL query that returns all free response data that should be categorized as pet-friendly or safe for children? Your query should use the `LIKE` or `ILIKE` keywords and `AND` or `OR` conditions to search for different phrases related to pet-friendly and safe-for-children properties. Feel free to compare your categorization results with `reason_for_purchase` column!

```
SELECT
    reason_for_purchase_words,
    reason_for_purchase
FROM
    terracotta.survey
WHERE
    reason_for_purchase_words ILIKE '%dog%'
    OR
    reason_for_purchase_words ILIKE '%cat%'
    OR
    reason_for_purchase_words ILIKE '%children%'
    OR
    reason_for_purchase_words ILIKE '%mom%'
    OR
    reason_for_purchase_words ILIKE '%toxic%'
    OR
    reason_for_purchase_words ILIKE '%safe%';
```

