Please answer the following questions with a SQL Query:

1. How Many Dogs Were Successfully Screened? (Screened defined as having a valid date in form_completion_date).

There are 12 dogs successfully screened.

2 A. How Many Dogs Are Eligible? (Eligible defined as having a value of 1 for all elig_question columns and having any value for the inelig_question columns aside from 1 (must not be missing))

There are 4 dogs that meet the criteria for eligibility.

2 B. If missing values could be considered for the ineligibility questions how would you rewrite this query?

This query would have been rewritten to include ineligibility questions marked NULL values, which would increase the total to 7 dogs.

3. How Many Dogs Are Ready To Be Enrolled In The Study? (Ready To Be Enrolled being defined as being Eligible and not having a value of 1 or 2 under withdrawal_status)

There are only 4 dogs which are ready to be enrolled in the study, based on their eligibility while not having a value of 1 or 2 in the withdrawal status field.

4. How Many Dogs Were Withdrawn From the Study? (Withdrawn being defined as having a value of 1 or 2 under withdrawal_status)

There are only 2 dogs who were withdrawn from the study based on their values under the withdrawal status field.

5. How many screened dogs are over the age of 5?

There are 9 dogs over the age of 5.

6. Write a query that designates the order in which those ready to enroll should be contacted (Those with the oldest form_completion_dates should be contacted first)

The query needed to satisfy this requirement concludes that the 4 dogs ready to be enrolled should be contacted in the following order:

id	*form_completion_date*	*Contact_order*
58	*2025-08-10*	*1*
4	*2025-08-25*	*2*
5	*2025-09-03*	*3*
40	*2025-09-15*	*4*

Tableau Assessment

1. Create an additional visual in tableau you feel would be helpful for those managing the study. Please describe why you feel the visual would be helpful.

In addition to providing metrics showing how many dogs were screened, eligible and ready for enrollment, as well as those who were withdrawn, visuals are created indicating the total number of dogs in each weight category (less than 25 LBS, or greater than 25 LBS). An additional visual is provided, consistent with the SQL query scenario of dogs whose age is greater than 5 years. This is useful in that there may be differential study outcomes based on dog age, and having a metric that tracks this provides a clearer picture of the demographics of those dogs in the study.

Furthermore, I have used the logic from the SQL queries to derive counts for those dogs who are either withdrawn, participating in the study, but not screened, screened but not eligible for participation, and those who are ready to be enrolled. This provides useful categorization for understanding where participants are at in the process, as well as identifying potential points of the enrollment and screening process that are weeding participants out, leading to increased efficiency in study planning. The pie chart shows a proportionate view for each status established by these constructed categories.

These categories were then mapped to the weight and age categories mentioned above, as well as to the number of study participants registered per hospital. Again, this provides important contextual information, not just in aggregate totals, but where the preponderance of study participants are within the screening and enrollment process. For example, we see Hospital A contains more study participants than all other hospitals together, with the majority of those being screened, but not eligible. This is helpful because the U of U engages with study participants/patients from a network of hospitals, as one example, and creating discrete categories shows which hospitals contribute most participants, helps identify outlier clinics and provides insight into the social makeup of the study by showing where they are most likely to go for medical assistance.

2. Are there any other data points you would recommend that the study track? If so, Why?

One other data point that might be useful is diet composition. While we know breed, weight and age, knowing the primary food source could be helpful in evaluating the overall health of the dog.
Additionally, I would look to increase focus on the null, or missing values, such as those identified more thoroughly in the SQL question 2 part 2. By evaluating missing data more closely, we can better identify where study participants fail to complete the enrollment and screening process. This leads to insight into study design, highlighting ways in which the process can be refined to ensure that there are properly representative study samples constructed.