INSTRUCTIONS

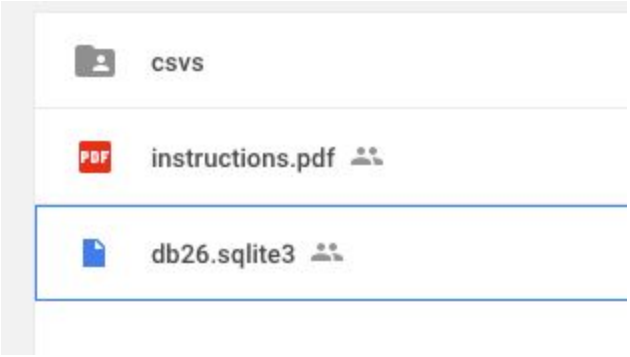
This set of exercises is meant to be done using SQL or Excel. If you have previous SQL knowledge, the simplest way to complete these exercises is to use an interface that can connect to a SQLite database (such as SQLite Studio). It is also possible to do these exercises in Excel, though they are optimized for SQL.

SQLite is a software library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. If you have experience with MySQL, PostgreSQL, Redshift, or other SQL variants, you should be familiar enough to write queries in SQLite. For information on commands and keywords for SQLite, see the following links:

- SQLite commands
- SQLite keywords

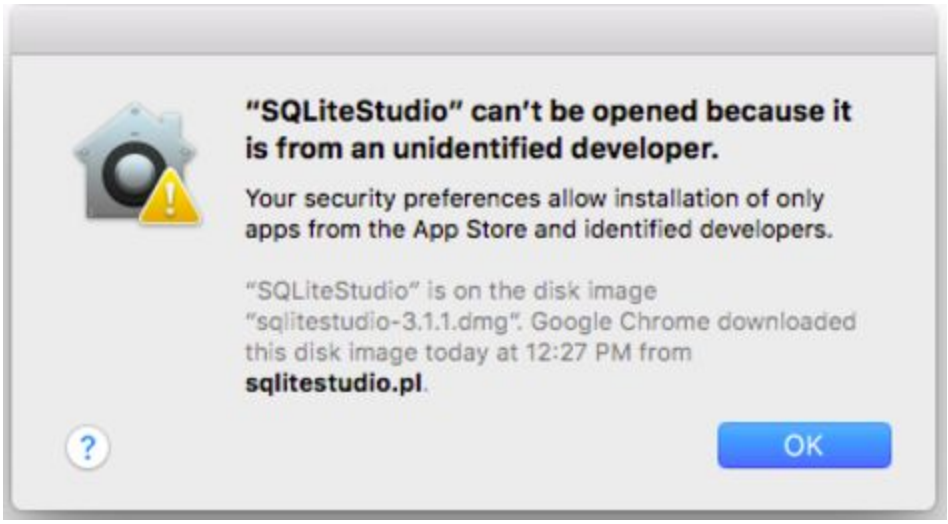
SQLite Studio Instructions

First, please make sure that you have downloaded the following SQL database file on to your hard drive:



Go to the following link and download the installer for SQLite Studio:

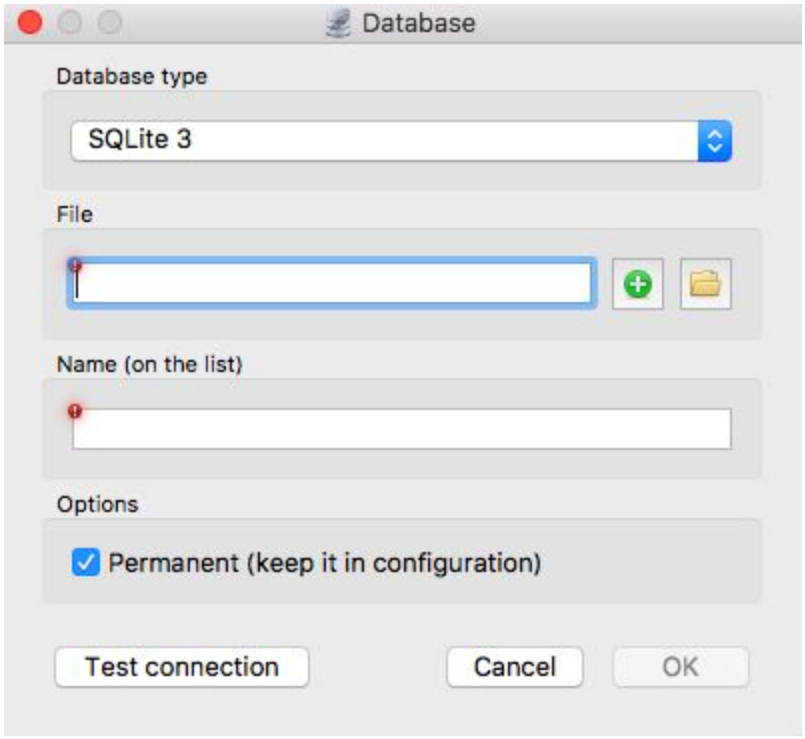
You may encounter this security message:



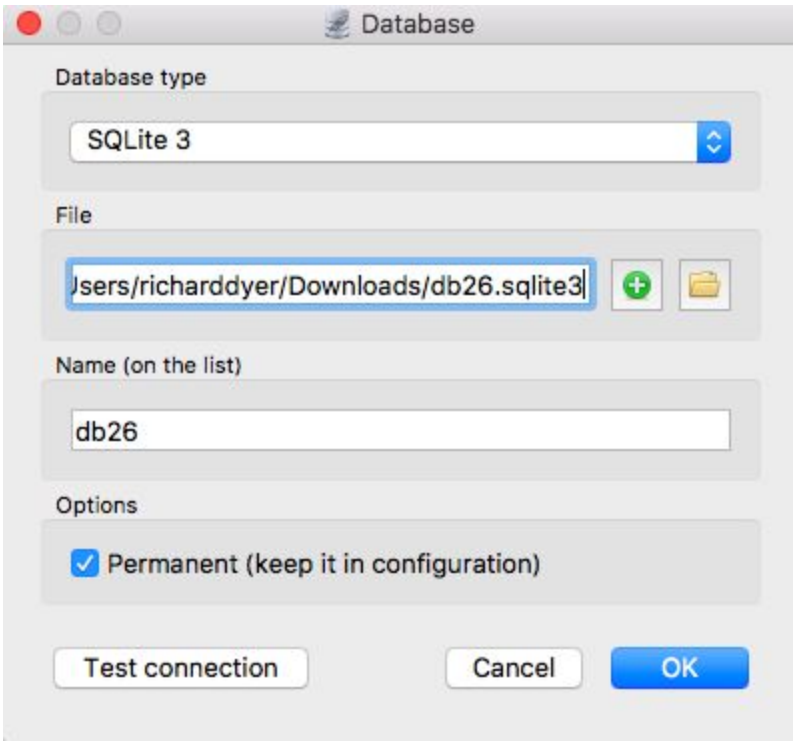
If you do get this message, go to System Preferences > Security and Privacy > and you should see the following screen:

Once SQLite Studio has been installed, you will need to connect to the database that you downloaded earlier. Once you have SQLite Studio open, click on the Add a Database button:

Select the SQLite 3 database type, then navigate to find the database that you downloaded.

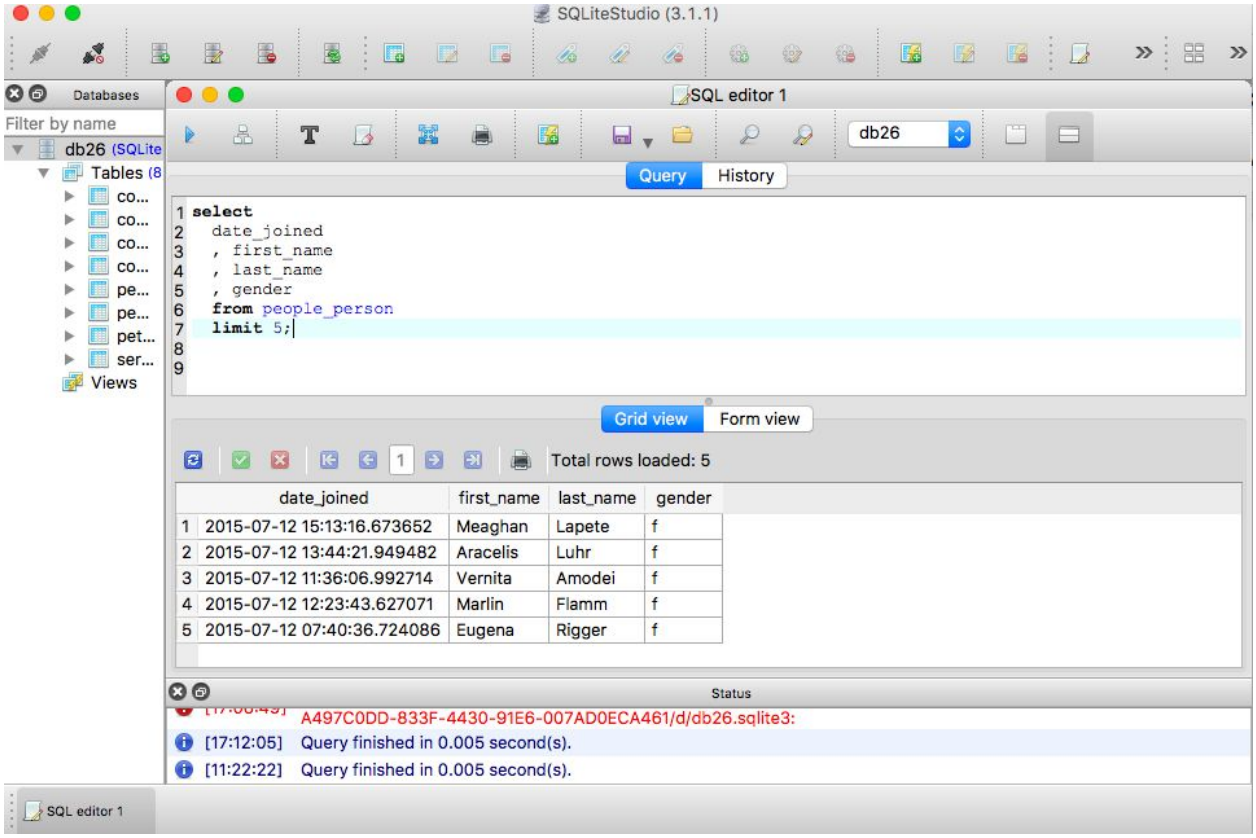


Once you have navigated to select the database, hit “OK” and you should now be connected:

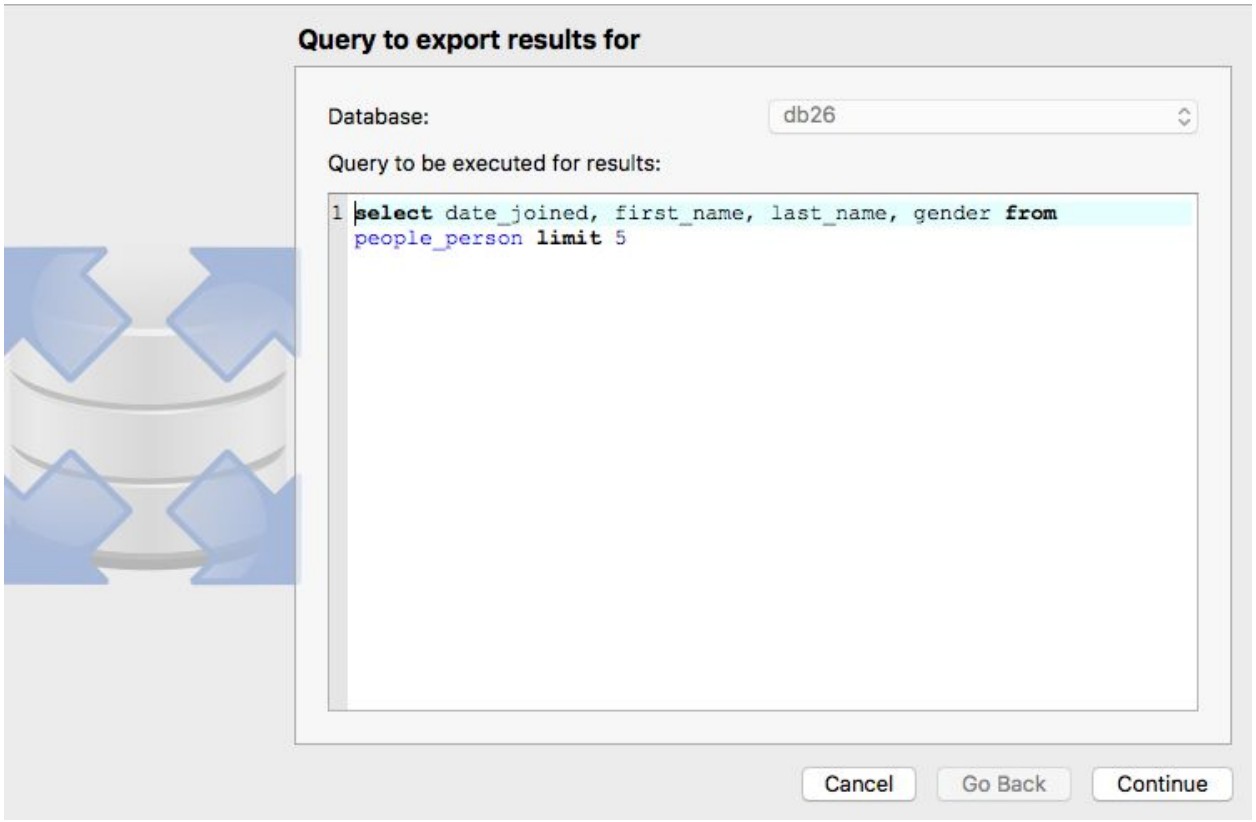


To query the database, click on the Open SQL Editor button:

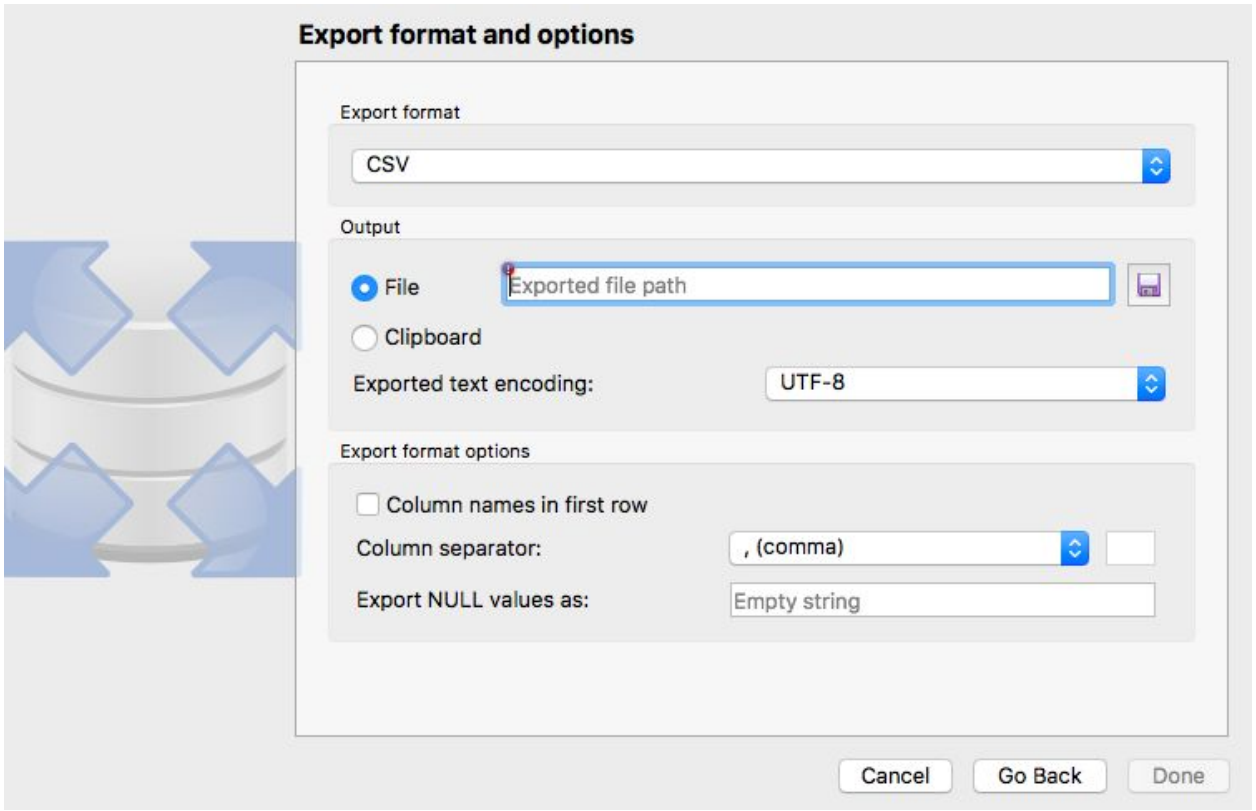
Simply type your query and click on the Execute Query button:



Lastly, you can export the results of a query by clicking on the Export button and following the wizard:



Select a file location to export your results to:



Excel Instructions

The databases are also included in CSV format to manipulate in Excel. If you choose this route, it will be easiest to create an Excel workbook with all the CSVs (tables) stored as tabs.

EXERCISES

For this exercise, imagine that Rover has acquired a small pet care start-up. As an analyst, you have been tasked with the responsibility of exploring their database. We have shared a SQLite database with you containing this (fabricated) data. The file is named db26.sqlite3 (see instructions above). Once you are comfortable connecting to the database with SQLite Studio or you have downloaded the CSVs in Excel, please proceed to the exercises. ***If you have trouble connecting to the database, please reach out with questions to our team, and we will be more than happy to help.***

You may submit your report in any format you choose; for example, you might submit a Google Doc, a PowerPoint presentation, an Excel workbook, etc. Assume your results will be shared with analysts and partners from marketing and operations. The latter will have a degree of analytical sophistication (i.e., they know some SQL and basic stats) but may need guidance in interpreting raw results or advanced techniques. Use this exercise as an opportunity to show off your communication skills and style. ***Please also ensure that you submit your underlying work.***

Lastly, we understand that everyone's circumstances are different so there is no hard deadline; please complete the exercise with whatever timing works best for you. That said, we have one ask: we constantly collect data on this exercise so we can calibrate it and tune it going forward, so if you are willing to share how much total time you put into it, we would be appreciative.

If needed, there is an appendix that details the tables in this database.

I. Exploring the Database

We begin by asking a few basic questions about the users of this platform. This first exercise is presented with answers so that you can diagnose any issues you might have connecting to or working with the data.

1. How many users have signed up?
   The answer is 64416.
2. How many users signed up prior to 2017-01-12 ?
   The answer is 35500.
3. What percentage of users have added pets?
   The answer is 80.44%.
4. Of those users, how many pets have they added on average?
   The answer is 1.496.
5. What percentage of pets play well with cats?
   The answer is 24.78%.

II. Conversations and Bookings

Some users can offer pet care services. When an owner needs pet care, they can create a conversation with another user that offers the service they are interested in. After exchanging some messages and possibly meeting in person, that conversation hopefully books. In that case, services are paid for and delivered. Occasionally, some conversations that have booked may be cancelled. Lastly, for un-cancelled bookings, both owners and sitters have the option of leaving a review. **In the following questions, we explore these concepts.**

1. What are the possible services and what is the average price per unit for each service type?
2. How many requests have there been for each service type and what percentage of those have booked? The percentage of those that have booked is called booking rate, i.e. what is the booking rate by service?
3. What are the cancellation rates for each service?
4. For un-cancelled bookings, is the owner or provider more likely to leave a review and which tends to leave better reviews? How would you narrate this finding to a business partner? Feel free to use whatever “proof” you need to share these findings (visualizations, graphs, talking points, etc.).

III. Recent Daily Booking Rate

The snapshot of this database was taken on 2017-07-11 at midnight and only contains data reflecting events prior to that date. A junior analyst is investigating daily booking rate during the days prior to the snapshot and is concerned about an apparent downward trend. You are tasked with helping them out.

1. First, let's reproduce their results. What is the daily booking rate (defined in Question II.2) for each of the 90 days prior to the snapshot? Is there a downward trend?
2. Can you narrate a reason why this trend exists? Is there a reason to be concerned?

IV. Analyzing Take Rate

In order to do the next exercise, you will need to understand the fee structure for this company. Each user has a fee associated with their account (recorded on people_person). If that user books as an owner, the company charges a service fee (in addition to the booking total) that is a percentage of the booking total (to a maximum of $50). Also, each service has a fee amount (recorded on services_service).

Before a provider receives their payment, the company takes a percentage of the booking total as dictated by that fee. As an example, suppose an owner has a fee amount of 5% and books with a service that has a fee amount of 15%. If the booking was for $100, then the owner would get charged $105 (adding the owner’s fee). The $5 owner fee would go to the company. An

additional $15 would also go to the company since the service had a 15% fee associated to it. The remaining $85 would go to the provider. To summarize:

Amount	***Description***
Booking Total	$100 (e.g. 4 walks at $25/each)
Owner Fee	$5 (5% of the booking total)
Gross Billings	$105 charged to the owner
Service Fee	$15 (15% of the booking total)
Net Revenue	$20 (owner fee + service fee)
Provider Payment	$85 earnings for the provider

1. In each month, what were the gross billings and net revenue? (Note: month is determined by when the request actually booked, i.e. booked_at).
2. Define take rate to be the percentage of gross billings that is net revenue. In the previous example, the take rate is slightly more than 19% since $20/$105 is approximately 0.1905. In each month, what was the aggregate take rate?
3. Did take rate trend up or trend down or remain unchanged over time?
4. If it did change, investigate why and provide an explanation. Be sure to provide additional data/charts/evidence that justify your explanation. Any claims should be backed by data.

V. New Conversation Flow

Internal documents indicate that this recently acquired company was performing many A/B tests; we would like to investigate one. This platform had a conversation page where owners and service providers could exchange messages as they organized their booking. The team thought this page could use a re-design and set out to improve its UI. A product manager then set up a test to measure the new page's effectiveness. On 2017-03-13 , an A/B test was launched. For those owners who sent a request, they would be randomly assigned to variant or holdout groups. Those users who are in the variant group would see the new conversation flow. However, those in the holdout group would see the old conversation flow. Providers would always see the old conversation flow.

1. Did conversations with the new conversation page book at a higher rate?
2. Is it statistically significant?
3. What would you recommend as next steps?

APPENDIX: TABLE OVERVIEWS

People_person

This table details each user on our site. This table may contain dog owners, dog sitters, or people who have not transacted on our site. Many of the fields on this table are self explanatory but we have detailed a few below.

- channel - This field reports how this user discovered our site when they signed up.
- date_joined - The timestamp for when this user signed up.
- fee - When a user books a service as a dog owner, we charge the owner a separate service fee that takes the form of a percentage of the booking total.

People_testsegmentation

Occasionally, this company would run an A/B test which required that users get placed in two groups. This table provides a log for experiments which require user-level segmentations. Many of the fields on this table are self explanatory but we have detailed a few below.

- person_id - This foreign key reports the people_person record that was segmented.
- test_name - Multiple tests were run on this site and all are logged on this table. Use this column to filter to the correct experiment.
- test_group - For the purposes of the experiment in test_name , the user given by person_id was segmented into the group named in this column (e.g., holdout , variant , A , B , etc.).
- added - The timestamp reporting the time when this user was segmented.

Pets_pet

This table details each pet that a user has added to their profile. One owner may have more than one pet, but not vice versa. Many of the fields on this table are self explanatory but we have detailed a few below.

- description - A short (lorem ipsum) description of the pet.
- plays_cats - If 1, then this pet plays well with cats.
- plays_children - If 1, then this pet plays well with children.
- plays_dogs - If 1, then this pet plays well with dogs.
- spayed_neutered - If 1, then this pet has been spayed or neutered.
- house_trained - If 1, then this pet is house trained.
- owner_id - This foreign key reports the people_person record for this pet’s owner.

Services_service

On our site, users may offer pet care services. This table stores a record for each service that is offered. Each user can offer more than one service, but not more than one of each type. Many of the fields on this table are self explanatory but we have detailed a few below.

- max_dogs - This number is the maximum number of pets this provider would prefer to care for.
- fee - When a user books with a service, we take a percentage of the booking total. This field reports the percentage.
- provider_id - This foreign key reports the people_person record for this service's provider.
- added - A timestamp for when this service became active.
- price - The price per unit booked.

Conversations_conversation

An owner can book a service provider by starting a conversation with them. This table stores a record for each conversation started on our platform. Many of the fields on this table are self explanatory but we have detailed a few below.

- start_date - This is the date for which pet care will first be needed.
- end_date - This is the last date for which pet care will be needed.
- units - This is the number of units of service that the owner is interested in booking.
- added - A timestamp for when this conversation was created.
- booking_total - This is the dollar amount (not including the owner's service fee) that this booking would cost.
- requester_id - This foreign key reports the people_person record for the pet owner that is requesting pet care.
- service_id - This foreign key reports the services_service record for the service that the pet owner is requesting.
- booked_at - If the request is booked, this timestamp reports when that occurred.
- cancelled_at - A booked request can be cancelled. In that case, this timestamp reports when that occurred.

Conversations_conversation_pets

Since a booking may involve many pets and many pets might have had many bookings, it is necessary to store this many-to-many relationship on a separate table. Many of the fields on this table are self explanatory but we have detailed a few below.

- conversation_id - A foreign key to a booking request on the conversations_converation table. If this conversation involves caring for more than one pets, then this conversation_id will occur on more than one row on this table (once for each pet).
- pet_id - A foreign key to a pet that will receive pet care during the corresponding conversation's booking.

Conversations_message

Each conversation consists of a series of messages. A conversation may contain many messages, but not vice versa. Many of the fields on this table are self explanatory but we have detailed a few below.

- conversation_id - This foreign key reports the conversation in conversations_conversation for which this message is apart of.
- sender_id - This foreign key reports the user in people_person that sent this message.

Conversations_review

If a booking occurs, then either participant can leave a review for the experience. This table records those reviews, which consist of a brief statement and a star rating. Many of the fields on this table are self explanatory but we have detailed a few below.

- conversation_id - This foreign key reports the booking in conversations_conversation for which this review pertains.
- reviewer_id - This foreign key reports the user in people_person that wrote this review.