Using Mockaroo.com to generate Data

Introduction

There is a neat website, www.mockaroo.com, which provides a way for us to generate fake data which can then be used for writing some sample queries. For week 5, you can either use your discussion project and generate some data or create three new tables with data.

Mockaroo will write all of the create table and insert statements needed to create your sample tables. In this tutorial, we will generate three tables and will also show you how to do this in a way that makes the tables link together. In other words, if one table references item_no in another table, that reference will be valid.

Sample Dataset

Let's create a dataset for ordering a single item. Here are the tables:

Customer	Order	Product
CustID (primary key)	OrderID (primary key)	ProductID (primary key)
Name	OrderDate	Description
Street	ProductID (ref product)	Price
City	CustID (ref customer)	
State		
Zipcode		
Telephone		
AvailableCredit		

Notice that we do have some foreign key constraints. For example, CustID in the order table references CustID in the customer table. We will use Mockaroo to make this work.

Mockaroo

The best thing about mockaroo is that it is free. Go to www.mockaroo.com and create a free account. You can purchase a plan as well, but for our purposes being able to generate 1000 rows of data for each table will be sufficient.

Once you have created your account, you can create a schema to hold all of your information. Click on the Schemas area next to mockaroo, as shown in Figure 1.

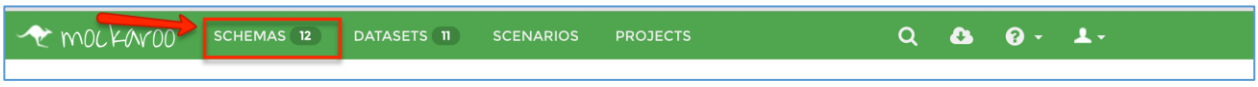


Figure 1

This will bring you to a screen similar to Figure 2 where you can create the first schema.

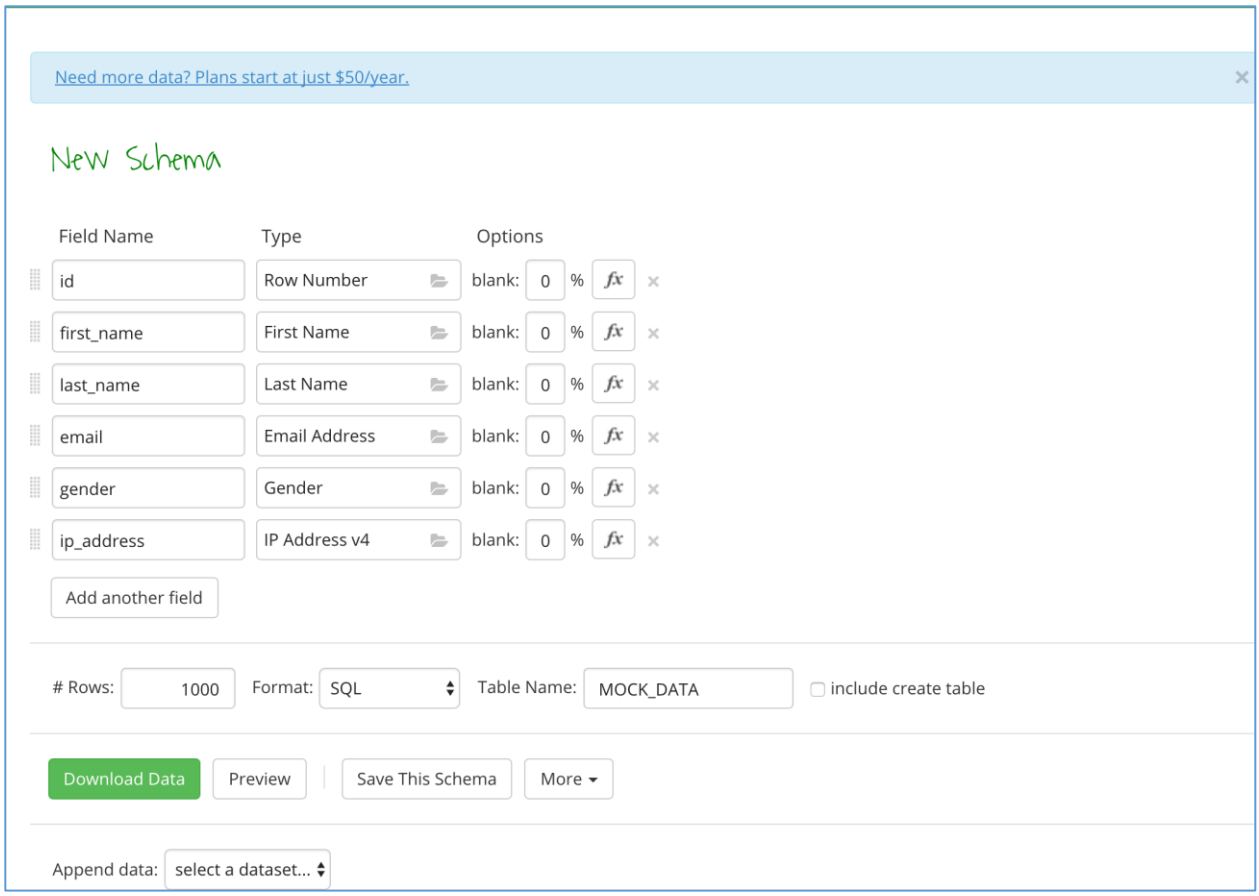


Figure 2

So, let's go ahead and create the customer schema. First, let's add the custid

Figure 3

Notice that we chose a sequence starting at 1000. This will give us something like an auto increment. Now, when I clicked on type, I received a whole list of types as shown in Figure 4.

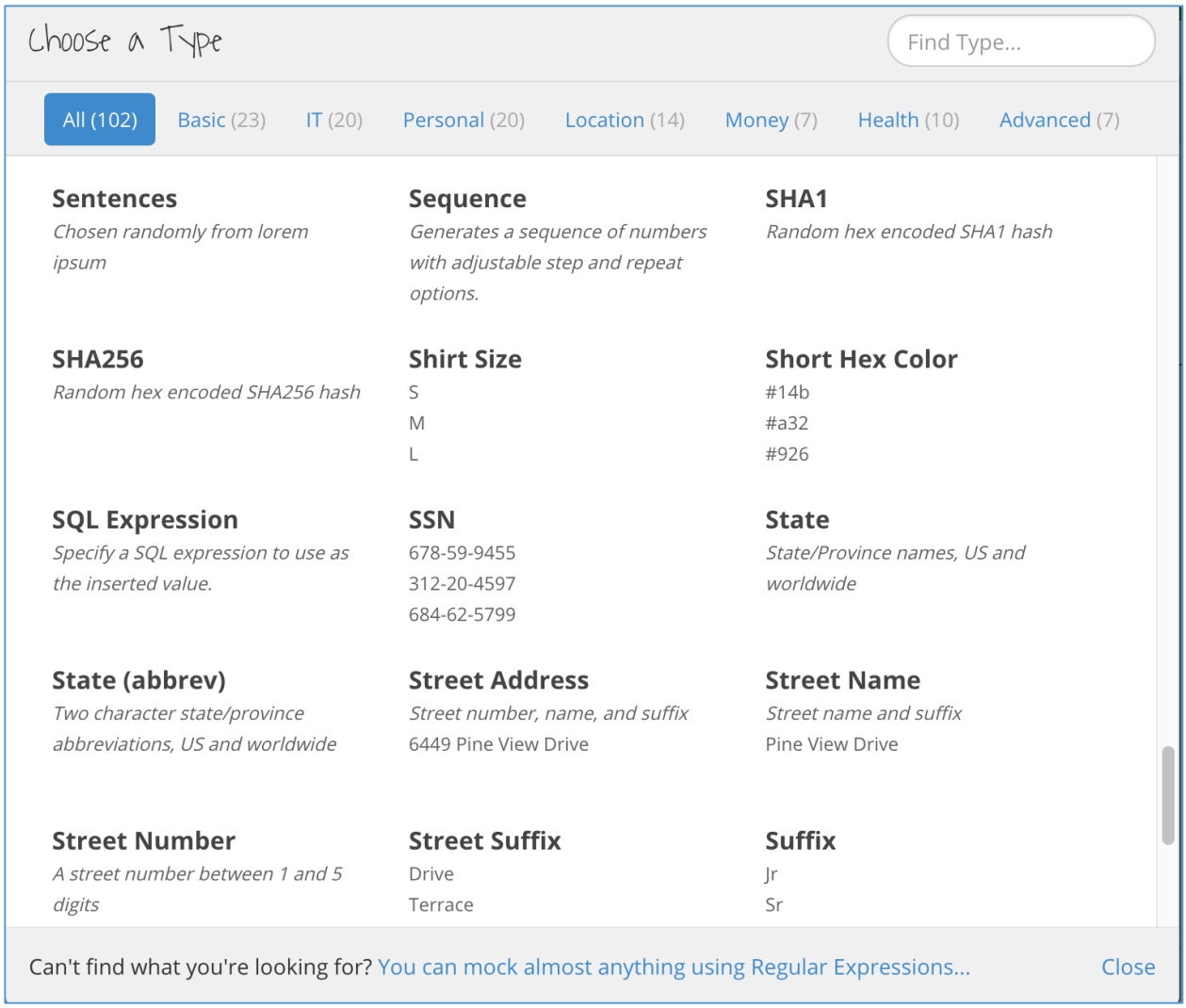


Figure 4

This is how we will get all of our types. Take a look around and see what types you might want to use.
Let's fill in the rest (see Figure 5.)

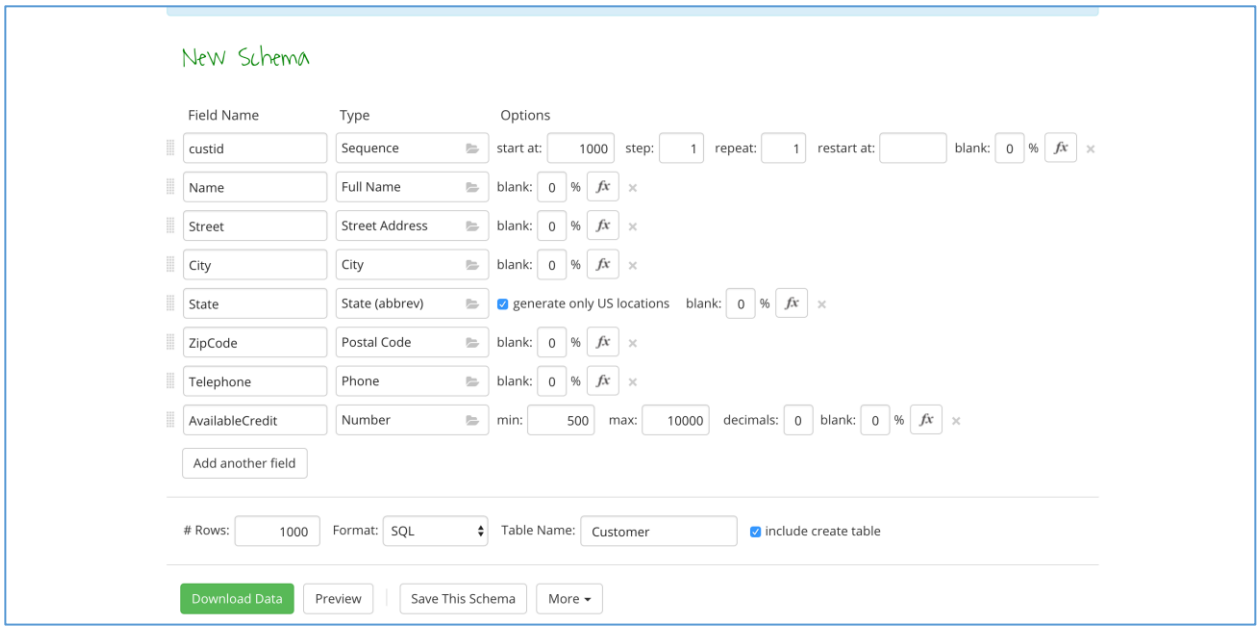


Figure 5

In order to get the references to work, we have to do a few tricks. After creating the schema and saving it, you want to download a CSV file with a header. We will upload this as a dataset shortly. Figure 6 shows where to change this:

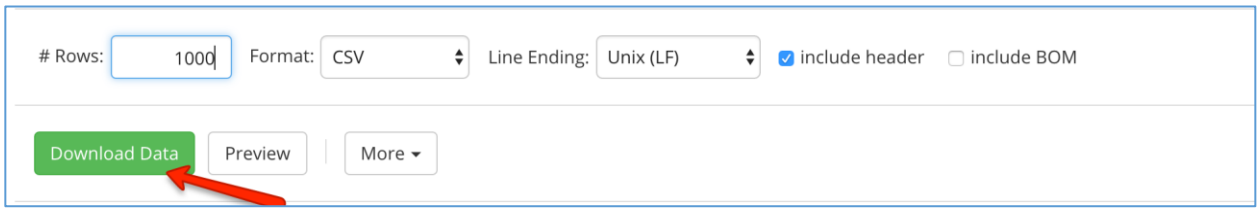
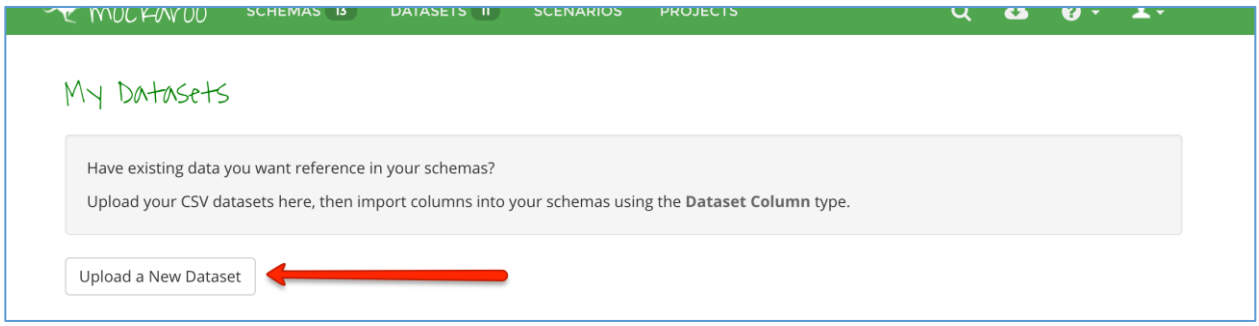


Figure 6

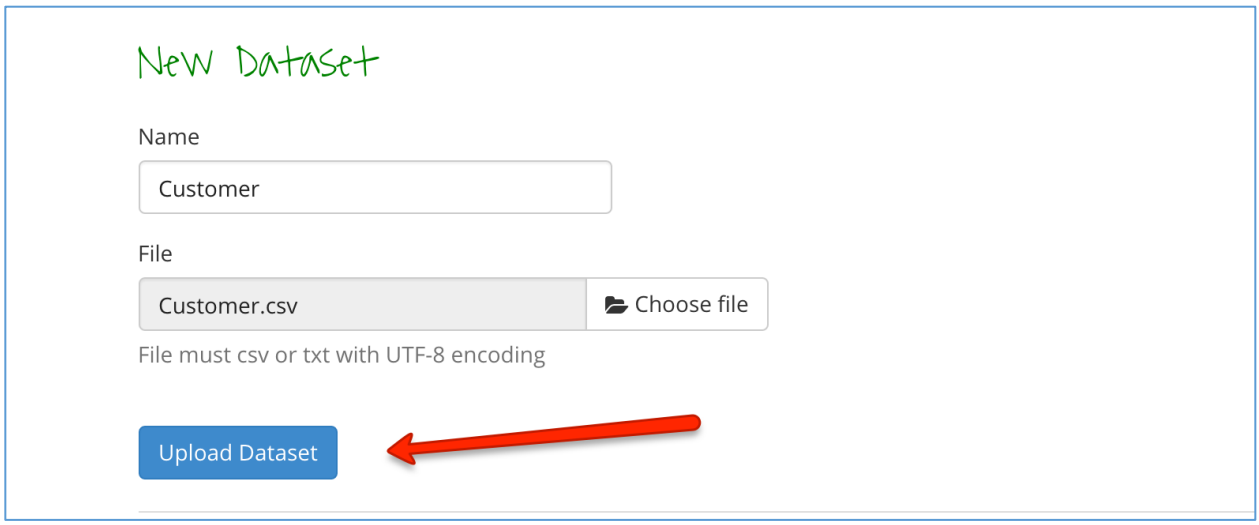
Once you click the download data button, then you should save the generated CSV file on your hard drive. Once this is done, then you will need to upload this as a dataset. To do this, choose dataset from the top list as shown in Figure 7:

Figure 7

Once on this screen, click on Upload a New Dataset:



Choose the csv file that you just created and name it the same as your table.



Once the dataset has been uploaded, switch back to the schema that you created (in my case customer.) I'm now going to generate an SQL file with create table statements but only for one record. Make sure that you click on the include create table checkbox (see Figure 10.)

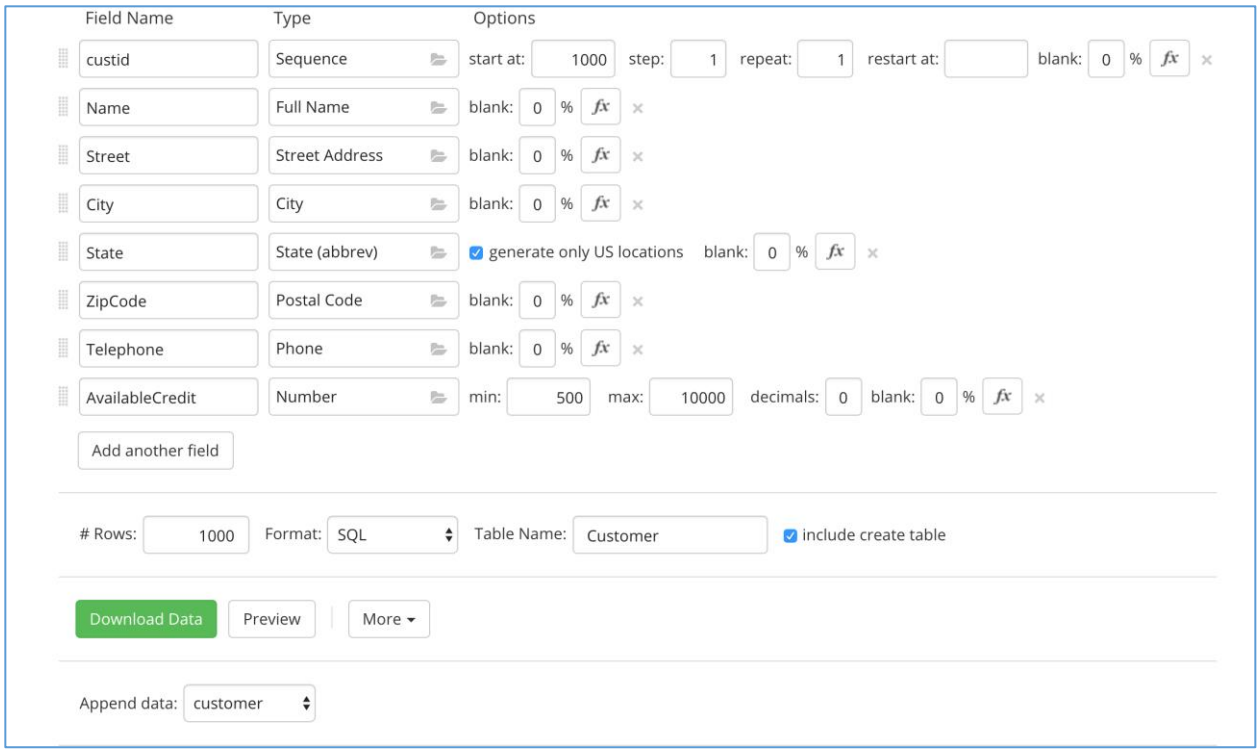


Figure 10

Now, in order to include the data that we created before, we will need to select your table name (customer in my case) as append (see Figure 11)

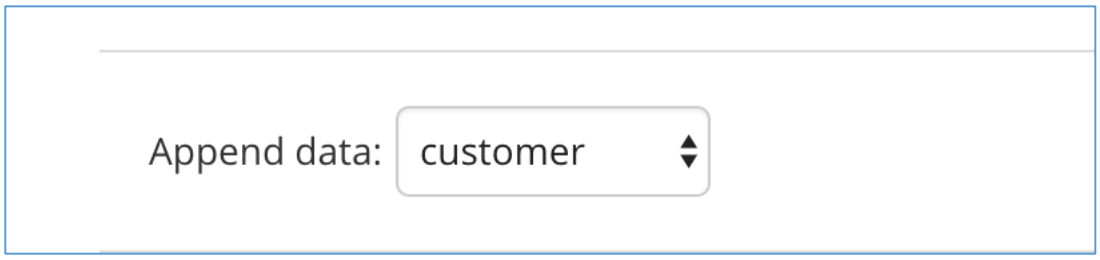


Figure 11

Click on download data. This will give you an SQL file. You may have to tweak the create table slightly, especially if using Oracle. You also should delete the first insert statement as well. You will also want to include a primary key.

Now that this is done and saved, we can create the product schema using the same technique. For the description, I just used drug names since the tool does accommodate this.

Now, for the third table, order, I do not need to save a CSV. However, I will be using the other two CSVs for my references. Notice in Figure 12 that I've chosen a dataset column for productId and custid.

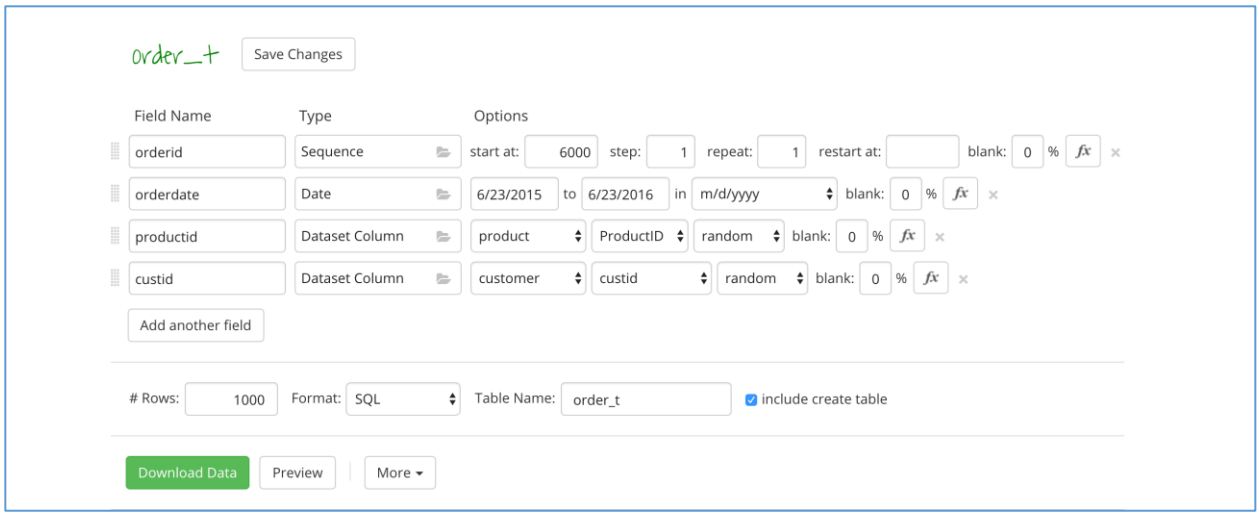


Figure 12

Once you download this, you will need to change the create table statement to include a primary key as well as foreign key references.

You can copy and paste each script into either MySQL or Oracle. Make sure that you create tables that will be referenced by other tables first (for example, product and customer)
You are now ready to create queries.