Contents

[1 Getting Started 3](#_Toc464053929)

[1.1 Accessing inSilico 3](#_Toc464053930)

[1.2 Trouble logging in? 3](#_Toc464053931)

[1.3 Changing your password 4](#_Toc464053932)

[1.4 Logging out and Session Timeout 6](#_Toc464053933)

[2 Introduction to InSilico 6](#_Toc464053934)

[2.1 What is inSilico supposed to do? 6](#_Toc464053935)

[2.2 For whom is this tool intended? 6](#_Toc464053936)

[2.3 inSilico Modules 6](#_Toc464053937)

[What is Library? 6](#_Toc464053938)

[What is Compose? 6](#_Toc464053939)

[What is Train? 7](#_Toc464053940)

[Validate 7](#_Toc464053941)

[What is Deploy? 7](#_Toc464053942)

[What is Visualize? 7](#_Toc464053943)

[3 Library 8](#_Toc464053944)

[3.1 Navigation 8](#_Toc464053945)

[3.2 inSilico Asset Types 9](#_Toc464053946)

[Tables 10](#_Toc464053947)

[Flows 10](#_Toc464053948)

[Models 10](#_Toc464053949)

[Predictions 10](#_Toc464053950)

[Dashboards 10](#_Toc464053951)

[3.3 Importing Data 10](#_Toc464053952)

[3.4 Table Preview 11](#_Toc464053953)

[3.5 Quick Actions 12](#_Toc464053954)

[3.6 Asset Actions 13](#_Toc464053955)

[Tables 13](#_Toc464053956)

[Flows 14](#_Toc464053957)

[Models 14](#_Toc464053958)

[Predictions 15](#_Toc464053959)

[Dashboards 15](#_Toc464053960)

[4 Compose 15](#_Toc464053961)

[4.1 Create a Flow 15](#_Toc464053962)

[4.2 Add Tables to a Flow 16](#_Toc464053963)

[4.3 Create a Step 17](#_Toc464053964)

[4.4 Step Transformations 17](#_Toc464053965)

[Formula 18](#_Toc464053966)

[Aggregate 21](#_Toc464053967)

[Left Join 21](#_Toc464053968)

[Cluster 23](#_Toc464053969)

[Query 24](#_Toc464053970)

[Train 25](#_Toc464053971)

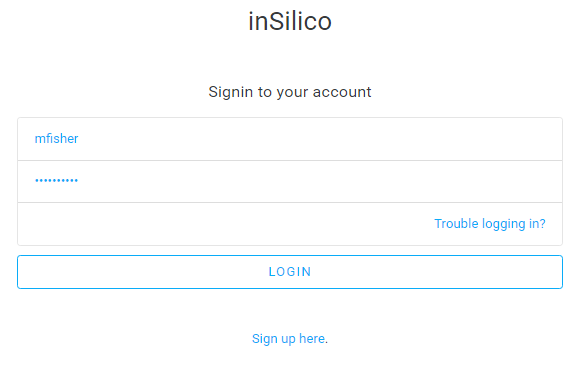
[5 Validate 25](#_Toc464053972)

[6 Report Feedback 26](#_Toc464053973)

# Getting Started

## Accessing inSilico

When you first navigate to inSilico, you will encounter the login screen, where you can enter your username and password if you already have an account or sign up via the “Sign up here” link:



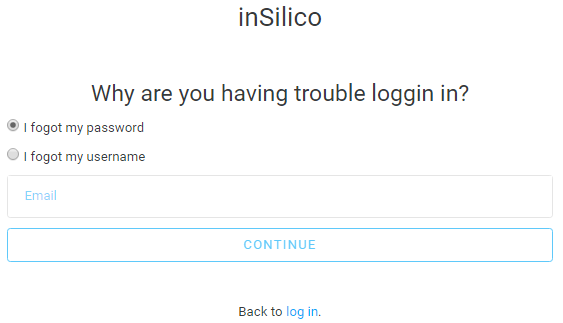
## Creating an inSilico Account

If you do not already have an account, you can create an account by completing the account creation form on the Signup page (found via the “Sign up here” link on the inSilico landing page). After submitting the form, you will not be able to access the platform until you confirm your account via an email that is sent to the specified email address. If you have any issues regarding account creation or difficulties with the confirmation email, please email our support team at [support@sentrana.com](mailto:support@sentrana.com).



## Trouble logging in?

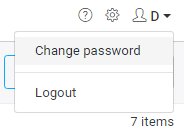
If you have already created an inSilico account and are unable to login, click on the “Trouble logging in?” link. Select the appropriate option and enter your account email address and click Continue:



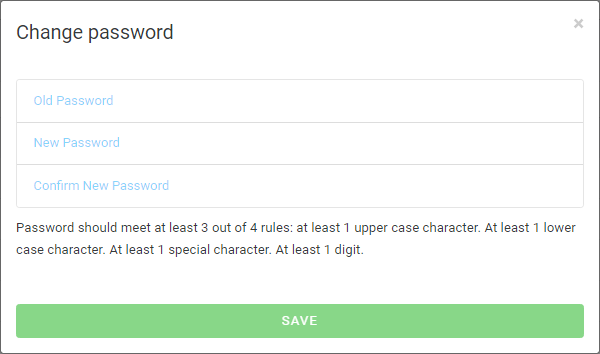
If you forgot your password, you will receive an email that contains a new inSilico randomly generated password. After you log in, you can change your password from the options menu. If you forgot your user name, you will receive an email containing your inSilico user name. If you are still having trouble logging in, please contact our support team at [support@sentrana.com](mailto:support@sentrana.com).

## Changing your password

Once you’ve successfully logged in, you can change your password by clicking on the gear icon at the top right of the screen, selecting ‘Change Password’ and following the instructions:



You will be asked to enter your Old Password and your New Password. Ensure that the New Password meets the password requirements.



You will receive an email notification that your password has been successfully changed. If you receive this email but have not changed your password yourself than contact the inSilico support team at [support@sentrana.com](mailto:support@sentrana.com).

## Logging out and Session Timeout

If you wish to logout from InSilico, simply hover the mouse over the gear icon at the top right of the screen and select ‘Logout’. Any unsaved changes will be discarded when you log out.

Please note that if your computer or device is idle 20 minutes, your session will expire and you will be automatically logged out.

# Introduction to InSilico

## What is inSilico supposed to do?

The purpose of inSilico is to facilitate "Do-it-Yourself" Big Data & Machine Learning. We intend to put into the hands of non-Data Scientists the ability Integrate data from many different tables or silos, manipulate that data, and then discover predictive patterns within the integrated data on their own without the need to write software programs or elicit help from programmers and Data Scientists.

## For whom is this tool intended?

## inSilico Modules

inSilico is comprised of six different modules; Library, Compose, Train, Validate, Deploy, Visualize. For this current version of inSilico, only Library, Compose and Validate are functional. Look for upcoming version of inSilico to expand and include additional modules.

### What is Library?

This the user's main point-of-entry into the inSilico suite, and gives the user a portal into all of the different assets that they can maintain, manage, and interact with on the inSilico suite.

### What is Compose?

Gives the user a visual, spreadsheet-like tool and their own sandbox environment in which they can specify a Flow, which is step-by-step process for manipulating data contained within any table through data transformation operations (this whole process is sometimes referred to as Feature Engineering), integrate data scattered across multiple tables, and cleanse the data (i.e., perform de-duplication). Composer automatically generates the software code needed to fulfill the step-by-step data transformation operations, and will automatically execute this software code on a massively parallelized cluster.

### What is Train?

Gives the user their own sandbox environment in which they can perform fully automated Machine Learning, where all the processing stages involved in Machine Learning (referred to as the Machine Learning Pipeline - MLP) and the configuration of parameters needed to drive each stage (these configuration parameters are referred to as Hyper-parameters), are automatically created and explored by inSilico to find the best configuration of MLP stages that ultimately produces the most valid/accurate AI Model). Cortex also gives the user the ability to immediately make Predictions from this newly discovered or any pre-existing AI model using any batch of input data as long as the data conforms to the input data fields expected by the AI model.

*Note: Train currently lives inside of Compose in this version. It will be moved to the Train tab in a future version.*

### 

### Validate

Gives the user the ability to test model performance against either a subset of the training data (e.g. testing out of bag performance via an 80-20 split or testing model performance on a subset of data of interest) or the entire dataset used for training as long as the new dataset has the exact same data structure (same number of columns, with data types, relation, etc.). This allows the user to compare actual and predicted responses for the executed dataset by mapping each source table and column to the corresponding new table and column.

### What is Deploy?

Gives the user the ability to deploy their AI model(s) so that inSilico either: (1) Automatically "wakes up" at pre-specified times to make new predictions using any new data that may have accumulated since the last round of predictions, or (2) Automatically detects a change in the data in the input tables and automatically makes new predictions using these changes (inSilico does not automatically re-train an AI Model - re-training must be initiated by a user). Robot also gives the user the ability to setup Automated Alerts if new predictions fall within some criteria specified by the user (these criteria can be setup as Rules).

*Note: Not included in current version. Will be implemented in future version of inSilico*

### What is Visualize?

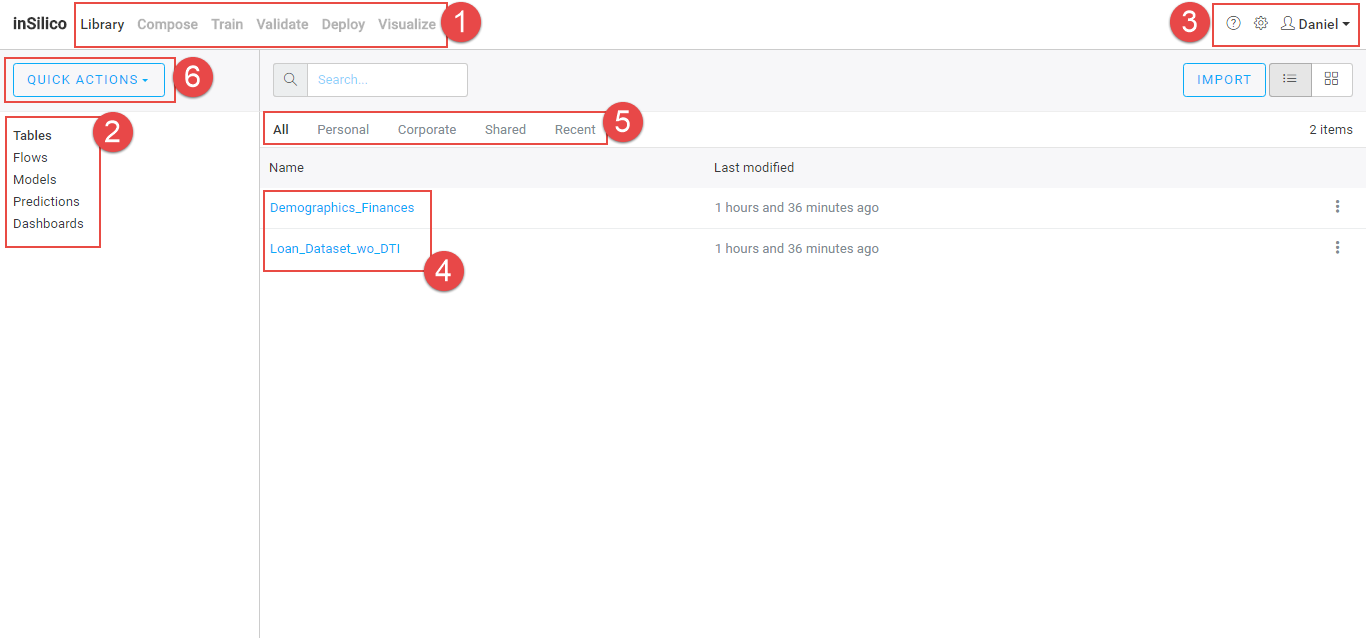
Gives the user a tool to perform Visual Analytics on the AI Models and any of the Data available through the user's Library. An AI Model provides a mathematical "formula" that Predicts, Classifies, or finds Association Links in information in response to the values of input variables (known as Predictor Variables). Corona allows the user to simulate what Predictions, Classifications, and/or Associations will be generated by the AI Model when the user supplies data for the input variables. Within Corona, the user can combine data from the Predictions/Classifications/Associations with any of the other Data contained within their library and create Dashboards, Reports, and Ad-Hoc Analyses similar to those provided by the leading Visual Analytics tools (with the difference against these leading tools being Corona empowers a user to visualize an AI Model, and not just visualize historical data).

*Note: Not included in current version. Will be implemented in future version of inSilico*

# Library

## Navigation

Library module is where you will manage all of your assets. The image below describes all of the important sections of this module/



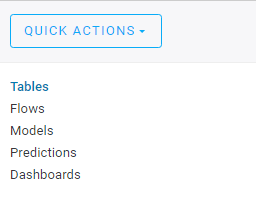
1. **inSilico Modules** – this tabs will appear throughout the application and allow you to quickly switch between modules
2. **inSilico Asset List** – the left panel contains the different inSilico assets. Selecting one of these will display as of those assets in the main area to the right.
3. The top-right contains the link to the help guide, asset summary and your account info. Such as change password and logout.
4. The main area will display the list of assets that you have selected in the left panel. In this case, Tables are displayed since Tables is selected in the left panel. Your selection will appear in bold.
5. **Asset Tag** – this section allows you to filter the selected asset list.
6. **All** – displays all assets
7. **Personal** – displays only assets you have ownership of
8. **Corporate** – displays assets that your company has ownership of
9. **Shared** – displays assets that were shared with you
10. **Recent** – shows the assets that were recently modified

Note: only All and Personal are implemented for version 1.4

1. **Quick Actions** – clicking on this displays a list of actions

## inSilico Asset Types

All of the inSilico assets will be stored in Library. You can navigate to each by using the left panel and selecting the asset you want to view.



### Tables

There are various types of Tables in inSilico. Regardless of the type they will all be stored here. These are all the tables of data, and their corresponding Meta-Data, to which the user has access and is permitted to utilize. Tables are used to create Flows, A.I. models or to create Dashboards & Reports.

The different types of tables are:

1. Source Table – Tables that you imported into inSilico
2. Derived Table – Tables that were generated as a result of a Transformation. This includes Tables that were created in Compose, Train and Validate

### Flows

The sequential step-by-step process of Data Operations is referred to as a Flow, or a "Directed Acyclic Graph" (DAG). We refer to each step of the Flow as a "Step" (or, also as a "Node" in the DAG). The Flow step consists of an input table and a transformation. The successful execution of this transformation results in a derived table.

### Models

After a model has been trained, it is stored as a Model asset. This asset contains the mathematical, statistical, or machine learning algorithms that are executed behind the scenes to generate Predictions. Models can be Edited (retrained using a *new* set of tables) or Refreshed (retrained using the same set of tables if, for example, the source tables have been updated).

### Predictions

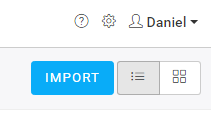
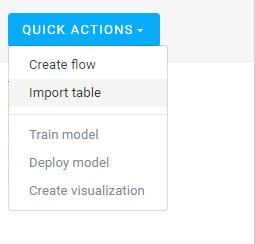
Predictions are the generated output tables of executing a model against a new dataset to generate a set of predictions. This is where the “insights” generated from a model are stored. At its core, Predictions are a special case of a Table, so the same set of actions that can be performed on a Table can be performed on a Prediction (e.g. Visualize, download, add/use in a flow, etc.).

### Dashboards

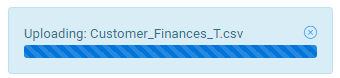
Dashboards include the various visualizations that are stored to visualize, interact with, and share data. Dashboards, similar to other BI dashboards, refer to some underlying data. In inSilico, the user can create Dashboard using any combination of Tables and Predictions. We plan on extending the Dashboard capabilities to include Models, giving the user the ability to interactively play with the Model to deepen understanding and build intuition.

## Importing Data

Data can be imported two different ways. By click on IMPORT on the right side of Library. Also, by clicking on QUICK ACTIONS and selecting Import table from the dropdown list

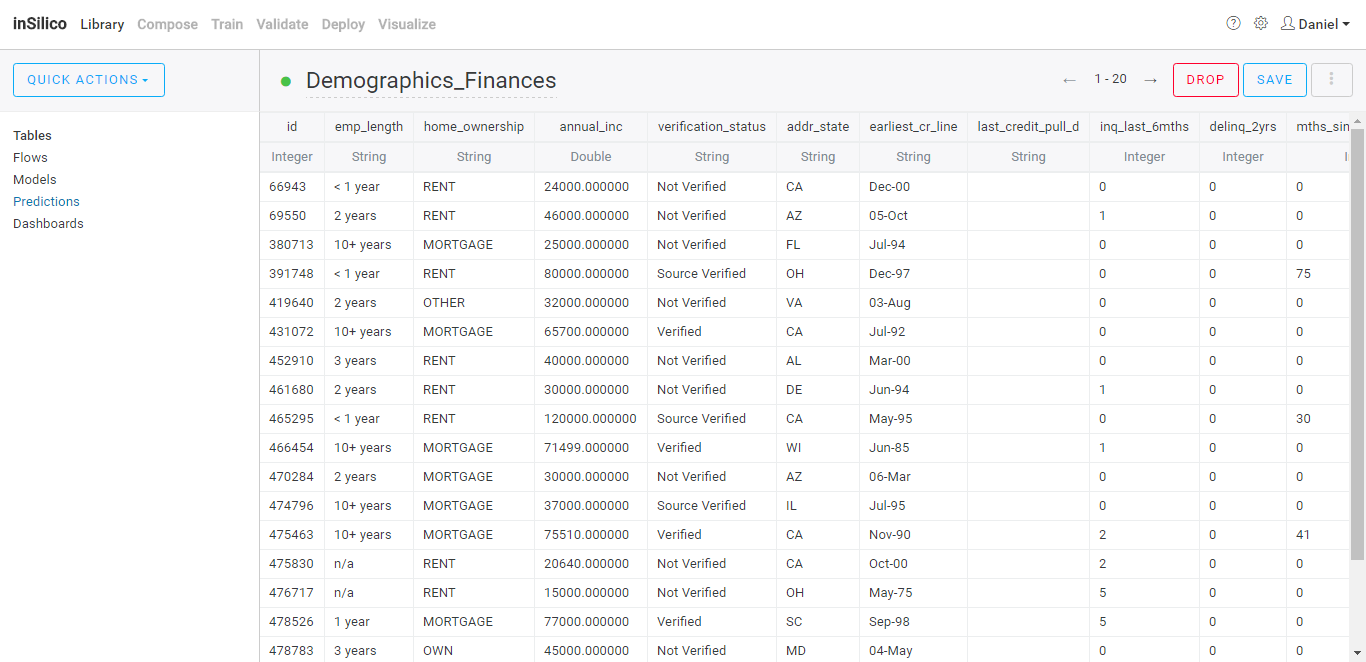
 

Both of these options will launch the same file import window. Select the table you wish to upload from your computer. You may upload multiple tables at the same time. Once you add the table, you’ll notice a blue box in the bottom right corner. This displays the table upload progress. If the table imports successfully, it will appear in your Table list.



## Table Preview

To preview a table from the Library, simply click on the table name and you will see a preview of the table. To change the name of the table, click on the table name. After you rename the table, hit enter on your keyboard to save.

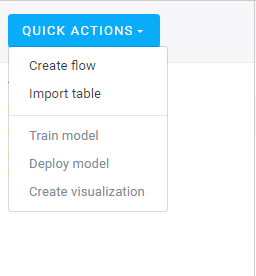


## Quick Actions

Quick Actions displays a list that allows you to quickly do multiple things.

1. Create flow – routes you to Compose to begin creating a new flow
2. Import table – launch file import window
3. Train model – routes you to Train to create a new model
4. Deploy model – routes you to Deploy to deploy a trained model
5. Create visualization – routes you to Visualize to create a dashboard using your tables and models

Note: only Create flow and Import table are implemented for version 1.4

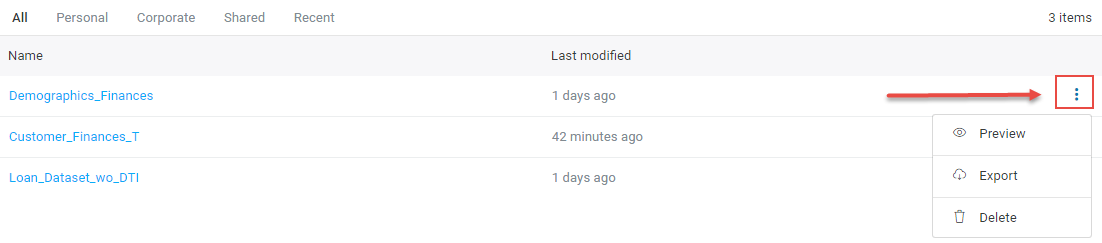


## Asset Actions

Each inSilico asset has a different set of actions you can take on them. You can display these actions by clicking on the icon with the vertical three dots.

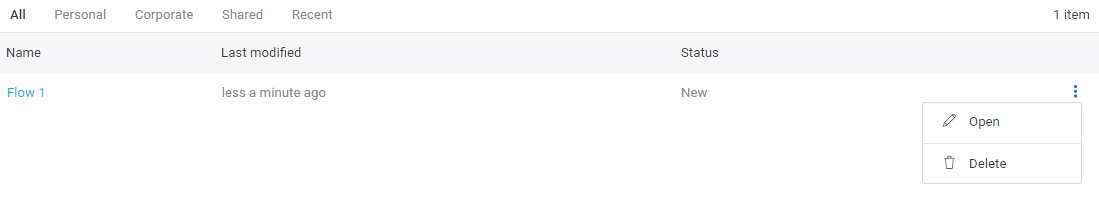
### Tables

The available actions for table are Preview, Export and Delete. Preview was discussed earlier. Clicking Export will prompt a download window and export the file as a .csv. Clicking on Delete will remove the table from your Library. Note that this may have an effect on any Flow that uses this table.



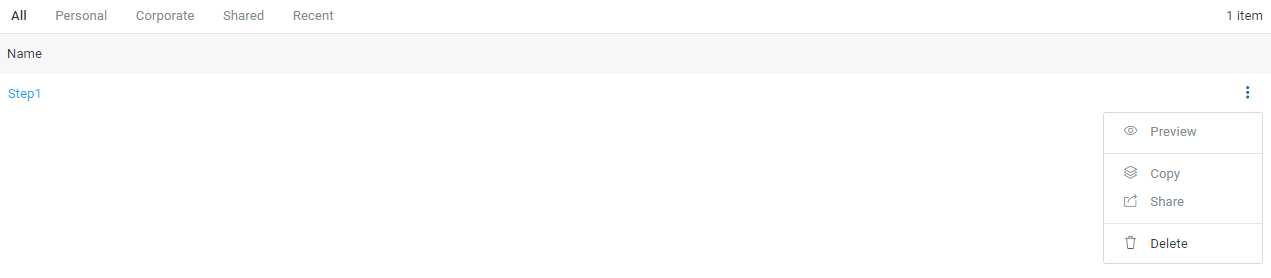
### Flows

The available actions for table are Open and Delete. Opening a Flow will route you to Compose with the Flow selected. Clicking on Delete will remove the Flow from your Library. Note that this may have an effect on any Model that uses this Flow.



### Models

Currently, you can only delete a Model in Library. Additionally, you can remove a Model, by deleting the defining Step in the Flow. To see the Model definition (underlying table, and Response/Predictor selectors), open the corresponding Flow and navigate to the Train step.



### Predictions

Currently, there is no capability to generate Predictions. Therefore, there are no Prediction assets and the Prediction tab in Library is left intentionally blank.

### Dashboards

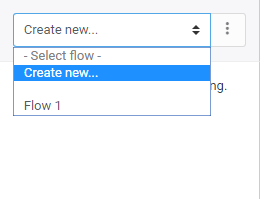
Currently, there is no capability to generate Dashboards. Therefore, there are no Dashboard assets and the Dashboard tab in Library is left intentionally blank.

# Compose

A Flow is an abstract mathematical concept that encapsulates a sequence of data operations, where the operand of the operation is actually a meta-data field. In order to make Big Data Analytics and the creation of Artificial Intelligence an easy process for non-Data Scientists, we want to help users perform the data preparation required prior to the invocation of Machine Learning as simple as manipulating and organizing data and worksheets in Excel. To achieve this, we will allow users to do a sequential step-by-step process of discrete data operations that are performed using Input Tables selected or uploaded by the user. The result of the operation is saved off as an Output Table (also referred to as a Derived Table). These data operations to the "raw" data sources and subsequent "saves" do not mutate any of the original source tables.

## Create a Flow

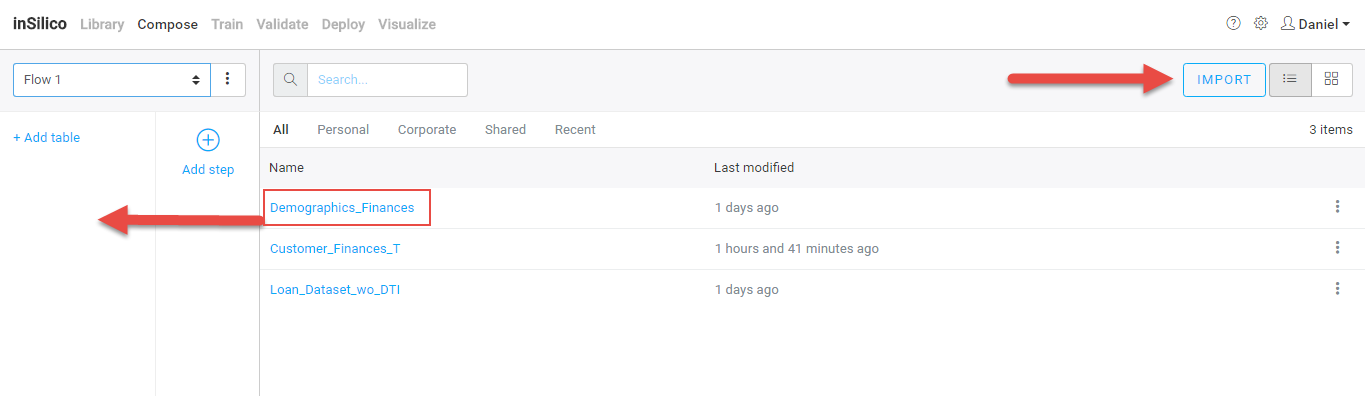
There are multiple ways to create a new Flow. Selecting Create flow from the Quick Actions button as highlighted earlier. Another way is to select the Compose tab. You will be directed to the Compose landing page where you can create a new flow by selecting it from the Flow dropdown list on the left



Once you select the option above, you will be prompted to give the Flow a name and a description. Flow name is required but description is optional. Once you’ve entered a name for your Flow, click the create button. Your flow will now be created and ready to begin using.

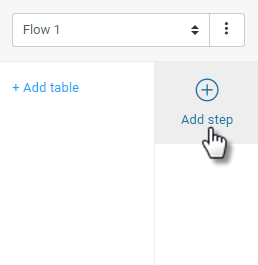
## Add Tables to a Flow

The first thing you will be asked to do when you create a Flow is to add tables. You can do this by dragging tables that already are in your Library from the right side to the Flow on the left. You can also add tables to the Flow that aren’t yet in your Library by clicking on IMPORT.



## Create a Step

Flows are made up of a series a Steps. A Step is where you execute transformations on tables to produced derived tables. Each Step can only contain 1 data engineering operation. Operations can be thought of as appending a new column to the existing table structure. When the Data Engineering operation executes successfully, it will be added to the Flow List and, optionally, the DERIVED folder in the LibraryTo add a Step, click on the “Add Step” icon in the left panel.

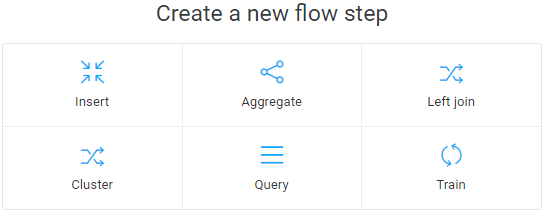


After you click Add Step, you will see a set of Transformers you can choose from.

## Step Transformations

A step consists of an input table(s) and a transformation taken on that table(s). The result creates a derived table. This table will appear in the left panel along with the other tables in the Flow. You can use the created derived table as an input table in any step.

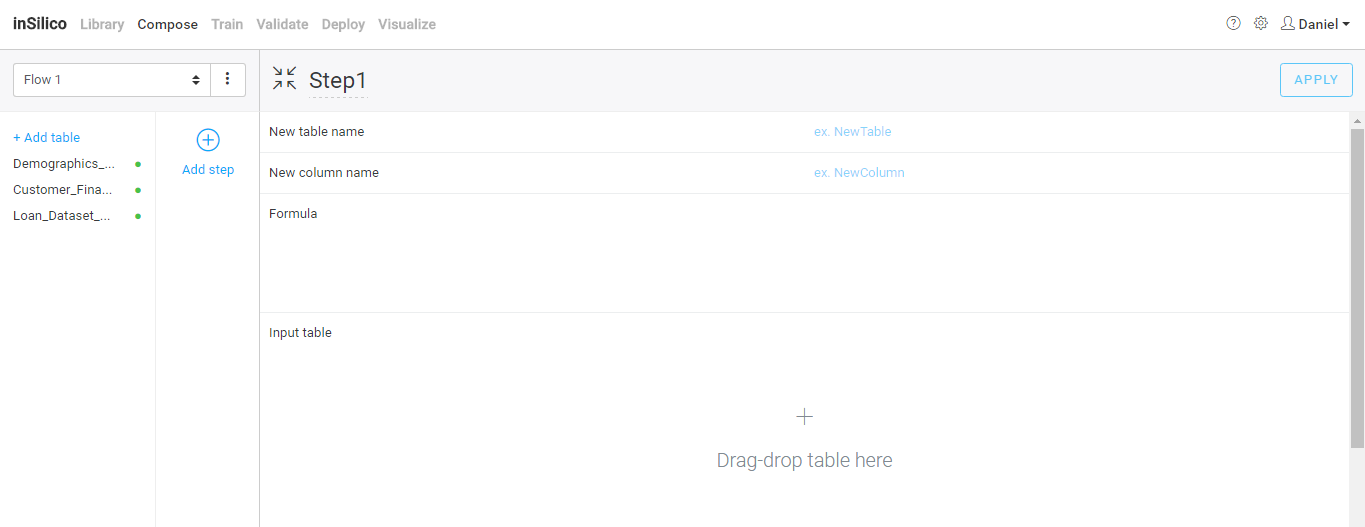
You can choose from six different transformations. Each transformation will have its own specific set of inputs. However, each will require you to name the output table.



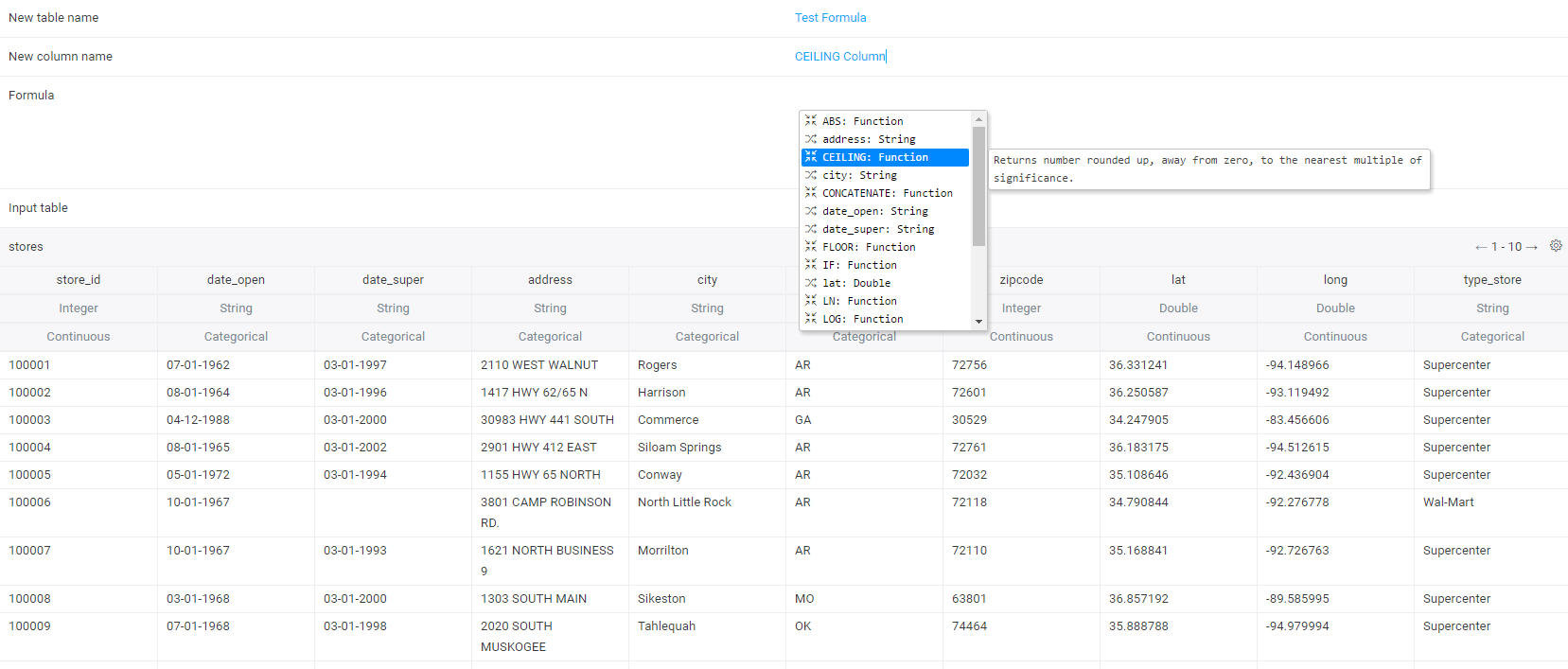
### 

### Formula

Formula allows the user to insert columns using Excel-line formulas. The user will input the formula into a formula input field and the system will auto-fill the formula to all the rows.



To begin, drag a table from the left panel to the table drop area on the right. The table data will now be viewable. Name both the output table and the new column. Then click on the formula area, a list of formulas will appear. NOTE: You don’t need to start the formula with an “=” as you would in Excel – just begin composing the Formula directly. As you mouse-over the formulas, each one will display an explanation to help you better understand the transformation. Once you select a formula, you can type out the formula just like you would in excel. Select the column you want and operator type. When are the input values have been specified, click APPLY and the transformation will be executed. If executed successfully, you will see the derived table with the newly appended column.

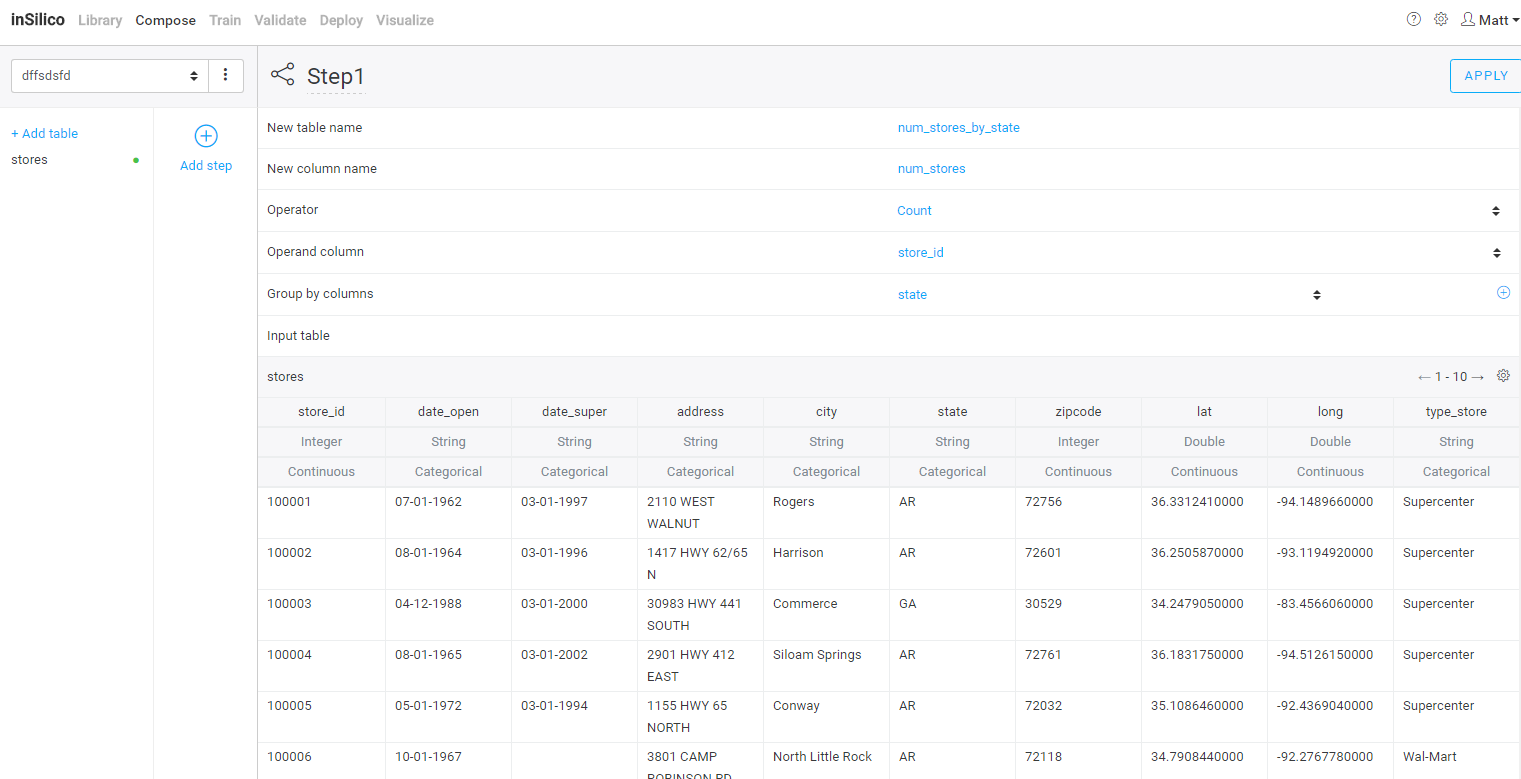


### 

### Aggregate

The Aggregate operator is equivalent to the GROUP BY operation in SQL and allows the user to “roll up” or aggregate data based by collapsing data based off of a column or multiple columns (i.e. equivalent of group by columns in SQL) and perform a mathematical operation on each set of rolled up data. For example, I want to figure out how many stores there are in each state, I would:

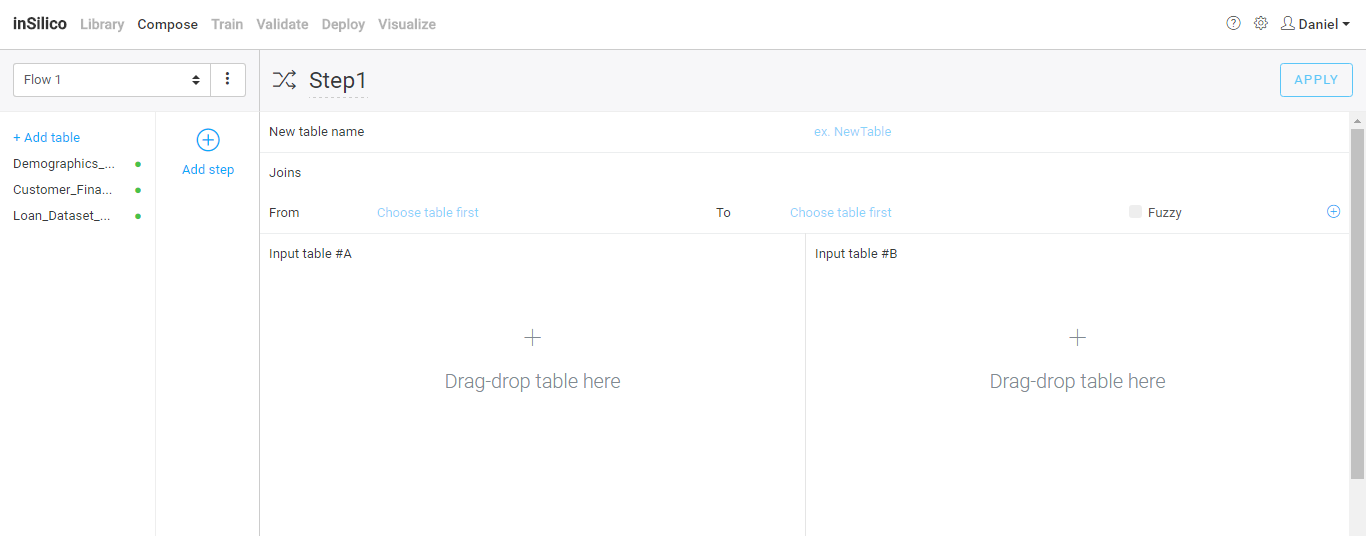
1. Name derive table and column
2. Set operator to Count
3. Set operand column (the column that is being used in the collapsing operation) to state\_id
4. Set group by column(s) (the key over which I’m rolling up the data) to state



### 

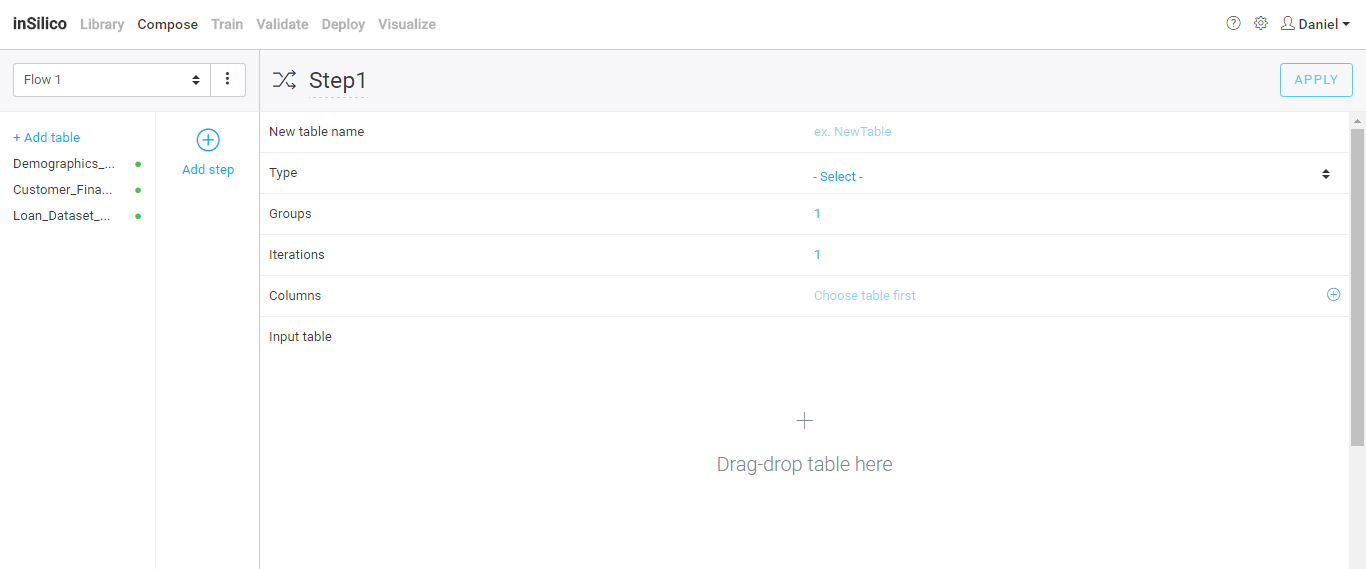
### Join (Left)

Join allows the user to execute left table joins. The user will specify the results table name and which columns to use as a key in the join operation. Currently, only left joins are supported with the left table (the table in the left drag-drop box) being used as the base and joined data from the right table are pulled into the resultant (joined) table. Where the right table does not have a matching key, null (empty) values will be inserted into the joined table. When the join is executed successfully, the results table will replace the two input tables on the step screen.



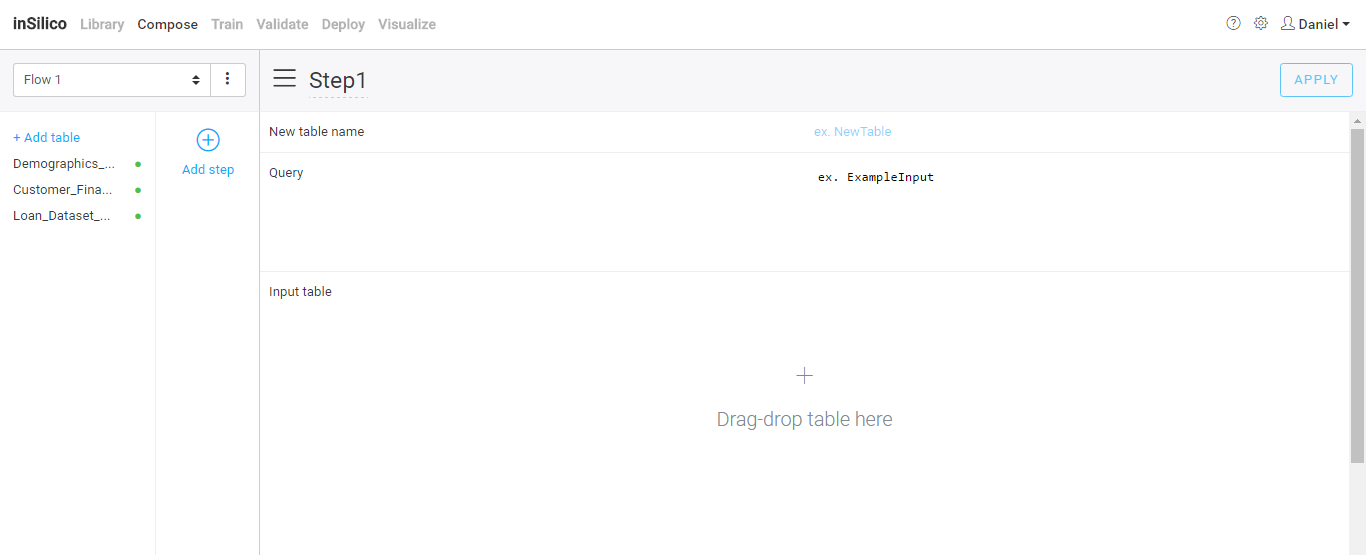
### Cluster

Clustering gives the user the ability to bucket data into distinct groups via an automatically applied machine learning algorithm. The user must specify the number of groups or clusters they expect there to be in the data (e.g. an example of this could be low-medium-high performers) and which column(s) they want the algorithm to use to sort the data points into distinct groups (e.g. in the previous example, sales and purchase frequency can be the defining characteristics). NOTE: the algorithm will always divide the data points into the specified number of groups; the number of groups and the defining columns are the most important and sensitive inputs the user must supply. There are two types of clustering: K-Means and Fuzzy-C-Means. K-Means will only give binary categorization (i.e. group 1 vs group 2) and Fuzzy-C-Means is defined on a continuum, representing like a particular data point is to a cluster (i.e. 0.85).



### Query

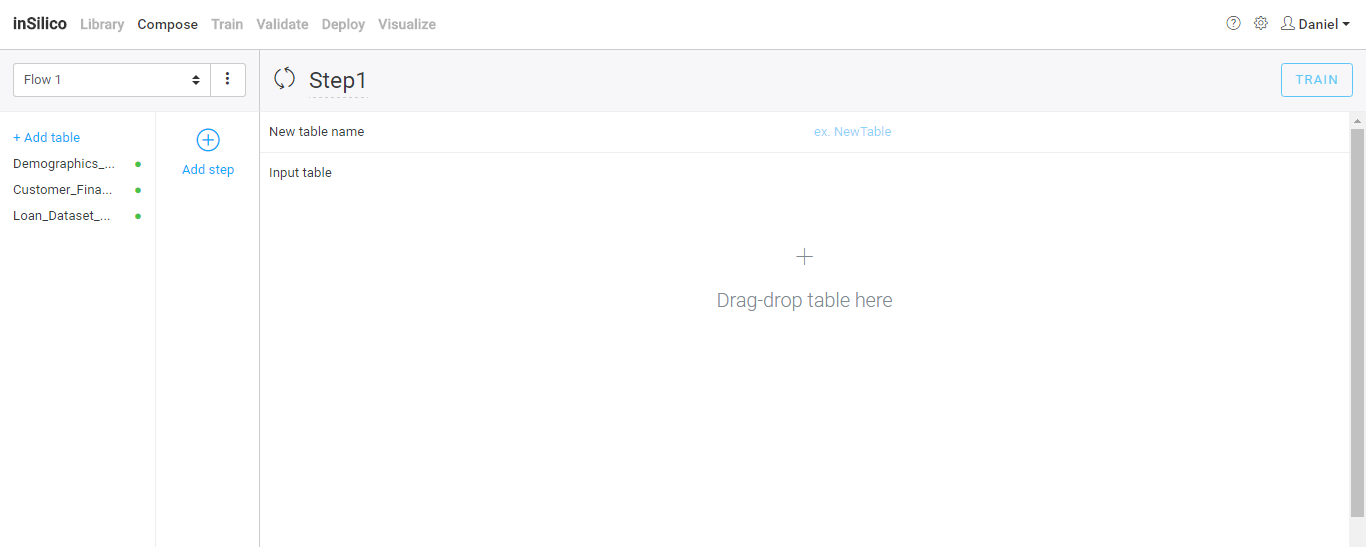
Query is the most open ended and perhaps one of the more powerful operators. It allows the user to inject any user-defined SQL select statement to define a new table. Refer to Redshift documentation for syntax and set valid of functions. Dragging a table into the drag-drop box, automatically prefills the query to: “SELECT \* FROM \_\_\_\_\_\_”. Proceed to modify the query using the autocomplete to assist in referring to columns.



### 

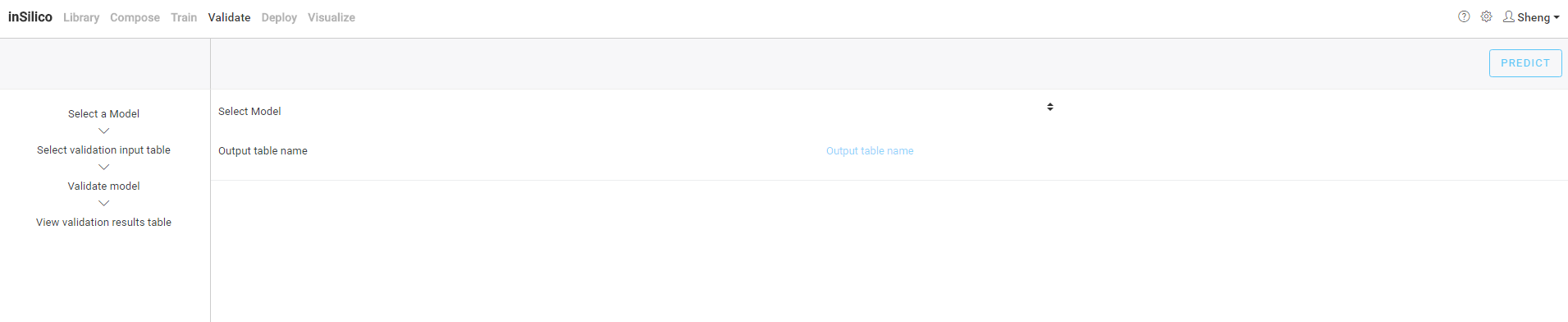
### Train

Train is perhaps the most important “operator” in Compose. It allows users to define a Model definition—given by the input matrix and the selection of predictor columns and response column. After dragging the table that you want to be the input matrix for your Model, you will see a preview of the table in the drag-drop box. By default, all columns are marked as “Ignore”. Using the dropdown selector for each column, select the set of Predictors and one Response column you want. NOTE: you must have at least one Predictor and one and only one Response. By clicking Train, the model training has been initiated and a view of counter of elapsed time and percent complete will be shown. At any time during the training process, you can review the Model definition by navigating to the Input Table tab. Upon successful completion of the model, the saved model asset will appear in Library. Removing an executed Train step will also delete the generated Model.



# Validate

This module enables you to test model performance against either a subset of the training data (e.g. testing out of bag performance via an 80-20 split or testing model performance on a subset of data of interest) or the entire dataset used for training as long as the new dataset has the exact same data structure (same number of columns, with data types, relation, etc.). This allows you to compare actual and predicted responses for the executed dataset by mapping each source table and column to the corresponding new table and column. First, select the model you want to Validate via the dropdown. After defining the output table name, you must map a table to each of the source tables used in the flow that generated the model. Navigate through all input table tabs to select the appropriate mapping. By default, the system will try to map source to new column names based on the column name (if validating a model using the training dataset, all columns should be mapped automatically once the table has been mapped). Clicking “Predict” initiates the model Prediction and upon successful execution, the output table saved (includes all source columns, including source response, and predicted response).



# Report Feedback

We encourage our Users to submit any feedback to [support@sentrana.com](mailto:support@sentrana.com). Please report any bugs you may encounter, general questions and features or enhancements you would like to see in future versions of inSilico.