**Art of Machine learning for beginners**

**Azure Machine Learning Studio**

Author: Balamurugan Balakreshnan

Date: Nov 30th 2018

Step by step guide to create Machine learning model using Azure Machine learning Studio

Contents

[Introduction 2](#_Toc530567396)

[Prerequisite 3](#_Toc530567397)

[Building Model 8](#_Toc530567398)

[Deploy Model 20](#_Toc530567399)

[Q&A 23](#_Toc530567400)

[Conclusion 23](#_Toc530567401)

[Next Steps 23](#_Toc530567402)

# Introduction

#### KDD Cup 2009: Customer relationship prediction

Customer Relationship Management (CRM) is a key element of modern marketing strategies. The KDD Cup 2009 offers the opportunity to work on large marketing databases from the French Telecom company Orange to predict the propensity of customers to switch provider (churn), buy new products or services (appetency), or buy upgrades or add-ons proposed to them to make the sale more profitable (up-selling).

The most practical way, in a CRM system, to build knowledge on customer is to produce scores. A score (the output of a model) is an evaluation for all instances of a target variable to explain (i.e. churn, appetency or up-selling). Tools which produce scores allow to project, on a given population, quantifiable information. The score is computed using input variables which describe instances. Scores are then used by the information system (IS), for example, to personalize the customer relationship. An industrial customer analysis platform able to build prediction models with a very large number of input variables has been developed by Orange Labs. This platform implements several processing methods for instances and variables selection, prediction and indexation based on an efficient model combined with variable selection regularization and model averaging method. The main characteristic of this platform is its ability to scale on very large datasets with hundreds of thousands of instances and thousands of variables. The rapid and robust detection of the variables that have most contributed to the output prediction can be a key factor in a marketing application.

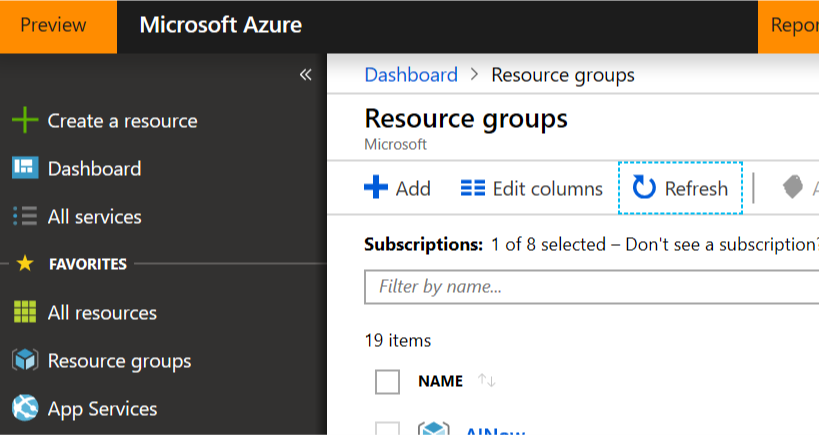
The challenge is to beat the in-house system developed by Orange Labs. It is an opportunity to prove that you can deal with a very large database, including heterogeneous noisy data (numerical and categorical variables), and unbalanced class distributions. Time efficiency is often a crucial point. Therefore part of the competition will be time-constrained to test the ability of the participants to deliver solutions quickly.

Details can be found in <https://www.kdd.org/kdd-cup/view/kdd-cup-2009>

The sample for our workshop today is from this URL. <https://gallery.azure.ai/Experiment/Binary-Classification-Customer-relationship-prediction-1>.

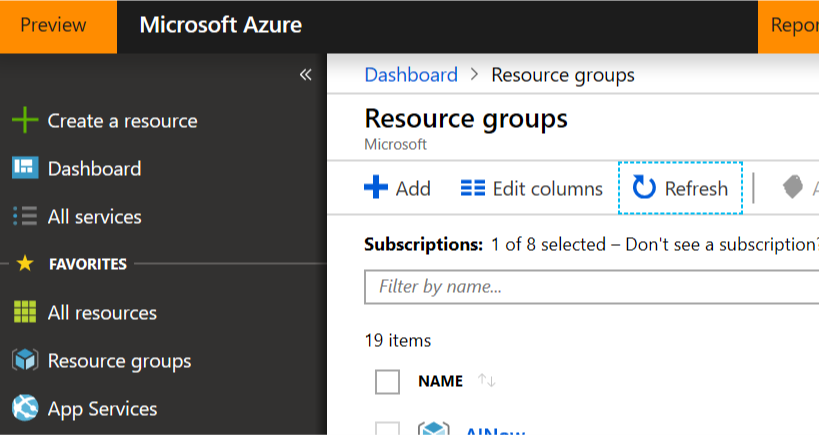
# Prerequisite

1. Azure Account
2. Login into Portal
   1. Open Browser preferable Edge, Chrome or Firefox
   2. https://portal.azure.com
   3. Once you have logged in.

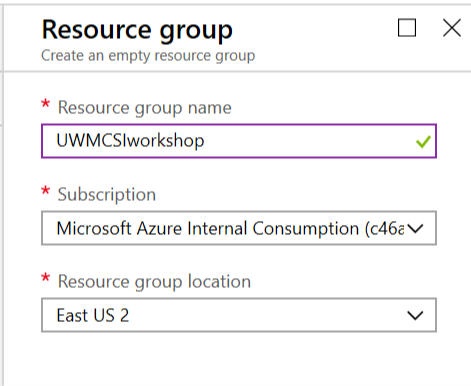


* 1. After login you should see the above screen

1. Create a resource Group

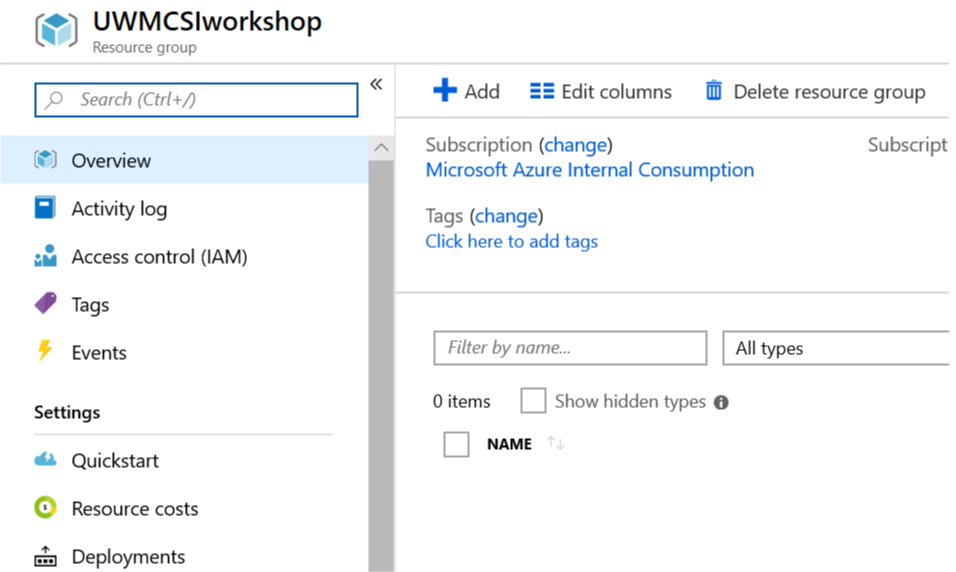


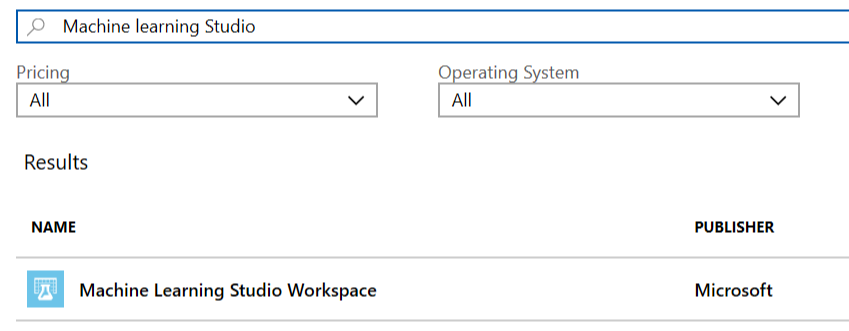
* 1. You should be able to see the above screen.
  2. Click on the Resource groups on the left Menu
  3. Then Click Add button to create a new resource group
  4. Resource group name: UWMCSIworkshop

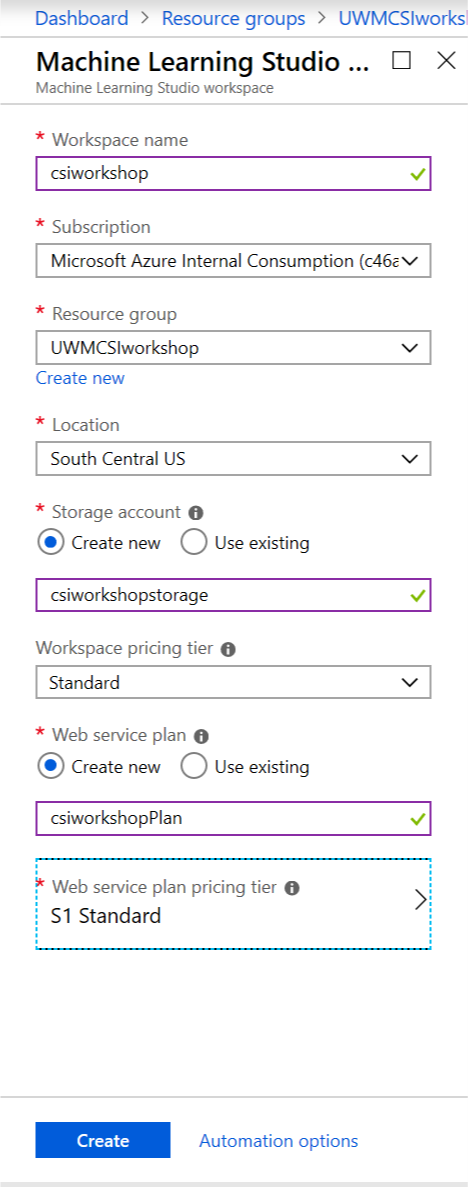


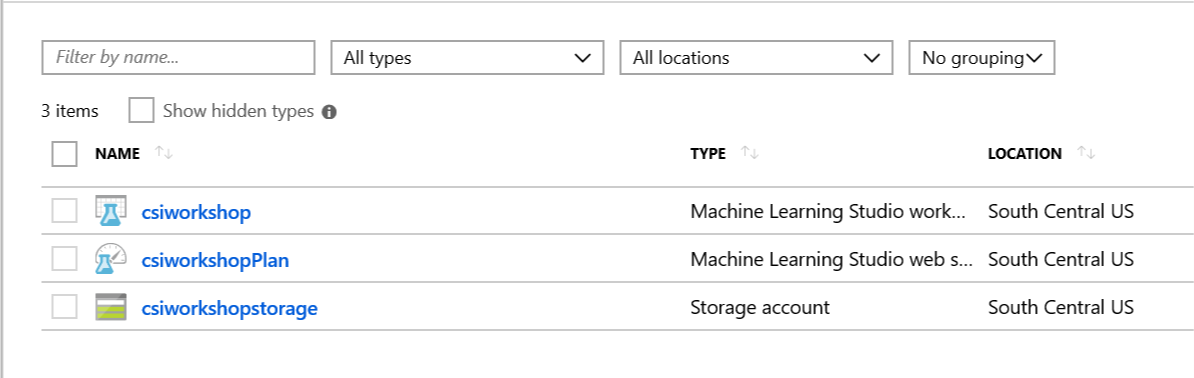
* 1. Select the appropriate subscription if you have many
  2. Select the region as East US 2 for workshop
  3. Then in the bottom of the Blade or Section click Create
  4. Wait for few minutes and you will now have a resource group

1. Create Azure machine learning workspace
   1. Now click the newly created resource group and that should open the resource blade.
   2. The resource blade should have nothing at this time

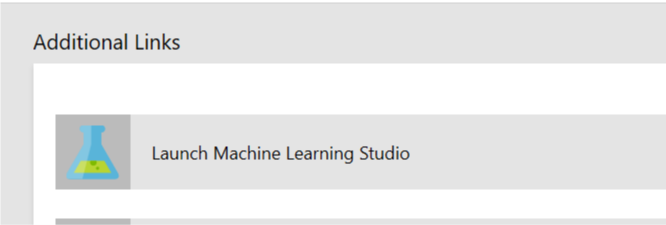


* 1. Click the Add Button
  2. On the search bar type Machine Learning Studio
  3. 
  4. Select Machine Learning Studio Workspace
  5. Then click create.
  6. Fill in the details as per the below picture



* 1. Once you fill the details like above then click Create button
  2. If the name for workspace already exist or for any other resource change the name appropriately
  3. Wait for few minutes until the resources are created
  4. Here are a sample of once the resources are created
  5. 
  6. The web service plan is for deploying the machine learning model as web api REST API

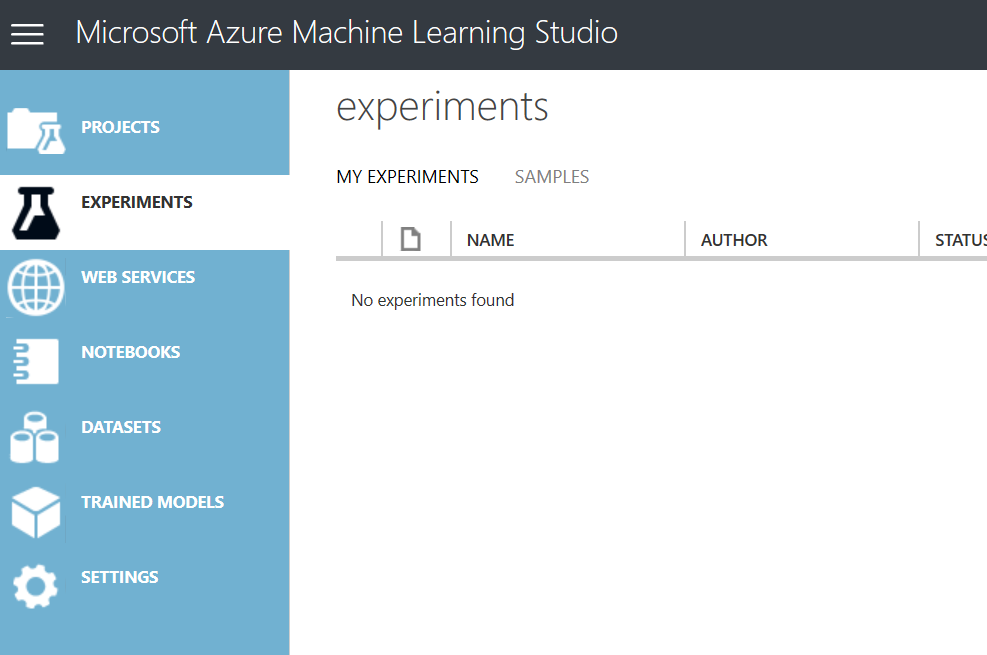
1. Now you can go into the Machine learning Studio
   1. From the above resource group click csiworkshop
   2. This should give you option like below
   3. In the Overview page you should be able to see

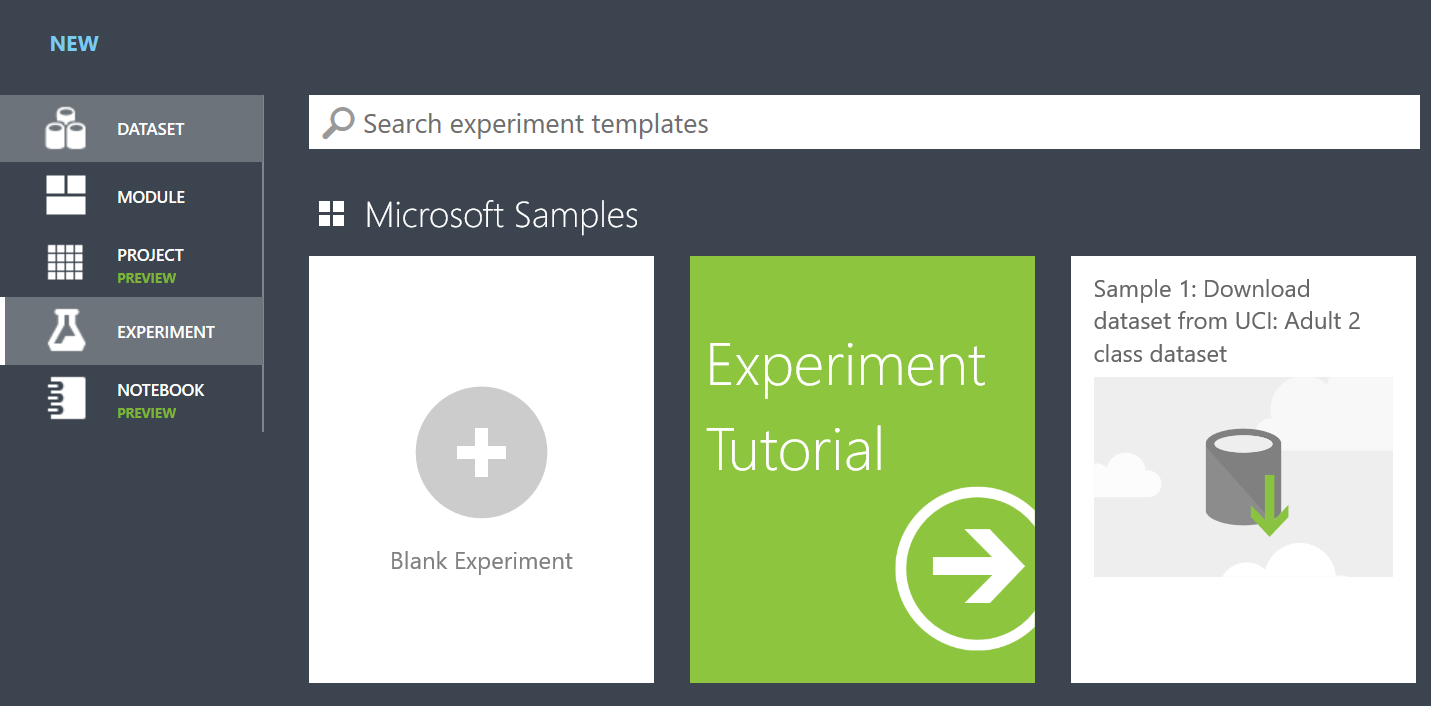


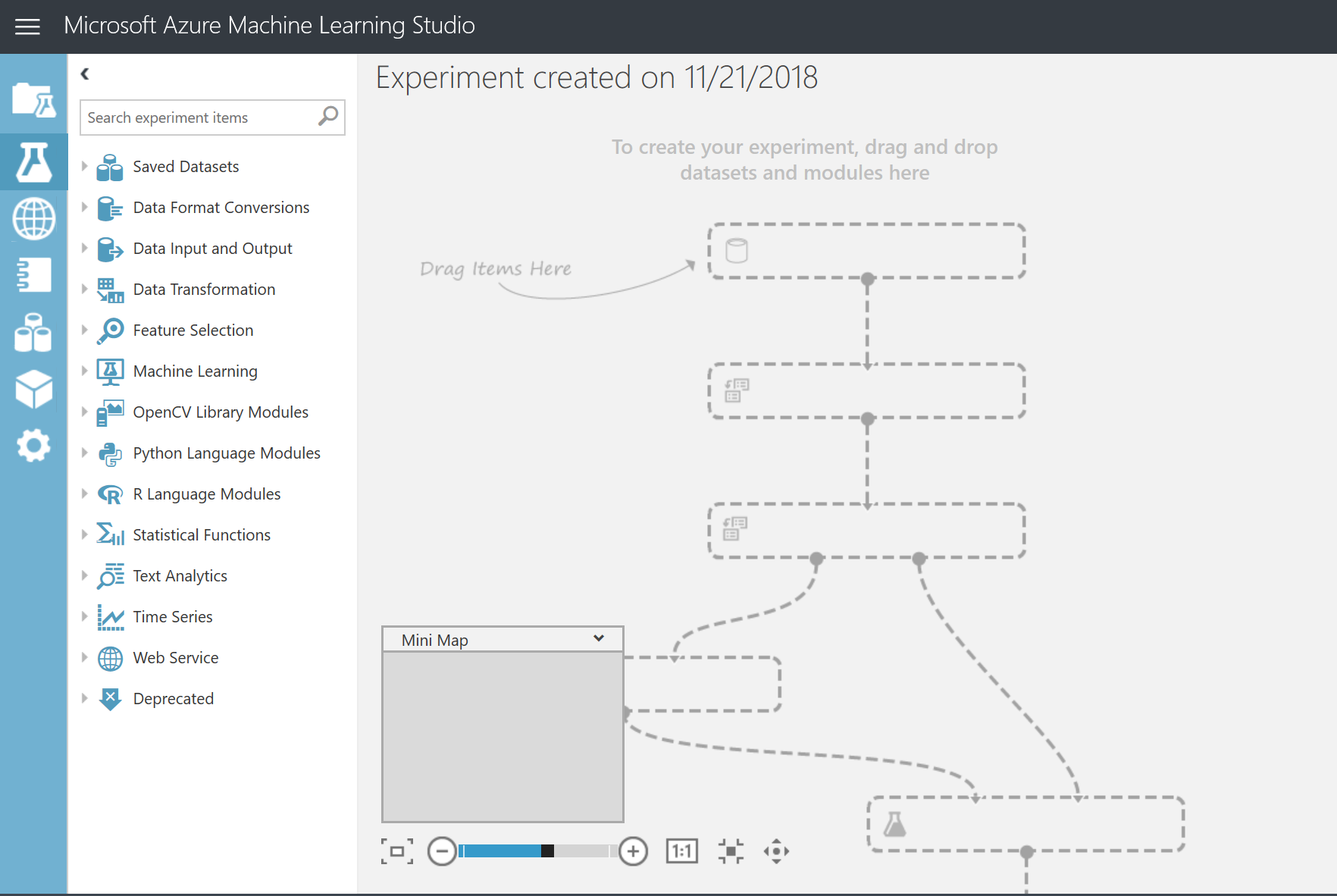
* 1. Now click on Launch Machine learning Studio and that will take you to the Machine learning Studio IDE
  2. Sign in with your login used for Azure Portal Access

# Building Model

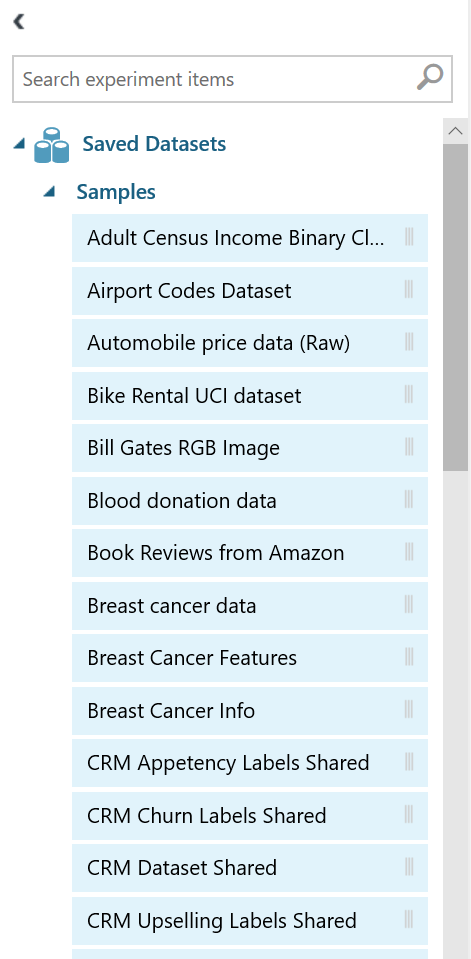
1. Log into the Machine learning studio workspace as per above instruction
2. Now you should see empty experiments.
3. The left side has Menu navigation with all the options.



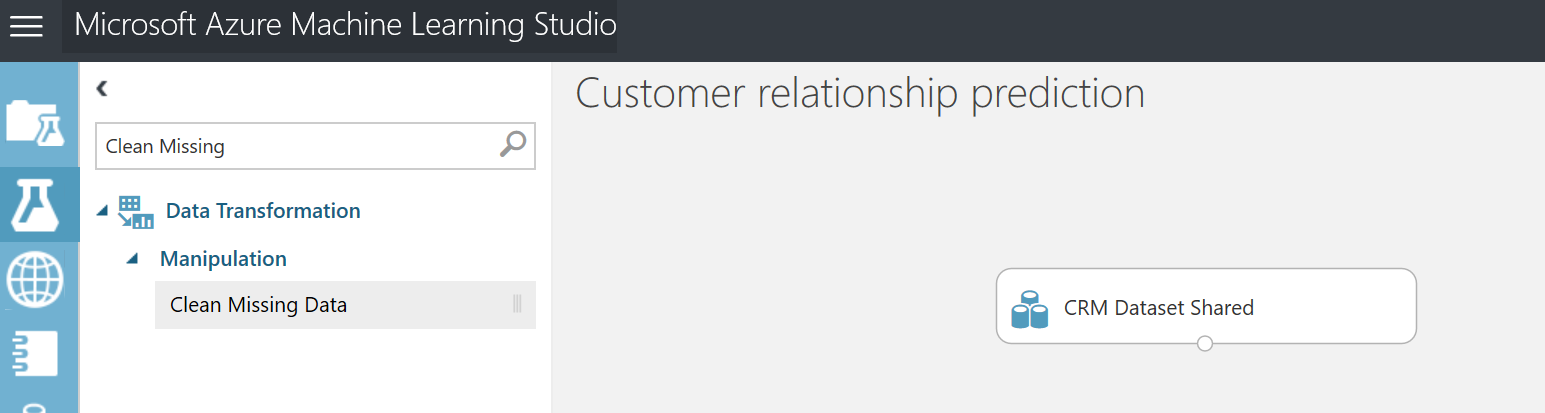
1. Here we can work both with drag and drop ide or notebooks as well.
2. for now, we are going to create experiment.
3. ON the left bottom there should be a + button
4. 
5. Click the new
6. Here is screen you should see
7. 
8. Select Blank Experiment
9. Here is what you should see



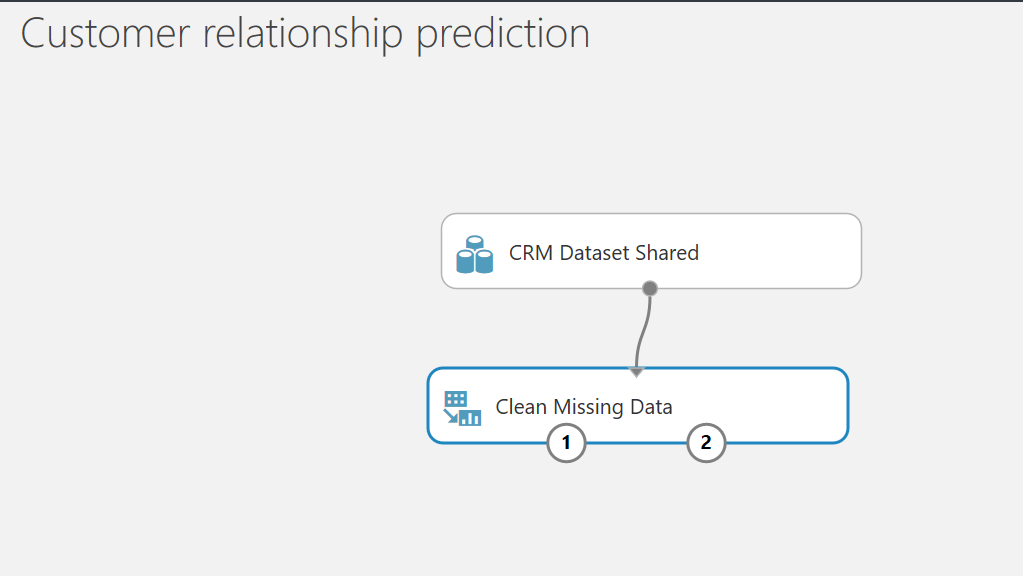
1. This is a blank experiment and we can now start are Model building.
2. First thing lets name the experiment as Customer relationship prediction
3. Now we need our data set to do our experiment. For the workshop I am using existing available public data set.
4. On the left menu expand the Saved Datasets
5. Then expand the samples



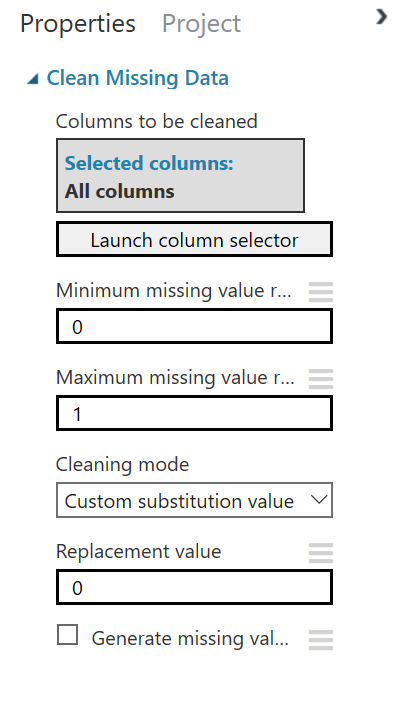
1. Now drag CRM Dataset shared and drop in the canvas
2. Now that we have the data, we need to do some cleaning.
3. On the search window type clean missing



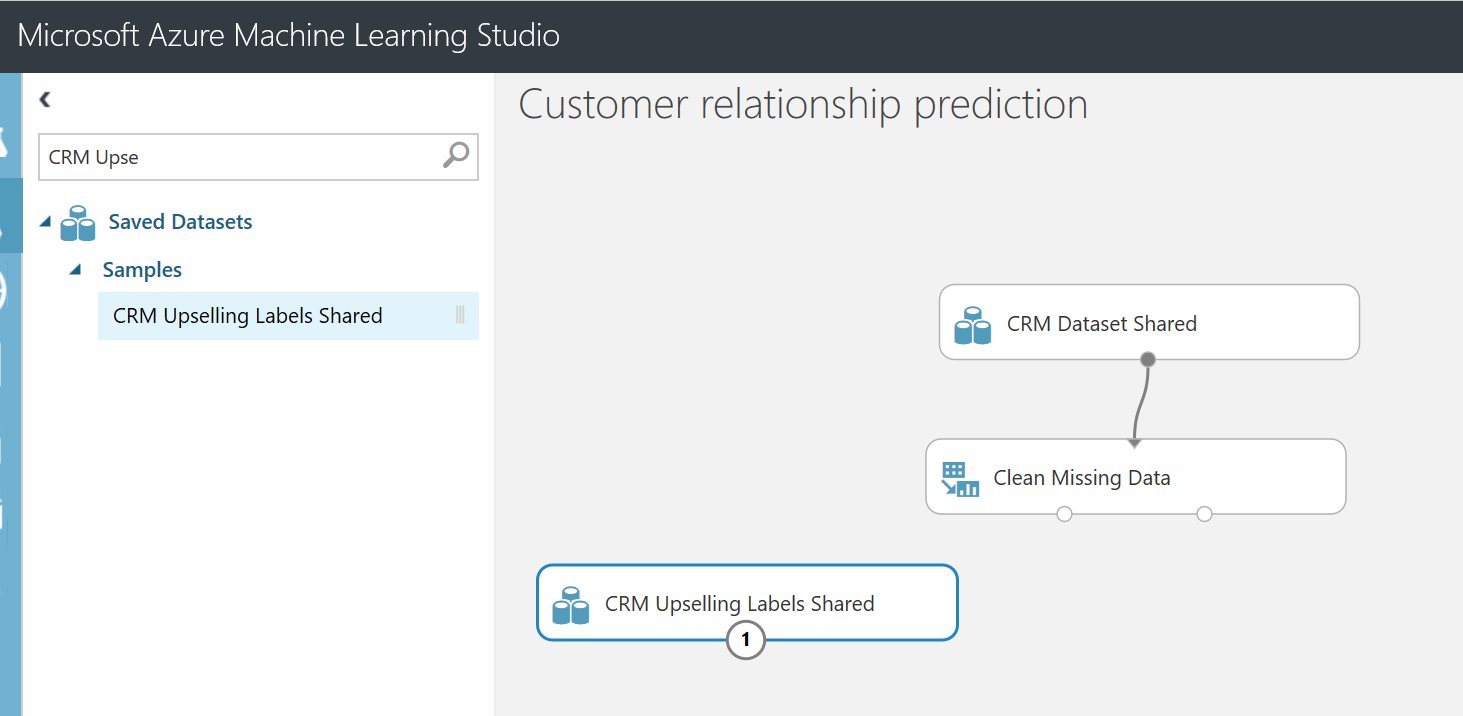
1. Now drag the Clean Missing Data and drop under the CRM Dataset Shared
2. Now connect both together



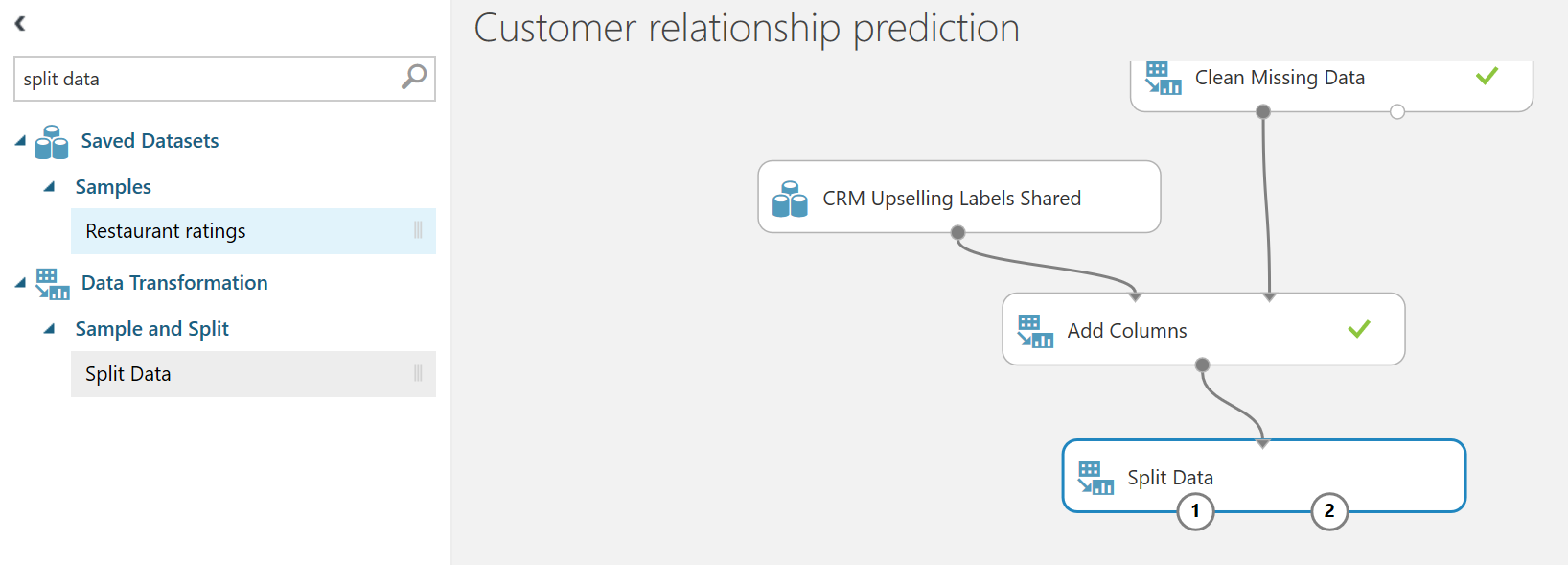
1. Now click on Clean Missing data and to the right you should be able to see properties for the cleaning that we can configure
2. Make sure you have all the data filled like the below picture



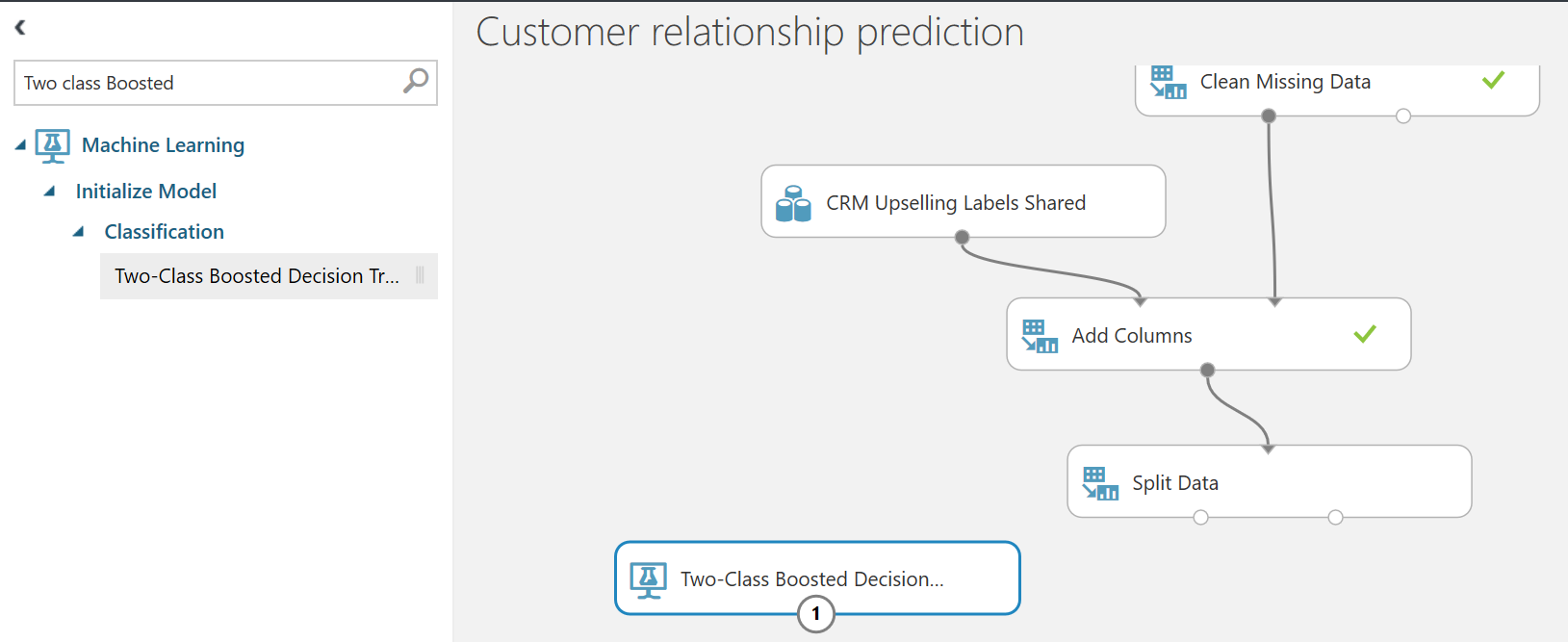
1. Now that we have cleaned the data set lets bring in another data set with some



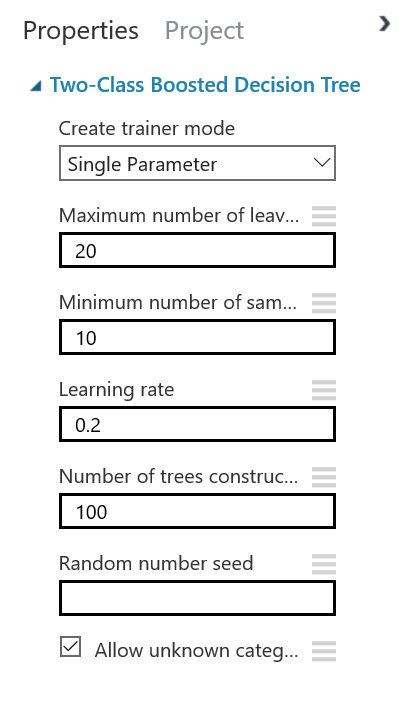
1. Search for CRM Upselling and drag and drop to canvas like above picture
2. The CRM Upselling labels has the labels for the dataset and the CRM Dataset Shared contains the features for the model to train
3. Now we can save the experiment and run it.
4. If you right click the bottom dot or connection in the data set there will be visualize option which allows us to view the sample from the dataset
5. If you check col1 which is label column is added into the dataset which has columns as va1 to var230
6. Now let’s split the data for training and testing. Usually out of 100%, 70% is used for Training and 30% is used for testing



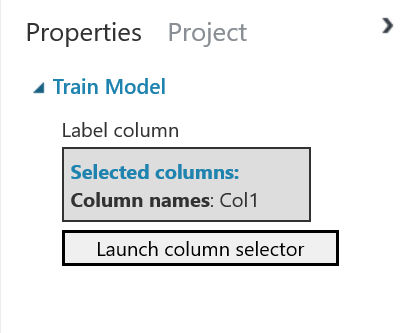
1. You can search for Split data on the search window and drag and drop in the canvas
2. Now let’s bring in the model
3. Search for Two Class Boosted and then drag and drop as per below picture



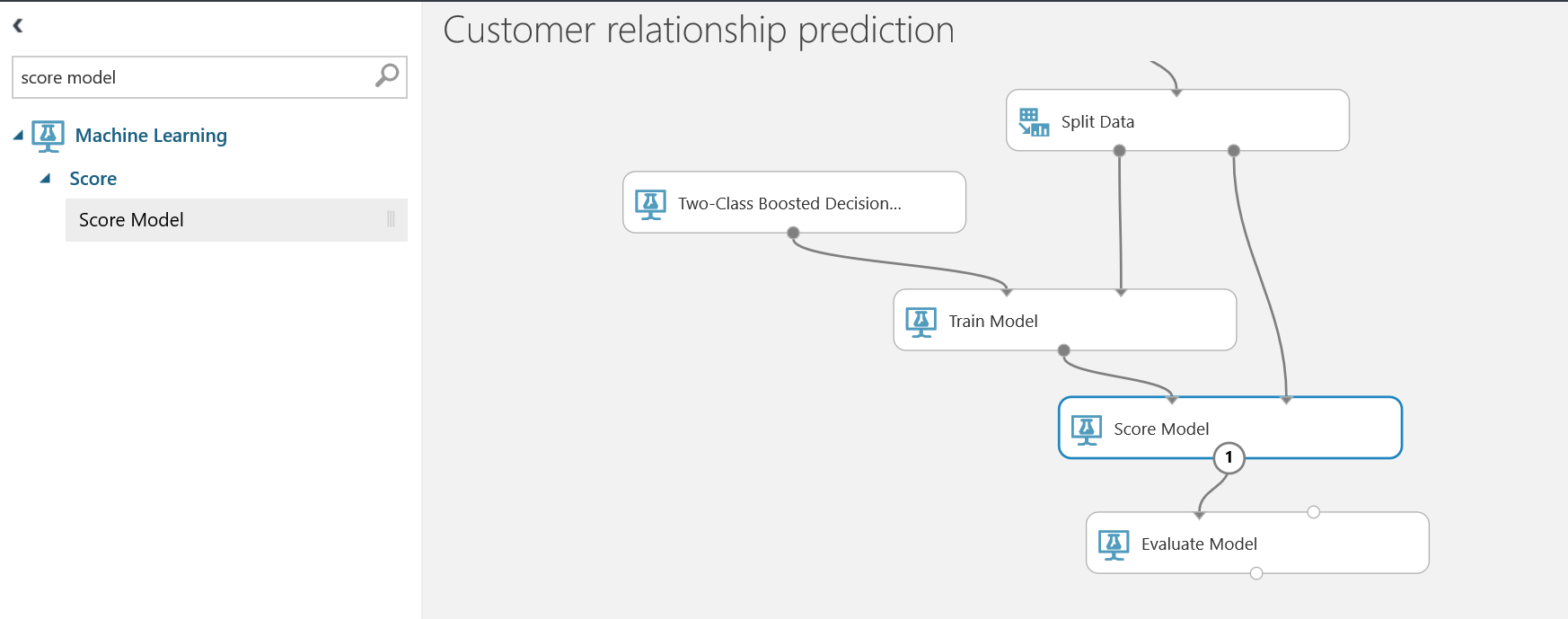
1. Now let’s search and bring Train Module
2. To set the parameters for the algorithm click Two-Class Boosted Decision and on the right side you should see the properties like below



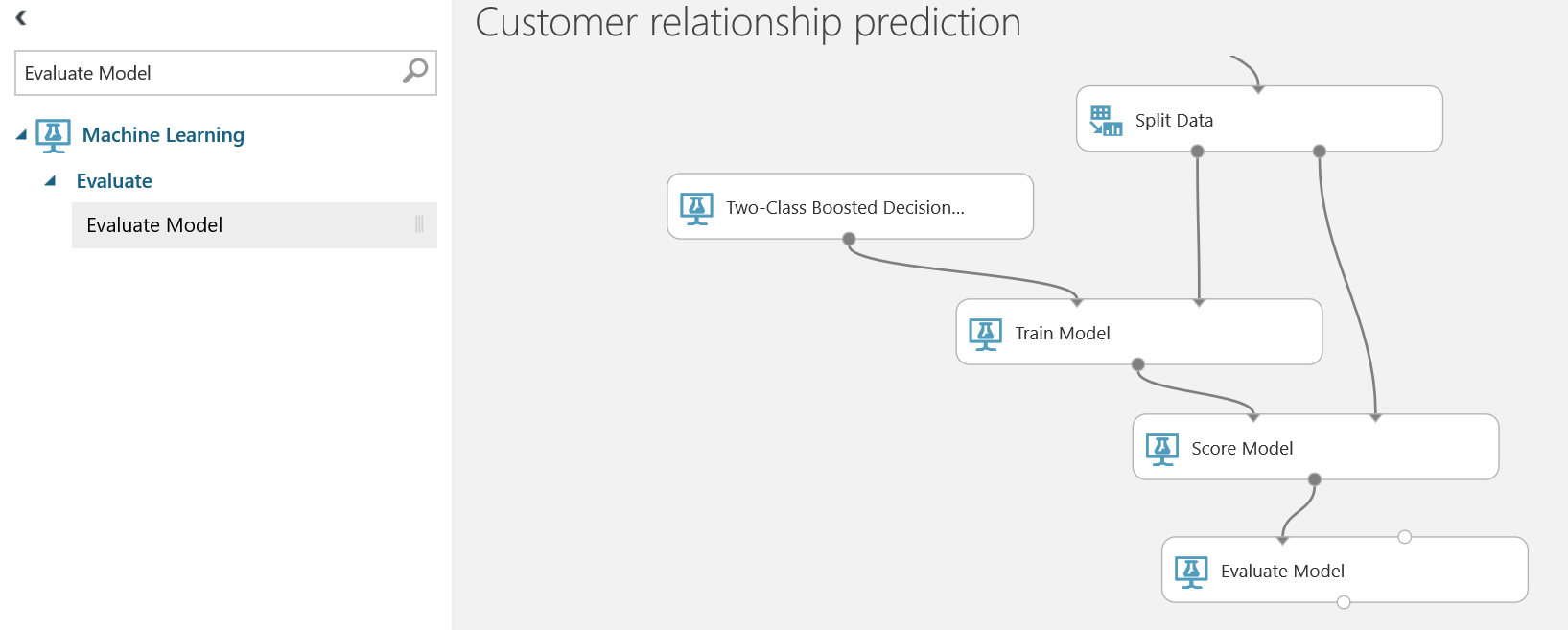
1. Now let’s bring the Train module and score as well
2. Once you got the Train Module connect the train module to Two-class boosted decision algorithm and then the other with split data first output.
3. Now Click the Train module and in the properties set the label column that will be trained for prediction. This is the column that the model will predict



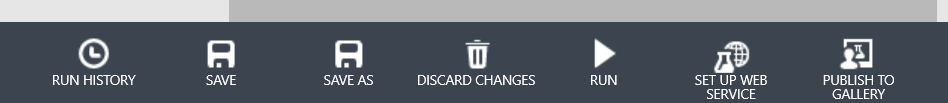
1. Now search and drag the score model



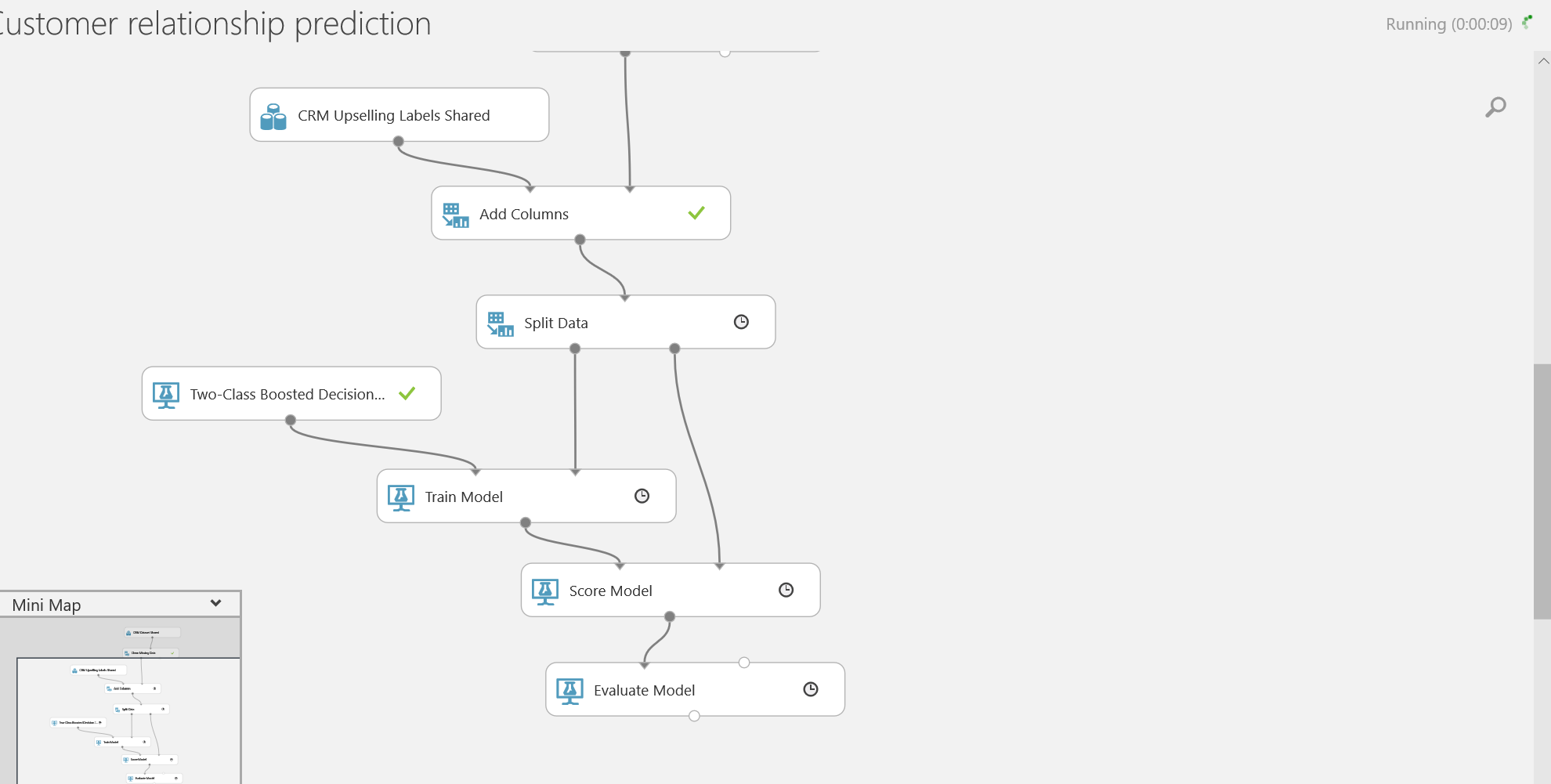
1. Connect the score model first input to Train model and the second input to the second output of split data into second input of score model
2. Now bring the Evaluate model and drop it in canvas



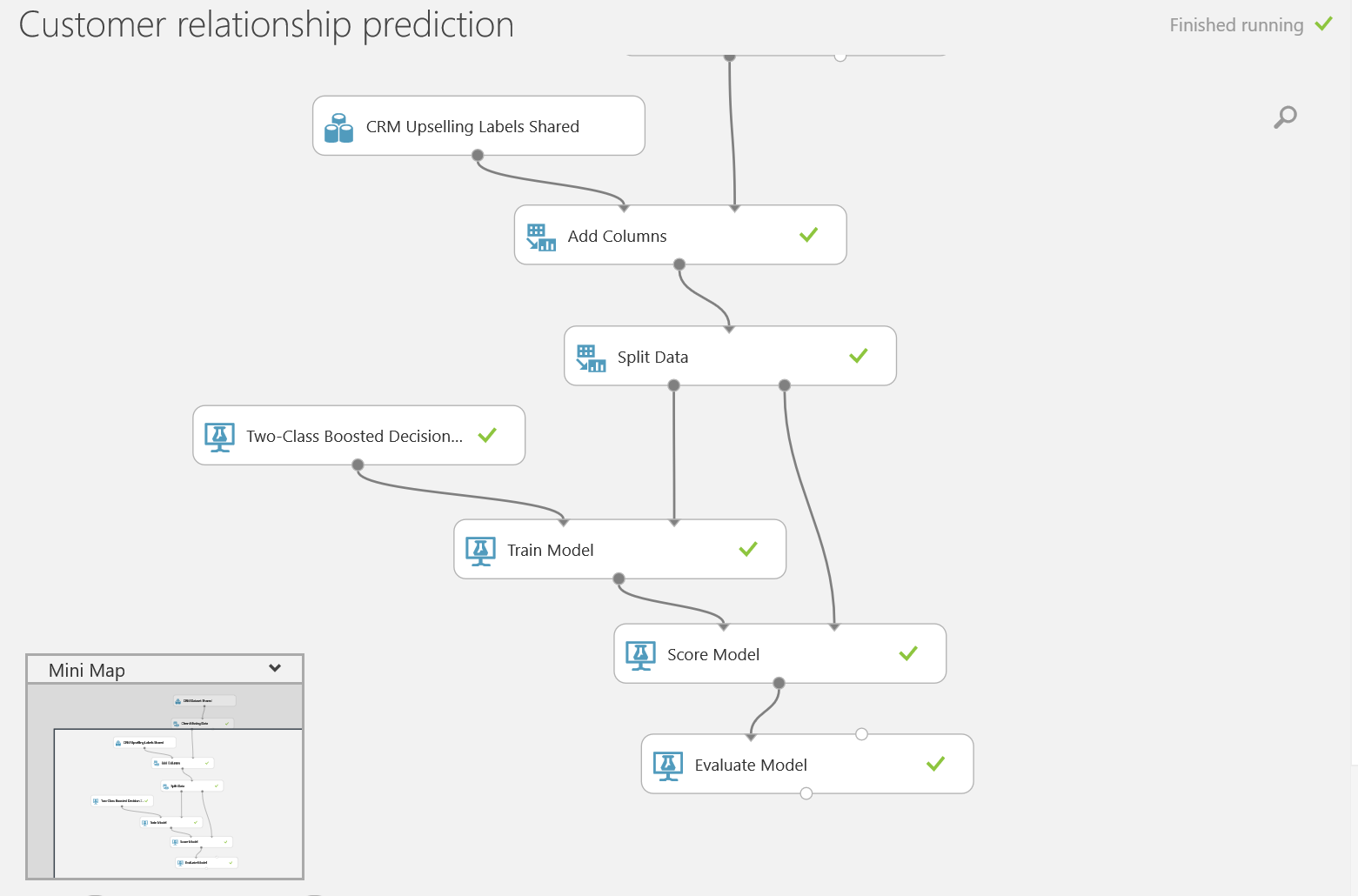
1. Connect the output from Score Model to Evaluate Model.
2. Save the model and then Run
3. Save and Run are both in the bottom of the IDE canvas



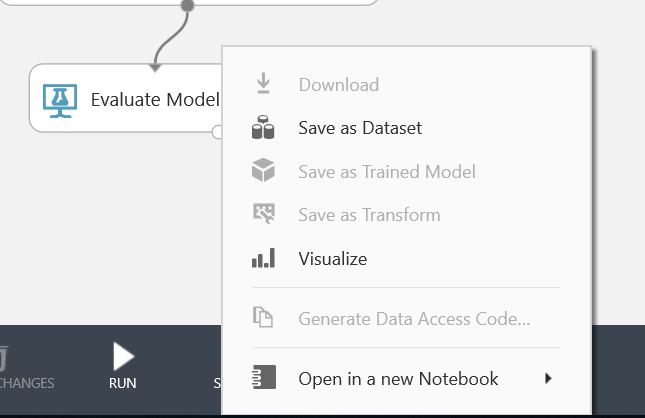
1. Click Run and then select Run.
2. Now the model should train and test the model



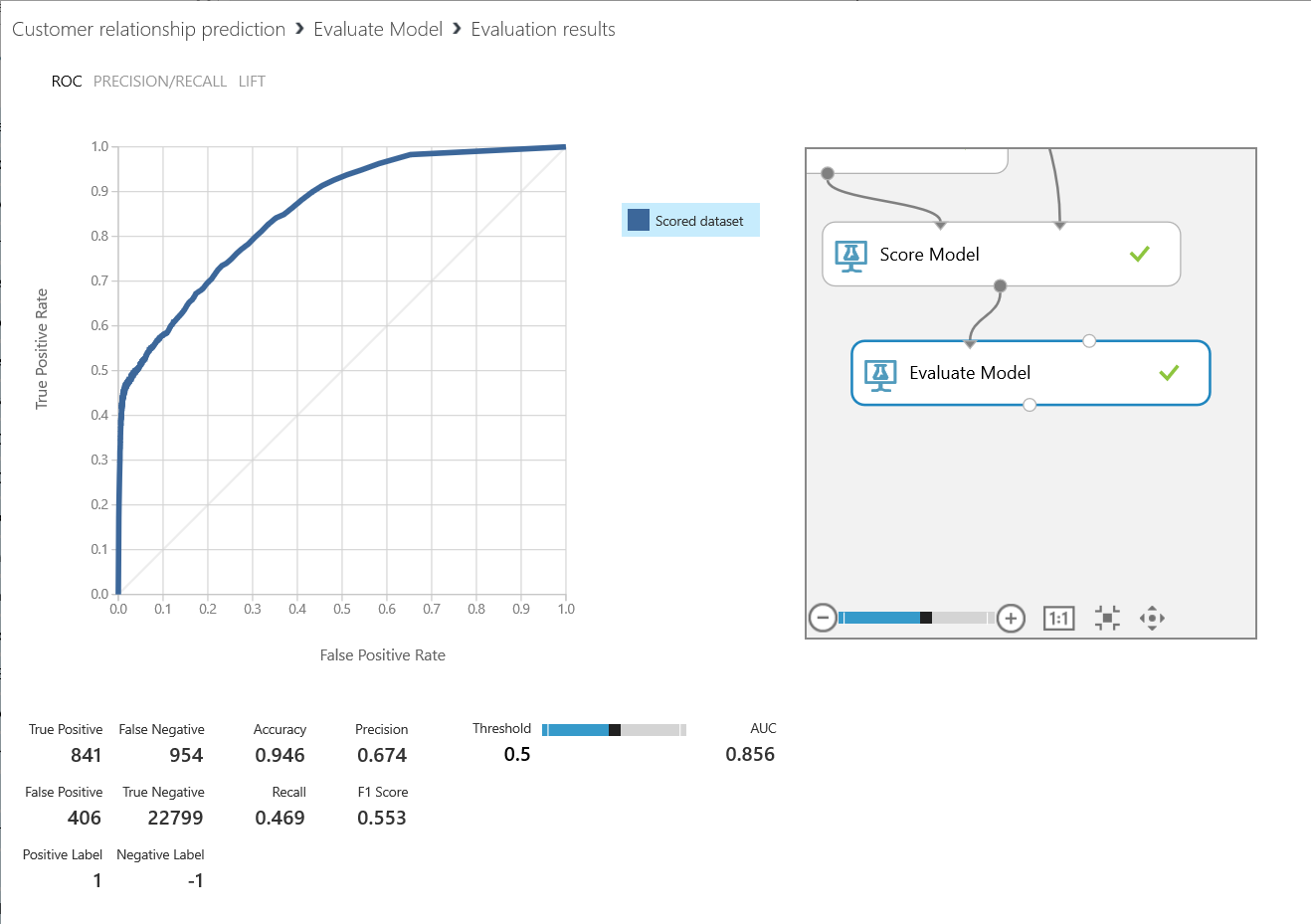
1. Once the model is run then we are going to check the output metrics of the model to see how the model performed.
2. The Model will output the accuracy or RUC and the predicted output with test data set to validate and see
3. Once the model shows finished in the upper right corner of center canvas



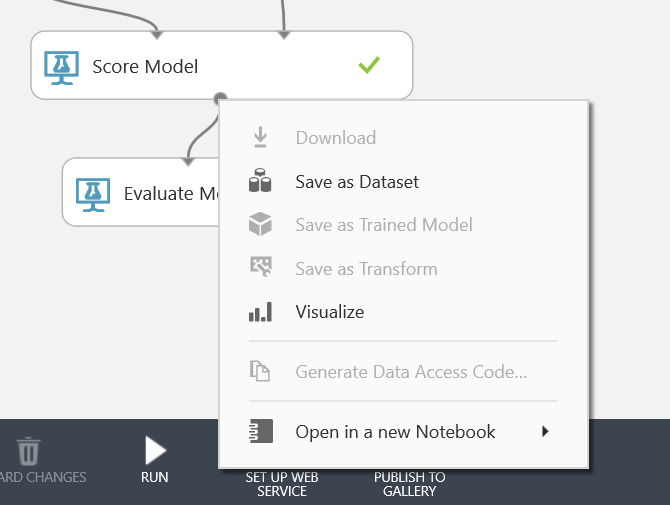
1. Now Click on the Bottom round of Evaluate Model



1. Click on Visulize
2. The output should be something like below



1. The left chart is ROC and shows the precision/recall,
2. The confusion matrix is show cased below as well.
3. The AUC is also displayed which shows almost 85% accurate is the model
4. To look at the predicted data go to Score Model and right click the bottom connector and then click visualize



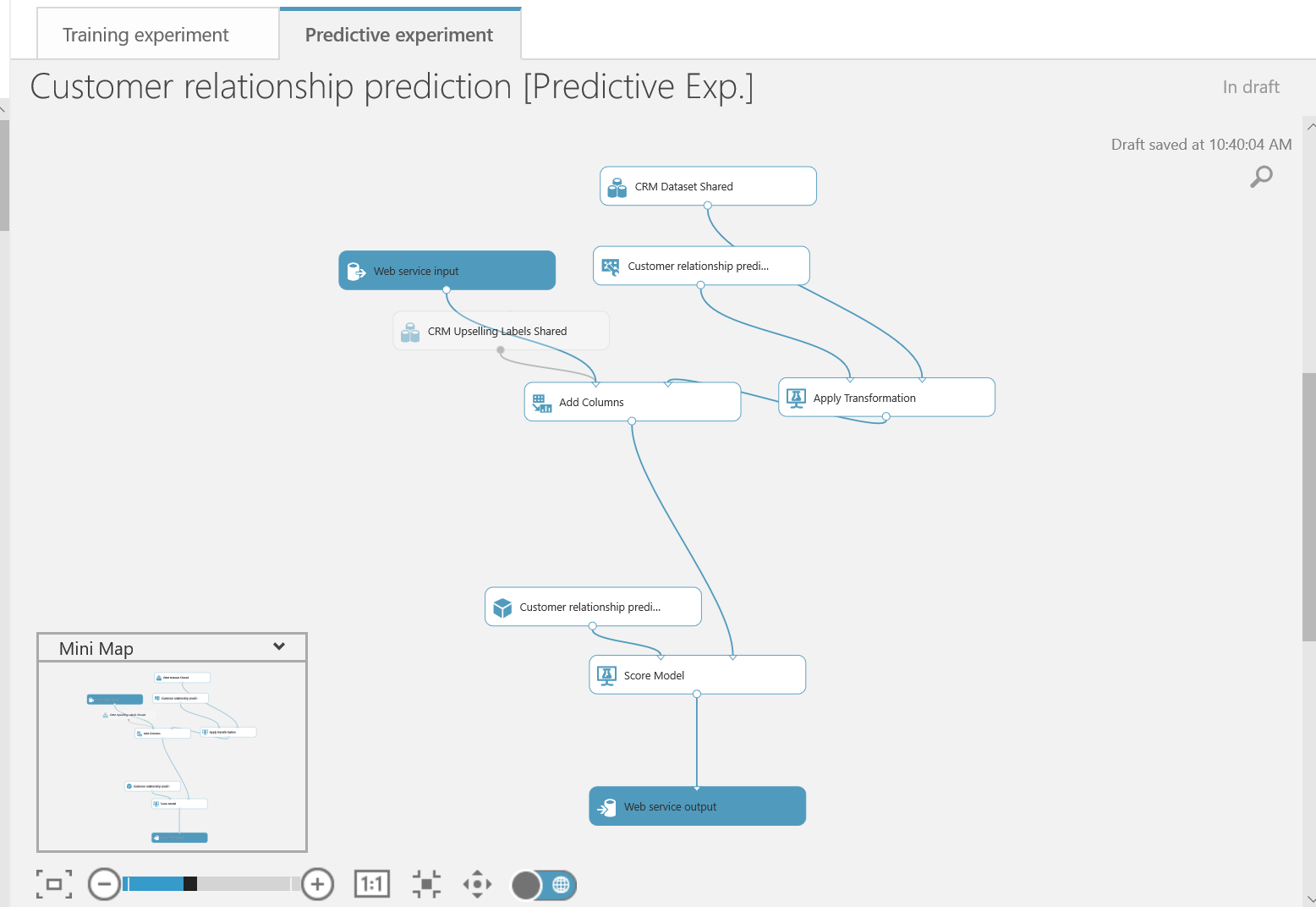
1. The dataset will display with predicted output label as well
2. If you have too many columns in the data set you might not see all the column.
3. To augment for not seeing all the columns we can write the data set back into blob or Azure SQL data base and then we can visualize the data.

# Deploy Model

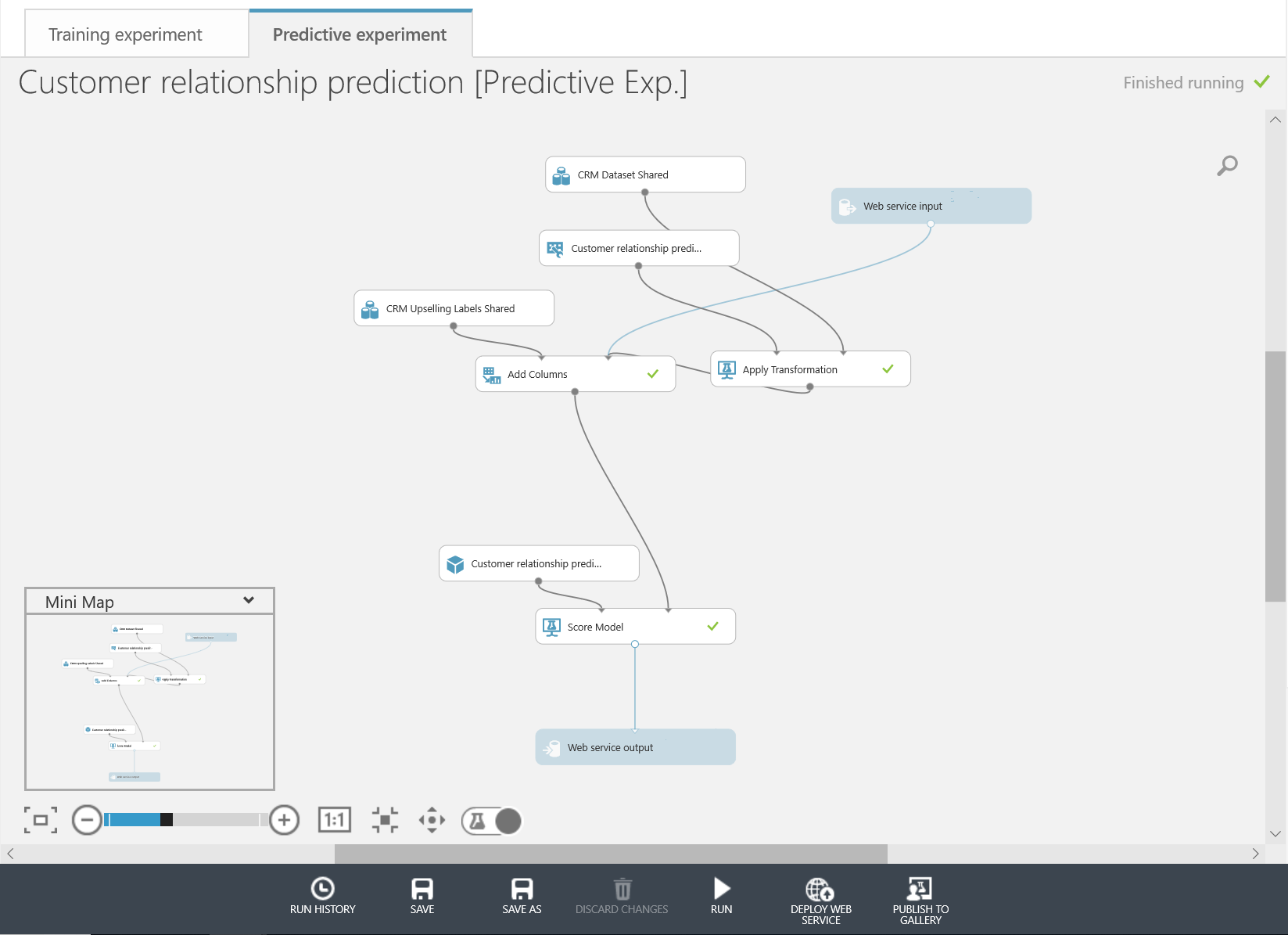
1. From the same experiment we did above can be deployed as web service.
2. In the Bottom where you see Set up Web service click that.
3. Then select predictive web service and now you should see the experiment rearranges the experiment to make it a rest service.
4. Select Predictive web service (recommend)



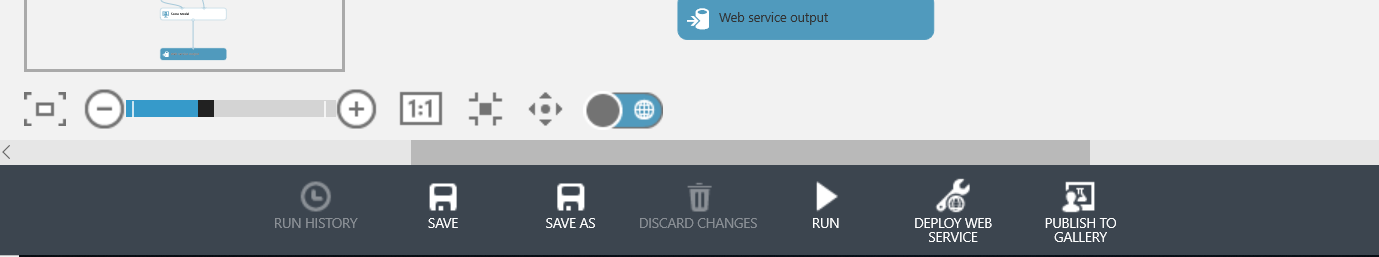
1. Once the predictive experiment is created then save the run the new predictive experiment before starting the deploying process



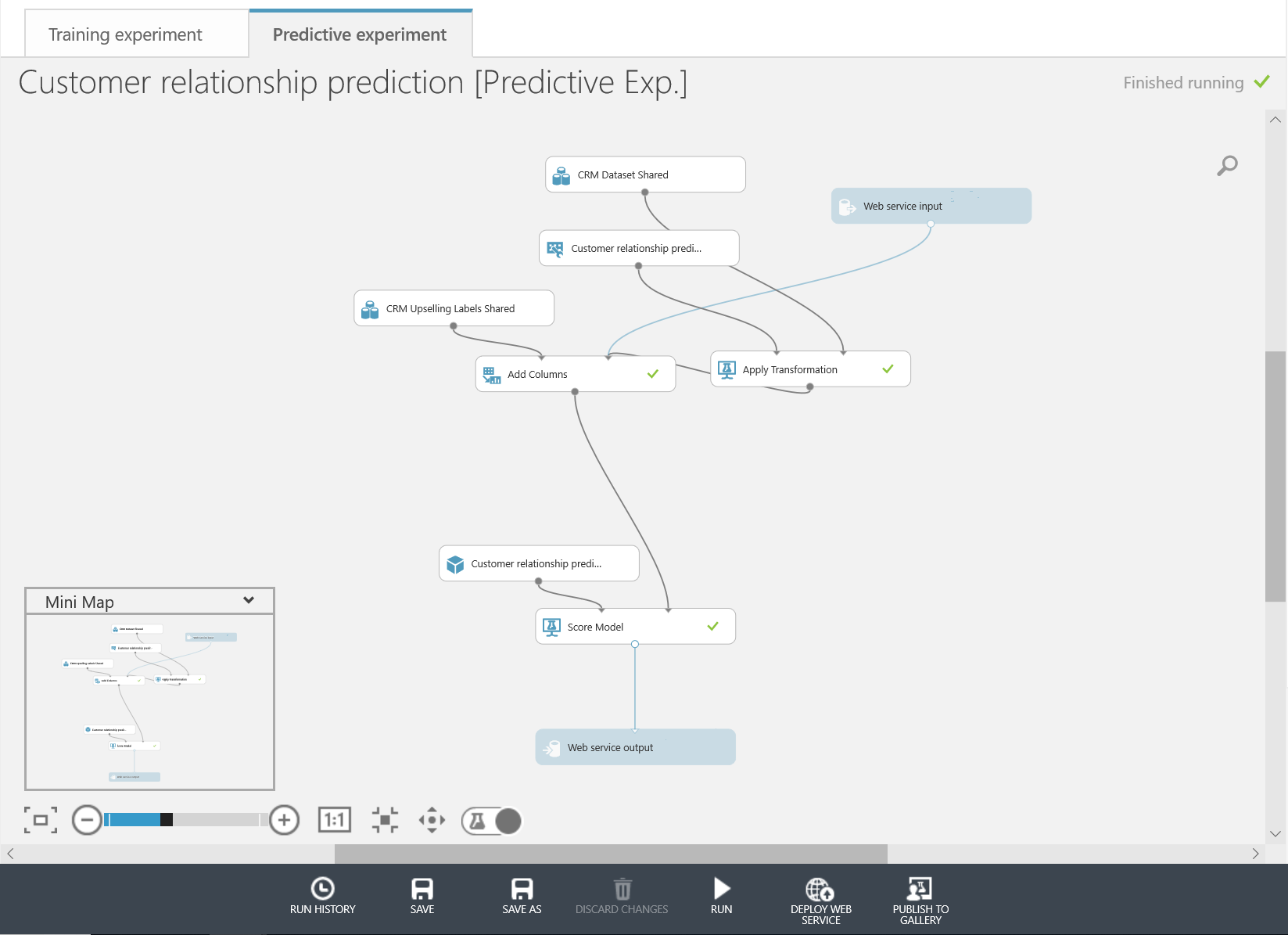
1. Delete the web service input from first input column in add columns to the second one



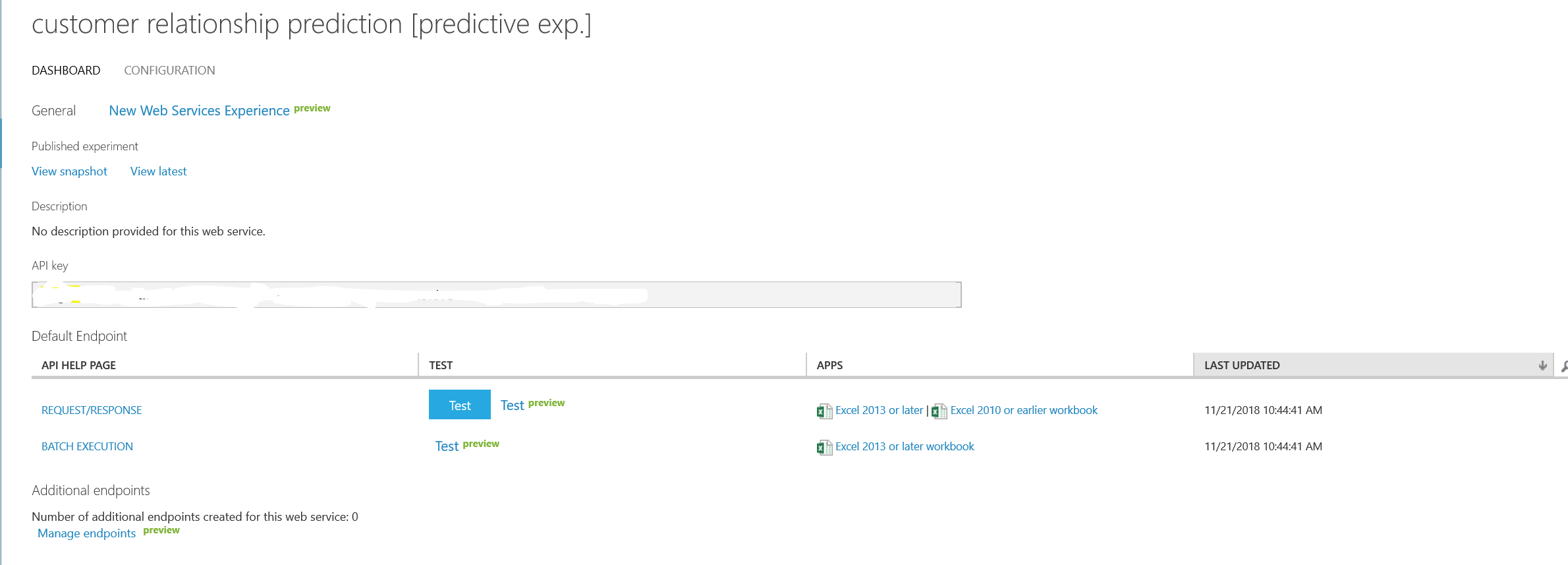
1. Because we have to send the features to get the label predicted.
2. In the bottom click Save and then Run



1. Once the experiment is saved and run like below



1. Click Deploy Web Service (Classic)
2. Now you should see the Test page for web service



1. There is also option to download a sample excel with ML plugin to test the model using Excel as well.
2. Download the excel.
3. Test the solution as well
4. The Excel application can produce sample data set as well.
5. Select sample dataset and then make sure select the input cell range for input data and specify the output cell where the predicted value will be displayed.
6. The other option is to go to test in web service and click the test button and test it there as well.
7. A

# Q&A

# Conclusion

Have fun in exploring more with different use case.

# Next Steps

I want everyone to find some simple use case and try the modelling end to end to give us feedback.