**Build model using IBM Watson Machine Learning**

In this how-to, you will go through the process of preparing data and building a predictive model using IBM Watson Machine Learning Modeler.

The business use case for this how-to is to assess the risk of a loan application, hence, approving the application and giving the loan to a customer or rejecting the application.  
The dataset used in this how-to is free, open-source and available here: <http://bml.io/W2SpyF>

https://bigml.com/user/BigML/gallery/dataset/4f89c38f1552686459000033

You can download the file at the following link:

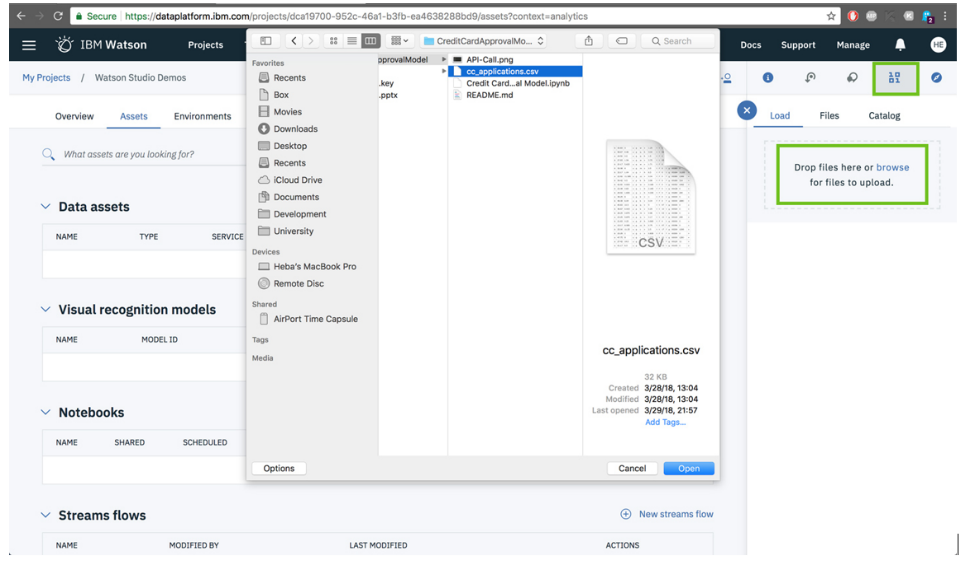
<https://github.com/Danizu/watsonstudioworkshop/blob/master/customers_credit_status.csv>

Perform a “save link as” to save the file on your computer.

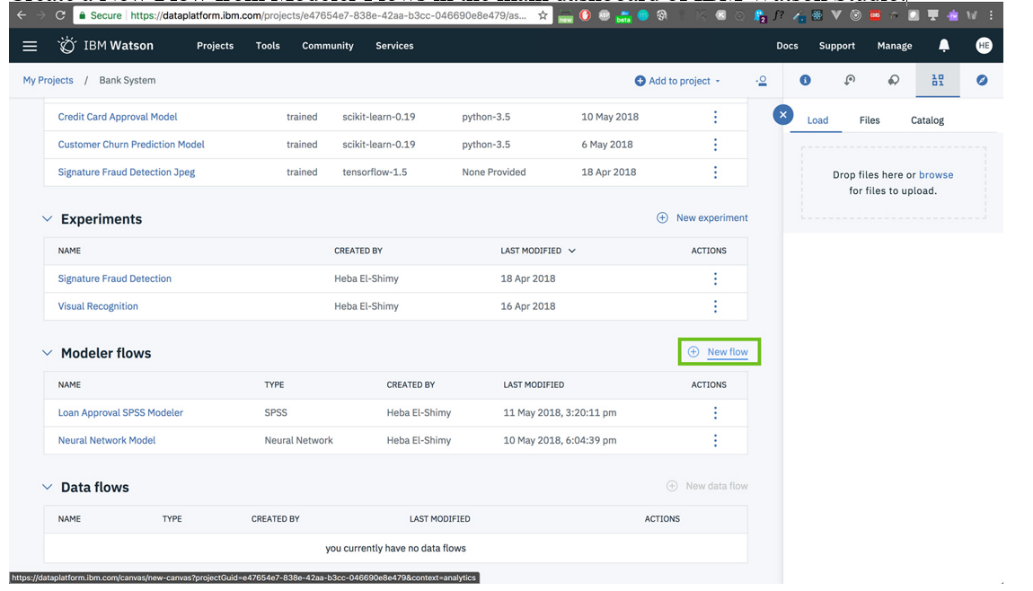
The dataset contains details about customers applying for loans. Examples of the details available are: checking status, duration, credit history, purpose, credit amount, savings status, employment, etc.

IBM Watson Machine Learning is available on [IBM Watson Studio](https://dataplatform.ibm.com/home?context=analytics) as one of many options to build predictive models.

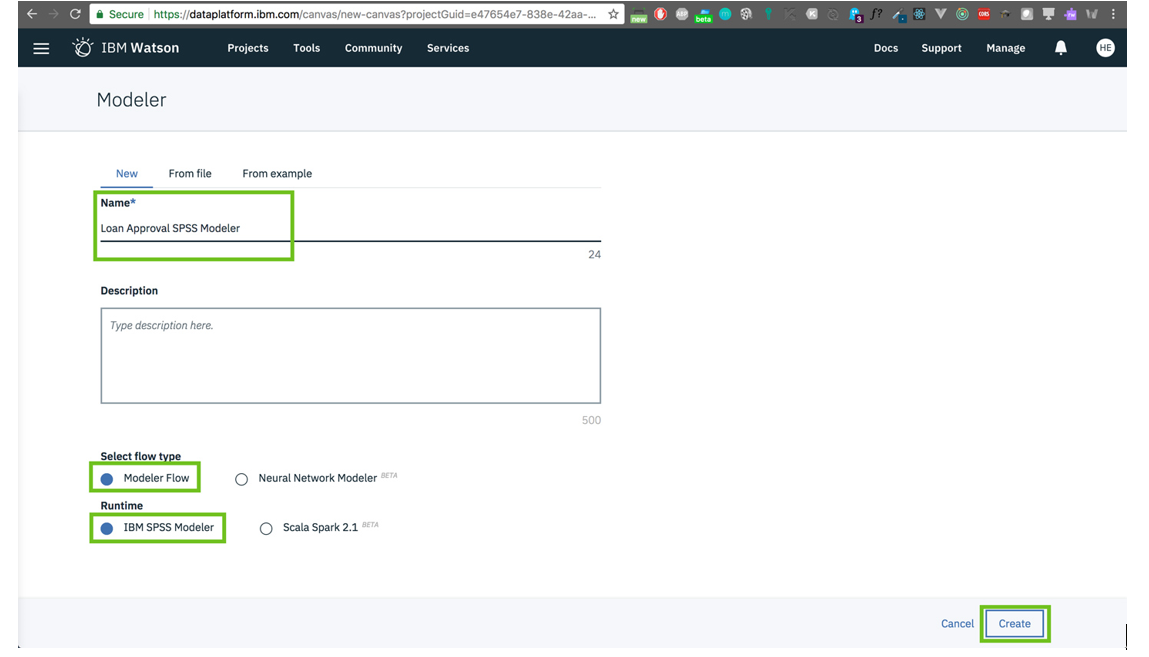
To upload the dataset to IBM Watson Studio and start working with it, refer to the screenshot below. **Note:** The name of the dataset in the screenshot is different.



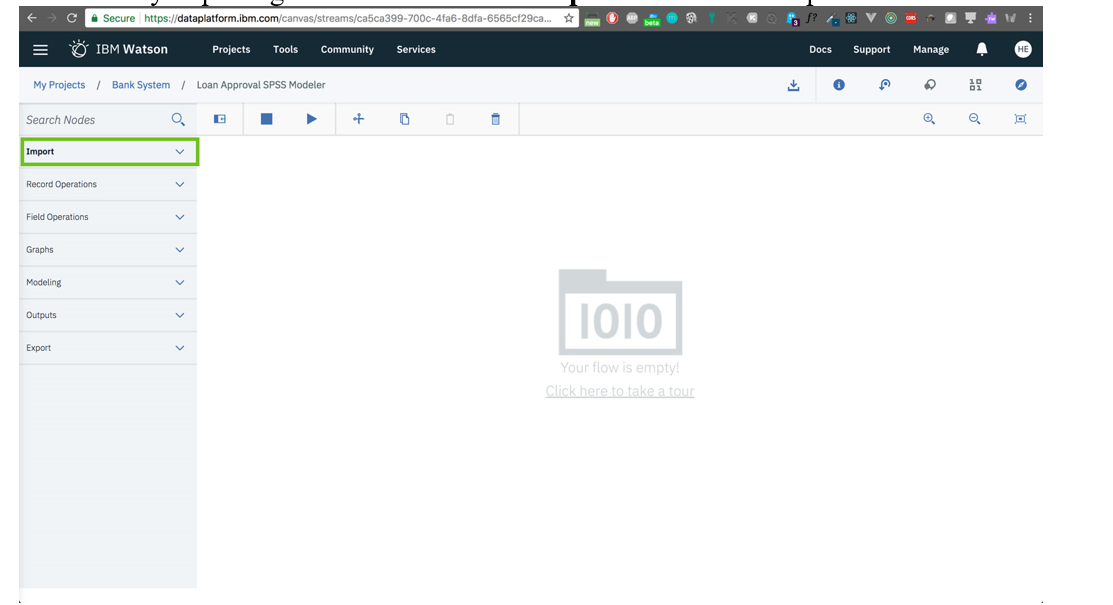
Once the file is uploaded it will show in the Data Asset section of the project

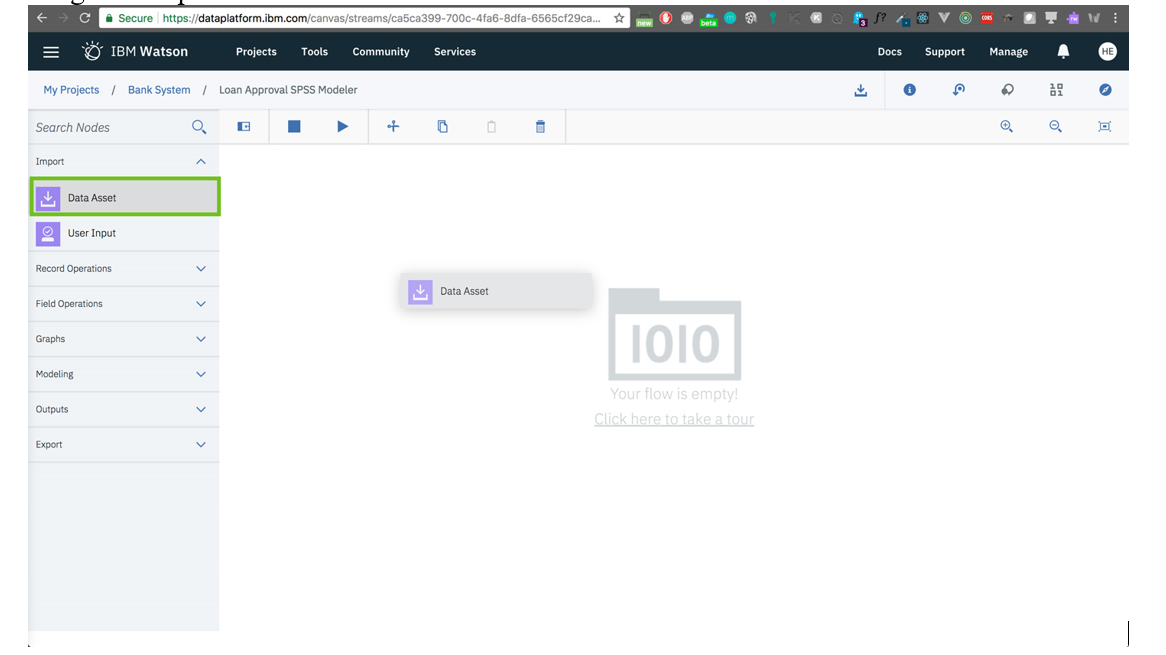
Create a **New Flow** from Modeler Flows in the main dashboard of IBM Watson Studio.  


Type a **Name** for your Modeler task, select **Modeler Flow** as flow type and select **IBM SPSS Modeler** as runtime.

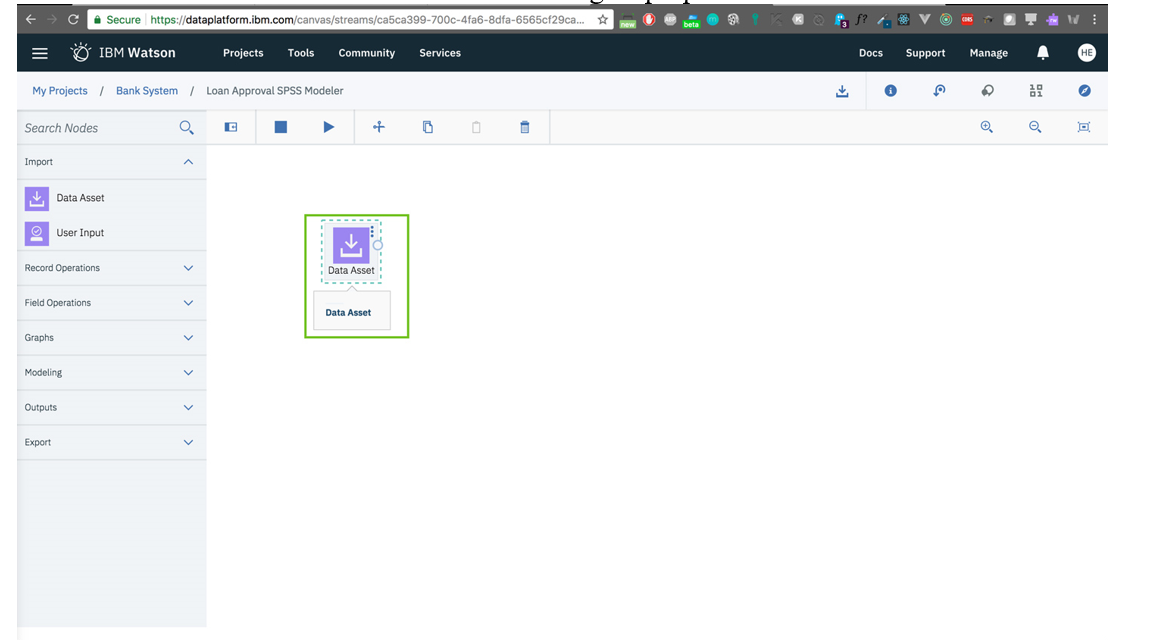


Let's start by importing our dataset. Click on **Import** on the left-side panel.

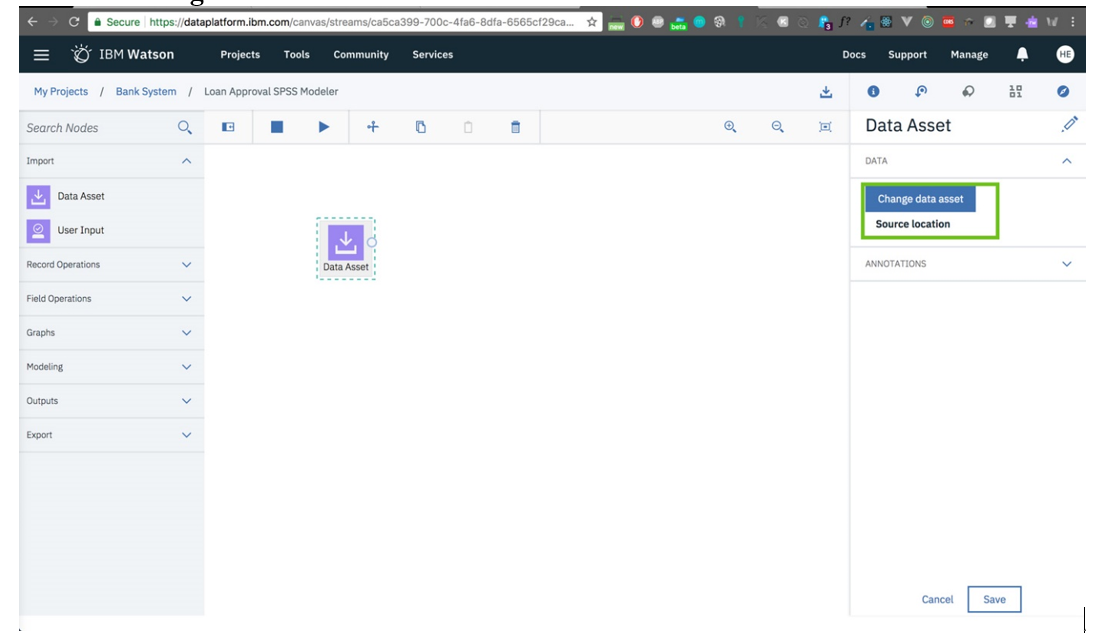


Drag and drop **Data Assets** node into the Modeler canvas.  


Double-click the node on the canvas to start editing its properties.

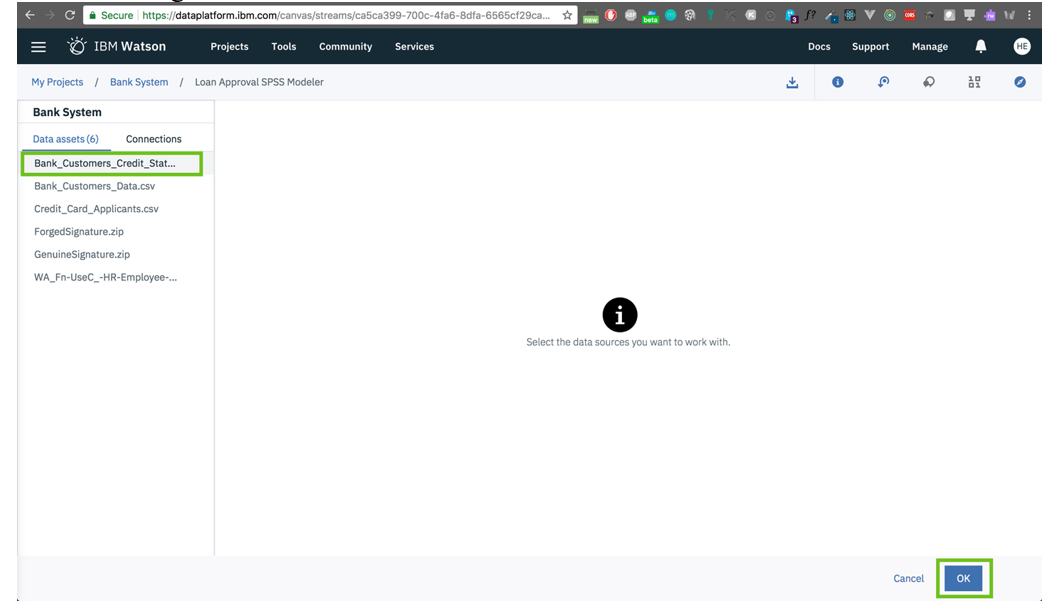


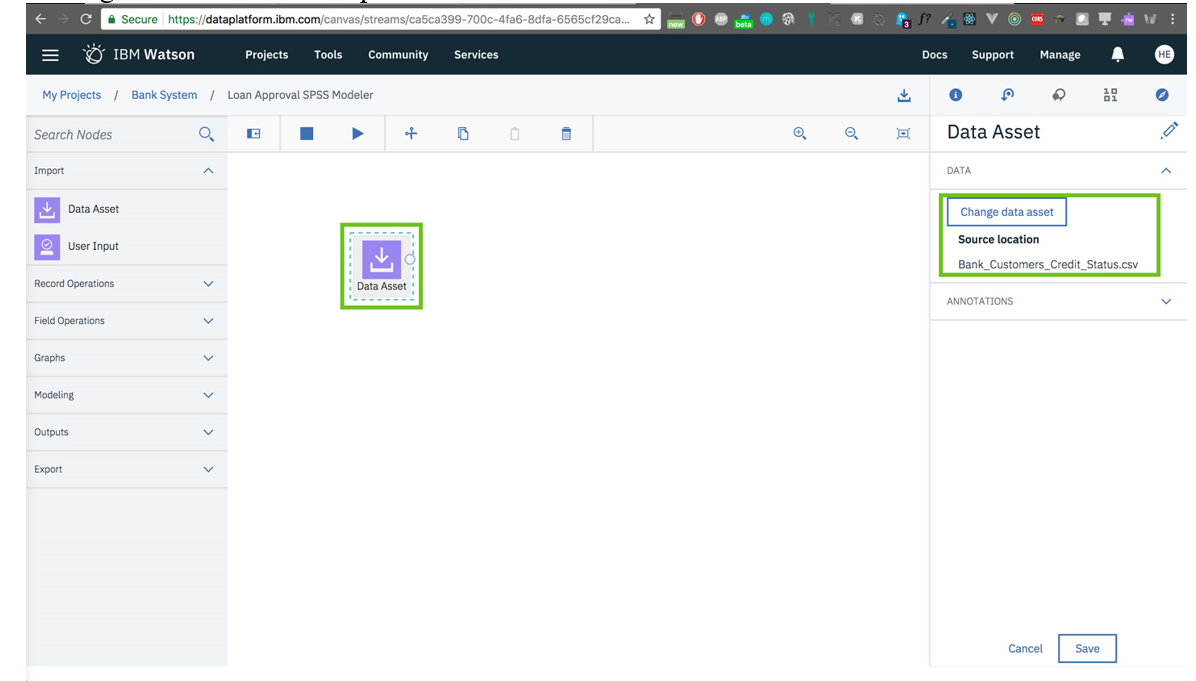
Click on **Change Data Asset** to define the source of our data.



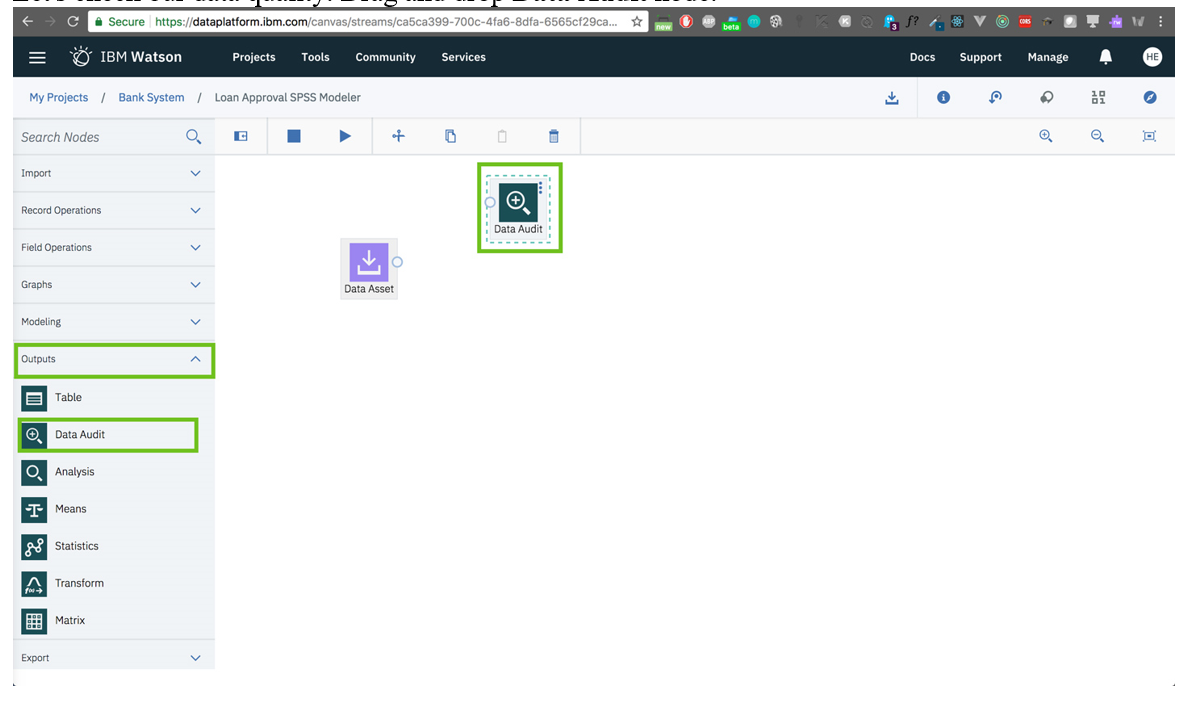
A list of all the data assets you have in this project will be listed. Let's select customers\_credit\_status.csv.

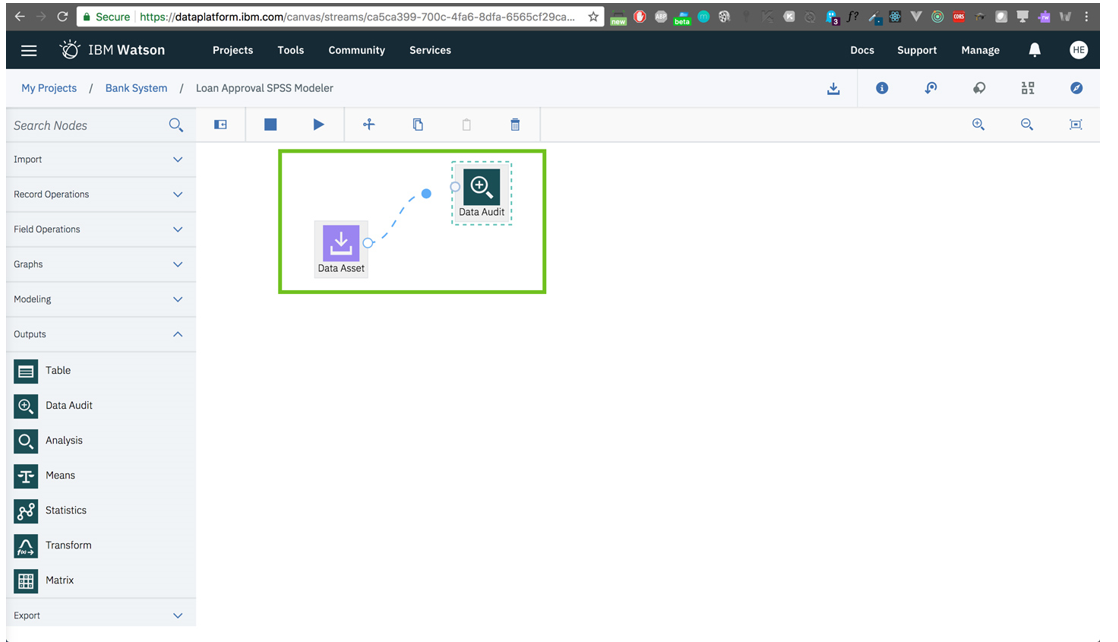
**Note:** Names of datasets in the screenshots may be different from what you have.

After selecting the dataset, click **Ok**.  


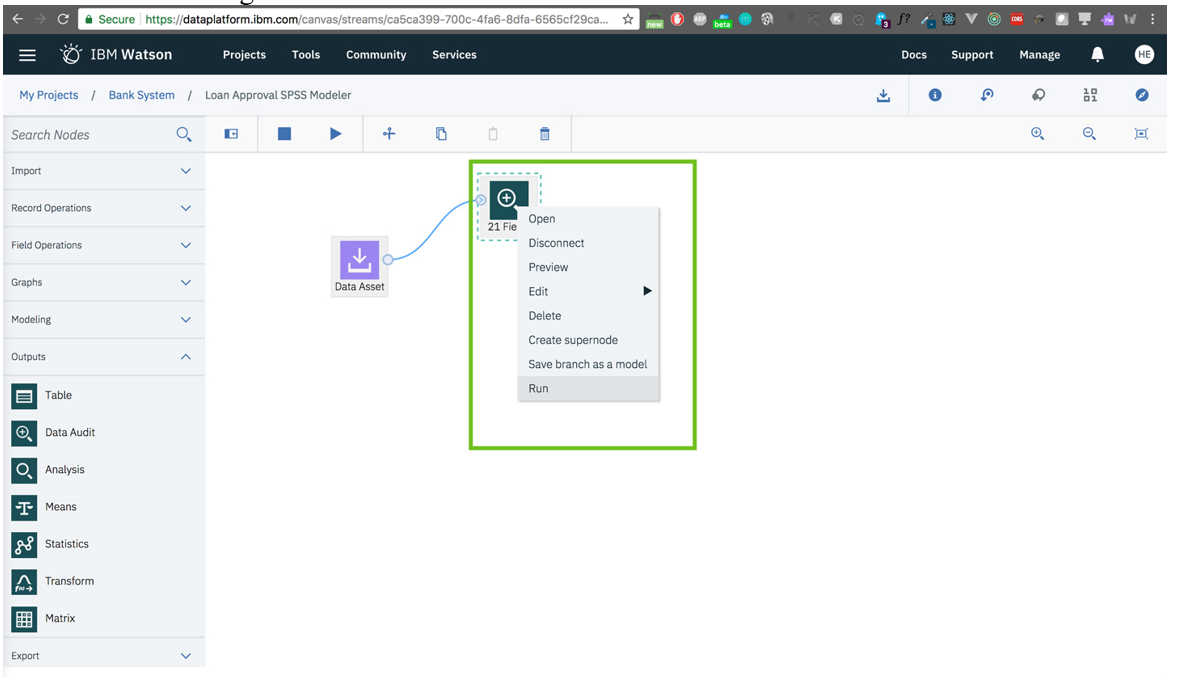
Making sure the dataset is imported. Click **Save**.  


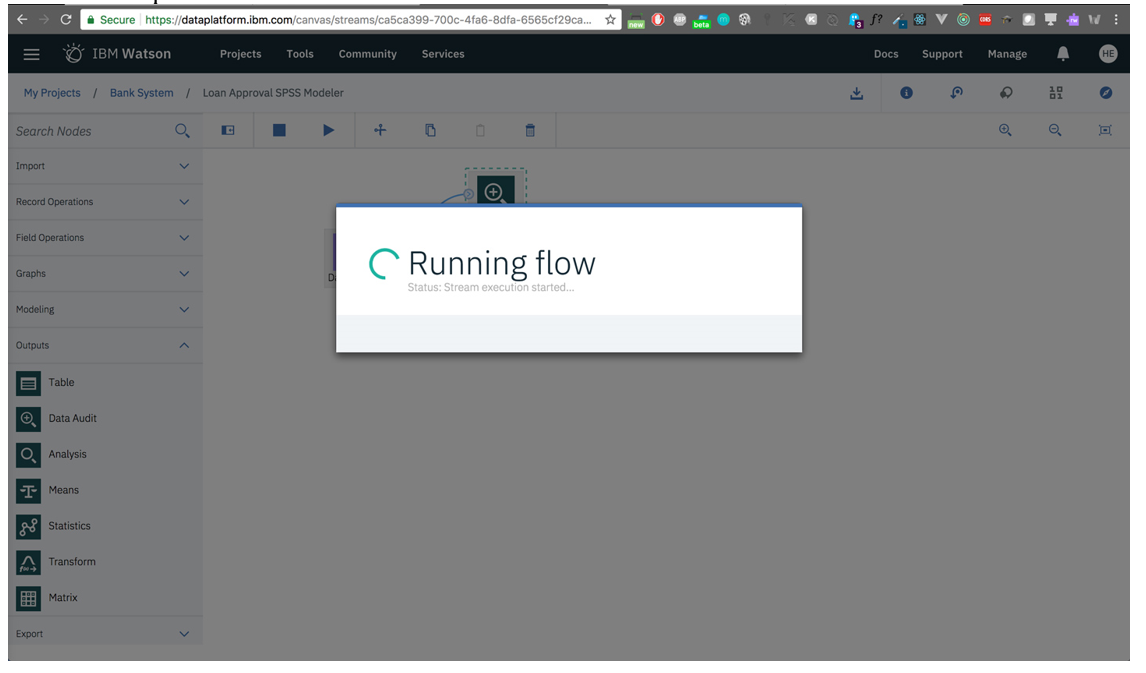
Let's check our data quality. Drag and drop **Data Audit** node.



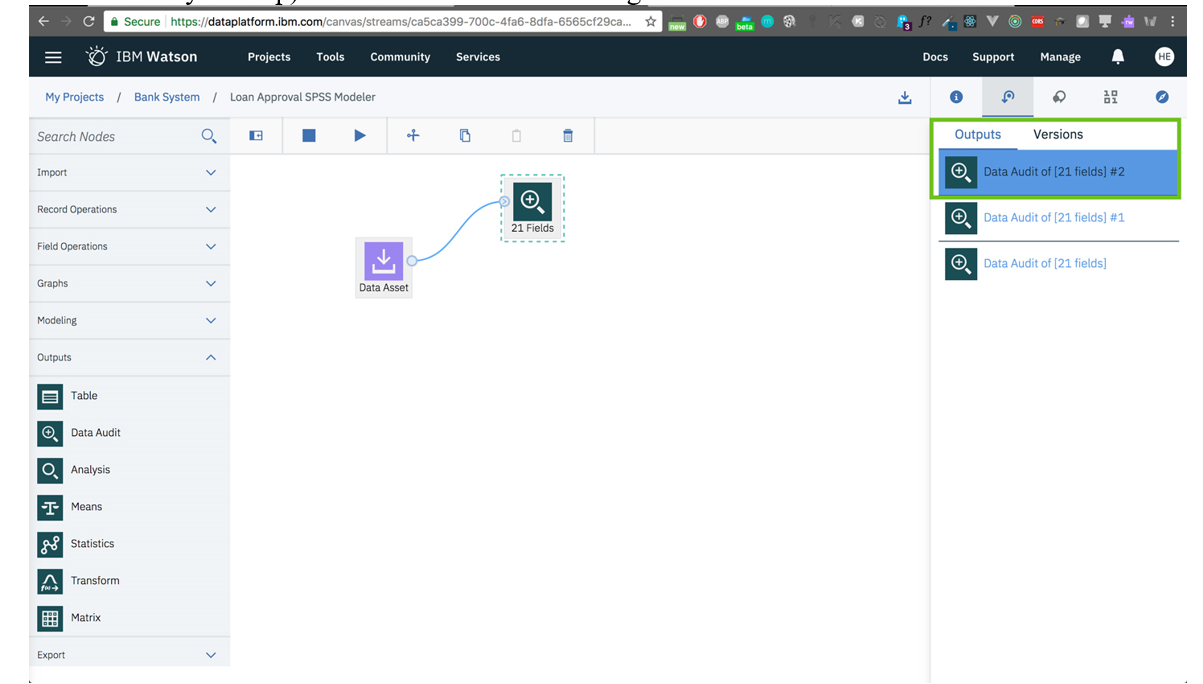
Connect the nodes.  


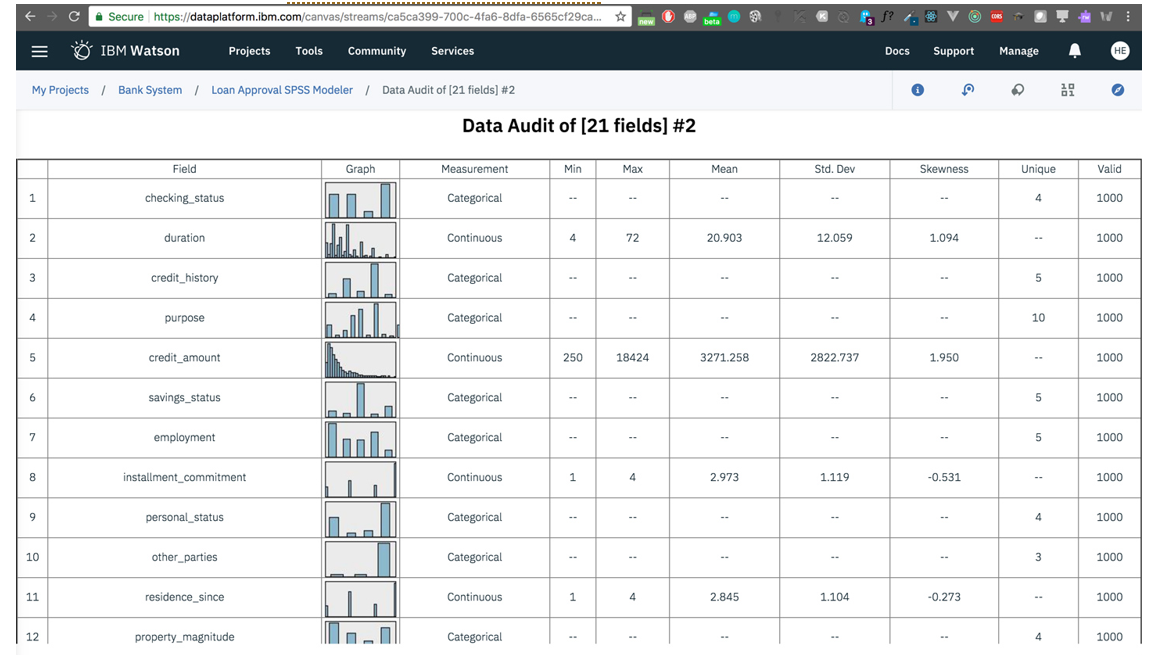
Let's run the flow to get the results of the data audit.

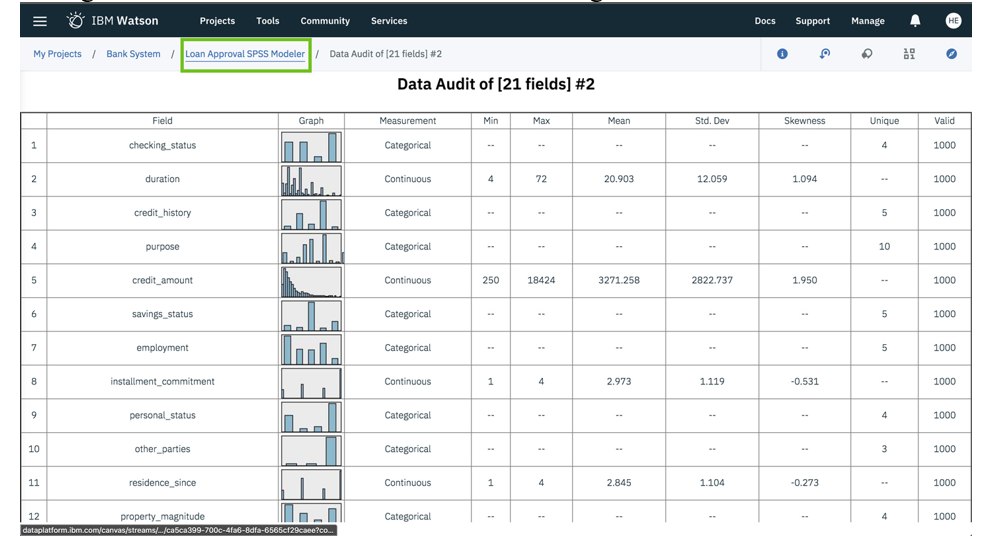


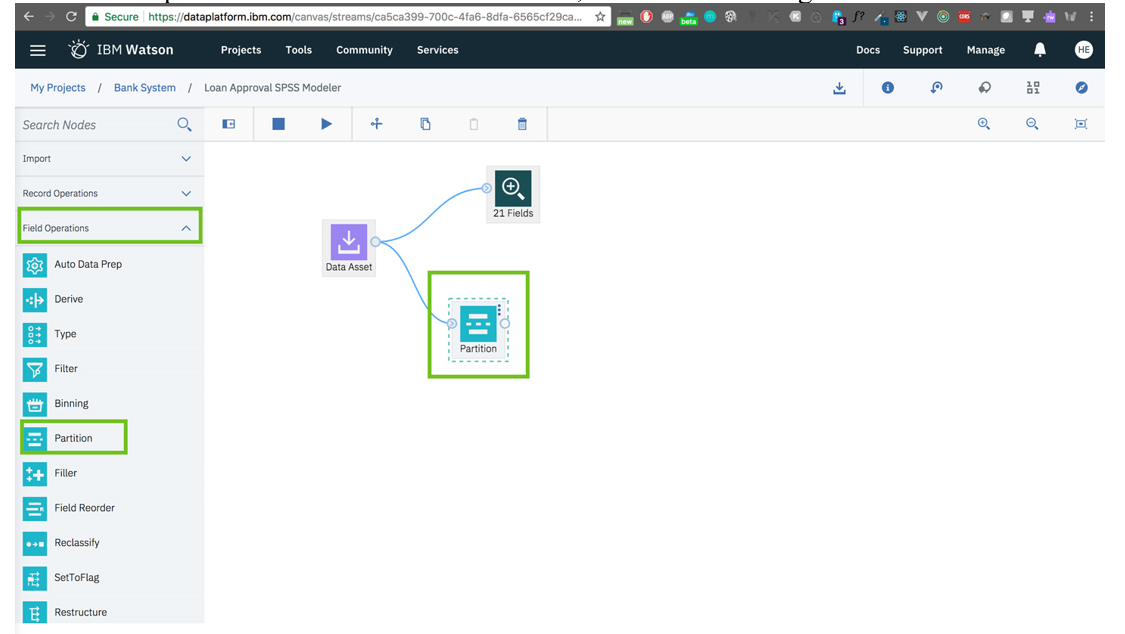
Wait for the process to finish.  


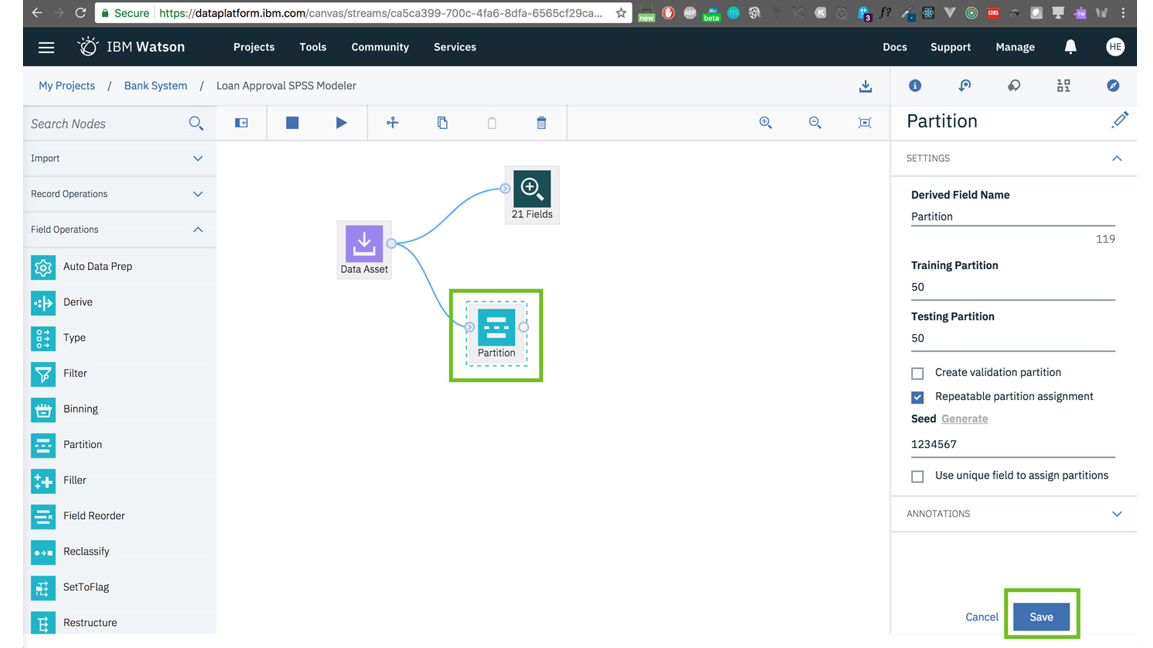
In the right-side panel, under the **Outputs** tab, let's select the most recent output (most recent is always on top) to view the results of running the flow so far.

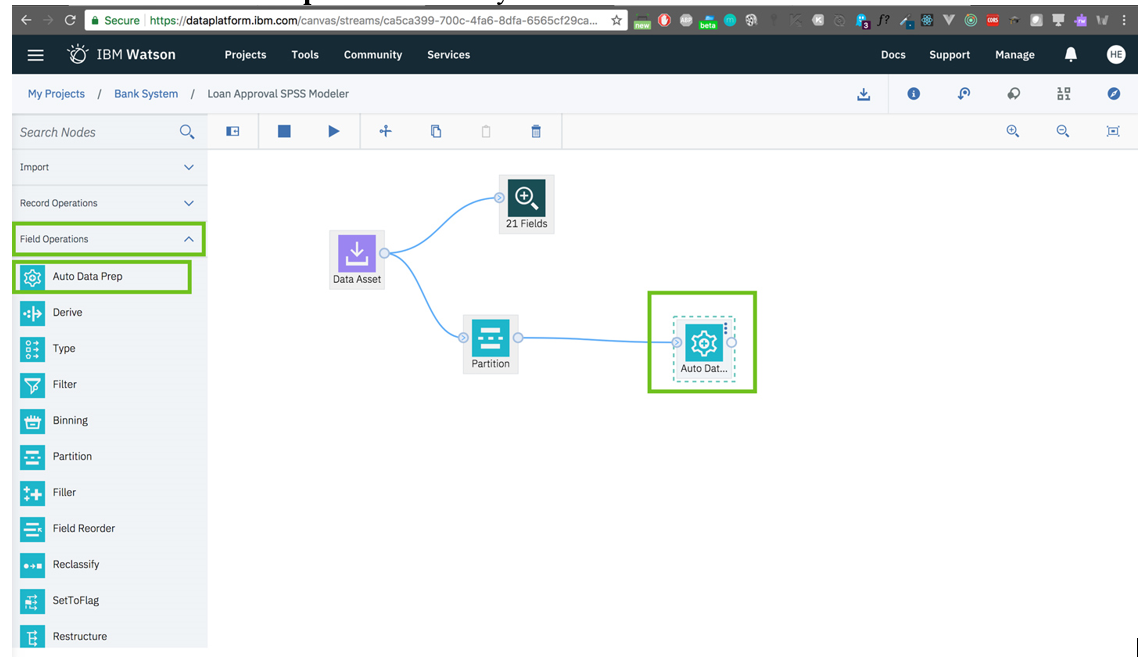


We get an idea about some statistics of the data and the distribution of the features. We find that the data needs some kind of normalization.  


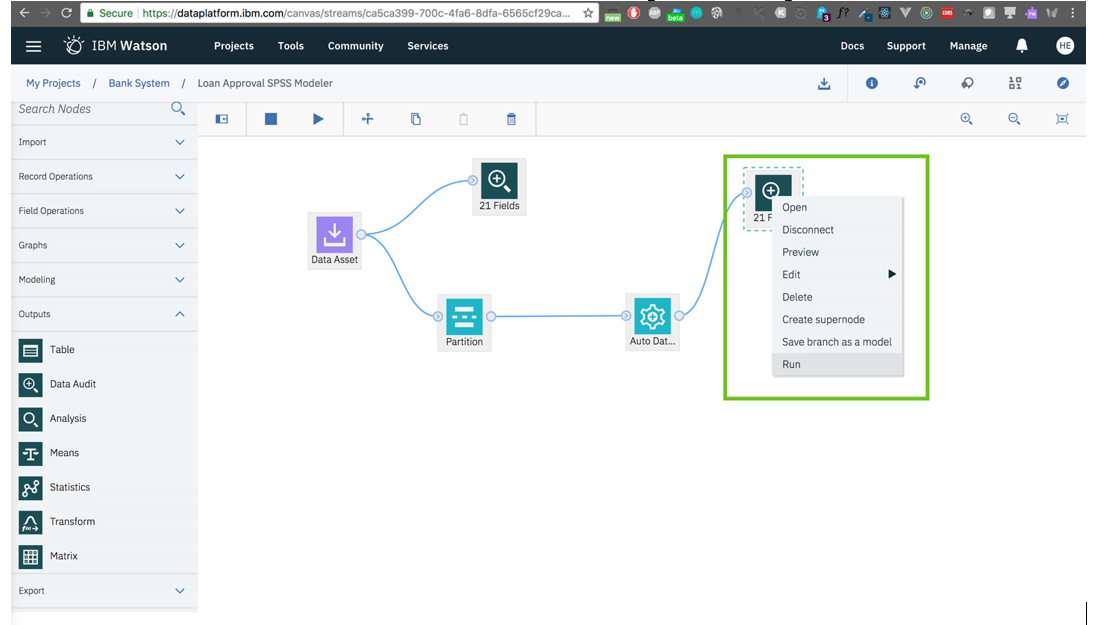
Let's go back to our Modeler and continue working there.  


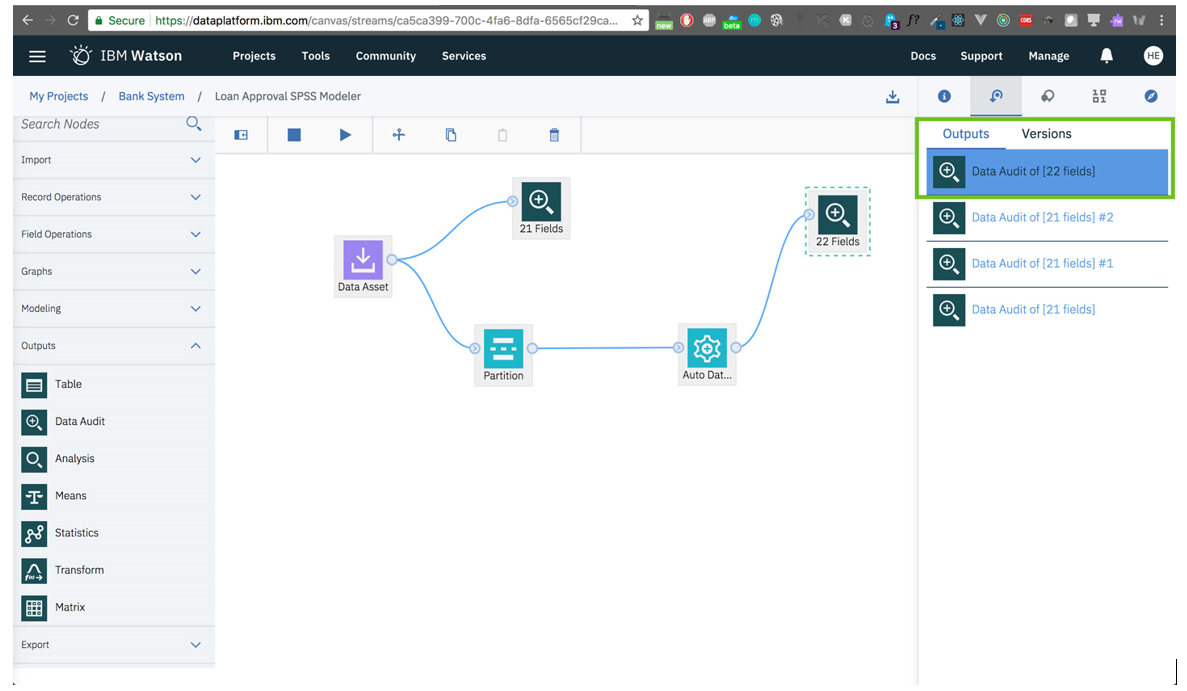
We want to split our data into Train and Test sets, we will do so using the **Partition** node.  


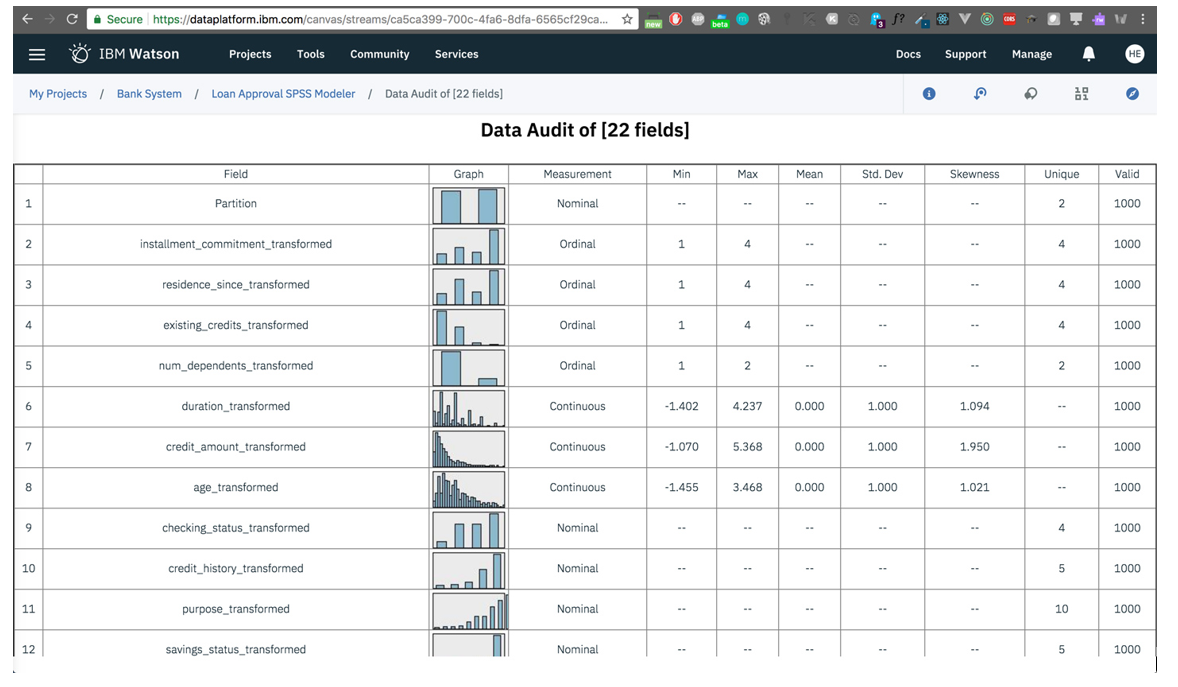
Double-clicking on the **Partition** node to explore its properties. We'll leave everything as it is.  


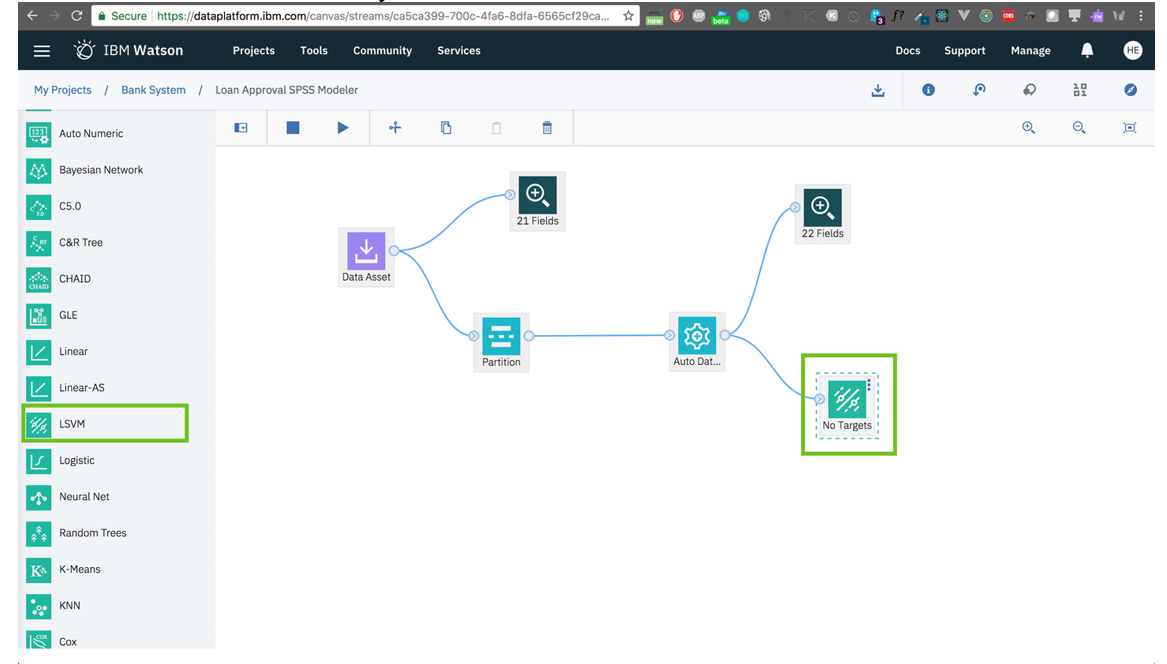
We'll add **Auto Data Prep** to automatically find issues in our dataset and fix it.  


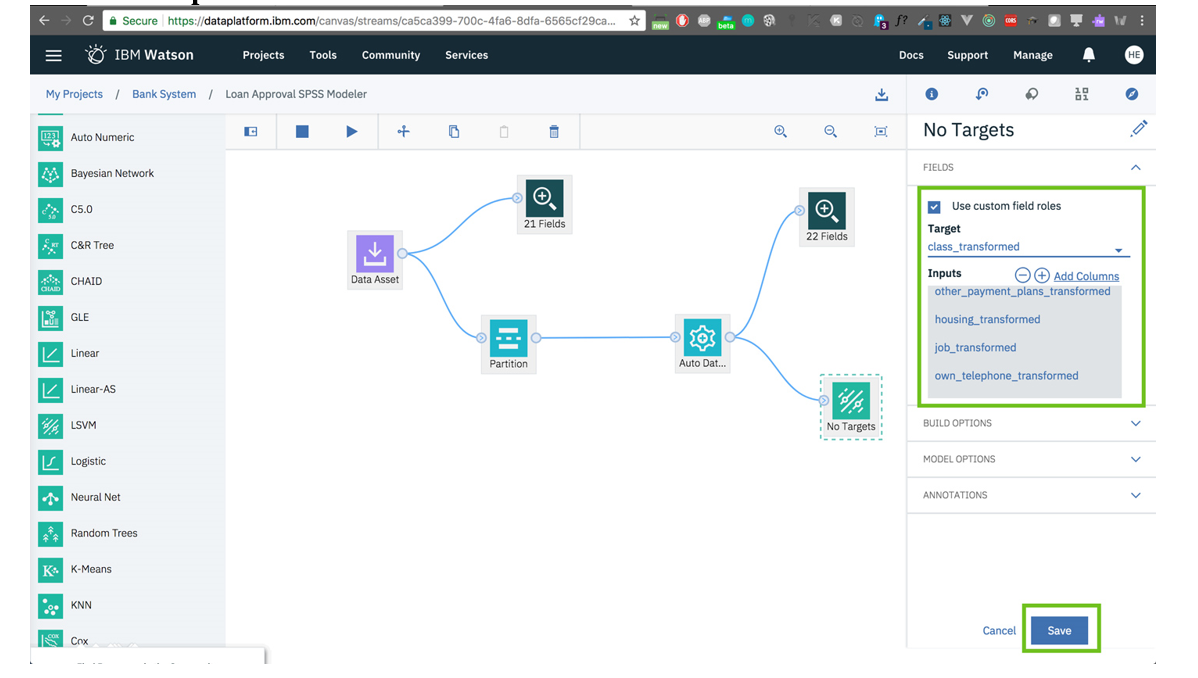
We'll do another **Data Audit** to check the results of the previous step.

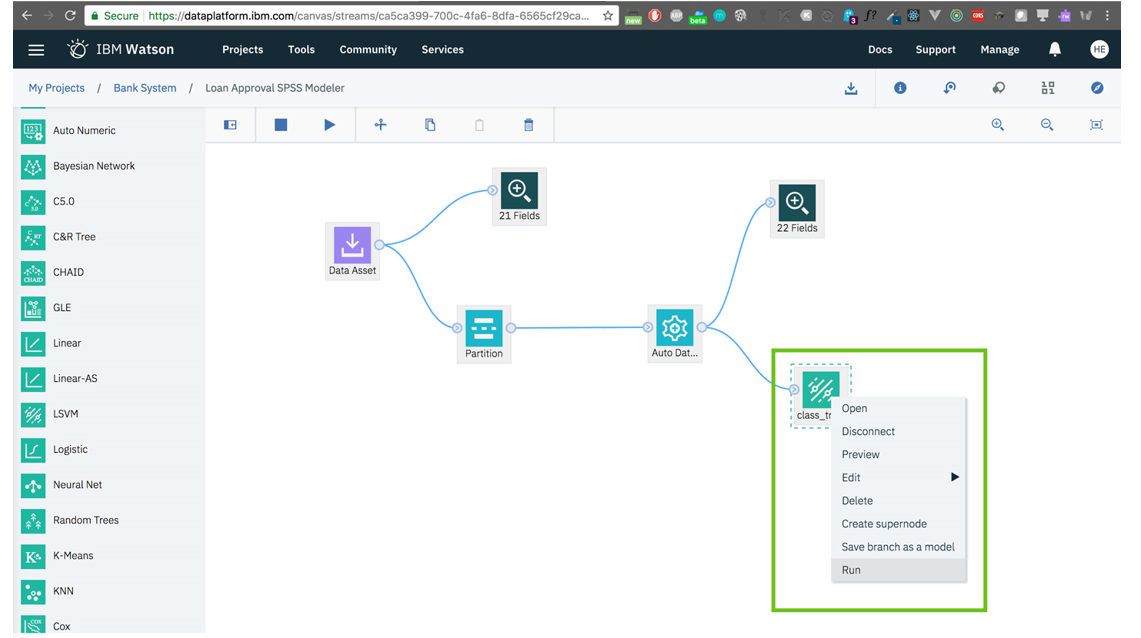


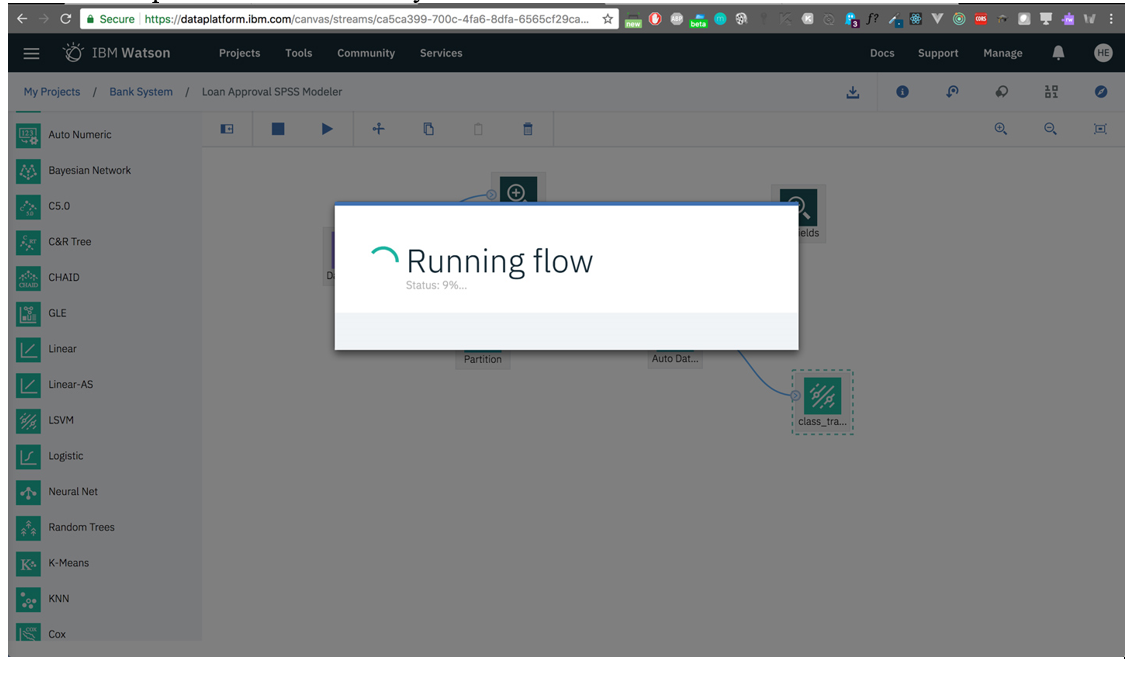
View the results.  


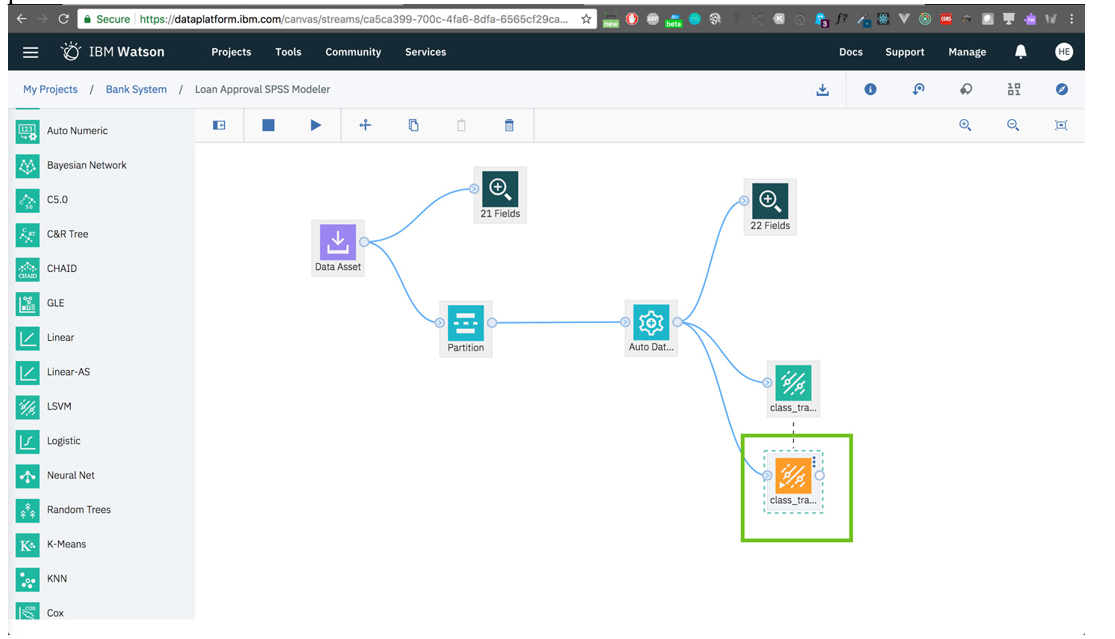
The data has been normalized and issues have been fixed.  


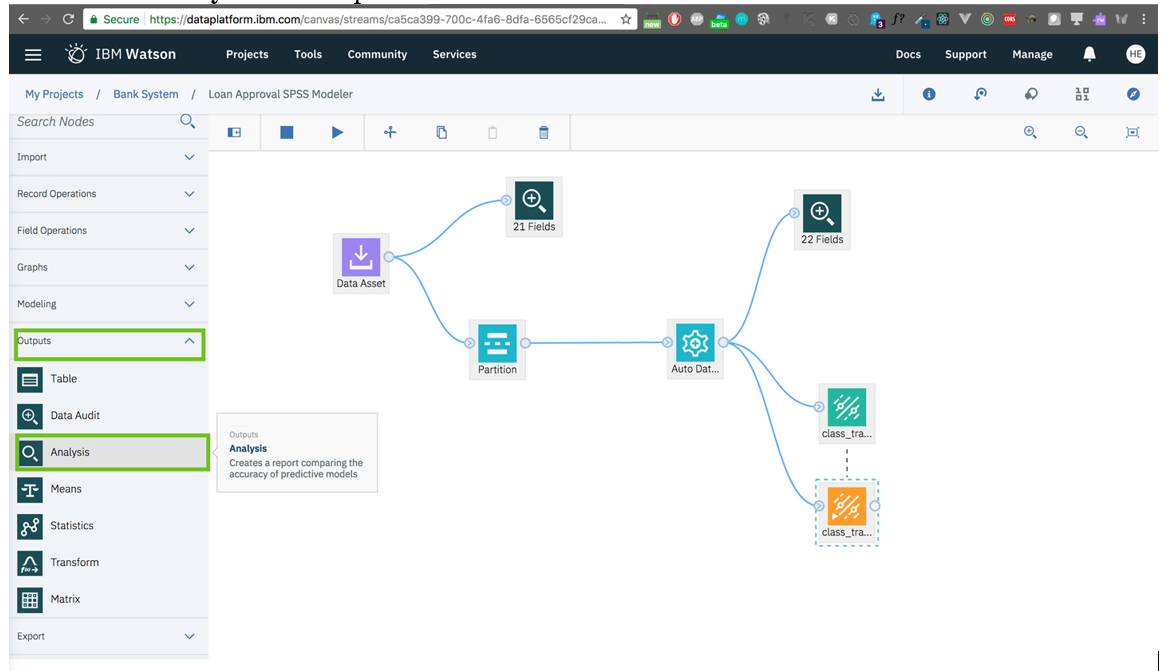
Now let's select a model, let's try **LSVM**.  


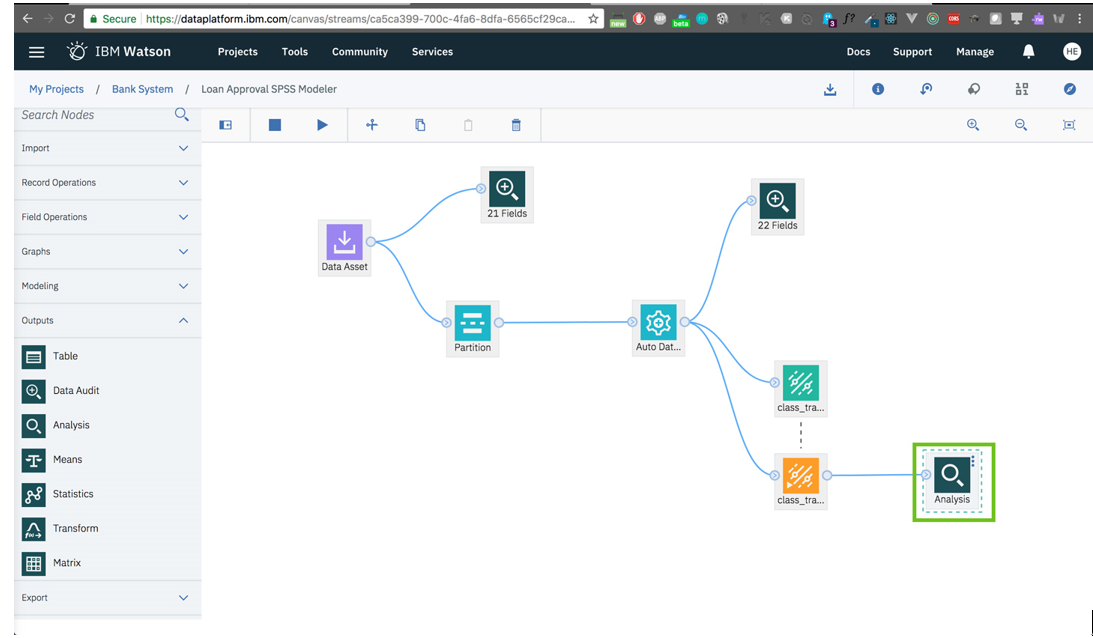
Double-clicking on the **LSVM** node to change its properties. Check **Use custom field roles**, then select class\_transformed as the target(labels) column and select all other columns as **Inputs**. Click **Save**.  


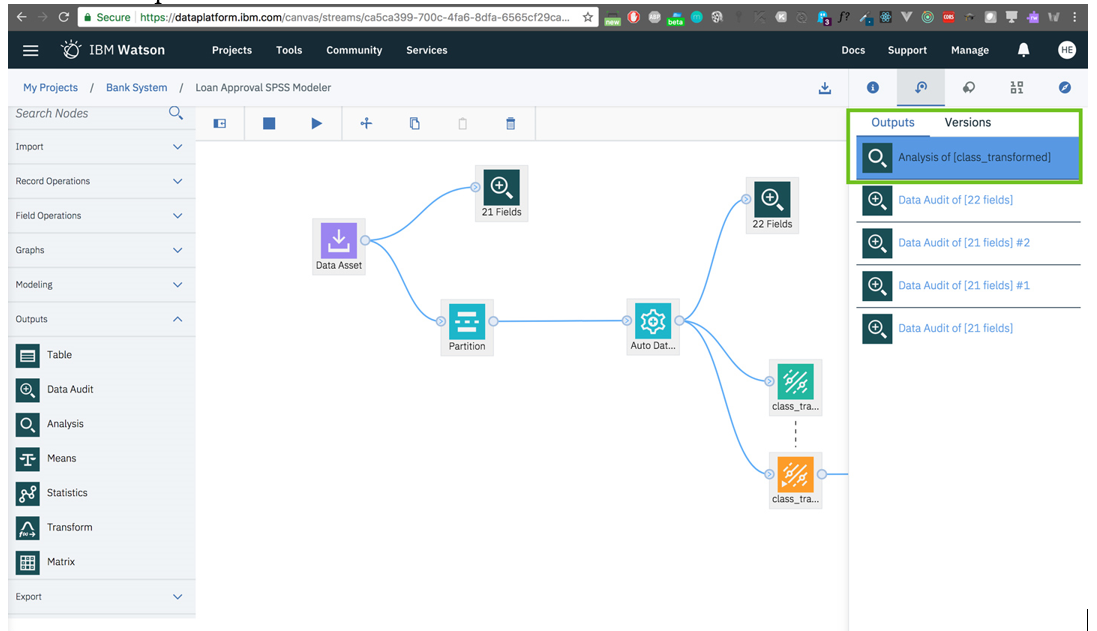
Let's run the flow.  


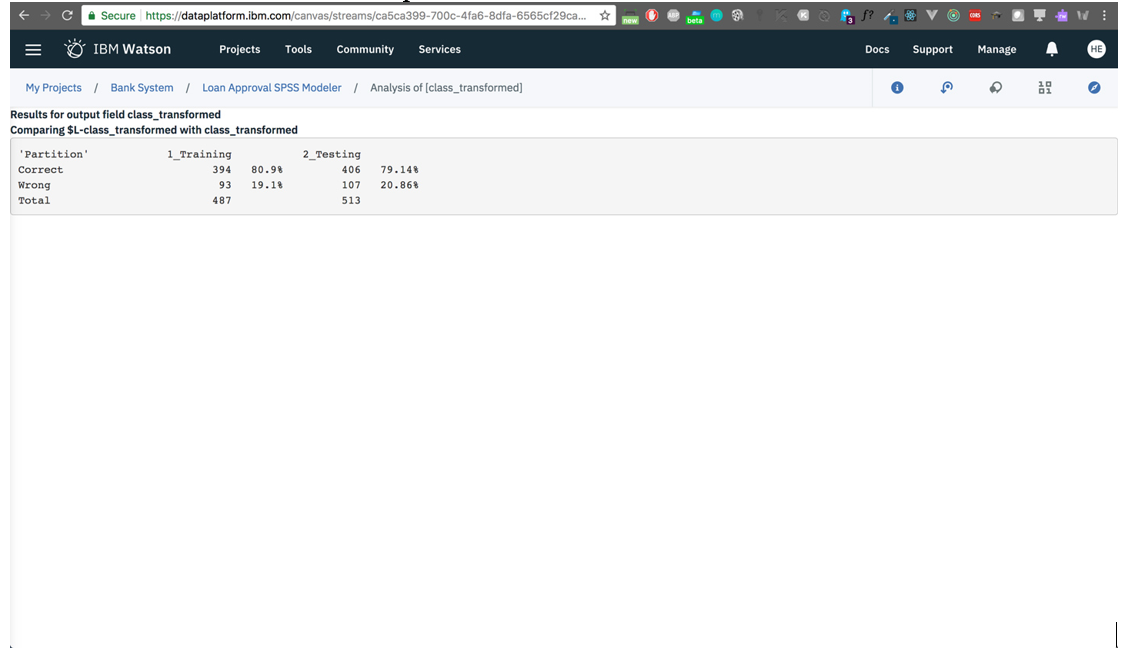
Wait for the process to finish. It may take a few minutes.  


After the model runs, it will produce a new node that holds information about the performance of the model.  


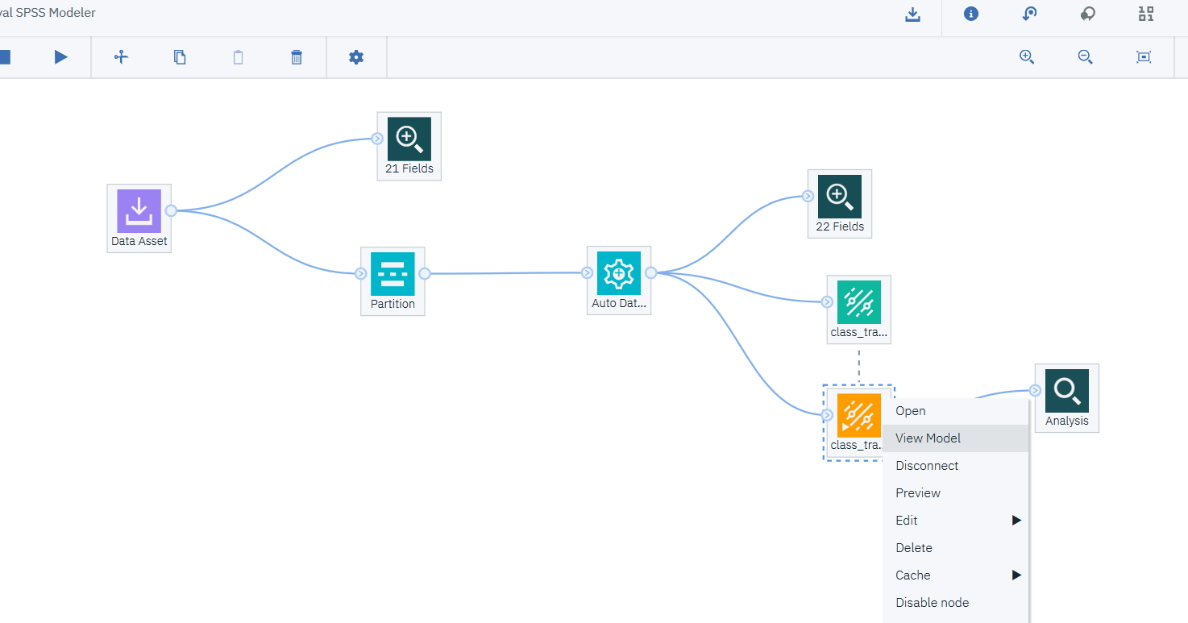
Let's add an **Analysis** node to peek into the model results node.  


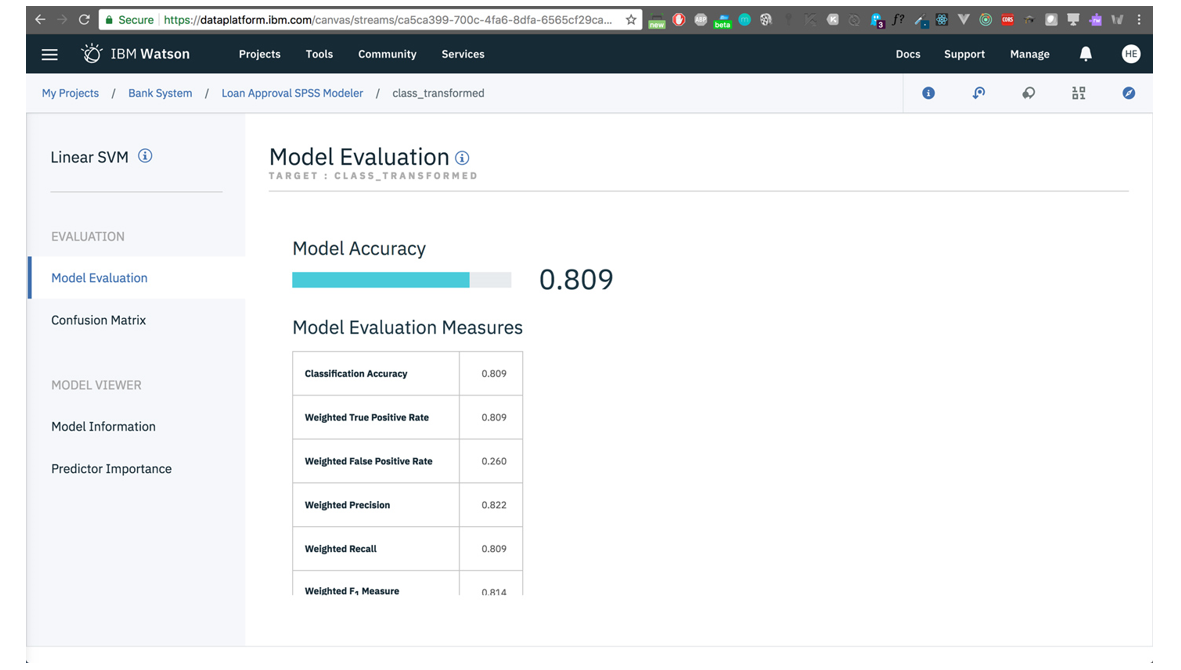
Connect the nodes.  


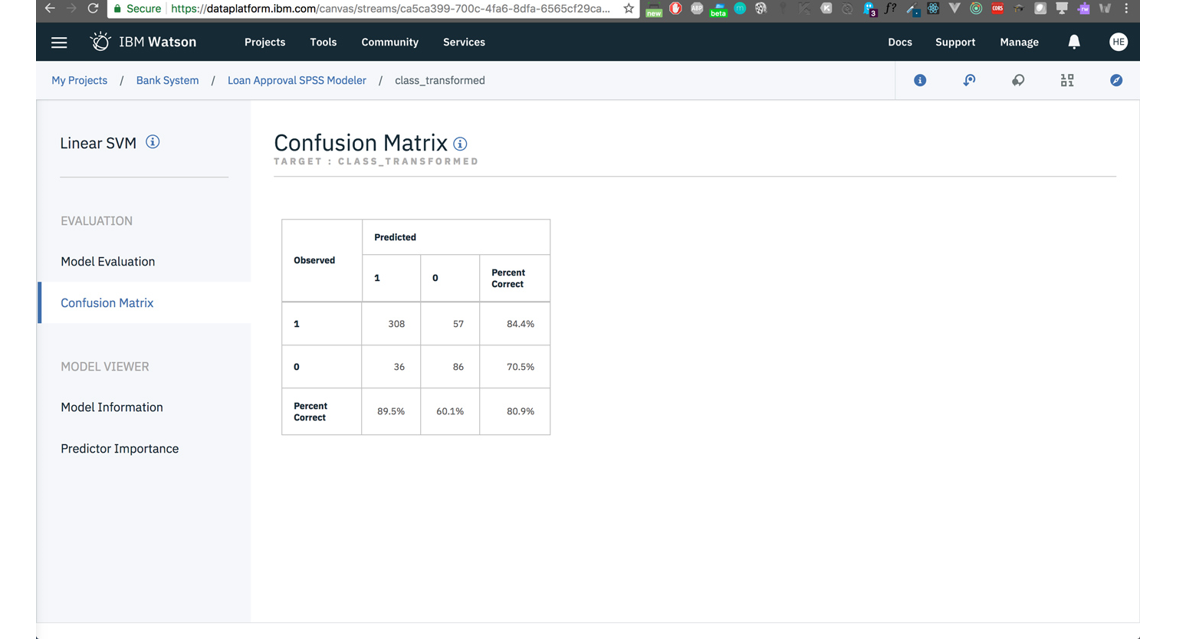
View the output.  


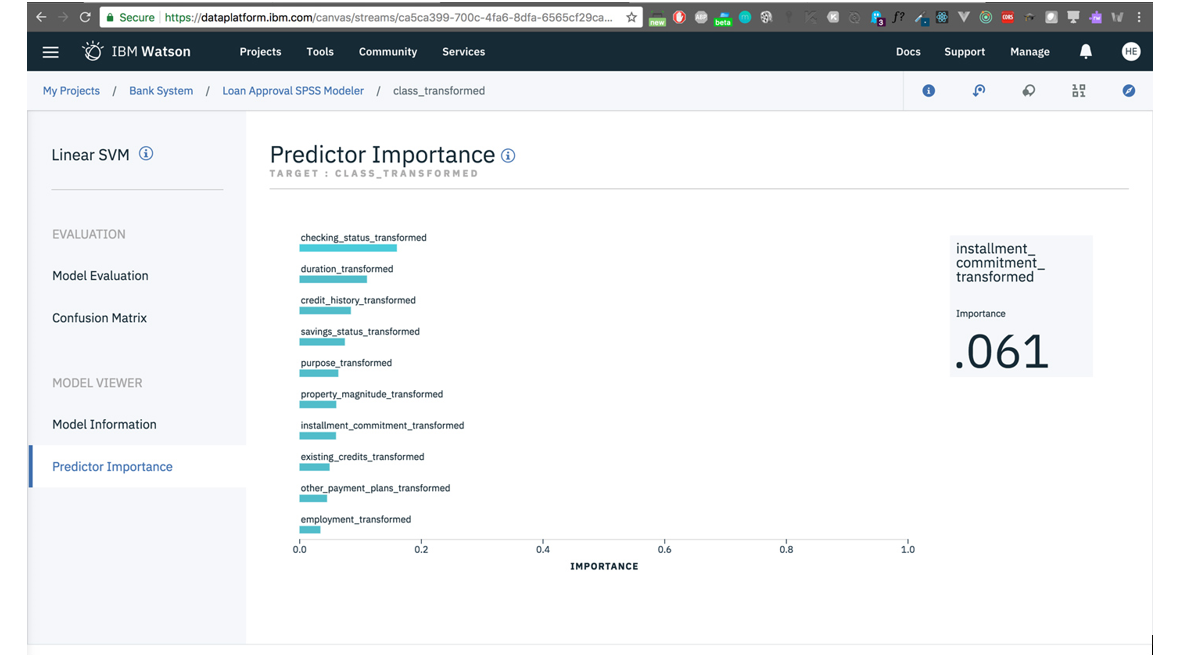
Information about how our model performed.  


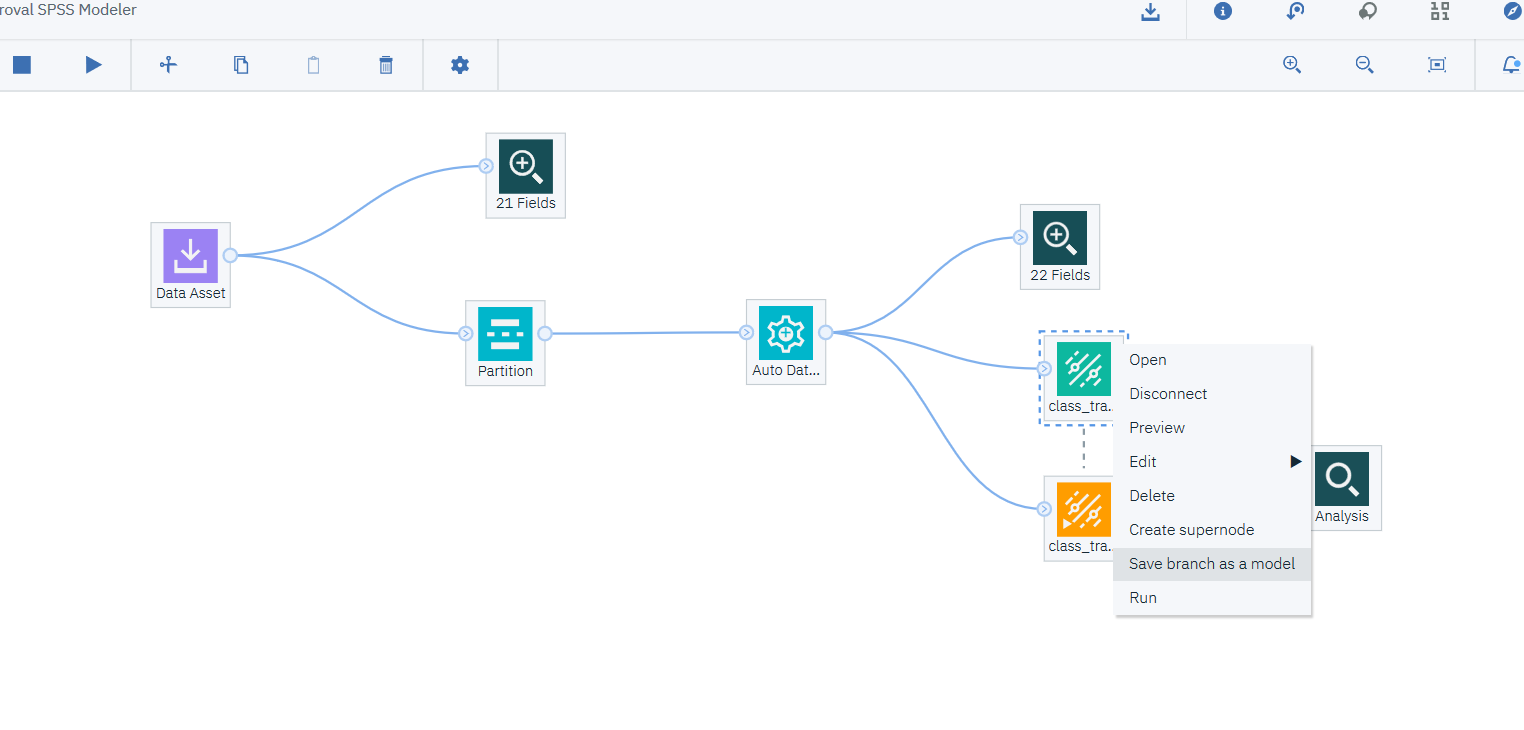
If we click on the **LSVM** model node again and choose **View model**, we will get more detailed information about different performance metrics.



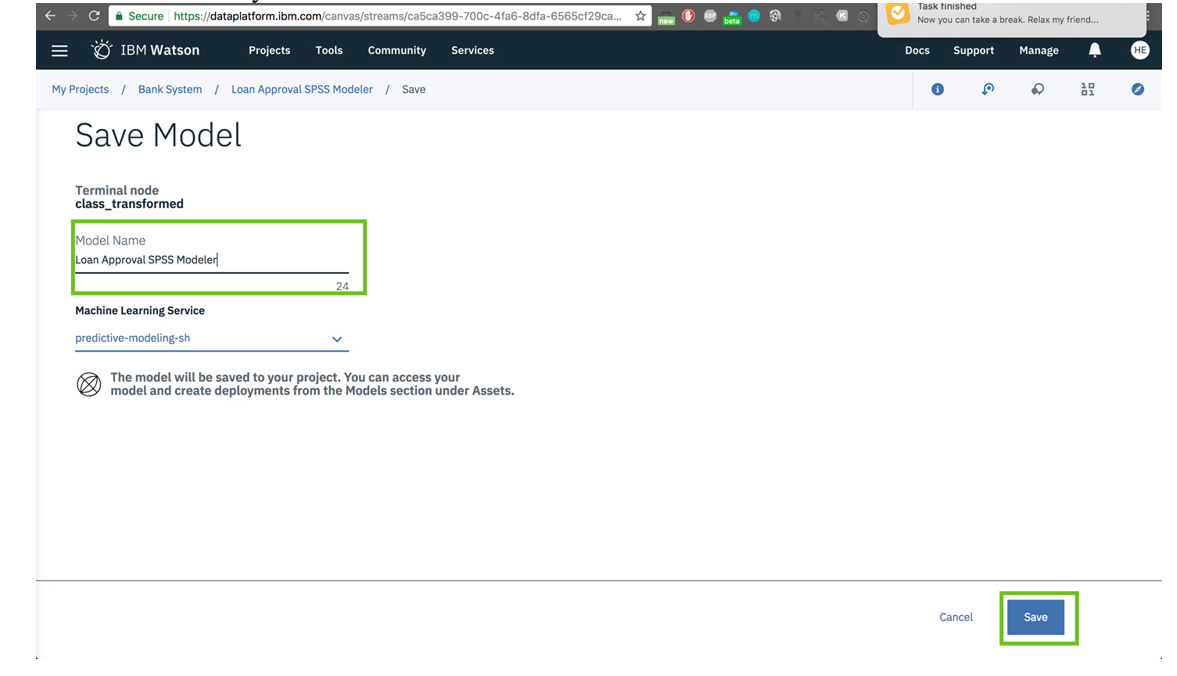




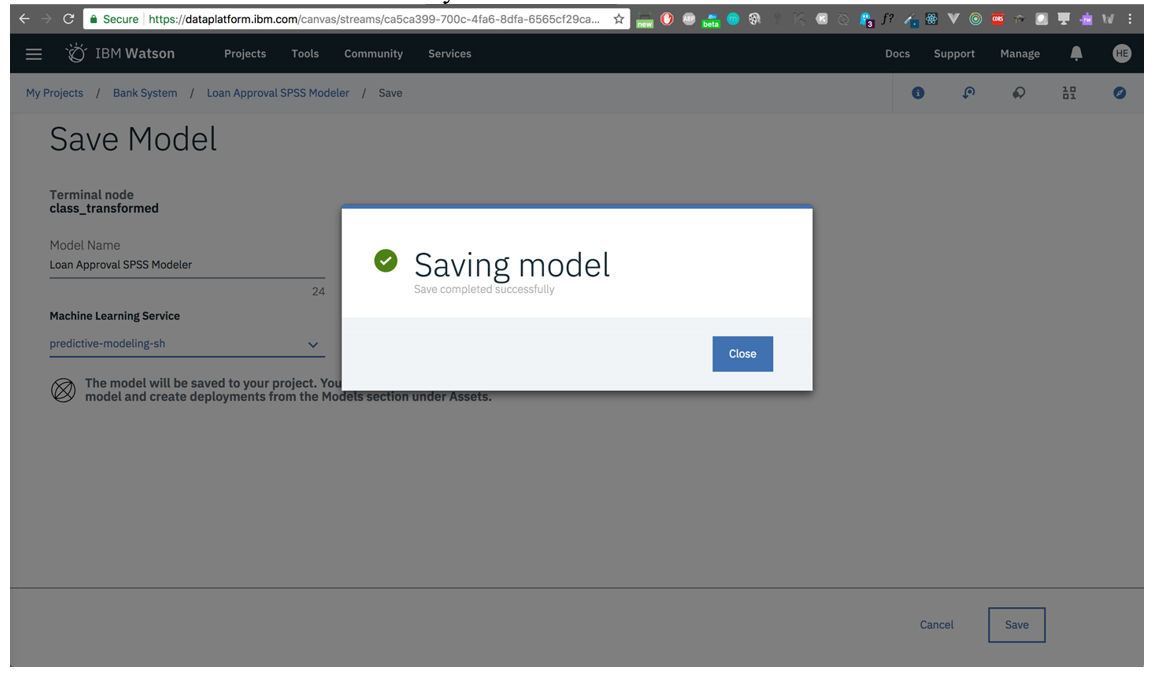


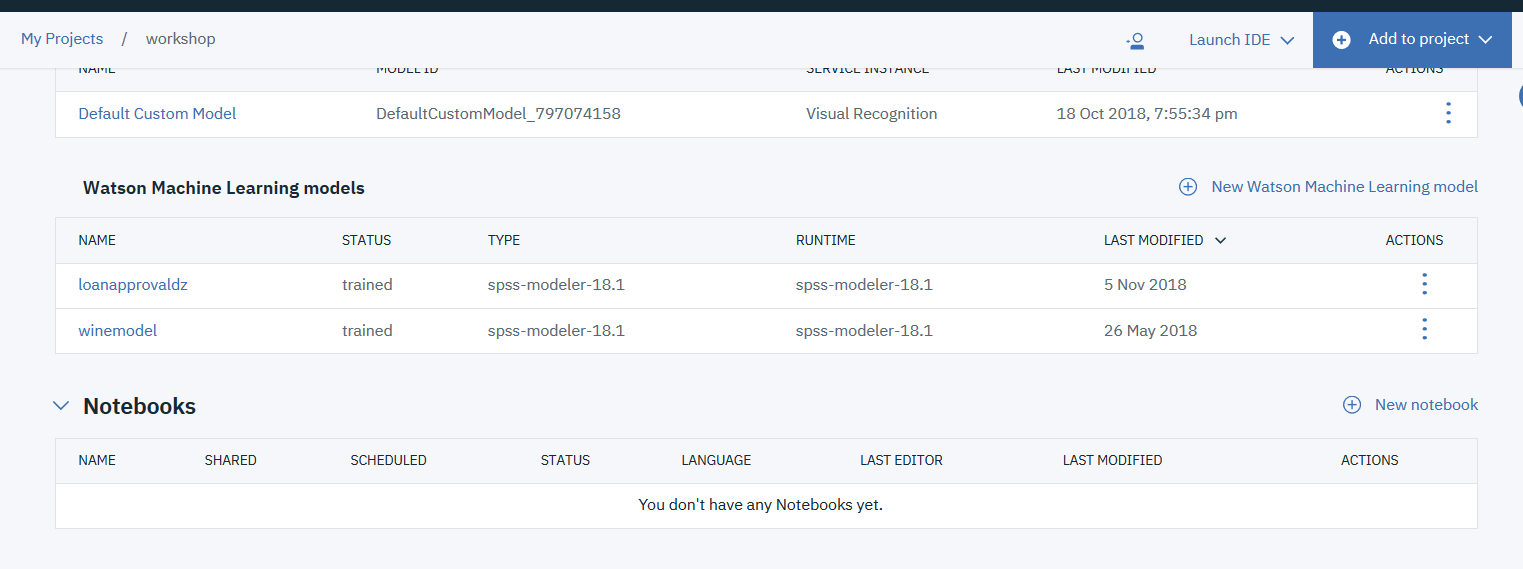
We're satisfied with our first model, it has better overall performance. Let's click on its node **LSVM** and select **Save branch as model**.  


Type a name for your model and a machine learning service should be detected and added automatically. Then click **Save**.

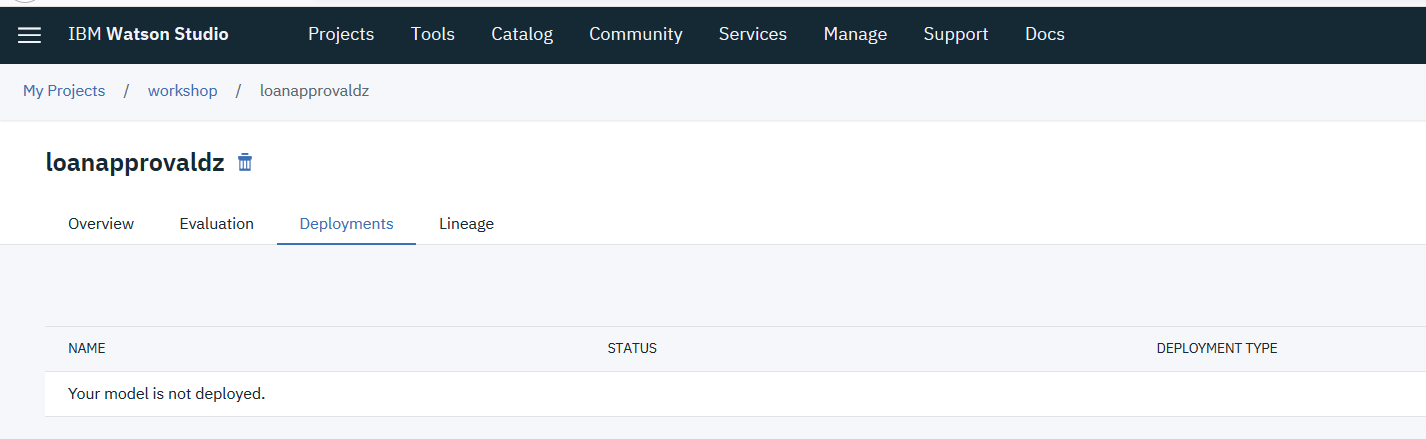


The model has been saved successfully.

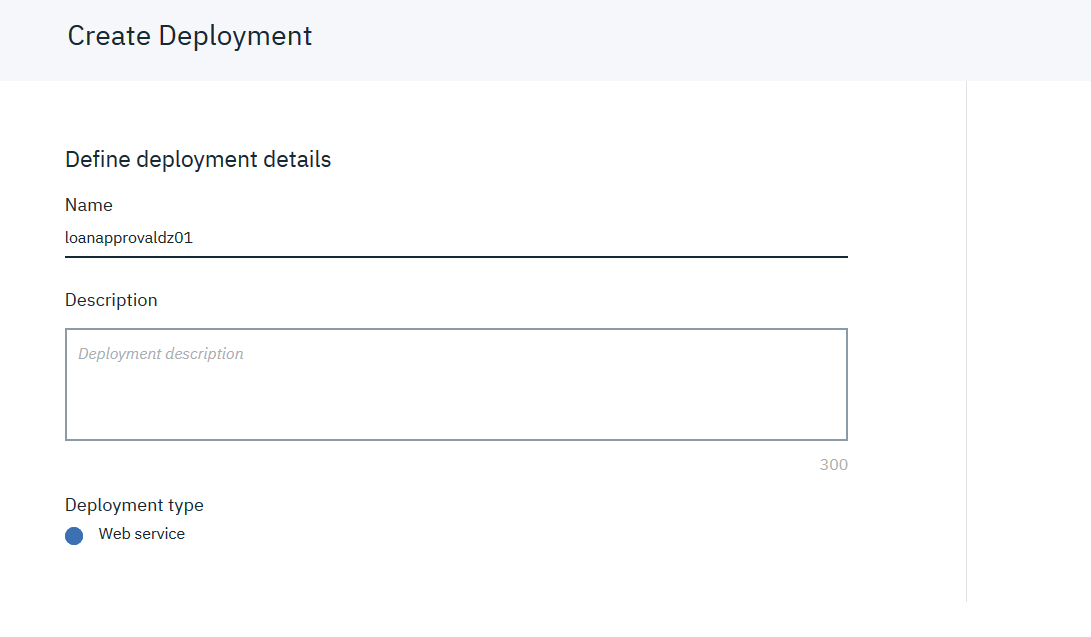


You can access it in the Watson Machine Learning model panel in the main dashboard.  


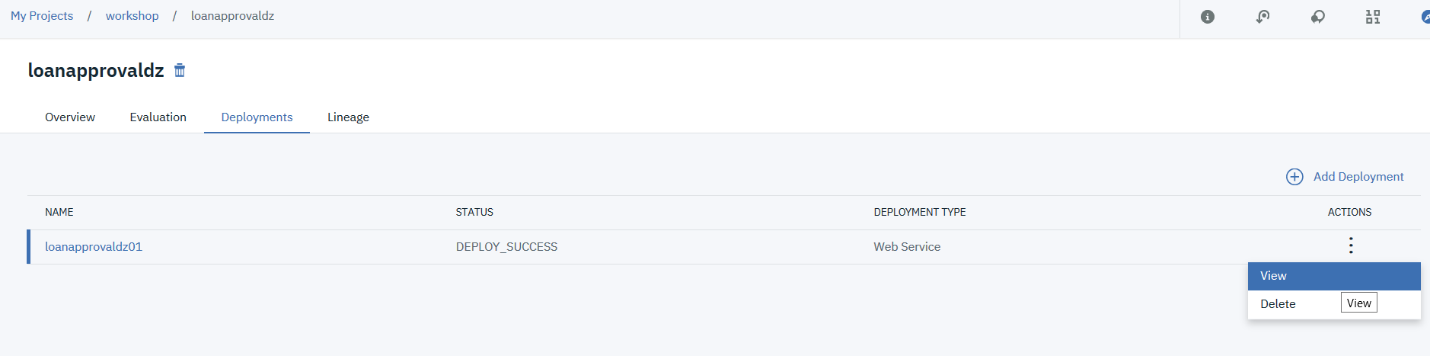
Click non the model e open the deployment tab.



Select add deployment.



Select save.



Select View and the implementation tab

