# Lab: SPSS Modeler in DSX

*Feb 28th, 2018*

*Author: Elena Lowery elowery@us.ibm.com*
*Updated by: Sidney Phoon yfphoon@us.ibm.com*

## Table of contents

# Contents

# Overview

In this lab you will learn how to implement analytics in the **SPSS Modeler** interface of Data Science Experience.

While knowledge of SPSS Modeler is not required to complete this lab, we don't explain the basics of working in Modeler UI. Please see SPSS Modeler documentation for detailed information:
https://www.ibm.com/support/knowledgecenter/SS3RA7_18.1.1/modeler.kc.doc/clementine/knowledge_center/product_landing.html

If you're completely new to analytics, this article provides a brief explanation of model types:
https://www.ibm.com/support/knowledgecenter/SS3RA7_18.1.1/modeler_mainhelp_client_ddita/clementine/getstart_mod_tech.html#getstart_mod_tech

# Required software, access, and files

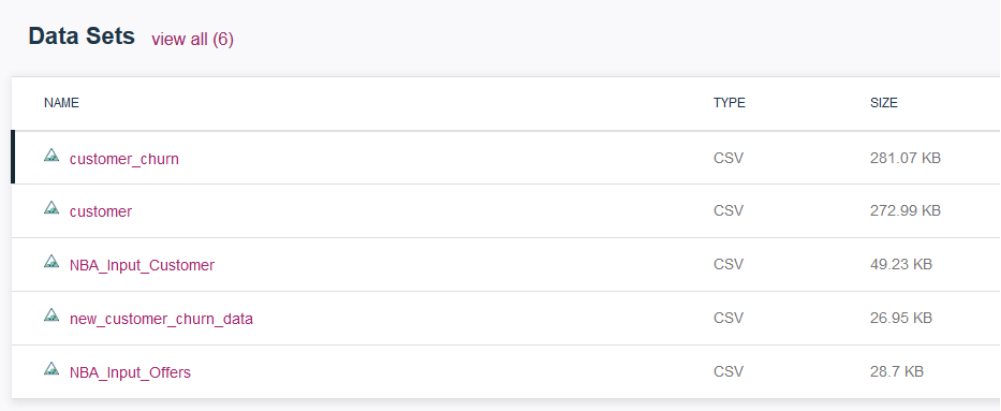
- To complete this lab, you will need access to a DSX Local cluster.
- You will also need to download files from this GitHub repository:
  https://github.com/SidneyPhoon/DSX_Local_Workshop
- Navigate to SPSS Modeler/Data in the GitHub repository and download the csv files
- Navigate to SPSS Modeler/streams in the GitHub repository and download the stream file

# Part 1: Create a DSX Project and Load Data

1. Log in to a **DSX Local cluster**.
2. Create a project. You can provide any name. In DSX Local, project names must be unique within in a cluster. For example, add initials to the name of your project.
   - If you already loaded *DSX_Local_Workshop* project, then you can add a Modeler Stream to it.
3. Switch to the **Assets** tab. Scroll to **Data Sets** and click on **add data set**.

4. Click **Browse** and import the csv files you have downloaded.

    - If you already imported the *DSX_Local_Project*, then all files are already loaded.



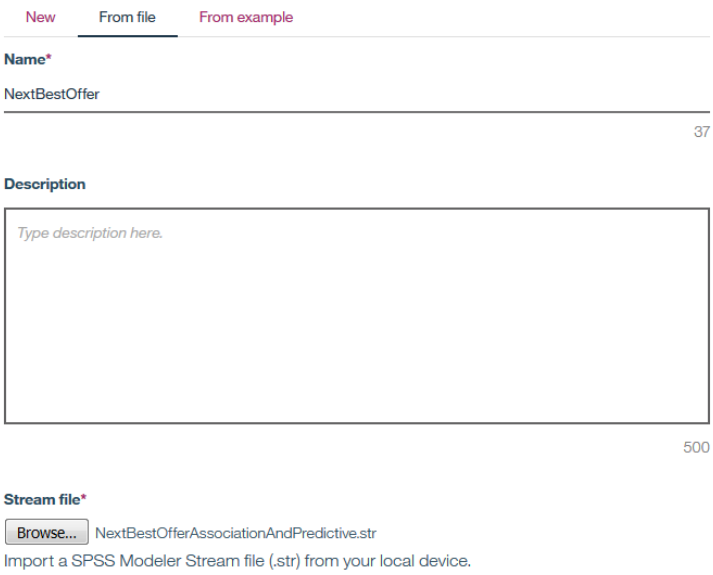

## Part 2: Import an existing SPSS Modeler Stream

In this section we will upload a stream that was previously created in SPSS Modeler desktop.

1. In the project scroll down to **SPSS Modeler** and click **add stream.**

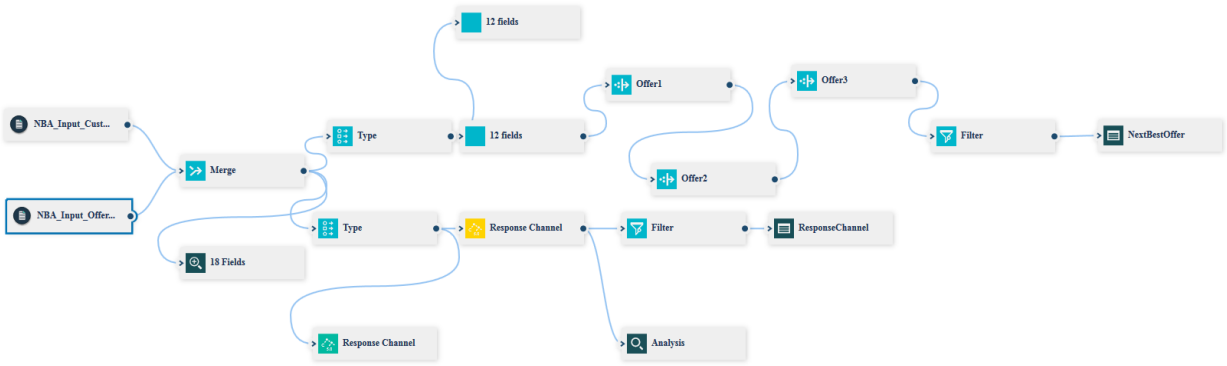
2. Enter a unique Stream name (for example, add your initials): *NextBestOffer_el.*

3. Select **From file** and select the *NextBestOfferAssociationAndPredictive.str* file you downloaded from GitHub. Click **Create**.




4. The imported stream will look similar to the following screenshot.

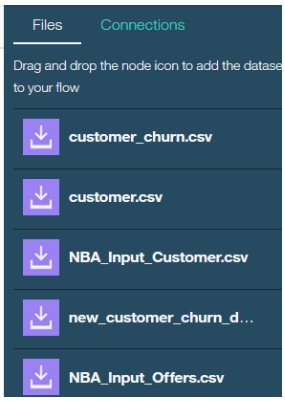
   This stream implements 2 use cases:
   - Recommend the top 3 offers (association model) for each customer
   - Recommend the best channel for contacting the customer (decision tree model)



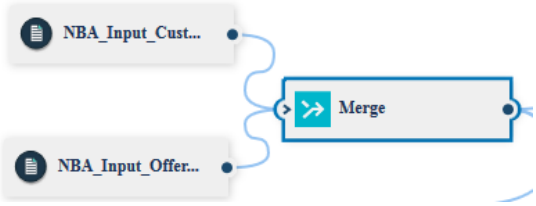

   *Note: No picture on the icon means that the node is not yet supported in the DSX/Modeler UI. However, the Modeler Stream will still run and execute this node. For example, in the stream we are using the association model is not yet implemented in the UI, but we can still run the stream.*

5. Before we can run this stream, we need to change the *Source* (input data) nodes. Delete the two *NBA_Input…* source nodes. In the imported stream the source nodes point to local file system.


6. Click on the **Data** icon and from the **Files** tab drag the two *NBA…* files onto the canvas.



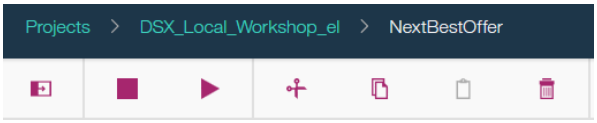

7. Connect the input files to the **Merge** node.



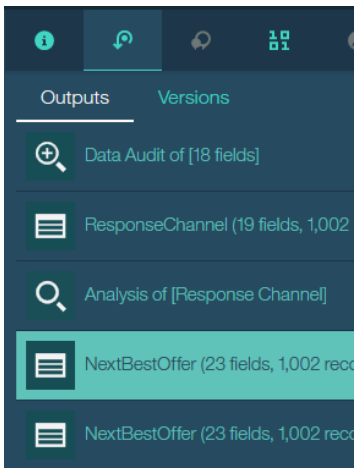
8. We are now ready to run the stream. You can run each individual branch of the stream by selecting the terminal (end) node and clicking the **Run** (arrow) icon.

   - For example, click on the Next Best Offer output field at the end of the main branch (on the right)


   - Or, you can run the entire stream, which will create output from all branches, by selecting the run icon (arrow) on the menu bar.

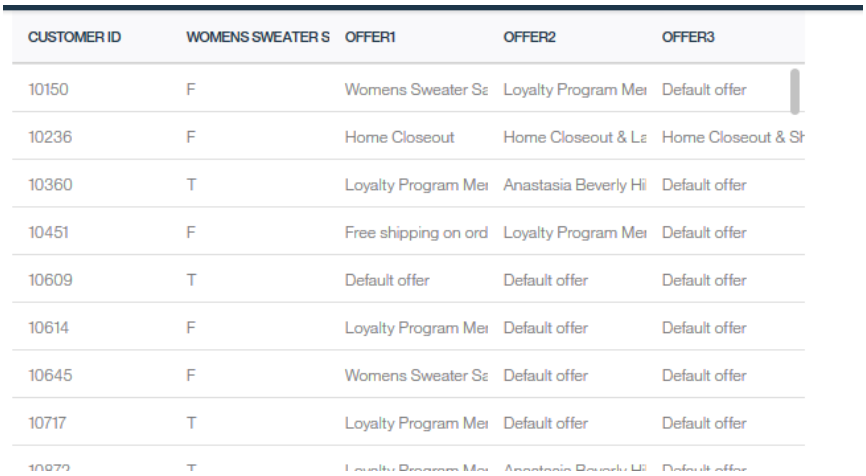



   - Output can be viewed by selecting the **Outputs** tab and double clicking on the output you'd like to view.



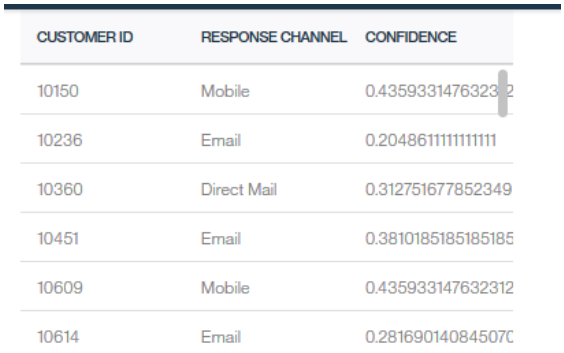
*NextBestOffer* and *ResponseChannel* outputs show the scoring results.

Double click on *NextBestOffer* to review recommendations for each customer.



| CUSTOMER ID | WOMENS SWEATER S | OFFER1 | OFFER2 | OFFER3 |
|---|---|---|---|---|
| 10150 | F | Womens Sweater Sa | Loyalty Program Mei | Default offer |
| 10236 | F | Home Closeout | Home Closeout & La | Home Closeout & Sh |
| 10360 | T | Loyalty Program Mei | Anastasia Beverly Hi | Default offer |
| 10451 | F | Free shipping on ord | Loyalty Program Mei | Default offer |
| 10609 | T | Default offer | Default offer | Default offer |
| 10614 | F | Loyalty Program Mei | Default offer | Default offer |
| 10645 | F | Womens Sweater Sa | Default offer | Default offer |
| 10717 | T | Loyalty Program Mei | Default offer | Default offer |

Double click on *ResponseChannel* to view recommendations for a marketing channel.



| CUSTOMER ID | RESPONSE CHANNEL | CONFIDENCE |
|---|---|---|
| 10150 | Mobile | 0.4359331476323 |
| 10236 | Email | 0.2048611111111111 |
| 10360 | Direct Mail | 0.312751677852349 |
| 10451 | Email | 0.3810185185185185 |
| 10609 | Mobile | 0.435933147632312 |
| 10614 | Email | 0.28169014084507( |

*Note: At this time SPSS stream deployment capabilities are not yet available in DSX Local. When deployment capabilities are available, you will be able to deploy the Modeler stream for batch and online scoring.*

After you're done reviewing the stream, navigate back to the project.

# Part 3: Create a Stream

In this section you will create a Modeler stream in DSX/Modeler UI. The process is the same as creating a Modeler stream in SPSS Modeler desktop, the only differences are in the UI look and feel and runtime environment (DSX).

As explained earlier, if you need more information about SPSS Modeler, refer to SPSS Modeler Knowledge Center documentation.

We provide instructions for the following use case:

- *Telco churn*: uses the same data and implements the same use case as Jupyter and Zeppelin Telco Churn notebooks.

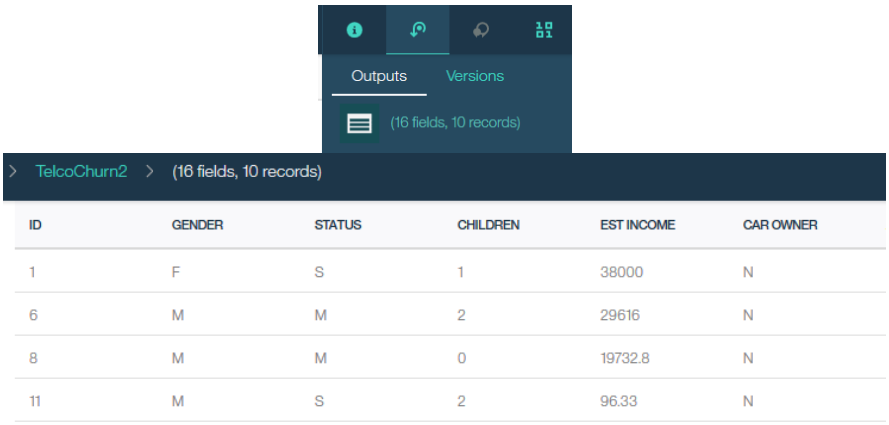
## Telco Churn

1. On the **Assets** tab scroll down to **SPSS Modeler** and click **add streams**.

2. Provide a unique stream name (for example, add your initials) *TelcoChurn_el*. Click **Create**.

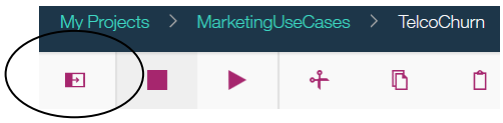
3. Drag and drop *customer* and *churn* data source to the canvas.


4. Right mouse click on the data sources in the canvas and select **Preview** to view the data. Preview is accessible through the **Outputs** tab.

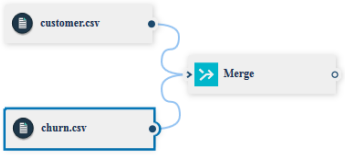



| ID | GENDER | STATUS | CHILDREN | EST INCOME | CAR OWNER |
|---|---|---|---|---|---|
| 1 | F | S | 1 | 38000 | N |
| 6 | M | M | 2 | 29616 | N |
| 8 | M | M | 0 | 19732.8 | N |
| 11 | M | S | 2 | 96.33 | N |

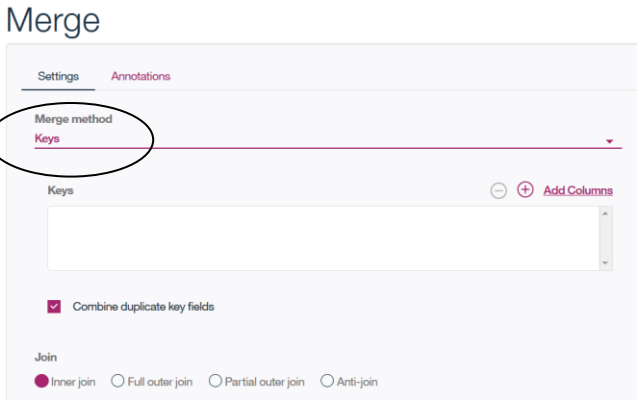
5. Click on the **Palette** icon and expand **Record Operations.**



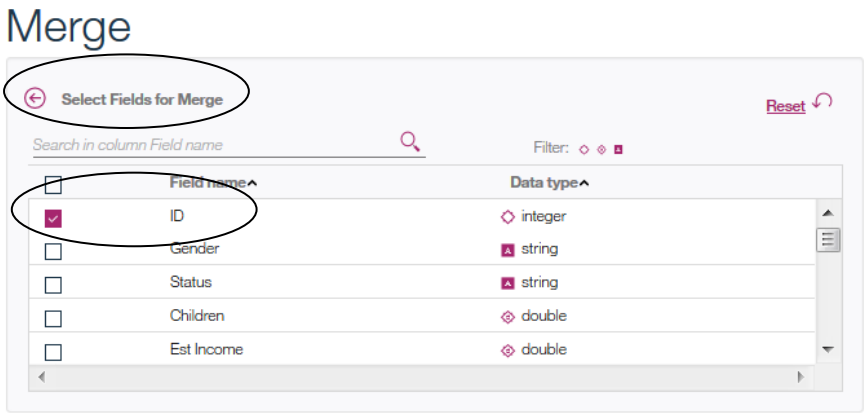

6. Add the **Merge** node then connect the *customer* and the churn data sources to it.



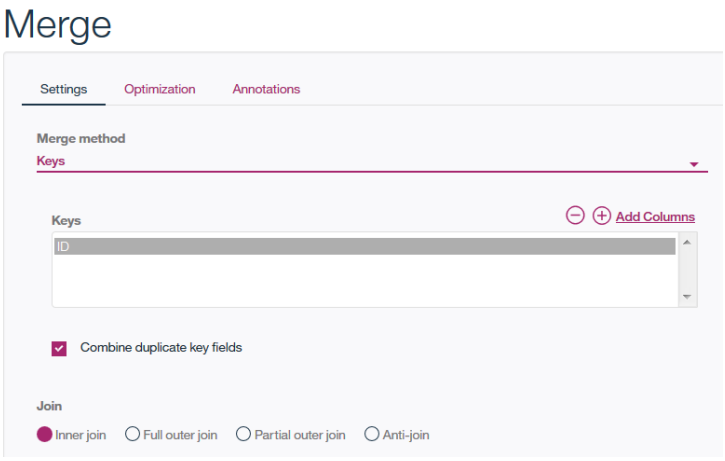

7. Double click on the **Merge** node. Select *Keys* as the **Merge method**.



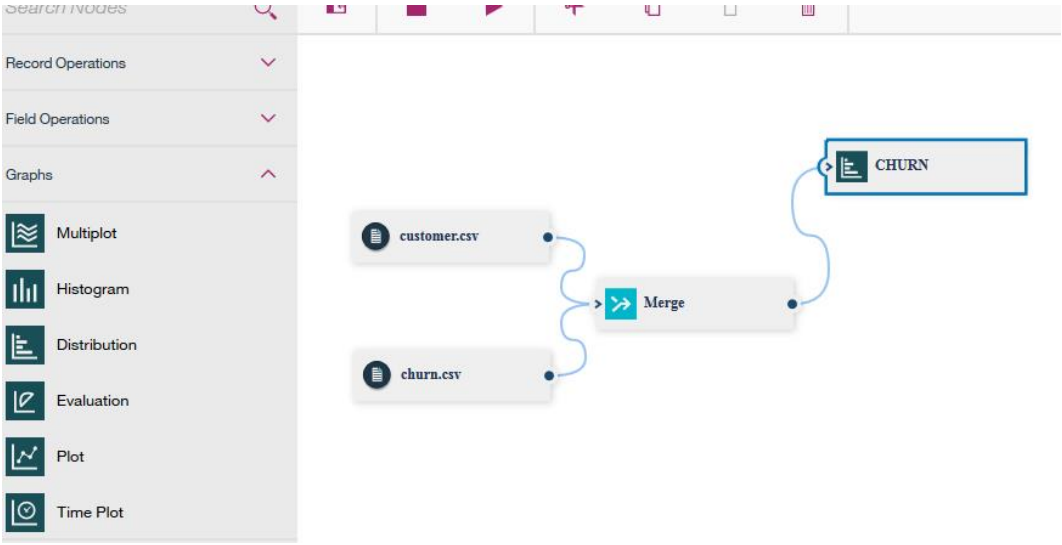
8. Click **Add Columns** and select the ID filed. Click **Select Fields for Merge** to return to the previous screen.



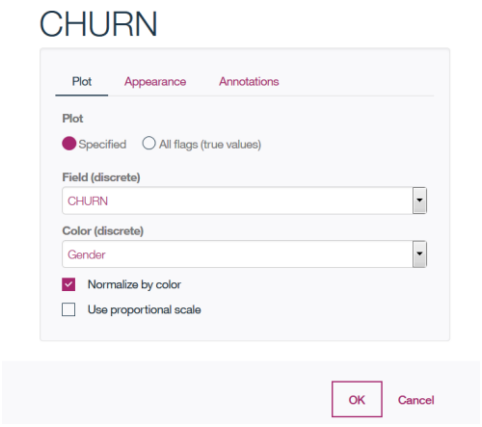

9. Now the Merge screen looks like the following screenshot. Click **OK**.



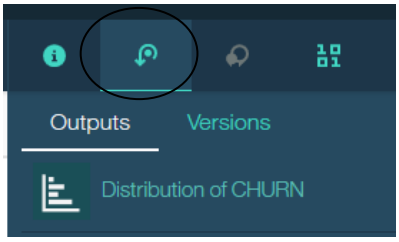
10. Next, you can connect the merged data to different types of graphs to get a better understanding of data. For example, you can add a **Distribution** graph and display *churn* by *gender*.

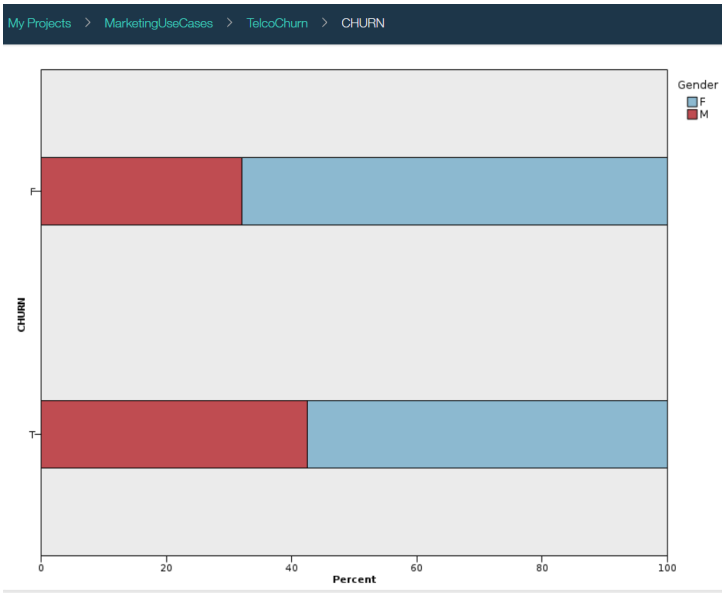



11. Double click on the **Distribution** node to edit it. Then right click and select **Run**.

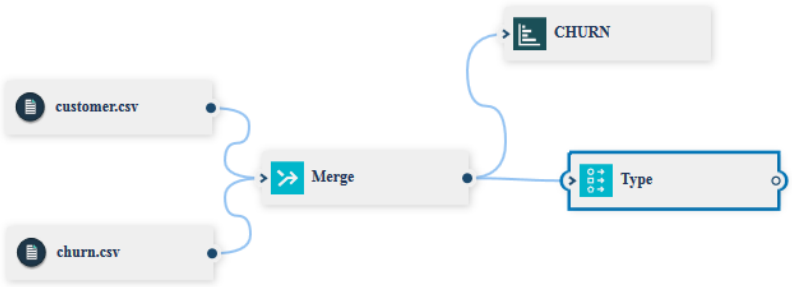


12. Output is shown in the **Ouputs** panel. Double click to display it.




13. Next, we are going to build a model for predicting churn. Add a **Type** node from the **Field operations** and connect it to the **Merge** node.

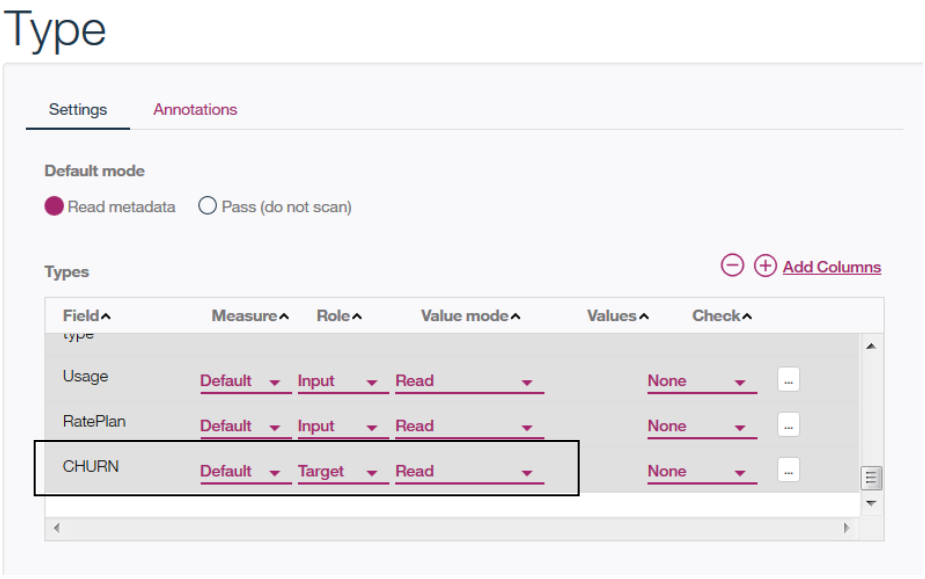



14. Double click on the **Type** node and click **Add Columns**. On this screen we are selecting the fields that will be used for modeling.

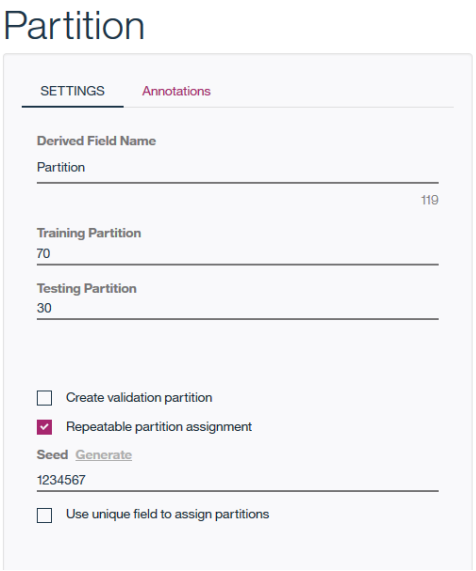
15. Select all fields with the exception of **ID** (because ID is not a predictor for customer churn). Return back to the main screen and change **Role** of *CHURN*

field to *Target* and **Measure** to *Flag* because that's the value we would like to predict.

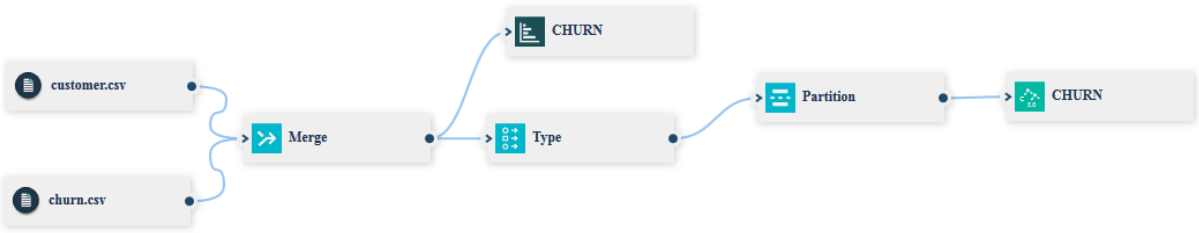
Click **OK**.




16. From the **Field Operations** add the **Partition** node and set the **Training Partition** to *70* and the **Testing Partition** to *30*.

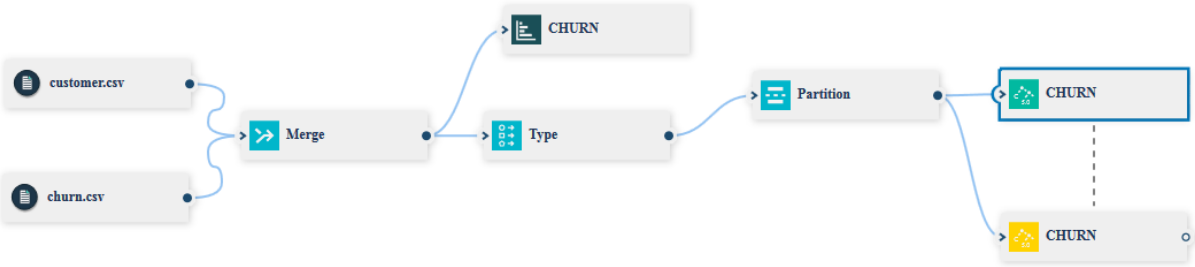



17. Add the **C5** node from the **Modeling** tab. The C5 is a popular decision tree algorithm.

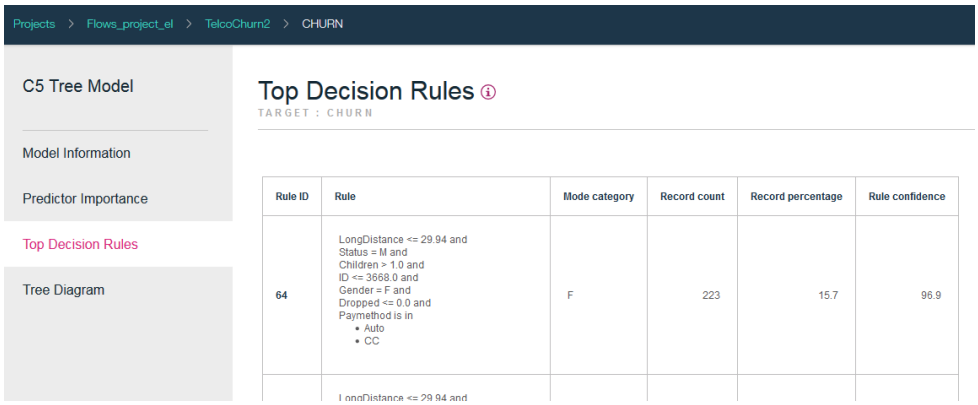



18. Right click on the **C5** node on the canvas (*CHURN*) and select **Run**.

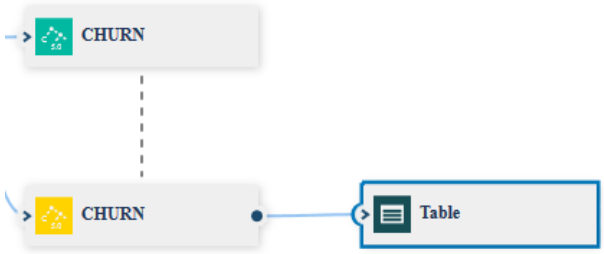
19. Model building will take a few minutes. When model building is done, you'll see a model node on the canvas.



20. Right click on the model nugget and select **View Model**. Explore the model. For example, **Top Decision Rules tab** shows the combination of predictors that result in specific customer churn value.

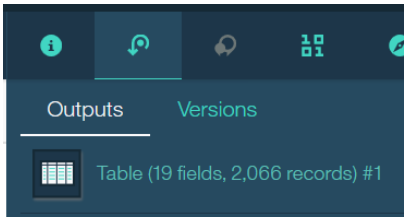



21. Finally, add a **Table** (from **Outputs** tab) and connect it to the model node.




22. You can run the entire stream by clicking the **Run** icon (arrow) in the menu bar. When the stream runs, it scores the data and writes the output to the **Table**.

**Table** output can be viewed by clicking on the **Table** in the **Outputs** view.



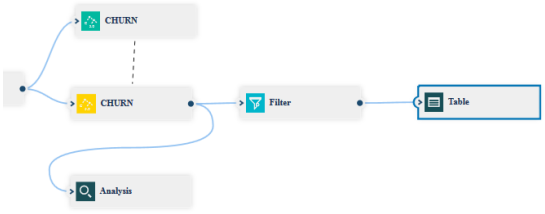
23. Scroll all the way to the right and you'll see two values generated by the model – the predicted value *($C-CHURN*) and the confidence in the prediction *($CC-CHURN*)

| RATEPLAN | $C-CHURN | $CC-CHURN |
|---|---|---|
| 3 | T | 0.99404761 |
| 2 | F | 0.99305555 |
| 3 | F | 0.95454545 |

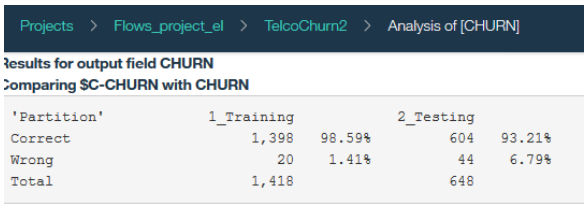
24. Optionally, you can add the **Analysis** (**Output** tab) the **Filter** node (**Field Operations** tab).

    - Connect the Analysis node to the model node. It shows the accuracy of the model (Run the stream and review the output)

    - Connect the Filter node to the model node before the Table. You can use it to filter out all fields with the exception of ID and the two fields generated by the model - *($C-CHURN*) and *($CC-CHURN*)




Analysis node output




Filtered table output

| ID | $C-CHURN | $CC-CHURN |
|---|---|---|
| 1 | T | 0.991596638655 |
| 6 | F | 0.988888888888 |
| 8 | F | 0.875 |

You have finished developing a model to predict customer churn.

*Note: At this time SPSS stream deployment capabilities are not yet available in DSX Local. When deployment capabilities are available, you will be able to deploy the Modeler stream for batch and online scoring.*