Lab: DSX Flows

October 30th, 2017

Author: Elena Lowery elowery@us.ibm.com

Table of contents

Contents

Overview

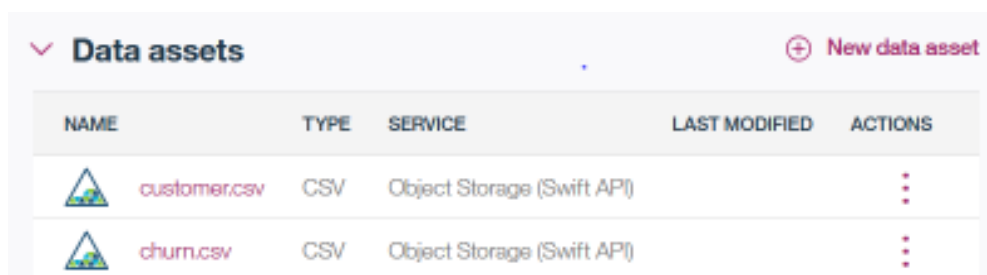In this lab you will learn how to implement analytics in the **Flows** interface of the Data Science Experience.

Required software, access, and files

To complete this lab, you will need an account in **DSX**:
https://apsportal.ibm.com/analytics

You will also need to download and unzip this GitHub repository:
https://github.com/elenalowery/DSX-Local-Telco-Churn

Part 1: Create a DSX Project and Load Data

1. Log in to DSX: https://apsportal.ibm.com/analytics

2. Create a project. You can provide any name.

3. Switch to the **Assets** tab and click on **New Data Asset**.

4. Click **Browse**. Navigate to the *data* folder of the unzipped GitHub repository and import *customer.csv* and *churn.csv*

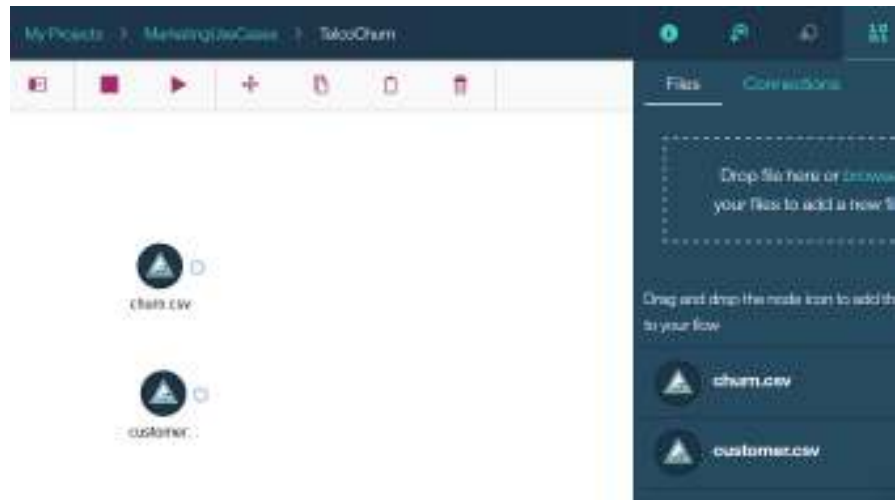| NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
customer.csv	CSV	Object Storage (Swift API)		
churn.csv	CSV	Object Storage (Swift API)		

Part 2: Create a Flow

1. On the Assets tab scroll down to **Flows** and click on **New Flow**
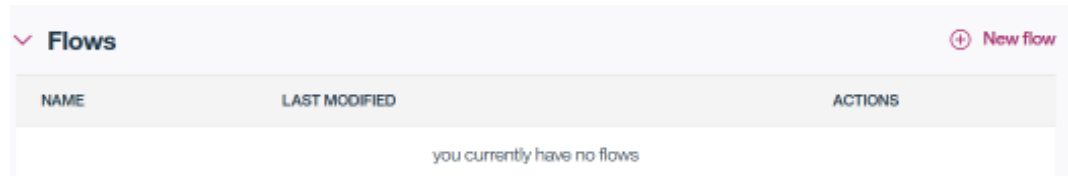
2. Name the flow *TelcoChurn* and make sure *IBM SPSS Modeler* runtime is selected. Click **Create Flow**.

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

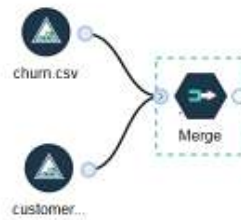4. Right mouse click on the data source in the canvas and select **Preview** to view the data.

5. Click on the **Palette** icon and expand **Record Operations.**
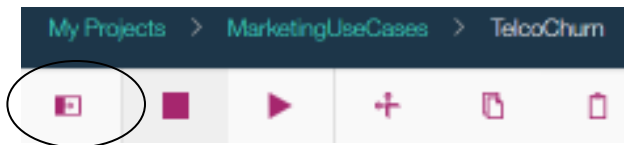
6. Add the **Merge** node then connect the *customer* and the churn data
   sources to it.

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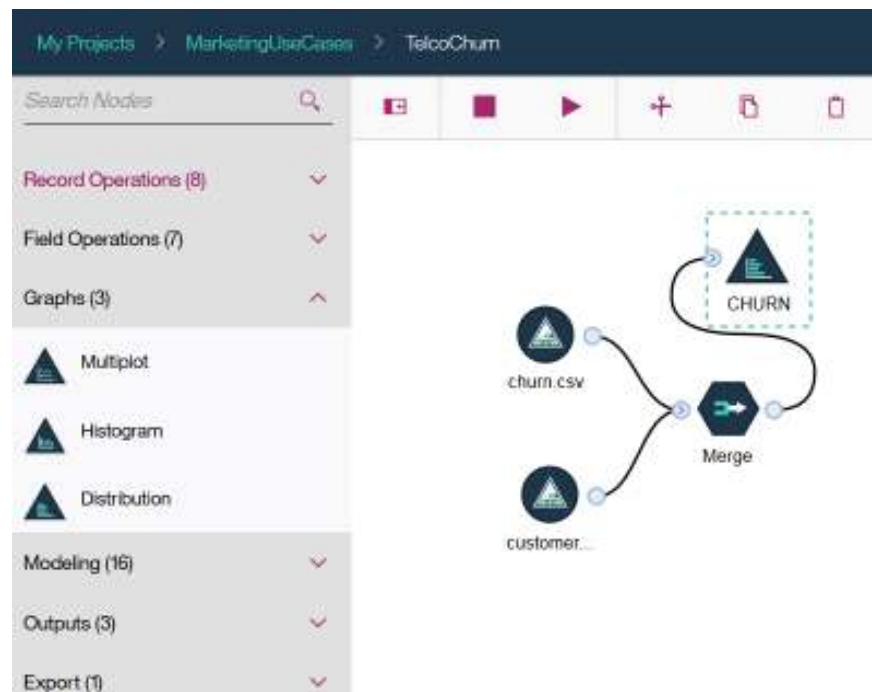
- Double click on the **Distribution** node to edit it. Then right click and select **Run**.

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

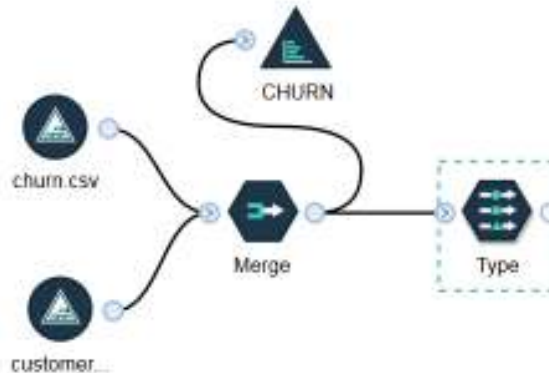11. Next, we are going to build a model for predicting churn. Add a **Type** node from the **Field operations** and connect it to the **Merge** node.

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

13. 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**.

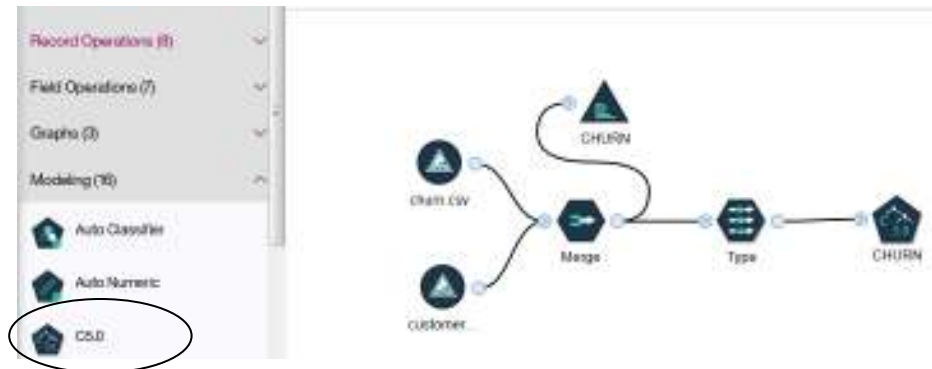14. Add the **C5** node from the Modeling tab. The C5 is a popular decision tree algorithm.

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

16. Model building will take a few minutes. When model building is done, you'll see a model nugget on the canvas.
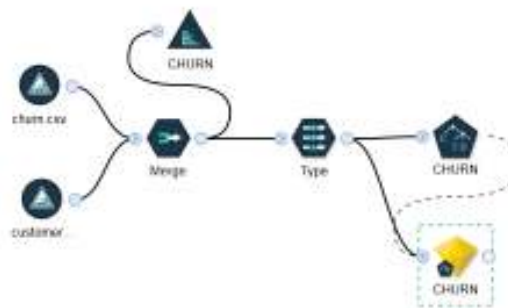
17. Right click on the model nugget and select **View Model**. Explore the model. For example, **Top Decision Rules tab** shows the combination of predictors that result in specific customer churn value.

18. Finally, add a **Table** (from **Outputs** tab) and an **Object Store** (from **Export** tab) nodes and connect them to the model nugget.
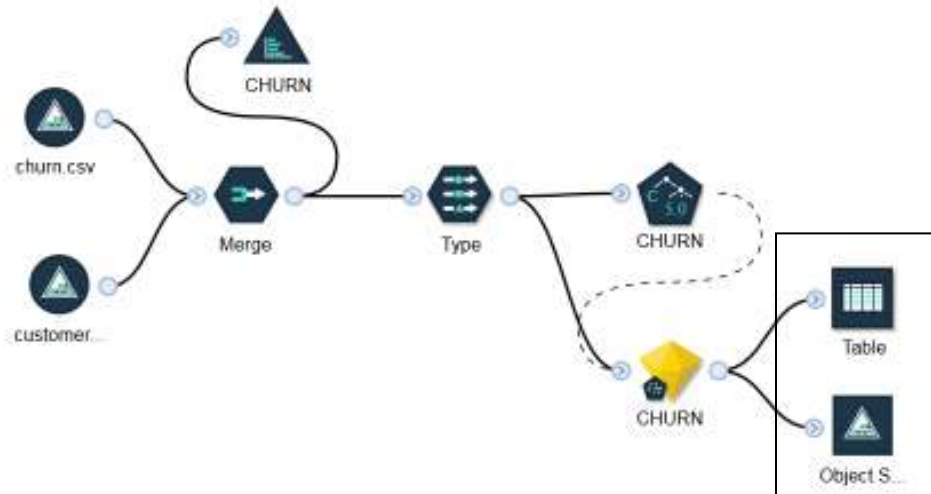
19. Double click on the **Object Store** and provide file name, for example, *ScoringOutput.csv*.

20. You can run the entire stream by clicking the **Run** icon (arrow). When the stream runs, it scores the data and writes it to the specified file in the **Object Store** as well as the visual output (**Table**).

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

- 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)*

21. So far we've built and scored the model using the training data. Models are typically evaluated with a different data set. If you wish, you can import the *customer_churn.csv* data set that's located in the data folder of the unzipped file from GitHub.

- After adding *customer_churn.csv* to the **Data Assets** on the Project page, connect it directly to the model nugget. You will first need to delete a connection from the **Type** node.

- You can also add the **Analysis** node from the **Output** tab and connect it to the model nugget.
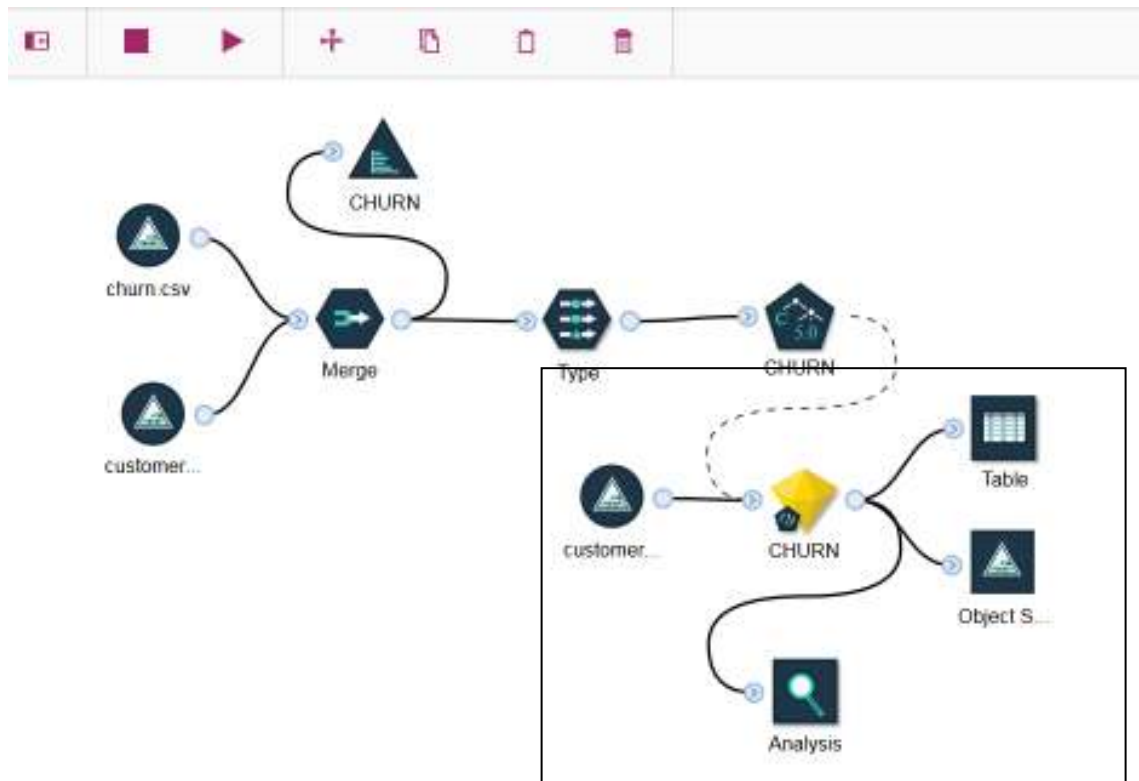
22. Right click on the **Analysis** node and select **Run**. The results are shown on the **Outputs** tab.

You have finished developing a model to predict customer churn. This model can be deployed in Watson Machine Learning for batch or real time scoring. Once deployed, the model can be integrated with Line of Business applications.