

IBM Cloud

Data Science Experience (DSX) Practical Hands-on Introduction

Lab Guide



Data Science



Machine Learning



Notices and Disclaimers

© Copyright IBM Corporation 2017.

The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. This information is based on current IBM product plans and strategy, which are subject to change by IBM without notice. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

IBM, the IBM logo and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Other company, product and service names may be trademarks or service marks of others

Document Revision History

Rev #	File Name	Date
1.0	DSX Hands-on Workshop.docx	11/1/2017

Prepared & Revised by:

Louis Frolio – louis.frolio@ibm.com

Table of Contents

Lab Environment Overview.....	5
Lesson 1: DSX Signup & Home Page	6
Lesson 1: Workflow Overview	7
Lesson 1: Instructions	8
Lesson 2: Jupyter Notebook.....	13
Lesson 2: Workflow Overview	14
Lesson 2: Instructions	15
Lesson 3: Machine Learning Flows	26
Lesson 3: Workflow Overview	27
Lesson 3: Instructions	28
Lesson 4: Watson Machine Learning	36
Lesson 4: Workflow Overview	37
Lesson 4: Instructions	38

Lab Environment Overview

Installed Software and Tools

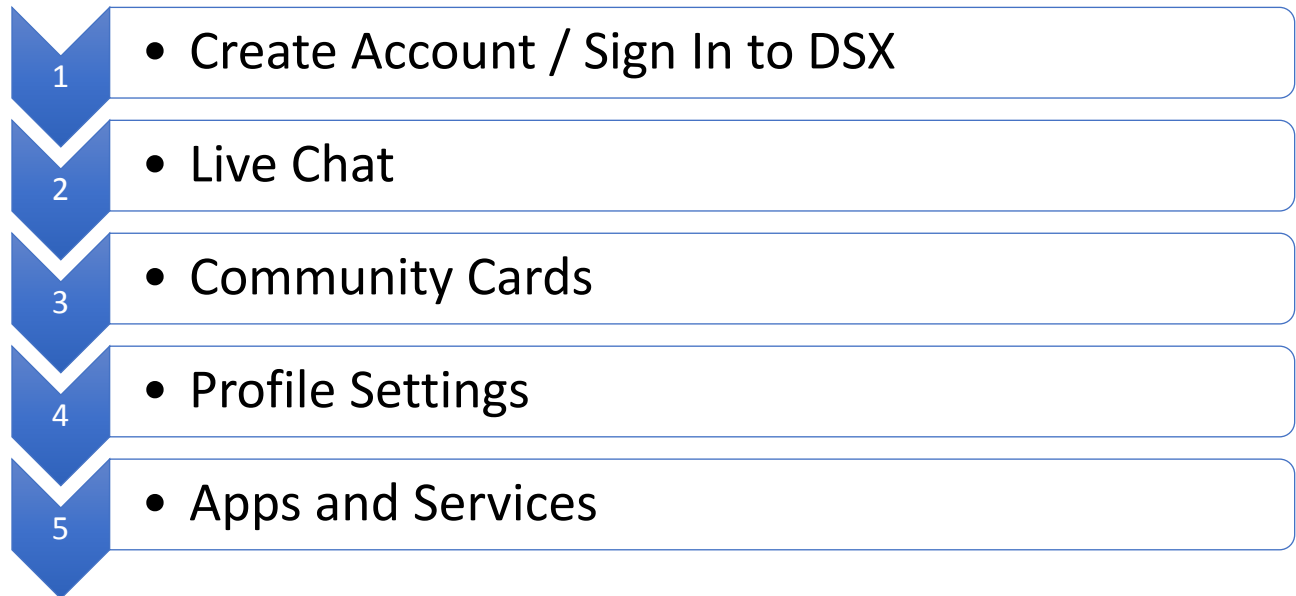
Software	Link
IBM Data Science Experience (DSX)	https://datascience.ibm.com/
IBM SPSS Statistics	http://www-03.ibm.com/software/products/no/spss-stats-base
Jupyter	http://jupyter.org/
GitHub	https://github.org/
R	https://www.r-project.org



Lesson 1: DSX Signup & Home Page

Purpose:	This lab introduces DSX, its sign up and walk-through of the features and functions starting at the Home Page.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Create/Sign In to DSX Account• Engage Live Chat• Differentiate Four Types of Community Cards• Explore Personal Profile, Apps/Services, and Integrations

Lesson 1: Workflow Overview

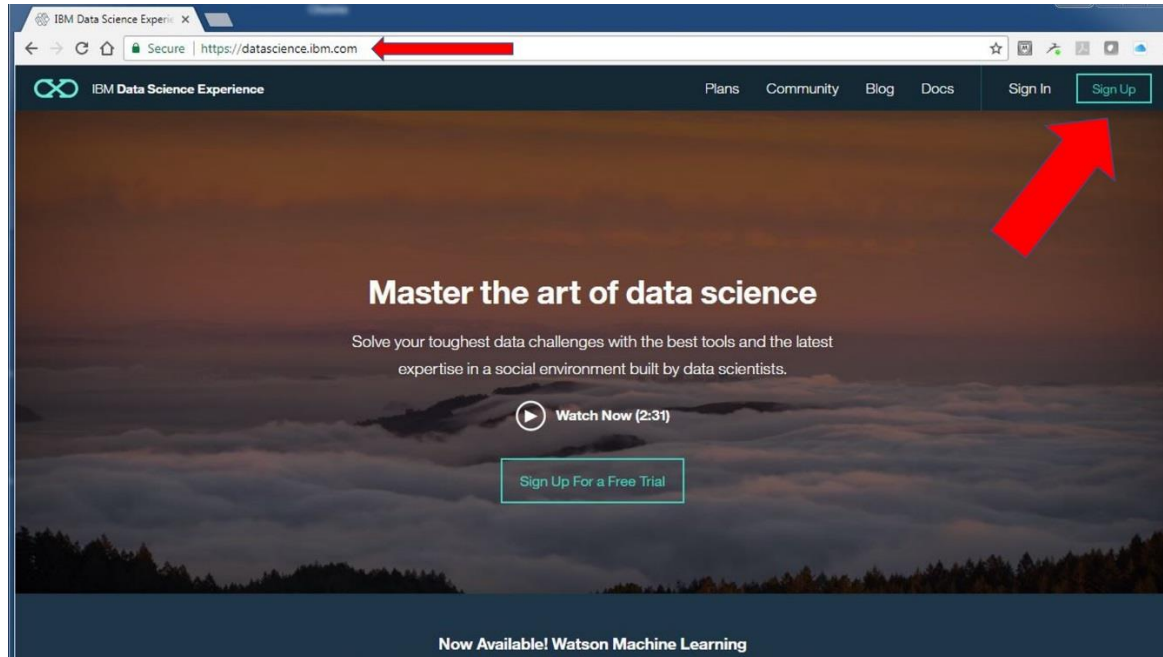


Lesson 1: Instructions

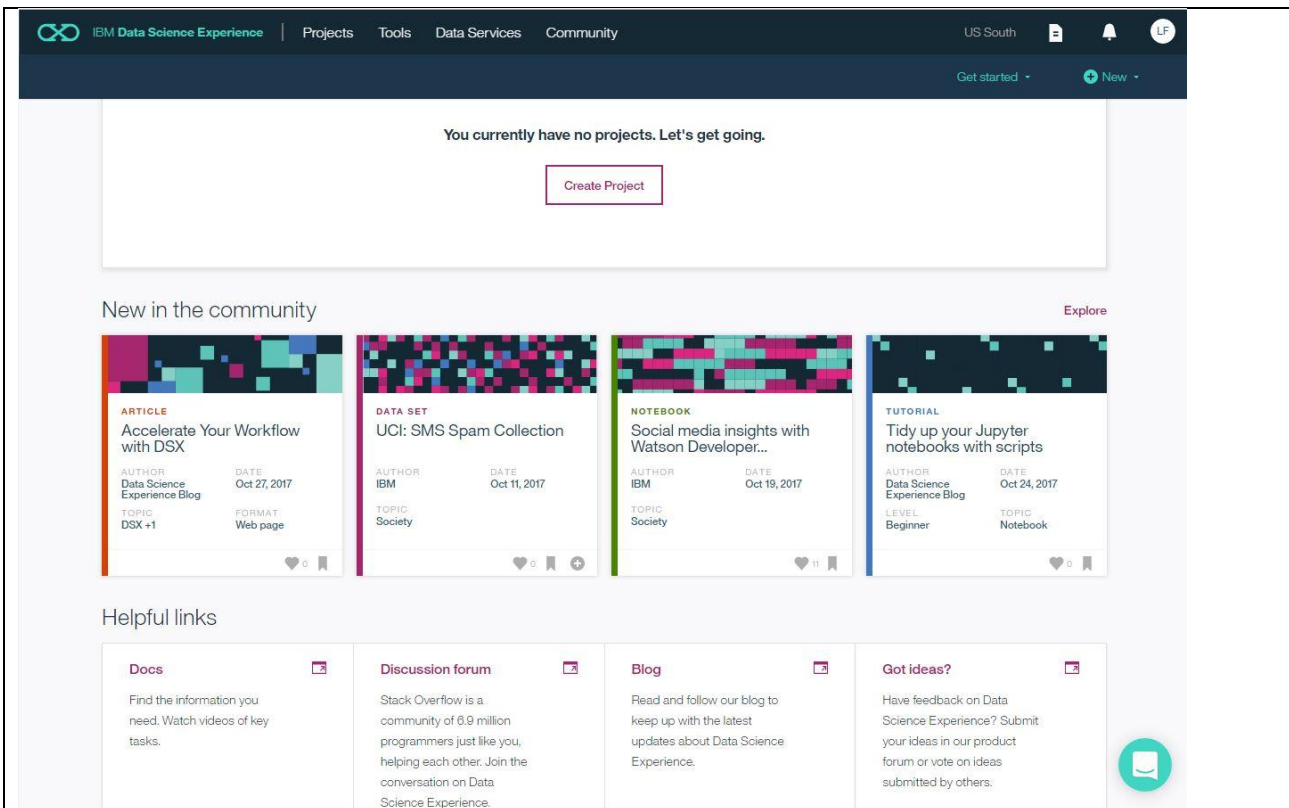
Action

1. Create Account/Sign In to DSX

- Open web browser and navigate to: <https://datascience.ibm.com>



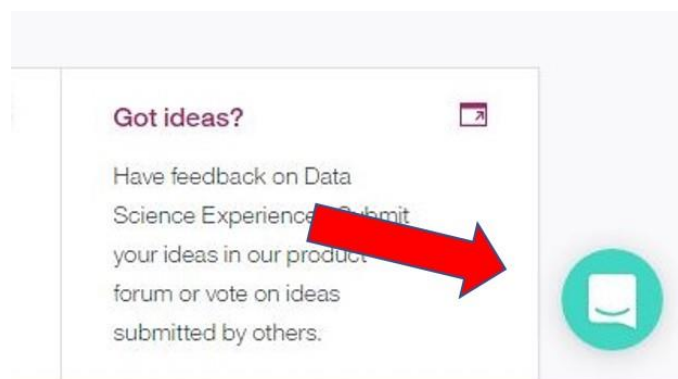
- Click on “Sign Up” and you will be prompted for several items of information. After a few moments of self-configuration, you will be brought to your new Home Page:

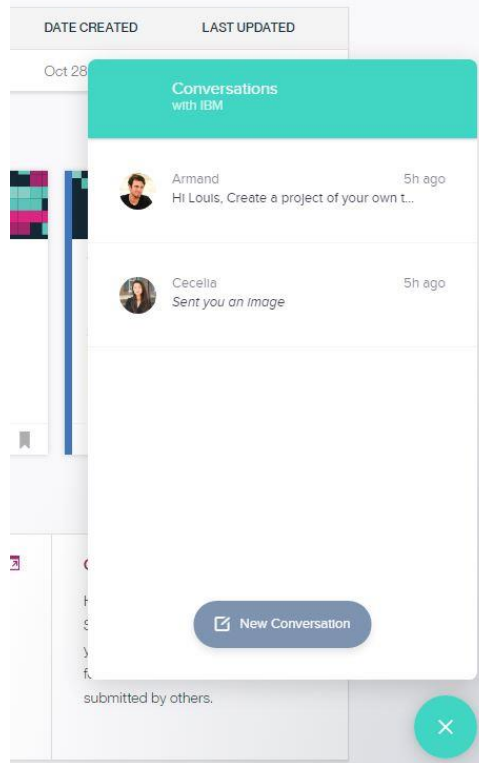


2. Live Chat

This is the home page of IBM Data Science Experience(DSX). Here you have all the tools that you need in a single place to **Learn, Create, and Collaborate**.

- On the bottom right-hand corner, you will see a **Live Chat** feature. Click on the **Chat** icon to launch Live Chat:



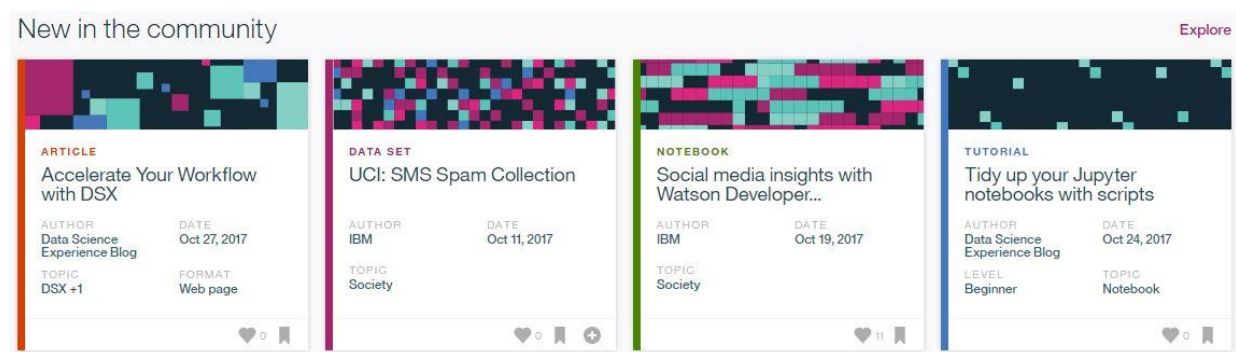


If you need assistance, you need only click on **New Conversation** to connect with a live person. Through this Live Chat feature, you can also continue conversations the next time you log into DSX.

We use feedback captured through **Live Chat** and the offerings instrumentation to guide our decisions in designing and developing **Data Science Experience**. We perform this analysis using DSX.

3. Community Cards

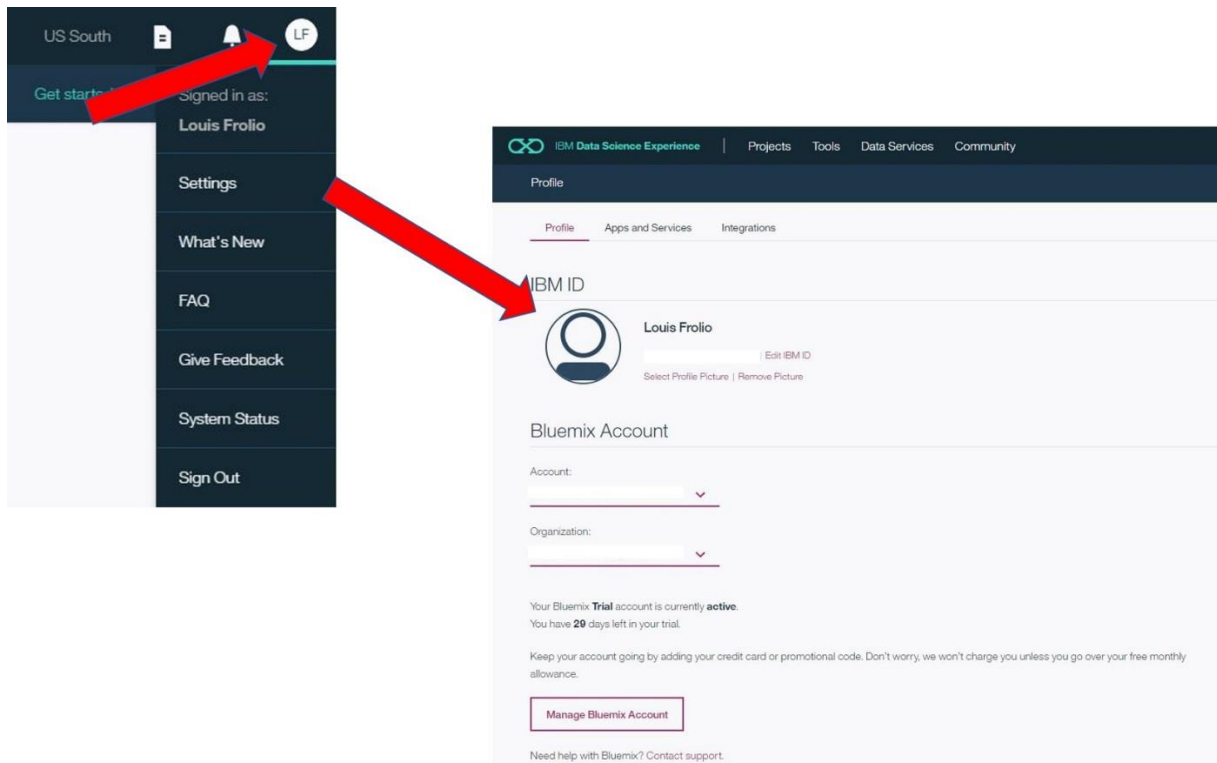
In the middle of the Home Page you see the **Community Cards**:



There are four types of cards – [Articles](#), [Data Sets](#), [Notebooks](#), and [Tutorials](#). These are designed to make it easier for you to learn about data science and experiment with its various tools and techniques.

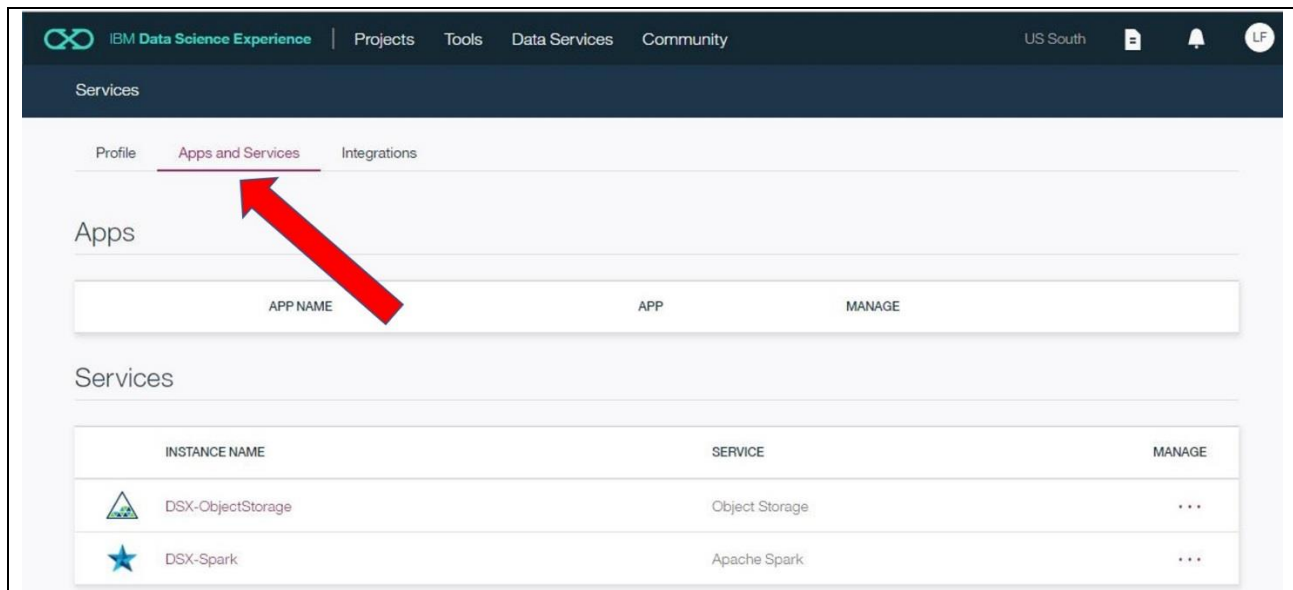
4. Profile Settings

- Click on [Settings](#) to look at your [Profile](#), [Apps and Services](#), and [Integrations](#). This is where you see the details of your Bluemix Account:



5. Apps and Services

- Click on [Apps and Services](#) to view all your current Bluemix account service instances:





Services

Profile **Apps and Services** Integrations

Apps

APP NAME	APP	MANAGE

Services

INSTANCE NAME	SERVICE	MANAGE
 DSX-ObjectStorage	Object Storage	...
 DSX-Spark	Apache Spark	...

Above is the default for the brand-new account, there is an instance of DSX Object Storage, and DSX Spark.

Integrations is where you configure DSX for GitHub integration.

End of Lesson 1

Lesson 2: Jupyter Notebook

Purpose:	This lesson introduces projects within DSX, their purpose, value, and how they are used to support collaboration. Also, Jupyter notebooks are introduced and used as part of a customer churn analysis using the R programming language.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Create and Configure DSX Project• Add Notebook Asset• Add Data Asset• Create Notebook Reference to Data Asset• Predict Customer Churn using Machine Learning Techniques• Evaluate Model Performance

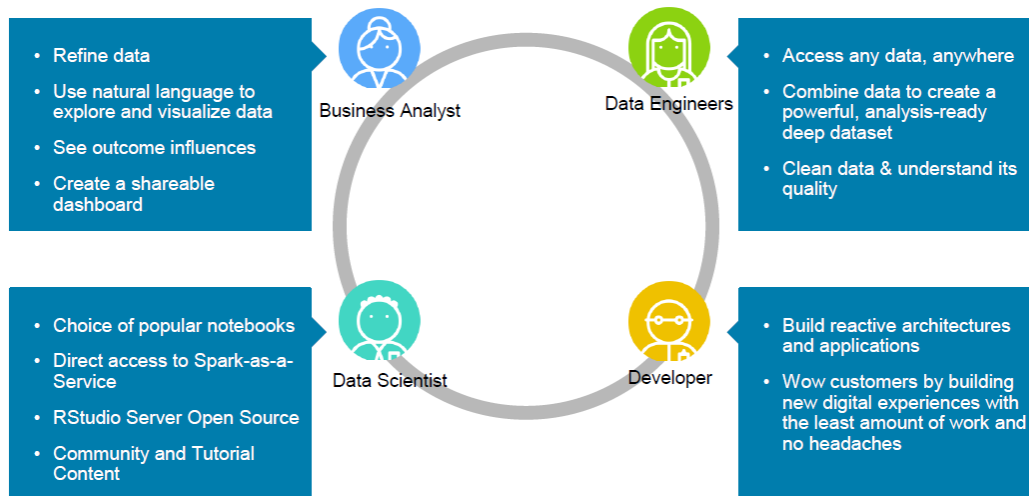
Lesson 2: Workflow Overview

- 1 • Project Overview
- 2 • Create New Project
- 3 • Create Notebook
- 4 • Load Data
- 5 • Bind Notebook to Data Asset
- 6 • Build and Evaluate Customer Churn Model

Lesson 2: Instructions

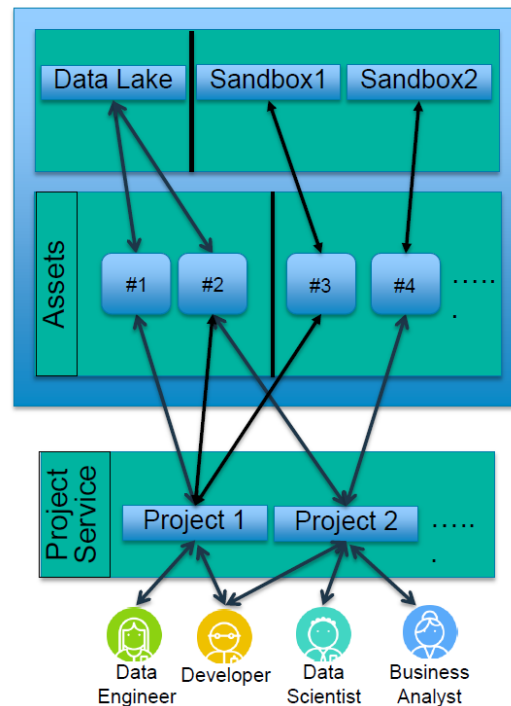
1. Project Overview

Data professionals need purpose-built, self-service communities that enable them to seamlessly collaborate across personas.



Projects make collaboration easier by:

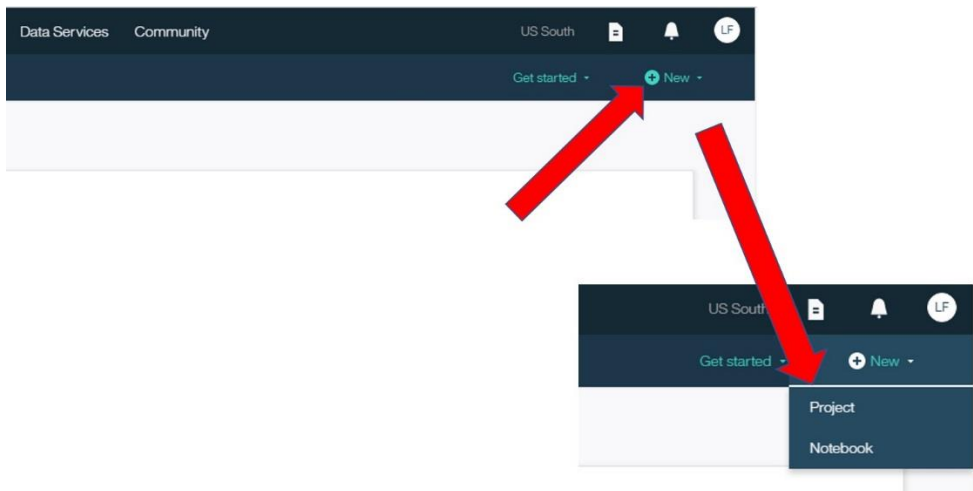
- Allowing different users and personas to share a set of assets
- Enabling users to collaborate and manage their notebooks, artifacts, plus more
- Providing three levels of rights: Viewers, Editors, and Admins



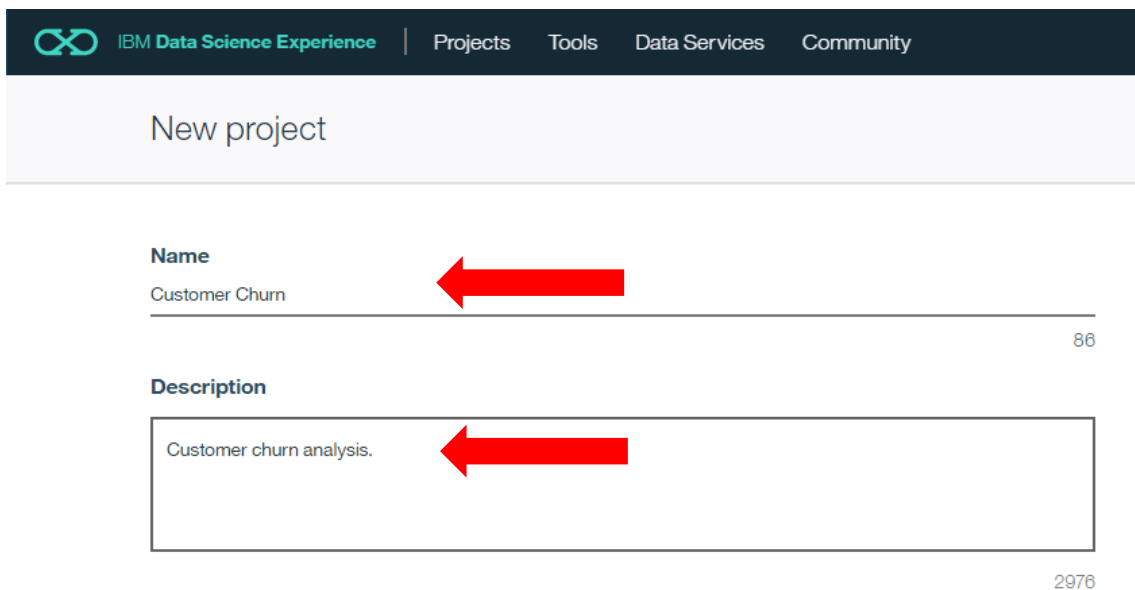
Action

2. Create New Project

- Navigate to <https://datascience.ibm.com>
- Login to DSX
- On the top right side, click **Create New and select project**



- Type the Project Name **Customer Churn**, add a meaningful description:



Name
Customer Churn

Description
Customer churn analysis.

Action

Ensure the defaults are selected as follows:

- Select your Spark Service **DSX-Spark**
- Select **Object Storage (Swift API)**
- Select Target Object Storage Instance **DSX-ObjectStorage**
- Default Target Container **CustomerChurn**

Spark service

DSX-Spark



If you associate the same Spark service with multiple projects, the Spark history server will display job history information for all the projects.

Storage type

☒ Object Storage (Swift API) ☐ Object Storage (S3 API) ☐ Cloud Object Storage (Beta)

Target object storage instance

DSX-ObjectStorage



Target container






CustomerChurn







243

- Click **Create**

Action

IBM Data Science Experience | [Projects](#) | [Tools](#) | [Data Services](#) | [Community](#)
US South   

[My Projects](#) > [CustomerChurn](#)

 Add to project
 





Overview
Assets
Bookmarks
Deployments
Collaborators
Settings

CustomerChurn

Last Updated: Oct 29 2017


0
0
1

Assets
Bookmarks
Collaborators

Date created
Oct 29 2017


Description
Customer churn analysis

Storage
0% of 5 GB used

Collaborators View all (1)
 Louis Frolo

Bookmarks View all (0)

Recent activity



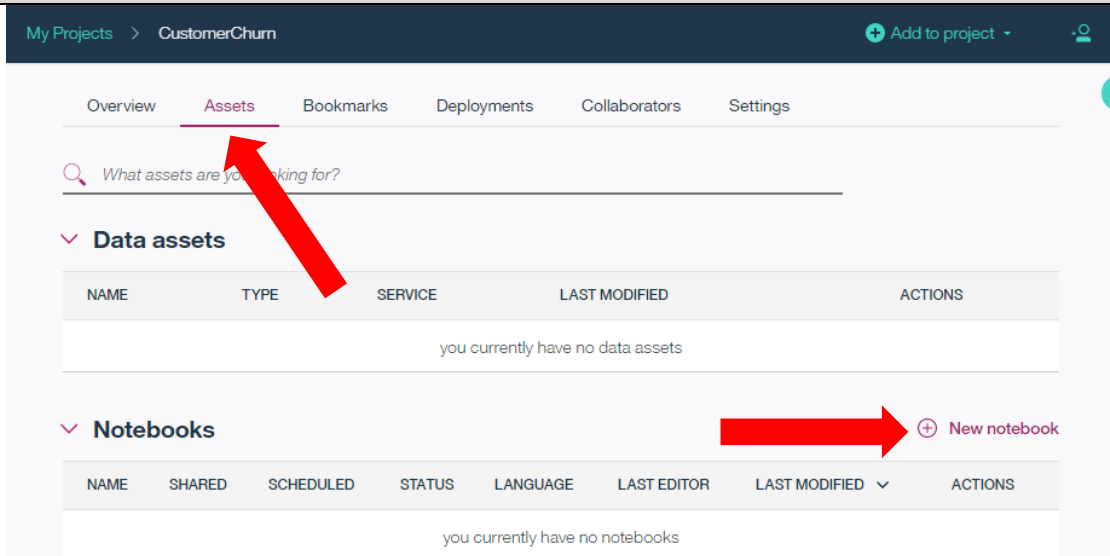
Alerts related to this project will show here when the project is active.

You now have a **Project** that is empty. You can use the tabs along the top to **add assets** to your project such as Connections, Notebooks, Data Assets, etc. You can also **add collaborators** to the Project.

3. Create Notebook

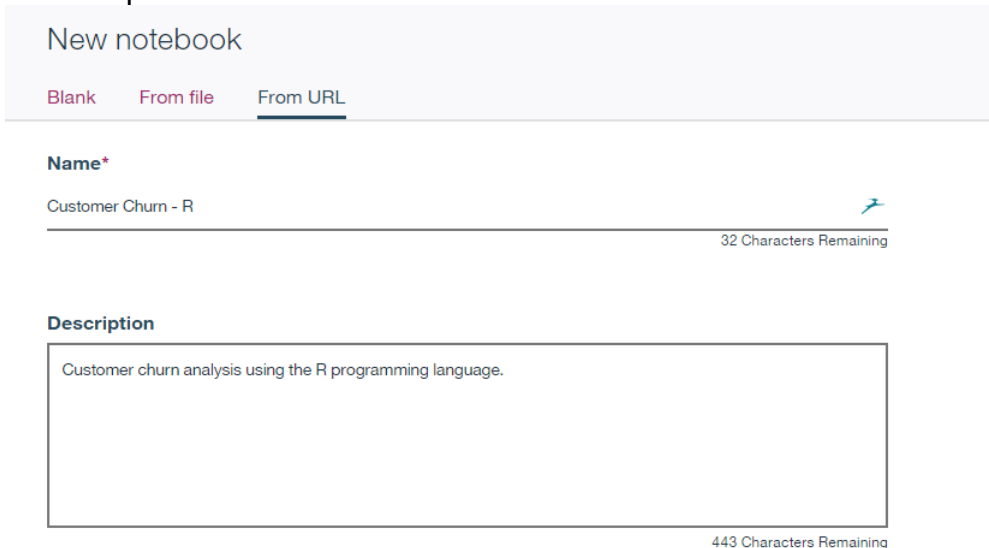
- Click **Assets**, then **Add Notebooks**

Action



The screenshot shows the IBM DSX interface for a project named 'CustomerChurn'. The 'Assets' tab is selected, and a red arrow points to the search bar. Below the search bar, there are sections for 'Data assets' and 'Notebooks'. The 'Notebooks' section has a red arrow pointing to the '+ New notebook' button.

- Choose **From URL** from the tab, give the notebook a name and meaningful description:



The screenshot shows the 'New notebook' form. The 'From URL' tab is selected. The 'Name' field is filled with 'Customer Churn - R' and has a character count of 32 Characters Remaining. The 'Description' field is filled with 'Customer churn analysis using the R programming language.' and has a character count of 443 Characters Remaining.

- In a separate browser window navigate to:
<https://github.com/team-wolfpack/DSX-Hands-on-Workshop>
- Click on Notebooks, right click on **CustomerChurn-R.ipynb** then choose **Copy link address**. Go back to the **DSX New Notebook** page.

Action

Paste URL into **Notebook URL** text box then choose **Create Notebook**:

Notebook URL*

`https://github.com/team-wolfpack/DSX-Hands-on-Workshop/blob/master/Notebooks/CustomerChurn-R.ipynb`

Spark service*

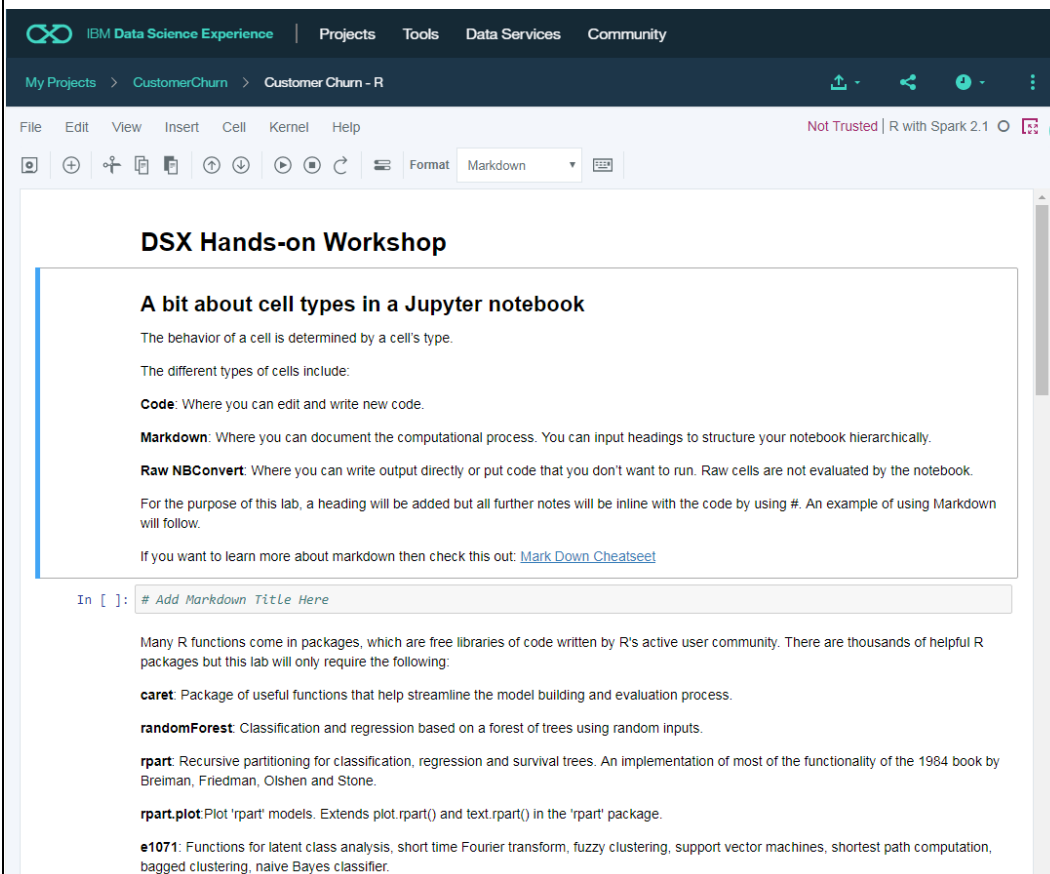
DSX-Spark

Associate this notebook with the Spark Service of your choice.

Cancel

Create Notebook

You should now see something like this:



The screenshot shows the IBM Data Science Experience Jupyter Notebook interface. The top navigation bar includes 'IBM Data Science Experience', 'Projects', 'Tools', 'Data Services', and 'Community'. The breadcrumb trail shows 'My Projects > CustomerChurn > Customer Churn - R'. The notebook title is 'DSX Hands-on Workshop'. The content area displays a markdown document titled 'A bit about cell types in a Jupyter notebook'. The text explains that cell behavior is determined by its type and lists three types: Code, Markdown, and Raw NBConvert. It also provides a link to a 'Mark Down Cheatsheet'. Below the text, there is a code cell with the following content:

```
In [ ]: # Add Markdown Title Here
```

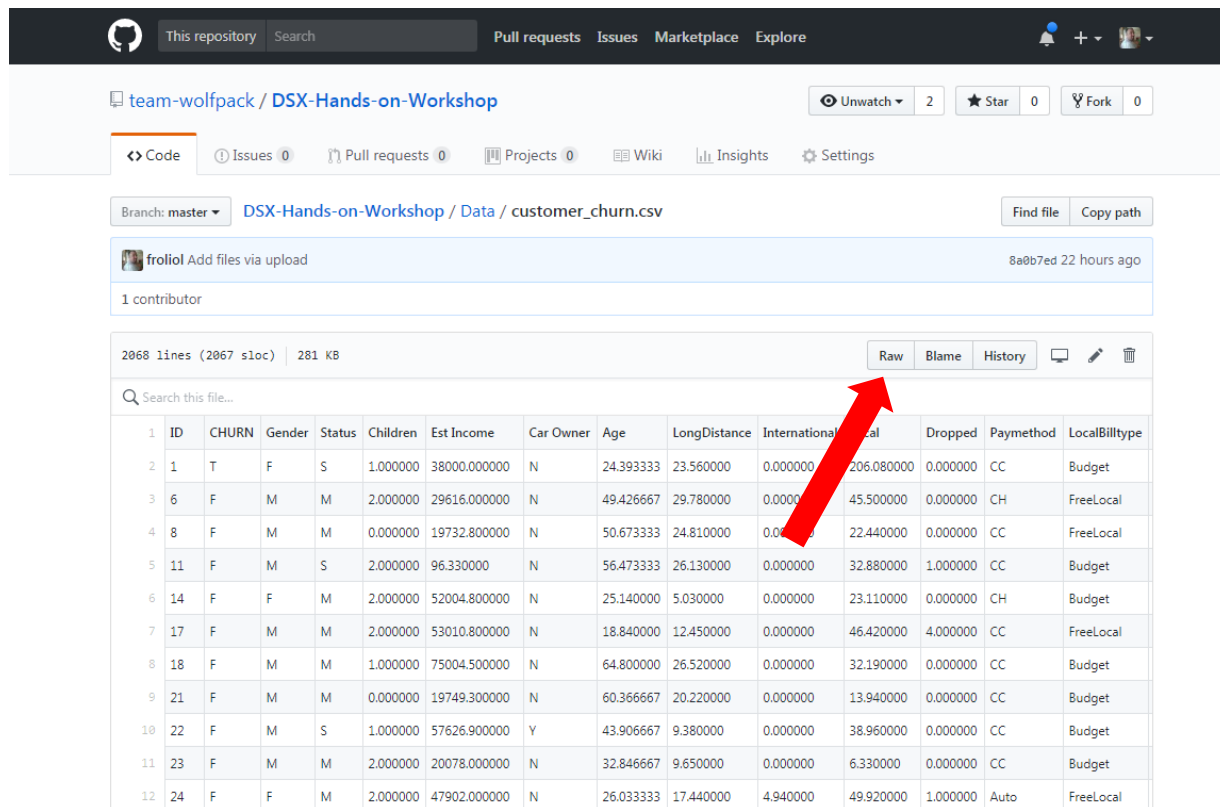
The code cell contains several R packages and their descriptions:

- caret**: Package of useful functions that help streamline the model building and evaluation process.
- randomForest**: Classification and regression based on a forest of trees using random inputs.
- rpart**: Recursive partitioning for classification, regression and survival trees. An implementation of most of the functionality of the 1984 book by Breiman, Friedman, Olshen and Stone.
- rpart.plot**: Plot 'rpart' models. Extends plot.rpart() and text.rpart() in the 'rpart' package.
- e1071**: Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier.

Action

4. Load Data

- In a separate browser, navigate once again to Github:
- <https://github.com/team-wolfpack/DSX-Hands-on-Workshop>
- Click on customer_churn.csv, you should see a tabular list of data. Click on the Raw button:



team-wolfpack / DSX-Hands-on-Workshop

Unwatch 2 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Branch: master DSX-Hands-on-Workshop / Data / customer_churn.csv Find file Copy path

frolol Add files via upload 8a0b7ed 22 hours ago

1 contributor

2068 lines (2067 sloc) 281 KB

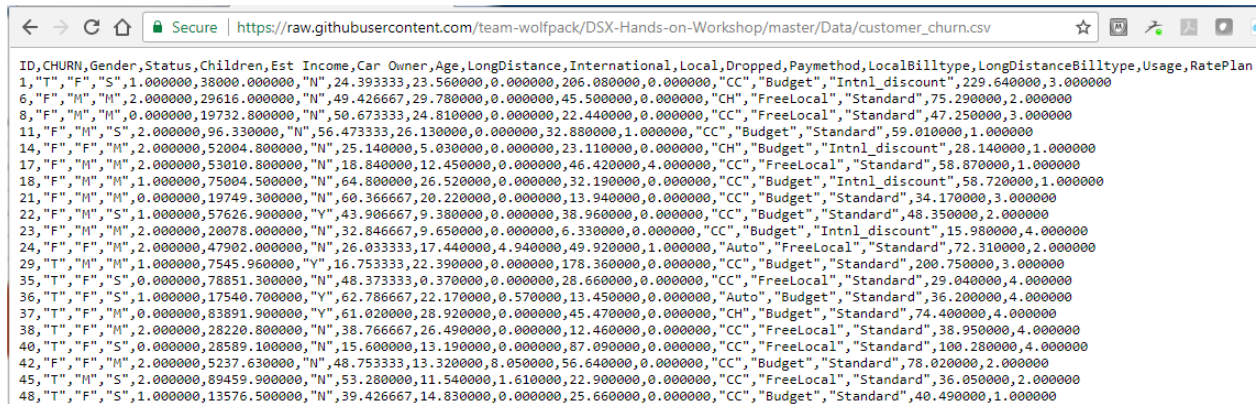
Raw Blame History

Search this file...

	ID	CHURN	Gender	Status	Children	Est Income	Car Owner	Age	LongDistance	International	Local	Dropped	Paymethod	LocalBilltype
1	1	T	F	S	1.000000	38000.000000	N	24.393333	23.560000	0.000000	206.080000	0.000000	CC	Budget
3	6	F	M	M	2.000000	29616.000000	N	49.426667	29.780000	0.000000	45.500000	0.000000	CH	FreeLocal
4	8	F	M	M	0.000000	19732.800000	N	50.673333	24.810000	0.000000	22.440000	0.000000	CC	FreeLocal
5	11	F	M	S	2.000000	96.330000	N	56.473333	26.130000	0.000000	32.880000	1.000000	CC	Budget
6	14	F	F	M	2.000000	52004.800000	N	25.140000	5.030000	0.000000	23.110000	0.000000	CH	Budget
7	17	F	M	M	2.000000	53010.800000	N	18.840000	12.450000	0.000000	46.420000	4.000000	CC	FreeLocal
8	18	F	M	M	1.000000	75004.500000	N	64.800000	26.520000	0.000000	32.190000	0.000000	CC	Budget
9	21	F	M	M	0.000000	19749.300000	N	60.366667	20.220000	0.000000	13.940000	0.000000	CC	Budget
10	22	F	M	S	1.000000	57626.900000	Y	43.906667	9.380000	0.000000	38.960000	0.000000	CC	Budget
11	23	F	M	M	2.000000	20078.000000	N	32.846667	9.650000	0.000000	6.330000	0.000000	CC	Budget
12	24	F	F	M	2.000000	47902.000000	N	26.033333	17.440000	4.940000	49.920000	1.000000	Auto	FreeLocal

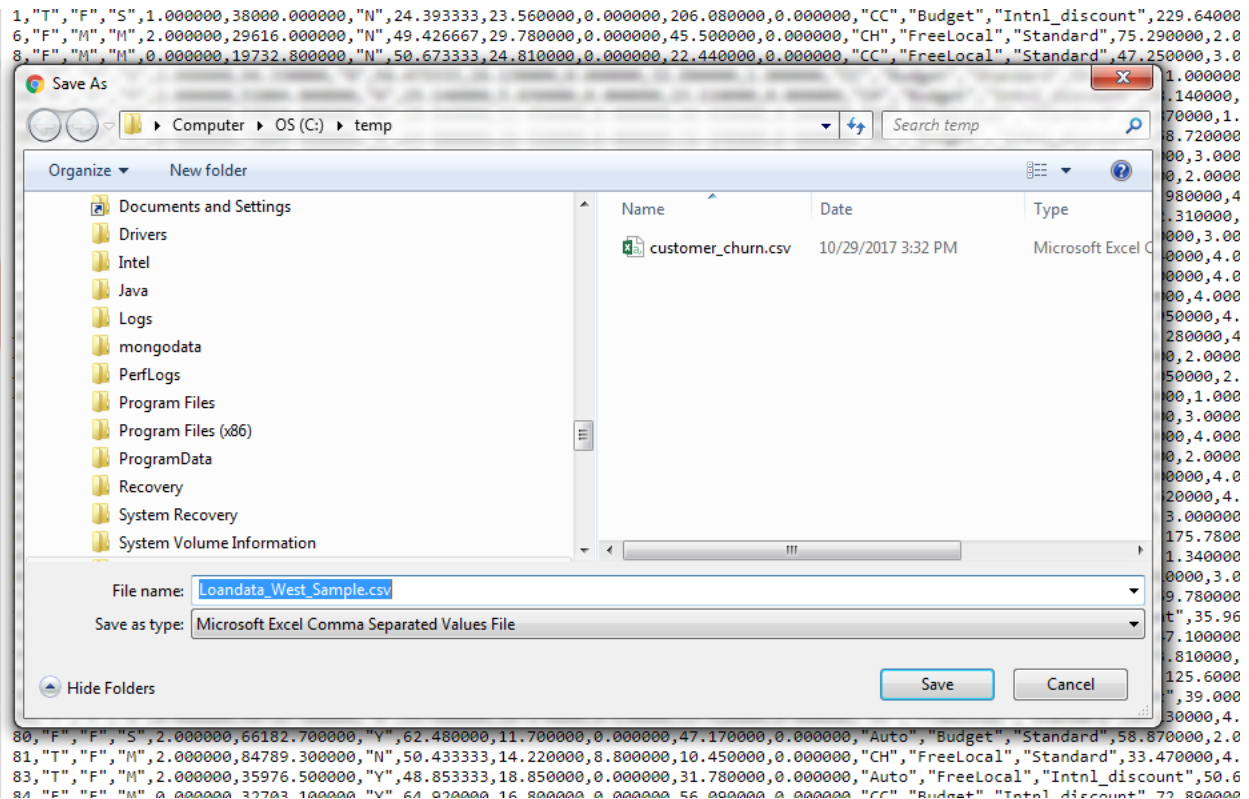
Action

You should now see the raw output from the csv file:



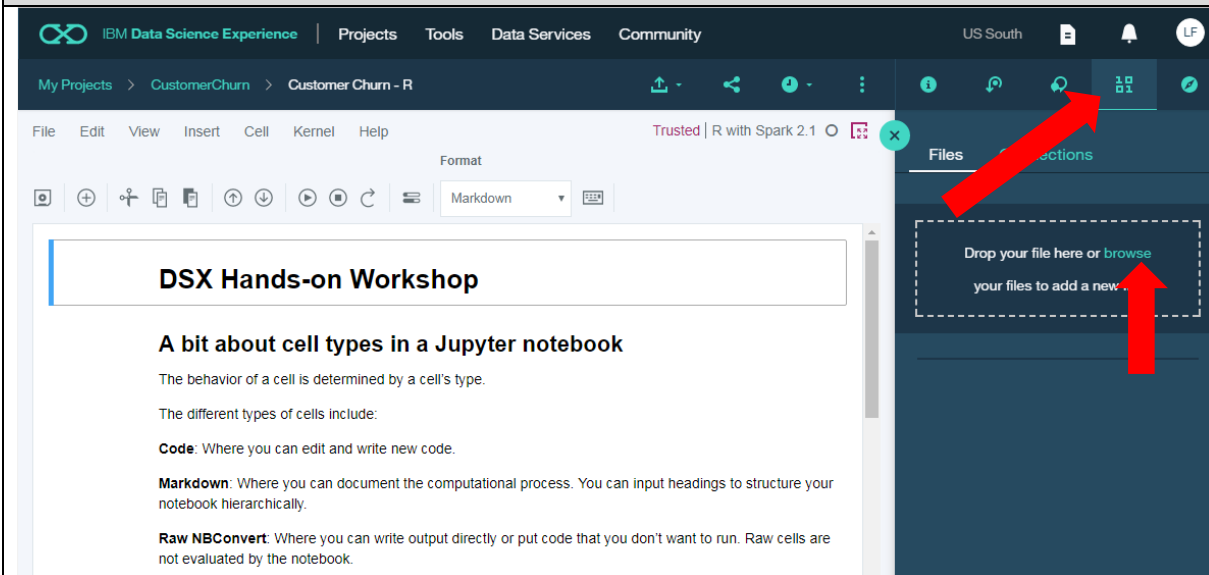
```
ID,CHURN,Gender,Status,Children,Est Income,Car Owner,Age,LongDistance,International,Local,Dropped,Paymethod,LocalBilltype,LongDistanceBilltype,Usage,RatePlan
1,"T","F","S",1.000000,38000.000000,"N",24.393333,23.560000,0.000000,206.080000,0.000000,"CC","Budget","Intl_discount",229.640000,3.000000
6,"F","M","M",2.000000,29616.000000,"N",49.426667,29.780000,0.000000,45.500000,0.000000,"CH","FreeLocal","Standard",75.290000,2.000000
8,"F","M","M",0.000000,19732.800000,"N",50.673333,24.810000,0.000000,22.440000,0.000000,"CC","FreeLocal","Standard",47.250000,3.000000
11,"F","M","S",2.000000,96.330000,"N",56.473333,26.130000,0.000000,32.880000,1.000000,"CC","Budget","Standard",59.010000,1.000000
14,"F","F","M",2.000000,52004.800000,"N",25.140000,5.030000,0.000000,23.110000,0.000000,"CH","Budget","Intl_discount",28.140000,1.000000
17,"F","M","M",2.000000,53010.800000,"N",18.840000,12.450000,0.000000,46.420000,4.000000,"CC","FreeLocal","Standard",58.870000,1.000000
18,"F","M","M",1.000000,75004.500000,"N",64.800000,26.520000,0.000000,32.190000,0.000000,"CC","Budget","Intl_discount",58.720000,1.000000
21,"F","M","M",0.000000,19749.300000,"N",60.366667,20.220000,0.000000,13.940000,0.000000,"CC","Budget","Standard",34.170000,3.000000
22,"F","M","S",1.000000,57626.900000,"Y",43.906667,9.380000,0.000000,38.960000,0.000000,"CC","Budget","Standard",48.350000,2.000000
23,"F","M","M",2.000000,20078.000000,"N",32.846667,9.650000,0.000000,6.330000,0.000000,"CC","Budget","Intl_discount",15.980000,4.000000
24,"F","F","M",2.000000,47902.000000,"Y",26.033333,17.440000,4.940000,49.920000,1.000000,"Auto","FreeLocal","Standard",72.310000,2.000000
29,"T","M","M",1.000000,17540.700000,"Y",16.753333,22.390000,0.000000,178.360000,0.000000,"CC","Budget","Standard",200.750000,3.000000
35,"T","F","S",0.000000,78851.300000,"Y",48.373333,0.370000,0.000000,28.660000,0.000000,"CC","FreeLocal","Standard",29.040000,4.000000
36,"T","M","M",1.000000,17540.700000,"Y",62.786667,22.170000,0.570000,13.450000,0.000000,"Auto","Budget","Standard",36.200000,4.000000
37,"T","F","M",0.000000,83891.900000,"Y",61.020000,28.920000,0.000000,45.470000,0.000000,"CH","Budget","Standard",74.400000,4.000000
38,"T","F","M",2.000000,28220.800000,"N",38.766667,26.490000,0.000000,12.460000,0.000000,"CC","FreeLocal","Standard",38.950000,4.000000
40,"T","F","S",0.000000,28589.100000,"N",15.600000,13.190000,0.000000,87.090000,0.000000,"CC","FreeLocal","Standard",100.280000,4.000000
42,"F","F","M",2.000000,5237.630000,"N",48.753333,13.320000,8.050000,56.640000,0.000000,"CC","Budget","Standard",78.020000,2.000000
45,"T","M","S",2.000000,89459.900000,"N",53.280000,11.540000,1.610000,22.900000,0.000000,"CC","FreeLocal","Standard",36.050000,2.000000
48,"T","F","S",1.000000,13576.500000,"N",39.426667,14.830000,0.000000,25.660000,0.000000,"CC","Budget","Standard",40.490000,1.000000
```

- Right click on the output and choose “Save as” and then save the output to your local file system with the file’s original name “customer_churn.csv.”

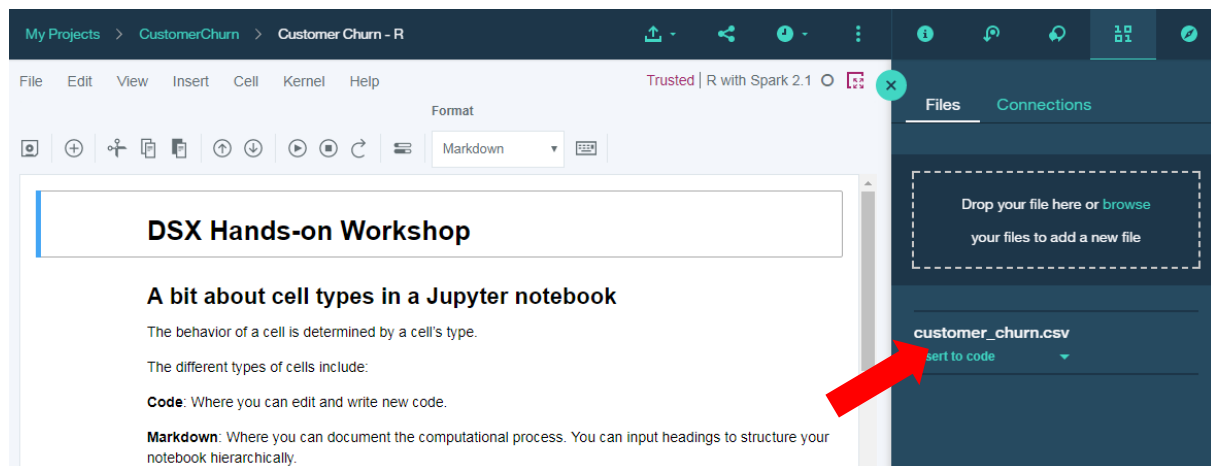


- Go back to the CustomerChurn-R notebook and then click on the Data icon at the top right of the screen:

Action



A new panel will be presented with Files highlighted. Click on browse, navigate to the customer_churn.csv file and select it. You should now see that the file has been imported into the project:



5. Bind Notebook to Data Asset

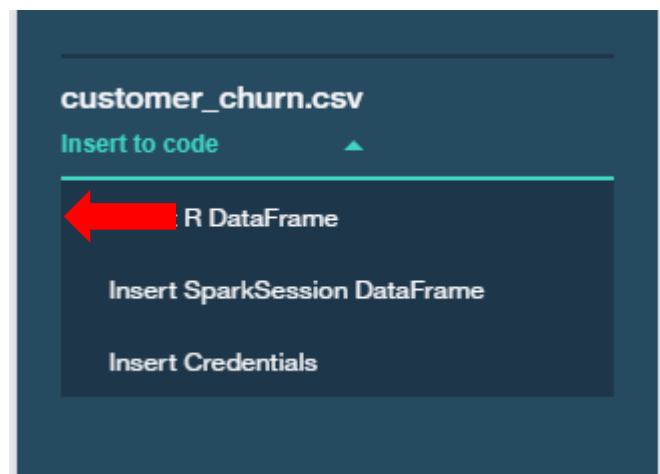
Although the data is part of the project the notebook has no reference to it. Let's now add a reference so that we can analyze the data.

Action

- In the notebook scroll down to the fifth cell:

```
In [ ]: # Placeholder for R Data Frame Auto-code
```

- Place the cursor in the cell and beneath the comment. Navigate the side panel where the data set is displayed, click on “insert into code”:



- Choose “Insert R DataFrame” to insert auto-generated code that will allow the notebook to access the data stored on Bluemix:

```
In [ ]: # Placeholder for R Data Frame Auto-code
# @hidden_cell
# This function accesses a file in your Object Storage. The definition contains your credentials.
# You might want to remove those credentials before you share your notebook.
getObjectStorageFileWithCredentials_d5fa59dc7e72461489ab1f8be43ed5a0 <- function(container, filename) {
  # This functions returns a textConnection object for a file
  # from Bluemix Object Storage.

  if(!require(httr)) install.packages('httr')
  if(!require(RCurl)) install.packages('RCurl')
  library(httr, RCurl)
  auth_url <- paste("https://identity.open.softlayer.com", '/v3/auth/tokens', sep= '')
  auth_args <- paste('/auth', 'identity', 'password', 'user', 'domain', 'id', '26746425e3e44h7383e
```

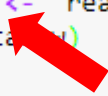
Look at the last line of the newly inserted data frame, particularly the name assigned to it:

```
df.data.1 <- read.csv(file = getObjectStorageFileWithCrede
head(df.data.1)
```


Action

Let's make it more friendly. Change "df.data.1" to "custDataRow."

```
custDataRow <- read.csv(file = getObjectStorage  
head(custDataRow))
```



6. Build and Evaluate Customer Churn Model

[Lesson 2 Continued in Jupyter Notebook](#)

Lesson 3: Machine Learning Flows

Purpose:	This lesson introduces Machine Learning Flows in DSX. Flows provide a graphical approach to machine learning like that of SPSS Modeler.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Create Machine Learning Flow• Import Data• Leverage Flows' Palette to Orchestrate Customer Churn Machine Learning Pipeline• Evaluate Customer Churn Model

Lesson 3: Workflow Overview

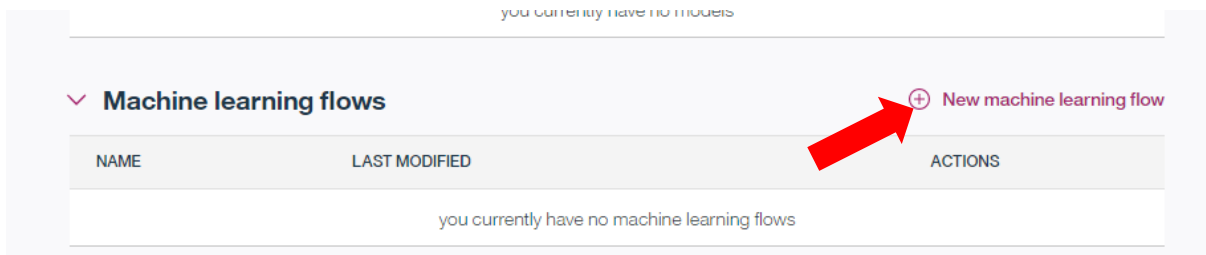
- 1 • Create Machine Learning Flow
- 2 • Add Data Asset
- 3 • Add & Configure Type Object
- 4 • Add & Configure Model Objects
- 5 • Run Flow to Create Nuggets
- 6 • Add & Configure Analysis Object - Measure Performance
- 7 • Add Second Model Technique to Flow

Lesson 3: Instructions

Action

1. Create Machine Learning Flow

- Navigate to CustomerChurn project page
- Click on “**New machine learning flow**”



- Choose “**Create flow**” on the top menu. Give the flow a meaningful name and description. For “Runtime” choose “**IBM SPSS Modeler**”:

My Projects > CustomerChurn > New Flow

New flow BETA

Create flow From file

Name*

CustomerChurn-Flow

Description

DSX machine learning flow for customer churn

Runtime

IBM SPSS Modeler

Cancel Create Flow

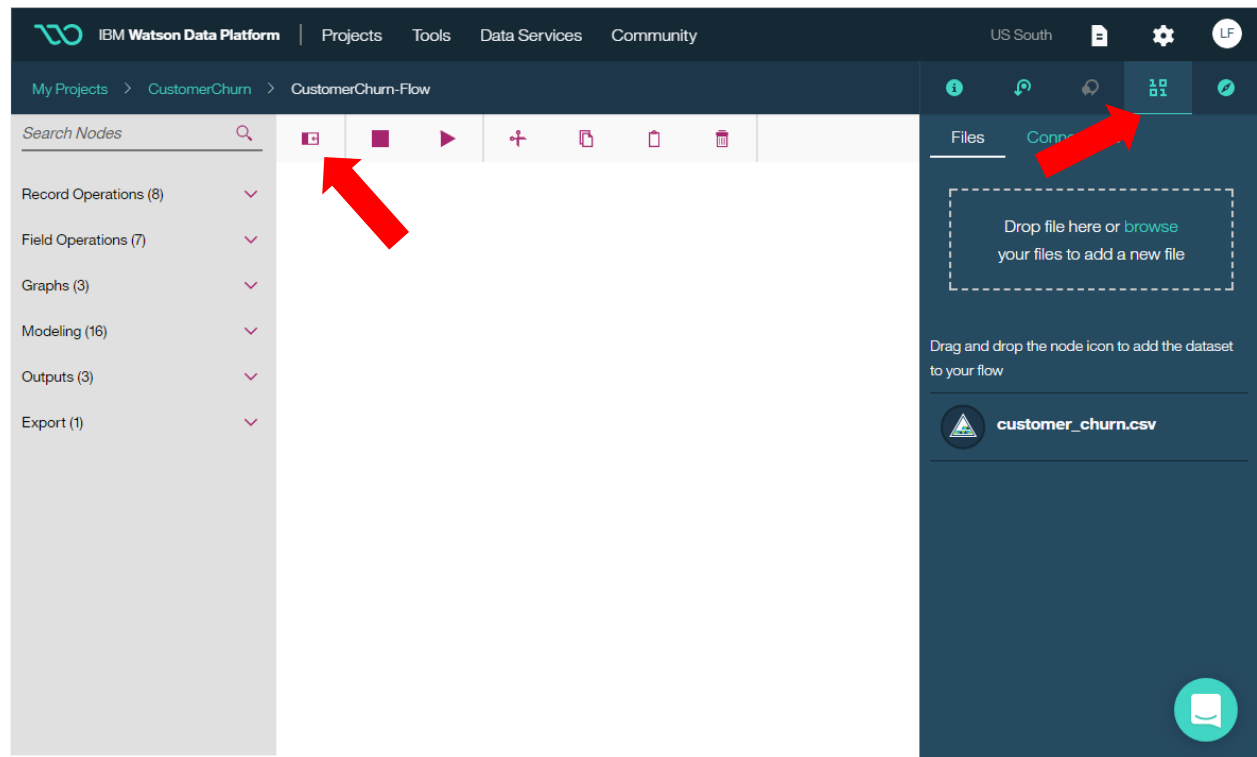
Action

Click on “Create flow.”

2. Add Data Asset

You should now see an empty workspace.

- On the top left click on the “Palette” icon, and on the top right click on the “Find and Add Data” icon.



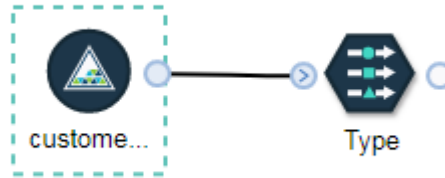
The palette represents the set of tools available for use with DSX flows. The menu of the right should look familiar.

- Let’s start by dragging and dropping the “[customer_churn.csv](#)” file onto the workspace.

3. Add & Configure Type Object

- From the palette, expand “[Field Operations](#)”, then drag and drop “[Type](#)” onto the workspace and to the right of “customer_churn.csv”. Connect the two objects:

Action



- Double click on “Type”, click on “**Add Columns**” then add all the columns. Click on the arrow back to the “Settings” page of the “Type” object. For the “CHURN” column, change its Role to that of “**Target.**” Leave the default for the remaining columns:

Type

Settings
Annotations

Default mode
☒ Read metadata
☐ Pass (do not scan)

Types

-
+
Add Columns

Field	Measure	Role	Value mode	Values
ID	Default	Input	Read	
CHURN	Default	Target	Read	
Gender	Def	Input	Read	
Status	Default	Input	Read	

OK
Cancel

- Click “OK” to exit.

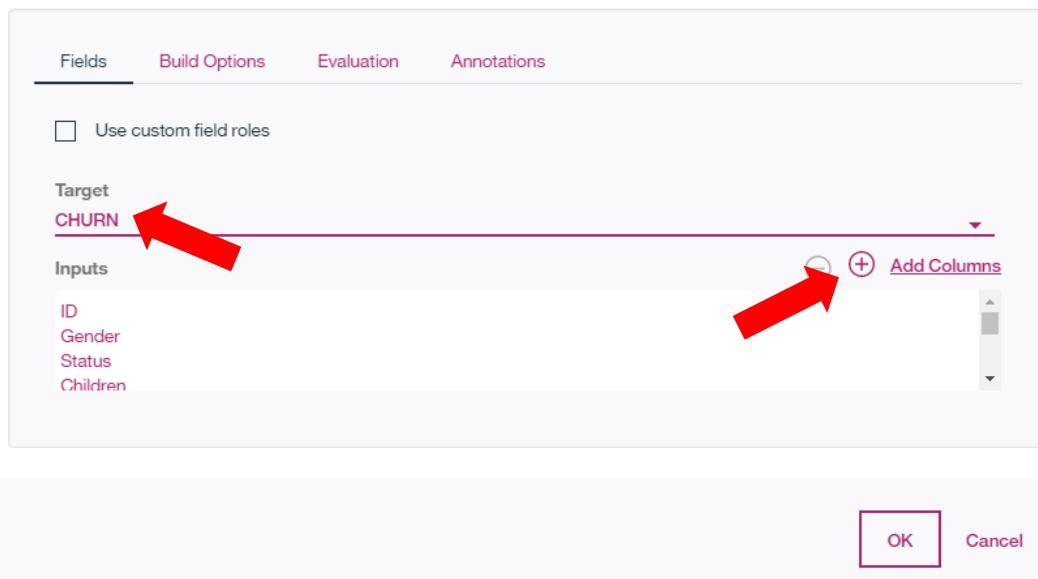
Action

4. Add & Configure Model Object

- From the palette, expand the “**Modeling**” branch then drag “**C&R Tree**” onto the workspace to the right of “Type.”
- Connect the two then double click on “C&R Tree” to edit its properties.

The “Target” should identify “CHURN” automatically:

CHURN



Fields Build Options Evaluation Annotations

☐ Use custom field roles

Target
CHURN

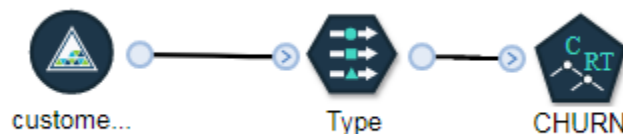
Inputs

ID
Gender
Status
Children

+ Add Columns

OK Cancel

- Click on “Add Columns.” Recall from the notebook exercise you were asked to jot down the top 10 fields that were identified as the greatest influencers. Choose those columns as inputs to the decision tree model. Click “OK” to return to the workspace:

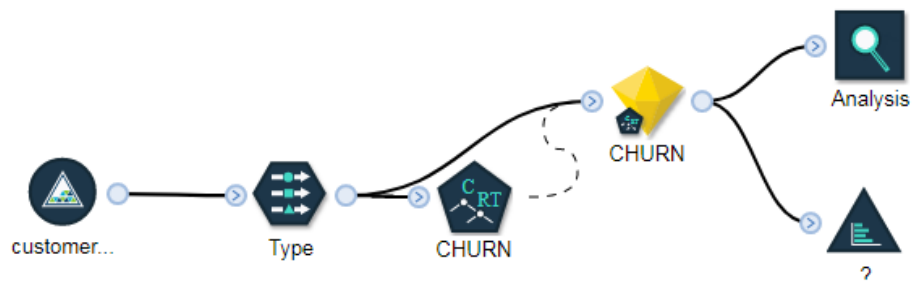


5. Run Flow to Create Nugget

- Run the flow by clicking on the “Run” icon at the top of the workspace.

You should see a new forth object on the workspace, this is called a nugget.

- From the palette add an “**Analysis**” object to the workspace, you will find it under the “Outputs” drop down. Also, from the “**Graphs**” drop down add a “**Distribution**” object to the workspace. Connect the nugget to each of them:



6. Add & Configure Analysis Object – Measure Model Performance

- Double click on “**Analysis**” and check off the four checkboxes, leave the rest as default:

Analysis

Settings

Annotations

☒ Coincidence matrices (categorical targets)
 ☒ Performance evaluation
 ☒ Evaluation metric (binary classifiers)
 ☒ Confidence figures (if available)

Threshold for pct. correct
 90

Improve accuracy multiplier
 2

☒ Separate by partition

Break down by

ID
 CHURN
 Gender
 Status

OK

Cancel

- Click “OK” to return to the workspace.
- Double click on “Analysis” and configure it as depicted below:

?

Plot

Appearance

Annotations

Plot

☒ Specified
 ☐ All flags (true values)

Field (discrete)

CHURN

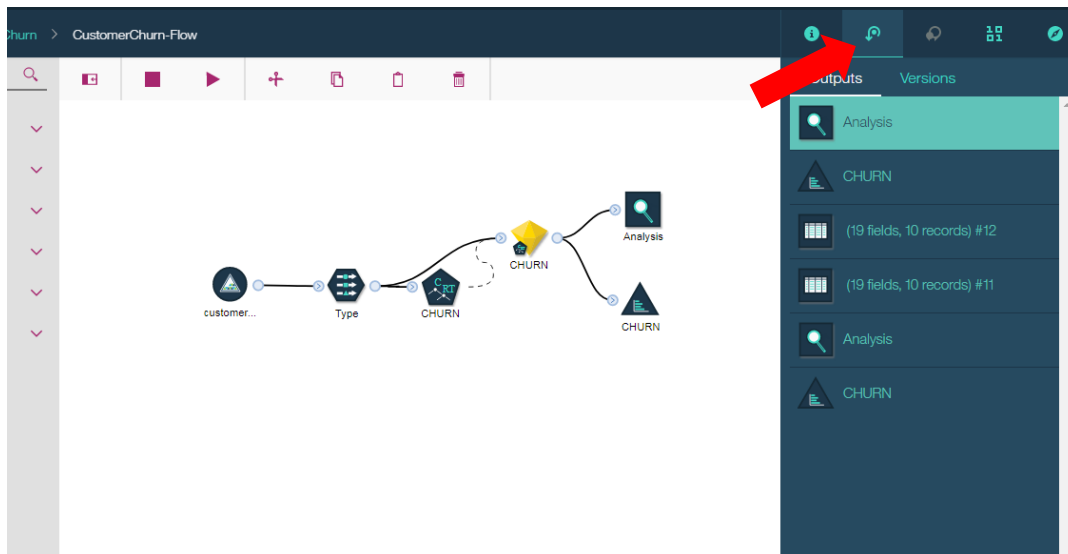
Color (discrete)

\$R-CHURN

☒ Normalize by color

☐ Use proportional scale

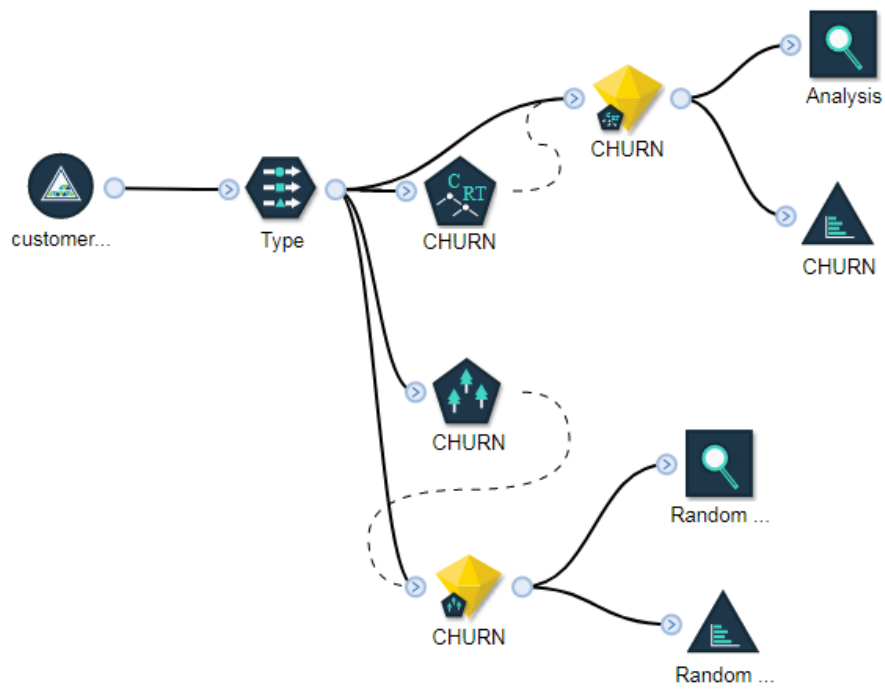
- Click on “OK” to return to the workspace.
- Run the flow again.
- On the right side of the workspace click on the “**Outputs and Versions**” icon to see the resulting analysis:



- Explore the results

7. Add Second Modeling Technique to Flow

- To the palette repeat the process for “**Random Trees**” that you did for “C&R Trees.” Your resulting workspace should look like the following:



- Explore the results.

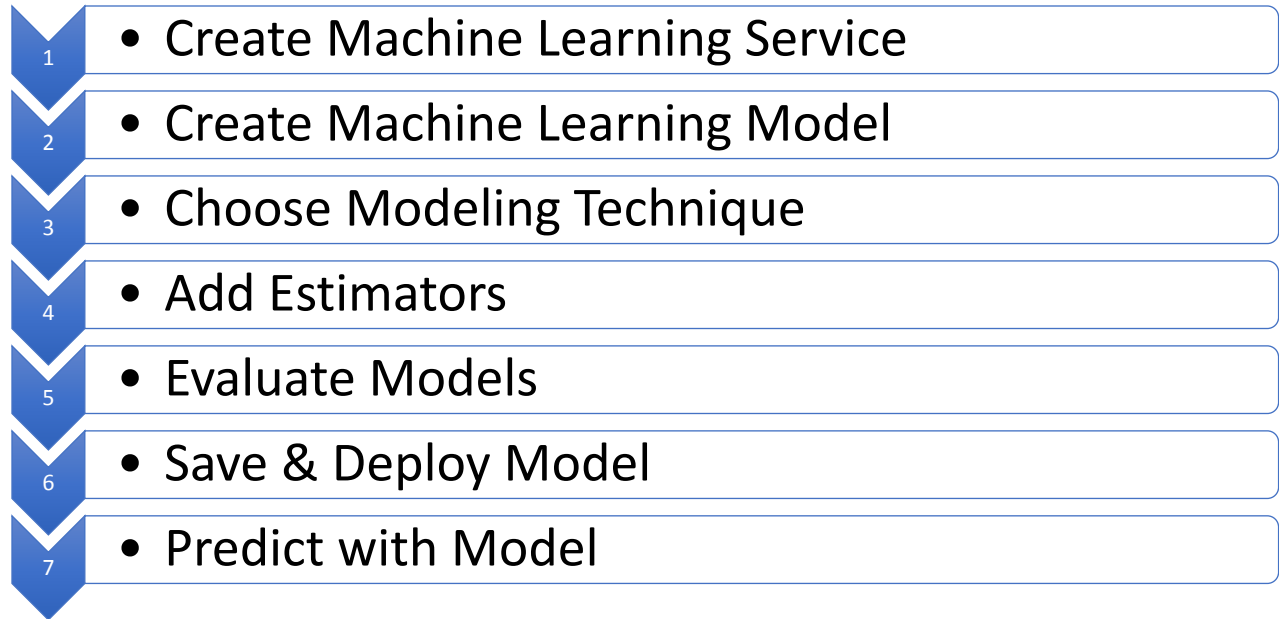
End of Lesson 3



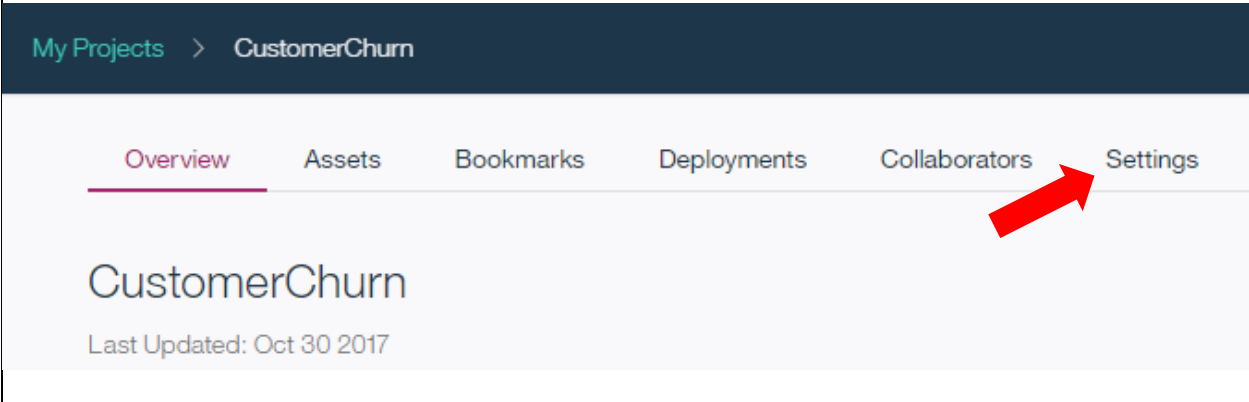
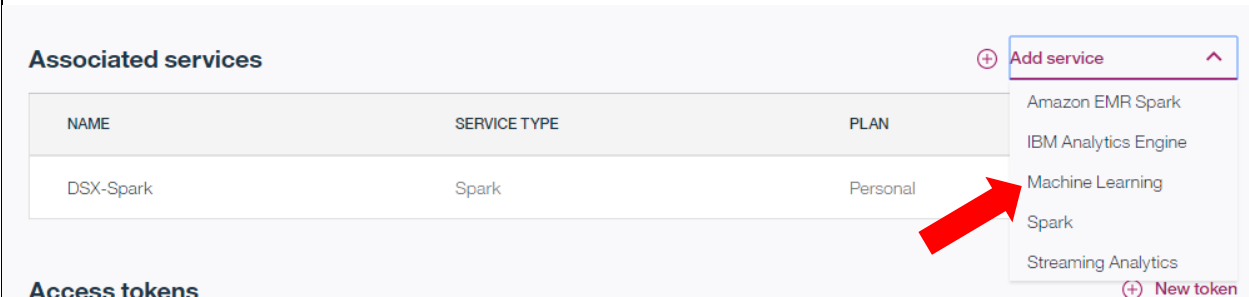
Lesson 4: Watson Machine Learning

Purpose:	This lab introduces Watson Machine Learning in DSX. Watson Machine Learning makes the task of machine learning easy with as little as a few clicks of the mouse.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Creation of requisite services to support Watson Machine Learning• Creation of Watson Machine Learning Models• Model Performance Evaluation• Deployment and Prediction of Model

Lesson 4: Workflow Overview



Lesson 4: Instructions

Action
<p>1. Create Machine Learning Service</p> <ul style="list-style-type: none"> Navigate to CustomerChurn project page At the top click on the “Settings” icon:

<p>Scroll to the middle of the page and click on “Add service” then choose “Machine Learning”:</p>

<ul style="list-style-type: none"> On the Machine Learning page make sure that the tab is set to “New”, for the plan choose “Free”:

Action

Existing
New

Machine Learning

IBM Watson Machine Learning is a full-service Bluemix offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

SPSS analytics platform features

SPSS streams management and deployment with realtime scoring and batch processing options.

Spark and Python Machine Learning features

Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch (beta) and streaming (beta).

Integration with Data Science Experience

Visit <http://datascience.ibm.com>. Create and train predictive analytics models with the best tools and the latest expertise in a social environment built by data scientists.

Pricing Plan: Monthly Process shown above reflect the: United States

Plan	Features	Pricing
<div style="display: flex; align-items: center;"> <div style="width: 15px; height: 15px; background-color: #c00000; border-radius: 50%; margin-right: 5px;"></div> Free </div>	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free

- Click on “[Create IBM Watson Machine Learning.](#)”
- At the confirmation page you can give your service a meaningful name:

Confirm Creation

Organization: louis@datatechblog.com

Plan

Free

Space

WatsonDataPlatform

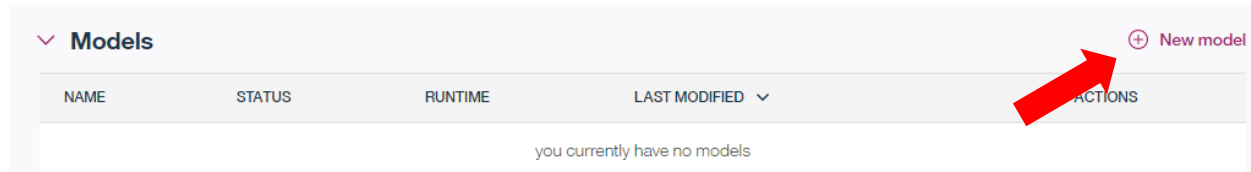
Service Name

DSX-Lab-Machine_Learning

- Click “Confirm” to create Watson Machine Learning Service.

2. Create Machine Learning Model

- In the Project click on “**Assets**” at the top of the window.
- In the middle of the page you will see “**Models**”, click on “New model”:



- In the “New model” window give your model a meaningful name and description, you should also see the machine learning service you just created. Click on “**Manual**” then “**Create**”:

New model BETA

Name
 CustomerChurn-WML 83

Description

Customer churn using Watson Machine Learning

256

Machine Learning Service
 DSX-Lab-Machine_Learning v

Spark Service
 DSX-Spark v

Automatic
 Prepare my data and create a model automatically



Manual
 Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) .

- When complete you will be prompted for a data asset, choose “customer_churn.csv”, then click “Next.”

Select data asset

The model builder currently supports CSV files.

NAME	TYPE	SERVICE
  customer_churn.csv	CSV	Object Storage (Swift API)

3. Choose Modeling Technique

- At the “**Select a Technique**” screen select “**CHURN**” as the “**Column value to predict**”, and for the feature columns choose the top 10 features identified in the Jupyter notebook lab. Also, make sure “**Binary Classification**” is highlighted:

Select a technique


Column value to predict (Label Col)


CHURN (String) ▼


Feature columns

Gender (String), Status (String), Children (Decimal), Est Income (Decimal), Age (Decimal), LongDistance (Decimal), Local (Decimal), Paymethod (String), Usage (Decimal) ✕

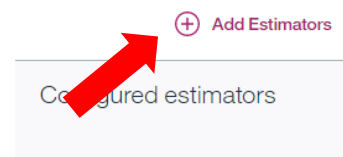
✓ Suggested technique.


Binary Classification
 Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.


Multiclass Classification
 Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.


Regression
 Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split




4. Add Estimators

- In the upper right-hand corner of the screen you will see “**Add Estimators**”, click on the icon. In the “Select estimator(s)” screen choose **Decision Tree Classifier**, and **Random Forest Classifier**:


Select estimator(s)

What type of estimator are you looking for?




Logistic Regression

Analyzes a data set in which there are one or more independent variables that determine one of two outcomes. Only binary L...




Decision Tree Classifier

Maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in...



Random Forest Classifier

Constructs multiple decision trees to produce the label that is a mode of each decision tree. It supports both binary and ...



Gradient Boosted Tree Classifier

Produces a classification prediction model in the form of an ensemble of decision trees. It only supports binary labels, a...

- Click “Add”:

Select a technique


You cannot change label column, feature columns, model type, or validation split after adding an estimator.
You must first delete all estimators in order to make changes to these attributes.

Column value to predict (Label Col)

CHURN (String)


Feature columns

Est Income (Decimal), Age (Decimal), LongDistance (Decimal), Status (S




Binary Classification

Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.



Multiclass Classification

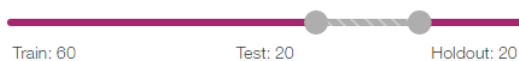
Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.



Regression


Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split




+ Add Estimators

Configured estimators



Decision Tree Classifier

Not Yet Trained



Random Forest Classifier

Not Yet Trained

- Click “Next” to train models. This will take 1-2 minutes with the data set we are using:

5. Evaluate Models

Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
<input checked="" type="radio"/>	RandomForestClassifier	Trained & Evaluated	Excellent	0.97826	0.97383	30 Oct 2017, 5:38 PM	...
<input type="radio"/>	DecisionTreeClassifier	Trained & Evaluated	Excellent	0.90436	0.87193	30 Oct 2017, 5:37 PM	...

[Close](#)
[Previous](#)
[Save](#)

6. Save & Deploy Model

- Pick which model you want to keep then click “Save.”

CustomerChurn-WML

[Overview](#)
[Evaluation](#)
[Deployments](#)
[Test](#)

Summary

Machine learning service

DSX-Lab-Machine_Learning

Runtime environment

spark-2.0

Training date

30 Oct 2017, 5:40 PM

Label column

CHURN

Latest version

15233fc3-28a0-485d-9f09-445bac5bcd11

Model builder details


[View](#)

The overview page provides useful information about the model. This includes the ability to deploy and predict with the model.

- Click on “Deployments” then “**Add Deployment**”:

CustomerChurn-WML

Overview Evaluation **Deployments** Test



+ Add Deployment

NAME	DEPLOYMENT TYPE	ACTIONS
Your model is not deployed.		

- For “**Select Deployment Type**” choose “**Online**” then give the deployment a useful name:

Deploy model

Deployment Type

Online 

Name

CustChurnRandForestModelDeployed

Close
Deploy

7. Predict with Model

- Click on “**Test**” to test the model.

The input features will be pre-populated, but you can change them to see different outcomes. Just be sure that the values you add are valid as per the data set. **See “Summary Statistics”** from the Jupyter notebook exercise:

CustomerChurn-WML

Overview Evaluation Deployments **Test**

The predictor simulates scoring your prediction data with an undeployed model.

Spark service

DSX-Spark 

Prediction input data

International

0

Local

206.08

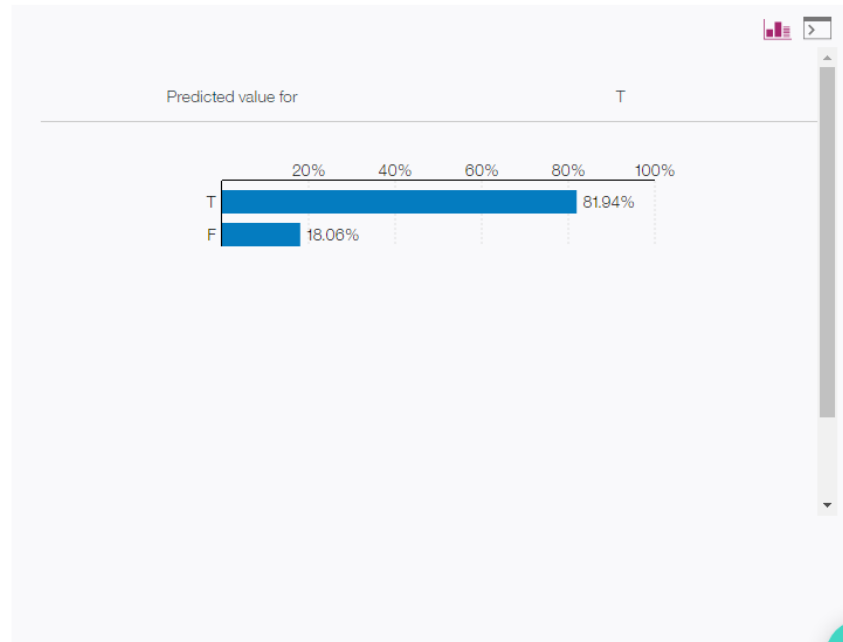
Dropped

0

Paymethod

CC

Predict



End of Lesson 4

End of Hands-on Workshop

Thank You