

IBM Cloud

Predicting Customer Churn

Watson Data Platform



Lab Guide





Notices and Disclaimers

© Copyright IBM Corporation 2018.

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
1.1	Predicting Customer Churn with Watson Data Platform Lab.docx	1/21/2018
1.2	Predicting Customer Church with Watson Data Platform Lab_V2.docx	4/2/2018

Prepared & Revised by:

Louis Frolio – louis.frolio@ibm.com

Loren Murphy – lmurphy@us.ibm.com



Table of Contents

Lab Environment Overview.....	5
Lesson 1: Watson Studio 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	23
Lesson 3: Workflow Overview	24
Lesson 3: Instructions	25
Lesson 4: Watson Machine Learning.....	36
Lesson 4: Workflow Overview	37
Lesson 4: Instructions	38



Lab Environment Overview

Software and Tools

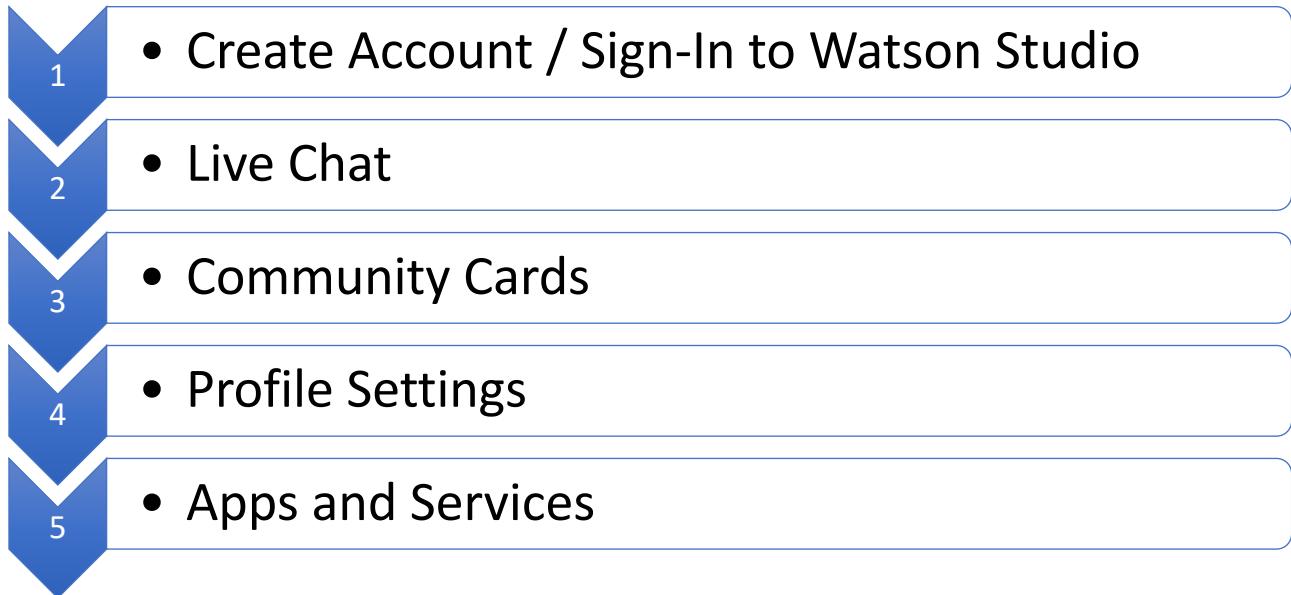
Software	Link
IBM Watson Studio	https://datascience.ibm.com/
GitHub	https://github.com/team-wolfpack



Lesson 1: Watson Studio Signup & Home Page

Purpose:	This lab introduces IBM Watson Studio, 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 Watson Studio Account• Engage Live Chat• Differentiate Four Types of Community Cards• Explore Personal Profile, Apps/Services, and Integrations

Lesson 1: Workflow Overview

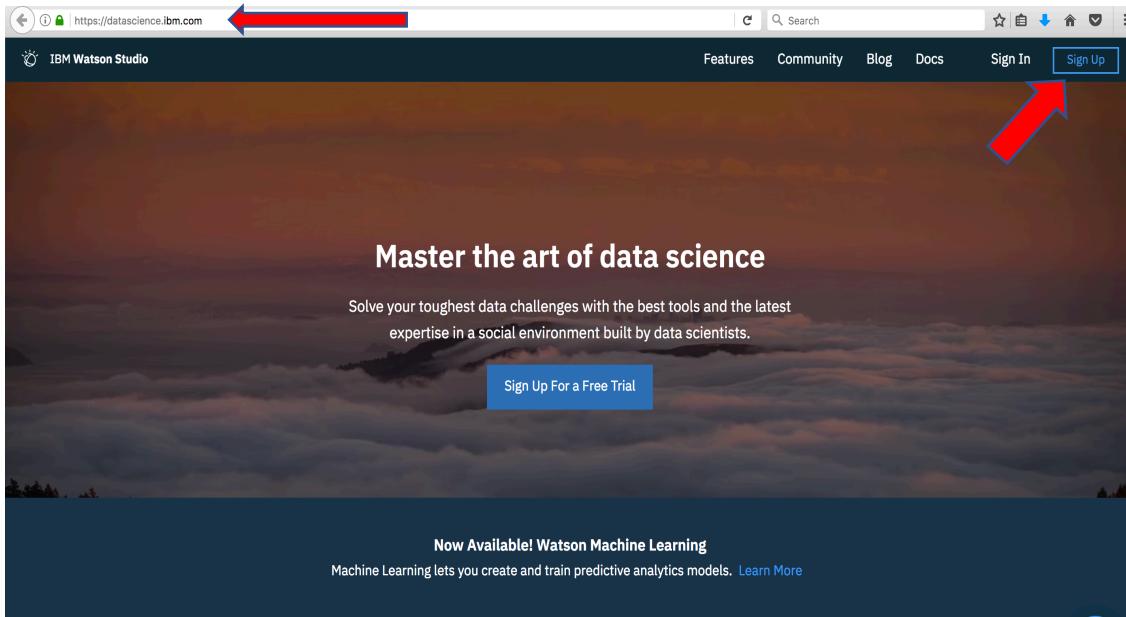


Lesson 1: Instructions

Action

1. Create Account/Sign In to Watson Studio

- 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:

IBM Watson Projects Tools Community Services US South Get started ▾

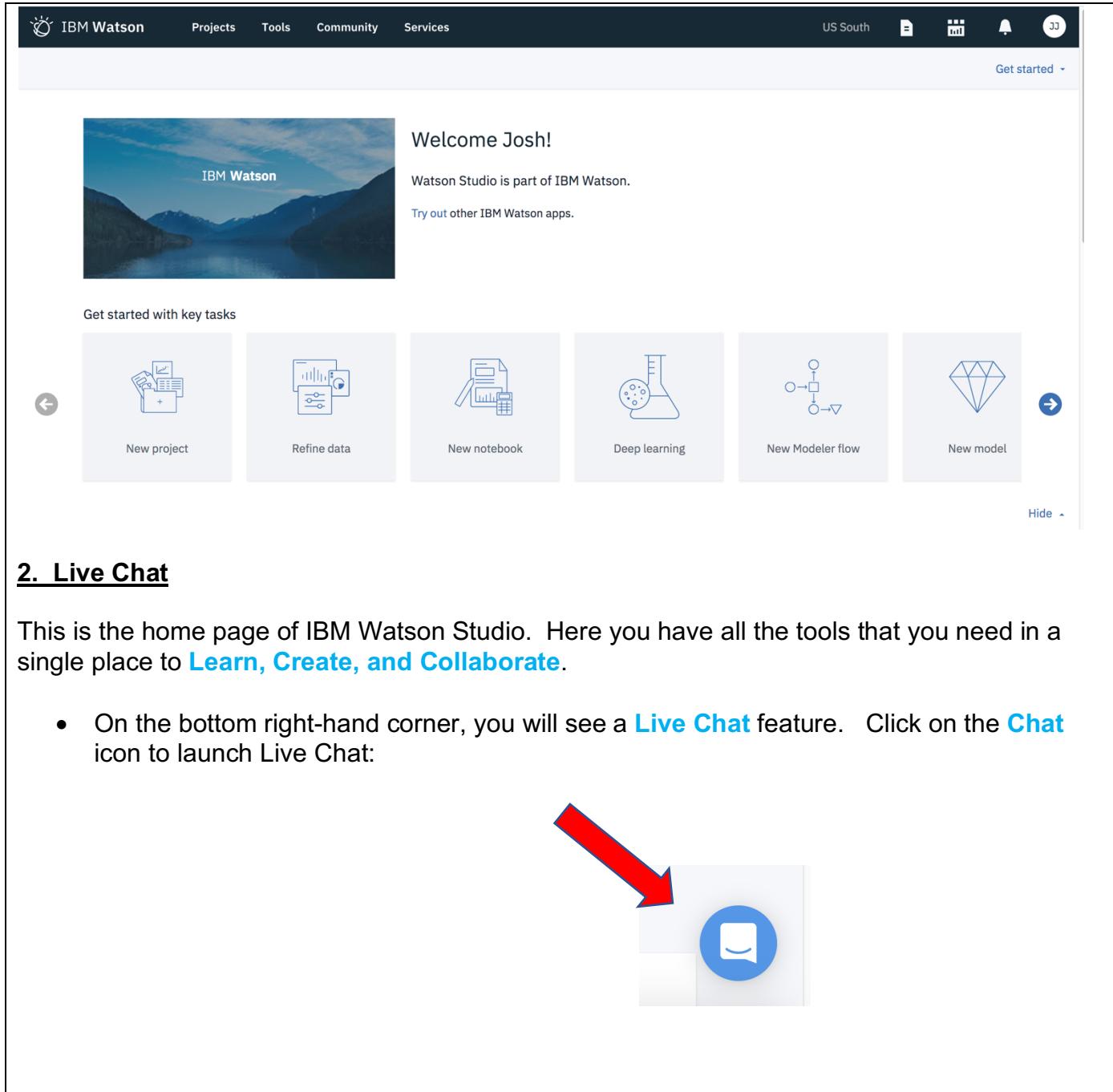
Welcome Josh!

Watson Studio is part of IBM Watson.

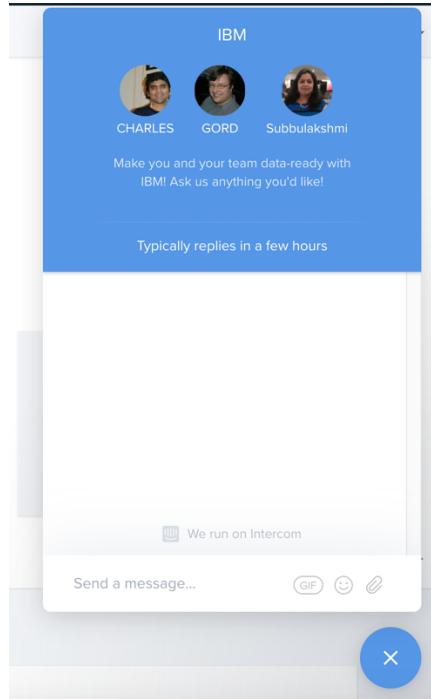
[Try out other IBM Watson apps.](#)

Get started with key tasks

New project Refine data New notebook Deep learning New Modeler flow New model Hide ▾



A screenshot of the IBM Watson Studio home page. At the top, there's a navigation bar with links for IBM Watson, Projects, Tools, Community, Services, US South, and Get started. Below the navigation is a welcome message "Welcome Josh!" with a note that Watson Studio is part of IBM Watson and a link to try other apps. A large image of a lake with mountains in the background is on the left. On the right, there's a section titled "Get started with key tasks" featuring six cards: "New project" (document icon), "Refine data" (document icon), "New notebook" (document icon), "Deep learning" (beaker icon), "New Modeler flow" (flowchart icon), and "New model" (diamond icon). A red arrow points to the "New model" card, which has a blue circular icon with a white speech bubble containing a smiley face.



If you need assistance, start typing your message in the **Send a Message** box to connect with a live person. Through this Live Chat feature, you can also continue conversations the next time you log into Watson Studio.

We use feedback captured through **Live Chat** and the offerings instrumentation to guide our decisions in designing and developing **Watson Studio**.

3. Community Cards

At the top of the Home Page click on **Community**:



Featured

Sort by: Featured ▾

ARTICLE Apple, IBM add machine learning to... AUTHOR TechCrunch DATE Mar 20, 2018 TOPIC Watson FORMAT Web page	ARTICLE Introducing IBM Watson Studio AUTHOR Armand Ruiz DATE Mar 20, 2018 TOPIC Watson FORMAT Web page	ARTICLE Webinar: April 11 - Thinking inside the box.... AUTHOR RStudio DATE Apr 02, 2018 TOPIC Data Science FORMAT Web page
--	---	---

All content

ARTICLE Webinar: April 11 - Thinking inside the box.... AUTHOR RStudio DATE Apr 02, 2018 TOPIC Data Science FORMAT Web page	NOTEBOOK Watson Assistant Workspace Analysis with... AUTHOR IBM DATE Apr 02, 2018 TOPIC Communications	TUTORIAL Build Deep Learning Architectures With... AUTHOR developerWorks TV DATE Apr 02, 2018 LEVEL Beginner TOPIC Deep Learning +2	NOTEBOOK Connect to Db2 Warehouse on Cloud and Db2... AUTHOR IBM DATE Mar 29, 2018 TOPIC Economy & Business
NOTEBOOK From scikit-learn Model to Cloud with WML... AUTHOR DATE Apr 02, 2018 TOPIC FORMAT Web page	NOTEBOOK Access MySQL with Python AUTHOR DATE Apr 02, 2018 TOPIC FORMAT Web page	ARTICLE Using shell scripts to control data flows... AUTHOR DATE Apr 02, 2018 TOPIC FORMAT Web page	ARTICLE Working with data flows using Watson Data... AUTHOR DATE Apr 02, 2018 TOPIC FORMAT Web page

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.

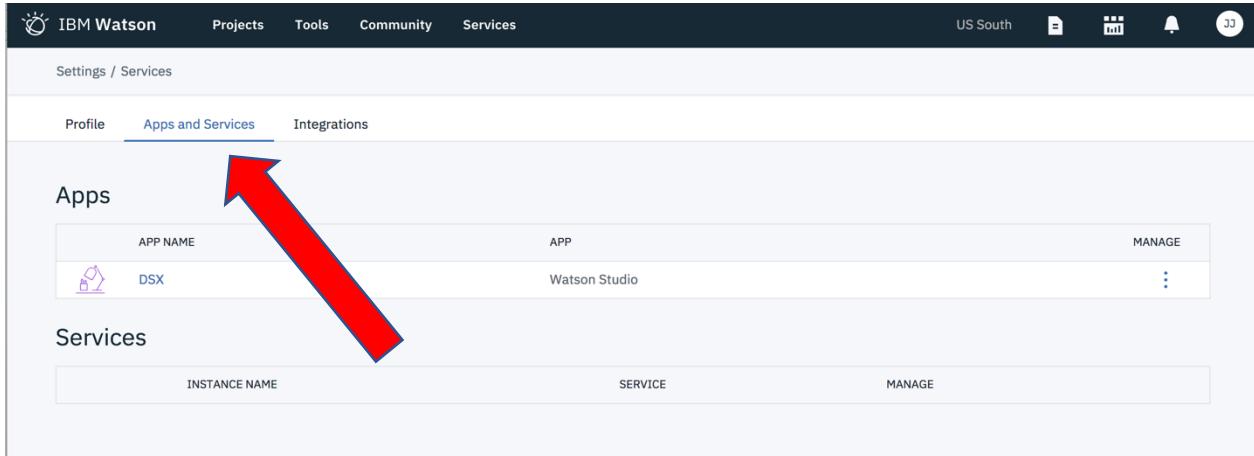
4. Profile Settings

- Click on **Settings** to look at your **Profile, Apps and Services, and Integrations**. This is where you see the details of your IBM Cloud Account:

The image shows two screenshots of the IBM Watson dashboard. The left screenshot shows the main dashboard with a navigation bar at the top and a sidebar on the left containing links like 'What's New', 'Add Other Apps', 'FAQ', 'Give Feedback', and 'Sign Out'. A red arrow points from the top right towards the 'Settings' link in the sidebar. The right screenshot shows the 'Settings / Profile' page. It has a header with tabs for 'Profile', 'Apps and Services', and 'Integrations'. Below the tabs, there's a section for 'BM ID' with a profile picture placeholder. Underneath, there's a section for 'IBM Cloud Account' with a dropdown menu showing 'Account: Josh Jones's Account'. Another red arrow points from the bottom right towards the 'Profile' tab in the header of the settings page.

5. Apps and Services

- Click on **Apps and Services** to view all your current IBM Cloud Apps and Services:



The screenshot shows the IBM Watson Cloud interface. At the top, there's a navigation bar with links for 'IBM Watson', 'Projects', 'Tools', 'Community', and 'Services'. On the right side of the bar are icons for 'US South', a file, a gear, a bell, and a user profile. Below the navigation bar, the page title is 'Settings / Services'. Underneath that, there are three tabs: 'Profile', 'Apps and Services' (which is underlined in blue), and 'Integrations'. The main content area is divided into two sections: 'Apps' and 'Services'. The 'Apps' section has a table with one row. The 'Services' section is currently empty. A large red arrow points from the left towards the 'Apps and Services' tab.

APP NAME	APP	MANAGE
DSX	Watson Studio	⋮

INSTANCE NAME	SERVICE	MANAGE

Above is the default for the brand-new account, there are no services or apps deployed other than Watson Studio.

Integrations is where you configure Watson Studio for GitHub integration.

End of Lesson 1



Lesson 2: Jupyter Notebook

Purpose:	This lesson introduces projects within Watson Studio, 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 Spark.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Create and Configure Watson Studio Project• Add Notebook Asset• Retrieve Data from External Repository• Predict Customer Churn using Machine Learning Techniques• Evaluate Model Performance

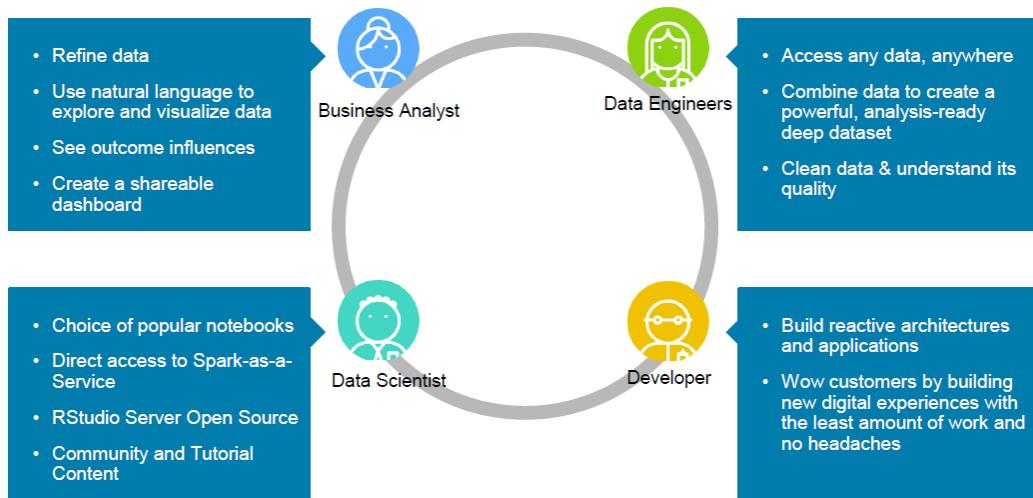
Lesson 2: Workflow Overview

- 1 • Project Overview
- 2 • Create New Project
- 3 • Create Notebook
- 4 • Load Data from Github Repo
- 5 • Create Spark DataFrames
- 6 • Rename Columns
- 7 • Explore Data
- 8 • Create Spark ML pipeline
- 9 • Create Random Forests & Decision Tree Models
- 10 • Evaluate & Invoke Models

Lesson 2: Instructions

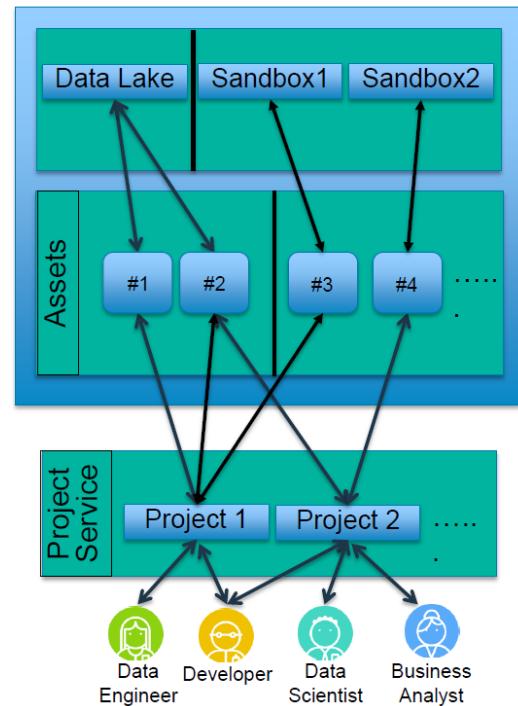
1. Project Overview

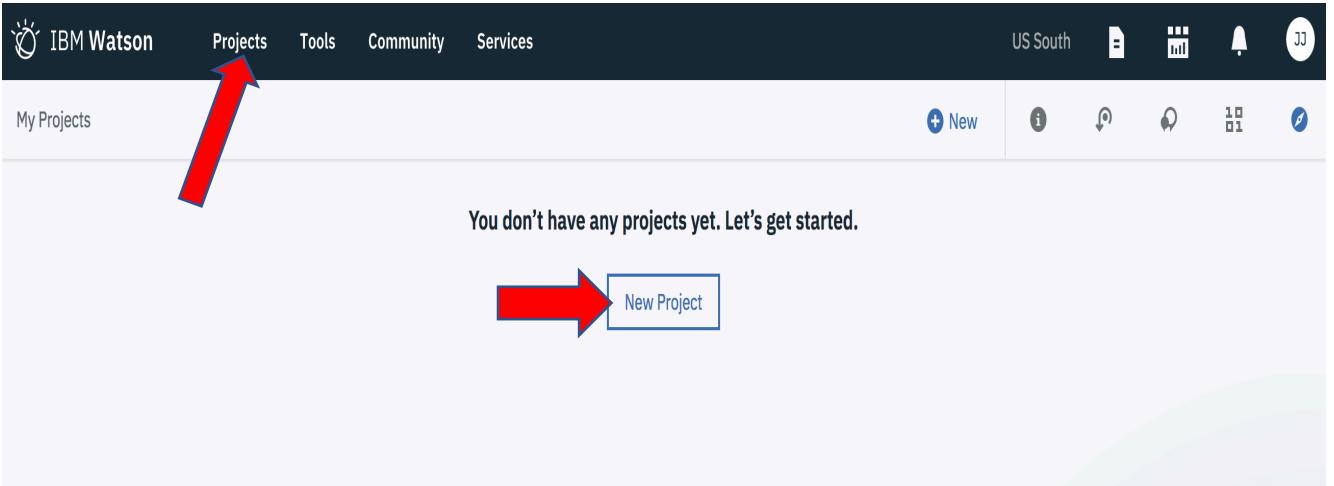
Data professionals need purpose-built, self-service communities that enable them to seamlessly collaborate across personas.



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
<ul style="list-style-type: none">• Navigate to https://datascience.ibm.com• Login to Watson Studio• On the top right side, click Projects and select New Project  <ul style="list-style-type: none">• Type the Project Name Customer Churn, add a meaningful description:

IBM Watson Projects Tools Community Services

New project

Define project details

Name 

Description 

86 2977

Choose project options

Restrict who can be a collaborator 

Project will include integration with Object Storage for storing project assets.

Define Storage:

- Click **Add**
- Choose **Lite** plan then **Create**
- Verify your options then **Confirm**
- Refresh**

Storage

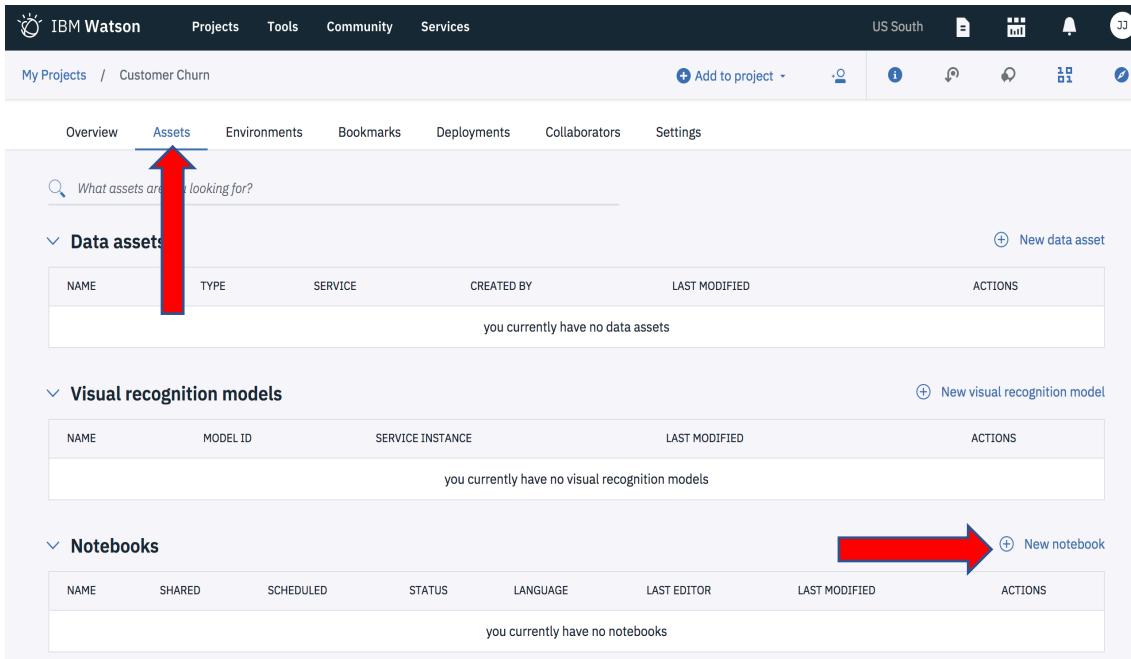
cloud-object-storage-pd

- Click **Create**

The screenshot shows the 'Customer Churn' project page in the IBM Watson interface. At the top, there are tabs for Overview, Assets, Environments, Bookmarks, Deployments, Collaborators, and Settings. The 'Assets' tab is highlighted with a red arrow. Below the tabs, the project name 'Customer Churn' is displayed, along with the last update date ('Apr 02 2018'). To the right, there are three summary metrics: 'Assets' (0), 'Bookmarks' (0), and 'Collaborators' (1). The 'Recent activity' section is empty, showing a placeholder icon and text: 'Alerts related to this project will show here when the project is active.' On the left, there are sections for 'Date created' ('Apr 02 2018'), 'Description' ('Customer churn analysis'), 'Storage' ('0% of 5 GB used'), and 'Collaborators' ('View all (1)'). A single collaborator, 'Josh Jones (Admin)', is listed.

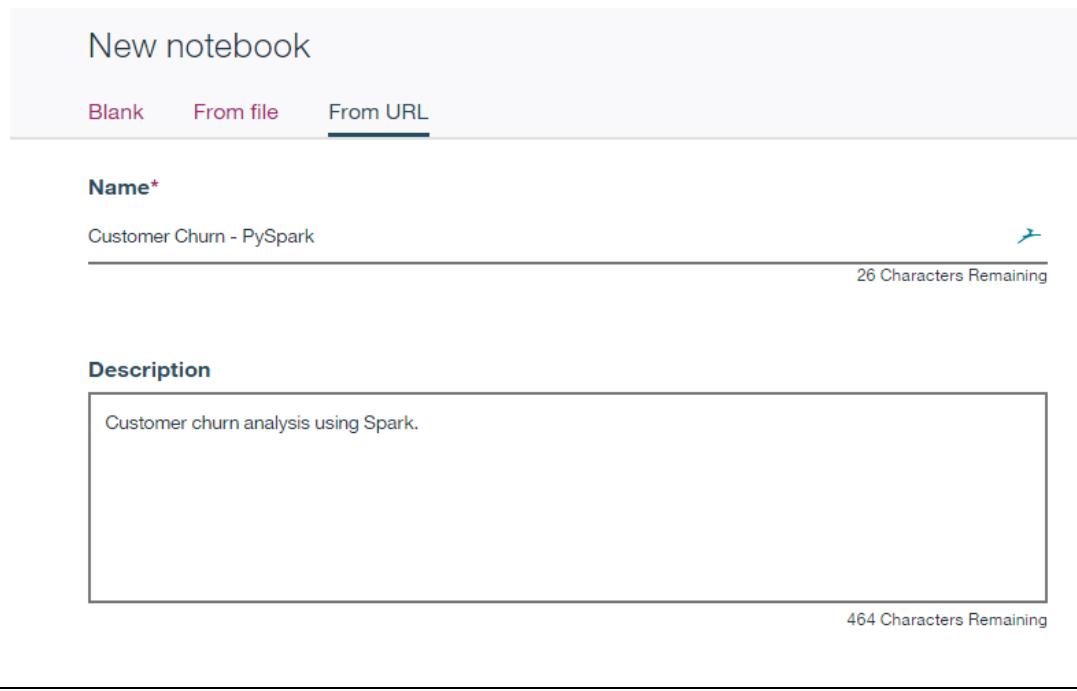
3. Create Notebook

- Click **Assets**, then **New Notebook**



The screenshot shows the IBM Watson interface with the 'Assets' tab selected. The main content area displays sections for 'Data assets', 'Visual recognition models', and 'Notebooks'. Under 'Notebooks', there is a '+ New notebook' button highlighted with a red arrow.

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



New notebook

Blank From file **From URL**

Name*

Customer Churn - PySpark

26 Characters Remaining

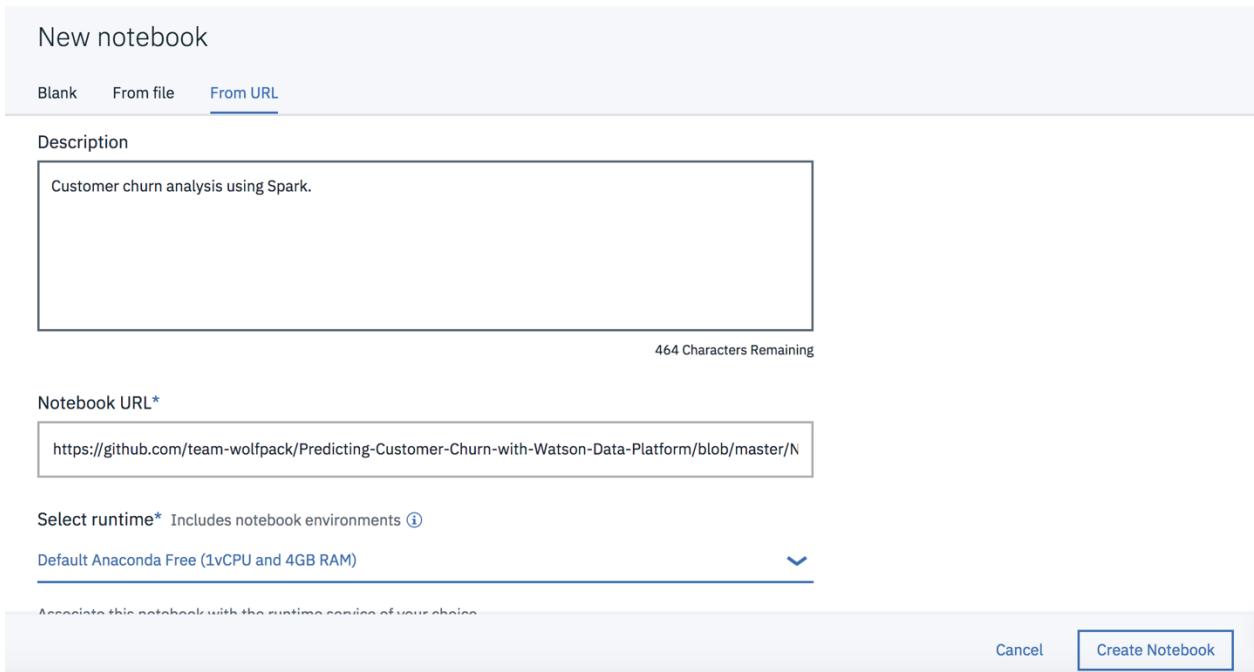
Description

Customer churn analysis using Spark.

464 Characters Remaining

- In a separate browser window navigate to:
[Predicting Customer Churn with Watson Data Platform](https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform)
 (https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform)
- Click on Notebooks, right click on **CustomerChurn-PySpark.ipynb** then choose **Copy link address**. Go back to the **Watson Studio New Notebook** page.

Paste URL into **Notebook URL** text box. Select **Default Anaconda Free (1vCPU and 4GB RAM)** as the runtime. Then click **Create Notebook**:



New notebook

Blank From file From URL

Description

Customer churn analysis using Spark.

464 Characters Remaining

Notebook URL*

https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform/blob/master/N

Select runtime* Includes notebook environments ⓘ

Default Anaconda Free (1vCPU and 4GB RAM)

Annotate this notebook with the runtime environment of your choice

Cancel Create Notebook

You should now see:

My Projects / Customer Churn / Customer Churn - PySpark

File Edit View Insert Cell Kernel Help

Not Trusted | Python 3.5

IBM WolfPack



Predicting Customer Churn with Watson Data Platform

Adapted from "Predict Customer Churn Use Case Implementation" by Sydney Phoon | <https://github.com/SidneyPhoon/IntroToNotebooksLab>

Table of contents

1. Step 1: Download the customer churn data



Lesson 2 Continued in [Customer Churn – PySpark] Notebook

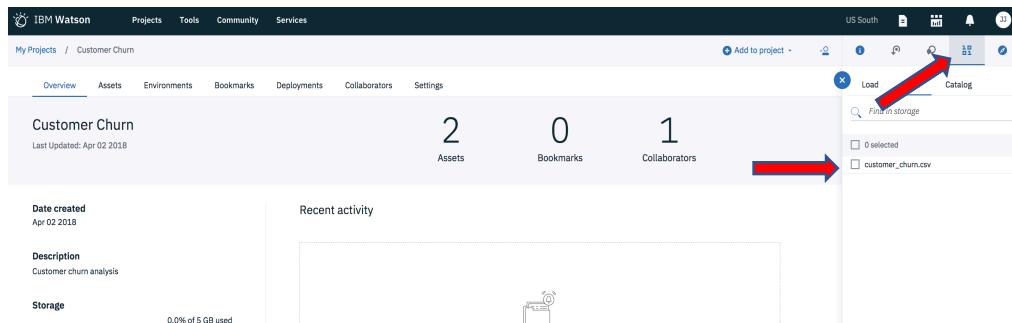
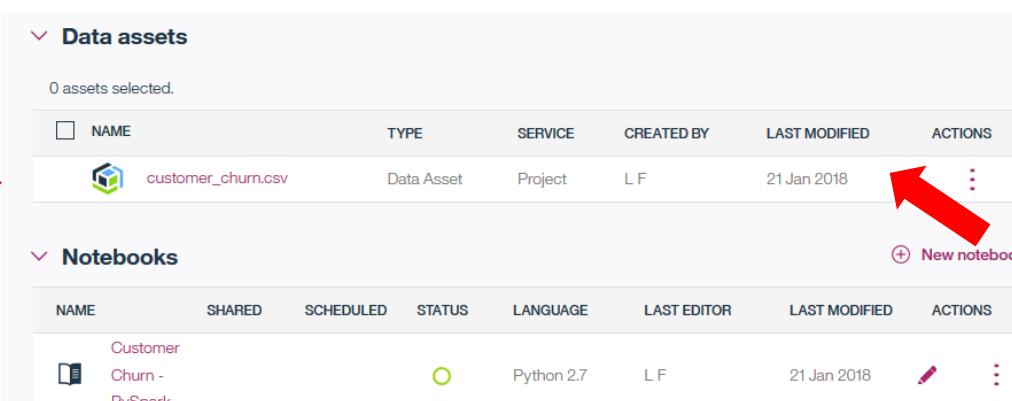
Lesson 3: Machine Learning Flows

Purpose:	This lesson introduces Machine Learning Flows in Watson Studio. 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 to Project
 - 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

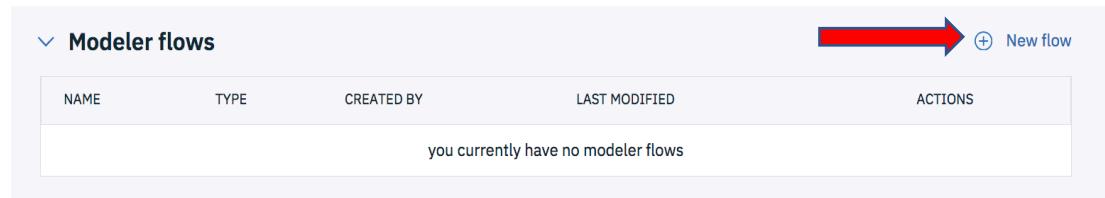
Lesson 3: Instructions

Action
<p>1. Load Data from Local File</p> <ul style="list-style-type: none"> In a separate browser navigate to: Customer Churn Data: https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform/tree/master/Data Download customer_churn_data.zip file, unzip and place customer_churn.csv in a folder on your computer. Go back to the Customer Churn project and then click on the Data icon at the top right of the screen: <p>A new panel will be presented with Files highlighted. Click on Load, navigate to the customer_churn.csv file and select it. You should now see that the file has been imported into the project under the Files tab</p>  <p>Navigate back to Assets and see the new “Data Asset”:</p> 

Action

2. Create Machine Learning Flow

- Navigate to Customer Churn project page
- Click on “**New flow**”

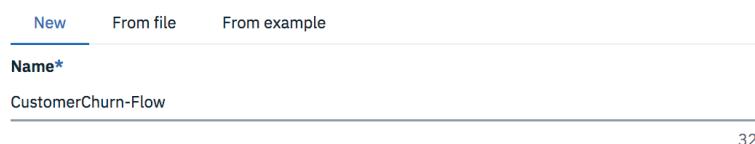


The screenshot shows a list titled "Modeler flows". At the top right, there is a blue button labeled "New flow" with a plus sign icon. A thick red arrow points to this button. Below the button, the list table has columns: NAME, TYPE, CREATED BY, LAST MODIFIED, and ACTIONS. A message at the bottom of the list says "you currently have no modeler flows".

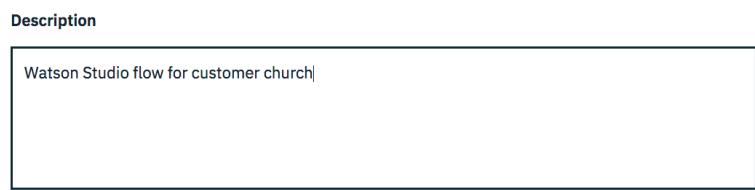
- Choose “**New**” on the top menu. Give the flow a meaningful name and description. For “Flow Type” choose “**Modeler Flow**”. For “Runtime” choose “**IBM SPSS Modeler**”:



The screenshot shows a form for creating a new flow. The title "Modeler" is displayed above the form. The "Name*" field is filled with "CustomerChurn-Flow". Below the name field, there is a "Description" section containing the text "Watson Studio flow for customer church".



The screenshot shows the "New" tab selected in the top navigation bar. The "Name*" field is filled with "CustomerChurn-Flow".



The screenshot shows the "Description" field containing the text "Watson Studio flow for customer church".

462

Select flow type

Modeler Flow Neural Network Modeler BETA

Runtime

IBM SPSS Modeler Scala Spark 2.0 BETA

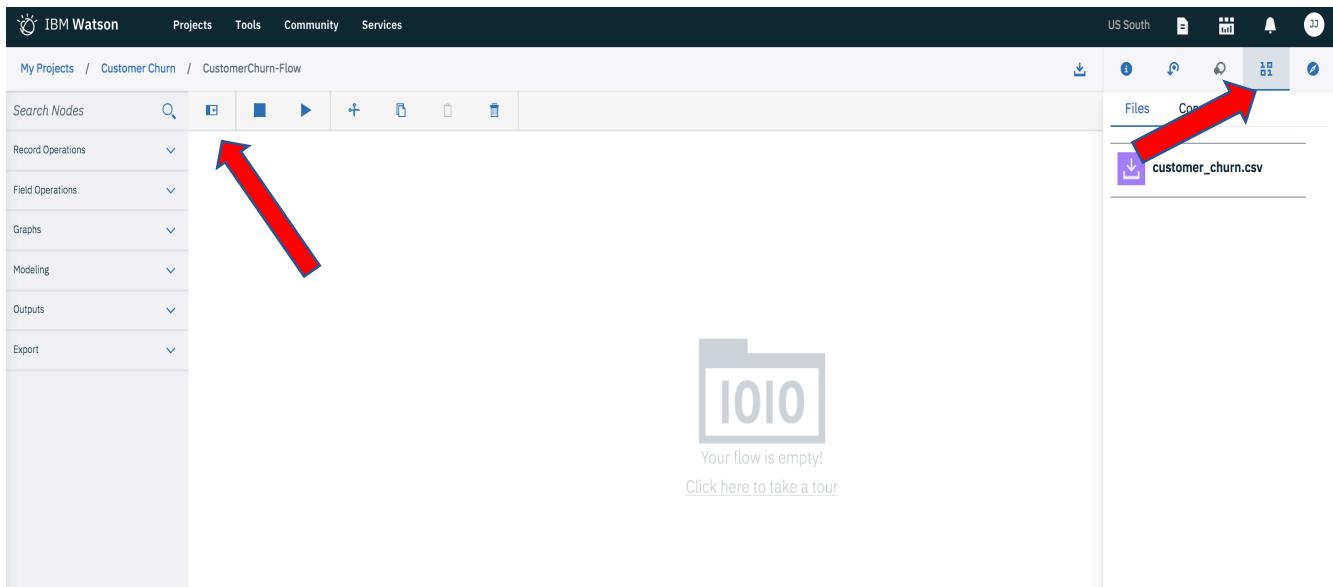
- Click on “**Create**”

Action

3. 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 Watson Studio 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:



- Double click on “Type”, click on “**Configure Types**” then “**Add Columns**”

Action

- Add all the columns except for “ID”.

Select Fields for Type

Field name	Data type
ID	integer
CHURN	string
Gender	string
Status	string
Children	unknown
Est Income	unknown
Car Owner	string
Age	unknown
LongDistance	unknown
International	unknown
Local	unknown
Dropped	unknown
Paymethod	string
LocalBiltype	string
LongDistanceBiltype	string

Cancel OK

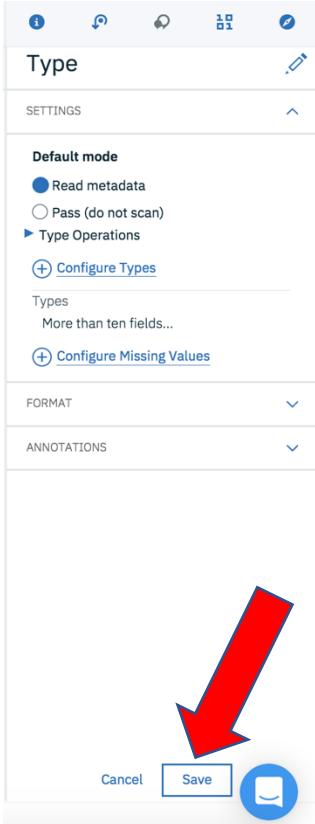
- Click **OK**
- For the “**CHURN**” column, change its Role to that of “**Target**.” Leave the default for the remaining columns:

Configure Types

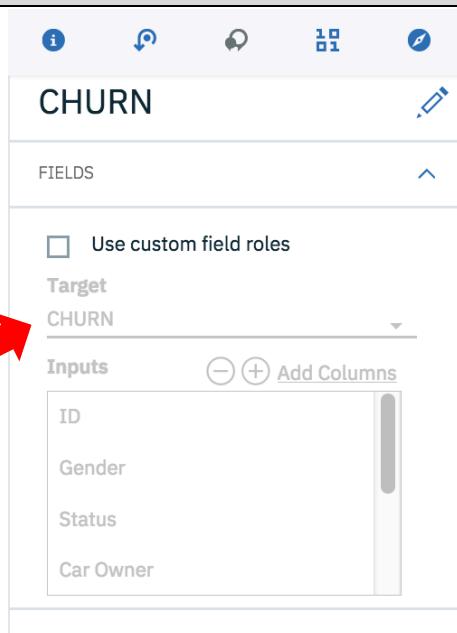
Read Values

Add Columns

Field	Measure	Role	Value mode	Values	Check
CHURN	Default	Target	Read	None	...
Gender	Default	Input	Read	None	...
Status	Default	Input	Read	None	...
Children	Default	Input	Read	None	...

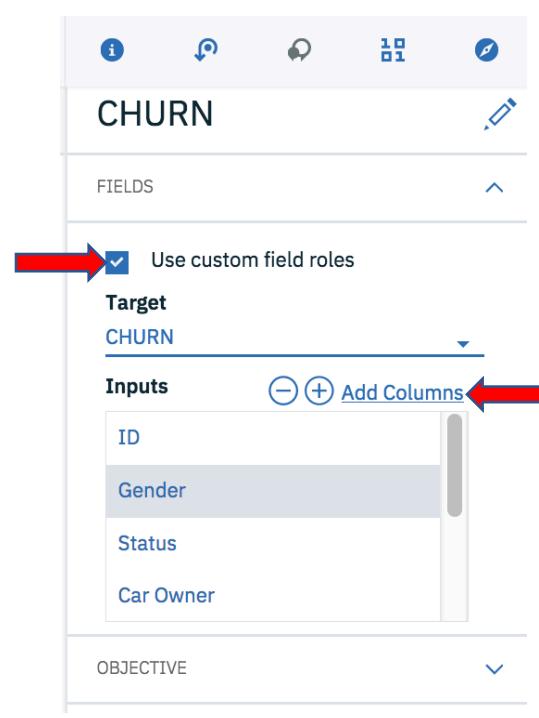
Action
<ul style="list-style-type: none"> • Click “OK”. • Click “Save” to exit 
4. Add & Configure Model Object
<ul style="list-style-type: none"> • From the palette, expand the “Modeling” branch then drag “C&R Tree” onto the workspace to the right of “Type.” • Connect the two then double click on “C&R Tree” to edit its properties. • The C&R Tree object should now say “CHURN”. Double click on this object. • Click on “FIELDS”, Target should be set to “CHURN”

Action

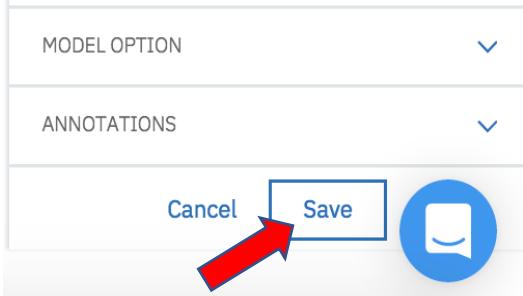
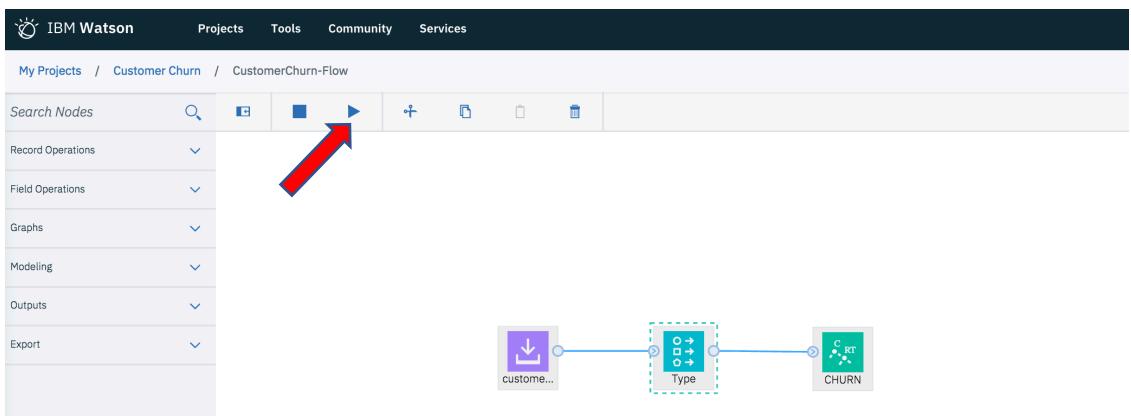


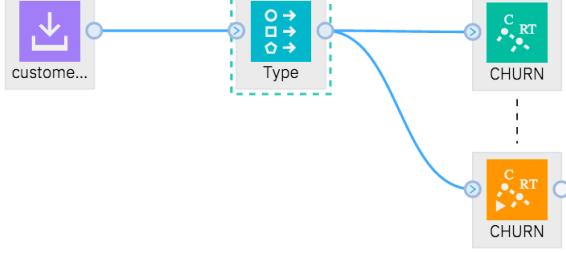
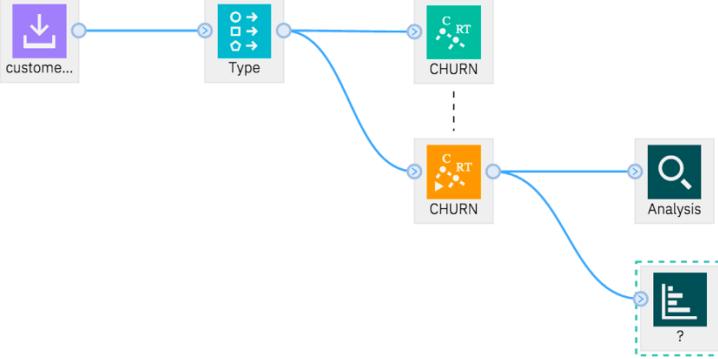
The screenshot shows the 'CHURN' dataset configuration in the Watson Studio workspace. The 'Target' field is set to 'CHURN'. The 'Inputs' section lists 'ID', 'Gender', 'Status', and 'Car Owner'. A red arrow points to the 'Use custom field roles' checkbox, which is currently unchecked.

- Check the “**Use custom field roles**” box. Click “**Add Columns.**” Recall from the notebook exercise you were asked to jot down the top 5 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:



The screenshot shows the 'CHURN' dataset configuration in the Watson Studio workspace. The 'Target' field is set to 'CHURN'. The 'Inputs' section lists 'ID', 'Gender', 'Status', and 'Car Owner'. A red arrow points to the 'Use custom field roles' checkbox, which is now checked. Another red arrow points to the '+ Add Columns' button in the 'Inputs' section.

Action
<ul style="list-style-type: none"> Click on “Save.” 
<ul style="list-style-type: none"> Your palette should resemble this: 
<h2><u>5. Run Flow to Create Nugget</u></h2> <ul style="list-style-type: none"> Run the flow by clicking on the “Run” icon at the top of the workspace. 
<p>You should see a new forth object on the workspace, this is called a nugget.</p>

Action

<ul style="list-style-type: none"> • 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:

<p><u>6. Add & Configure Analysis Object – Measure Model Performance</u></p> <ul style="list-style-type: none"> • Double click on “Analysis” and check off the four checkboxes, leave the rest as default:

Action

Analysis

[Edit](#)

SETTINGS

- Coincidence matrices (for symbolic targets)
- Performance evaluation
- Evaluation metric (AUC & Gini, binary classifiers only)
- Confidence figures (if available)

Threshold for pct. correct
90

Improve accuracy multiplier
2

Find predicted/predictor fields using

- Model output field metadata
- Field name format (for example, '\$<x>-<target field>')
- Separate by partition
- User defined analysis

[Configure Analysis](#)

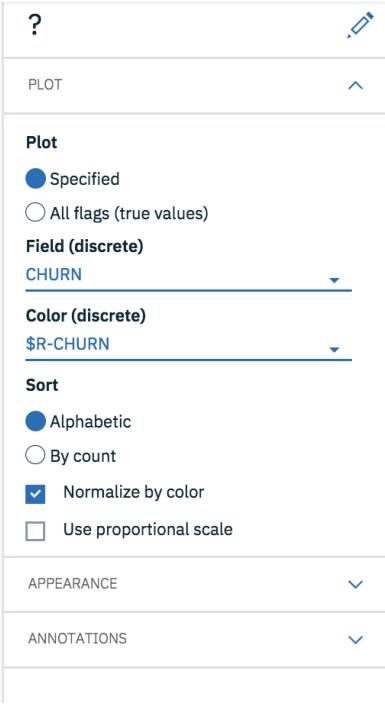
Break down analysis by fields

[-](#) [+](#) [Add Columns](#)

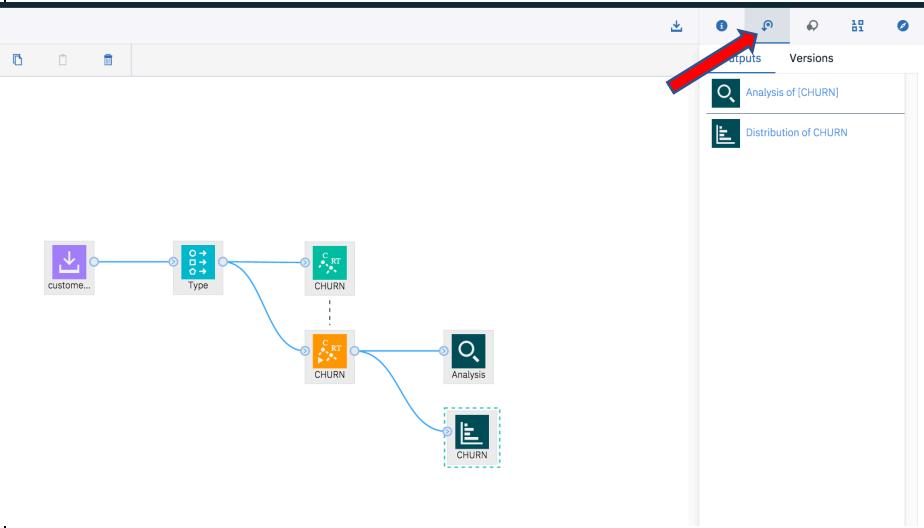
Cancel [Save](#) 

- Click “**Save**” to return to the workspace.
- Double click on the “**Distribution**” object and configure it as depicted below:

Action



- Click on “Save” 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:

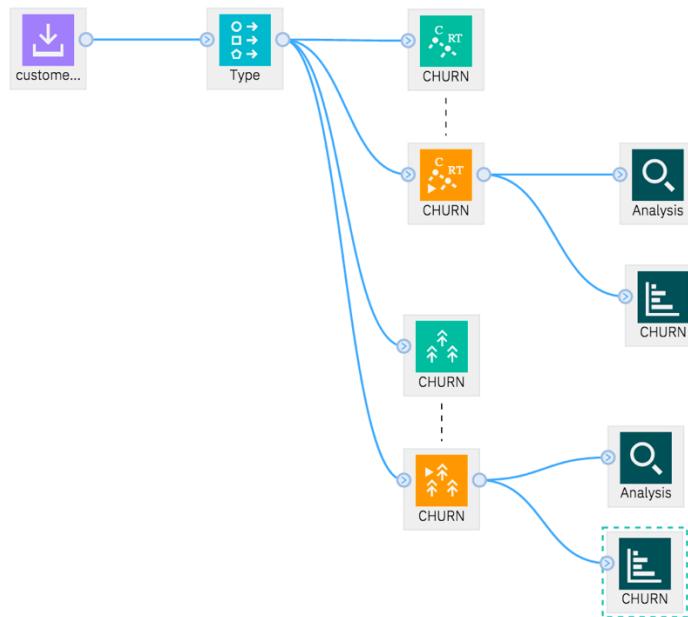


- Explore the results

Action

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

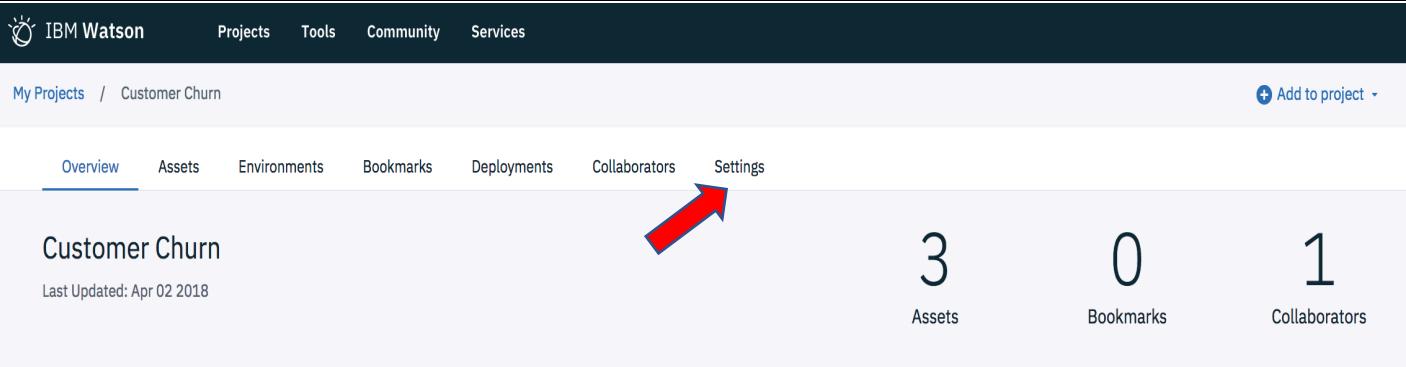
- 1 • Create Machine Learning Service
- 2 • Create Machine Learning Model
- 3 • Choose Modeling Technique
- 4 • Add Estimators
- 5 • Evaluate Models
- 6 • Save & Deploy Model
- 7 • Predict with Model

Lesson 4: Instructions

Action

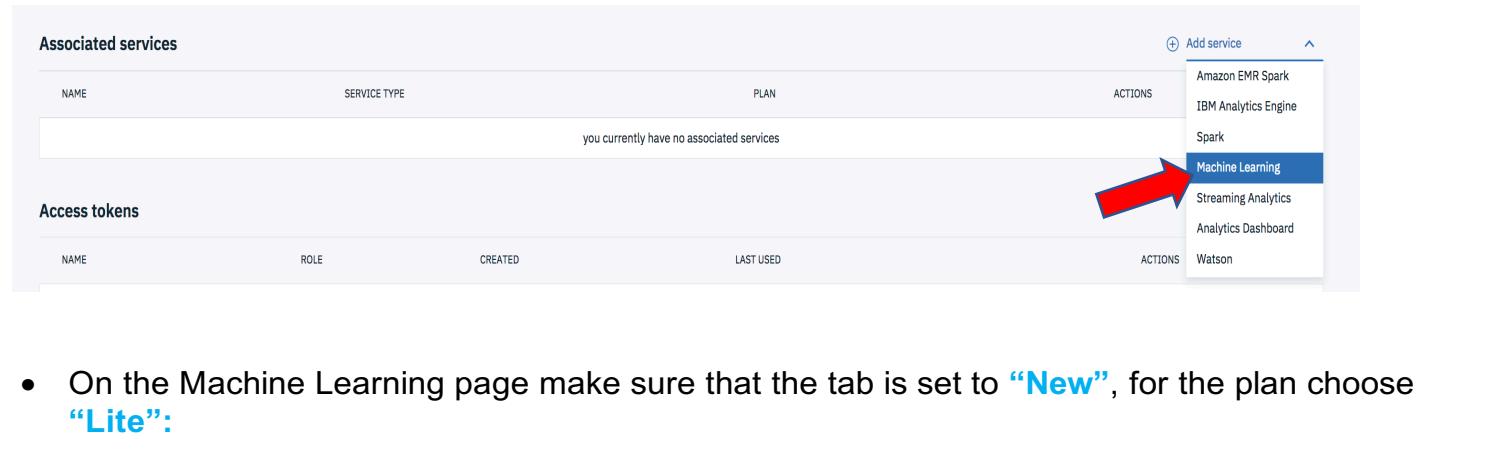
1. Create Machine Learning Service

- Navigate to Customer Churn project page
- At the top click on the “**Settings**” icon:



The screenshot shows the IBM Watson Project Overview page. At the top, there's a navigation bar with links for IBM Watson, Projects, Tools, Community, and Services. Below that, a breadcrumb navigation shows 'My Projects / Customer Churn'. On the right, there's a 'Add to project' button. The main area has tabs for Overview, Assets, Environments, Bookmarks, Deployments, Collaborators, and Settings. The 'Settings' tab is underlined and has a red arrow pointing to it. Below the tabs, the project name 'Customer Churn' is displayed along with its creation date ('Last Updated: Apr 02 2018'). To the right, there are summary numbers: 3 Assets, 0 Bookmarks, and 1 Collaborator. Further down, there are sections for 'Date created' (Apr 02 2018) and 'Recent activity'.

Scroll to the middle of the page and click on “**Add service**” then choose “**Machine Learning**”:



The screenshot shows the 'Associated services' section. It includes tables for 'Associated services' and 'Access tokens'. To the right, there's a 'Add service' dropdown menu with a list of options: Amazon EMR Spark, IBM Analytics Engine, Spark, Machine Learning (which is highlighted with a red arrow), Streaming Analytics, Analytics Dashboard, and Watson.

- On the Machine Learning page make sure that the tab is set to “**New**”, for the plan choose “**Lite**”:

Action								
<h2>Machine Learning</h2> <p>Existing New</p>  <h3>Machine Learning</h3> <p>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.</p>	<h3>Features</h3> <p>SPSS analytics platform features SPSS streams management and deployment with realtime scoring and batch processing options.</p> <p>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.</p>	<p>Spark and Python Machine Learning features Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch and streaming.</p>						
<p>Pricing Plan: Monthly Process shown above reflect the: United States</p> <table border="1"> <thead> <tr> <th>Plan</th> <th>Features</th> <th>Pricing</th> </tr> </thead> <tbody> <tr> <td> Lite</td> <td>Service instance (5 models per instance) 5,000 predictions 5 compute hours</td> <td>Free</td> </tr> </tbody> </table> 			Plan	Features	Pricing	 Lite	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free
Plan	Features	Pricing						
 Lite	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free						

- Click on “**Create**”
- At the confirmation page you can give your service a meaningful name:

Action
<h1>Confirm Creation</h1> <p>Organization: louisfrolio@gmail.com</p> <p>Plan Lite</p> <p>Space dev</p> <p>Service name dsx-wml-lab</p> <p style="text-align: right;">Cancel Confirm</p>

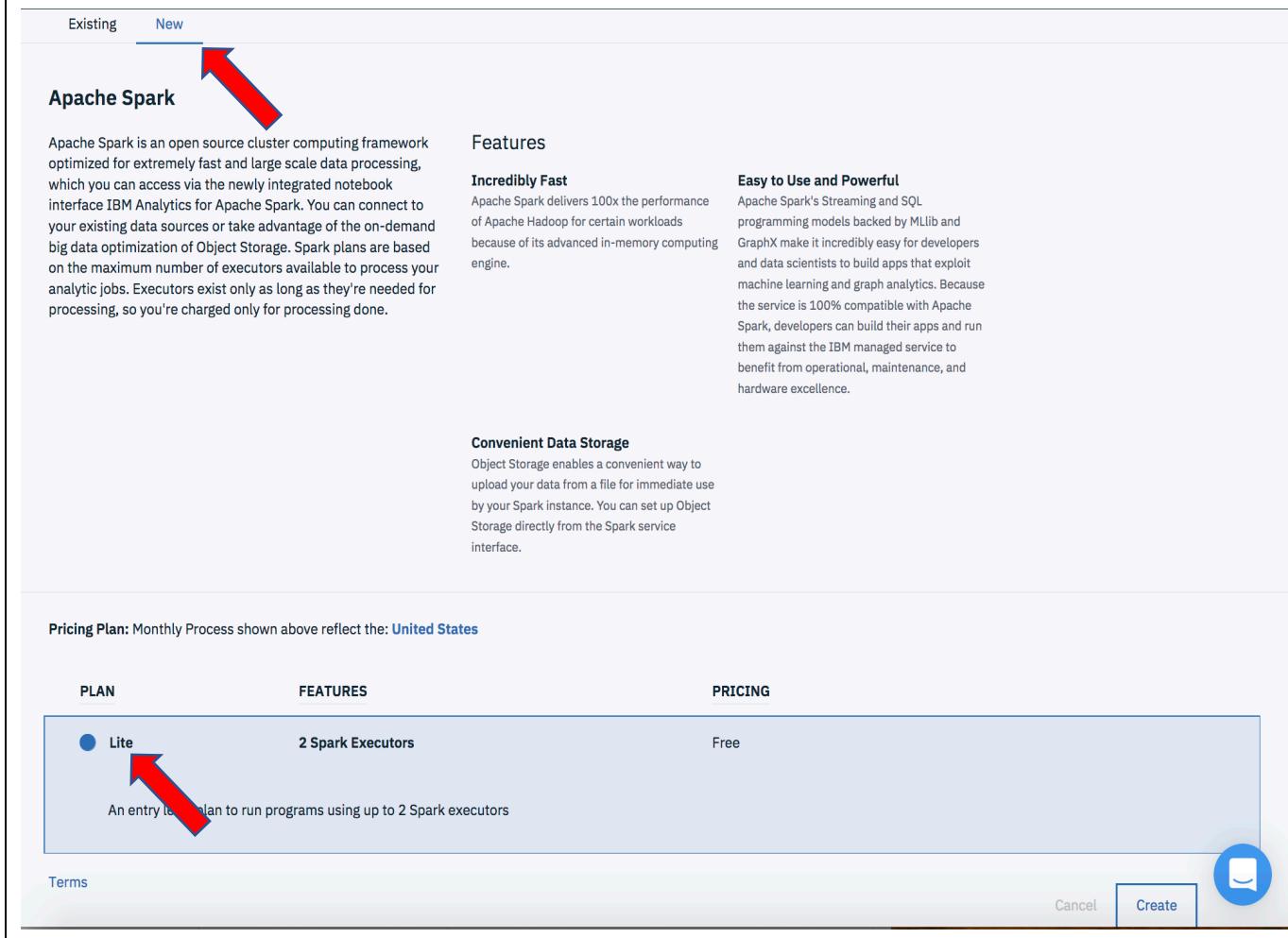
- Click “**Confirm**” to create Watson Machine Learning Service.

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**”:

Action

• In the “New model” window, associate an **IBM Analytics for Apache Spark instance**. On the Machine Learning page make sure that the tab is set to “**New**”, for the plan choose “**Lite**”. Select **Create**.

Pricing Plan: Monthly Process shown above reflect the: [United States](#)

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	2 Spark Executors An entry level plan to run programs using up to 2 Spark executors	Free

Cancel Create 

• At the confirmation page you can give your service a meaningful name. Click **Confirm**

Action

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

New model BETA

Define model details

Name

CustomerChurn-WML

83

Description

Customer churn using Watson Machine Learning

256

Machine Learning Service

[predictive-modeling-ts](#)

Select model type

Model builder From file From sample

Spark Service

spark-bp

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](#) or design an [Modeler flow](#).

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

Action								
<p>Select data asset</p> <p>The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.</p> <p> <input type="text" value="What asset are you looking for?"/></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 5px;">NAME</th> <th style="text-align: left; padding: 5px;">TYPE</th> <th style="text-align: left; padding: 5px;">SERVICE</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;"> customer_churn.csv</td> <td style="padding: 5px;">Data Asset</td> <td style="padding: 5px;">Project</td> </tr> </tbody> </table> <p style="text-align: right; margin-top: -10px;"> <input style="border: 1px solid #ccc; padding: 2px 5px; margin-right: 5px;" type="button" value="Close"/> <input style="border: 1px solid #0070C0; background-color: #e6f2ff; color: #0070C0; padding: 2px 5px;" type="button" value="Next"/> </p>			NAME	TYPE	SERVICE	 customer_churn.csv	Data Asset	Project
NAME	TYPE	SERVICE						
 customer_churn.csv	Data Asset	Project						

3. Choose Modeling Technique

- At the “**Select a Technique**” screen select “**CHURN**” as the “**Column value to predict**”, and select the following feature columns: Gender, Status, Children, Est Income, Car Owner, Paymethod, LongDistanceBilltype, Usage, RatePlan
- Make sure “**Binary Classification**” is highlighted.

Action
<p>Select a technique</p> <p>Column value to predict (Label Col) CHURN (String)</p> <p>Feature columns Gender (String), Status (String), Children (Decimal), Est Income (Decimal), Car Owner (String), Paymethod (String), LongDistanceBilbype (String), Usage (Decimal), RatePlan (Decimal) </p> <p> Suggested technique.</p> <div style="display: flex; justify-content: space-around; margin-top: 10px;"> <div style="border: 1px solid #ccc; padding: 5px; width: 25%;">  Binary Classification Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories. </div> <div style="border: 1px solid #ccc; padding: 5px; width: 25%;">  Multiclass Classification Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories. </div> <div style="border: 1px solid #ccc; padding: 5px; width: 25%;">  Regression Predict values from a continuous set of values. Choose if your label column contains a large number of values. </div> </div> <p>Configured estimators </p> <p>Add Estimators </p> <p>Validation Split</p> <div style="display: flex; align-items: center; margin-top: 10px;"> Train: 60 Test: 20 Holdout: 20  </div>

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**:

Action

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...

[Cancel](#)

[Add](#)

- Click “**Add**”

Action

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

Gender (String), Status (String), Children (Decimal), Est Income (Decimal), C

 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.

 Add Estimators

Configured estimators

	Decision Tree Classifier	
	Random Forest Classifier	

Validation Split



[Close](#) [Previous](#) [Next](#)

- 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 type="radio"/>	RandomForestClassifier	Trained & Evaluated	Excellent	0.94129	0.91694	3 Apr 2018, 10:28 AM
<input checked="" type="radio"/>	DecisionTreeClassifier	Trained & Evaluated	Excellent	0.90718	0.86968	3 Apr 2018, 10:28 AM

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

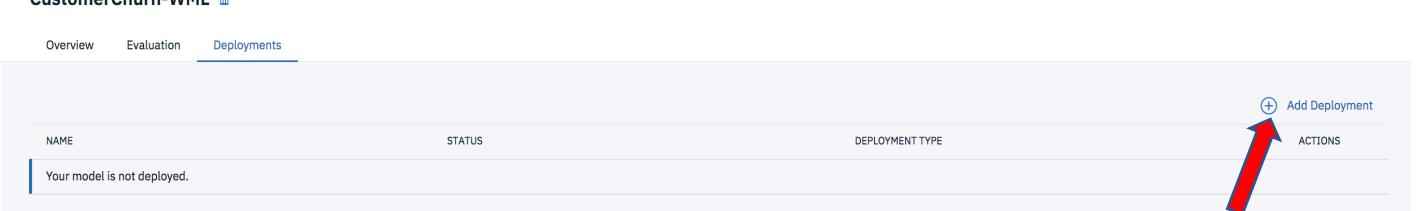
6. Save & Deploy Model

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

Action															
	<p>CustomerChurn-WML </p> <p>Overview Evaluation Deployments</p> <p>Summary</p> <table border="1"> <tbody> <tr> <td>Machine learning service</td> <td>predictive-modeling-ts</td> </tr> <tr> <td>Model Type</td> <td>wml-1.1</td> </tr> <tr> <td>Runtime environment</td> <td>spark-2.0</td> </tr> <tr> <td>Training date</td> <td>3 Apr 2018, 10:30 AM</td> </tr> <tr> <td>Label column</td> <td>CHURN</td> </tr> <tr> <td>Latest version</td> <td>3bc1d323-ab9a-40d4-8d72-0728d7e3ea0d</td> </tr> <tr> <td>Model builder details</td> <td>View</td> </tr> </tbody> </table> <p>Input Schema</p>	Machine learning service	predictive-modeling-ts	Model Type	wml-1.1	Runtime environment	spark-2.0	Training date	3 Apr 2018, 10:30 AM	Label column	CHURN	Latest version	3bc1d323-ab9a-40d4-8d72-0728d7e3ea0d	Model builder details	View
Machine learning service	predictive-modeling-ts														
Model Type	wml-1.1														
Runtime environment	spark-2.0														
Training date	3 Apr 2018, 10:30 AM														
Label column	CHURN														
Latest version	3bc1d323-ab9a-40d4-8d72-0728d7e3ea0d														
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**”:



- For deployment type choose “**Web Service**” then give the deployment a useful name:

Action

Create Deployment

[Web Service](#)

[Batch Prediction](#)

[Real-time Streaming Predictions](#)

Name

CustChurnRandForestDeployed

Description

Deployed Random Forests model to predict customer churn|

244

- Click “[Save](#)”

7. Predict with Model

- Choose newly created deployed model:

CustomerChurn-WML

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

[+ Add Deployment](#)

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
CustChurnRandForestDeployed	DEPLOY_SUCCESS	Web Service	⋮

- 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:](#)

Action

CustChurnRandForestDeployed

Overview Implementation Test

Enter input data

ID
1

Gender
F

Status
S

Children
1

Est Income

Predict

Predicted value for CHURN

T 81.59%
F 18.41%



End of Lesson 4

End of Hands-on Workshop

Thank You