

IBM Watson

Predicting Customer Churn

Watson Studio



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
1.3	Predicting Customer Churn with Watson Studio.docx	4/23/2018

Prepared & Revised by:

Louis Frolio – louis.frolio@ibm.com

Loren Murphy – lrmurphy@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	22
Lesson 3: Workflow Overview	22
Lesson 3: Instructions	24
Lesson 4: Watson Machine Learning	34
Lesson 4: Workflow Overview	35
Lesson 4: Instructions	36

Lab Environment Overview

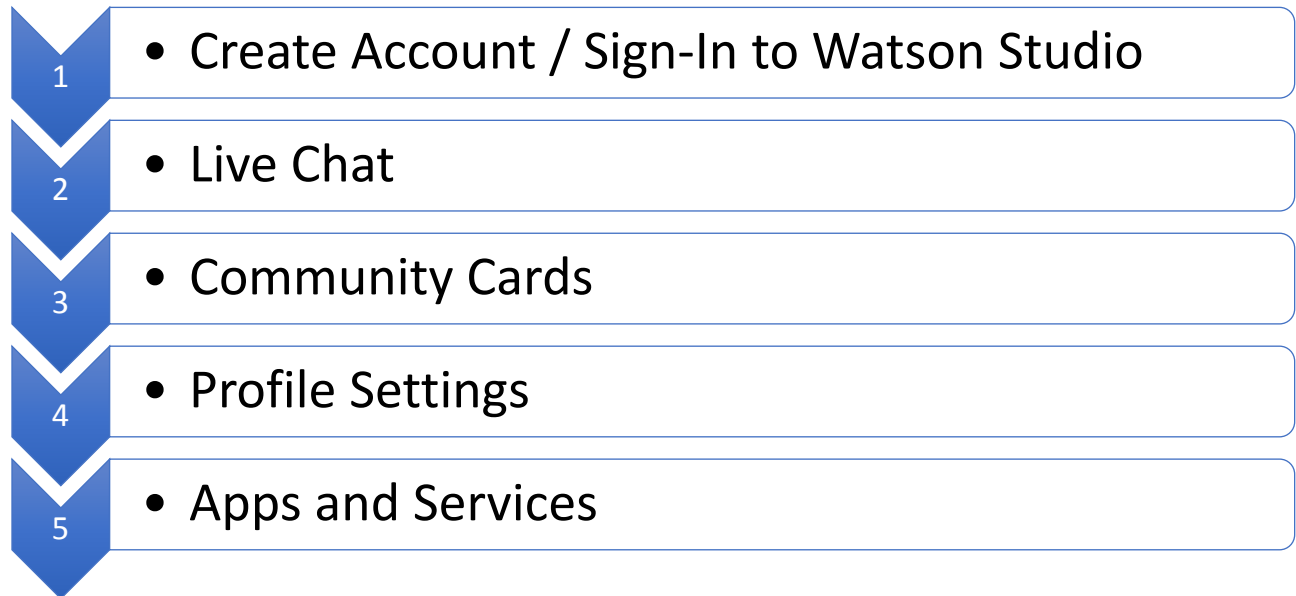
Software and Tools

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

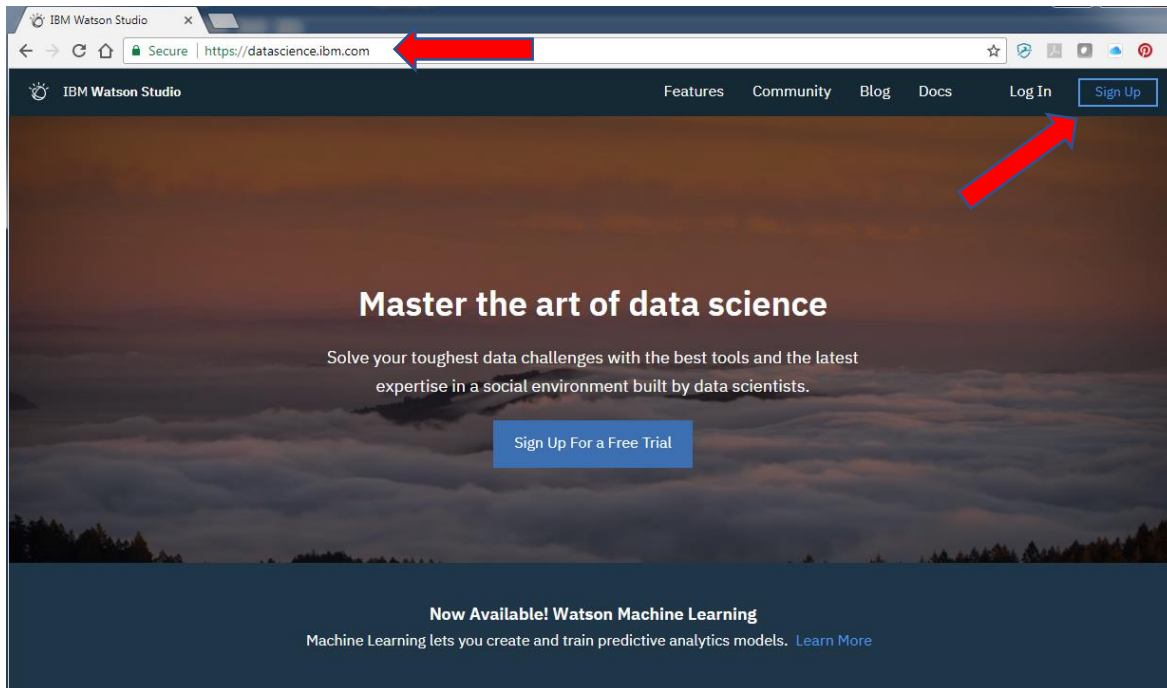


Lesson 1: Instructions

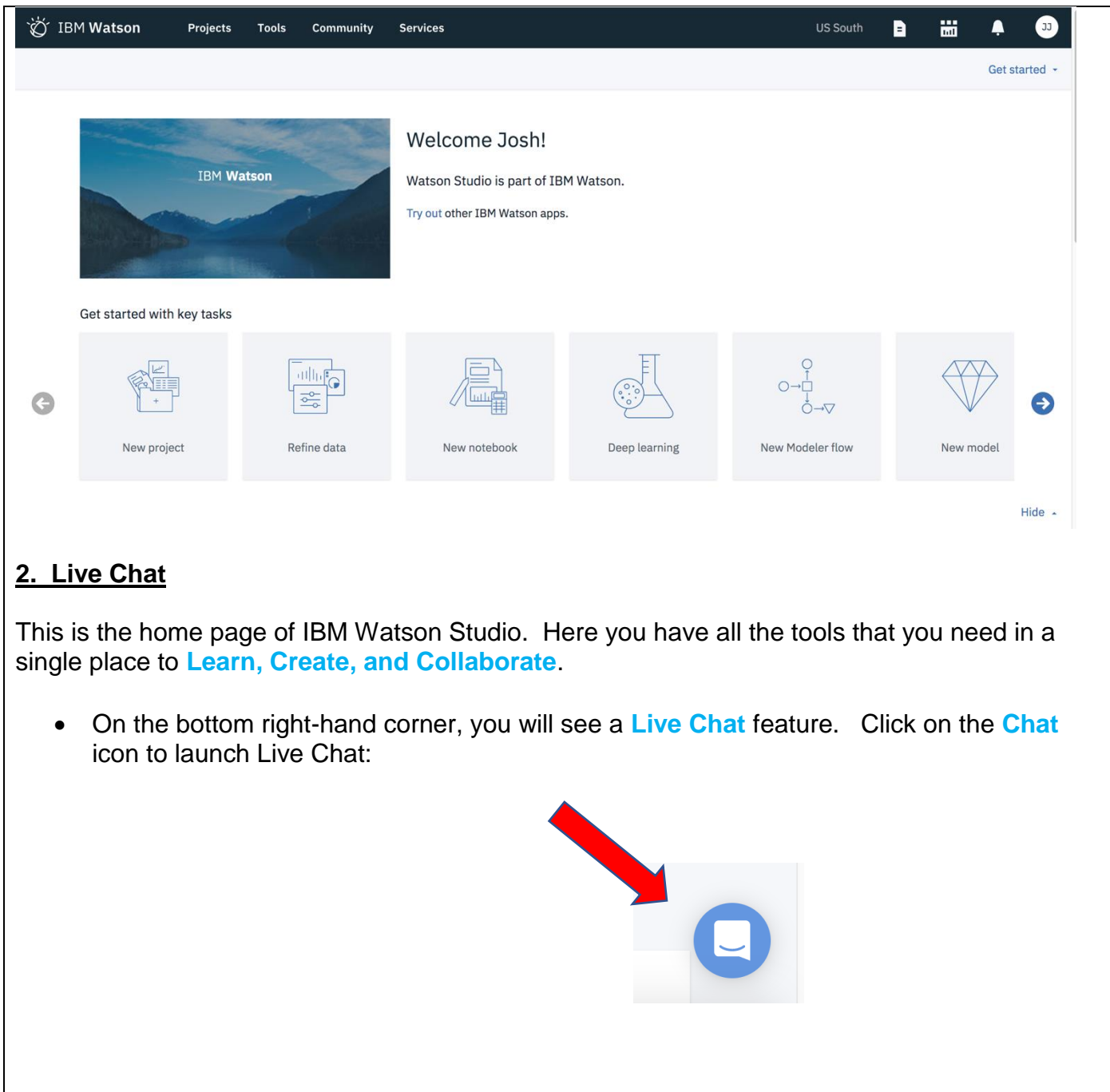
Action

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

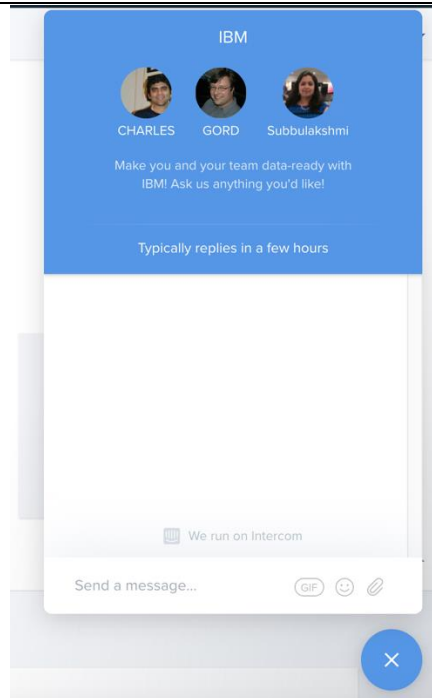
New project Refine data New notebook Deep learning New Modeler flow New model

Live Chat

2. Live Chat

This is the home page of IBM Watson Studio. 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:

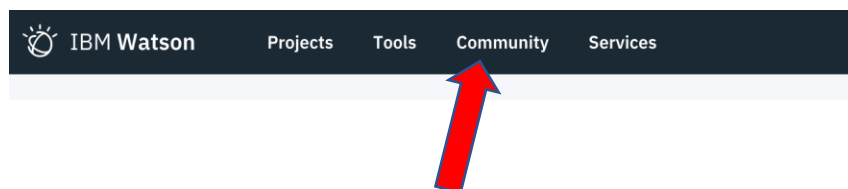


If you need assistance, start typing your message in the **Send a Message** box to connect with a live person. Through this Live Chat feature, you can also continue conversations the next time you log into Watson Studio.

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
 1

ARTICLE
 Introducing IBM Watson Studio
 AUTHOR: Armand Ruiz
 DATE: Mar 20, 2018
 TOPIC: Watson
 FORMAT: Web page
 8

ARTICLE
 Webinar: April 11 - Thinking inside the box:...
 AUTHOR: RStudio
 DATE: Apr 02, 2018
 TOPIC: Data Science
 FORMAT: Web page
 0

All content

ARTICLE
 Webinar: April 11 - Thinking inside the box:...
 AUTHOR: RStudio
 DATE: Apr 02, 2018
 TOPIC: Data Science
 FORMAT: Web page
 0

NOTEBOOK
 Watson Assistant Workspace Analysis with...
 AUTHOR: IBM
 DATE: Apr 02, 2018
 TOPIC: Communications
 4

TUTORIAL
 Build Deep Learning Architectures With...
 AUTHOR: developerWorks TV
 DATE: Apr 02, 2018
 LEVEL: Beginner
 TOPIC: Deep Learning +2
 0

NOTEBOOK
 Connect to Db2 Warehouse on Cloud and Db2...
 AUTHOR: IBM
 DATE: Mar 29, 2018
 TOPIC: Economy & Business
 8

NOTEBOOK
 From scikit-learn Model to Cloud with WML...

NOTEBOOK
 Access MySQL with Python

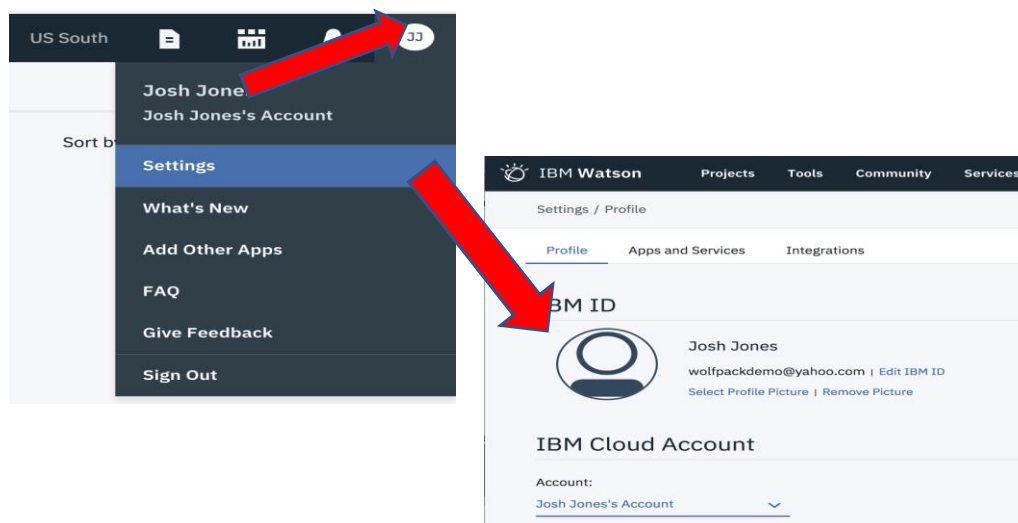
ARTICLE
 Using shell scripts to control data flows...

ARTICLE
 Working with data flows using Watson Data...

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

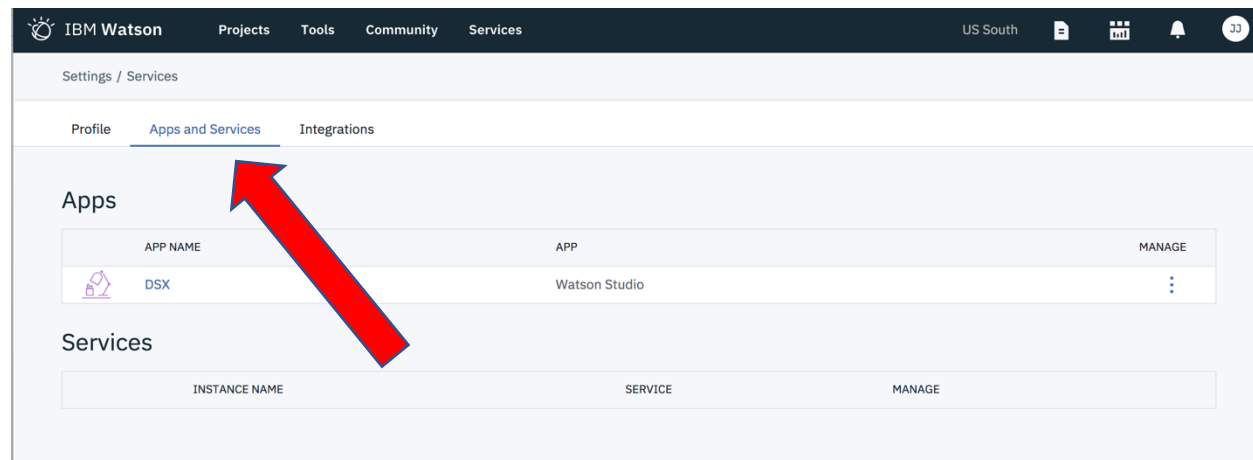
- Click on **Settings** to look at your **Profile**, **Apps and Services**, and **Integrations**. This is where you see the details of your IBM Cloud Account:



The image shows two parts of the IBM Watson interface. On the left, a dark sidebar menu is open, showing options like 'Settings', 'What's New', 'Add Other Apps', 'FAQ', 'Give Feedback', and 'Sign Out'. A red arrow points from the 'Settings' option to the right. On the right, the 'Settings / Profile' page is shown. It has tabs for 'Profile', 'Apps and Services', and 'Integrations'. The 'Profile' tab is active, displaying the user's IBM ID (Josh Jones), email (wolfpackdemo@yahoo.com), and account name (Josh Jones's Account). Another red arrow points from the 'Settings' menu item to the 'Profile' tab.

5. Apps and Services

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



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

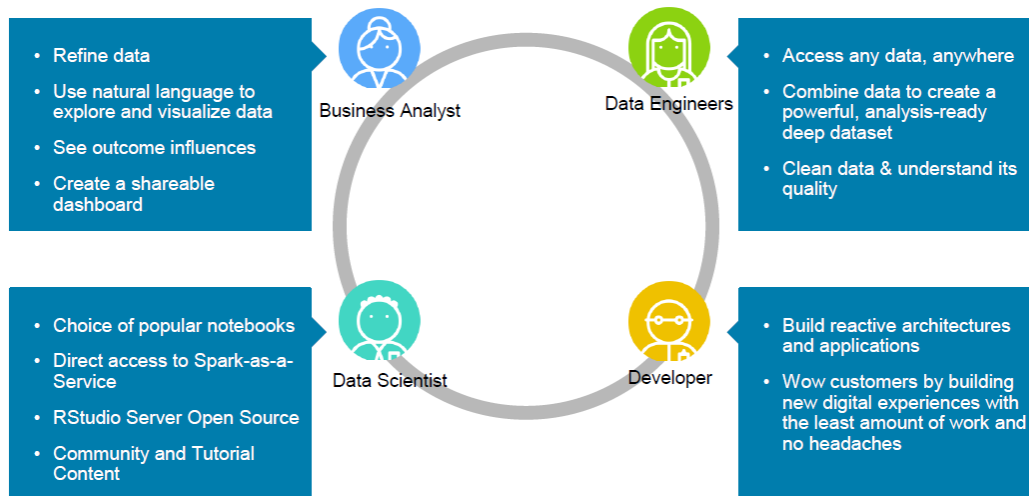
Lesson 2: Workflow Overview

- 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

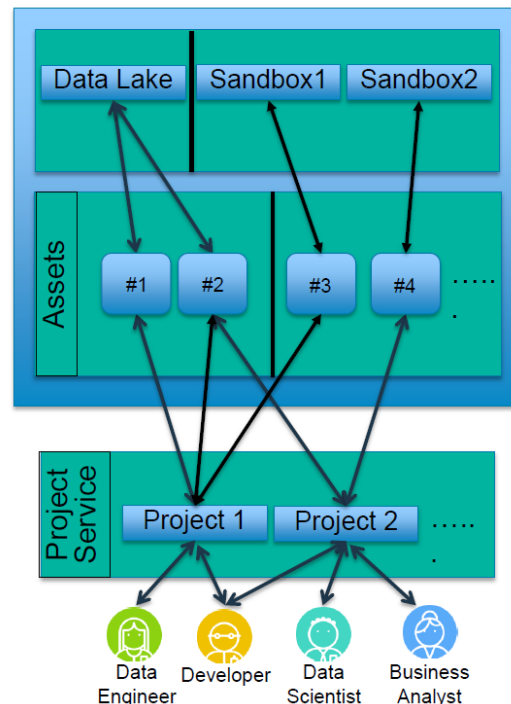
1. Project Overview

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



Projects make collaboration easier by:

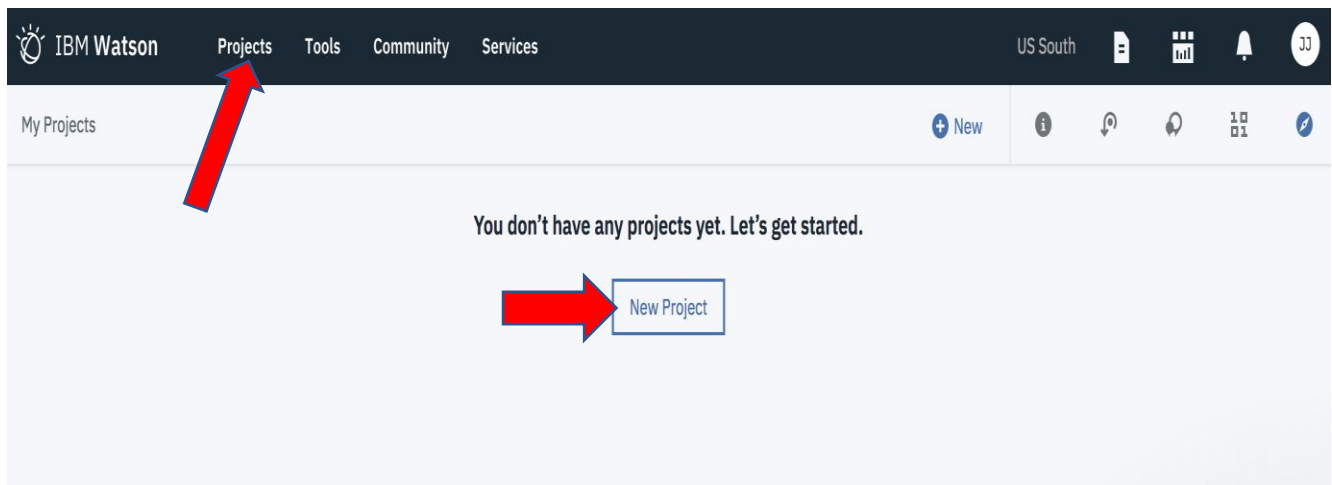
- Allowing different users and personas to share a set of assets
- Enabling users to collaborate and manage their notebooks, artifacts, plus more
- Providing three levels of rights: Viewers, Editors, and Admins



Action

2. Create New Project

- Navigate to <https://datascience.ibm.com>
- Login to Watson Studio
- On the top right side, click **Projects** and select **New Project**



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

New project

Define project details

Name

Customer Churn

86

Description

Customer churn analysis

2977

Choose project options

☐ Restrict who can be a collaborator ⓘ

Project will include integration with [Object Storage](#) for storing project assets.

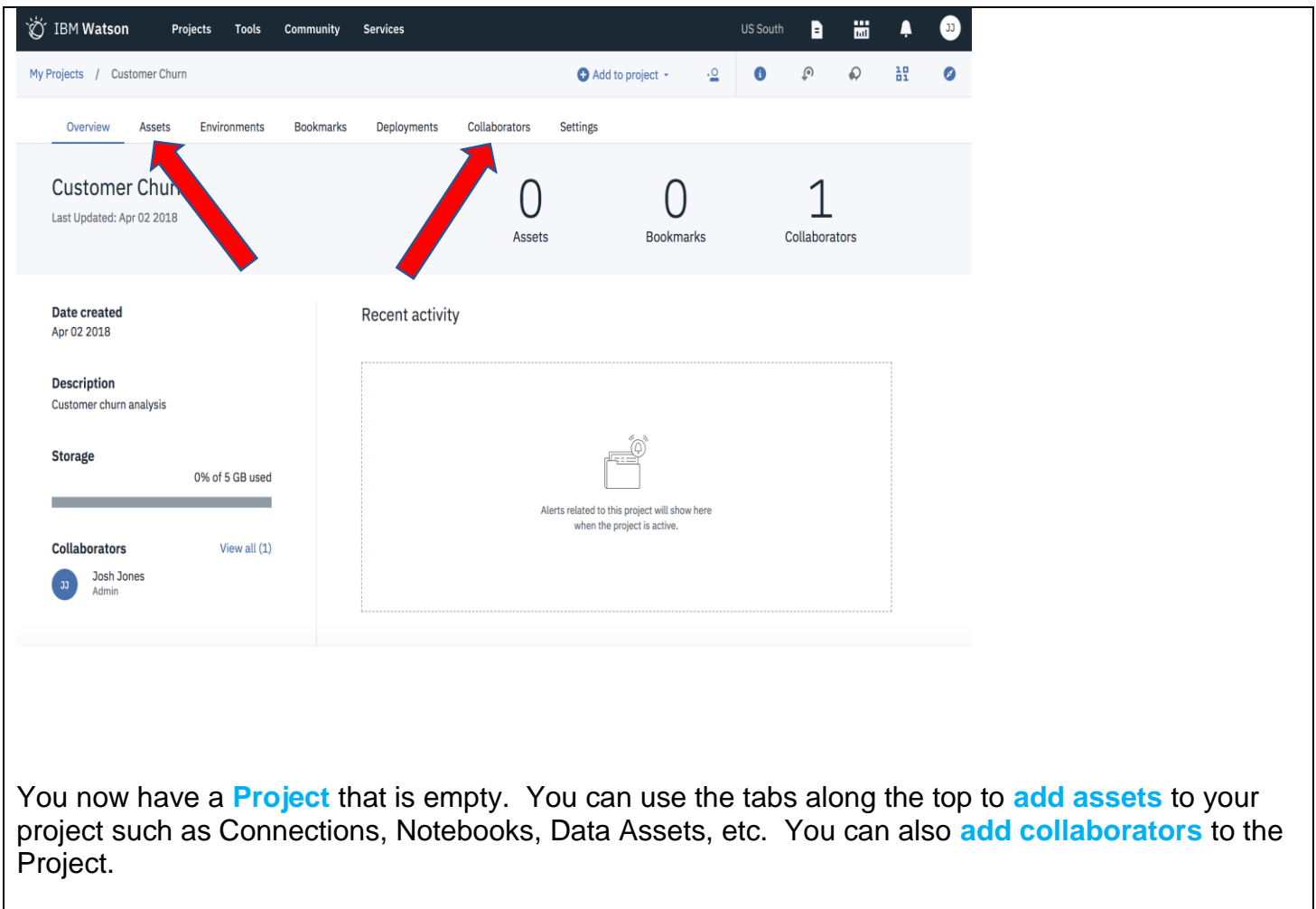
Define Storage:

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

Storage

cloud-object-storage-pd

- Click **Create**

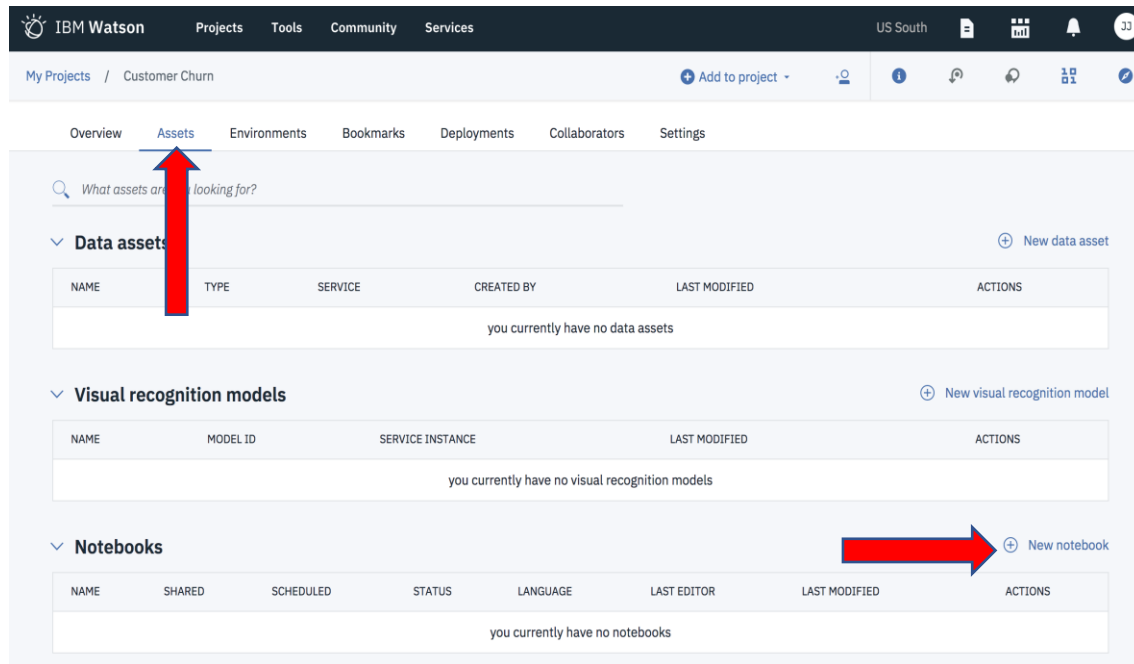


The screenshot shows the IBM Watson Project interface for a project named "Customer Churn". The top navigation bar includes "IBM Watson", "Projects", "Tools", "Community", and "Services". The project breadcrumb is "My Projects / Customer Churn". The main navigation tabs are "Overview", "Assets", "Environments", "Bookmarks", "Deployments", "Collaborators", and "Settings". The "Assets" tab is highlighted with a red arrow. The "Collaborators" tab is also highlighted with a red arrow. The "Assets" tab shows 0 Assets, 0 Bookmarks, and 1 Collaborators. The "Collaborators" tab shows 1 Collaborator (Josh Jones, Admin). The "Overview" tab shows the project name "Customer Churn", last updated "Apr 02 2018", a description "Customer churn analysis", storage usage "0% of 5 GB used", and a list of collaborators. The "Recent activity" section is empty, showing a message: "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.

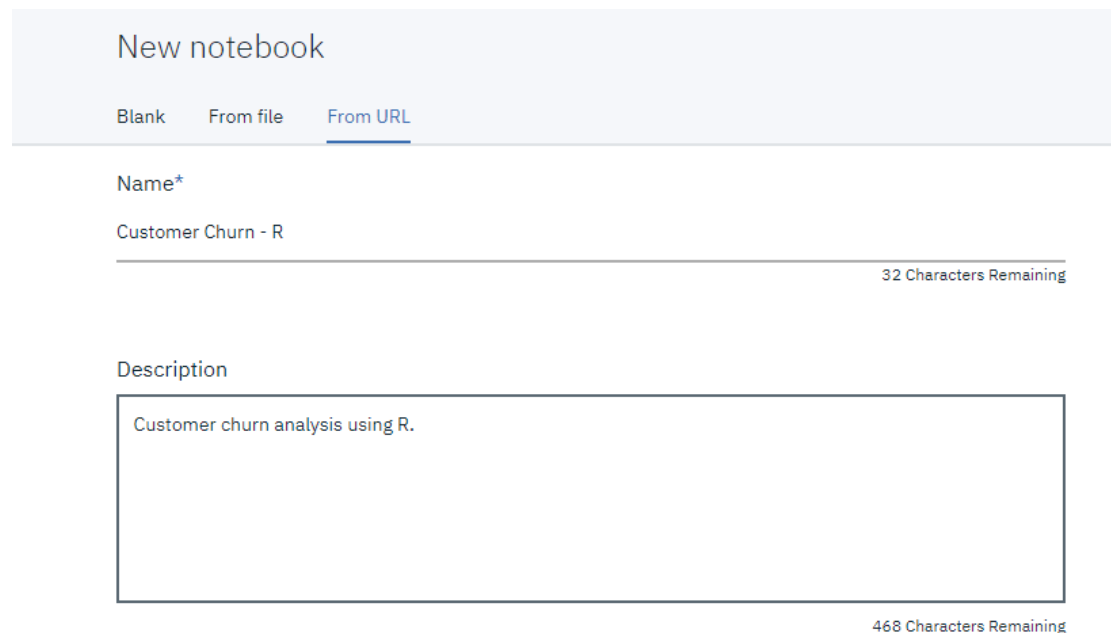
3. Create Notebook

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



The screenshot shows the IBM Watson interface. At the top, there's a navigation bar with 'IBM Watson' and tabs for 'Projects', 'Tools', 'Community', and 'Services'. Below this, a breadcrumb trail shows 'My Projects / Customer Churn'. The main content area has a tab bar with 'Overview', 'Assets', 'Environments', 'Bookmarks', 'Deployments', 'Collaborators', and 'Settings'. The 'Assets' tab is selected, and a red arrow points to it. Below the tab bar, there's a search bar and three sections: 'Data assets', 'Visual recognition models', and 'Notebooks'. Each section has a table with columns and a '+ New' button. A red arrow points to the '+ New notebook' button in the 'Notebooks' section.

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



The screenshot shows the 'New notebook' form. At the top, there's a title 'New notebook'. Below it, there are three tabs: 'Blank', 'From file', and 'From URL'. The 'From URL' tab is selected. Below the tabs, there's a 'Name*' field with the text 'Customer Churn - R' and a '32 Characters Remaining' indicator. Below the name field, there's a 'Description' field with the text 'Customer churn analysis using R.' and a '468 Characters Remaining' indicator.

- 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 **Customer Churn-R.ipynb** then choose **Copy link address**. Go back to the **Watson Studio New Notebook** page.

Paste URL into **Notebook URL** text box. Select **Default R Environment XS (2 vCPU and 8GB RAM)** as the runtime. Then click **Create Notebook**:

Notebook URL*

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

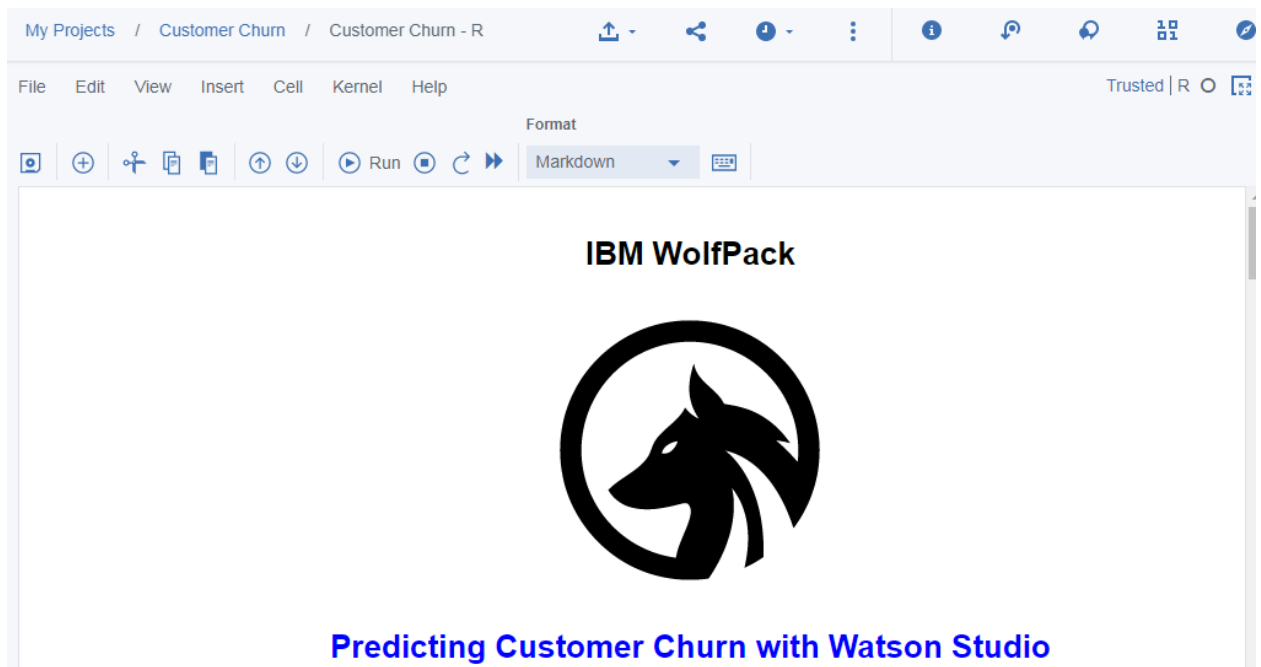
Select runtime* Includes notebook environments ⓘ

Default R Environment XS (2 vCPU and 8 GB RAM) 

The selected Runtime has 2 vCPU and 8 GB RAM and consumes 1 capacity units per hour.

[Learn more](#) about capacity unit hours and Watson Studio pricing plans.

You should now see:



Lesson 2 Continued in [Customer Churn – R] 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

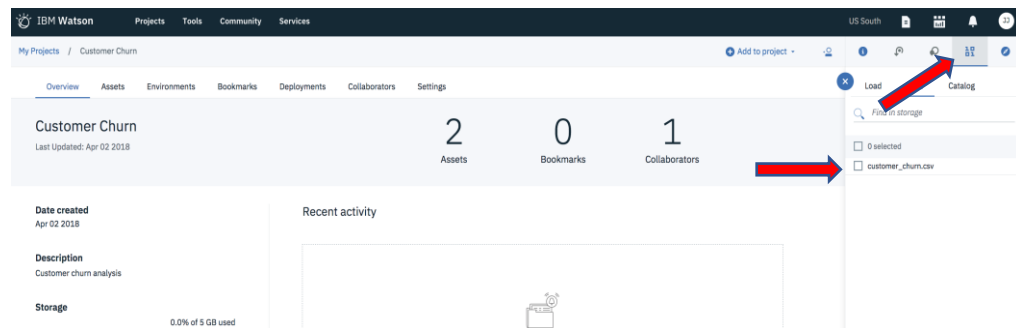
Lesson 3: Instructions

Action

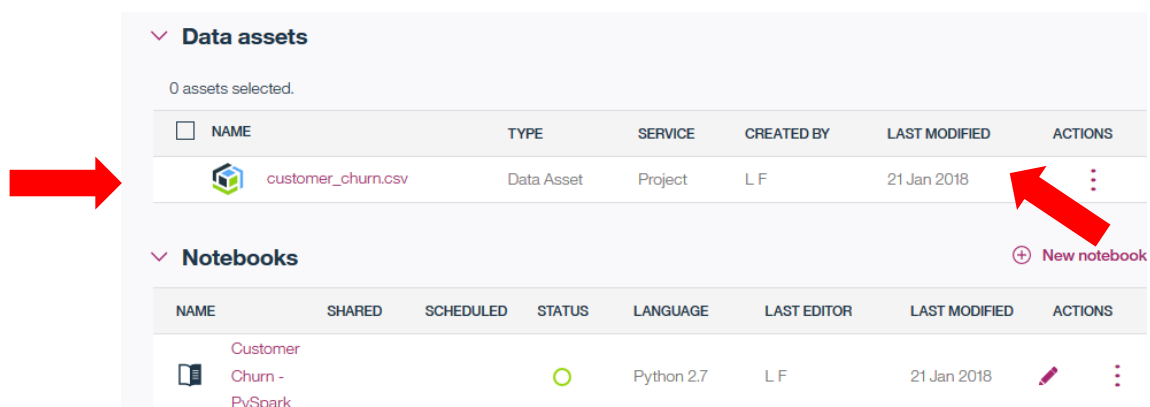
1. Load Data from Local File

- In a separate browser navigate to: [Customer Churn Data:](https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform/tree/master/Data)
<https://github.com/team-wolfpack/Predicting-Customer-Churn-with-Watson-Data-Platform/tree/master/Data>
- Download **customer_churn_data.csv** and save it to 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:

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




Navigate back to **Assets** and see the new “Data Asset”:



2. Create Machine Learning Flow

Action

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

▼ **Modeler flows**  [+ New flow](#)

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
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**”:

Modeler

[New](#) [From file](#) [From example](#)

Name*

CustomerChurn-Flow 32

Description

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

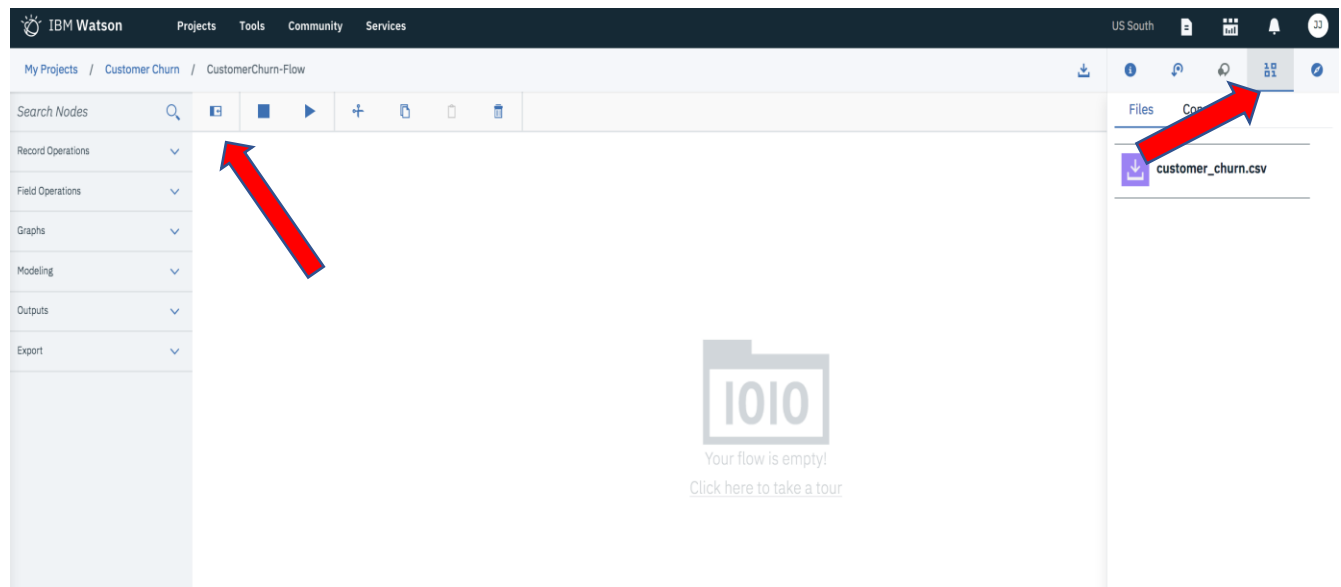
- Click on “**Create**”

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

<input type="checkbox"/> Field name ^	Data type ^
<input type="checkbox"/> ID	integer
<input checked="" type="checkbox"/> CHURN	string
<input checked="" type="checkbox"/> Gender	string
<input checked="" type="checkbox"/> Status	string
<input checked="" type="checkbox"/> Children	unknown
<input checked="" type="checkbox"/> Est Income	unknown
<input checked="" type="checkbox"/> Car Owner	string
<input checked="" type="checkbox"/> Age	unknown
<input checked="" type="checkbox"/> LongDistance	unknown
<input checked="" type="checkbox"/> International	unknown
<input checked="" type="checkbox"/> Local	unknown
<input checked="" type="checkbox"/> Dropped	unknown
<input checked="" type="checkbox"/> Paymethod	string
<input checked="" type="checkbox"/> LocalBilltype	string
<input checked="" type="checkbox"/> LongDistanceBilltype	string

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

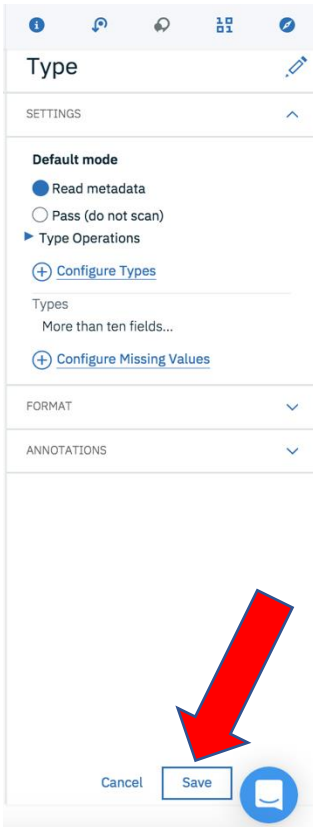
Configure Types

Types

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

- Click **“OK”**.
- Click **“Save”** to exit



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 C&R Tree object should now say **“CHURN”**. Double click on this object.
- Click on **“FIELDS”**, Target should be set to **“CHURN”**

CHURN

FIELDS

☐ Use custom field roles

Target

CHURN

Inputs

ID

Gender

Status


Car Owner

+

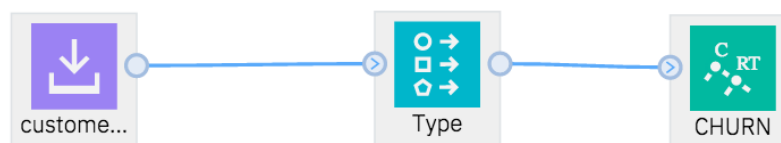
+

Add Columns

- Click on “**Save.**”



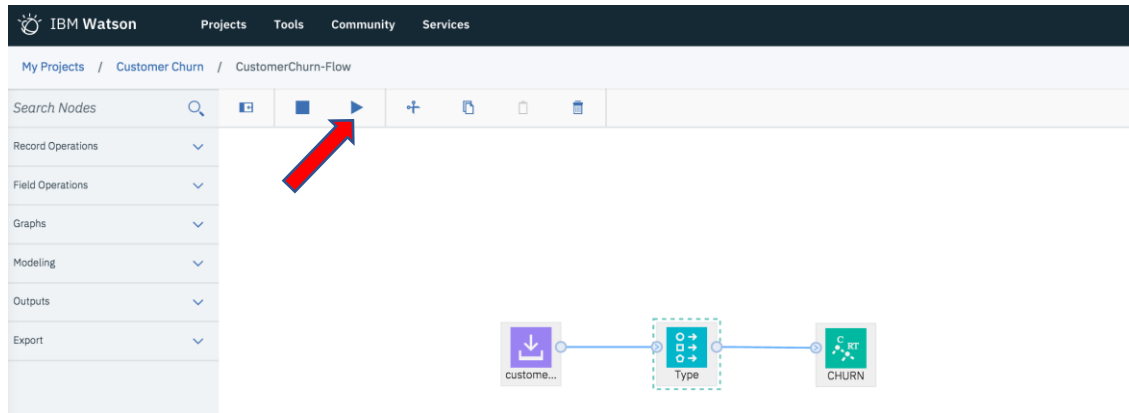
- Your palette should resemble this:



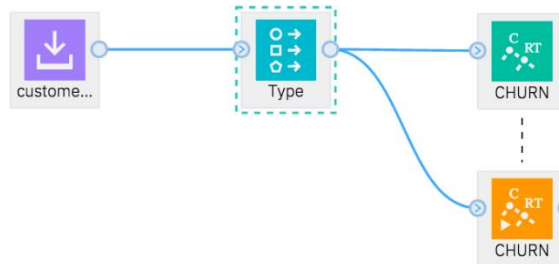
Action

5. Run Flow to Create Nugget

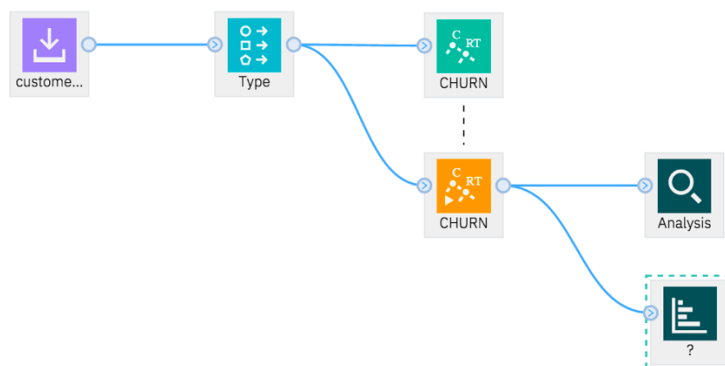
- Run the flow by clicking on the **“Run”** icon at the top of the workspace.



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



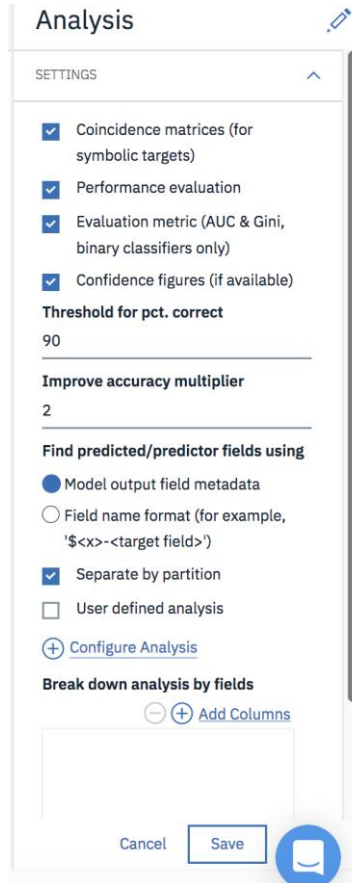
- From the palette add an **“Analysis”** object to the workspace, you will find it under the **“Outputs”** drop down. Also, from the **“Graphs”** drop down add a **“Distribution”** object to the workspace. Connect the nugget to each of them:



Action

6. Add & Configure Analysis Object – Measure Model Performance

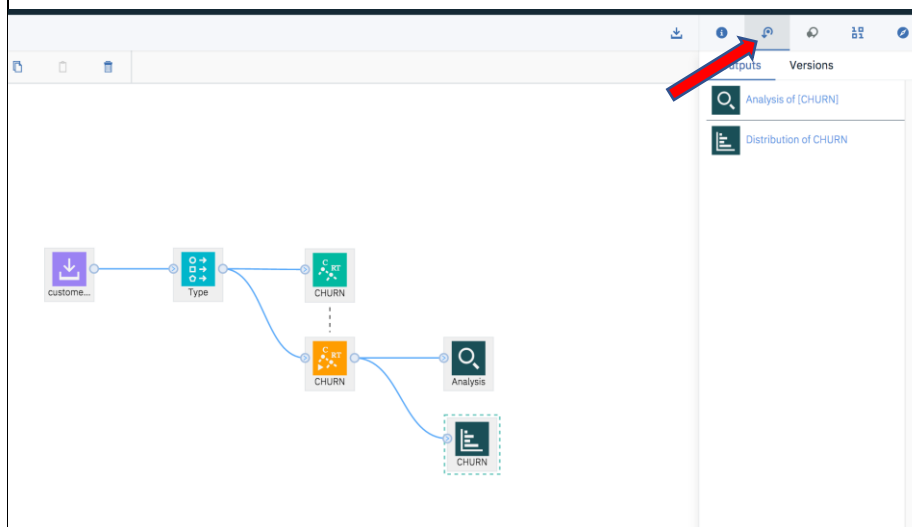
- Double click on “**Analysis**” and check off the four checkboxes, leave the rest as default:



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

Action	
	?
	PLOT
	Plot
	<input checked="" type="radio"/> Specified
	<input type="radio"/> All flags (true values)
	Field (discrete)
	CHURN
	Color (discrete)
	\$R-CHURN
	Sort
<input checked="" type="radio"/> Alphabetic	
<input type="radio"/> By count	
<input checked="" type="checkbox"/> Normalize by color	
<input type="checkbox"/> Use proportional scale	
APPEARANCE	▼
ANNOTATIONS	▼

- 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:

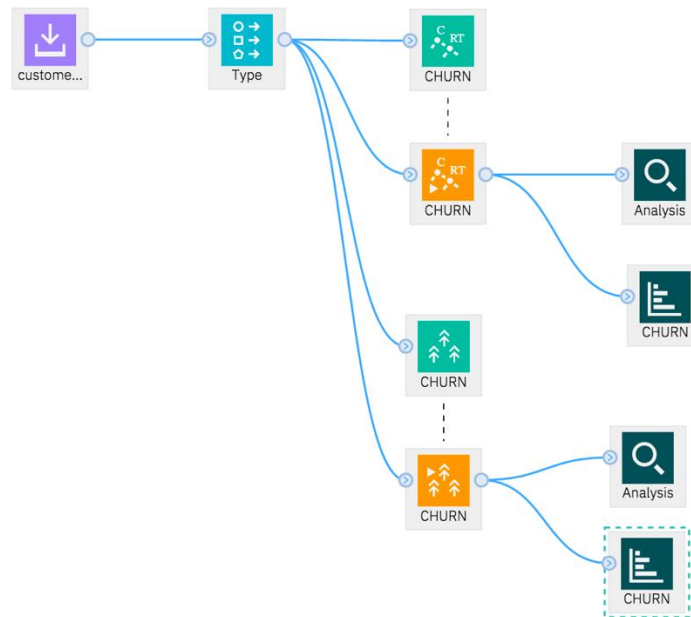


- Explore the results

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:



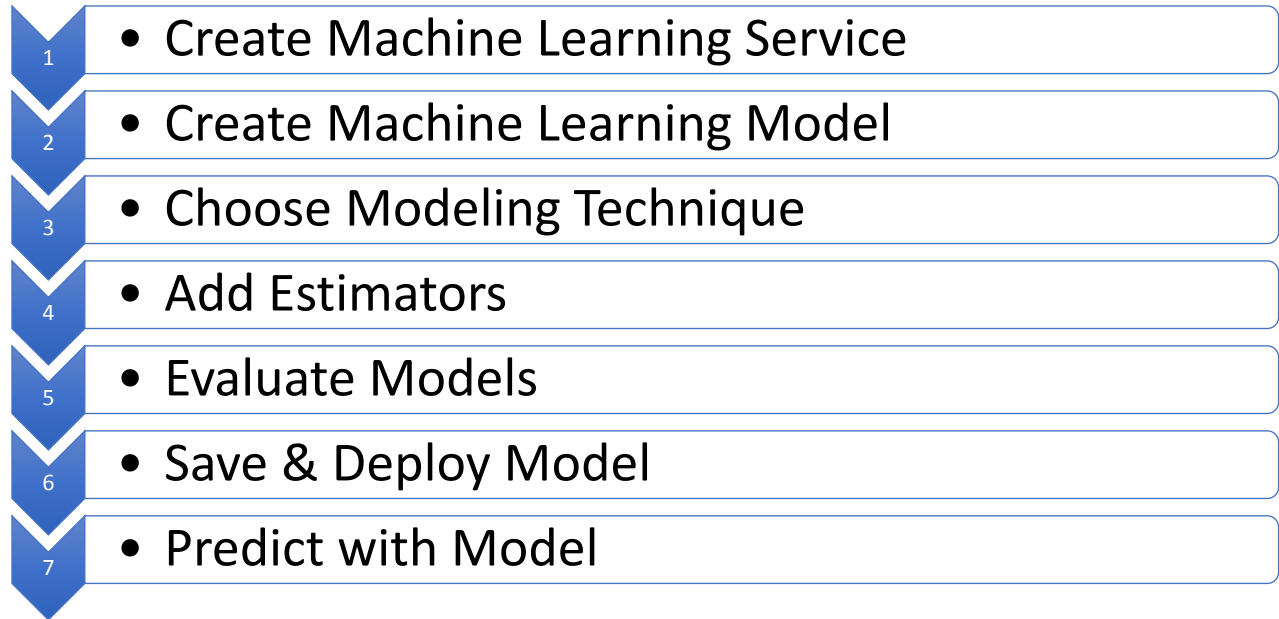
- Explore the results.

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

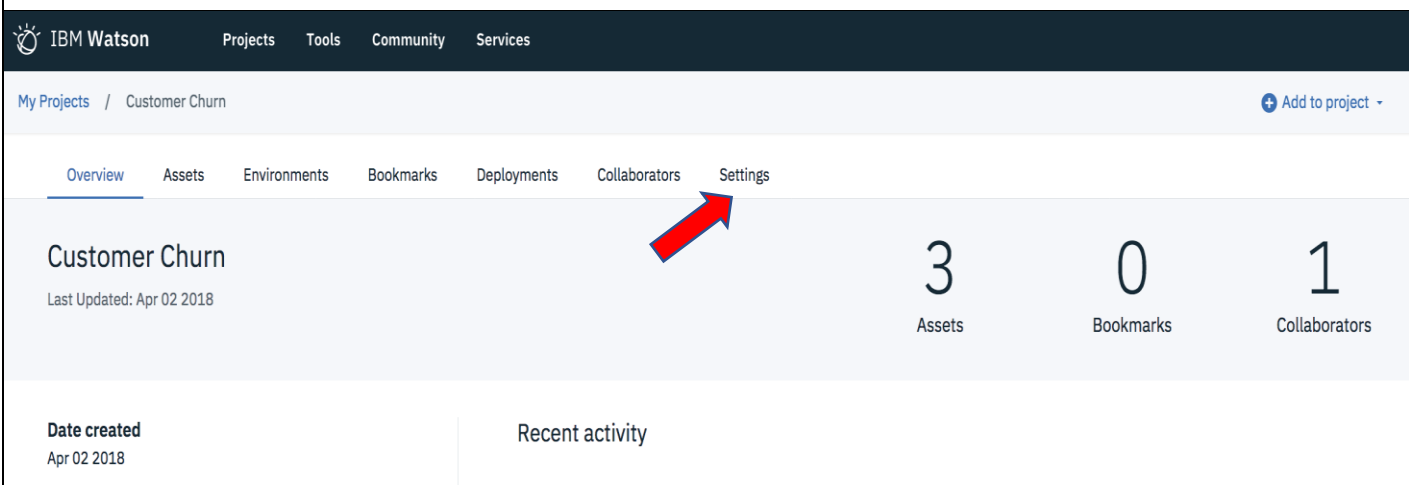


Lesson 4: Instructions

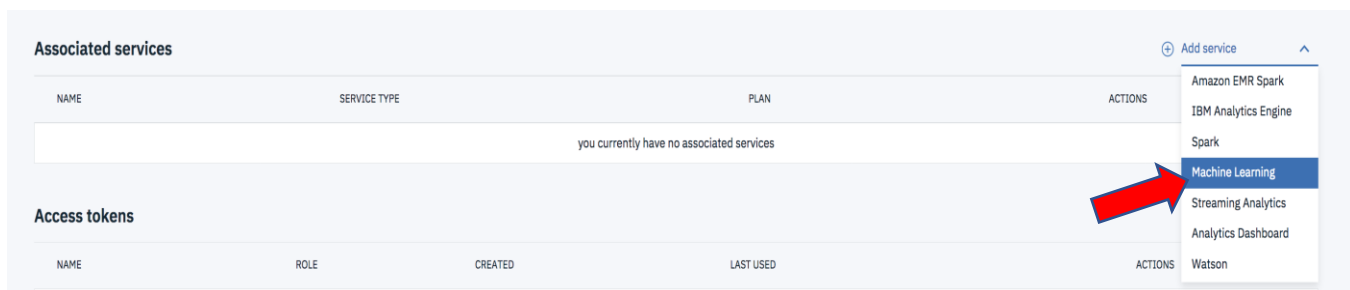
Action

1. Create Machine Learning Service

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



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



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

Action

Machine Learning

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 and streaming.

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



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

Confirm Creation

Organization: louisfrolio@gmail.com

Plan

Lite



Space

dev



Service name

dsx-wml-lab

Cancel

Confirm

- 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

✓ Models

[+ New model](#)

NAME	STATUS	TYPE	RUNTIME	LAST MODIFIED	ACTIONS
------	--------	------	---------	---------------	---------

you currently have no models

- 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**.

Existing New

Apache Spark

Apache Spark is an open source cluster computing framework optimized for extremely fast and large scale data processing, which you can access via the newly integrated notebook interface IBM Analytics for Apache Spark. You can connect to your existing data sources or take advantage of the on-demand big data optimization of Object Storage. Spark plans are based on the maximum number of executors available to process your analytic jobs. Executors exist only as long as they're needed for processing, so you're charged only for processing done.

Features

Incredibly Fast

Apache Spark delivers 100x the performance of Apache Hadoop for certain workloads because of its advanced in-memory computing engine.

Easy to Use and Powerful

Apache Spark's Streaming and SQL programming models backed by MLlib and GraphX make it incredibly easy for developers and data scientists to build apps that exploit machine learning and graph analytics. Because the service is 100% compatible with Apache Spark, developers can build their apps and run them against the IBM managed service to benefit from operational, maintenance, and hardware excellence.

Convenient Data Storage

Object Storage enables a convenient way to upload your data from a file for immediate use by your Spark instance. You can set up Object Storage directly from the Spark service interface.

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

PLAN	FEATURES	PRICING
------	----------	---------

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

[Terms](#)
[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**”:

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


Action

Select data asset

[+ Add Data Assets](#)

The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.

What asset are you looking for?

NAME	TYPE	SERVICE
 customer_churn.csv	Data Asset	Project

Close Next

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

Select a technique

Column value to predict (Label Col)

CHURN (String) 

Feature columns

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

Car Owner (String), Paymethod (String), LongDistanceBilltype (String), 

Usage (Decimal), RatePlan (Decimal) 

 Add Estimators

Configured estimators

 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

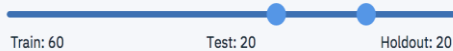
Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.



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**:

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.



Validation Split

Train: 60 Test: 20 Holdout: 20

Close Previous Next

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 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

CustomerChurn-WML

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

Summary

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

Input Schema

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](#)

NAME	STATUS	DEPLOYMENT TYPE	ACTIONS
Your model is not deployed.			
			+ 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 church

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

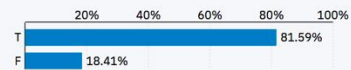
1

Est Income

Predict

Predicted value for CHURN

T



End of Lesson 4

End of Hands-on Workshop

Thank You