

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 Churn with Watson Data Platform Lab_V2.docx	4/2/2018
1.3	Predicting Customer Churn with Watson Studio.docx	4/23/2018
1.4	Predicting Customer Churn with Watson Studio.docx	5/1/2018
1.5	Predicting Customer Churn with Watson Studio.docx	5/10/2018

Prepared & Revised by:

Louis Frolio – louis.folio@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: Data Refinery	14
Lesson 2: Workflow Overview	14
Lesson 2: Instructions	15
Lesson 3: Jupyter Notebook	34
Lesson 3: Workflow Overview	35
Lesson 3: Instructions	36
Lesson 4: Machine Learning Flows	41
Lesson 4: Workflow Overview	41
Lesson 4: Instructions	42
Lesson 5: Watson Machine Learning	52
Lesson 5: Workflow Overview	53
Lesson 5: Instructions	54



Lab Environment Overview

Software and Tools

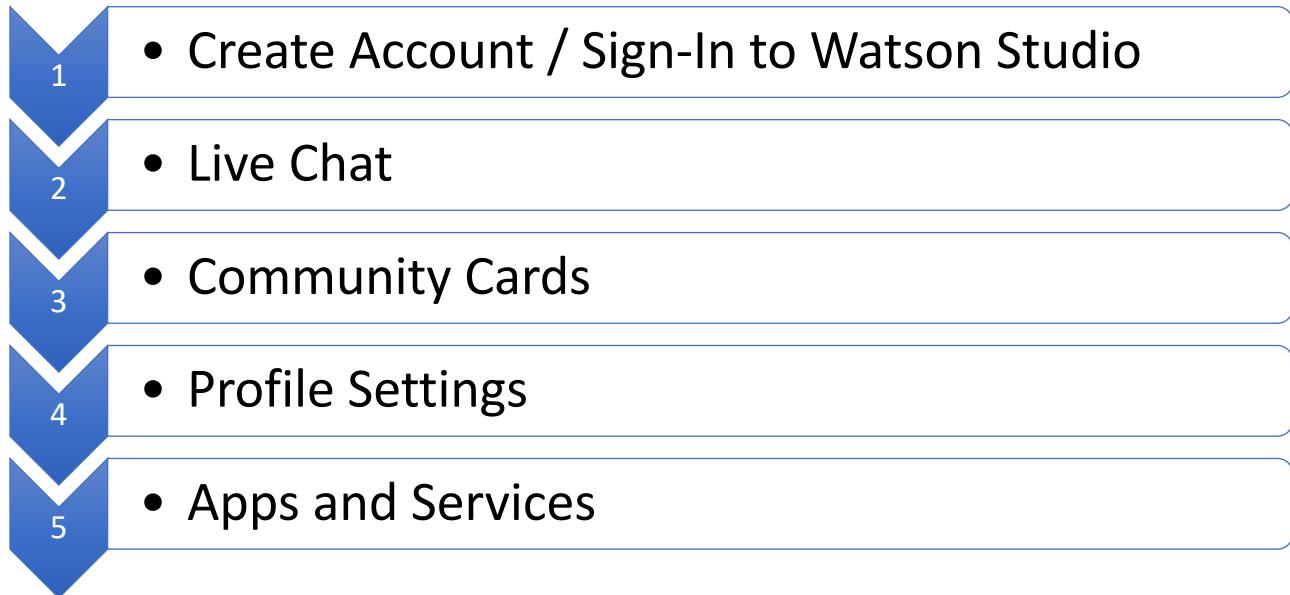
Software	Link
IBM Watson Studio	https://datascience.ibm.com
GitHub	https://github.com/team-wolfpack



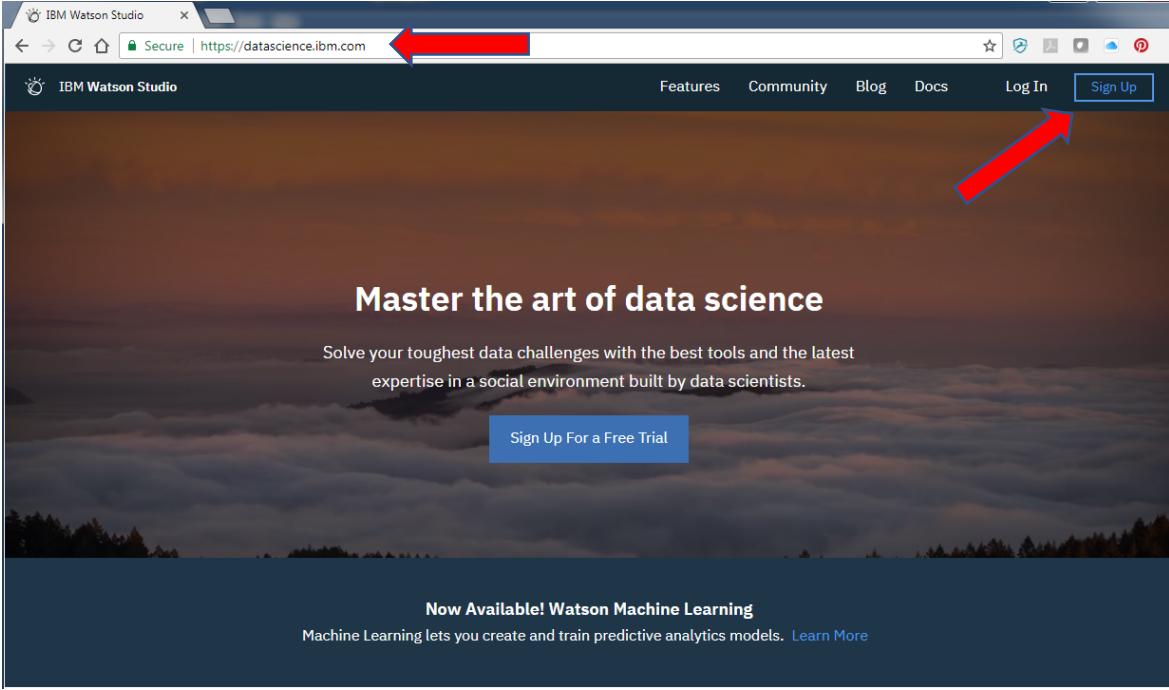
Lesson 1: Watson Studio Signup & Home Page

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
<p>1. <u>Create Account/Sign In to Watson Studio</u></p> <ul style="list-style-type: none">• Open web browser and navigate to: https://datascience.ibm.com  <p>Master the art of data science</p> <p>Solve your toughest data challenges with the best tools and the latest expertise in a social environment built by data scientists.</p> <p>Sign Up For a Free Trial</p> <p>Now Available! Watson Machine Learning Machine Learning lets you create and train predictive analytics models. Learn More</p>

Add other Watson apps


Knowledge Catalog
 Discover, catalog, and securely share enterprise data.
[Try it for free](#)


Watson Studio
 Embed AI and machine learning into your business. Create custom models using your own data.
[Try it for free](#)

- Select the **Lite** Plan and click **Create**

Watson Studio

Features

Use what you know, learn what you don't
 Start from a tutorial, start from a sample, or start from scratch. Tap into the power of the best of open source (RStudio, Jupyter Notebooks) and Watson services for flexible model creation. Use Python, R, or Scala. Stop downloading and configuring analysis environments and start getting insights.

Power on demand
 Enterprise-scale features on demand. From data exploration and preparation, to enterprise-scale performance. Manage your data, your analytical assets, and your projects in a secured cloud environment.

Be a founding member
 Take advantage of shared data sets, notebooks, models, and tutorials. Share your work with your team and your peers across job roles. Join a vibrant community of data scientists, developers, and domain experts across industries, functions, and organization types.

Collaborate for better outcomes
 Work with your peers on projects to find better solutions together. Share your knowledge and your work easily with visualizations and code – and help fuel the advancement of data science and AI for all.

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

PLAN	FEATURES	PRICING
 Lite	1 authorized user 50 capacity unit-hours monthly limit 1 free small compute environment with 1 vCPU and 4 GB RAM (does not require capacity unit-hours)	Free

The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment.

- Select **Confirm**

Confirm Creation

Organization: wolfpackdemo2@gmail.com

Plan

Lite

Space

dev

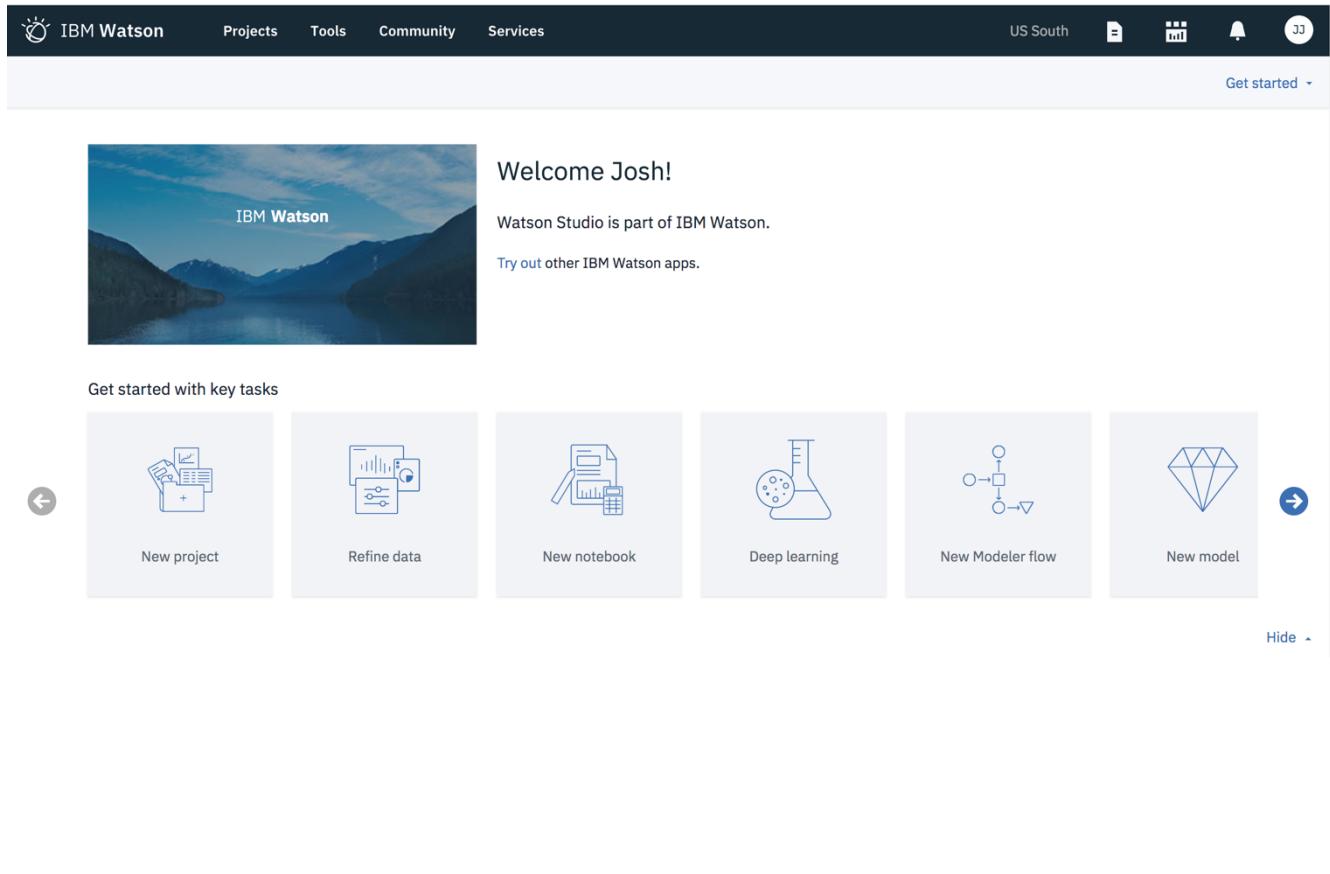
Service name

data-science-experience-ch

Cancel

Confirm

- You should now see your Watson Studio Homepage



The screenshot shows the Watson Studio homepage. At the top, there is a navigation bar with links for IBM Watson, Projects, Tools, Community, Services, US South, and a user profile icon. Below the navigation bar, a banner features a scenic landscape with mountains and water, and the text "Welcome Josh! Watson Studio is part of IBM Watson. Try out other IBM Watson apps." A "Get started" button is located in the top right corner of the banner area.

On the left side, there is a sidebar titled "Get started with key tasks" containing six cards:

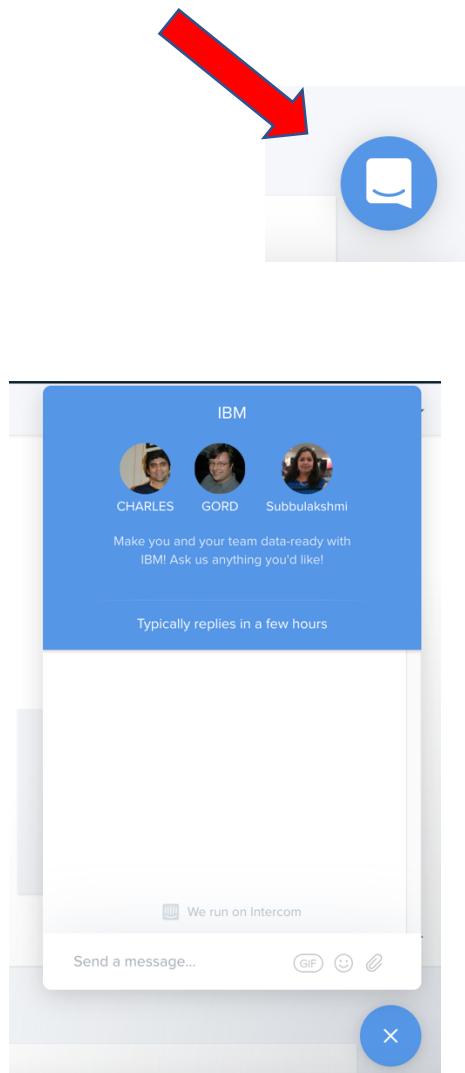
- New project (Icon: document with plus sign)
- Refine data (Icon: document with circular arrows)
- New notebook (Icon: document with chart)
- Deep learning (Icon: flask with brain)
- New Modeler flow (Icon: diamond shape with arrows)
- New model (Icon: diamond shape)

At the bottom right of the sidebar, there is a "Hide" button.

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:

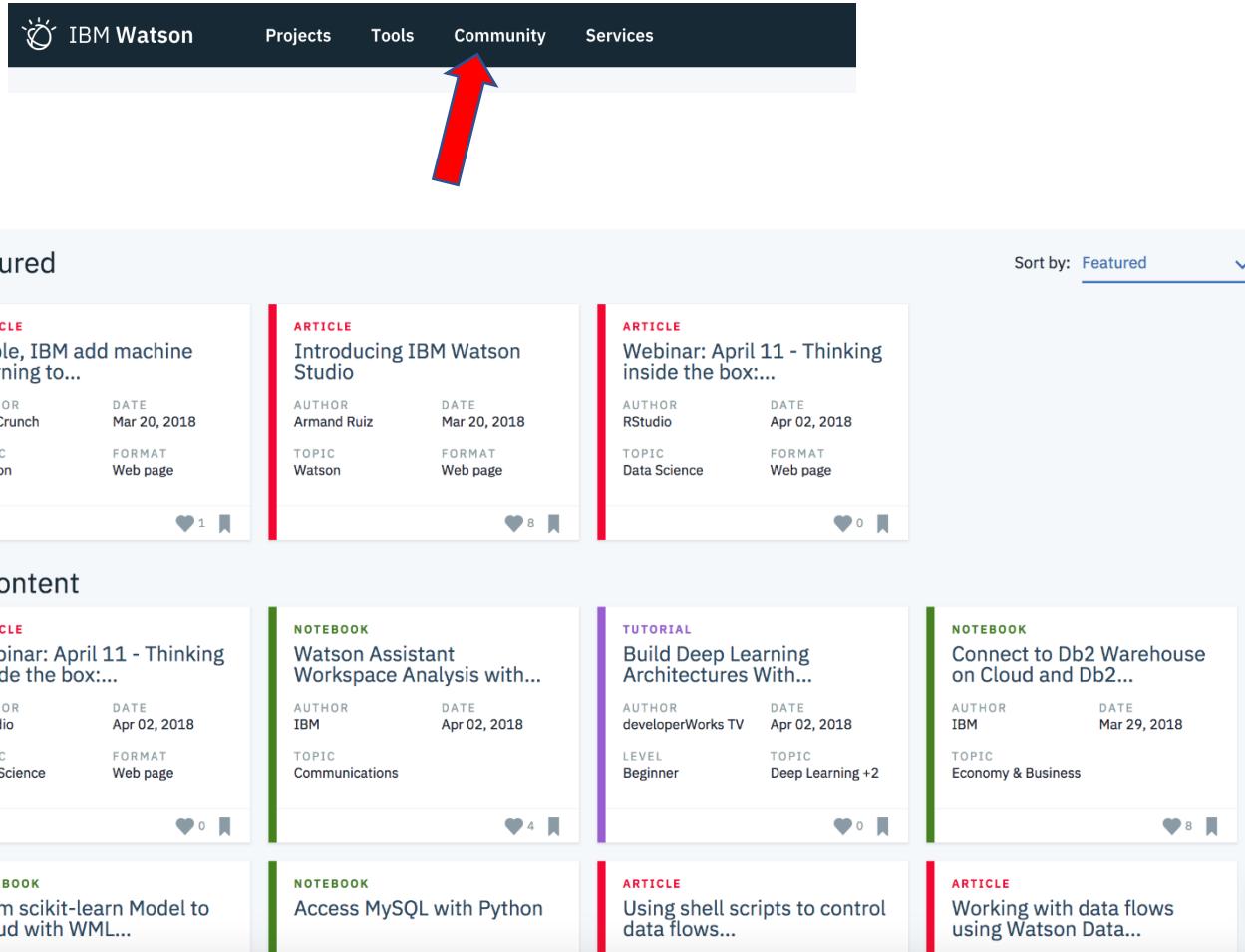


If you need assistance, start typing your message in the [Send a Message](#) box to connect with a live person. Through this Live Chat feature, you can also continue conversations the next time you log into Watson Studio.

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**:



Sort by: Featured ▾

ARTICLE	Introducing IBM Watson Studio	ARTICLE	Webinar: April 11 - Thinking inside the box:...
AUTHOR TechCrunch	DATE Mar 20, 2018	AUTHOR Armand Ruiz	DATE Mar 20, 2018
TOPIC Watson	FORMAT Web page	AUTHOR RStudio	DATE Apr 02, 2018
	1	8	0

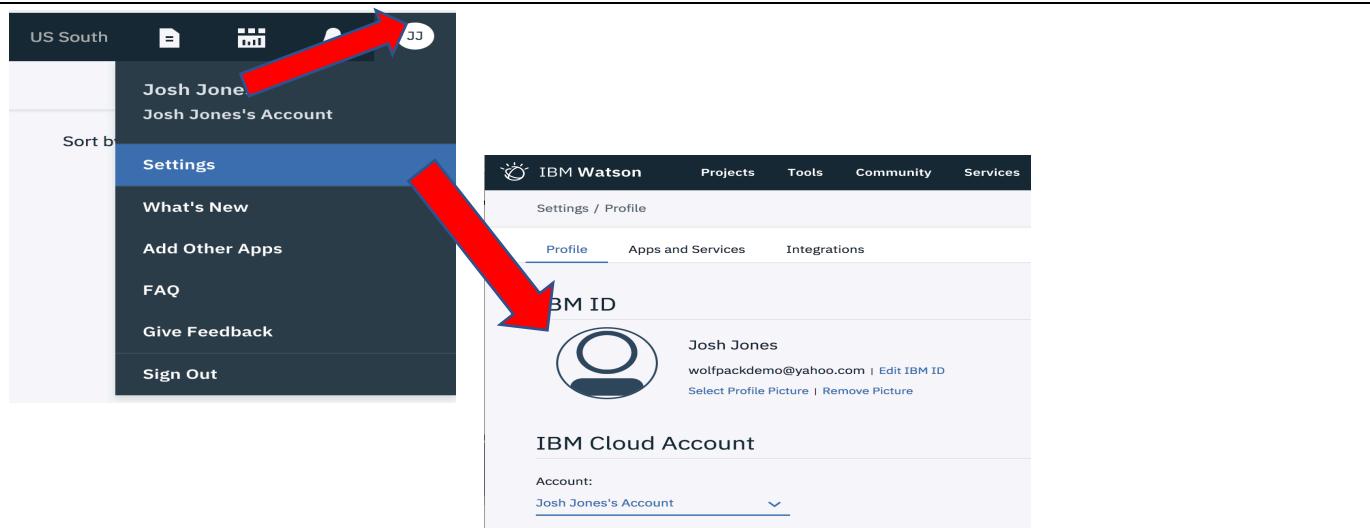
ARTICLE	Watson Assistant Workspace Analysis with...	TUTORIAL	Build Deep Learning Architectures With...	NOTEBOOK	Connect to Db2 Warehouse on Cloud and Db2...
AUTHOR RStudio	DATE Apr 02, 2018	AUTHOR developerWorks TV	DATE Apr 02, 2018	AUTHOR IBM	DATE Mar 29, 2018
TOPIC Data Science	FORMAT Web page	LEVEL Beginner	TOPIC Deep Learning +2	TOPIC Communications	TOPIC Economy & Business
	0	4	0	8	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

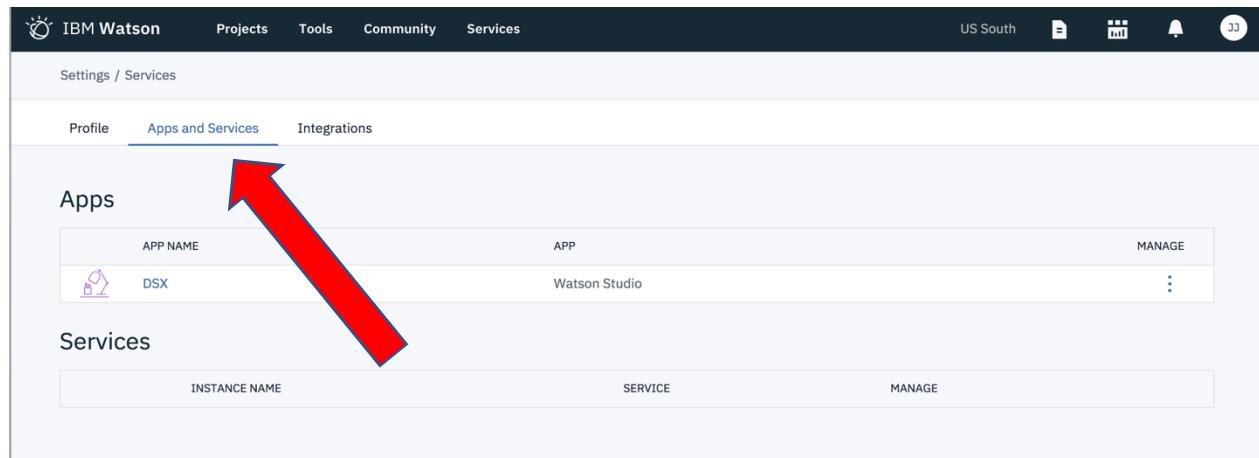
- Click on **Settings** to look at your **Profile, Apps and Services, and Integrations**. This is where you see the details of your IBM Cloud Account:



The screenshot shows the IBM Watson Settings / Profile page. At the top, there's a navigation bar with 'US South' and other icons. Below it, the user's name 'Josh Jones' and account 'Josh Jones's Account' are displayed. A red arrow points from the top right towards the user profile picture. The main content area has a sidebar with links like 'Settings', 'What's New', 'Add Other Apps', 'FAQ', 'Give Feedback', and 'Sign Out'. To the right of the sidebar, there's a 'Profile' section with a placeholder profile picture, the user's name 'Josh Jones', and their email 'wolfpackdemo@yahoo.com'. Below that is the 'IBM Cloud Account' section, which shows the account name 'Josh Jones's Account'. A red arrow points from the bottom left towards the 'Apps and Services' tab.

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 Settings / Services page. The top navigation bar includes 'IBM Watson', 'Projects', 'Tools', 'Community', 'Services', 'US South', and other icons. Below the navigation is a 'Settings / Services' header. The main content area has three tabs: 'Profile' (selected), 'Apps and Services' (highlighted with a blue underline), and 'Integrations'. A large red arrow points from the bottom left towards the 'Apps and Services' tab. Under the 'Apps' section, there's a table with one row for 'DSX' (APP NAME) which is 'Watson Studio'. Under the 'Services' section, there's a table with one row for 'INSTANCE NAME' (SERVICE) which is 'Watson Studio'.

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

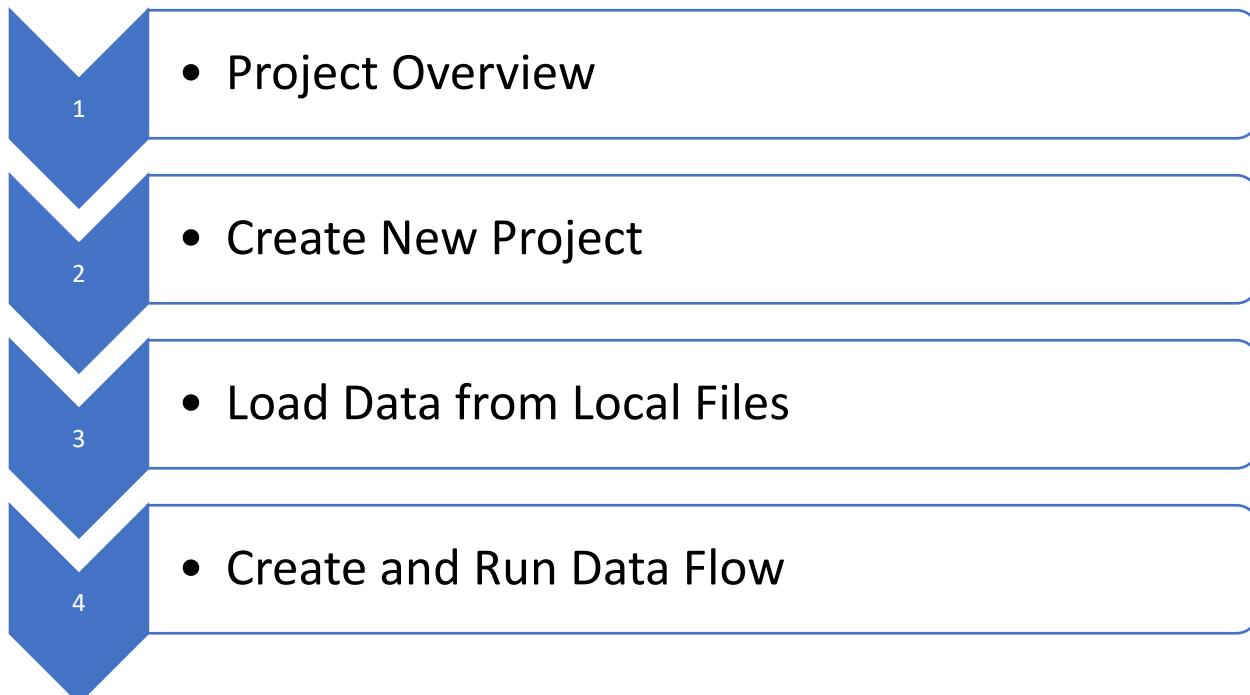
Integrations is where you configure Watson Studio for GitHub integration.

End of Lesson 1

Lesson 2: Data Refinery

Purpose:	This lesson introduces projects within Watson Studio, their purpose, value, and how they are used to support collaboration. Also, data flows are introduced and used to cleanse and transform the data before processing.
Tasks:	Tasks you will complete in this lab exercise include: <ul style="list-style-type: none">• Create and Configure Watson Studio Project• Add Data Flow Asset• Cleanse and Transform Data• Run Data Flow

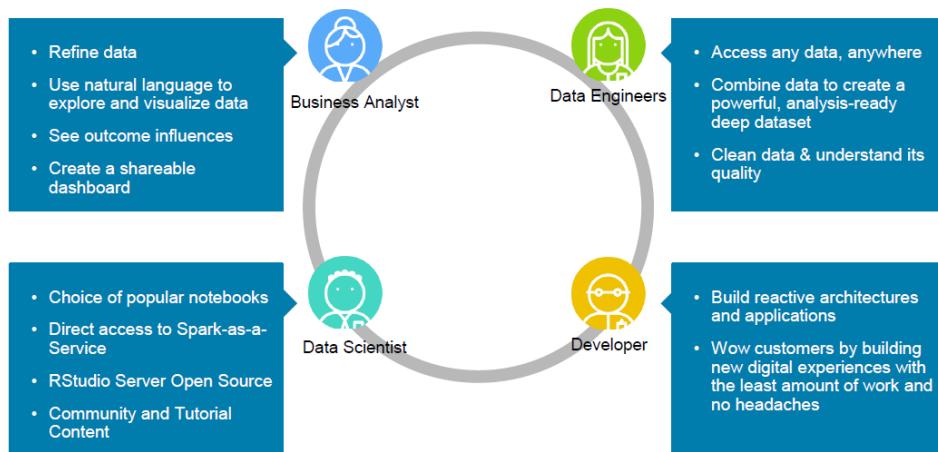
Lesson 2: Workflow Overview



Lesson 2: Instructions

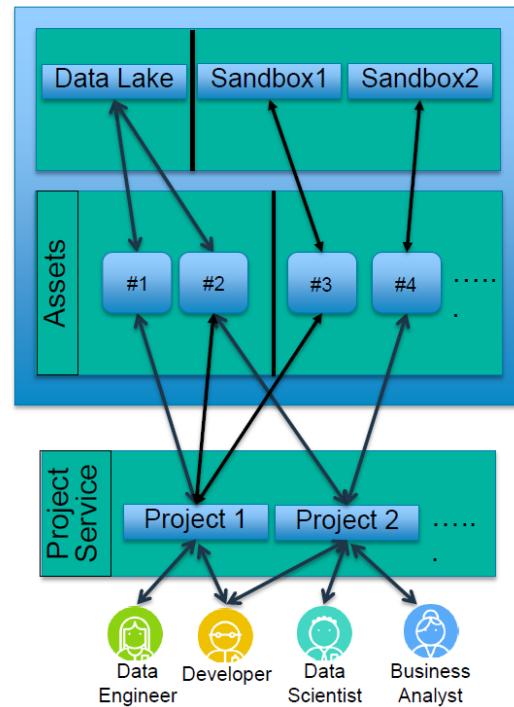
1. Project Overview

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



Projects make collaboration easier by:

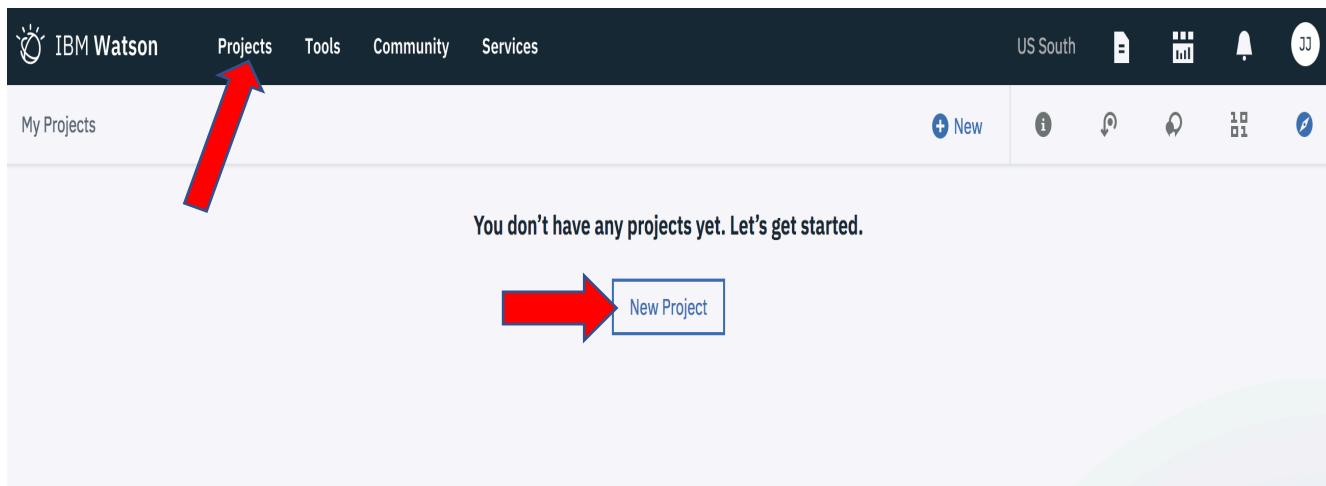
- Allowing different users and personas to share a set of assets
- Enabling users to collaborate and manage their notebooks, artifacts, plus more
- Providing three levels of rights: Viewers, Editors, and Admins



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

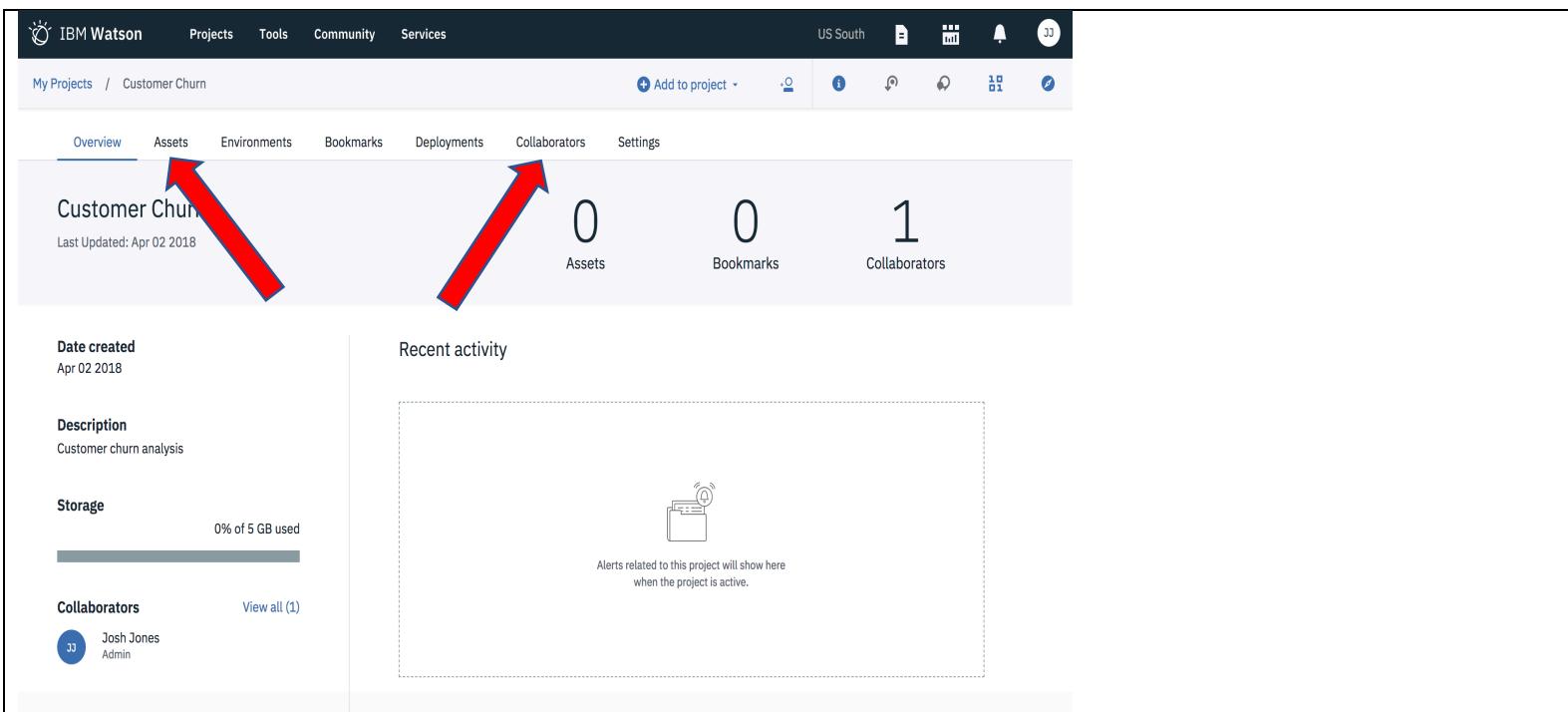
Define Storage:

- 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. At the top, there's a navigation bar with 'IBM Watson', 'Projects', 'Tools', 'Community', 'Services', 'US South', and various icons. Below the navigation is a header with 'My Projects / Customer Churn' and a search bar. The main content area has tabs: 'Overview' (selected), 'Assets', 'Environments', 'Bookmarks', 'Deployments', 'Collaborators', and 'Settings'. The 'Overview' section displays the project name 'Customer Churn', last updated date 'Apr 02 2018', and summary statistics: 0 Assets, 0 Bookmarks, and 1 Collaborator. It also shows 'Recent activity' with a placeholder message about alerts. On the left, there are sections for 'Date created' (Apr 02 2018), 'Description' (Customer churn analysis), 'Storage' (0% of 5 GB used), and 'Collaborators' (Josh Jones, Admin).

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. Load Data from Local File

- In a separate browser navigate to: [Customer Churn Data:](#)
<https://github.com/team-wolfpack/Wolfpack-Technical-Event-Content/tree/master/AI and Data Science/Machine Learning for Customer Churn/Data>
- Download **customer.csv** and **churn.csv** and save both files 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.csv and churn.csv files and select them.

IBM Watson Projects Tools Catalog Community Services Docs Support Manage LM

My Projects / Customer Churn + Add to project

Overview Assets Environments Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Data assets

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data assets					

Load Files Catalog

Drop files here or [browse](#) for files to upload.

- You should now see that the customer and churn data files have been imported into the project under the **Files** and they are listed under **Data Assets**

IBM Watson Projects Tools Catalog Community Services Docs Support Manage LM

My Projects / Customer Churn + Add to project

Overview Assets Environments Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Data assets

0 assets selected.

<input type="checkbox"/> NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
churn.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:44 pm	
customer.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:21 pm	

Find in storage

0 selected
 customer.csv
 churn.csv

4. Create and Run Data Flow

- Under **Assets**, click **New Data Flow**

IBM Watson Projects Tools Community Services Docs Support Manage L1

My Projects / Customer Churn Add to project

Overview Assets Environments Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Modeler flows New flow

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no modeler flows				

Data flows New data flow

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
you currently have no data flows				

- Click **customer.csv** and select **Add** in the lower right corner

IBM Watson Projects Tools Community Services

My Projects / Customer Churn / Data Refinery

Customer Churn

Data assets (2) Connections

churn.csv

customer.csv

Cancel Add

- A preview of customer data should now be displayed, under the **Data** tab. Only the first 1000 rows are displayed.

IBM Watson Projects Tools Catalog Community Services

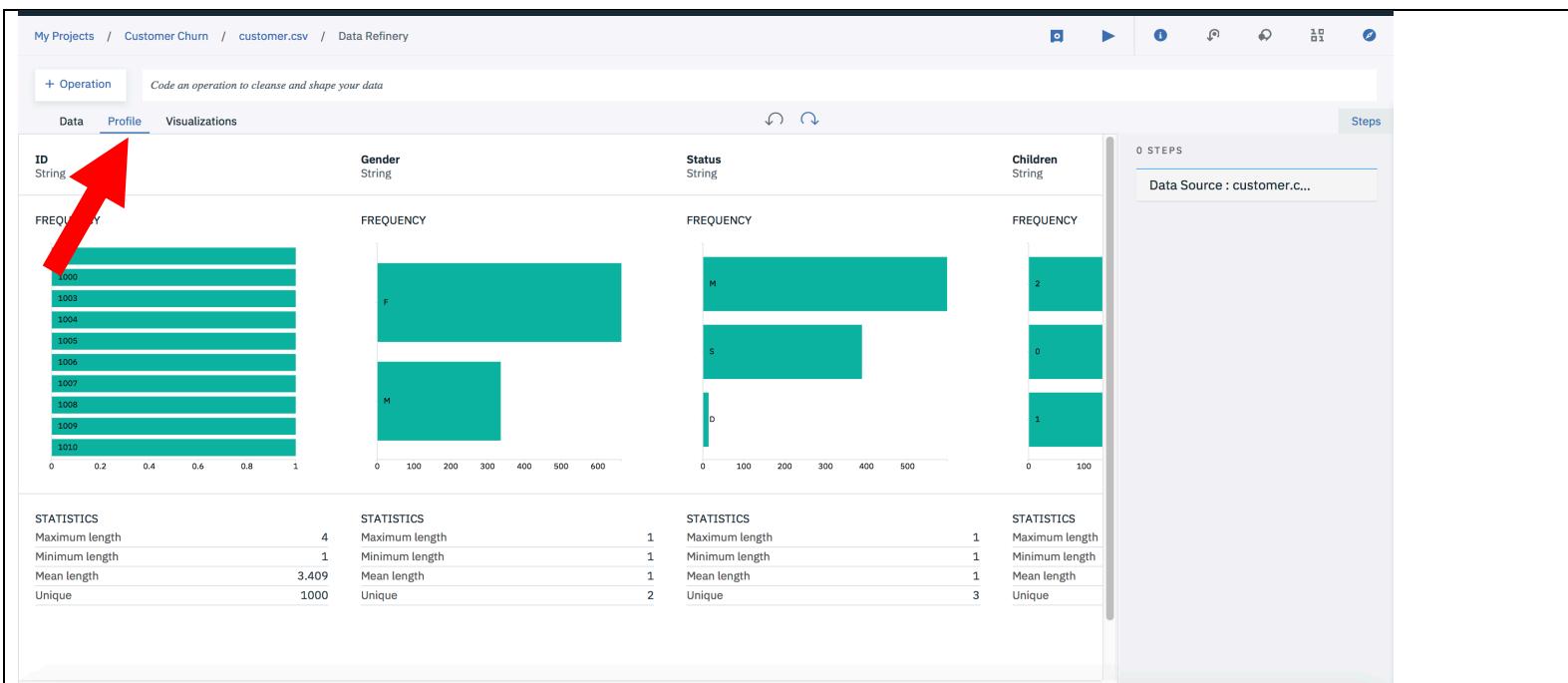
My Projects / Customer Churn / customer.csv / Data Refinery

+ Operation Code an operation to cleanse and shape your data

Data Profile Visualizations

ID	Gender	Marital Status	Children	Est Income	Car Owner	Age	LongDistance
1	F	S	1	38000	N	24.393333	23.56
2	M	M	2	29616	N	49.426667	29.78
3	M	M	0	19732.8	N	50.673333	24.81
4	M	S	2	96.33	N	56.473333	26.13
5	F	M	2	52004.8	N	25.14	5.03
6	M	M	2	53010.8	N	18.84	12.45
7	M	M	1	75004.5	N	64.8	26.52

- Select the **Profile** tab. Information about your data, such as frequency, is displayed



- Return to the **Data tab** and click on the **Age** column. Select the **3 dots** beside the column name and select **Convert Column Type**. Click **Decimal**. Notice a grey circle appeared beside Decimal. Based upon the data values, Watson recommended the data type.

+ Operation

Code an operation to cleanse and shape your data

Data

Profile

Visualizations

Steps

	Est Income String	Car Owner String	Age String	LongDistance String		International String	Local String	Dropped String	Paymeth String
1	38000	N	24.393333		Remove				
2	29616	N	49.426667		Remove duplicates				
3	19732.8	N	50.673333		Remove empty rows				
4	96.33	N	56.473333		Sort ascending				
5	52004.8	N	25.14		Sort descending				
6	53010.8	N	18.84		Substitute				
7	75004.5	N	64.8						
8	19749.3	N	60.366667		CONVERT C...				
9	57626.9	Y	43.906667						
10	20078	N	32.846667	TEXT					
11	47902	N	26.033333		View All				
12	7545.96	Y	16.753333	22.39		Date			
13	78851.3	N	48.373333	0.37		Decimal			
14	17540.7	Y	62.786667	22.17		Integer			
15	83891.9	Y	61.02	28.92		String			
16	28220.8	N	38.766667	26.49		Timestamp			
17	28589.1	N	15.6	13.19					
18	5237.63	N	48.753333	13.32					
19	89459.9	N	53.28	11.54					

- The **Age** column is now of data type **Decimal** and our conversion action now appears on the left-hand side under **Steps**. Each transformation performed on the data is tracked under Steps

Children String	Est Income String	Car Owner String	Age Decimal	LongDistance String	Steps
1	38000	N	24.393333	23.56	
2	29616	N	49.426667	29.78	
0	19732.8	N	50.673333	24.81	
2	96.33	N	56.473333	26.13	
2	52004.8	N	25.14	5.03	1 STEPS
2	53010.8	N	18.84	12.45	Data Source : customer.c...
1	75004.5	N	64.8	26.52	Convert column type JUST ADDED
0	19749.3	N	60.366667	20.22	Converted Age from String to Decimal
1	57626.9	Y	43.906667	9.38	

- Next, click the **Martial Status** column. Hover over the column name until an edit pencil icon appears. Change the column name to **Status**

+ Operation

Code an operation to cleanse and shape your data

Data Profile Visualizations

	ID String	Gender String	Marital Status String	⋮
1	1	F	S	
2	6	M	M	
3	8	M	M	
4	11	M	S	
5	14	F	M	
6	17	M	M	
7	18	M	M	
8	21	M	M	
9	22	M	S	

	ID String	Gender String	Status String
1	1	F	S
2	6	M	M
3	8	M	M
4	11	M	S
5	14	F	M
6	17	M	M
7	18	M	M
8	21	M	M

- Click **Operation** in the upper left corner.

IBM Watson

Projects Tools Catalog Community Services Docs Support Manage

My Projects / Customer Churn / customer.csv / Data Refinery

x Operation Code an operation to cleanse and shape your data

FREQUENTLY USED

- Calculate
- Convert column type
- Filter
- Math
- Remove
- Rename
- Sort ascending
- Sort descending

Steps

2 STEPS

Data Source : customer.c...
Convert column type
Converted Age from String to Decimal
Rename column
JUST ADDED
Renamed column Marital Status to Status

ID String	Gender String	Status String	Children String	Est Income String	Car Owner String
1	F	S	1	38000	N
6	M	M	2	29616	N
8	M	M	0	19732.8	N
11	M	S	2	96.33	N
14	F	M	2	52004.8	N
17	M	M	2	53010.8	N
18	M	M	1	75004.5	N
21	M	M	0	19749.3	N
22	M	S	1	57626.9	Y

- Under the **Organize** section, click **Join**.

x Operation Code an operation to cleanse and shape your data

Search operations 

ID	Gender	Status
String	String	String
1	F	S
6	M	M
8	M	M
11	M	S
14	F	M
17	M	M
18	M	M
21	M	M
22	M	S
23	M	M
24	F	M
29	M	M
35	F	S
36	F	S
37	F	M
38	F	M
40	F	S
42	F	M
45	M	S

Filter

Math

Remove

Rename

Sort ascending

Sort descending

Substitute

Text

CLEANSE 

Convert column value to missing

Remove duplicates

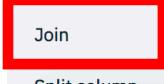
Remove empty rows

Replace missing values

Replace substring

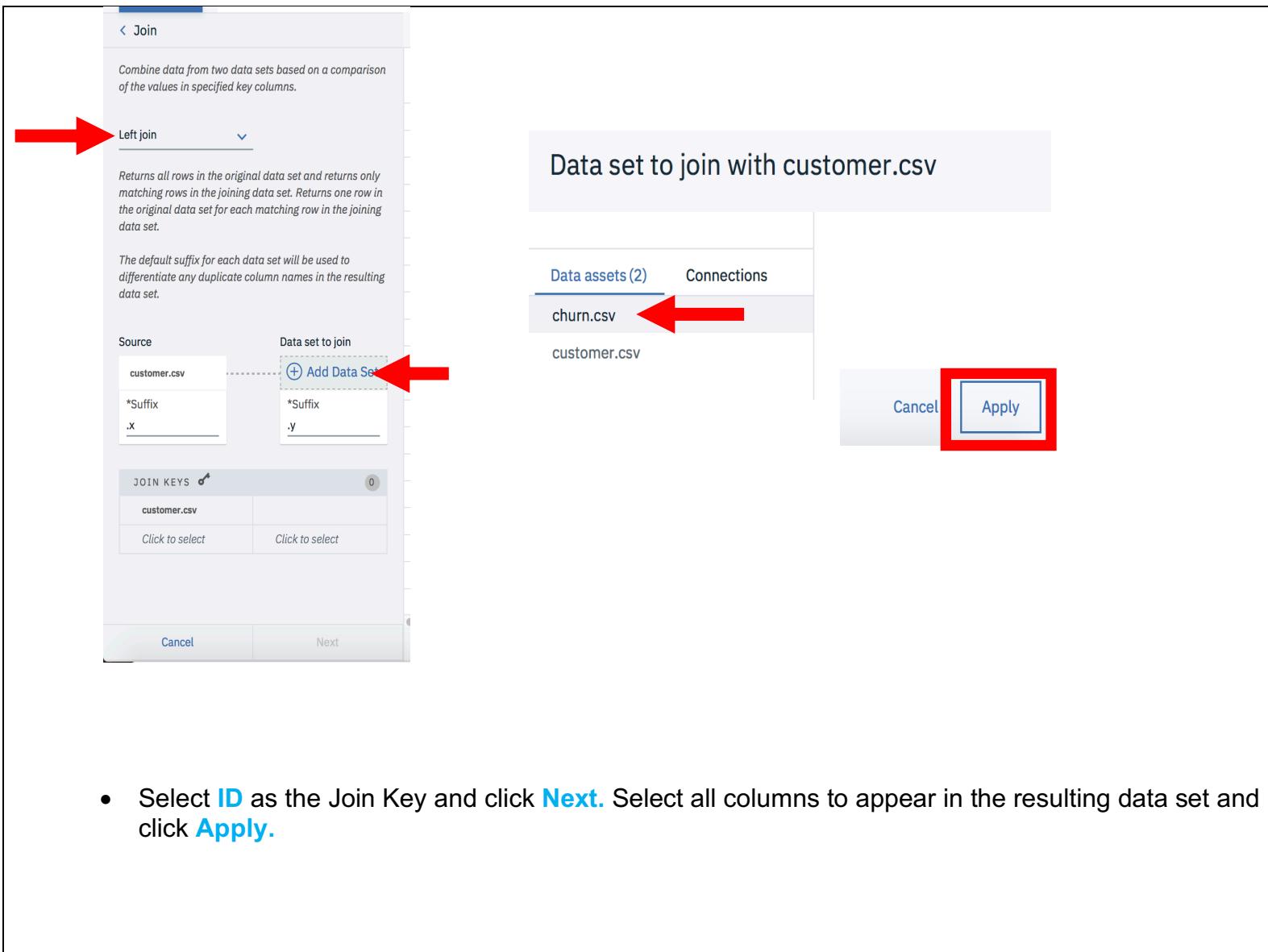
ORGANIZE 

Concatenate

Join 

Split column

- Select **Left Join** and click **Add Data Set**. Select **Churn.csv**, as the data set to join, and click **Apply**



The screenshot shows the 'Join' step configuration in the Watson Studio Data Catalog. A red arrow points to the 'Left join' dropdown menu. Another red arrow points to the 'Add Data Set' button in the 'Data set to join' section. A third red arrow points to the 'churn.csv' asset in the 'Data assets (2)' list. A fourth red box highlights the 'Apply' button at the bottom right of the dialog.

Join

Combine data from two data sets based on a comparison of the values in specified key columns.

Left join

Returns all rows in the original data set and returns only matching rows in the joining data set. Returns one row in the original data set for each matching row in the joining data set.

The default suffix for each data set will be used to differentiate any duplicate column names in the resulting data set.

Source Data set to join

customer.csv + Add Data Set

*Suffix .x .y

JOIN KEYS 0

customer.csv Click to select Click to select

Data assets (2) Connections

churn.csv customer.csv

Cancel Apply

Data set to join with customer.csv

- Select **ID** as the Join Key and click **Next**. Select all columns to appear in the resulting data set and click **Apply**.

Join

Combine data from two data sets based on a comparison of the values in specified key columns.

Left join

Returns all rows in the original data set and returns only matching rows in the joining data set. Returns one row in the original data set for each matching row in the joining data set.

The default suffix for each data set will be used to differentiate any duplicate column names in the resulting data set.

Source Data set to join

customer.csv	churn.csv
*Suffix .x	*Suffix .y

JOIN KEYS  1

customer.csv	churn.csv
ID	 ID

 Add Join Key

Cancel **Next**

Join

Select the columns in the resulting data set

- Clear all selections
- ID
- Gender
- Status
- Children
- Est Income
- Car Owner
- Age
- LongDistance
- International
- Local
- Dropped
- Paymethod
- LocalBilltype
- LongDistanceBilltype

Back

Apply

- The **Churn** column is now joined to the customer data set. And the **join operation** step is listed.

My Projects / Customer Churn / customer.csv / Data Refinery

+ Operation Code an operation to cleanse and shape your data

Data Profile Visualizations

	ocal ring	Dropped String	Paymethod	LocalBiltype	LongDistanceBi...	Usage	RatePlan	CHURN
1	16.08	0	CC	Budget	Intl_discount	229.64	3	T
2	3.5	0	CH	FreeLocal	Standard	75.29	2	F
3	3.44	0	CC	FreeLocal	Standard	47.25	3	F
4	3.88	1	CC	Budget	Standard	59.01	1	F
5	3.11	0	CH	Budget	Intl_discount	28.14	1	F
6	3.42	4	CC	FreeLocal	Standard	58.87	1	F
7	3.19	0	CC	Budget	Intl_discount	58.72	1	F
8	3.94	0	CC	Budget	Standard	34.17	3	F
9	3.96	0	CC	Budget	Standard	48.35	2	F
10	33	0	CC	Budget	Intl_discount	15.98	4	F
11	3.92	1	Auto	FreeLocal	Standard	72.31	2	F
12	3.86	0	CC	Budget	Standard	200.75	3	T
13	3.66	0	CC	FreeLocal	Standard	29.04	4	T

Steps

3 STEPS

- Data Source : customer.c...
- Rename column
- Renamed column Marital Status to Status
- Convert column type
- Converted Age from String to Decimal
- Join JUST ADDED
- left-joined data from churn.csv based on columns ID,ID

- Click the **Save icon** in the upper right to save the data flow

My Projects / Customer Churn / customer.csv / Data Refinery

+ Operation Code an operation to cleanse and shape your data

Details Help

- Before we run the data flow, lets visualize the data. Select the **Visualizations** tab in the upper left

My Projects / Customer Churn / customer.csv_flow / Data Refinery

+ Operation Code an operation to cleanse and shape your data

Data Profile Visualizations

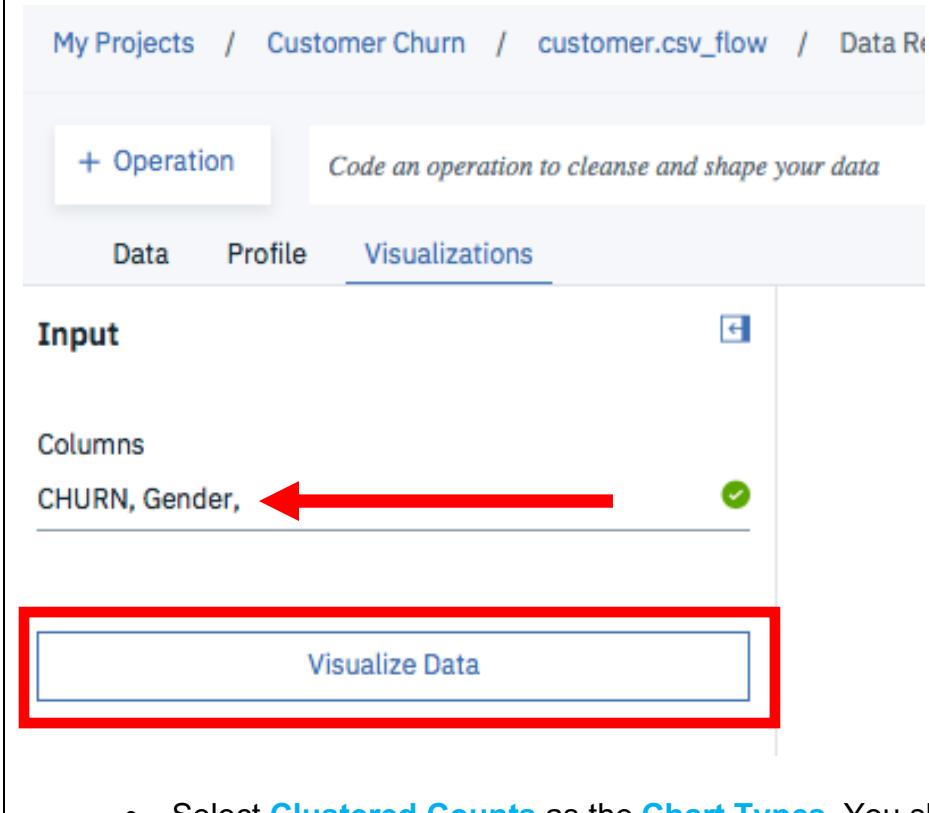
Input

Columns

Use commas to separate columns

Visualize Data

- Select **Churn** and **Gender** as the columns. Click **Visualize Data**



My Projects / Customer Churn / customer.csv_flow / Data Re

+ Operation *Code an operation to cleanse and shape your data*

Data Profile Visualizations

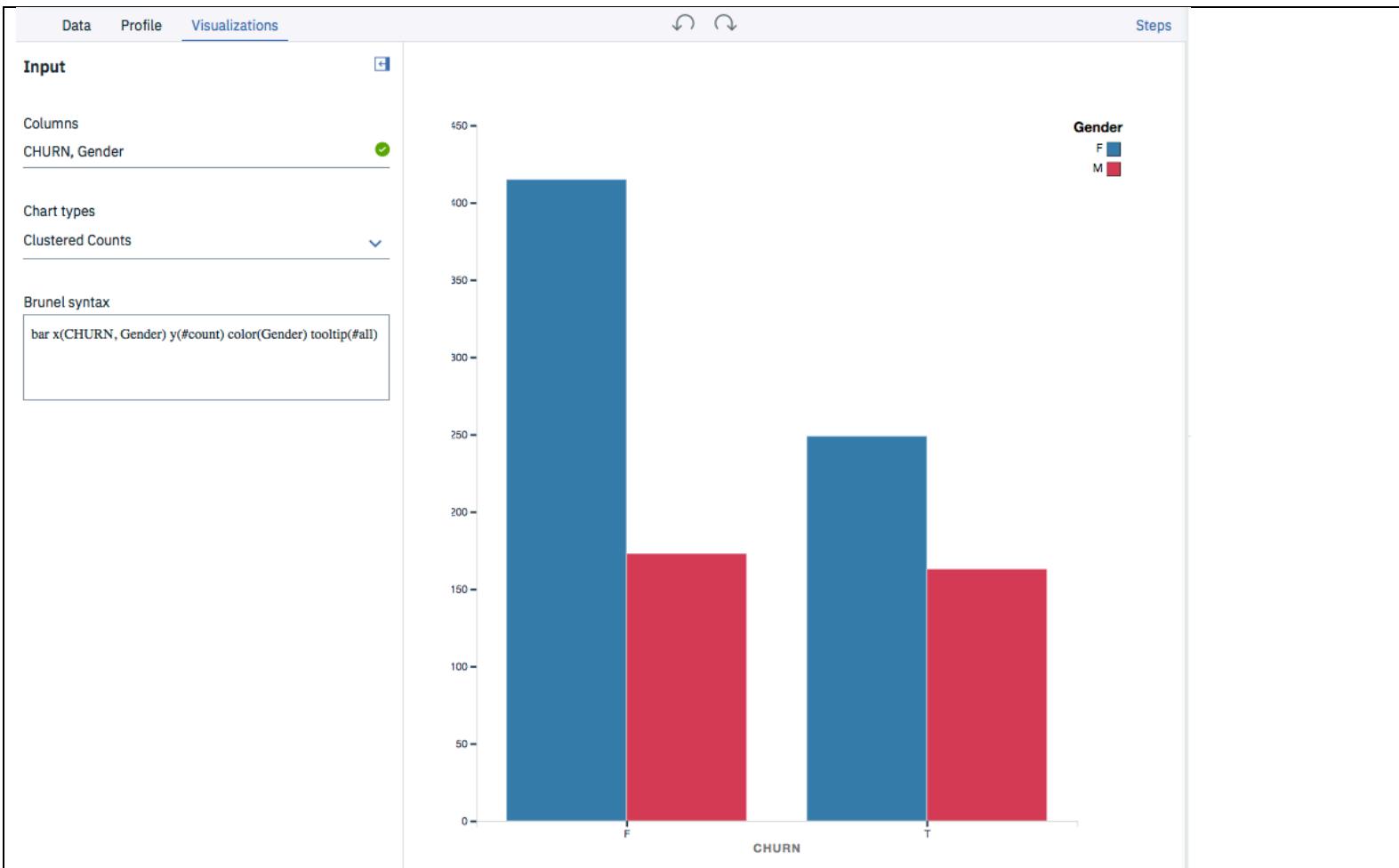
Input

Columns

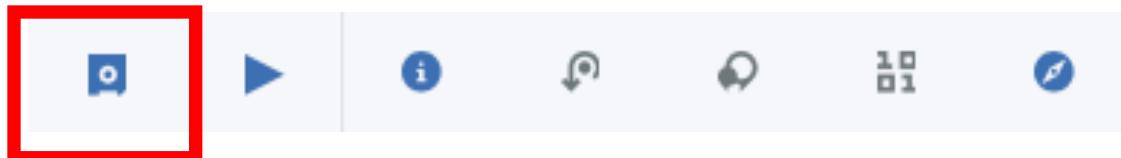
CHURN, Gender, 

Visualize Data

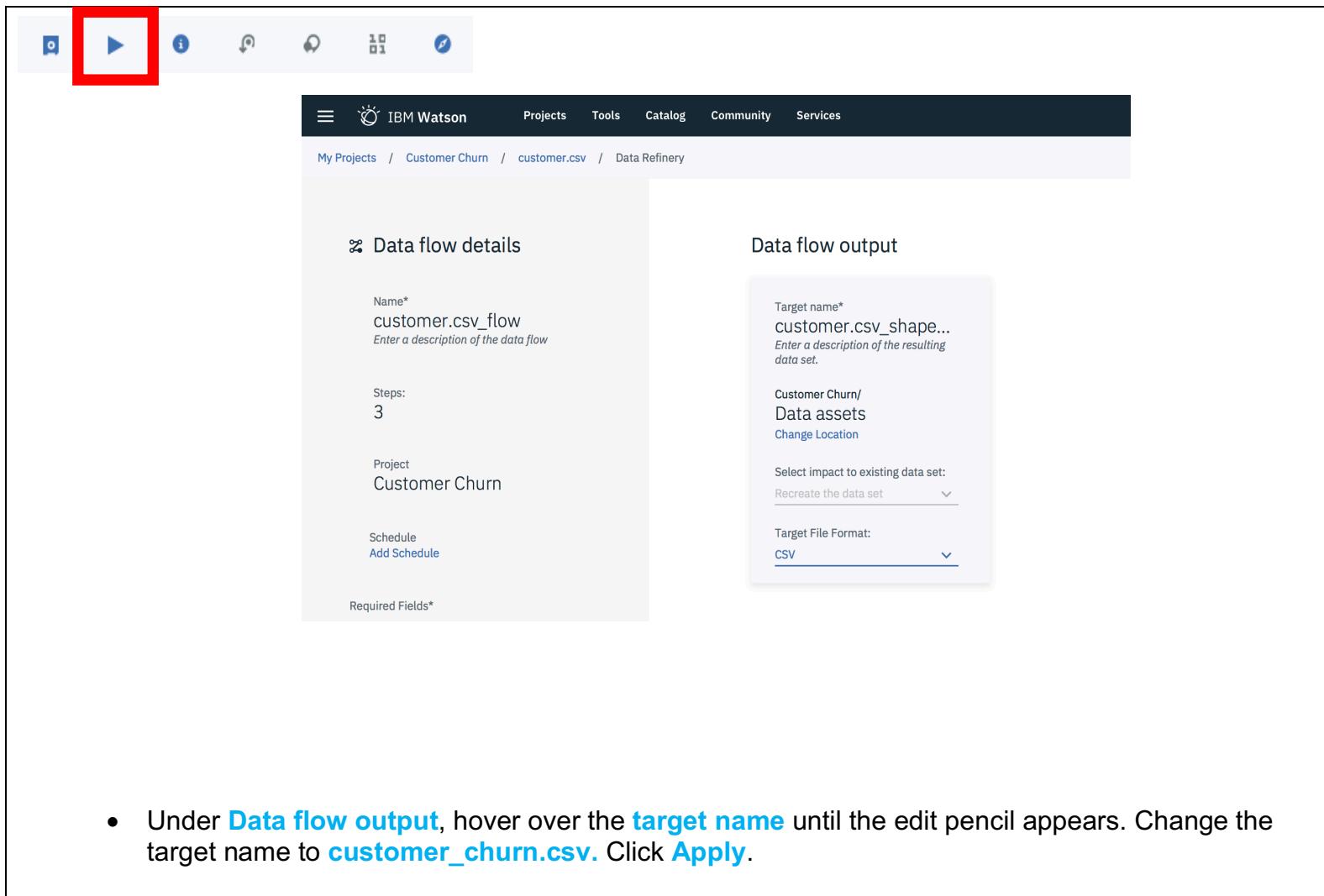
- Select **Clustered Counts** as the **Chart Types**. You should see the following. Notice the Brunel Syntax for the visualization is provided.



- Click the **Save icon** in the upper right to save the data flow



- Click the **Run icon** in the upper right to run to the data flow. You should see the following.



The screenshot shows the IBM Watson Data Refinery interface. At the top, there is a navigation bar with icons for cloud, play, info, refresh, and search. Below the navigation bar is a dark header bar with the IBM Watson logo and links for Projects, Tools, Catalog, Community, and Services. The main content area has a breadcrumb trail: My Projects / Customer Churn / customer.csv / Data Refinery. On the left, under 'Data flow details', there are fields for Name* (customer.csv_flow), Steps: 3, Project (Customer Churn), and Schedule (Add Schedule). On the right, under 'Data flow output', there are fields for Target name* (customer.csv_shape...), Select impact to existing data set (Recreate the data set), and Target File Format (CSV). A note at the bottom says 'Required Fields*'.

- Under **Data flow output**, hover over the **target name** until the edit pencil appears. Change the target name to **customer_churn.csv**. Click **Apply**.

Data flow output

Target name*
Customer.... 

Enter a description of the resulting data set.

Customer Churn/ Data assets

Change Location

Select impact to existing data set:
 Recreate the data set

Target File Format:
CSV

Data flow output

Target name*
customer_churn.csv| 

Enter a description of the resulting data set.

82 300

Apply 

Customer Churn/ Data assets

Change Location

Select impact to existing data set:
 Recreate the data set

Target File Format:
CSV

- Keep location as **Data assets**. You should now have the following. Click **Save and Run** in the lower right corner

Data flow details

Name*
Customer.csv_flow

Enter a description of the data flow

Steps:
 3

Project
 Customer Churn

Schedule
[Add Schedule](#)

Required Fields*

Data flow output

Target name*
customer_churn.csv

Enter a description of the resulting data set.

Customer Churn/ Data assets

Change Location

Select impact to existing data set:
 Recreate the data set

Target File Format:
CSV

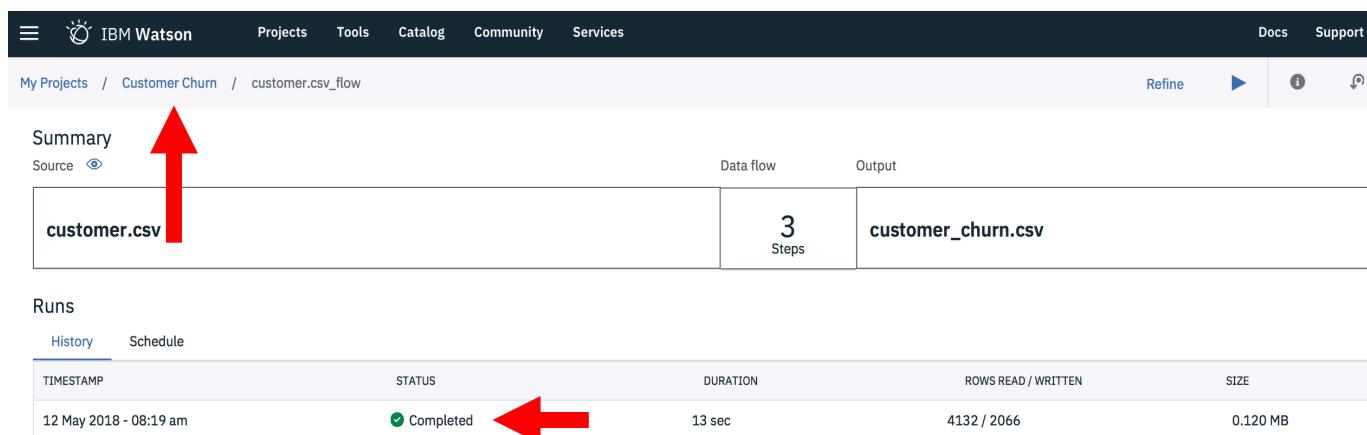
- Select **View Flow**, as the next step, to be redirected to the **Summary** page.

What's next?

Your data flow is currently running. You can view its progress on the Summary and Runs page. When the flow completes, you can view its output from there too.

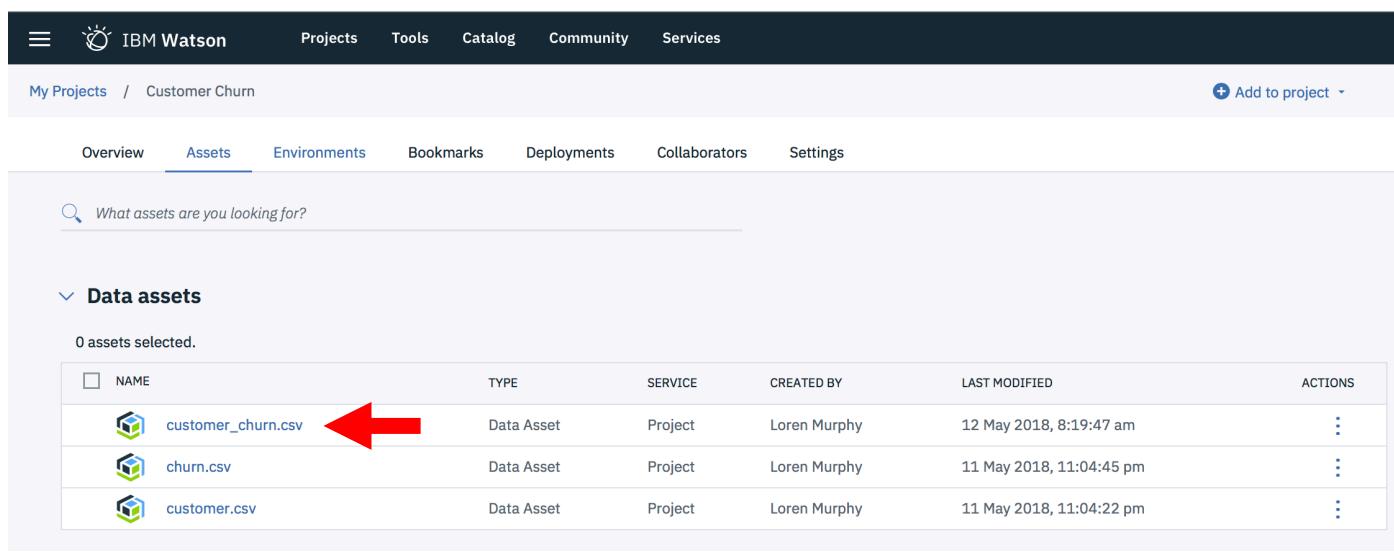


- Within a few seconds, the flow's status will indicate **Completed**. Select **Customer Churn** in the upper left to be taken back to your project workspace.



The screenshot shows the IBM Watson Project workspace. At the top, there's a navigation bar with links for IBM Watson, Projects, Tools, Catalog, Community, Services, Docs, and Support. Below the navigation bar, the URL is shown as My Projects / Customer Churn / customer.csv_flow. On the left, there's a sidebar with a 'Summary' section containing a 'Source' link and a preview of the 'customer.csv' file. A red arrow points upwards from the 'customer.csv' preview towards the 'Summary' section. To the right of the source preview, there's a 'Data flow' section showing 3 Steps and an 'Output' section labeled 'customer_churn.csv'. Below this, the 'Runs' section shows a table of completed runs. One run is listed: '12 May 2018 - 08:19 am' with a status of 'Completed' (indicated by a green checkmark) and a duration of '13 sec'. A red arrow points to the 'Completed' status. The table also includes columns for 'TIMESTAMP', 'STATUS', 'DURATION', 'ROWS READ / WRITTEN', and 'SIZE'.

- The **customer_churn.csv** file will now appear under **Data Assets**



The screenshot shows the IBM Watson Project workspace with the 'Assets' tab selected in the navigation bar. A red arrow points down to the 'Data assets' section. This section shows a table of data assets. The first asset listed is 'customer_churn.csv', which is identified by a red arrow. The table has columns for 'NAME', 'TYPE', 'SERVICE', 'CREATED BY', 'LAST MODIFIED', and 'ACTIONS'. Other assets listed are 'churn.csv' and 'customer.csv'.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
customer_churn.csv	Data Asset	Project	Loren Murphy	12 May 2018, 8:19:47 am	⋮
churn.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:45 pm	⋮
customer.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:22 pm	⋮



Lesson 3: Jupyter Notebook

Purpose:	This lesson introduces Jupyter notebooks and are used as part of a customer churn analysis using R.
Tasks:	<p>Tasks you will complete in this lab exercise include:</p> <ul style="list-style-type: none">• Add Notebook Asset• Retrieve Data from External Repository• Predict Customer Churn using Machine Learning Techniques• Evaluate Model Performance

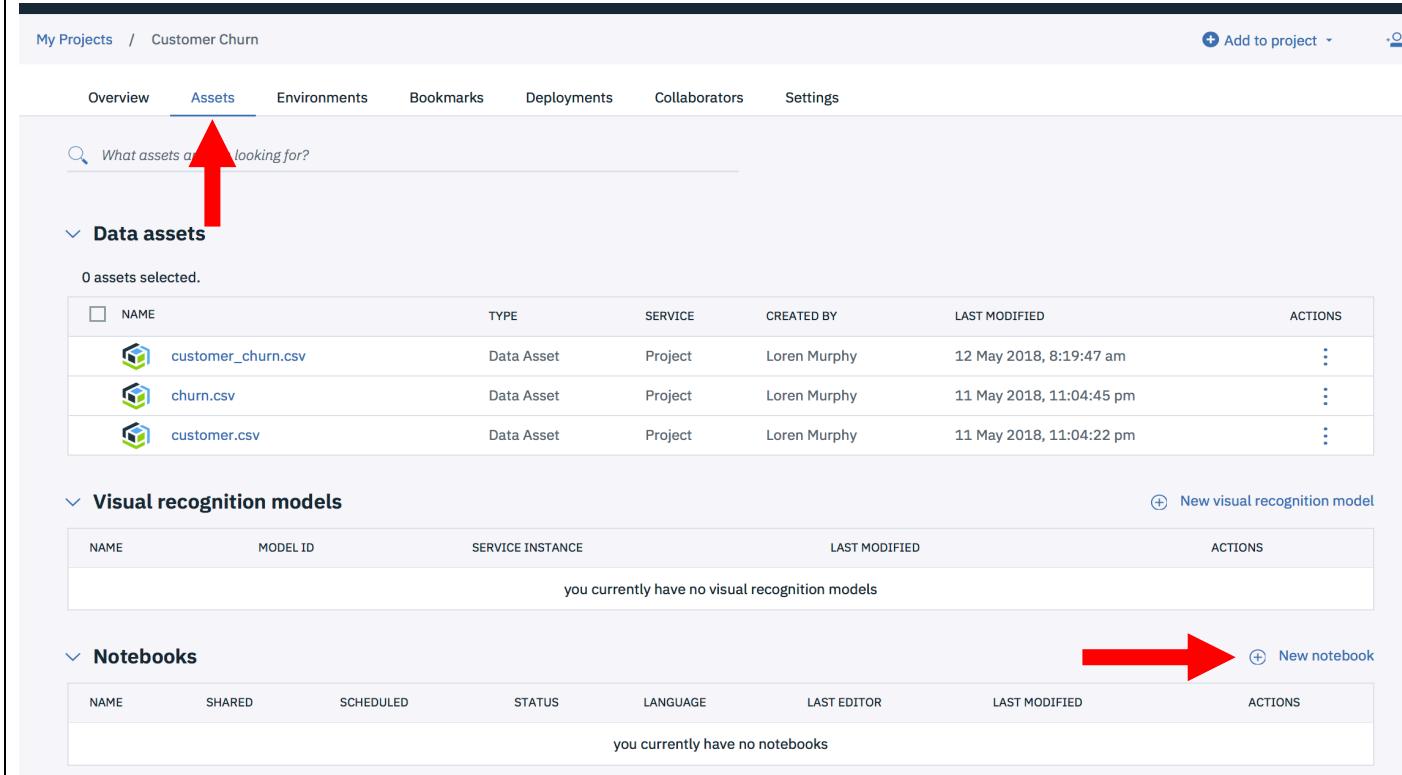
Lesson 3: Workflow Overview

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

Lesson 3: Instructions

1. Create Notebook

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



My Projects / Customer Churn + Add to project ⋮

Overview Assets Environments Bookmarks Deployments Collaborators Settings

What assets are you looking for?

Data assets

0 assets selected.

NAME	TYPE	SERVICE	CREATED BY	LAST MODIFIED	ACTIONS
customer_churn.csv	Data Asset	Project	Loren Murphy	12 May 2018, 8:19:47 am	⋮
churn.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:45 pm	⋮
customer.csv	Data Asset	Project	Loren Murphy	11 May 2018, 11:04:22 pm	⋮

Visual recognition models

+ New visual recognition model

NAME	MODEL ID	SERVICE INSTANCE	LAST MODIFIED	ACTIONS
you currently have no visual recognition models				

Notebooks

+ New notebook

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks							

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

New notebook

Blank From file From URL

Name*

Customer Churn - R

32 Characters Remaining

Description

Customer churn analysis using R.

468 Characters Remaining

- In a separate browser window navigate to:

[Machine Learning for Customer Churn](#)

[https://github.com/team-wolfpack/Wolfpack-Technical-Event-Content/tree/master/AI and Data Science/ Machine Learning for Customer Churn](https://github.com/team-wolfpack/Wolfpack-Technical-Event-Content/tree/master/AI%20and%20Data%20Science%2F%20Machine%20Learning%20for%20Customer%20Churn)

- 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 [\(i\)](#)

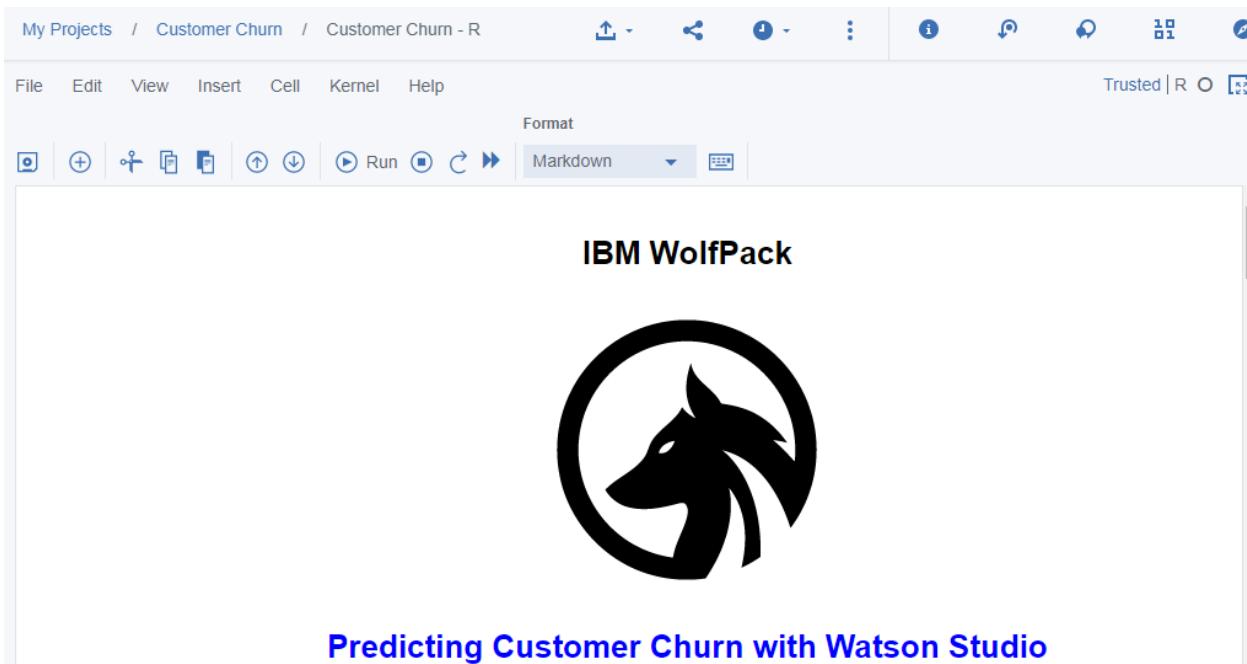
Default R Environment XS (2 vCPU and 8 GB RAM)



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:



The screenshot shows a Jupyter Notebook interface with the following details:

- Header:** My Projects / Customer Churn / Customer Churn - R
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Help, Trusted | R O
- Format Bar:** Markdown
- Content:**
 - Section Header:** IBM WolfPack
 - Image:** A black silhouette of a wolf's head inside a circle.
 - Text:** Predicting Customer Churn with Watson Studio

- Scroll down in the notebook until **Step 3: Read Customer_Churn Data into R Dataframe / Simple Analysis**. Click on the notebook cell.

Step 3: Read Customer_Churn Data into R Dataframe / Simple Analysis

```
# Insert customer_churn.csv file into code HERE!!!
```

- Click on the **Data Icon** in the upper right corner. Under the **Files** tab, click the **Insert to code** dropdown for **customer_churn.csv** file. Select **Insert R DataFrame**.

IBM Watson Projects Tools Catalog Community Services Docs Support Manage LM

My Projects / Customer Churn / Customer Churn

File Edit View Insert Cell Kernel Help Not Trusted | R O

set.seed(3842)

Step 3: Read Customer_Churn Data into R Dataframe / Simple Analysis

In [2]: # Insert customer_churn.csv file into code HERE!!!

In [16]: # Primary data set row count
cat(sprintf("[custDataRaw] has %d rows:\n", nrow(custDataRaw)))
[custDataRaw] has 2066 rows:

In [17]: # Summary Stats for entire data set
summary(custDataRaw)

Files Connections

Drop your file here or browse your files to add a new file

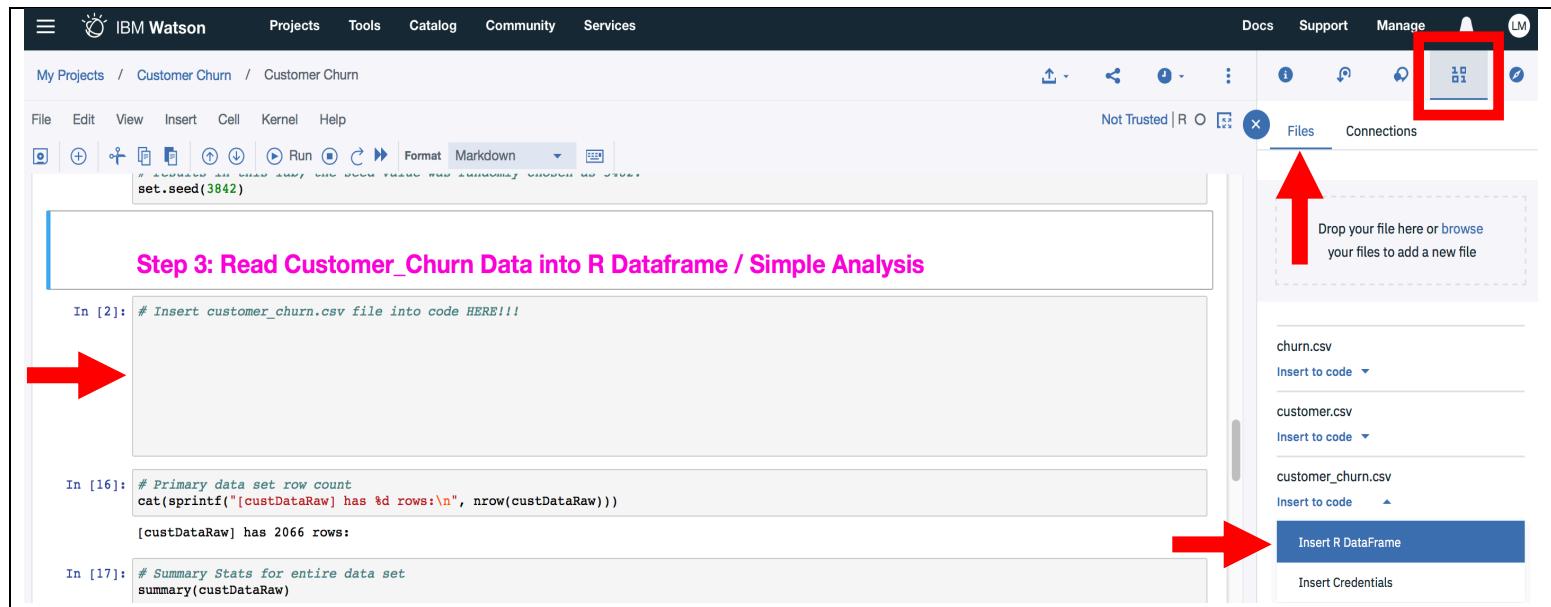
churn.csv
Insert to code ▾

customer.csv
Insert to code ▾

customer_churn.csv
Insert to code ▾

Insert R DataFrame

Insert Credentials



- The R DataFrame code is automatically generated and should now appear within the notebook cell

Step 3: Read Customer_Churn Data into R Dataframe / Simple Analysis

```
# Insert customer_churn.csv file into code HERE!!!
library("aws.s3")

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share your notebook.
Sys.setenv("AWS_ACCESS_KEY_ID" = "e1428a9a59da4701a9a6dba582ec95a0", "AWS_SECRET_ACCESS_KEY" = "d0594edbaba71cafccff929835fafa08a1030bdc1f457a4ab")
url <- "s3-api.us-geo.objectstorage.service.networklayer.com"
bucket <- "customerchurn-donotdelete-pr-sh12y1mzd0umqm"
headers <- list(`x-amz-content-sha256`="e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855")

obj <- s3HTTP(
  verb = "GET",
  bucket = bucket,
  headers = headers,
  path = "data_asset/customer_churn_N6zh3vK5T46YBzOyFOx5lw.csv",
  key = Sys.getenv("AWS_ACCESS_KEY_ID"),
  secret = Sys.getenv("AWS_SECRET_ACCESS_KEY"),
  check_region = FALSE,
  base_url = url)

df.data.1 <- read.csv(text = rawToChar(obj$content))
head(df.data.1)
```



- Change the name of the R DataFrame from df.data.1 to **custDataRaw**. It should look like the following

```
custDataRaw <- read.csv(text = rawToChar(obj$content))
head(custDataRaw)
```



- Your final notebook cell should look like the following. The Object Storage connection information will be different. Run the cell to view the first few rows of data.

Step 3: Read Customer_Churn Data into R Dataframe / Simple Analysis

```
# Insert customer_churn.csv file into code HERE!!
library("aws.s3")

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share your notebook.
Sys.setenv("AWS_ACCESS_KEY_ID" = "e1428a9a59da4701a9a6dba582ec95a0", "AWS_SECRET_ACCESS_KEY" = "d0594edbaba71cafccf929835fafaf08a1030bdc1f457a4ab")
url <- "s3-api.us-geo.objectstorage.service.networklayer.com"
bucket <- "customerchurn-donotdelete-pr-sbl2ylmzd0umqm"
headers <- list(`x-amz-content-sha256`="e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855")

obj <- s3HTTP(
  verb = "GET",
  bucket = bucket,
  headers = headers,
  path = "data_asset/customer_churn_N6zH3vK5T46YBzOyFOx5lw.csv",
  key = Sys.getenv("AWS_ACCESS_KEY_ID"),
  secret = Sys.getenv("AWS_SECRET_ACCESS_KEY"),
  check_region = FALSE,
  base_url = url)

custDataRaw <- read.csv(text = rawToChar(obj$content))
head(custDataRaw)
```

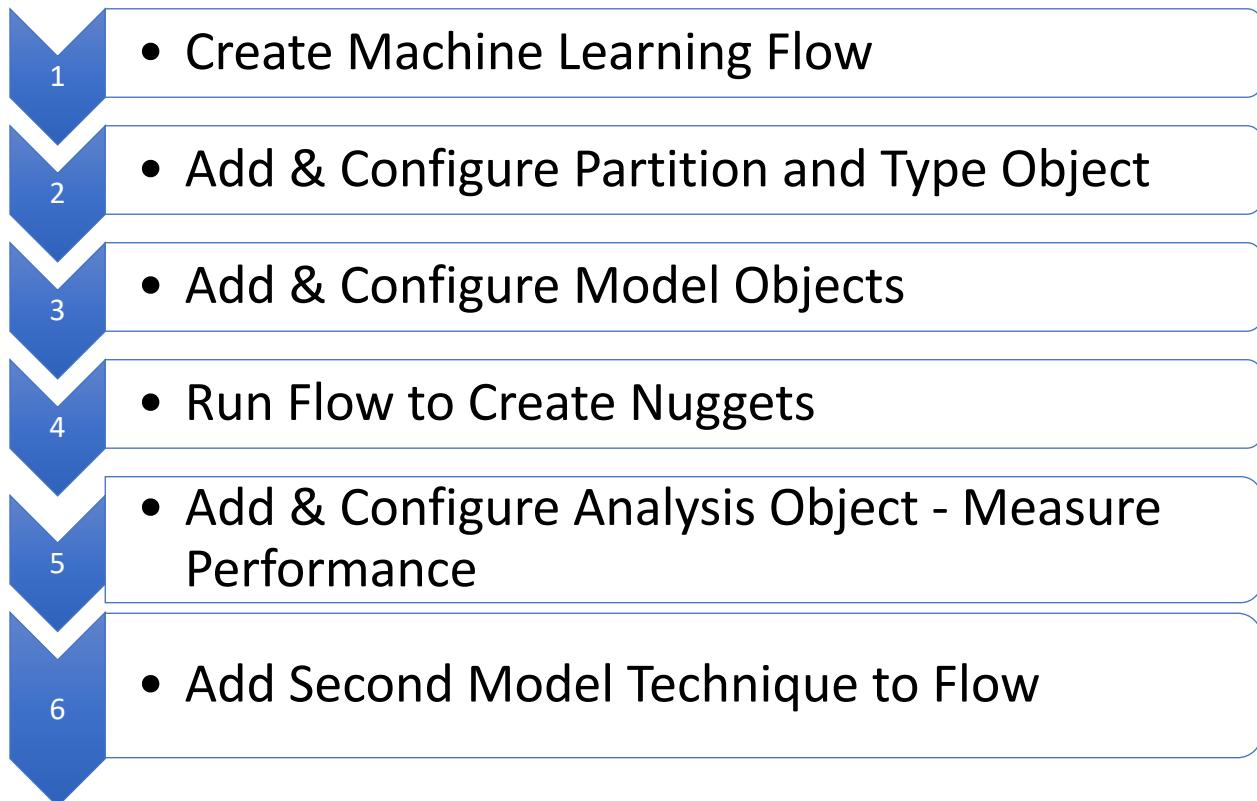
ID	Gender	Status	Children	Est.Income	Car.Owner	Age	LongDistance	International	Local	Dropped	Payoutmethod	LocalBilltype	LongDistanceBilltype	Usage	RatePlan	CHURN
1	F	S	1	38000.00	N	24.39333	23.56	0	206.08	0	CC	Budget	Intl_discount	229.64	3	TRUE
1003	F	M	0	55221.00	N	43.00000	0.49	0	15.20	0	CC	FreeLocal	Standard	15.69	1	FALSE
1004	F	M	1	8073.11	N	46.00000	28.70	0	60.35	0	Auto	FreeLocal	Standard	89.05	4	FALSE
1018	F	S	0	95786.80	Y	52.64667	21.33	0	87.46	0	CC	Budget	Standard	108.79	1	FALSE
1020	M	S	0	90321.60	N	55.11333	0.87	0	11.52	0	CC	FreeLocal	Intl_discount	12.39	3	TRUE
1030	M	M	2	29616.00	N	49.42667	29.78	0	45.50	0	CH	FreeLocal	Standard	75.29	4	FALSE

Lesson 3 Continued in [Customer Churn – R] Notebook

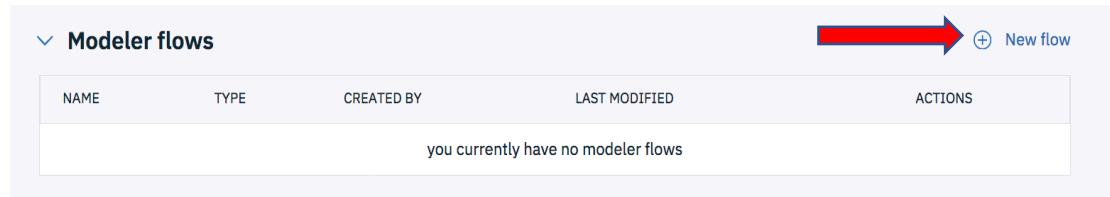
Lesson 4: Machine Learning Flows

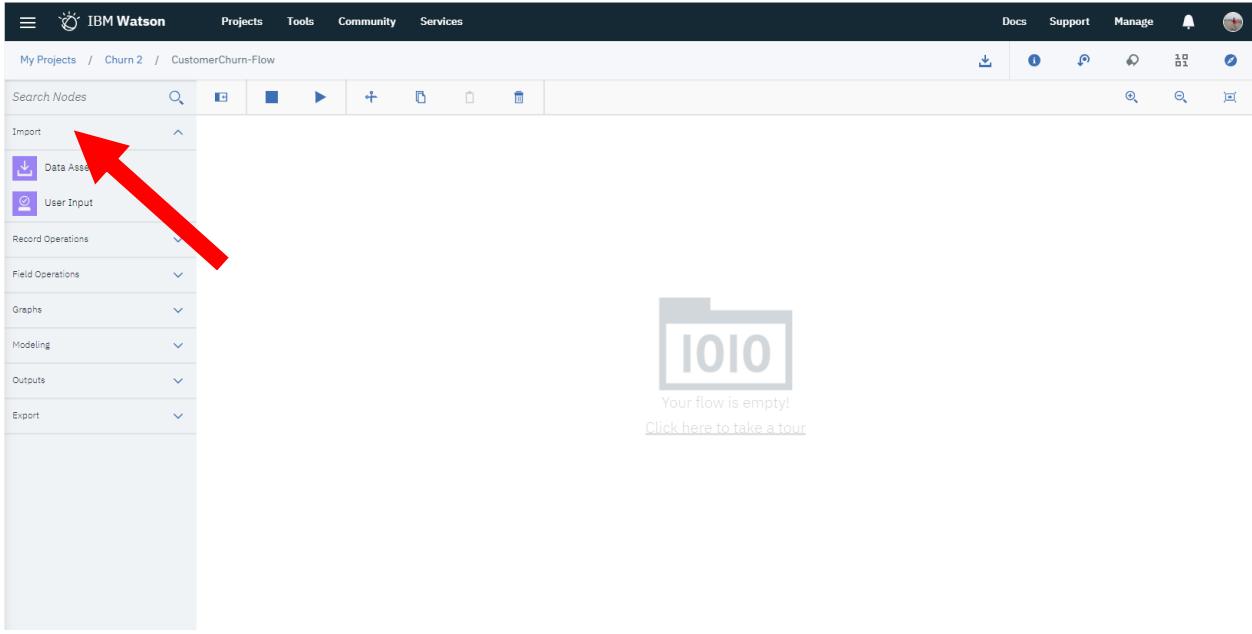
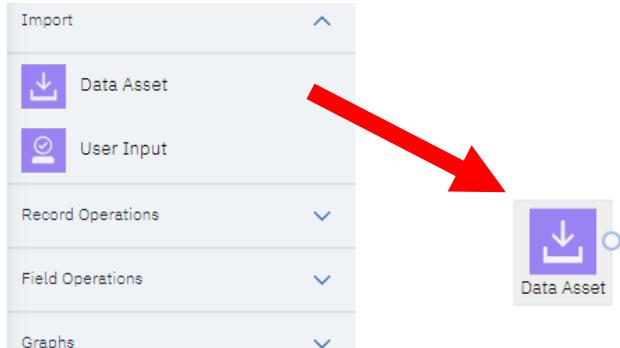
Purpose:	This lesson introduces Machine Learning Flows in Watson Studio. Flows provide a graphical approach to machine learning like that of SPSS Modeler.
Tasks:	Tasks you will complete in this lab exercise include: <ul style="list-style-type: none">• Create Machine Learning Flow• Import Data• Leverage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

Lesson 4: Workflow Overview



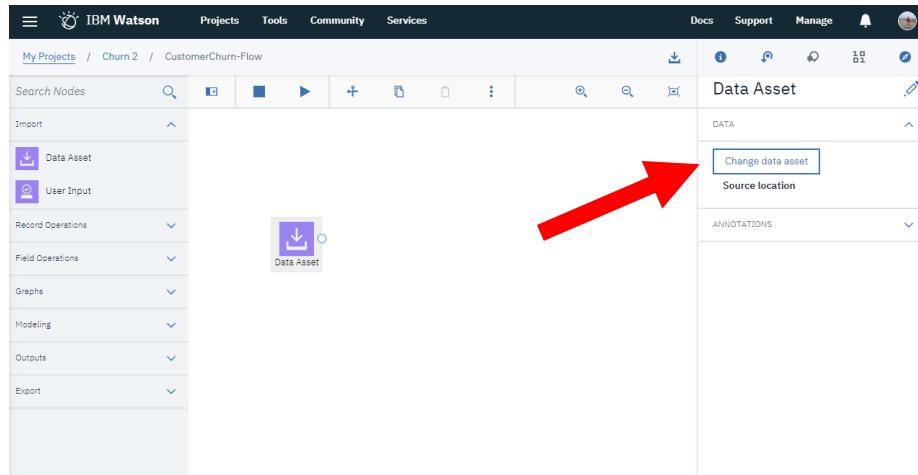
Lesson 4: Instructions

Action
1. Create Machine Learning Flow
<ul style="list-style-type: none"> • Navigate to Customer Churn project page • Click on “New flow” 
<ul style="list-style-type: none"> • 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”: <div style="border: 1px solid #ccc; padding: 10px;"> <p>Modeler</p> </div> <div style="border: 1px solid #ccc; padding: 10px; margin-top: 10px;"> <p>New From file From example</p> <p>Name* CustomerChurn-Flow</p> <p>Description Watson Studio flow for customer church </p> <p>Select flow type <input checked="" type="radio"/> Modeler Flow <input type="radio"/> Neural Network Modeler <small>BETA</small> </p> <p>Runtime <input checked="" type="radio"/> IBM SPSS Modeler <input type="radio"/> Scala Spark 2.0 <small>BETA</small> </p> </div> <ul style="list-style-type: none"> • Click on “Create”

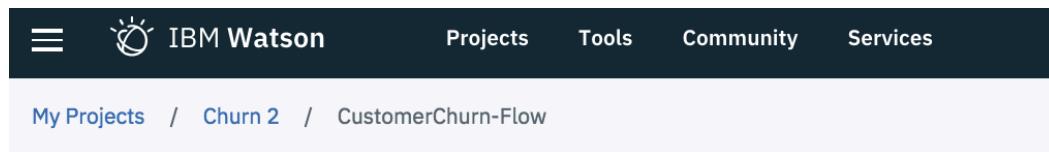
Action
<p>3. Add Data Asset</p> <p>You should now see an empty workspace.</p> <ul style="list-style-type: none"> On the top left click on the “Palette” icon, then click on the “Import” the icon. 
<p>The palette represents the set of tools available for use with Watson Studio flows. The menu of the right should look familiar.</p> <ul style="list-style-type: none"> Let's start by dragging and dropping the “Data Asset” node onto the workspace. 

Action

- Double click on the “**Data Asset**” node. In the top right-hand corner click on the “**Change data asset**” icon.



- At the next dialog click on the “Data assets” icon. Choose the “customer_churn.csv” file then click on OK to select that dataset then “Save”.



My Projects / Churn 2 / CustomerChurn-Flow

Churn 2

Data assets (3)

churn.csv

customer2.csv

customer_churn.csv

Action

3. Add & Configure Partition and Type Object

- From the palette, expand “**Field Operations**”, then drag and drop “**Partition**” onto the workspace and to the right of “Data Asset”. Connect the two objects:



- Double click on “**Partition**” then “**Settings**”.
- Set “**Training Partition**” to 80 and “**Testing Partition**” to 20. Click on Save.

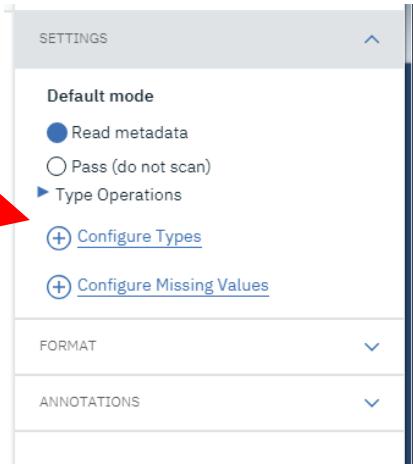
Node	Value
Training Partition	80
Testing Partition	20
Create validation partition	<input type="checkbox"/>
Repeatable partition assignment	<input checked="" type="checkbox"/>
Seed	Generate
1234567	
Use unique field to assign partitions	<input type="checkbox"/>

- Next, drag and drop a “**Type**” node onto the canvas. Connect it to the “**Partition**” node.

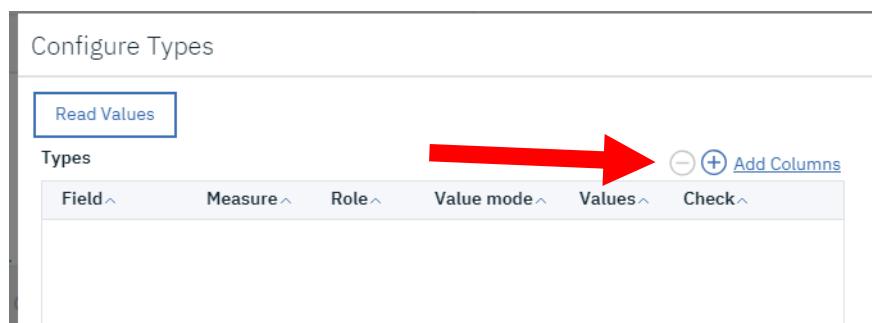


- Double click on “**Type**”. On the top right click on “**Configure Types**”:

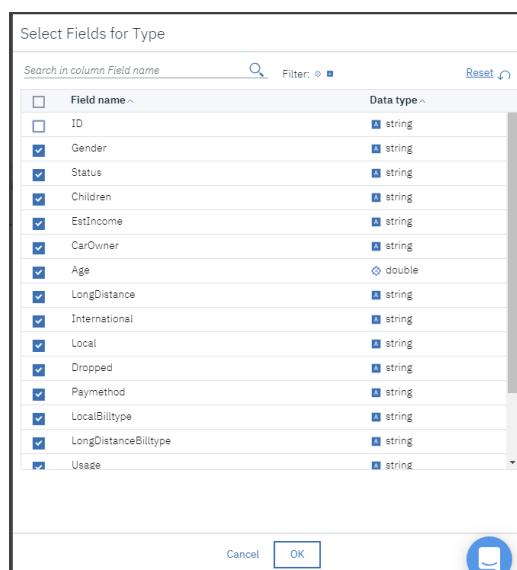
Action



- Click on “Add Columns”:



- Choose all of the fields except for “ID” then click on OK.



Action

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

Configure Types

Read Values

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	...

- 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”

Action

CHURN

FIELDS

Use custom field roles

Target
CHURN

Inputs (-) (+) Add Columns

- ID
- Gender
- Status
- Car Owner

OBJECTIVE



- Click on “**Save.**”

MODEL OPTION

ANNOTATIONS

Cancel Save 



- Your palette should resemble this:



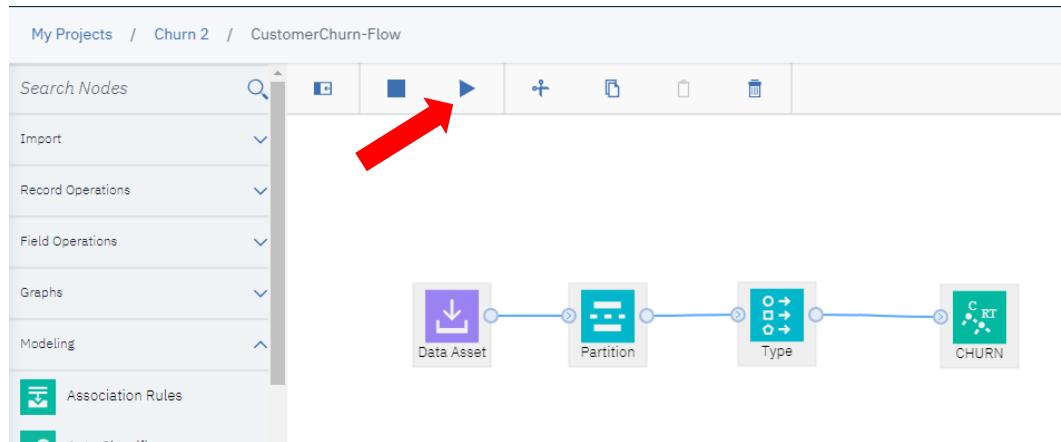
```

graph LR
    DA[Data Asset] --> P[Partition]
    P --> T[Type]
    T --> CHURN[CHURN]
  
```

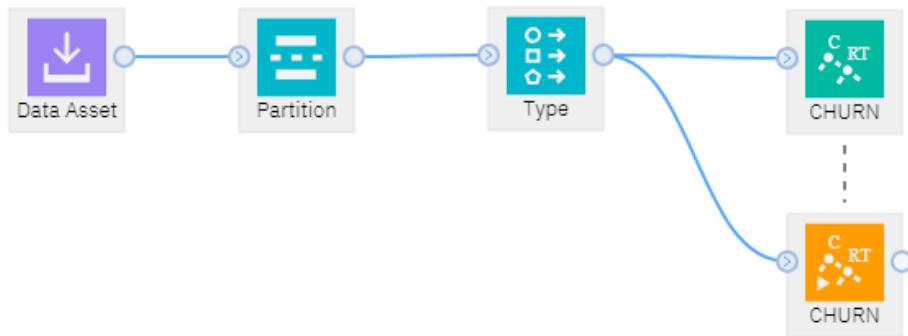
Action

5. Run Flow to Create Nugget

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



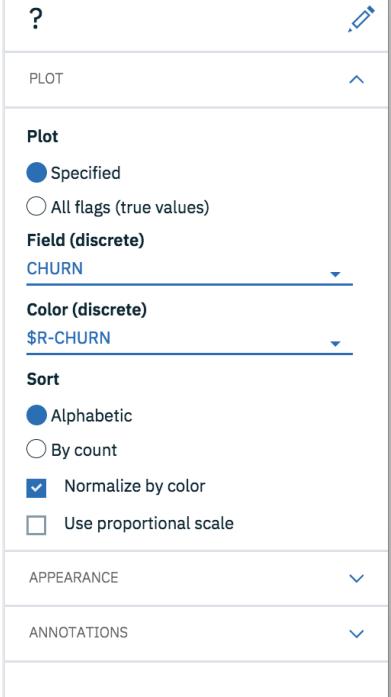
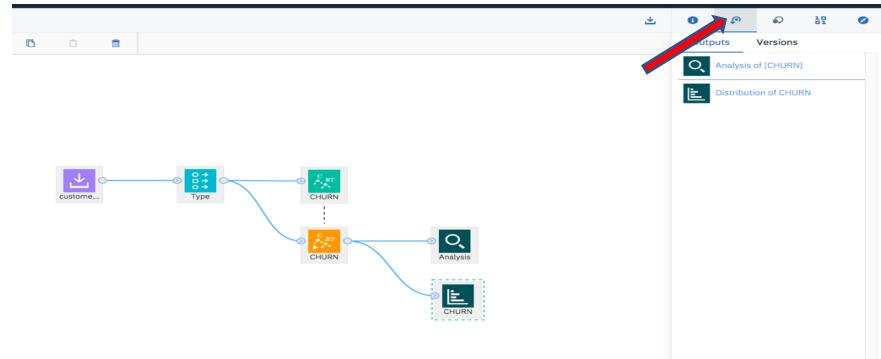
You should see a new forth object on the workspace, this is called a nugget.



6. Add & Configure Analysis Object – Measure Model Performance

- 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:

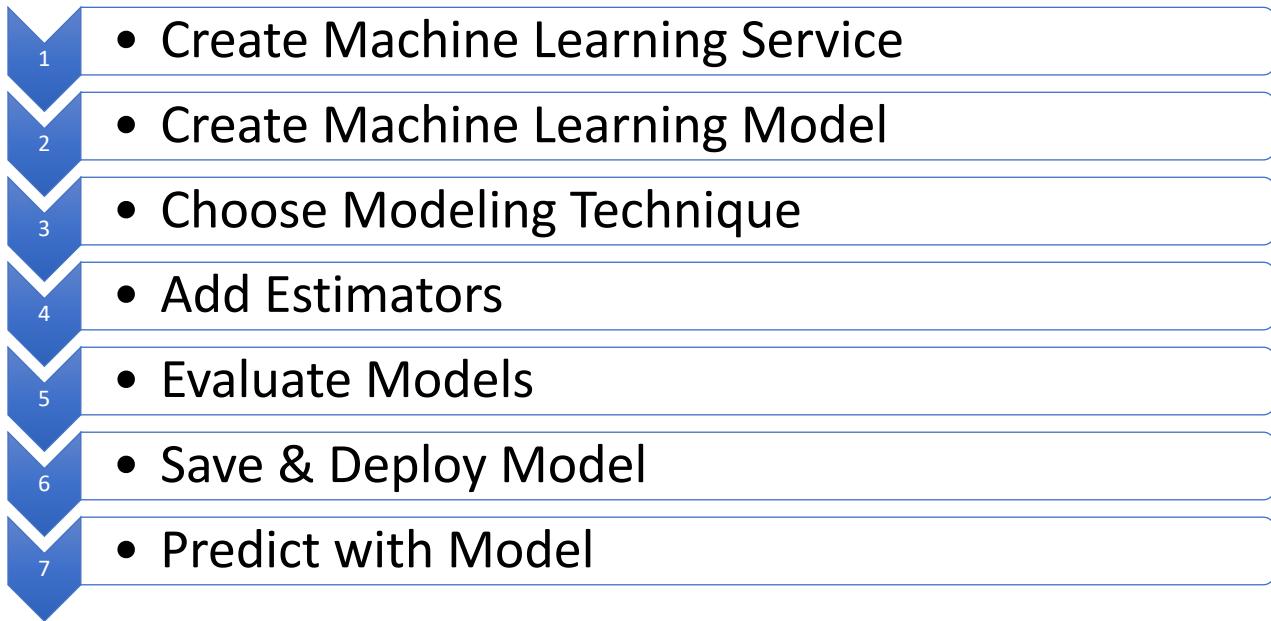
Action
<pre> graph LR DA[Data Asset] --> P[Partition] P --> T[Type] T --> CHURN1[CHURN] CHURN1 -.-> CHURN2[CHURN] CHURN2 --> A[Analysis] A --> D[?] </pre>
<ul style="list-style-type: none"> Double click on “Analysis” and check off the four checkboxes, leave the rest as default:
<ul style="list-style-type: none"> Click “Save” to return to the workspace. Double click on the “Distribution” object and configure it as depicted below:

Action

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

<ul style="list-style-type: none"> Explore the results
<h2>End of Lesson 4</h2>

Lesson 5: 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 5: Workflow Overview

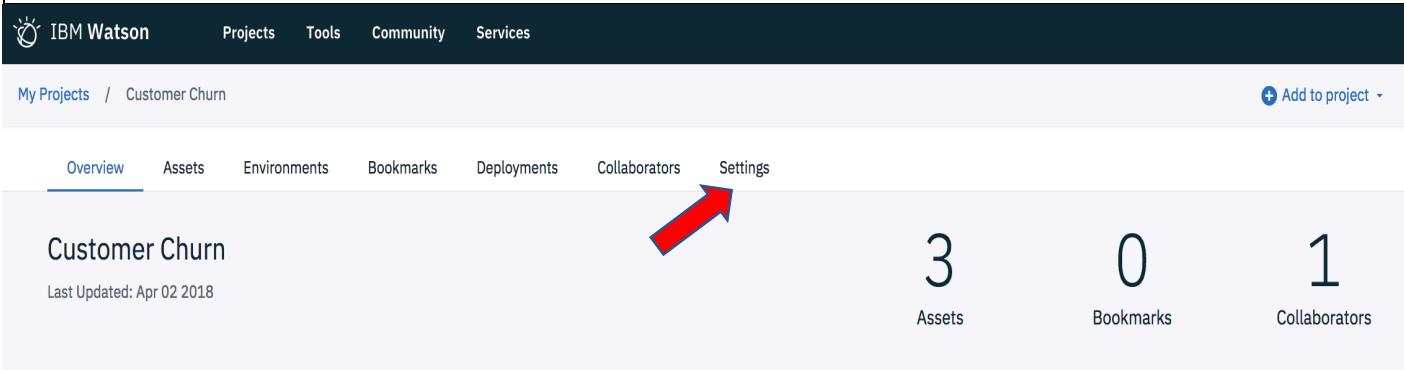


Lesson 5: Instructions

Action

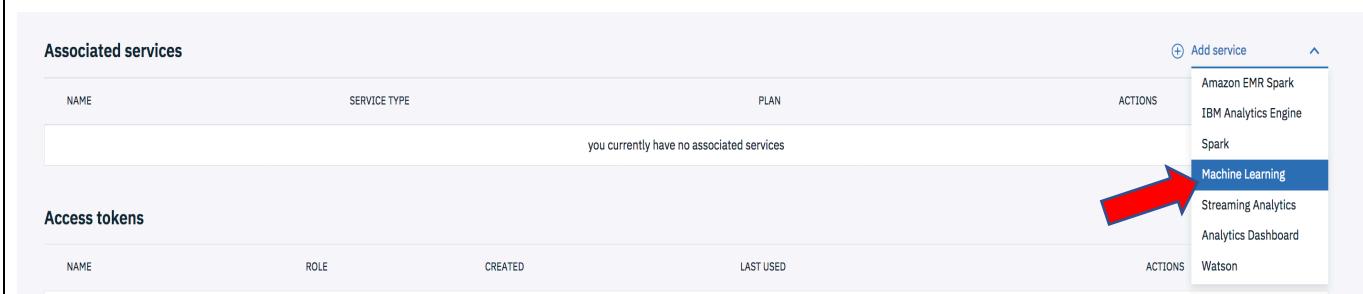
1. Create Machine Learning Service

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



The screenshot shows the IBM Watson Project Overview page for a project named "Customer Churn". The top navigation bar includes links for "IBM Watson", "Projects", "Tools", "Community", and "Services". Below the navigation, the project name "Customer Churn" is displayed along with its last update date, "Apr 02 2018". On the right side, there are summary counts: 3 Assets, 0 Bookmarks, and 1 Collaborator. The main content area has tabs for "Overview", "Assets", "Environments", "Bookmarks", "Deployments", "Collaborators", and "Settings". A red arrow points to the "Settings" tab. Below the tabs, there are sections for "Associated services" and "Access tokens".

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



The screenshot shows the "Associated services" section of the project page. It includes a table for "Associated services" and a table for "Access tokens". To the right of the tables is a "Actions" column with a "Add service" button. A dropdown menu is open, listing several service options: "Amazon EMR Spark", "IBM Analytics Engine", "Spark", "Machine Learning", "Streaming Analytics", "Analytics Dashboard", and "Watson". A red arrow points to the "Machine Learning" option, which is highlighted with a blue background.

- 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>								
<table border="0"> <tr> <td style="vertical-align: top;"> <p>Features</p> <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> </td> <td style="vertical-align: top;"> <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> </td> </tr> </table>			<p>Features</p> <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>Features</p> <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" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding-bottom: 5px;">Plan</th> <th style="text-align: left; padding-bottom: 5px;">Features</th> <th style="text-align: left; padding-bottom: 5px;">Pricing</th> </tr> </thead> <tbody> <tr> <td style="text-align: center; vertical-align: middle;">  Lite  </td> <td style="text-align: center; vertical-align: middle;"> Service instance (5 models per instance) 5,000 predictions 5 compute hours </td> <td style="text-align: center; vertical-align: middle;"> 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						
<ul style="list-style-type: none"> • 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

Models

NAME	STATUS	TYPE	RUNTIME	LAST MODIFIED	ACTIONS
you currently have no models					

New model

- 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](#)

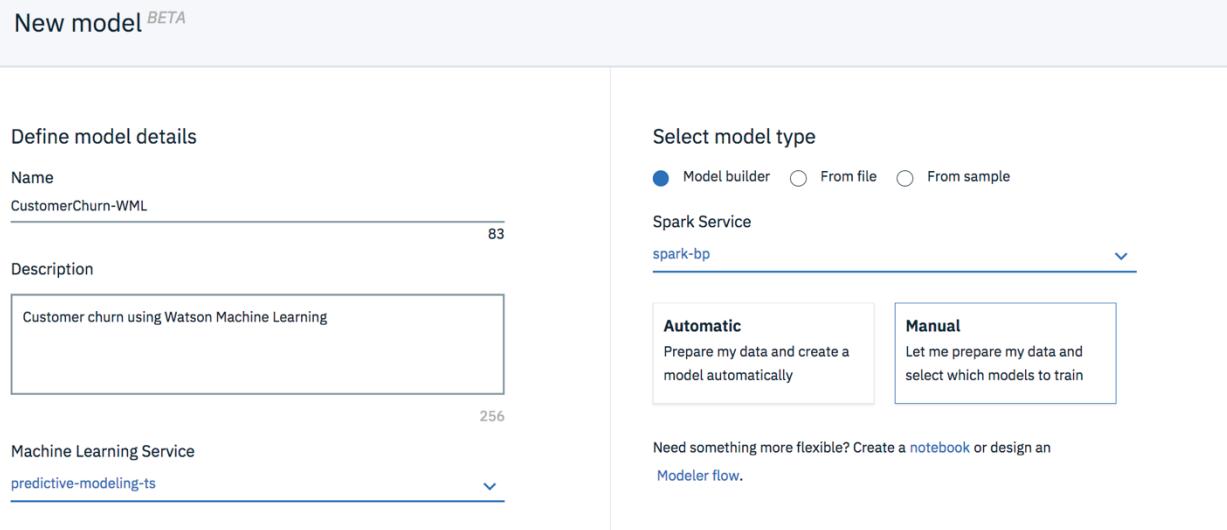
PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	2 Spark Executors An entry level plan to run programs using up to 2 Spark executors	Free

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**”:



New model BETA

Define model details

Name
CustomerChurn-WML

Description
Customer churn using Watson Machine Learning

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

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

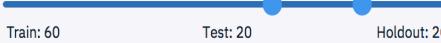
Select a technique

Column value to predict (Label Col)
CHURN (String)

Feature columns
Gender (String), Status (String), Children (Decimal), Est Income (Decimal),
Car Owner (String), Paymethod (String), LongDistanceBilbype (String),
Usage (Decimal), RatePlan (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

 Train: 60 Test: 20 Holdout: 20

Configured estimators 

 Add Estimators

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

- 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

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															
	CustomerChurn-WML  <div style="border: 1px solid #ccc; padding: 10px;"> <p>Overview Evaluation Deployments</p> <h3>Summary</h3> <table border="1" style="width: 100%; border-collapse: collapse;"> <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> </table> <h3>Input Schema</h3> </div>	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**”:

CustomerChurn-WML  <div style="border: 1px solid #ccc; padding: 10px;"> <p>Overview Evaluation Deployments</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>NAME</th> <th>STATUS</th> <th>DEPLOYMENT TYPE</th> <th>ACTIONS</th> </tr> </thead> <tbody> <tr> <td colspan="4">Your model is not deployed.</td> </tr> </tbody> </table> </div>				NAME	STATUS	DEPLOYMENT TYPE	ACTIONS	Your model is not deployed.			
NAME	STATUS	DEPLOYMENT TYPE	ACTIONS								
Your model is not deployed.											

- 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

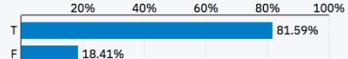
Gender
F

Status
S

Children
1

Est Income

Predicted value for CHURN



End of Lesson 5

End of Hands-on Workshop

Thank You