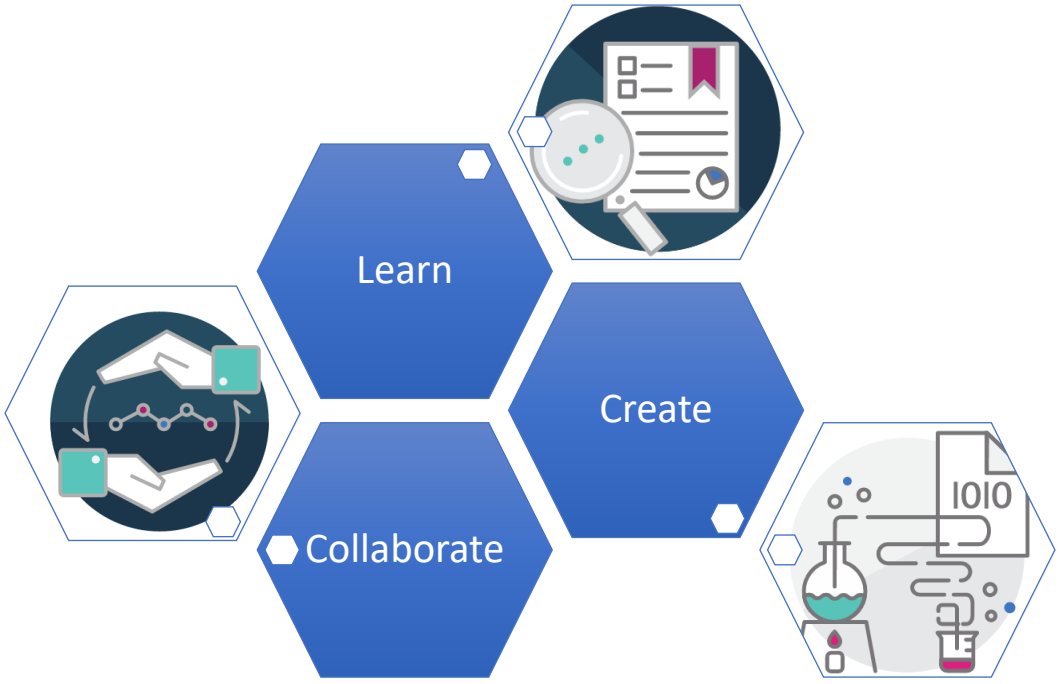
IBM Watson

Predicting Customer Churn

Watson Studio

Lab Guide



Notices and Disclaimers

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.frolio@ibm.com
Loren Murphy – lrmurphy@us.ibm.com

Table of Contents

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:	Tasks you will complete in this lab exercise include: • 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

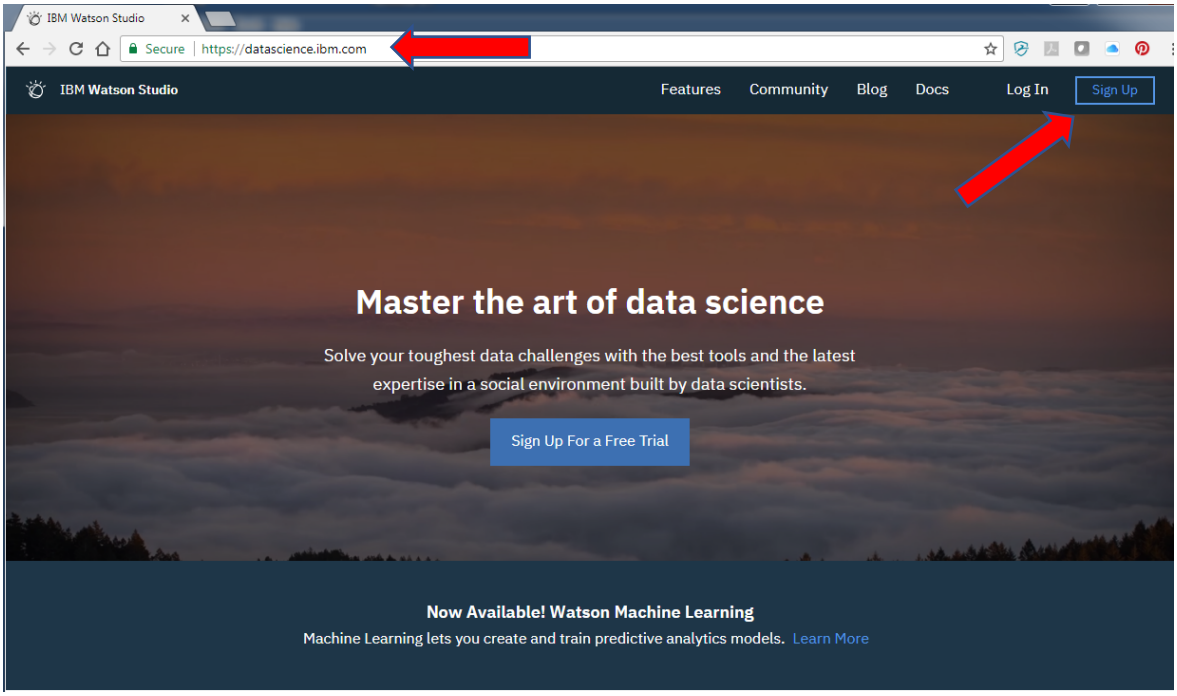
1. Create Account / Sign-In to Watson Studio
2. Live Chat
3. Community Cards
4. Profile Settings
5. Apps and Services

Lesson 1: Instructions

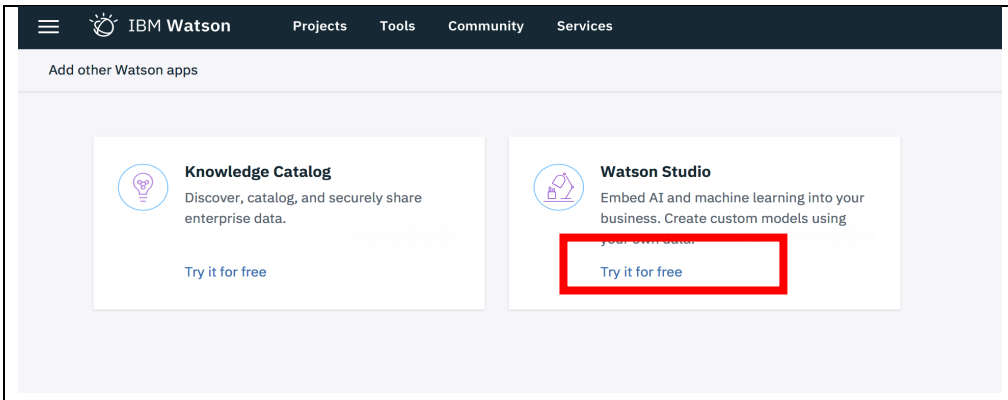
Action

1. **Create Account/Sign In to Watson Studio**

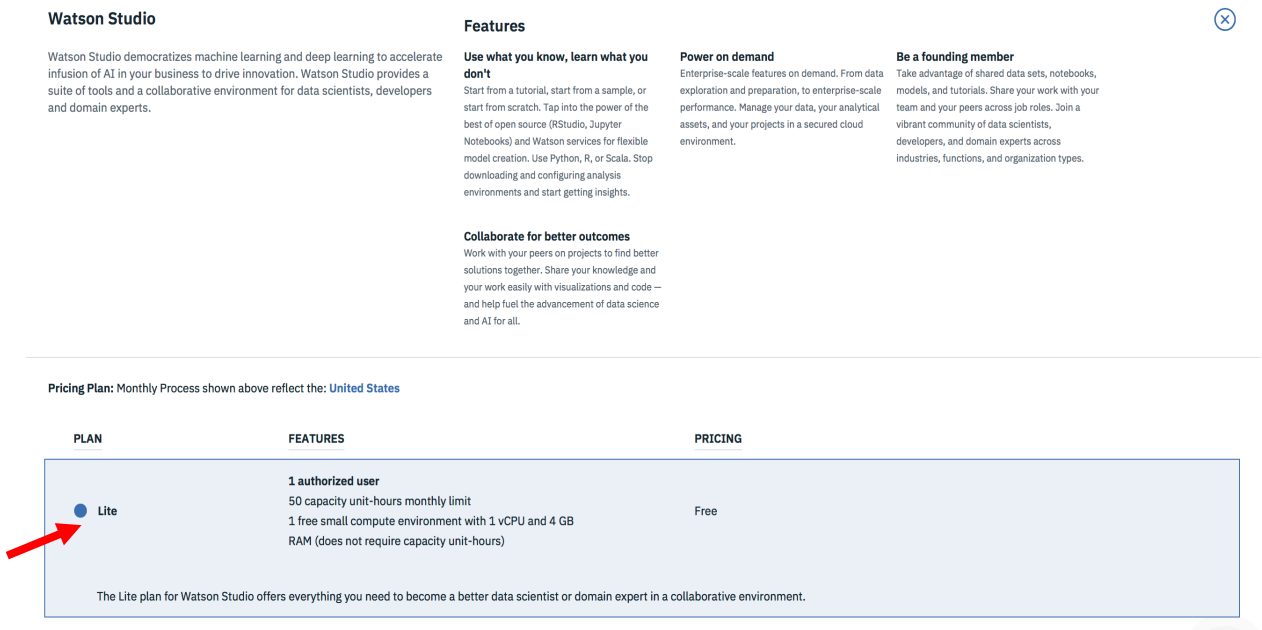
- Open web browser and navigate to: **https://datascience.ibm.com**



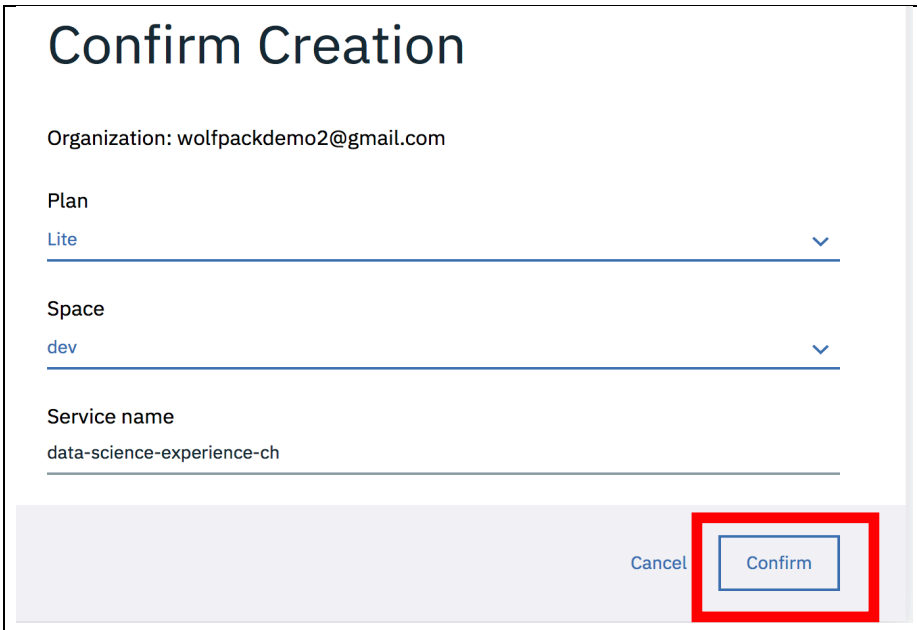
- Click on “Sign Up” and you will be prompted for several items of information. After a few moments of self-configuration, you will be brought to the following page.
- Click on **Watson Studio.**



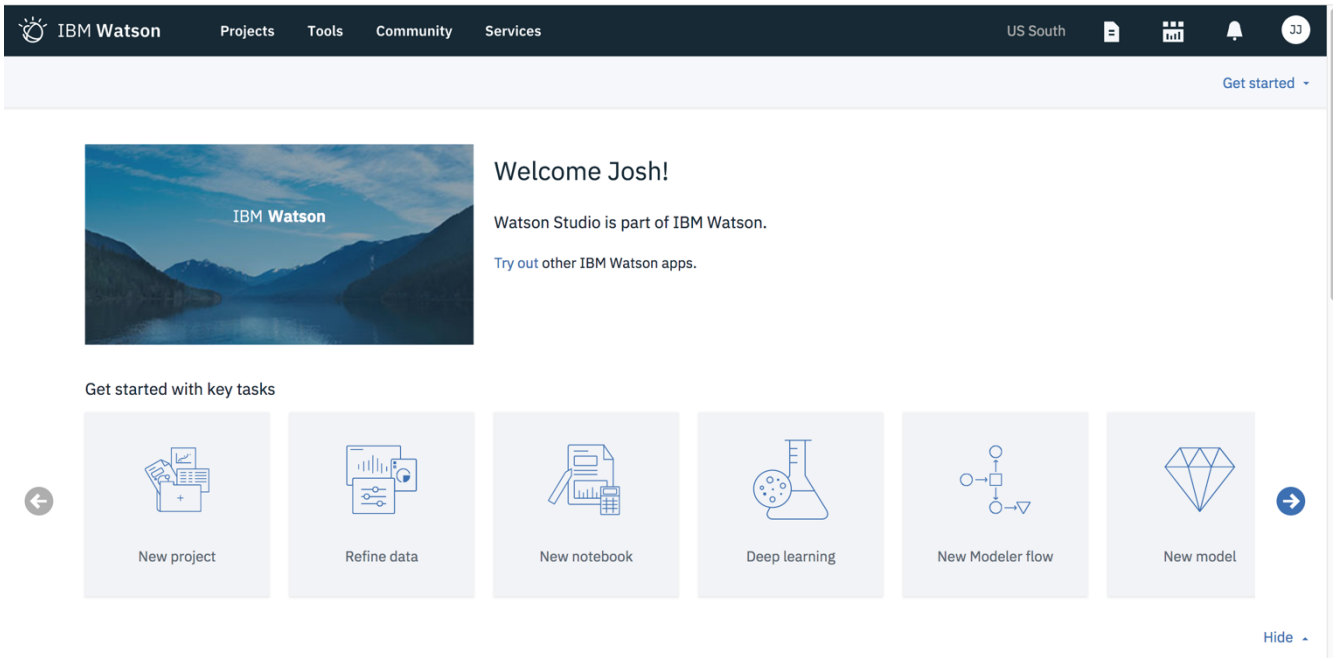
- Select the **Lite** Plan and click **Create**



- Select **Confirm**



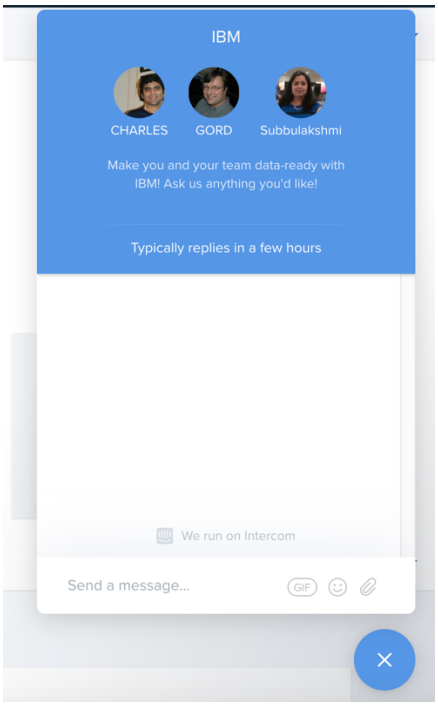
- You should now see your Watson Studio Homepage



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:

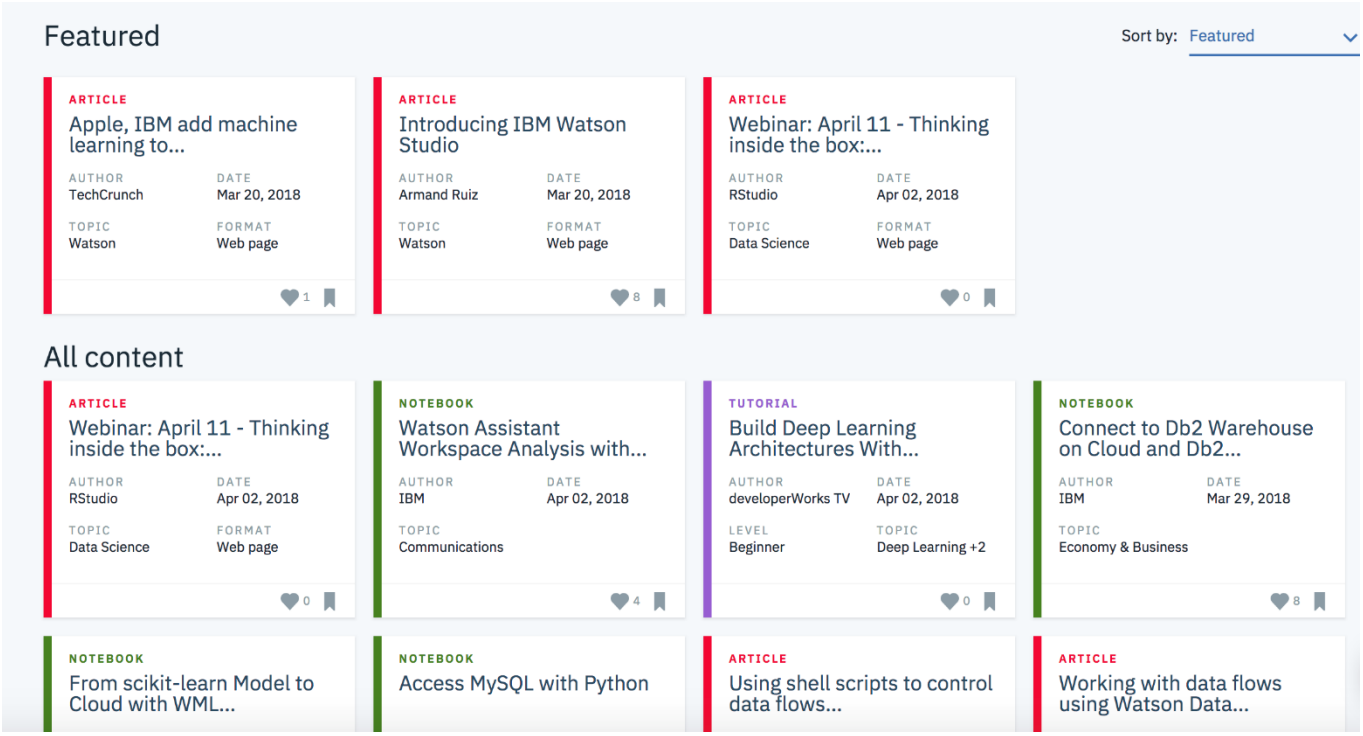


If you need assistance, start typing your message in the **Send a Message** box to connect with a live person. Through this Live Chat feature, you can also continue conversations the next time you log into Watson Studio.

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

3. Community Cards

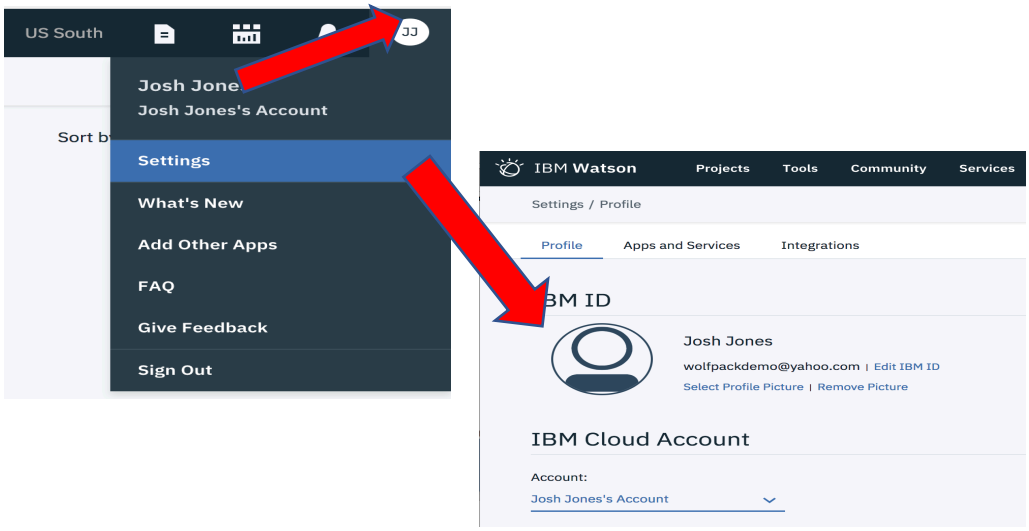
At the top of the Home Page click on **Community**:



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.

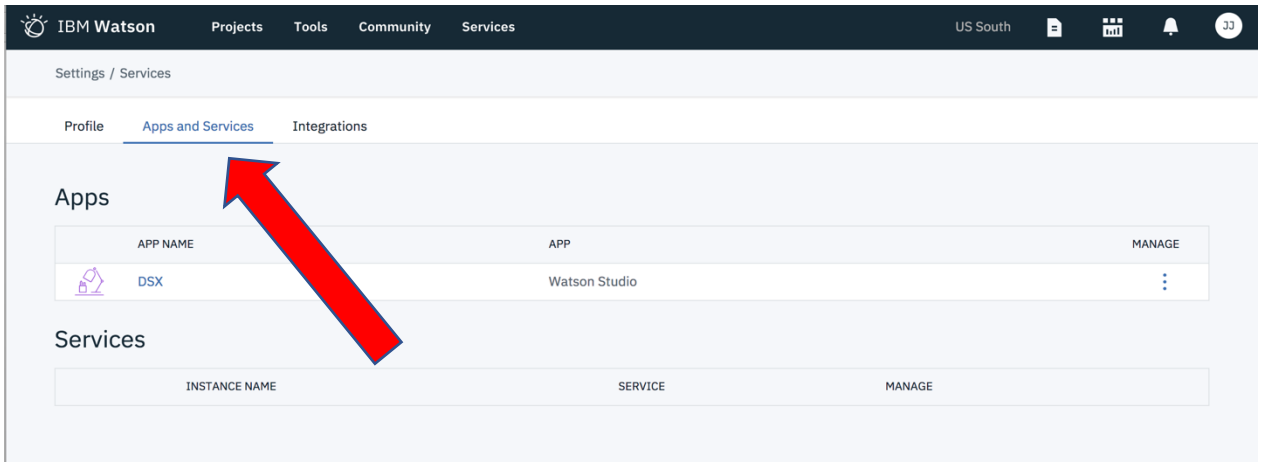
4. Profile Settings

- 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:



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.

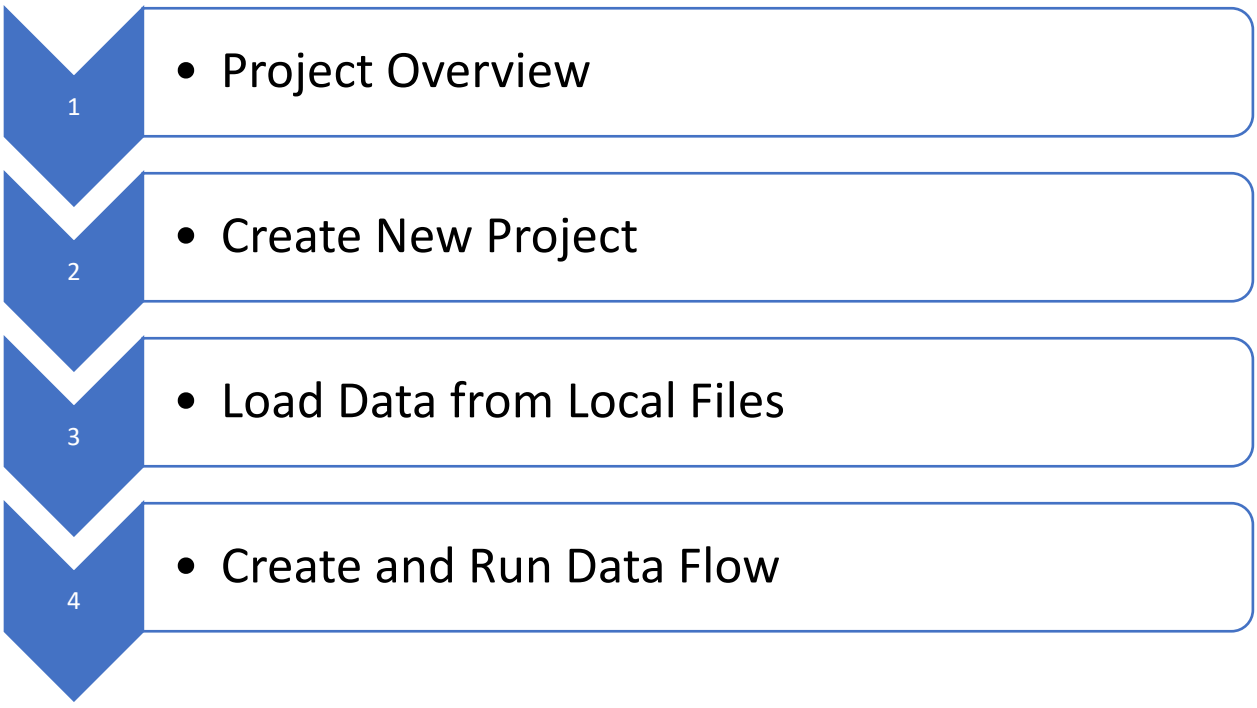
End of Lesson 1

Lesson 2: 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: • Create and Configure Watson Studio Project • Add Data Flow Asset • Cleanse and Transform Data • Run Data Flow

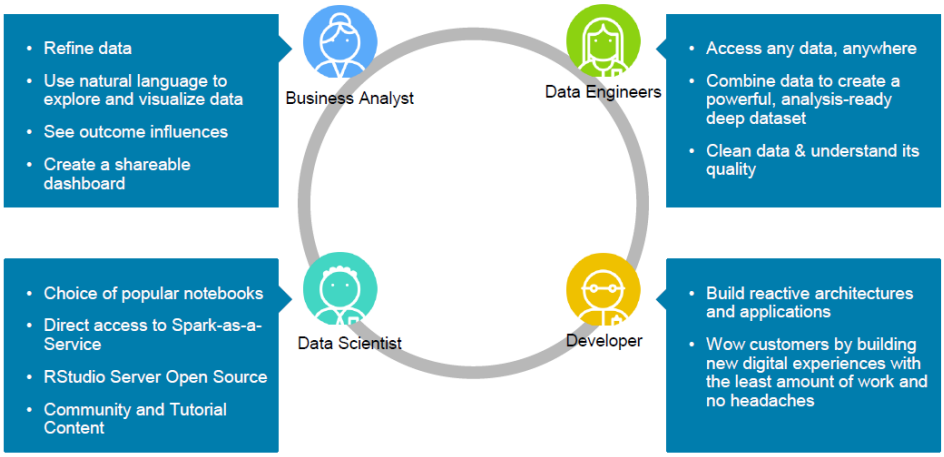
Lesson 2: Workflow Overview



Lesson 2: Instructions

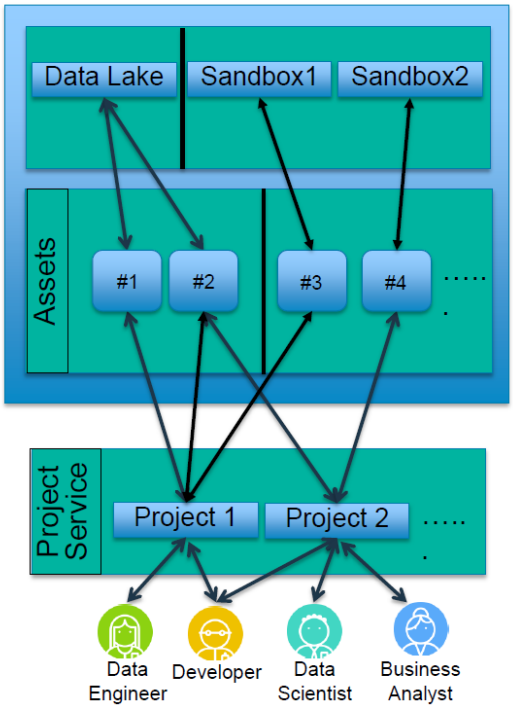
1. Project Overview

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



Projects make collaboration easier by:

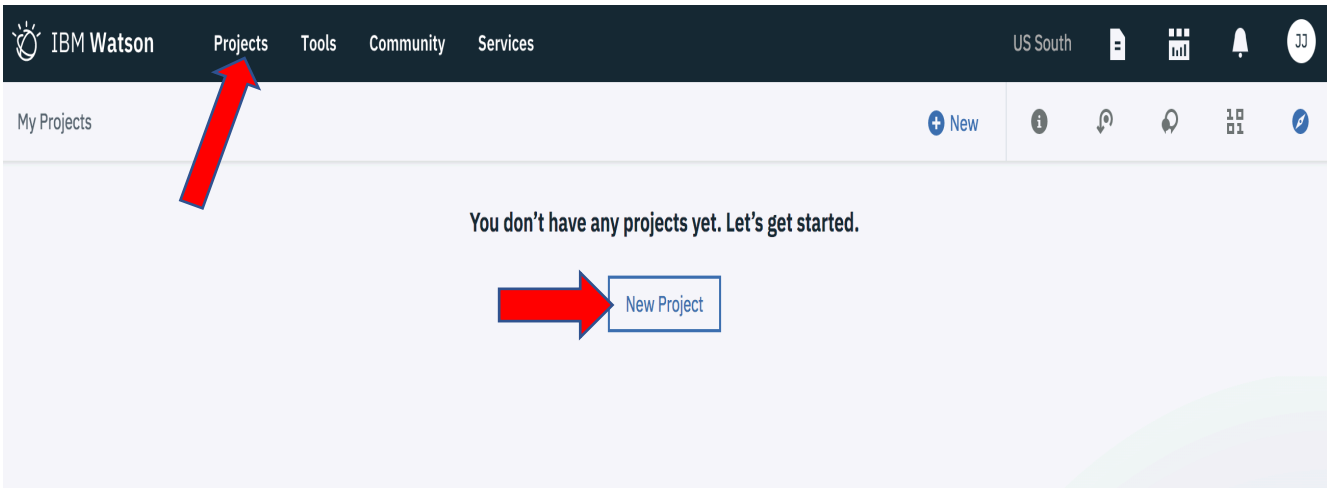
- Allowing different users and personas to share a set of assets
- Enabling users to collaborate and manage their notebooks, artifacts, plus more
- Providing three levels of rights: Viewers, Editors, and Admins



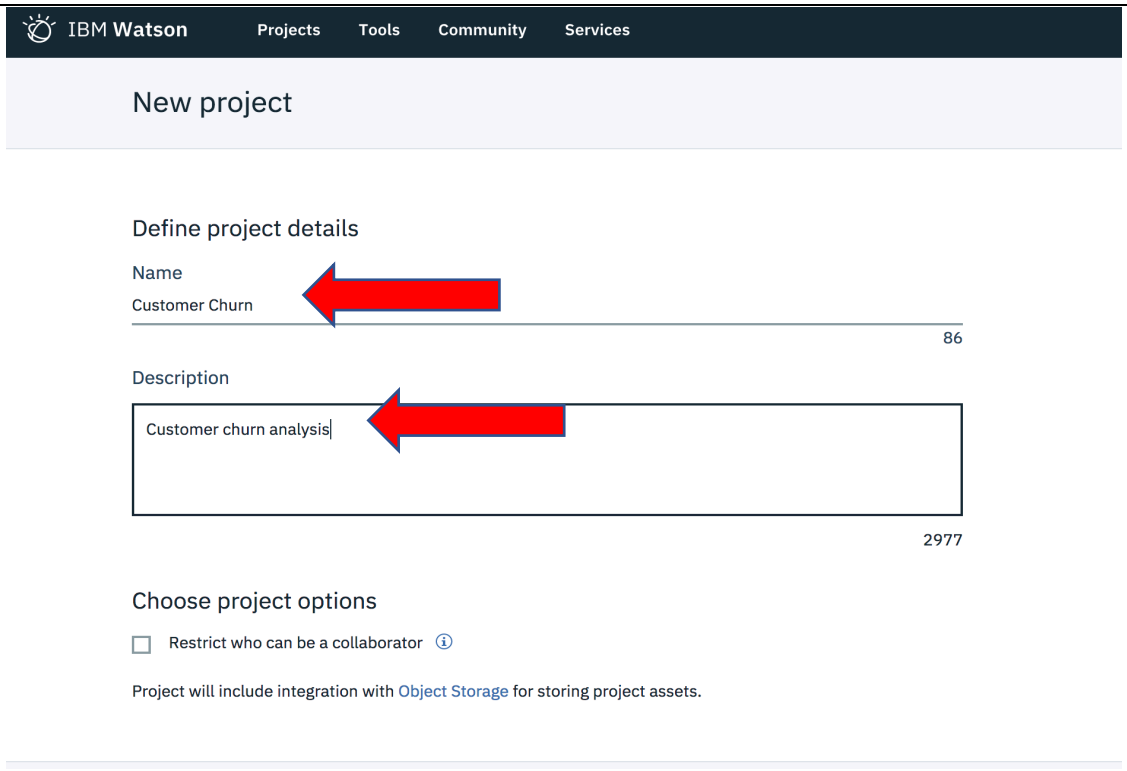
Action

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**

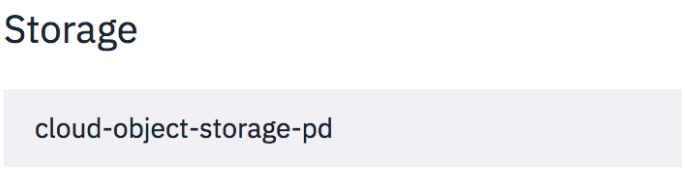


- Type the Project Name **Customer Churn,** add a meaningful description:

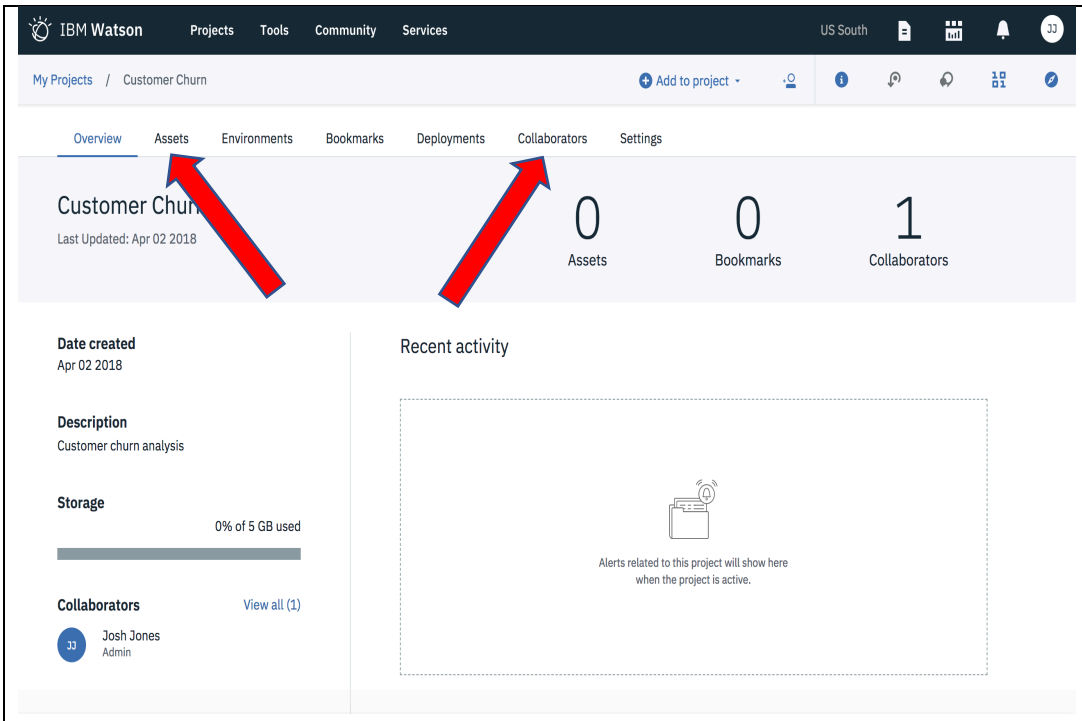


Define Storage:

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



- Click **Create**



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.

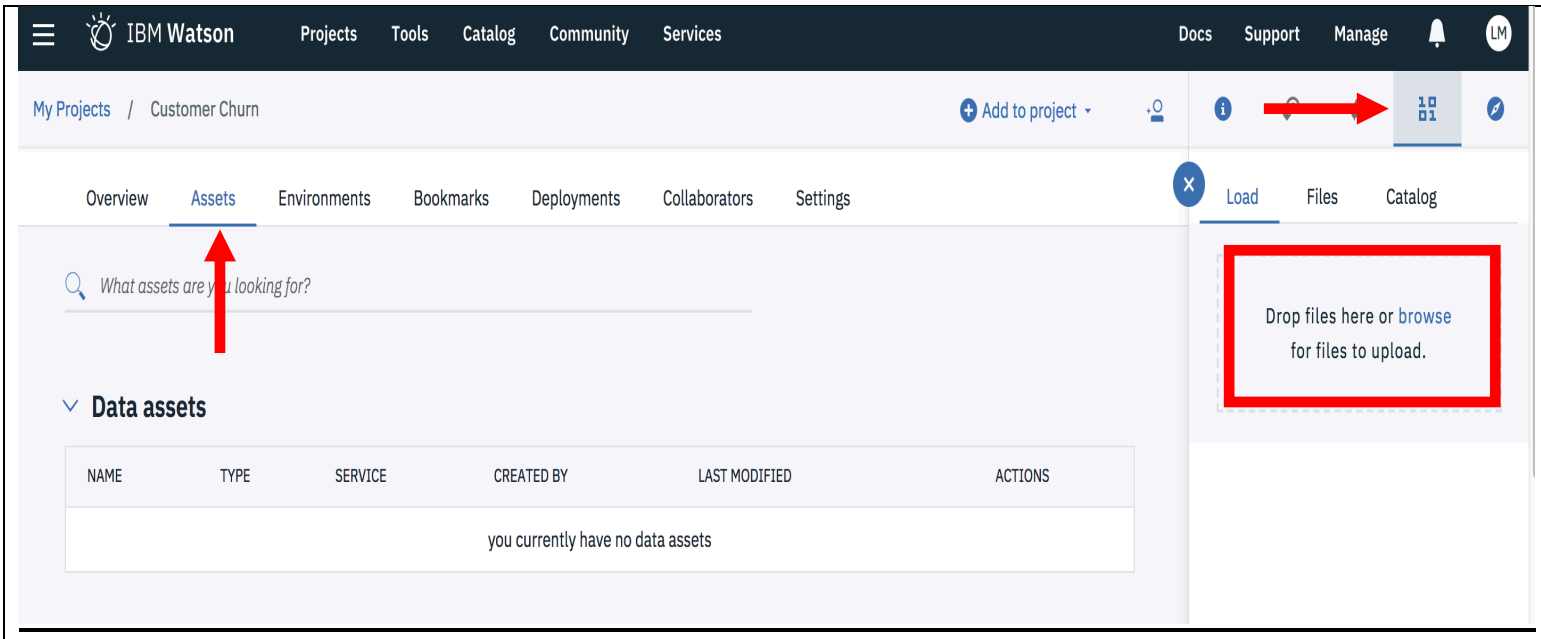
3. Load Data from Local File

- In a separate browser navigate to: Customer Churn Data:

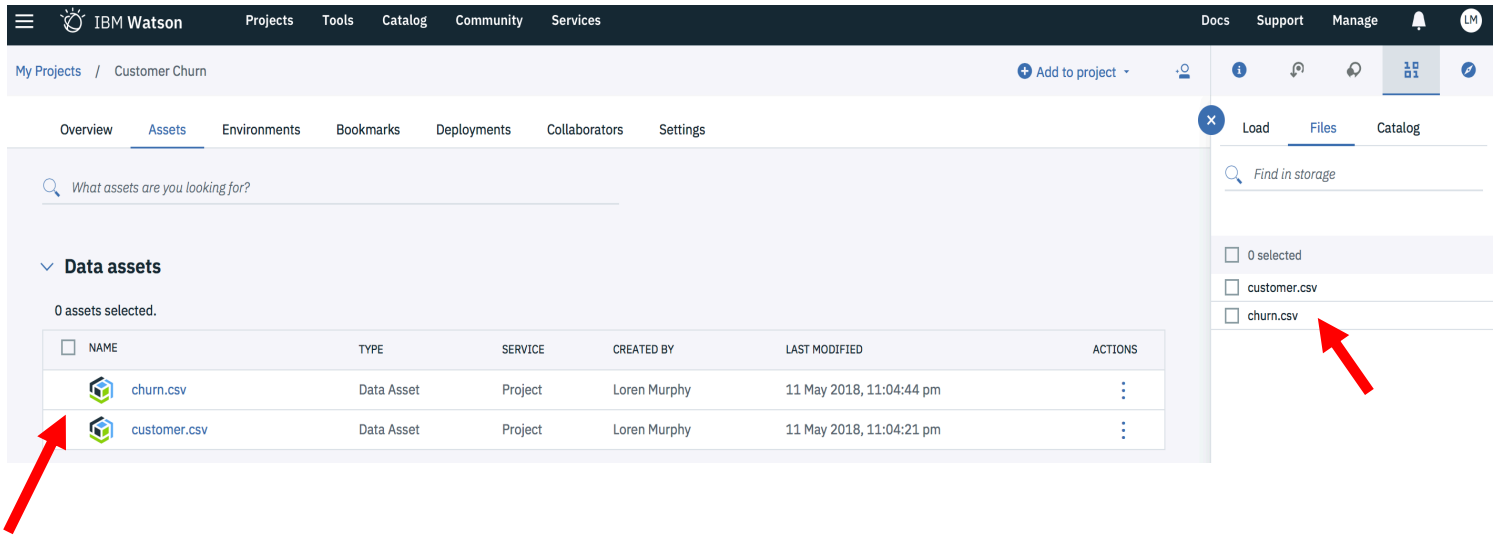
  https://github.com/team-wolfpack/Wolfpack-Technical-Event-Content/tree/master/AI and Data Science/ Machine Learning for Customer Churn/Data

- 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.

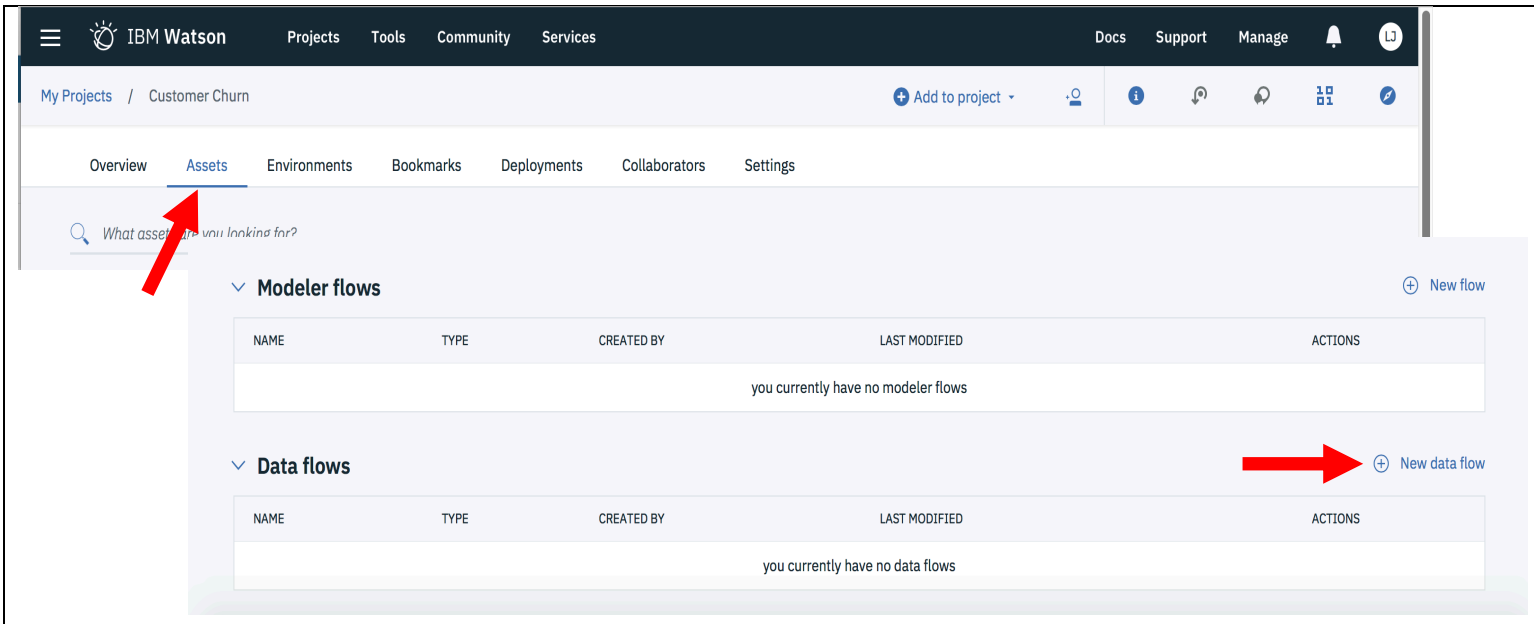


- You should now see that the customer and churn data files have been imported into the project under the **Files** and they are listed under **Data Assets**

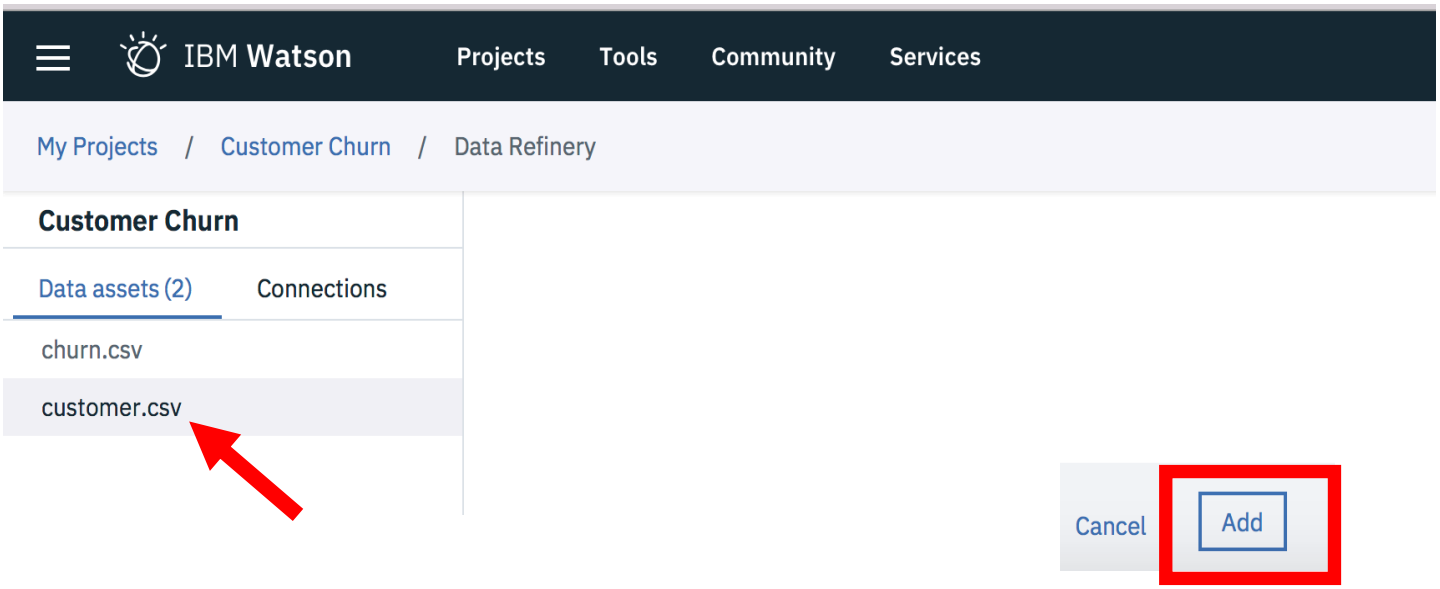


4. Create and Run Data Flow

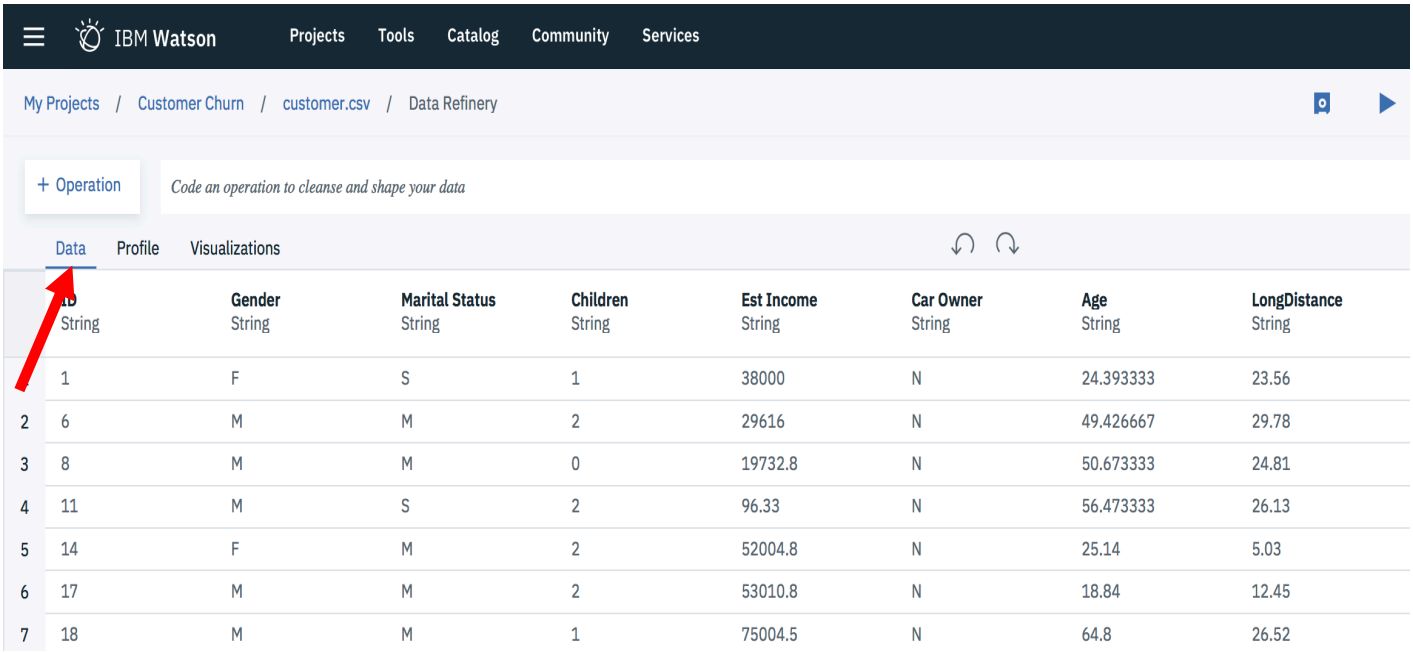
- Under **Assets**, click **New Data Flow**



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

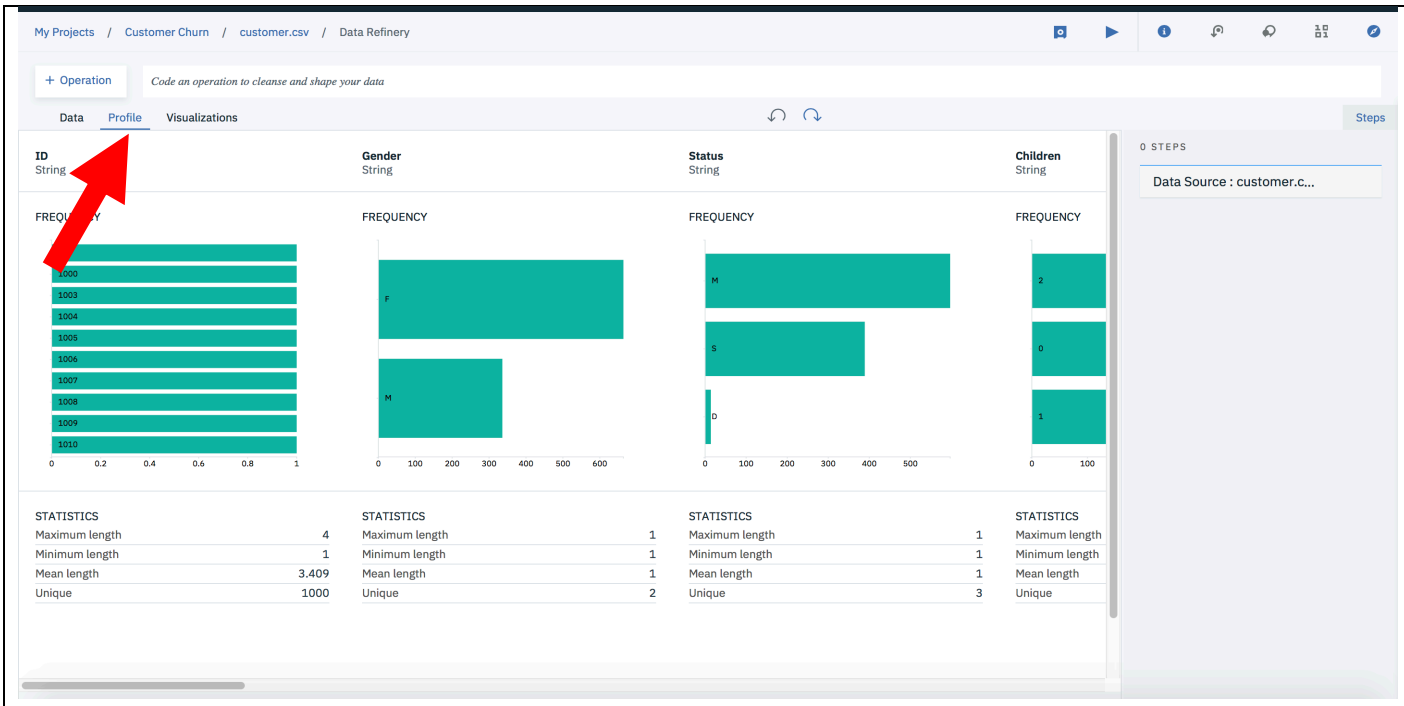


- A preview of customer data should now be displayed, under the **Data** tab. Only the first 1000 rows are displayed.

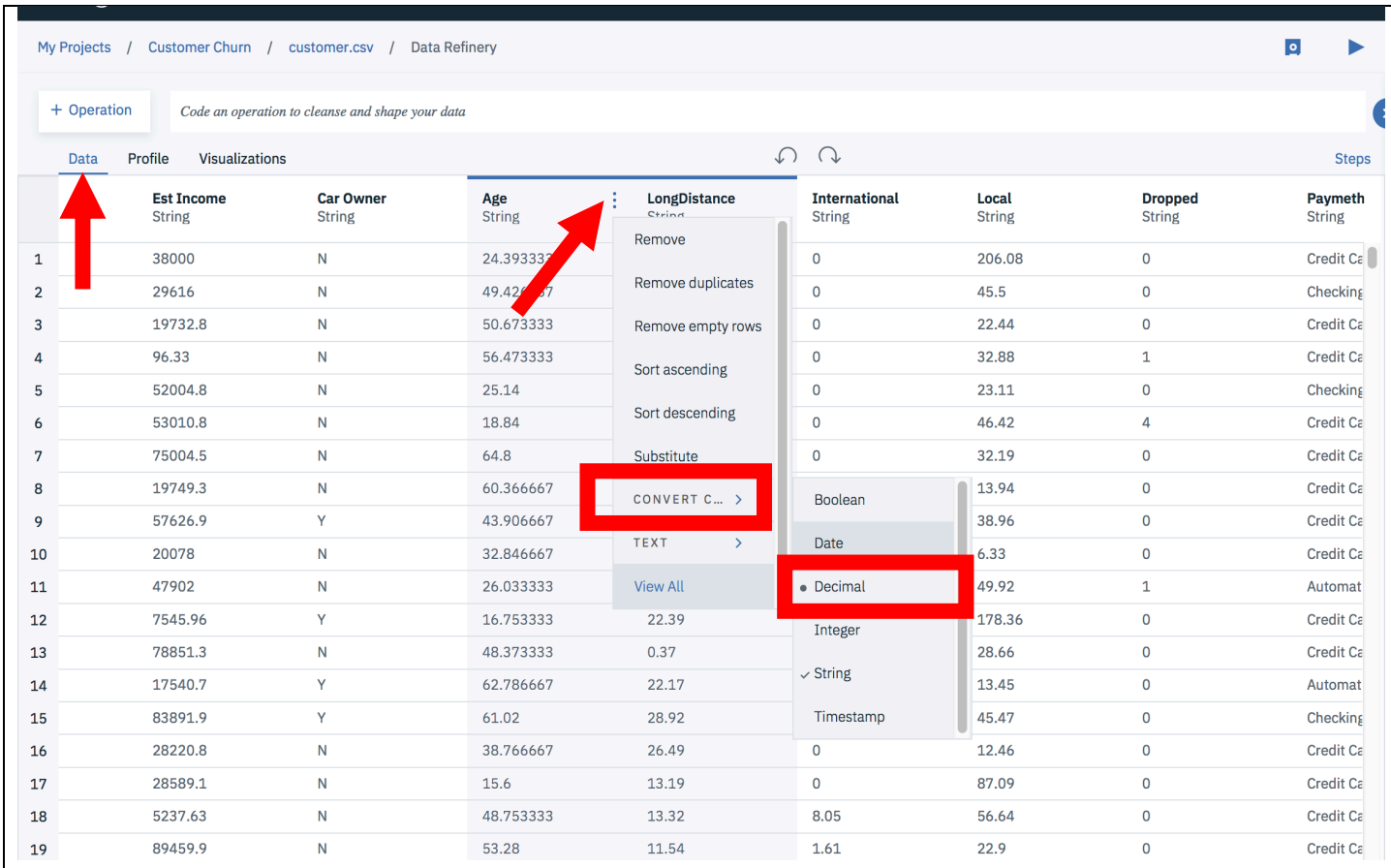


	ID String	Gender String	Marital Status String	Children String	Est Income String	Car Owner String	Age String	LongDistance String
1	1	F	S	1	38000	N	24.393333	23.56
2	6	M	M	2	29616	N	49.426667	29.78
3	8	M	M	0	19732.8	N	50.673333	24.81
4	11	M	S	2	96.33	N	56.473333	26.13
5	14	F	M	2	52004.8	N	25.14	5.03
6	17	M	M	2	53010.8	N	18.84	12.45
7	18	M	M	1	75004.5	N	64.8	26.52

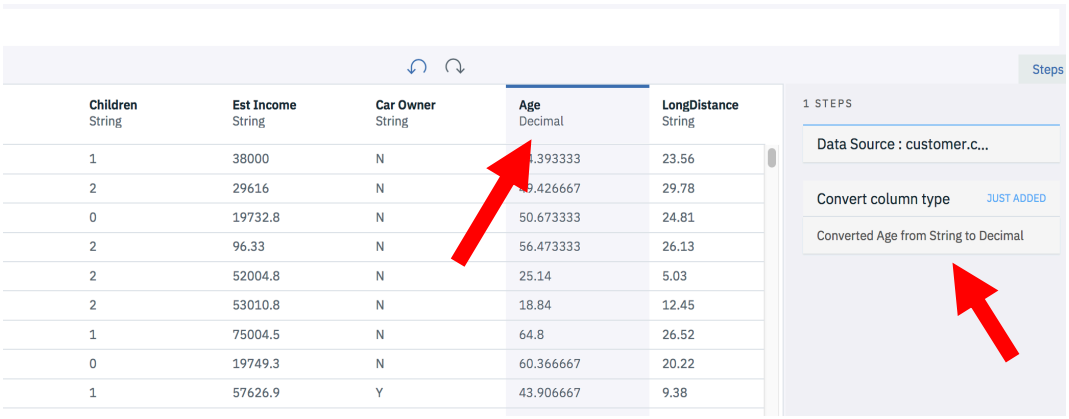
- Select the **Profile** tab. Information about your data, such as frequency, is displayed



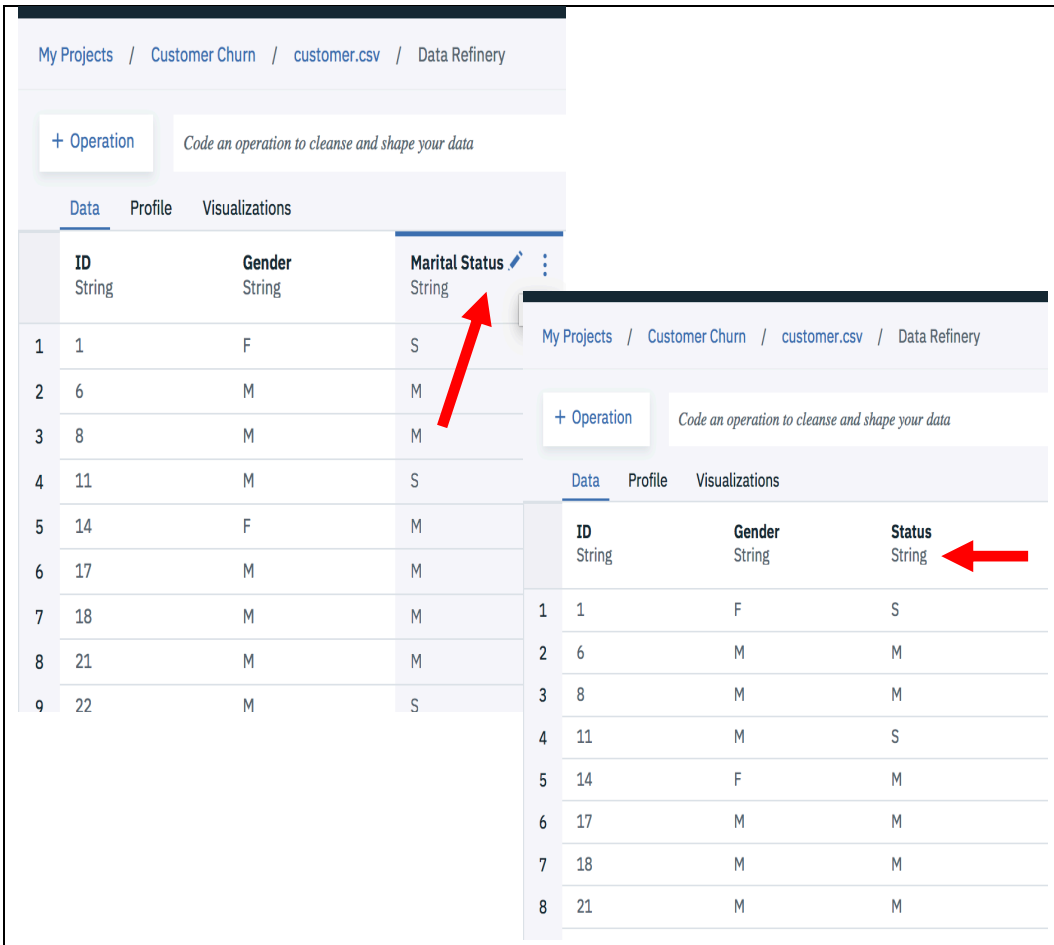
- Return to the **Data tab** and click on the **Age** column. Select the **3 dots** beside the column name and select **Convert Column Type.** Click **Decimal.** Notice a grey circle appeared beside Decimal. Based upon the data values, Watson recommended the data type.



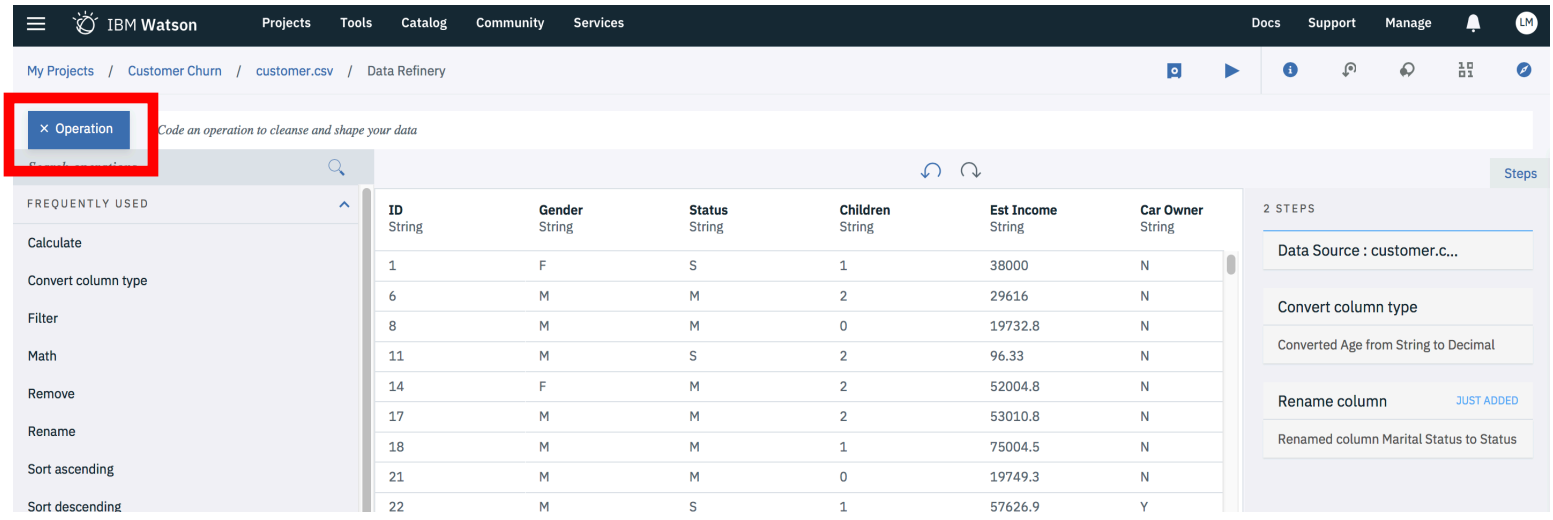
- The **Age** column is now of data type **Decimal** and our conversion action now appears on the left-hand side under **Steps.** Each transformation performed on the data is tracked under Steps



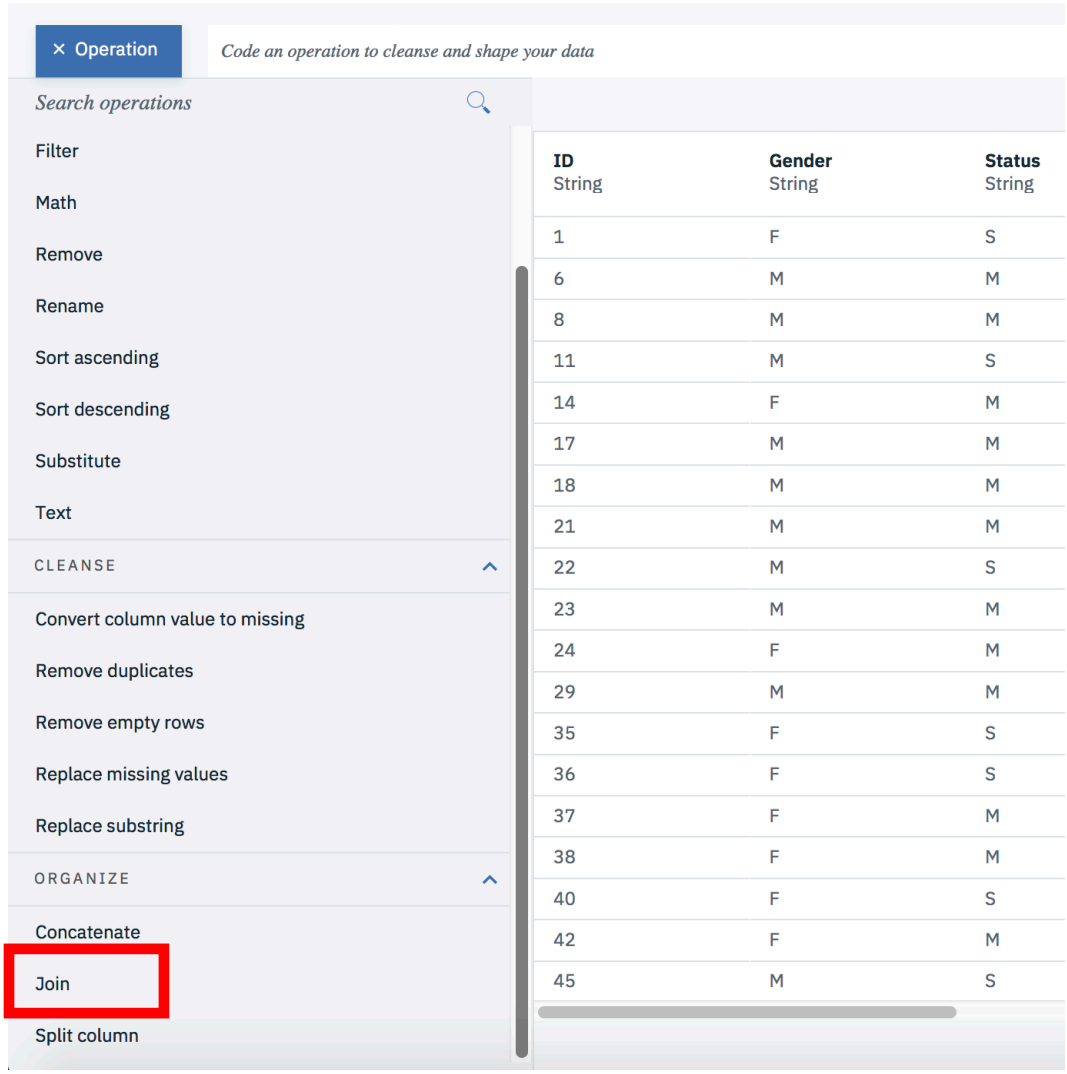
- Next, click the **Martial Status** column. Hover over the column name until an edit pencil icon appears. Change the column name to **Status**



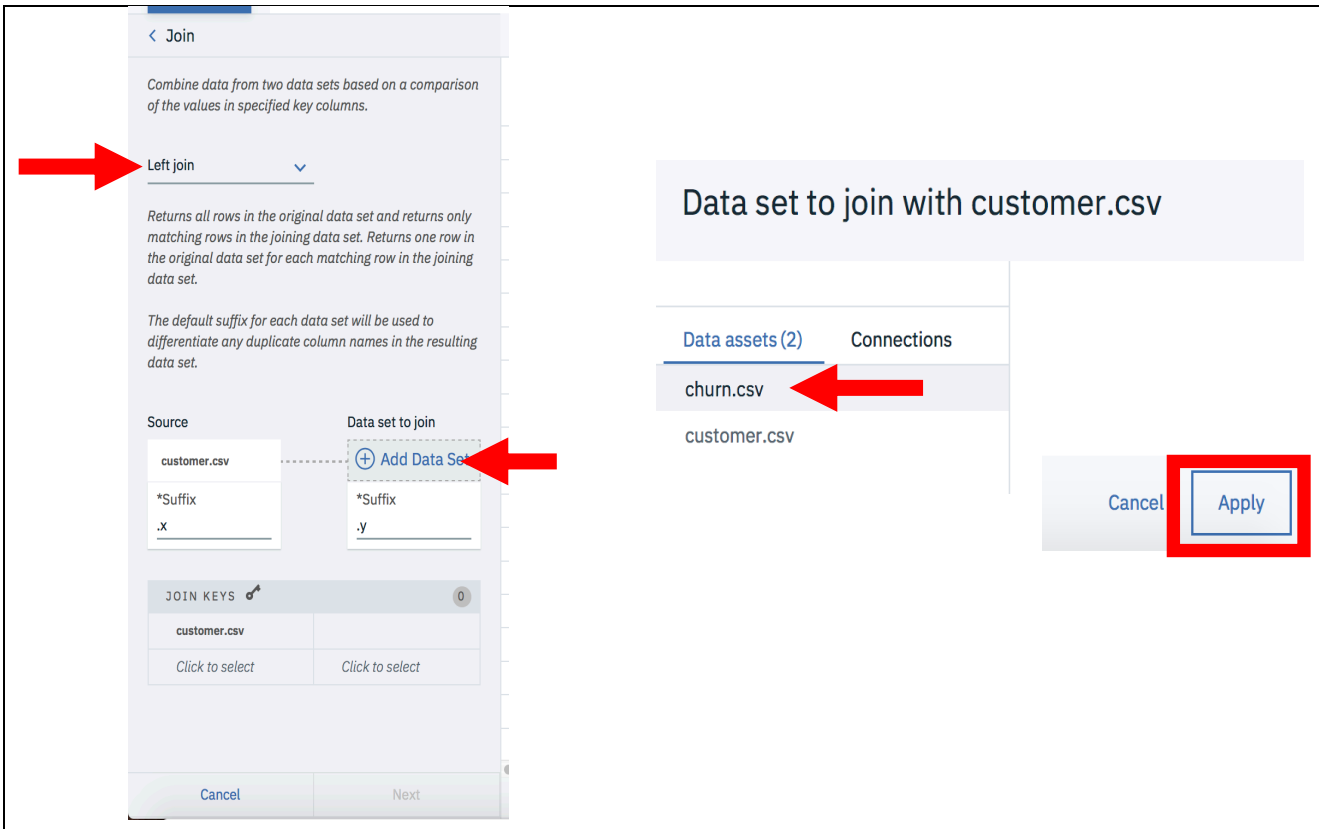
- Click **Operation** in the upper left corner.



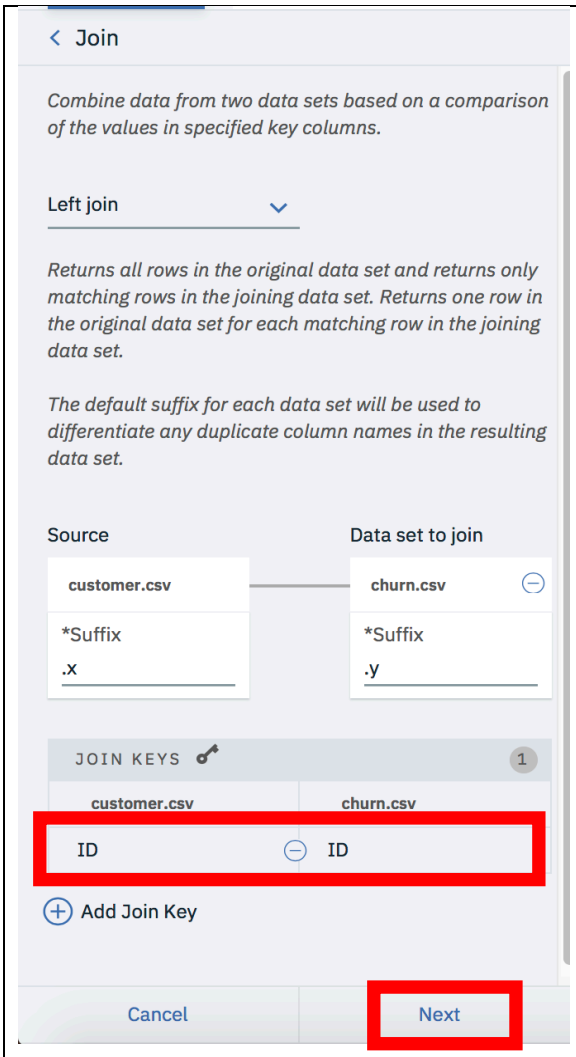
- Under the **Organize** section, click **Join**.

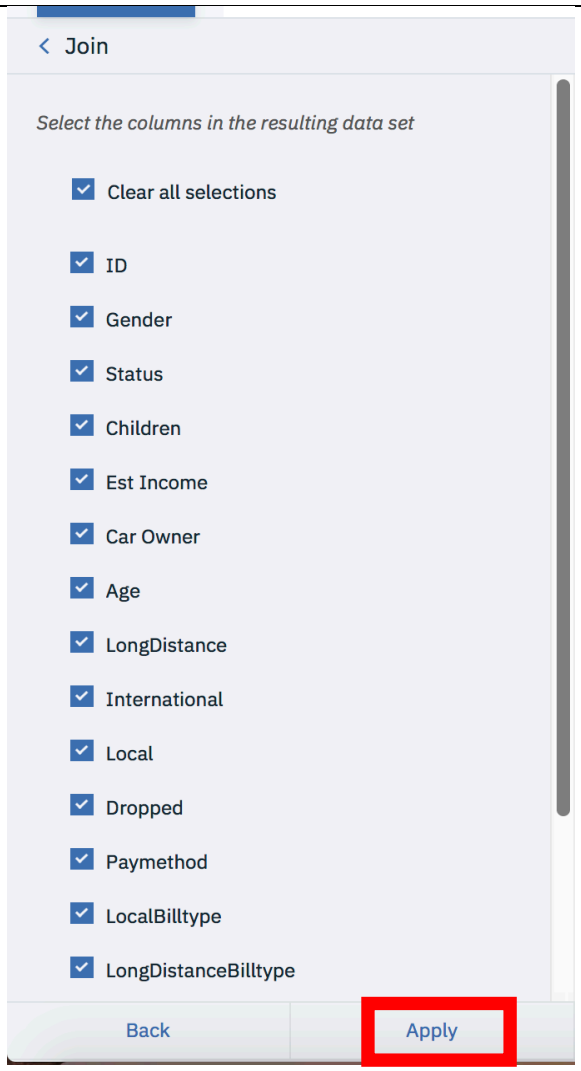


- Select **Left Join** and click **Add Data Set**. Select **Churn.csv**, as the data set to join, and click **Apply**

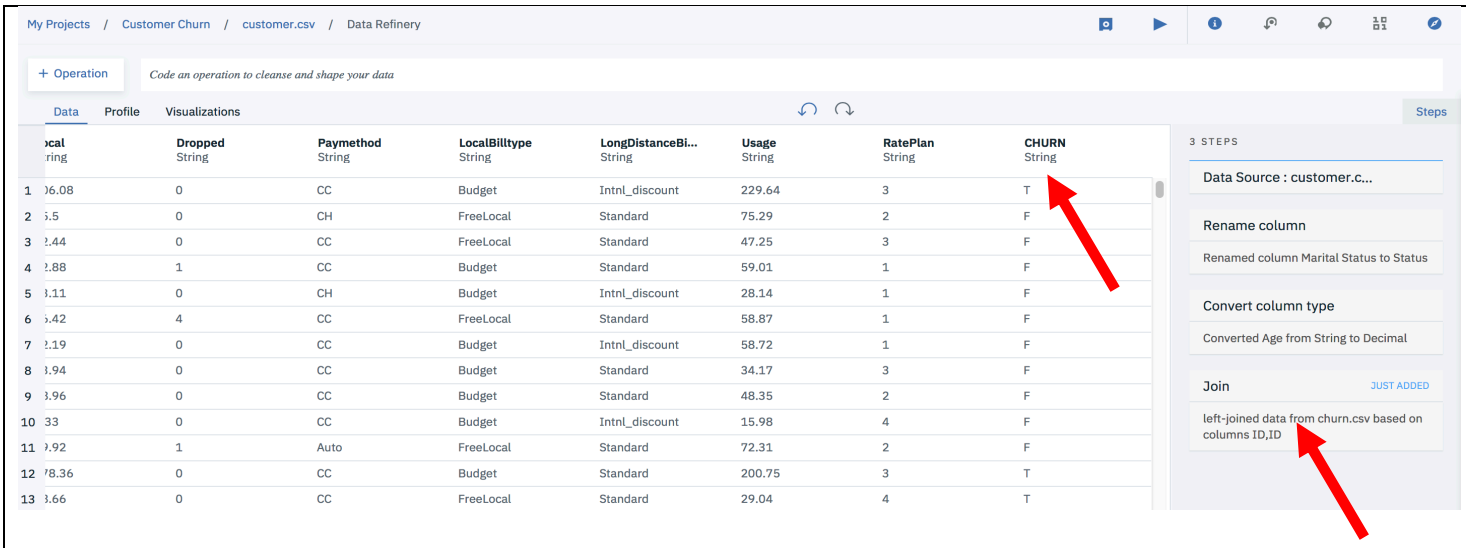


- Select **ID** as the Join Key and click **Next.** Select all columns to appear in the resulting data set and click **Apply.**

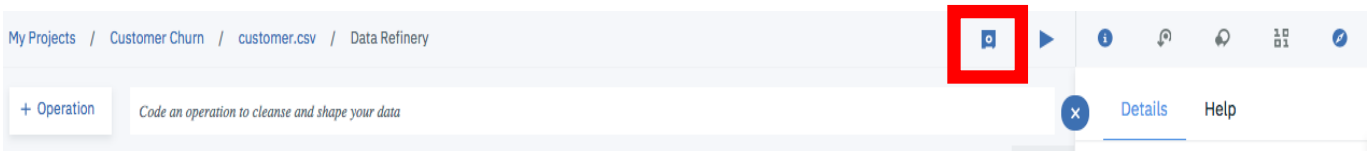




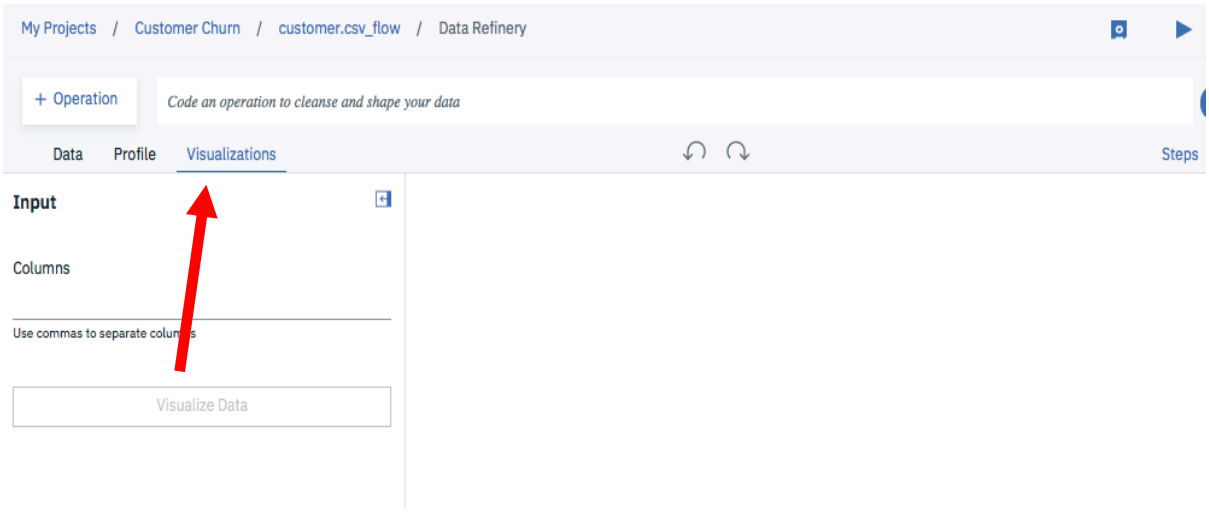
- The **Churn** column is now joined to the customer data set. And the **join operation** step is listed.



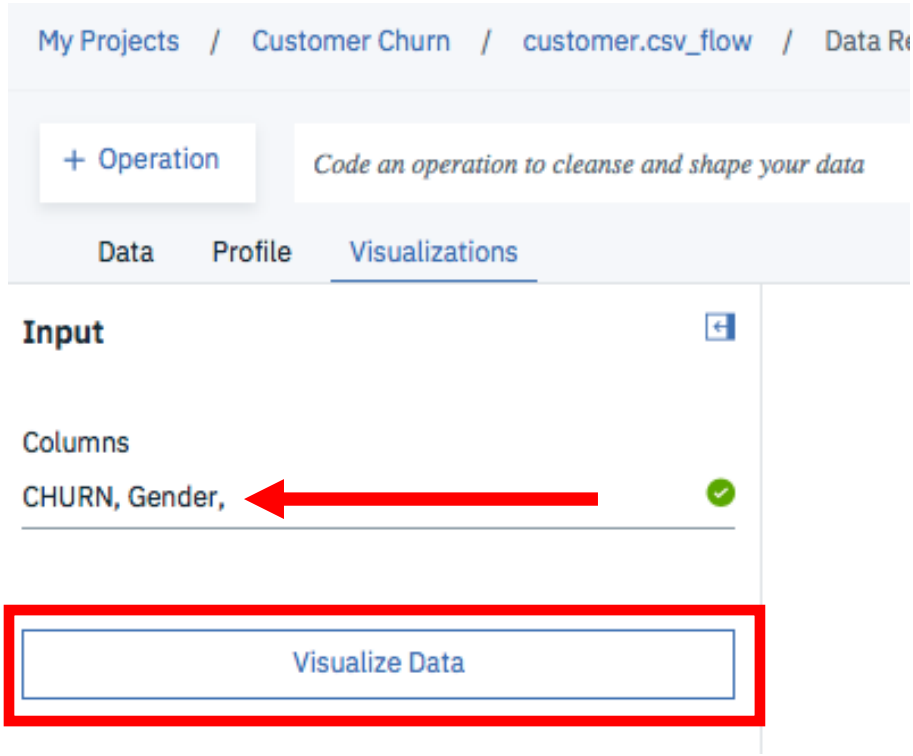
- Click the **Save icon** in the upper right to save the data flow



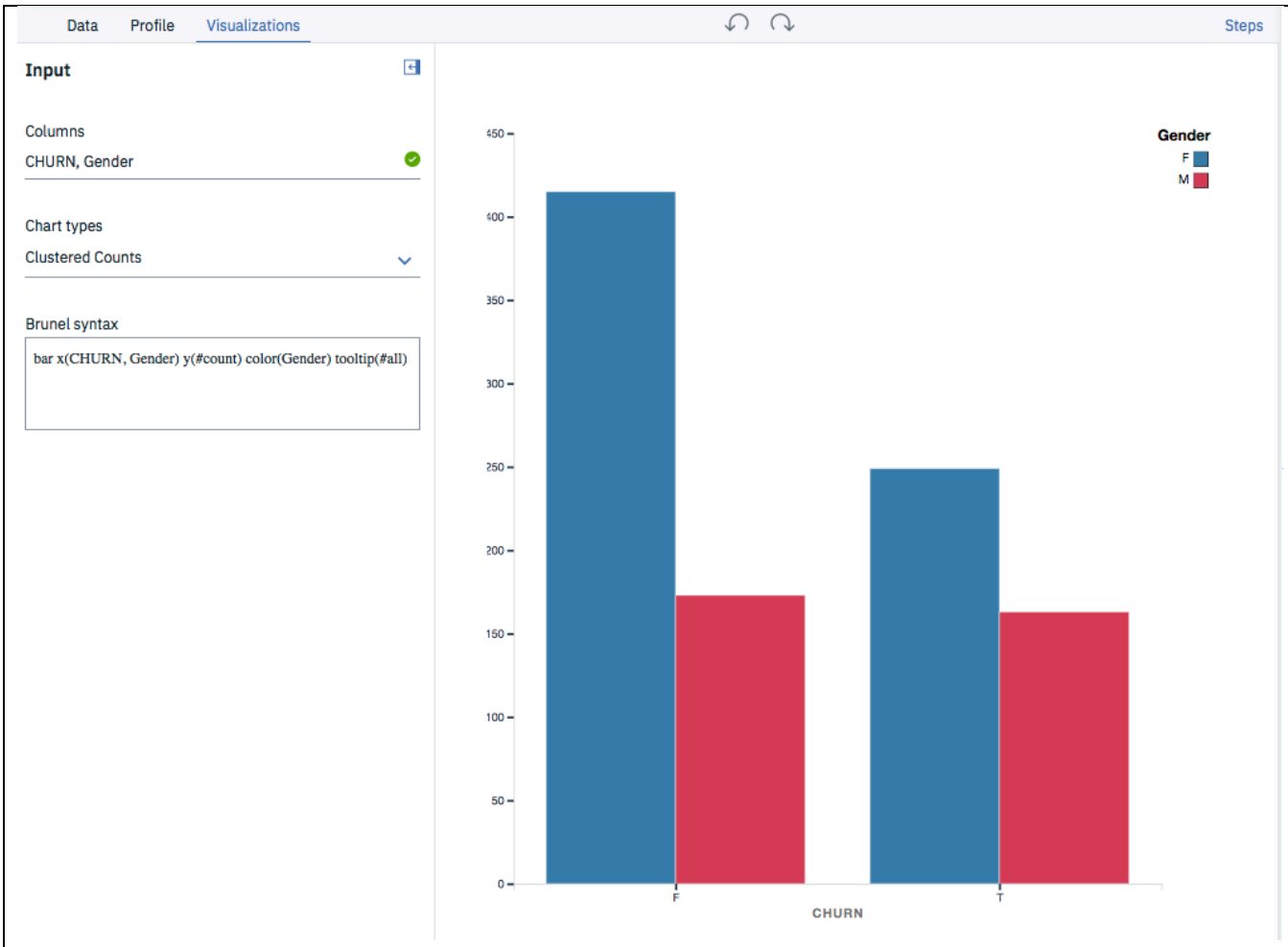
- Before we run the data flow, lets visualize the data. Select the **Visualizations** tab in the upper left



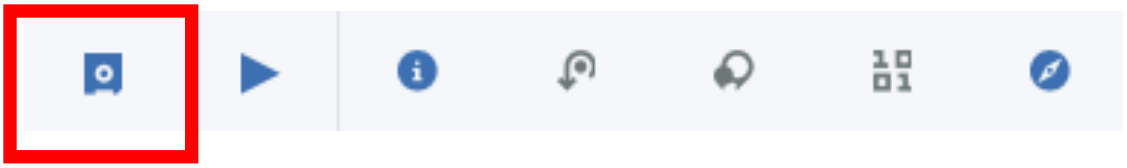
- Select **Churn** and **Gender** as the columns. Click **Visualize Data**



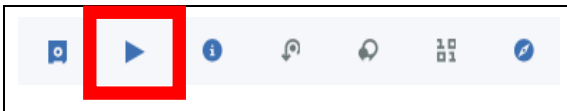
- Select **Clustered Counts** as the **Chart Types**. You should see the following. Notice the Brunel Syntax for the visualization is provided.

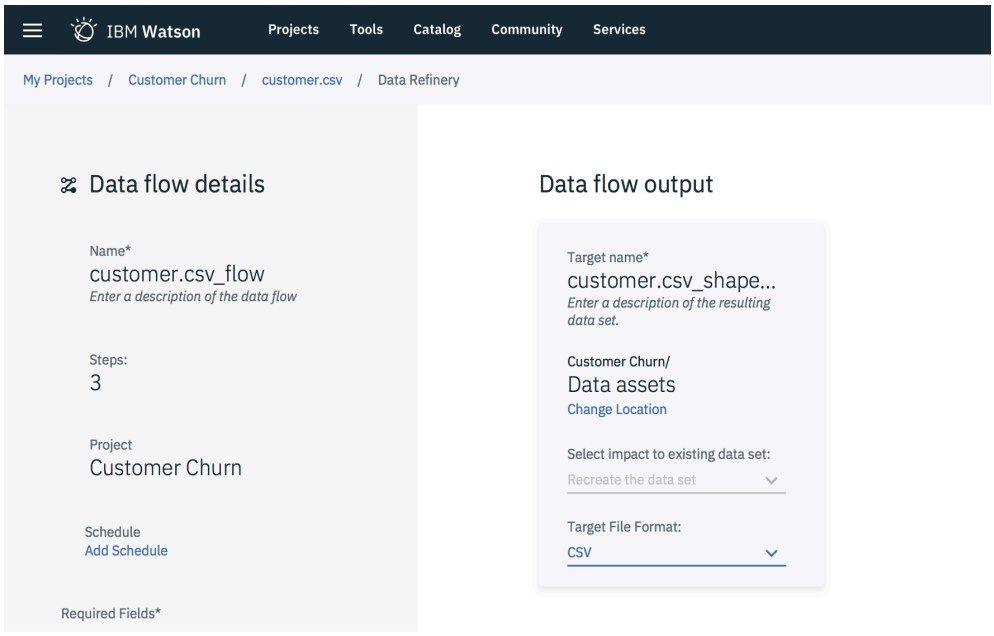


- Click the **Save icon** in the upper right to save the data flow

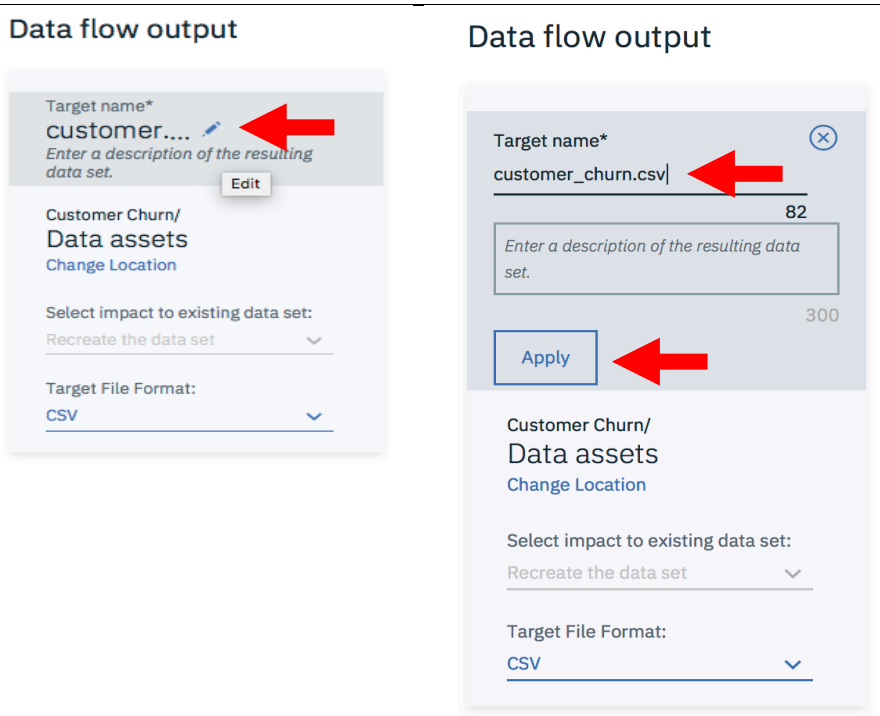


- Click the **Run icon** in the upper right to run to the data flow. You should see the following.

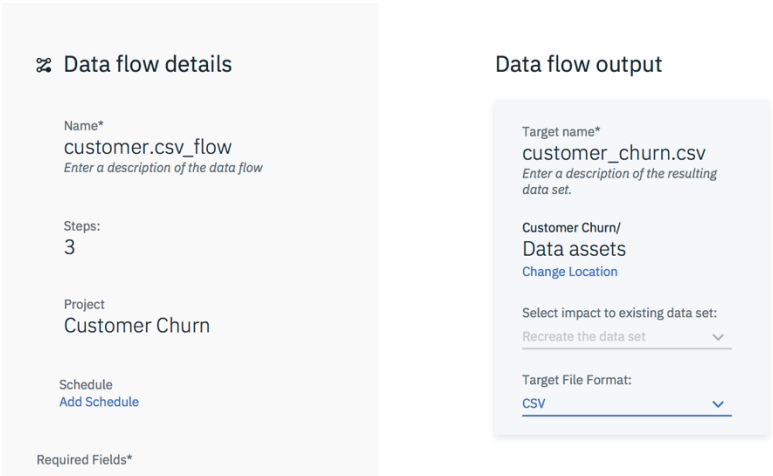




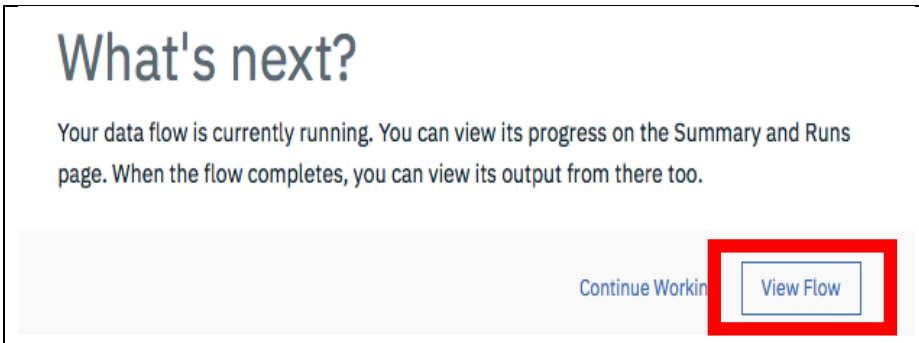
- Under **Data flow output**, hover over the **target name** until the edit pencil appears. Change the target name to **customer_churn.csv.** Click **Apply**.



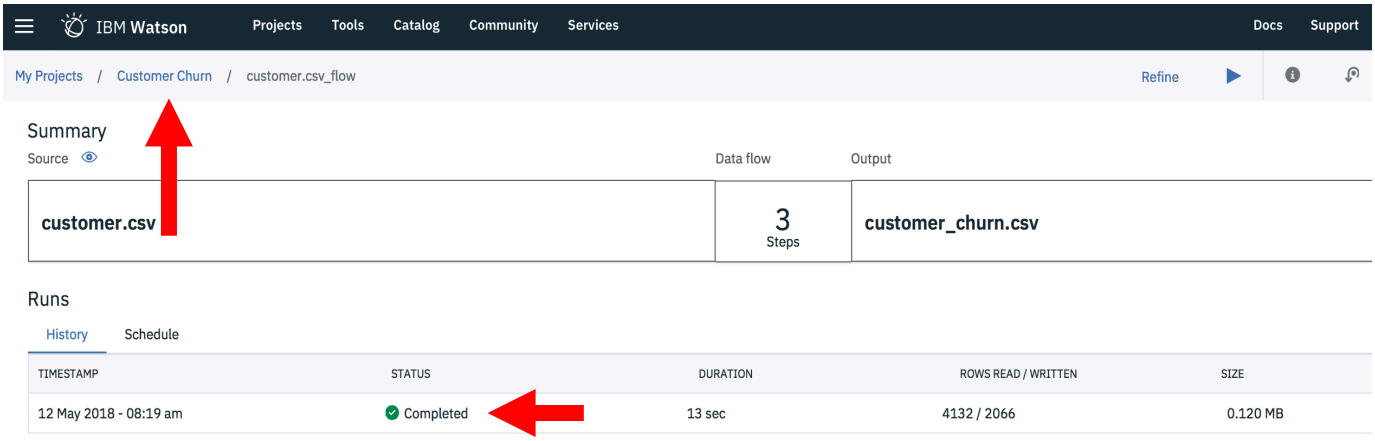
- Keep location as **Data assets**. You should now have the following. Click **Save and Run** in the lower right corner



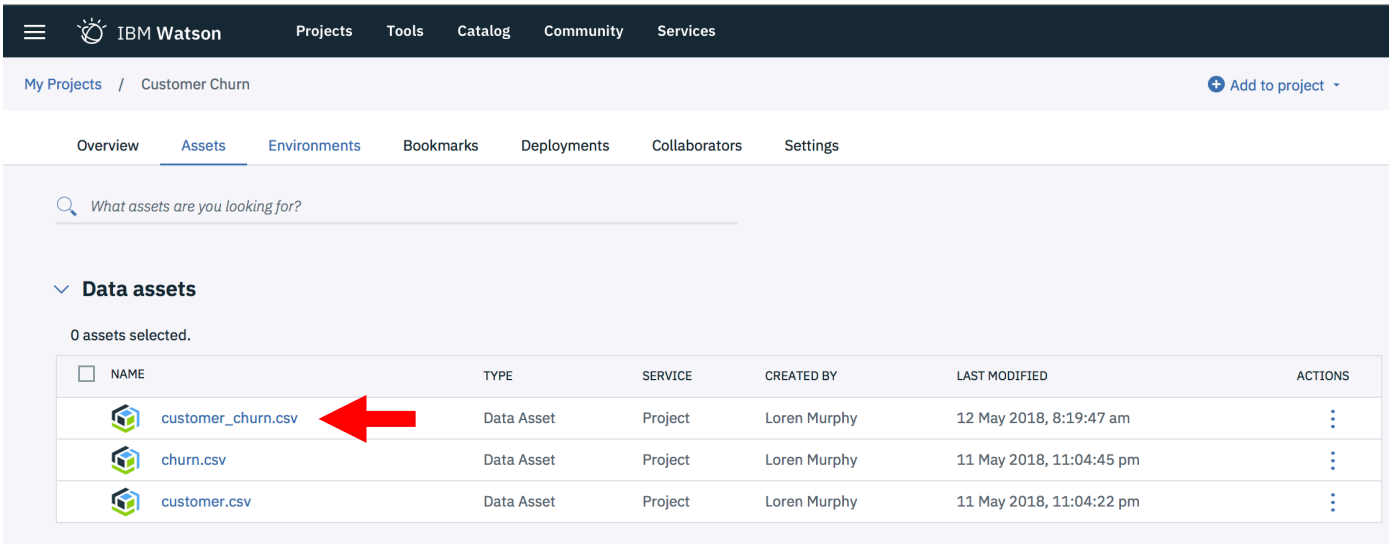
- Select **View Flow**, as the next step, to be redirected to the **Summary** page.



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



Lesson 3: Jupyter Notebook

Purpose:	This lesson introduces Jupyter notebooks and are used as part of a customer churn analysis using R.

Tasks:	Tasks you will complete in this lab exercise include: • Add Notebook Asset • Retrieve Data from External Repository • Predict Customer Churn using Machine Learning Techniques • Evaluate Model Performance

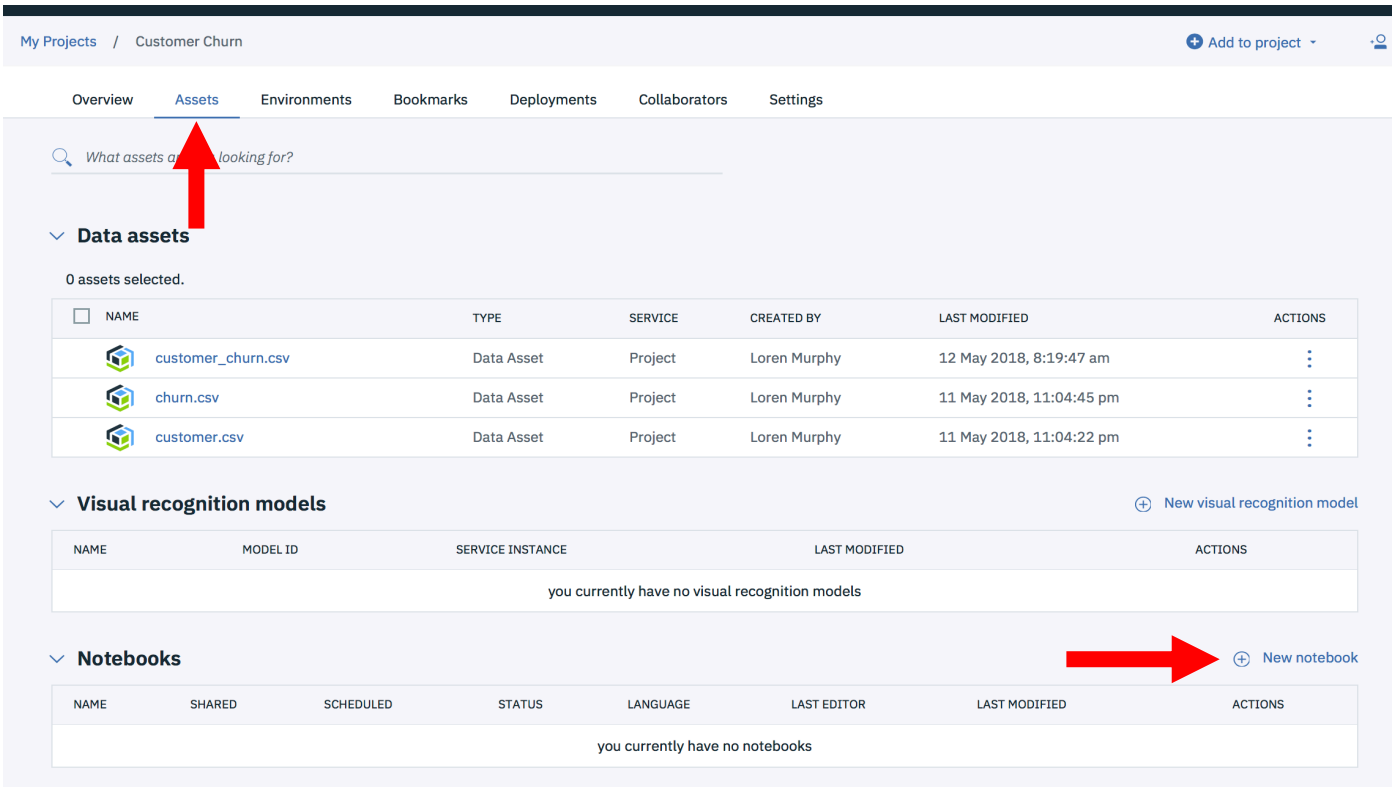
Lesson 3: Workflow Overview

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

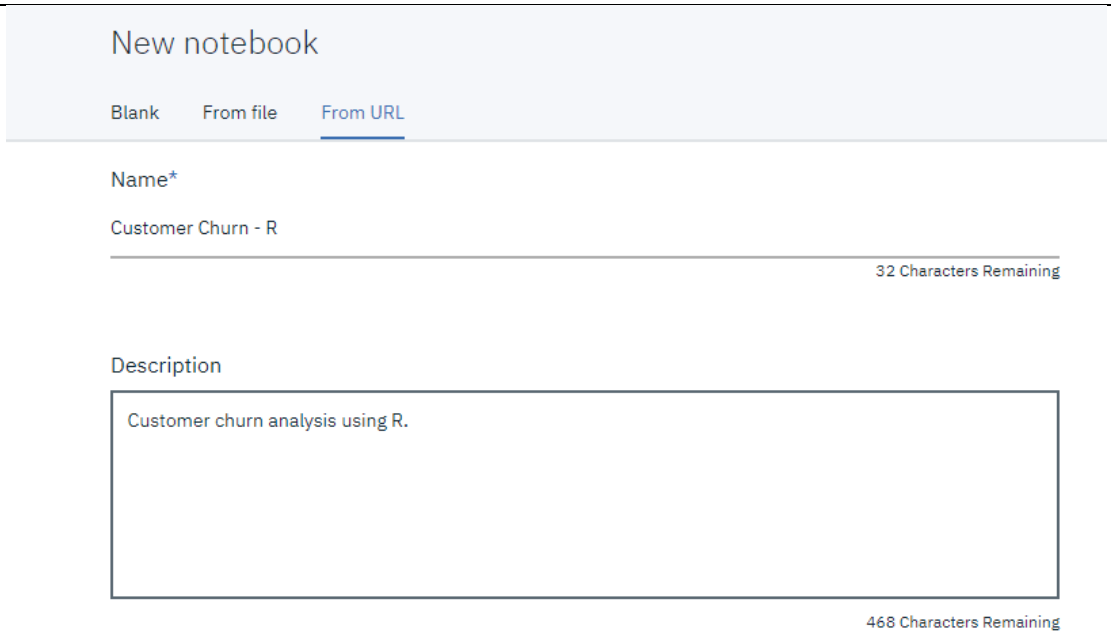
Lesson 3: Instructions

1. Create Notebook

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



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

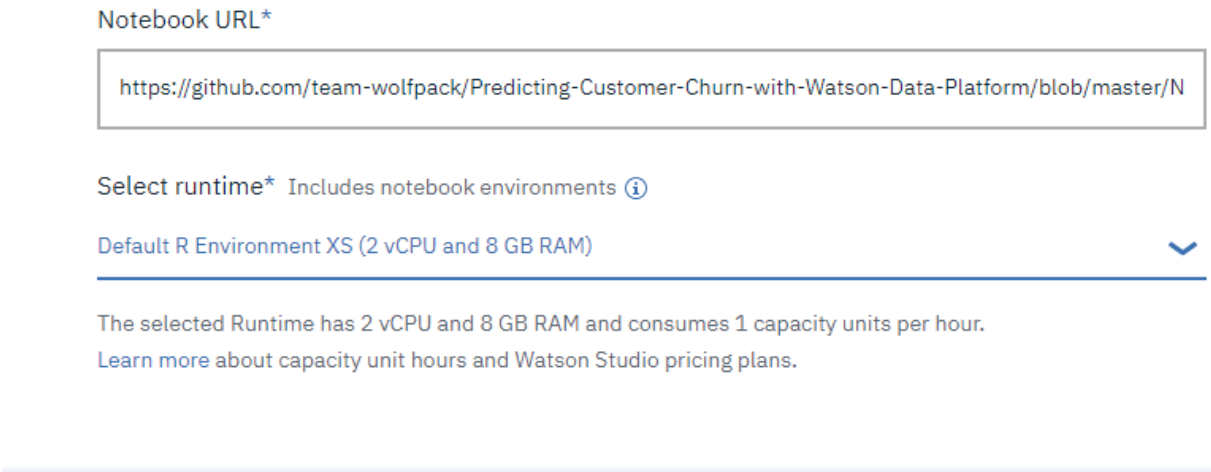


- In a separate browser window navigate to:
[Machine Learning for Customer Churn]

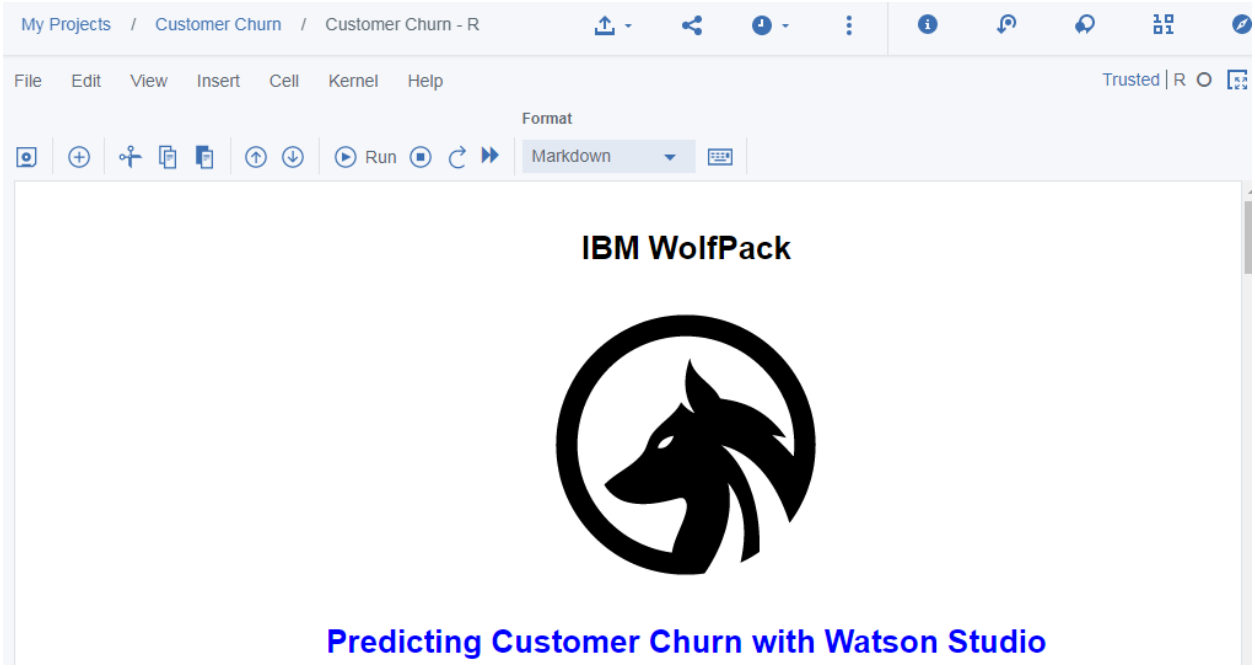
https://github.com/team-wolfpack/Wolfpack-Technical-Event-Content/tree/master/AI and Data Science/ Machine Learning for Customer Churn

- Click on **Notebooks**, right click on **Customer Churn-R.ipynb** then choose **Copy link address**. Go back to the **Watson Studio New Notebook** page.

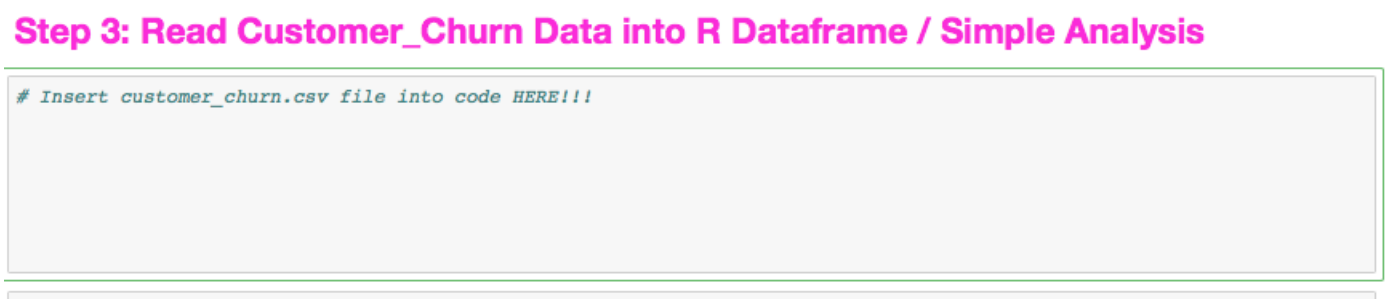
Paste URL into **Notebook URL** text box. Select **Default R Environment XS (2 vCPU and 8GB RAM))** as the runtime. Then click **Create Notebook**:



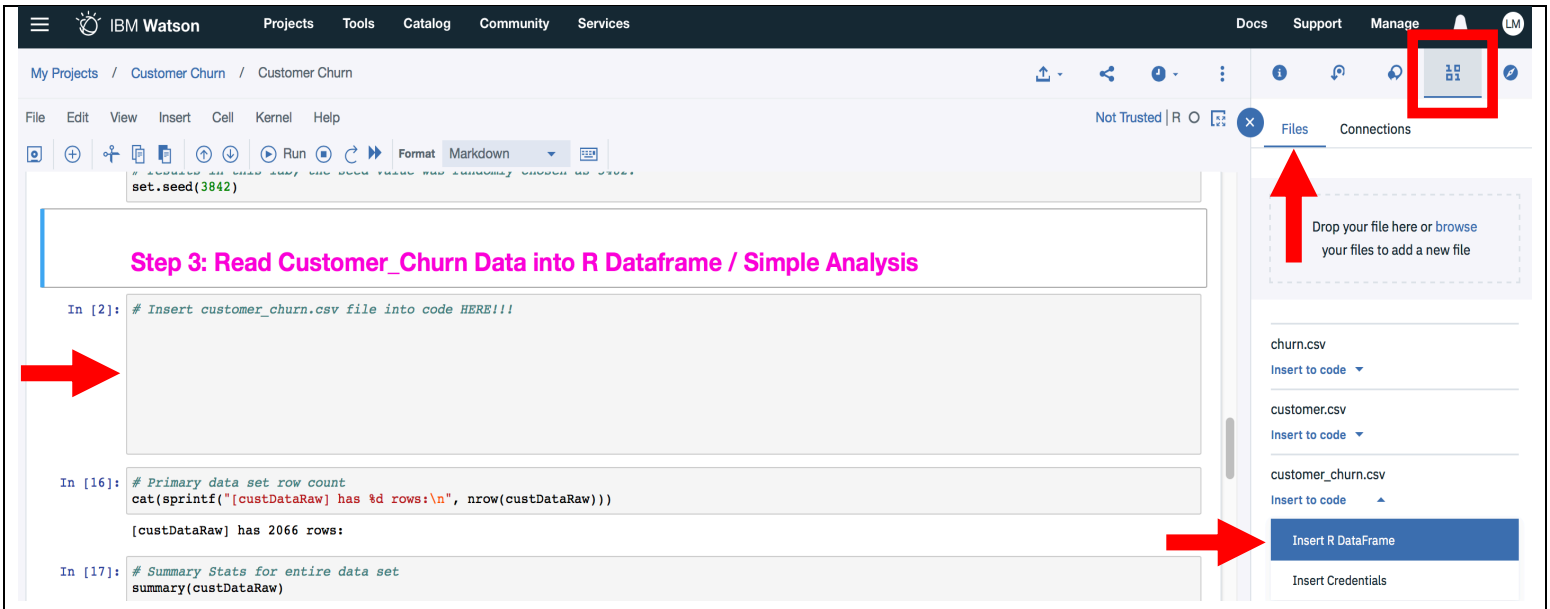
- You should now see:



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



- 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**.



- 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" = "d0594edbaba71cafcff929835fafa08a1030bdc1f457a4ab")
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_N6zH3vK5T46YBzOyFOx51w.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" = "d0594edbaba71cafcff929835fafa08a1030bdc1f457a4ab")
url <- "s3-api.us-geo.objectstorage.service.networklayer.com"
bucket <- "customerchurn-donotdelete-pr-sb12ylmzd0umqm"
headers <- list(`x-amz-content-sha256`="e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855")

obj <- s3HTTP(
    verb = "GET",
    bucket = bucket,
    headers = headers,
    path = "data_asset/customer_churn_N6zH3vK5T46YBzOyFOx51w.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	Paymethod	LocalBilltype	LongDistanceBilltype	Usage	RatePlan	CHURN
1	F	S	1	38000.00	N	24.39333	23.56	0	206.08	0	CC	Budget	Intnl_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	Intnl_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: • Create Machine Learning Flow • Import Data • Leverage Flows' Palette to Orchestrate Customer Churn Machine Learning Pipeline • Evaluate Customer Churn Model

Lesson 4: Workflow Overview

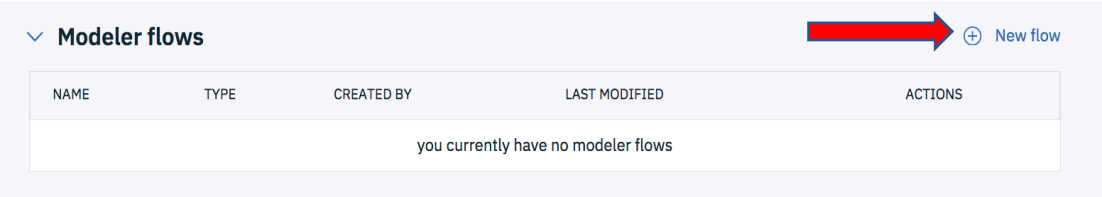
1. Create Machine Learning Flow
2. Add & Configure Partition and Type Object
3. Add & Configure Model Objects
4. Run Flow to Create Nuggets
5. Add & Configure Analysis Object - Measure Performance
6. Add Second Model Technique to Flow

Lesson 4: Instructions

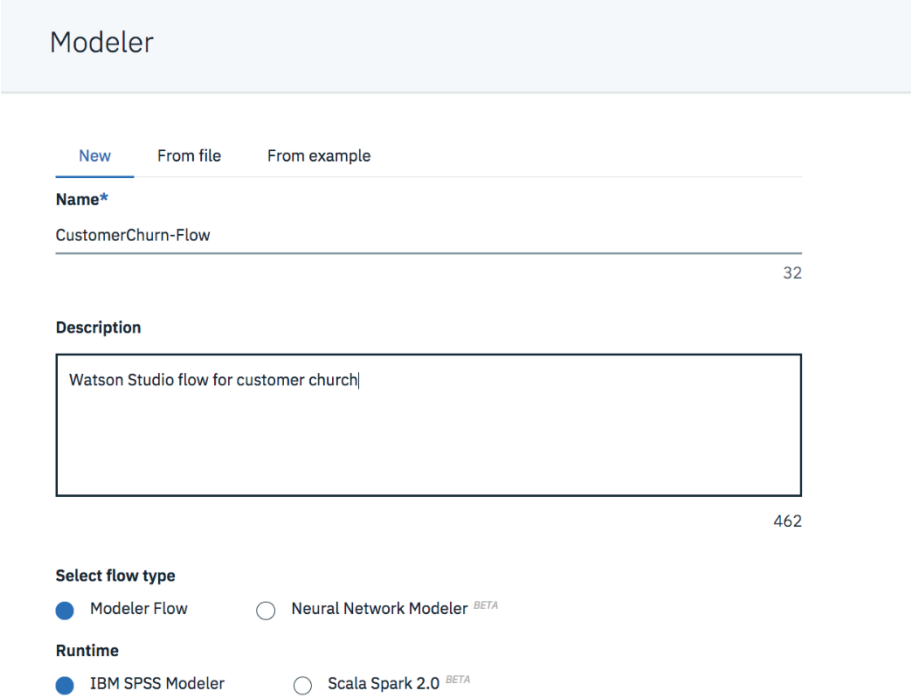
Action

1. Create Machine Learning Flow

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



- 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**":



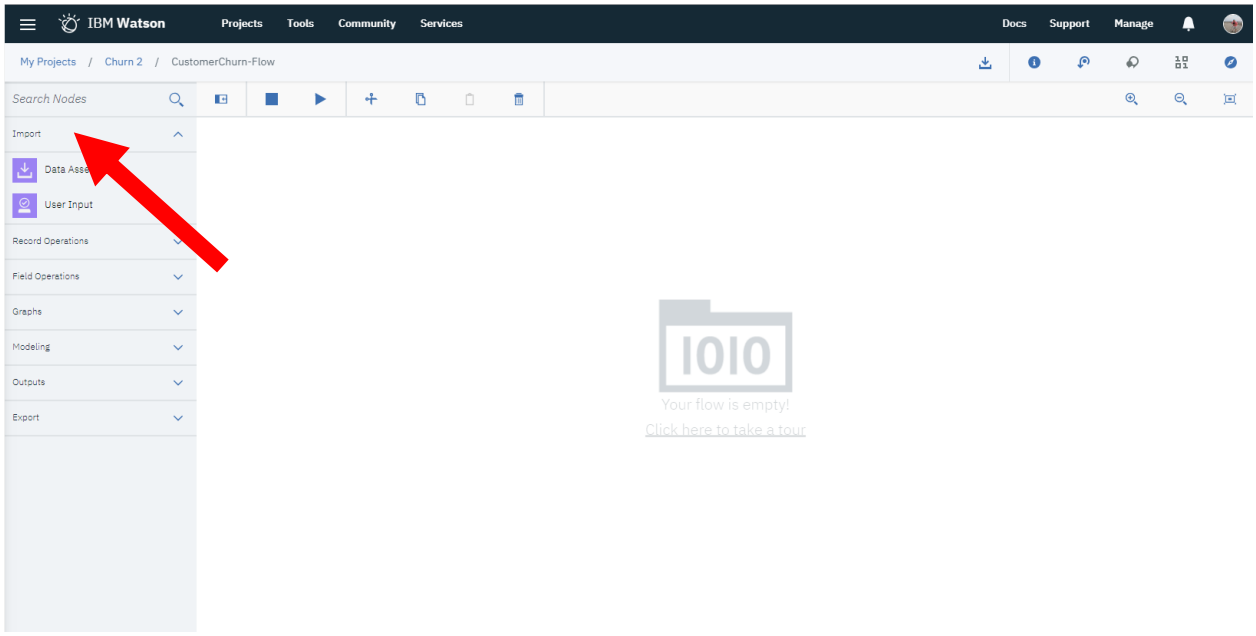
- Click on "**Create**"

Action

3. Add Data Asset

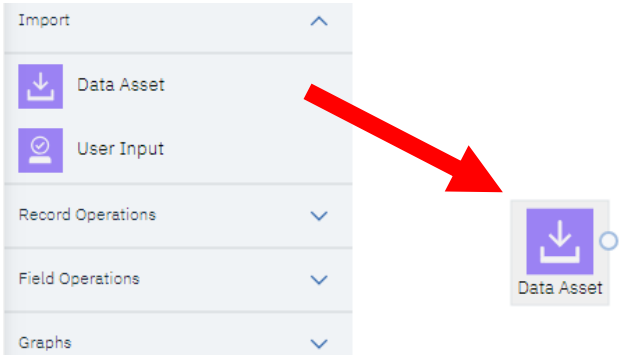
You should now see an empty workspace.

- On the top left click on the "**Palette**" icon, then click on the "**Import**" the icon.



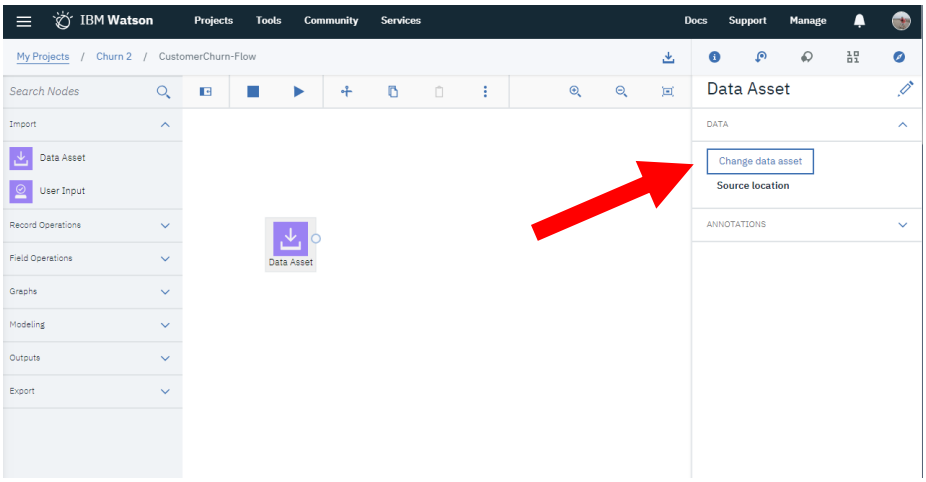
The palette represents the set of tools available for use with Watson Studio flows. The menu of the right should look familiar.

- Let's start by dragging and dropping the "**Data Asset**" node onto the workspace.

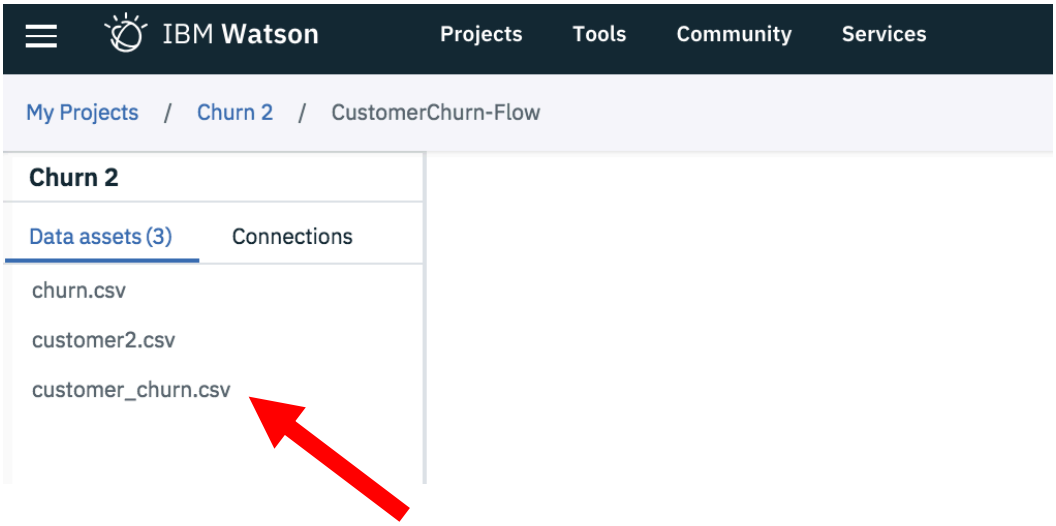


Action

- 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”.

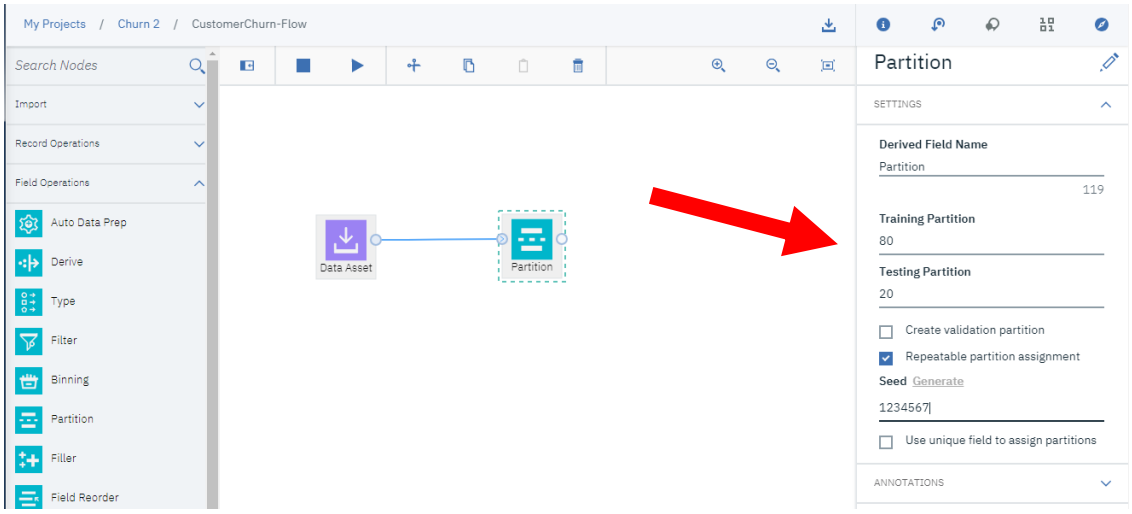


Action

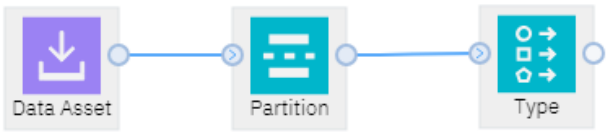
3. Add & Configure Partition and Type Object

- 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.

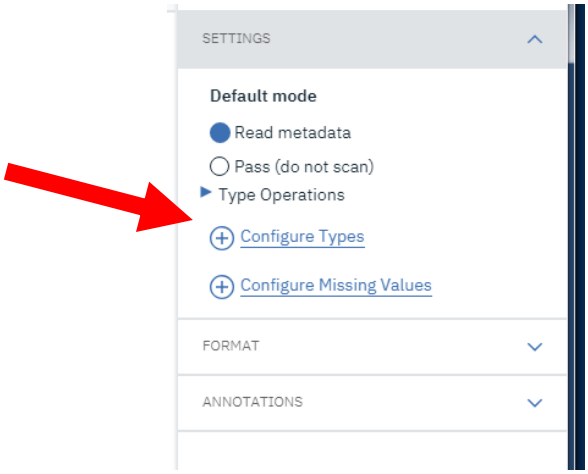


- Next, drag and drop a **"Type"** node onto the canvas. Connect it to the **"Partition"** node.

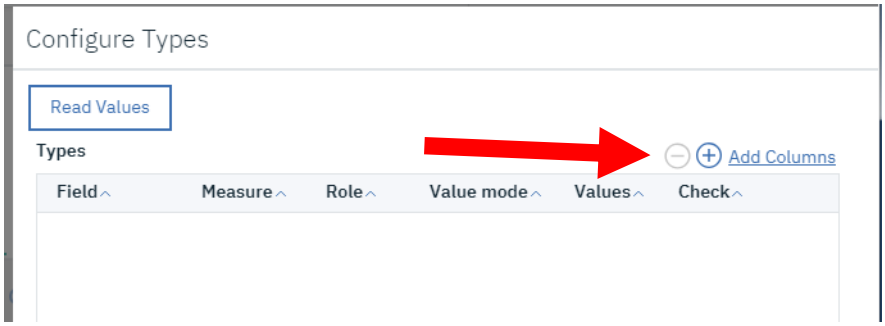


- Double click on **"Type"**. On the top right click on **"Configure Types"**:

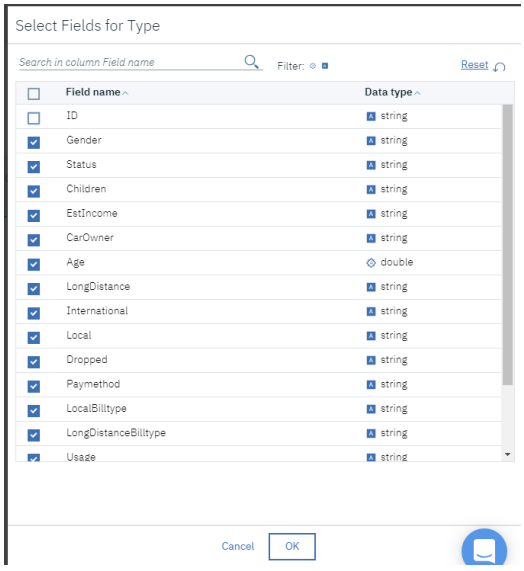
Action



- Click on **"Add Columns"**:

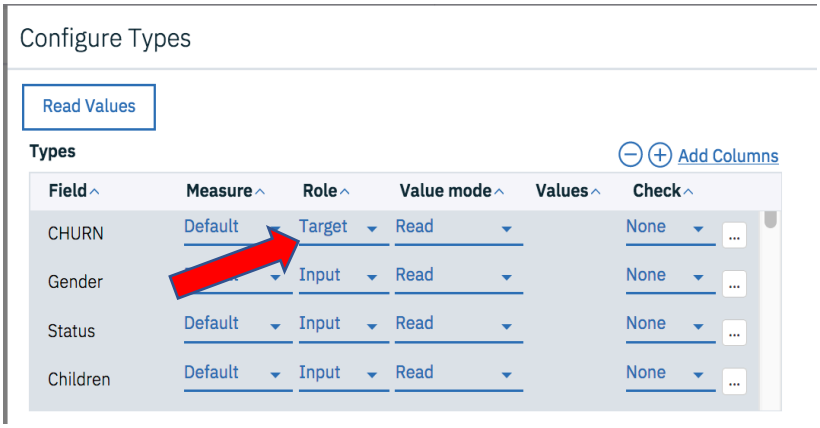


- Choose all of the fields except for "ID" then click on OK.



Action

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

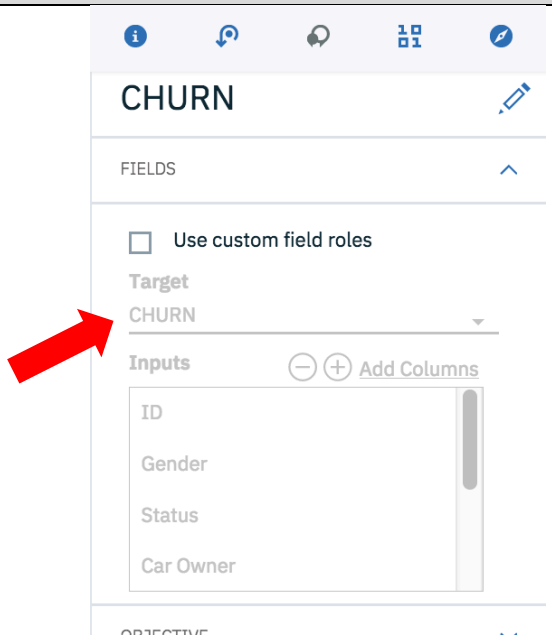


- Click **"OK"**.
- Click **"Save"** to exit

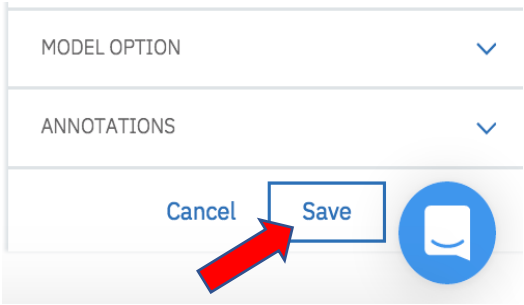
4. Add & Configure Model Object

- 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



- Click on "**Save.**"

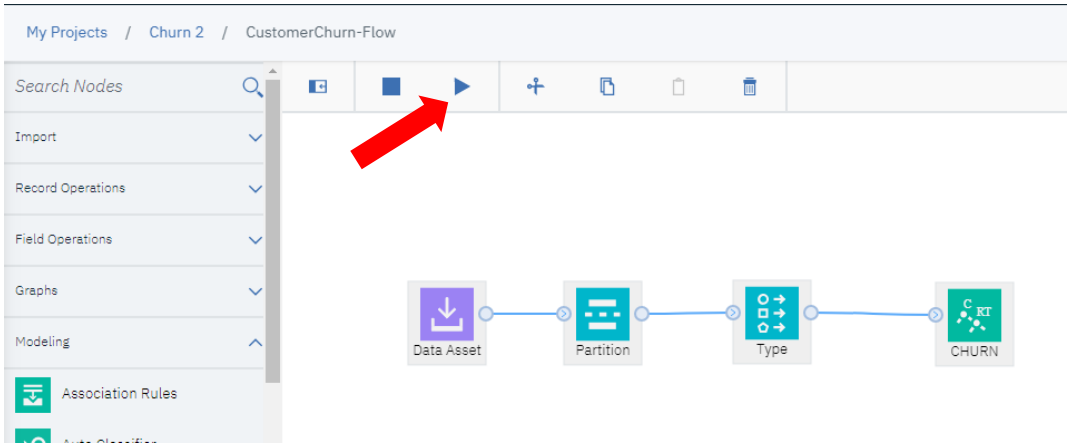


- Your palette should resemble this:

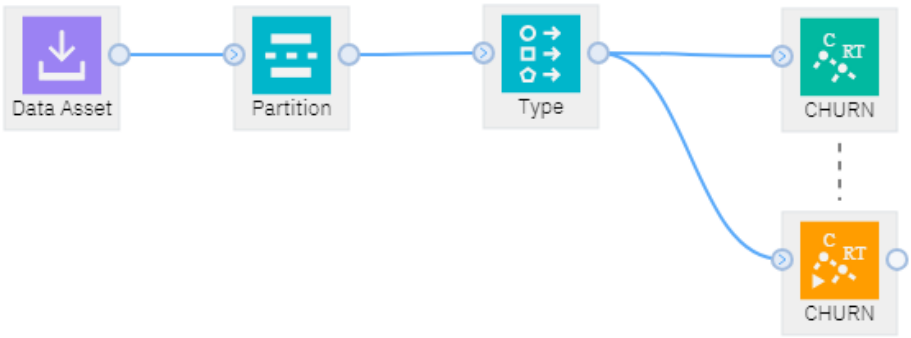
Action

5. Run Flow to Create Nugget

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



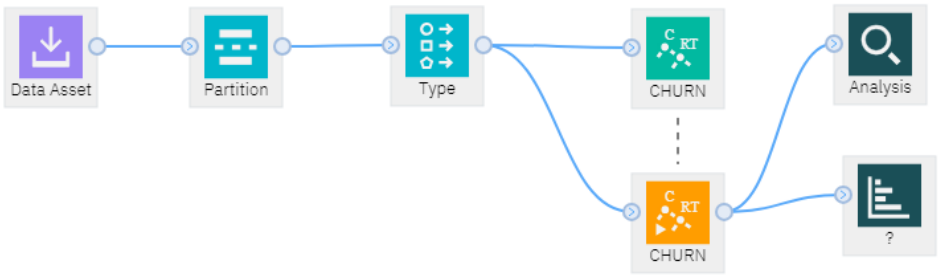
You should see a new forth object on the workspace, this is called a nugget.



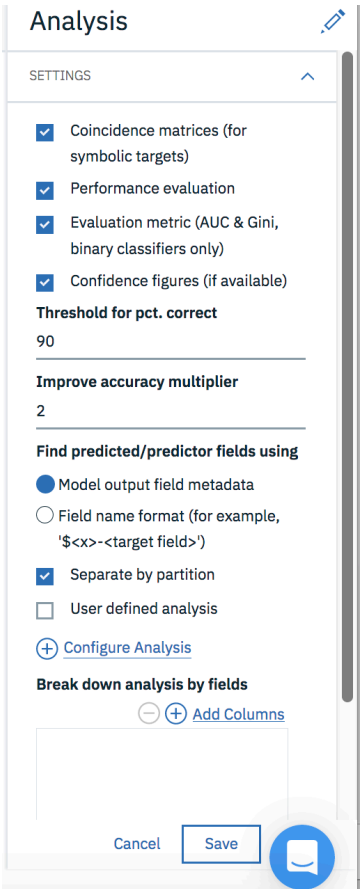
6. Add & Configure Analysis Object – Measure Model Performance

- 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

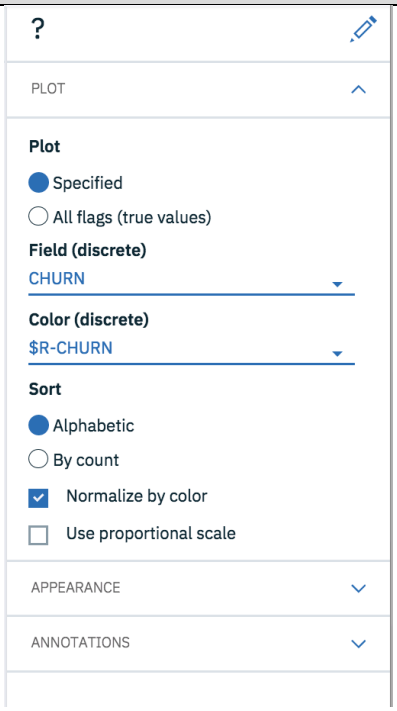


- Double click on "**Analysis**" and check off the four checkboxes, leave the rest as default:

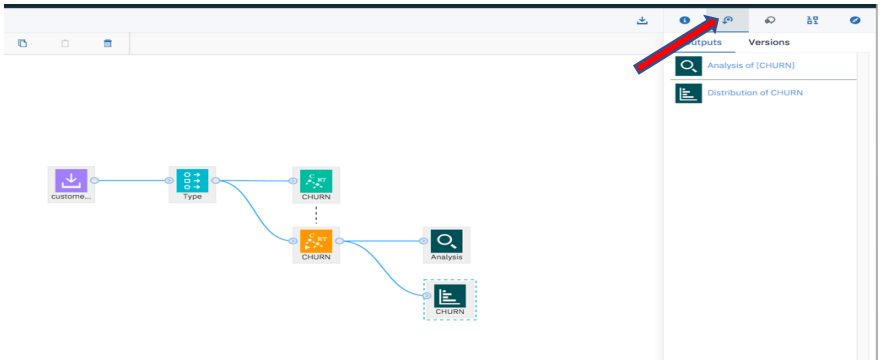


- Click **"Save"** to return to the workspace.
- Double click on the **"Distribution"** object and configure it as depicted below:

Action



- 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

End of Lesson 4

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:	Tasks you will complete in this lab exercise include: • 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

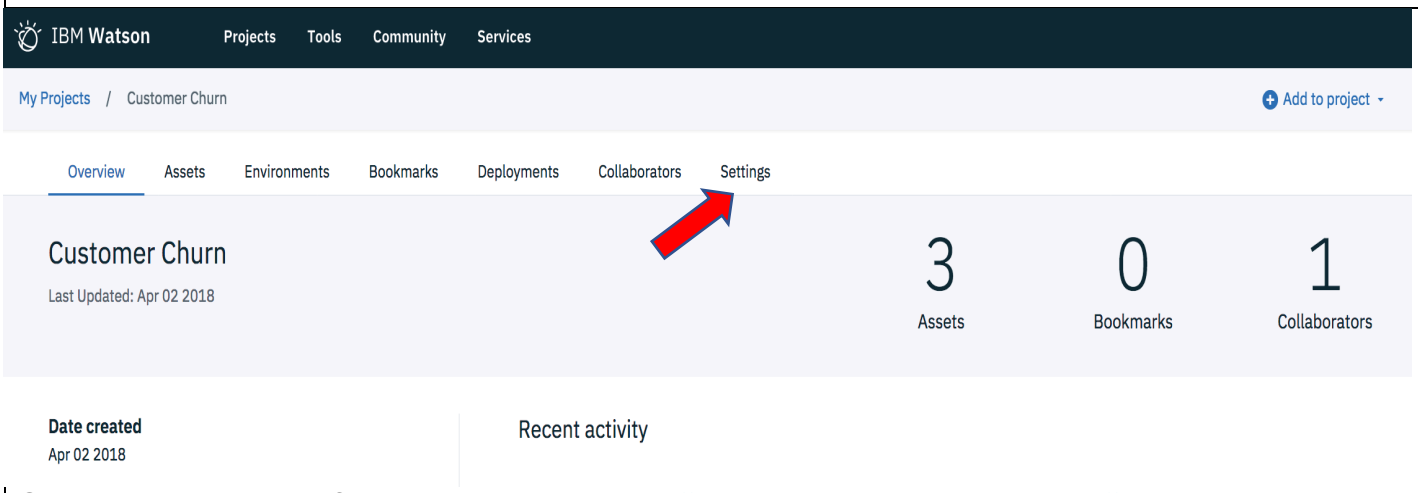
1. Create Machine Learning Service
2. Create Machine Learning Model
3. Choose Modeling Technique
4. Add Estimators
5. Evaluate Models
6. Save & Deploy Model
7. Predict with Model

Lesson 5: Instructions

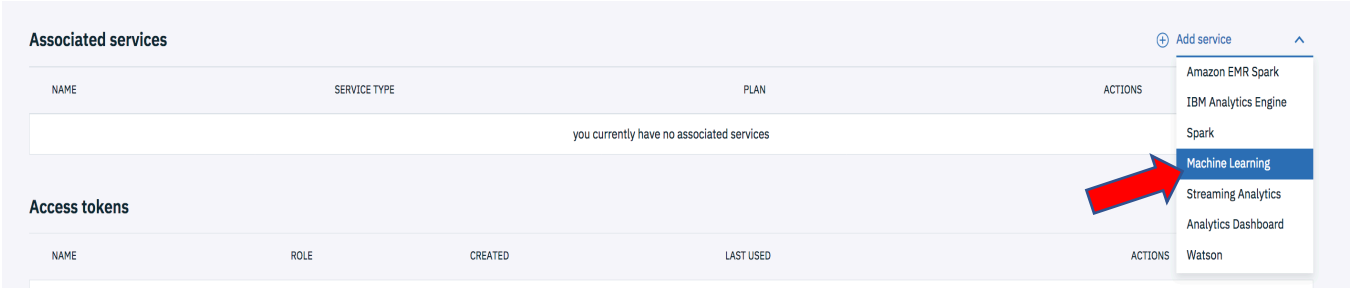
Action

1. Create Machine Learning Service

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

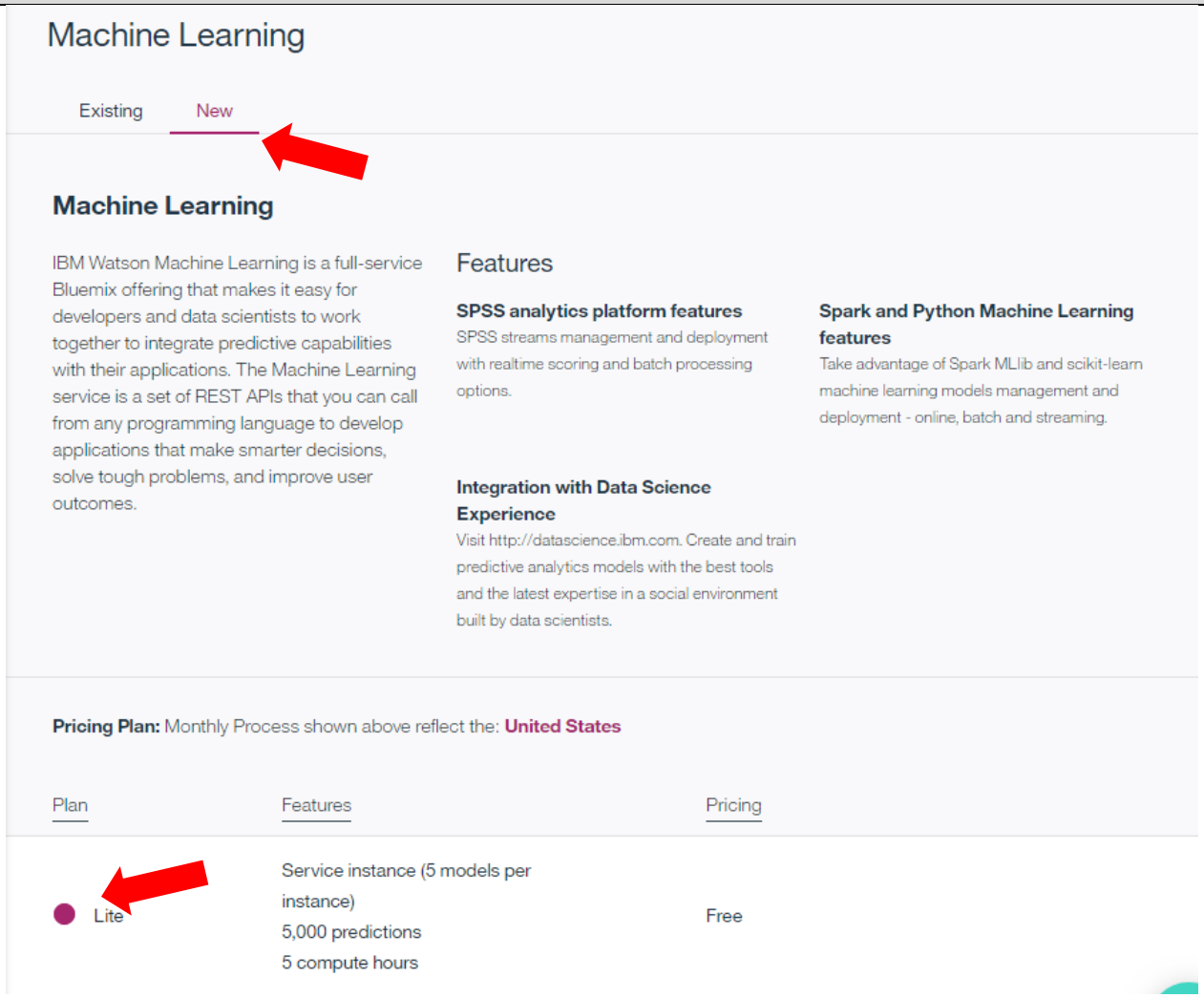


Scroll to the middle of the page and click on **"Add service"** then choose **"Machine Learning":**



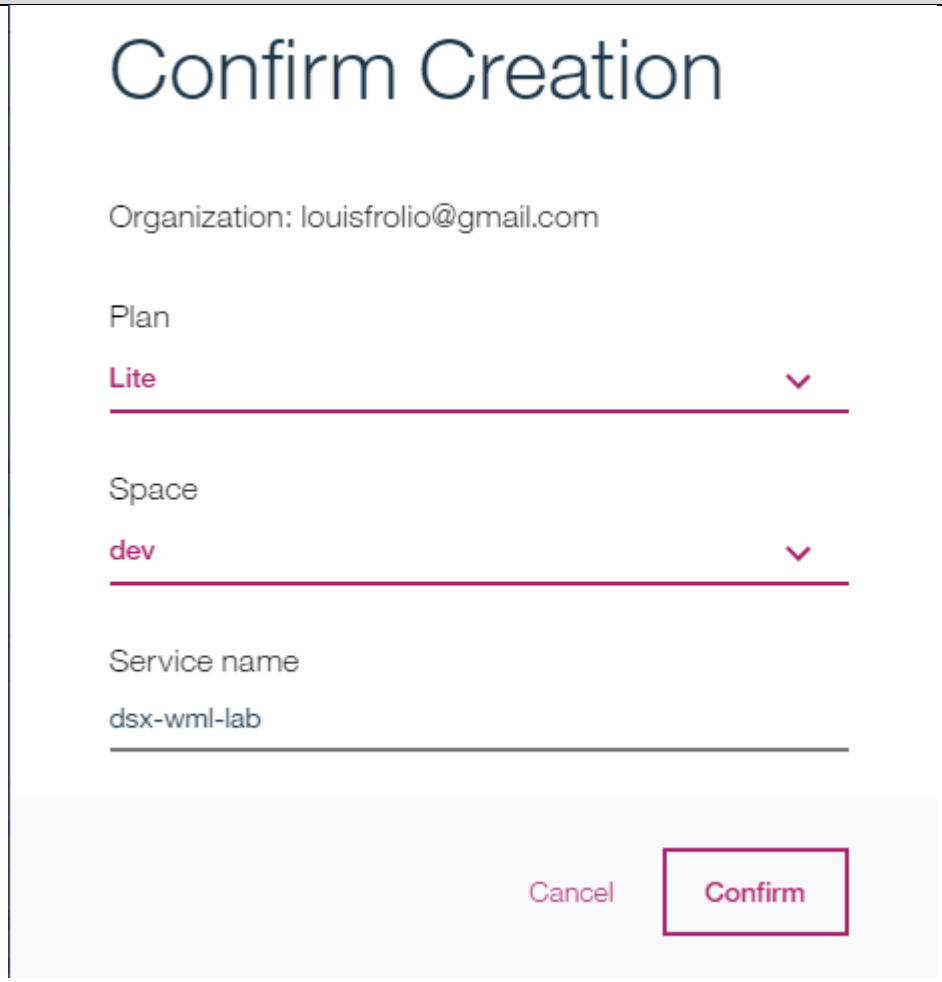
- On the Machine Learning page make sure that the tab is set to **"New"**, for the plan choose **"Lite":**

Action



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

Action

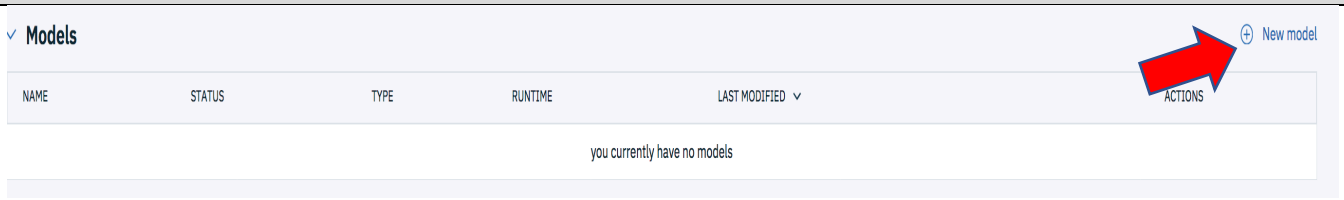


- Click **"Confirm"** to create Watson Machine Learning Service.

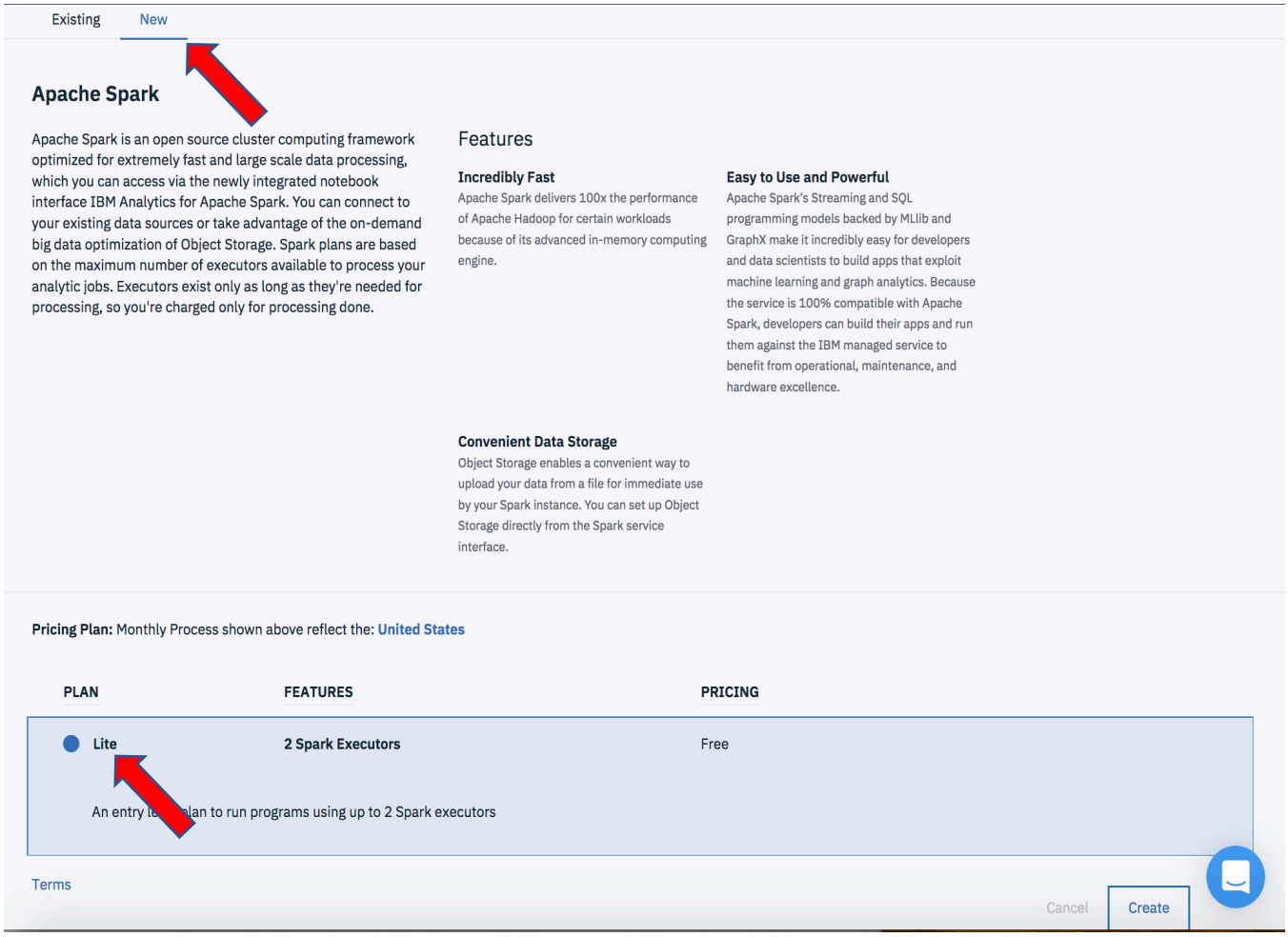
2. Create Machine Learning Model

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

Action



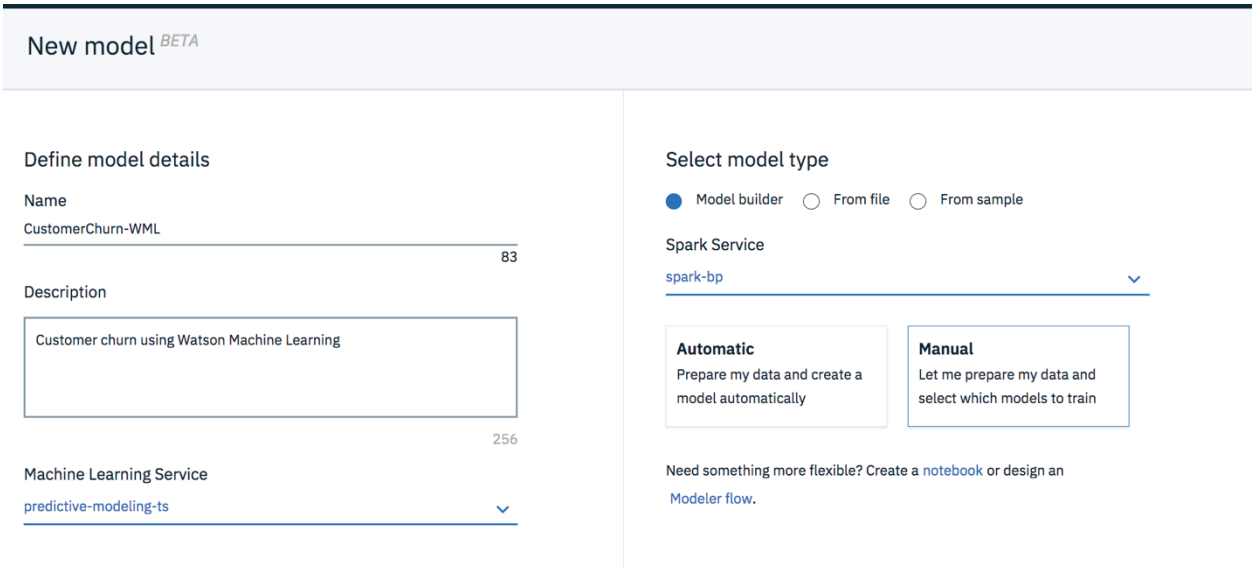
- In the “New model” window, associate an **IBM Analytics for Apache Spark instance.** On the Machine Learning page make sure that the tab is set to **“New”**, for the plan choose **“Lite”.** Select **Create.**



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

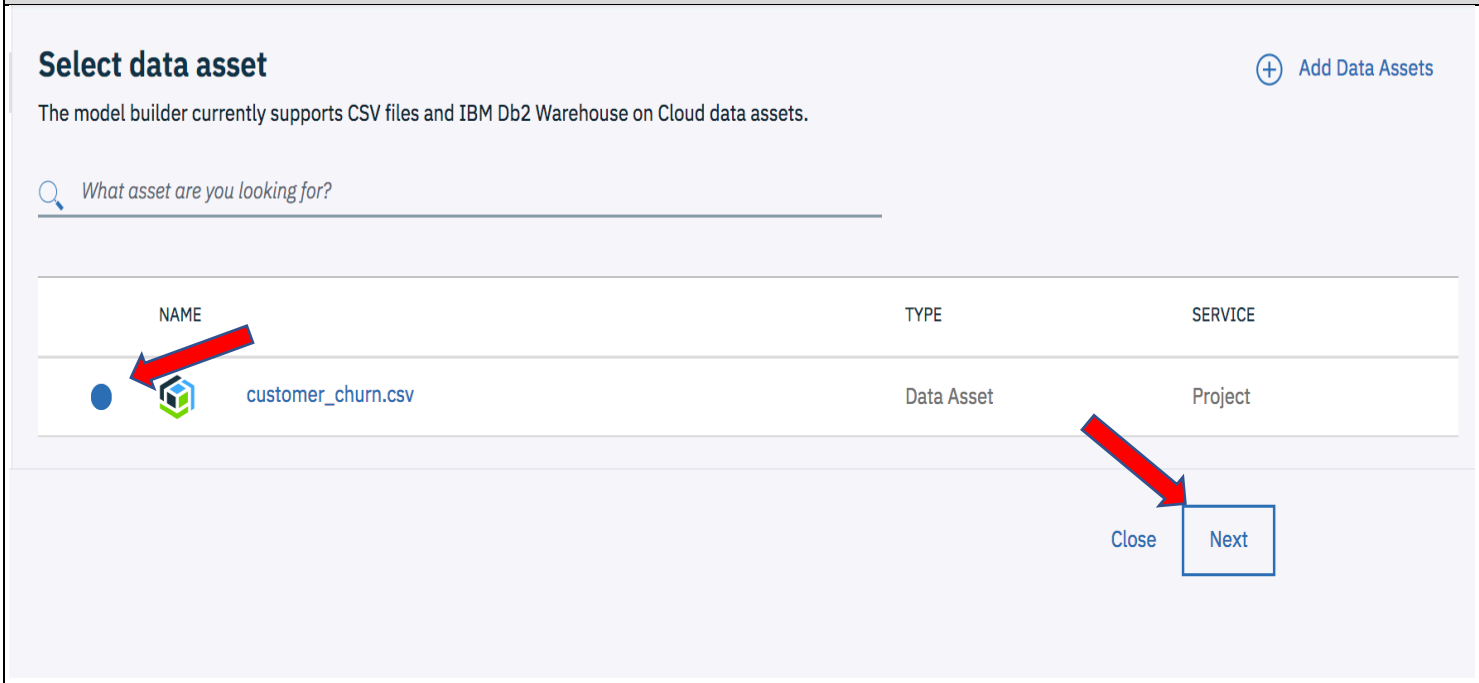
Action

- 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.**”

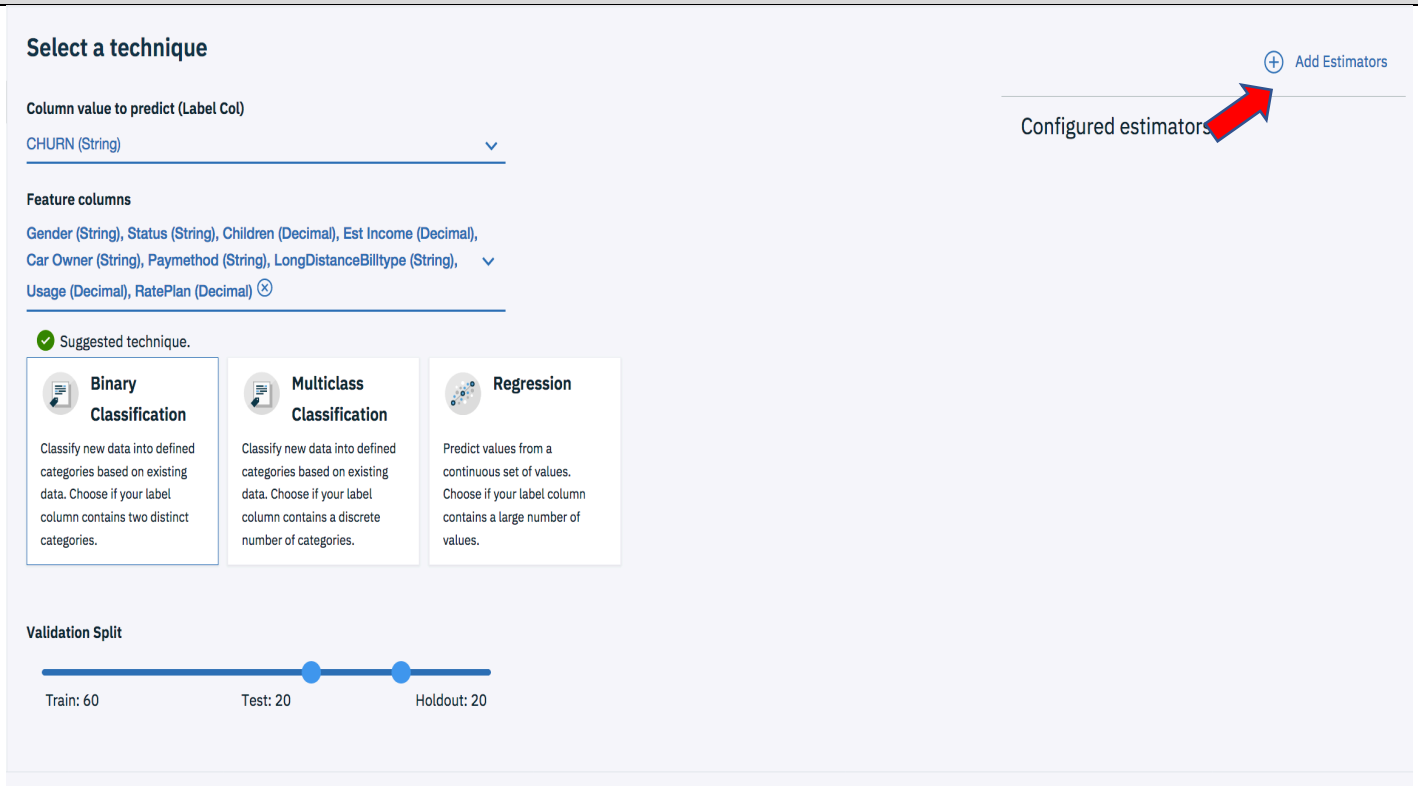
Action



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.

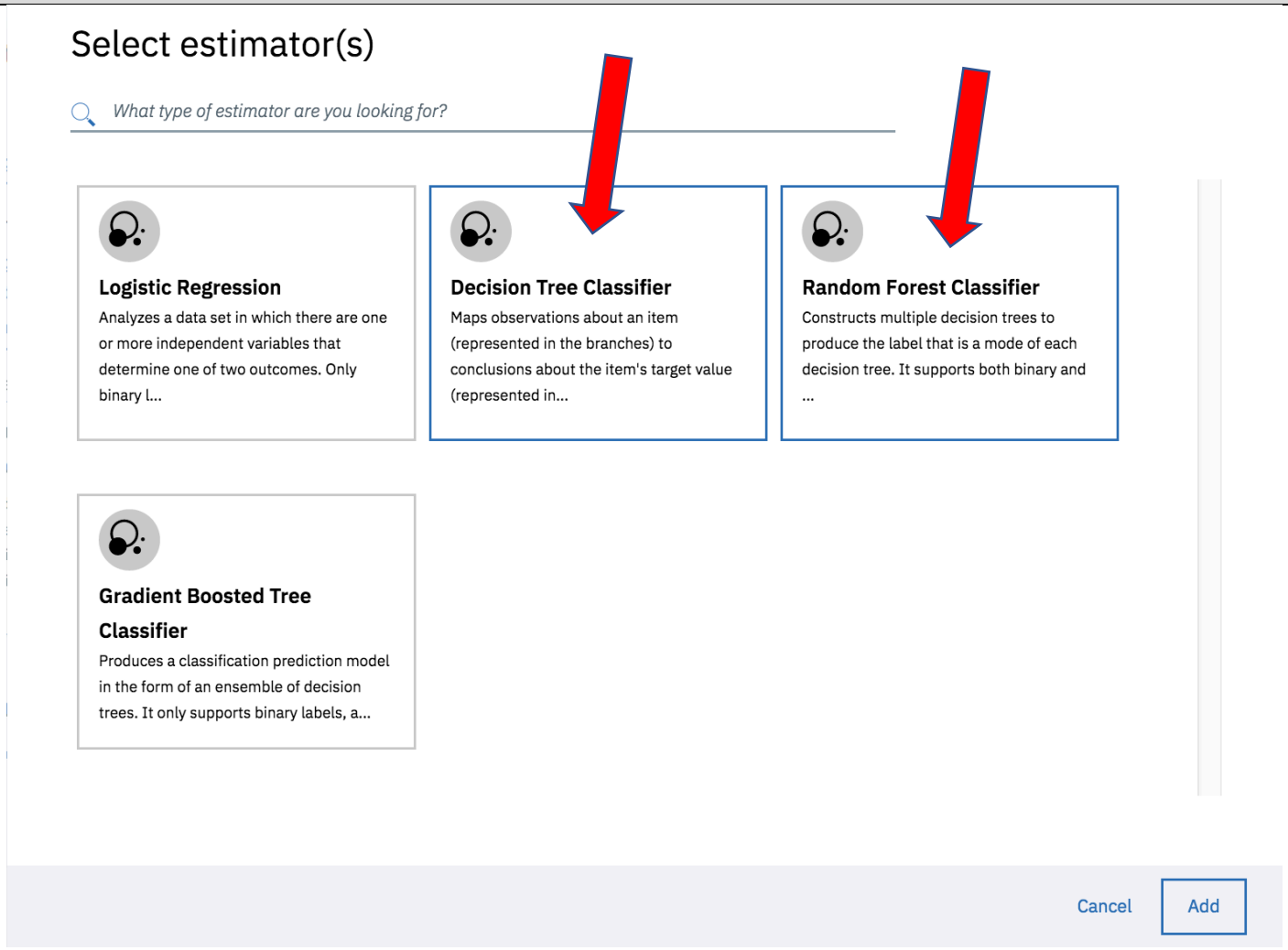
Action



4. Add Estimators

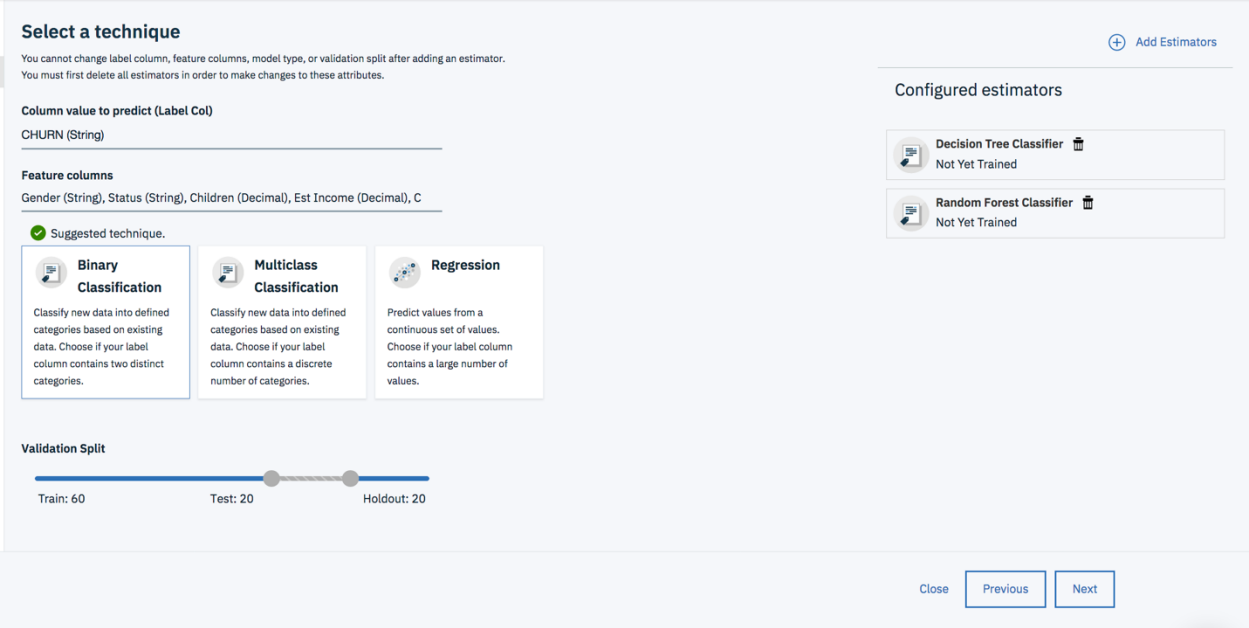
- In the upper right-hand corner of the screen you will see "**Add Estimators**", click on the icon. In the "Select estimator(s)" screen choose **Decision Tree Classifier**, and **Random Forest Classifier**:

Action



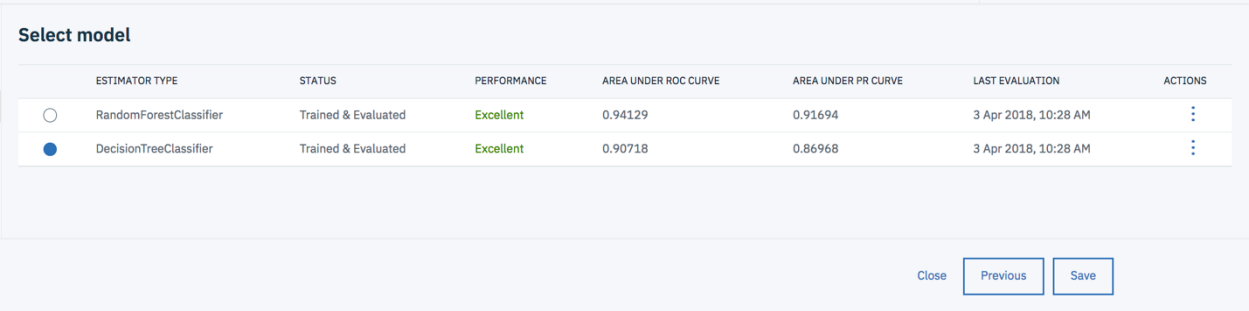
- Click **"Add"**

Action



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

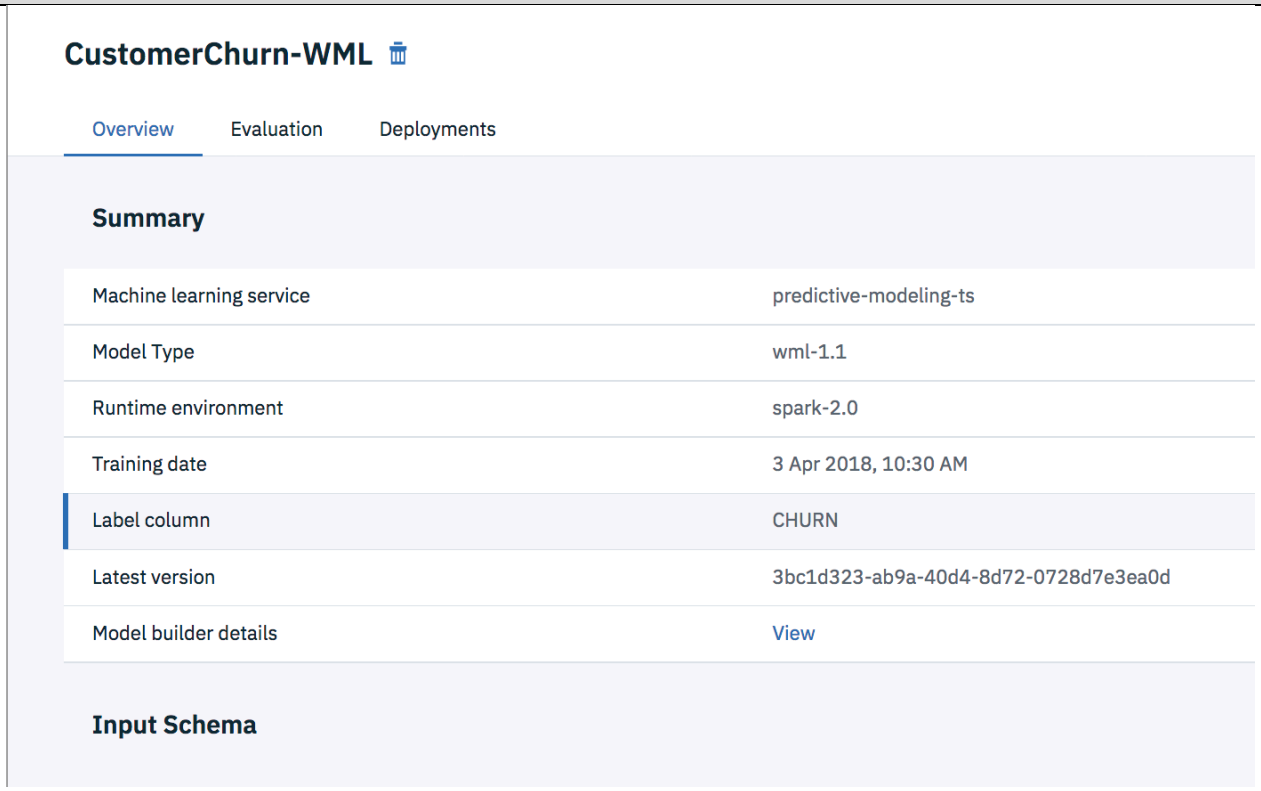
5. Evaluate Models



6. Save & Deploy Model

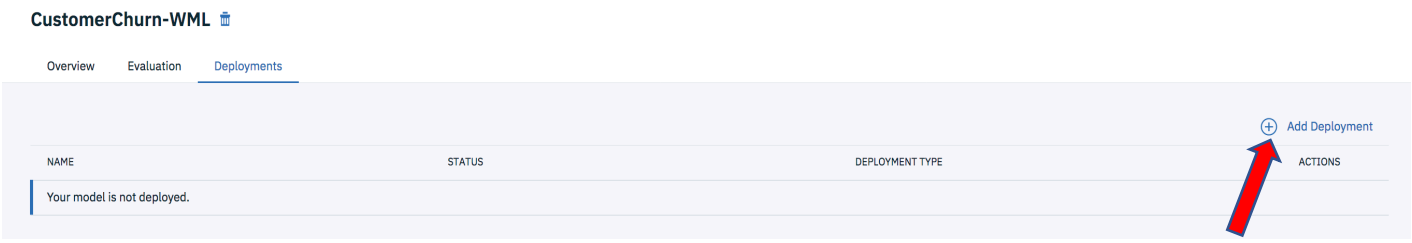
- Pick which model you want to keep then click **"Save:"**

Action



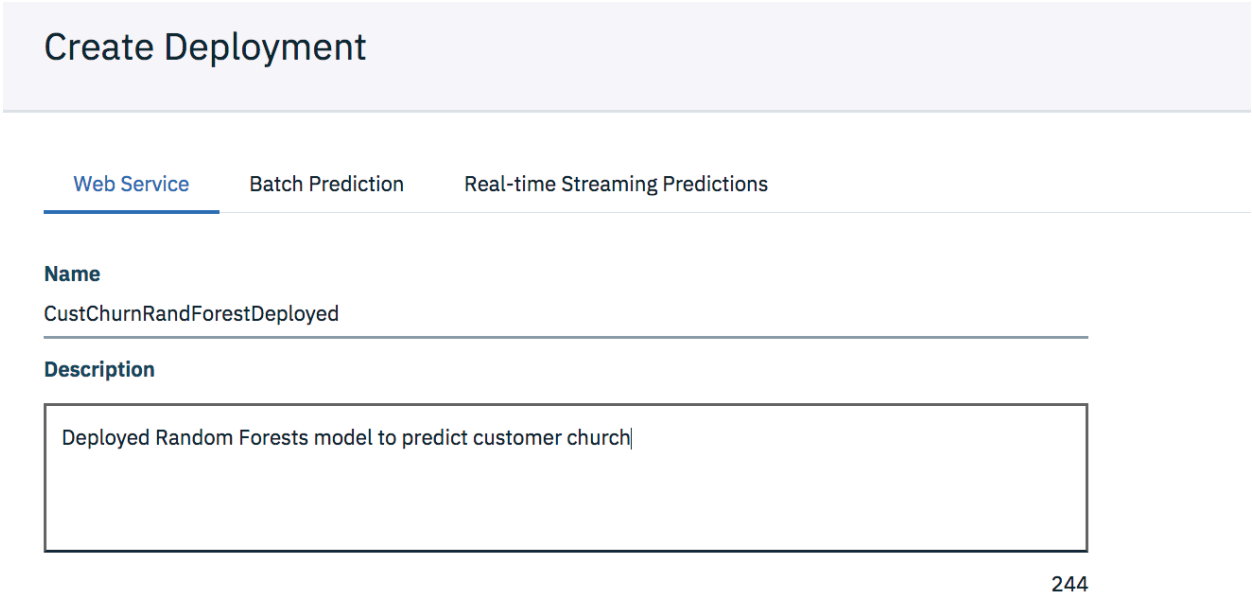
The overview page provides useful information about the model. This includes the ability to deploy and predict with the model.

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



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

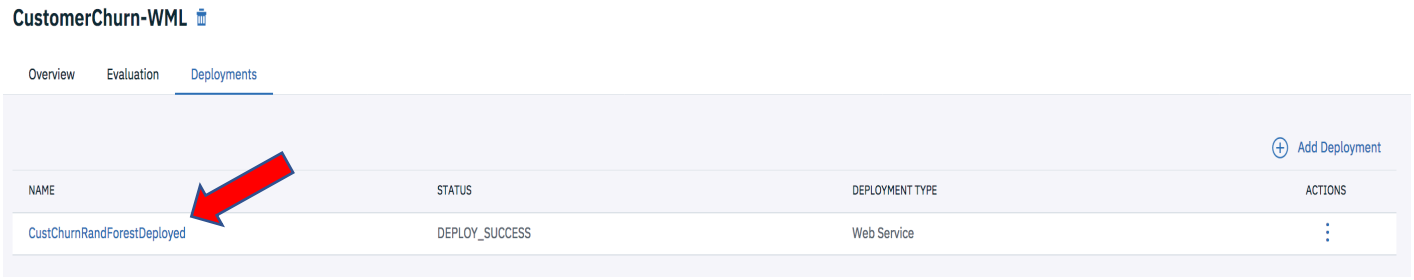
Action



- Click "**Save**"

7. Predict with Model

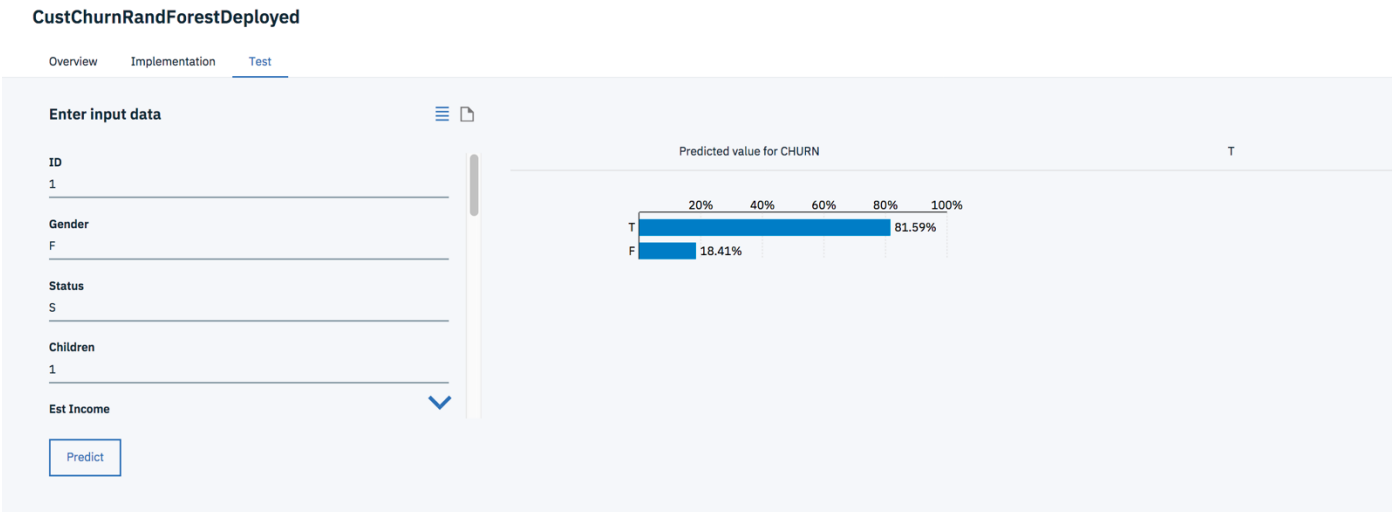
- Choose newly created deployed model:



- Click on "**Test**" to test the model.

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

Action



End of Lesson 5

End of Hands-on Workshop

Thank You