



# Lab: Getting Started with DSXL

*Feb 14th, 2018*

*Author: Sidney Phoon, [yfphoon@us.ibm.com](mailto:yfphoon@us.ibm.com)*



**Table of contents**

Contents

**Overview ..... 1**

**Required software, access, and files ..... 1**

**Part 1: Login to Data Science Experience Local ..... 1**

**Part 2: Create a Jupyter Notebook..... 5**

## Overview

---

In this lab you will learn how to navigate within DSXL and create a new Project.

## Required software, access, and files

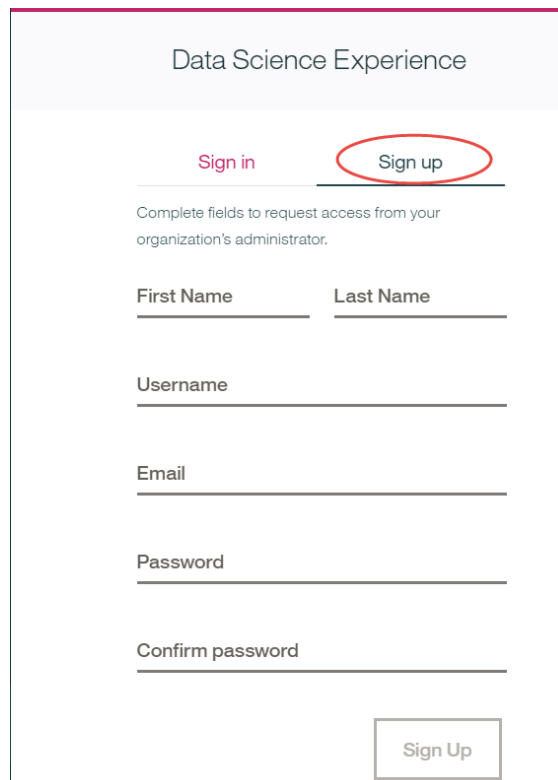
---

- To complete this lab, you will need access to a DSX Local cluster.
- You will also need to download and unzip this GitHub repository:  
[https://github.com/SidneyPhoon/DSX\\_Local\\_Workshop](https://github.com/SidneyPhoon/DSX_Local_Workshop)

## Part 1: Login to Data Science Experience Local

---

1. In a Firefox or Chrome browser, go to your assigned DSXL cluster, e.g  
<https://x.x.xxx.xxx/auth/login/login.html>
2. Click "Sign up" to sign up for an account. Your instructor will approve your account and you can begin the lab



The screenshot shows the 'Data Science Experience' sign-up interface. At the top, there are two links: 'Sign in' and 'Sign up'. The 'Sign up' link is circled in red. Below these links, a message states: 'Complete fields to request access from your organization's administrator.' The form contains several input fields: 'First Name', 'Last Name', 'Username', 'Email', 'Password', and 'Confirm password'. At the bottom right of the form is a 'Sign Up' button.

3. Sign into DSXL with your username

4. In the landing page, explore the Community and View Projects

IBM Data Science Experience Local

Click More to view community for sharing analytics

More

Sample notebooks

NOTEBOOK - JUPYTER

Learn basics about notebooks and...

AUTHOR

IBM

DATE

May 30, 2016

TOPIC

Environment

SOURCE

External

NOTEBOOK - JUPYTER

Modeling Weather Geographies using...

AUTHOR

IBM

DATE

Dec 12, 2017

TOPIC

Weather

SOURCE

External

NOTEBOOK - JUPYTER

Train and predict with Scala machine...

AUTHOR

IBM

DATE

Mar 07, 2017

TOPIC

Transportation

SOURCE

External

NOTEBOOK - JUPYTER

Use Python to load data and run...

AUTHOR

IBM

DATE

May 15, 2016

TOPIC

Transportation

SOURCE

Self-Contained

Click "View all" to list all projects

View all

Recently updated projects

My Projects

New Project

NAME	PROJECT TYPE	ROLE	COLLABORATORS	LAST UPDATED
dsx-samples	-	Viewer	-	2018-02-14

5. Click “**New Project**” to create a “**Blank**” new project.

IBM Data Science Experience Local

Projects > Create Project

Create Project

Blank

From File

From Github

Name\*

Demo

This name is valid

96

Description

Type your description here

3000

☐

Library Project

6. Click **"Assets"** to see all the assets in the project

Projects > Demo

Demo

Description not available.

Date created

Wed Feb 14 2018

0

Assets

5

Environments

0

Jobs

0

Data Sources

Collaborators

View all (1)

user2 user

admin



Recent Assets




NAME	ASSET TYPE	LAST MODIFIED
you have no recently modified assets		

Lab: Getting Started with DSXL

3



7. Click "**Data Sets**" and then "**add data set**" to upload a csv file into DSXL

 IBM Data Science Experience Local 

Projects > Demo > Data Sets   

All Notebooks RStudio Models SPSS Modeler Streams Scripts **Data Sets** Other Files Published Assets


**Data Sets** (0)

All   **add data set**

NAME	TYPE	SIZE	LAST MODIFIED	DATA SOURCE
you have no data sets				

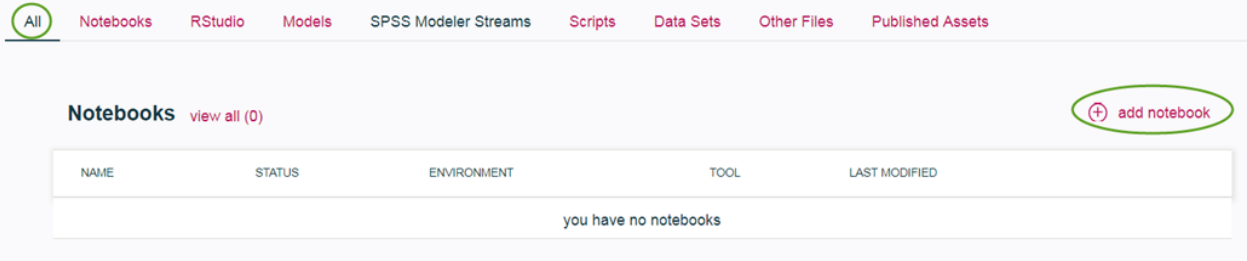
8. Load the **customer\_churn.csv**

The screenshot displays the IBM Watson Studio interface. At the top, there are two tabs: 'Local File' (selected) and 'Remote Data Set'. Below the tabs is a large dashed box containing an upward arrow icon and the text 'Drag and drop your files here'. A red-bordered box highlights the text 'Select from your local file system'. Below this, a horizontal navigation bar contains several tabs: 'All', 'Notebooks', 'RStudio', 'Models', 'SPSS Modeler Streams', 'Scripts', 'Data Sets' (selected), 'Other Files', and 'Published Assets'. The main content area shows a table titled 'Data Sets (1)' with a dropdown menu set to 'All'. The table has five columns: 'NAME', 'TYPE', 'SIZE', 'LAST MODIFIED', and 'DATA SOURCE'. It contains one row with the data set 'customer\_churn', which is a CSV file of 200.93 KB, last modified on 02-14-2018, and sourced from 'Local File'.

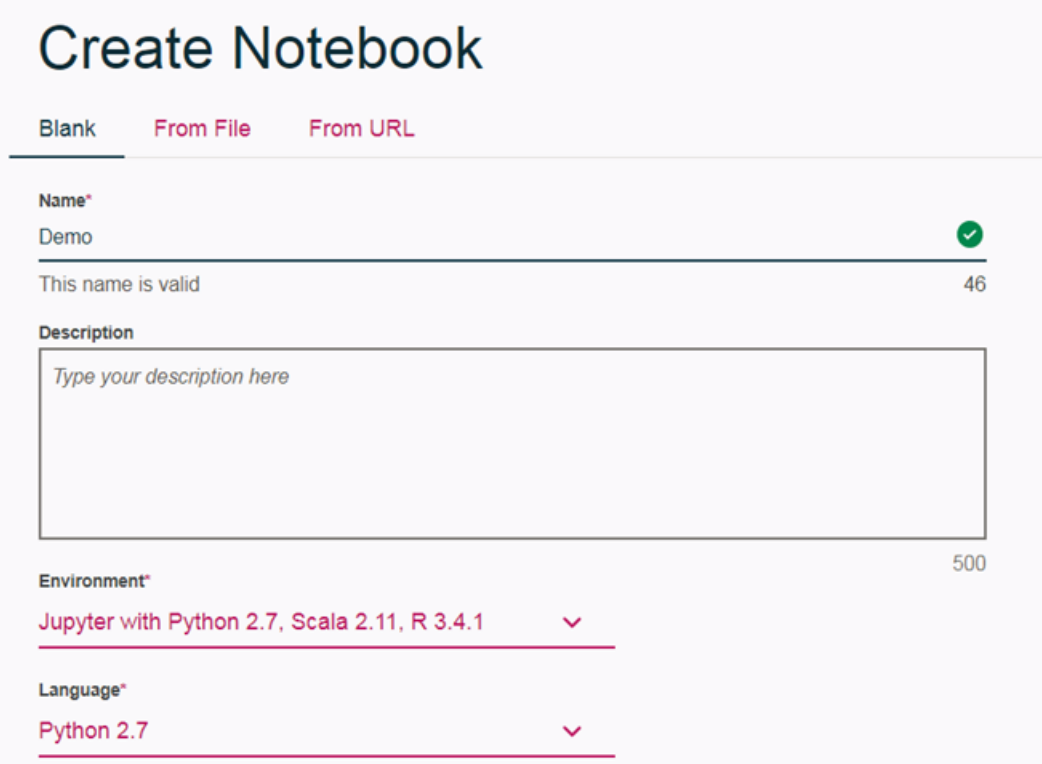
NAME	TYPE	SIZE	LAST MODIFIED	DATA SOURCE
 customer_churn	CSV	200.93 KB	02-14-2018	Local File

## Part 2: Create a Jupyter Notebook

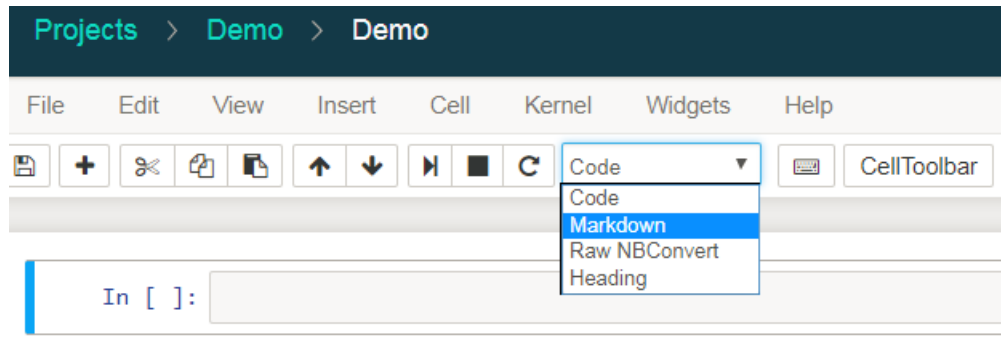
1. Within the project you have created in Part 1, click “**All**”, and “**add notebook**” to add a blank notebook.



Take the default settings and click “**Create**”



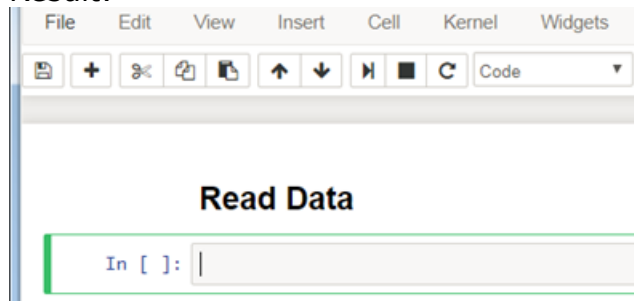
2. Define a **markdown** cell: place the cursor in the first code cell and change the cell type to **markdown**.



Enter **## Read Data** into the markdown cell and click the run icon.

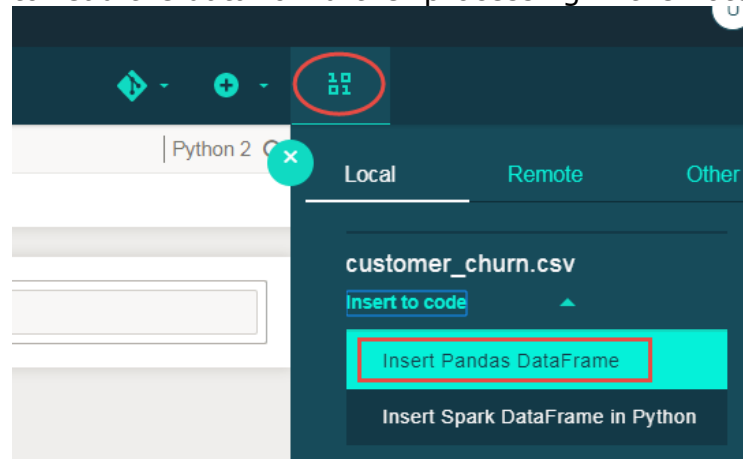


Result:



**Markdown cheatsheet:** <https://datascience.ibm.com/docs/content/analyze-data/markd-jupyter.html?context=analytics>

3. **Read data** into the notebook: Click the **"Find Data"** icon, click **insert to code** to insert the data as a Pandas dataframe. This will generate the code to read the data for further processing in the notebook.





#### 4. Run the generated code cell

File

Edit

View

Insert

Cell

Kernel

Widgets

Help

+

⌕

⬆

⬇

⏮

⏭

↺

Code

⌵

CellToolbar

Read Data

In [1]:

```
import os, pandas as pd
# Add asset from file system
df_data_1 = pd.read_csv(os.environ['DSX_PROJECT_DIR']+'/datasets/customer_churn.csv')
df_data_1.head()
```

Out[1]:

	CHURN	Gender	Status	Children	EstIncome	CarOwner	Age	LongDistance	International	Local	Dro
0	T	F	S	1	38000.00	N	24	23.56	0.0	206.08	0
1	F	M	M	2	29616.00	N	49	29.78	0.0	45.50	0
2	F	M	M	0	19732.80	N	50	24.81	0.0	22.44	0
3	F	M	S	2	96.33	N	56	26.13	0.0	32.88	1
4	F	F	M	2	52004.80	N	25	5.03	0.0	23.11	0