Introduction to Notebooks for SPSS Professionals

*Notebooks and Lab Instructions by*

*Sidney Phoon, Mokhtar Kandil, Rich Tarro, and Elena Lowery*

*April 19, 2017*

# Overview

In this lab you will learn the basics of working with Jupyter notebooks in IBM Data Science Experience.

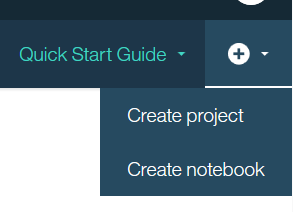
# Required software, files, and user ids

To complete this lab, you will need:

* User id for **IBM Data Science Experience (DSX):** http://datascience.ibm.com/
* Lab files provided in **IntroductionToNotebooksForSPSS.zip** (unzipped to C:\ IntroductionToNotebooksForSPSS)

# Part 1: Create a DSX project and import sample notebook

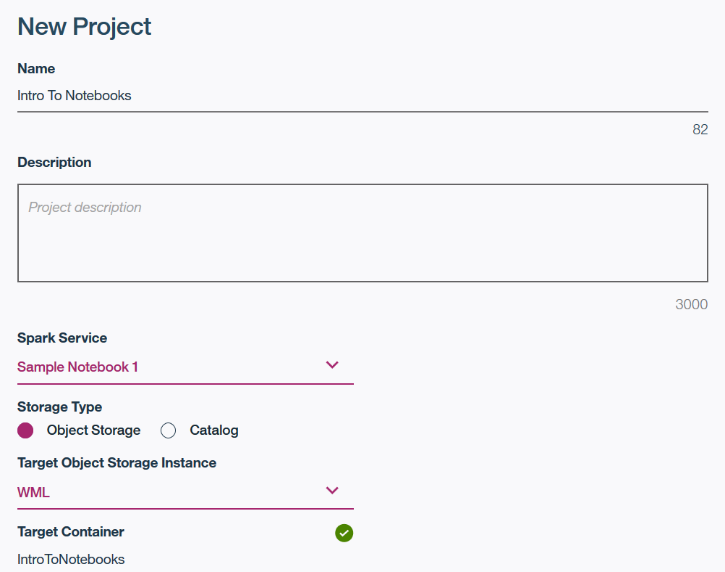
1. Login to DSX**:** http://datascience.ibm.com/
2. In the top right corner click on the + icon and select **Create Project**



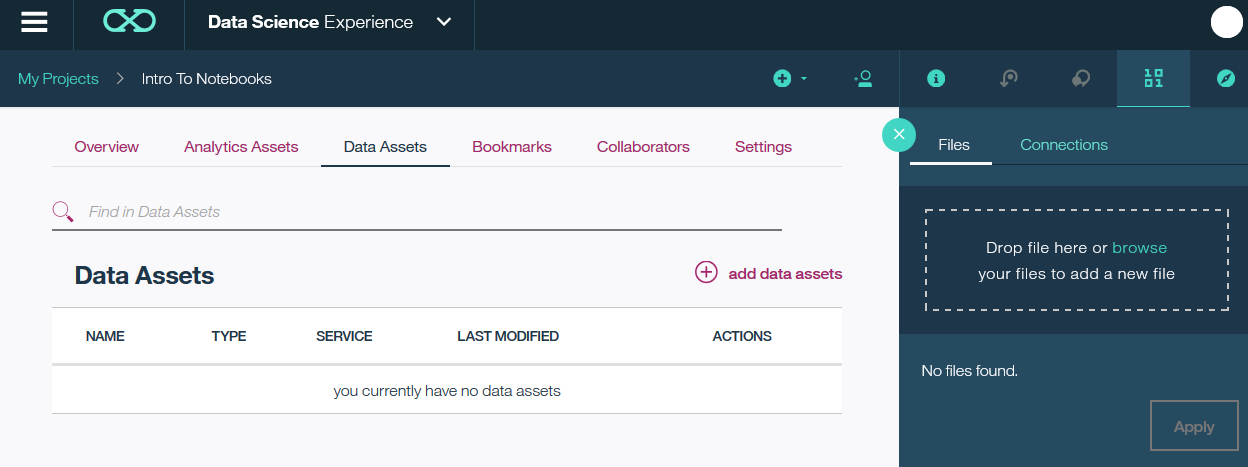
1. Enter project name ***Intro To Notebooks***. Make sure *Object Storage* is selected as **Storage type**.

Your **Spark Service** and **Target Object Storage** values will be different. Check with the instructor if they are blank – you may need to set them up if this is the first time that you are using DSX.

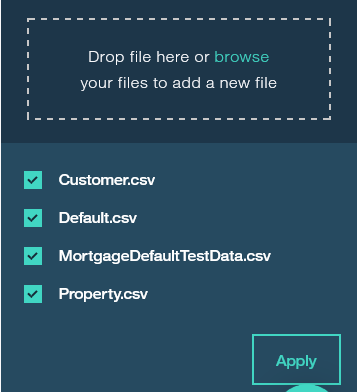
Click **Create**.

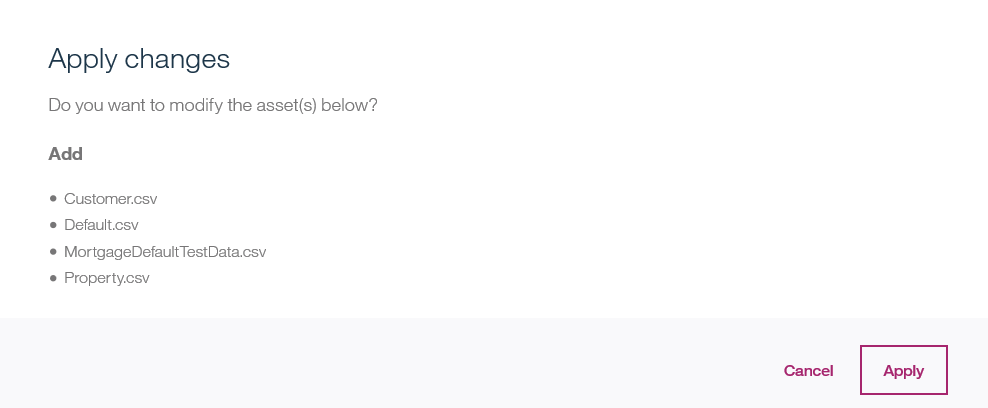


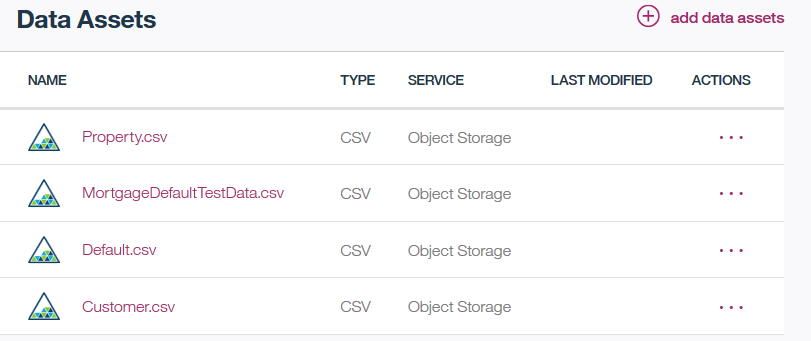
1. Click on the **Data Assets** tab, then click **browse**



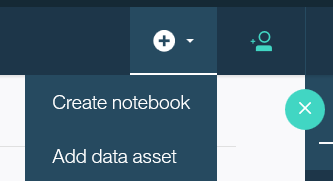
1. Browse to each file in the C:\IntroductionToNotebooksForSPSS\Data folder. Then select Apply.







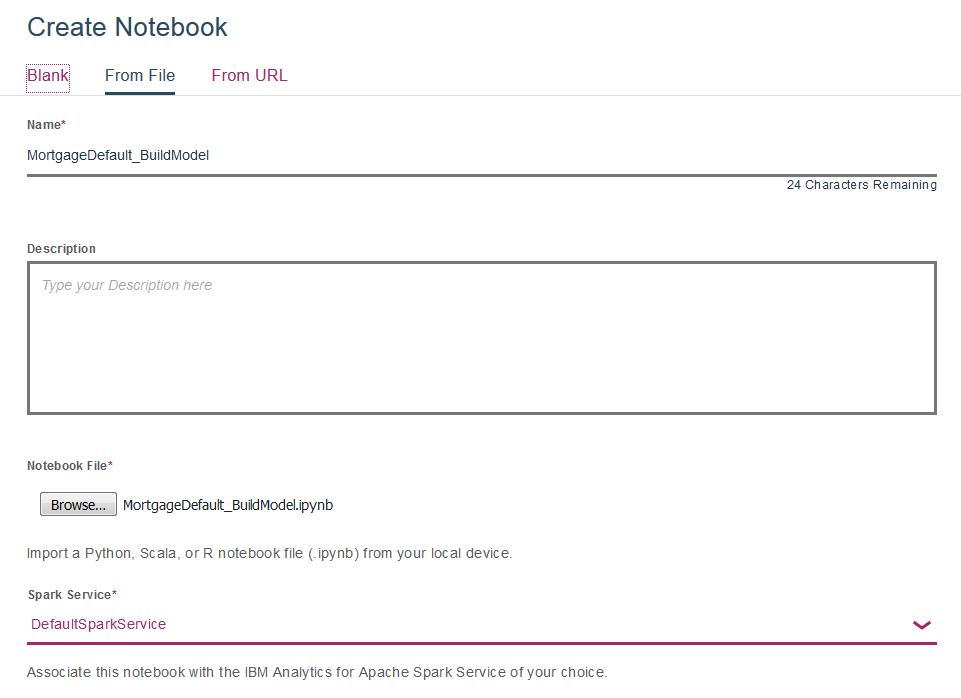
1. Switch to the **Analytic Assets** tab Click on the + icon and select **Create Notebook**



1. Click the *From File* tab. Enter name *MortgageDefault\_BuildModel* and Browse to the notebook file: *C:\IntroductionToNotebooksForSPSS\Notebooks\ MortgageDefault\_BuildModel.ipynb*

Your *Spark Service name* will be different.

Click **Create Notebook**.



By default, the notebook opens in the edit mode. Continue this lab by following the instructions in the notebook.

*After you have finished working on the model building notebook, follow these instructions to add and test the model scoring notebook.*

1. Create another notebook From File
2. Enter name *MortgageDefault\_Scoring* and Browse to the notebook file: *C:\IntroductionToNotebooksForSPSS\Notebooks\ MortgageDefault\_Scoring.ipynb*

Click **Create Notebook** and follow instructions in the notebook to

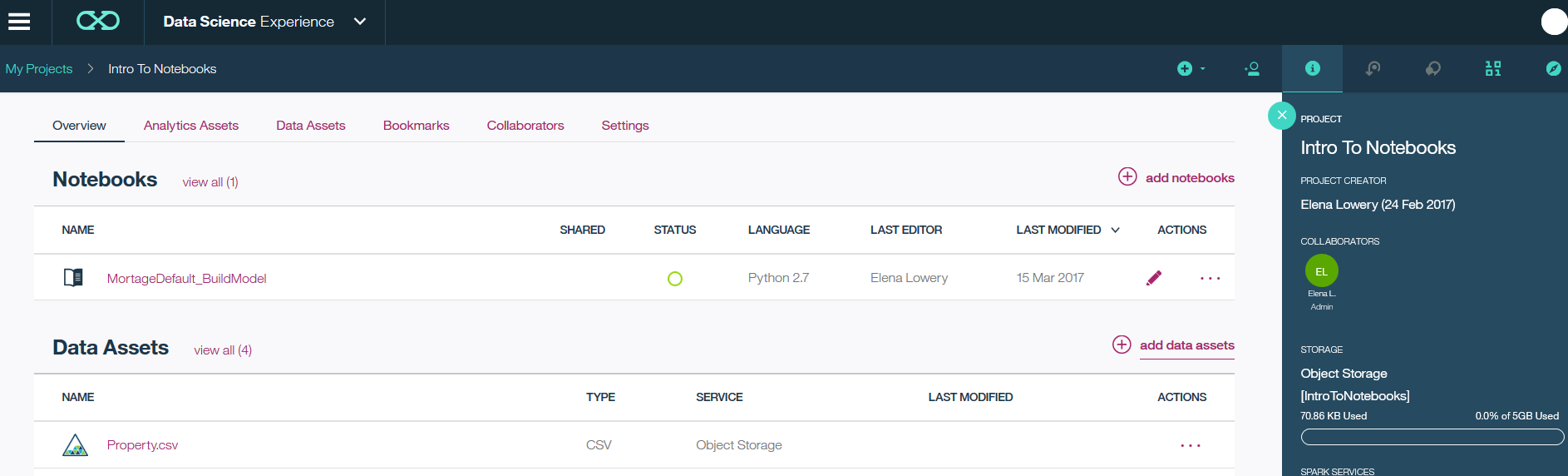
complete the lab.

# Appendix A: Change Object Storage Access

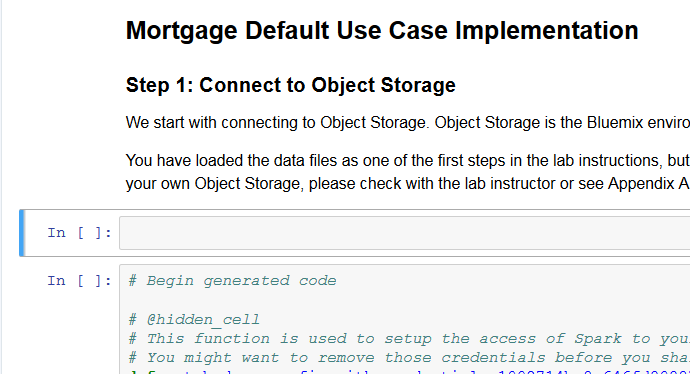
This section explains how you can change the provided notebook to read files from your project’s Object Storage (by default the notebook points to one of the instructors’ Object Storage).

1. On the project page click on the Information (i) icon and take a note of the Object Storage name.

*If you followed instructions in the lab, it will be the same as shown in the screenshot. If you gave your project a different name, it may be different.*



1. Open the *MortgageDefault\_BuildModel* notebook and change to Edit mode.
2. Add a new cell above the 1st code cell: position cursor in the 1st code sell and select **Insert -> Insert Cell Above**

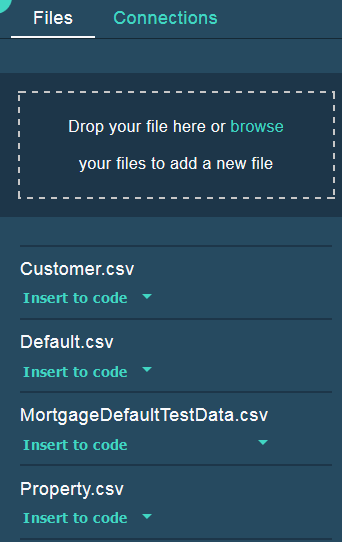


New cell

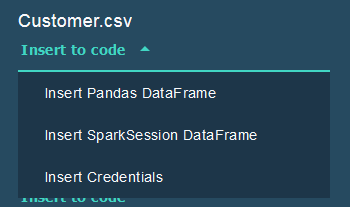
1. Click on the **Find and Add Data** icon in the top right corner



Notice that all data files are displayed.

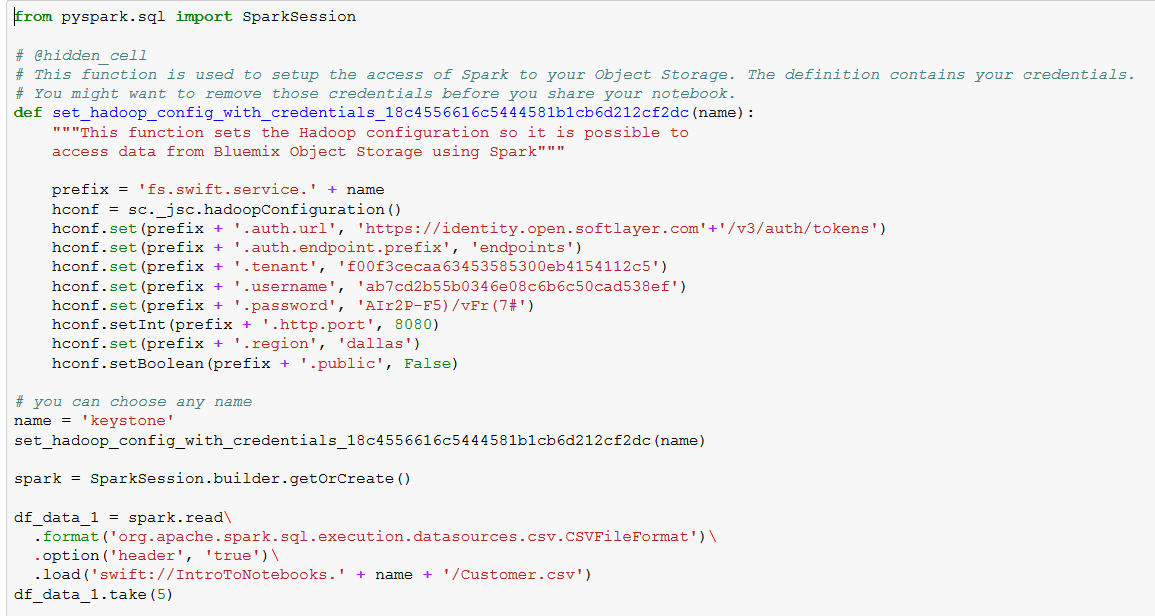


1. Position cursor in the newly created cell, then select *Customer.csv* and click **Insert to code -> Insert SparkSession DataFrame**.

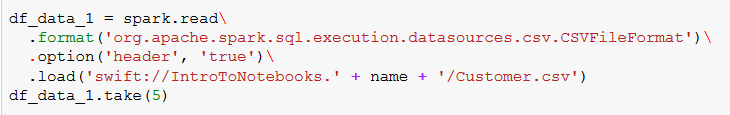


1. Notice that the generated code is similar to the code that was provided in the notebook. The main differences are credentials and tenant information. Also, we consolidated data access in a different cell.

Take a note of the Object Storage name – you may need to replace it in the 2nd cell if you used a different project name.



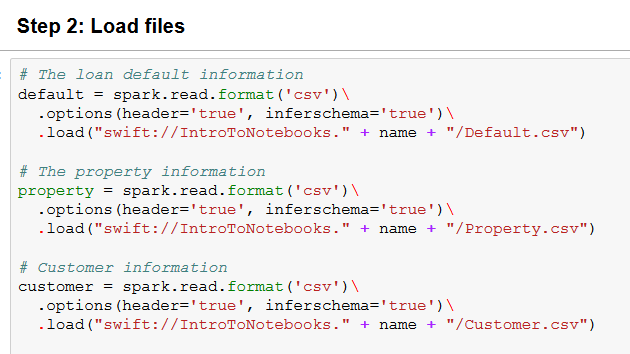
1. Delete code after *spark=SparkSession.builder.getOrCreate()*



1. Comment out or delete the cell below - the shortcut is to select everything in the cell and press Ctrl+/



1. Review the next cell to make sure that Object Storage doesn’t need to be changed. If it does, replace reference to Object Storage (in this example *IntroToNotebooks*) in this and the last cell.





You have finished modifying the notebook to point you’re your project Object Storage.