Lab-1: Setup Environment

Introduction

This lab will set up the Watson Studio environment for subsequent labs and introduce you to the Project and Community features of Watson Studio. Watson Studio is an integrated platform of tools, services, data, and meta-data to help companies and agencies accelerate their shift to be data driven organizations. The platform enables data professionals such as data scientists, data engineers, business analysts, and application developers collaboratively work with data to build, train, deploy machine learning and deep learning models at scale to infuse AI into business to drive innovation. Watson Studio is designed to support the development and deployment of data and analytics assets for the enterprise.

End-to-End Data Science

The general flow of the End to End Data Science PoT will be guided by the activities shown in Figure 1- End to End Flow. This lab will focus on the Create Project, Research Topics, and Connect to Data activities.

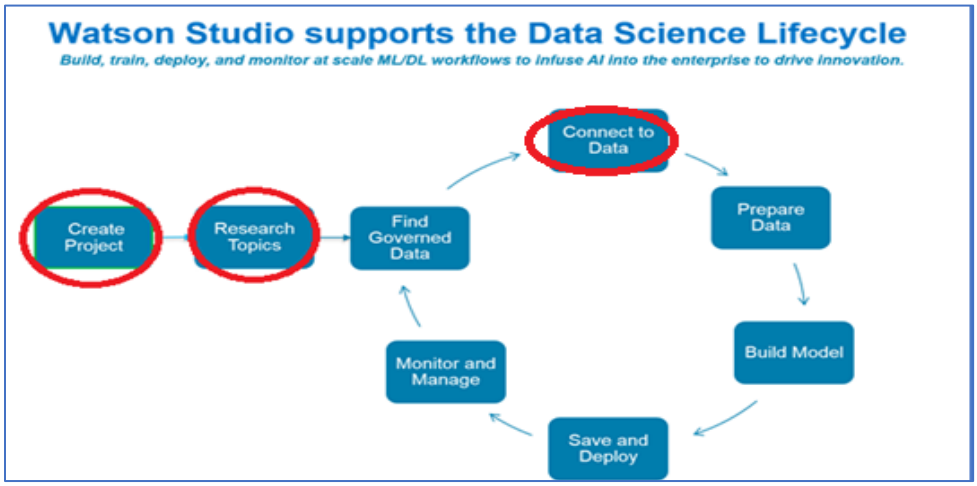


Figure 1- End to End Flow

Objectives

The goal of this lab is to familiarize the user with the Project and Community features of Watson Studio, and to set up the environment for subsequent labs. Projects are a core component of Watson Studio. Projects enable you to organize your analytic and data assets in one place. Projects are also the home base for collaboration. Colleagues can be added as collaborators on a project with administrator, editor, or viewer access.

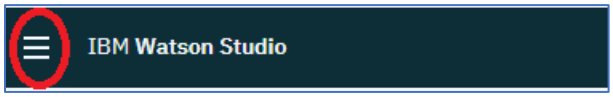
The Community feature of Watson Studio provides built-in learning. Articles, Notebooks, Tutorials, Datasets, and Papers are curated from well-known sources and provided as "Community Cards". These artifacts can be bookmarked in Projects for easy reference. The Community feature supports the "Research Topics" activity in the end-to-end process shown above.

After completing this lab, you will be familiar with these features of Watson Studio.

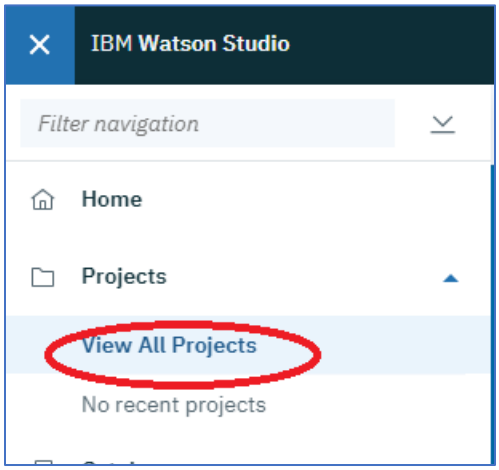
1. Create a project
2. Create an object storage instance and associate it with the project
3. Associate an existing Watson Machine Learning service instance with the project
4. Add a collaborator to the project
5. Research topics by searching the Community
6. Add data assets to the project

Create a Project

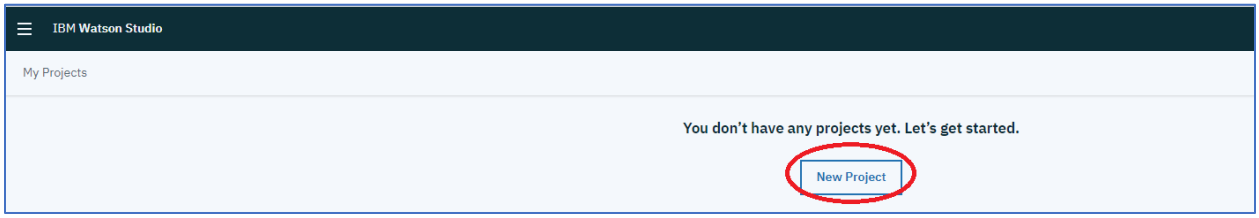
1. Log into your Watson Studio account at datascience.ibm.com, then click on the hamburger icon .



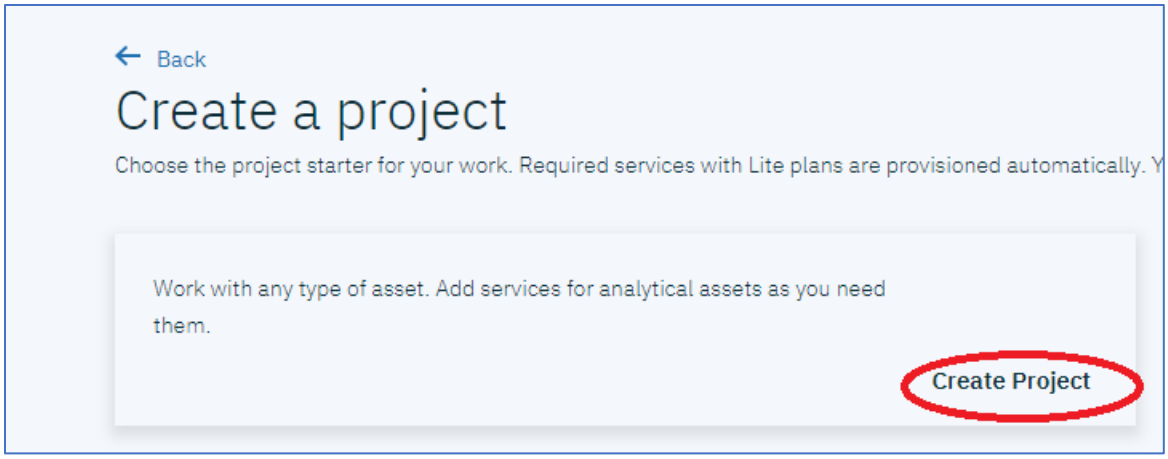
2. Click on **View All Projects**



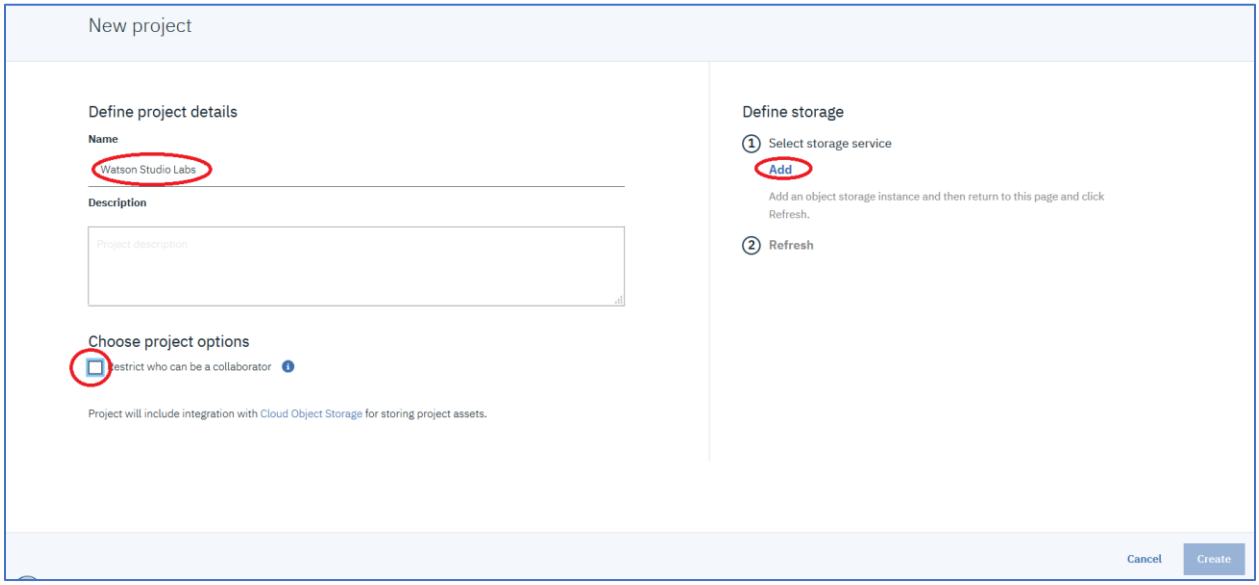
3. Click on **New Project**.



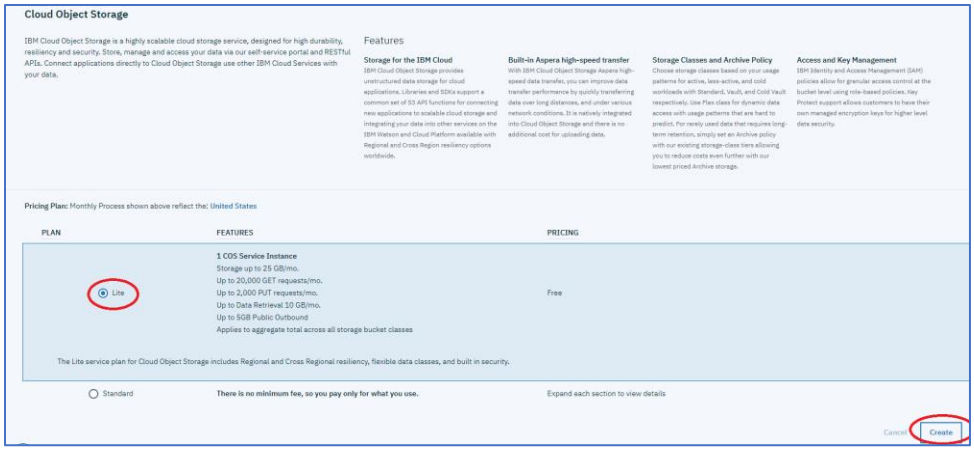
4. Hover the mouse over **Standard,** and then click **Create Project**.



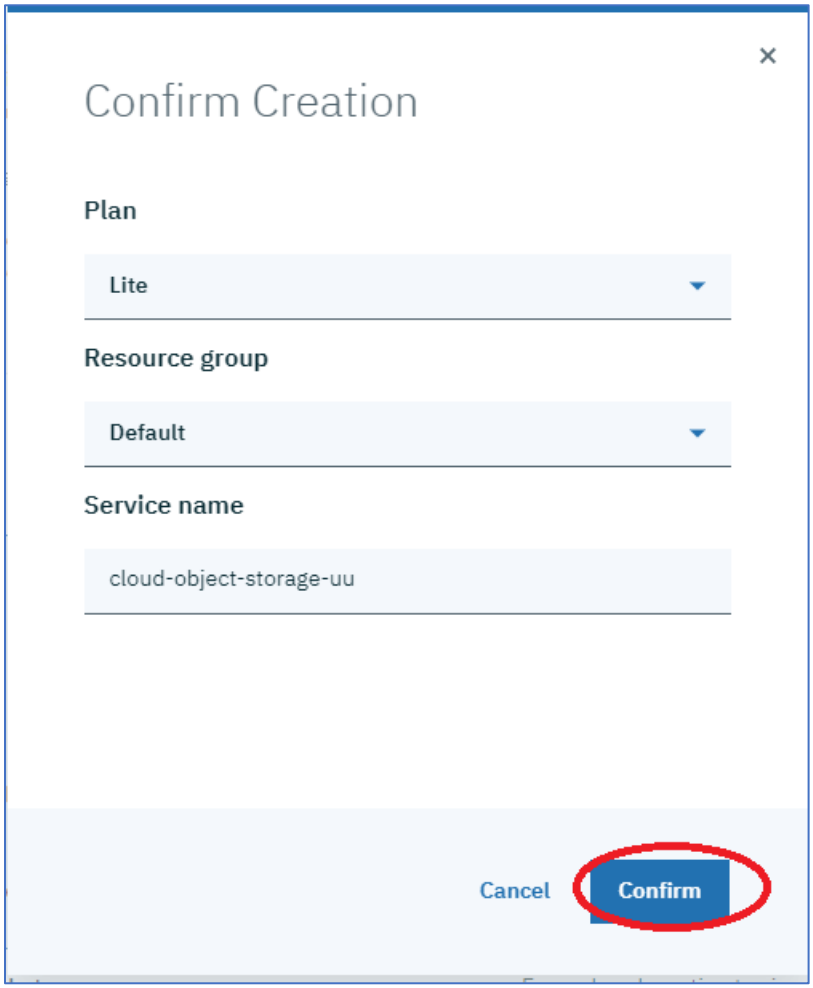
5. Enter “Watson Studio Labs” for the **Name,** optionally enter a **Description,** make sure to uncheck **Restrict who can be a collaborator** (if it’s checked), and in **Define Storage** click on **Add** to add an object storage instance.



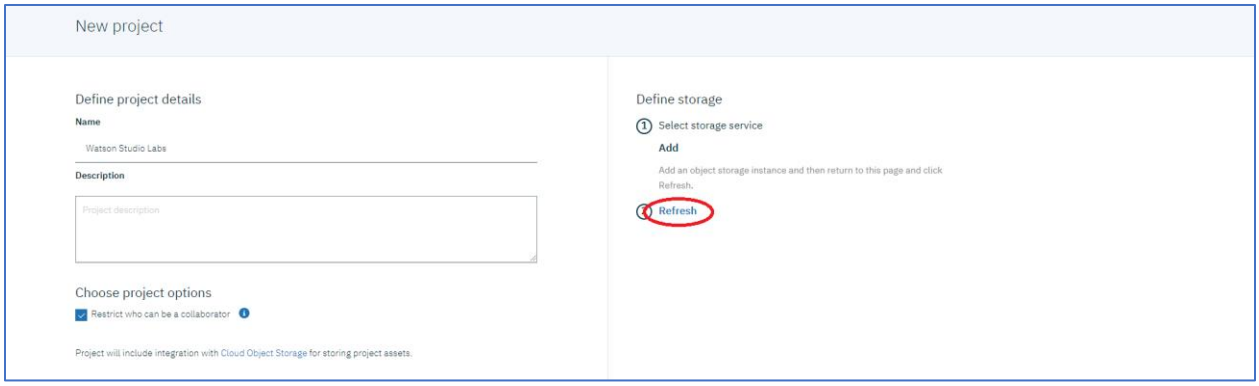
6. Click on **Lite,** and then click on **Create**



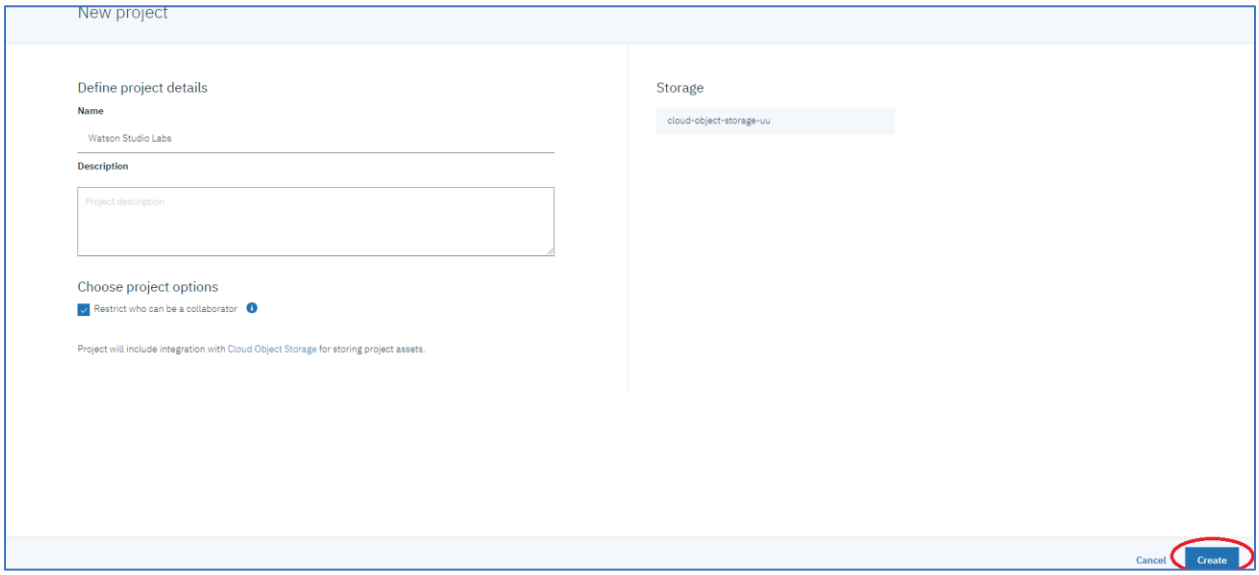
7. Click **Confirm**.



8. Click **Refresh**.

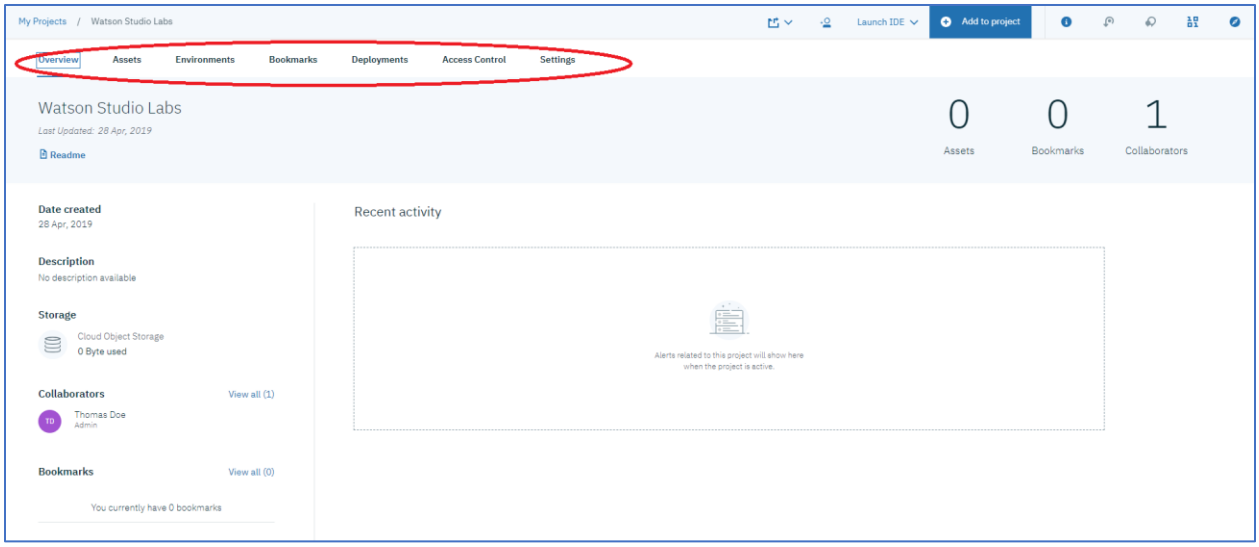


9. Click **Create**.



10. The Project **Overview** page is shown. This page provides summarized information about the project. In addition to the Overview page, are six other pages described below.

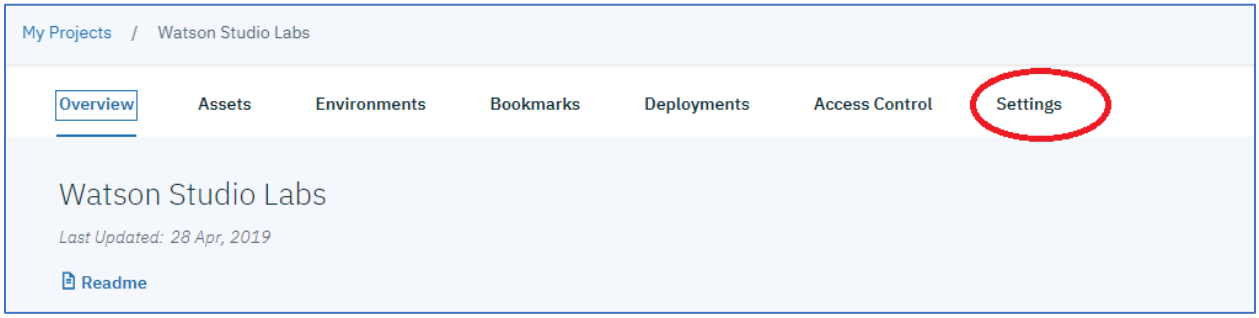
- **a. Assets Page –** Analytics and Data assets can be added to the project from this page.
- **b. Environments Page -** Provides information on the current notebook environments that are defined, lists the active notebook environments currently running, and enables users to create custom notebook environments.
- **c. Bookmarks Page -** Lists artifacts from the Community that are bookmarked in this project.
- **d. Deployments Page –** Lists the deployed models
- **e. Access Control –** Lists the project collaborators and enables users to add/remove collaborators.
- **f. Settings –** Enables users to view and set project attributes.



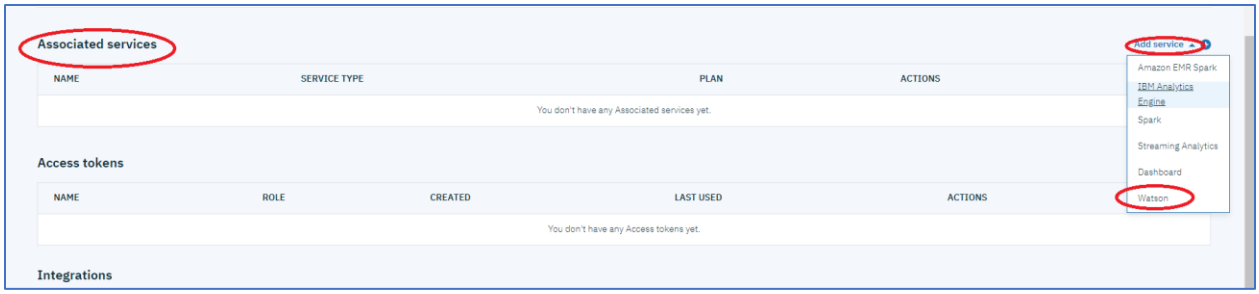
Associate a Watson Machine Learning Service to the Project

To save and deploy machine learning models, a Watson Machine Learning service must be created (if one doesn't exist) and added to our project.

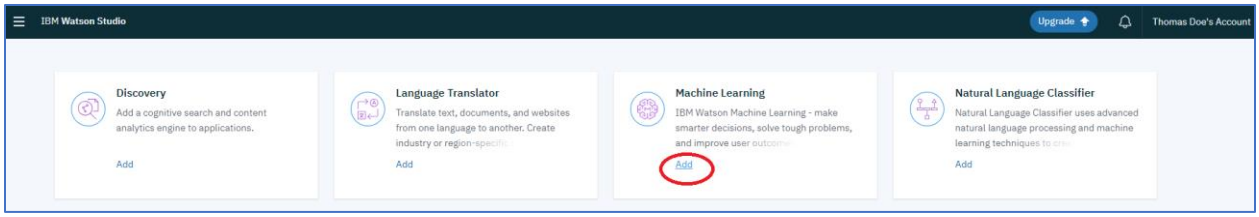
1. Click on **Settings** to navigate to the Project **Settings** page.



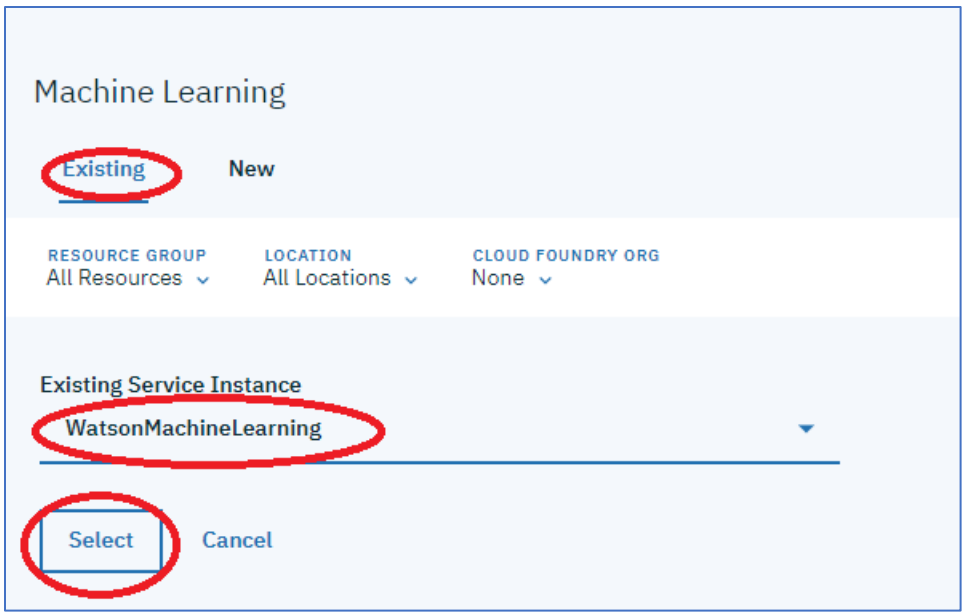
2. Scroll down to **Associated Services,** click on **Add service,** click on **Watson**.



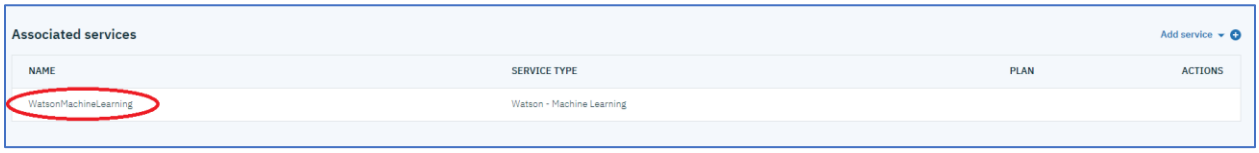
3. Click on **Add** in the **Machine Learning** tile.



4. Select **Existing**, select **WatsonMachineLearning** for the **Existing Service Instance**, and click on **Select**.



5. The **WatsonMachineLearning** service is associated with the project.

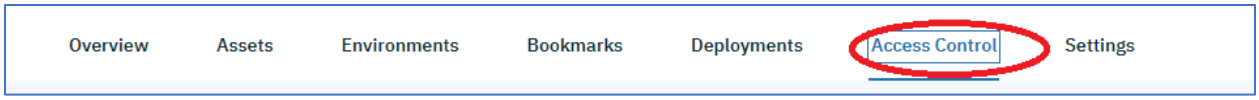


Add a Project Collaborator

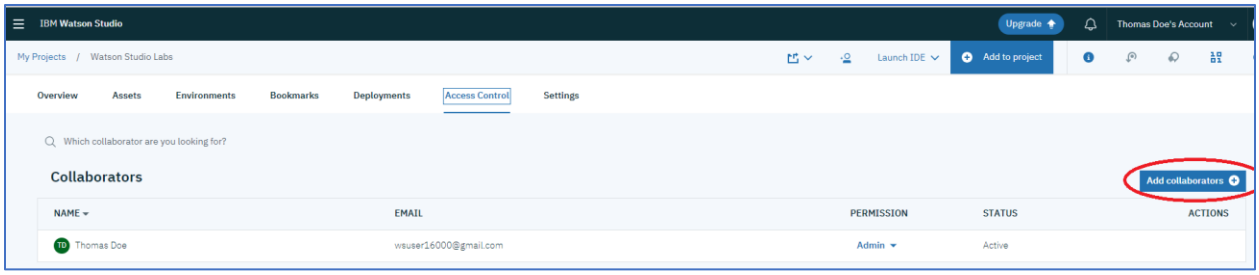
Colleagues can gain access to a project's data and analytic assets by being made a collaborator. Permissions are based on the assigned role. The roles are administrator, editor, and viewer.

We will add a collaborator with a role of **Editor.**

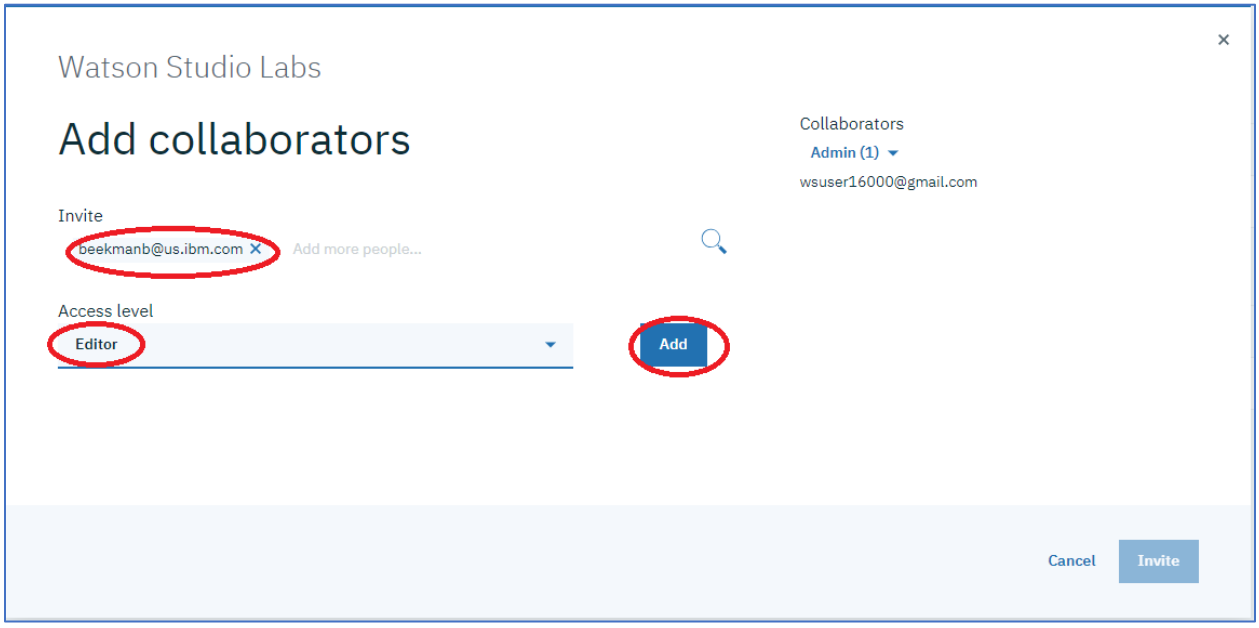
1. Click on the **Access Control** tab.



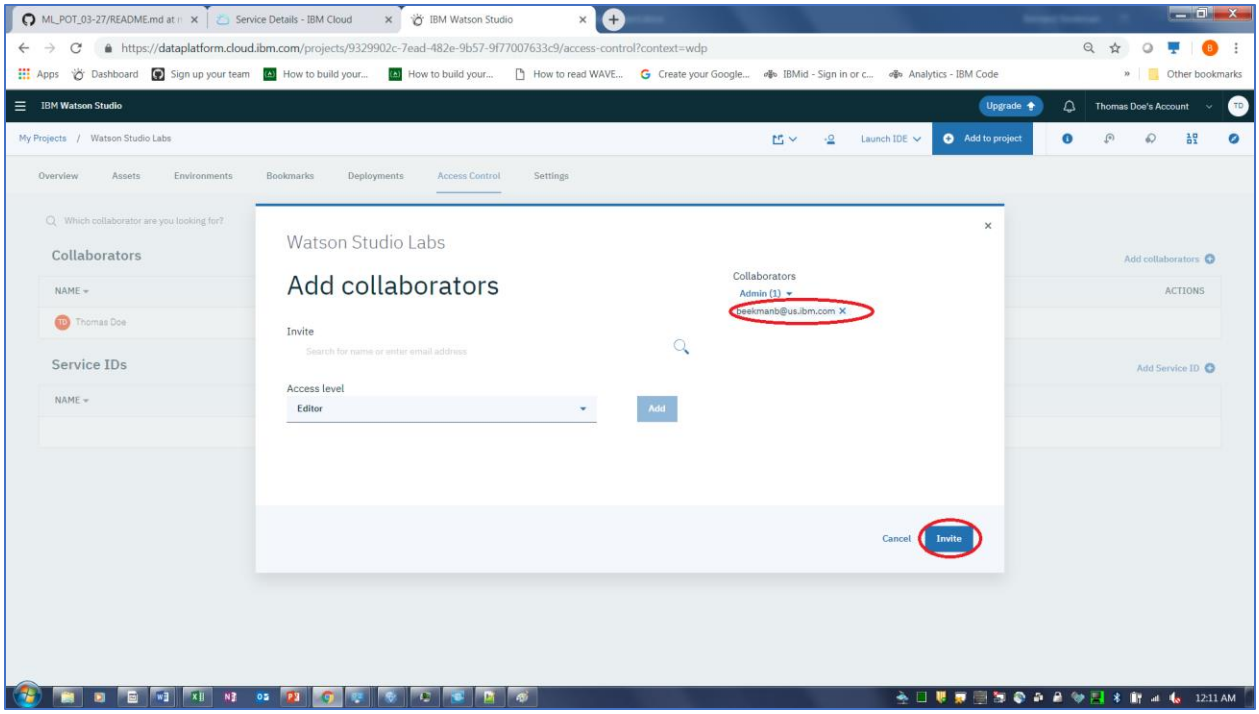
2. Click on **Add collaborators**.



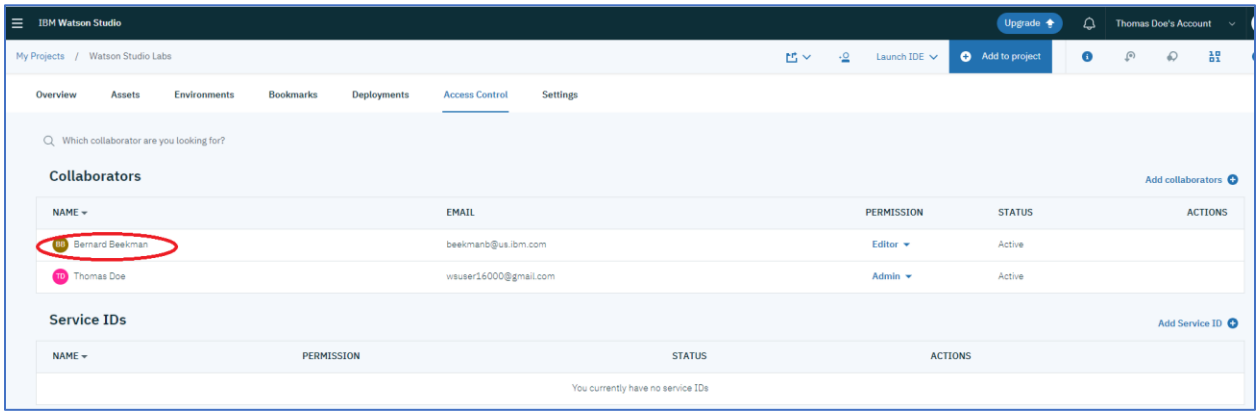
3. For **Invite**, enter the **PROJECT COLLABORATOR** e-mail address on the index card handed out by the instructor, press Tab key, select Editor from the **Access Level** dropdown, and click on **Add**.



4. The collaborator is added to the list of Collaborators on the right-hand side. Click on **Invite**.



5. The collaborator is added.

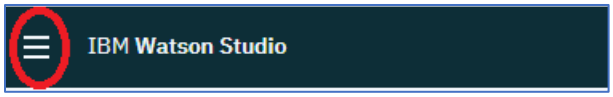


Research Topics

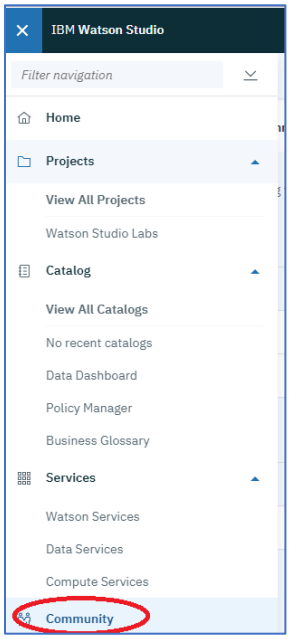
As you work on a data analysis project, you may need to do some research to help find a solution. Watson Studio provides a built-in learning capability, accessed via the **Community** option, that contains articles, sample notebooks, tutorials, sample datasets, and papers on a variety of topics. These are curated on a regular basis to provide up-to-date materials.

For the lab exercise, assume that you are interested in learning how to use the Watson Machine Learning api to save and deploy a machine learning mode. We will look for a sample notebook that demonstrates this capability and bookmark this capability in our project.

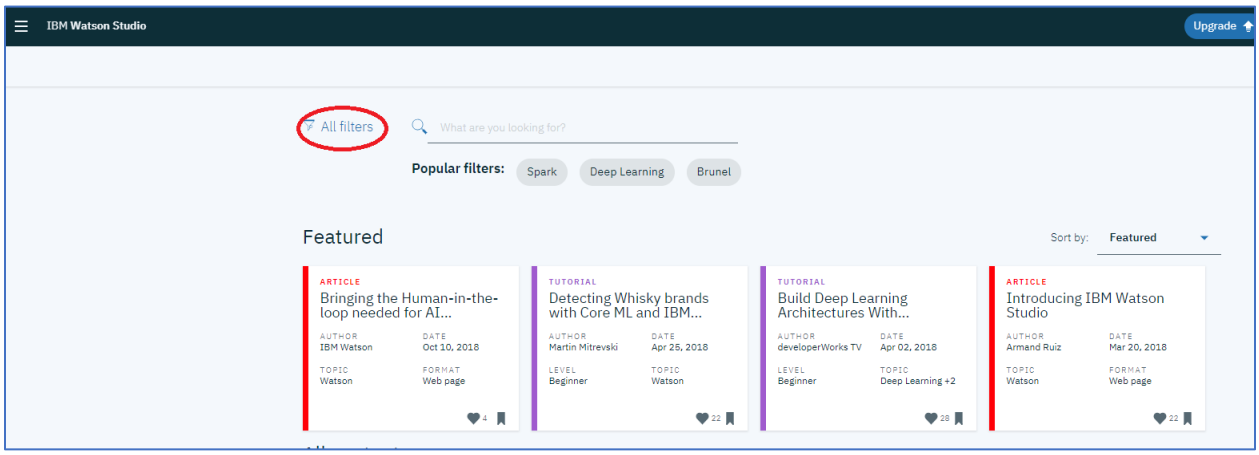
1. Click on the ☰ icon.



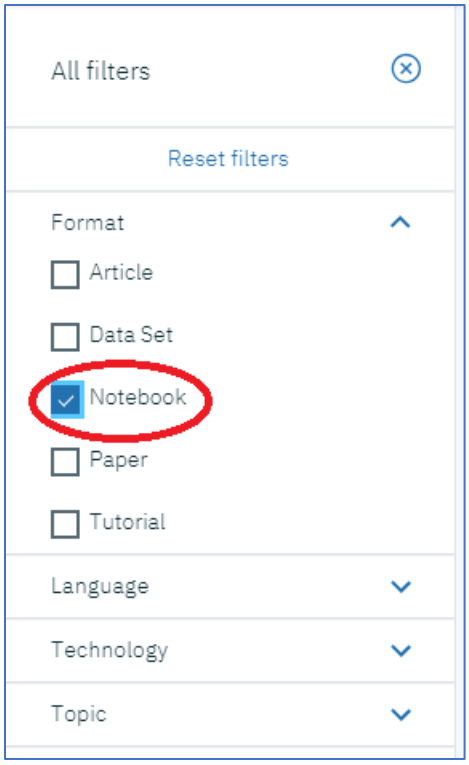
2. Click on **Community**.



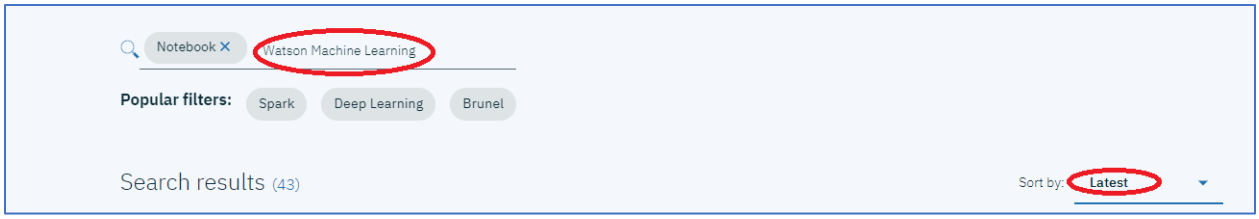
3. The Community Cards are displayed. Click on **All filters**.



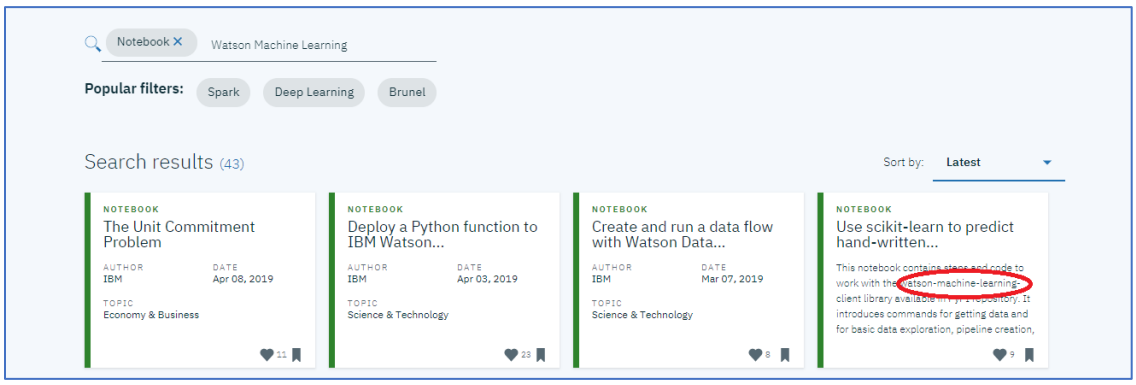
4. Click on **Notebook**.



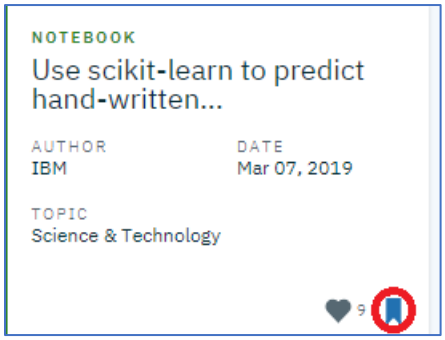
5. Enter **Watson Machine Learning** in the **Search** area. Click on the **Sort by** dropdown and change it to **Latest** to get the most up-to-date information.



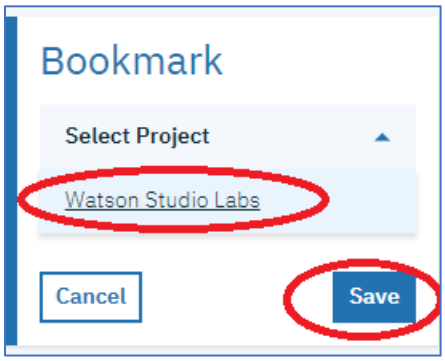
6. The Community view is updated. Hover the mouse over the Community Card "Use scikit-learn to predict hand-written …" The descriptive text refers to the Watson Machine Learning api. This notebook appears to be a good candidate for having code demonstrating the use of the Watson Machine Learning api. Let's bookmark the notebook in the project.



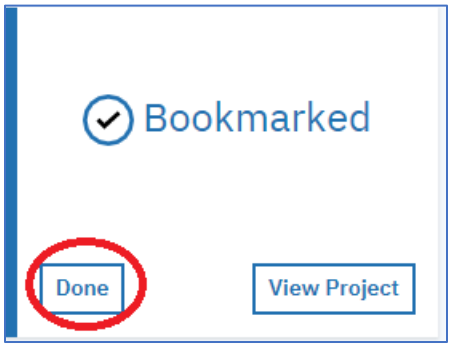
7. Click on the bookmark icon.



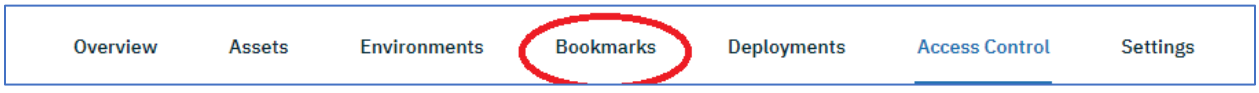
8. Click on the **Watson Studio Labs** project and click **Save**.



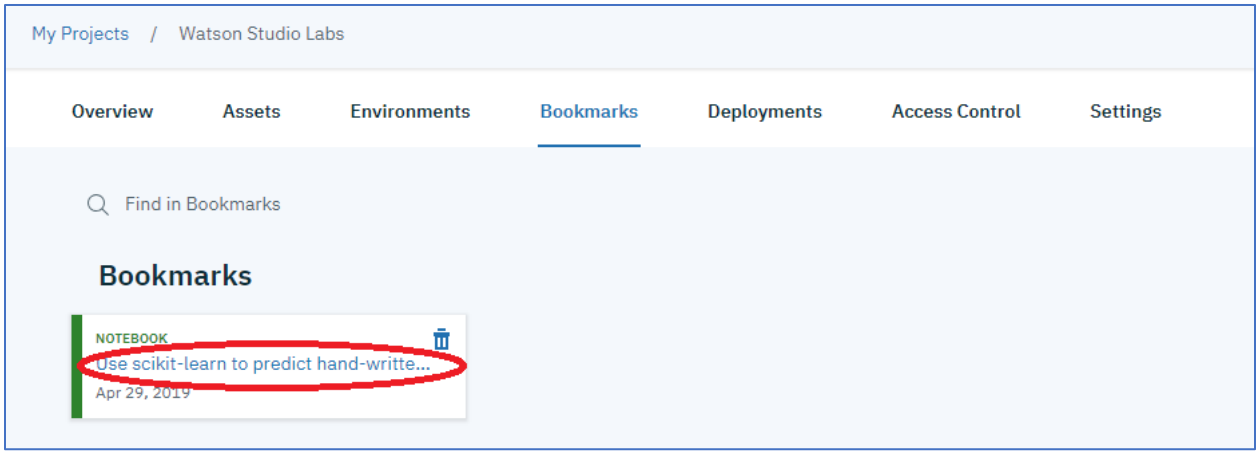
9. Click on **Done**.



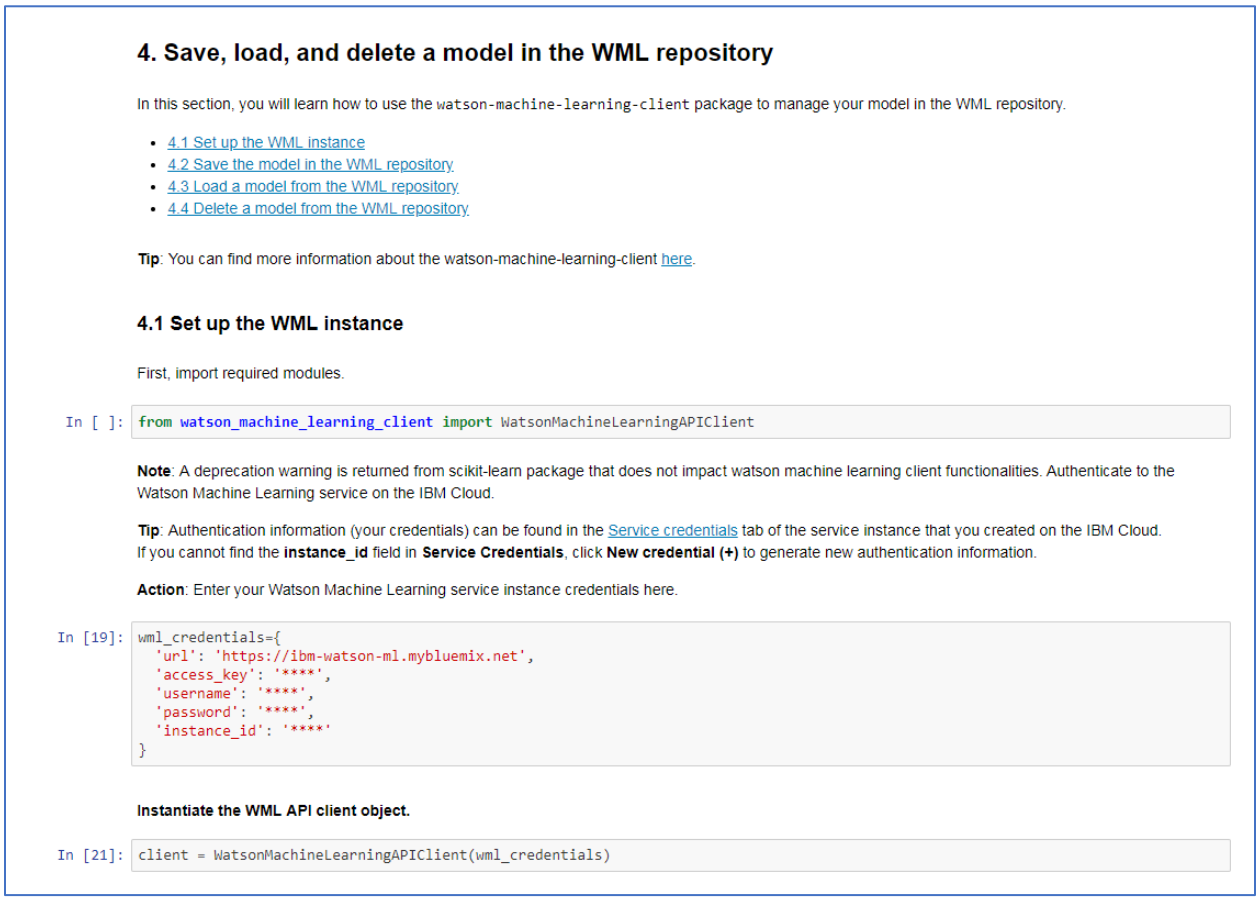
10. Close the Watson Studio tab corresponding to the **Community**. Click on the **Bookmark** tab.



11. Click on the notebook.



12. Scroll to Section 4 to see the information on Watson Machine Learning apis.



4. Save, load, and delete a model in the WML repository

In this section, you will learn how to use the `watson-machine-learning-client` package to manage your model in the WML repository.

- 4.1 Set up the WML instance
- 4.2 Save the model in the WML repository
- 4.3 Load a model from the WML repository
- 4.4 Delete a model from the WML repository

Tip: You can find more information about the watson-machine-learning-client here.

4.1 Set up the WML instance

First, import required modules.

In [ ]:
```
from watson_machine_learning_client import WatsonMachineLearningAPIClient
```

Note: A deprecation warning is returned from scikit-learn package that does not impact watson machine learning client functionalities. Authenticate to the Watson Machine Learning service on the IBM Cloud.

Tip: Authentication information (your credentials) can be found in the Service credentials tab of the service instance that you created on the IBM Cloud. If you cannot find the **instance_id** field in **Service Credentials**, click **New credential (+)** to generate new authentication information.

Action: Enter your Watson Machine Learning service instance credentials here.

In [19]:
```
wml_credentials={
  'url': 'https://ibm-watson-ml.mybluemix.net',
  'access_key': '****',
  'username': '****',
  'password': '****',
  'instance_id': '****'
}
```

Instantiate the WML API client object.

In [21]:
```
client = WatsonMachineLearningAPIClient(wml_credentials)
```

Download Files

We will download a zip file that contains 2 data files that we will use in subsequent labs.

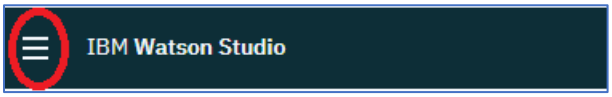
1. Download the trafficking_data.zip from the github repository. Click on trafficking_data.zip, to download the zipped file.

2. Extract the file contents. You should have two files extracted, (1) Categories.csv, and (2) Occupation.csv.

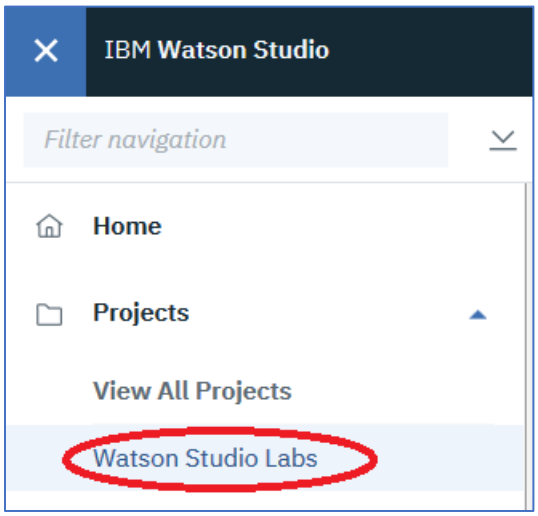
Connect to Data

In this section, we will upload the 2 files to object storage, add a connection asset and a connected data asset to the project. This will result in 4 Data Assets.

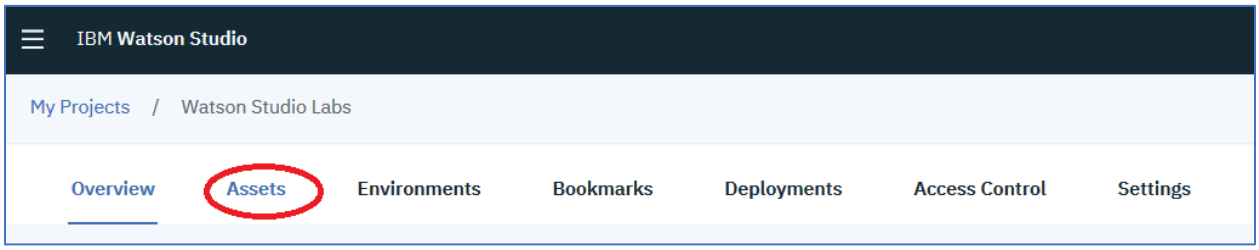
1. Click on the hamburger icon ☰



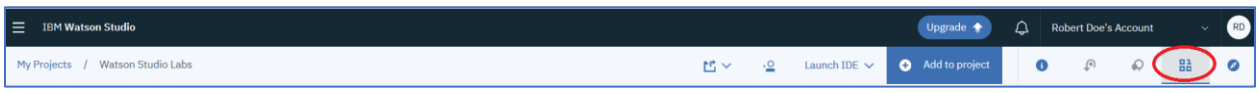
2. Click on **Watson Studio Labs.**



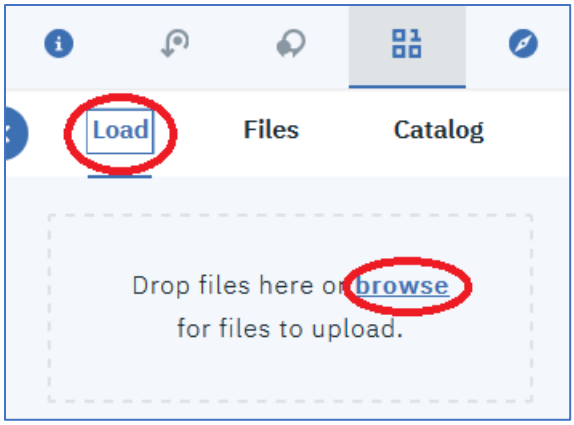
3. Click on **Assets**.



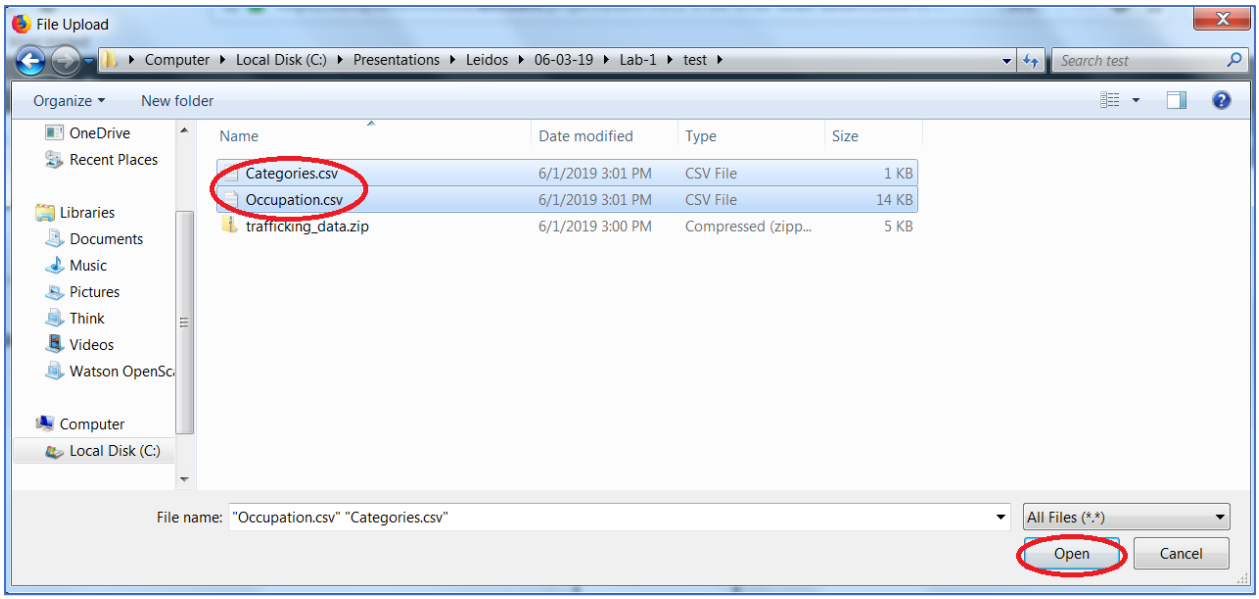
4. Click on the "one-zero" icon .



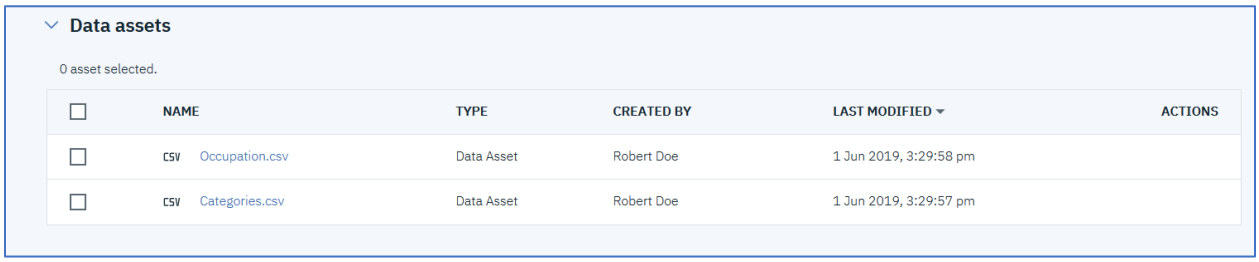
5. Click on **Load** and then click on **browse**. If you don't see **Load** and **browse** click on the again.



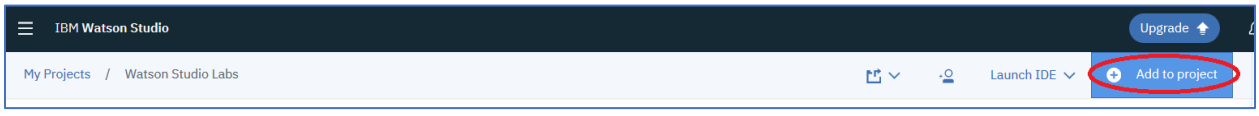
6. Navigate to the folder where you downloaded the **Categories.csv** and **Occupation.csv** files. Click on those files and click **Open.**



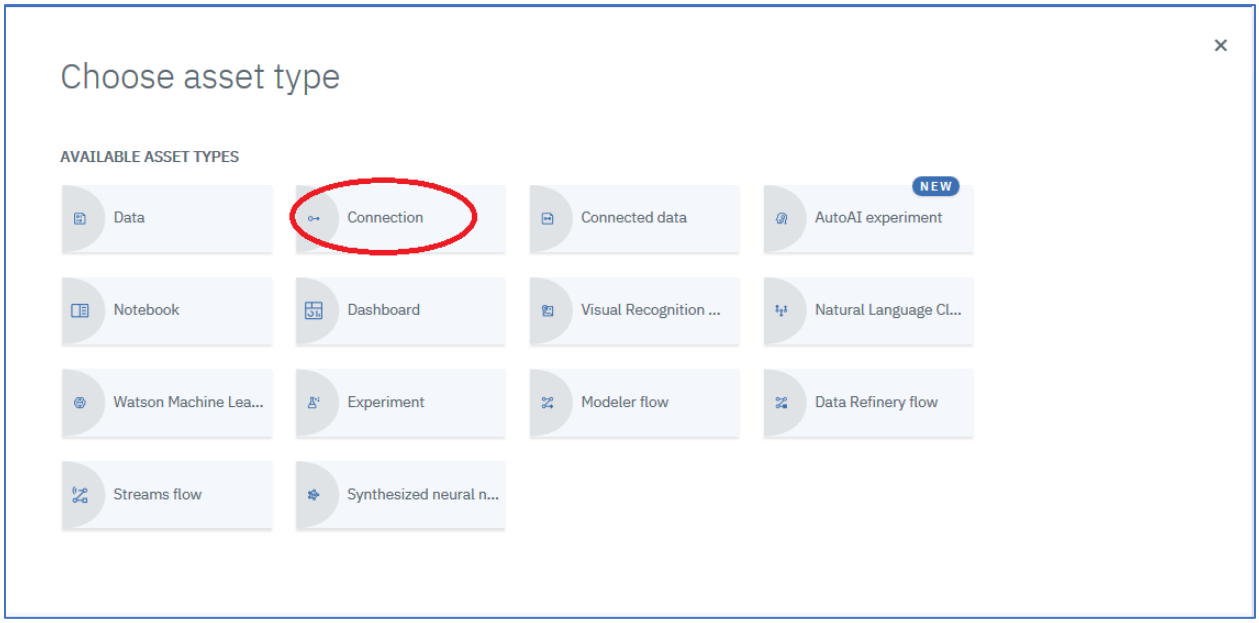
7. Two data assets are added to the project.



8. Click on **Add to project**.

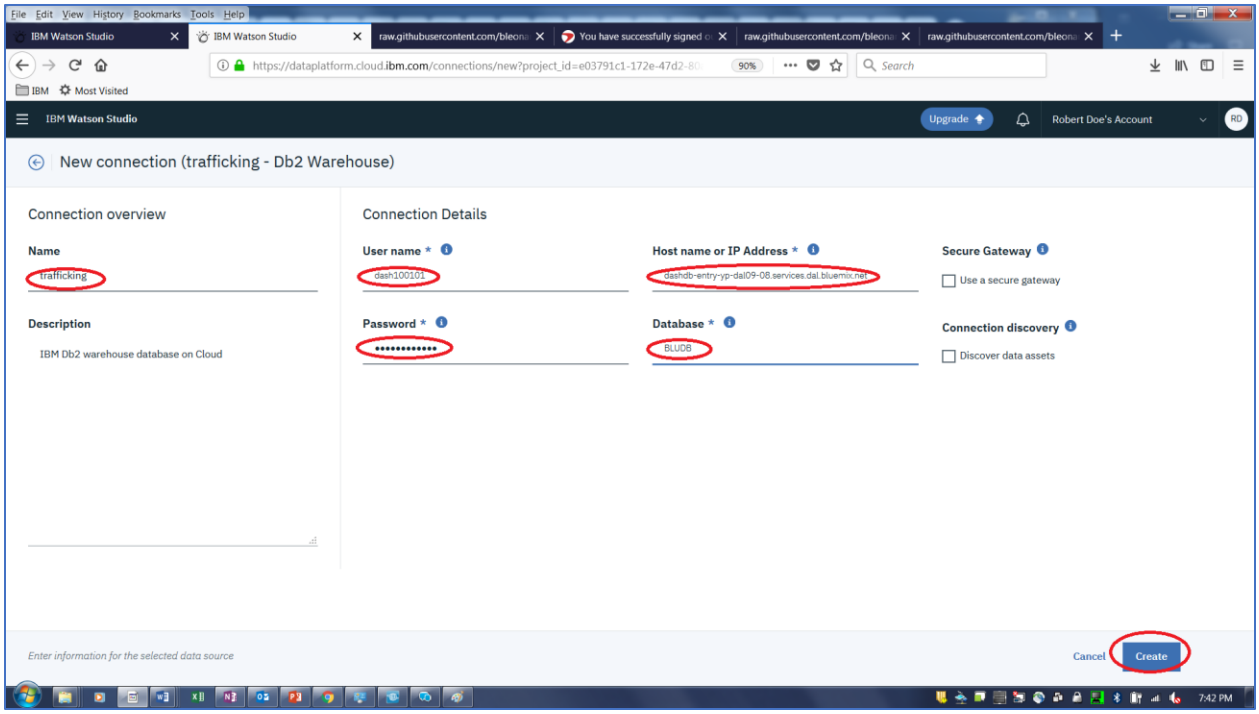


9. Click on **Connection**.

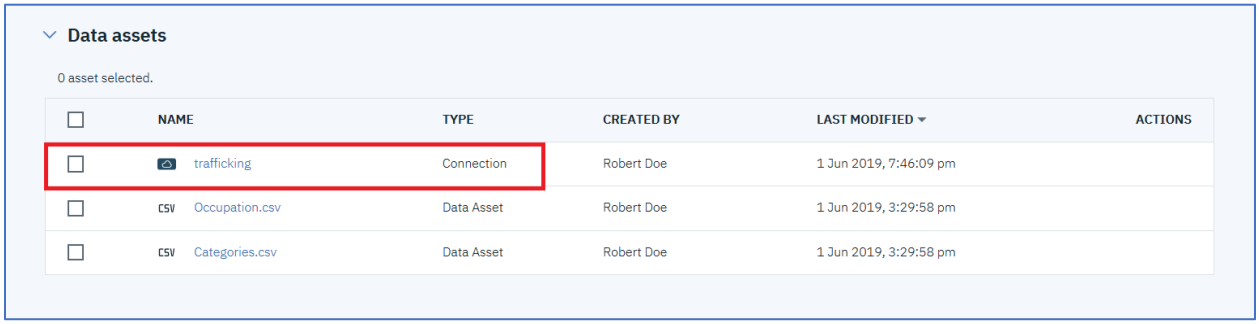


10. Enter "trafficking" for the **Name** of the connection. Click on Access DB2 Credentials to obtain the DB2 Credentials. You should have an index card that provides your assigned database. If not, ask your instructor. Locate the credentials that match your assigned database. The example below assumes that Database number 2 was on the index card.

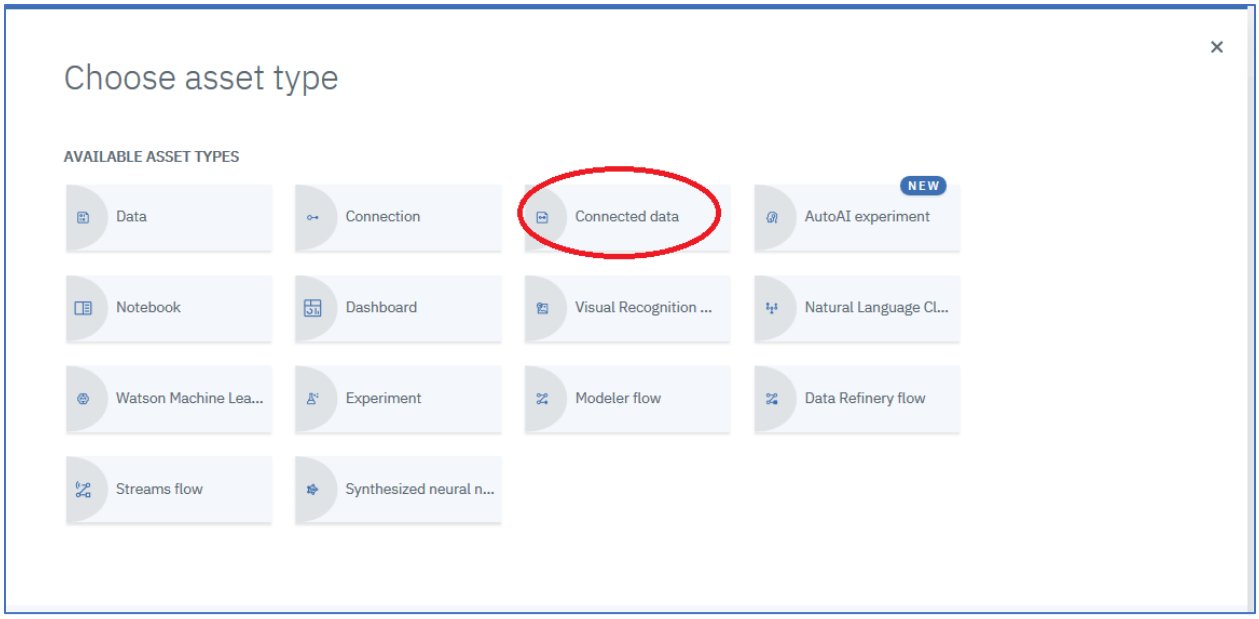
11. Cut and paste the values for **hostname**, **password**, **db**, and **username** into the corresponding entry areas on the screen. Click on **Create**.



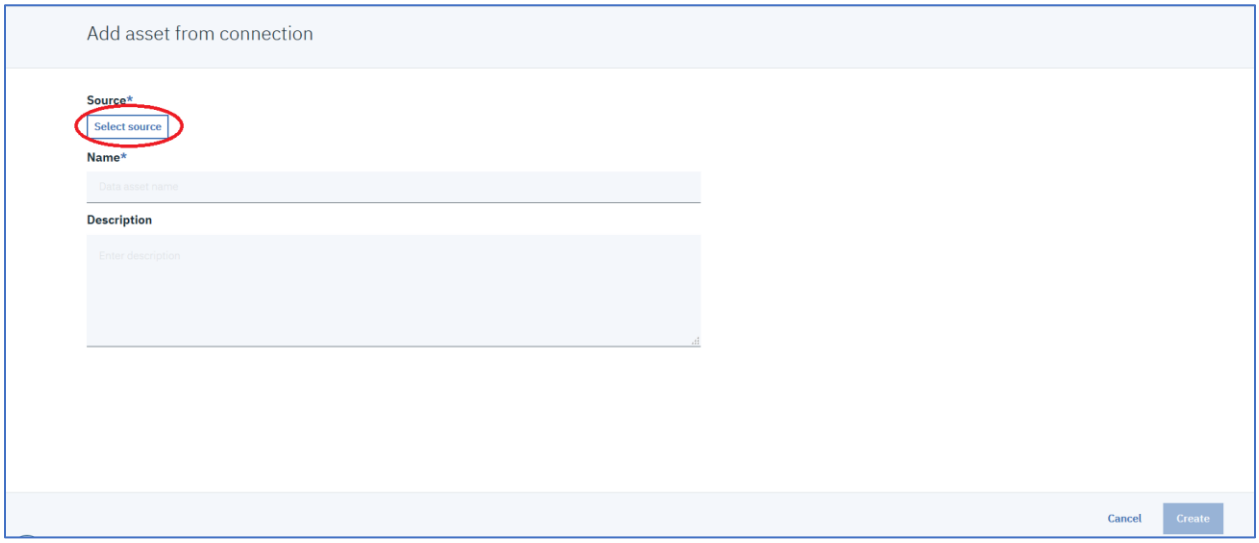
12. The connection is added.



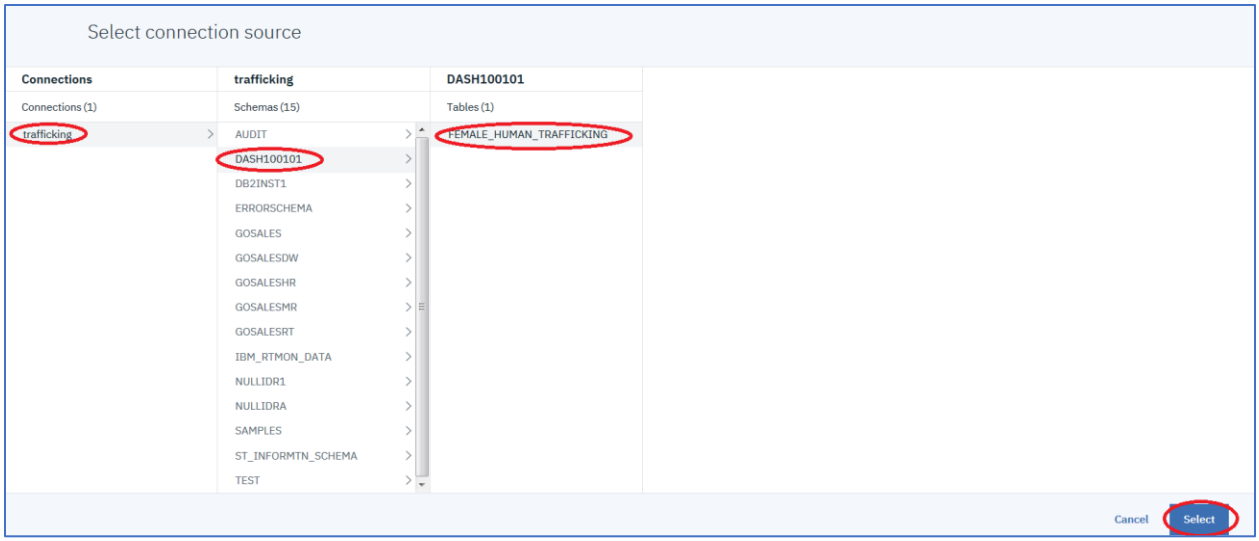
13. Click on **Add to project**.



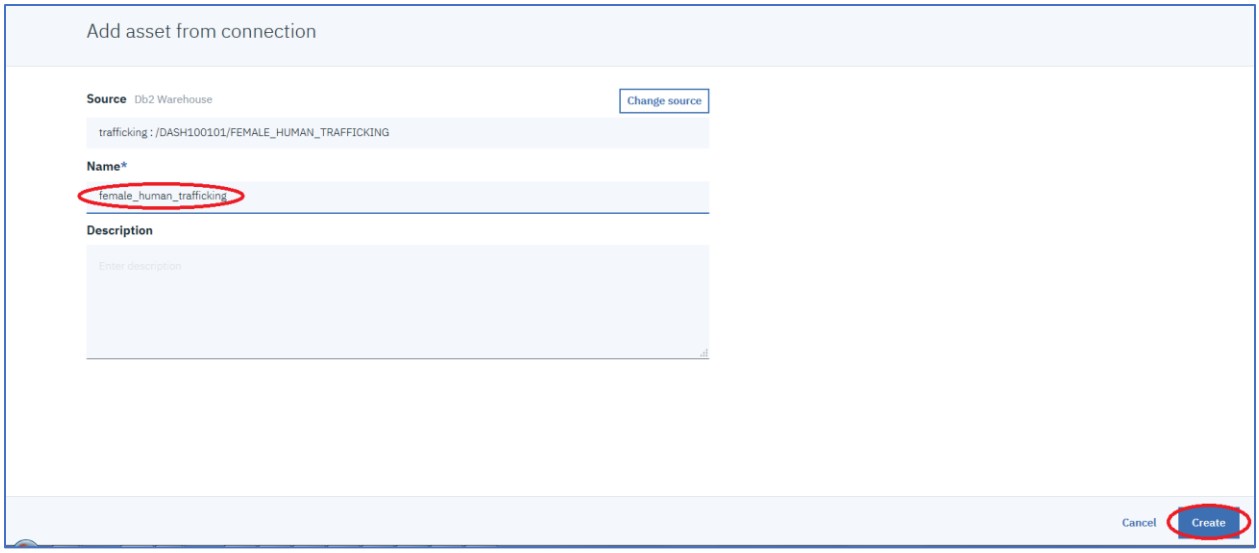
14. Click on **Select Source**.



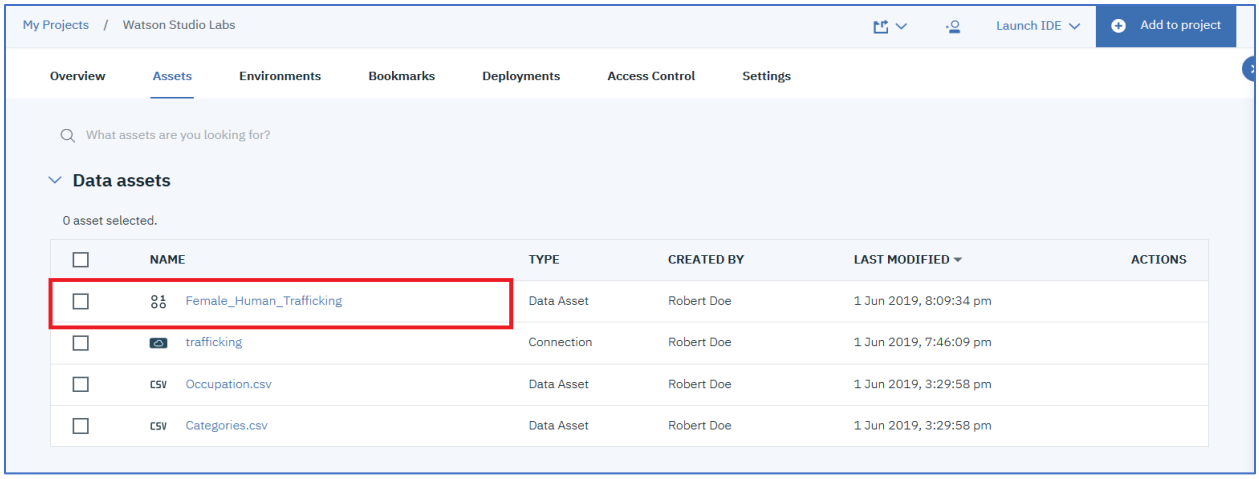
15. Click on the **trafficking** connection, click on the **Schema name** (should be DASHnnnnn where "nnnnn" is a number), click on FEMALE_HUMAN_TRAFFICKING, and then click **Select**.



16. Enter the **female_human_trafficking** for the **Name** of the Connected asset and click **Create**.



17. The Connected asset is added to the project.



You have completed Lab-1!

- ✓ Created a project
- ✓ Created an object storage instance and associate it with the project
- ✓ Associated an existing Watson Machine Learning service instance with the project
- ✓ Added a collaborator to the project
- ✓ Researched topics by searching the Community
- ✓ Added data assets to the project