

Introduction to Watson Knowledge Catalog

Introduction

This lab will introduce Watson Knowledge Catalog. Watson Knowledge Catalog is a secure enterprise catalog to discover, catalog and govern your data/models with greater efficiency. The catalog is underpinned by a central repository of metadata describing all the information managed by the platform. Users will be able to securely share data with their colleagues more easily, regardless of what the data is, where it is stored, or how they intend to use it. In this way, the intelligent asset catalog will unlock the value held within that data across user groups helping organizations use this key asset to its full potential.

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 **Find Governed Data** activity and demonstrate **Connect to Data**.

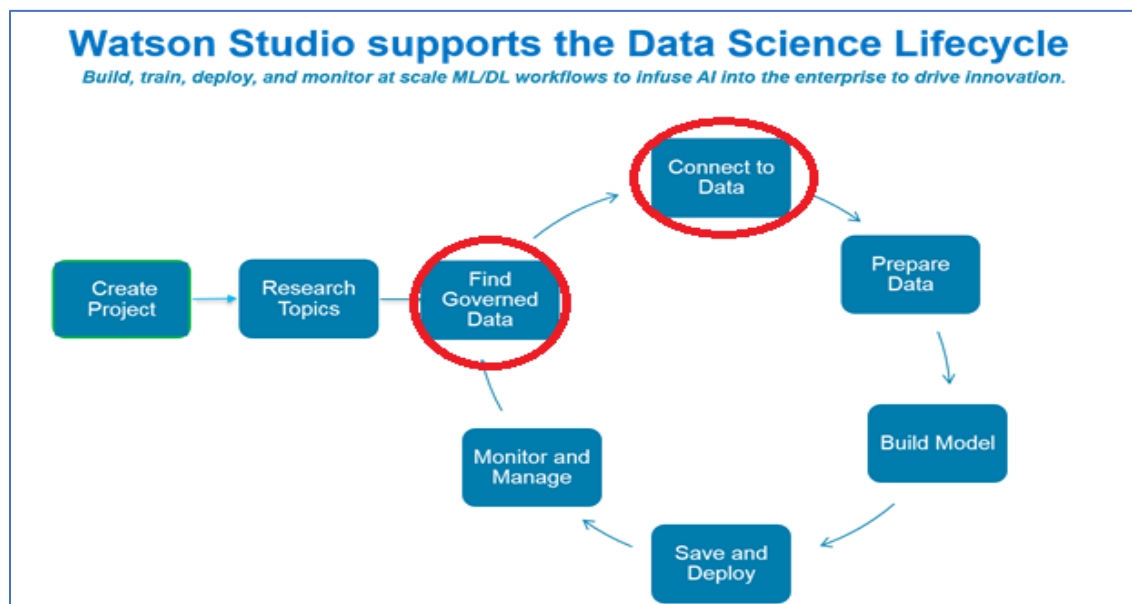


Figure 1- End to End Flow

Objectives

The goal of the lab is for the users to gain familiarity with the features of the Watson Knowledge Catalog. We will perform the following catalog tasks:

1. Create a governed catalog
2. Add a member to the catalog
3. Add Data Assets to the catalog
4. Search the catalog
5. Edit/Review/Profile a structured Data Asset
6. Demonstrate access control features

7. Demonstrate policy creation and enforcement (optional)
8. Push the Data Assets to the project set up in Lab-1

Download Files

Step 1: Download the trafficking_data.zip from the github repository

1. Click [here](#), to download the zipped file.
2. Extract the file contents. You should have two files extracted, (1) Categories.csv, and (2) Occupation.csv.

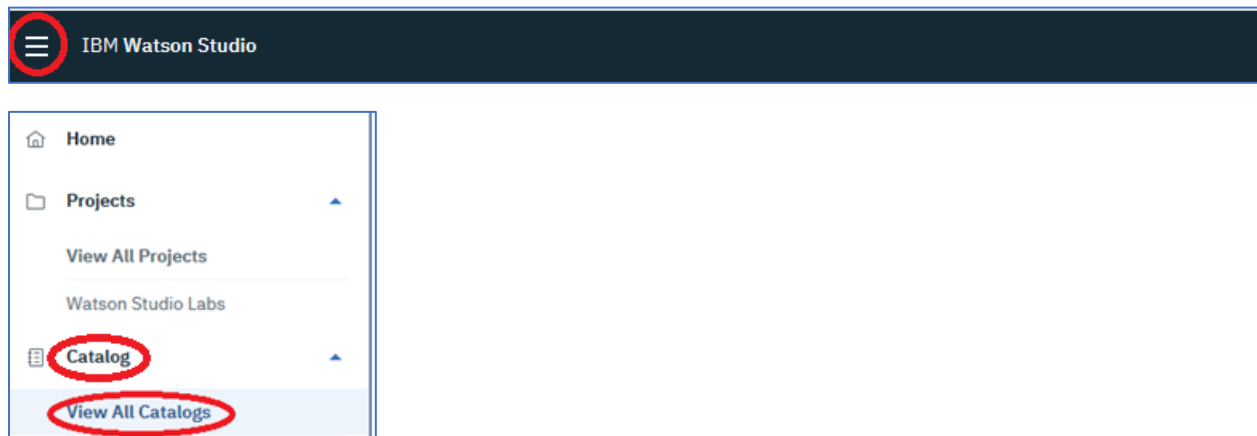
Perform Catalog Tasks

Step 1 – Create a Governed Catalog

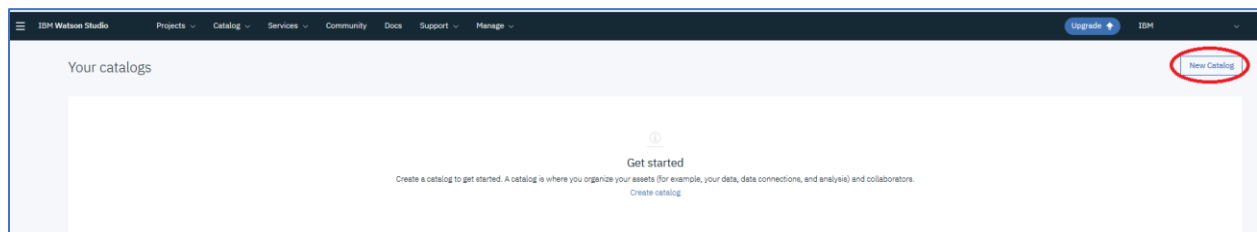
1. Click on the **IBM Watson Studio** browser tab to return to Watson Studio.
2. Click on **Watson Studio** to go to the home screen.



3. Click on  icon, click on **Catalog**, and then **View All Catalogs**.



4. Click on **New Catalog**.



5. Enter **Watson Studio Labs** for the **Catalog** name. Click on **Permanently enforce data policies**. You will get a pop-up as shown below. Click on **OK**.

IBM Watson Studio

Projects Catalog Services Community Docs Support Manage

New catalog

Details

Name*
Watson Studio Labs

Description

IBM Cloud Object Storage ⓘ
This is for storing files that are associated with assets.

Object storage instance*
cloud-object-storage-cv

Or click here to create an additional instance. Click here to refresh the inst

Permanently enforce data policies

☐ Enforce data policies
Can't be undone. You can't disable policy enforcement for a catalog after you enable it.

⚠ Permanently enforce data policies

Are you sure you want to enforce data policies for this catalog? Enforcement is permanent! If you're not sure, uncheck the Enforce data policies check box. When you're ready to start enforcing policies, go to the catalog Settings page.

Cancel OK

Note: By default, access to data assets in a catalog are only restricted by the privacy settings of the data assets. Privacy settings and policy rules can limit which members of the catalog can view and use the assets. You can implement data policies to restrict access to data based on the contents of the data. Data policies help you control data access and ensure that the right people can access the right data. Selecting the option to **enforce data policies** enables the enforcement of data policy rules to allow or deny access to a data asset or mask, anonymize and redact data at the data asset field level.

Setting this option for a catalog is a good best practice. Once it is enabled it cannot be undone, but it does not restrict or impede any functionality, it provides additional security measures to protect data assets.

6. Click **Create** to create the catalog.

New catalog

Details

Name*
Watson Studio Labs

Description

IBM Cloud Object Storage ⓘ
This is for storing files that are associated with assets.

Object storage instance*
cloud-object-storage-cv

Or click here to create an additional instance. Click here to refresh the instances.

Permanently enforce data policies

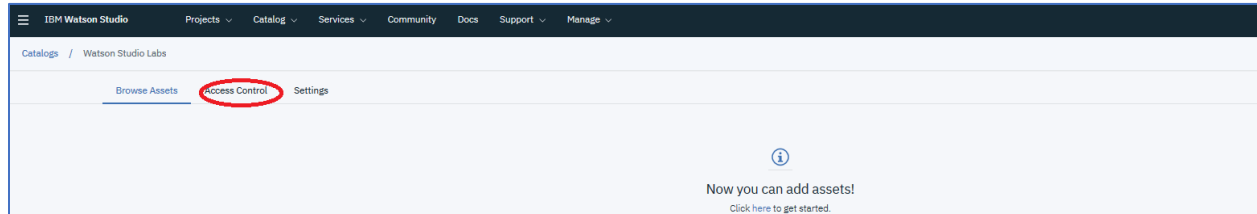
☒ Enforce data policies
Can't be undone. You can't disable policy enforcement for a catalog after you enable it.

Cancel Create

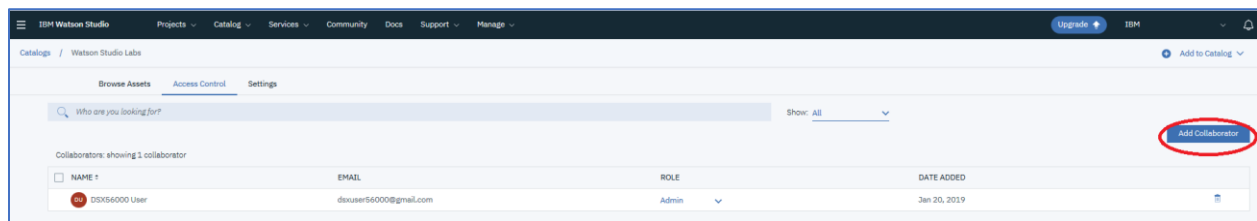
Step 2: Add a Member to the Catalog

To demonstrate the Catalog access control capability, we will add the **CATALOG COLLABORATOR**, which was provided by the lab instructor on an index card, as a member of the catalog. For illustrative purposes, we are using wsuserblb@gmail.com in this document. This may or may not be the e-mail address for the **CATALOG COLLABORATOR** on the index card. Note that the name of this user is “John Doe”.

1. Click on **Access Control**.



2. Click on **Add Collaborator**.



3. Enter the first name of the CATALOG COLLABORATOR from the index card. The list should populate with the full name of the CATALOG COLLABORATOR and the CATALOG COLLABORATOR's e-mail address. For illustrative purposes we are showing John Doe with e-mail address wsuserblb@gmail.com. You may have been provided with a different name, and password. Click on the **John Doe** entry in the list box.

×

Add Collaborators as Viewers

Collaborators (0)

John

John Doe

wsuserblb@gmail.com

Having trouble adding collaborators?

?

Cancel

Add

4. Leave the Viewer as the **Role** and click on **Add**.

Add Collaborators as Viewers

Collaborators (1)

John Doe ✕

Role: Viewer

▼



Having trouble adding collaborators?

?

Cancel

Add

5. “John Doe” is added as a member of the catalog.

Collaborators: showing 2 collaborators				
<input type="checkbox"/>	NAME ↕	EMAIL	ROLE	DATE ADDED
	DSX59000 User	dsxuser59000@gmail.com	Admin ▼	Jan 30, 2019
	John Doe	wsuserbb@gmail.com	Viewer ▼	Feb 5, 2019

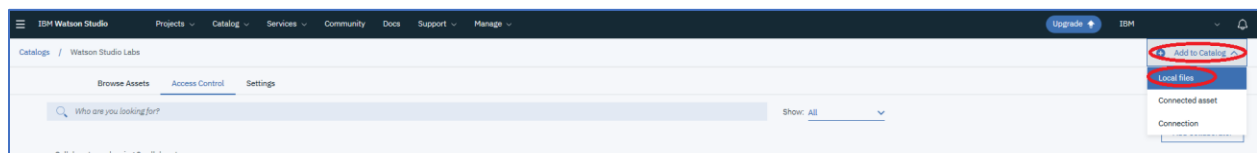
Step 3: Add Data Assets to the Catalog

In this step, we will add two local files, a connection, and a connected data asset. The local files we will add are the “Occupation.csv” file, and the “Categories.csv” file. We will set up a connection to a DB2 Warehouse service. We will add a connected data asset that will point to the FEMALE_HUMAN_TRAFFICKING table in the DB2 Warehouse service instance.

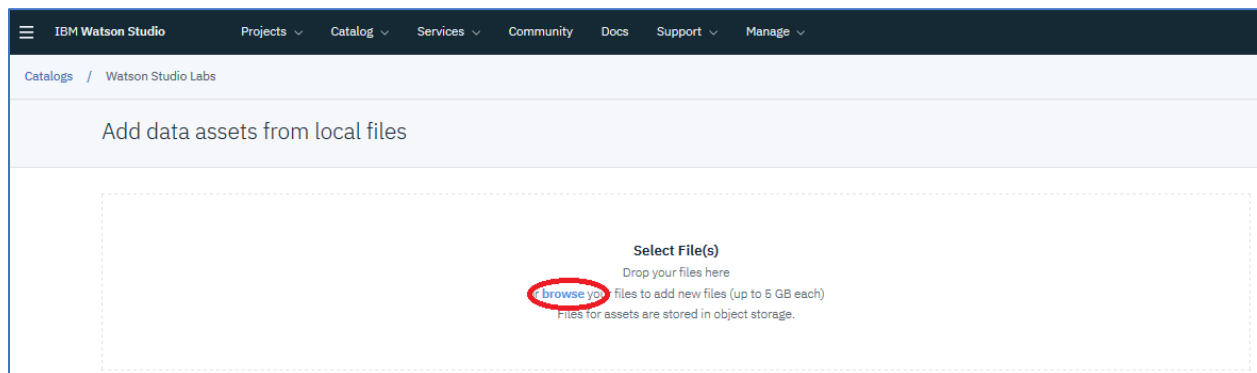
This step will introduce you to the three methods available to discover and catalog data assets; **Local files**, **Connected asset** and **Connection**. You will use these methods to catalog data assets into the newly created Knowledge Catalog and then tag them for users to easily find them, understand their content and make them available throughout IBM Watson Studio, for use during data preparation and within models, dashboards and notebooks as we will see in our lab exercises.

Add Local Files

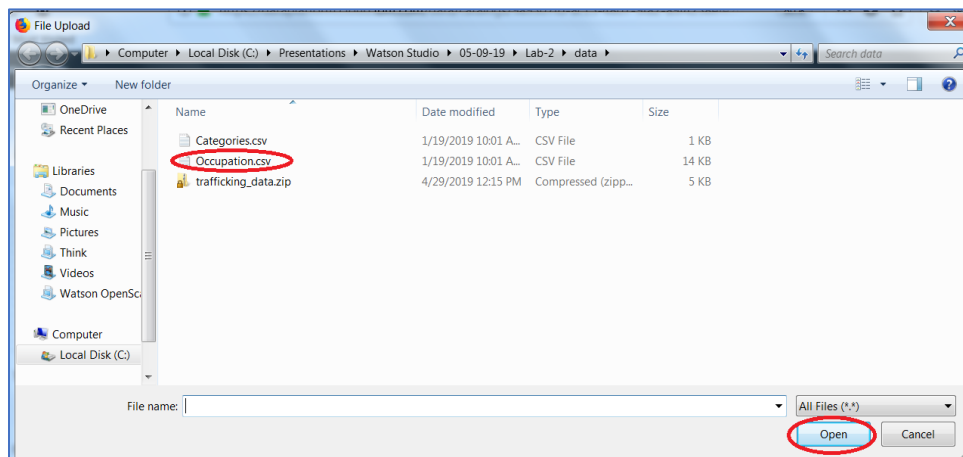
1. Click on **Add to Catalog** and click on **Local Files**.



2. Click on **browse** to navigate to the folder where the “Occupation.csv” file is stored.



3. Click on **Occupation.csv** and then click **Open**.



4. We can add tags to the newly added data asset to more easily search for it. Add two tags, enter “occupations”, click on the + icon, and enter “jobs” and click on the + icon. Click on **Private** for the **Privacy**. Click on **Add**.

Catalogs / Watson Studio Labs

Add data assets from local files

Selected Files (1)* [Edit name and format](#)

Continue adding files in the drop zone below or [browse](#) to select files

Asset Name	Format
Occupation.csv	Microsoft Excel

Target

Watson Studio Labs - Watson Studio Labs_DataCatalog

Description

Business Terms

Tags

Start typing to add values

occupations X jobs X

Privacy

☐ Public ☒ Private

Only the asset owner and asset members can find or use the asset.

Members

[Add members](#)

[Cancel](#) [Add](#)

Stay in the catalog until loading is complete! If you leave the catalog, the incomplete asset will be deleted.

5. The Occupation.csv asset metadata is displayed. Note a number of features on the panel. The top part of the page contains the search area. This consists of an entry field to type search text and three filters, one by data type, one by tags, and one by business term (filters shown in red below). Note the business term filter is disabled since we have not defined any business terms. Underneath the search area, are 3 tabs: **Watson Recommended**, **Highly Rated**, and **Recently added**. The square below contains assets that are displayed when one of these tabs are selected (shown in blue below). Note this area can be collapsed. At the bottom is the list of assets (shown in green).

IBM Watson Studio

Catalogs / Watson Studio Labs

[Filter](#) [Any type](#) [Any business terms](#) [Any tag](#) [Clear all](#)

Watson Recommended **Highly Rated** **Recently Added** [Collapse](#)

Data Asset
Occupation.csv

Owner: Andrew Doe
Added: Apr 30, 2019 8:36 PM
Tags: occup... jobs
☆☆☆☆☆ 0 reviews

Showing 1 of 1 item

<input type="checkbox"/>	Name	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/>	Occupation.csv	Andrew Doe	occupations jobs		Data Asset	Apr 30, 2019

6. Follow the same procedure (1-4) above to add the local file “Categories.csv” file to the catalog. For step 4, enter the tag “job types” and then click on the “+” icon, enter the tag “job categories” and then click on the “+” icon, leave the **Privacy** as Public, and then click on **Add**.

Catalogs / Watson Studio Labs

Add data assets from local files

Selected Files (1)* [Edit name and format](#)

Continue adding files in the drop zone below or browse to select files

Asset Name	Format
Categories.csv	Microsoft Excel

Target

Watson Studio Labs - Watson Studio Labs_DataCatalog

Description

Business Terms

Search Business Terms

Tags

Start typing to add values

job types X job categories X

Privacy

☒ Public ☐ Private

All catalog members can find and use the asset.

Members

[Add members](#)

Stay in the catalog until loading is complete! If you leave the catalog, the incomplete asset will be deleted.

[Cancel](#) [Add](#)

7. The Categories.csv asset is displayed

What assets are you looking for?

Filter Any type Any business term Any tag Clear all

Watson Recommends

Data Asset
Categories.csv

Owner: Andrew Doe
Added: Apr 30, 2019 9:03 PM
Tags: job ty... job ca...

☆☆☆☆☆ 0 reviews

Highly Rated

Data Asset
Occupation.csv

Owner: Andrew Doe
Added: Apr 30, 2019 8:36 PM
Tags: occup... jobs

☆☆☆☆☆ 0 reviews

Showing 2 of 2 items

<input type="checkbox"/>	Name	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/>	Categories.csv	AD Andrew Doe	job types job categor...		Data Asset	Apr 30, 2019
<input type="checkbox"/>	Occupation.csv	AD Andrew Doe	occupations jobs		Data Asset	Apr 30, 2019

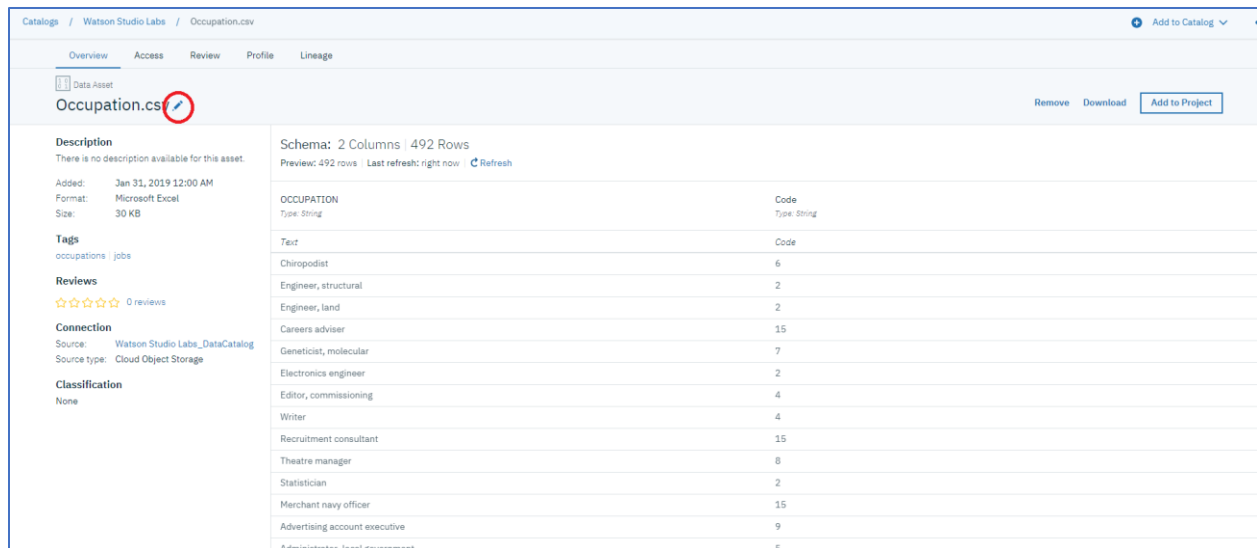
8. Change the name of the Occupation.csv asset to be just Occupation by clicking on the **Occupation.csv** asset.

Available Assets

Showing 2 of 2 assets

<input type="checkbox"/>	NAME	OWNER	TAGS	TYPE	DATE ADDED	
<input type="checkbox"/>	Categories.csv	DU DSX59000 User	job types job categories	Data Asset	Jan 30, 2019	⋮
<input checked="" type="checkbox"/>	Occupation.csv	DU DSX59000 User	occupations jobs	Data Asset	Jan 30, 2019	⋮

9. Hover the mouse on the **Occupation.csv**. Click the “pencil” icon when it appears.



Catalogs / Watson Studio Labs / Occupation.csv

Overview Access Review Profile Lineage

Data Asset

Occupation.csv

Remove Download Add to Project

Description
There is no description available for this asset.

Added: Jan 31, 2019 12:00 AM
Format: Microsoft Excel
Size: 30 KB

Tags
occupations jobs

Reviews
☆☆☆☆ 0 reviews

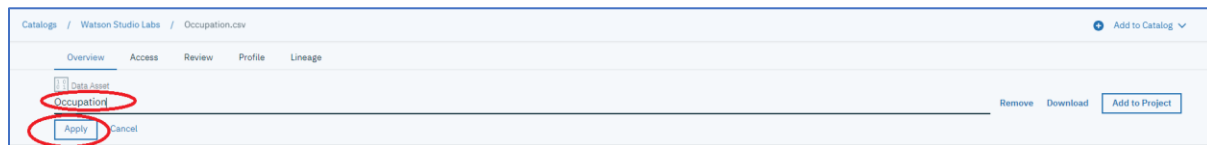
Connection
Source: Watson Studio Labs_DataCatalog
Source type: Cloud Object Storage

Classification
None

Schema: 2 Columns | 492 Rows
Preview: 492 rows | Last refresh: right now | Refresh

OCCUPATION	Code
Text	Type: String
Text	Code
Chiroprapist	6
Engineer, structural	2
Engineer, land	2
Careers adviser	15
Geneticist, molecular	7
Electronics engineer	2
Editor, commissioning	4
Writer	4
Recruitment consultant	15
Theatre manager	8
Statistician	2
Merchant navy officer	15
Advertising account executive	9
Administrator, local government	6

10. Change the name to Occupation and click **Apply**.



Catalogs / Watson Studio Labs / Occupation.csv

Overview Access Review Profile Lineage

Data Asset

Occupation.csv

Remove Download Add to Project

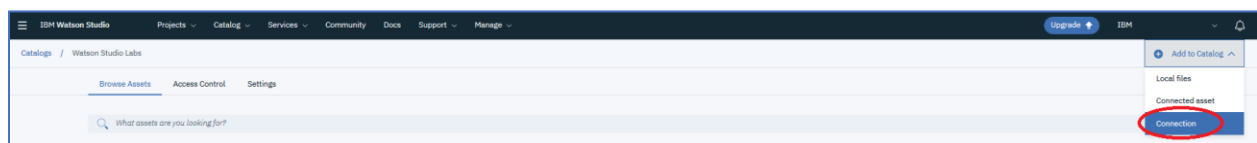
Apply Cancel

11. Click on **Watson Studio Labs** to return to the catalog display.

12. Follow steps 8-10 to change the Categories.csv asset name to be Categories.

Add Connection

1. Click on **Add to Catalog**, and then click on **Connection**.



IBM Watson Studio Projects Catalog Services Community Docs Support Manage

Upgrade IBM

Catalogs / Watson Studio Labs

Browse Assets Access Control Settings

What assets are you looking for?

Add to Catalog

Local files

Connected asset

Connection

2. The Connection page is displayed with the various connectivity options. Click on **DB2 Warehouse**.

New connection

IBM services

BigInsights HDFS	Cloud Object Storage	Cloud Object Storage (infrastructure)	Cloudant
Cognos Analytics	Compose for MySQL	Compose for PostgreSQL	Db2
Db2 Big SQL	Db2 for i	Db2 for z/OS	Db2 Hosted
Db2 on Cloud	Db2 Warehouse	Informix	Object Storage OpenStack Swift (infrastructure)
PureData for Analytics			

Third-party services

Amazon Redshift	Amazon S3	Apache Hive	Cloudera Impala
Dropbox	FTP	Google BigQuery	Google Cloud Storage
Hortonworks HDFS	Looker	Microsoft Azure Data Lake Store	Microsoft Azure SQL Database
Microsoft SQL Server	MySQL	Oracle	Pivotal Greenplum

Select a data source to begin Cancel

- Enter “trafficking” for the **Name** of the connection. Click [here](#) to obtain the DB2 Warehouse Credentials. You should have an index card that provides your assigned database warehouse. If not, ask your instructor. Locate the credentials that match your assigned database. Cut and paste the values for username, hostname, password, and database into the corresponding entry areas on the screen. Click on **Create**.

Connection overview

Name
trafficking

Description
IBM Db2 fully-managed cloud SQL database

Connection Details

Username *
cxd96032j

Hostname or IP Address *
c8d9db-ten-ssbx-yp-dal09-03.services.dal.ibmcloud.net

Secure Gateway
☐ Use a secure gateway

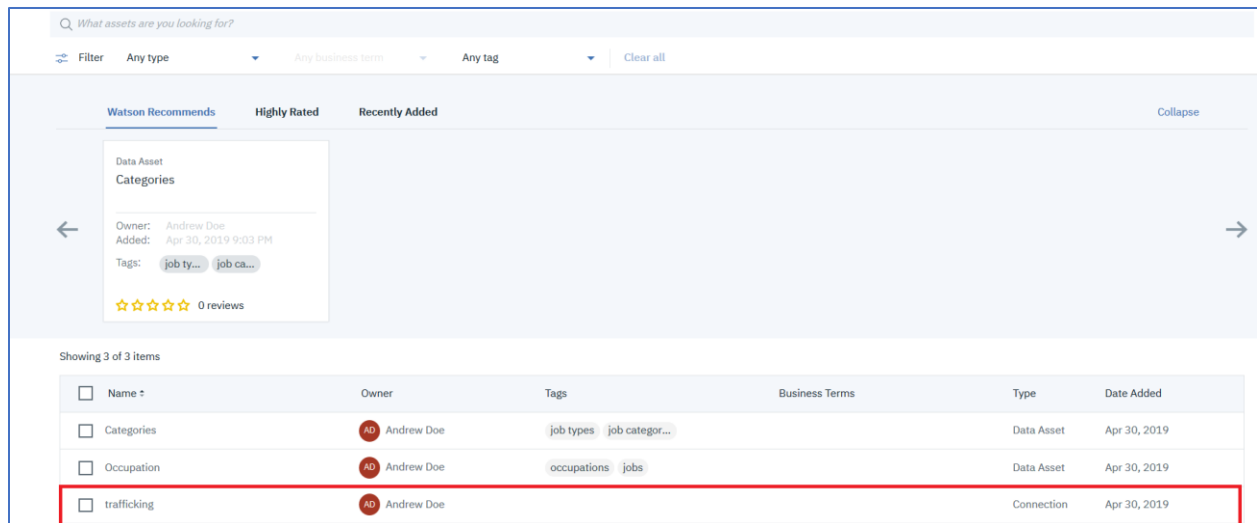
Password *

Database *
BLUDB

Connection discovery
☐ Discover data assets

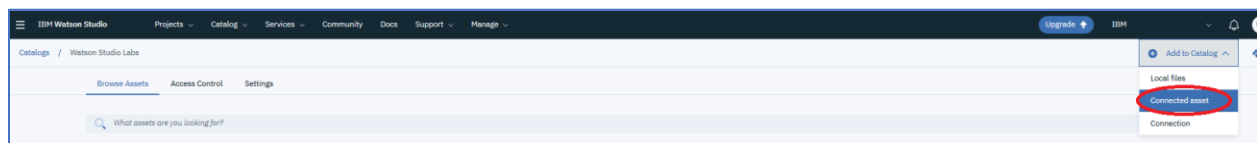
Enter information for the selected data source Cancel **Create**

- The connection asset is displayed.

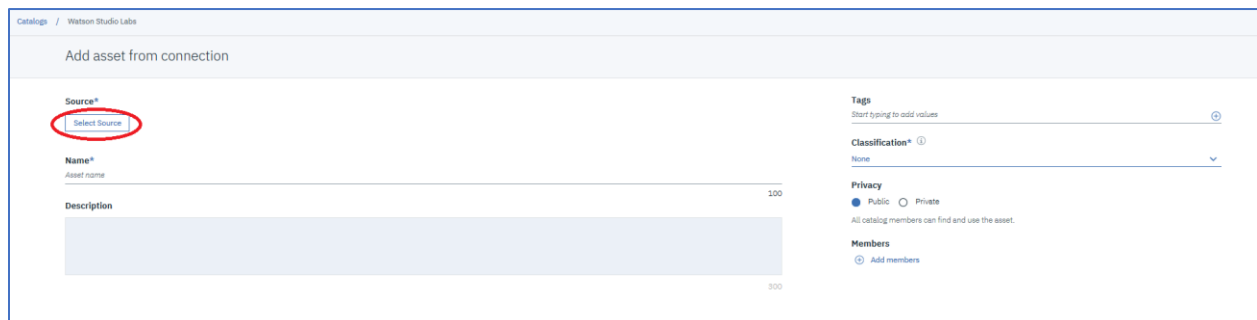


Add Connected Data

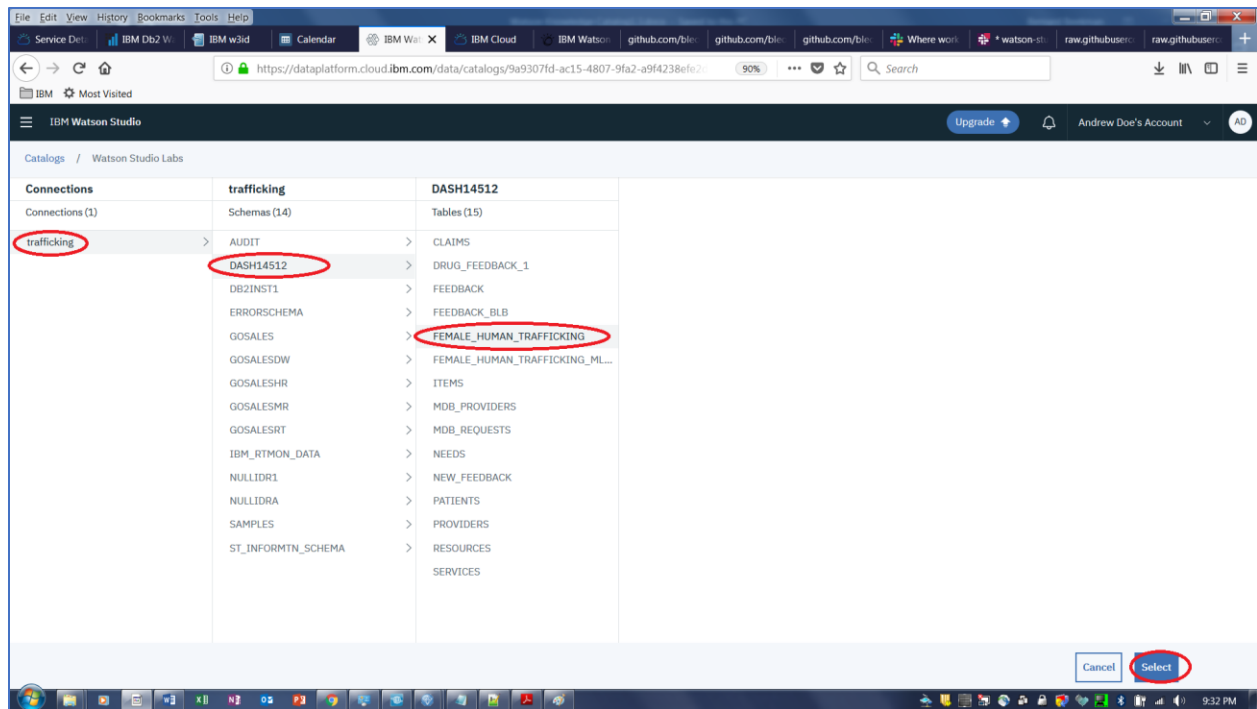
1. We have added a connection to the DB2 Warehouse service. We will now add a Connected Data Asset to point to the FEMALE_HUMAN_TRAFFICKING table in the warehouse. Click on **Add to Catalog** and click on **Connected asset**.



2. Click on **Select Source**.



3. 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**.



- Enter `female_human_trafficking` for **Name**, enter “trafficking” for **Tags**, click the “+” icon, enter “female human trafficking” for **Tags**, click the “+” icon. Click on **Private**, for Privacy, and click on **Add members**. The privacy setting of “Private” specifies that only catalog members on the access control list for this data asset will have access to the data asset.

The screenshot shows the 'Add asset from connection' form in IBM Watson Studio. The form is divided into several sections:

- Source:** Db2 Warehouse (with a 'Change source' button).
- Connection Path:** trafficking : /DASH100288/FEMALE_HUMAN_TRAFFICKING.
- Name*:** female_human_trafficking.
- Description:** A large text area for additional information.
- Business Terms:** A search bar for business terms.
- Tags:** A list of tags including 'trafficking' and 'female human traffi...'.
- Privacy:** Radio buttons for 'Public' and 'Private' (selected).
- Members:** An 'Add members' button.

 The 'Select' button at the bottom right is also highlighted.

- Enter John in the **Members** field. The list of matching users is displayed in the member list. Click on **John Doe**.

×

Add Members

Members (0)

John

John Doe	wsuserblb@gmail.com
----------	---------------------

Cancel

Add

- Click on **Add** to add John Doe to the access control list for the female_human_trafficking asset.

Add Members

Members (1)

John Doe X

You can only add people who are in the catalog.

Cancel

Add

- Click on **Add**.

Add asset from connection

Source Db2 Warehouse

Change source

trafficking : /DASH14512/FEMALE_HUMAN_TRAFFICKING

Name*

female_human_trafficking

Description

Business Terms

Search Business Terms

Tags

Start typing to add values

trafficking X

female human traffi... X

Privacy

☐ Public
☒ Private

Only the asset owner and asset members can find or use the asset.

Members

Add members

Name

John Doe

Cancel

Add

- All of the assets have now been added to the catalog. The catalog display should appear as below.

What assets are you looking for?

Filter
Any type
Any business term
Any tag
Clear all

Watson Recommends

Highly Rated

Recently Added

Collapse

Data Asset Categories

Owner: Andrew Doe

Added: Apr 30, 2019 9:03 PM

Tags: job ty... job ca...

☆☆☆☆☆ 0 reviews

Connection trafficking

Owner: Andrew Doe

Added: Apr 30, 2019 9:26 PM

☆☆☆☆☆ 0 reviews

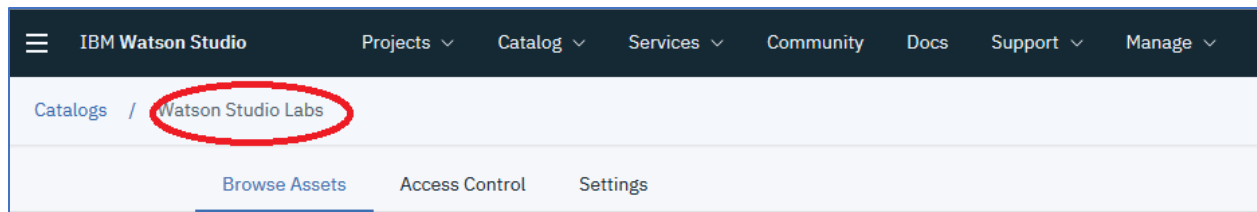
Showing 4 of 4 items

<input type="checkbox"/>	Name *	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/>	Categories	Andrew Doe	job types job categor...		Data Asset	Apr 30, 2019
<input type="checkbox"/>	Occupation	Andrew Doe	occupations jobs		Data Asset	Apr 30, 2019
<input type="checkbox"/>	female_human_trafficking	Andrew Doe	trafficking female hu...		Data Asset	Apr 30, 2019
<input type="checkbox"/>	trafficking	Andrew Doe			Connection	Apr 30, 2019

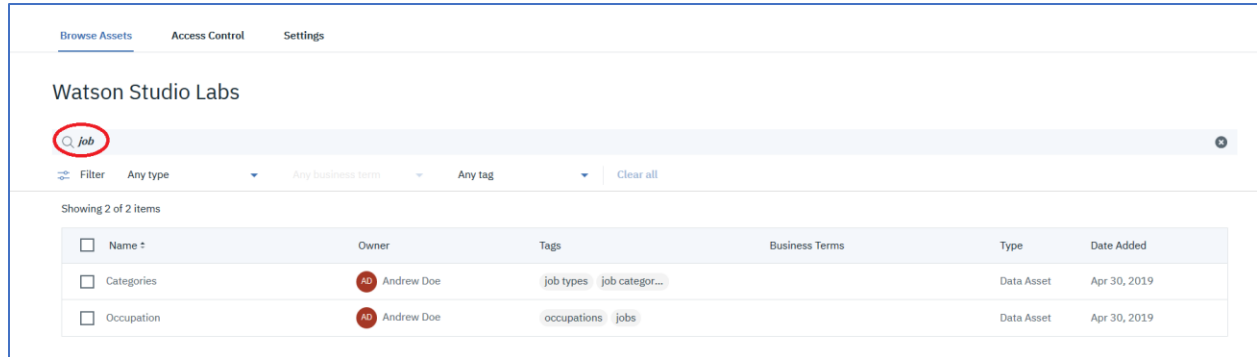
Step 4: Search the Catalog

Several options exist for searching the assets in the catalog. We can enter text in the search input text field and the system will match against asset names and tags. We can then further filter the search results by selecting the Asset Type, and/or the Tags. Or we can directly select the Asset Type or Tags without entering anything in the search text field.

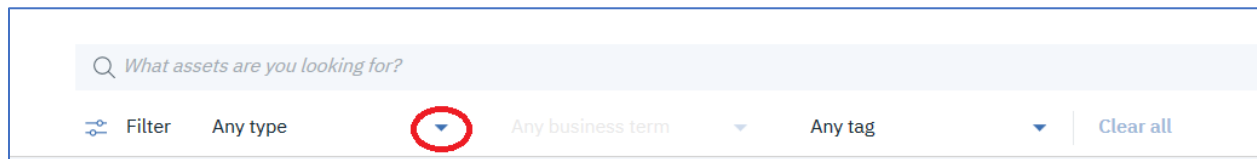
- Click on **Watson Studio Labs** to return to the Catalog home screen.



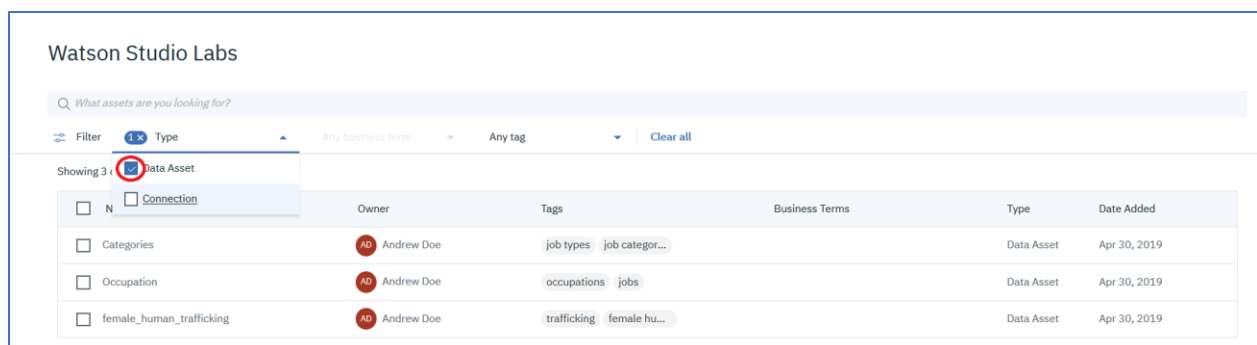
2. Enter **job** in the **search text field**. Data assets are displayed that contain the word or letters **job** in their Name, Description, or tags.



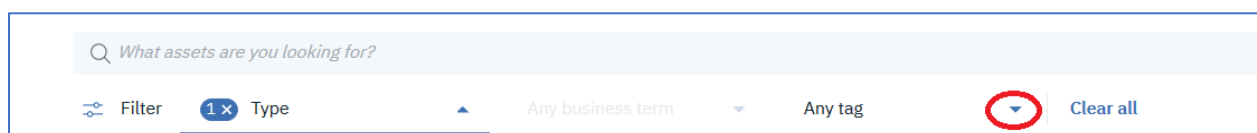
3. Clear the **search text field**. Click on the down arrow next to **Any Type**.



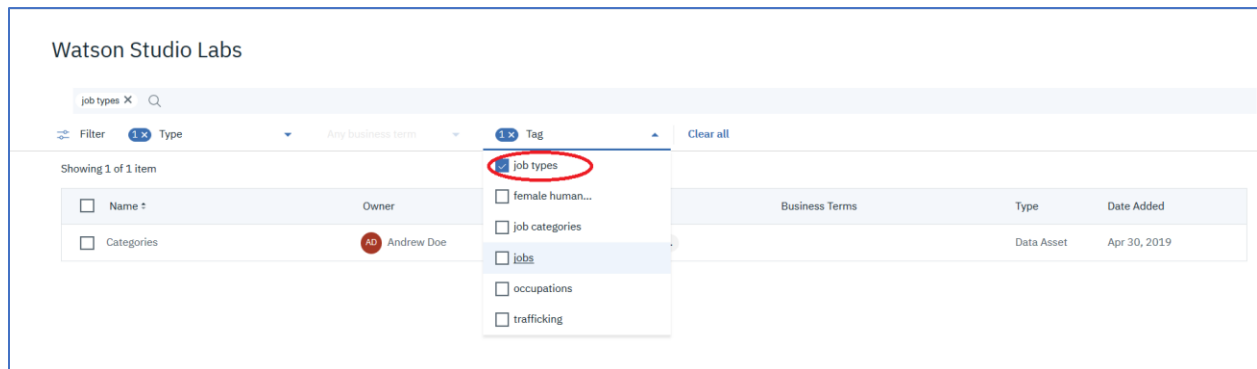
4. Click on the check box next to **Data Asset**. The search results list the 3 Data Assets. The 1 Connected asset, and the 2 uploaded files. Note: Other asset types that can be cataloged include, Connections, Models, Notebooks, and Dashboards.



5. Click on the down arrow next to **Any Tag**.



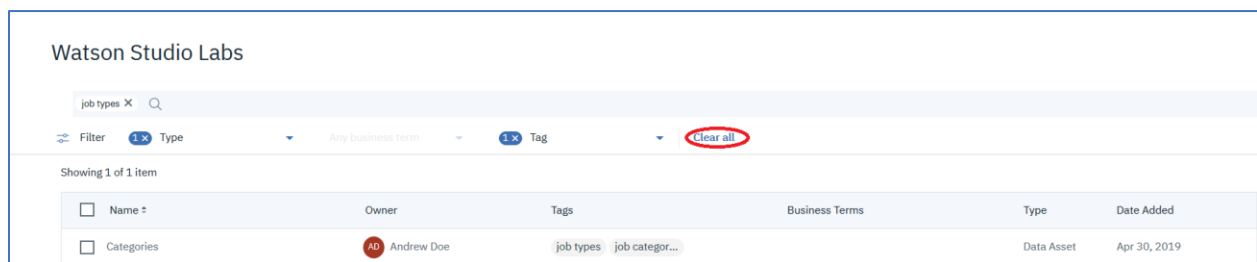
- Click on the check box next to **job types**. The Categories Data Asset is listed.



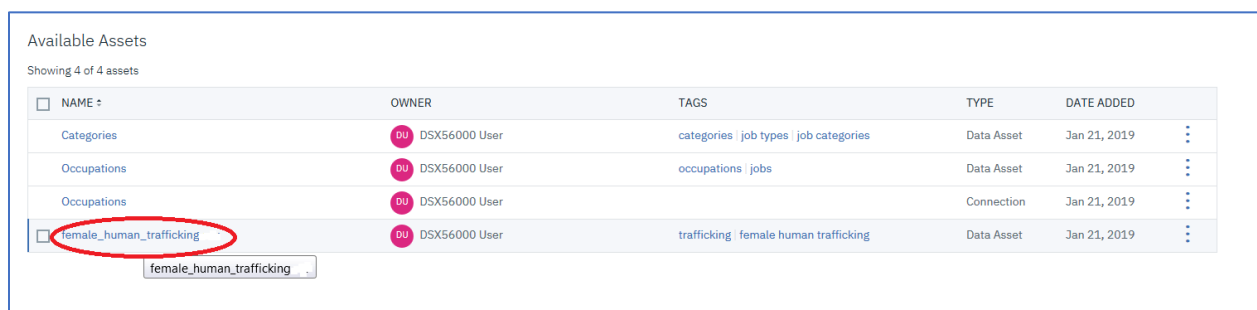
Step 5: Edit/Review/Profile Data Asset

In this step, we will drill down on the female_human_trafficking data asset to show how to edit the metadata describing the asset, to provide a rating and review of the asset, and to profile the asset.

- Click on **Clear all**.



- Click on the female_human_trafficking asset.



- The metadata and data contents of the asset are displayed as shown below. Toward the top of the screen, are 5 tabs, circled in red below. The panel shown below corresponds to the **Overview** tab. The **Access** tab allows the user to edit the access control list for this asset. The **Review** tab enables the user to provide a rating and descriptive review of the asset. These social features aid in searching for relevant and useful assets. The **Profile** tab invokes an auto-profiling capability where the columns of the data asset are classified based on machine learning scoring into one of many classifiers that are built into the

platform. This capability will be shown below. Finally, the **Lineage** tab tracks the changes made to the data asset over its lifecycle. This capability is not available in the lite catalog offering so will not be exercised in this lab.

Catalogs / Watson Studio Labs / female_human_trafficking

Overview Access Review Profile Lineage

DATA ASSET
female_human_trafficking

Remove Download Add to Project

Description
There is no description available for this asset.
Added: Apr 30, 2019 9:43 PM
Format: application/octet-stream
Size: 324 KB

Business Terms
There are no terms available for this asset.

Tags
trafficking female human trafficking

Reviews
★★★★★ 0 reviews

Connection
Source: trafficking
Source type: Db2 Warehouse

Schema: 26 Columns 1085 Rows
Preview: 1000 rows Last refresh: just now Refresh

INTERNA...	VETTING_L...	DESCRIPT...	NAME	GENDER	BIRTH_D...	BIRTH_COUN...	BIRTH_COUNTRY_C...	OCCUPATI...	ADDRESS
Type: Smallint	Type: Smallint	Type: String	Type: String	Type: String	Type: Date	Type: String	Type: String	Type: String	Type: String
785	20	NA	Sara Nunez	F	1978-12-28	Ghana	GH	Industrial/produ	7369 Kevin H
786	100	NA	Sara Schultz	F	1983-08-10	Ghana	GH	Ranger/warden	1064 Jorge M
787	100	NA	Mackenzie Emi C	F	2000-11-13	Ghana	GH	Food technologis	2296 Anthon
788	100	NA	Rea Ray	F	1996-01-13	Ghana	GH	TEFL teacher	7285 David M
789	100	NA	Becky Anthony	F	1996-01-23	Ghana	GH	Lighting technici	96566 Nancy
790	100	NA	Emily McBride	F	1996-11-27	Ghana	GH	Education officer	11227 Hood
791	100	NA	Allison Stanley	F	1994-12-04	Ghana	GH	Psychotherapist,	91579 Micher
792	100	NA	Elle King	F	1971-03-29	Ghana	GH	Photographer	6283 Cindy P

The actions circled in blue will **Remove** the asset from the catalog, **Download** the asset to the local file system, or add the asset to a catalog enabled project (**Add to Project**). Hovering over the **Name** of the data asset, the **Description**, the **Business Terms**, or the **Tags** will popup an Edit icon that allows the user to edit these entities.

4. Click on the **Review** tab.

Catalogs / Watson Studio Labs / female_human_trafficking

Overview Access Review Profile Lineage

5. Click on the fourth star to rate this asset. Enter “Good data set for demonstrating data profiling capabilities” in the **My Review** text field. Click on the **Submit** button.

Catalogs / Watson Studio Labs / female_human_trafficking

Overview Access Review Profile Lineage

DATA ASSET
female_human_trafficking

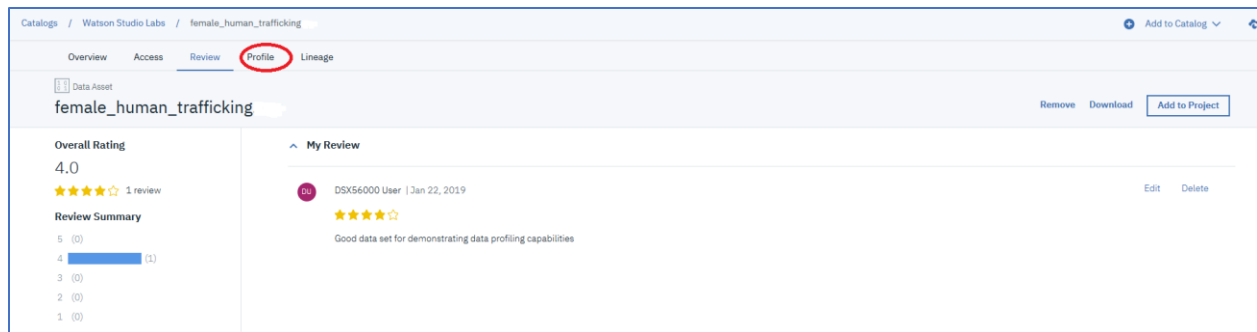
Remove Download Add to Project

Overall Rating
0.0
★★★★★ 0 reviews

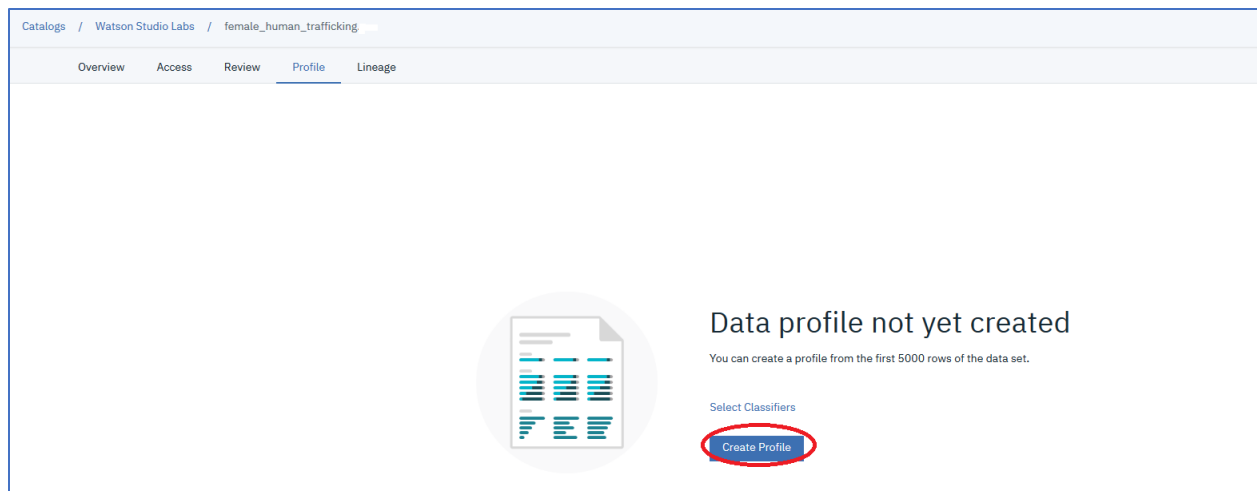
Review Summary
5 (0)
4 (0)
3 (0)
2 (0)
1 (0)

My Review
DSX56000 User | Jan 22, 2019
★★★★★
Good data set for demonstrating data profiling capabilities
Cancel Submit

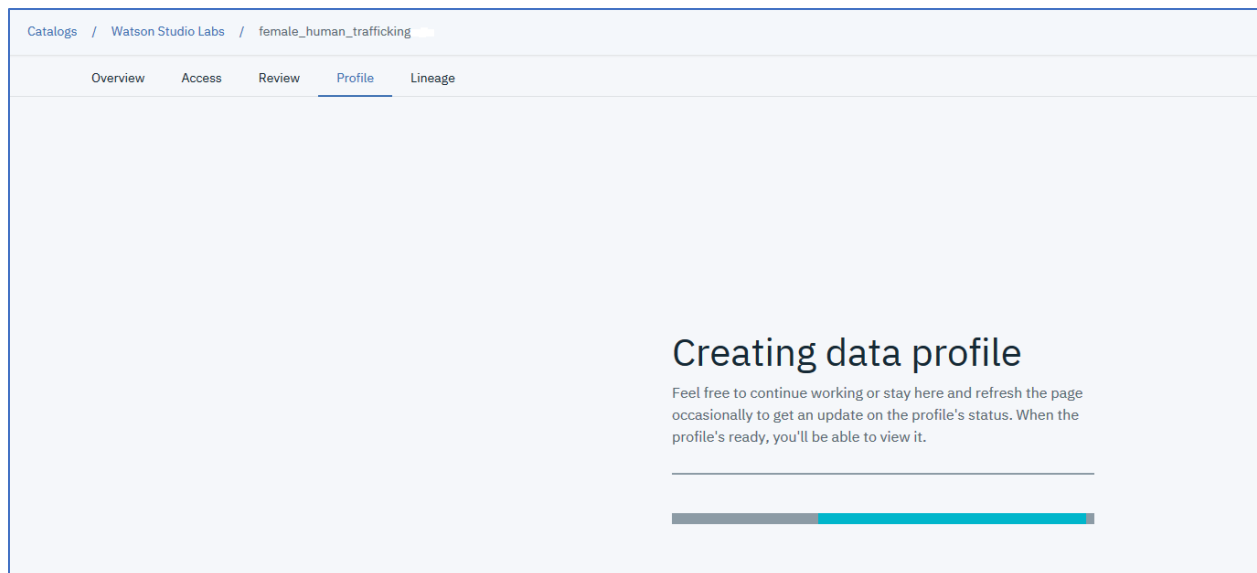
6. The review of the asset is displayed. Click on the Profile tab.



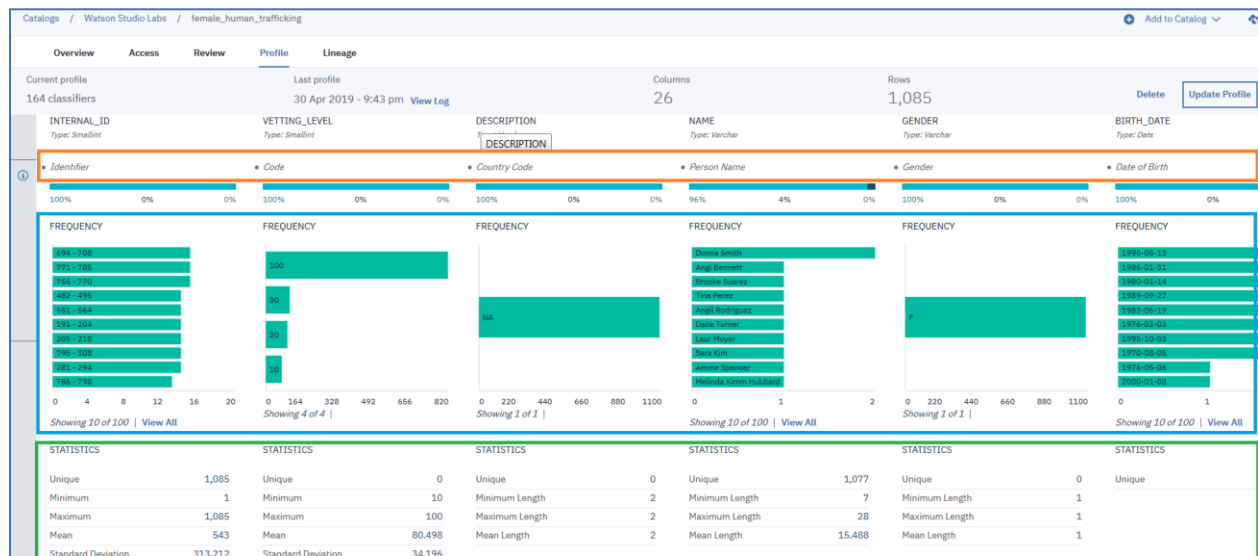
7. Click on **Create Profile**. If a profile was already created, it would be displayed instead of this prompt screen.



8. This step could take some time. Please wait a minute or so, and then click on browser refresh.



9. The profile information is displayed below using several screenshots, based on horizontal scrolling, to display many of the columns of the data asset. The highlighted red box contains the inferred column classifications. The highlighted blue box contains counts of the number of occurrences of each value (for discrete fields), and a histogram for continuous fields. The green box contains statistics calculated from the contents of the columns.



The classifications of the columns are listed below:

- INTERNAL_ID – Identifier
- VETTING_LEVEL – Code
- DESCRIPTION – Country Code – this is an error due to the description input as NA. We can correct this misclassification.
- NAME: PersonName – the classification accurately determines that this field is the name of a person.
- GENDER: Gender
- BIRTH_DATE: Date of Birth – the classification accurately determines that this field represents a date of birth.

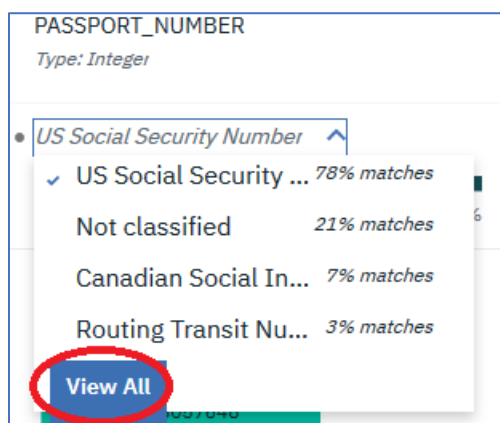
- COUNTRY_VISITED_COUNT – Code – this is an error. It’s not a code. This can be re-classified if necessary.
- ARRIVAL_AIRPORT_COUNTRY_CODE – Country Code – correctly classified.
- ARRIVAL_AIRPORT_IATA: Airport Code – the classification accurately determines that this field represents an Airport Code.

There are more columns in the data set but this gives a representative sample of the profiling capability. The classification of the columns becomes important when we create and enforce security policies.

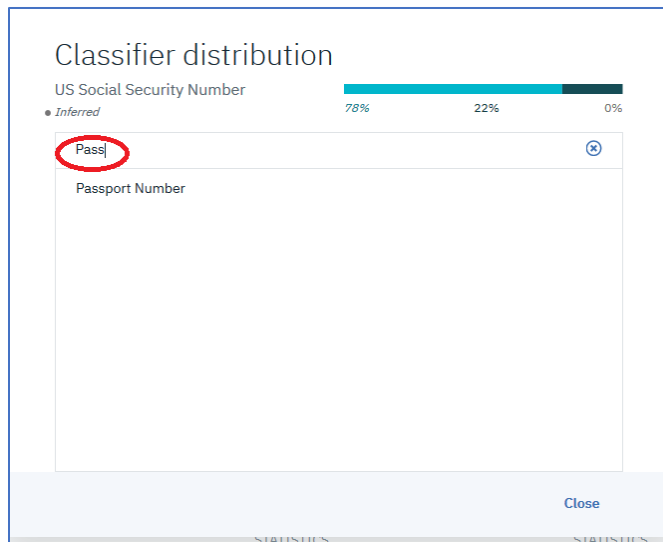
10. Let’s fix the classification for passport. Hover the mouse over the US Social Security Number in the **PASSPORT_NUMBER** Column and click on the  icon.



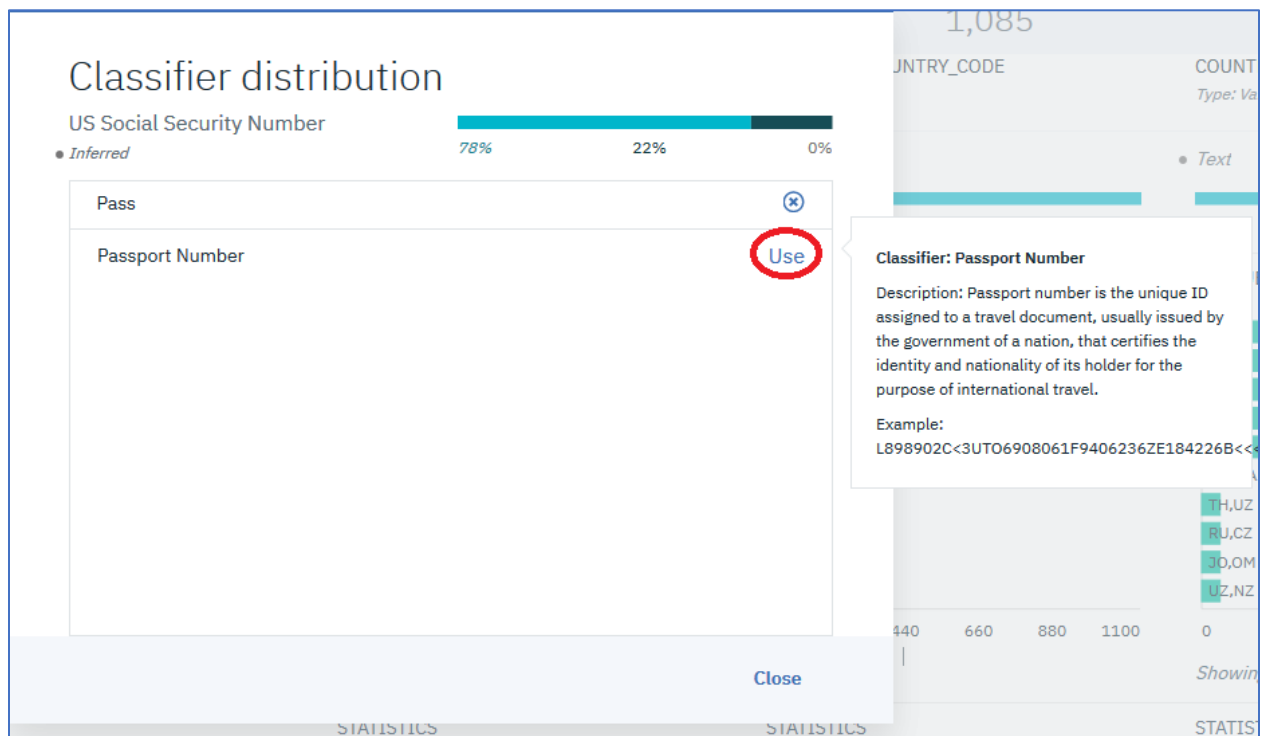
11. Click on **View All**.



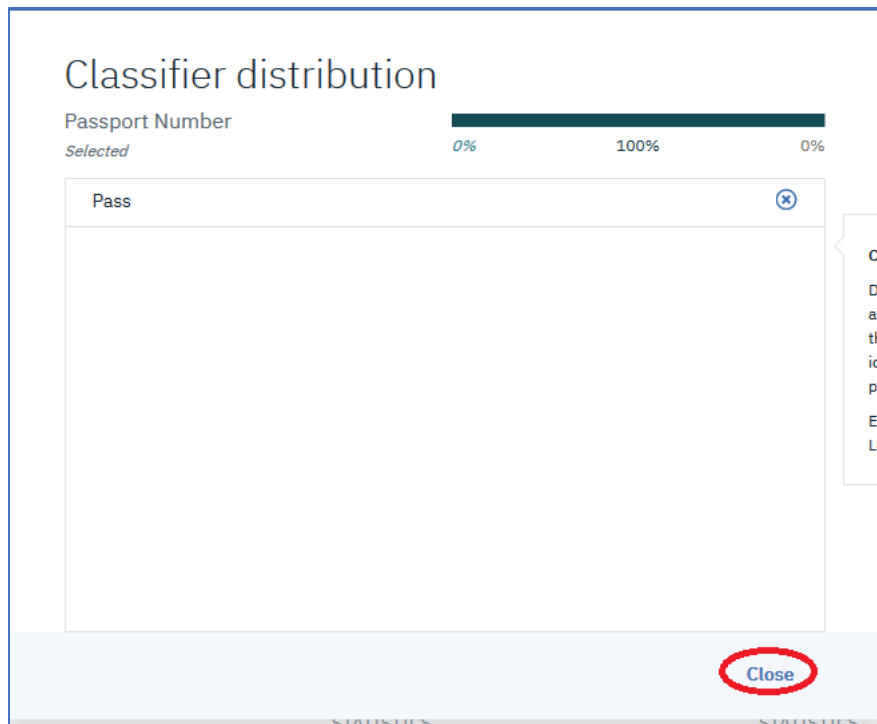
12. Type “Pass” in the **Search** field.



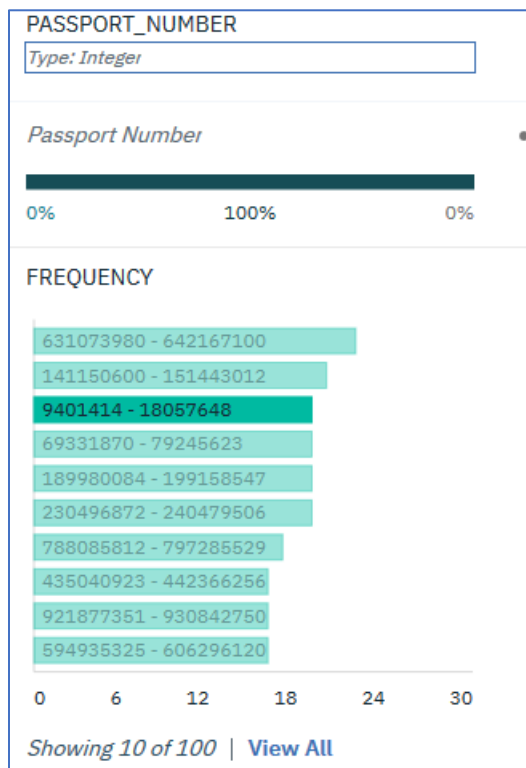
13. Hover over Passport Number and click on **Use**.



14. Click **Close**.



15. The column classification has been changed.



Step 6: Access Control

All Catalog members can access assets with a privacy setting of Public. Assets that have a privacy setting of Private are secured based on the asset access control list. A summary of the access control settings for our Watson Studio Labs catalog and the contained Data Assets is shown below.

Watson Studio Lab Catalog members:

1. Catalog owner (logged on user)
2. E-mail of user provided by lab instructor (used here for illustrative purposes - wsuserblb@gmail.com)

Data Asset Privacy:

1. female_human_trafficking (Private) ACL Members:
 - a. asset owner (logged on user)
 - b. e-mail of user provided by lab instructor (used here for illustrative purposes - wsuserblb@gmail.com)
2. Occupations (Private) ACL Members:
 - a. asset owner (logged on user)
3. Categories (Public)
4. trafficking (Connection) (Public)

We would expect that if we log in as wsuserblb@gmail.com, that only 3 Data Assets would appear. The two public Data Assets, and the Private Data Asset where wsuserblb@gmail.com is on the access control list. This user would not have access to the Occupations asset, since it is a Private asset, and the user is NOT on the access control list for this asset. To verify complete the following steps.

1. Open a different browser (if you are using FireFox, then open Chrome or vice versa).
2. Type in datascience.ibm.com for the URL to navigate to.
3. Click on **Sign into IBM Watson Studio Cloud**



4. Login using wsuserblb@gmail.com for the user name and click **Continue**.

Log in to IBM

IBMid [Forgot IBMId?](#)

☐ Remember me [i](#)

Don't have an account? [Create an IBMId](#)

5. Enter **wsPassw0rd** for the **Password** and click **Log in**.

Log in to IBM

Logging in as wsuserblb@gmail.com [Not you?](#)

Password [Forgot password?](#)

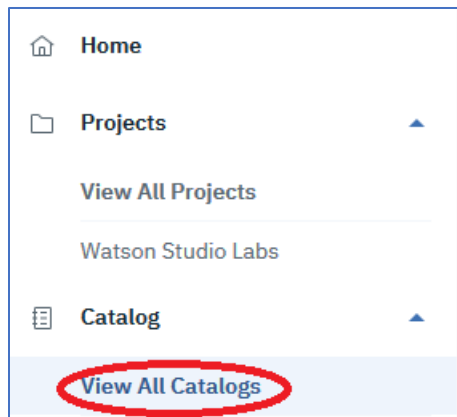
[i](#)

☐ Remember me [i](#)

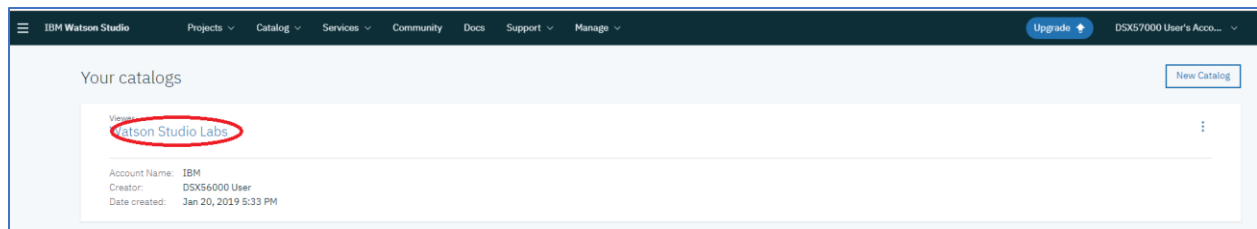
Don't have an account? [Create an IBMId](#)

Need help? [Contact the IBMId help desk](#)

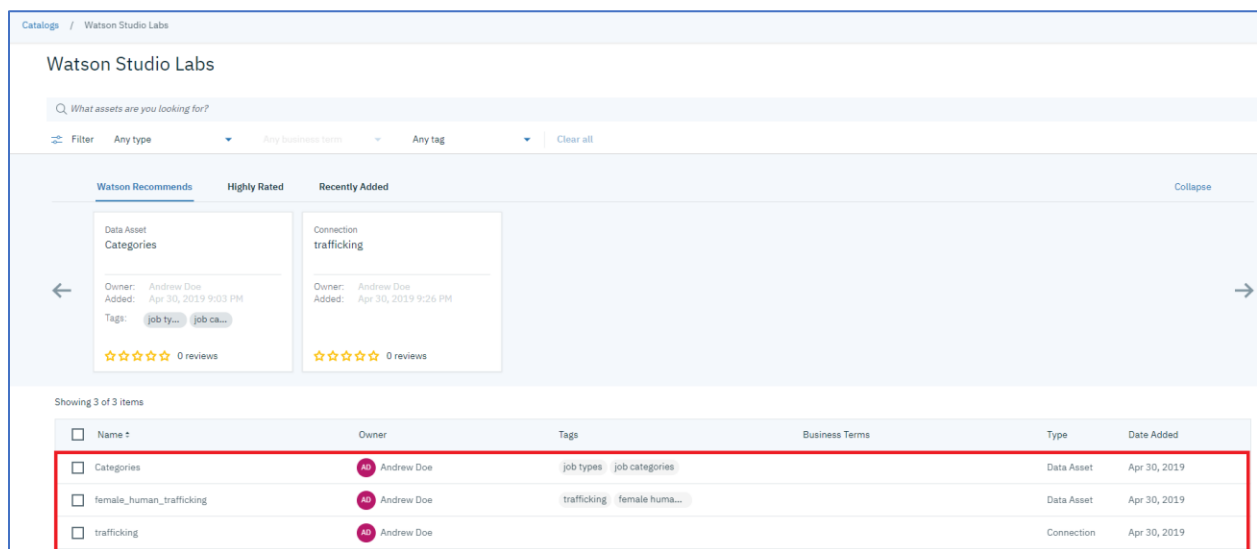
6. Click on  icon. Click on **Catalog** and then **View All Catalogs**



7. Scroll down to find the **Watson Studio Labs** catalog that corresponds to your username. Note that there will be 1 catalog per lab participant.




8. The Watson Studio Labs catalog is displayed. Three Data Assets are listed as expected.

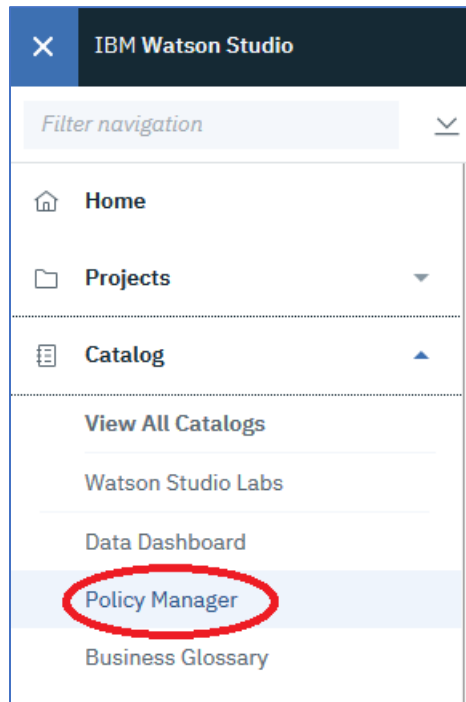


9. Leave this browser at this screen for Step 7.

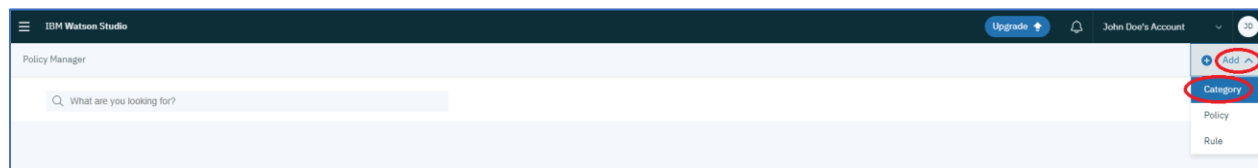
Step 7: Create and Enforce Policy (optional)

In this step, we will create a policy that will anonymize the SSN and Passport fields when user wsuserblb@gmail.com tries to access the female_human_trafficking data asset.

1. Return to the browser where **your** Watson Studio id is logged in.
2. Click on the  icon. Click on **Catalog** and then **Policy Manager**.



3. Click on **Add** and then **Category**.



4. Type in **Privacy** for the **Name** of the Category, and optionally a **Description**, and click on **Create**.

New category

Details

Name*

Privacy

Description

Policies that enforce privacy

Cancel Create

5. Click on **Add** and then **Policy** to add a policy to the Privacy Category.

IBM Watson Studio

Policy Manager / Privacy

Upgrade

DSX59000 User's Acco...

DU

Category

Privacy

Description

Policies that enforce privacy

Creator: DSX59000 User Created: Feb 5, 2019 Contains: 0 published policies and 0 rules

Add

Delete

Category

Policy

Rule

6. Type in PII for the Policy **Name**, type in “Limit access to PII data” for the Policy **Description**, click on **create a new rule**.

New policy

Details

Name*

PII

Type*

Access

Category*

Privacy

Description

Limit access to PII data

Rules

No rules in policy!

To add rules to this policy,

create a new rule select from existing rules

Cancel Save as draft Publish

7. Enter “Redact Government ID” for the rule **Name**, enter “Redact a Government ID” for the rule **Description**, for the rule click in the entry fields to select the following:

If **Data Class** contains any

Government Identities


And

If **User Name** contains any
John Doe

then **anonymize data** where **in columns containing**
Government Identities

The screenshot displays the IBM Watson Studio Rule Builder interface. On the left, the 'Details' panel shows the rule name 'Redact Government ID', type 'Access', and business definition 'Redact a Government ID'. The main 'Rule builder' panel is divided into 'Criteria' and 'Action' sections. The 'Criteria' section, labeled 'CONDITION 1', contains two conditions connected by an 'And' operator: 'Data class' is 'Government Identities X' with the operator 'contains any', and 'User Name' is 'John Doe X' with the operator 'contains any'. The 'Action' section, labeled 'Action 1', shows the action 'anonymize data' with the operator 'in columns containing' and the target 'Government Identities X'. Red circles highlight the key configuration elements: 'Redact Government ID', 'Redact a Government ID', 'Data class', 'Government Identities X', 'contains any', 'User Name', 'John Doe X', 'contains any', 'anonymize data', 'in columns containing', and 'Government Identities X'.

8. Select how to anonymize data: **Substitute**, and then click **Create**.


Action * 

then **anonymize data** in columns containing

Government Identities X

Select how to anonymize data:

Redact
BEFORE
452-821-1120
Replace data with Xs.

Substitute 
BEFORE
452-821-1120
Replace data with values that don't match the original format.

Mask
BEFORE
452-821-1120
Replace data with similarly formatted values.

[Visit the documentation](#) to learn more about data anonymization.

Cancel **Create**

9. Click on **Publish**.

New policy


Details


Name*
PII

Type*
Access

Category*
Privacy

Description
Limit access to PII Data

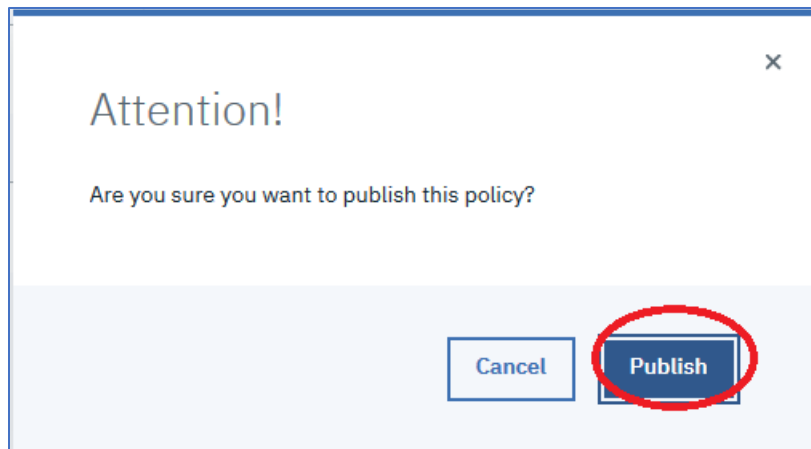
Rules 

 Add

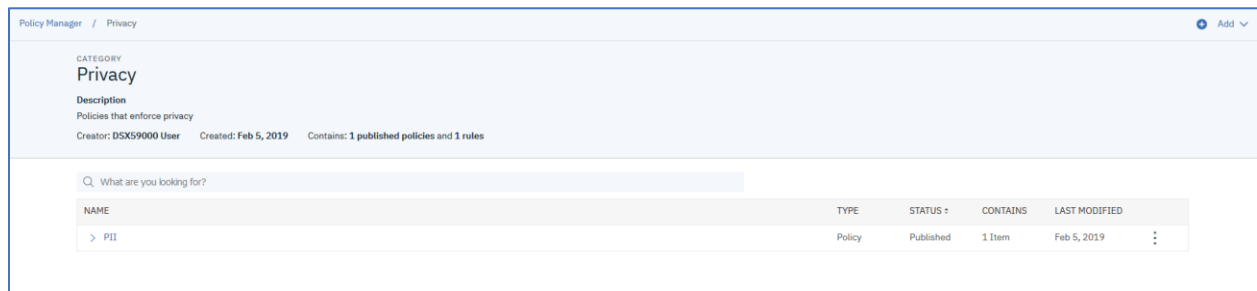
Redact Government ID
Redact a Government ID

Cancel Save as draft **Publish**

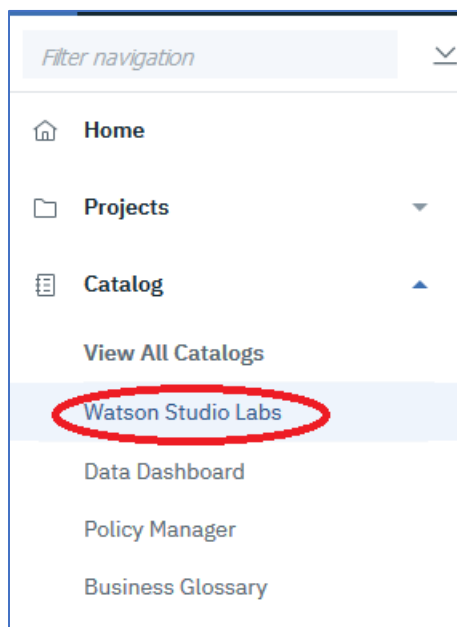
10. Click on **Publish**.



11. The PII policy has been added.



12. Click on the  icon, then click on **Catalog** and then **Watson Studio Labs**.



13. Switch back to the browser with the wsuserblb@gmail.com logged in. Click on the **female_human_trafficking** asset.

Available Assets

Showing 3 of 3 assets

<input type="checkbox"/> NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories.csv	DSX59000 User		Data Asset	Jan 31, 2019
female_human_trafficking	DSX59000 User	trafficking female human trafficking	Data Asset	Jan 31, 2019
trafficking	DSX59000 User		Connection	Jan 31, 2019

14. The female_human_trafficking asset is displayed. Wait for the anonymization to complete. Note that 2 columns are anonymized.

Data Asset

female_human_trafficking

Description
There is no description available for this asset.

Added: Jan 31, 2019 10:02 AM
Format: application/octet-stream
Size: 347 KB

Tags
trafficking | female human trafficking

Reviews
0 reviews

Connection
Source: Watson Studio Labs_DataCatalog
Source type: Cloud Object Storage

Classification
Personally Identifiable Information
Personally identifiable information (PII) is defined as any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another can be considered PII.

Schema: 26 Columns | 1085 Rows | **2 Columns anonymized**
Preview: 1000 rows | Last refresh: 22 seconds ago | Refresh

DATE	BIRTH_COUNT...	BIRTH_COUNTRY_CODE	OCCUPATION	ADDRESS	SSN	PASSPORT_NUMBER
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
th	Country Name	Country Code	Text	Text	US Social ...	Passport Number
15	Ghana	GH	Engineer, land	824 Kristin Grv, /	afe55d1d355c3:	1c9da91e1e20863dd850
19	Ghana	GH	Editor, commissi	1148 Wang Fall :	77a0daa42ec7d	12d38855ed107e7cc5dd
16	Ghana	GH	Merchant navy of	9486 Pratt Wall,	669061087d6d:	c43ed0283a3def7031d8:
17	Ghana	GH	Paramedic	0890 Johnson Tr	997b59e501b2e	179abee5ba608418154d
18	Ghana	GH	Surveyor, buildin	2315 Brittany Cr	70329b83b40ct	84524ccc3c5c6590600e:
24	Ghana	GH	Waste managem	88811 Donald Pr	d2f2236f52407:	a730ae13f5ed96f71e904
23	Ghana	GH	Doctor, general p	9150 Donald Rpx	d2c2d41163d8f	ced1617be1d70e44421c
02	Ghana	GH	Forest/woodland	1355 Lopez Villa	62007942c2b0c	8c8debd401b6b6d954b
12	Ghana	GH	Land/geomatics :	86792 Amy Vlgs,	08f8dd9f9ba89t	a43f1d6c9cacdfa82a1a1
10	Ghana	GH	Oncologist	108 Erin Via, New	f8b871f6e058e:	f289be62078ebbe457c6:
07	Ghana	GH	Veterinary surger	79572 Schmidt E	f2006c1d30df3:	624a9605774a0cfd98aa

15. Close the browser tab for user wsuserblb.

Step 8: Push the data assets to the project

This step will add the cataloged data assets to our Watson Studio project to be used in subsequent labs. Return to your Watson Studio account.

1. Click on **Watson Studio Labs** to go to the Catalog home screen.

IBM Watson Studio

Projects | Catalog | Services | Community | Docs | Support | Manage

Catalogs / **Watson Studio Labs**

Browse Assets | Access Control | Settings

2. Click on the checkbox to the left of the **female_human_trafficking** asset and the **trafficking** asset. Click on **Add to Project**.

Add to Project + Remove 🗑️		2 of 4 items selected				Cancel
Name ▴	Owner	Tags	Business Terms	Type	Date Added	
<input type="checkbox"/> Categories	PD Peter Doe	job types job categor...		Data Asset	May 22, 2019	
<input type="checkbox"/> Occupation	PD Peter Doe	occupations jobs		Data Asset	May 22, 2019	
<input checked="" type="checkbox"/> female_human_trafficking	PD Peter Doe	trafficking female hu...		Data Asset	May 22, 2019	
<input checked="" type="checkbox"/> trafficking	PD Peter Doe			Connection	May 22, 2019	

3. Select the Watson Studio Labs as the **Target**. Click on **Add**.

Catalogs / Watson Studio Labs

Add to Project

Target*
Watson Studio Labs ▾

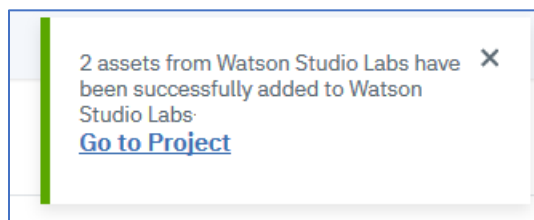
Selected assets (4)

ASSET NAME	CONNECTION	
Categories	Occupations	⊙
Occupations	Occupations	⊙
Occupations	Occupations	⊙
female_human_trafficking	Watson Studio Labs_DataCatalog	⊙

Connections to be added (1)
The following connections will be added to the project to successfully add the assets.
Occupations
IBM Db2 fully-managed cloud SQL database ⊙

Cancel Add

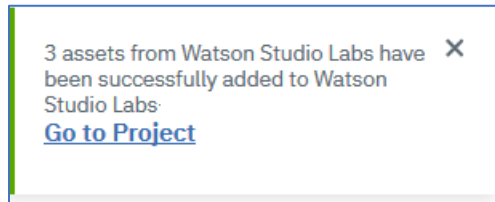
4. Wait for the message that 2 assets have been added to the project in the top right.




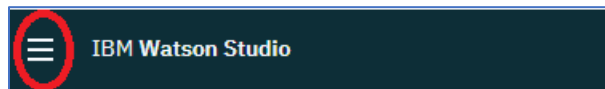
5. Click on **Categories**, and **Occupation**, and then click again on **Add to Project**. Select **Watson Studio Labs** as the **Target** and click **Add**.

Add to Project + Remove 🗑️		2 of 4 items selected				Cancel
Name ▴	Owner	Tags	Business Terms	Type	Date Added	
<input checked="" type="checkbox"/> Categories	PD Peter Doe	job types job categor...		Data Asset	May 22, 2019	
<input checked="" type="checkbox"/> Occupation	PD Peter Doe	occupations jobs		Data Asset	May 22, 2019	
<input type="checkbox"/> female_human_trafficking	PD Peter Doe	trafficking female hu...		Data Asset	May 22, 2019	
<input type="checkbox"/> trafficking	PD Peter Doe			Connection	May 22, 2019	

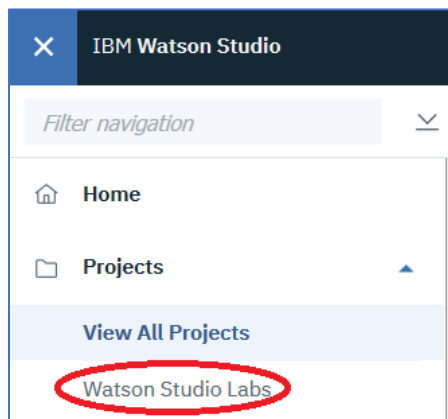
- Wait for the message that **3** assets have been added to the project in the top right. Two data assets and a connection to the catalog object storage.



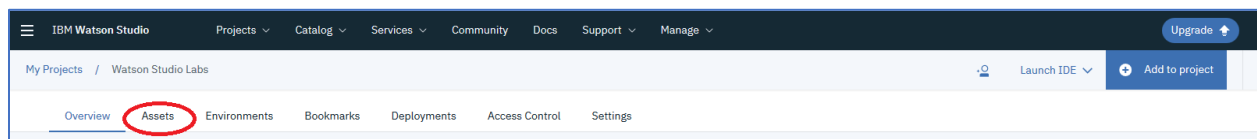
- Click on the hamburger icon .



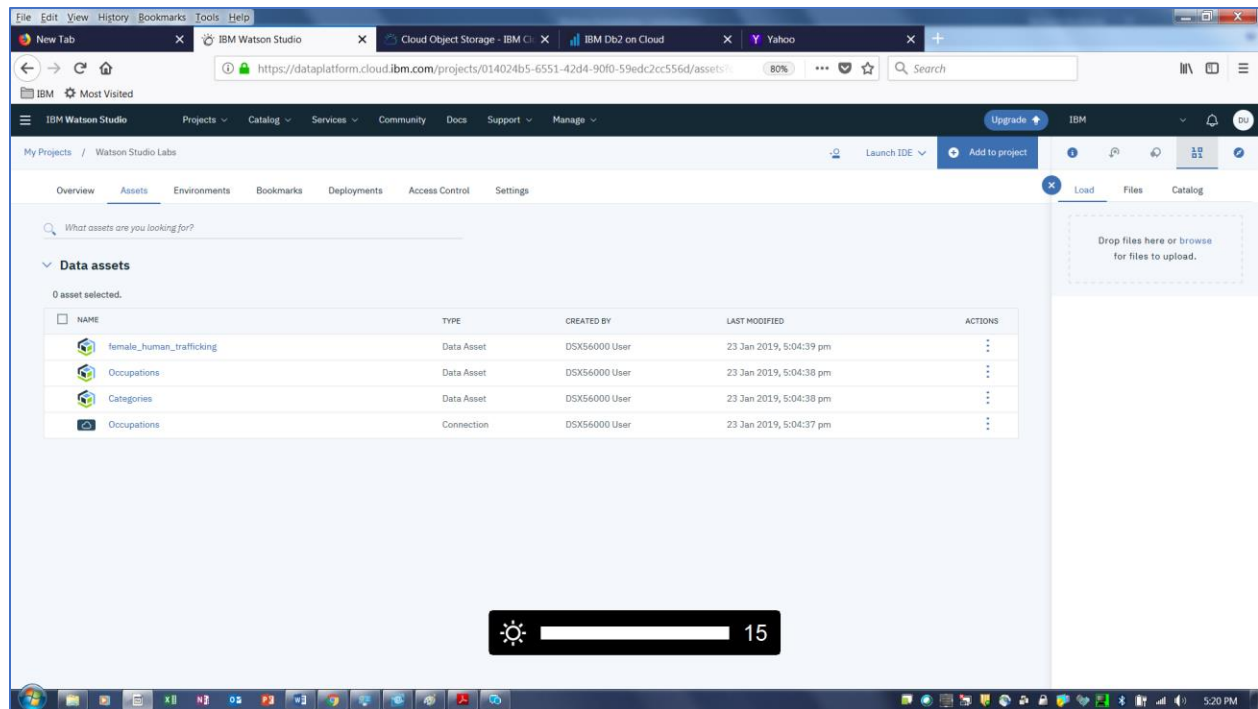
- Click on **Watson Studio Labs** under **Projects**.



- Click on the **Assets** tab.



- The Data assets are now incorporated in the project.



You have completed Lab-2!

- ✓ Created a governed catalog
- ✓ Added a member to the catalog
- ✓ Added Data Assets to the catalog
- ✓ Searched the catalog
- ✓ Edited/Reviewed/Profiled a structured Data Asset
- ✓ Demonstrated access control features
- ✓ Demonstrated policy creation and enforcement (optional)
- ✓ Pushed the Data Assets to the project set up in Lab-1