

# 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.

## Objectives

The goal of the lab is for the users to gain familiarity with the features of the Watson Knowledge Catalog. Before exercising the catalog functionality, the user will need to set up the environment needed to demonstrate the catalog features. This will include downloading the data assets, creating a database service, uploading two of the data assets to the database, and adding a user to the Watson Studio account. Once this is completed, 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 the prerequisites.

## Set Up Environment

### Step 1: Download the [trafficking\\_data.zip](#) from the github repository

1. Click on [trafficking\\_data.zip](#), to download the zipped file.
2. Extract the file contents. You should have three files extracted, (1) female\_human\_trafficking.csv, (2) Categories.csv, and (3) Occupation.csv.

### Step 2: Create a DB2 on Cloud service and Load the file contents into 2 Database tables.

1. Click on [DB2](#) to create the **DB2 on Cloud** service.
2. Enter Trafficking\_DB as the **Service Name** and then click on **Create**.

← View all

**Db2**  
Lite • IBM

A fully-managed cloud SQL database. Powered by a turbo-charged Db2 engine.

[View Docs](#) [Terms](#)

AUTHOR IBM  
PUBLISHED 12/28/2018  
TYPE Service  
LOCATION Sydney, Frankfurt, London, Dallas

**Service name:**  
Trafficking\_DB

**Choose a region/location to deploy in:**  
Dallas

**Choose an organization:**  
DSXUser56000@gmail.com

**Choose a space:**  
dev

**Email:**  
email@example.com

Please provide an email address where we can reach you for updates

**Features**

- Fast, Reliable & Robust**  
Uses Db2 technology for enterprise-level OLTP performance.
- Fully managed, safe, and secure**  
Includes daily backups for 14 days, at-rest database encryption,

Need Help? [Contact IBM Cloud Support](#) [Estimate Monthly Cost](#) [Cost Calculator](#) [Create](#)

3. Click on Trafficking\_DB to open the service.

Cloud Foundry Services

Name	Region	CF Org	CF Space	Plan	Service Offering
Trafficking_DB	Dallas	DSXUser56000@g...	dev	Lite	dashDB For Transa...

Services

Name	Location	Resource Group	Plan	Details	Service Offering
KnowledgeCatalog	Dallas	Default	Lite	Provisioned	Knowledge Catalog
WatsonStudio	Dallas	Default	Lite	Provisioned	Watson Studio
cloud-object-storage-cv	global	Default	Lite	Provisioned	Cloud Object Storage

4. We will need the Service Credentials for the DB2 on Cloud service later in the lab. The next few steps will generate the Service Credentials. You will then cut and paste these credentials into a text editor for later use. Click on **Service Credentials**.

Manage

**Service credentials**

Connections

Dashboard / Trafficking\_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

[Open Console](#)

**Getting Started**

**Where can I find my credentials?**  
Get your username and password by clicking the "Service Credentials" link to the left and selecting "New Credentials".

[Getting Started](#)

**Need Help?**

Use IBM dW Answers to view recently asked questions or ask your own. Still unable to find an answer? Submit a Bluemix Support Ticket to our team.

[IBM dW Answers](#) [Support Ticket](#)

5. Click on **New credential**.


Dashboard /

Trafficking\_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

**Service credentials**

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service. [Learn more](#)

**Service credentials** **New credential** 


Click **New credentials** to create a set of credentials for this instance

6. Click on **Add**.

×

## Add new credential

Service credentials-1

**Add Inline Configuration Parameters (Optional):** 

Provide service-specific configuration parameters in a valid JSON object

**Choose File...**

**Cancel** **Add**

7. Click on **View Credentials**.

Dashboard / Trafficking\_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

**Service credentials**

Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service. [Learn more](#)

**Service credentials** [New credential +](#)

Items per page 10 | 1-1 of 1 items 1 of 1 pages < 1 >

KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	<a href="#">View credentials</a>

8. Cut and paste the credential information into a text editor. We will need the username, password, host, and db fields to set up a connection to the database service.

KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	<a href="#">View credentials</a>

```

{
  "password": "2154100118f8s27",
  "https_url": "https://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "port": 50000,
  "ssldsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "host": "dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "jdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "uri": "db2://xc42672:2154100118f8s27@dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "db": "BLUDB",
  "dsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "username": "xc42672",
  "ssljdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB:sslConnection=true;"
}

```

9. We are now going to load the Categories and Occupation file contents into the database service that we just created. The DB2 on Cloud service console provides a built-in load capability. Click on **Manage**.

**Manage**

Service credentials

Connections

Trafficking\_DB

Location: Dallas Org: DSXUser56000@gmail.com Space: dev

Items per page 10 | 1-1 of 1 items 1 of 1 pages < 1 >

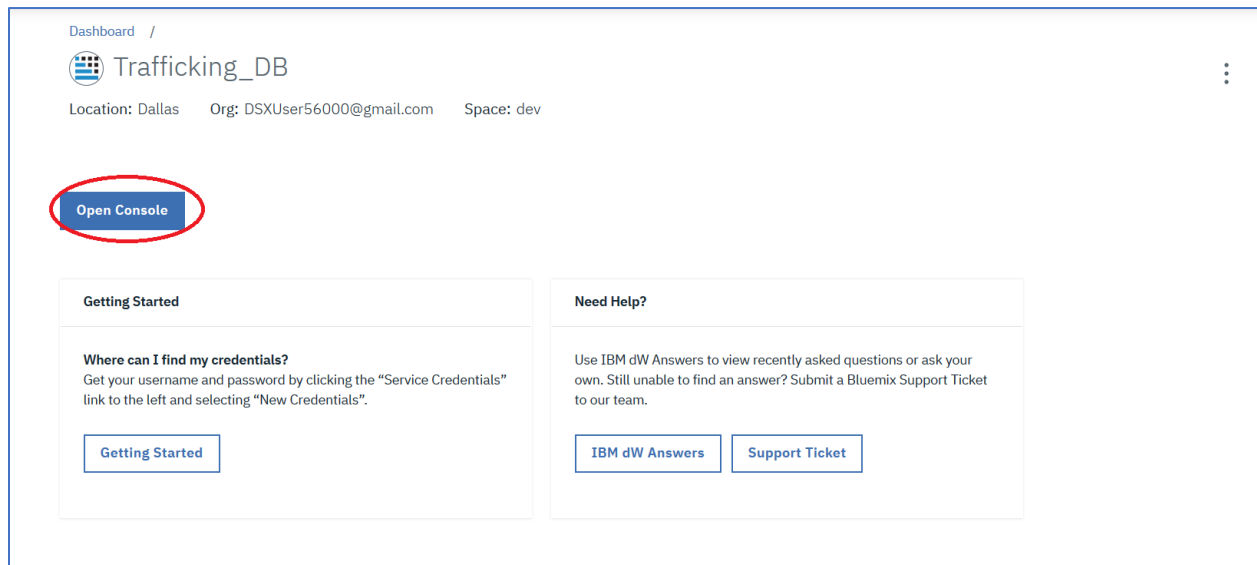
KEY NAME	DATE CREATED	ACTIONS
<input type="checkbox"/> Service credentials-1	JAN 19, 2019 - 03:00:51 PM	<a href="#">View credentials</a>

```

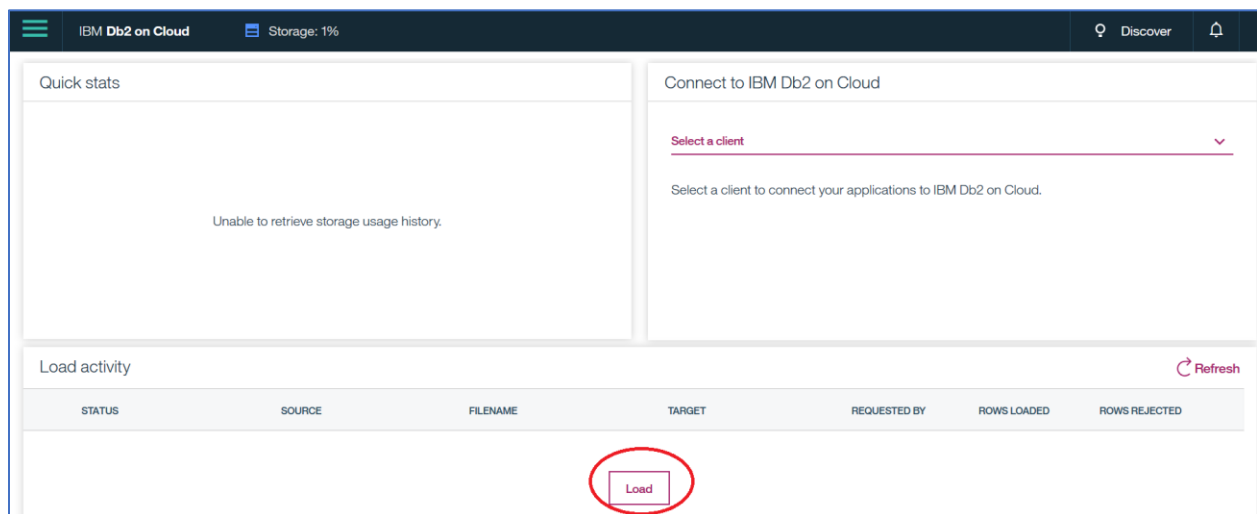
{
  "password": "2154100118f8s27",
  "https_url": "https://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "port": 50000,
  "ssldsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "host": "dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net",
  "jdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "uri": "db2://xc42672:2154100118f8s27@dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB",
  "db": "BLUDB",
  "dsn": "DATABASE=BLUDB;HOSTNAME=dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net;PORT=50000;PROTOCOL=TCP;SSL=TRUE;SECURITY=SSL;",
  "username": "xc42672",
  "ssljdbcurl": "jdbc:db2://dashdb-txn-sbox-yp-dal09-03.services.dal.bluemix.net:50000/BLUDB:sslConnection=true;"
}

```

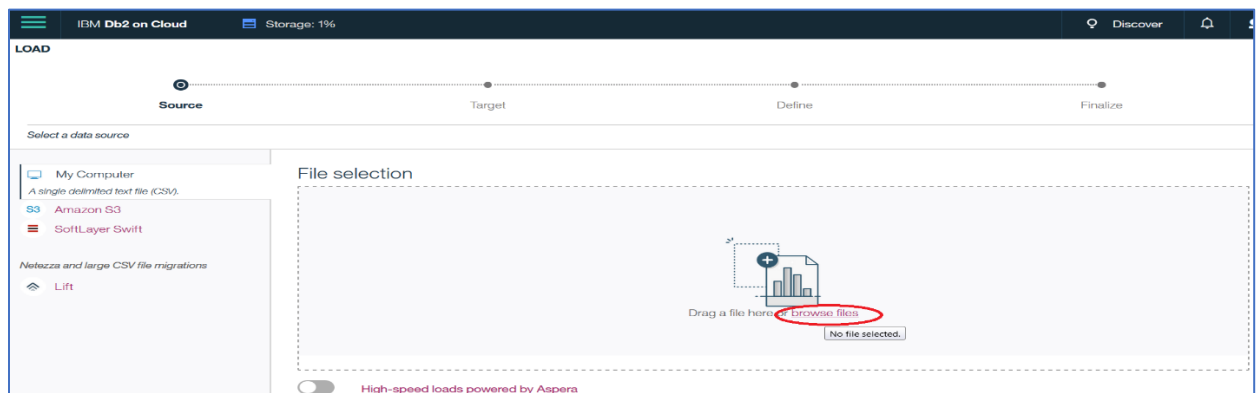
## 10. Click on **Open Console**



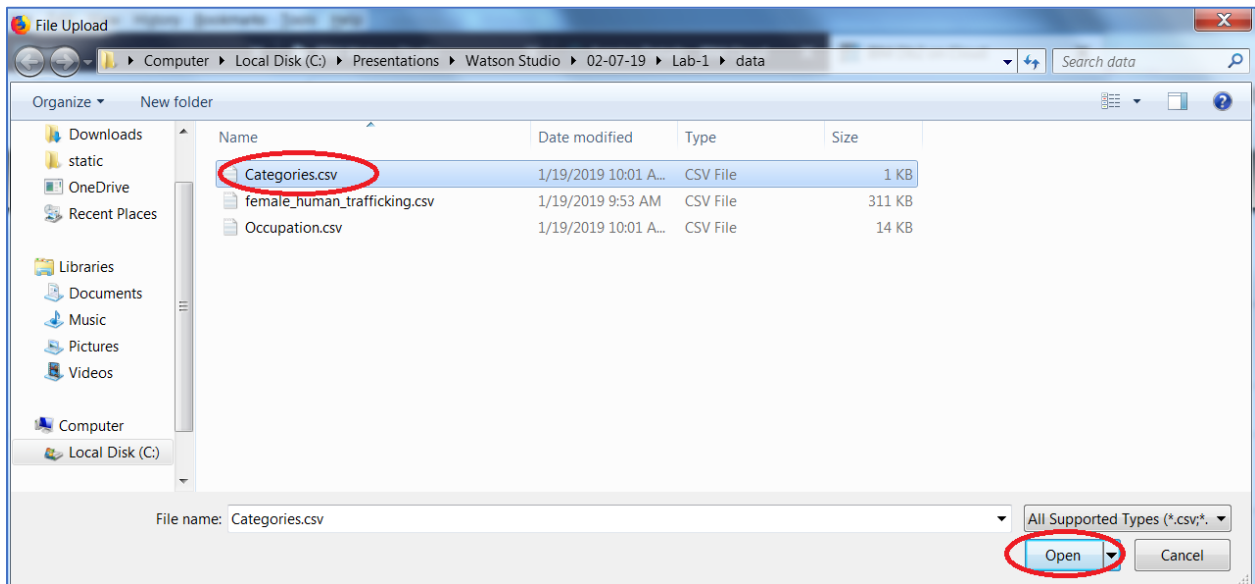
## 11. Click on **Load**.



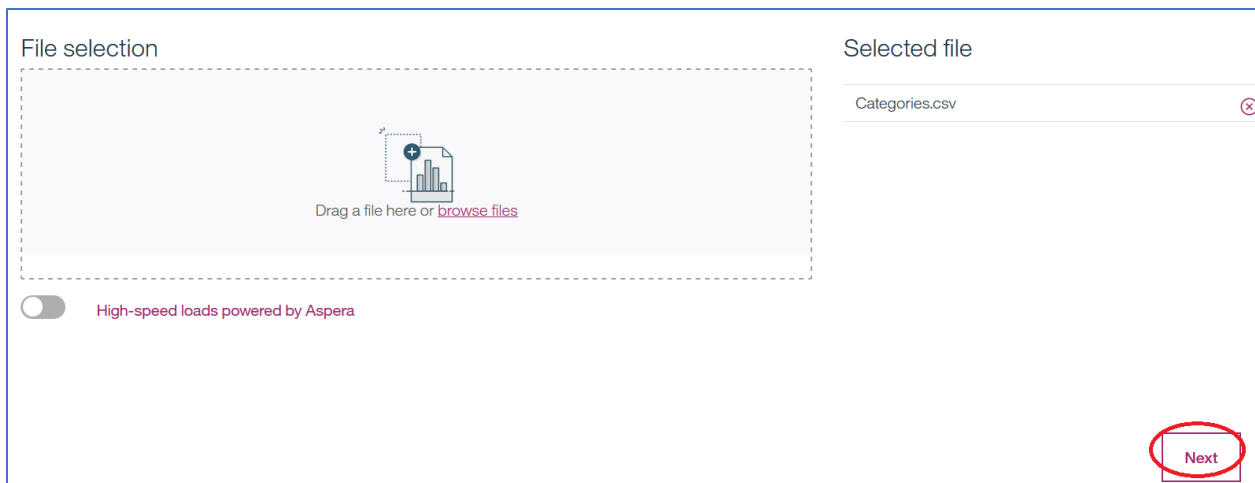
## 12. Click on **browse files**.



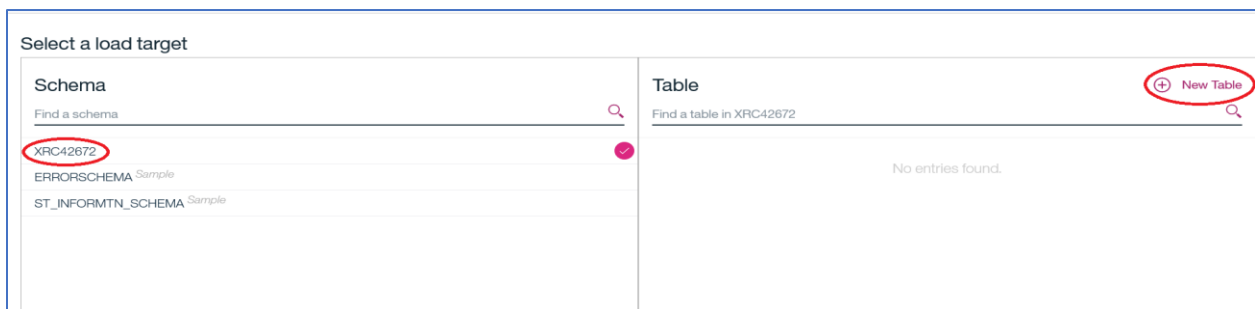
13. Navigate to the folder where you downloaded the trafficking files. Select **Categories.csv**, and then click on **Open**.



14. Click on **Next**.



15. Click on the schema that doesn't have Sample appearing next to it. It should have 3 characters followed by a number. After clicking on the schema, click on **New Table**.



16. Enter CATEGORIES for the name of the database table, and then click **Create** to create the CATEGORIES table.

Select a load target

Schema	Table	Create a new Table
Find a schema	Find a table in XRC42672	CATEGORIES
XRC42672	No entries found.	Create
ERRORSCHEMA <i>Sample</i>		
ST_INFORMTN_SCHEMA <i>Sample</i>		

17. Click **Next**.

Select a load target

Schema	Table	Table definition
Find a schema	Find a table in XRC42672	CATEGORIES
XRC42672	CATEGORIES	No statistics available
ERRORSCHEMA <i>Sample</i>		Specify the table columns in the next step.
ST_INFORMTN_SCHEMA <i>Sample</i>		

Back Next

18. Click **Next**.

Code page (character encoding): 1208 (UTF-8) Separator: , Header in first row: ☒ Time & date format: Time Detect data types: ☒

Code	Category
SMALLINT	VARCHAR(22)
1	Sports/Travel
2	Engineering
3	Information Technology
4	Journalism
5	Government
6	Medical
7	Science
8	Arts
9	Advertising
10	Legal

Back Next

19. Click **Begin Load**.

Source Target Define Finalize

You are loading the file **Categories.csv** into **XRC42672.CATEGORIES**

### Review settings

#### Summary

Code page: 1208 (Default)

Separator: , (Default)

Header in first row: Yes (Default)

Time format: HH:MM:SS (Default)

Date format: YYYY-MM-DD (Default)

Timestamp format: YYYY-MM-DD HH:MM:SS (Default)

String delimiter: "(Default)

#### Option

Maximum number of warnings

1000

Back **Begin Load**

20. The 15 records in the Categories.csv file are successfully loaded. Click on the canvas outside of the Notifications popup.

LOAD

Load details

COMPLETE

My computer

Categories.csv

Target

XRC42672.CATEGORIES

Status

Settings

15

15

0

Rows read

Rows loaded

Rows rejected

Start time

01/19/2019 3:21:51 PM

End time

01/19/2019 3:21:54 PM

The data load job succeeded.

You can now work with your data.

Errors

No errors

Notifications

My computer

Categories.csv

Target

XRC42672.CATEGORIES

View Table

Clear All

21. Click on **Load More Data**.

LOAD

Load details

COMPLETE

My computer

Categories.csv

Target

XRC42672.CATEGORIES

Status

Settings

15

15

0

Rows read

Rows loaded

Rows rejected

Start time

01/19/2019 3:21:51 PM

End time

01/19/2019 3:21:54 PM

The data load job succeeded.

You can now work with your data.

Errors

Warnings

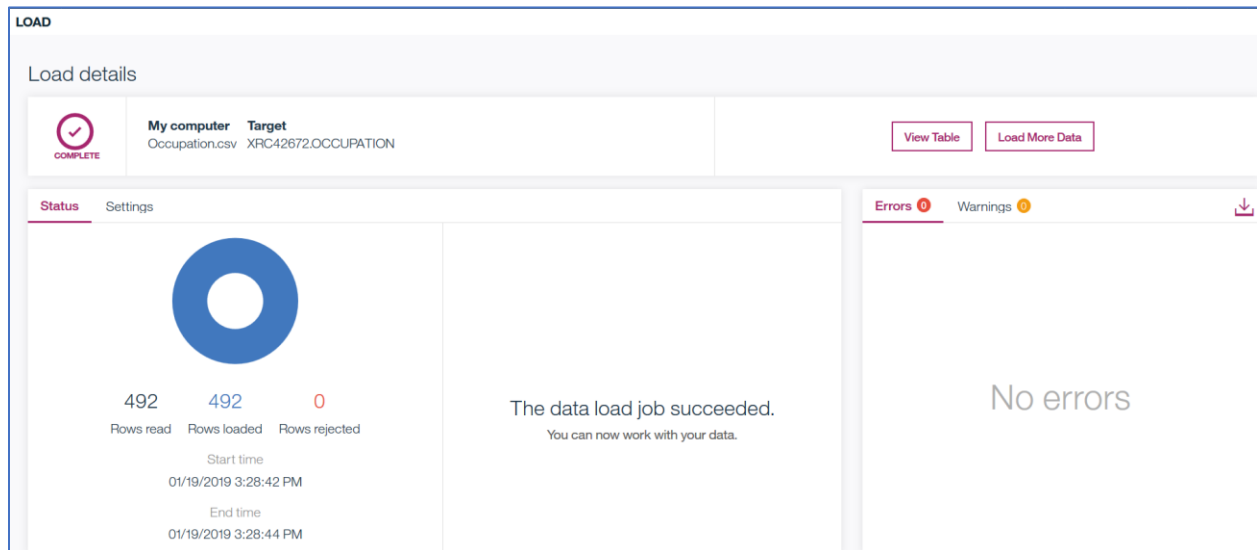
No errors

View Table

**Load More Data**



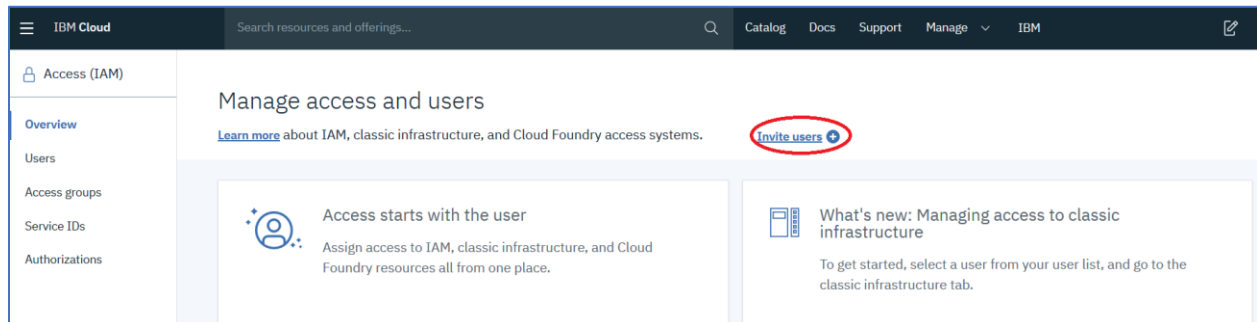
22. Repeat steps 12-20 to create an OCCUPATION database table and to load the contents of the Occupation.csv file into the table. This should result in 492 records from the Occupation.csv file to be successfully loaded into the OCCUPATION database table.



### Step 3: Add a User to the Watson Studio account

To demonstrate the access control features of Watson Knowledge Catalog, add a user to your account.

1. Click on [Manage Users](#), and then click on **Invite users**.



2. In the **Users** section, enter [DSX57000@gmail.com](mailto:DSX57000@gmail.com) for the **Email** address.

The screenshot shows the 'Invite users' form. Under the 'Users' section, it says 'You can invite one or more users. All users invited together are granted the same access.' The 'Email address' field is highlighted with a red circle and contains the text 'DSXUser57000@gmail.com'.

3. In the **Services** section, select **Resources** for the **Assign access to**, and select **All Identity and Access enabled services** for **Services**.

Access

Assign access for all invited users. The default is no access, but at least one type of access must be assigned.

Services ⓘ

An initial IAM policy can be assigned with this invite. You can edit this policy or assign additional access after the user accepts the invite. All fields are required.

Assign access to ⓘ

Resource

Services

All Identity and Access enabled services ⓘ

This option automatically grants access to new services when added.

4. Scroll down and assign **Reader** for the **Assign service access roles**, and then click on **Invite users**.

**Assign platform access roles**

These roles enable users to complete tasks on platform resources, such as creating instances, connecting instances to apps, and assigning user permissions.

**Assign platform access roles**

☐ Administrator  
As an administrator, you can perform all platform actions based on the resource this role is being assigned, including assigning access policies to other users.

☐ Editor  
As an editor, you can perform all platform actions except for managing the account and assigning access policies.

☐ Operator  
As an operator, you can perform platform actions required to configure and operate service instances, such as viewing a service's dashboard.

☐ Viewer  
As a viewer, you can view service instances, but you can't modify them.

**Assign service access roles**

These roles allow a user access for using the service and performing service API calls.

**Assign service access roles**

☐ Manager  
As a manager, you have permissions beyond the writer role to complete privileged actions as defined by the service. In addition, you can create and edit service-specific resources.

☐ Writer  
As a writer, you have permissions beyond the reader role, including creating and editing service-specific resources.

☒ Reader  
As a reader, you can perform read-only actions within a service such as viewing service-specific resources.

Cloud Foundry access ⓘ

Cancel Invite users

5. [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com) is added as a user in the Watson Studio account.

Users

Account: IBM

Use the View by option to change your view of the user list by selecting the user grouping. Depending on your access, you might see users grouped at the account level or by Cloud Foundry org or user hierarchy for classic infrastructure access.

View by: Account users

Filter

Filter by name or email Status

Invite users

User	Email	Status
DSX56000 User owner self	DSXUser56000@gmail.com	Active
DSX57000 User	DSXUser57000@gmail.com	Active

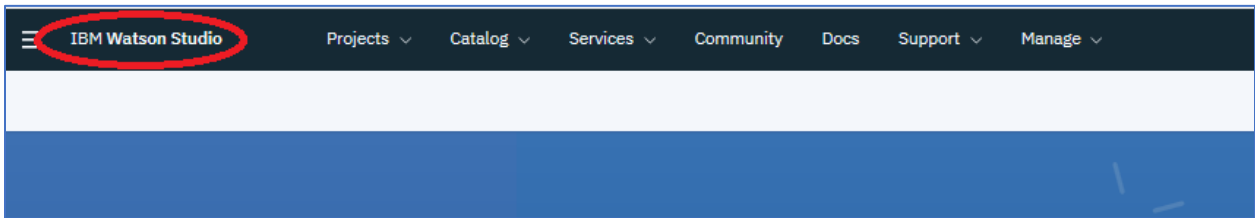
Users per page: 25 | 1-2 of 2 items

1 of 1 pages

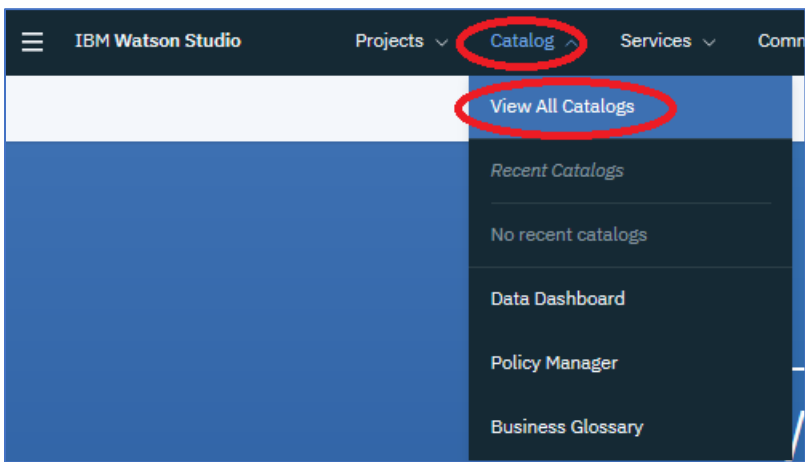
## Perform Catalog Tasks

### Step 1 – Create a Governed Catalog

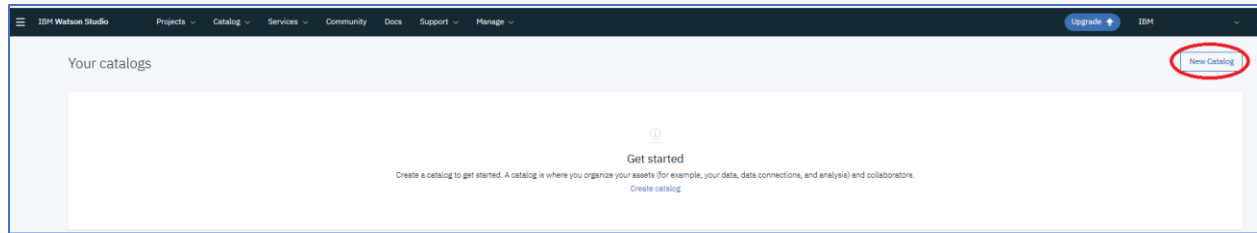
1. Click on the browser tab to return to Watson Studio.
2. Click on Watson Studio to go to the home screen.



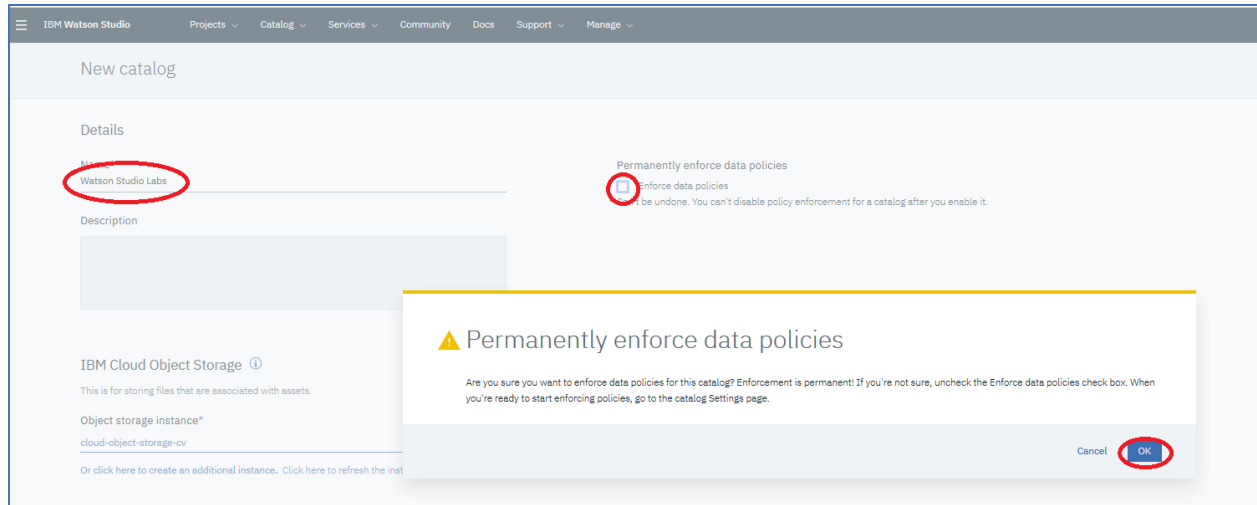
3. Click on **Catalog** and then **View All Catalogs**.



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**.



**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.

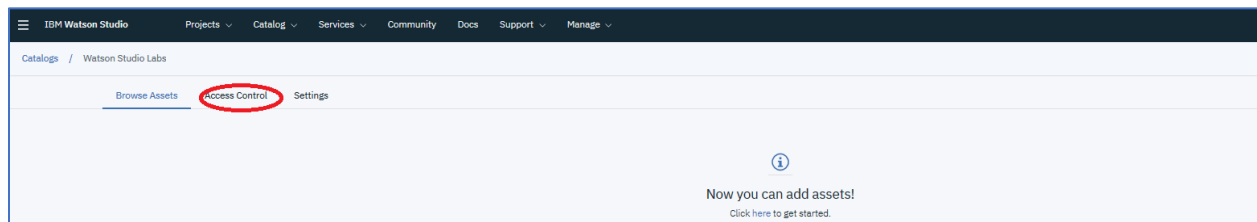
Permanently enforce data policies  
☒ Enforce data policies  
Can't be undone. You can't disable policy enforcement for a catalog after you enable it.

Cancel Create

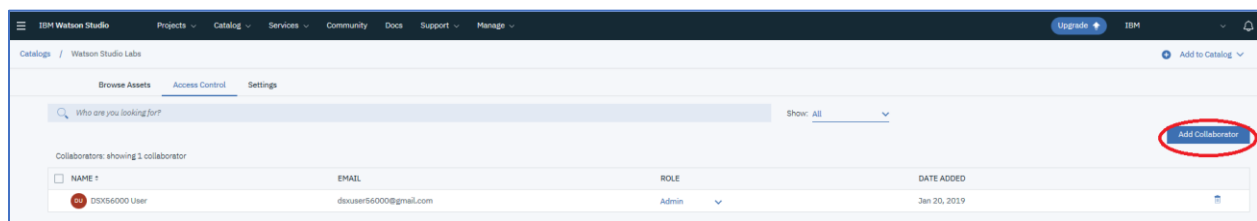
## Step 2: Add a Member to the Catalog

To demonstrate the Catalog access control capability, we will add the [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com) as a member of the catalog.

1. Click on **Access Control**.



2. Click on **Add Collaborator**.



3. Enter DSX57000 as the **Collaborator**. DSX57000 should show up in the list box. Click on DSX57000 User.

## Add Collaborators as Viewers

Collaborators (0)

DSX57000

NAME	EMAIL
DSX57000 User	dsxuser57000@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)

DSX57000 User X

Role: Viewer

Having trouble adding collaborators? ⓘ

Cancel Add

5. DSX57000 is added to the catalog.

IBM Watson Studio

Catalogs / Watson Studio Labs

Who are you looking for?

Show: All

Add Collaborator

Collaborators: showing 2 collaborators

NAME	EMAIL	ROLE	DATE ADDED
DSX56000 User	dsxuser56000@gmail.com	Admin	Jan 20, 2019
DSX57000 User	dsxuser57000@gmail.com	Viewer	Jan 21, 2019

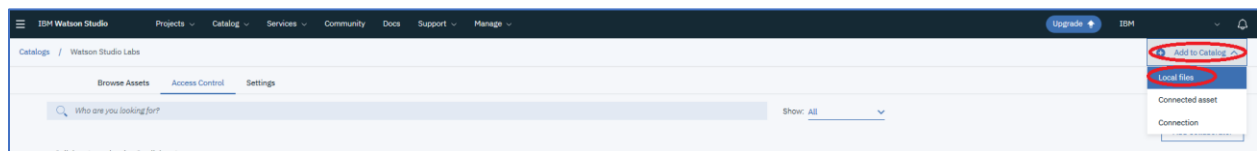
### Step 3: Add Data Assets to the Catalog

In this step, we will add a local file, a connection, and two connected data assets. The local file we will add is the “female\_human\_trafficking.csv” file. We will set up a connection to the DB2

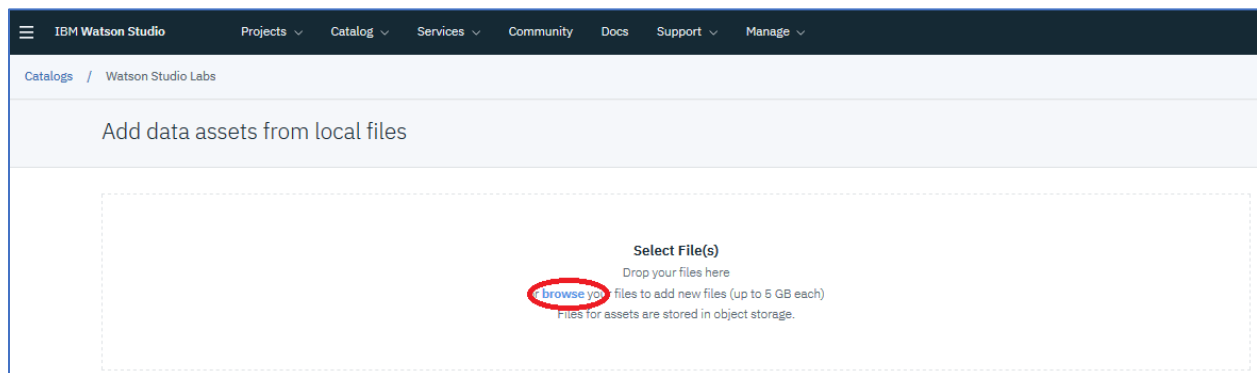
on Cloud database service that we created earlier (this is where we need the credentials). The two connected data assets will point to the CATEGORIES and OCCUPATION tables.

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.

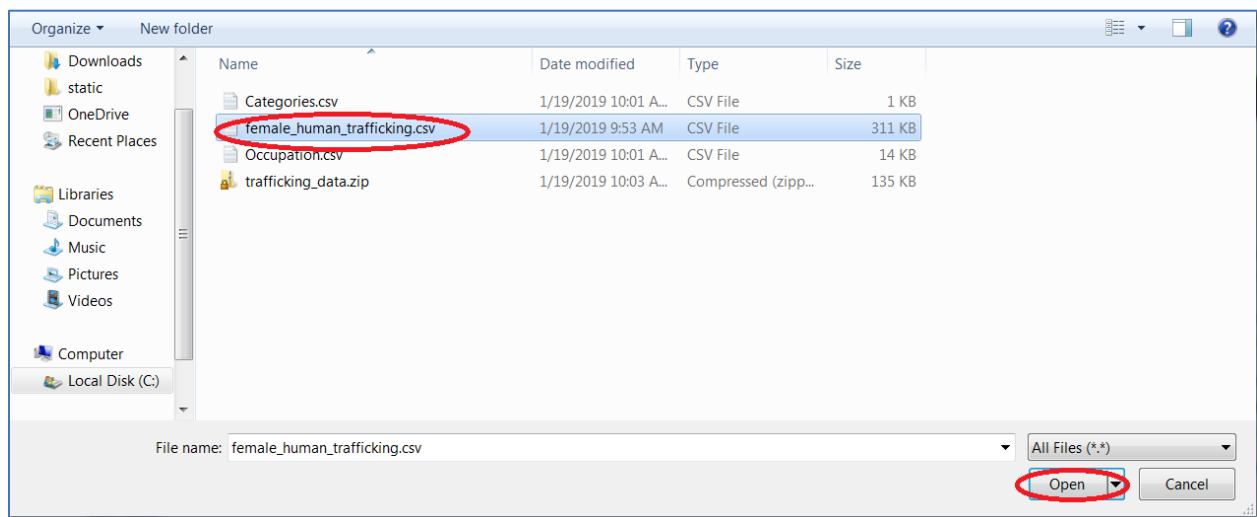
1. Click on **Add to Catalog** and click on **Local Files**.



2. Click on **browse** to navigate to the folder where the “female\_human\_trafficking.csv” file is stored.



3. Click on **female\_human\_trafficking.csv** and then click **Open**.



4. We can add tags to the newly added data asset to more easily search for it. Enter “trafficking” for the **Tags** and click on the “+” icon, and then enter “female human trafficking”, and click on the “+” icon. The **Tags** field should appear as below.

IBM Watson Studio

Catalogs / Watson Studio Labs

Add data assets from local files

Selected Files (1)\*

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

ASSET NAME	FORMAT
female_human_trafficking.csv	Microsoft Excel

Target

Watson Studio Labs - Watson Studio Labs\_DataCatalog

Description

Tags

Start typing to add values

trafficking X female human traffi... X

Classification\*

None

Privacy

☒ Public ☐ Private

All catalog members can find and use the asset.

Members

[Add members](#)

5. Select “Personally Identifiable Information” for **Classification**, 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.

IBM Watson Studio

Catalogs / Watson Studio Labs

Add data assets from local files

Select File(s)

Drop your files here

or browse your files to add new files (up to 5 GB each)

Files for assets are stored in object storage.

Target

Watson Studio Labs - Watson Studio Labs\_DataCatalog

Description

Tags

Start typing to add values

trafficking X female human traffi... X

Classification\*

Personally Identifiable Information

Privacy

☐ Public ☒ Private

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

Members

[Add members](#)

6. Enter DSX in the **Members** field. The list of matching users is displayed in the member list. Select the DSX57000 user.

## Add Members

Members (0)

DSX

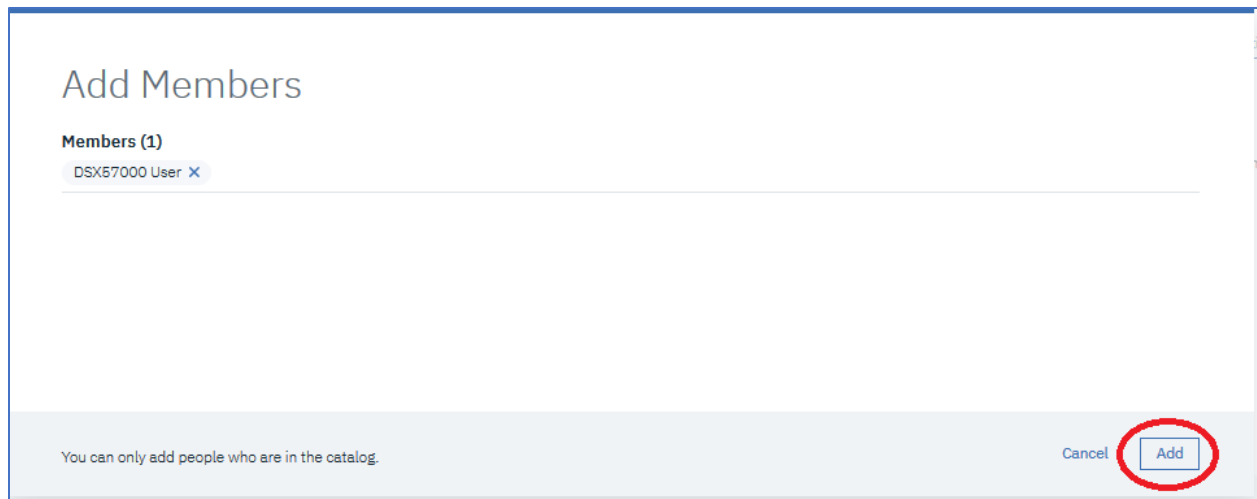
DSX56000 User	dsxuser56000@gmail.com
DSX57000 User	dsxuser57000@gmail.com

You can only add people who are in the catalog.

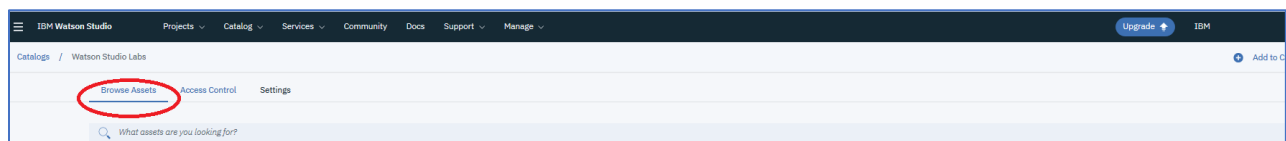
Cancel Add

7. Click on **Add**.

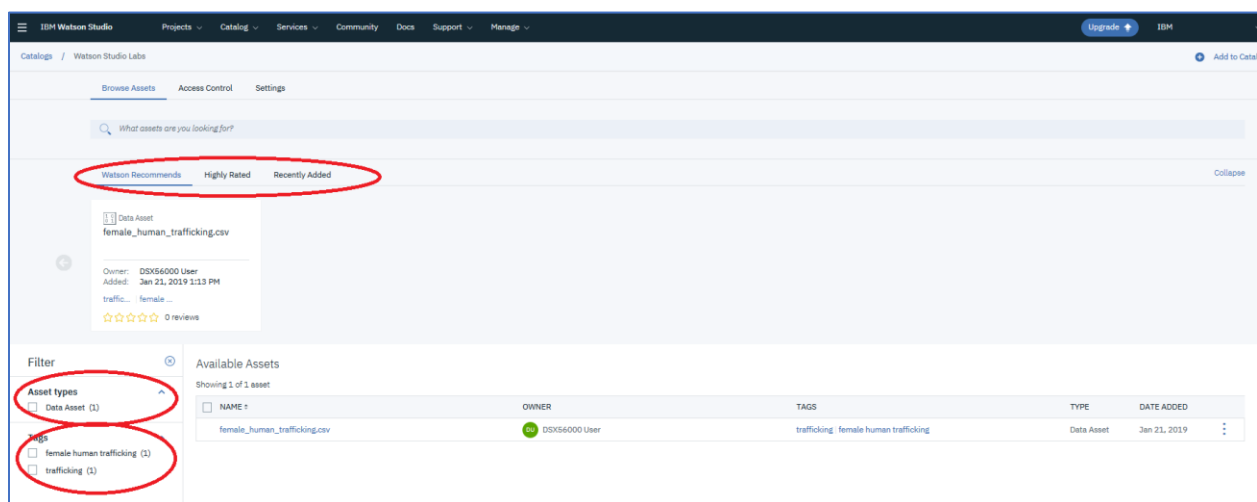




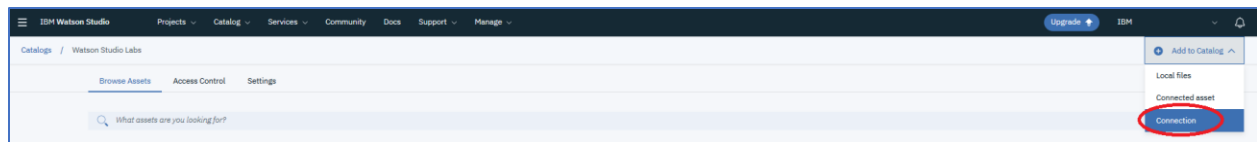
8. Click on **Browse Assets**.



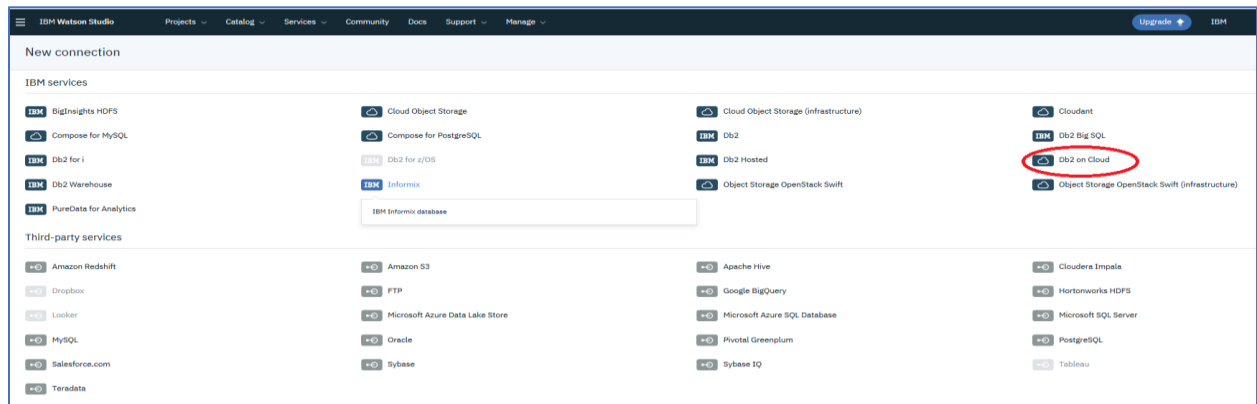
9. The female\_human\_trafficking.csv asset is displayed. Note a number of features on the panel. At the top are 3 tabs: Watson Recommended, Highly Rated, and Recently added. The square below contains assets that are displayed when one of these tabs are selected. On the left side in the Filter area are shown filters by **Asset Type**, and **Tags**. As we add data assets, the tags associated with those assets will be displayed, and an additional asset type will be recorded. The number next to each tag entry indicates how many assets contains the specific tag. The number next to the Asset Type entry indicates how many assets exist for the specific type.



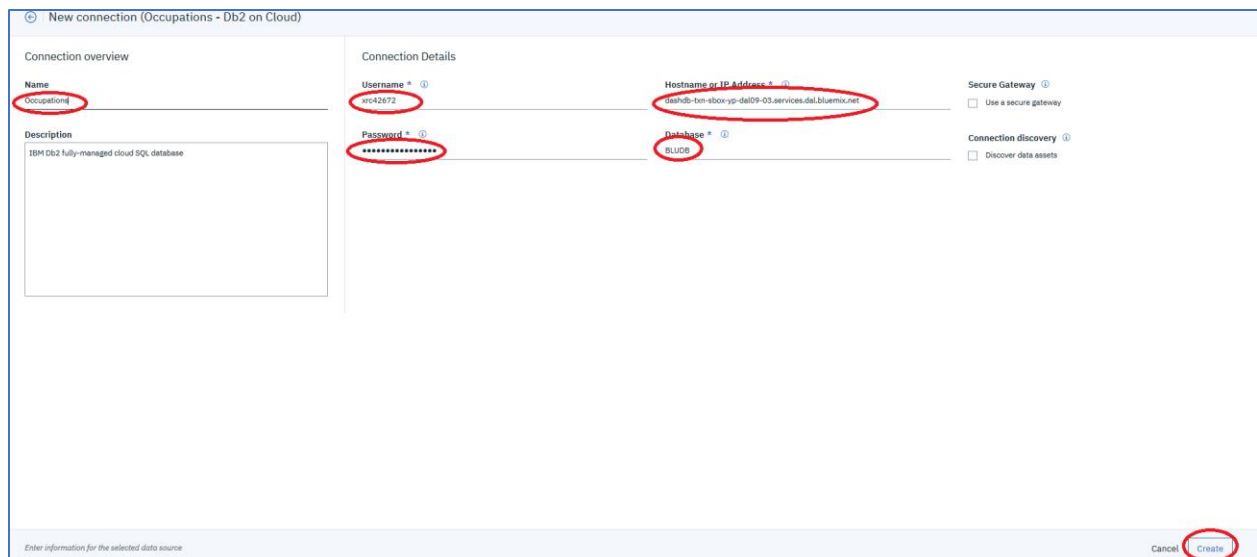
10. Click on **Add to Catalog**, and then click on **Connection**.



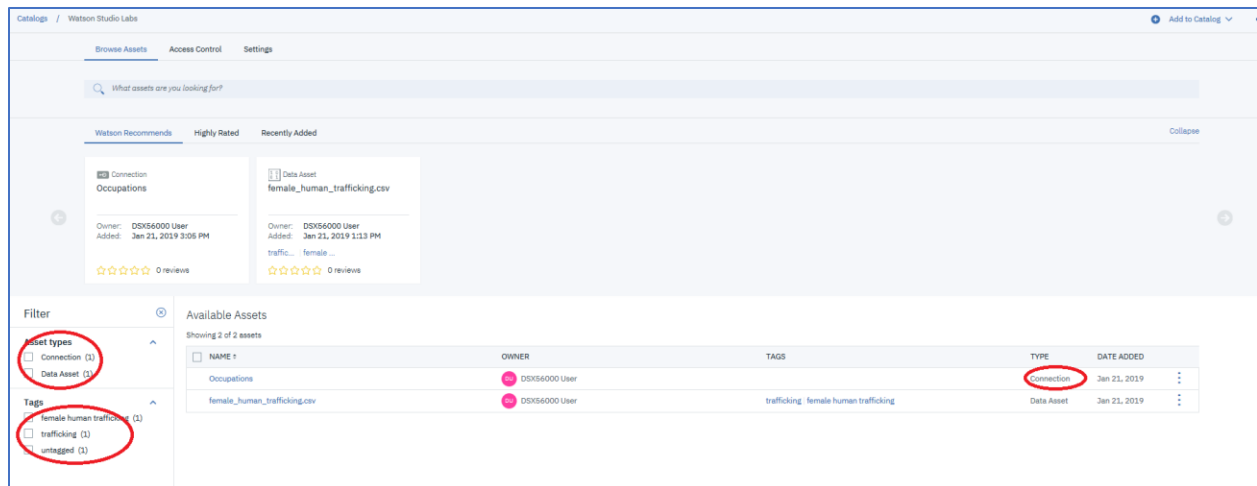
11. Click on **DB2 on Cloud**.



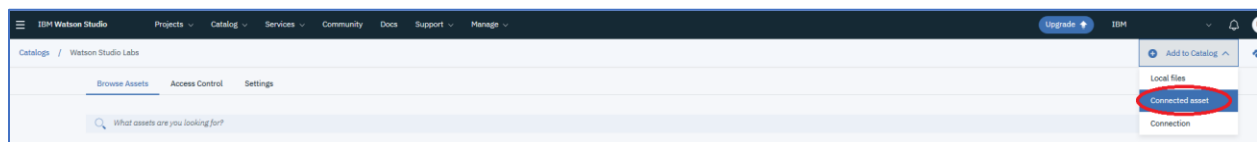
12. Enter a name for the “Occupations” for the **Name** of the connection. Cut and paste the credentials Username, Host, Password, and Database that you saved earlier into the corresponding fields. Click on **Create**.



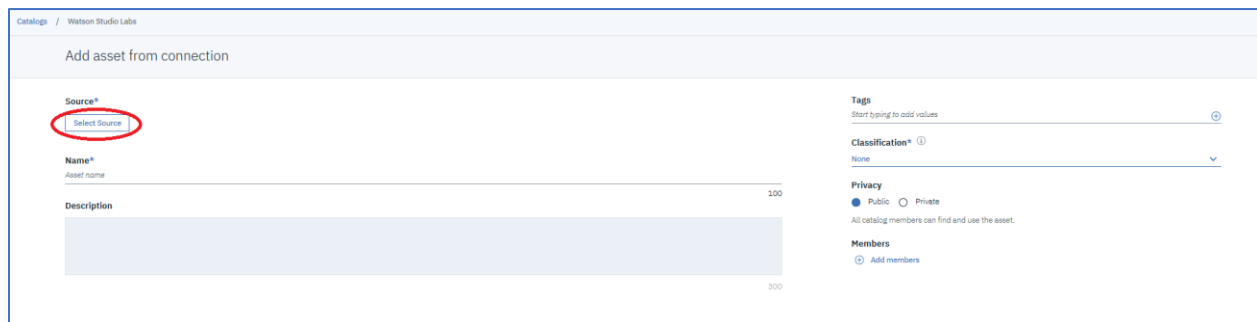
13. The connection asset is displayed. Note that a new Asset Type (Connection) is added in the Filter-Asset Type area, as well as in the Type description for the data asset. Also, note an untagged entry has been created in the Filter-Tags area.



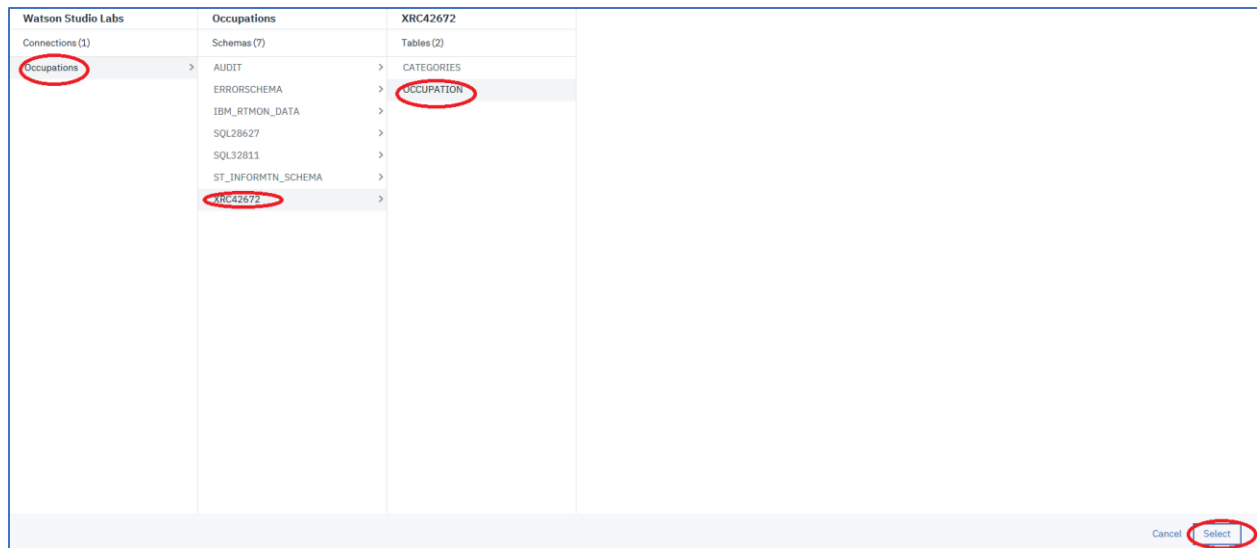
14. We will now add two connected data assets corresponding to the CATEGORIES and OCCUPATION tables. We will use the connection data asset just created to establish a connection to these tables. Click on **Add to Catalog** and click on **Connected asset**.



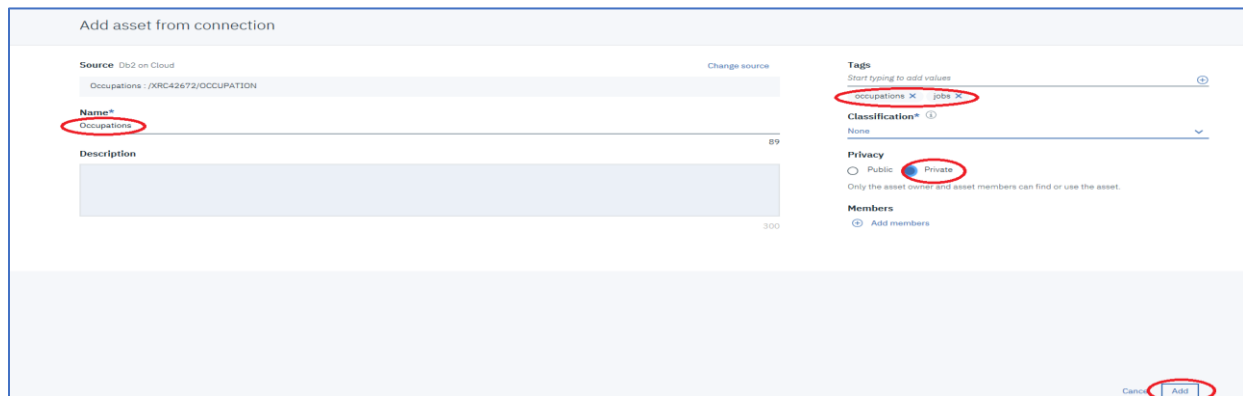
15. Click on **Source**



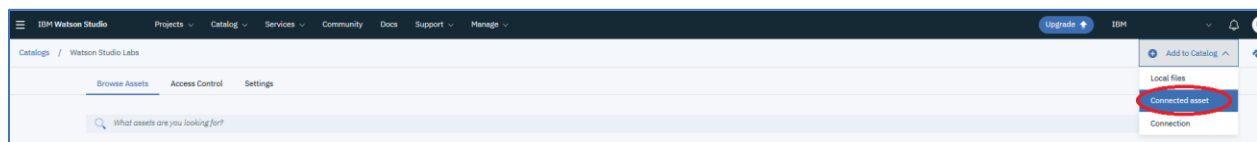
16. Click on the **Occupations** connection, click on the **Schema name**, click on OCCUPATION, and then click **Select**.



- Enter **Occupations** for the **Name** of the Connected asset. Add two tags, “occupations”, click on the + icon, and enter “jobs” and click on the + icon. Click on **Private** for the **Privacy** (note we didn’t add DSXUser57000 to the access control list for this asset). Click on **Add**.



- Click **Add to Catalog**, then click on **Connected asset**.



- Click on **Source**.

Catalogs / Watson Studio Labs

Add asset from connection

**Source\***  
Select Source

**Name\***  
Asset name

**Description**

**Tags**  
Start typing to add values

**Classification\***  
None

**Privacy**  
Public Private

**Members**  
Add members

20. Click on **Occupations**, the **Schema Name**, **CATEGORIES**, and **Select**.

Watson Studio Labs

Connections (1)	Occupations	Schemas (7)	Tables (2)
Occupations	AUDIT	CATEGORIES	OCCUPATION
	ERRORSCHEMA		
	IBM_RTHMON_DATA		
	SQL28627		
	SQL32811		
	ST_INFORMATION_SCHEMA		
	XRC42672		

Cancel Select

21. Enter **Categories** for the name of the **Connected asset**, enter “job types” and then click on the “+” icon, enter “job categories” and then click on the “+” icon, and then click on **Add**.

Add asset from connection

**Source** Db2 on Cloud

Change source

Occupations : /XRC42672/CATEGORIES

**Name\***  
Categories

**Description**

**Tags**  
Start typing to add values  
job types X job categories X

**Classification\***  
None

**Privacy**  
Public Private

**Members**  
Add members

Cancel Add

22. The connected assets are displayed below. Again, the **Asset types**, and **Tags** information in the **Filter** area are updated based on the addition of the two Connected assets.

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Data Asset	Jan 21, 2019
Connections	DSX56000 User		Connection	Jan 21, 2019
female_human_trafficking.csv	DSX56000 User	trafficking   female human trafficking	Data Asset	Jan 21, 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 the asset name 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.

1. 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.

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Data Asset	Jan 21, 2019

3. Clear the **search text field**. Click on the check box next to **Data Asset** in the Filter area. The search results list the 3 Data Assets. The 2 Connected assets, and the 1 uploaded file. Note: Other asset types that can be cataloged include, Connections, Models, Notebooks, and Dashboards.

The screenshot shows the IBM Watson Studio Data Catalog interface. At the top, there's a search bar with the placeholder text 'What assets are you looking for?'. Below the search bar, there are tabs for 'Watson Recommends', 'Highly Rated', and 'Recently Added'. The main content area displays four asset cards: 'Data Asset Categories', 'Data Asset Occupations', 'Connection Occupations', and 'Data Asset female\_human\_trafficking.csv'. Each card shows the owner 'DSX56000 User', the date added, and a star rating. On the left, there's a 'Filter' sidebar. Under 'Asset types', 'Data Asset (3)' is selected and circled in red. Under 'Tags', several tags are listed, including 'categories (1)', 'female human trafficking (1)', 'job categories (1)', 'job types (1)', 'jobs (1)', 'occupations (1)', and 'trafficking (1)'. The 'Available Assets' section shows a table with 3 assets:

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Data Asset	Jan 21, 2019
female_human_trafficking.csv	DSX56000 User	trafficking   female human trafficking	Data Asset	Jan 21, 2019

4. Click on the Tag **job types** in the Filter area. The Categories Data Asset is listed. Note that the Filter area has been updated to only list Asset Types, and Tags that correspond to the search result.

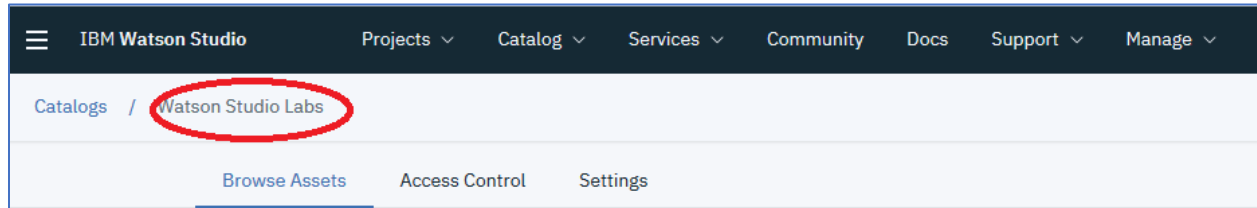
The screenshot shows the IBM Watson Studio Data Catalog interface with the search filter updated to 'job types'. The 'Filter' sidebar on the left shows 'Asset types' with 'Data Asset (1)' selected. Under 'Tags', 'job types' is selected and circled in red. The 'Available Assets' section now shows only one asset:

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019

## Step 5: Edit/Review/Profile Data Asset

In this step, we will drill down on the `female_human_trafficking.csv` 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.

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



2. Click on the `female_human_trafficking.csv` asset.

Available Assets

Showing 4 of 4 assets

<input type="checkbox"/> NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Data Asset	Jan 21, 2019
Occupations	DSX56000 User		Connection	Jan 21, 2019
<input checked="" type="checkbox"/> female_human_trafficking.csv	DSX56000 User	trafficking   female human trafficking	Data Asset	Jan 21, 2019

female\_human\_trafficking.csv

3. The metadata and data contents of the asset is 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 a number of 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.csv

Overview Access Review Profile Lineage

female\_human\_trafficking.csv

Remove Download Add to Project

**Description**  
Travel data to be examined for trafficking  
Added: Jan 21, 2019 1:13 PM  
Format: Microsoft Excel

**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.

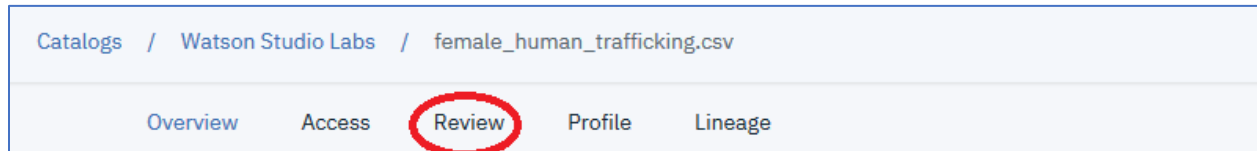
Schema: 26 Columns  
Preview: 1000 rows Last refresh: 2 hours ago Refresh

INTERNAL_ID	VETTING_LEVEL	DESCRIPTION	NAME	GENDER	BIRTH_DATE	BIRTH_COUNT...	BIRTH_COUNTRY_CODE	OCCUPATION	ADDRESS	SSN	PASSPORT
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
1	100	NA	Kathleen Bailey	F	1985-11-27	Ghana	GH	Chiropractist	79421 Jordan Cr	886-71-9055	24047950
2	30	NA	Kelly Robbins	F	1971-01-17	Pakistan	PK	Engineer, structu	5808 Jacobs Uni	141-10-9199	44906403
3	100	NA	Linda Stewart	F	1997-04-15	Ghana	GH	Engineer, land	824 Kristin Giv, f	011-46-5304	63921781
4	30	NA	Stacey Courtney	F	1984-03-27	Ghana	GH	Careers adviser	322 Hutchinson	691-25-2647	21283844
5	30	NA	Erika Patie Fowle	F	1998-06-18	Ghana	GH	Geneticist, molec	07695 Michael V	165-33-0802	16514173
6	100	NA	Ashlee Fisher	F	1981-01-05	Ghana	GH	Electronics engin	4838 Cassandra	072-16-8742	51342811
7	100	NA	Jacqueline Clark	F	1976-01-19	Ghana	GH	Editor, commissi	1148 Wang Fall	634-03-1462	74270955
8	30	NA	Sheil Chapman	F	1997-08-24	Ghana	GH	Writer	72346 Wilson Fo	004-07-1446	17000264
9	10	NA	Sabrina Debra W	F	1977-02-21	Haiti	HT	Recruitment con	251 John SH, Fir	233-53-1156	93421404
10	100	NA	Ashley Gibson	F	1986-04-06	Ghana	GH	Theatre manager	17894 Gary Grev	565-33-0366	3699030
11	100	NA	Tammy Sara Gan	F	1995-08-17	Ghana	GH	Statistician	2902 Raemusser	069-16-0401	13644854
12	10	NA	Sandry Santiago	F	1971-12-16	Ghana	GH	Merchant navy of	9486 Pratt Wall,	778-07-9880	62555365
13	100	NA	Becky Cynthia St	F	1993-11-05	Ghana	GH	Advertising acco	6362 William Cll	612-01-4772	41124738

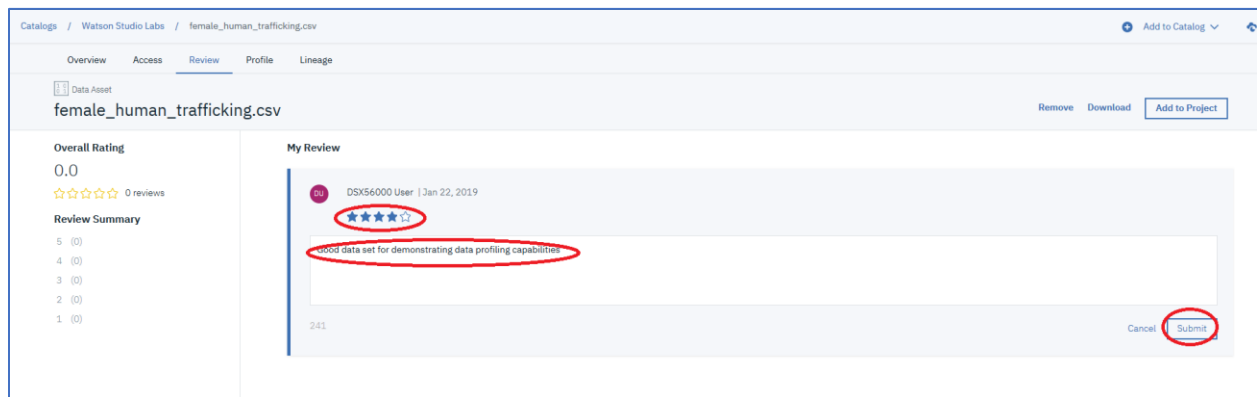


The actions circled in blue will **Remove** the asset from the catalog, **Download** the asset to the local file system, or add the asset to a catalog enable project (**Add to Project**). Hovering over the name of the data asset, the **Description**, the **Tags**, or the **Classification** will popup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

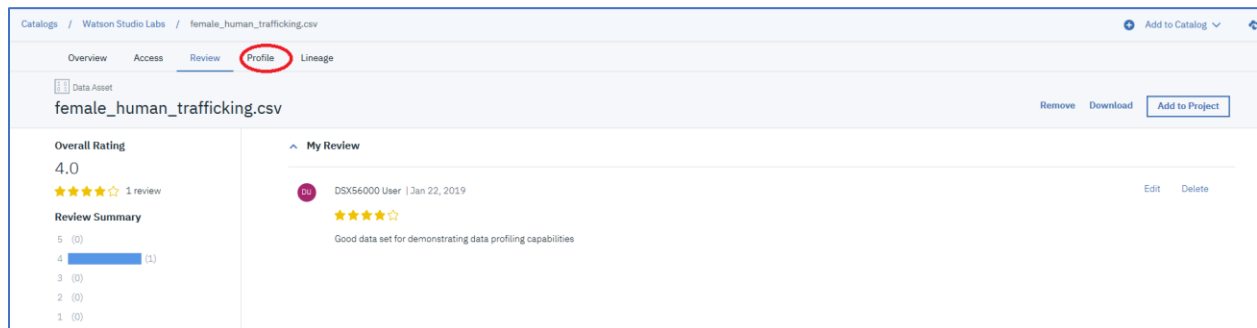
4. Click on the **Review** tab.



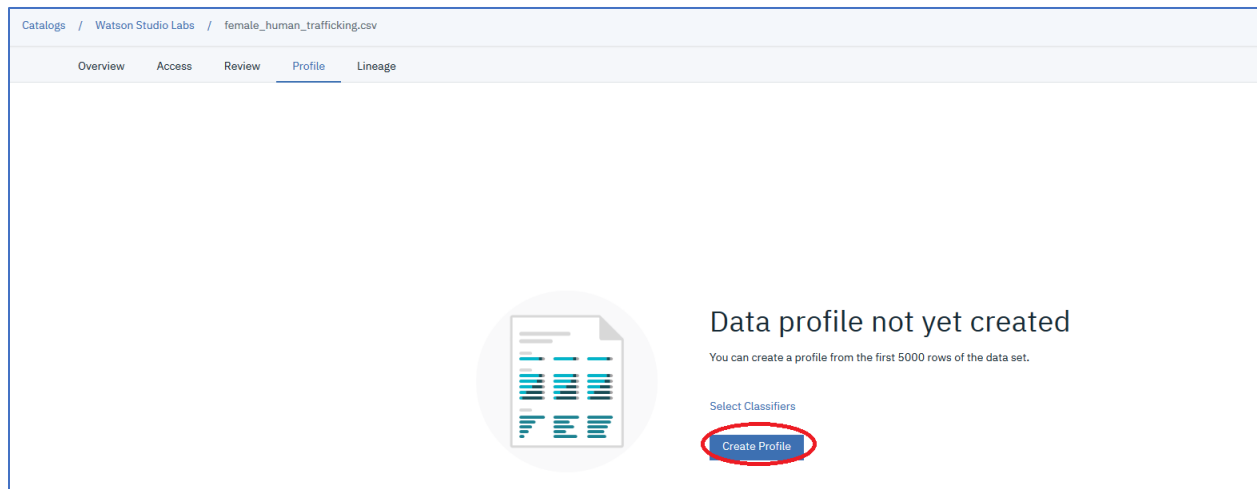
5. Click on the fourth star to rate this asset. Enter “Good data set for demonstrating data profiling capabilities” in the **My Review** text field. Click on the **Submit** button.



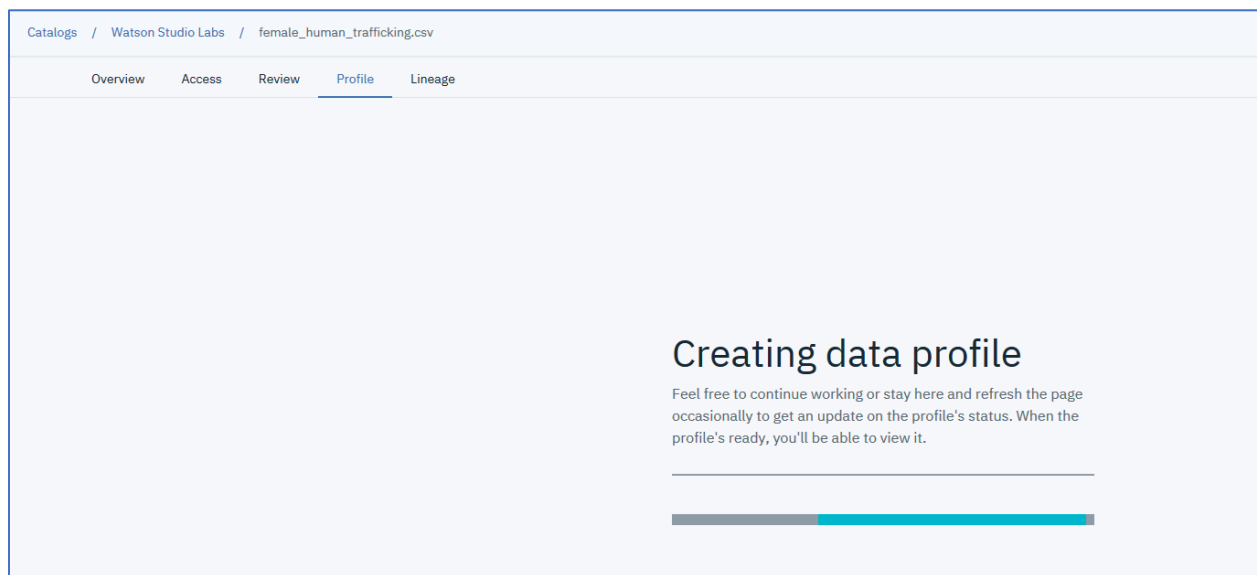
6. The review of the asset is displayed. Click on the Profile tab.



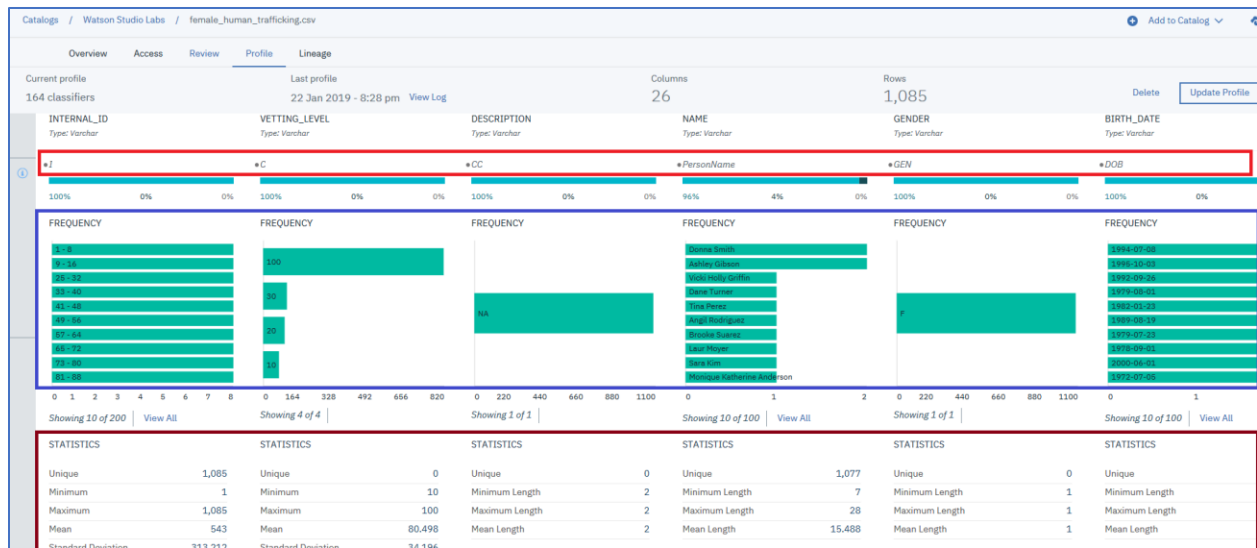
7. Click on **Create Profile**. If a profile was already created, it would be displayed instead of this prompt screen.



8. Wait a minute or so, and then click on browser refresh.

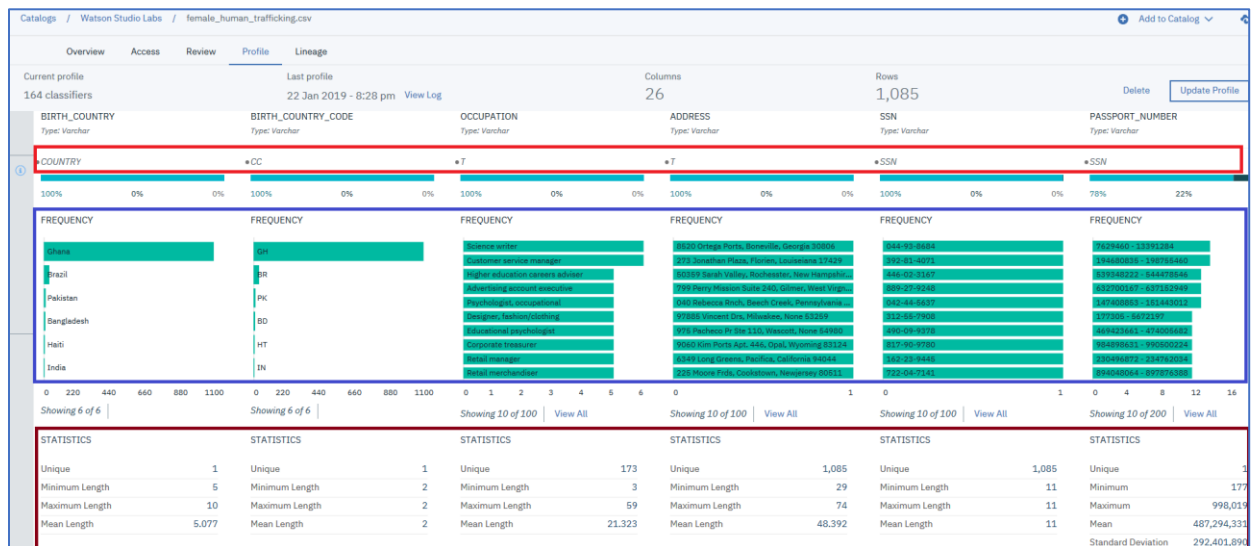


9. The profile information is displayed below in several screenshots based on horizontal scrolling to display many of the columns of the data asset. The highlighted red box contains the column classifications. The highlighted blue box contains counts of the number of occurrences of each value (for discrete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



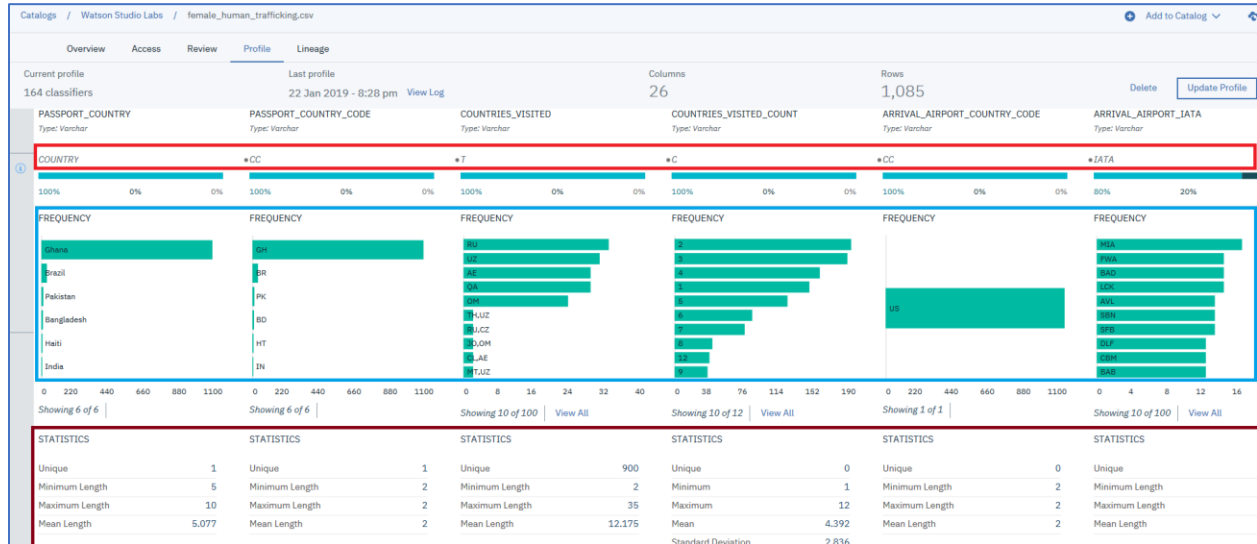
The classifications of the columns are listed below:

- INTERNAL\_ID – I (Identifier)
- VETTING\_LEVEL – C (Code)
- DESCRIPTION – CC (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.
- BIRTH\_DATE: DOB – the classification accurately determines that this field represents a date of birth.



- BIRTH\_COUNTRY: COUNTRY- correctly classified as a country
- BIRTH\_COUNTRY\_CODE – CC (Country Code ) - correctly classified
- OCCUPATION – T (Text field) – there is not a classifier for occupations so it determines that this is a text string.

- ADDRESS: T (Text Field)
- SSN:SSN – correctly classifies the column as a social security number.
- PASSPORT\_NUMBER – incorrectly classifies as a social security number. This can be corrected, although it may be useful to treat as a SSN for privacy reasons.



- PASSPORT\_COUNTRY: COUNTRY- correctly classified as a country
- PASSPORT\_COUNTRY\_CODE: CC(Country Code ) - correctly classified
- COUNTRIES\_VISITED – T (Text field)
- COUNTRY\_VISITED\_COUNT – C (Code) – this is an error. It's not a code. This can be re-classified if necessary.
- ARRIVAL\_AIRPORT\_COUNTRY\_CODE: COUNTRY –
- ARRIVAL\_AIRPORT\_IATA: IATA – the classification accurately determines that this field represents an IATA.

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

## 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:

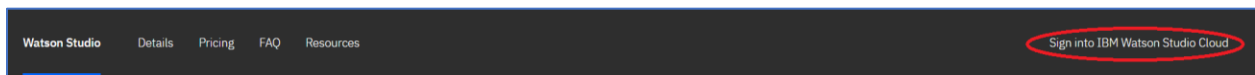
1. Catalog owner (logged on user)
2. [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com)

Data Asset Privacy:

1. female\_human\_trafficking.csv (Private) ACL Members:
  - a. asset owner (logged on user)
  - b. [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com)
2. Occupations (Connected Asset) (Private) ACL Members:
  - a. asset owner (logged on user)
3. Categories (Connected Asset) (Public)
4. Occupations(Connection) (Public)

We would expect that if we log in as [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com), that only 3 Data Assets would appear. The two public Data Assets, and the Private Data Asset where DSXUser57000@gmail is on the access control list. This user would not have access to the Occupations (Connected 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.

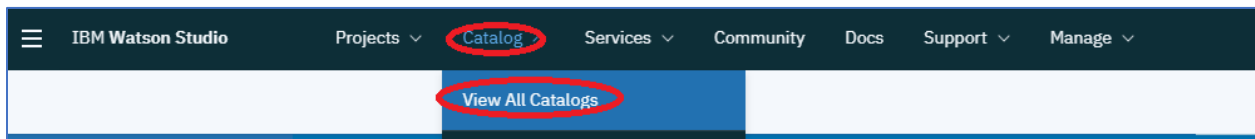
1. Open a different browser (if you are using FireFox, then open Chrome or vice versa).
2. Type in [datascience.ibm.com](https://datascience.ibm.com) for the URL to navigate to.
3. Click on **Sign into IBM Watson Studio Cloud**



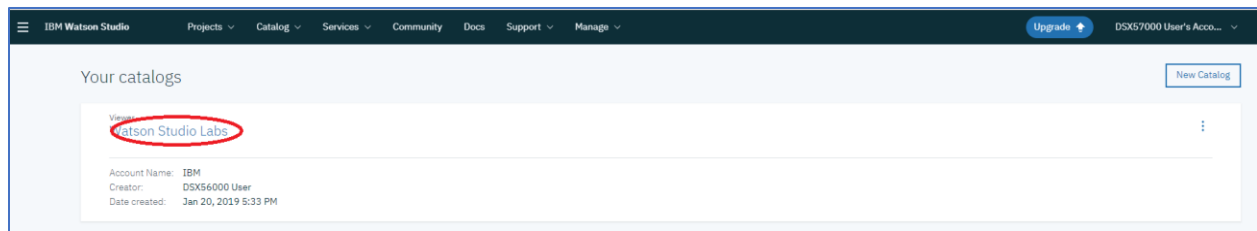
4. Login using [DSXUser57000@gmail.com](mailto:DSXUser57000@gmail.com) for the user name and DSXPasswOrd57000! for the password.

The image shows the IBM login page. On the left, there are links for 'Don't have an IBMId? Create an account', 'Log in with your company credentials (SSO)', and 'Need help? Contact the IBMId Helpdesk'. On the right, there is a login form with fields for 'IBMId' (containing 'DSXUser57000@gmail.com') and 'Password' (containing 'DSXPasswOrd57000!'). There are also links for 'Forgot IBMId?' and 'Forgot password?'. At the bottom right, there is a 'Remember me' checkbox and a 'Log in' button, which is circled in red.

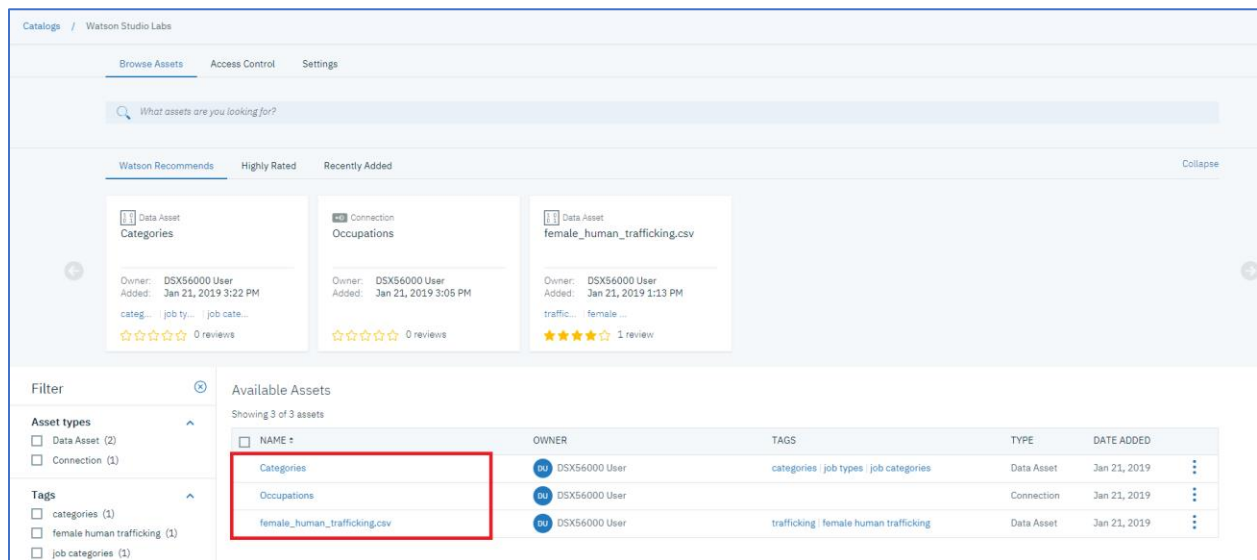
5. Click on **Catalog** and then **View All Catalogs**.



6. Click on **Watson Studio Labs**.



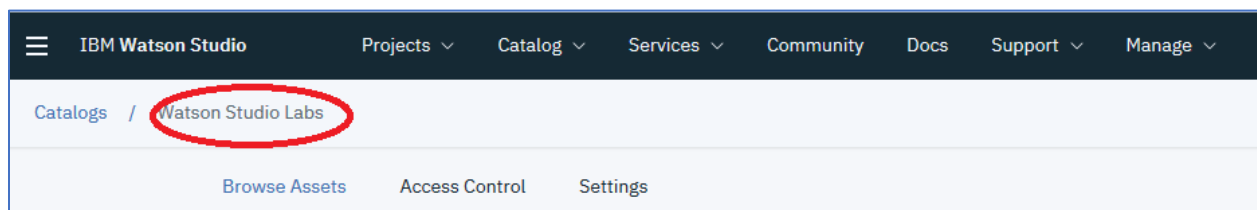
7. The Watson Studio Labs catalog is displayed. Three Data Assets are listed as expected.



## 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.

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



2. Hover the cursor to the left of each data asset and then click on the checkbox. Click on **Add to Project**.

Catalogs / Watson Studio Labs

Data Asset  
Occupations

Owner: DSX56000 User  
Added: Jan 23, 2019 11:11 AM

occup... | jobs

☆☆☆☆☆ 0 reviews

Data Asset  
Categories

Owner: DSX56000 User  
Added: Jan 21, 2019 3:22 PM

categ... | job ty... | job cate...

☆☆☆☆☆ 0 reviews

Connection  
Occupations

Owner: DSX56000 User  
Added: Jan 21, 2019 3:05 PM

☆☆☆☆☆ 0 reviews

Data Asset  
female\_human\_trafficking.csv

Owner: DSX56000 User  
Added: Jan 21, 2019 1:13 PM

traffic... | female ...

☆☆☆☆☆ 1 review

Filter

Asset types

- ☐ Data Asset (3)
- ☐ Connection (1)

Tags

- ☐ categories (1)
- ☐ female human trafficking (1)
- ☐ job categories (1)

Available Assets

Selected 4 of 4 assets

Add to Project Remove

NAME	OWNER	TAGS	TYPE	DATE ADDED
Categories	DSX56000 User	categories   job types   job categories	Data Asset	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Connection	Jan 21, 2019
Occupations	DSX56000 User	occupations   jobs	Data Asset	Jan 23, 2019
female_human_trafficking.csv	DSX56000 User	trafficking   female human trafficking	Data Asset	Jan 21, 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.csv	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. Click on **Projects** and then **Watson Studio Labs**.

IBM Watson Studio

Projects Catalog Services Community Docs Support Manage

Catalogs / Watson Studio Labs

View All Projects

Recent Projects

Watson Studio Labs

Data Asset  
female\_hun...

Data Asset  
Categories

Connection  
Occupations

5. Click on the **Assets** tab.

IBM Watson Studio

Projects Catalog Services Community Docs Support Manage Upgrade

My Projects / Watson Studio Labs

Launch IDE Add to project

Overview Assets Environments Bookmarks Deployments Access Control Settings

6. The Data assets are now incorporated in the project.

FileEditViewHistoryBookmarksToolsHelp

New TabIBM Watson StudioCloud Object Storage - IBM CIBM Db2 on CloudYahoo

https://dataplatfom.cloud.ibm.com/projects/014024b5-6551-42d4-90f0-59edc2cc556d/assets/80%

IBM Watson StudioProjectsCatalogServicesCommunityDocsSupportManageUpgradeIBM





My Projects / Watson Studio LabsLaunch IDEAdd to project

OverviewAssetsEnvironmentsBookmarksDeploymentsAccess ControlSettings

What assets are you looking for?

Data assets

0 asset selected.

<input type="checkbox"/>	NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
<input type="checkbox"/>	 female_human_trafficking.csv	Data Asset	DSX56000 User	23 Jan 2019, 5:04:39 pm	<div></div>
<input type="checkbox"/>	 Occupations	Data Asset	DSX56000 User	23 Jan 2019, 5:04:38 pm	<div></div>
<input type="checkbox"/>	 Categories	Data Asset	DSX56000 User	23 Jan 2019, 5:04:38 pm	<div></div>
<input type="checkbox"/>	 Occupations	Connection	DSX56000 User	23 Jan 2019, 5:04:37 pm	<div></div>

Drop files here or browse for files to upload.

15