IBM Cloud

A Hands-on Introduction to IBM Cloud Private for Data

Lab Guide

© Copyright IBM Corp. 2018. All rights reserved

Notices and Disclaimers

© Copyright IBM Corporation 2018.

The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. This information is based on current IBM product plans and strategy, which are subject to change by IBM without notice. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

IBM, the IBM logo and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

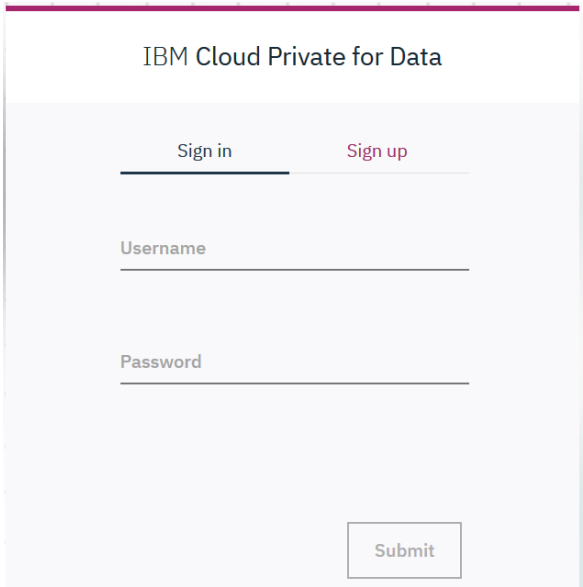
Other company, product and service names may be trademarks or service marks of others

Table of Contents

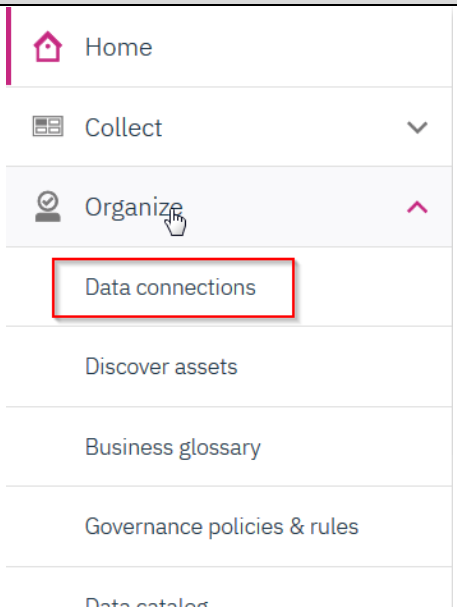
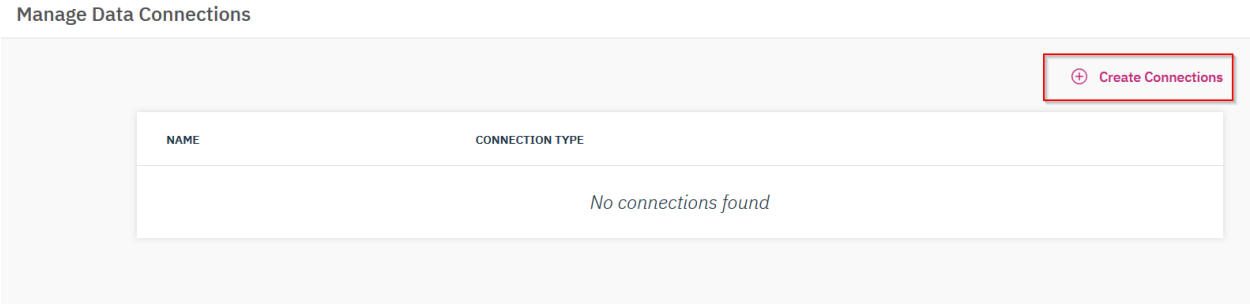
Section 1: Connect to and Acquire Data in ICPD 4
Section 2: Discover and Browse Data 8
Section 3: Transform Data 12
Section 4: Building and Deploying a Machine Learning Model 22

© Copyright IBM Corp. 2018. All rights reserved

Section 1: Connect to and Acquire Data in ICPD

Step	Action
1	**<u>Login to ICPD</u>** a. On a browser (latest version of Firefox recommended), navigate to ICP4Data from the following URL to login to it using the following credentials: URL: https://12.44.195.73:31843/auth/login/zen-login.html Username: *Provided by your instructor* Password: clusterb 
2	**<u>Setup your Connection</u>** In this step, you will setup a connection to a Db2 database that contains the banking data you will be working with today. a. On the left navigation bar click on **Organize -> Data Connections**

© Copyright IBM Corp. 2018. All rights reserved

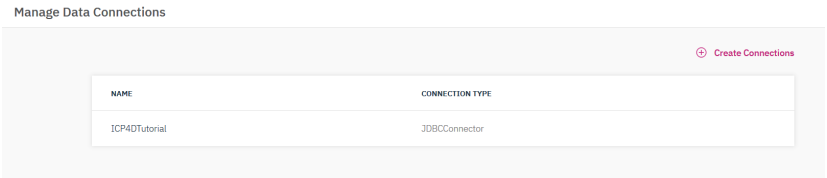
Step	Action
	 b. Click on "**Create Connections**"  c. On the connections form, enter the following details:

Database Type	DB2
Connection Type	Db2
Name	ICP4DLab
Database	BANK
Username	db2inst1
JDBC Host	***Provided by your Instructor***
Password	Password
JDBC Port	50000

© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	Create connection **Choose connection** Db2 **Name** * ICP4DTutorial **Description** A tutorial to showcase self service data preparation and Data Science Machine Learning -37 **Database/Location** * BANK **Username** * db2inst1 **JDBC Host** * <IP Address> **Password** * •••••••• **JDBC Port** * 50000 d. Once the details are entered, one can test the connection by clicking "Test Connection" in the bottom section of the page. Cancel Test Connection Save Connection The system reports that the connection was tested successfully. **Information -** Connection tested successfully e. Save the connection by clicking on "**Save Connection".** Cancel Test Connection Save Connection

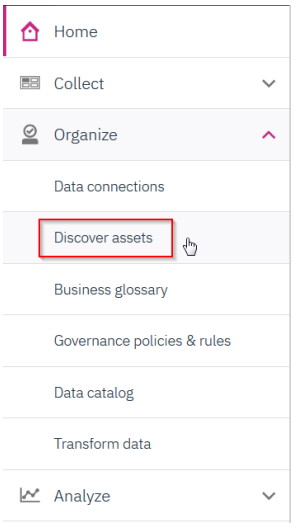
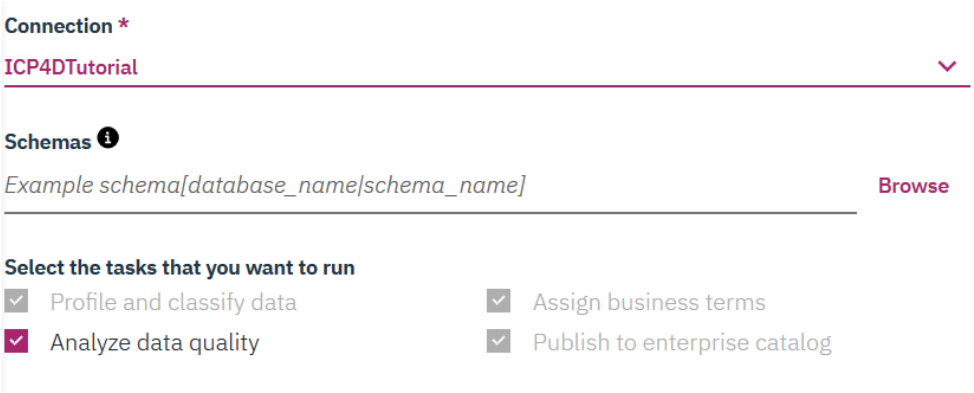
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	f. Confirm that the new connection appears in the manage connection console 

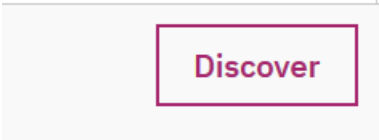
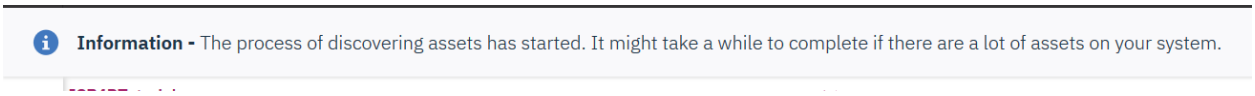
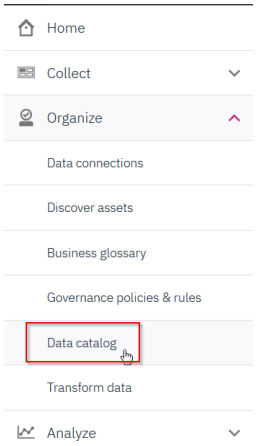
© Copyright IBM Corp. 2018. All rights reserved

Section 2: Discover and Browse Data

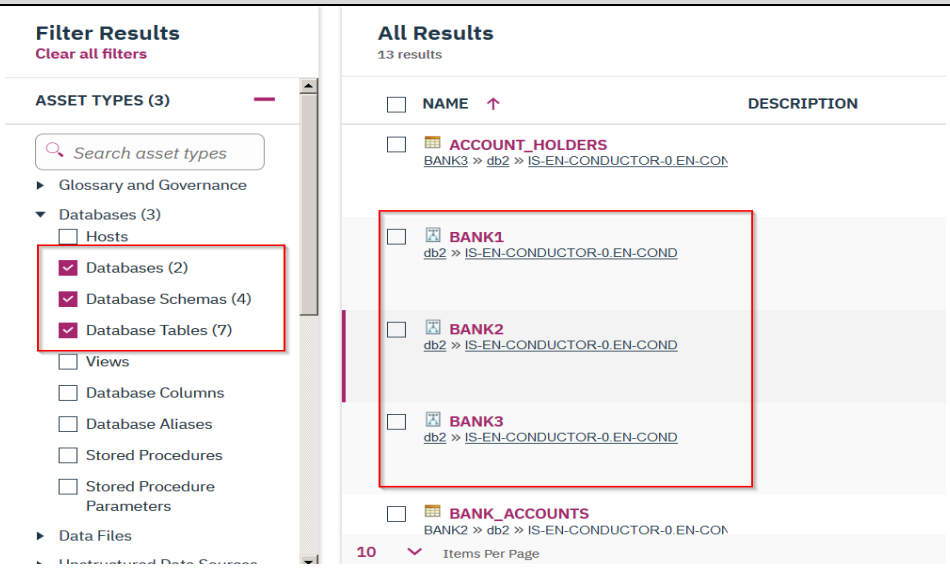
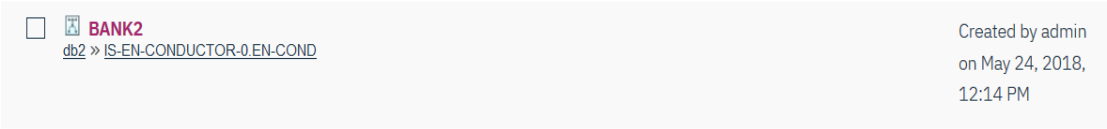
We will now discover the assets in this connection using the Discovery Services. These discovery services are key to cataloging data across the enterprise for better management in terms of search govern and use this data in various ML initiatives.

Step	Action
1	**<u>Automated Data Discovery</u>** a. On the left navigation bar click on **Organize -> Discover Assets**  b. In the form presented, choose the connection that we just created and select all the checkboxes on the last section. 

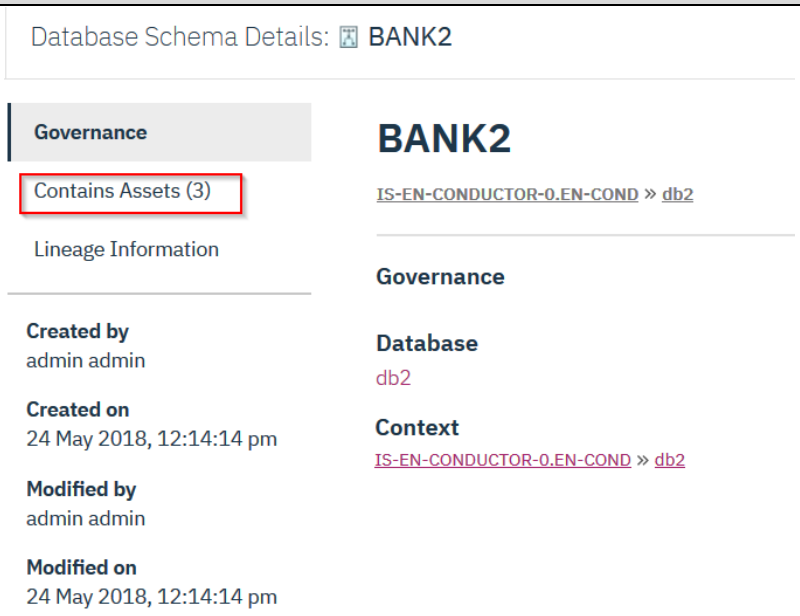
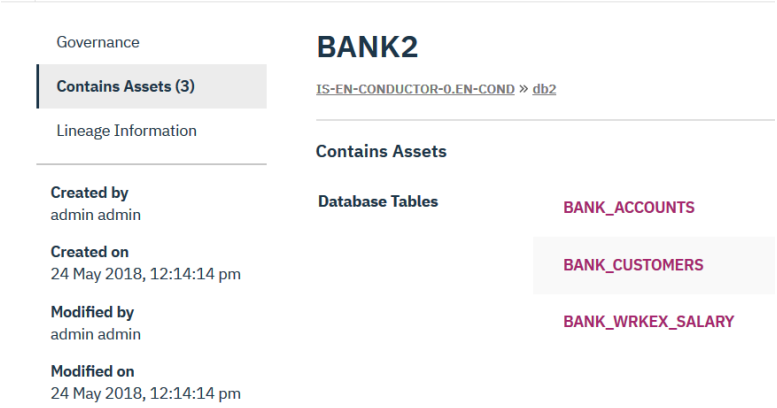
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	c. Click on Discover button on the bottom right to start the discovery Process  The system responds back with a message that the discover process has started.  Once the discovery process is completed, there are several ways the data can be browsed.
2	**<u>Browse Through the Catalog</u>** a. On the left navigation bar click on **Organize -> Data Catalog**  b. In the Data Catalog Console, Expand **Databases** and select Databases, Database Schemas, Database Tables to reveal what was discovered. The sample database had Three Schemas BANK1, BANK2, BANK3 and various tables underneath them.

© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 c. Let's explore one of the Schema "BANK2" since that is the one which would use for our next section. Click on "BANK2" on the right side. Notice that various attributes of the database are captured here along with the link to the assets contained within this database. 

© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 d. Click on “Contains Assets(3)” to explore the assets which are nothing but the tables that this schema contains. 
3	**Browse Through Enterprise Search- TBD**

© Copyright IBM Corp. 2018. All rights reserved

Section 3: Transform Data

In this module, we will use the data transformation capabilities of ICP for Data to join two tables to get the data in better shape. We would be join two tables from BANK2 schema. These two tables are BANK_ACCOUNTS and BANK_CUSTOMER. Here are the columns in both the tables.

BANK_ACCOUNTS

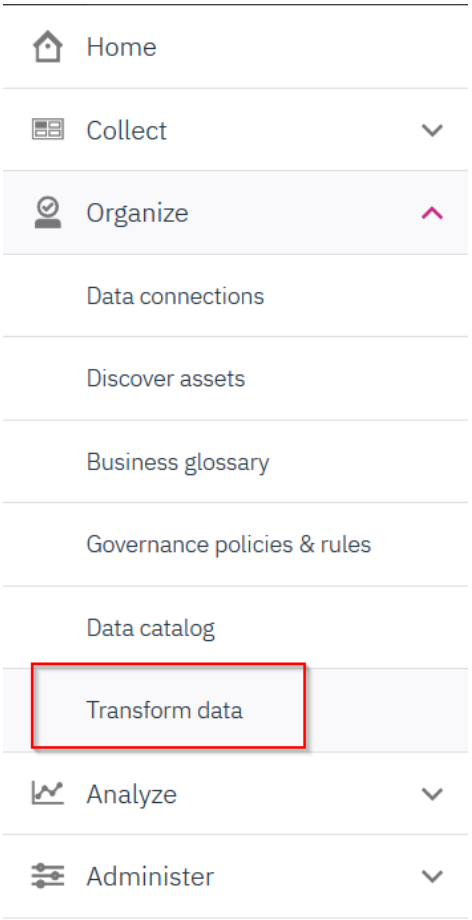
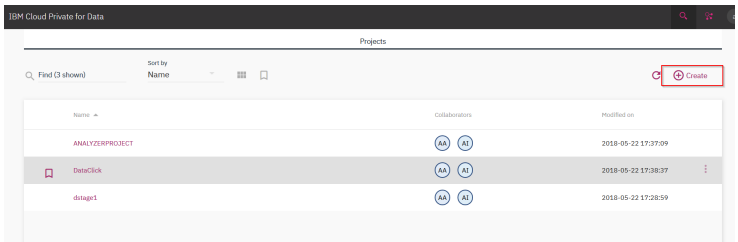
TABLE_NAME	COLUMN_NAME	DATA_TYPE	TYPE_NAME
BANK_ACCOUNTS	ACCOUNT_ID	4	INTEGER
BANK_ACCOUNTS	CUSTOMER_ID	4	INTEGER
BANK_ACCOUNTS	ACCOUNT_TYPE	1	CHAR
BANK_ACCOUNTS	ACCOUNT_BALANCE	3	DECIMAL
BANK_ACCOUNTS	JOINT_ACCOUNT_HOLDER	1	CHAR
BANK_ACCOUNTS	BANKCARD	1	CHAR
BANK_ACCOUNTS	ONLINE_ACCESS	1	CHAR
BANK_ACCOUNTS	CARDNB	12	VARCHAR
BANK_ACCOUNTS	RTN	12	VARCHAR

BANK_CUSTOMERS

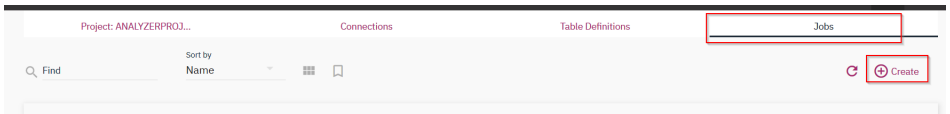
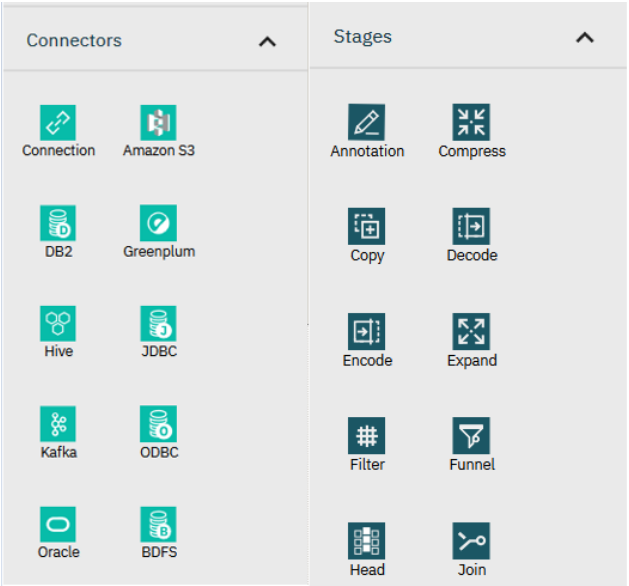
TABLE_NAME	COLUMN_NAME	DATA_TYPE	TYPE_NAME
BANK_CUSTOMERS	CUSTOMER_ID	4	INTEGER
BANK_CUSTOMERS	NAME	1	CHAR
BANK_CUSTOMERS	ADDRESS	1	CHAR
BANK_CUSTOMERS	ZIP	1	CHAR
BANK_CUSTOMERS	CREDIT_RATING	4	INTEGER
BANK_CUSTOMERS	AGE	7	REAL
BANK_CUSTOMERS	GENDER	1	CHAR
BANK_CUSTOMERS	MARITAL_STATUS	1	CHAR
BANK_CUSTOMERS	PROFESSION	1	CHAR
BANK_CUSTOMERS	NBR_YEARS_CLI	7	REAL
BANK_CUSTOMERS	EMAIL	12	VARCHAR
BANK_CUSTOMERS	CCN	12	VARCHAR
BANK_CUSTOMERS	PHONE1	12	VARCHAR
BANK_CUSTOMERS	PHONE2	12	VARCHAR
BANK_CUSTOMERS	CC	1	CHAR
BANK_CUSTOMERS	CONTACT	12	VARCHAR

We would join the tables on CUSTOMER_ID column and filter out some of the columns which may not be needed for my ML project. We would skip EMAIL and PHONE1 for this exercise.

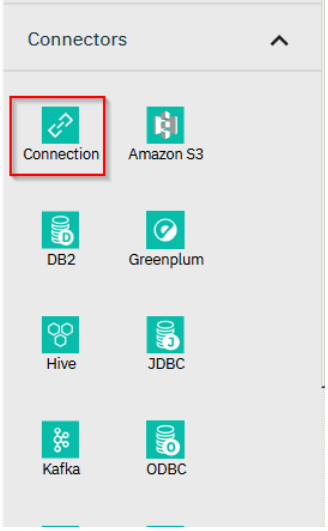
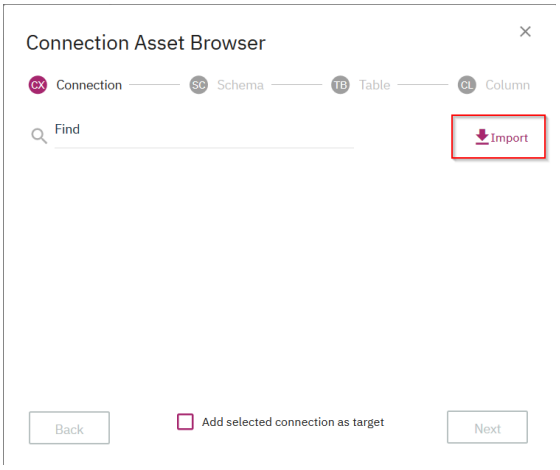
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
1	**<u>Create a Data Transformation Job</u>** d. On the left navigation bar click on **Organize -> Transform data**  e. On the project console, there are already few projects available. We would use one of them and add our transform job there. If you wanted to create a brand new project, you could have done so by clicking on "Create" on the top right corner. For this exercise, however, we will use the existing ANALYZEPROJECT and add transform job there. 

© Copyright IBM Corp. 2018. All rights reserved

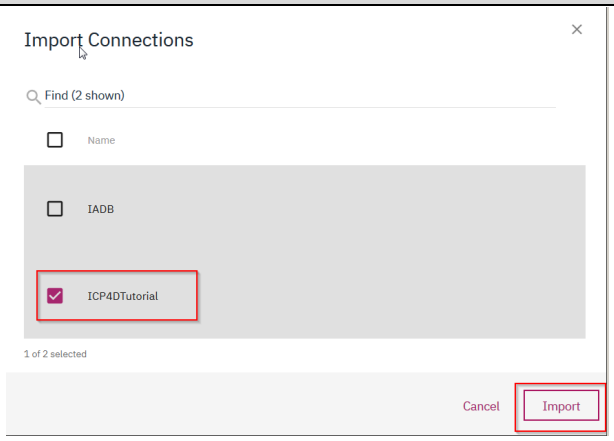
Step	Action
	f. Click on the “Jobs” tab at the top, then click “Create” 
2	**<u>Create the Data Transformation Job with the Data Flow Designer</u>** After clicking “Create”, The Data Flow Designer is loaded with pallet on the side and the designer canvas on the right. There are various connectors and stages which you can choose and build your transformation Job.  a. Our first step will be to create our data connections. On the connections panel, click on the “Connection” icon to select it and then click on the canvas to add it.

© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 b. As the connection connector is added, a dialog box is started. Click on import button and select ICP4DTutorial from the list. Once selected, click “Import” 

© Copyright IBM Corp. 2018. All rights reserved

Step	Action

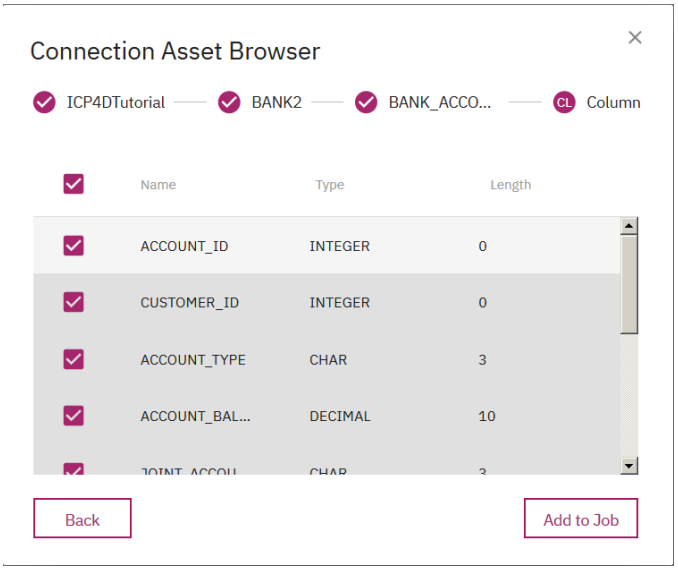


c. Select the connection you previously create and Click “Next” so the wizard moves to Schema Selection Page.

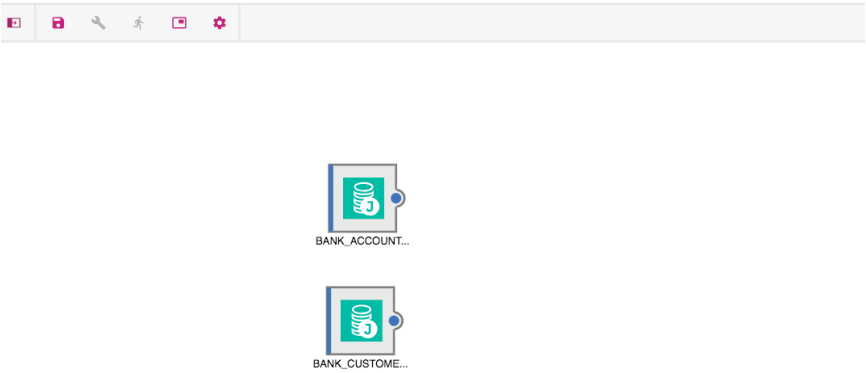
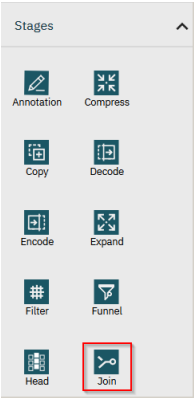
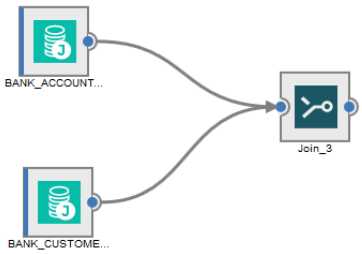
d. Select the “BANK2” schema and Click “Next”, so the wizard moves to Table Selection Page

e. Select the “BANK_ACCOUNTS” table and click “Next” to list the table’s columns.

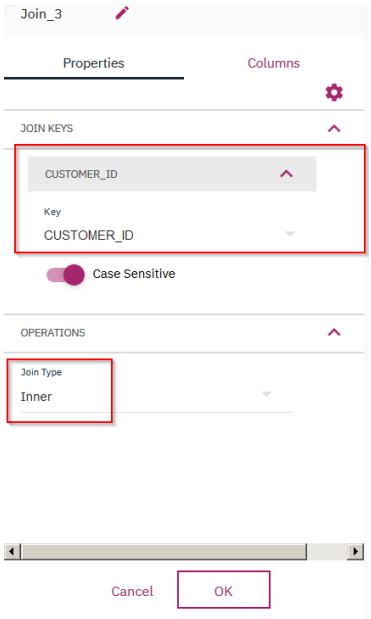
f. Since we will be using all the columns, you can simply select “Add to Job”, as shown below.



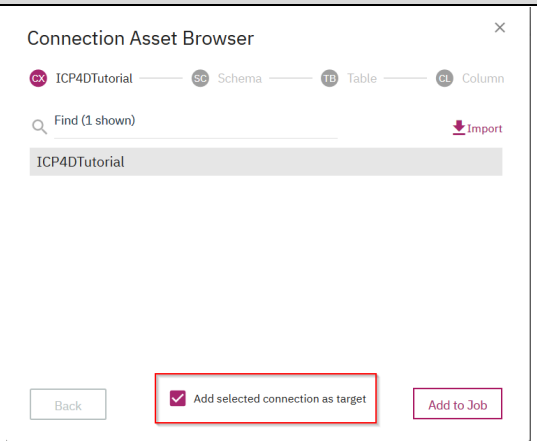
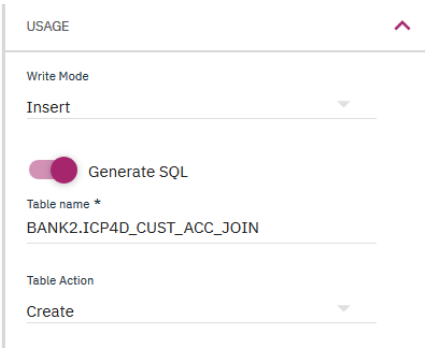
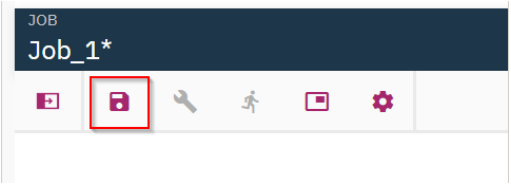
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	g. Repeat steps a-f for the BANK_CUSTOMER table. Your canvas should now look like this.  h. Select the "Join" stage from the palette and add it to the canvas.  i. Once the join stage is added, connect these three icons. Once connected, the canvas would look like this. 

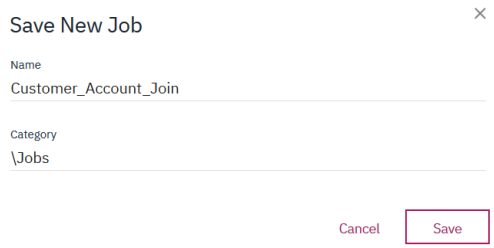
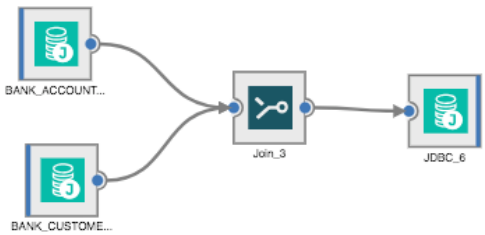
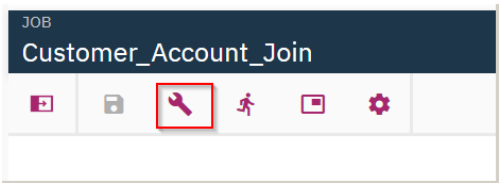
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	j. Double Click on Join to reveal additional properties for the join. On the first tab, We see that the stage has automatically detected that CUSTOMER_ID column is common between tables and suggest to go for an inner join. Click OK to agree to the selections.  k. Now we need to persist the result of this join in a table. For that we again use the "connection" connector as before, only this time we would make it as a target connection. Once marked as show, select on "Add to Job" to add it to the canvas.

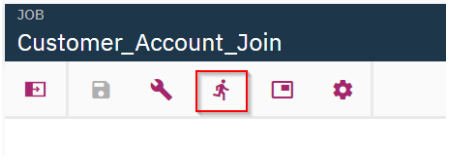
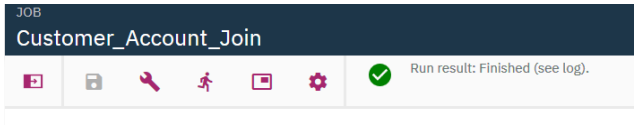
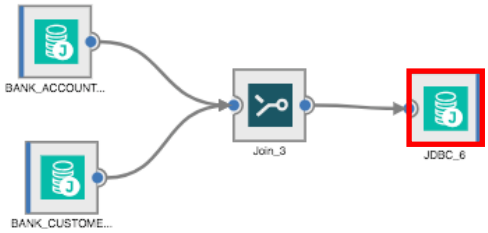
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 l. Once added, double click on this target connection to view its properties. Give a name to the generated table "**BANK2.ICP4D_CUST_ACC_JOIN_**<*your initials*>" and choose Table Action as "**Create".** Click "Ok" to save the changes to the connection.  m. At this time, a basic data transformation job is ready. Let's save the job by clicking on the highlighted button on the toolbar on the top. A dialog box pops up, lets Give a name "Customer_Account_Join_<*your initials*>" to the Job and click "**Save**" 

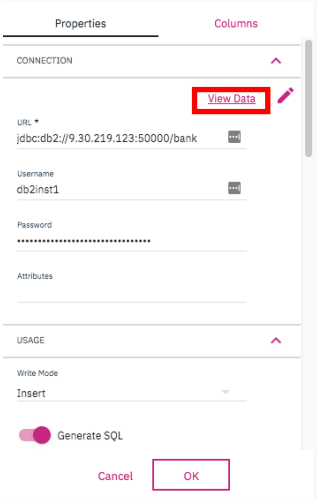
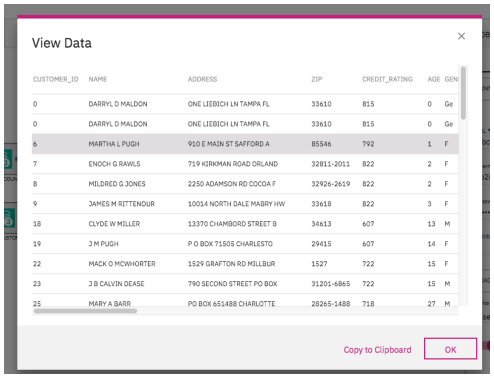
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 n. The final job design should look like this. 
3	**<u>Execute the Transformation Job</u>** a. Once the transformation job has been saved, the Compile and Run buttons are enabled. We will first compile the job by clicking on the wrench icon.  b. We can now run the job by clicking on the Run icon (note: it may take a minute or two for job to run)

© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 
4	**<u>View the Results of the Job</u>** a. To view the final table created by the transformation job, double click on the output stage from the canvas to display the properties page.  b. One the stage's properties, click on "View Data". This will query the contents of the newly created table.

© Copyright IBM Corp. 2018. All rights reserved

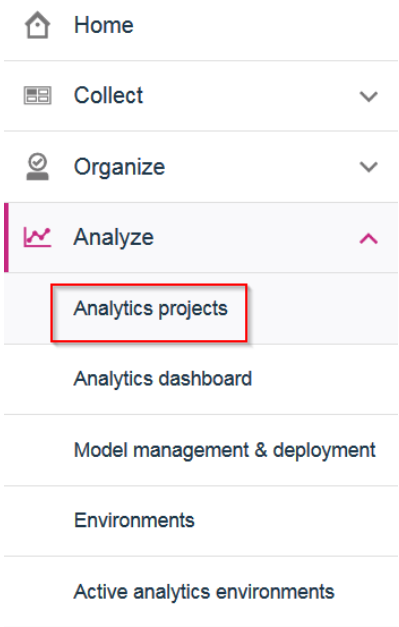
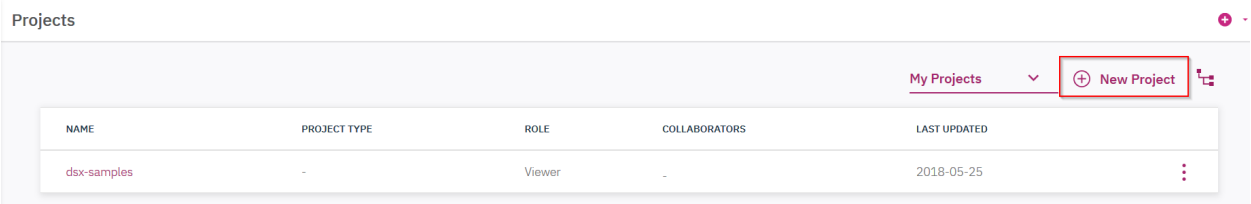
Step	Action
	 

Section 4: Building and Deploying a Machine Learning Model

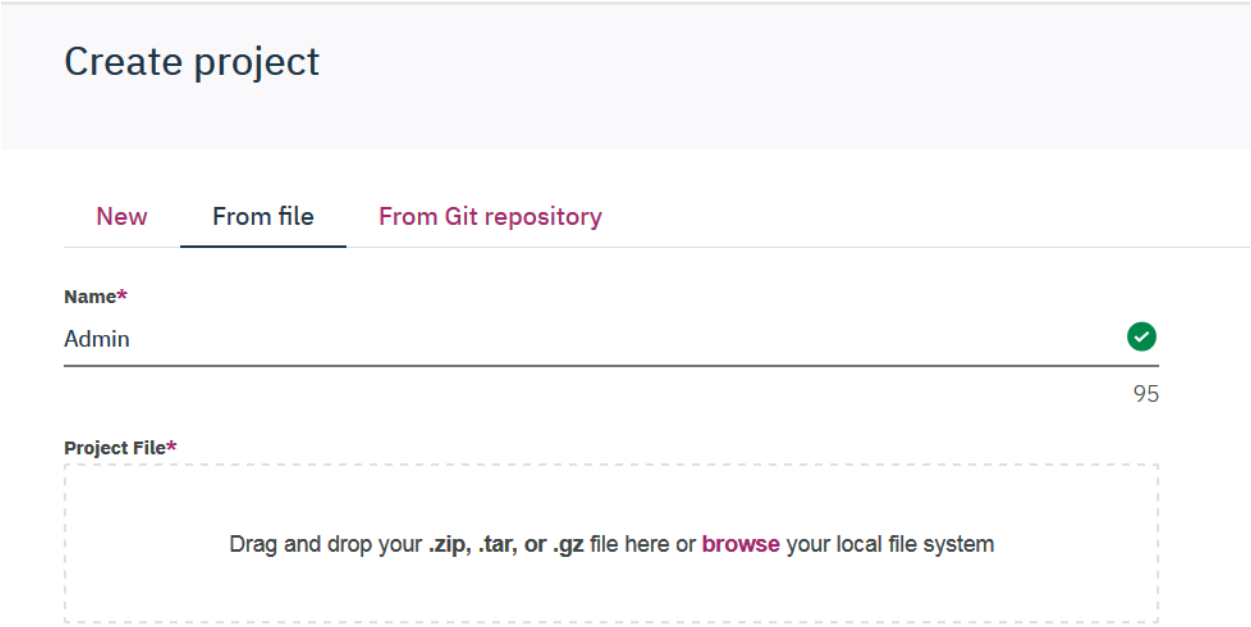
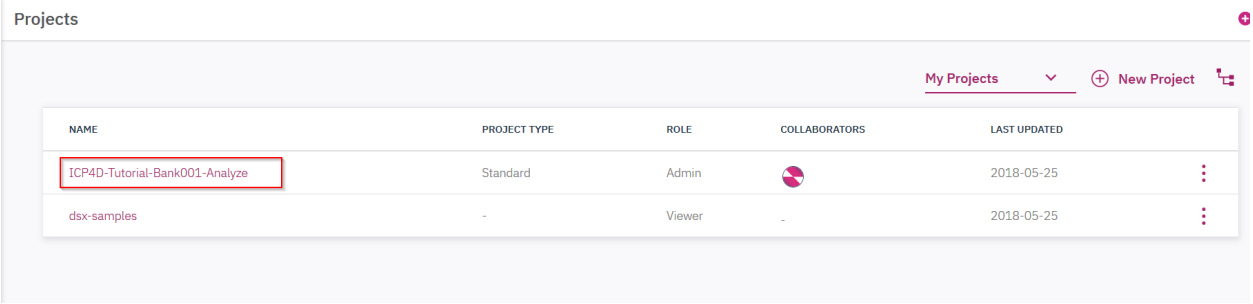
ICP4D provides proven tools and technologies for you to realize the full potential of your data. In this module, we would build a simple machine learning model and connect it to our sample dataset hosted on the database. The model would be used to make some predictions later on.

Step	Action
1	**<u>Create an Analytics Project</u>** __a. To navigate to the Analytics Project console, click on the left nav bar and select **Analyze -> Analytics Projects**

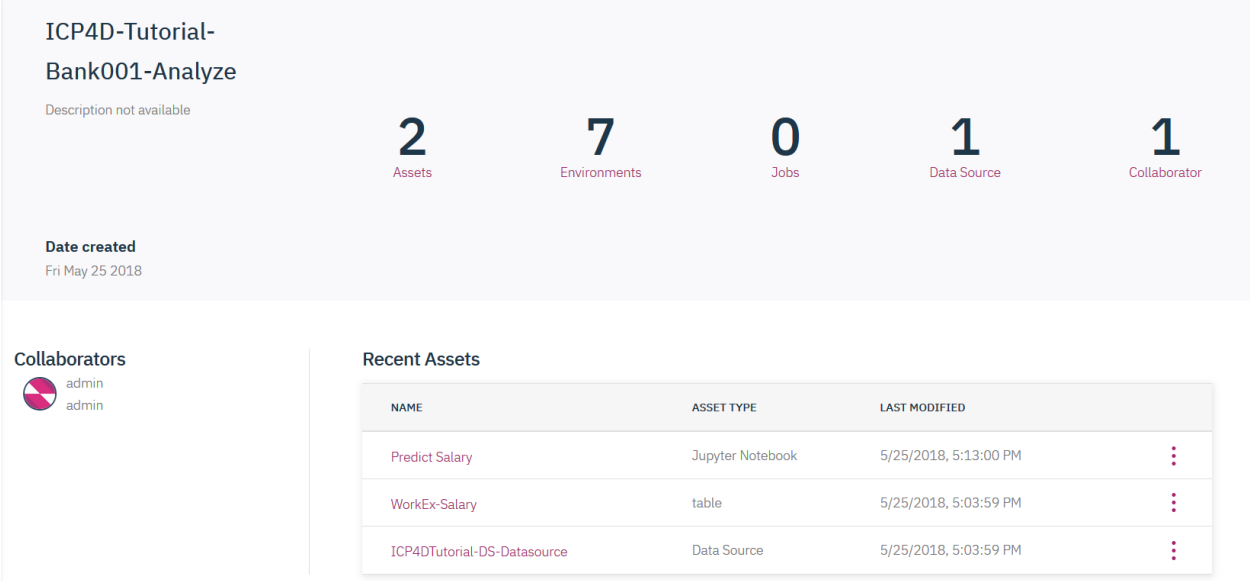
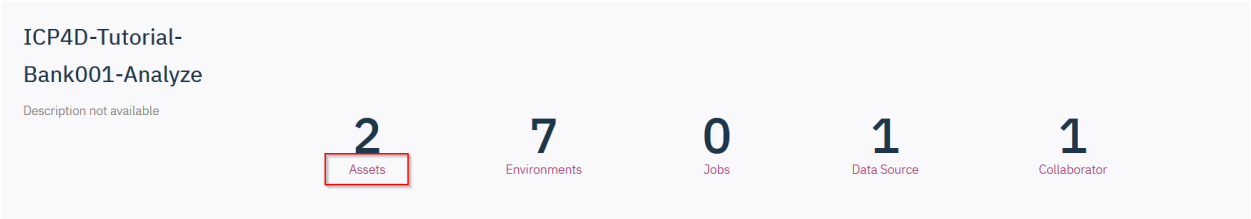
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 __b. On the Project Console, click on New Project button to add a new project screen.  __c. In this tutorial, we will import a preexisting project which primarily contains a Jupyter note book which would connect to our sample dataset and train a simple linear regression model based on the dataset. Since we are going to import an existing project, on the Create Project page, navigate to the "From File" tab. __d. A project import file is already available at the following link. Download the ICP4D-Tutorial-Bank001-Analyze.zip file to you local computer. Drag and Drop the file on this tab and system would import all the assets from the file. For the project's name, use "Predict Salary_*<your initials>".*

© Copyright IBM Corp. 2018. All rights reserved

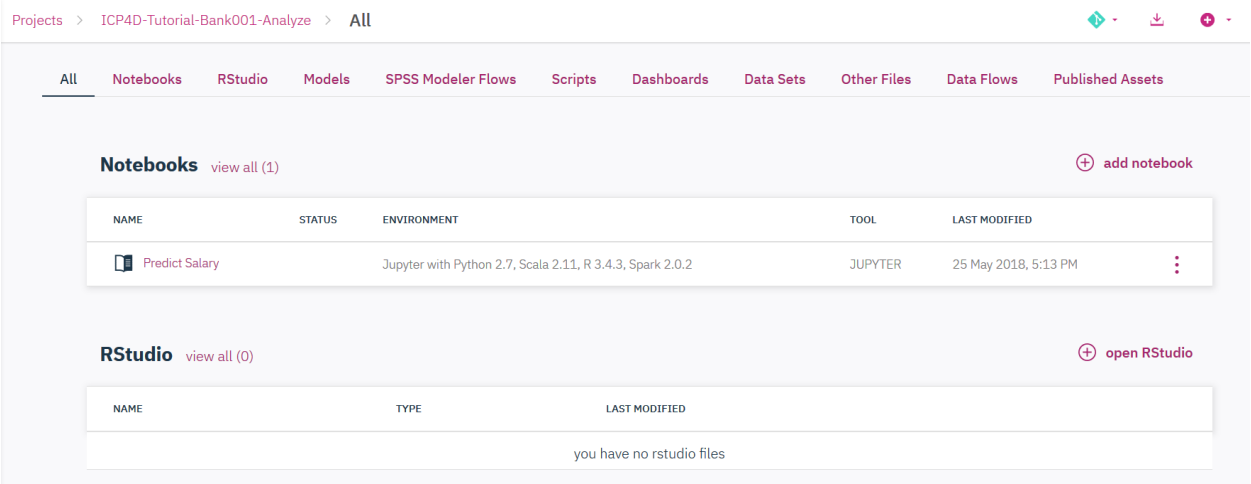
Step	Action
	 __e. Once import is complete, you should be able to see a project created on the project console.  __f. Click on the name and subsequent page shows the dashboard for various artifacts that were imported.

© Copyright IBM Corp. 2018. All rights reserved

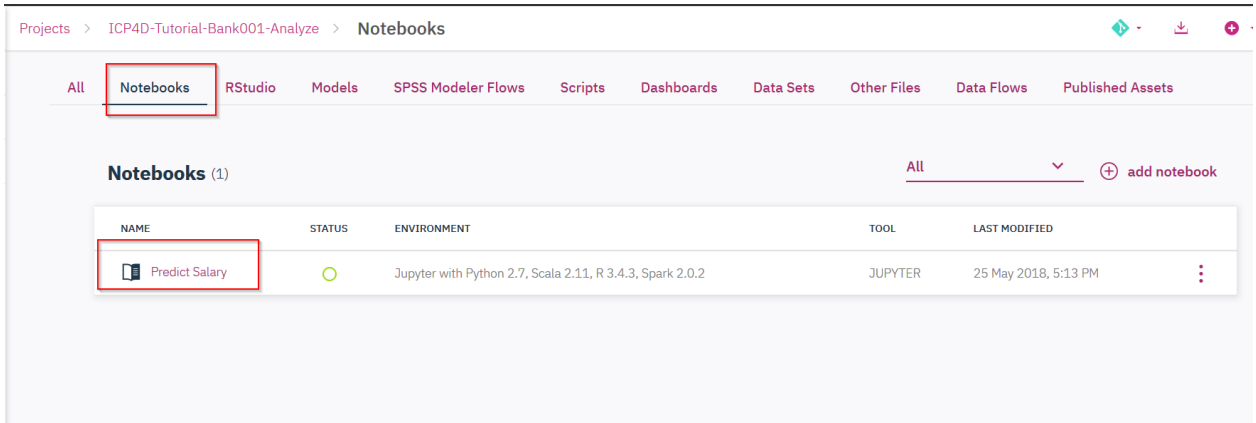
Step	Action
	
2	**<u>Explore and Run a Data Science Notebook</u>** __a. Click on Assets to see more details around it.  __b. The page shows all the assets that this project can contain. These could be various Notebooks, Models , Scripts, DataSets. In this tutorial, we have imported a Jupyter Notebook which when run would create a ML model for you to publish.

© Copyright IBM Corp. 2018. All rights reserved

Step	Action

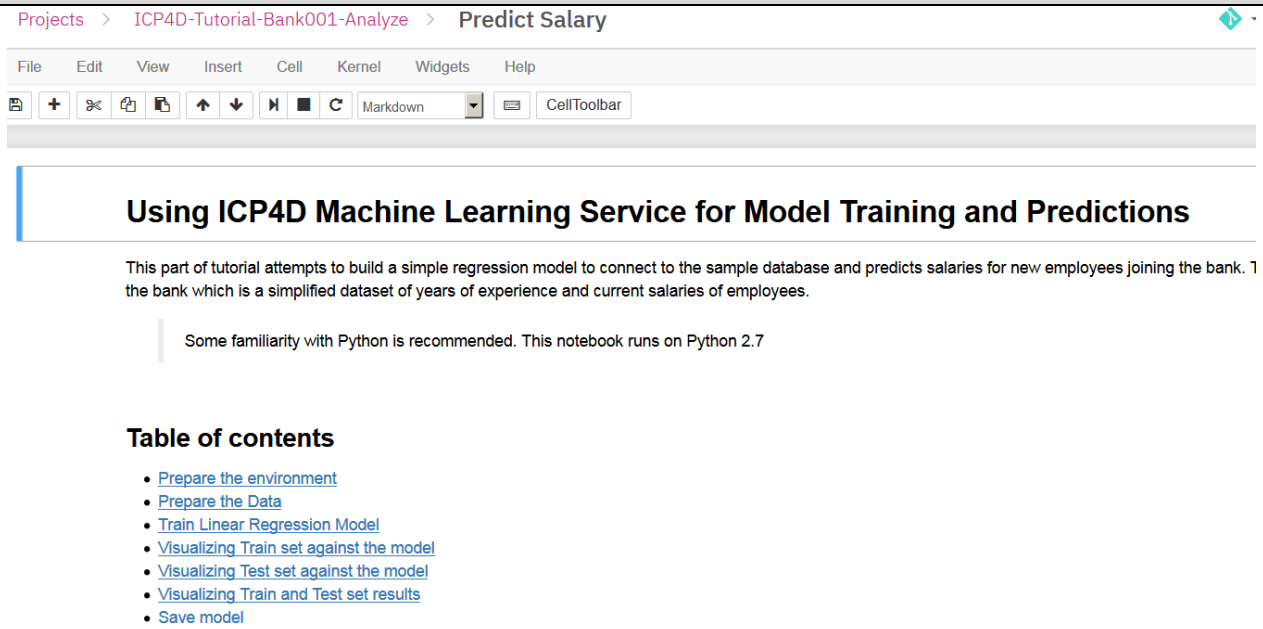


__c. Click on Notebooks Tab on top and you would see Predict Salary as one of the notebook. The details shows the environment it is meant to run in.

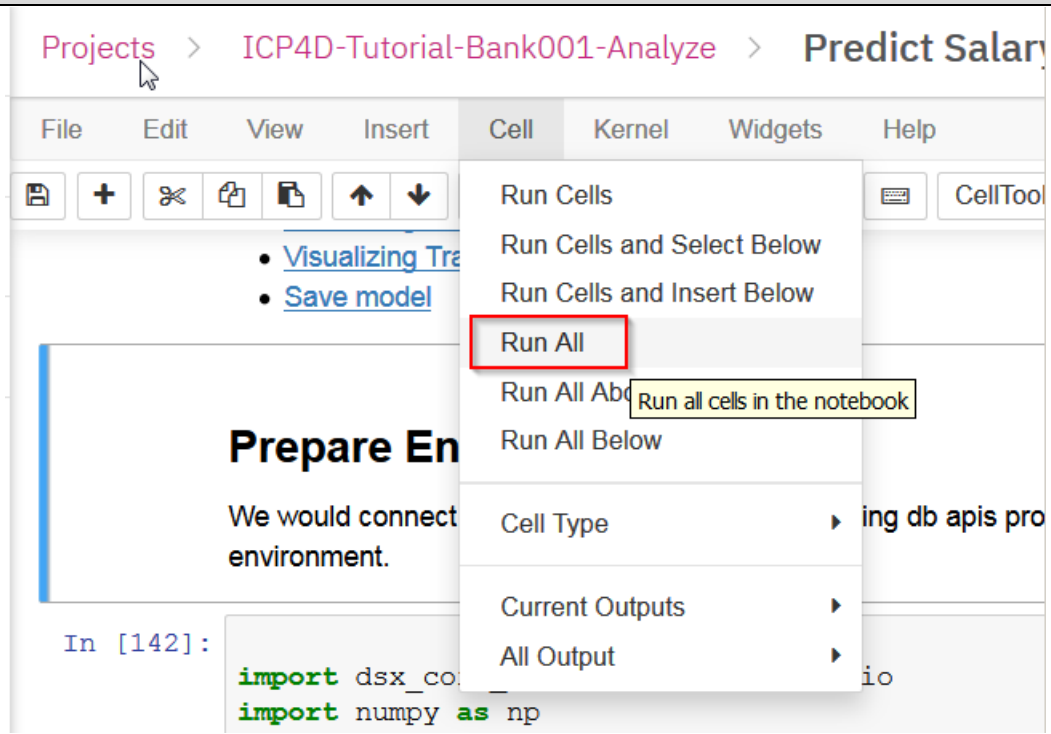


__d. Let's open the notebook. The notebook is already well annotated with documentation on each section as we build a Simple Linear Regression Model.

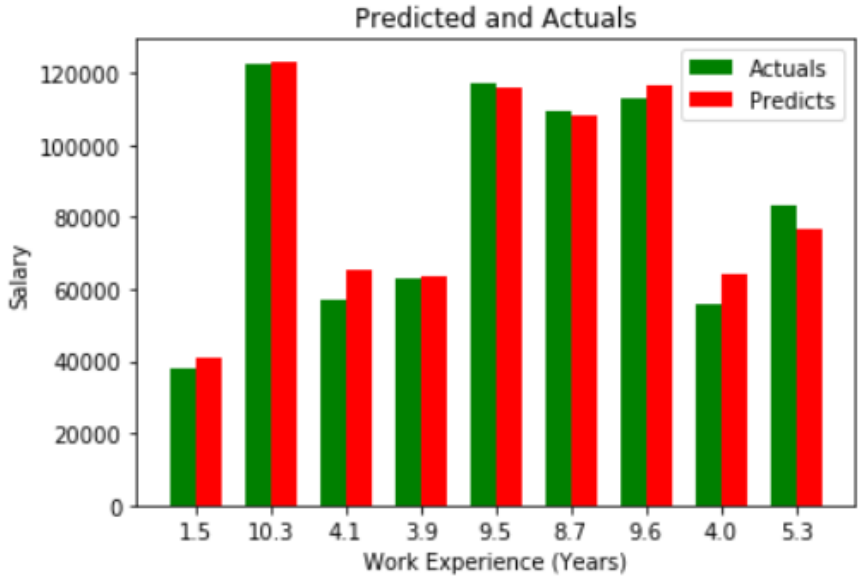
© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	 __e. Before running the notebook, please replace the Database Connection IP address in “Prepare Environment” section. __f. The notebook provides mechanism to run all the cells in one go or run only the selected cell which is a great way to test and understand the model as its been prepared, analyzed and tested. For Running all the cells Click on Cell->Run All

© Copyright IBM Corp. 2018. All rights reserved

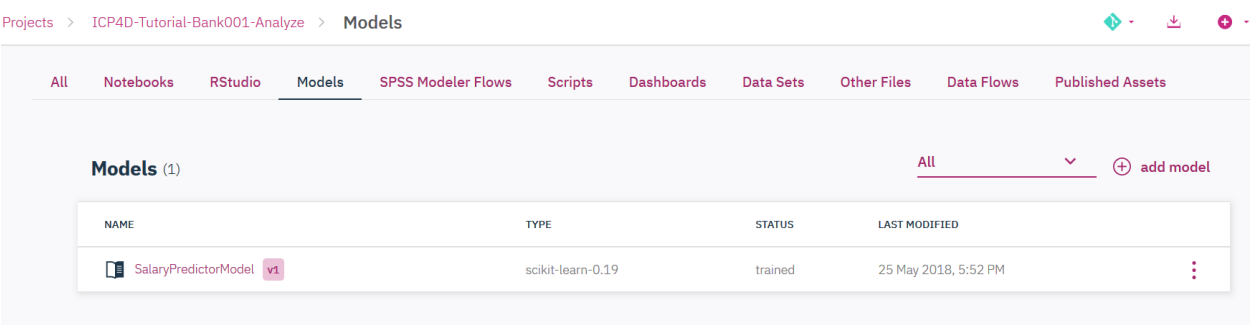
Step	Action
	 __g. Once the notebook finished running, You should be able to see plots of Training and Test data against the trained model and see the comparison of test data actual results vs test data predicted results.

© Copyright IBM Corp. 2018. All rights reserved

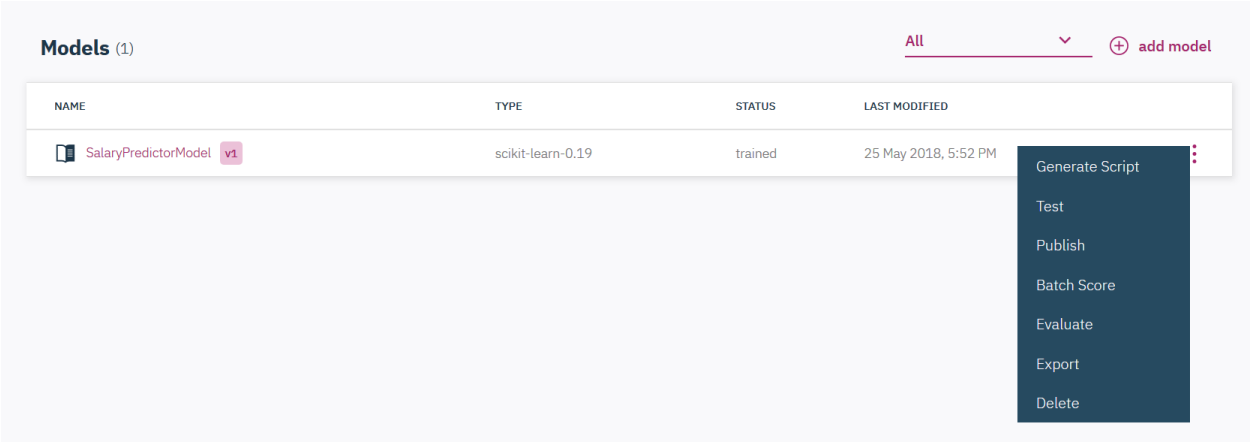
Step	Action
	
3	**<u>Examine and Test the Model</u>** __a. As part of the Notebook Run, the Model was also saved with in the project. Click on the Project name on the top section to go back to project asset __b. On the project details page, Click on “Models” Tab

© Copyright IBM Corp. 2018. All rights reserved

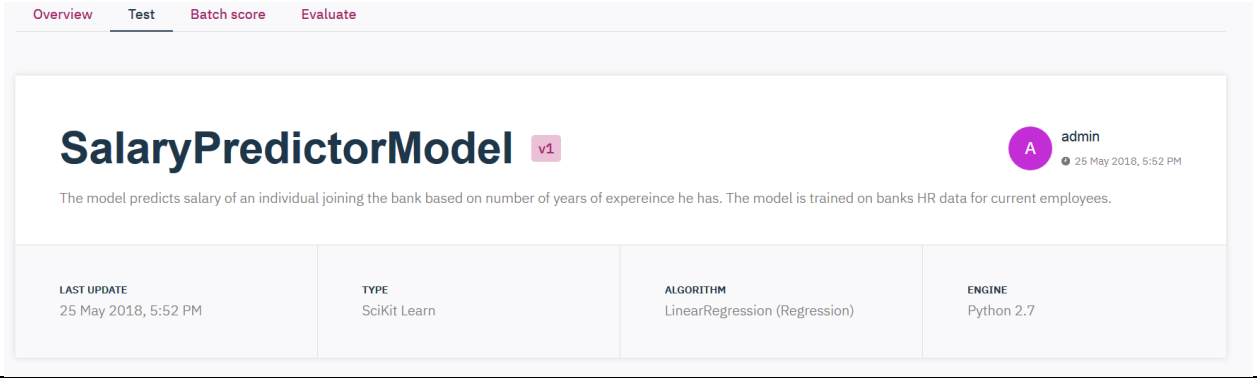
Step	Action



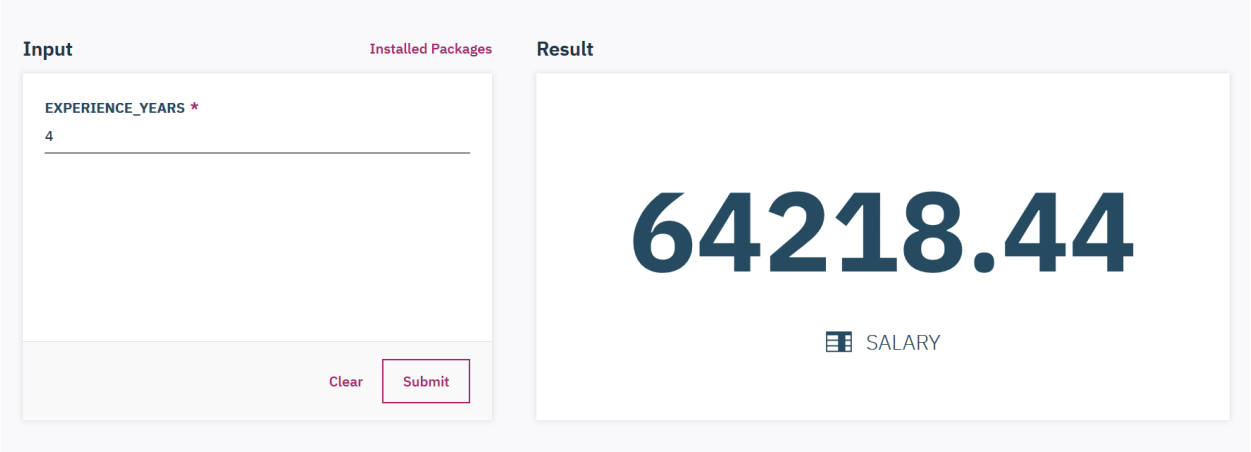
__c. ICP4D provides various options such as Test, Score and Publish this model to enterprise catalog for business users and IT to use for building apps or leveraging it for predictions. Clicking on the Options icon on the extreme right provides those options.



__d. Let's close this lab by quickly testing this ML model with some data. Click on "Test" from the available options and you can see details about the model. You can see that this is a Linear Regression Model built with Python 2.7 engine.



© Copyright IBM Corp. 2018. All rights reserved

Step	Action
	In the below section, you can enter the input data to get the salary predictions. A sample prediction for salary for 4 years of work experience is shown below. 
4	**<u>Deploy and Test the TensorFlow Model</u>**

Thank you for your time today!!

The IBM Wolfpack Team

© Copyright IBM Corp. 2018. All rights reserved

© Copyright IBM Corp. 2018. All rights reserved

© Copyright IBM Corp. 2018. All rights reserved