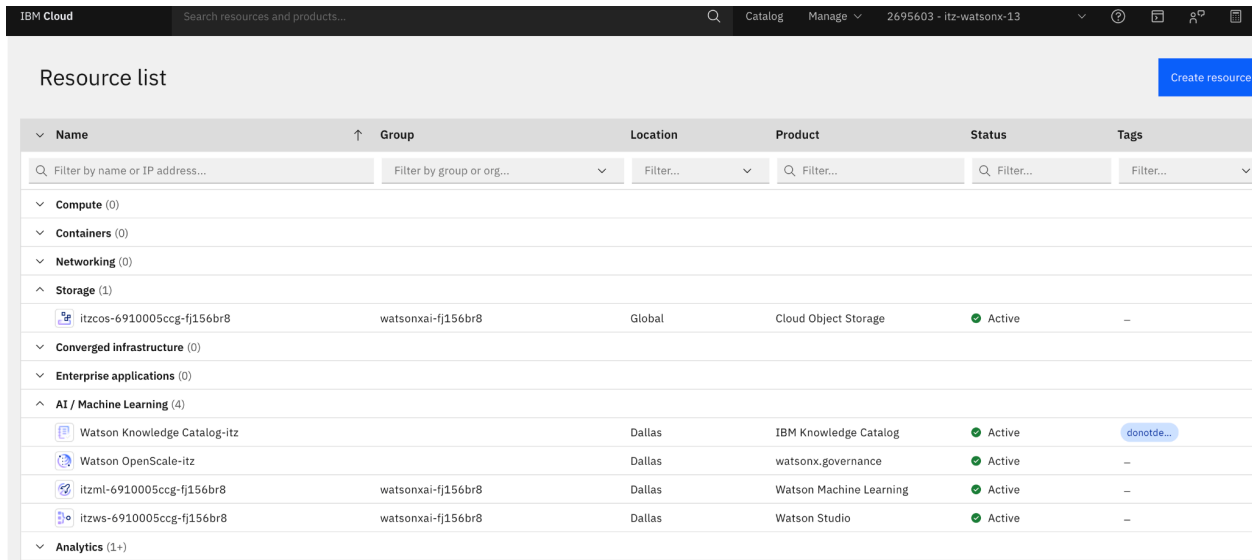


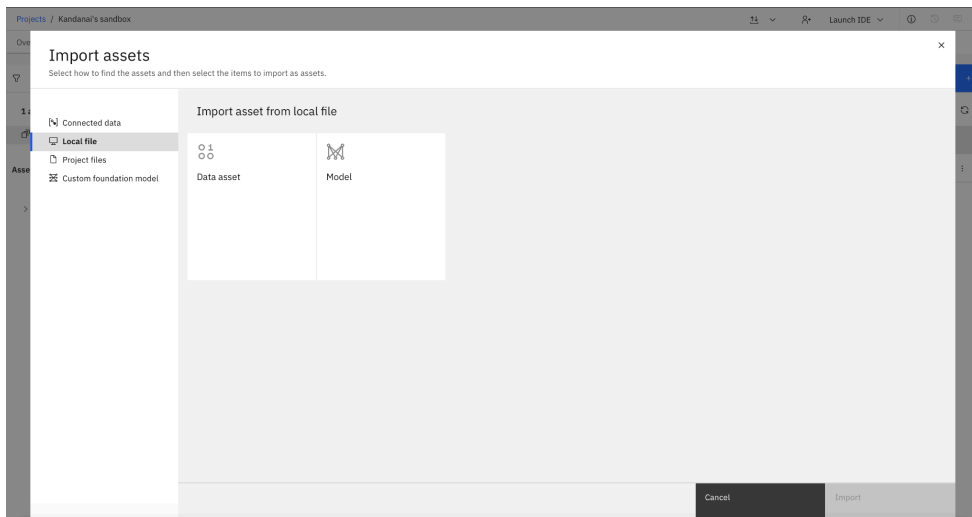
LAB 1 Data Visualization and Creating Connections

Ensure you have the sufficient resources required to do the lab (See Lab0)

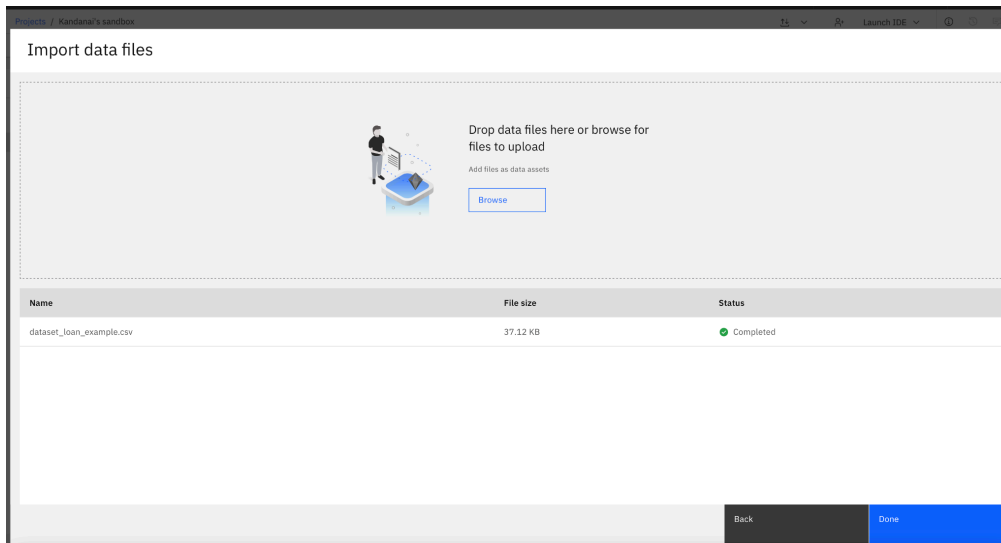


Name	Group	Location	Product	Status	Tags
Compute (0)					
Containers (0)					
Networking (0)					
Storage (1)					
itzc0s-6910005ccg-fj156br8	watsonxai-fj156br8	Global	Cloud Object Storage	Active	—
Converged infrastructure (0)					
Enterprise applications (0)					
AI / Machine Learning (4)					
Watson Knowledge Catalog-itz		Dallas	IBM Knowledge Catalog	Active	donotde...
Watson OpenScale-itz		Dallas	watsonx.governance	Active	—
itzml-6910005ccg-fj156br8	watsonxai-fj156br8	Dallas	Watson Machine Learning	Active	—
itzws-6910005ccg-fj156br8	watsonxai-fj156br8	Dallas	Watson Studio	Active	—
Analytics (1+)					

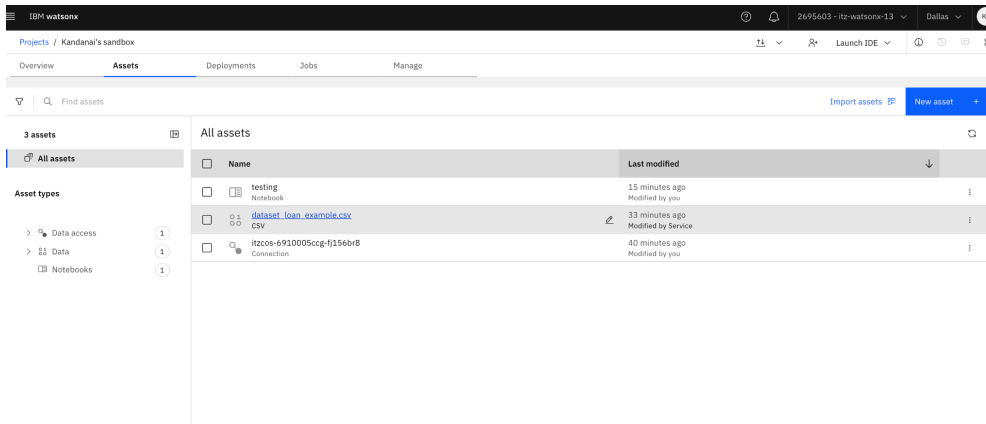
Firstly, locate your project and click on the import asset in the watsonx. Here our objective is to import a local CSV file as an asset to your sandbox project. Click on Data asset below:



Drag and drop the provided dataset_loan_example_v2.csv, then click on Done. Once Done, this will import the csv file to your project.



If you click on your dataset dataset_loan_example_v2.csv



You will be able to

- 1) View the dataset (Your Raw data)
- 2) Data Profile to look at the distribution of the data ie. is the problem imbalance
- 3) Do some visualization ie. plotting graphs to perform visualizations

Projects / Kandana's sandbox / dataset_loan_example.csv

Prepare data

Preview asset

Profile

Visualization

Feature group β

Columns: 13 | Sample rows: 614

Last refresh: 4 seconds ago

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount
LP001002	Male	No	0	Graduate	No	5849	0	
LP001003	Male	Yes	1	Graduate	No	4583	1508	128
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120
LP001008	Male	No	0	Graduate	No	6000	0	141
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267
LP001013	Male	Yes	0	Not Graduate	No	2333	1516	95
LP001014	Male	Yes	3+	Graduate	No	3036	2504	158
LP001018	Male	Yes	2	Graduate	No	4006	1526	168
LP001020	Male	Yes	1	Graduate	No	12841	10968	349
LP001024	Male	Yes	2	Graduate	No	3200	700	70
LP001027	Male	Yes	2	Graduate		2500	1840	109
LP001028	Male	Yes	2	Graduate	No	3073	8106	200
LP001029	Male	No	0	Graduate	No	1853	2840	114
LP001030	Male	Yes	2	Graduate	No	1299	1086	17
LP001032	Male	No	0	Graduate	No	4950	0	125
LP001034	Male	No	1	Not Graduate	No	3596	0	100
LP001036	Female	No	0	Graduate	No	3510	0	76
LP001038	Male	Yes	0	Not Graduate	No	4887	0	133
LP001041	Male	Yes	0	Graduate		2600	3500	115
LP001043	Male	Yes	0	Not Graduate	No	7660	0	104
LP001046	Male	Yes	1	Graduate	No	5955	5625	315

About this asset

Name

dataset_loan_example.csv

csv

Description

What's the purpose of this asset?

Tags

Add tags to make assets easier to find.

Last modified

Now by Kandana Lee

Created on

Sep 25, 2024 by Kandana Lee

After visualizing the data let's try importing the data into watson studio (Jupyter Notebook)

Projects / Kandana's sandbox

Launch IDE

What do you want to do?

Select a task based on your goal. You'll use a tool to create an asset for that goal.

All

Prepare data

Work with models

Automate model lifecycles

Search for a task or tool

Recents

Connect to a data source

with Connection

Work with data and models in Python or R notebooks

Learn more

with Jupyter notebook editor

Prepare data

Connect to a data source

with Connection

Ground gen AI with vectorized documents

with Vector indexes

Prepare and visualize data

with Data Refinery

Define reusable sets of parameters

with Parameter set

Enter any name to create the notebook and select the Runtime to be “Runtime 24.1 on Python 3.11 XS (2 VCPU 8 GB RAM)”

Work with data and models in Python or R notebooks

Define the details to create a notebook asset and open it in the Jupyter notebook editor tool.

+ New

Sample

Local file

URL

Define details

Name

testing

Description (optional)

What's the purpose of this notebook

Define configuration

Select runtime

Runtime 24.1 on Python 3.11 XS (2 vCPU 8 GB RAM)

The selected runtime has 2 vCPU and 8 GB RAM.
It consumes 1 capacity unit per hour.
[Learn more](#) about capacity unit hours and Watson Studio pricing plans.

Language

☒ Python 3.11

Cancel

Create

Here you will see the jupyter notebook being provisioned. Click on the `</>` icon and read data. Here select the csv file in the data asset you have just uploaded.

IBM watsonx

Projects / Kandana's sandbox / testing

Not Trusted Memory:148 / 8192 MB Python 3.11

Edit View Run Kernel Help

Code

[]: ...

Code Snippets

Make your data science workflow more simple and productive with Code Snippets.

[Data Ingestion]

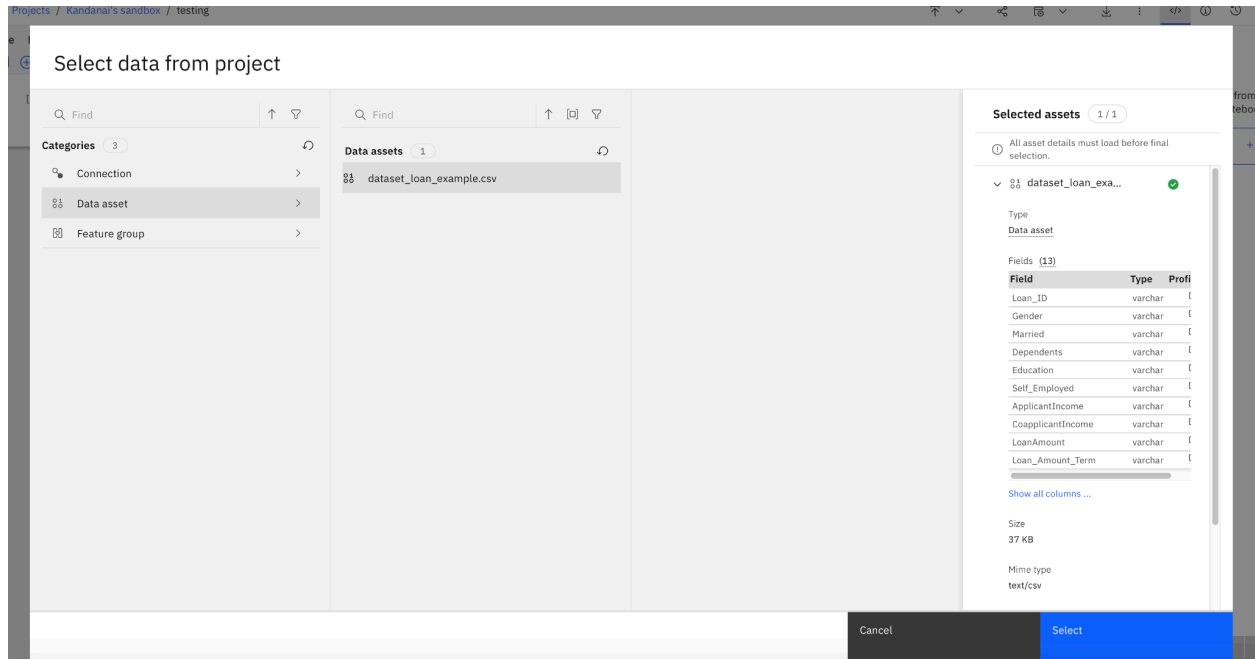
Read data

Generate a code snippet to load data from a data asset or connection into your notebook.

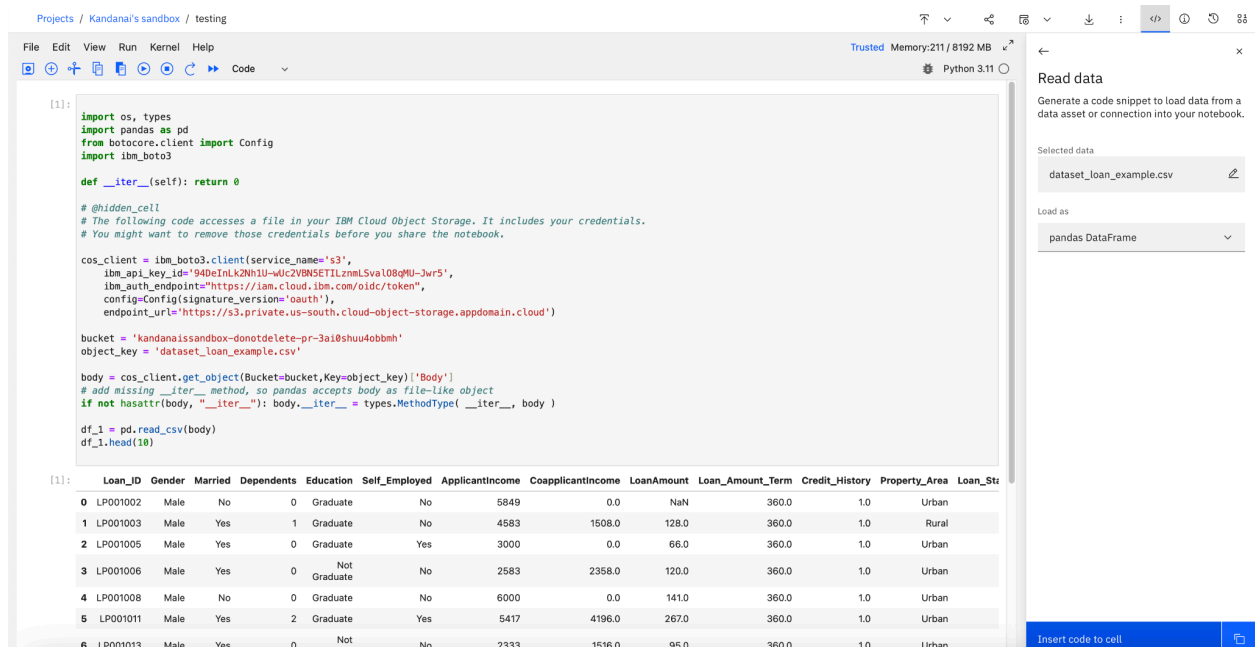
[Prompt engineering]

Add a sample prompt

Add a prompt from the Samples collection.

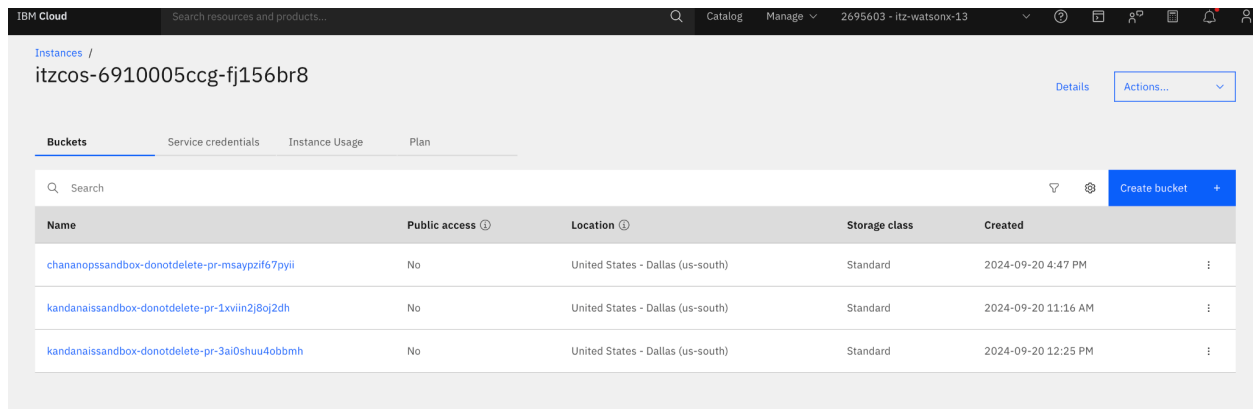


Watson studio will automatically generate code to create connections with the data asset and write code to create pandas dataframe. Click insert to cell and run the cell.



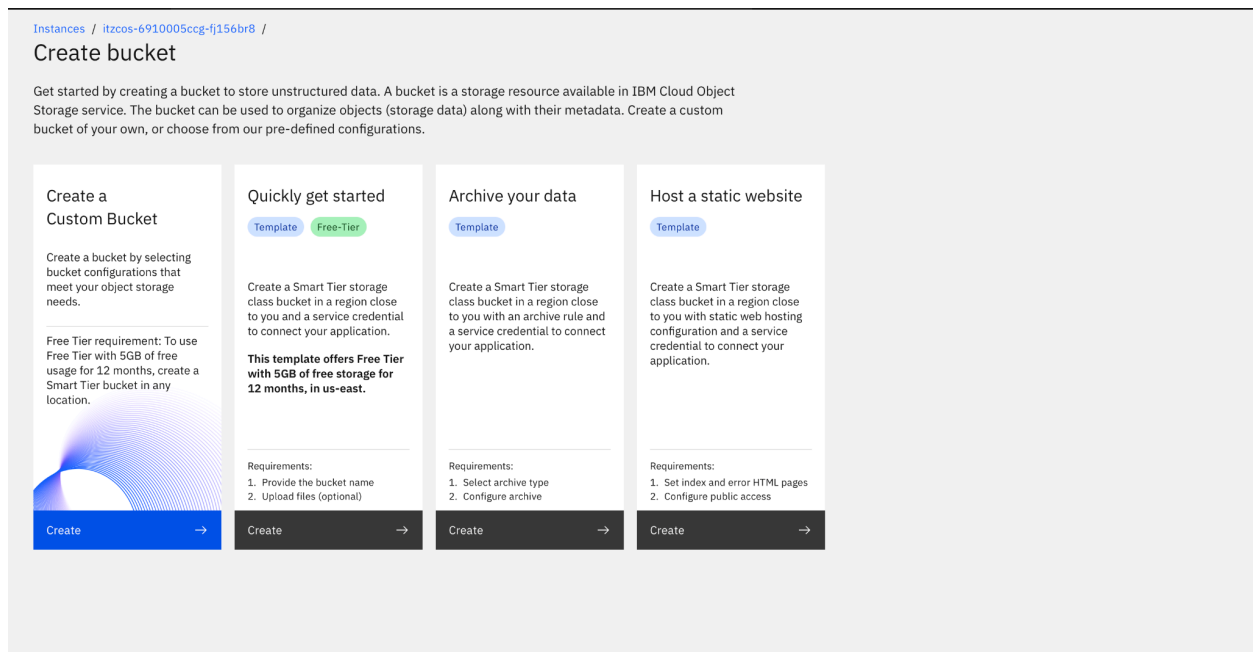
LAB 1 Data Visualization and Creating Connections (Optional)

Optionally, instead of uploading the data directly into the project, we can create a cloud storage and create connections to that cloud storage. Here we will create an IBM Cloud Object Storage and repeat the steps before.



Name	Public access	Location	Storage class	Created
chananopssandbox-donotdelete-pr-msaypzif67pyii	No	United States - Dallas (us-south)	Standard	2024-09-20 4:47 PM
kandanaissandbox-donotdelete-pr-1xvlin2j8oj2dh	No	United States - Dallas (us-south)	Standard	2024-09-20 11:16 AM
kandanaissandbox-donotdelete-pr-3ai0shuu4obbmh	No	United States - Dallas (us-south)	Standard	2024-09-20 12:25 PM

Create a custom bucket



Instances / itzcos-6910005ccg-fj156br8 / Create bucket

Get started by creating a bucket to store unstructured data. A bucket is a storage resource available in IBM Cloud Object Storage service. The bucket can be used to organize objects (storage data) along with their metadata. Create a custom bucket of your own, or choose from our pre-defined configurations.

Create a Custom Bucket

Create a bucket by selecting bucket configurations that meet your object storage needs.

Free Tier requirement: To use Free Tier with 5GB of free usage for 12 months, create a Smart Tier bucket in any location.

Create →

Quickly get started

Free-Tier

Create a Smart Tier storage class bucket in a region close to you and a service credential to connect your application.

This template offers Free Tier with 5GB of free storage for 12 months, in us-east.

Requirements:

1. Provide the bucket name
2. Upload files (optional)

Create →

Archive your data

Template

Create a Smart Tier storage class bucket in a region close to you with an archive rule and a service credential to connect your application.

Requirements:

1. Select archive type
2. Configure archive

Create →

Host a static website

Template

Create a Smart Tier storage class bucket in a region close to you with static web hosting configuration and a service credential to connect your application.

Requirements:

1. Set index and error HTML pages
2. Configure public access

Create →

Give an a name to your custom bucket and ensure it is selected in the United States region (us-south)

Unique bucket name


ibm-enablement-bucket-1

×

- Must be unique across the **whole** IBM Cloud Object Storage system
- Do not use any personal information (any part of a name, address, financial or security accounts or SSN)
- Must start and end in alphanumeric characters (3 to 63)
- Characters allowed: lowercase, numbers and nonconsecutive dots and hyphens

Resiliency

A bucket's resiliency is defined by the endpoint used to create it.

[View options](#) 

Cross Region
Highest availability



Regional
Best performance



Single Site
Data sovereignty


 Resiliency cannot be modified after provisioning

Location

Select a location where you would like your data to be physically stored.


[View options](#)

United States - Dallas (us-south)

 Location cannot be modified after provisioning

Storage class

Choose storage class based on how often you expect to read the stored data. Pricing varies for each option.

[View pricing](#) 

Smart Tier Free-Tier enabled

Smart Tier automatically gives you the lowest storage rate

Standard

For active workloads that require higher performance and

Cancel

Create bucket

Upload the provided csv in the repo to the cloud storage bucket.

Instances / itzcos-6910005ccg-fj156br8 /

ibm-enablement-bucket-1

Objects

Configuration

Permissions

If you're seeing more usage than expected, versions count towards your usage or you may have incomplete uploads [Learn more](#)

ibm-enablement-buc... /

Prefix filter

<input type="checkbox"/>	Object name	Archived ¹	Size	Last
<div><div><div><div></div><div>Objects</div></div><div><div>Drag and drop files (objects) to upload. An object is your data in fixed form.</div><div>Drag and drop files (objects) here or click to upload</div></div></div></div>				

Upload

Choose upload type

☒ Standard transfer

200 MB size limit - Ideal for smaller files such as html, images, and documents.

☐ Enable large file (200 MB+) uploads from the web on this bucket.

☐ Aspera high-speed transfer

No size limit - Recommended for video and audio files, or large files and folders. Installation required.

[Install Aspera](#)

Upload files (objects)

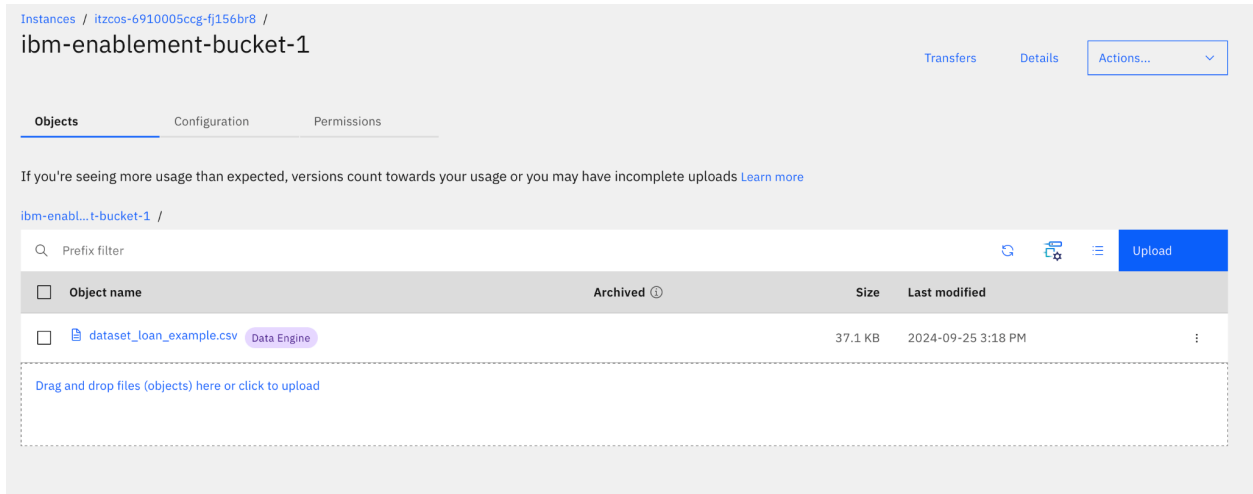
Files with duplicated names are replaced. Exclude personal information such as name or address. No special characters: /\"?:?<>*&*|

Drag and drop files and folders or click to upload

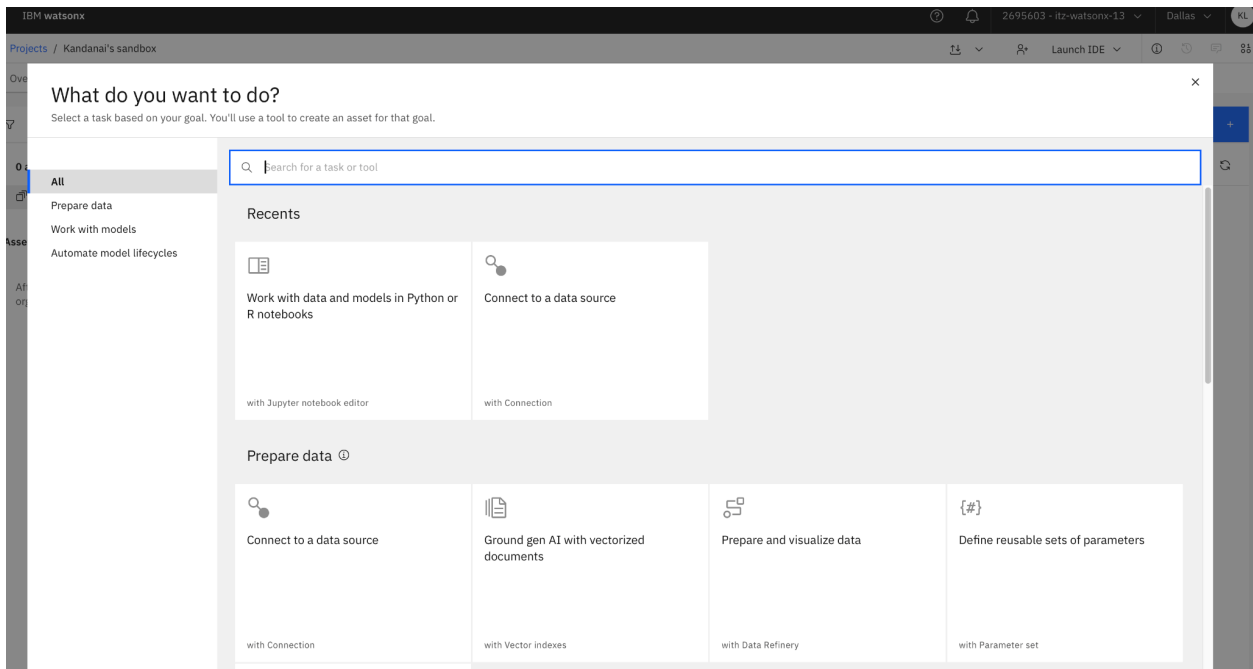
Upload files

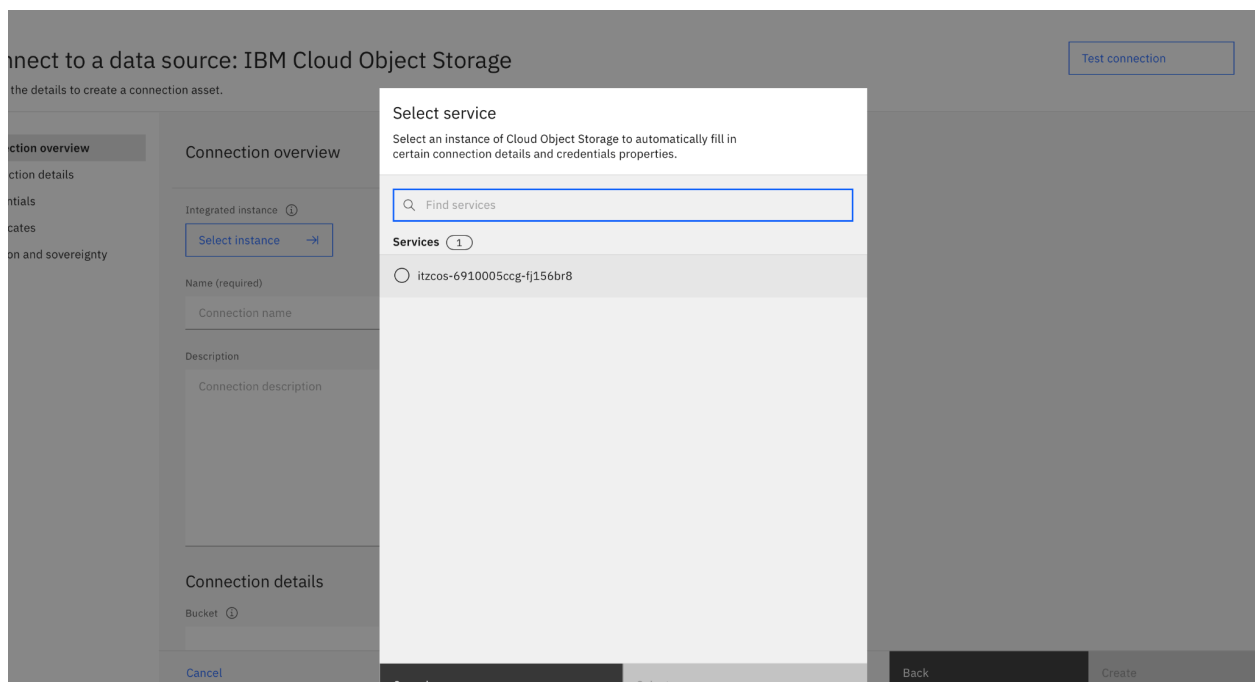
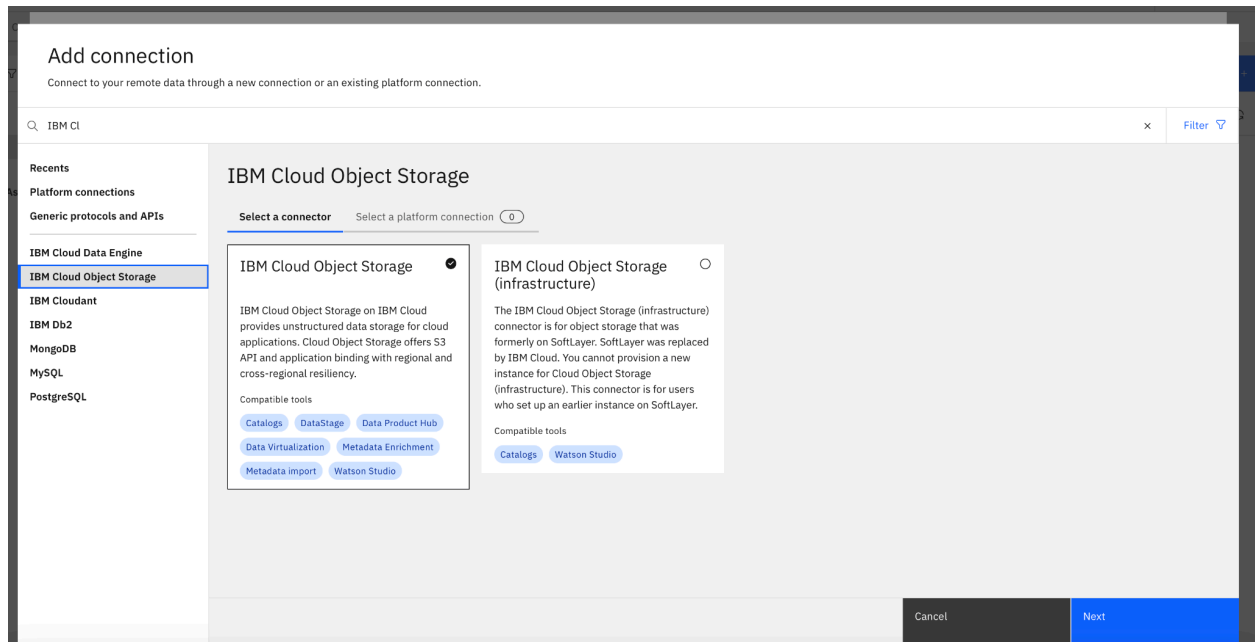
Upload folders

Organize your objects (advanced) ^



Next, return to the watsonx.ai interface and create a connection. Please select the datasource to be IBM Cloud Object Storage.





Choose the corresponding connection and enter your bucket name. Please Test your connection. If the test is successful please save to create your connection.

Projects / Kandanaï's sandbox / itzcos-6910005ccg-fj156br8

Edit connection: IBM Cloud Object Storage

Review the connection information

The test was successful. Click Save to update the connection information.

Connection overview [Show more](#)

Name (required)
itzcos-6910005ccg-fj156br8

Description
Connection description

Connection details

Bucket ⓘ
ibm-enablement-bucket-1

Login URL (required) ⓘ
s3.us-south.cloud-object-storage.appdomain.cloud

Cancel Save

Projects / Kandanaï's sandbox

Overview **Assets** Deployments Jobs Manage


Find assets [Import assets](#) [New asset](#) +

1 assets

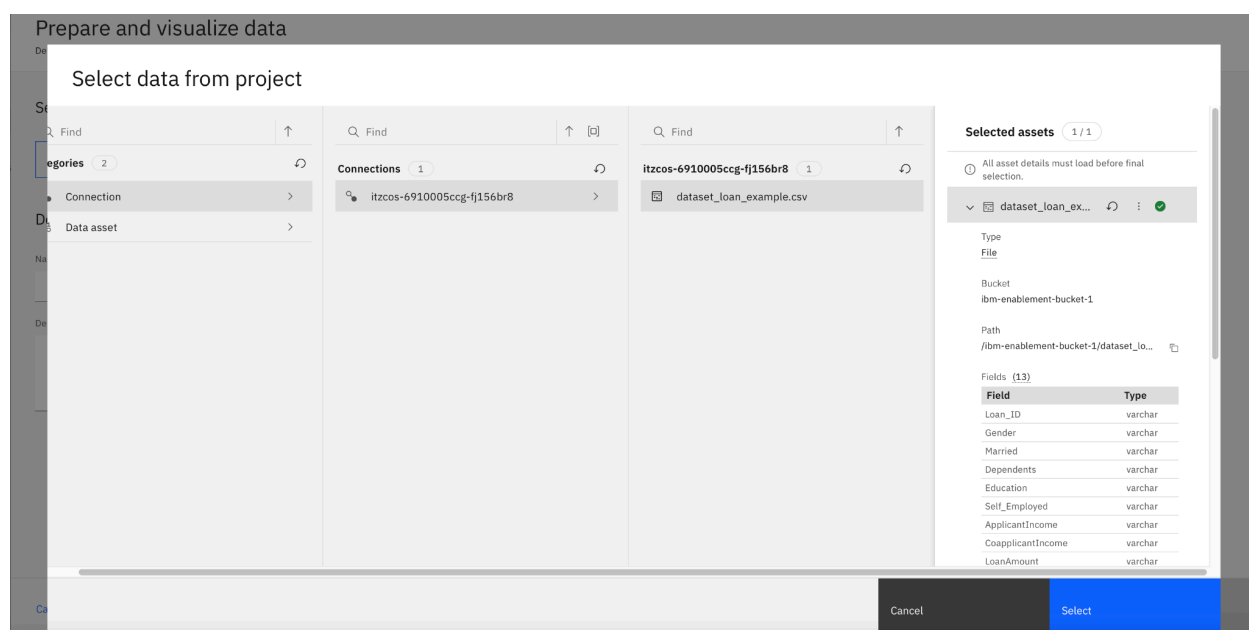
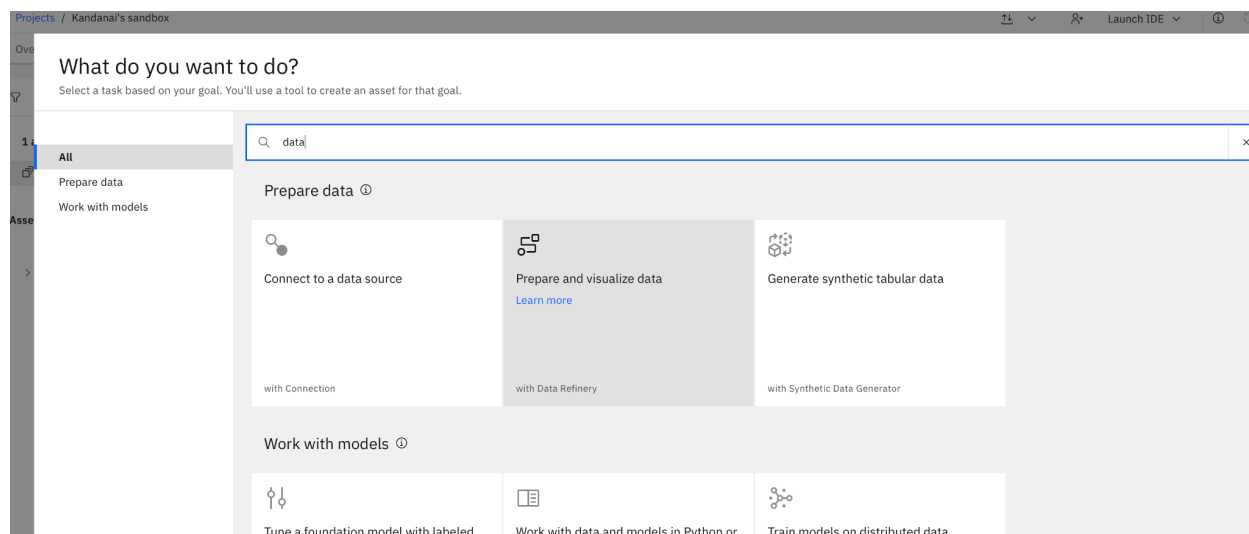
All assets

Asset types

> Data access 1

<input type="checkbox"/>	Name	Last modified	
<input type="checkbox"/>	 itzcos-6910005ccg-fj156br8 Connection	Now Modified by you	⋮

Now let us repeat the process of visualizing the data, but this time via connections. Please create a new asset for the data refinery selecting your uploaded data.



Once established create a visualization and view the data profile as before to perform basic EDA.

Prepare and visualize data

Define the details to create a Data Refinery flow asset and open it in the Data Refinery tool.

Select data

Selected asset



dataset_loan_example.csv



Define details

Name

visualization

Description (optional)

Enter a description

Cancel

Back

Create

IBM watsonx

2695603 - itz-watsonx-13 Dallas KL

Projects / Kandana's sandbox / Data Refinery

Steps (1)

Data source

itzcos-6910005ccg-fj156br8/ibm-enablement-bucket-1/dataset_loan_example.csv

1. Convert column type

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Auto-generated

Use a code template to add a step

	Data	Profile	Visualizations				
	Loan_ID String	Gender String	Married String	Dependents String	Education String	Self_Employed String	Appr Inte
1	LP001002	Male	No	0	Graduate	No	584
2	LP001003	Male	Yes	1	Graduate	No	458
3	LP001005	Male	Yes	0	Graduate	Yes	300
4	LP001006	Male	Yes	0	Not Graduate	No	258
5	LP001008	Male	No	0	Graduate	No	600
6	LP001011	Male	Yes	2	Graduate	Yes	541
7	LP001013	Male	Yes	0	Not Graduate	No	233
8	LP001014	Male	Yes	3+	Graduate	No	303
9	LP001018	Male	Yes	2	Graduate	No	400
10	LP001020	Male	Yes	1	Graduate	No	128
11	LP001024	Male	Yes	2	Graduate	No	320
12	LP001027	Male	Yes	2	Graduate	--	250
13	LP001028	Male	Yes	2	Graduate	No	307
14	LP001029	Male	No	0	Graduate	No	185
15	LP001030	Male	Yes	2	Graduate	No	129
16	LP001032	Male	No	0	Graduate	No	495
17	LP001034	Male	No	1	Not Graduate	No	359
18	LP001036	Female	No	0	Graduate	No	351
19	LP001038	Male	Yes	0	Not Graduate	No	488
20	LP001041	Male	Yes	0	Graduate	--	260

Configure

Viewing: 614 rows, 13 columns

Full data set: 614 rows, 13 columns

About this asset

Name

visualization

Data Refinery flow

Description

What is the purpose of this Data Refinery flow?

Asset details

Steps: 1

Associated assets

Source: itzcos-6910005ccg-fj156br...

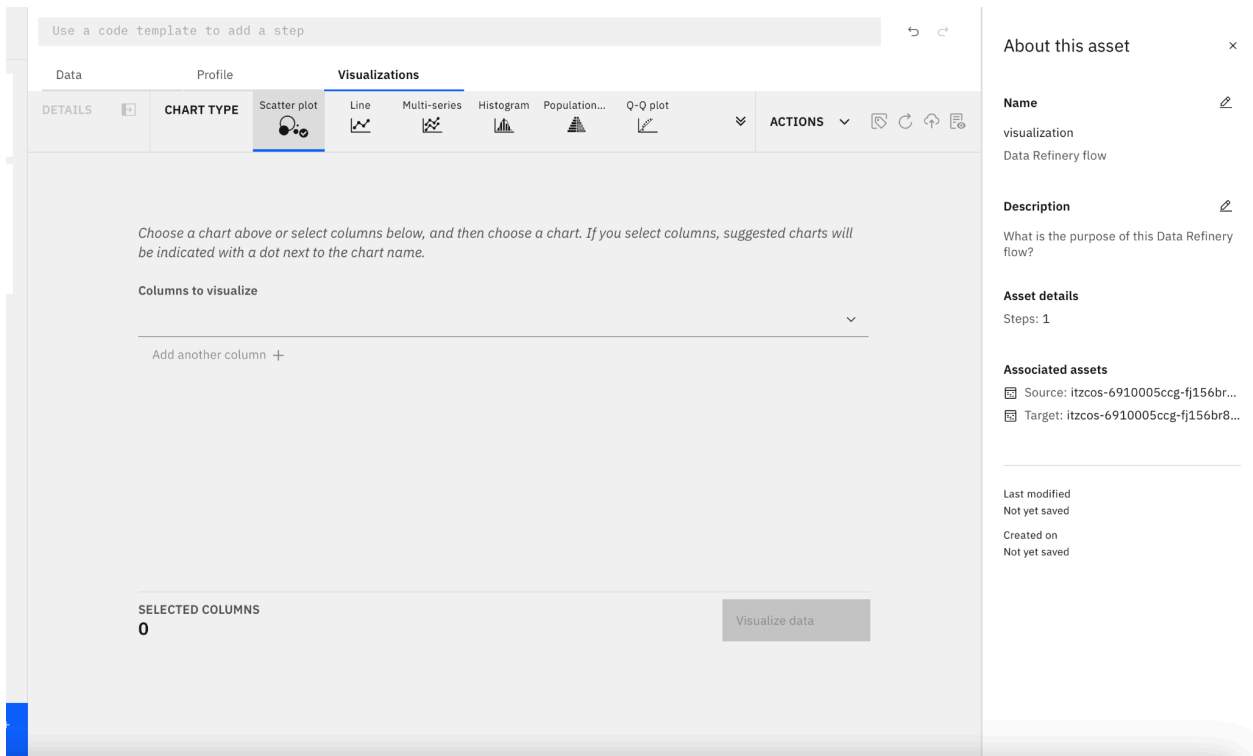
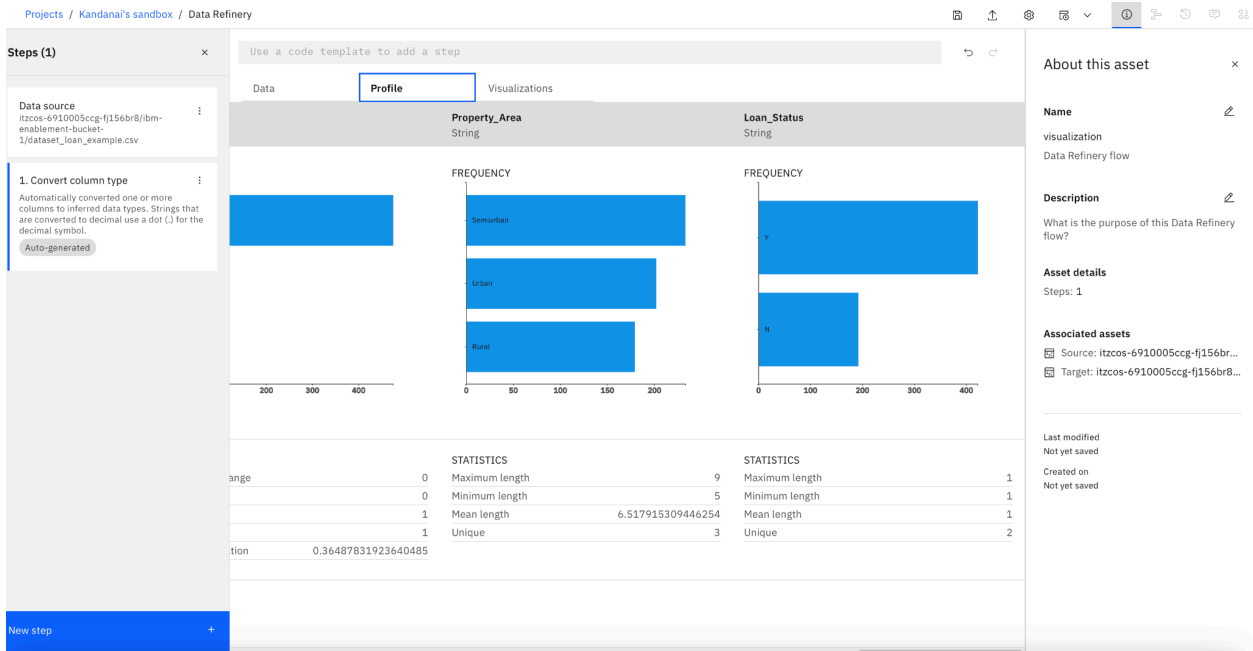
Target: itzcos-6910005ccg-fj156br8...

Last modified

Not yet saved

Created on

Not yet saved



Lastly, as we have done before, we can import the data frame through connections. And import the data frame in from IBM Cloud Object Storage.

The screenshot displays a Jupyter Notebook environment with a 'Select data from project' dialog box open. The dialog shows a search for 'dataset_loan_example.csv' within a connection named 'itzcos-6910005ccg-fj156br8'. The selected asset details show it is a file in the 'ibm-enablement-bucket-1' bucket, with a path of '/ibm-enablement-bucket-1/dataset_loan_example.csv'. The fields table lists 13 columns: Loan_ID, Gender, Married, Dependents, Education, Self_Employed, ApplicantIncome, CoapplicantIncome, LoanAmount, Loan_Amount_Term, Credit_History, Property_Area, and Loan_Status.

Below the dialog, the notebook code cell shows the import logic using the boto3 client to retrieve the CSV file from IBM Cloud Object Storage and load it into a pandas DataFrame.

```
2):  
import types  
import pandas as pd  
import ibm_boto3  
from botocore.client import Config  
  
def __iter__(self): return 0  
  
## @hidden_cell  
## The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.  
## You might want to remove those credentials before you share your notebook.  
  
itzcos_6910005ccg_fj156br8_client = ibm_boto3.client(  
    service_name='s3',  
    ibm_api_key_id='xL8tF83EnbC-FPFd2lvdrrz5rYYf09ZzP8o54_t-nZaj',  
    ibm_service_instance_id='crn:vl:bluemix:public:cloud-object-storage:global:a/98ff78eb326f477f8447f94142661697:05ef5cf0-21a9-4183-a307-793f612a8f01::',  
    ibm_auth_endpoint='https://iam.cloud.ibm.com/identity/token',  
    config=Config(signature_version='oauth'),  
    endpoint_url='https://s3.us-south.cloud-object-storage.appdomain.cloud'  
)  
  
bucket = 'ibm-enablement-bucket-1'  
object_key = 'dataset_loan_example.csv'  
  
body = itzcos_6910005ccg_fj156br8_client.get_object(Bucket=bucket, Key=object_key)['Body']  
# add missing __iter__ method, so pandas accepts body as file-like object  
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType(__iter__, body)  
  
df_0 = pd.read_csv(body)  
df_0.head(10)
```

The output of the code cell shows the first 10 rows of the DataFrame:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	Urban	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	Rural	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	Urban	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	Urban	
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	Urban	
5	LP001011	Male	Yes	2	Graduate	Yes	5417	4196.0	267.0	360.0	1.0	Urban	

On the right side of the notebook, the 'Read data' sidebar is visible, showing the selected data source and the option to load it as a 'pandas DataFrame'.