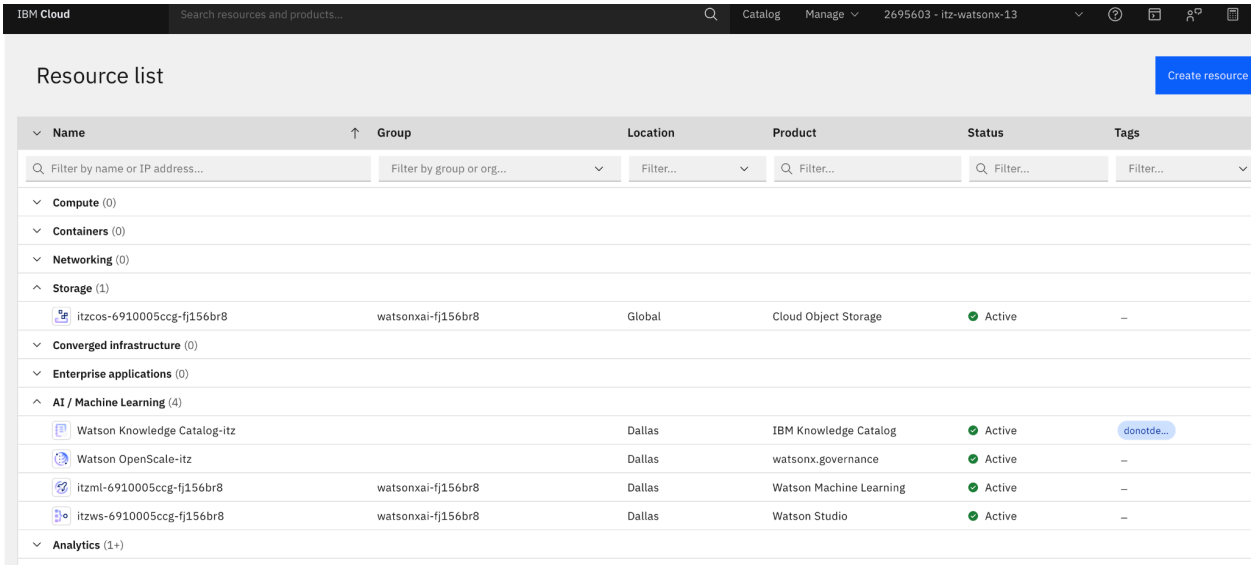
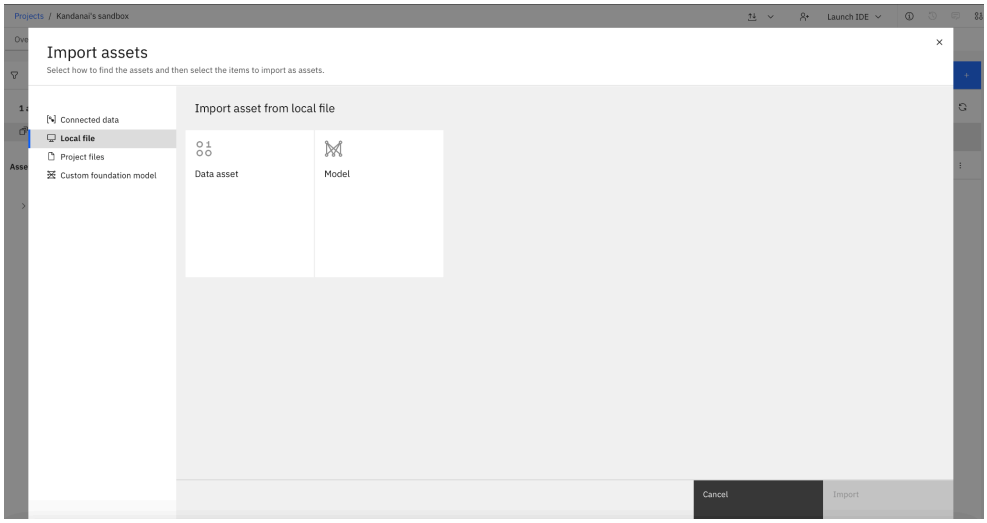
LAB 1 Data Visualization and Creating Connections

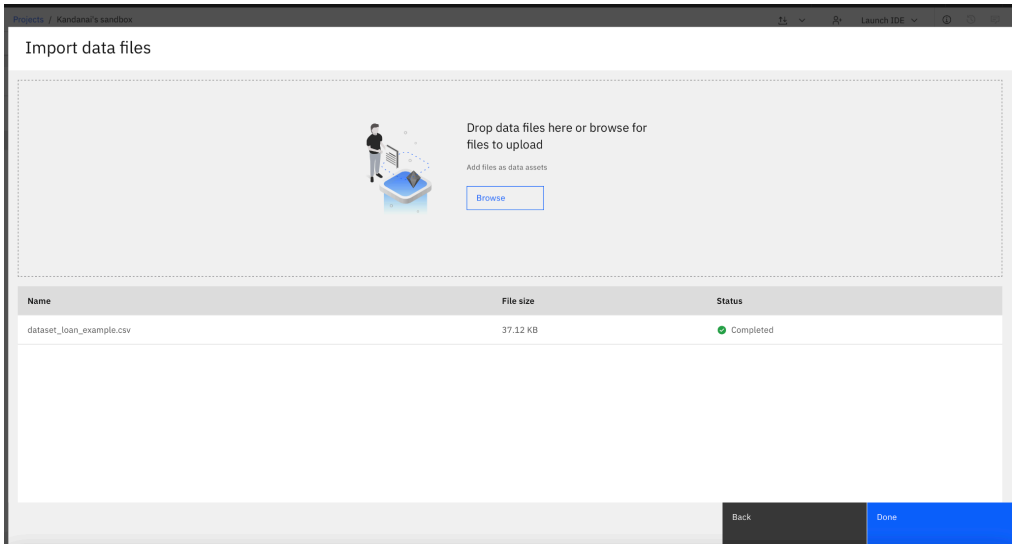
Ensure you have the sufficient resources required to do the lab (See Lab0)



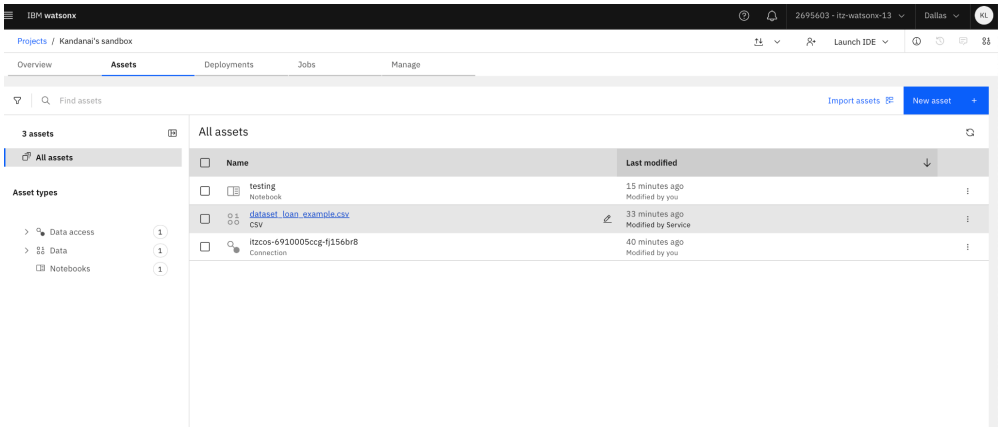
Firstly, locate your project and click on the import asset in the watsonx. Here our objective is to import a local CSV file as an asset to your sandbox project. Click on Data asset below:



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.

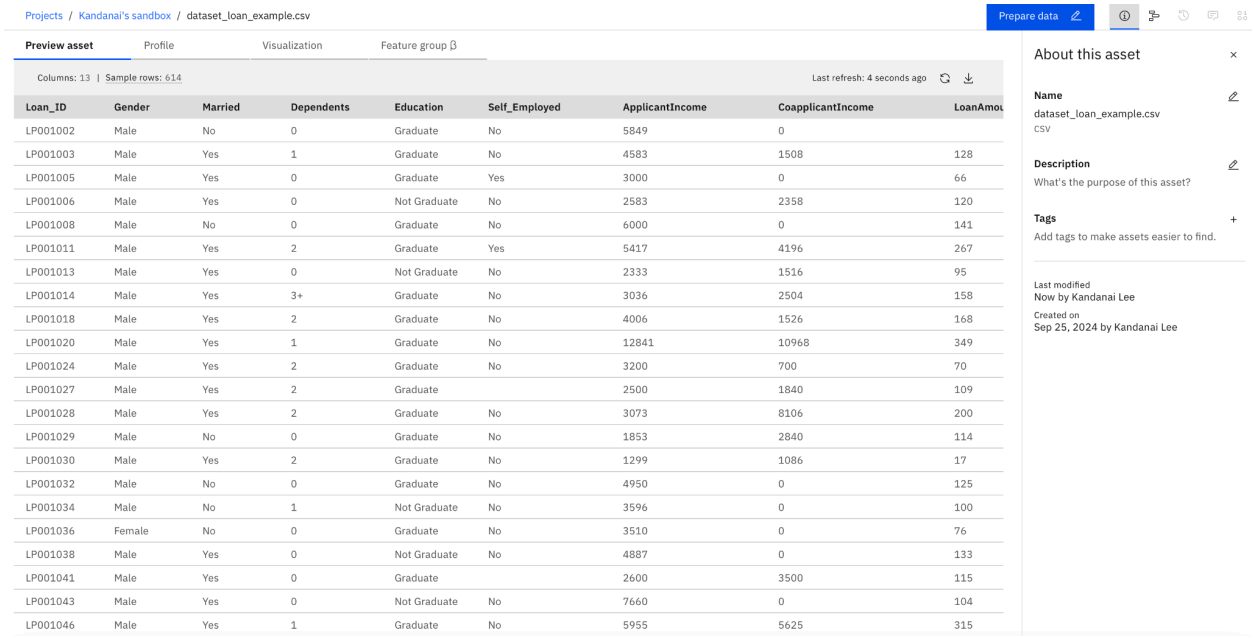


If you click on your dataset dataset_loan_example_v2.csv



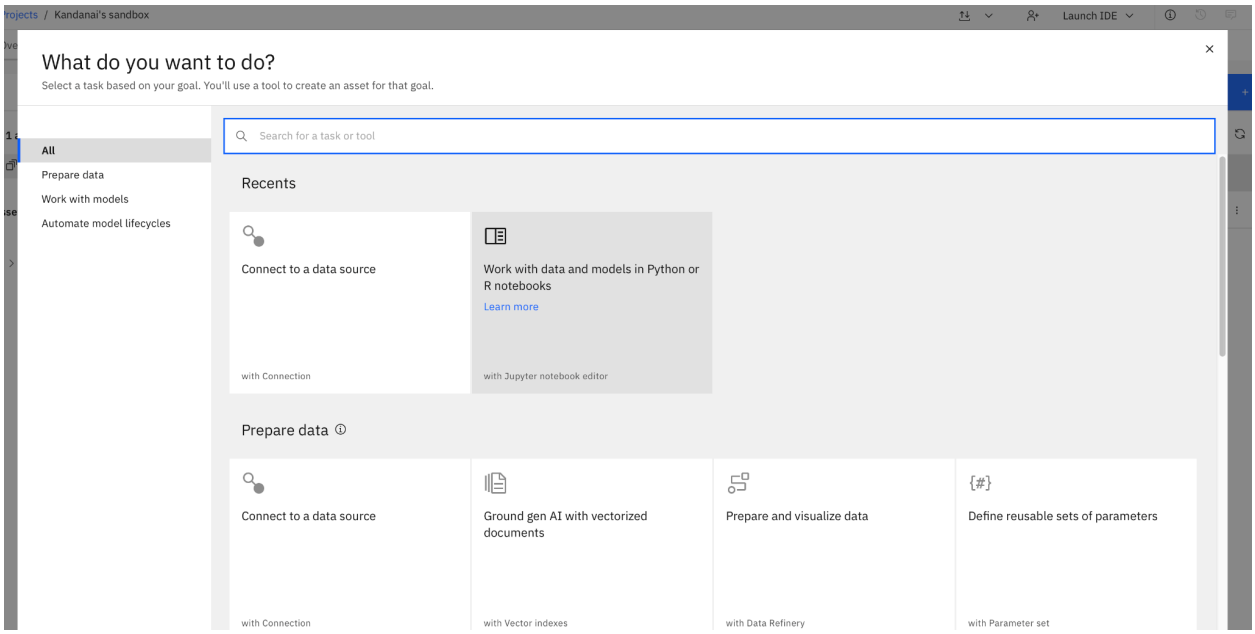
You will be able to

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

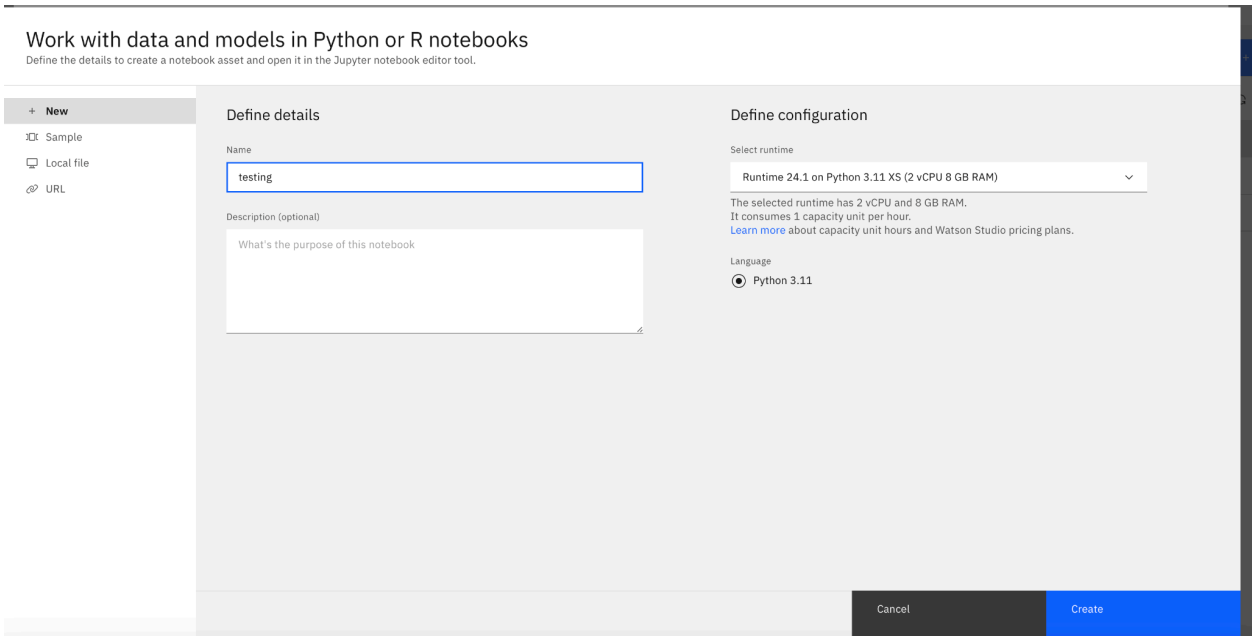


Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmou
LP001002	Male	No	0	Graduate	No	5849	0	
LP001003	Male	Yes	1	Graduate	No	4583	1508	128
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120
LP001008	Male	No	0	Graduate	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

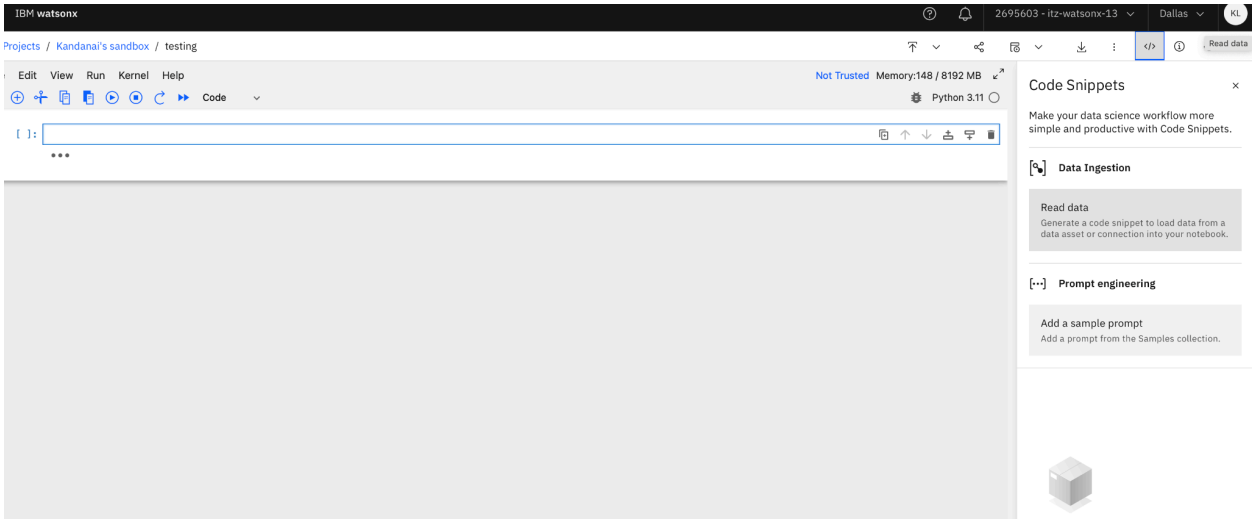
After visualizing the data let's try importing the data into watson studio (Jupyter Notebook)

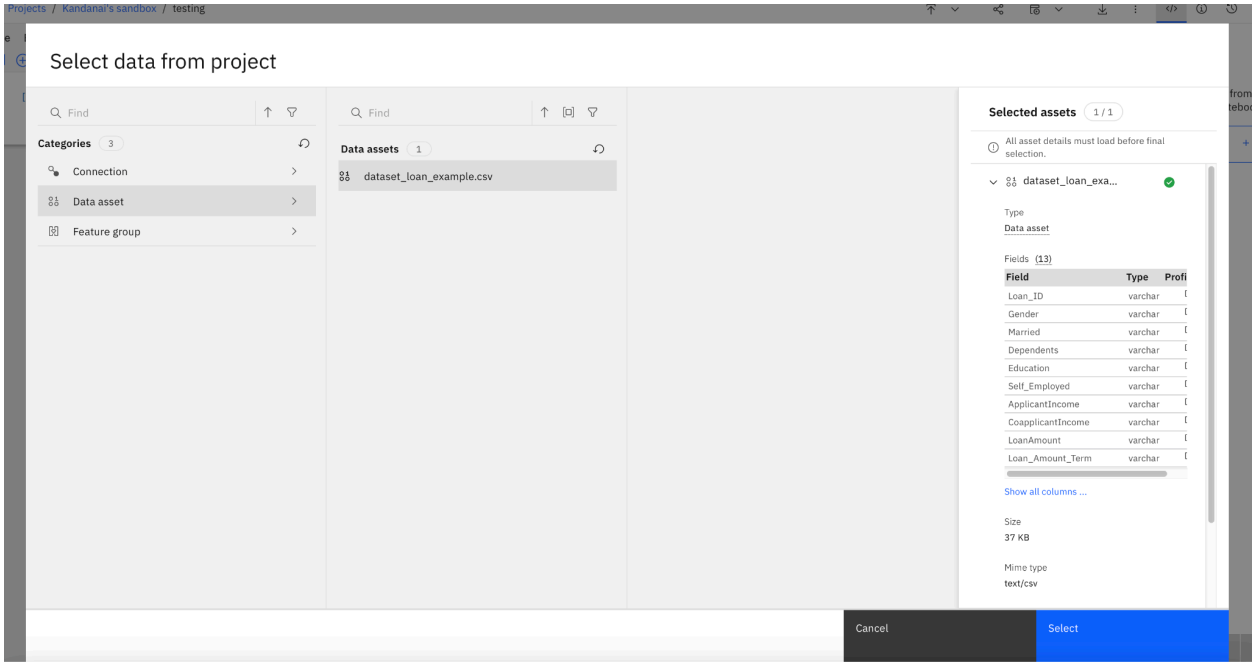


Enter any name to create the notebook and select the Runtime to be “Runtume 24.1 on Python 3.11 XS (2 VCPU 8 GB RAM)

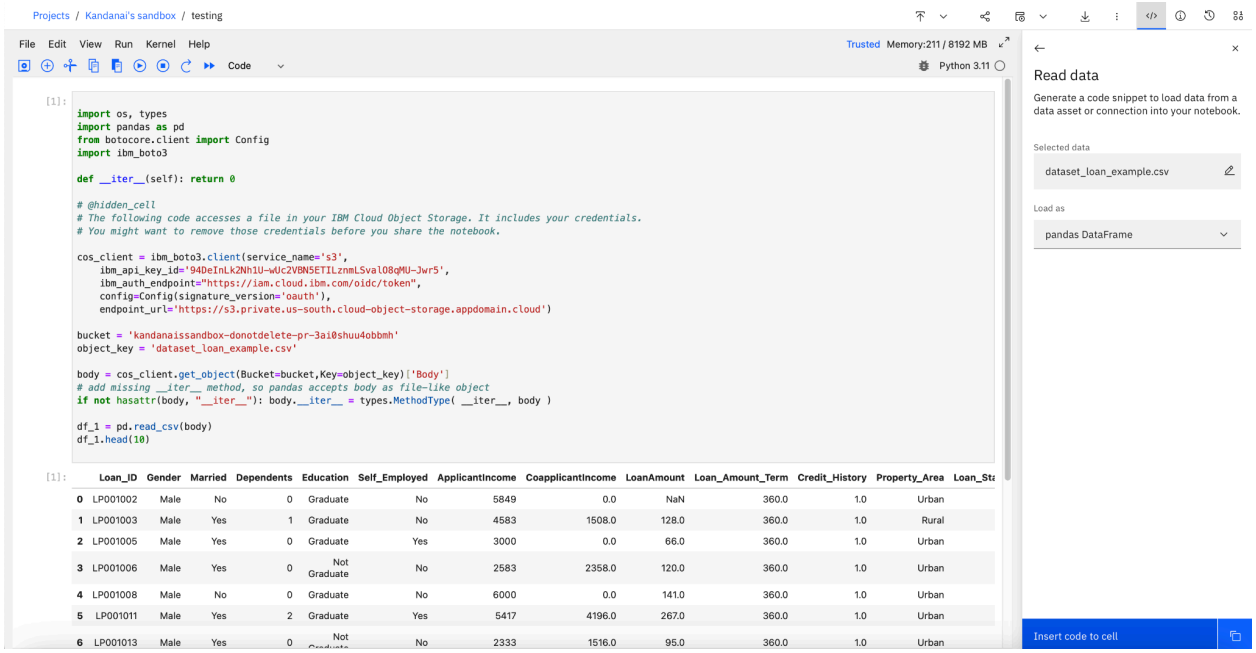


Here you will see the jupyter notebook being provisioned. Click on the </> icon and read data. Here select the csv file in the data asset you have just uploaded.



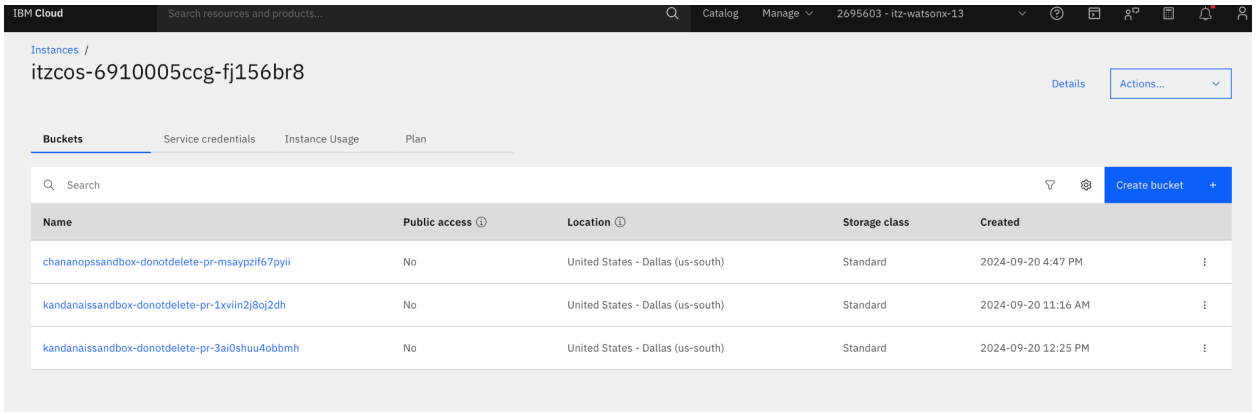


Watson studio will automatically generate code to create connections with the data asset and write code to create pandas dataframe. Click insert to cell and run the cell.

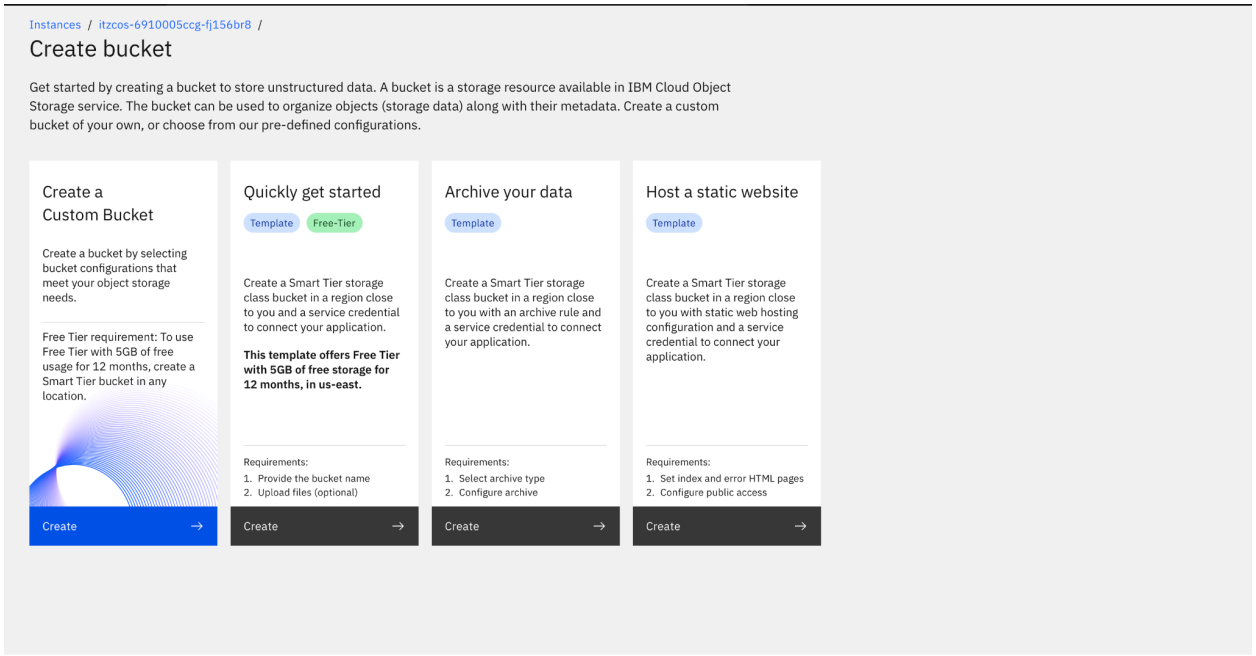


LAB 1 Data Visualization and Creating Connections (Optional)

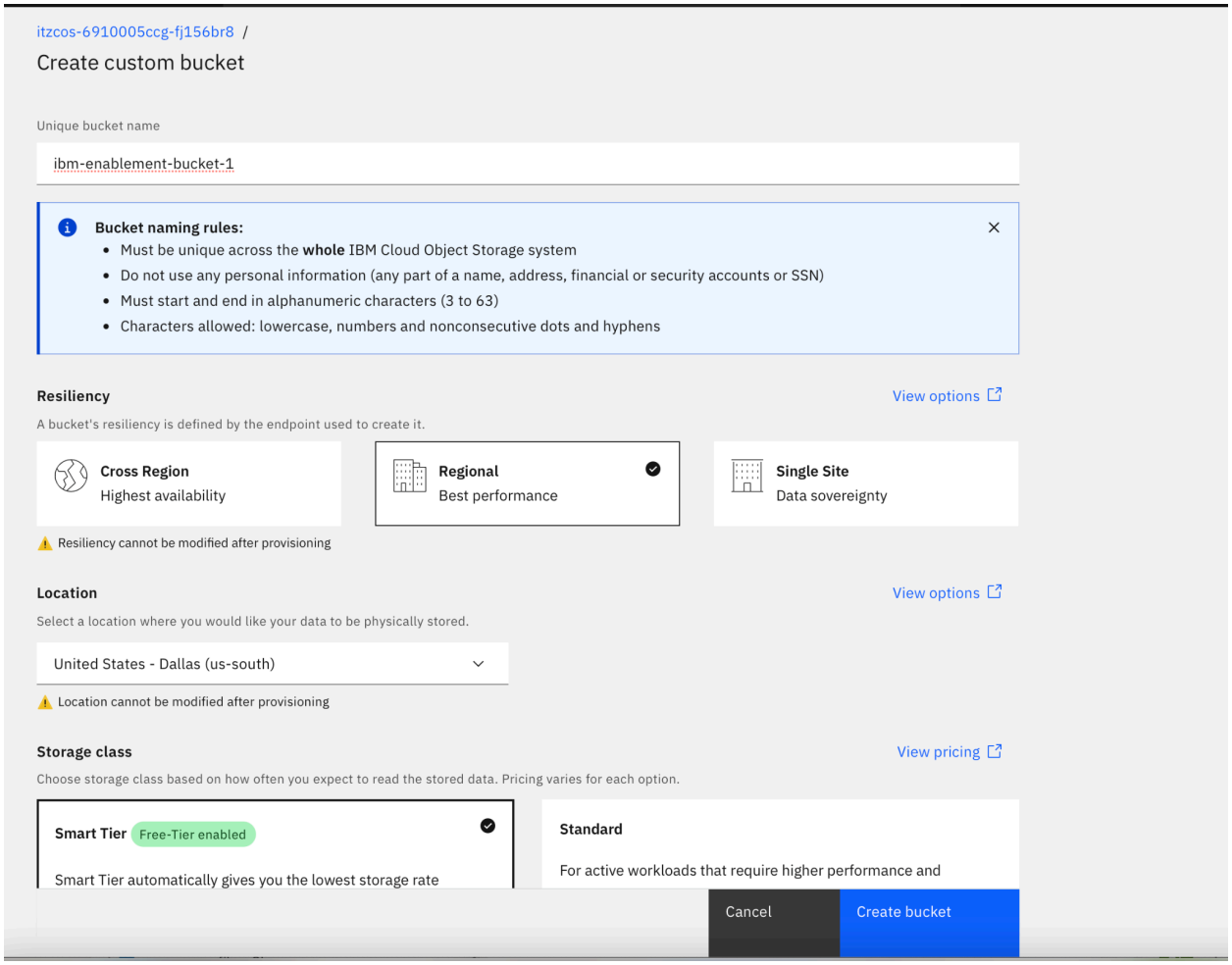
Optionally, instead of uploading the data directly into the project, we can create a cloud storage and create connections to that cloud storage. Here we will create an IBM Cloud Object Storage and repeat the steps before.



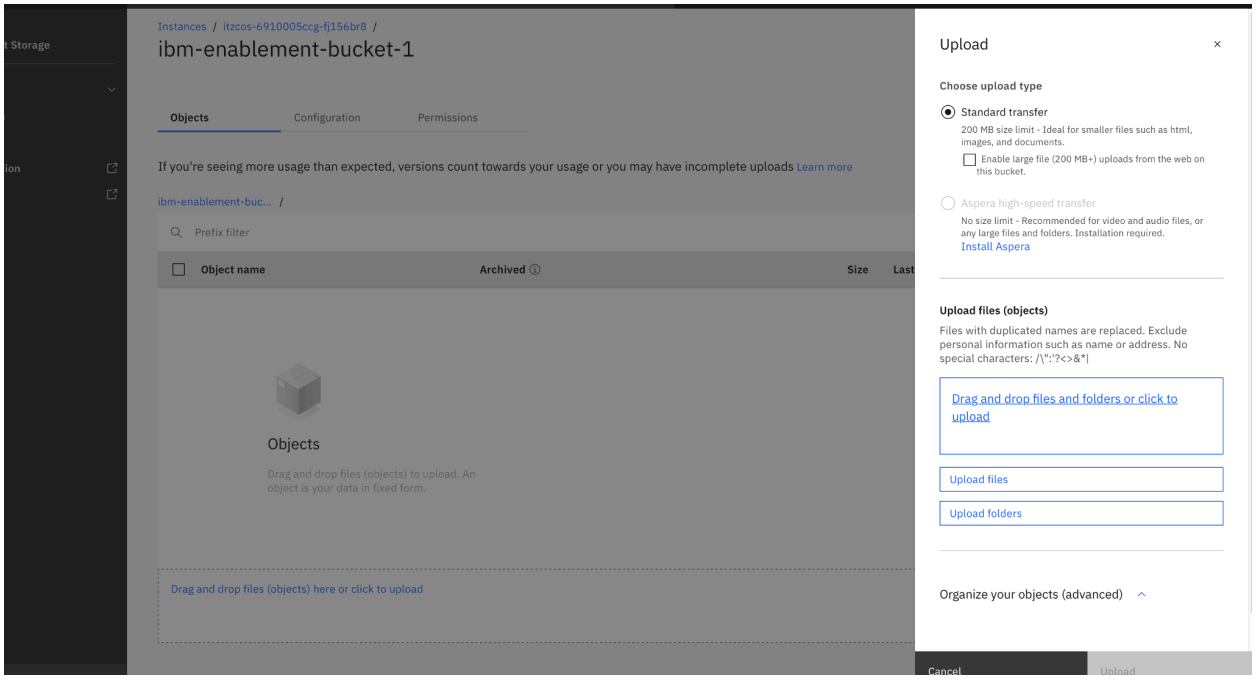
Create a custom bucket

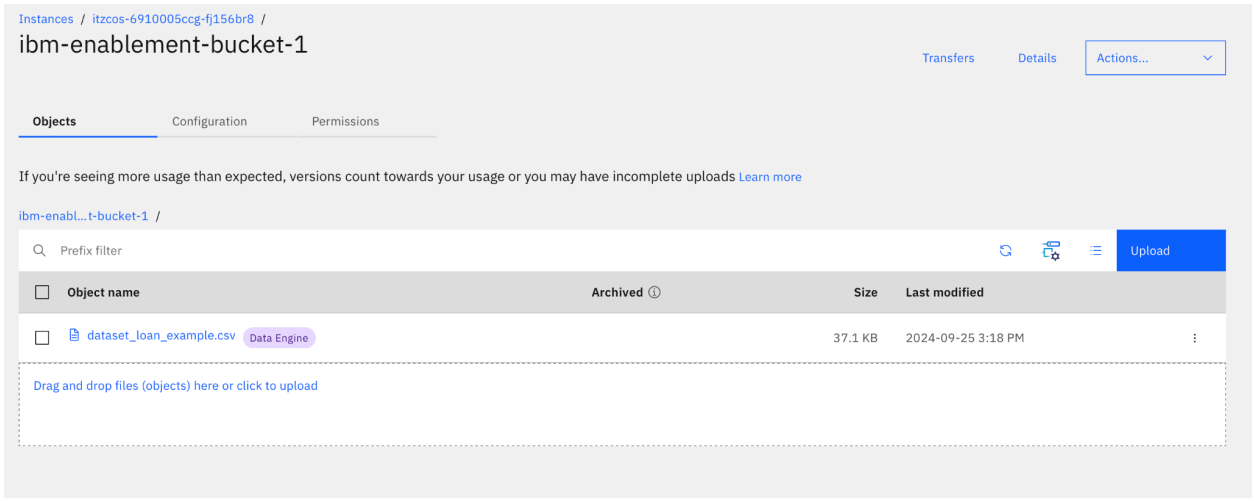


Give an a name to your custom bucket and ensure it is selected in the United States region (us-south)

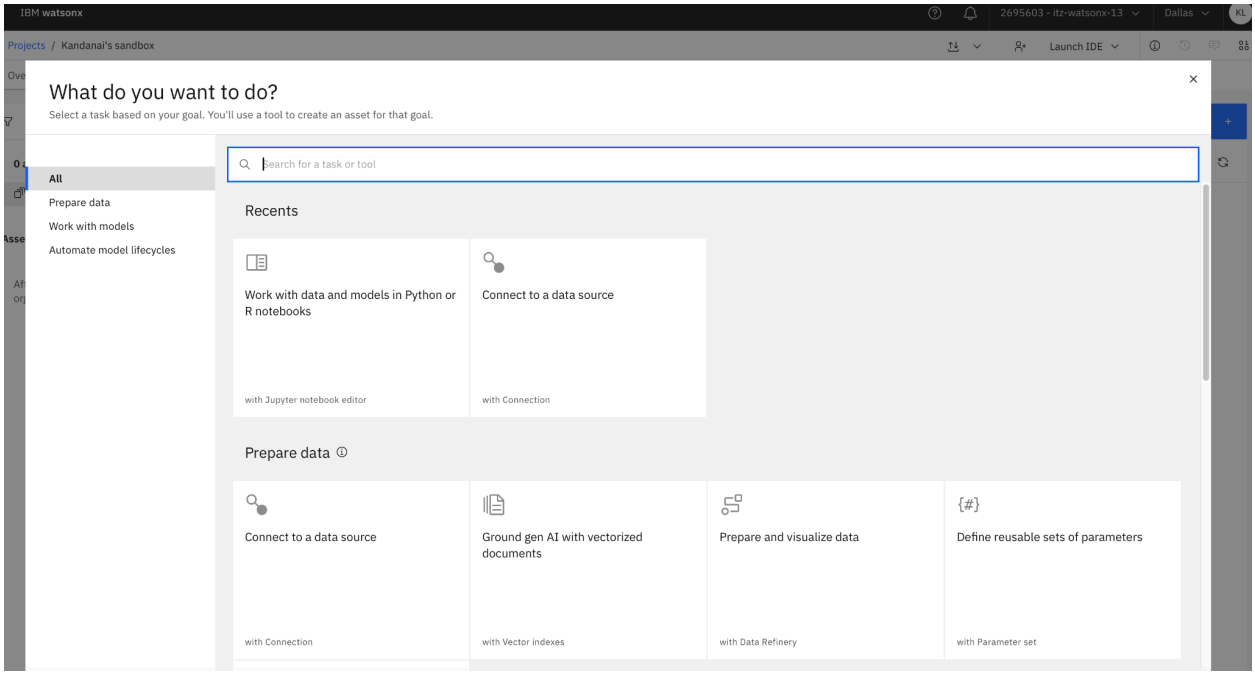


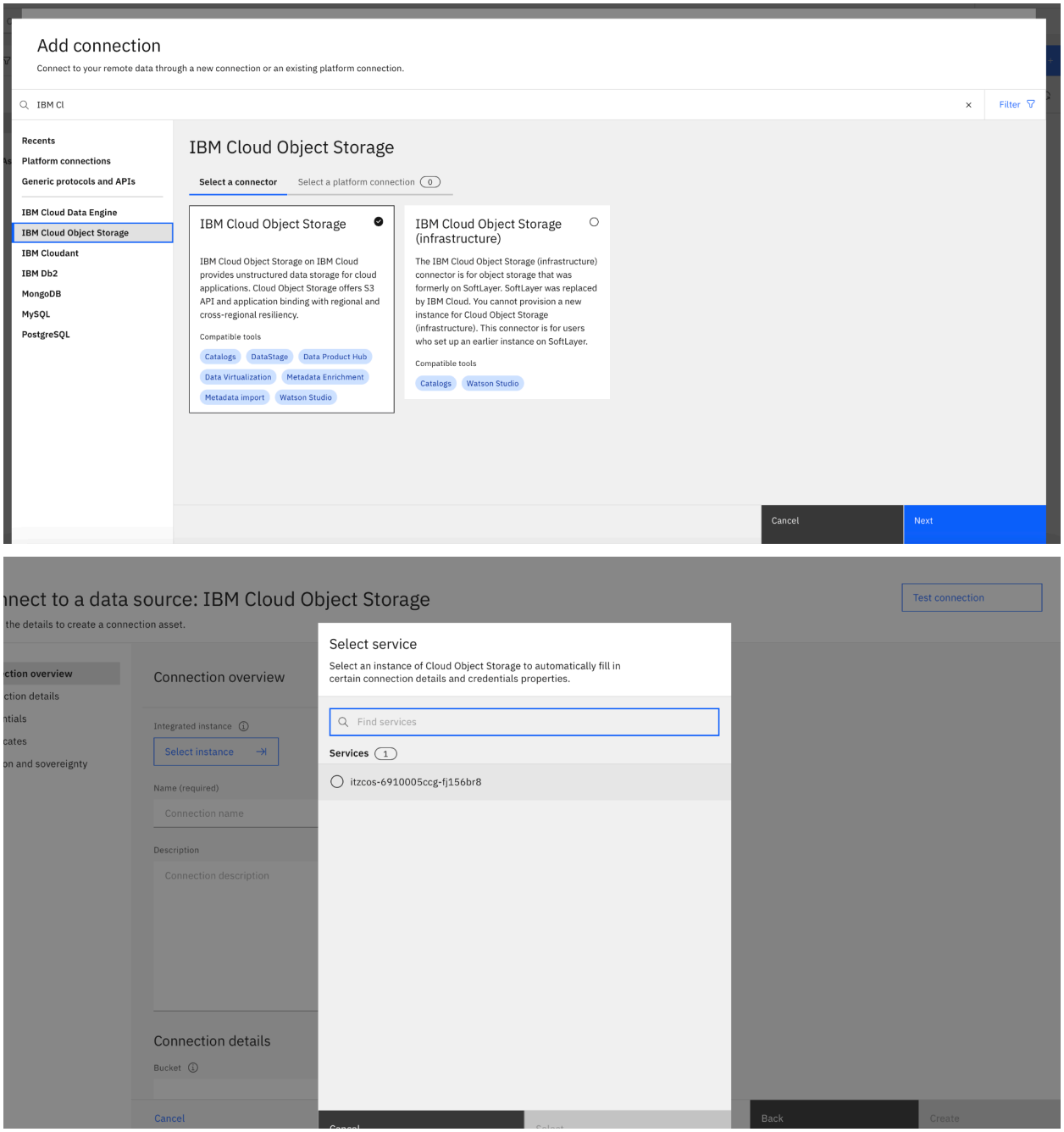
Upload the provided csv in the repo to the cloud storage bucket.



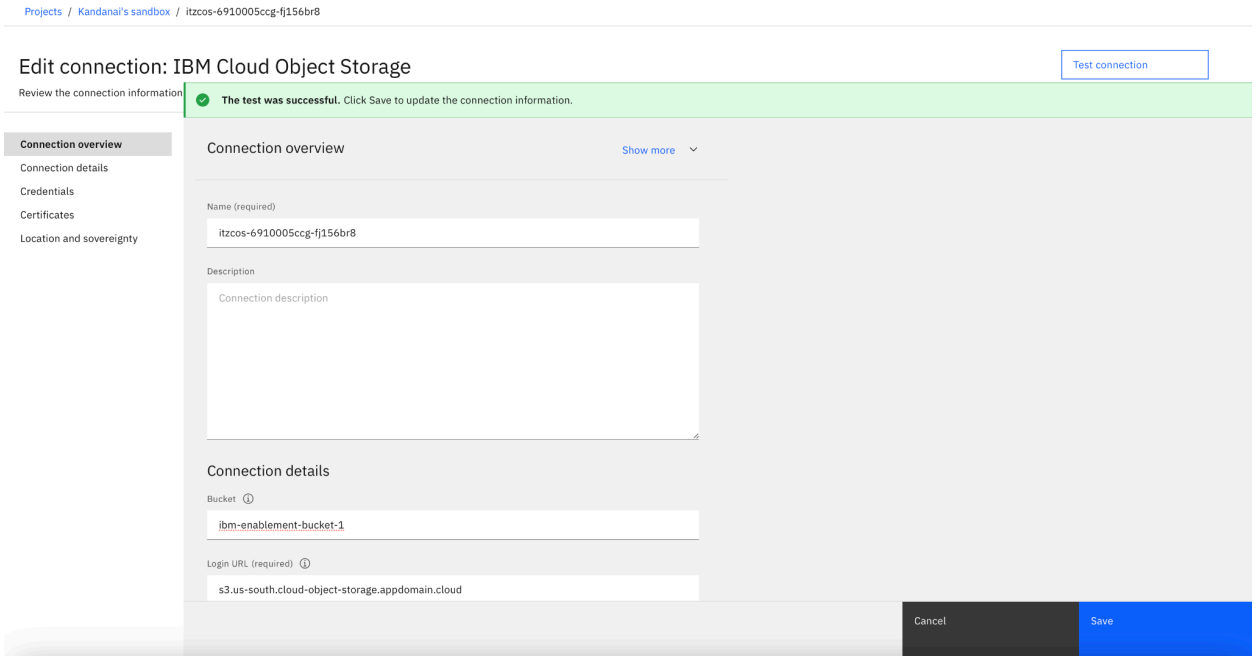


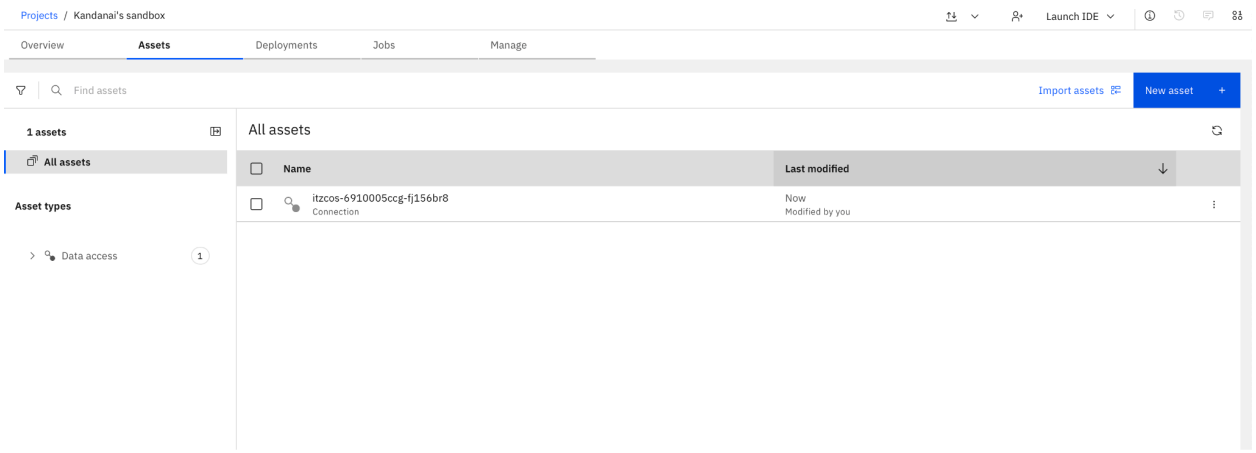
Next, return to the watsonx.ai interface and create a connection. Please select the datasource to be IBM Cloud Object Storage.



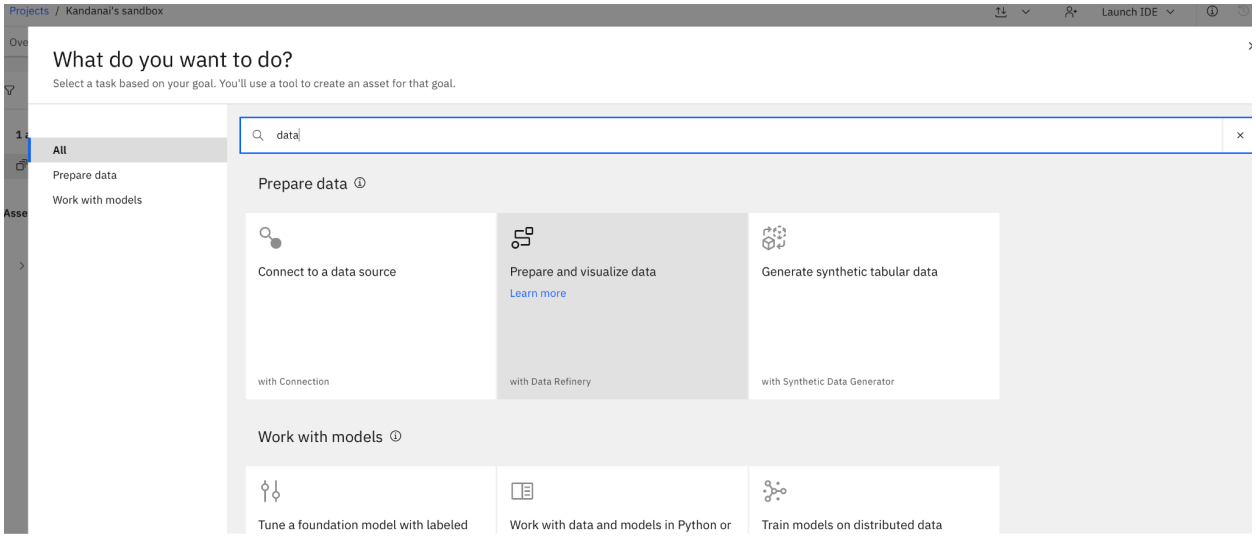


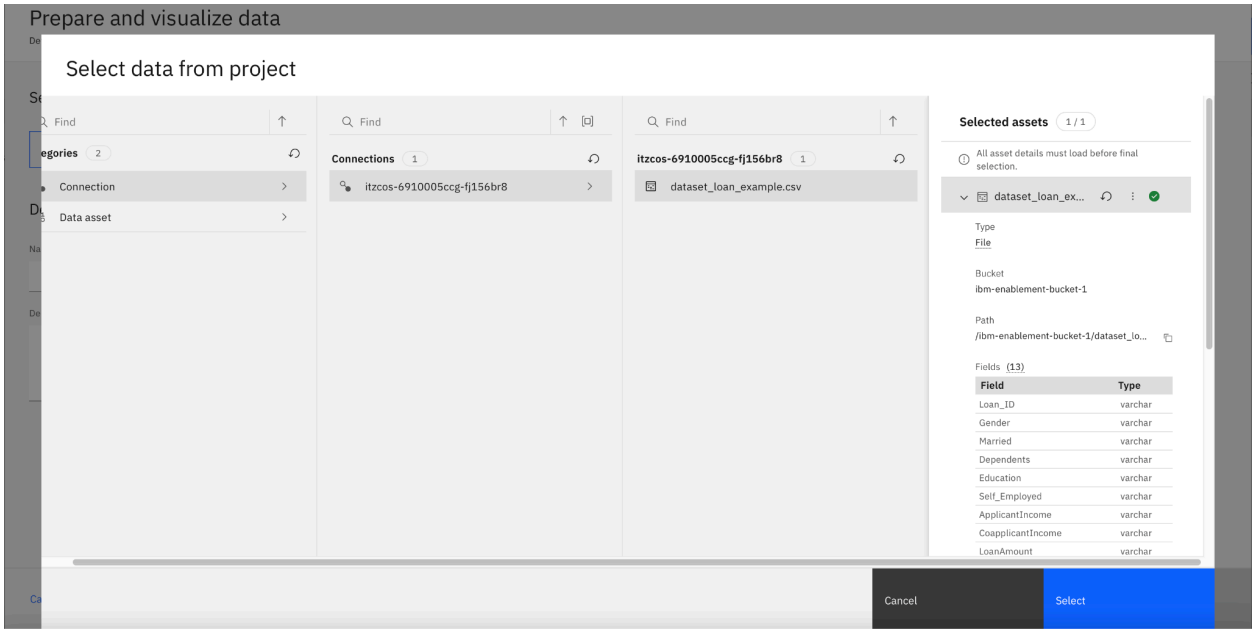
Choose the corresponding connection and enter your bucket name. Please Test your connection. If the test is successful please save to create your connection.



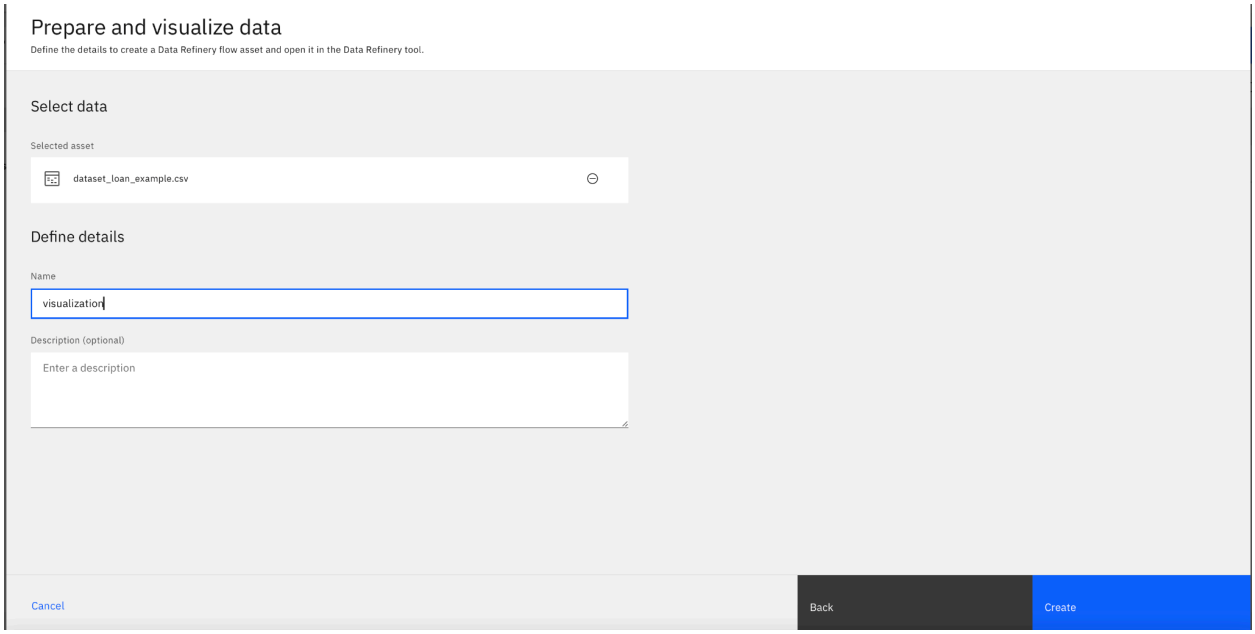


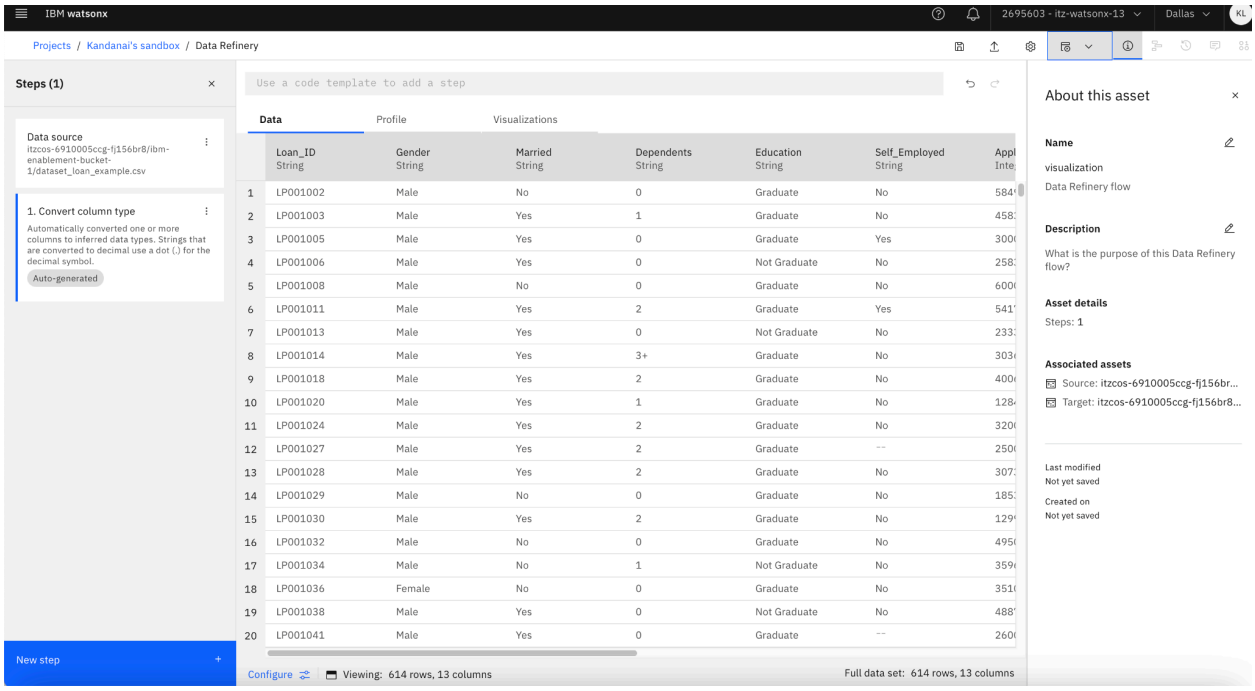
Now let us repeat the process of visualizing the data, but this time via connections. Please create a new asset for the data refinery selecting your uploaded data.

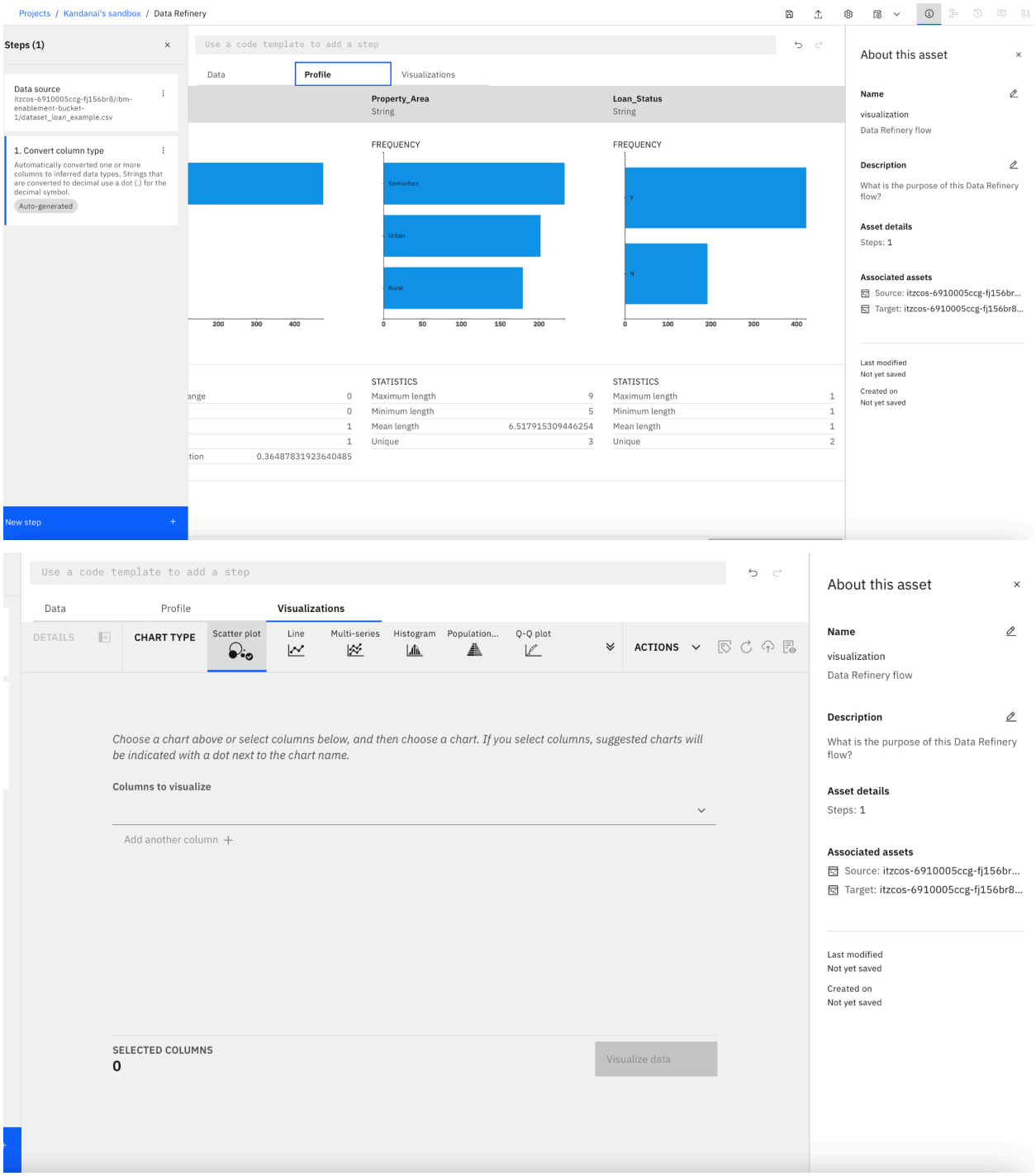




Once established create a visualization and view the data profile as before to perform basic EDA.


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 / Kandanai'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
New step
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
1 LP001002 Male No 0 Graduate No
2 LP001003 Male Yes 1 Graduate No
3 LP001005 Male Yes 0 Graduate Yes
4 LP001006 Male Yes 0 Not Graduate No
5 LP001008 Male No 0 Graduate No
6 LP001011 Male Yes 2 Graduate Yes
7 LP001013 Male Yes 0 Not Graduate No
8 LP001014 Male Yes 3+ Graduate No
9 LP001018 Male Yes 2 Graduate No
10 LP001020 Male Yes 1 Graduate No
11 LP001024 Male Yes 2 Graduate No
12 LP001027 Male Yes 2 Graduate
13 LP001028 Male Yes 2 Graduate No
14 LP001029 Male No 0 Graduate No
15 LP001030 Male Yes 2 Graduate No
16 LP001032 Male No 0 Graduate No
17 LP001034 Male No 1 Not Graduate No
18 LP001036 Female No 0 Graduate No
19 LP001038 Male Yes 0 Not Graduate No
20 LP001041 Male Yes 0 Graduate
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


Projects / Kandanai'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
New step
Use a code template to add a step
Data
Profile
Visualizations
Property_Area
String
Loan_Status
String
FREQUENCY
Semiurban
Urban
Rural
Y
N
STATISTICS
Maximum length 9
Minimum length 5
Mean length 6.517915309446254
Unique 3
Maximum length 1
Minimum length 1
Mean length 1
Unique 2
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
Use a code template to add a step
Data
Profile
Visualizations
DETAILS
CHART TYPE
Scatter plot
Line
Multi-series
Histogram
Population...
Q-Q plot
ACTIONS
Choose a chart above or select columns below, and then choose a chart. If you select columns, suggested charts will be indicated with a dot next to the chart name.
Columns to visualize
Add another column
SELECTED COLUMNS
0
Visualize data
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.