How to perform Data Science Analysis using Elyra?

Yiwen Li

About us

IBM – Center for Open Source Data and AI Technologies (CODAIT)

Yiwen Li
Data Scientist &
Developer Advocate

Yiwen.Li@ibm.com

github.com/yil532

linkedin.com/in/yiwenli

Overview

- Data Science Pipeline
- What is CODAIT?
- An overview of the Data Asset eXchange (DAX)
- A deep dive into Elyra and its features
- Demo showcasing DAX & Elyra
- How to get involved

Data Science Process

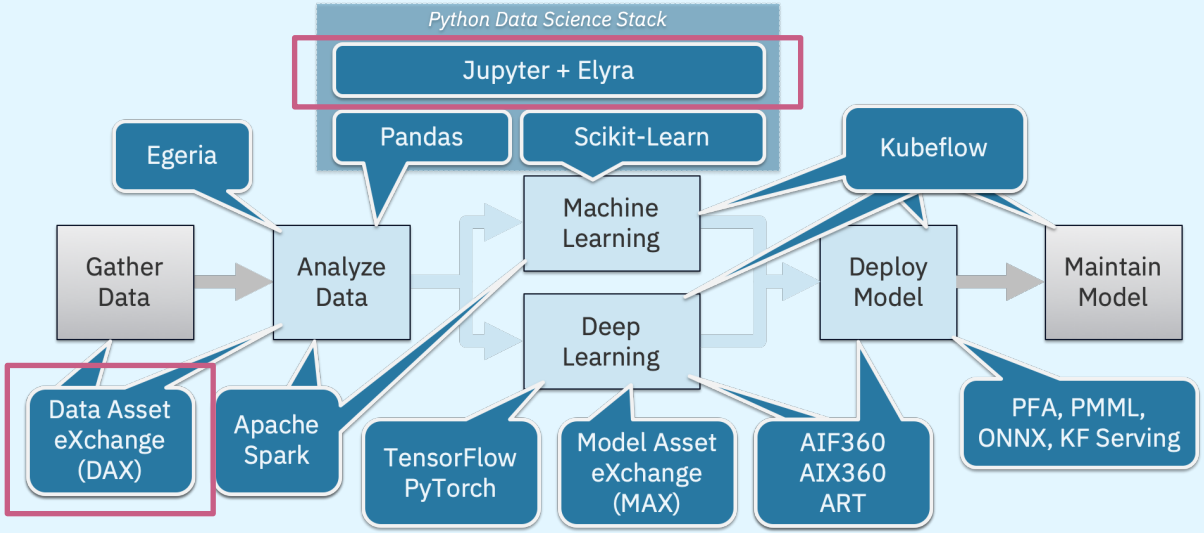
Center for Open Source Data & AI Technologies

CODAIT

Open Source @IBM

codait.org

- CODAIT aims to make AI solutions dramatically easier to create, deploy, and manage in the enterprise.
- We contribute to and advocate for the open-source technologies that are foundational to IBM's AI offerings.
- 30+ open-source developers!



bit.ly/belpy-dax

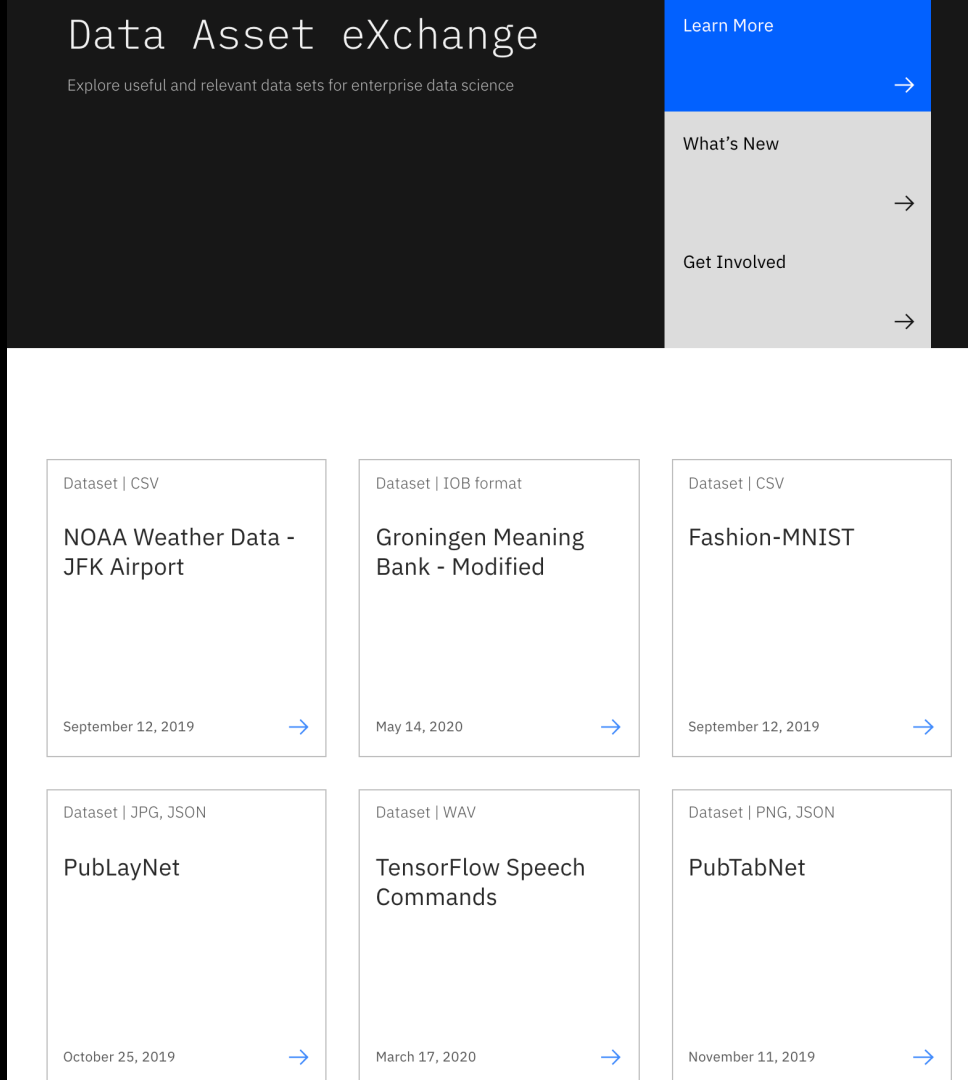
Data Asset eXchange

Data Asset Exchange offers high-quality datasets with clearly-defined open data licenses in standardized formats, according to IBM.

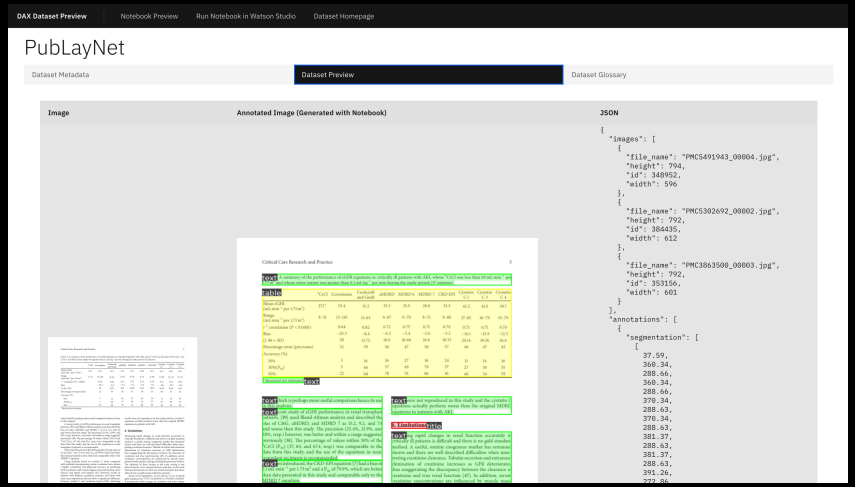
- Vetted data.
- Exclusive access to IBM Research datasets that have been used in creating popular AI products like Debater System, Entity Recognition, and so on.
- Datasets with open data licenses for both business applications and advancing core science.
- Packaged with tutorials that shows how to read and analyze data. As well as, train machine or deep learning models on IBM Cloud using IBM Cloud AI services as well as multi-cloud AI open-sourced tools.

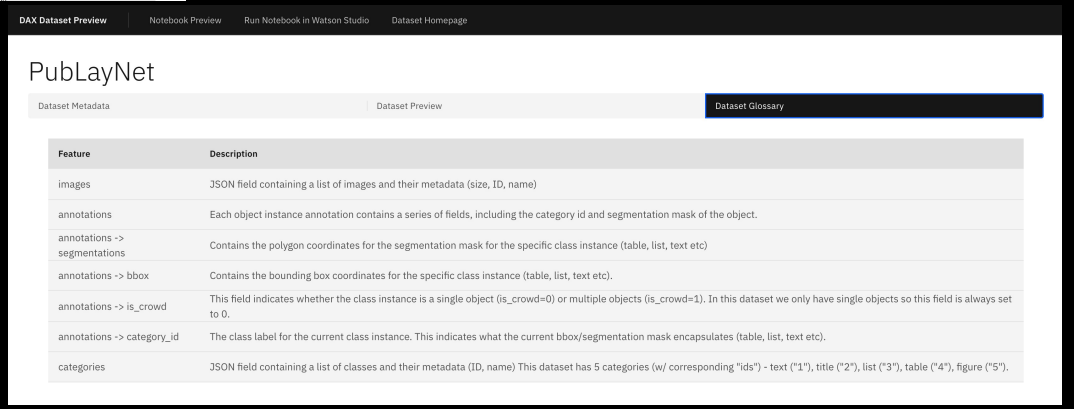
ibm.biz/data-exchange

bit.ly/belpy-dax



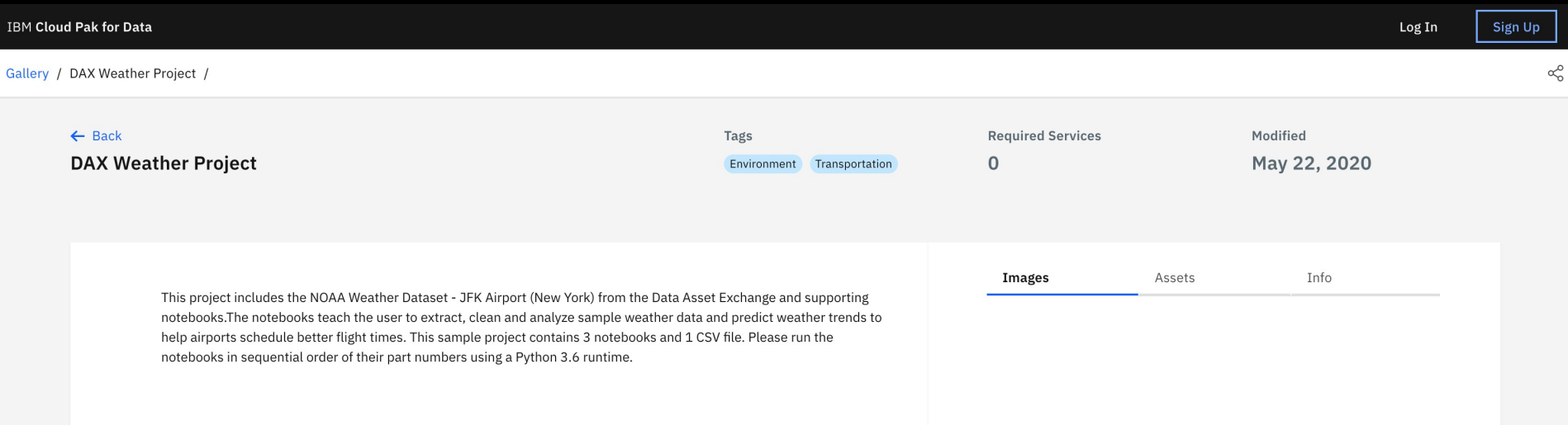
Data Preview and Data Glossary





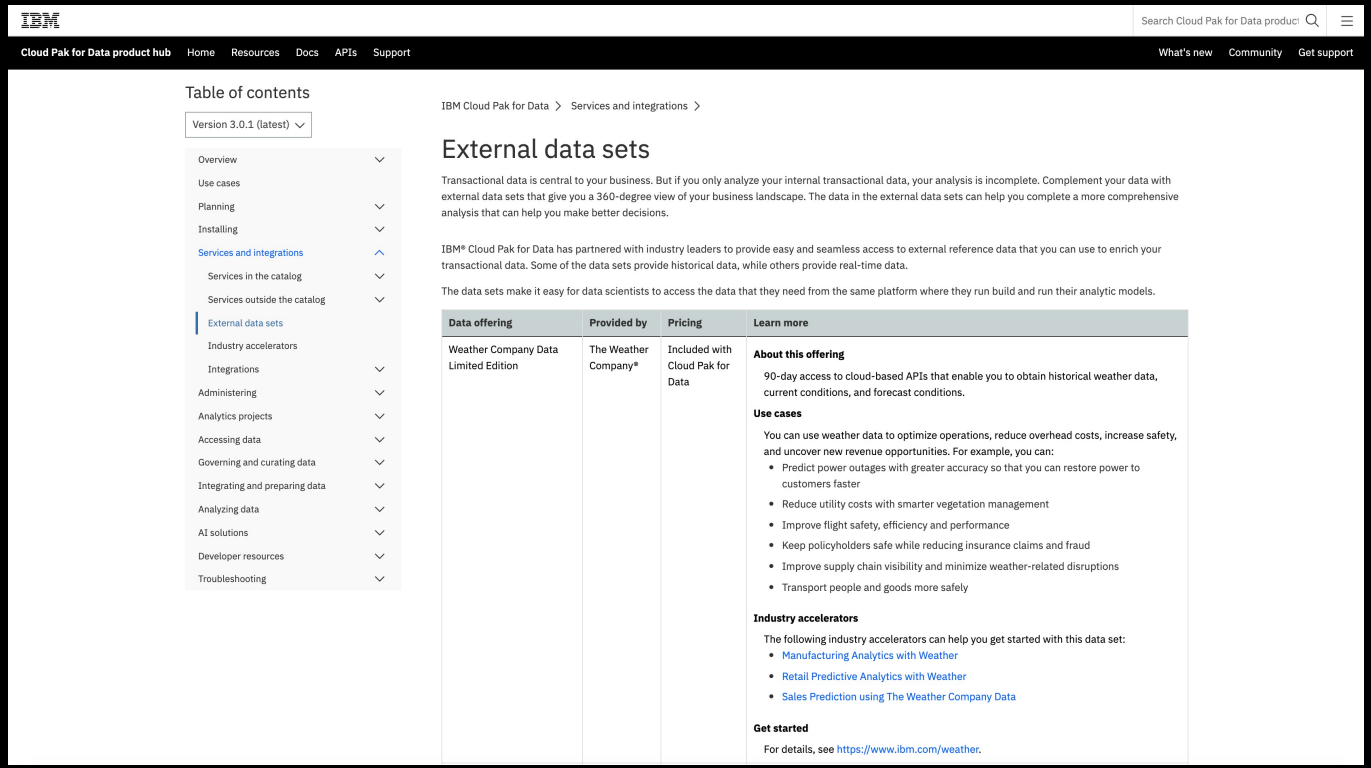
bit.ly/belpy-dax

Access notebook in Watson Studio



bit.ly/belpy-dax

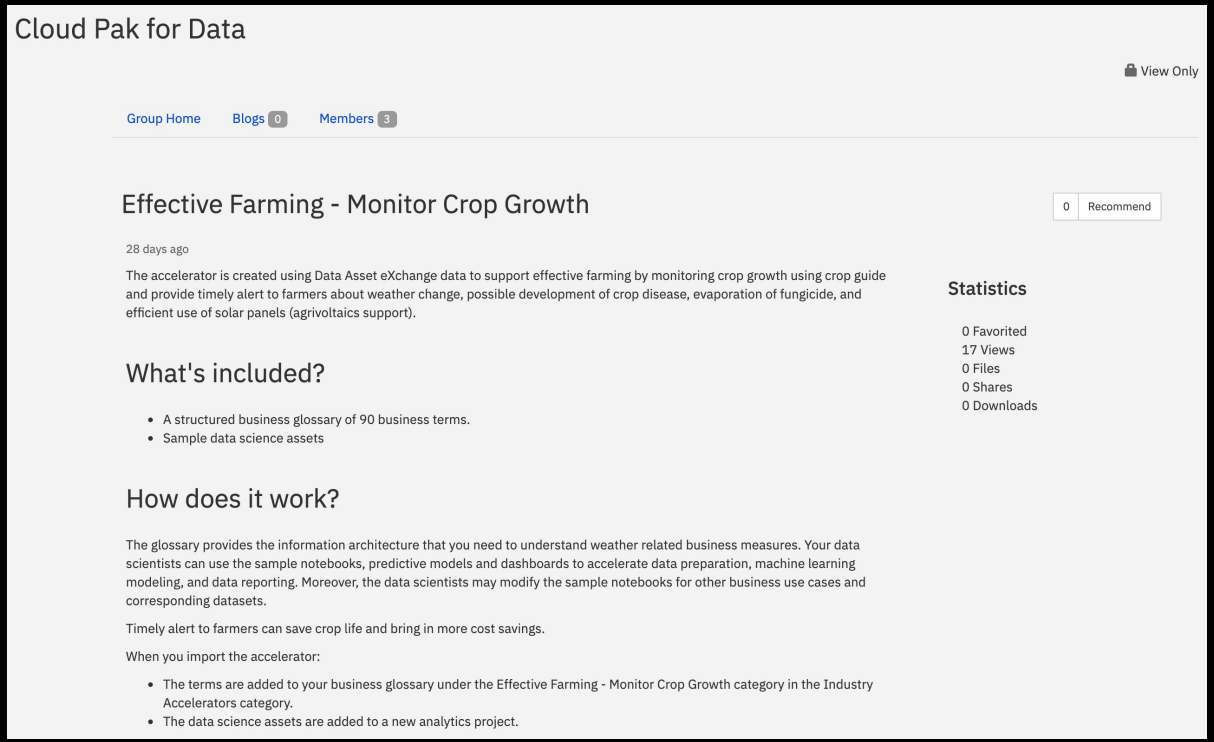
Access from Cloud Pak for Data



https://www.ibm.com/support/producthub/icpdata/docs/content/SSQNUZ_current/svc-nav/data-sets.html

bit.ly/belpy-dax

Industrial Accelerator - Cloud Pak for Data



https://community.ibm.com/community/user/cloudpakfordata/viewdocument/effective-farming-monitor-crop-gr

bit.ly/belpy-dax

What is Elyra?

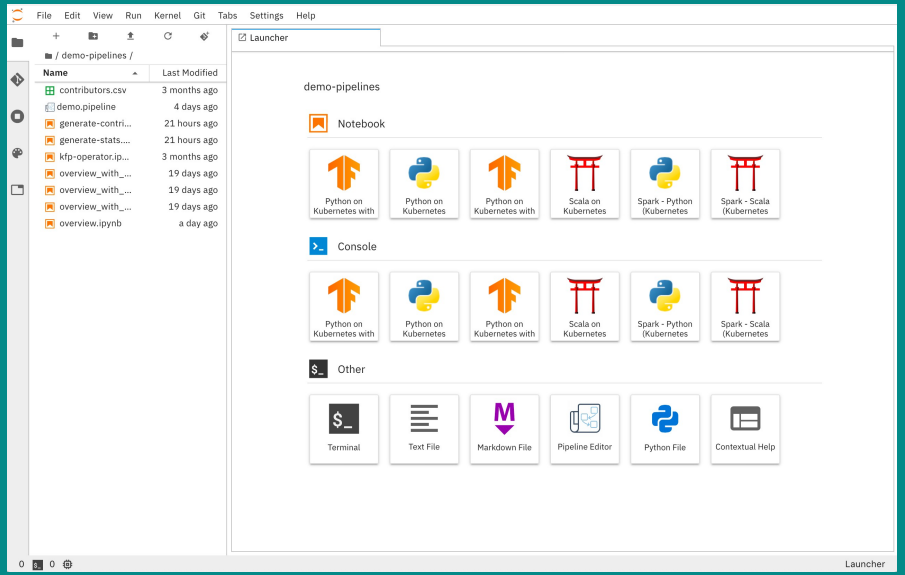
Elyra is a set of AI centric extensions to JupyterLab. It aims to help data scientists, machine learning engineers and AI developer's through the model development life cycle complexities.

Elyra on GitHub

https://github.com/elyra-ai/elyra

Elyra's Documentation

https://elyra.readthedocs.io/en/latest/



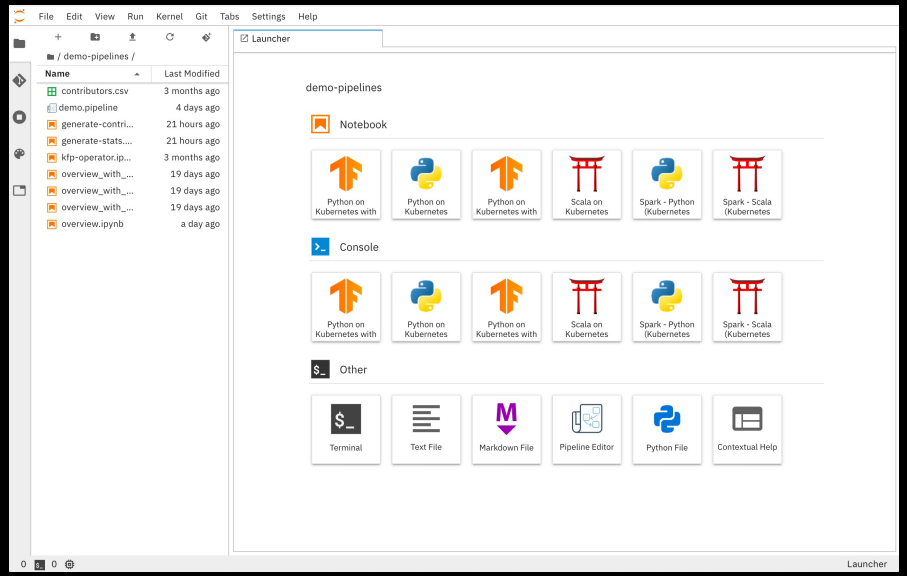
bit.ly/belpy-dax

What is Elyra?

Elyra is a set of
AI centric extensions for JupyterLab

Elyra was officially announced as an open source project by IBM on April 29th.

The name Elyra is a word play with one of the Jupyter moons “Elara” where we introduce the “y” from “Jupyter” to make it “Elyra”



bit.ly/belpy-dax

Elyra Core Features

Notebook Pipelines editor

Visual editor for building notebook-based AI pipelines, enabling the conversion of multiple notebooks into batch jobs or workflows.

Notebook as batch jobs

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

Code Snippets

Easy creation and insertion of reusable code snippets for the various languages

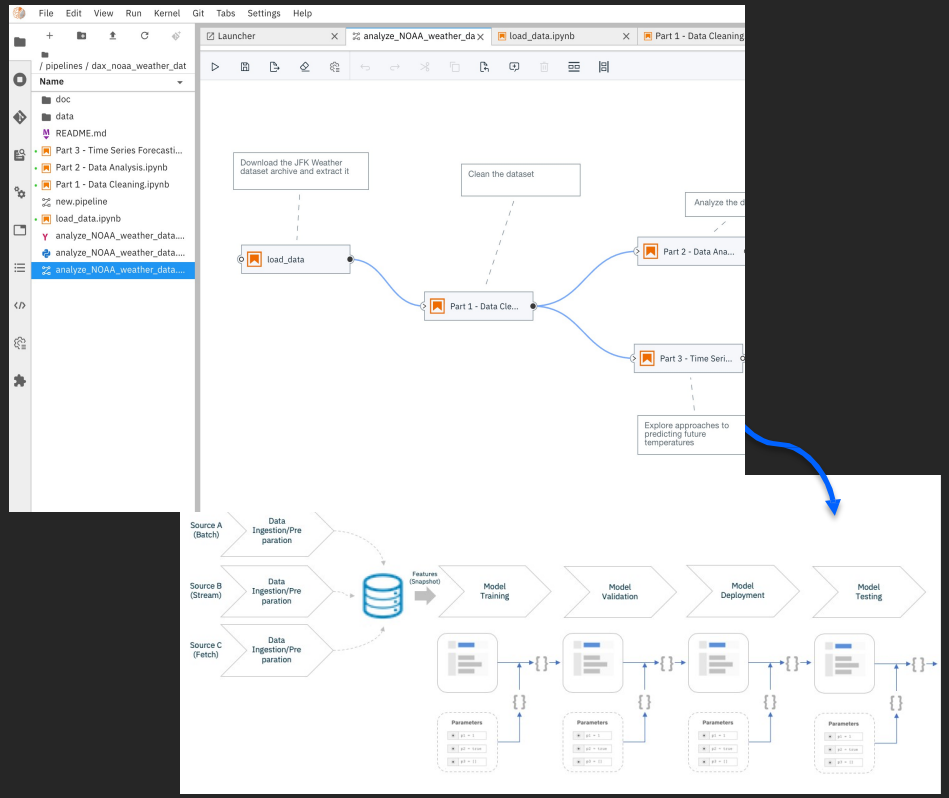
Git integration

Track project changes and share amongteammates

Python script execution

Edit and execute python scripts against local or cloud-based resources

Notebook Pipelines



bit.ly/belpy-dax

Elyra Core Features

Notebook Pipelines editor

Visual editor for building notebook-based AI pipelines, enabling the conversion of multiple notebooks into batch jobs or workflows.

Notebook as batch jobs

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

Code Snippets

Easy creation and insertion of reusable code snippets for the various languages

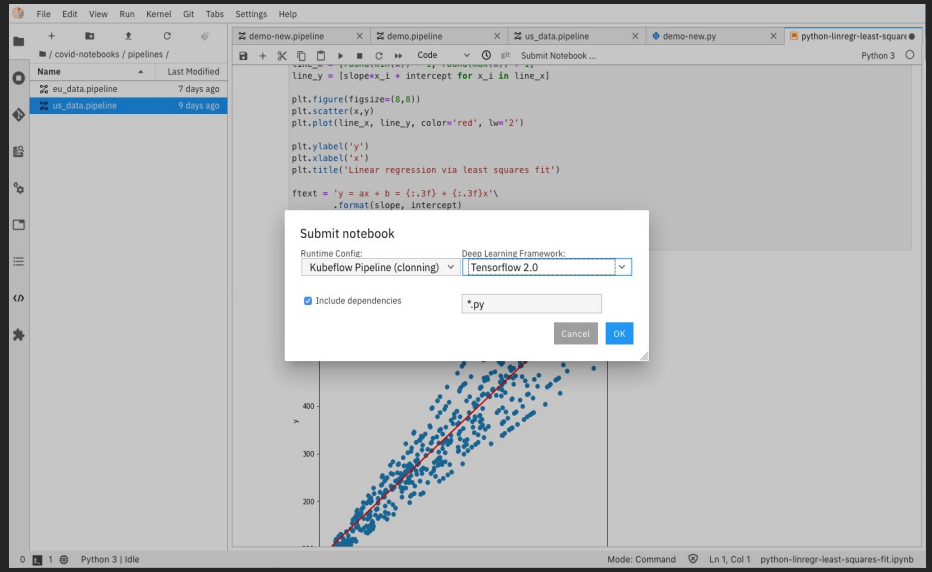
Git integration

Track project changes and share among teammates

Python script execution

Edit and execute python scripts against local or cloud-based resources

Notebook as batch jobs



bit.ly/belpy-dax

Elyra Core Features

Notebook Pipelines editor

Visual editor for building notebook-based AI pipelines, enabling the conversion of multiple notebooks into batch jobs or workflows.

Notebook as batch jobs

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

Code Snippets

Easy creation and insertion of reusable code snippets for the various languages

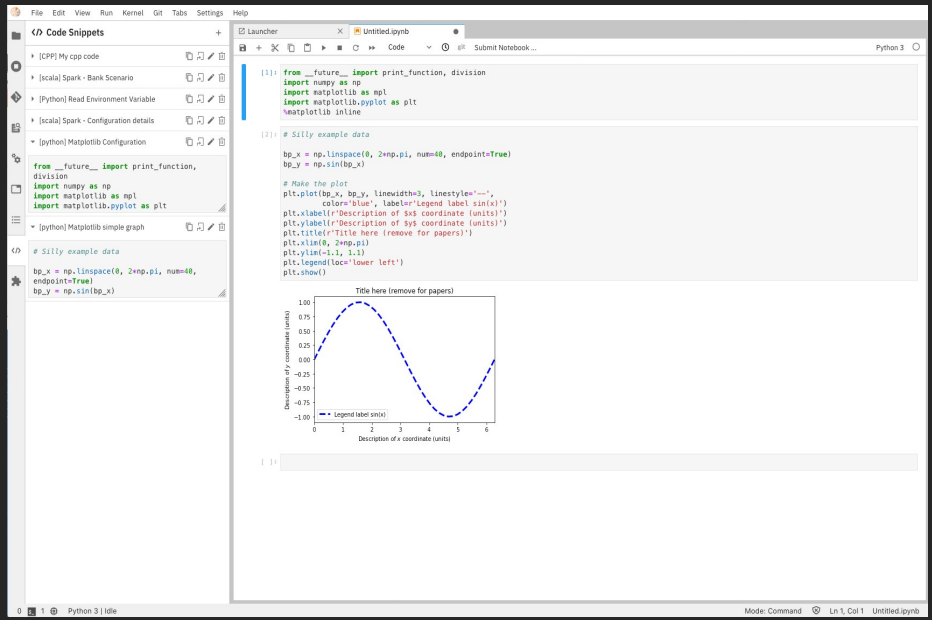
Git integration

Track project changes and share among teammates

Python script execution

Edit and execute python scripts against local or cloud-based resources

Code Snippets



bit.ly/belpy-dax

Elyra Core Features

Notebook Pipelines editor

Visual editor for building notebook-based AI pipelines, enabling the conversion of multiple notebooks into batch jobs or workflows.

Notebook as batch jobs

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

Code Snippets

Easy creation and insertion of reusable code snippets for the various languages

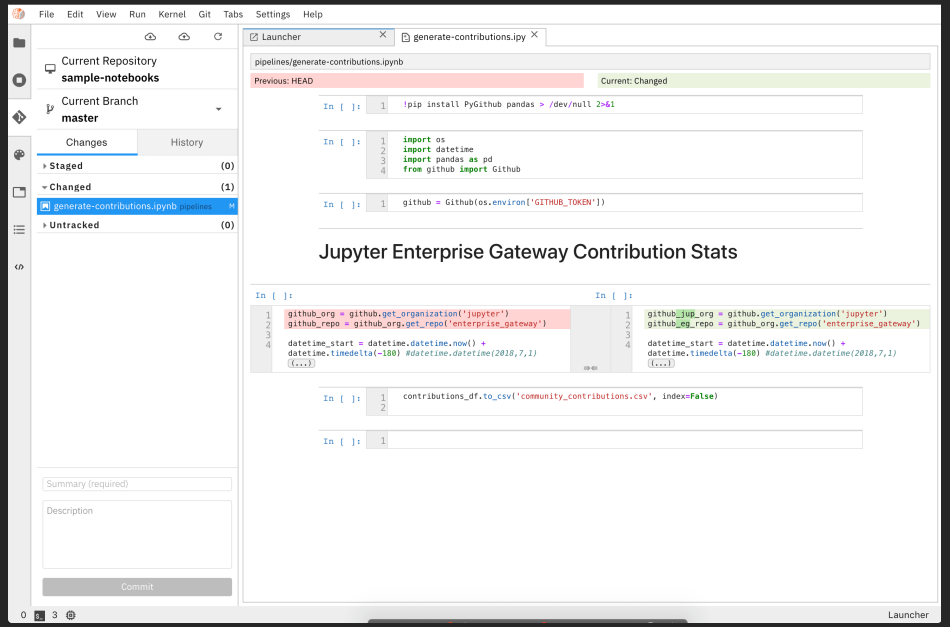
Git integration

Track project changes and share among teammates

Python script execution

Edit and execute python scripts against local or cloud-based resources

Git integration



bit.ly/belpy-dax

Elyra Core Features

Notebook Pipelines editor

Visual editor for building notebook-based AI pipelines, enabling the conversion of multiple notebooks into batch jobs or workflows.

Notebook as batch jobs

Elyra extends the notebook UI to simplify the submission of notebooks as a batch job for model training

Code Snippets

Easy creation and insertion of reusable code snippets for the various languages

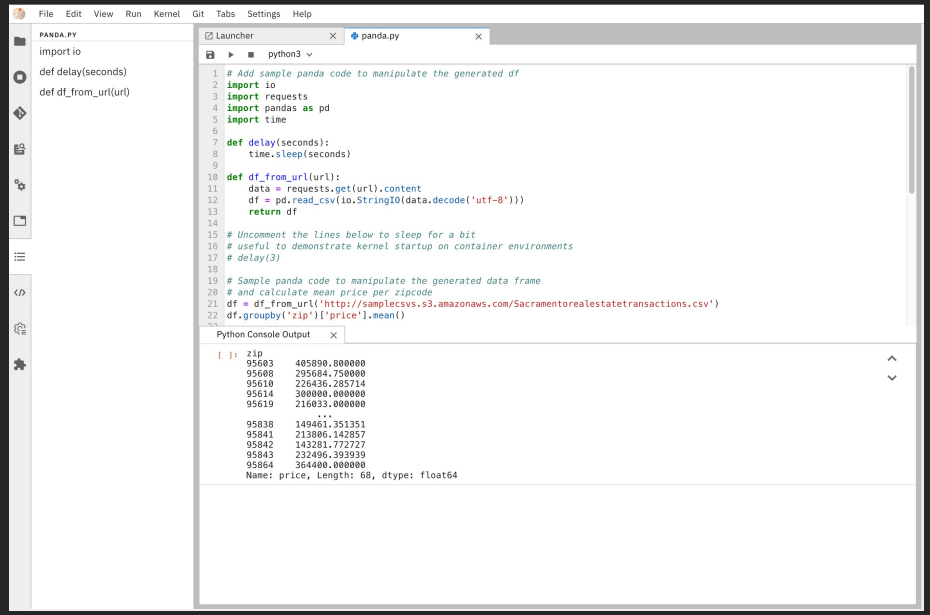
Git integration

Track project changes and share among teammates

Python script execution

Edit and execute python scripts against local or cloud-based resources

Python Script editor



bit.ly/belpy-dax

Getting Started

What are the pre-requisitesto run?

- NodeJS 12+
- Python 3.X
- Anaconda (optional)
- JupyterLab Support
  - JupyterLab 1.X is supported on Elyra 0.10.x and below
  - JupyterLab 2.X is supported on Elyra 1.0.0 and above
- KubeFlow Installation (optional)

Install Elyra

To install Elyra:

$ pip install elyra==1.1.0 && jupyter lab build Or:

$ pip install --upgrade elyra && jupyter lab build

To verify installation:

$ jupyter serverextension list And

$ jupyter labextension list

Starting Elyra:

$ jupyter lab

bit.ly/belpy-dax

Get involved

Getting started with Elyra

https://elyra.readthedocs.io/en/latest/getting_started/installation.html

Elyra's Github

https://github.com/elyra-ai/elyra

Data Asset eXchange

https://developer.ibm.com/exchanges/data/

DAX notebooks Github

https://github.com/elyra-ai/examples/tree/master/pipelines/dax_noaa_weather_data

Contributing to these projects

- Bug reports
- Enhancement requests
- Code reviews

bit.ly/belpy-dax

Data Asset eXchange

Explore useful and relevant data sets for enterprise data science

Related Links

Slides: http://bit.ly/pycon_elyra

Elyra Github: https://github.com/elyra-ai/elyra

DAX Asset eXchange: http://ibm.biz/data-exchange

Elyra demo Github: https://github.com/elyra-ai/examples/

Sign up for IBM Cloud: https://ibm.biz/BdqVxW

Model Asset eXchange: http://ibm.biz/model-exchange

bit.ly/belpy-dax