HPCC SYSTEMS

KSU HACKATHON

Mining the Value of Property Assessment Data

Conducting property analysis is a valuable way to understand the value of the properties in the marketplace. It is also important when making any financial decisions on whether to buy, hold, or sell.

In this hackathon, you will work with the real world property assessment data to conduct property analysis, leveraging the distributed computing environment of HPCC Systems and ECL Cloud IDE from LexisNexis Risk Solutions. A data dictionary will also be provided for a better understanding of the property assessment data.

recording_date	registry_number	sale_date	sale_price	separate_utilities	sewer	site_type	state_code	street_code
2018-10-09T00:00:00Z	009N130296	2018-10-03T00:00:00Z	438990	C	Y	A	1001	61660
2018-12-11T00:00:00Z	009N130298	2018-11-27T00:00:00Z	350000	C	Y	A	1001	61660
2018-11-28T00:00:00Z	009N130299	2018-10-17T00:00:00Z	448000	C	Y	A	1001	61660

Partial Sample Dataset

Project Evaluation:

- **You will be rated on how you break down the problem, how you understand the data, how you shape the data for analysis and then the steps you take to analyze it.**

- HPCC Systems Mentor team will base on below criterial to measure the project performance:

1. Please record each step clearly in the source code in ECL Cloud IDE
2. Report what's discovered in each step in the final presentation.
3. Play each step as a demo in ECL Cloud IDE in the final presentation.

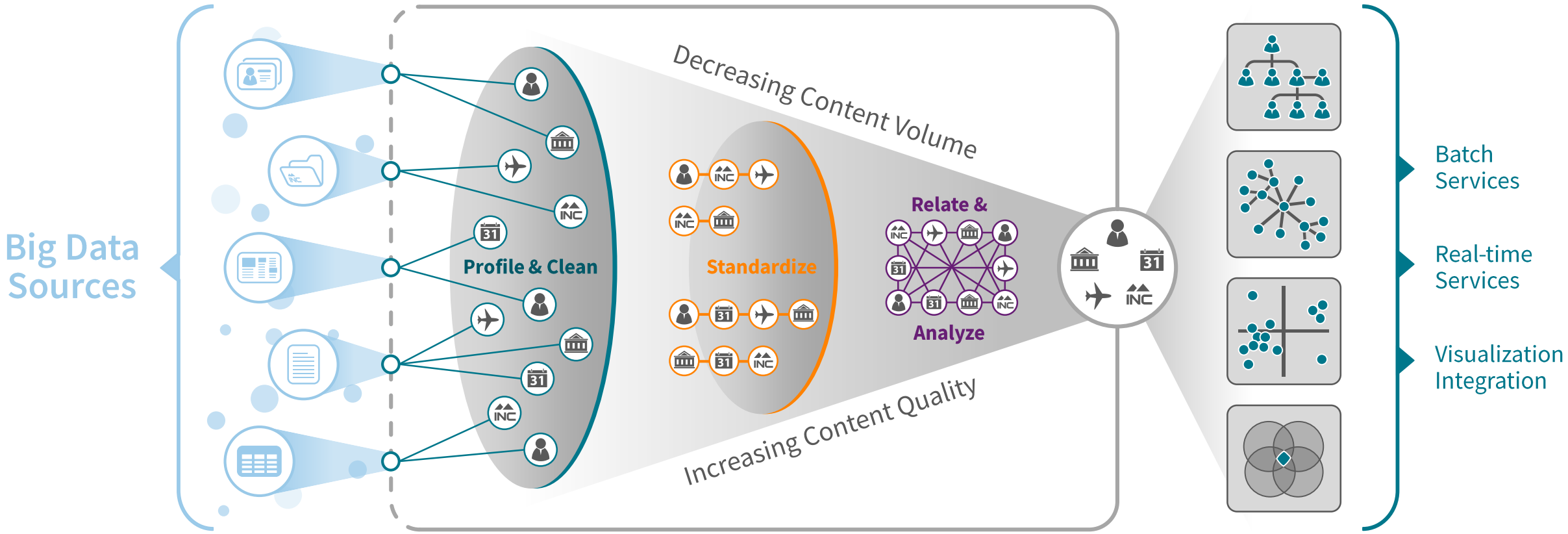
For example, if you execute a data profiling, please record this step clearly in the source code in ECL Cloud IDE and report what's discovered in the final presentation.

Process is more important than the result

What you will need in this hackathon?

HPCC Systems (Small to Big Data) ETL

Integrate > Profile > Clean > Standardize > Aggregate > Machine Learning > Deliver



Unstructured and Structured Content

HPCC Systems Data Lake

Analysis Applications

Machine Learning on HPCC Systems Platform

Core Bundles

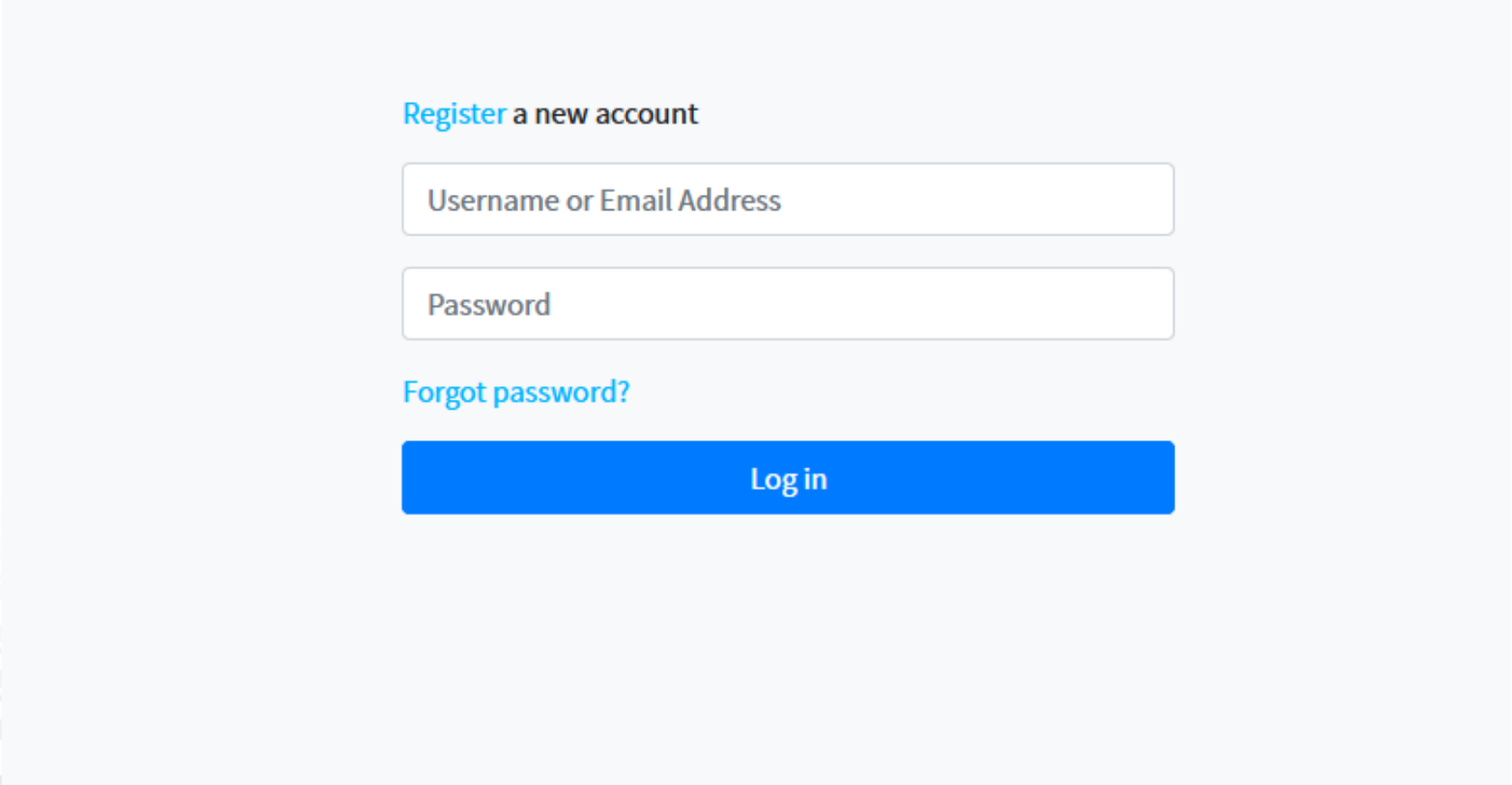
- ML_Core
- PBblas

Algorithm Bundles

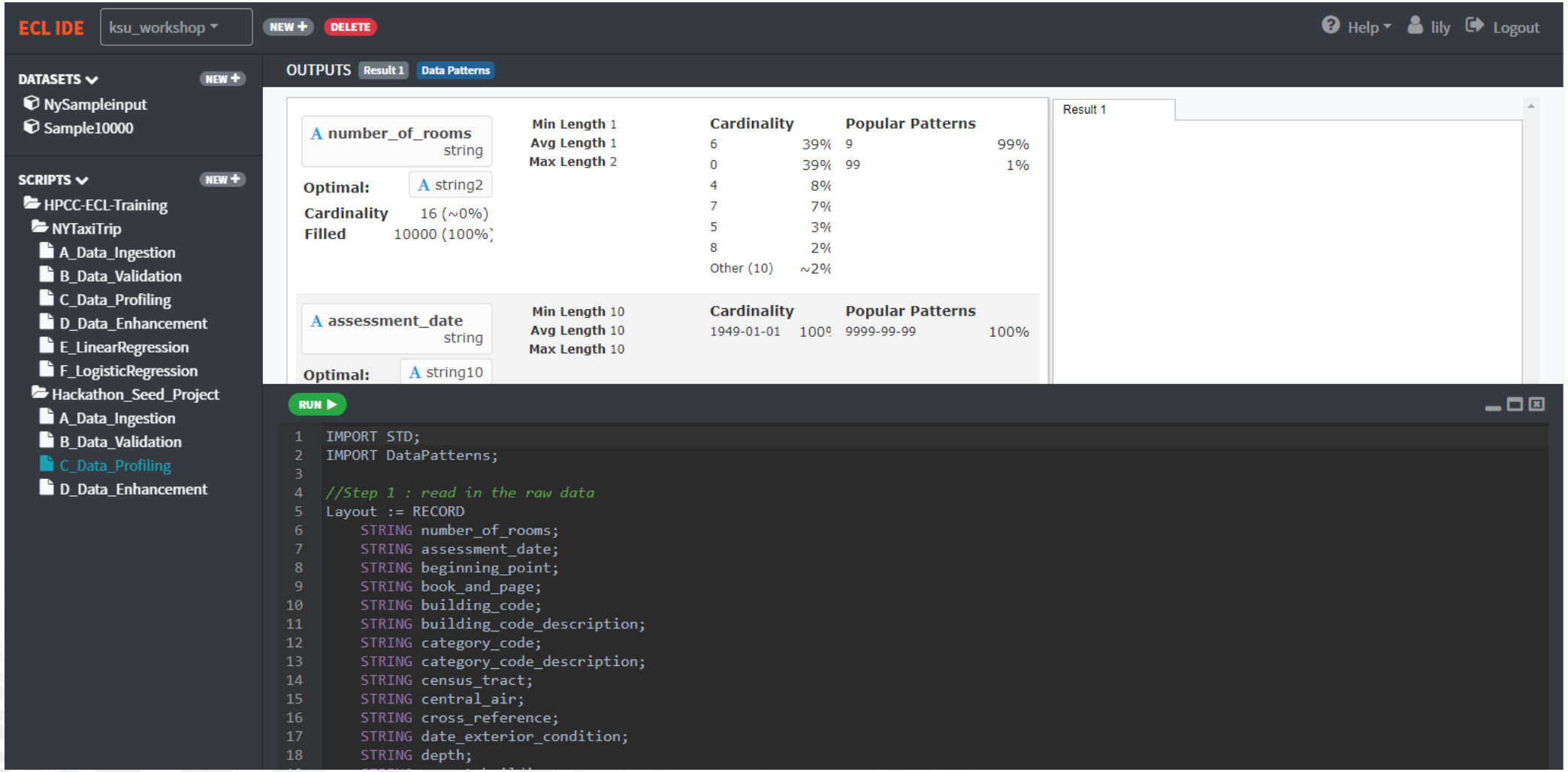
- LinearRegression
- LogisticRegression
- Text Vectors
- GLM
- K-Means
- SVM
- LearningTrees

ECL Cloud IDE

https://ide.hpccsystems.com/



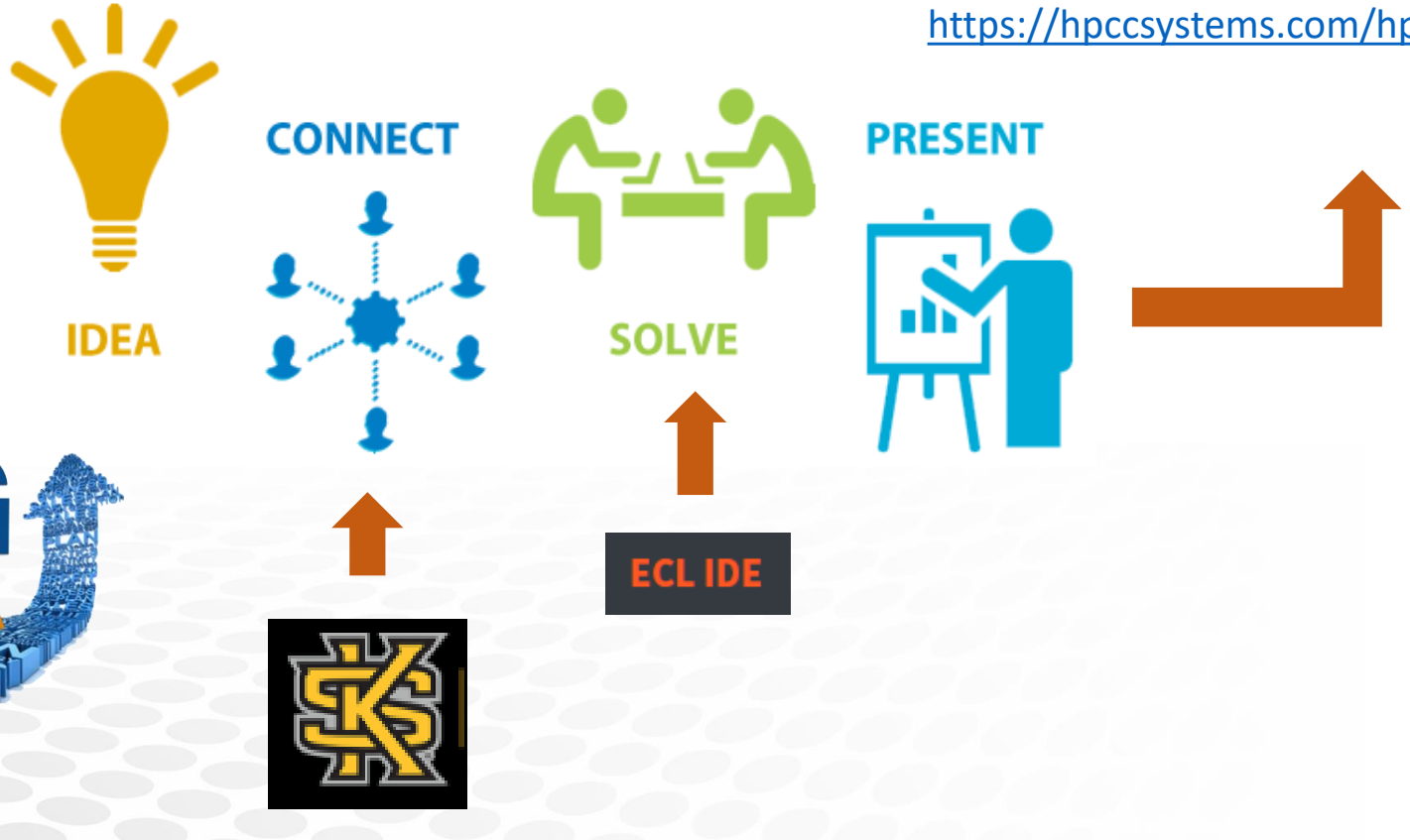
ECL Cloud IDE



ECL Cloud IDE WorkSpace

HPCC Systems KSU Hackathon 2019

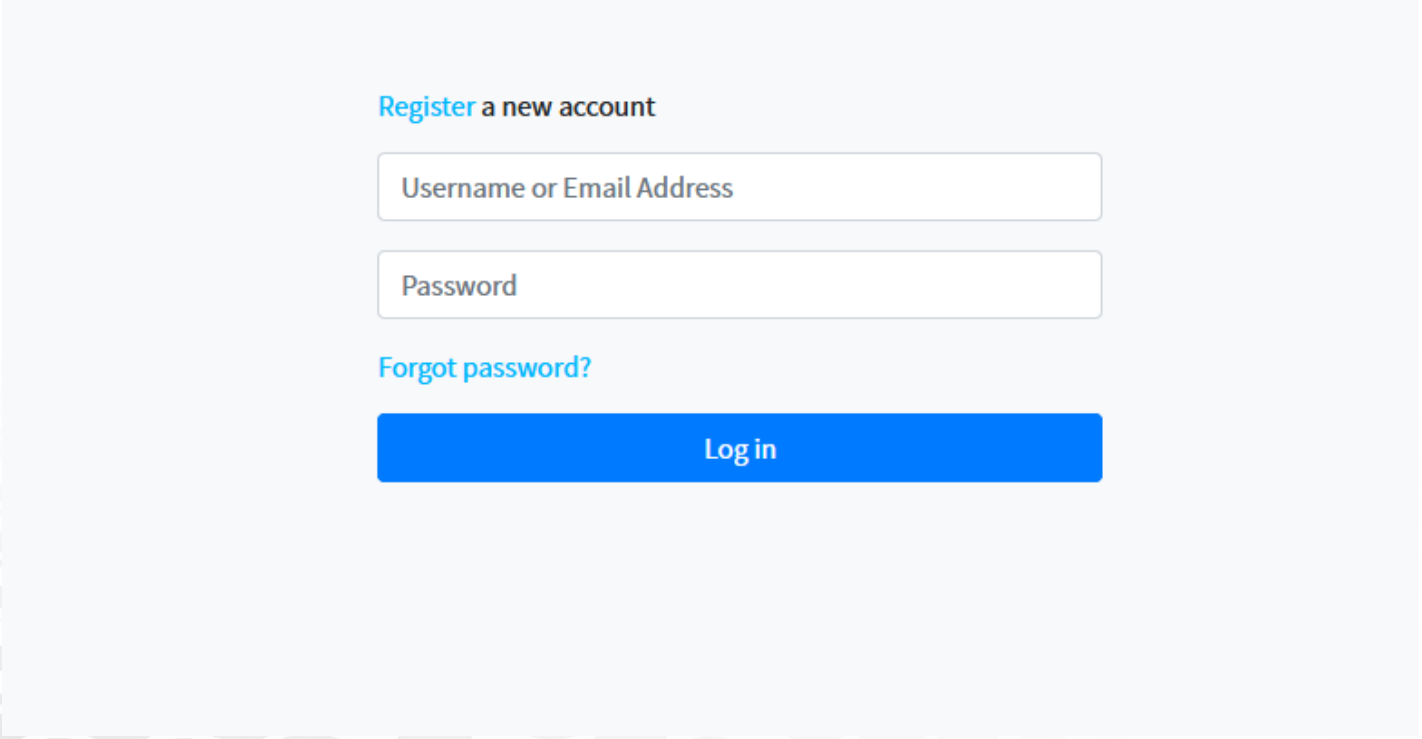
https://hpccsystems.com/hpccsummit2019



Hackathon Tips

1. Register to become a member of ECL Cloud IDE on KSU Campus

https://ide.hpccsystems.com/auth/login



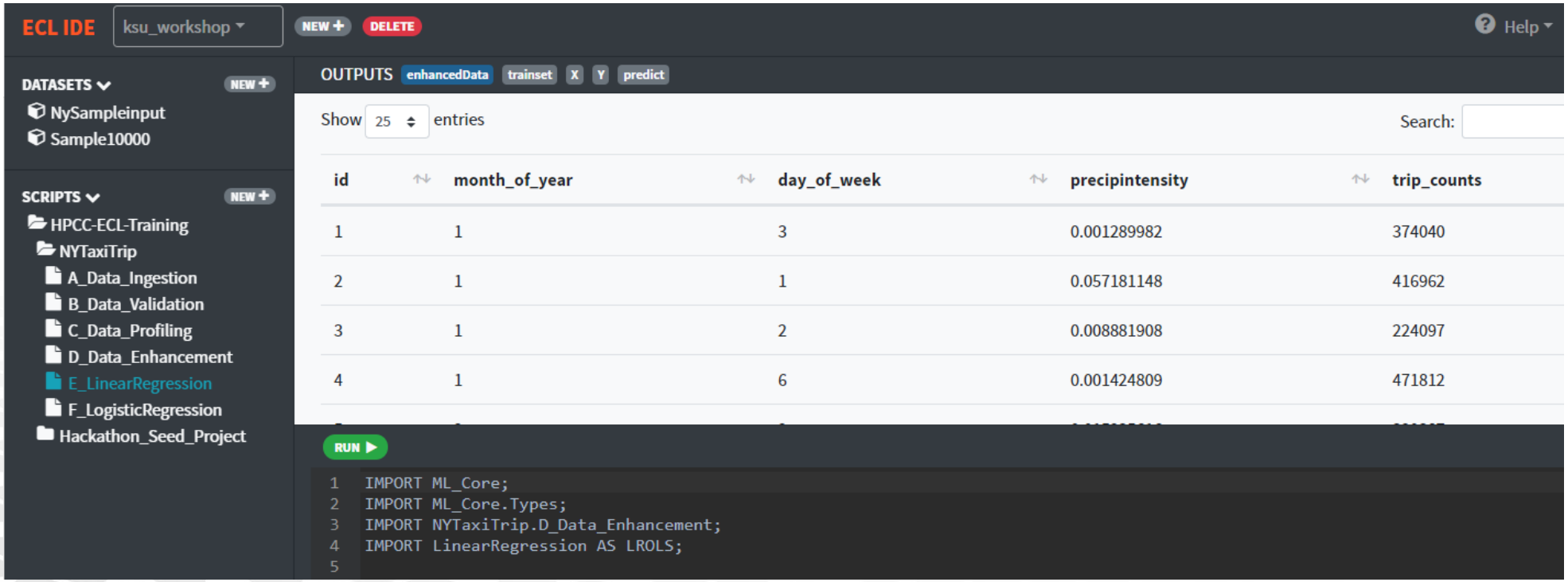
Hackathon Tips – cont.

2. Copy the KSU_Hackathon WorkSpace into your ECL Cloud IDE by open below link in your browser:

https://ide.hpccsystems.com/workspaces/share/fe09f0c6-62b4-4489-9e24-7f5a68c18e4b

Hackathon Tips – cont.

3. Try NYTaxiTrip examples in HPCC-ECL-Training folder to rewind the workshop



Hackathon Tips – cont.

- NOTE:

The name convention to refer the file uploaded to your workspace is

'~USERNMAE::WORKSPACENAME::RAWFILENAME‘

Example:

If your username is Mike, you created a worksplace 'HPCCSystems‘ and uploaded the file 'test.csv‘ to the workspace.

To use the dataset in ECL Cloud IDE, the directory to refer the file should be '~Mike::HPCCSystems::test.csv‘.

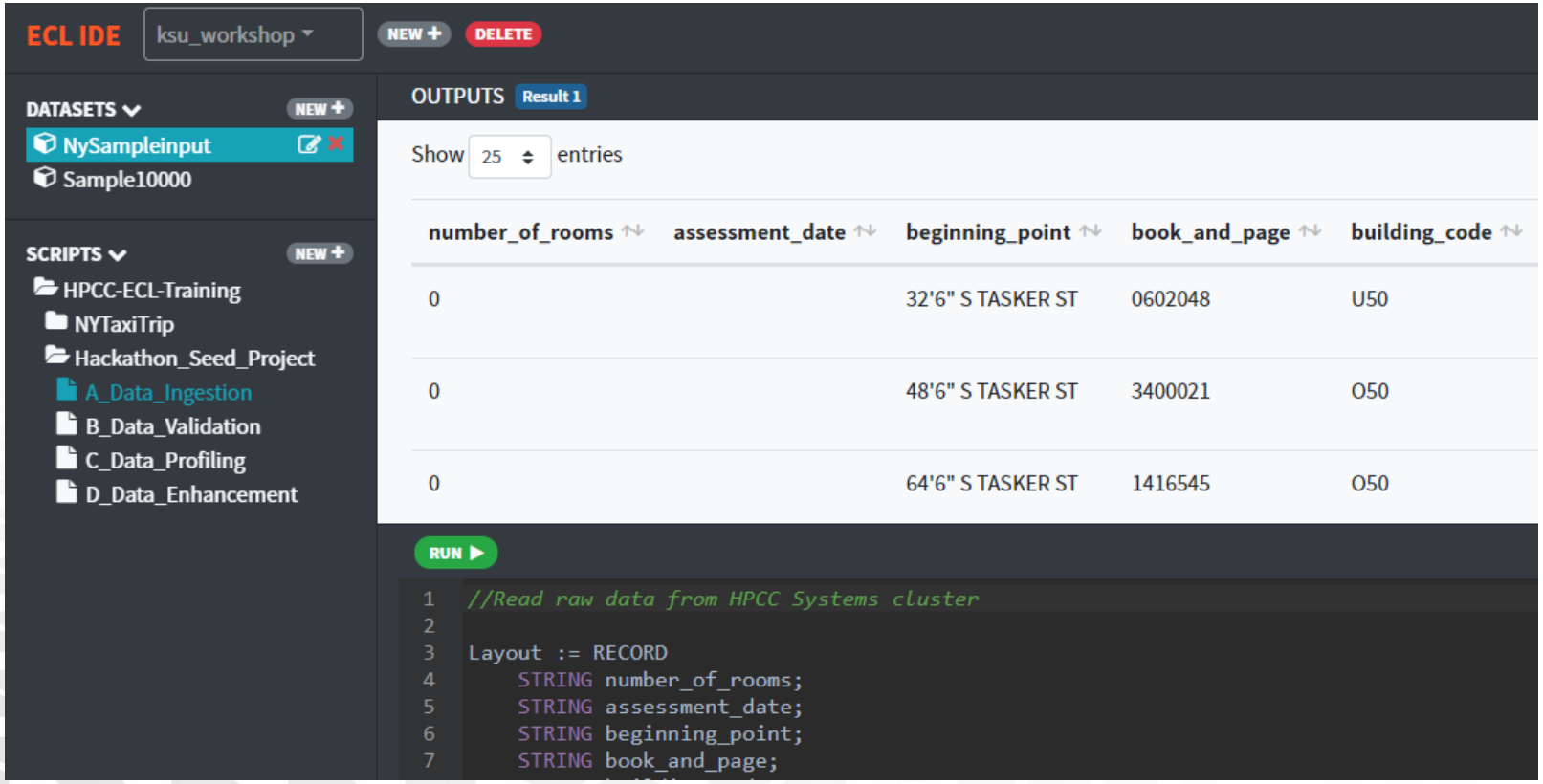
To ingest the dataset in ECL Cloud IDE, you can use DATASET function, such as:

raw := DATASET('~Mike::HPCCSystems::test.csv‘, Layout, CSV(HEADING(1)));

Hackathon Tips – cont.

4. Try the Hackathon_Seed_Project in ECL Cloud IDE.

This project shows a few sample steps that help you familiar with the Property Assessment sample dataset -- Sample10000. It's uploaded to the shared WorkSpace KSU_Workshop.



Hackathon Tips – cont.

- **Note:**
- The sample dataset Sample10000 only includes 10,000 records of the original Property Assessment dataset. Your solution should apply to the original Property Assessment Dataset as final result.

- The original Property Assessment dataset can be accessed as below in A_Data_Ingestions file:

```
propertyDS := DATASET('~ksu::hackathon::opa_properties_public.csv', Layout,
                                                                    CSV(HEADING(1)));
```

Hackathon Tips – cont.

5. Understand Property Assessment Data via Data Dictionary:

   https://github.com/lilyclemson/KSU_Hachathon2019

Hackathon Tips – cont.

6. Other HPCC Systems KSU hackathon related ECL Code examples and slides are available at Git repository KSU_Hackathon2019 :

https://github.com/lilyclemson/KSU_Hachathon2019

Hackathon Tips – cont.

7. Welcome to join our #hpcc Channel to interact with mentors and ask questions via below slack space:

   #ksuccsehackathon: ksuccsehackathon.slack.com

Helpful Links

- Introduction of HPCC System
  https://hpccsystems.com/about
- ECL CheatSheet:
  https://github.com/hpcc-systems/HPCC-ECL-Training/tree/master/CheatSheet
- Introduction of HPCC Systems Machine Learning Library
  https://hpccsystems.com/download/free-modules/machine-learning-library
- ECL Machine Learning Examples:
  https://github.com/hpcc-systems/HPCC-ECL-Training/tree/master/NYTaxiTrip
  https://github.com/lilyclemson/HPCC-ECL-Training/tree/master/StockTrade
- Other Documentations
  https://hpccsystems.com/training/documentation
- Opportunity to attend HPCC-Systems-Summit-2019
  https://hpccsystems.com/community/events/hpcc-systems-summit-2019


LexisNexis®
RISK SOLUTIONS