Cortex
ANALYTICS SIMULATION

Credit Risk Scenario – Round 2

SAS Viya 4.0

Last Update: February 23, 2022

ERPsimLab
HEC MONTRĒAL
Serious games to learn enterprise systems and business analytics

Table of Contents

- Round Overview
- Access SAS Viya
- Explore the Game
- Play Round 2: Improve Pipeline
- Score New Data
- Explore/Visualize/Export Results

© 2004-2022, ERPsim Lab, HEC Montréal.

Round 1 Recap

In Round 1, your goal was to maximize the total net profit generated, using the Risk Modeling Add-on for SAS Visual Data Mining and Machine Learning and developing a scorecard.

Round 2 Overview

In Round 2, your goal is still to maximize net profit generated. To do so, you will learn from data through the Machine Learning Model Approach in the SAS Analytics Life Cycle, where you will predict defaults in the loan applications without the rule-based programming such as a scorecard.

© 2004-2022, ERPsim Lab, HEC Montréal.

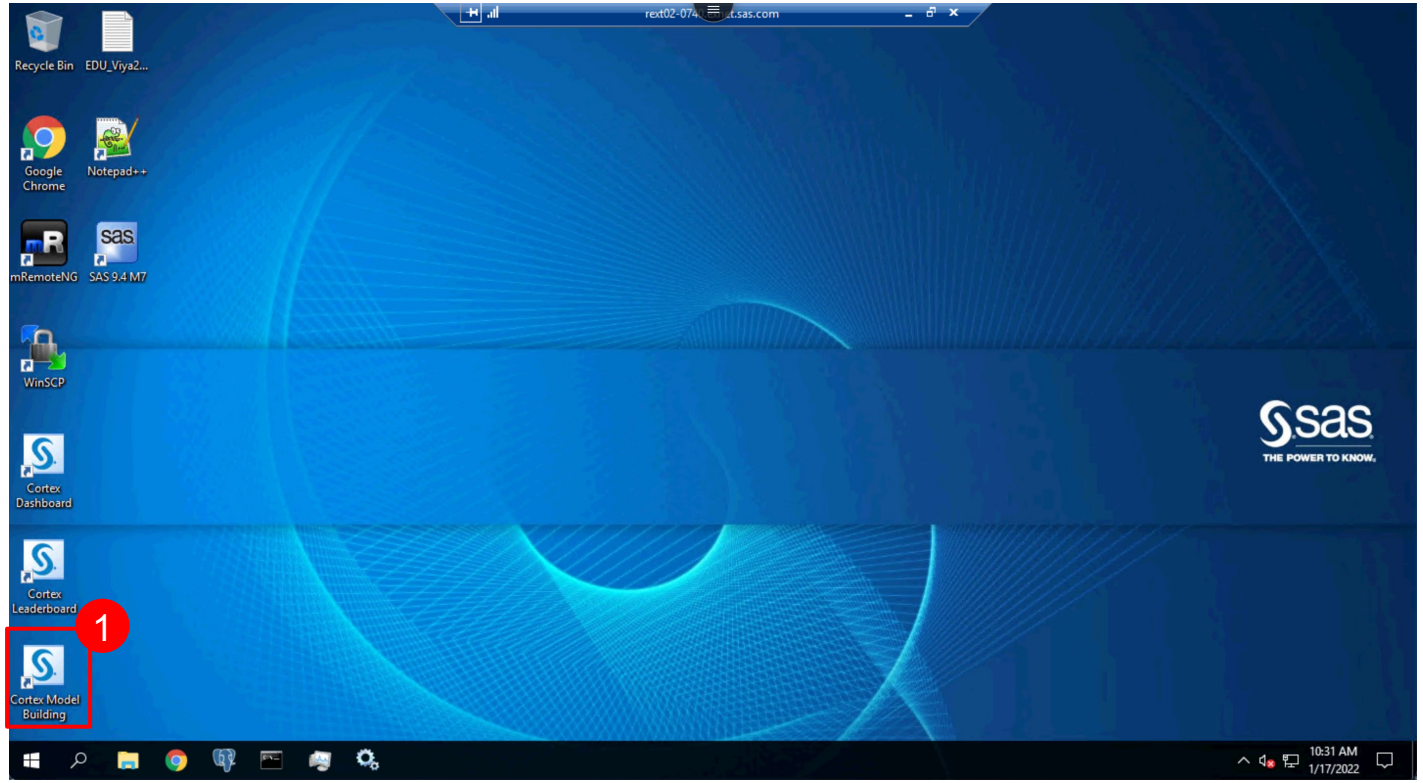
Access SAS Viya

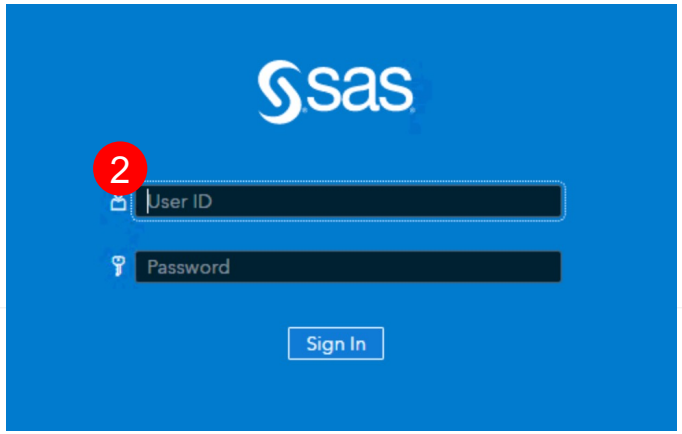
1. On your virtual desktop, click on ‘Cortex Model Building’ icon

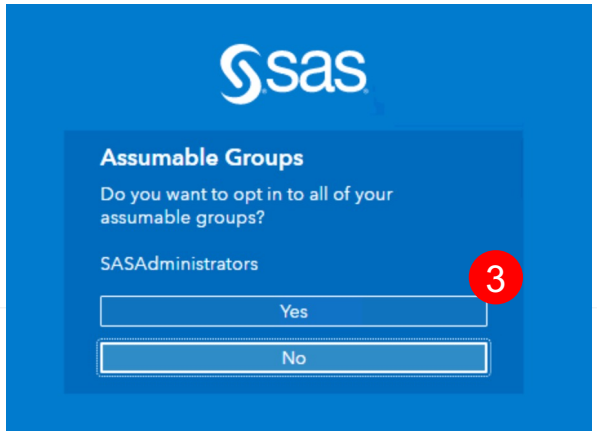
2. Enter your User ID and password:
   - User ID: student
   - Password: Metadata0

NOTE: Use the above credentials unless otherwise specified by your instructor

3. Select ‘Yes’ when asked about ‘Assumable Groups’



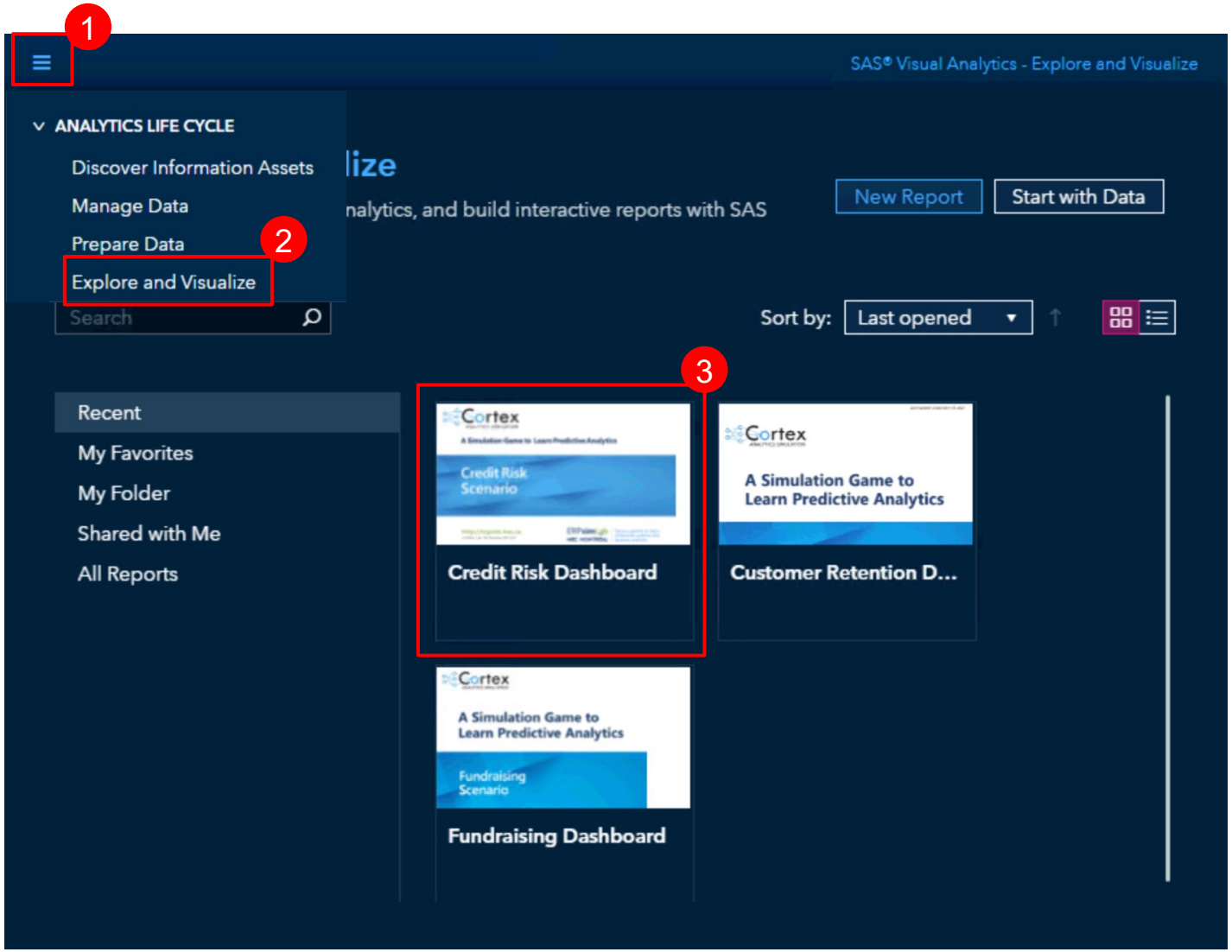




© 2004-2022, ERPsim Lab, HEC Montréal.

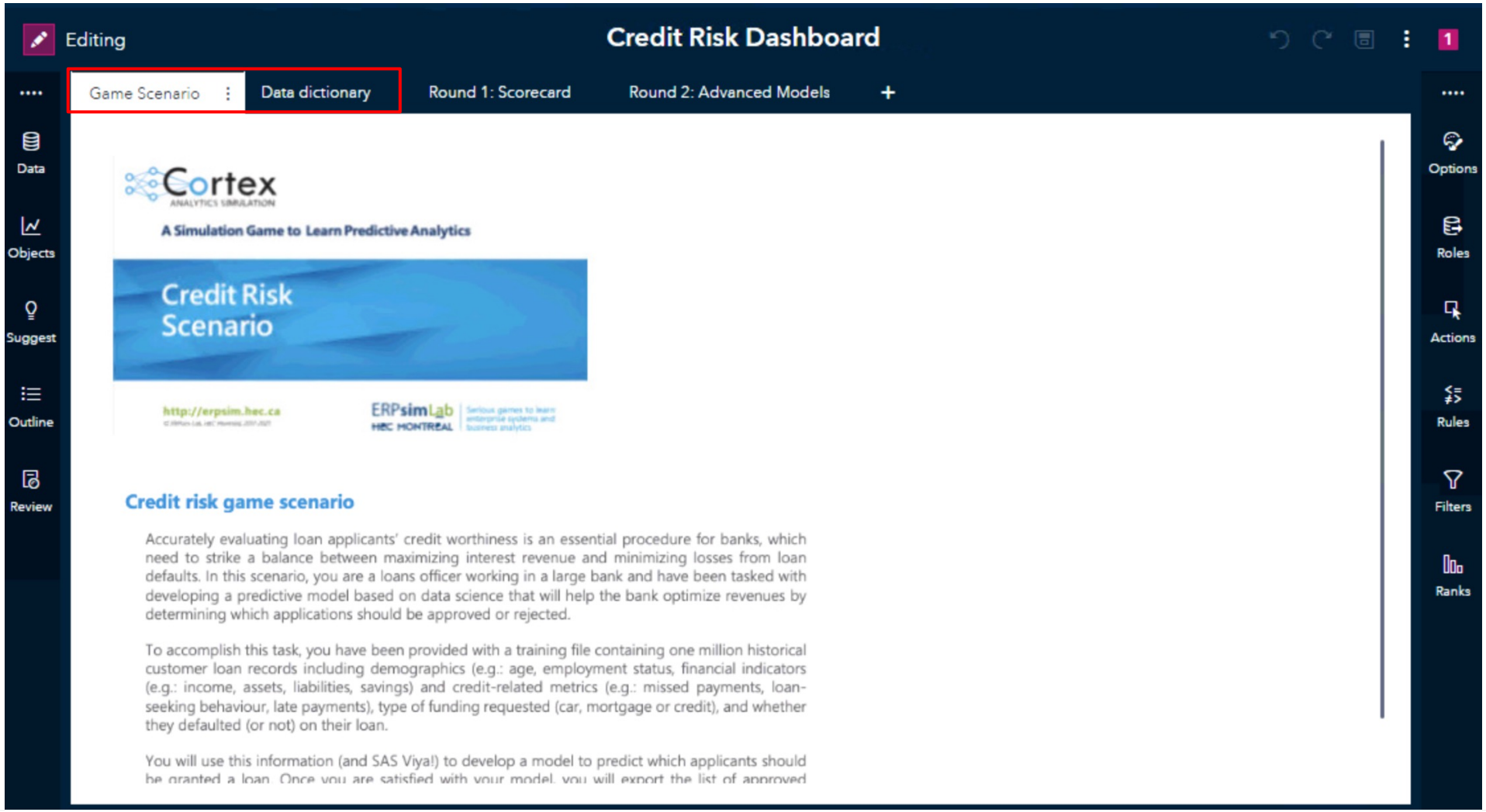
Explore the Game

1. In SAS Drive, click on the menu (hamburger icon) on the top left corner
2. Select 'Explore and Visualize'
3. Double click on 'Credit Risk Dashboard'



© 2004-2022, ERPsim Lab, HEC Montréal.

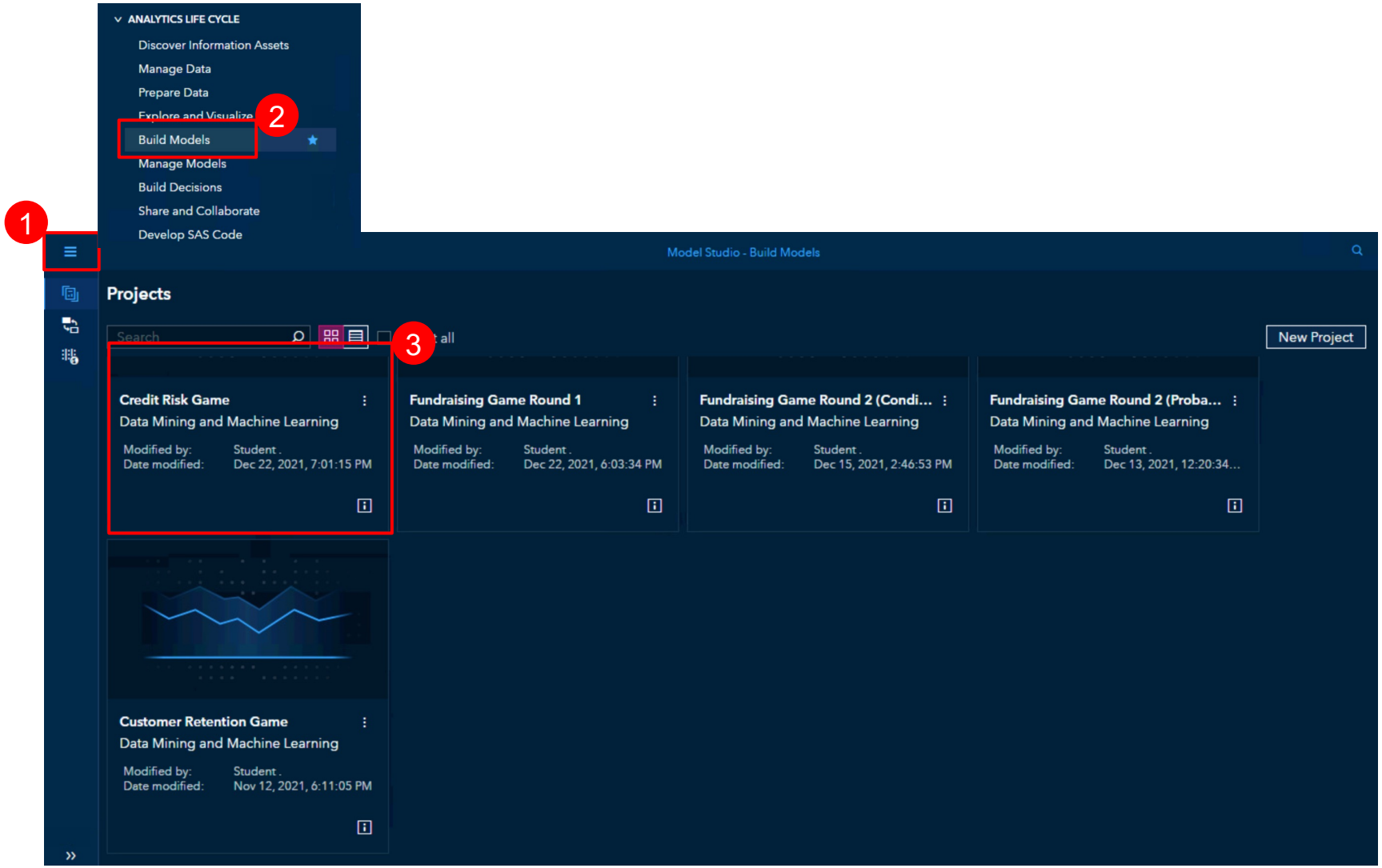
Use the first two tabs (Game Scenario and Data dictionary) to explore the game scenario and data



© 2004-2022, ERPsim Lab, HEC Montréal.

Open Project:
Credit Risk Game

1. In SAS Drive, click on the menu (hamburger icon) on the top left corner
2. Select 'Build Models'
3. Double click on 'Credit Risk Game' project



© 2004-2022, ERPsim Lab, HEC Montréal.

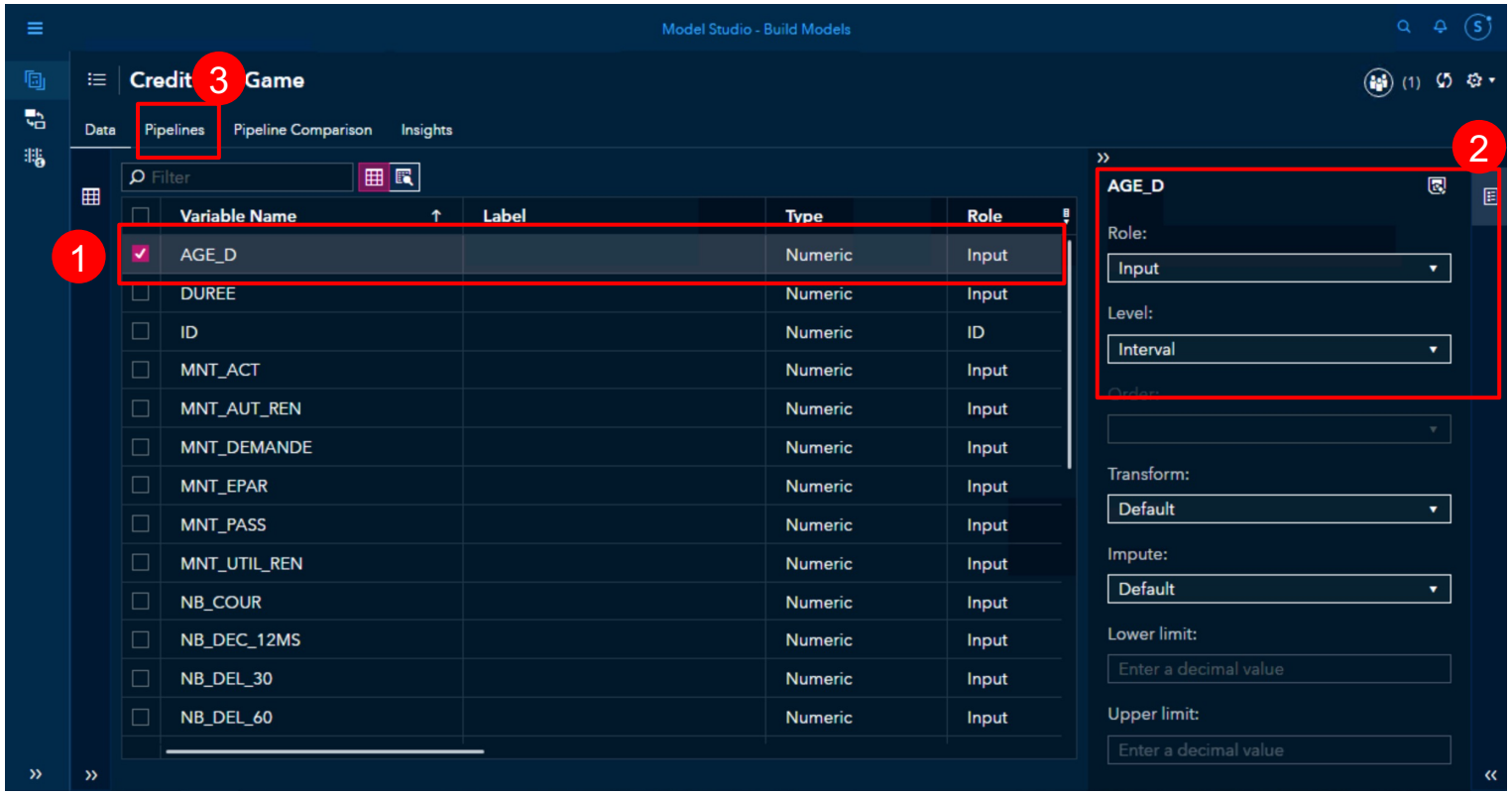
Assign Variable Roles & Levels

1. Select variables (e.g., age)
2. If needed, change their role and level in the right-side panel

NOTE 1: Remember to deselect a variable before selecting the next when modifying them individually

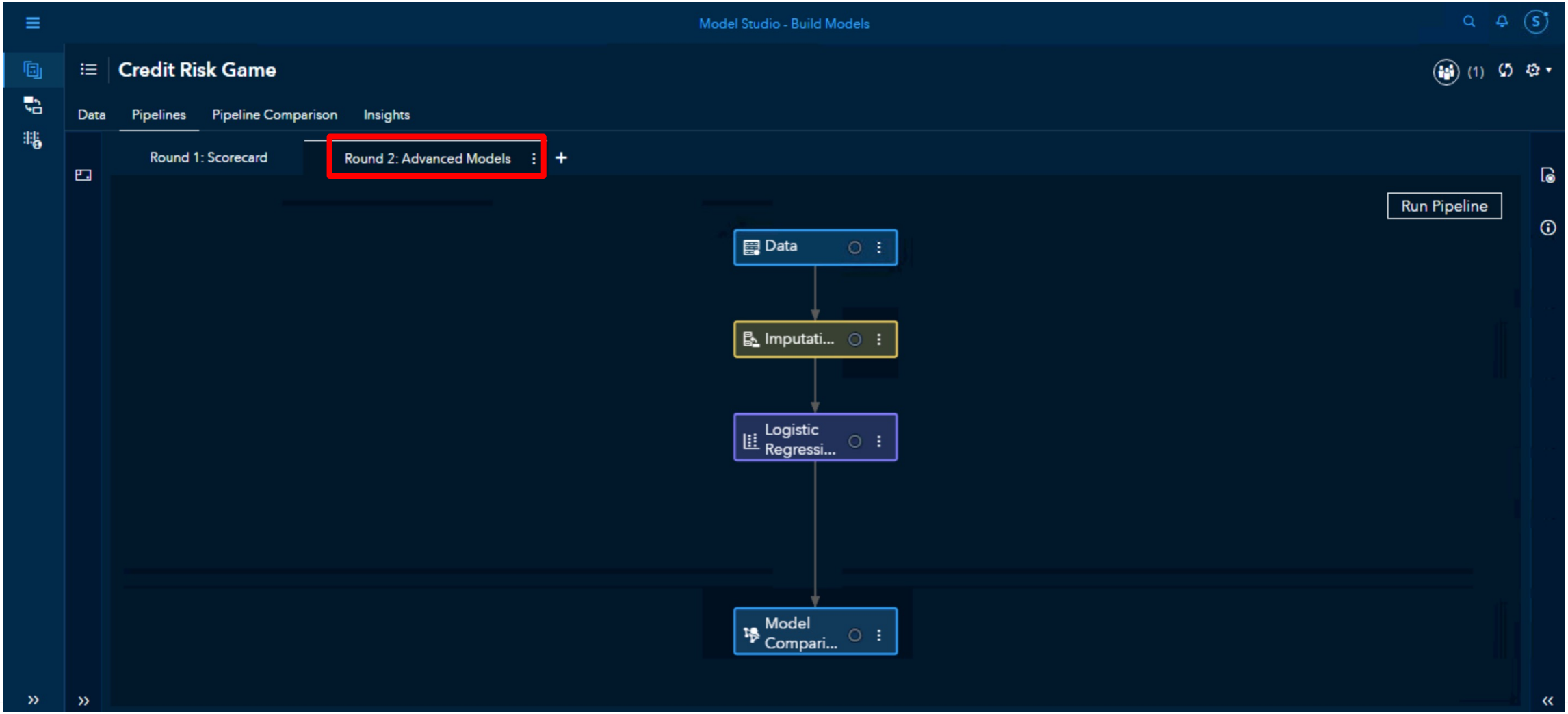
NOTE 2: 'target_0' should stay the target variable (scroll down the to see the variable)

3. Once done, click on the 'Pipelines' tab



© 2004-2022, ERPsim Lab, HEC Montréal.

Click on the tabs to view pre-built pipelines for each round (e.g., this is the pipeline for "Round 2: Advanced Models")



© 2004-2022, ERPsim Lab, HEC Montréal.

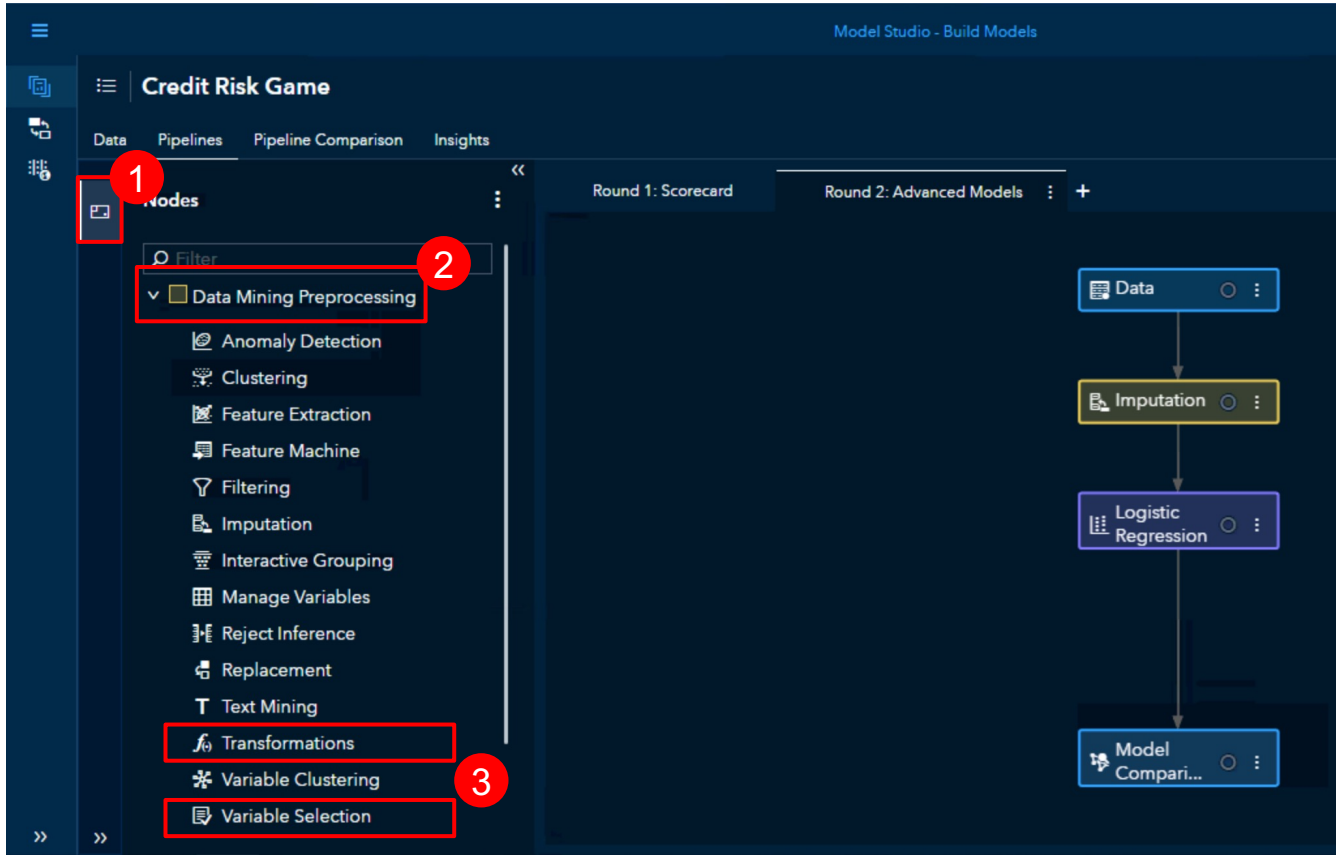
Improve Pipeline: Data preprocessing

Besides 'Imputation' node that already exists in the pipeline, you could also try adding other preprocessing nodes to your pipeline such as 'Transformations' or 'Variable Selection'. To do so:

1. Click the Nodes icon to expand the available node options

2. Select 'Data Mining Preprocessing'

3. Drag and Drop the desired node(s) to the pipeline (where appropriate depending on the node and its intended use)



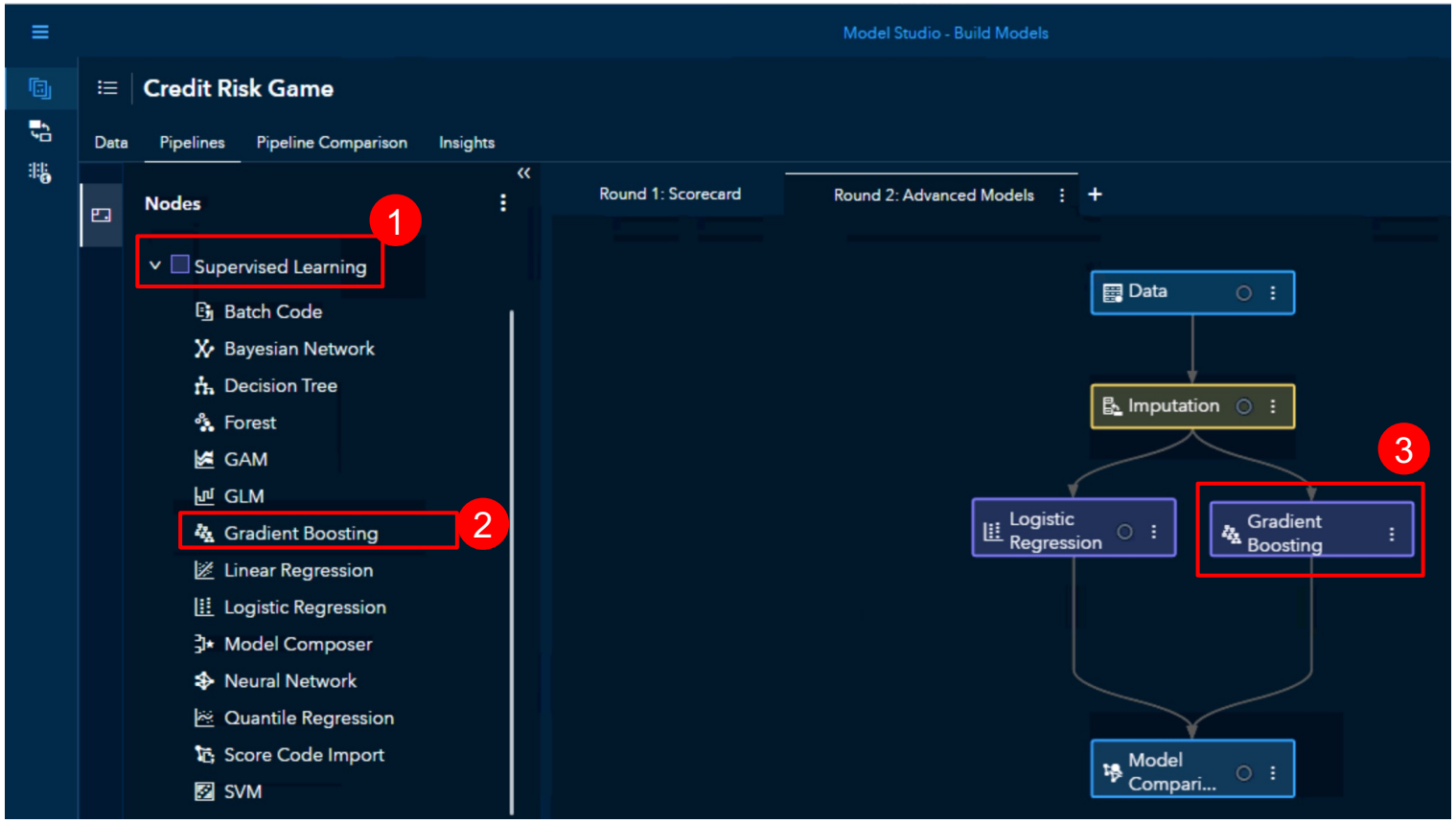
© 2004-2022, ERPsim Lab, HEC Montréal.

Improve Pipeline: Supervised Learning

Depending on the selected models, the 'run' time could vary significantly. E.g., execution of Neural Nets or Gradient Boosting, might take longer to finish!

You could try adding other Machine Learning models to the pipeline. To do so:

1. Select 'Supervised Learning'
2. Select the model you would like to add (e.g., Gradient Boosting)
3. Drag and drop it to the pipeline (here under the Imputation node)



© 2004-2022, ERPsim Lab, HEC Montréal.

Improve Pipeline: Model Properties

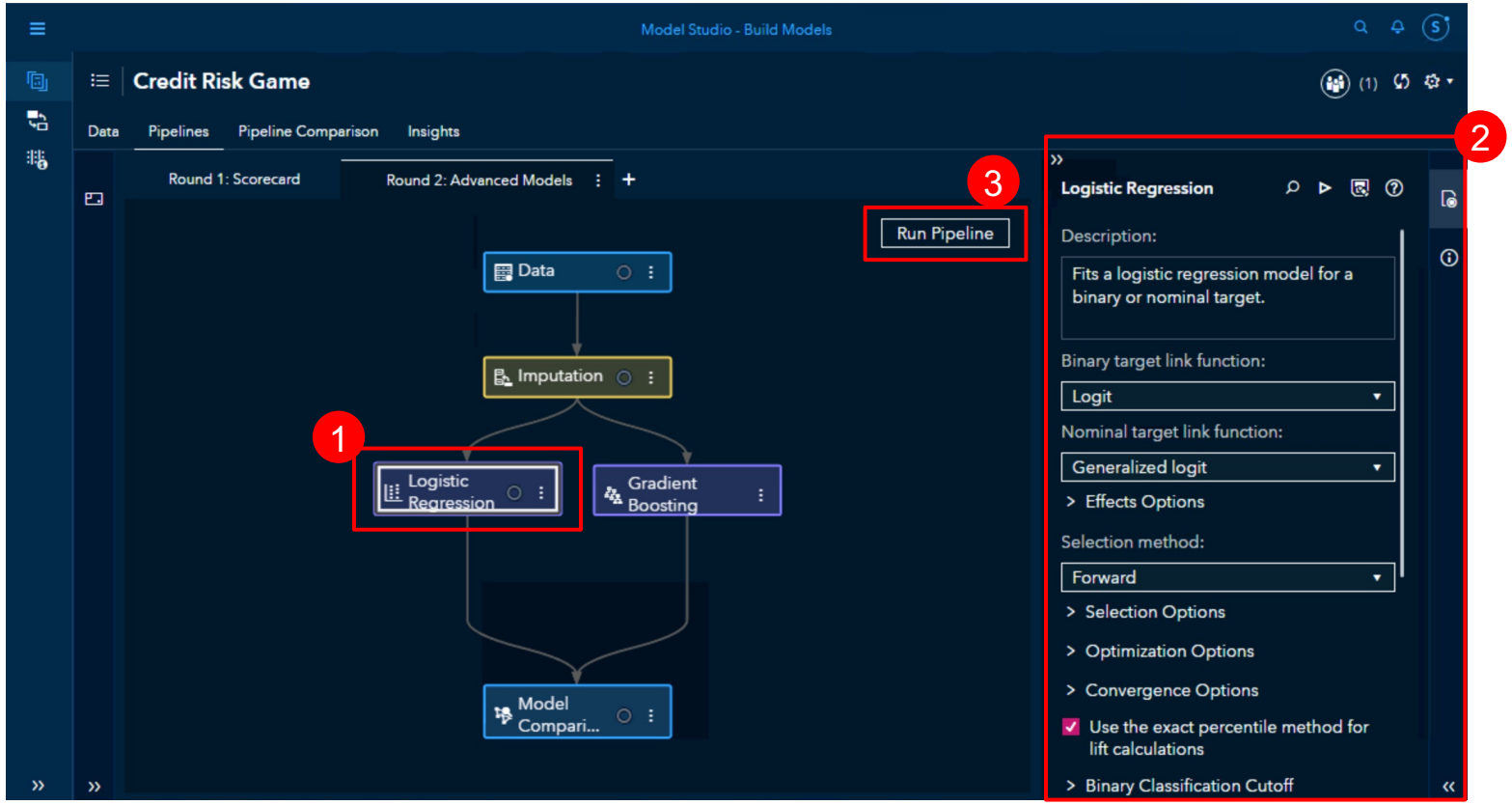
To improve your pipeline, you could also modify the properties of your selected Machine Learning models. To do so:

1. Click on the model you wish to modify (e.g., Logistic Regression)

2. Change their properties on the right-hand side panel

3. Once done with all the changes, Run Pipeline

NOTE: Green check marks appear on nodes when the run is complete.

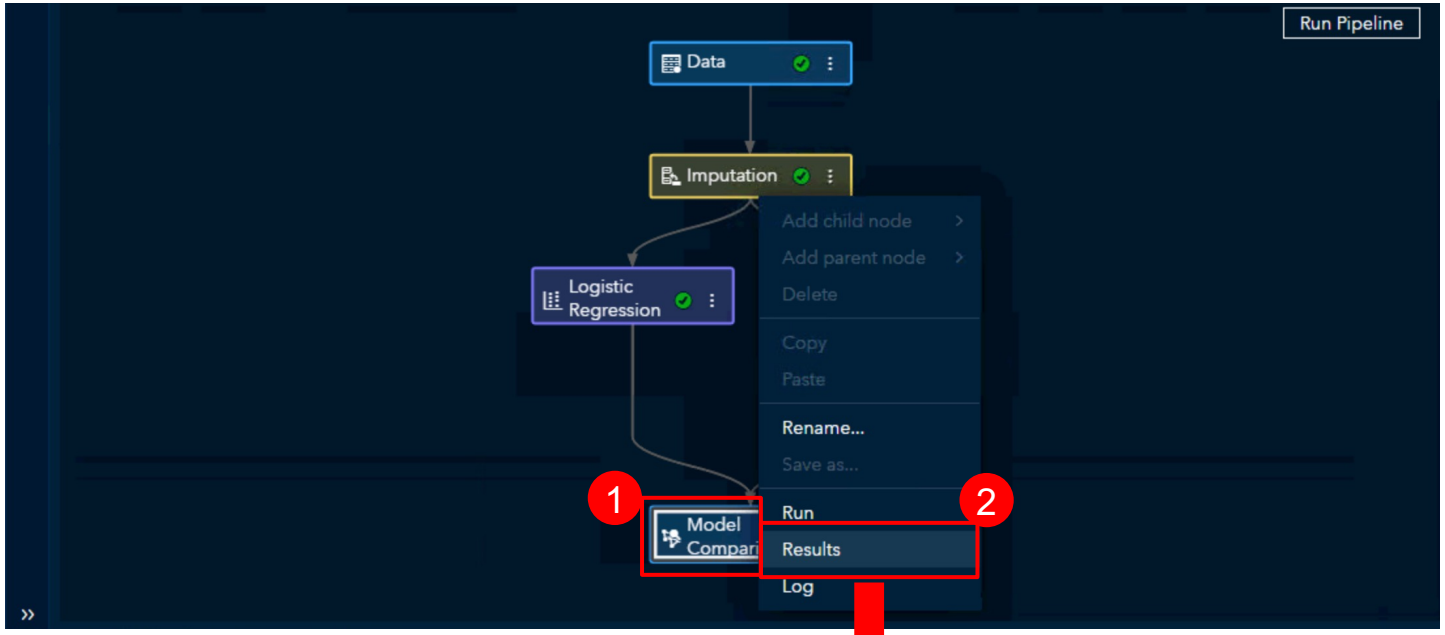


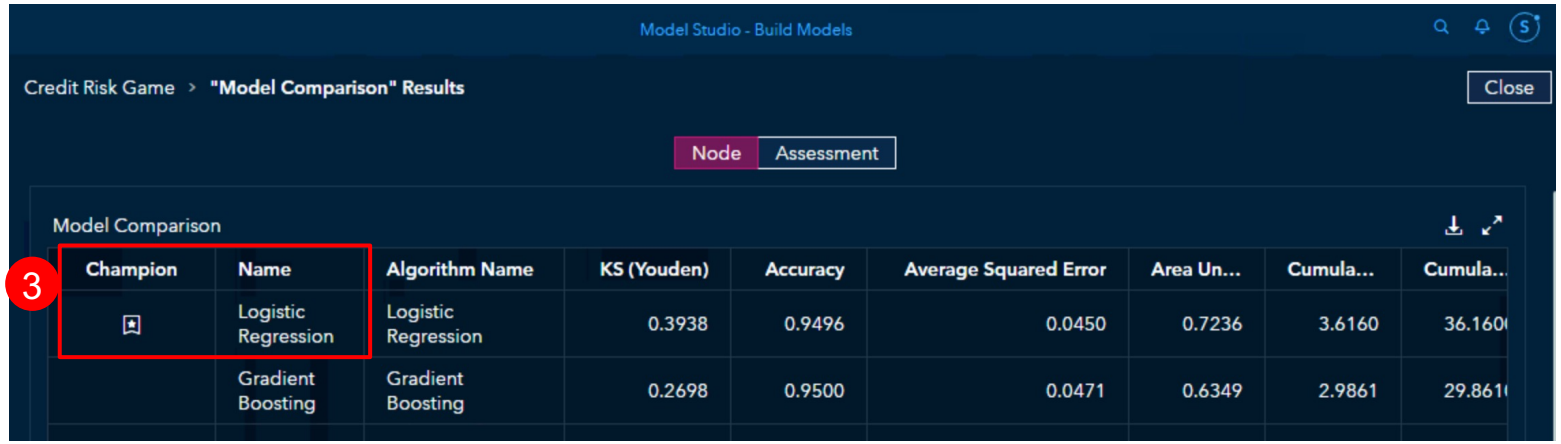
© 2004-2022, ERPsim Lab, HEC Montréal.

Explore Results: Model Comparison

Once run is complete, explore the results of different nodes. First, let's check which model performed best. To do so:

1. Right-click on the 'Model Comparison' node
2. Select 'Results'
3. See the Champion Model (here Logistics Regression)





Champion	Name	Algorithm Name	KS (Youden)	Accuracy	Average Squared Error	Area Un...	Cumula...	Cumula...
	Logistic Regression	Logistic Regression	0.3938	0.9496	0.0450	0.7236	3.6160	36.160
	Gradient Boosting	Gradient Boosting	0.2698	0.9500	0.0471	0.6349	2.9861	29.861

© 2004-2022, ERPsim Lab, HEC Montréal.

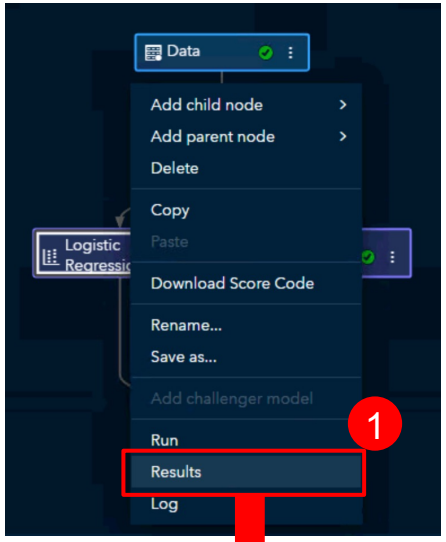
Explore Results: Logistic Regression

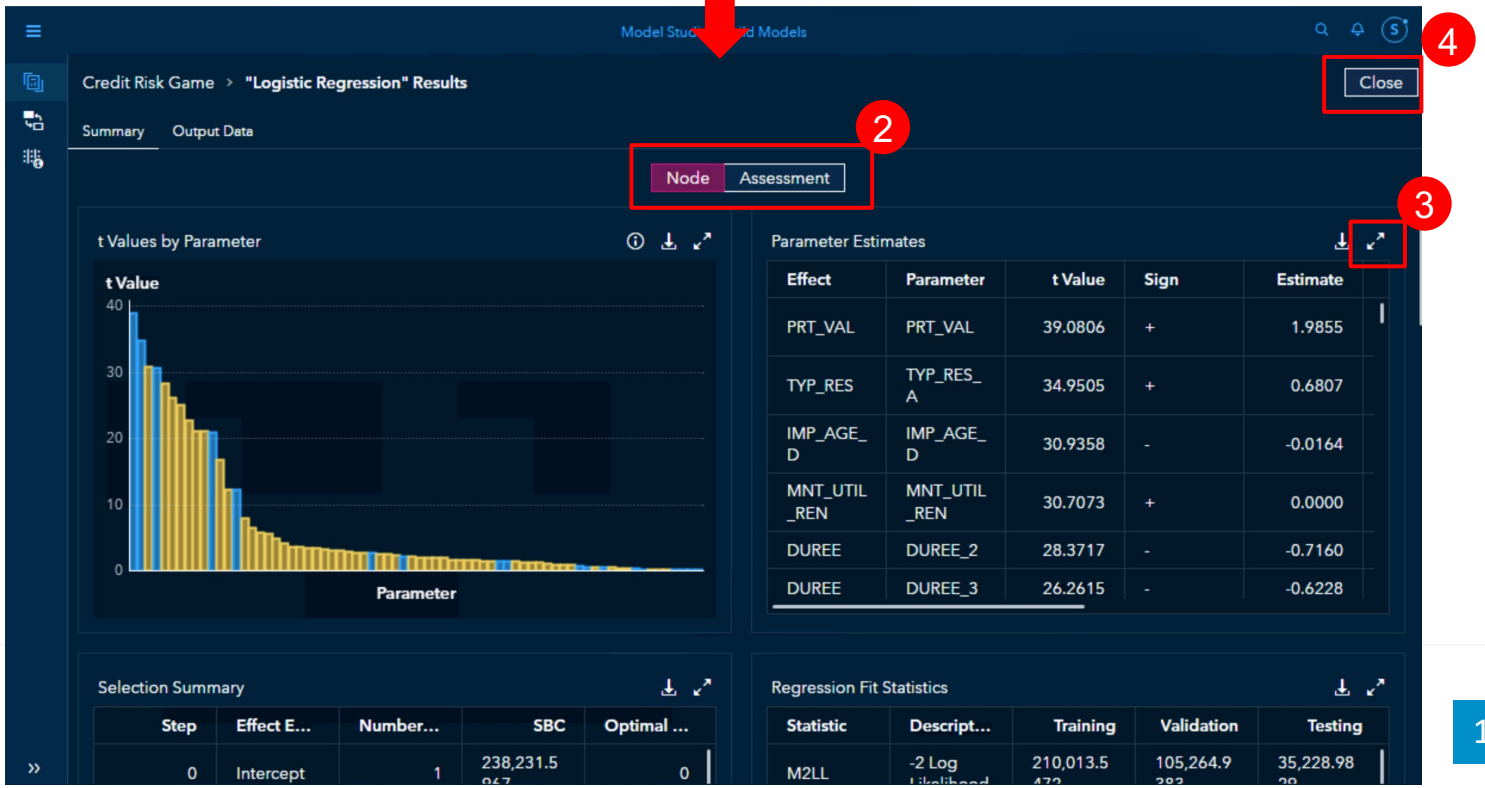
1. Right-click on the 'Logistic Regression' node and select 'Results'

2. Check different statistics, outputs, and reports under 'Node' and 'Assessment' tabs

3. 'Expand' each results window as needed

4. Once done, 'Close' the results to return to the pipeline





© 2004-2022, ERPsim Lab, HEC Montréal.

Compare Pipelines

1. Add and build more pipelines by clicking on '+' sign beside pipelines

2. Use 'Pipeline Comparison' tab to find out which pipeline has the best result

NOTE: Pipeline Comparison displays the pipelines and models in the order of the best performing.

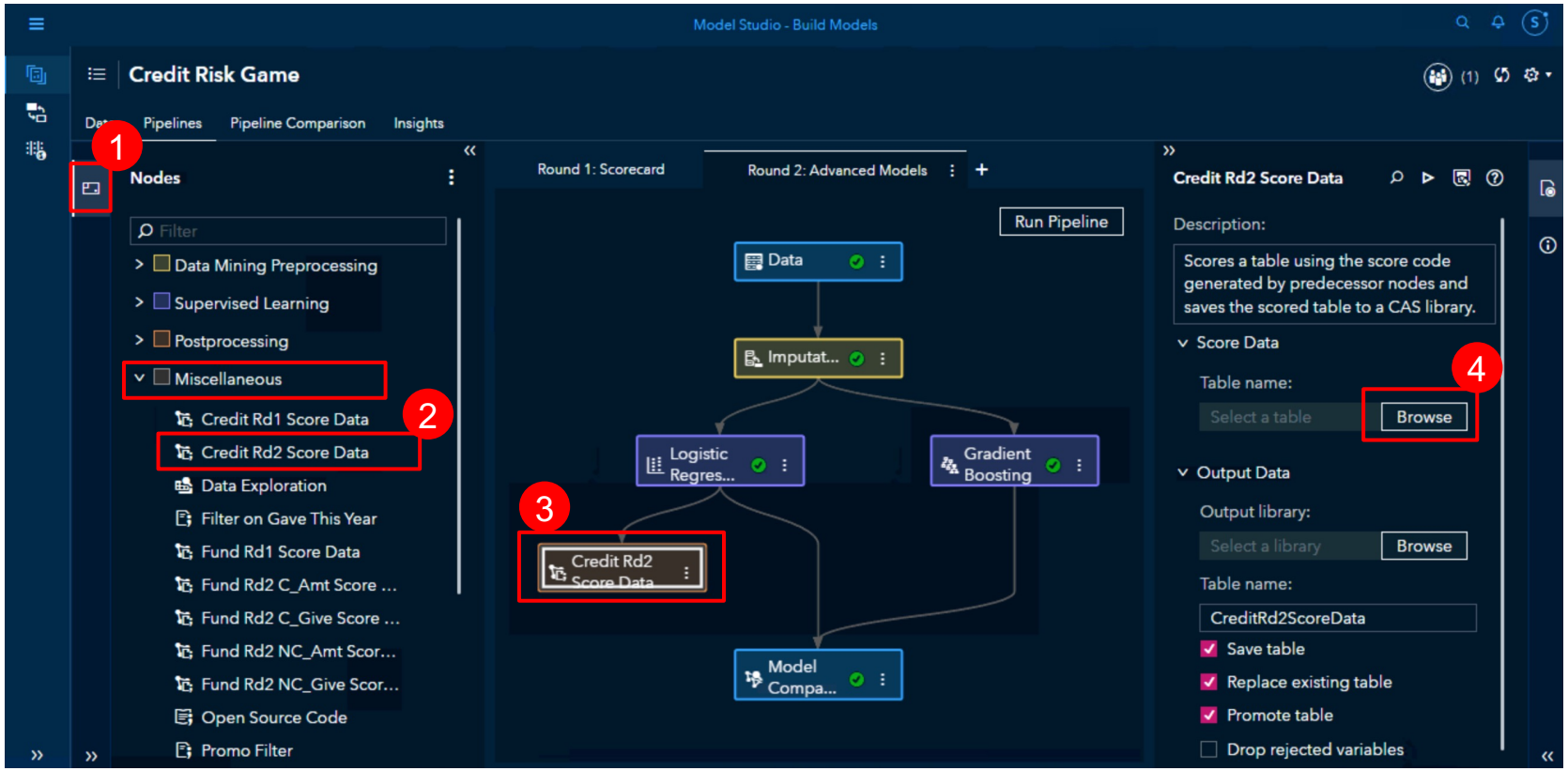
3. Under 'Insight', you can review the comparative reports for multiple pipelines and models.

© 2004-2022, ERPsim Lab, HEC Montréal.

Score Data

In this step, you will use your model to perform scoring of the 1,000,000 applicant records, assigned to you by your bank, in your capacity as the loan officer.

1. Click the 'Nodes' icon to expand the available node options
2. Under 'Miscellaneous', select 'Credit Rd2 Score Data' and release it under the champion model (here Logistic Regression)
3. Click on the node
4. Under 'Table name', click on 'Browse'



© 2004-2022, ERPsim Lab, HEC Montréal.

Load Score Data

1. Navigate to:

   Data Sources/cas-shared-default/public and select: 'CREDITRISK_SCOREDATA.sashdat'

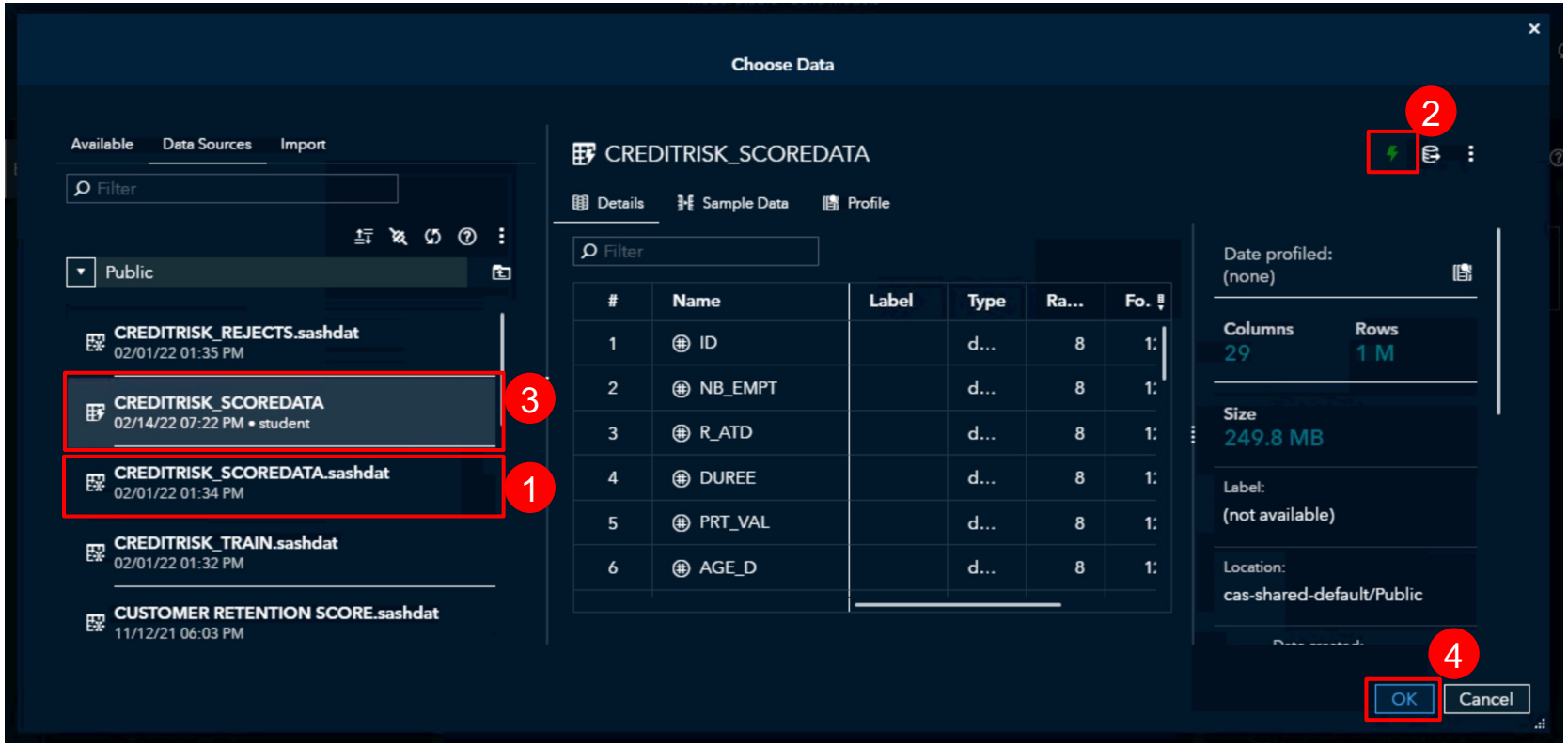
NOTE 1: If you cannot see a file or folder, use the Filter box to search

2. Click on the lightning bolt icon on the top right to load the dataset to memory

NOTE 2: The icon will turn green after data is loaded to memory

3. Select the loaded: CREDITRISK_SCOREDATA

4. Click on 'OK'



© 2004-2022, ERPsim Lab, HEC Montréal.

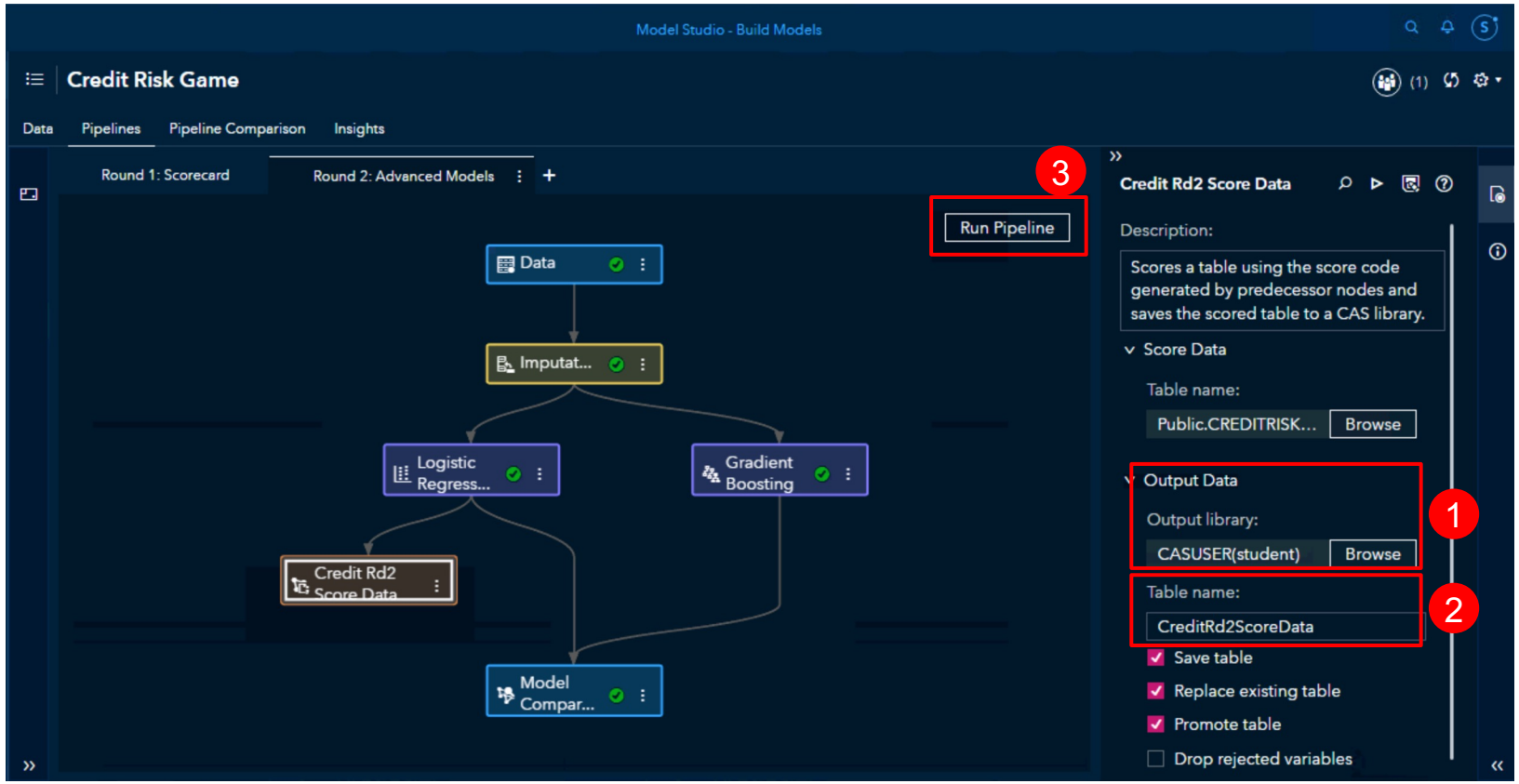
Select Output Library

1. Under ‘Output library’, click on ‘Browse’ and set the path to:

   cas-shared-default/CASUSER(student)

2. **Do not rename** the ‘Table name’

3. Run the ‘Pipeline’ again

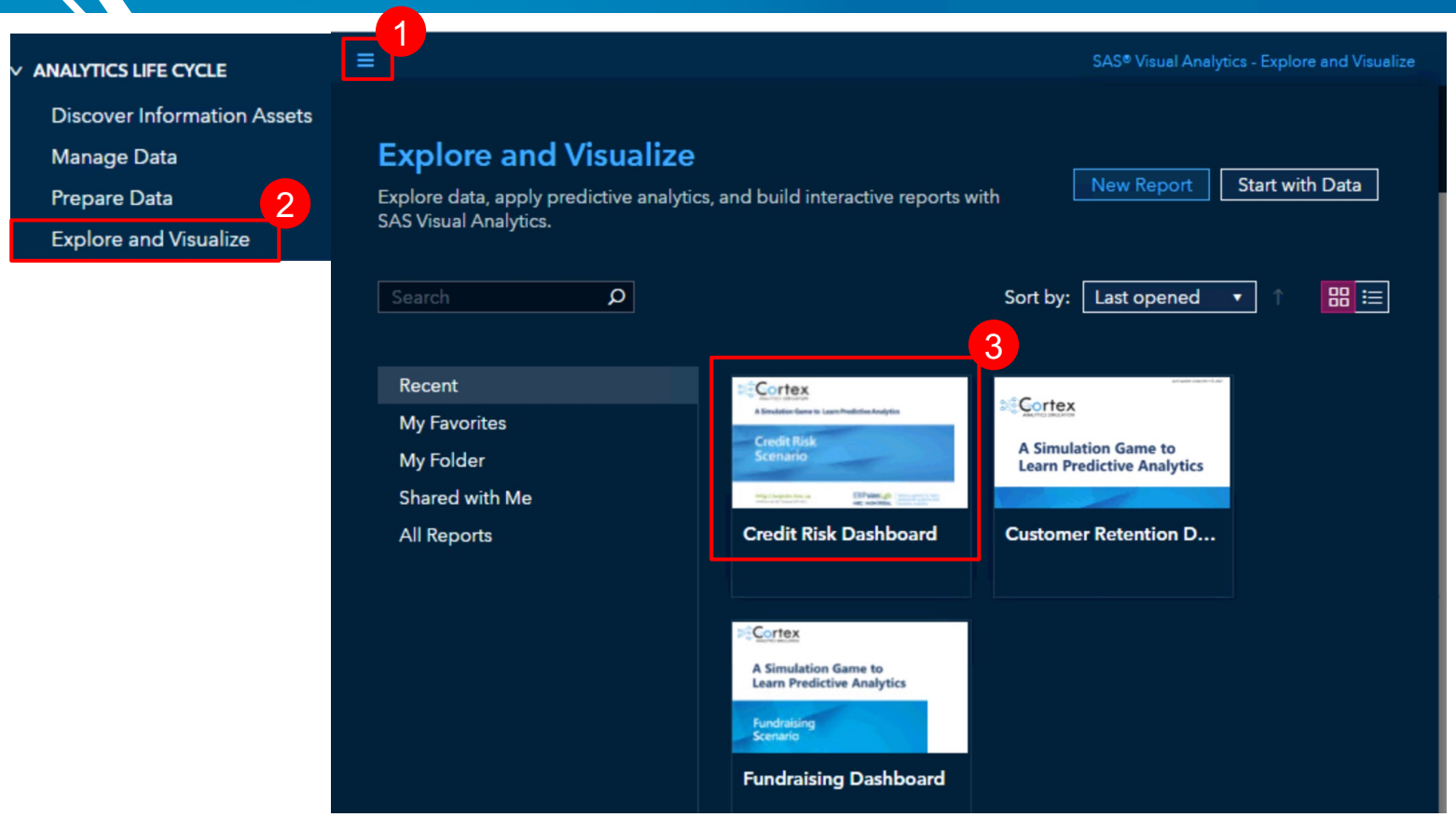


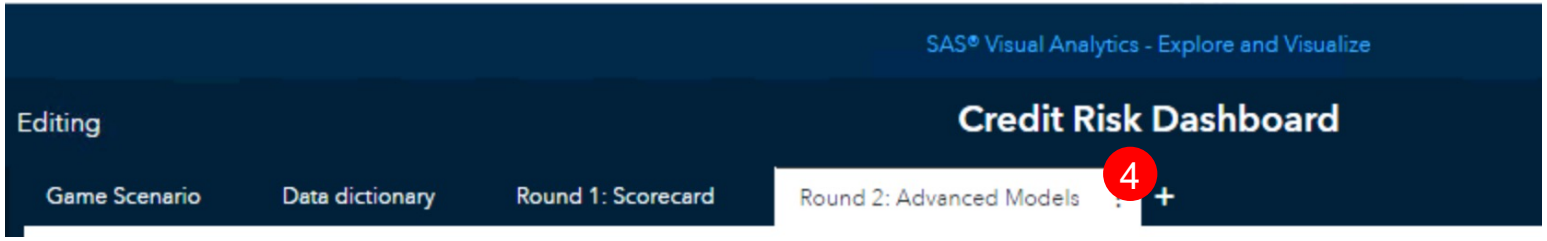
© 2004-2022, ERPsim Lab, HEC Montréal.

Reopen: Credit Risk Dashboard

Once the run is complete for the score node:

1. Click on the menu (hamburger icon) on the top left corner
2. Select 'Explore and Visualize'
3. Open the 'Credit Risk Dashboard'
4. Click on the tab corresponding to the round you are playing (e.g., 'Round 2: Advanced Models')

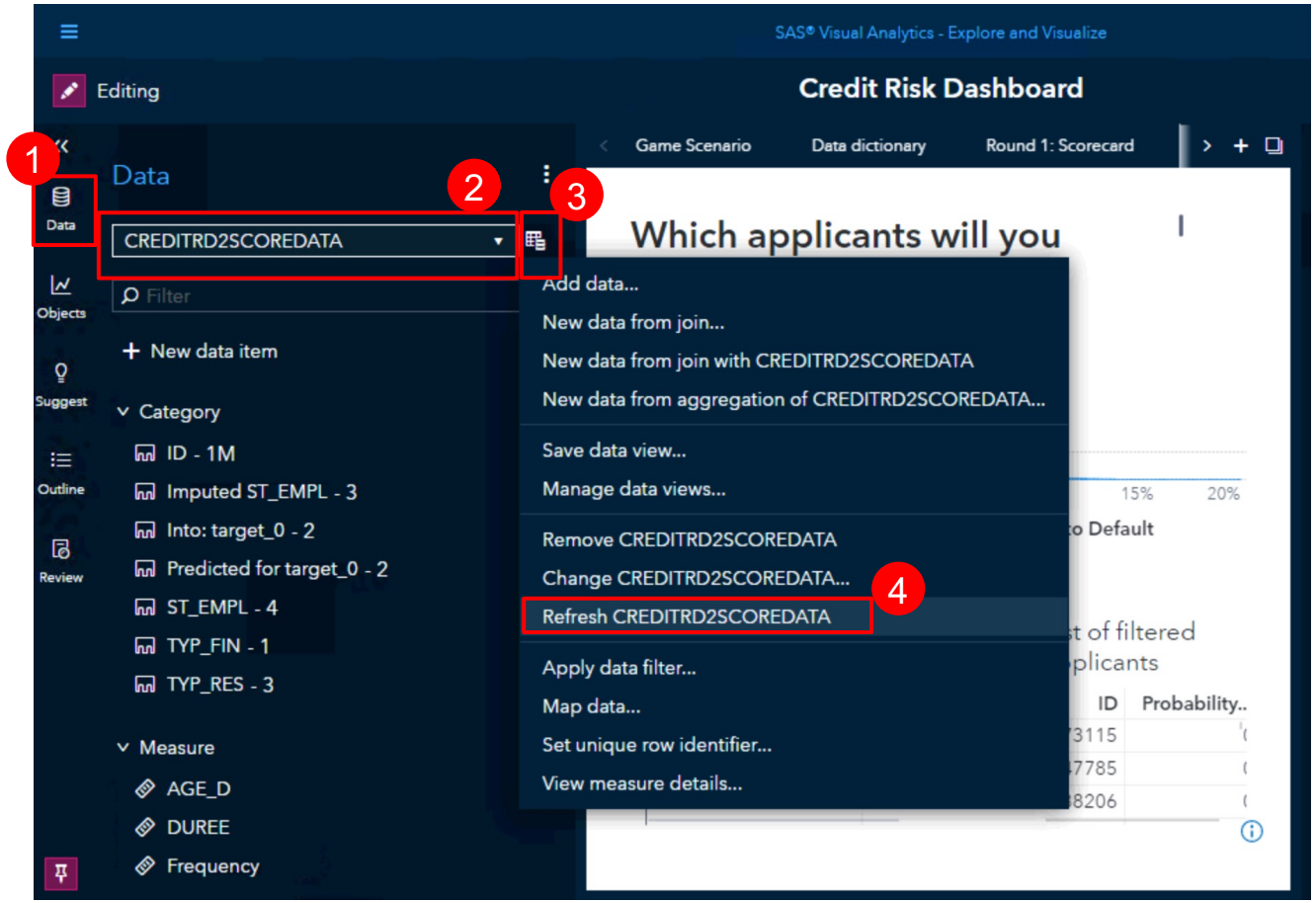




© 2004-2022, ERPsim Lab, HEC Montréal.

Refresh Data

1. Click on the 'Data' icon
2. Under 'Data' open the drop-down list and choose: 'CREDITRD2SCOREDATA'
3. Click on 'Actions'
4. Click on: 'Refresh CREDITRD2SCOREDATA'
5. Close data by re-clicking on 'Data' icon from step 1



NOTE: Every time you run a new model and come back to 'Explore and Visualize', make sure you refresh your data.

© 2004-2022, ERPsim Lab, HEC Montréal.

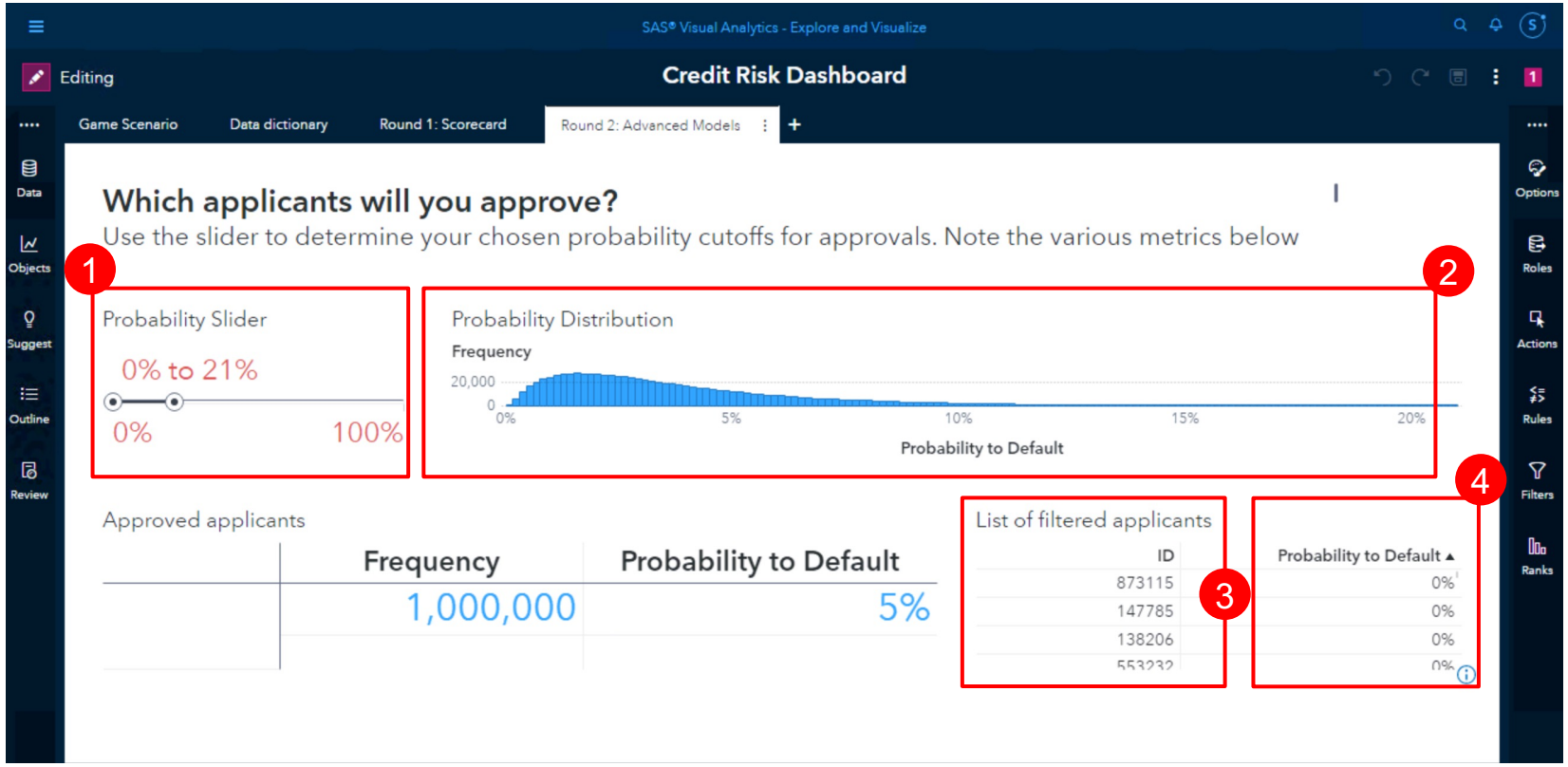
Select Applicants

1. Use the slider to detemine the applicants you'd like to approve based on their probability to default.

2. As you change this slider, note that the 'Probability Distribution' will adjust accordingly.

3. Under the 'List of filtered applicants', you will be able to select the individuals you approve for the loan.

4. Right-click on 'Probability to Default' column



© 2004-2022, ERPsim Lab, HEC Montréal.

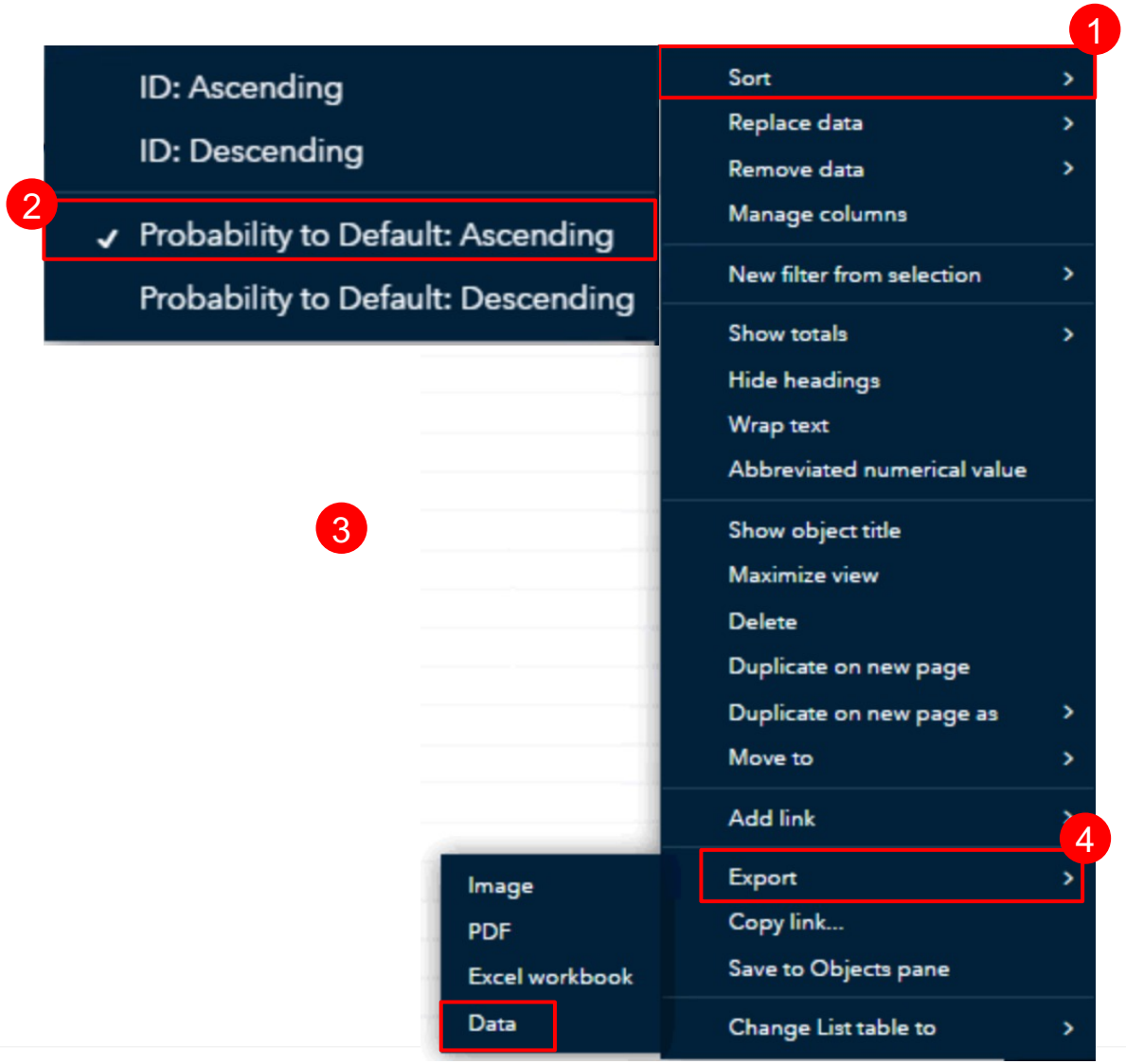
Sort and Export

1. Click on ‘Sort’

2. Choose: ‘Probability to Default: Ascending’

3. Right-click on any column

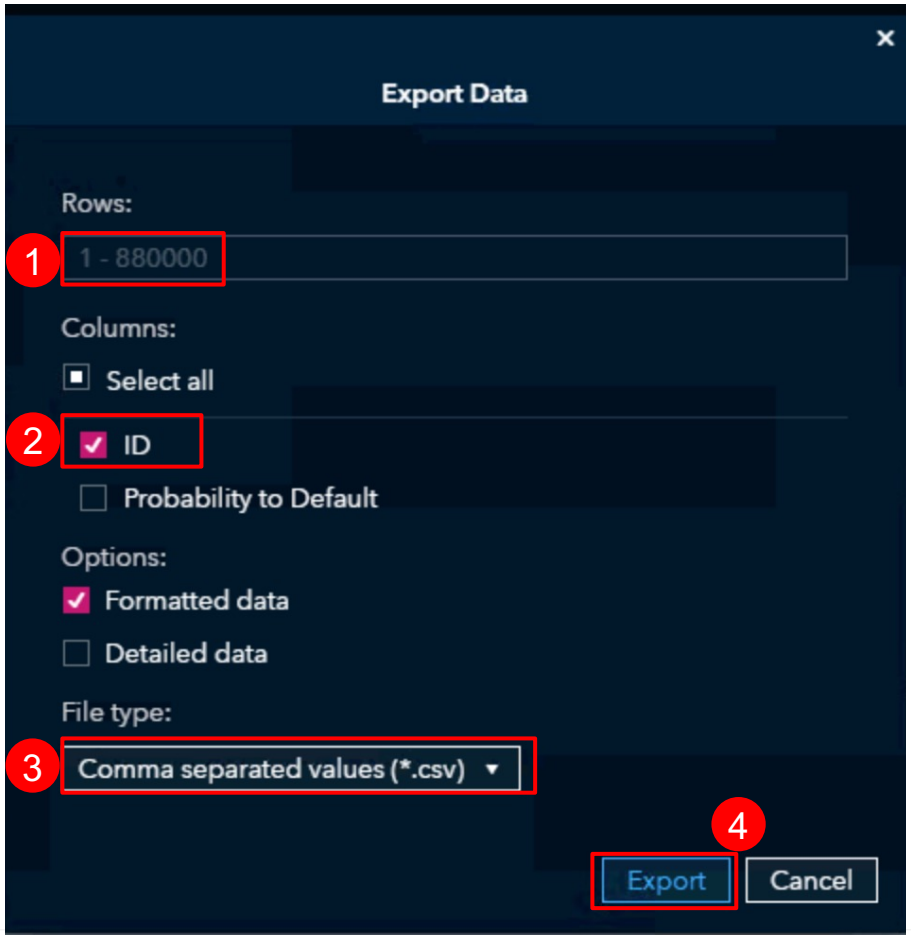
4. Click on ‘Export’ and ‘Data’



© 2004-2022, ERPsim Lab, HEC Montréal.

Export to .CSV

1. You can see/modify the rows being exported
2. Under column, select only 'ID'
3. Change the 'File type' to: Comma separated values (*.csv)
4. Click on 'Export'



NOTE: Your .csv file will be in the Downloads folder of the Virtual Machine.

© 2004-2022, ERPsim Lab, HEC Montréal.

You are now ready to upload your solution for Round 2 to the Game Leaderboard!

© 2004-2022, ERPsim Lab, HEC Montréal.