



Credit Risk Scenario – Round 1

SAS Viya 4.0

Last Update: February 23, 2022

ERPsimLab
HEC MONTREAL

Serious games to learn enterprise
systems and business analytics

Table of Contents

- [**Overview and Access SAS Viya**](#)
- [**Explore the Game**](#)
- [**Play Round 1: Scorecard Node**](#)
- [**Score New Data**](#)
- [**Explore/Visualize/Export Results**](#)
- [**Add Rejected Applicants**](#)

Round 1 Overview

Your goal is to develop a credit scorecard using SAS Viya's Risk Modelling nodes to assess and approve loan applicants.

You will refine your model(s) so that you can maximize net profit for your financial institution.

You will then add a dataset of rejected applicants to address the issue of sample selection bias.

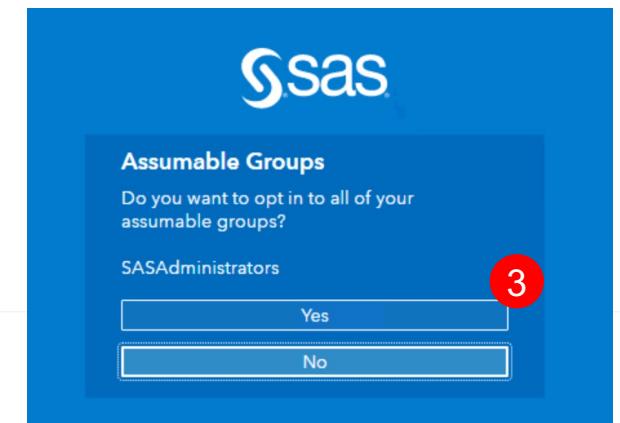
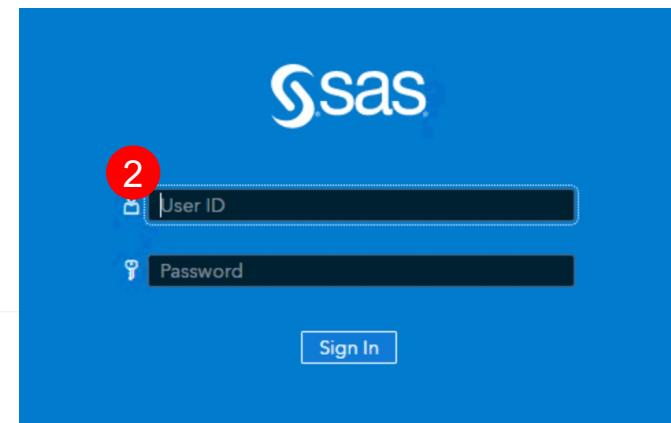
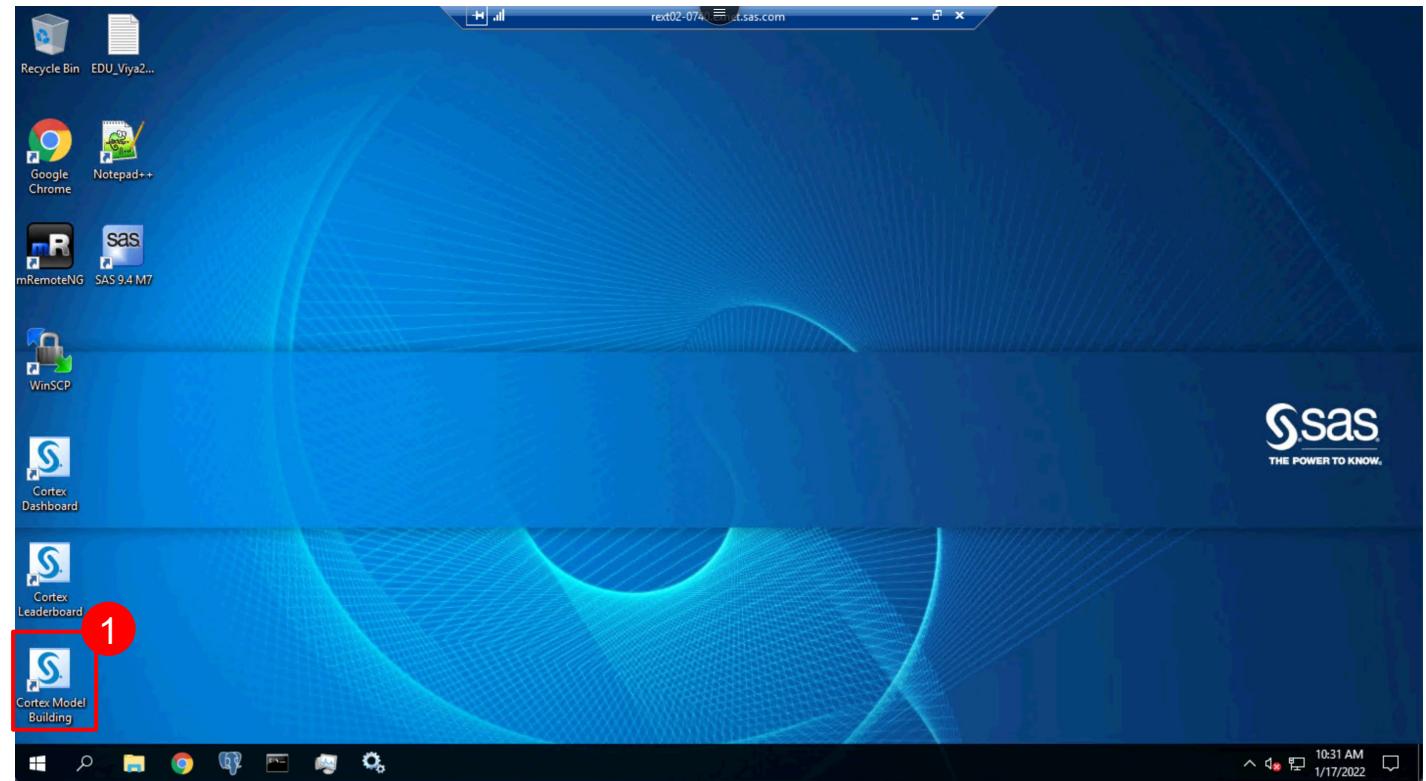


Access SAS Viya

1. On your virtual desktop, click on 'Cortex Model Building' icon
2. Enter your User ID and password:
 - User ID: [student](#)
 - Password: [Metadata0](#)

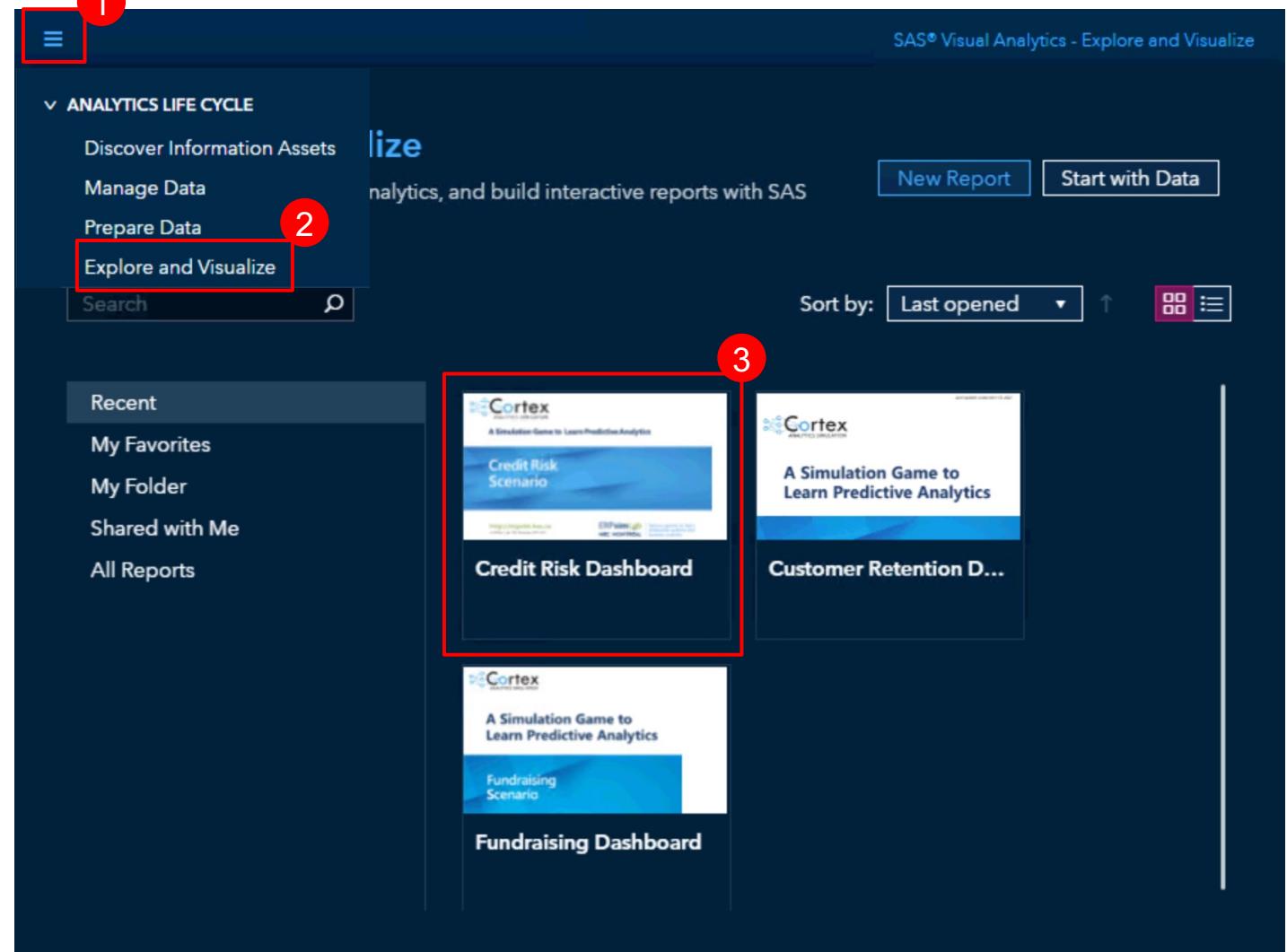
NOTE: Use the above credentials unless otherwise specified by your instructor

3. Select 'Yes' when asked about 'Assumable Groups'



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'



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

Credit Risk Dashboard

Editing

Game Scenario Data dictionary Round 1: Scorecard Round 2: Advanced Models +

Cortex
ANALYTICS SIMULATION

A Simulation Game to Learn Predictive Analytics

Credit Risk Scenario

<http://erpsim.hec.ca> **ERPsimLab** Serious games to learn enterprise systems and business analytics. © HEC Montréal, 2004-2021

Credit risk game scenario

Accurately evaluating loan applicants' credit worthiness is an essential procedure for banks, which need to strike a balance between maximizing interest revenue and minimizing losses from loan defaults. In this scenario, you are a loans officer working in a large bank and have been tasked with developing a predictive model based on data science that will help the bank optimize revenues by determining which applications should be approved or rejected.

To accomplish this task, you have been provided with a training file containing one million historical customer loan records including demographics (e.g.: age, employment status, financial indicators (e.g.: income, assets, liabilities, savings) and credit-related metrics (e.g.: missed payments, loan-seeking behaviour, late payments), type of funding requested (car, mortgage or credit), and whether they defaulted (or not) on their loan.

You will use this information (and SAS Viya!) to develop a model to predict which applicants should be granted a loan. Once you are satisfied with your model, you will export the list of approved

...

1

...

Options

Roles

Actions

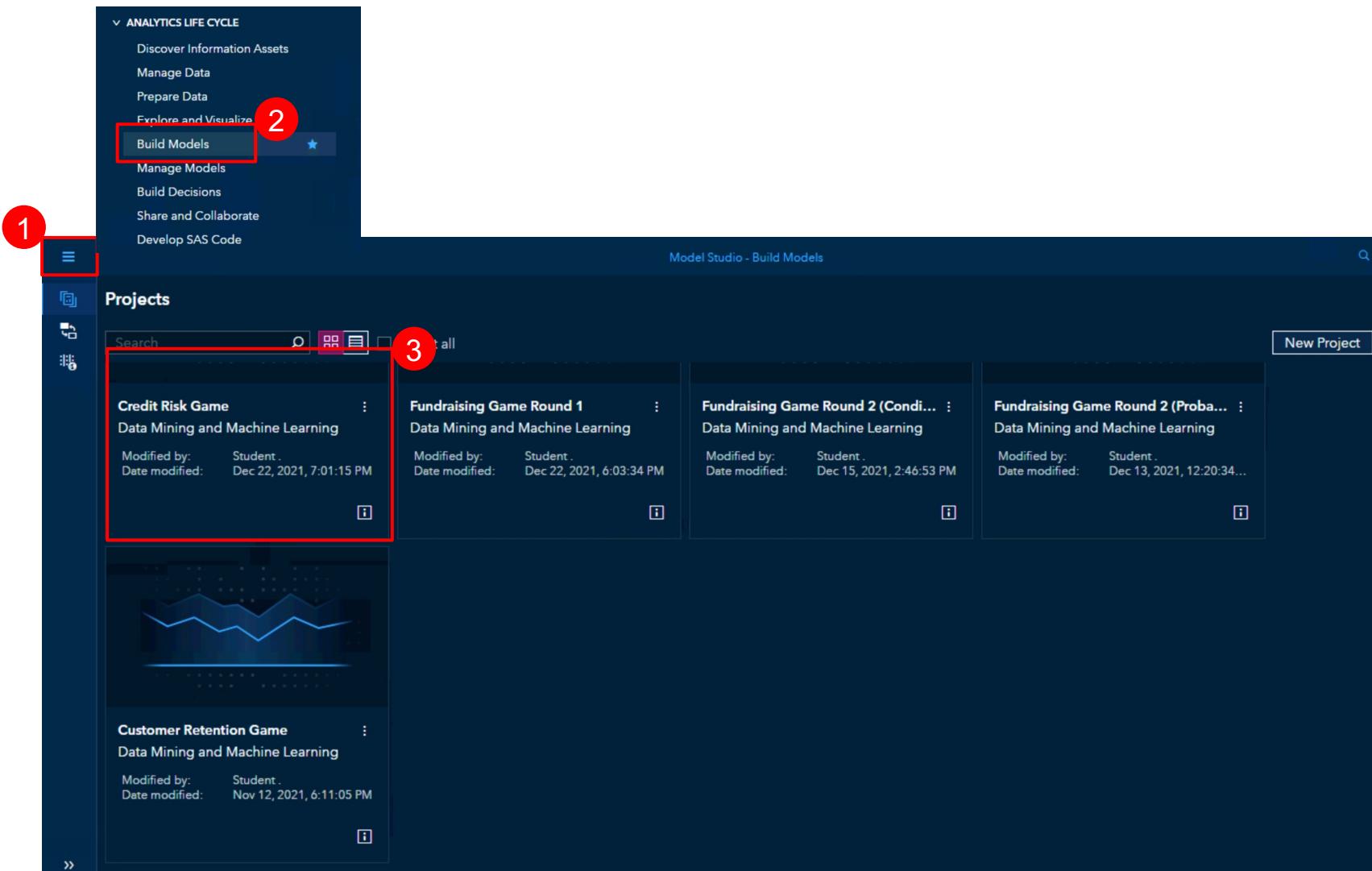
Rules

Filters

Ranks

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



Project Settings

To view or change the default data partitions (train, validate, test):

1. Click on the 'Project Settings'
2. Select 'Partition Data'
3. View/change the percentage allocated to each set (e.g., 80-20-0)

NOTE 1: There is no need to have a test set since you will have access to another data set (i.e., the game leaderboard) that will score your performance and give you an understanding on how your model will perform in real world. You can keep a test set here, but you will have less data to train and validate your models.

4. If made any changes, click **Save**

NOTE 2: Changing the project settings will be applied to **all the pipelines** of the project.

Model Studio - Build Models

Credit Risk Game

Data Pipelines Pipeline Comparison Insights

Variable Name Label Type

CREDITRISK_TRAIN

Project Settings

Batch API Project logs

Project Settings

Partition Data

Event-Based Sampling

Node Configuration

Rules

Output Library

Logging

Create partition variable

Note: These settings are active only when a partition variable is not set within the data. Using a data source with a pre-defined partition variable or manually selecting a partition variable will override these settings.

Method:

Stratify

Training: 80 80.00%

Validation: 20 20.00%

Test: 0 0.00%

Save Cancel

Assign Variable Roles & Levels

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

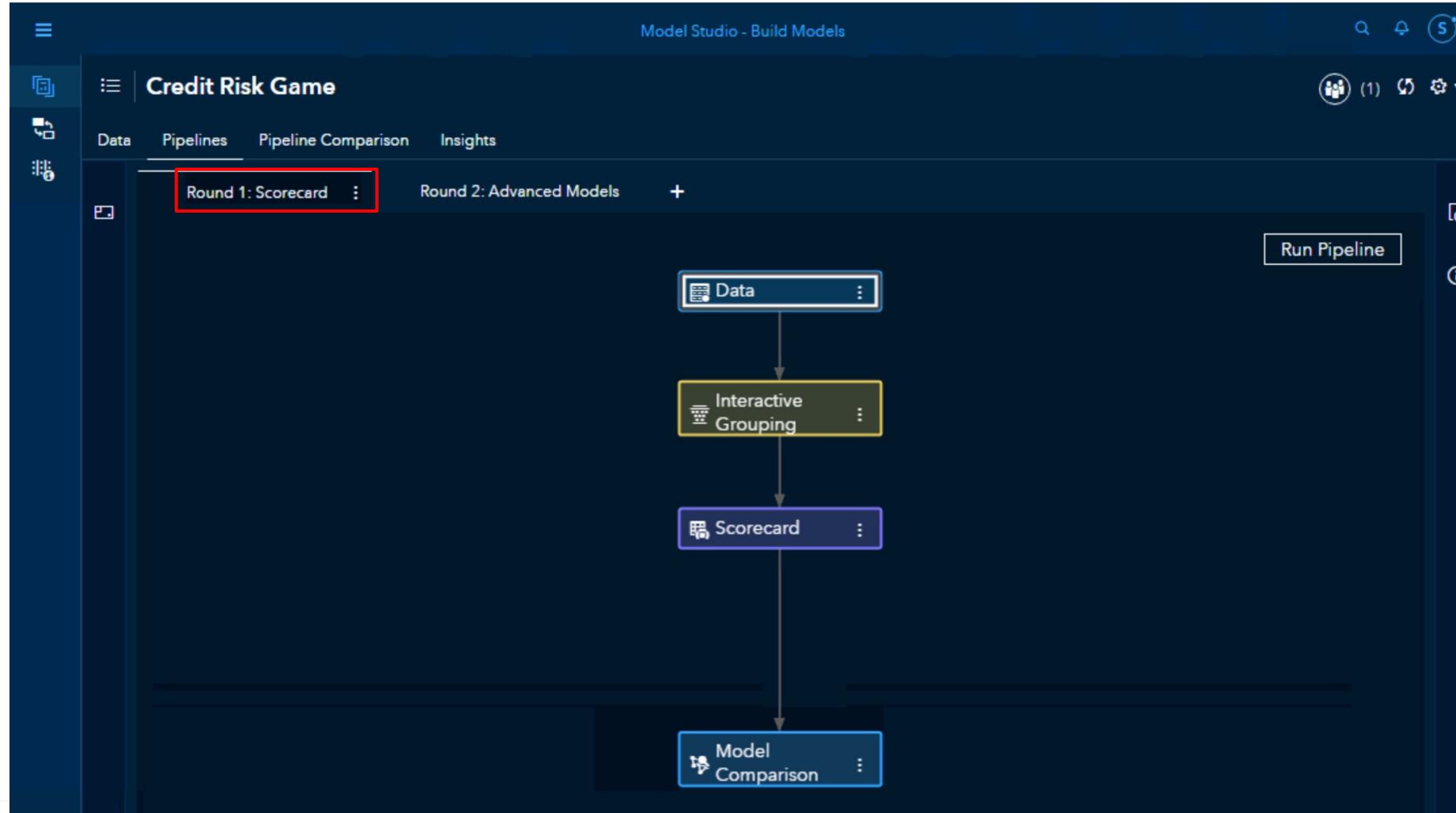
NOTE 1: Remember to [deselect](#) a variable before selecting the next when modifying them individually

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

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

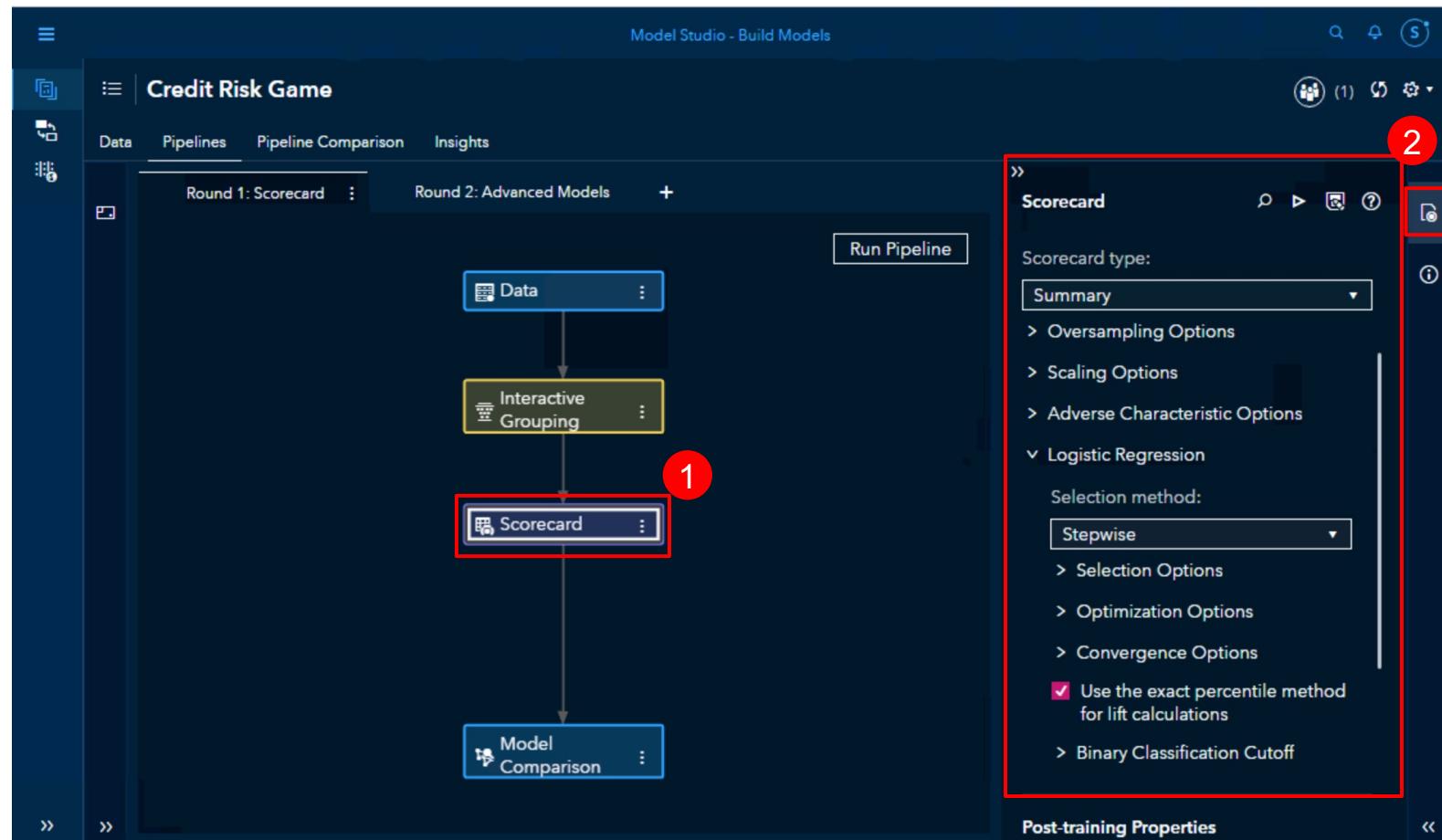
Variable Name	Label	Type	Role
AGE_D		Numeric	Input
DUREE		Numeric	Input
ID		Numeric	ID
MNT_ACT		Numeric	Input
MNT_AUT_REN		Numeric	Input
MNT_DEMANDE		Numeric	Input
MNT_EPAR		Numeric	Input
MNT_PASS		Numeric	Input
MNT_UTIL_REN		Numeric	Input
NB_COUR		Numeric	Input
NB_DEC_12MS		Numeric	Input
NB_DEL_30		Numeric	Input
NB_DEL_60		Numeric	Input

Click on the tabs to view pre-built pipelines for each round (e.g., this is the pipeline for "Round 1: Scorecard")



Improve Pipeline

1. Click on a node to see options on the right-hand side
2. If not open by default, click on 'Node Options' to open it:
 - You should try to improve the properties of the node

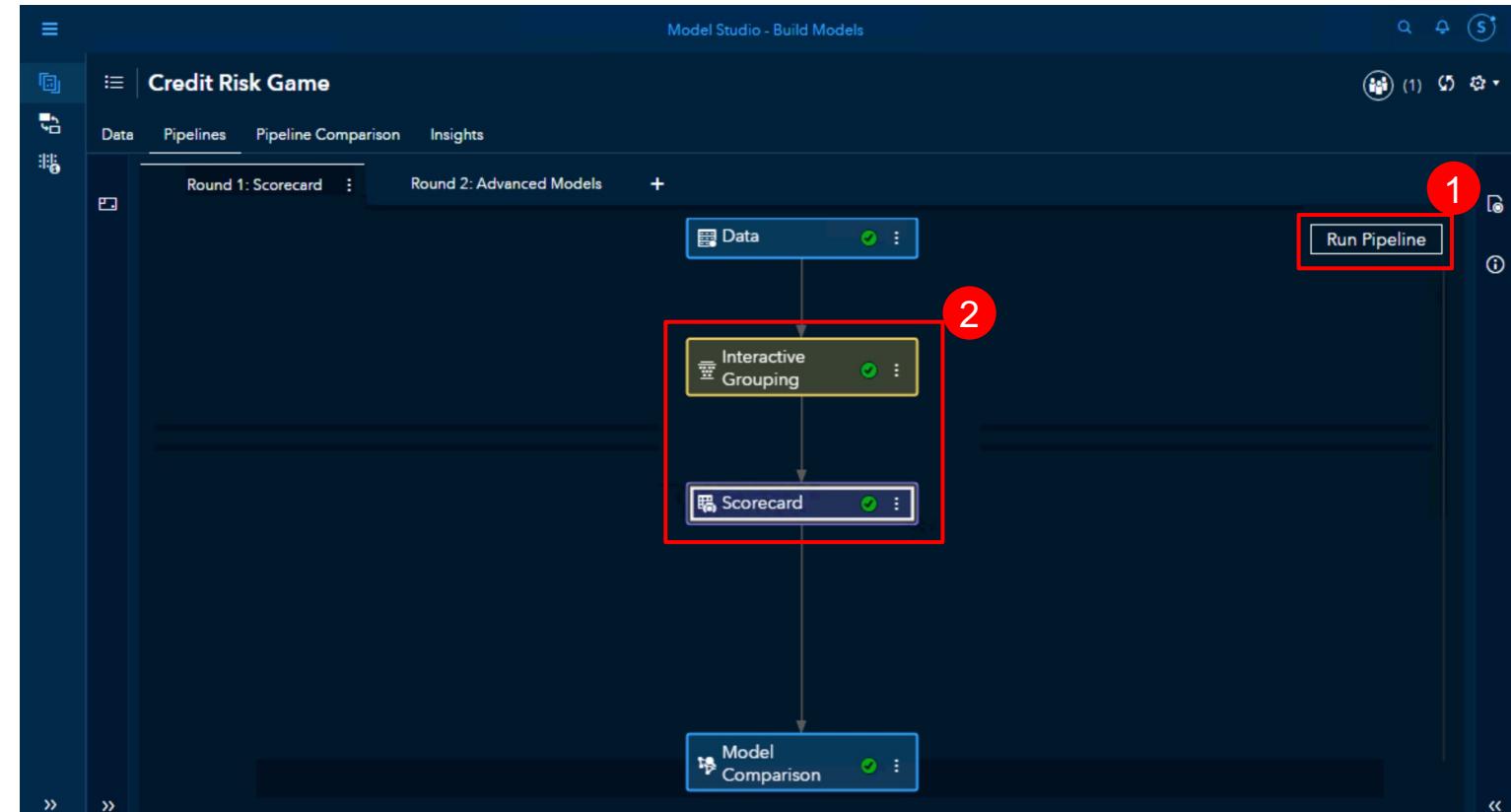


Run Pipeline

1. Click on 'Run Pipeline'

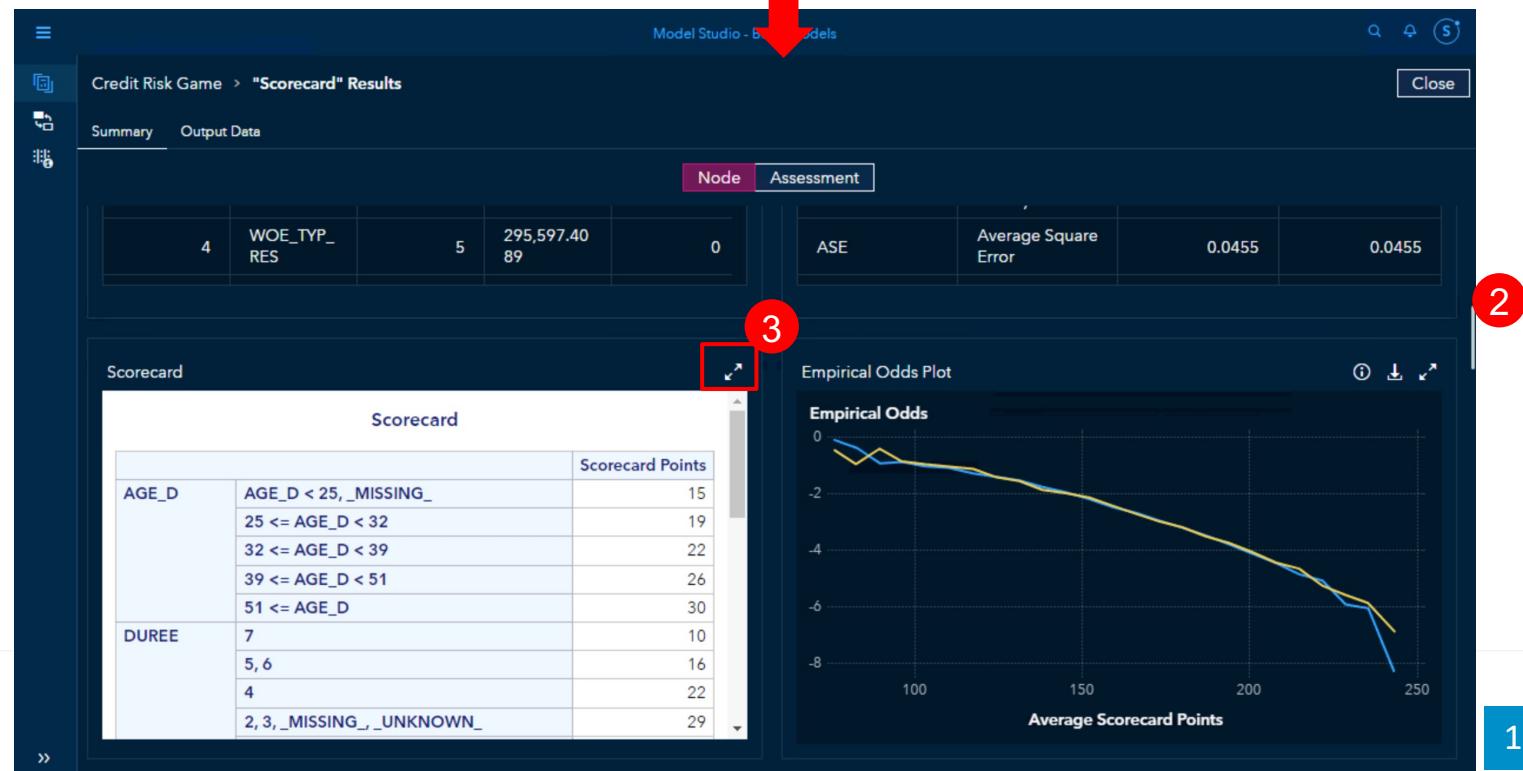
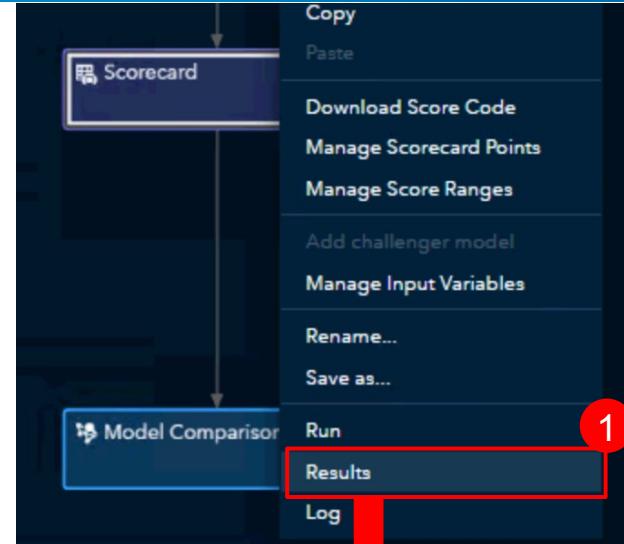
NOTE: Green check marks appear on nodes when the run is complete.

2. Once run is complete,
explore the results of the
'Interactive Grouping' and
'Scorecard'



Explore Results: Scorecard

1. Right-click on the 'Scorecard' node and select 'Results'
2. Scroll down a bit to find the third window on the left-hand side (Scorecard window)
3. Expand the window



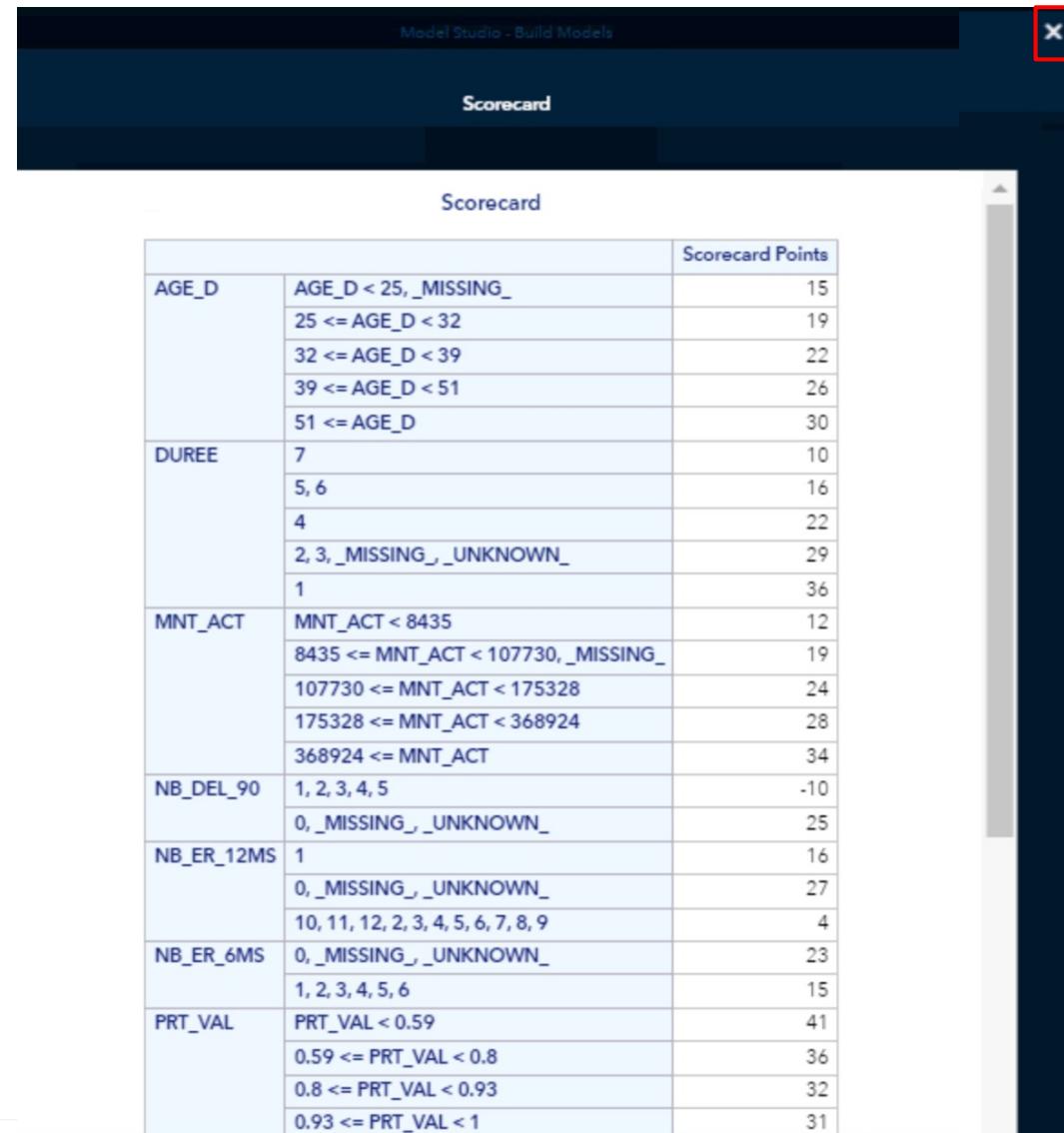
Explore Results: Scorecard (Cont.)

This is where you can see how the model has assigned points to each category.

Since this is your first run, take some time to look at these and other results.

You can return to the node again to change its parameters.

Click on 'Close' (top right) to return to your results page.

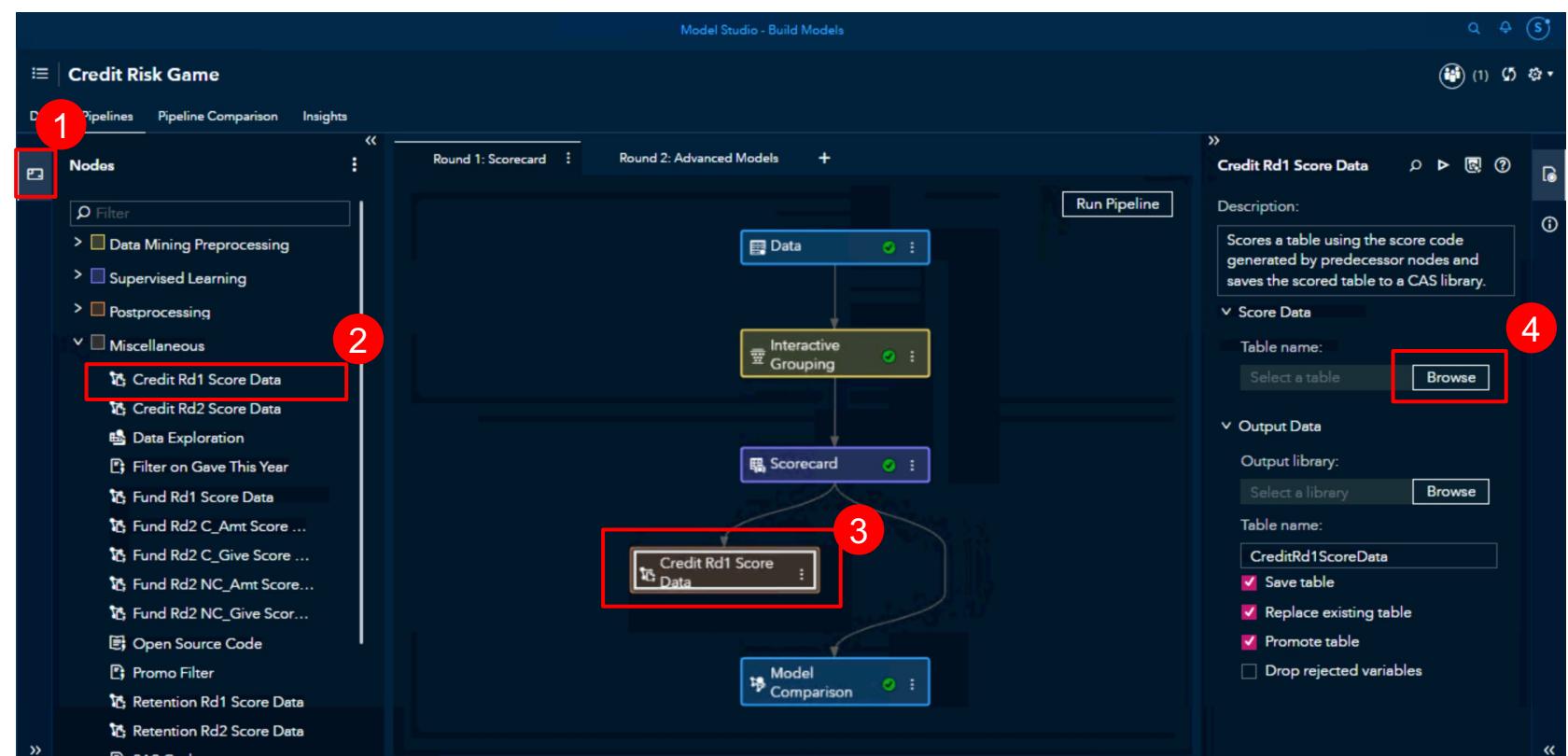


Scorecard		
		Scorecard Points
AGE_D	AGE_D < 25, _MISSING_	15
	25 <= AGE_D < 32	19
	32 <= AGE_D < 39	22
	39 <= AGE_D < 51	26
	51 <= AGE_D	30
DUREE	7	10
	5, 6	16
	4	22
	2, 3, _MISSING_, _UNKNOWN_	29
	1	36
MNT_ACT	MNT_ACT < 8435	12
	8435 <= MNT_ACT < 107730, _MISSING_	19
	107730 <= MNT_ACT < 175328	24
	175328 <= MNT_ACT < 368924	28
	368924 <= MNT_ACT	34
NB_DEL_90	1, 2, 3, 4, 5	-10
	0, _MISSING_, _UNKNOWN_	25
NB_ER_12MS	1	16
	0, _MISSING_, _UNKNOWN_	27
	10, 11, 12, 2, 3, 4, 5, 6, 7, 8, 9	4
NB_ER_6MS	0, _MISSING_, _UNKNOWN_	23
	1, 2, 3, 4, 5, 6	15
PRT_VAL	PRT_VAL < 0.59	41
	0.59 <= PRT_VAL < 0.8	36
	0.8 <= PRT_VAL < 0.93	32
	0.93 <= PRT_VAL < 1	31

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 Rd1 Score Data' and drag-and-drop it onto/under the 'Scorecard' node
3. Click on the node
4. Under 'Table name', click on 'Browse'



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

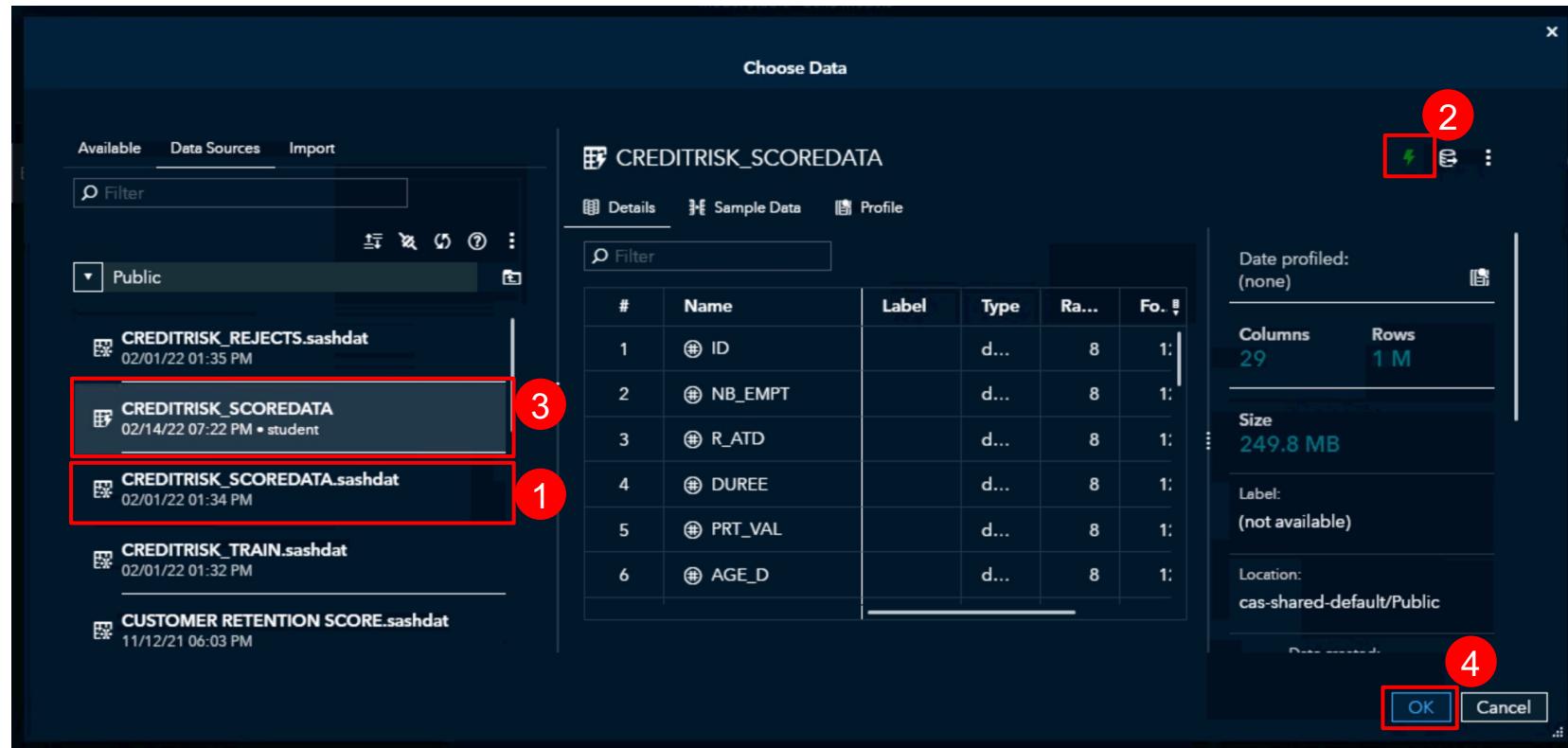
2. Click on the lightning bolt icon on the top right to load the dataset to memory

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

3. Select the loaded:

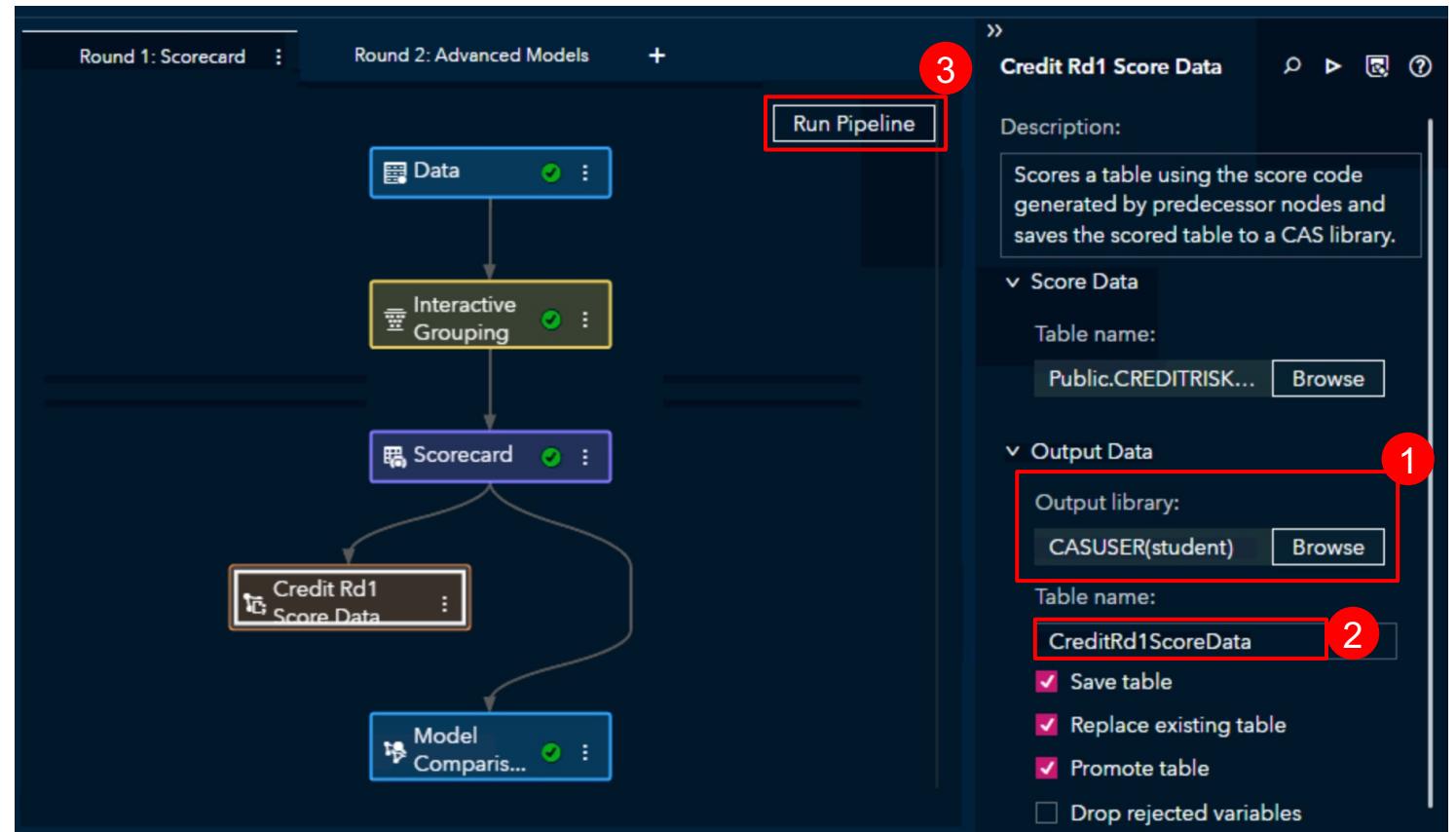
CREDITRISK_SCOREDATA

4. Click on 'OK'



Select Output Library

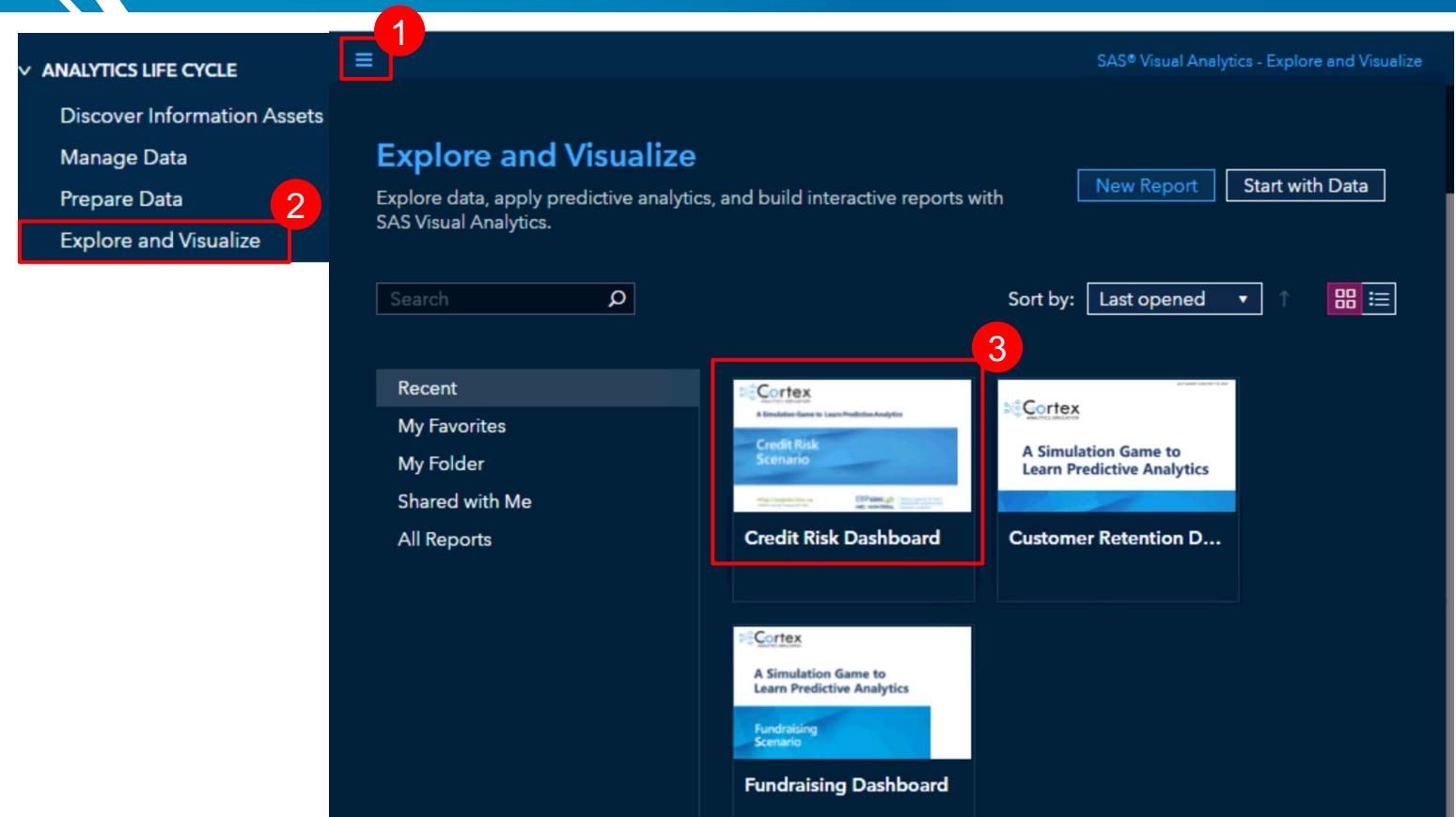
1. Under 'Output library', click on 'Browse' and set the path to: cas-shared-default/CASUSER(student)
2. Do not rename the 'Table name'
3. Run the 'Pipeline' again



Reopen: Credit Risk Dashboard

Once the run is complete for the score node:

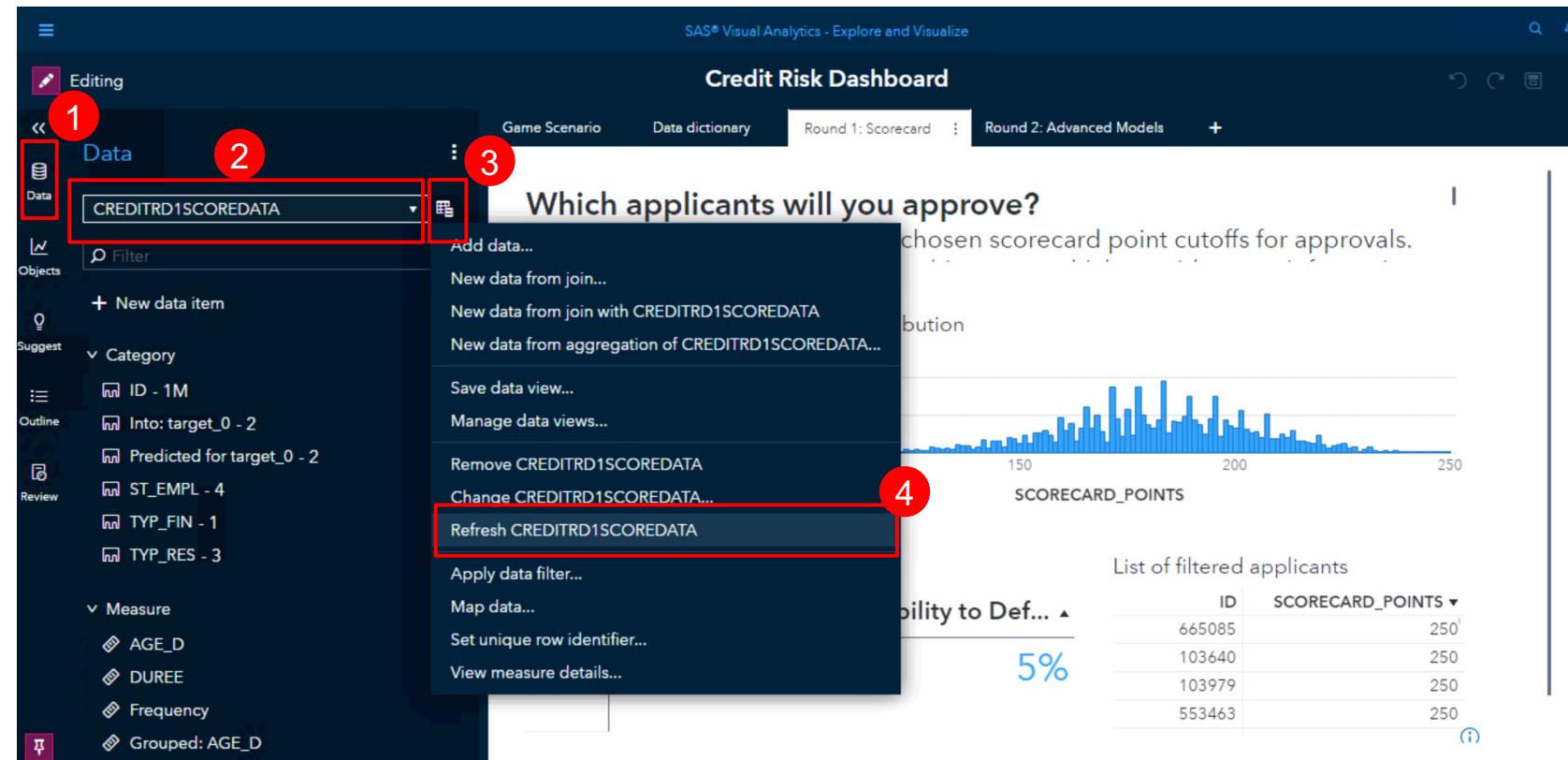
1. Click on the [menu](#) (hamburger icon) on the top left corner
2. Select '[Explore and Visualize](#)'
3. Open the '[Credit Risk Dashboard](#)'
4. Click on the tab corresponding to the round you are playing (e.g., '[Round 1: Scorecard](#)')



Refresh Data

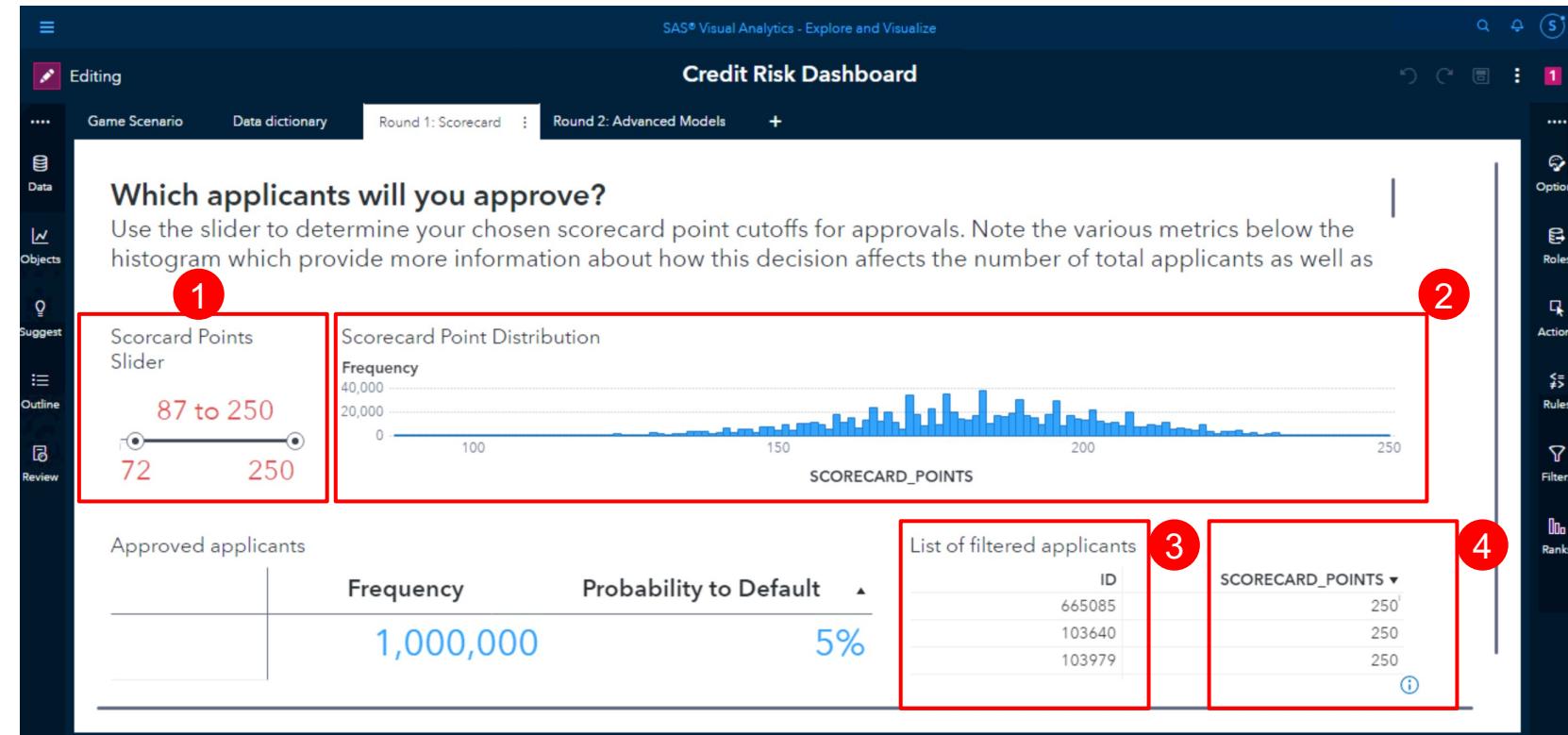
1. Click on the 'Data' icon
2. Under 'Data' open the drop-down list and choose: **CREDITRD1SCOREDATA**
3. Click on 'Actions'
4. Click on: 'Refresh CREDITRD1SCOREDATA'
5. Close data by re-clicking on 'Data' icon from step 1

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



Select Applicants

1. Use the [slider](#) to determine the applicants you'd like to approve based on their scorecard scores.
2. As you change this slider, note that the ['Scorecard point Distribution'](#) (frequency of approved applicants and potential defaults) 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 ['SCORECARD_POINTS'](#) column



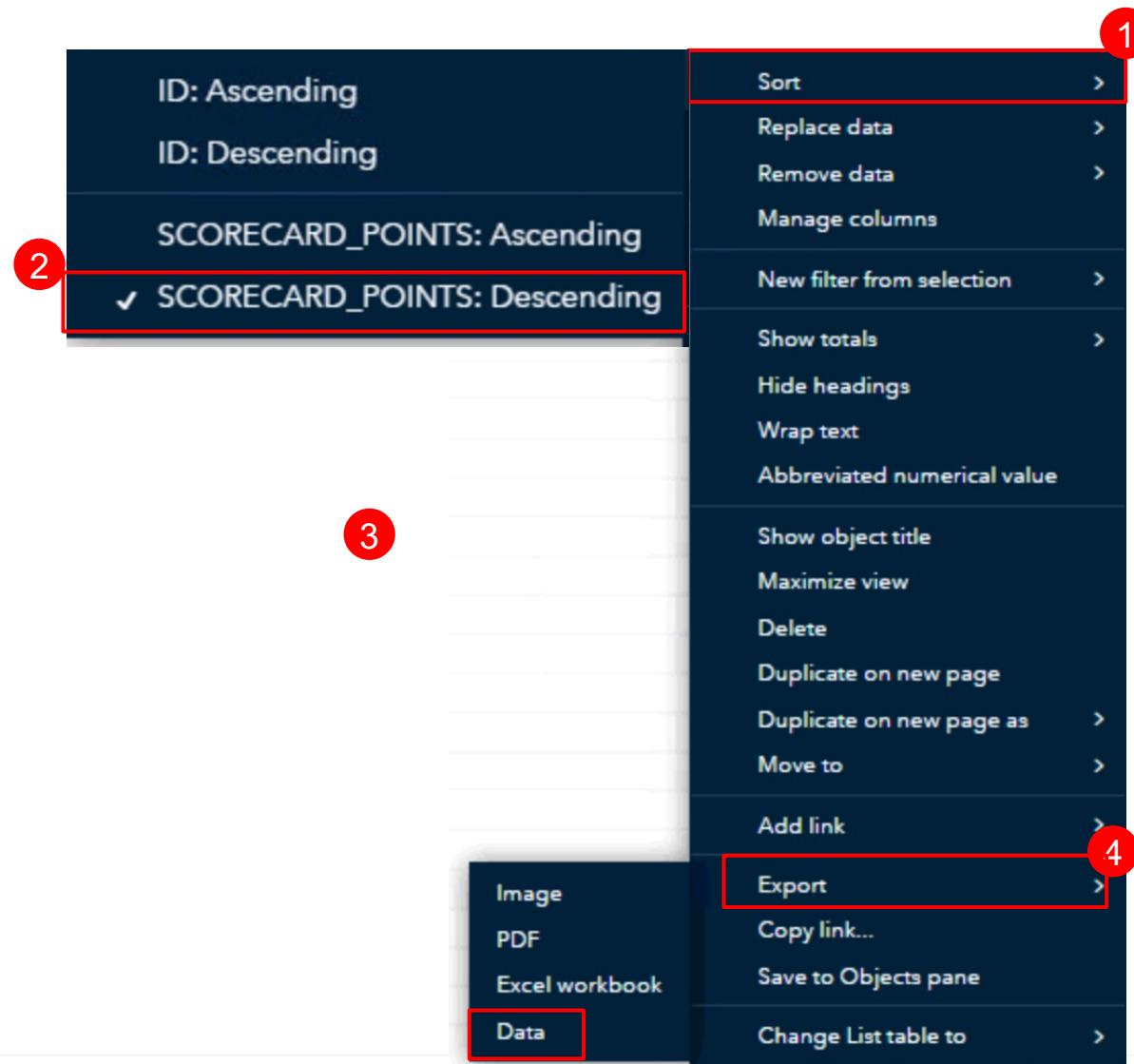
Sort and Export

1. Click on 'Sort'

2. Choose: 'SCORECARD_POINTS:
Descending'

3. Right-click on any column

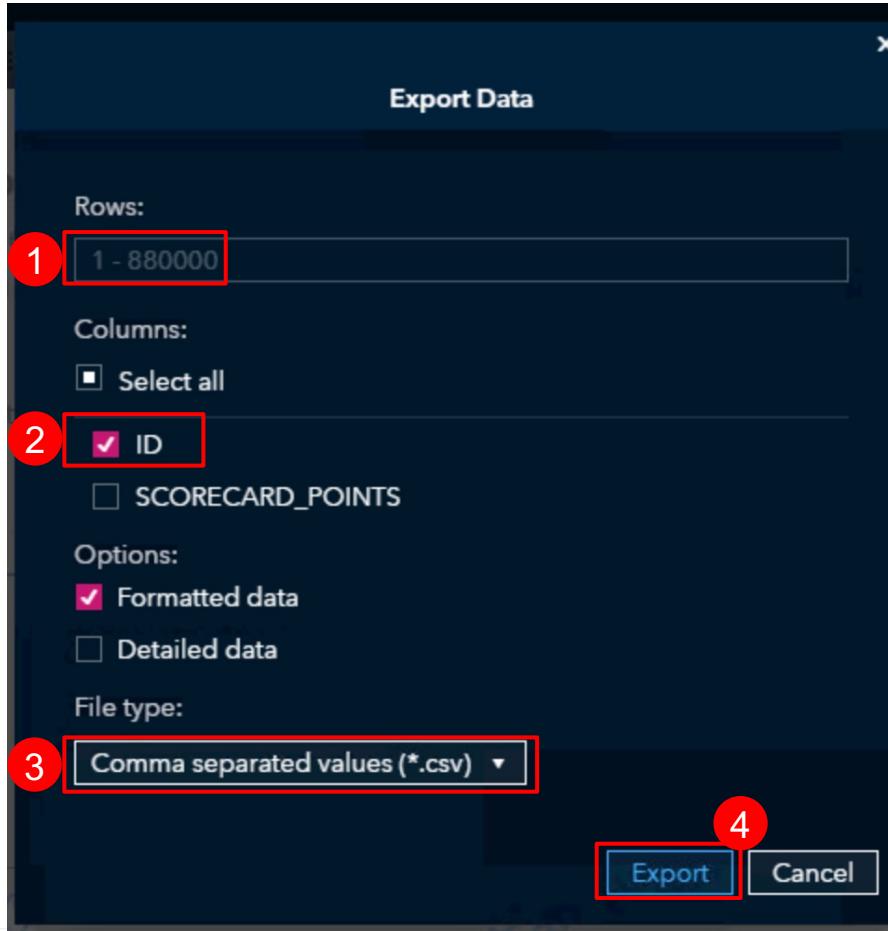
4. Click on 'Export' and 'Data'



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.





**You are now ready to upload
your solution (csv file)
for Round 1 to
the game leaderboard!**

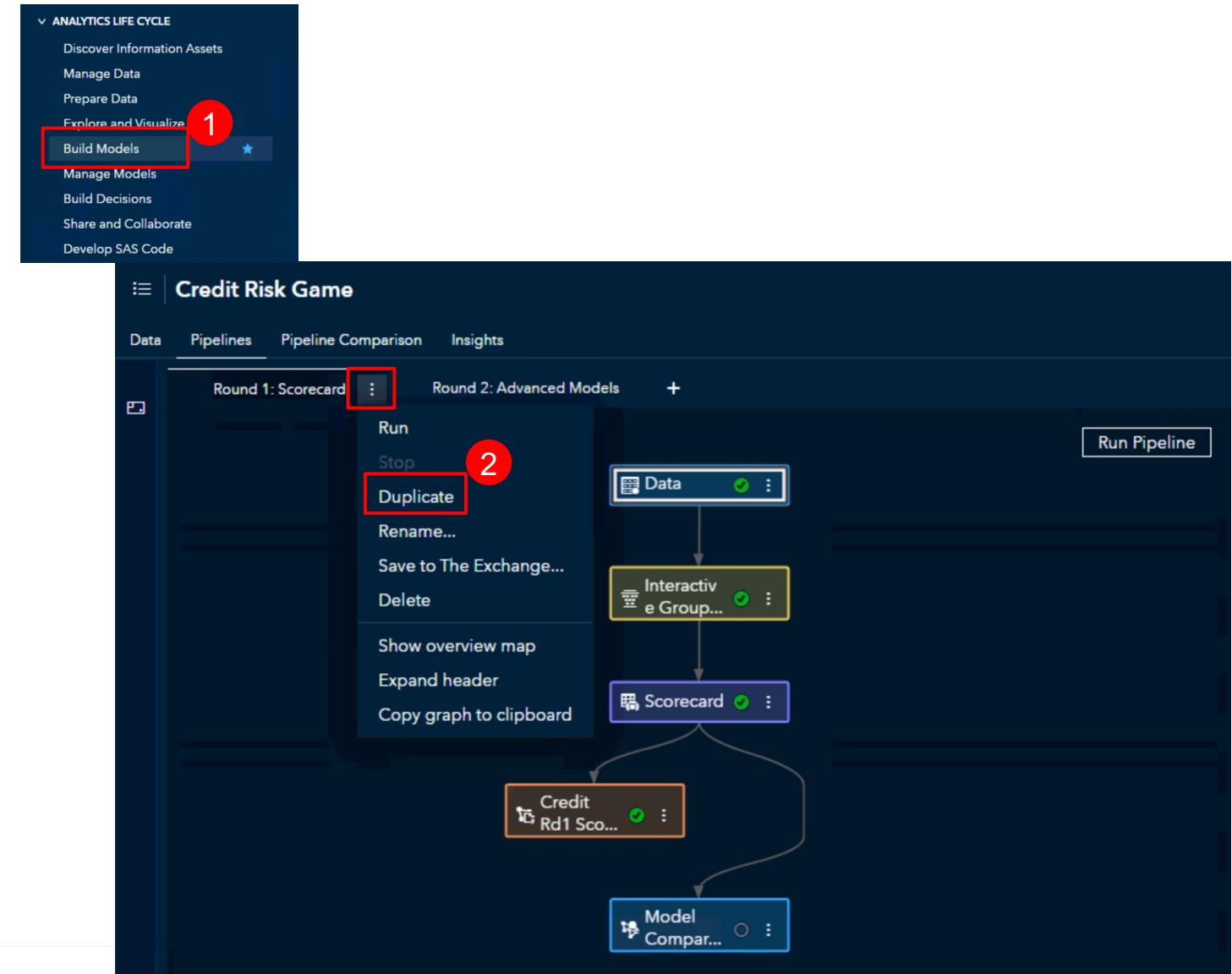
Round 1

- **Add Dataset of Rejected Applicants**

Duplicate and Rename Pipeline

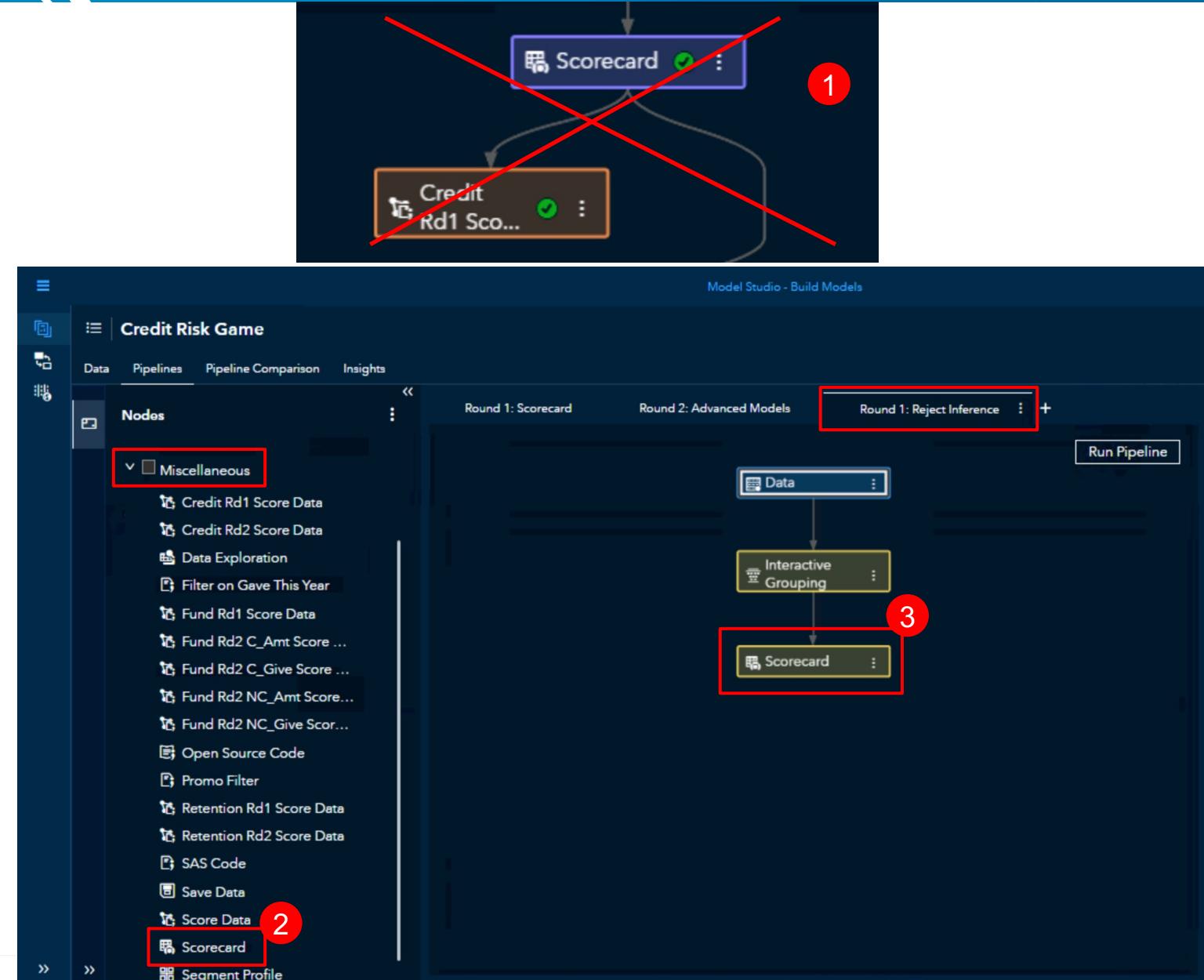
1. Return to 'Build Models' area
(click on the main menu or the hamburger icon)
2. Duplicate your pipeline to
keep the original one if you
wish

NOTE: You can rename your new pipeline to
'Round 1: Reject Inference' or another name.



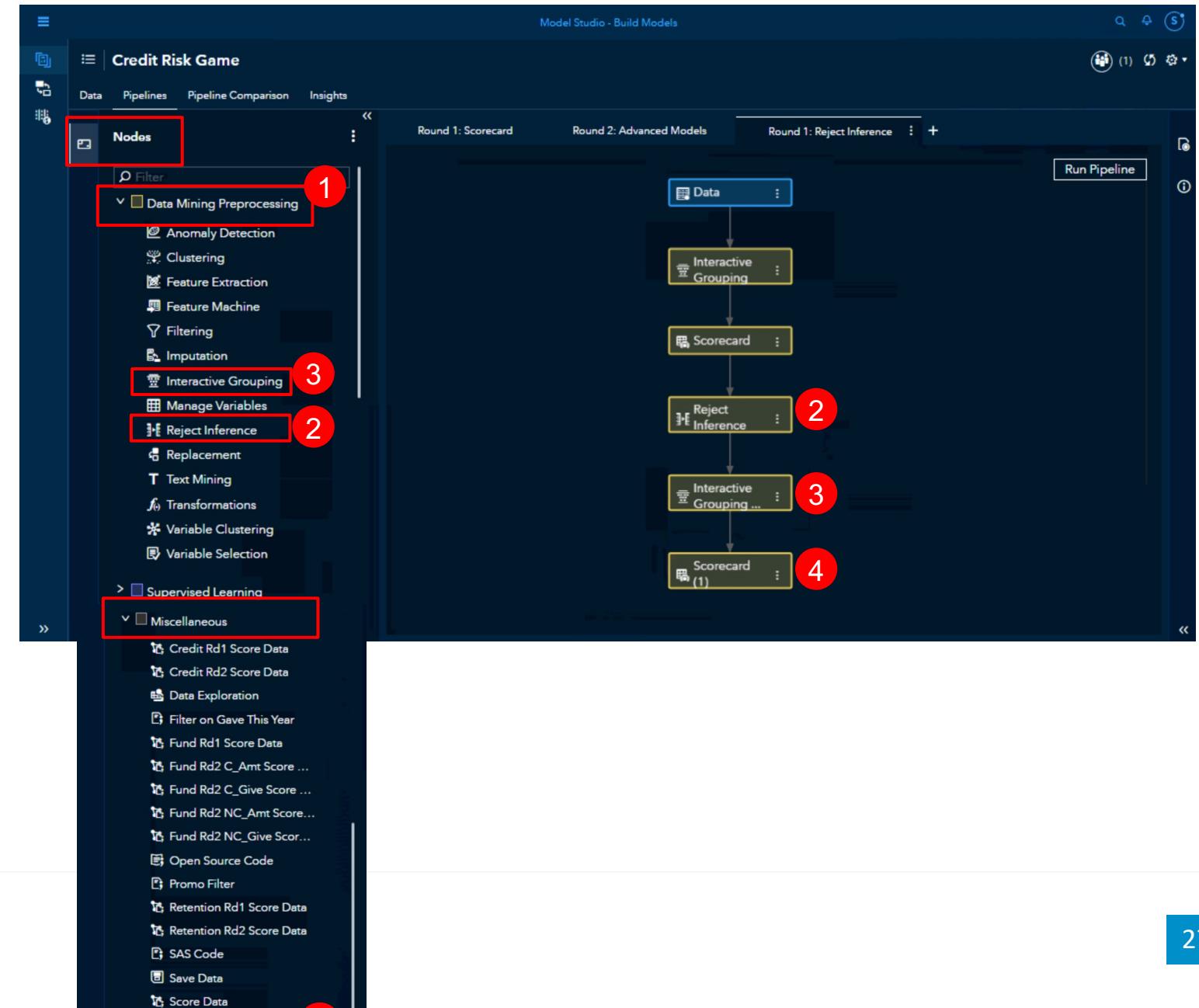
Re-Build the Pipeline

1. Delete 'Scorecard' and 'Score Data' nodes from the pipeline (right-click and delete)
2. Under 'Miscellaneous', select 'Scorecard'
3. Drag & drop it onto the "Interactive Grouping" node



Re-Build the Pipeline (Cont.)

1. Under Nodes, choose 'Data Mining Preprocessing'
2. Select 'Reject Inference' node and drag and drop it onto the 'Scorecard' node
3. Add another 'Interactive Grouping' Node
4. Under 'Miscellaneous', drag and drop another 'Scorecard' Node



NOTE: The final pipeline should look like the image in this slide.

Reject Inference

1. Click on the 'Reject Inference' node
2. Under Score data table name, browse to:

Data Sources >cas-shared-default > Public
CREDITRISK_REJECTS

NOTE: if data is not loaded, then click on the lightning icon to load the data first. This may happen the first time that you access this database.

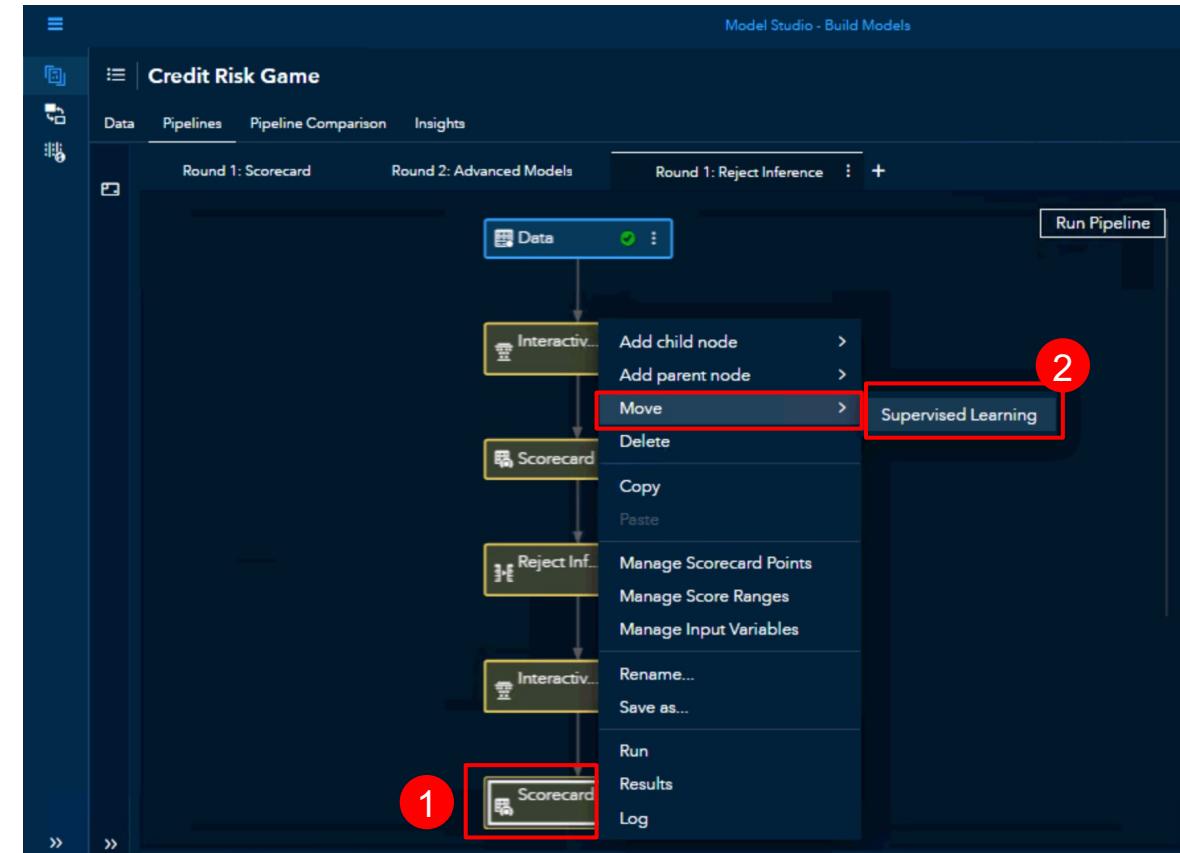
3. 'Run Pipeline' and explore your new Scorecard results

In this step, we are adding the dataset of rejected applicants to the 'Reject Inference' node. This will reduce the bias, since our dataset now includes both the data on accepted applicants (the original 'Data' node) and rejected applicants ('Reject Inference' node).



Move Scorecard to Supervised Learning

1. After the run is complete, right-click on the second 'Scorecard' node
2. Move the node to 'Supervised Learning'



The Pipeline Flow

Important Note:

Once you move the Scorecard to Supervised Learning, you need to make **some (any!) change** to the properties of the **second 'Interactive Grouping' node**, (grouping options, binning options, limits, etc.) to force the system to run it.

Otherwise, running the node will look like it's running, but nothing will happen.

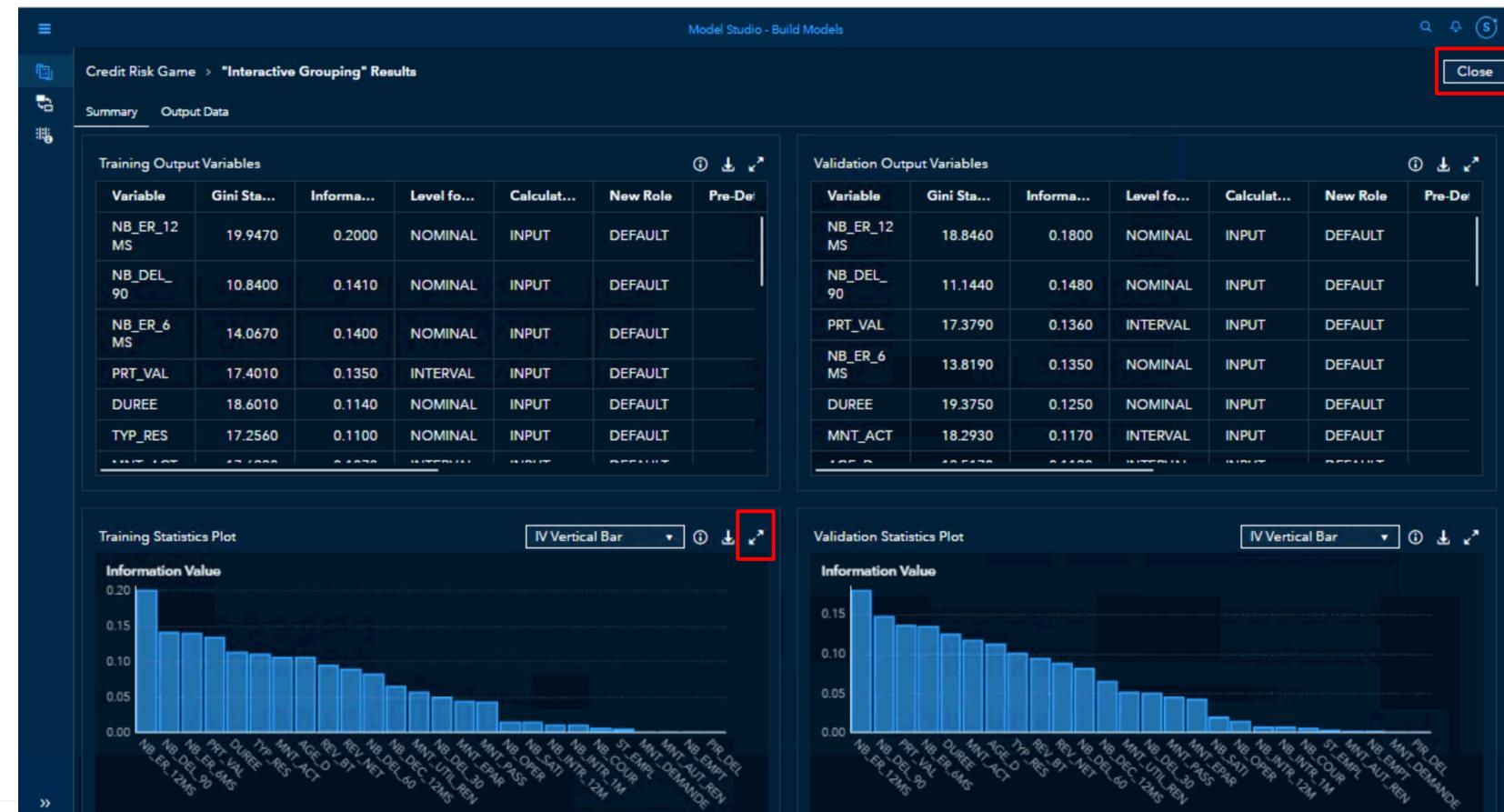
Once you made some changes, run the pipeline.

Depending on the nodes selected, the 'run' time may vary significantly. Please do not use any of the autotune features in this simulation.



Explore Results/Improve Models

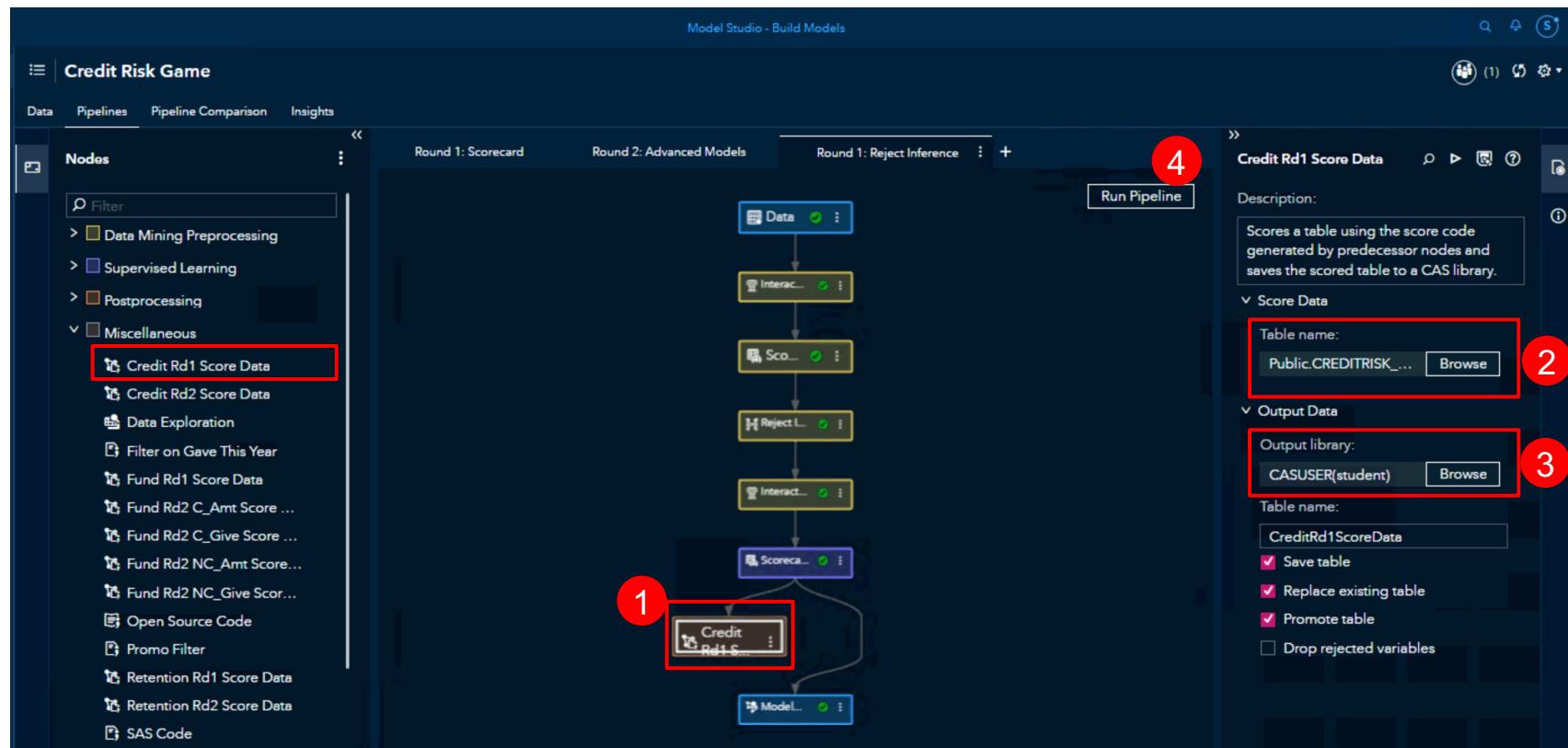
- Node Options have many controls that will impact the accuracy of selected models
- Right-clicking on each node provides detailed results
- Bins in Interactive Grouping can be adjusted by right-clicking on the node
- Explore the variables that have been accepted or rejected
- Expand any results window for more information and close the results once done



Score Data

Same as slides 15, 16, and 17:

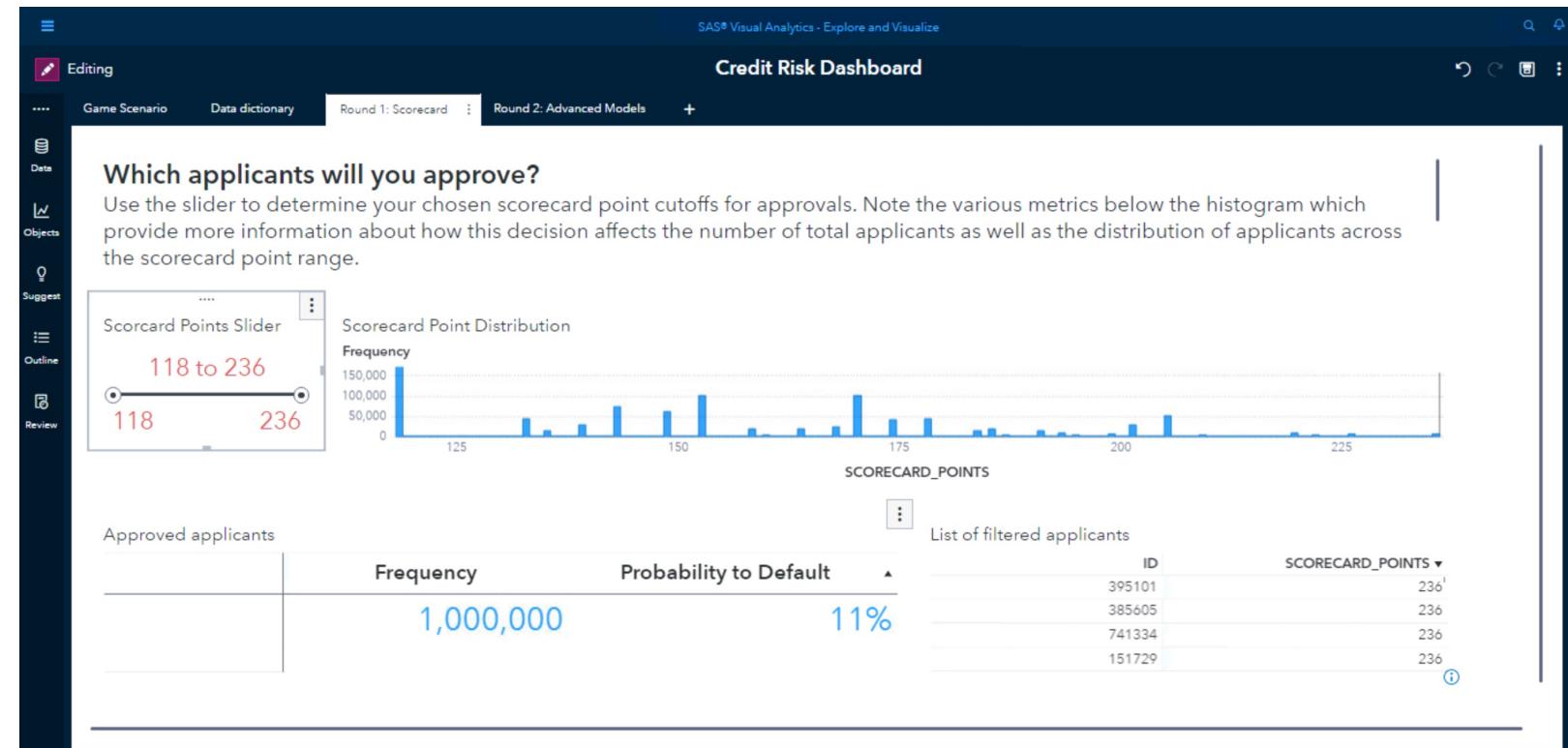
1. Add 'Credit Rd1 Score Data' node onto your 'Scorecard' node
2. Browse to Score data:
[CREDITRISK_SCOREDATA](#)
3. Set the Output library: Data Sources > [CASUSER\(student\)](#)
4. Run the 'Score Data' node or the whole pipeline again



Ready to Submit?

To make another submission, follow the same steps as your initial submission (slides 18 to 23):

- Go to '[Explore and Visualize](#)'
- Open the Credit Risk Dashboard and choose the tab for '[Round 1: Scorecard](#)'
- Refresh your data to: [Refresh CREDITRD1SCOREDATA](#)
- Select, sort and export your data to .csv





**You are now ready to upload
your solution (csv file)
for Round 1 to
the game leaderboard!**