



# How to play Cortex on SAS Viya

---

**ROUND 2: TWO-STAGE MODELING**

---

[erpsim.hec.ca](http://erpsim.hec.ca)

**ERPsimLab**  
**HEC MONTRÉAL**

Serious games to learn  
enterprise systems and  
business analytics

Last Update: July 24, 2020

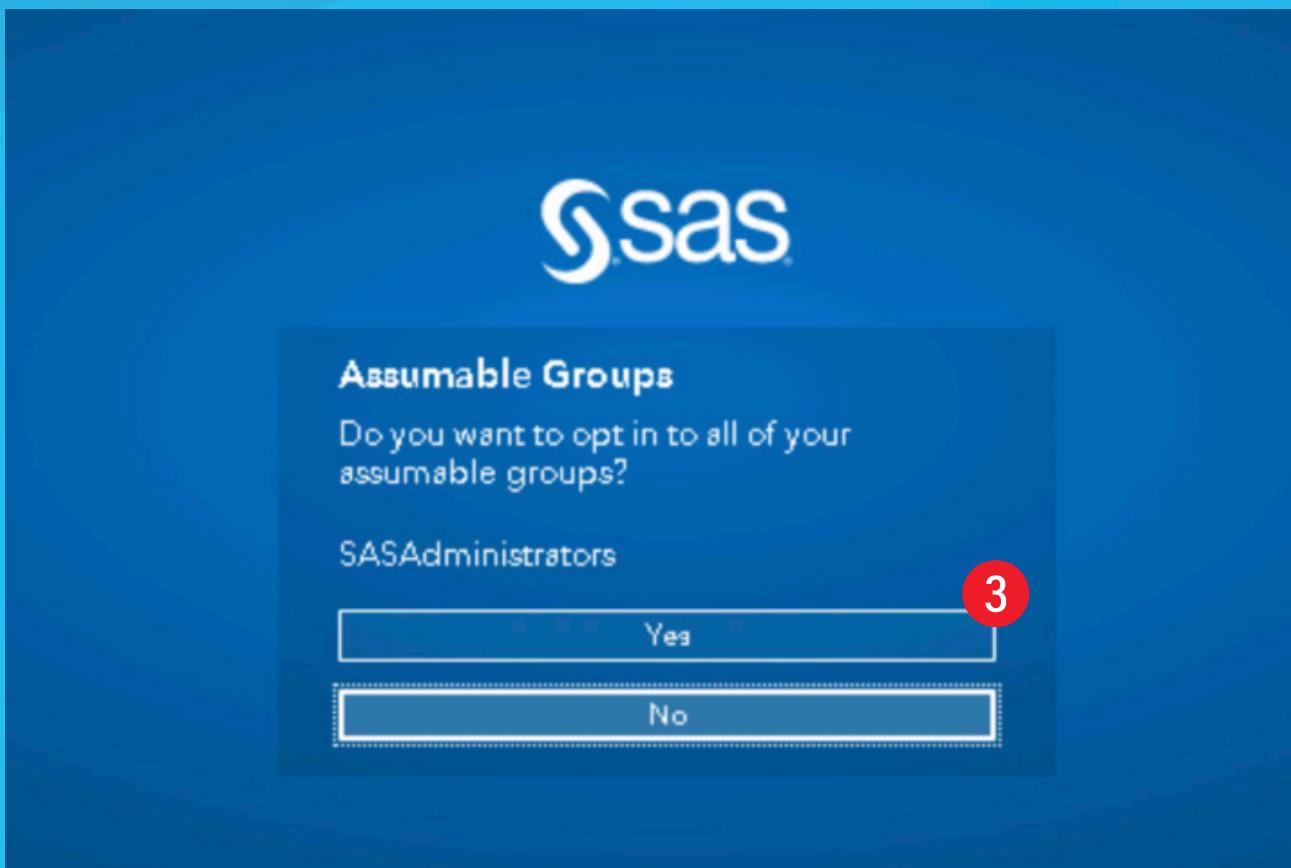
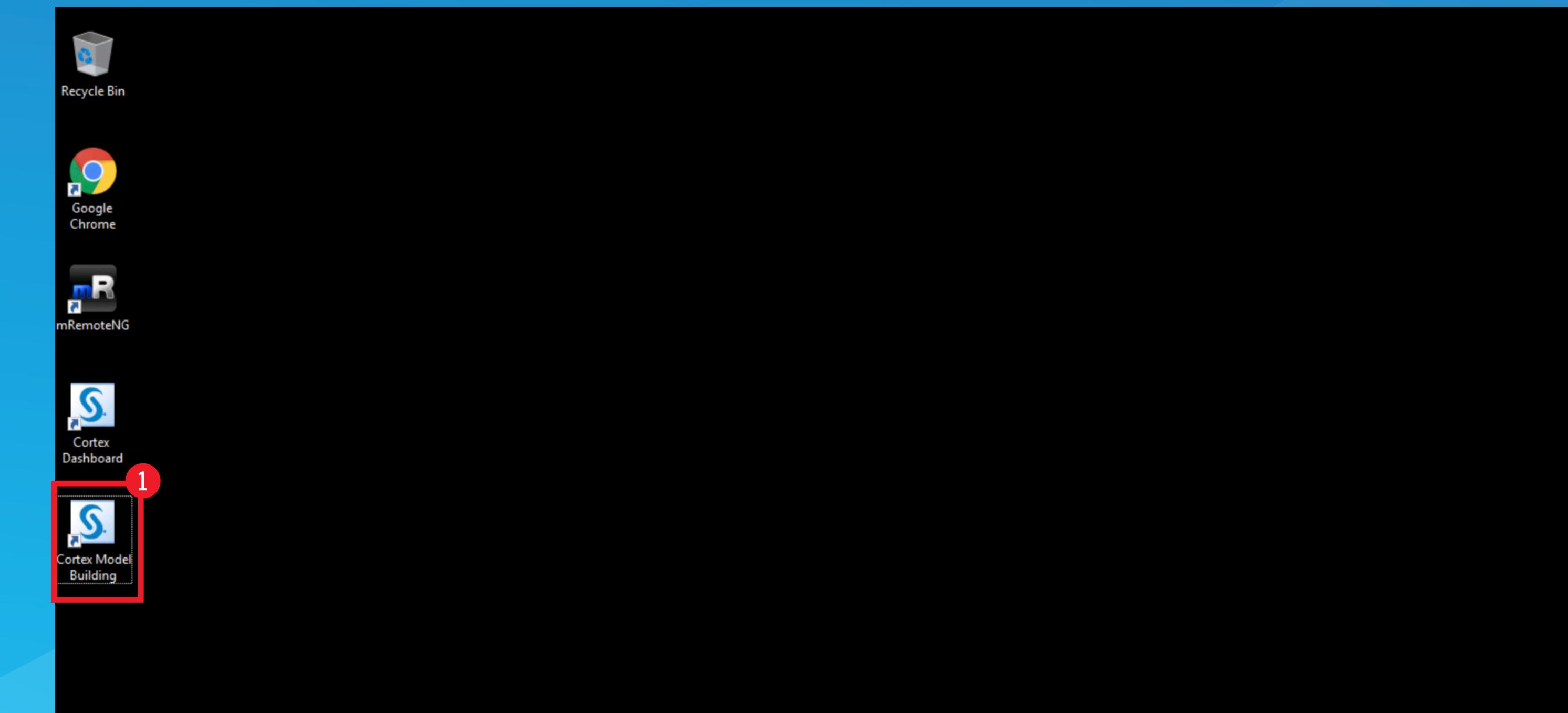
# Access SAS Viya

1. On your desktop, click on 'Cortex Model Building' icon
2. Enter your User ID and Password\*
3. Select 'Yes' when asked about 'Assumable Groups'

\* Please use the following user ID and password unless otherwise specified by your instructor:

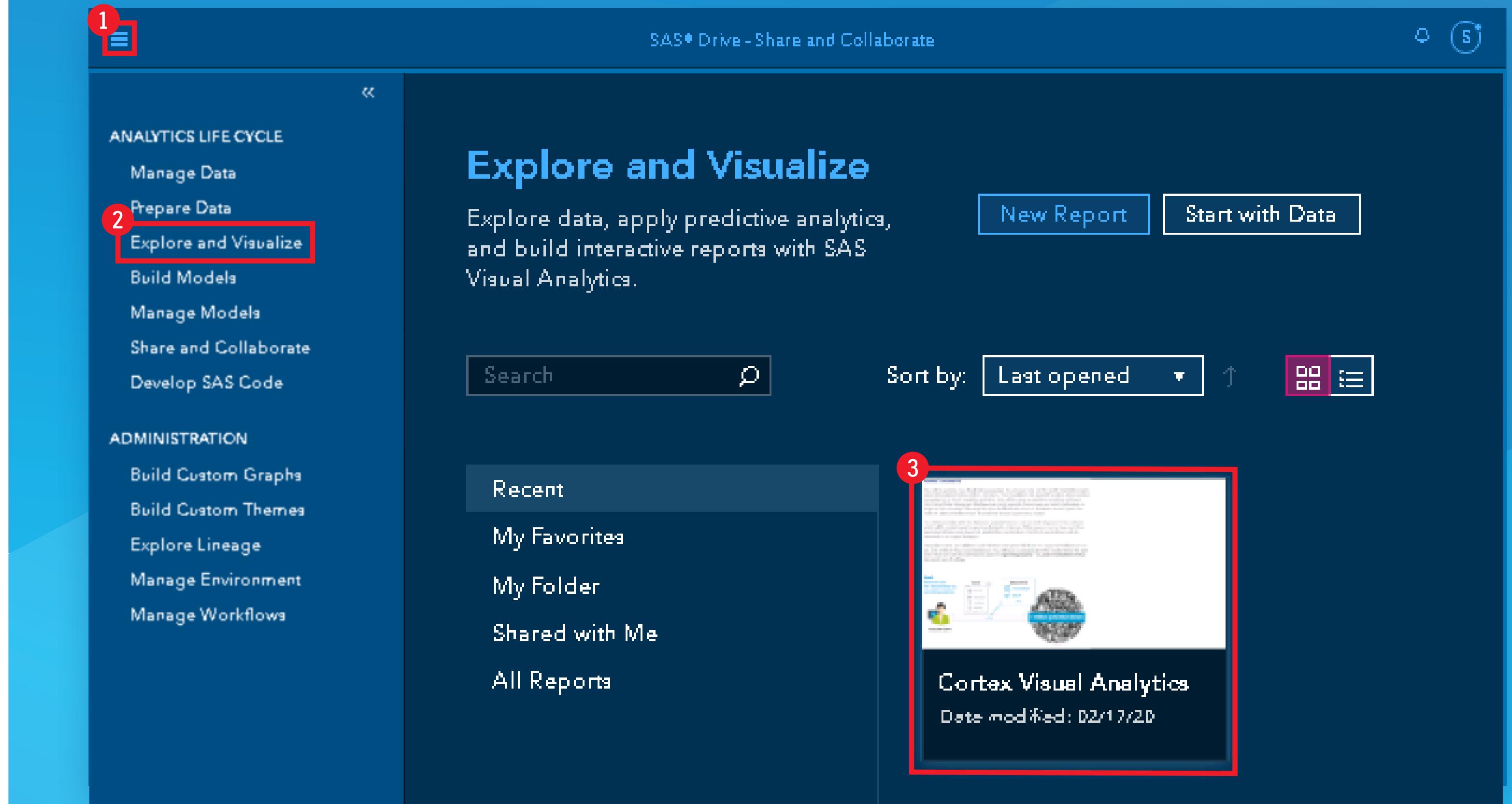
**User ID:** student

**Password:** Metadata0



# Explore the Game

1. In SAS Drive, click on the hamburger icon at the top left corner
2. Select 'Explore and Visualize'
3. Double click on 'Cortex Visual Analytics'





Use the tabs 'Game Scenario' and 'Data Dictionary' to explore the game scenario and round2 tabs to explore the game data (RD2: Explanation Conditional AmtThisYear, RD2: Explanation Conditional GiveThisYear).

**Editing**

Cortex Visual Analytics

Game Scenario Data Dictionary RD1: Explanantion Amt This Year RD1: Donors RD2: Explanation Conditional Give This Year RD2: Explanation Conditional Amt This Year RD2: Donors +

**Game's scenario**

You will be working on a fundraising campaign for a 12-year old, not-for-profit charitable organization (foundation) with a million members. The foundation has decided to add a direct contact campaign to its list of marketing activities. You will be using a predictive modeling software (SAS Visual Data Mining and Machine Learning) to predict how many and which individuals to target in the campaign. The objective is to fundraise the most in donation amount given the costs of calling members (sum of predicted amount given minus costs).

You will be provided with the dataset of potential donors, and pre-built diagrams in the software, which will fit models based on previous behavior of donors (if they gave or not or how much they gave) and will also score donors to predict this year donation. The list of scored donors will be exported to an output file/report.

Using this output, you will have to decide how many potential donors to target and will have to create a list of IDs of those potential donors. You will have to upload/submit the created list to the platform which will rank the submissions based on operating surplus – i.e., sum of donations minus the total cost of calling.

**Goal:**  
Maximize the net raised funds (i.e., operating surplus)

**DATA**

ANALYTICS

How many?  
Who?  
...

1 million potential donors

CALLING COST

**Available information**

In order to play the game and make decisions, you will have access to a dataset of 1 million potential donors as well as the costs associated to calling them (given to you by your teacher):

**Table 1 Members' Data (please see table 3 for a full list of variables)**

Category of data	Example
Membership activity (historical info)	Minimum, maximum, & total donation, ...
Demographic data	Age, Gender, ...
Socioeconomic status	Salary, education...
Previous behaviour (last year & this year)	If members gave and how much

**Table 2 Example of a Cost Schedule\***

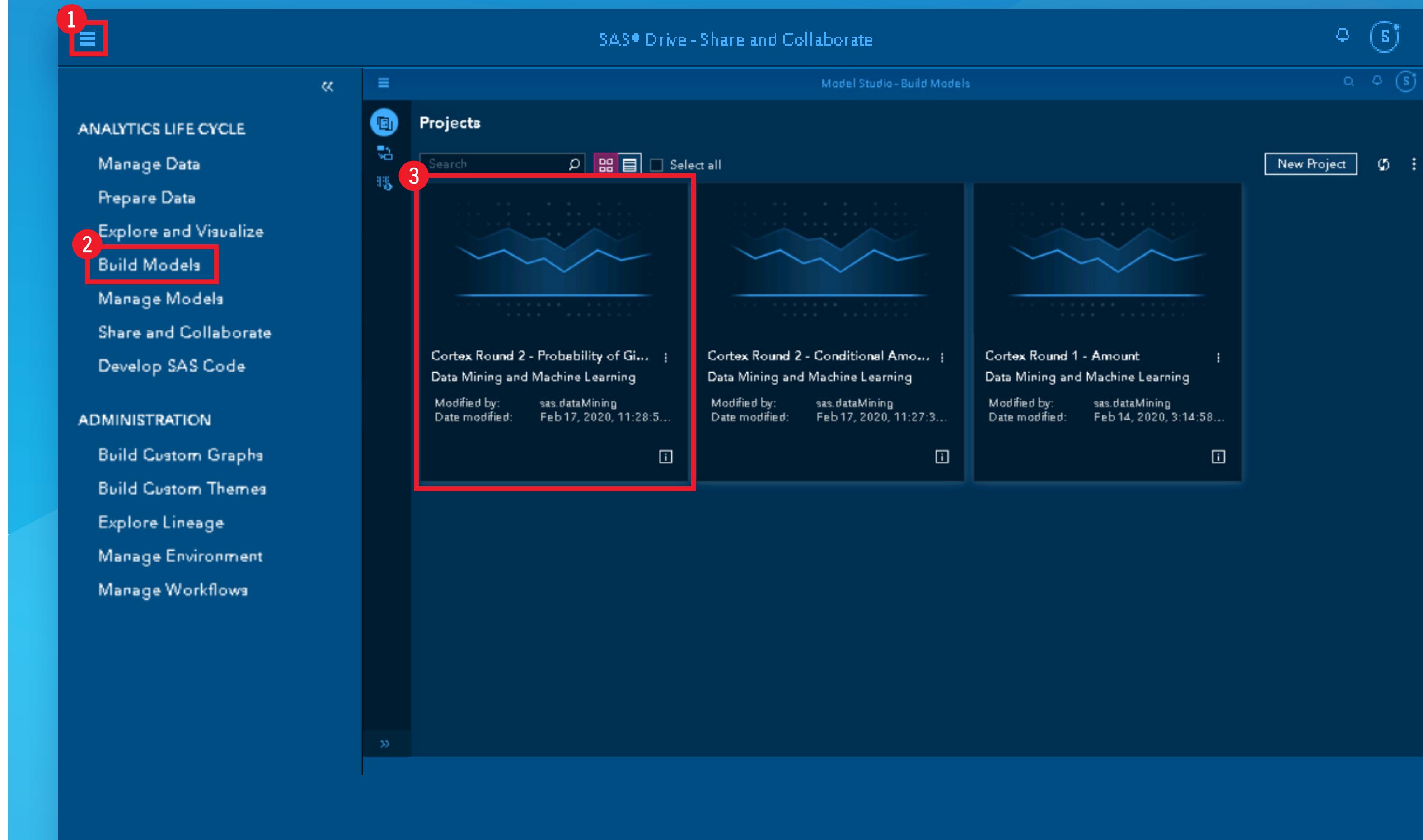
Number of contacted members	Cost per person
0 - 60,000	5\$/person
> 60,000	25\$/person

\* Please note that the cost schedule could vary from game to game. Make sure that you are using the correct cost schedule communicated to you by your teacher.

## STAGE 1

# Open Project: Cortex Round 2 - Probability of Giving

1. In SAS Drive, click on the hamburger icon at the top left corner
2. Select 'Build Models'
3. Double click on 'Round 2 – Probability of Giving'



# Assign Variable Roles

To change or reject some variable roles:

1. Select the variable and change their roles on the right-side panel

**NOTE:** choose variables one at the time by deselecting one before selecting the next

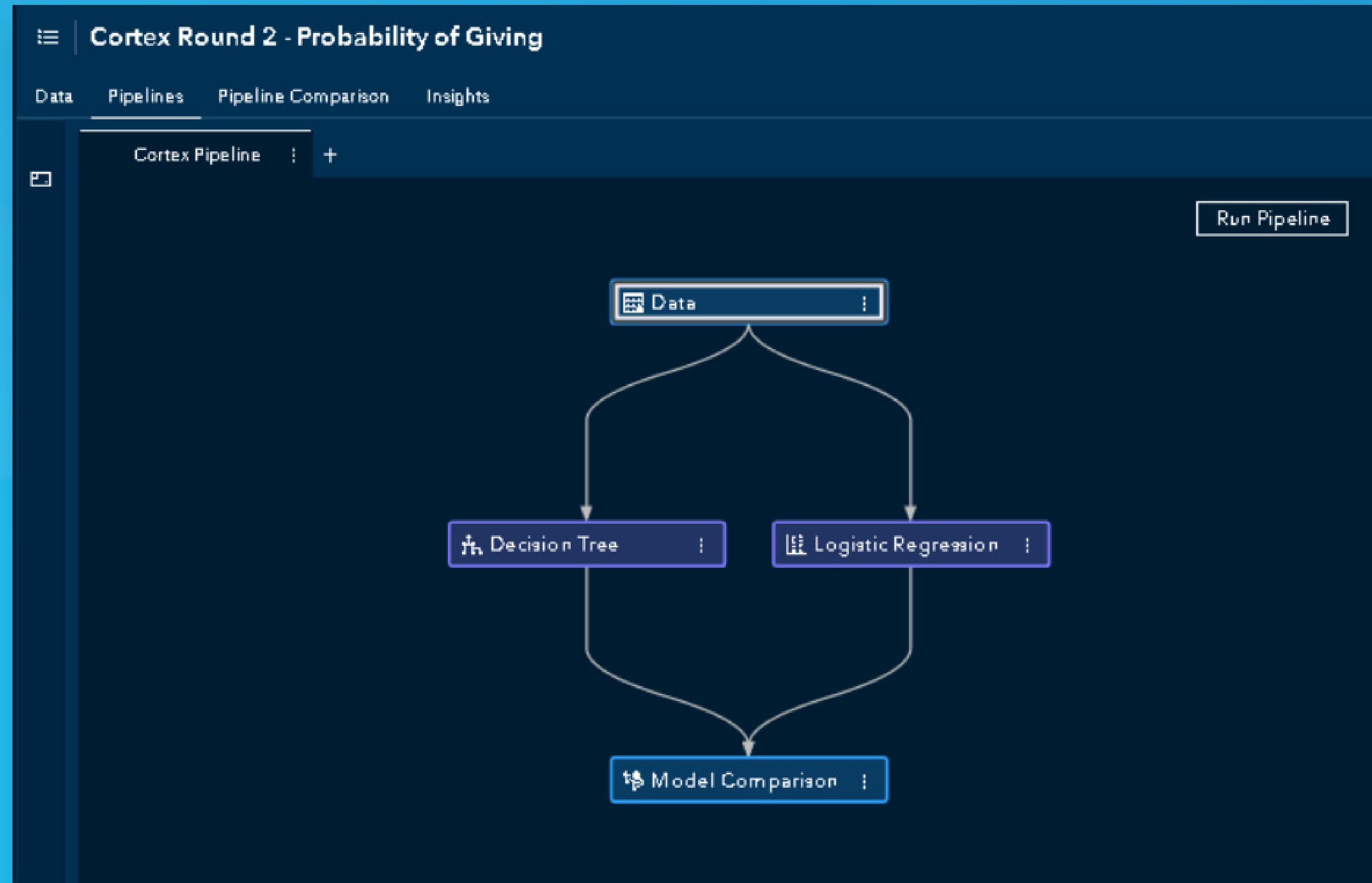
2. Once done with changing variable role, click on the 'Pipelines' tab

**NOTE:** 'GaveThisYear' should stay the target Variable and 'AmtThisYear' should stay rejected

Variable Name	Type	Role	Level	Order	Comment	Number of Levels
Age	Numeric	Input	Nominal	Default		75
AmtLastYear	Numeric	Input	Nominal	Default		49
AmtThisYear	Numeric	Rejected	Nominal	Default		49
City	Character	Input	Nominal	Default		4
Contact	Numeric	Input	Binary	Default		2
Education	Character	Input	Nominal	Default		3
FirstName	Character	ID	Nominal	Default	The variable exceeds the maximum number of levels cutoff value.	>254
Frequency	Numeric	Input	Nominal	Default		10
GaveLastYear	Numeric	Input	Binary	Default		2
GaveThisYear	Numeric	Target	Binary	Default		2
ID	Numeric	ID	Interval	Default		>254
LastName	Character	ID	Nominal	Default	The variable exceeds the maximum number of levels cutoff value.	>254
MaxGift	Numeric	Input	Nominal	Default		48
MinGift	Numeric	Input	Nominal	Default		48
NbActivities	Numeric	Input	Nominal	Default		11

# Pipeline for Round 2 – Stage1:

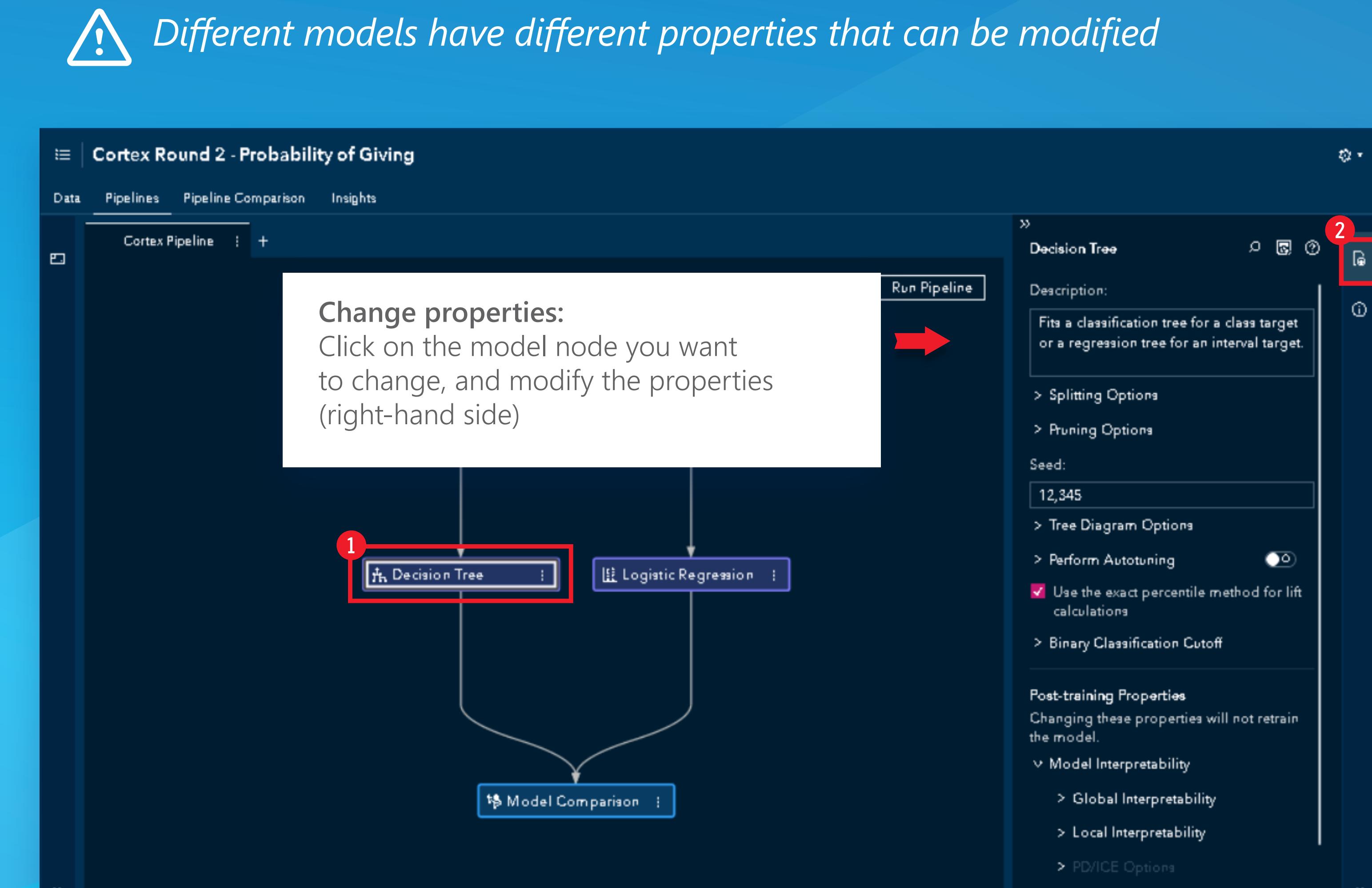
*Here is a pre-built pipeline that you can start with*



# Improve the Default Models

To open the properties of each model:

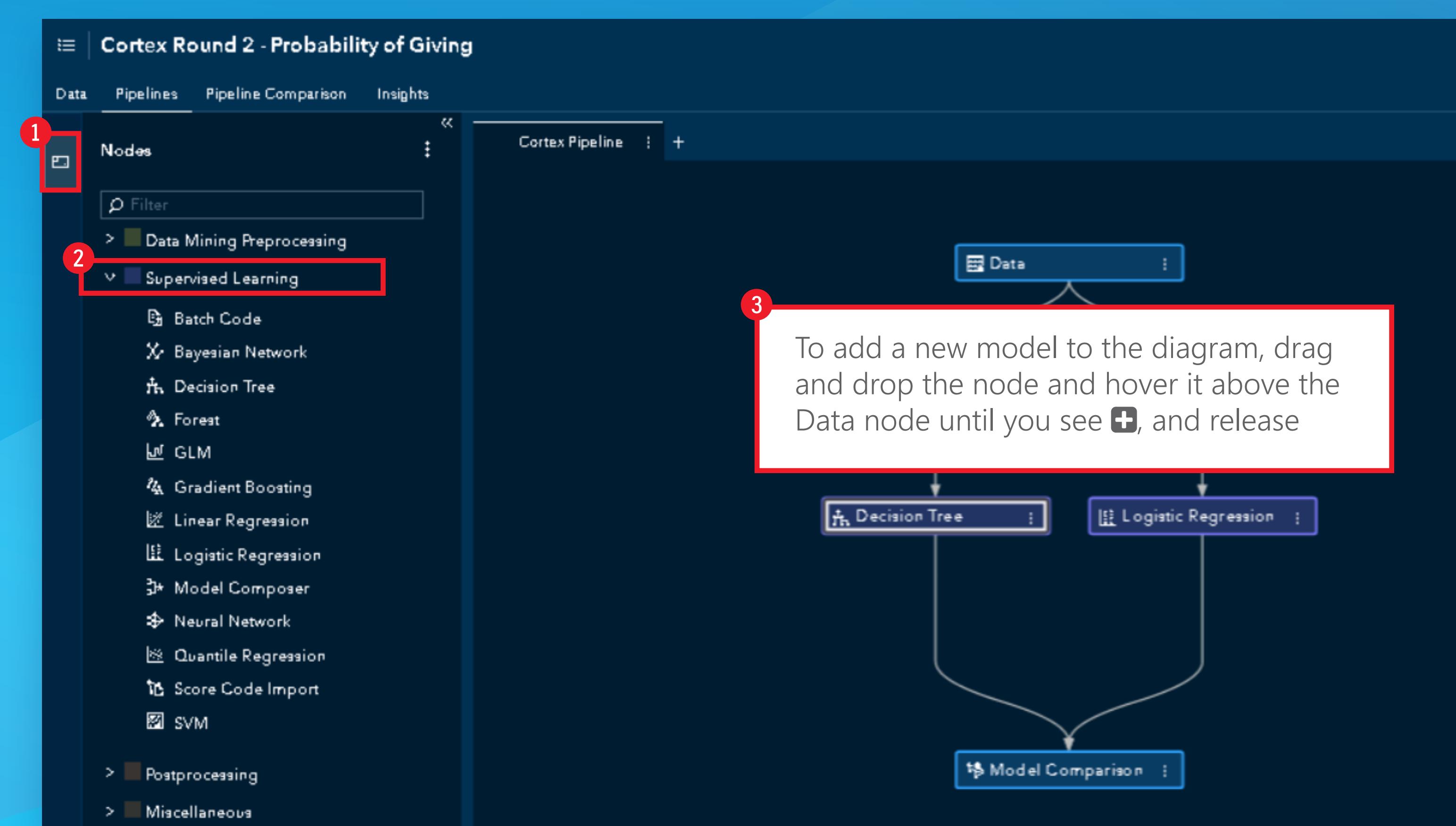
1. Click on the model's node to see the node options on the right-hand side.
2. To learn more about the nodes click on 'Node' details:
  - You should try to improve models' properties.
  - Remember your mission is to better predict who to call!



# Improve the Default Pipeline

1. Click on the 'Nodes' icon on the left to reveal additional nodes
2. Under 'Supervised Learning', find other predictive models that you can use
3. Drag the selected model to the data node and release

**NOTE:** 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!



# Run the Pipeline/ See Results

1. Click on 'Run Pipeline'

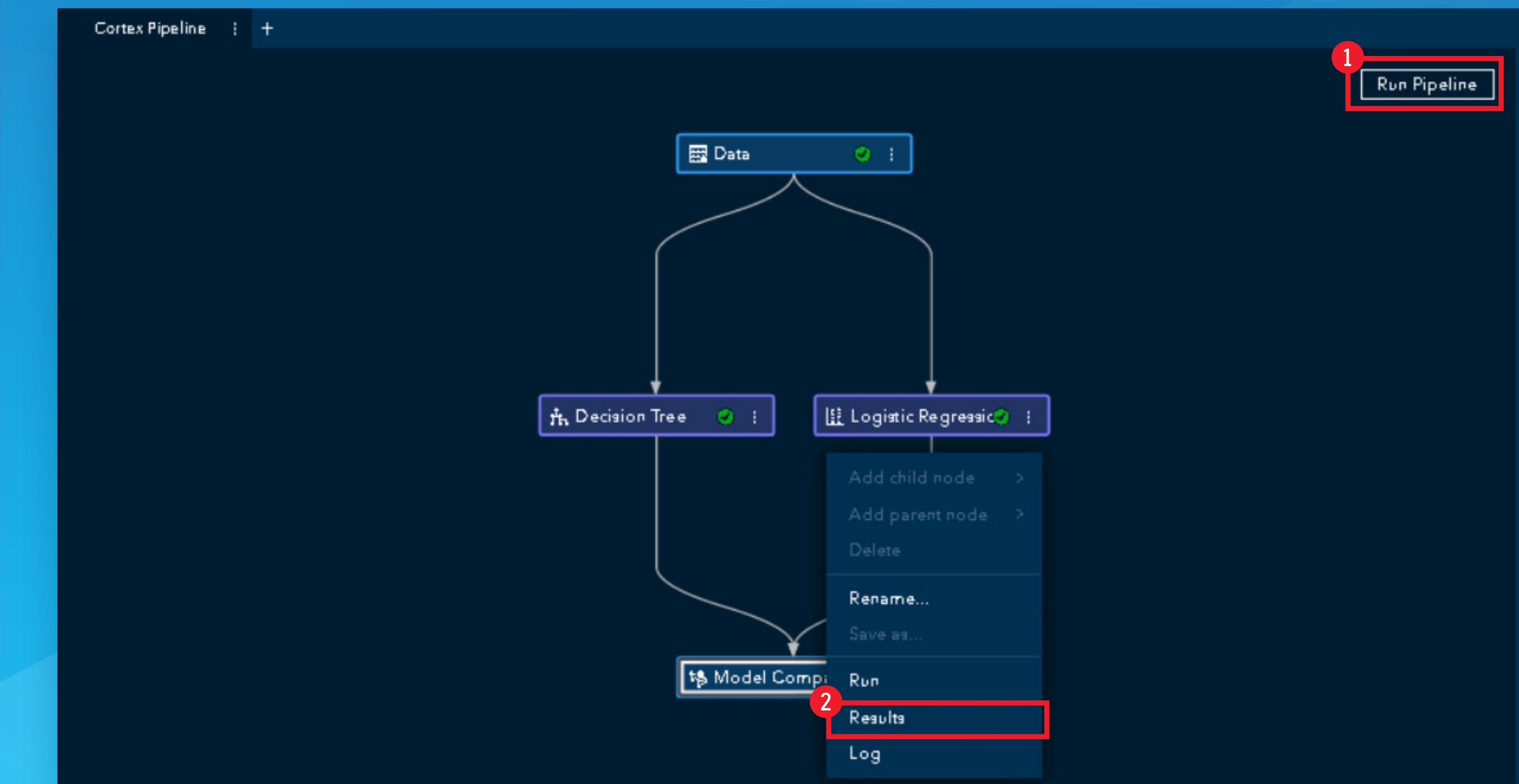
**NOTE:** green check marks appear on nodes when run is complete

2. When run is complete, right-click on the 'Model Comparison' node and click on 'Results'

3. In 'Model Comparison Results' see the Champion Model

**NOTE:** the model with the lower 'Misclassification rate' is chosen

4. Click on 'Close', to return to your pipeline page



The screenshot shows the 'Model Comparison Results' table. A red box labeled '3' highlights the 'Champion' column header. A red box labeled '4' highlights the 'Close' button at the top right. The table lists two models: 'Decision Tree' and 'Logistic Regression', along with their algorithm names and misclassification rates. The 'Misclassification Rate (Event)' column is highlighted by a red box.

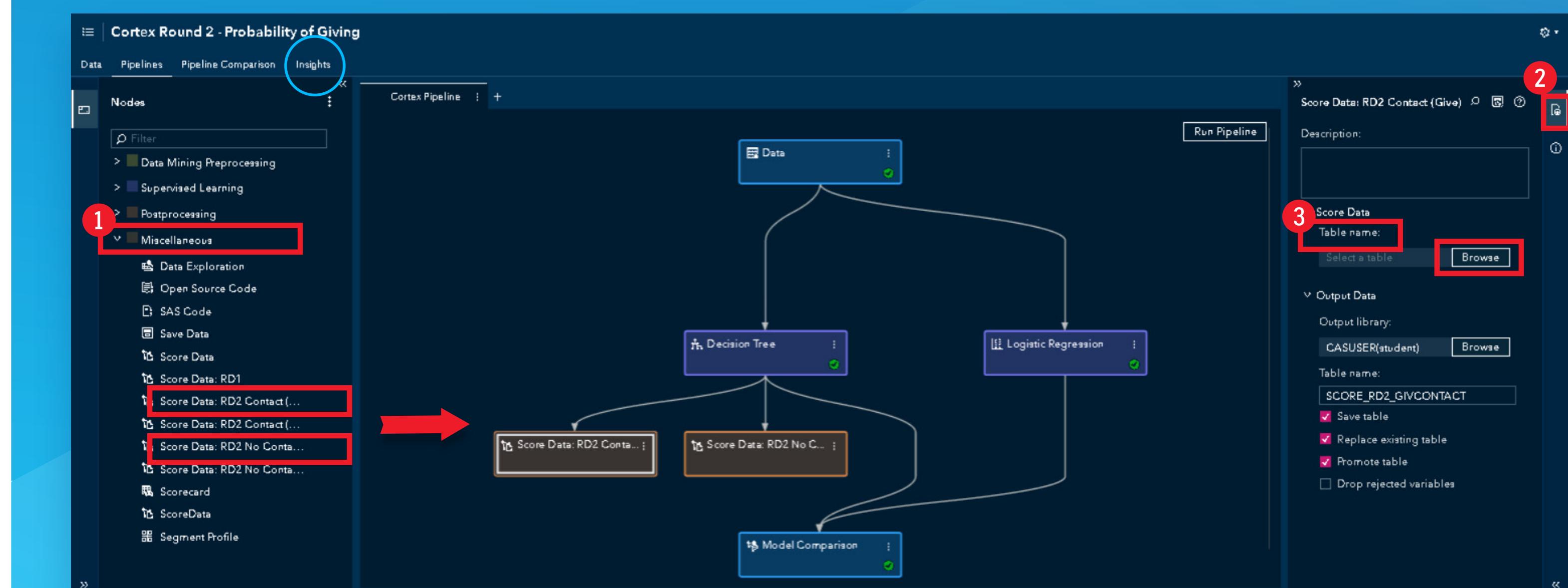
Champion	Name	Algorithm Name	Misclassification Rate (Event)	Misclassification Rate
Decision Tree	Decision Tree	Decision Tree	0.1456	0.1456
Logistic Regression	Logistic Regression	Logistic Regression	0.1487	0.1487

# Score New Data

1. Under 'Miscellaneous', select the 'Score Data: RD2 Contact (Give)' and 'Score Data: RD2 No Contact (Give)' nodes and drag them to the Champion model node (here decision tree) & release
2. Click on 'Score Data: RD2 Contact (Give)' and open the right-hand side pane
3. In the node option pane, under Score data, in the table name, click on browse

**NOTE:** You should not change the output table name

**⚠️ Use the 'Score Data: RD1' node on the champion model to score your data in your private directory**



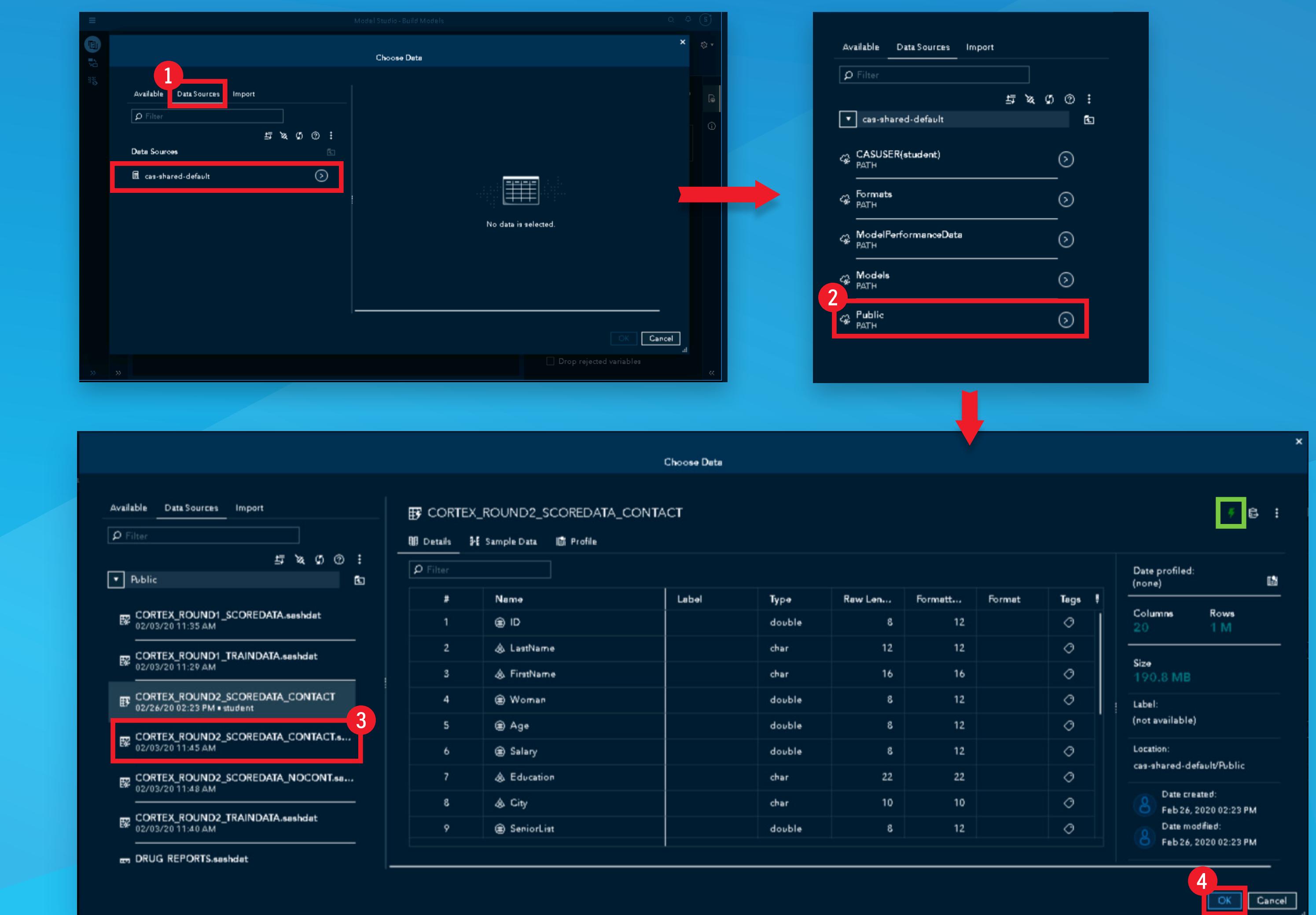
**NOTE:** Before scoring your data, it is recommended to evaluate the performance of your models using 'Insights' tab

# Load 'Contact Score' Data

1. Under 'Data sources', select 'Cas-shared-default'
2. Click on 'Public'
3. Select 'CORTEX\_ROUND2\_SCOREDATA\_CONTACT.sashdat'

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

4. Click on 'OK'

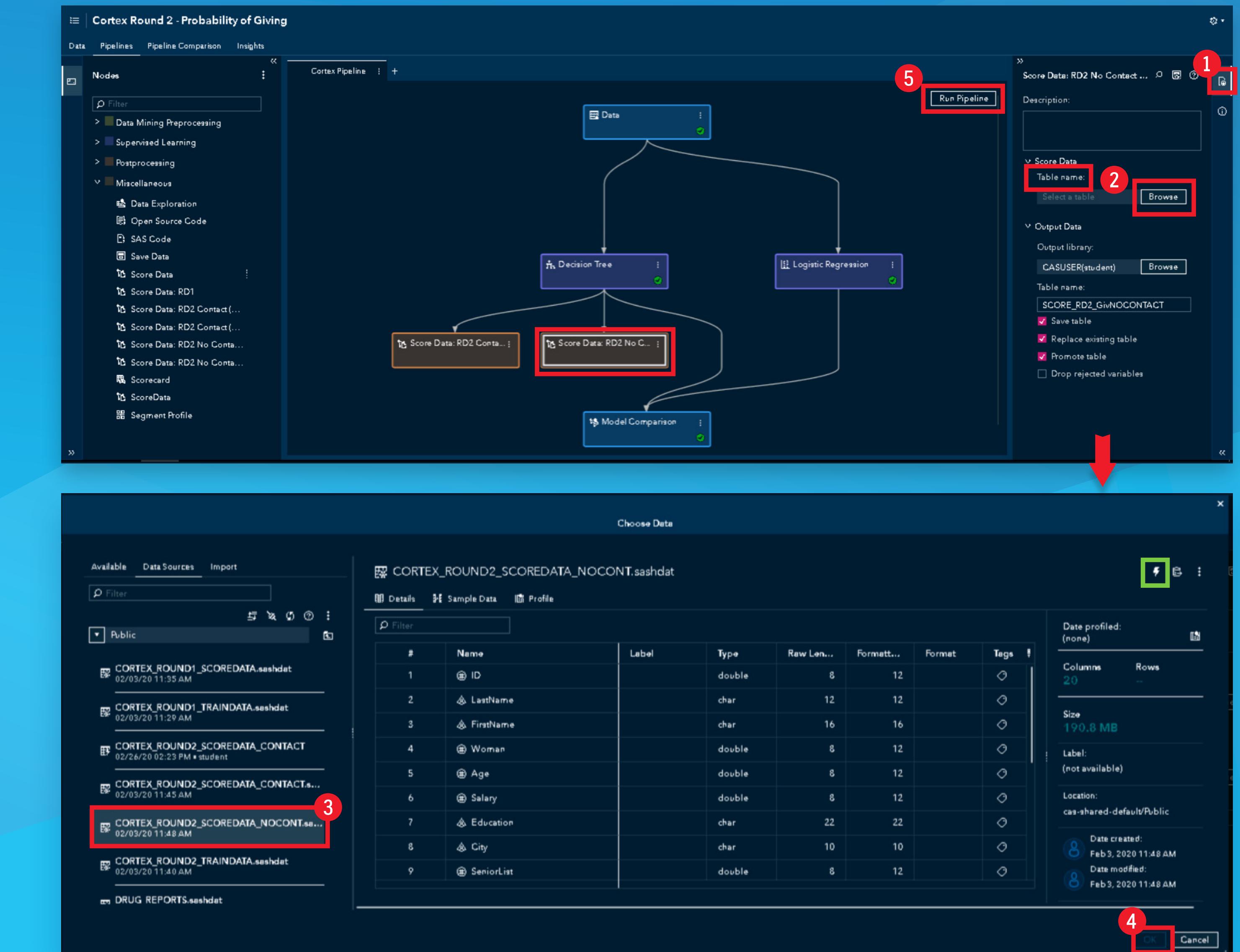


# Load 'No Contact Score' Data

1. Click on 'Score Data: RD2 No Contact (Give)' and open the right-hand side panel
2. In the node option pane, under Score data, in the table name, click on browse
3. Under 'Data sources', select 'Cas-shared-default' , click on 'Public' and Select 'CORTEX\_ROUND2\_SCOREDATA\_NOCONTACT.sashdat'

**NOTE:** if data is not loaded, then click on the lightning icon to load the data first.

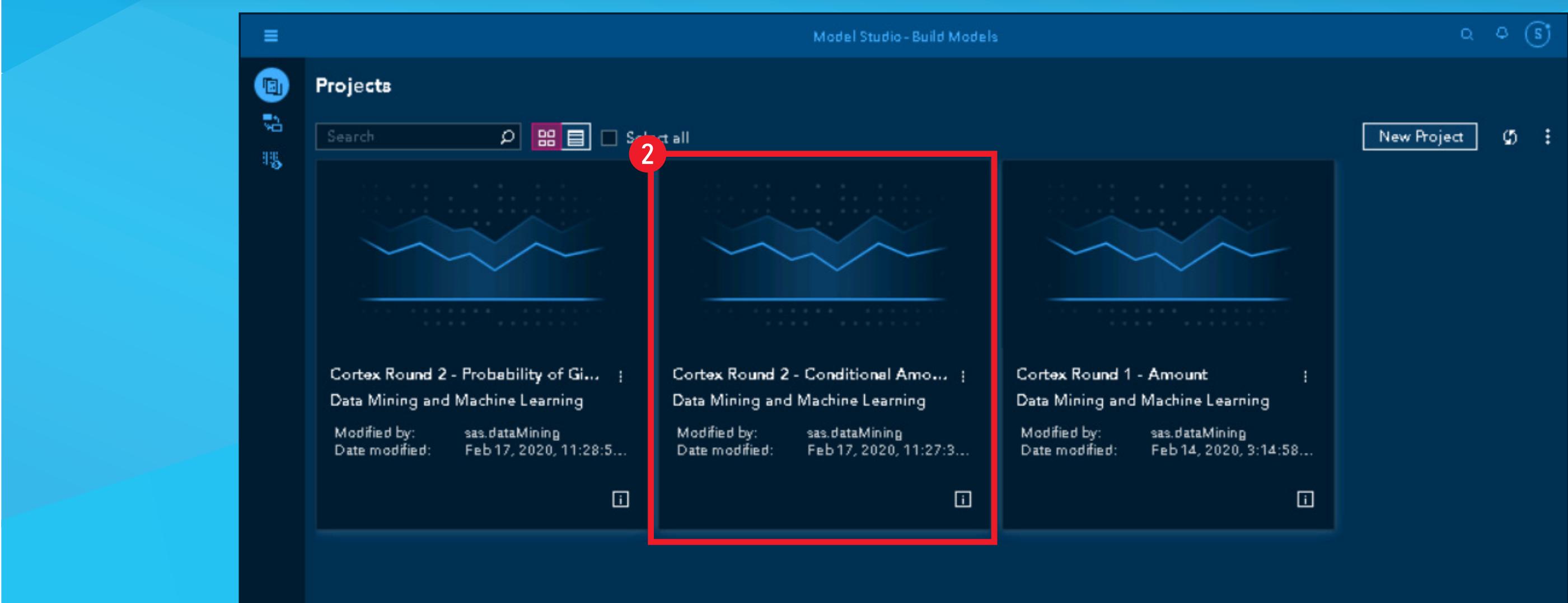
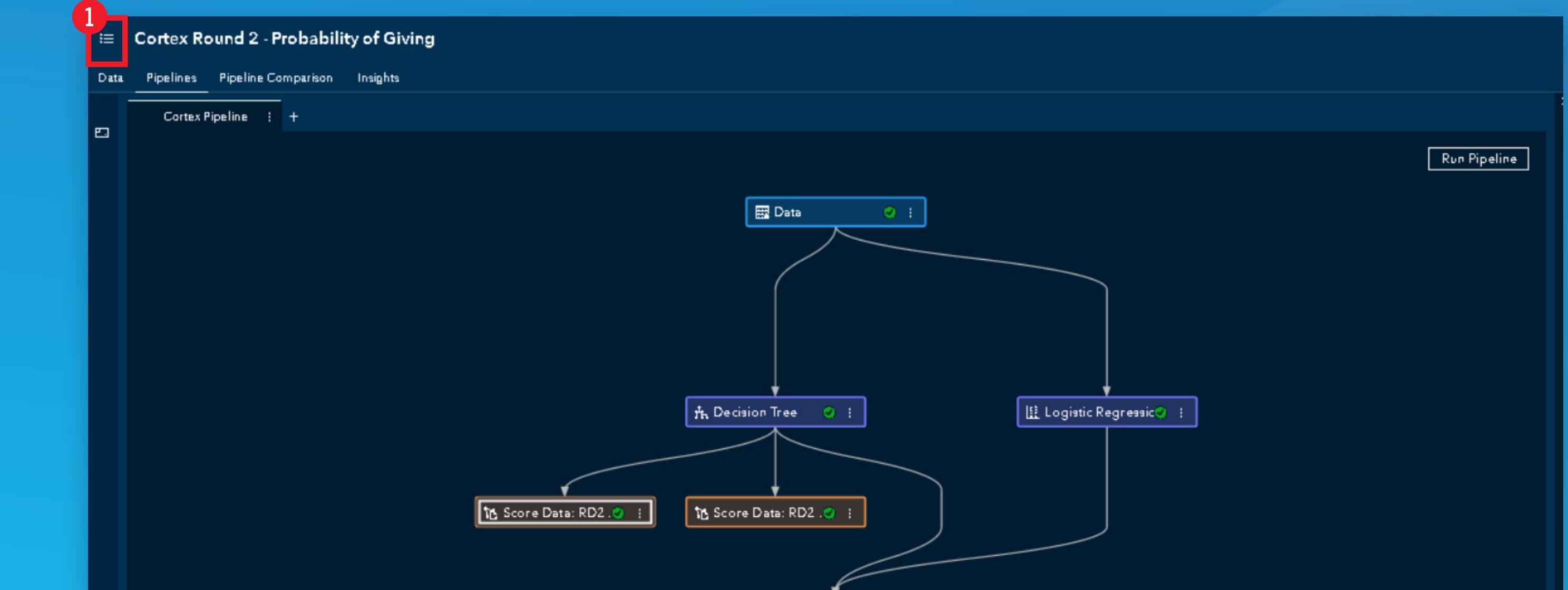
4. Click on 'OK'
5. Run Pipeline



## STAGE 2

# Open Project: Cortex Round 2 – Conditional Amount

1. Click on 'view all projects' icon
2. Double click to open 'Round 2 – Conditional Amount'



# Assign Variable Roles

To change or reject some variable roles:

1. Select the variable and change their roles on the right-side panel

**NOTE:** choose variables one at the time by deselecting one before selecting the next

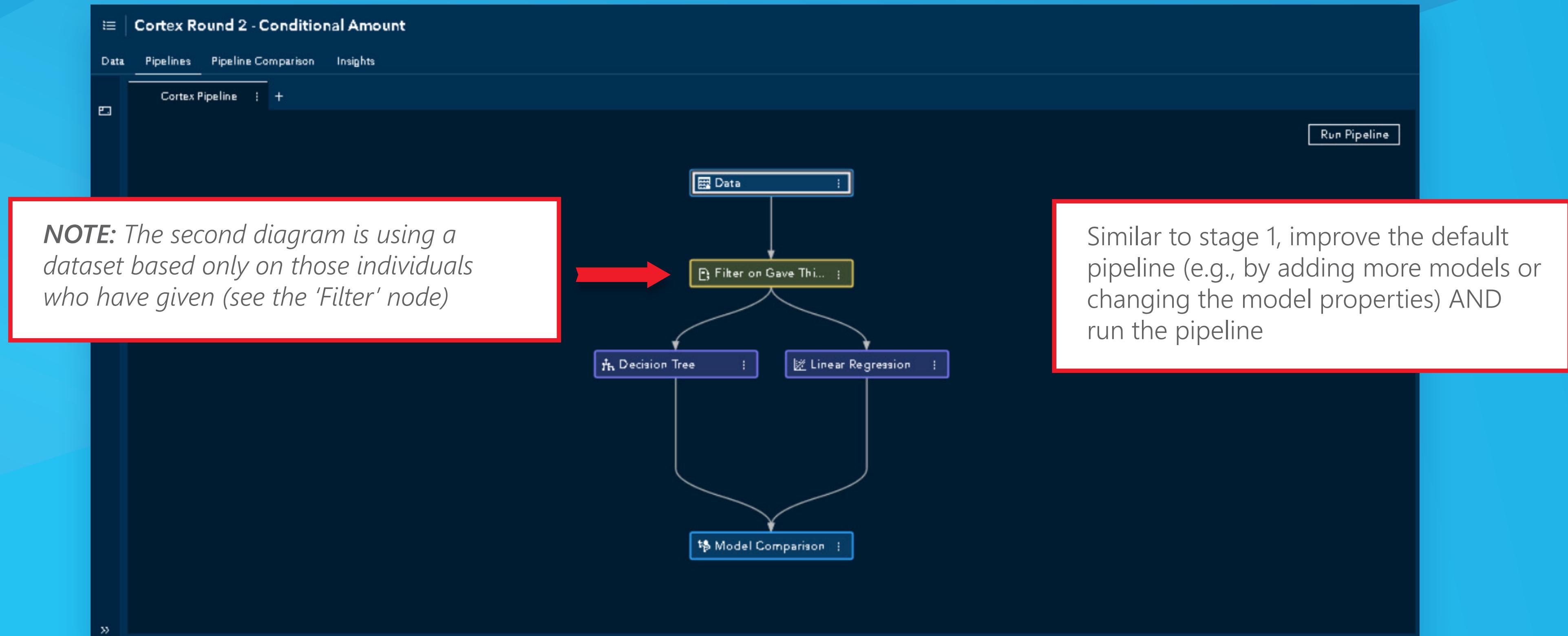
2. Once done with changing variable role, click on the 'Pipelines' tab

**NOTE:** 'AmtThisYear' should stay the target Variable and 'GaveThisYear' should stay rejected

Variable Name	Type	Role	Level	Order	Comment	Number
Age	Numeric	Input	Interval	Default		75
AmtLastYear	Numeric	Input	Interval	Default		49
AmtThisYear	Numeric	Target	Interval	Default		49
City	Character	Input	Nominal	Default		4
Contact	Numeric	Input	Binary	Default		2
Education	Character	Input	Nominal	Default		3
FirstName	Character	ID	Nominal	Default	The variable exceeds the maximum number of levels cutoff value.	>254
Frequency	Numeric	Input	Interval	Default		10
GaveLastYear	Numeric	Input	Binary	Default		2
GaveThisYear	Numeric	Rejected	Binary	Default		2
ID	Numeric	ID	Interval	Default		>254
LastName	Character	ID	Nominal	Default	The variable exceeds the maximum number of levels cutoff value.	>254
MaxGift	Numeric	Input	Interval	Default		48
MinGift	Numeric	Input	Interval	Default		48
NbActivities	Numeric	Input	Interval	Default		11

# Pipeline for Round 2 – Stage2:

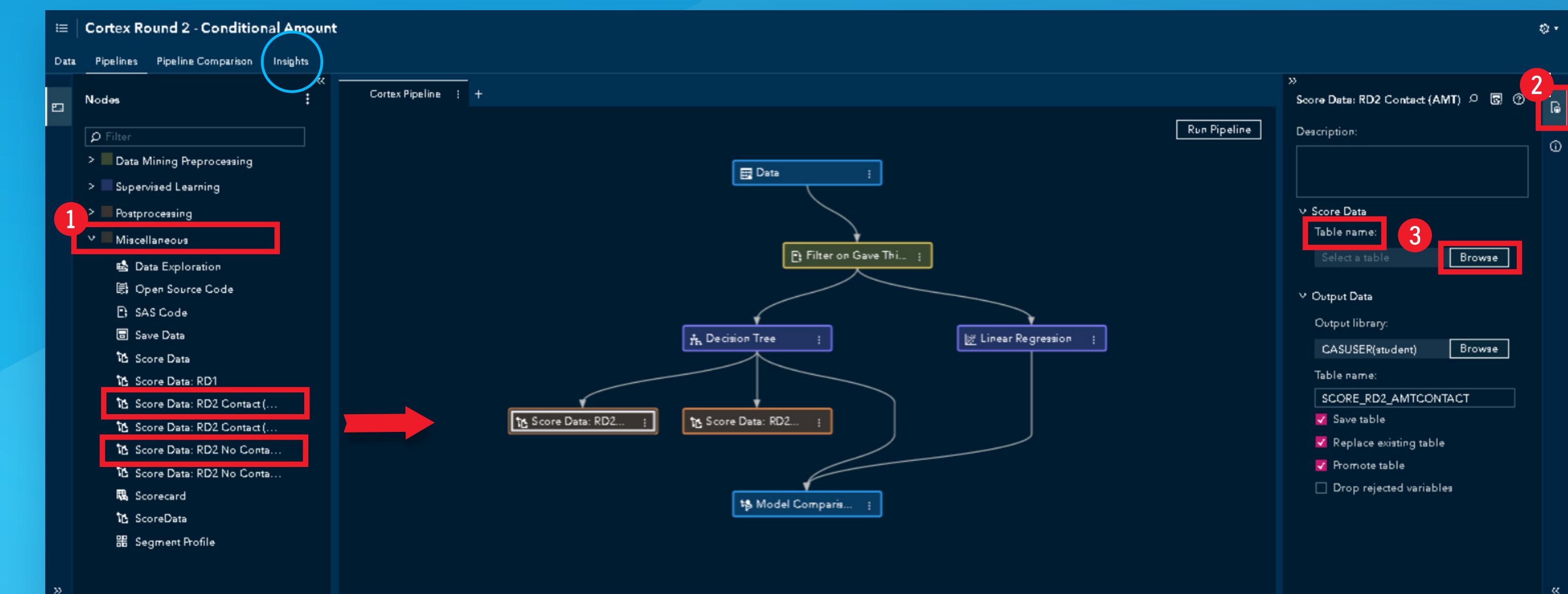
*Here is a pre-built pipeline that you can start with*



# Score New Data

1. Under 'Miscellaneous', select the 'Score Data: RD2 Contact (Amt)' and 'Score Data: RD2 No Contact (Amt)' nodes and drag them to the Champion model node (here decision tree) & release
2. Click on 'Score Data: RD2 Contact (Amt)' and open the right-hand side panel
3. In the node option pane, under Score data, in the table name, click on browse

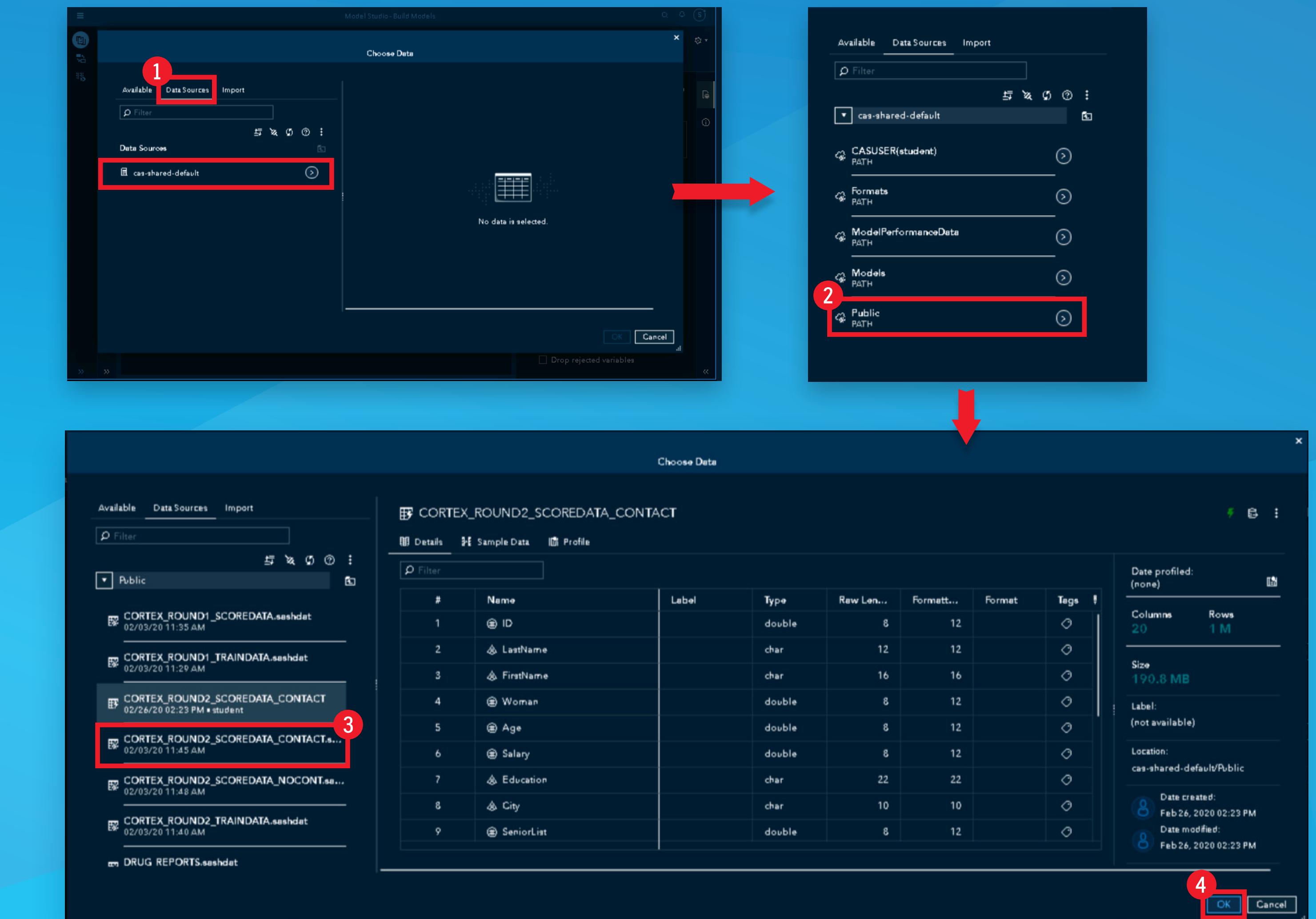
**NOTE:** You should not change the name of the output table



**NOTE:** Before scoring your data, it is recommended to evaluate the performance of your models using 'Insights' tab

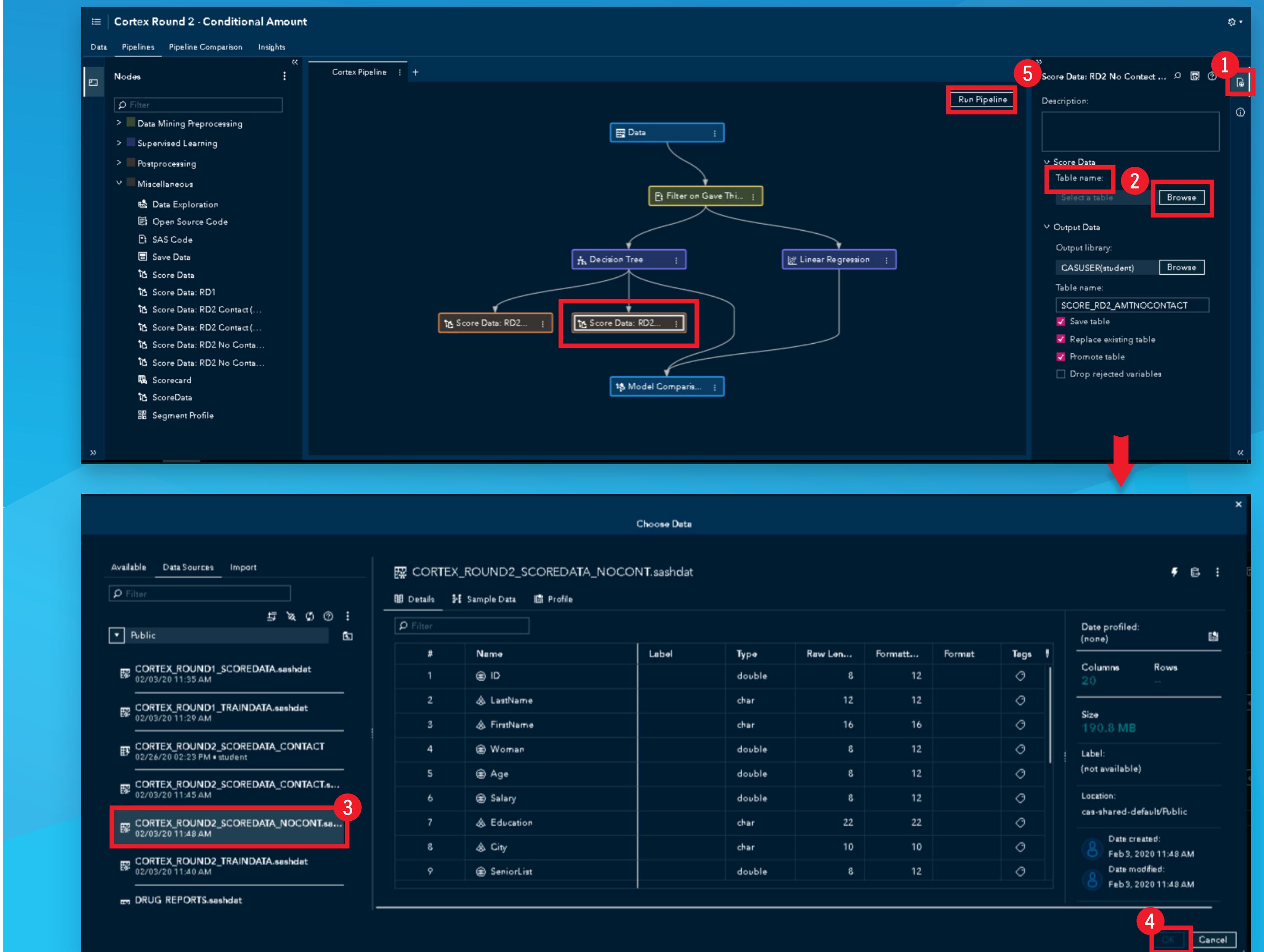
# Load 'Contact Score' Data

1. Under 'Data sources', select 'Cas-shared-default'
2. Click on 'Public'
3. Select 'CORTEX\_ROUND2\_SCOREDATA\_CONTACT'
4. Click on 'OK'



# Load 'No Contact Score' Data

1. Click on 'Score Data: RD2 No Contact (Give)' and open the right-hand side panel
2. In the node option pane, under Score data, in the table name, click on browse:
3. Under 'Data sources', select 'Cas-shared-default' , click on 'Public' and Select 'CORTEX\_ROUND2\_SCOREDATA\_NOCONTACT'
4. Click on 'OK'
5. Run Pipeline



# Re-open Cortex Visual Analytics Report

1. Click on the hamburger icon on the top left corner
2. Select 'Explore and Visualize'
3. Click on 'RD2: Donors' tab

The screenshot shows the SAS Visual Analytics interface. On the left, a sidebar titled 'ANALYTICS LIFE CYCLE' contains options: Manage Data, Prepare Data, **Explore and Visualize**, Build Models, Manage Models, Share and Collaborate, Develop SAS Code, ADMINISTRATION, Build Custom Graphs, Build Custom Themes, Explore Lineage, Manage Environment, and Manage Workflows. A red box labeled '1' highlights the hamburger icon at the top left of the sidebar. Another red box labeled '2' highlights the 'Explore and Visualize' option. A third red box labeled '3' highlights the 'RD2: Donors' tab in the top navigation bar. The main area displays a report titled 'Cortex Visual Analytics' with sections for 'Uplift Cutoff' (set to 0.00 to 35.77), 'Past donations(Gave)' (a 2x2 grid with values 0 and 1), and a large table titled 'Uplift Cutoff' with columns: ID, GiveLastYear, AmtLastYear, and UpLift. The table lists various donor IDs and their corresponding values.

ID	GiveLastYear	AmtLastYear	UpLift
2888556	1	20	35.76
2020200	1	10	35.76
2731579	1	10	35.76
2323228	1	40	35.76
2330142	1	20	35.76
2316469	0	0	35.76
2177633	1	30	35.76
2079261	0	0	35.76
2509806	0	0	35.76
2212981	0	0	35.75
2182540	0	0	35.75
2871745	1	20	35.75
2973108	1	250	35.75
2344587	0	0	35.75
2572751	0	0	35.74
2957189	0	0	35.74
2149093	0	0	35.74
2988236	0	0	35.74

# Explore & Visualize Data

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

The screenshot shows the Cortex Visual Analytics interface. On the left, a sidebar has 'Editing' at the top, followed by 'Data' (with a red box around it), 'Objects', 'Suggest', 'Outline', 'UpLift', 'Aggregated Measure', and 'Frequency Percent'. A dropdown menu is open under 'Data', showing '2 Stage Donor Prediction' (also with a red box around it), followed by 'CORTEX\_ROUND1\_TRAINDATA', 'CORTEX\_ROUND2\_TRAINDATA', 'RD 2 : Stage 1', 'RD 2 : Stage 2', 'SCORE\_RD1\_AMT', 'SCORE\_RD2\_AMTCONTACT', 'SCORE\_RD2\_AMTNOCONTACT', 'SCORE\_RD2\_GIVCONTACT', 'SCORE\_RD2\_GIVNOCONTACT', '+ Add data...', 'UpLift', 'Aggregated Measure', and 'Frequency Percent'. Step 4, 'Refresh 2 Stage Donor Prediction', is highlighted with a red box. To the right, there's a header with tabs: 'Cortex Visual Analytics', 'RD1: Donors', 'RD2: Explanation Conditional Give This Year', 'RD2: Explanation Conditional Amt This Year', 'RD2: Donors', and a '+' button. Below the tabs, there's a section titled 'Past donations (Gave)' with a table showing rows for 0 and 1. Further down, there are tables for 'GaveLastYear', 'AmtLastYear', and 'UpLift'. The 'UpLift' table includes columns for 'UpLift', '2.13', '4.49', '10.84', '2.13', '15.40', '2.13', '2.13', '2.13', '2.96', '18.93', '7.57', '12.78', '4.51', '2.13', '14.78', '2.13', and '2.13'. The 'UpLift' column has a tooltip '2.13' with an info icon.

# Report of Potential Donors

**Cortex Visual Analytics**

The screenshot shows the Cortex Visual Analytics interface. On the left, there's a sidebar with icons for Editing, Game Scenario, Data Dictionary, RD1: Explanation AmtThisYear, RD1: Donors, RD2: Explanation Conditional Give This Year, RD2: Explanation Conditional AmtThisYear, RD2: Donors, and a plus sign. The main area has tabs for RD1: Explanation AmtThisYear, RD1: Donors, RD2: Explanation Conditional Give This Year, RD2: Explanation Conditional AmtThisYear, RD2: Donors, and a plus sign. The RD1: Donors tab is active.

**Uplift Cutoff:** A slider from -66.98 to 0.00 to 35.77 with a value of 260.29.

**Past donations (Gave):** A 2x2 grid showing counts of 0 and 1.

**Donor List:**

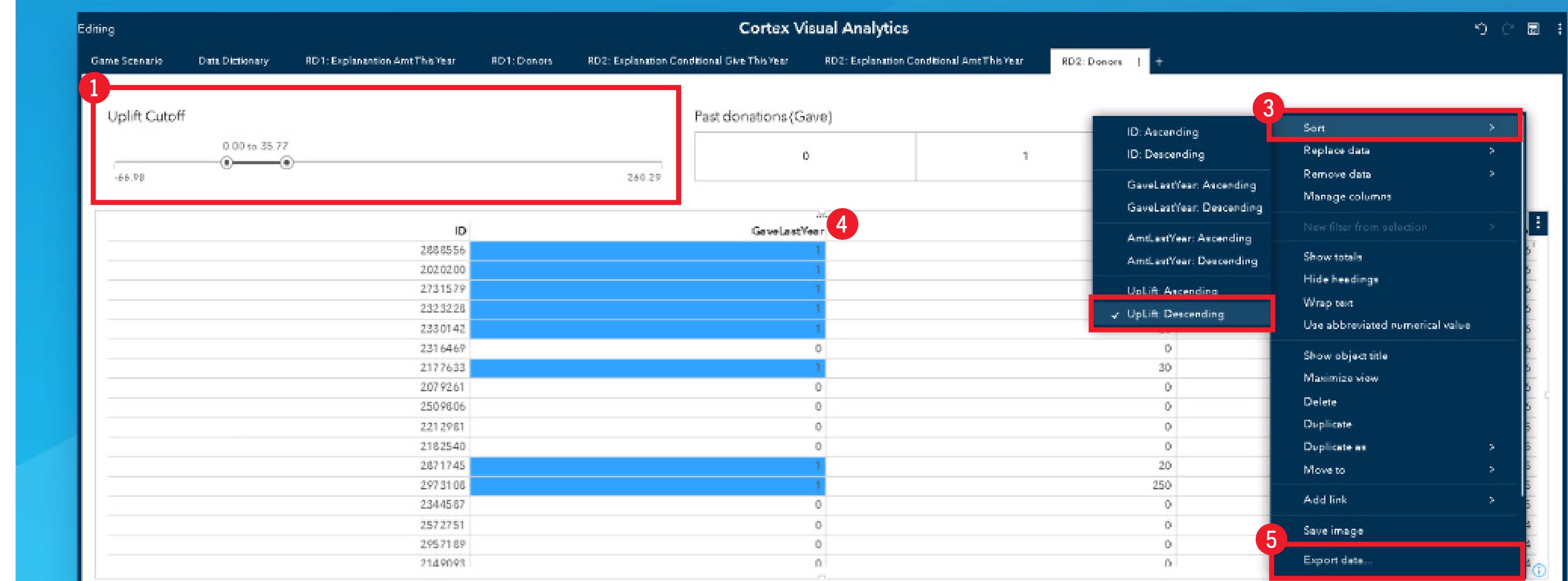
ID	GaveLastYear	AmtLastYear	Uplift
3000000	0	0	2.13
2999999	0	0	4.49
2999998	0	0	10.84
2999997	0	0	2.13
2999996	0	0	15.40
2999995	0	0	2.13
2999994	0	0	2.13
2999993	0	0	2.13
2999992	0	0	2.96
2999991	0	0	18.93
2999990	0	0	7.57
2999989	0	0	12.78
2999988	0	0	4.51
2999987	0	0	2.13
2999986	0	0	14.78
2999985	0	0	2.13
2999984	0	0	2.13



- Expected Donation with a call** = probability ( $p$ ) \* amount ( $m$ ): probability multiply by amount if the individual is contacted
- Expected Donation without a call** = probability ( $p$ ) \* amount ( $m$ ): probability multiply by amount if the individual is not contacted
- UPLIFT:** The difference between those two values of  $p*m$ , which is the value generated from the call

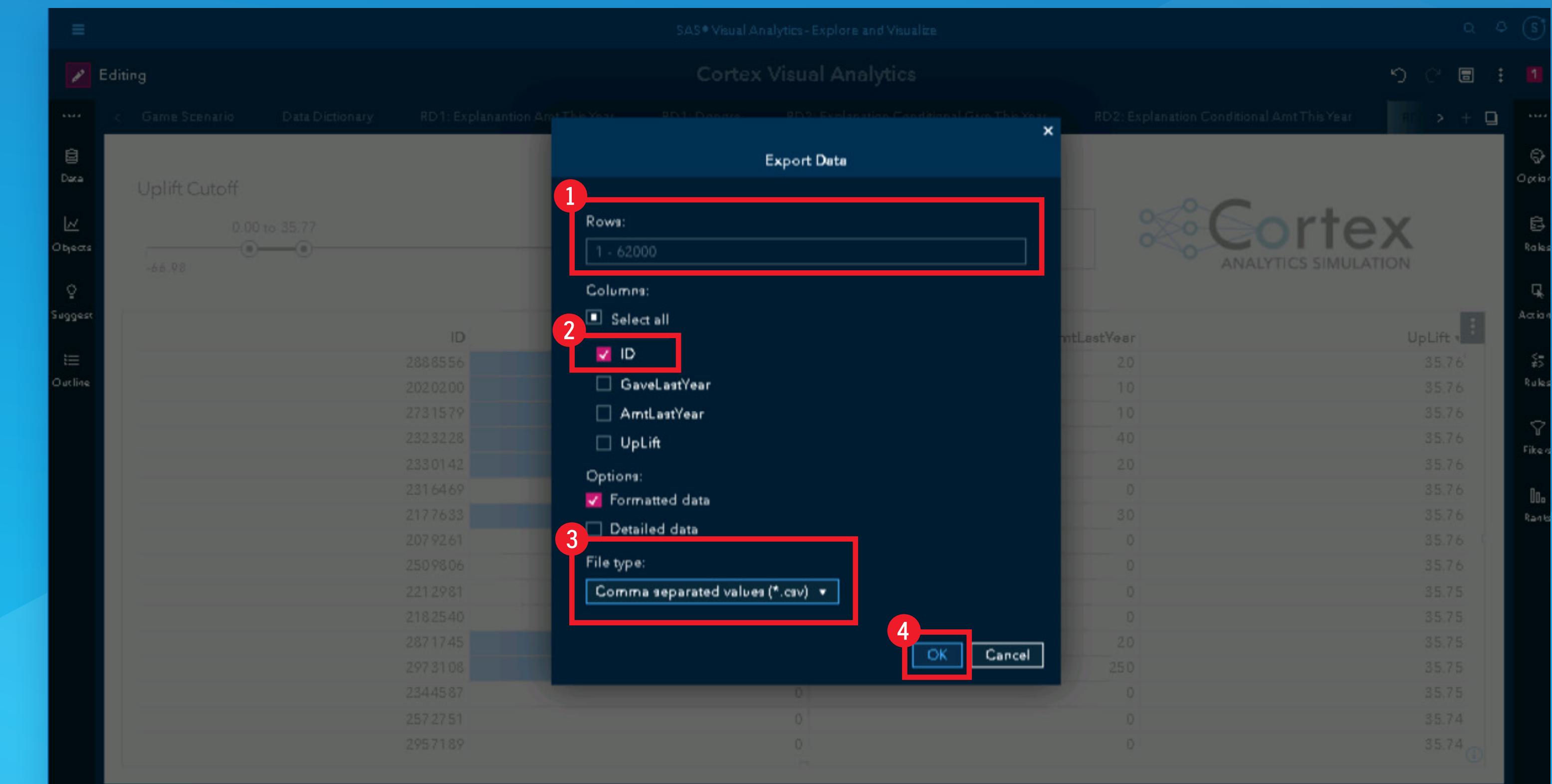
# Select Donors

1. Click on the 'Uplift Cutoff' and move the slider to choose your cutoff point
2. Right click on Uplift column
3. Click on 'Sort' and choose 'Uplift: Descending'
4. Right click on any column
5. Select 'Export Data'



# Export to .csv

1. Choose the rows you wish to export (e.g., 1-62000)
2. Select only the 'ID' column
3. Change the 'File type' to: Comma-separated values (\*.csv)
4. Click on 'OK'
5. Your .csv file will be in Download folder on the computer





**You are now ready to upload  
your solution (csv file)  
for Round 2 to  
the game leaderboard!**

Please refer to the Job Aids for a quick start on the game leaderboard