



## Predicting Graduate School Admissions at iLink University

### SAS Guided (Mini) Demo, Part 1

#### Purpose

This *SAS Guided (Mini) Demo* gets you started with ethical data analysis in SAS Model Studio. In seeking to select the fall 2025 class for the Masters in Analytics Program at iLink University – a completely made-up institution for higher learning – you’ll use the characteristics of successful applicants of the past. Then – in Part 2 – you are confronted with an important question: *just because you can run a model, should you?* Towards answering that important question, you’ll learn how built-in fairness and bias assessment tools in SAS Model Studio help you refine your modeling and get you closer to a solution that is both equitable and empirically rigorous.

#### Scenario

In this *SAS Guided (Mini) Demo, Part 1*, we’ll pretend that we’re all on our first day as Predictive Modelers in the Admissions Department at iLink University. Moreover, our modeling will help decide who gets into the fall 2025 M.S. in Analytics Program. So – congrats on the new job – and thank you for joining us on this analytics adventure!

#### Learning Objectives

We’ll cover a lot of ground in this *SAS Guided (Mini) Demo*, including:

- Creating SAS Model Studio projects
- Running a series of competing predictive models and evaluating which performs “best”
- Choosing a champion model based upon your chosen fit statistics (it may just be the default)

#### Software

This *SAS Guided Demo* uses SAS Viya, version 2025.03 Stable. Other versions can be used, but some screenshots will differ.

#### Prerequisites

A basic understanding of predictive modeling and machine learning is helpful. Additionally, some experience using the SAS Viya ecosystem is useful, but not required.

## Table of Contents

Predicting Graduate School Admissions at iLink University .....	1
Purpose .....	1
Scenario.....	1
Learning Objectives.....	1
Software .....	1
Prerequisites .....	1
Predicting Graduate School Admissions .....	3
Overview .....	3
Part 1: Business as Usual.....	3
Overview .....	3
Create a SAS Model Studio Project .....	4
Explore the Data.....	8
Run Predictive Models .....	14
Crown a champion model .....	21
Feel a bit cheated because I cut this short? .....	22

# Predicting Graduate School Admissions

## Overview

Welcome to iLink University! As a recently hired Predictive Modeler in our Admissions Department, you will help us decide who gets into our fall 2025 M.S. in Analytics Program. Yay!

As this is your first day on the job, I'll guide you through some useful Visual Data Mining and Machine Learning tools in SAS Model Studio. In Part 1 – you'll examine the data and select the incoming 2025 class based upon the tried-and-true process we've used in previous years. Put simply: we use the characteristics from successful applicants to determine the incoming class. In terms of modeling based upon our historical data, you'll use a variable called *Admissions* – which is a 0/1 variable that indicates whether we offered that student admissions to our program. This 0/1 variable can also be interpreted as “Yes” / “No” and the binary nature of this outcome makes it very useful for predictive modeling and machine learning tools.

In Part 2, a colleague will show you some new + useful features in SAS Model Studio, which fall under the category of **Fairness and Bias assessment** tools. These exciting features allow us to ask the question: “just because we can run a model, should we?” As you learned from your coursework, using historical data to predict new cases can allow existing bias to persist. And since we want the fairest admissions process possible, you will use the Fairness and Bias assessment tools to adjust your model, if needed.

Let's get started!

## Part 1: Business as Usual

### Overview

Ready, set, go!

I know you're ready to get to work, because you're a Hacker at heart. But before just plunging into the modeling, we should start with a better understanding of the data. For the modeling, we have 1000 admission decisions from the past 5 years. Let's understand the data a bit better, by examining the data dictionary below:

Variable	Label	Definition
Admitted	Admitted (Yes=1)	When Admitted = 1, the student is offered admissions into the iLink University M.S. in Analytics Program
Analytics_Work_Experience	Analytics Work Experience	Number of years working in the field of analytics.
Country_Region	Country Region	Region of the world applying from

Cultural_Identity	Cultural Identity	Cultural identity
Gender	Gender Identity or Gender at Birth	Gender identity or gender at birth
ID	Application ID	Application ID
Legacy_Admission	Legacy Admission	Legacy admission means that either (1) the student's parents attended the university or (2) they previously completed another degree at iLink University.
Mission_Statement	Mission Statement	Optional mission statement (maximum of 100 words)
Standardized_Test_Score	Standardized Test Score	Standardize test score (Z-score)
Strength_of_Recommendation	Strength of Recommendations	Overall strength of recommendations (0 to 5, higher is better)
Undergrad_Degree	Undergraduate Degree Category	Undergraduate degree category
Years_Work_Experience	Years Work Experience	Total years of work experience, all fields

*Admissions* is the outcome variable that we really care about. As noted, it is a binary variable that can be considered a yes/no variable. The other variables can be used to help us select the top candidates for admissions in the incoming class, based upon historical precedent.

With a better understanding of the data, the rest of the project flows with steps the data analyst in you (generally) knows by heart:

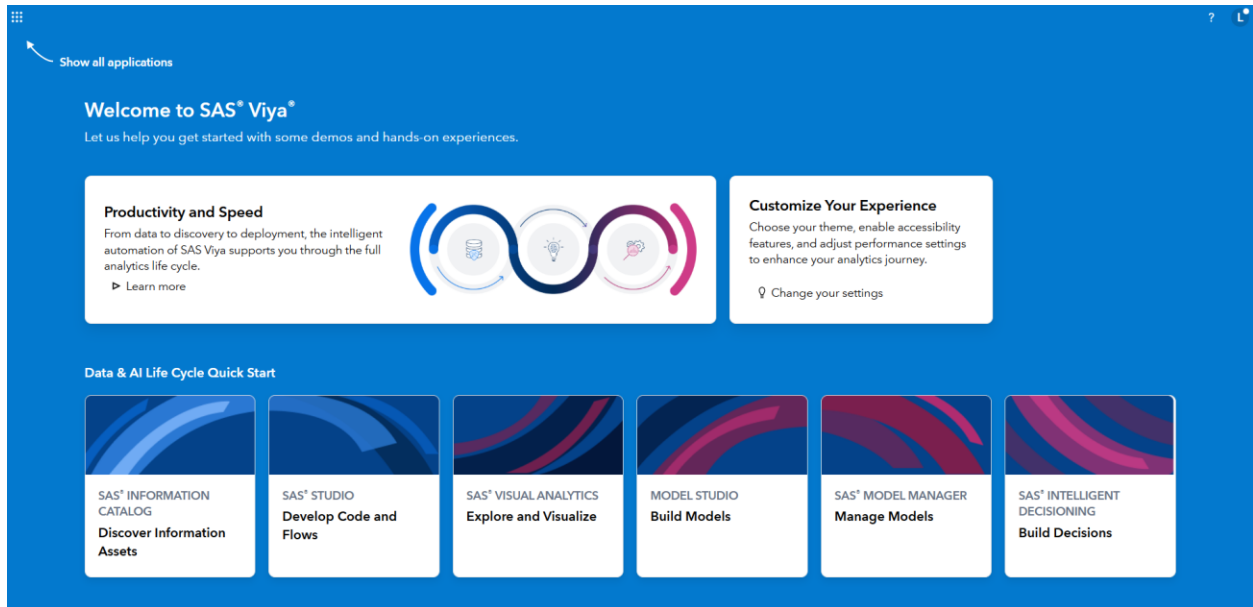
- Start a SAS Model Studio project.
- Explore the data and get to know it A LOT better.
- Run a bunch of predictive models using the Visual Data Mining and Machine Learning tools in SAS Viya.
- Find the best model predicting the historical sample. Crown it the champion.


Let's do it!

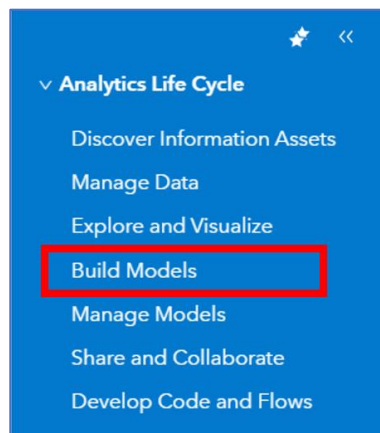
### Create a SAS Model Studio Project

A good data adventure often starts with software.

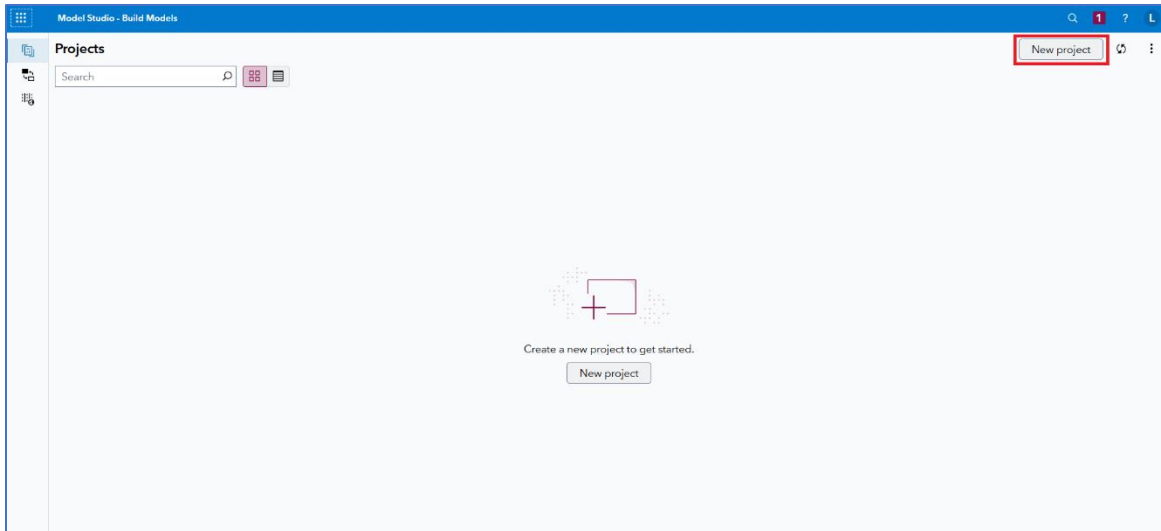
- As such, log in to your SAS Viya Environment.
- Once in, you'll land on the **Welcome to SAS Viya** page and it will look like the following:



- But, we won't linger here long – as we want to get right into **SAS Model Studio**. From the top left corner, select the **Applications menu**  and click **Build Models** from the **Analytics Life Cycle** options.



- Welcome to **SAS Model Studio**! Your environment may already have shared projects out there. For now, ignore them. Instead find and click that **New Project** button:



- The **New Project** window will step you through setting up a new project. To start, give it a useful Name, such as *Admissions Analysis*. Keep the **Type** of project at the default, which is *Data Mining and Machine Learning*. Next, find the **Browse** button, under the **Template** section. Those first three steps, succinctly summarized:

- Click that **Browse** button – and check out all those pre-built templates! When you're happy with your research, find that **Intermediate template for class target template** and select it. Then click **OK**:

**Browse Templates**

Filter

Template Name	Description	Owner	Last Modified
Feature engineering template	Data mining pipeline that performs feature engineering.	SAS Pipeline	March 14, 2025 at 10:30:51 PM
Intermediate template for class target	Data mining pipeline that extends the basic template for a class target by adding a stepwise logistic regression model and a decision tree.	SAS Pipeline <sup>1</sup>	March 14, 2025 at 10:30:52 PM
Intermediate template for interval target	Data mining pipeline that extends the basic template for an interval target by adding a stepwise linear regression model and a decision tree.	SAS Pipeline	March 14, 2025 at 10:30:52 PM

2 OK Cancel

- Sidebar: why “class target”? Well, Admissions is our target variable for modeling. It has two values 1=yes, we admitted them and 0=no, we did not extend an offer. A yes/no variables opens up a bunch of useful machine learning tools to us.
- Now let’s add some data to your **New Project**. Under that **Data** section, click that **Browse** button:

**New Project**

Name: \*  
Admissions Analysis

Type: \*  
Data Mining and Machine Learning

Template:  
Intermediate template for class target Browse

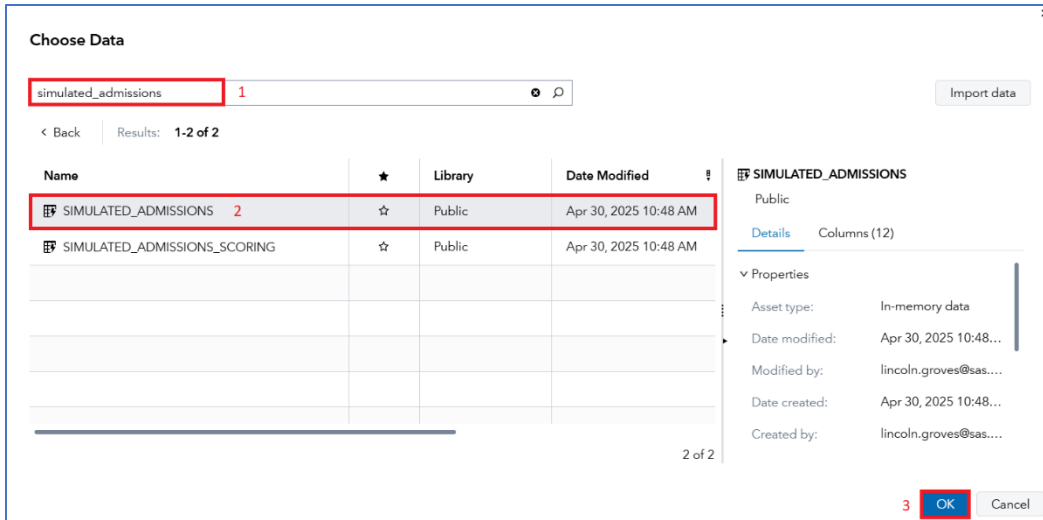
Data:  
Browse

Description:

Advanced

Save Cancel

- In the Choose Data window, find the *Simulated\_Admissions* data set. You can either scroll through the list of available data or type *simulated\_admissions* in the **Search bar**. Then select *Simulated\_Admissions* and then **OK**. Those 3 steps:



- And your **New Project** details should be good to go! Click **Save** so that your new project can be created:

**New Project**

Name: \*  
Admissions Analysis

Type: \*  
Data Mining and Machine Learning

Template:  
Intermediate template for class target Browse

Data: \*  
Public.SIMULATED\_ADMISSIONS Browse

Description:

Advanced

Save Cancel

- And your first SAS Model Studio project is all set up. Great job!

## Explore the Data

We resume our adventure with a SAS Model Studio project officially created. Next up is to explore the data, just to make sure we really understand what we're modeling.



- If you've never been here before, let me be the first to say: *Welcome to SAS Model Studio!* You'll love it! And you will always land on the **Data** tab. Why? Because this a great place to get to know your data a bit better – rather than just launching into the model (because, as they say: garbage in = garbage out in modeling!). You can scroll to the right to examine all the variable attributes. For example, check out the **Admitted** variable and scroll right:

Model Studio - Build Models

< Projects

Admissions Analysis

Data Pipelines Pipeline Comparison Insights

⚠ You must assign a variable with the role of Target in order to run a pipeline.

Filter

<input type="checkbox"/>	Variable Name ↑	Assess for Bias	Level	Order	Comment	Number of Lev...	Missing	Minimum	Maximum	Mean
<input checked="" type="checkbox"/>	Admitted 2		Binary	Default		2	0.0000	0.0000	1.0000	0.2000
<input type="checkbox"/>	Analytics_Work_Exp erience		Binary	Default		2	0.0000	0.0000	1.0000	0.5750
<input type="checkbox"/>	Country_Region		Nominal	Default		3	0.0000			
<input type="checkbox"/>	Cultural_Identity		Nominal	Default		5	0.0000			
<input type="checkbox"/>	Gender		Binary	Default		2	0.0000			
<input type="checkbox"/>	ID		Interval	Default		>254	0.0000	1.0000	1,000.0000	500.5000
<input type="checkbox"/>	Legacy_Admission		Binary	Default		2	0.0000	0.0000	1.0000	0.1140
<input type="checkbox"/>	Mission_Statement		Nominal	Default		>254	15.0000			
<input type="checkbox"/>	Standardized_Test_5 core		Interval	Default		>254	0.0000	-3.1765	3.1263	0.0948
<input type="checkbox"/>	Strength_of_Recom mendation		Interval	Default		>254	0.0000	0.5081	5.0000	3.3612
<input type="checkbox"/>	Undergrad_Degree		Nominal	Default		5	0.0000			
<input type="checkbox"/>	Years_Work_Experie nce		Interval	Default		>254	0.0000	0.0000	8.2294	2.5288

- From this succinct summary, we can see that *Admitted* has two values – which are 0 and 1 – with an average of 0.2000 or 20%. So, 20% of the applicants are admitted to the iLink University. That's helpful information!
- As noted by the warning, we must assign the outcome variable:

⚠ You must assign a variable with the role of Target in order to run a pipeline.

- And you know what, we're already there! Select **Admitted** and ensure that the **Properties** tab is active. Then you can change the **Role** to **Target**. Like so:

Projects

### Admissions Analysis

Data Pipelines Pipeline Comparison Insights

Filter

Variable Name	Label	Type	Role	Assess for Bias	Level
<input checked="" type="checkbox"/> Admitted	Admitted (Yes=1)	Numeric	Target		Binary
<input type="checkbox"/> Analytics_Work_Experience	Analytics Work Experience	Numeric	Input		Binary
<input type="checkbox"/> Country_Region	Country Region	Character	Input		Nominal
<input type="checkbox"/> Cultural_Identity	Cultural Identity	Character	Input		Nominal
<input type="checkbox"/> Gender	Gender Identity or Gender at Birth	Character	Input		Binary
<input type="checkbox"/> ID	Application ID	Numeric	ID		Interval
<input type="checkbox"/> Legacy_Admission	Legacy Admission	Numeric	Input		Binary
<input type="checkbox"/> Mission_Statement	Mission Statement	Character	Text		Nominal
<input type="checkbox"/> Standardized_Test_Score	Standardized Test Score	Numeric	Input		Interval
<input type="checkbox"/> Strength_of_Recommendation	Strength of Recommendations	Numeric	Input		Interval

**Admitted**

Role: Target

Level: Binary

Specify the Target Event Level

Order: Default

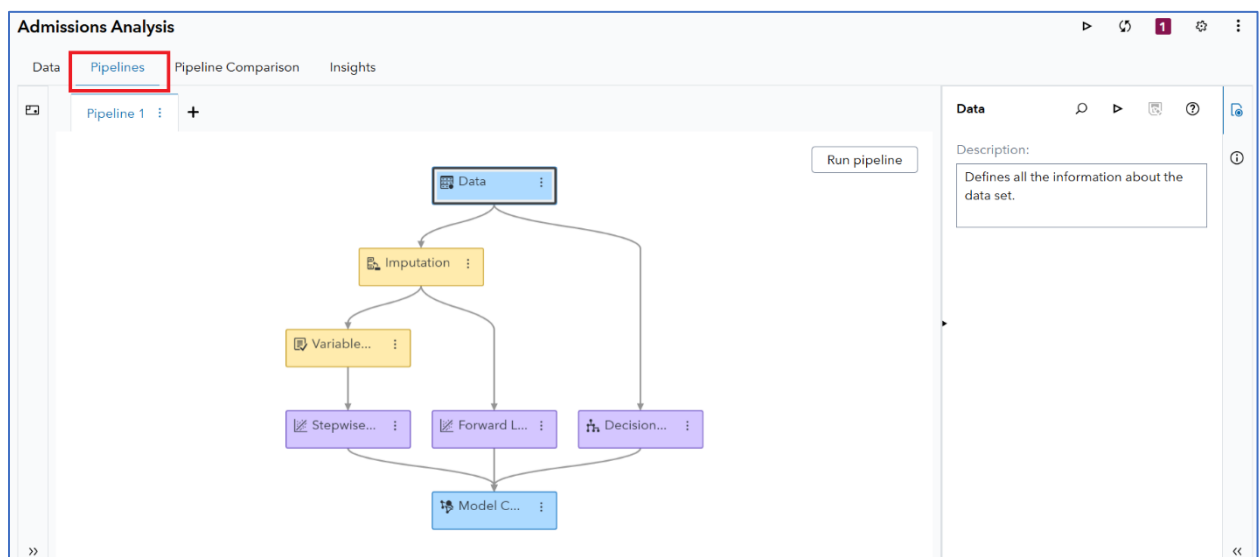
Transform:

Impute:

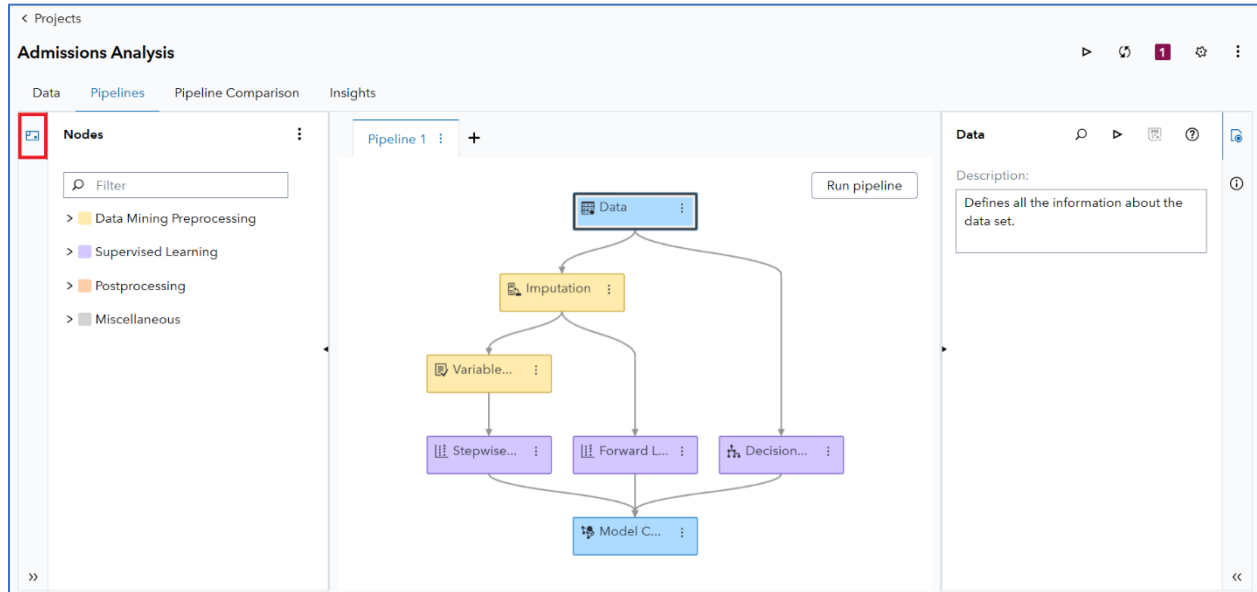
- Warning goes away! Yay! Spend a bit more time examining the variables, just so you get a little more comfortable. I don't see anything else that needs to be adjusted right now:

<input type="checkbox"/>	Variable Name	Label	Type	Role	Assess for Bias	Level	Order
<input type="checkbox"/>	Admitted	Admitted (Yes=1)	Numeric	Target		Binary	Default
<input type="checkbox"/>	Analytics_Work_Experience	Analytics Work Experience	Numeric	Input		Binary	Default
<input type="checkbox"/>	Country_Region	Country Region	Character	Input		Nominal	Default
<input type="checkbox"/>	Cultural_Identity	Cultural Identity	Character	Input		Nominal	Default
<input type="checkbox"/>	Gender	Gender Identity or Gender at Birth	Character	Input		Binary	Default
<input type="checkbox"/>	ID	Application ID	Numeric	ID		Interval	Default
<input type="checkbox"/>	Legacy_Admission	Legacy Admission	Numeric	Input		Binary	Default
<input type="checkbox"/>	Mission_Statement	Mission Statement	Character	Text		Nominal	Default
<input type="checkbox"/>	Standardized_Test_Score	Standardized Test Score	Numeric	Input		Interval	Default
<input type="checkbox"/>	Strength_of_Recommendation	Strength of Recommendations	Numeric	Input		Interval	Default
<input type="checkbox"/>	Undergrad_Degree	Undergraduate Degree Category	Character	Input		Nominal	Default
<input type="checkbox"/>	Years_Work_Experience	Years Work Experience	Numeric	Input		Interval	Default

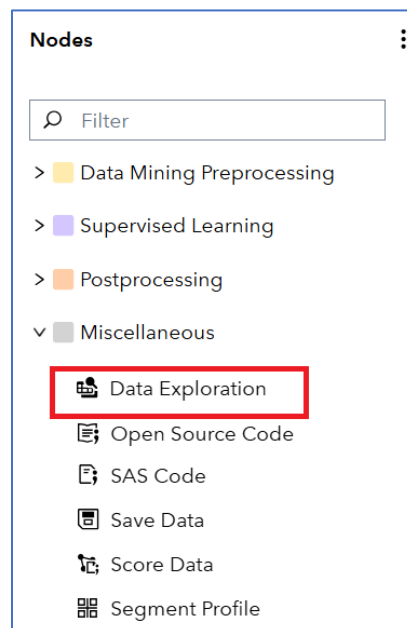
- Let's proceed to the **Pipeline** tab to unveil your first **SAS Model Studio** pipeline:



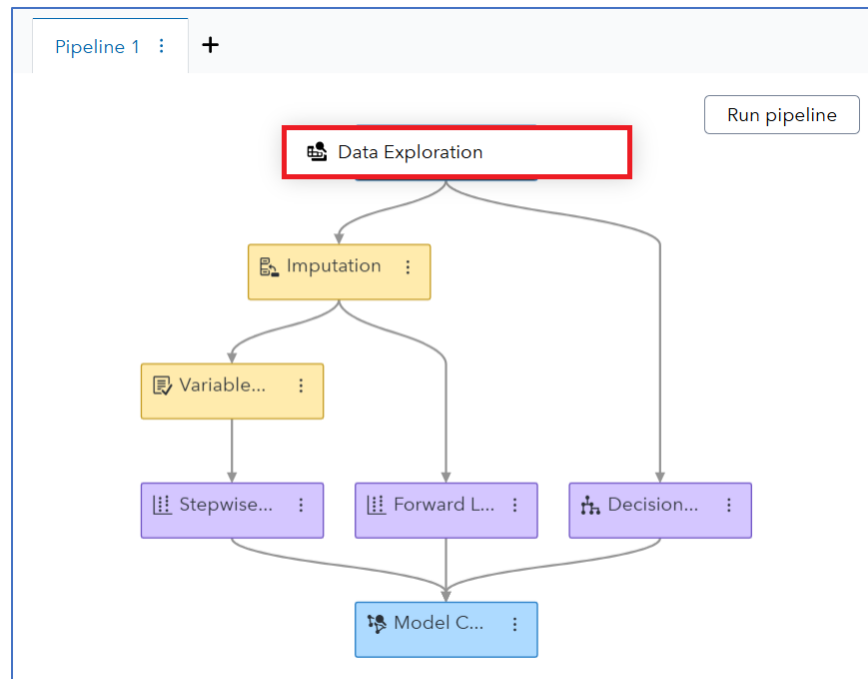
- Now we could just click **Run pipeline** and move straight on to the results from the three models from our **Intermediate Template** (which are Stepwise Logistic Regression, Forward Logistic Regression, and Decision Tree, respectively). But let's explore the data just a bit further. In your pipeline, find the **Nodes** icon and expand it:



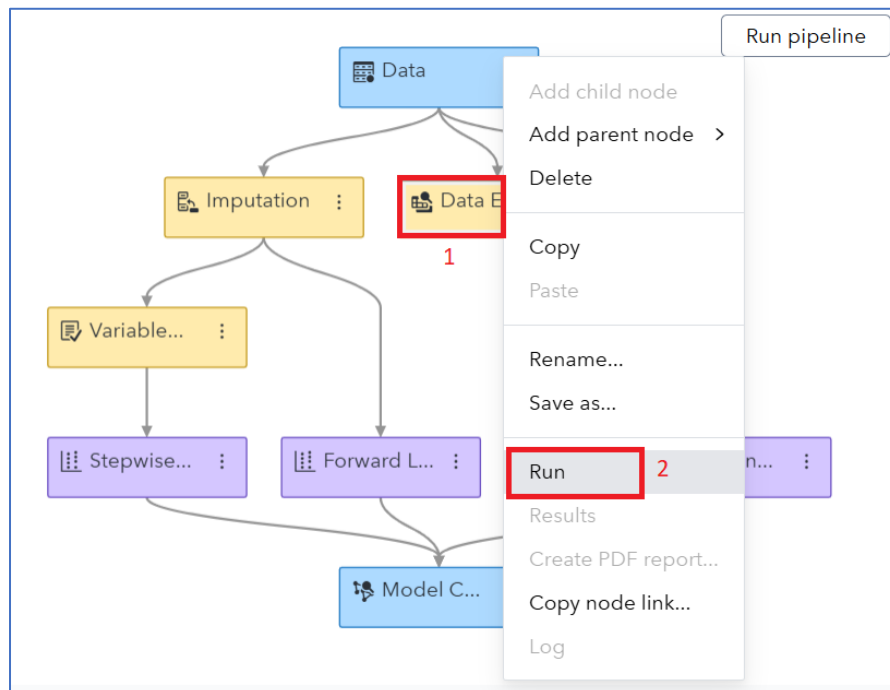
- Please explore the **Nodes** options just a little bit. Then find the **Data Exploration** node under the **Miscellaneous** section, here:



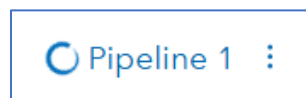
- Go ahead and drag-and-drop that **Data Exploration** node on top of the **Data** node, like so:



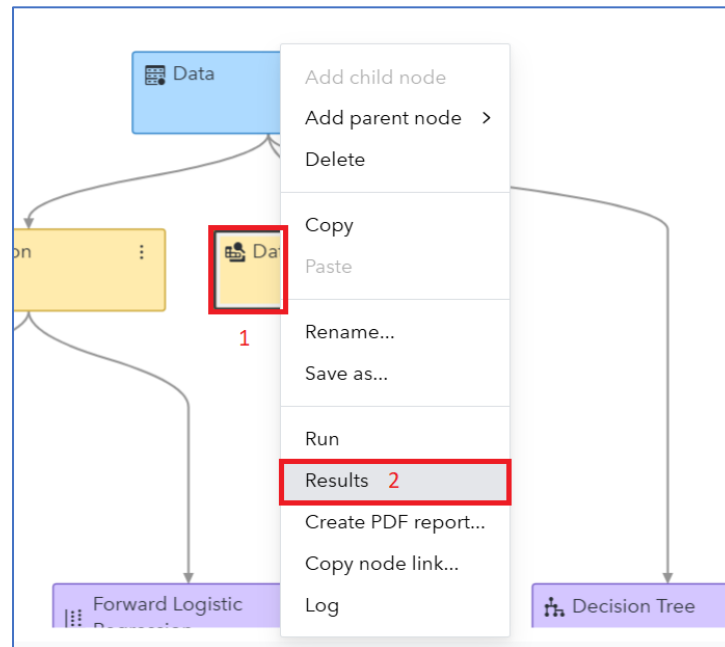
- Now look at that expanded pipeline. Right-click on the **Data Exploration** node and select **Run**:



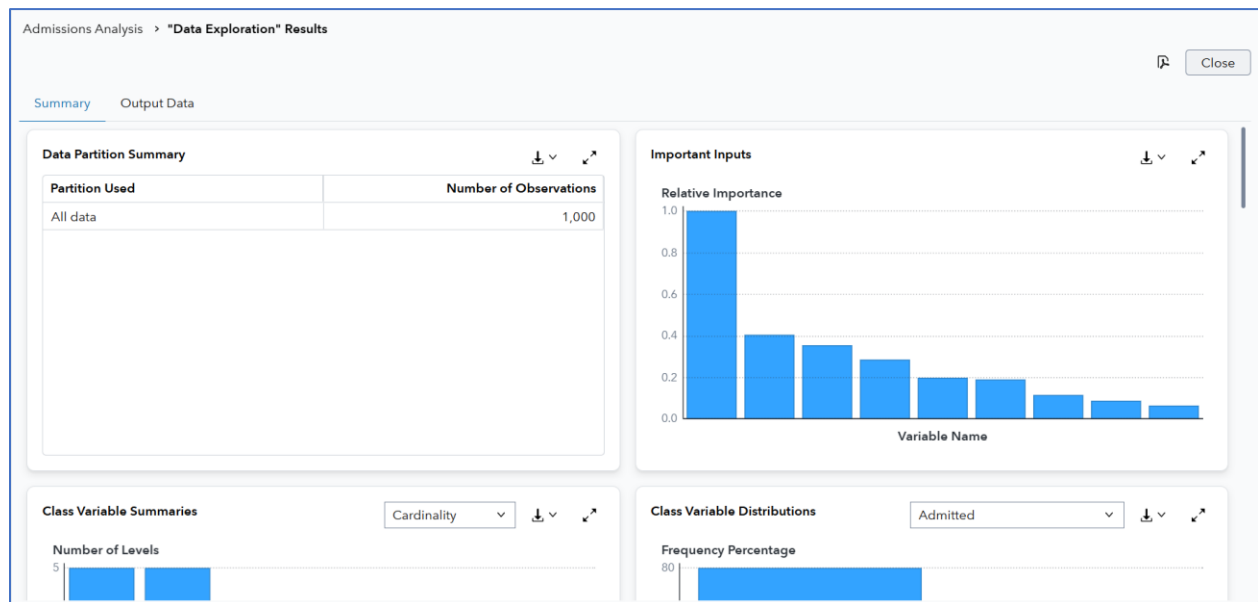
- The swirling magic means things are happening:



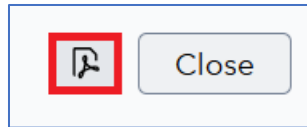
- A green checkmark means that the node has finished running. Right-click on the **Data Exploration** node when it is finished and select **Results**:



- A new window pops up with a lot of results to get you further acquainted with your data:



- Check it all out! There is a lot of really useful information in this summary, including missing values, correlations, interval moments, etc. So, spend a little bit of time here. And if you'd like to print a copy of the output to share with family, you're just a click away. Like here:



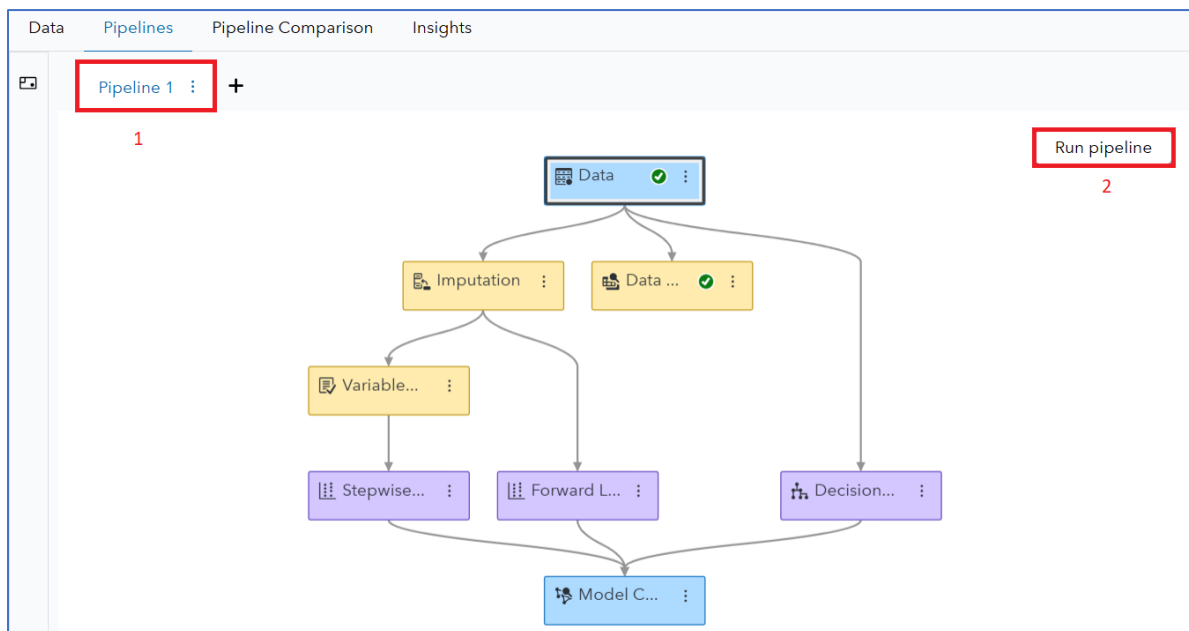
- Click that **Close** button when you're satiated. And congratulations – the **Explore the Data** section is completed!

## Run Predictive Models

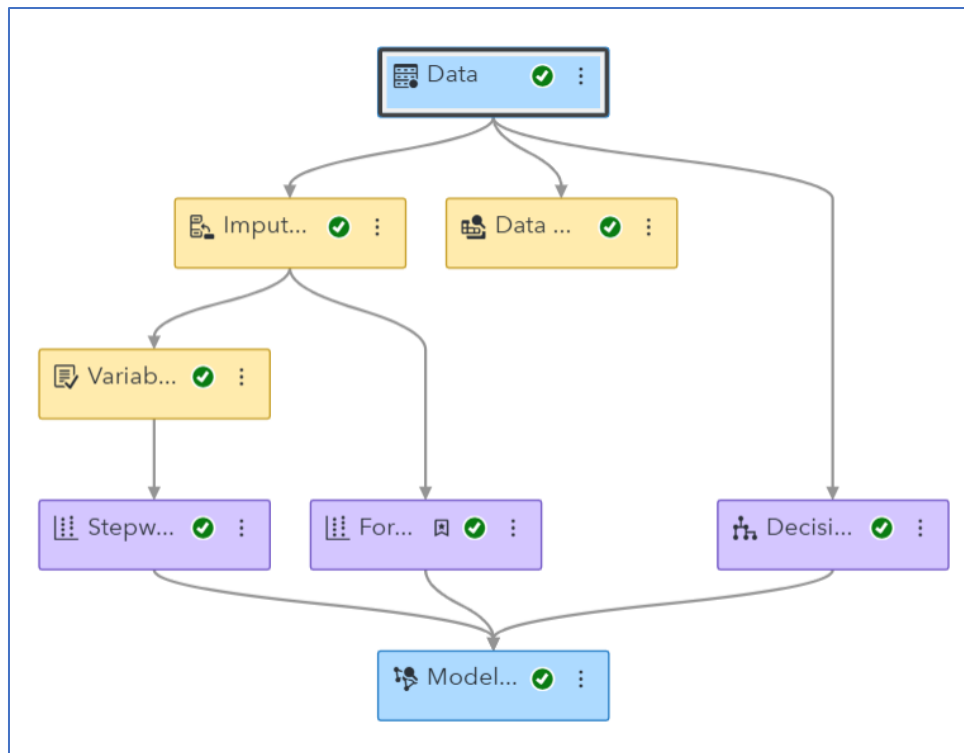
Let's be honest. Modeling is fun. Particularly with no-code tools, like SAS Model Studio. We can just select a bunch of models, hit a button, grab a coffee, and then come back to interpret the models. We can then make a couple of adjustments. Then it's run... rinse... repeat.

Let's do exactly that!

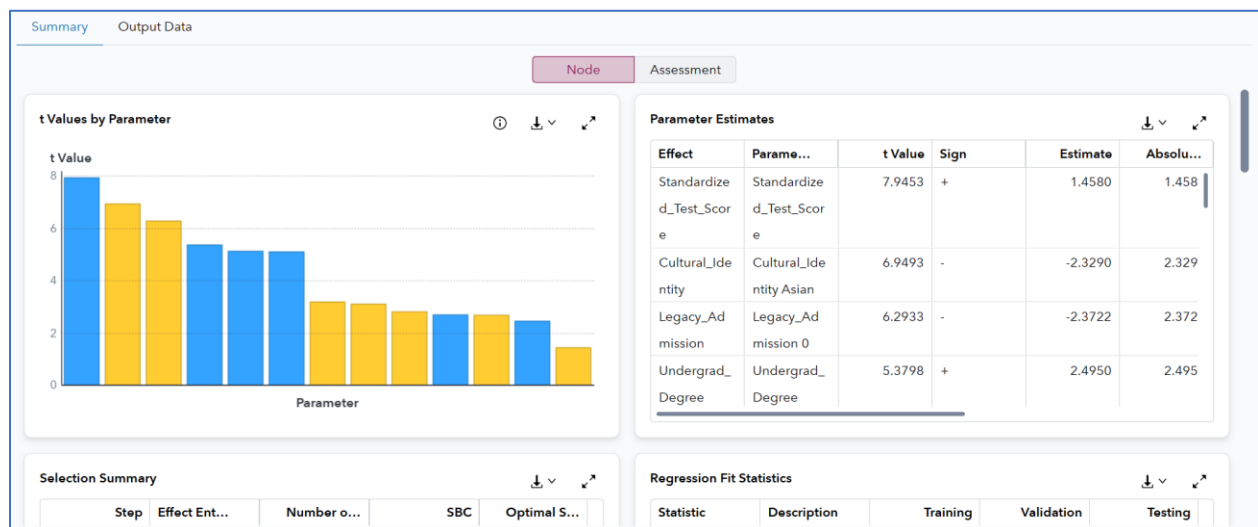
- If you've somehow navigated elsewhere, return to *Pipeline 1*. And click **Run pipeline** to run the models with their default settings:



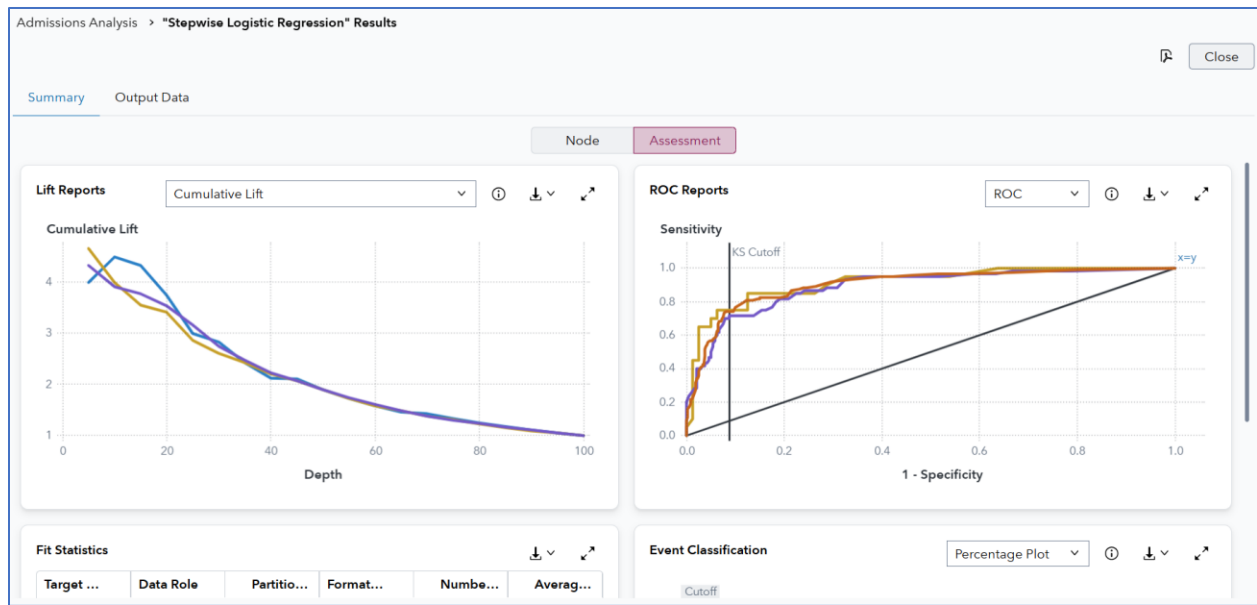
- Green checkmarks everywhere means that (modeling) life is good:



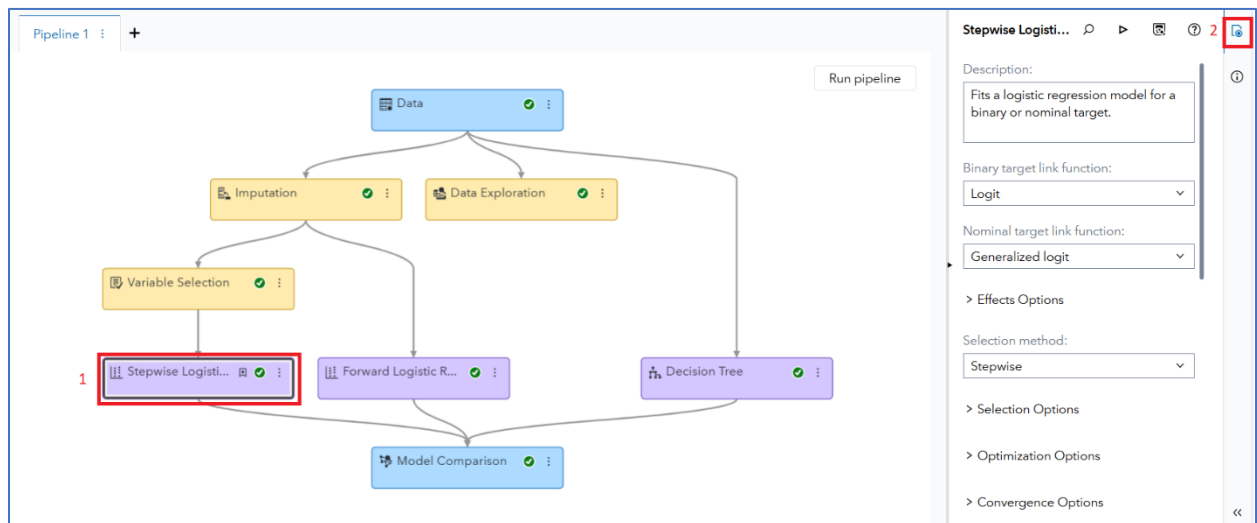
- To explore the models individually, right click on a modeling node and select **Results**. For example, here are the **Results** for the **Stepwise Logistic Regression** model:



- The **Node** tab gives you traditional statistical measures like parameter estimates, selection summary (if applicable), regression fit statistics, and score code.
- Click **Assessment** for an interactive dashboard of MANY model assessment fit statistics and tools:

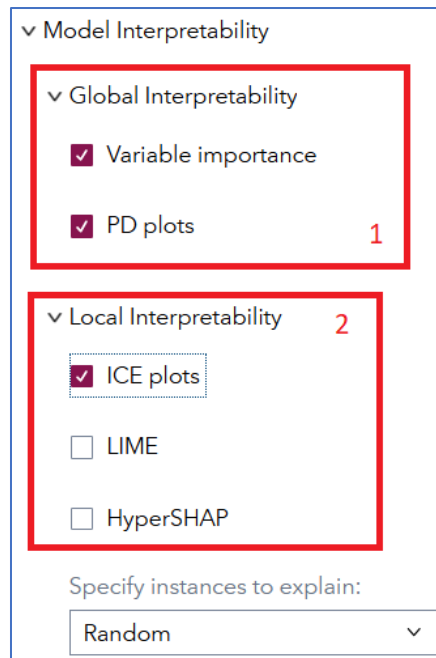


- Want to see even more measures of how your model is performing? And make your model a little more easy to explain? Well, SAS Model Studio also has you covered! Click on the **Stepwise Logistic Regression** node and activate the **Node options**, like so:



- Scroll down to **Post-training Properties**. Under **Model Interpretability**, expand the **Global Interpretability** and **Local Interpretability** sections. Then choose the following options:

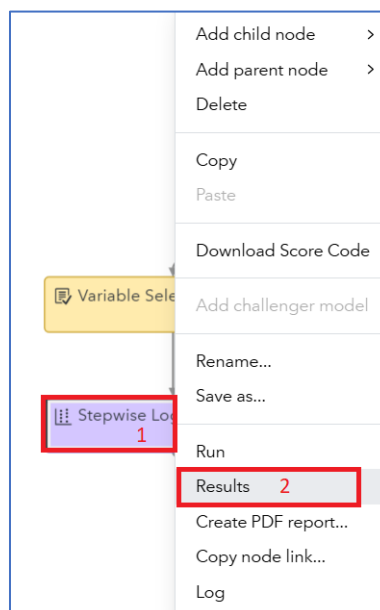




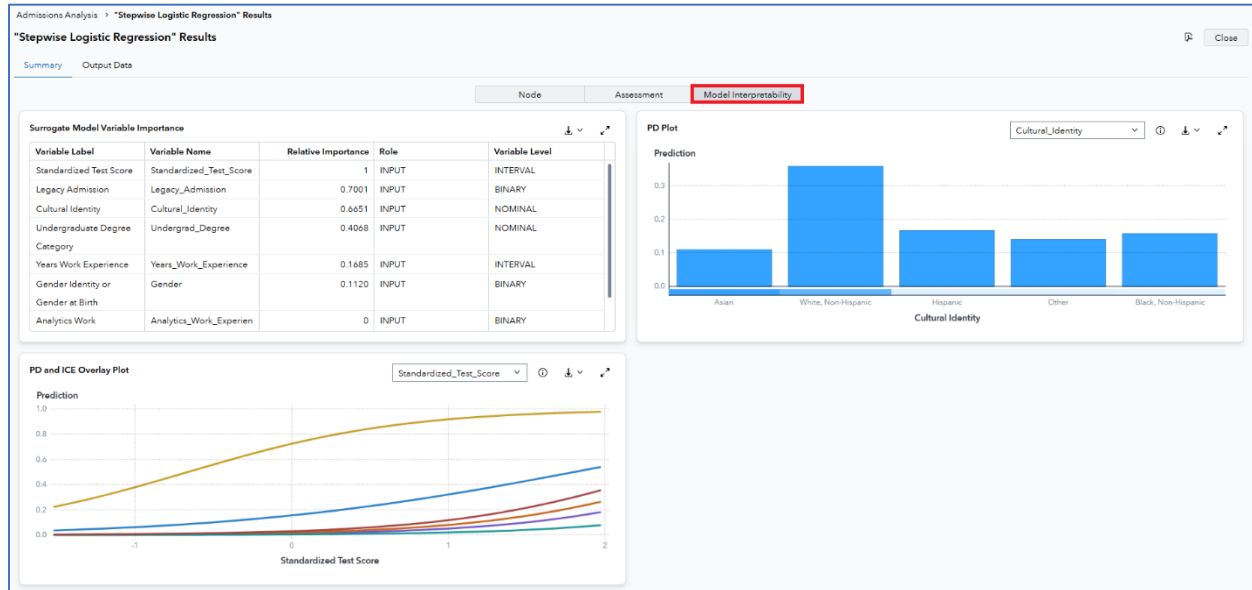
- After selecting those **Model Interpretability** options, you'll see that the **Stepwise Logistic Regression** node has reset itself:



- Click **Run pipeline** to just run it all again. When it's finished, open the **Results** to the **Stepwise Logistic Regression** model with two clicks:



- You've now got a new tab – **Model Interpretability**! Here you can examine (1) which variables are most important in determining admissions via **Surrogate Model Variable Importance**, (2) how admissions varies across variables in the final model in the **PD Plot**, and (3) how admissions varies by groups in the **PD and ICE Overlay Plots**.



- That's a whole lot of results to digest. So, spend a bit of time doing exactly that: digesting. And click **Close** when you're done with the Stepwise Logistic Regression model.
- You may think, *do I need to explore all the models individually and then choose the "best" model in this pipeline?* Nope. You sure don't. The **Model Comparison** node handles that for you. So right-click on that node and select **Results** to compare the three models in the pipeline:

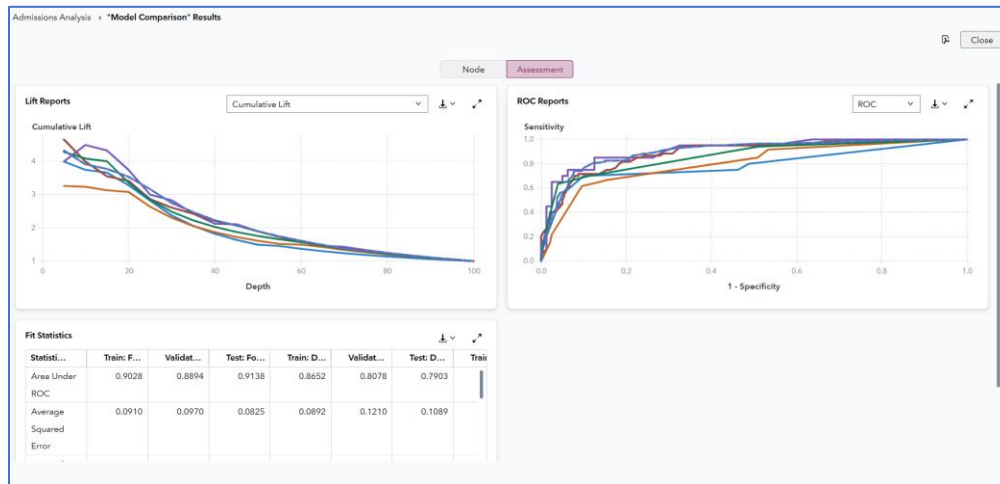
The screenshot shows the 'Model Comparison' Results window. It contains a table of model performance metrics and a properties section.

Cha...	Name	Algorithm Name	↓ KS (Youden)	Accuracy	Averag...	Area U...	Cumula...	Cumula...	Cutoff	Data Role	Depth	F1 Sc
★	Forward Logistic Regression	Logistic Regression	0.7250	0.9100	0.0825	0.9138	4.5000	45	0.5000	TEST	10	0.7
	Stepwise Logistic Regression	Logistic Regression	0.7250	0.9100	0.0825	0.9138	4.5000	45	0.5000	TEST	10	0.7
	Decision Tree	Decision Tree	0.6000	0.8700	0.1089	0.7903	3.7500	37.5000	0.5000	TEST	10	0.6

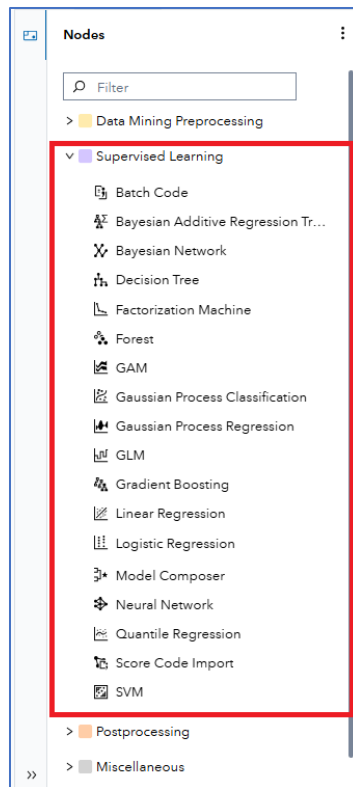
**Properties**

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)
selectionCriteriaInterval	Average squared error
selectionTable	Test
selectionDepth	10
cutoff	0.50

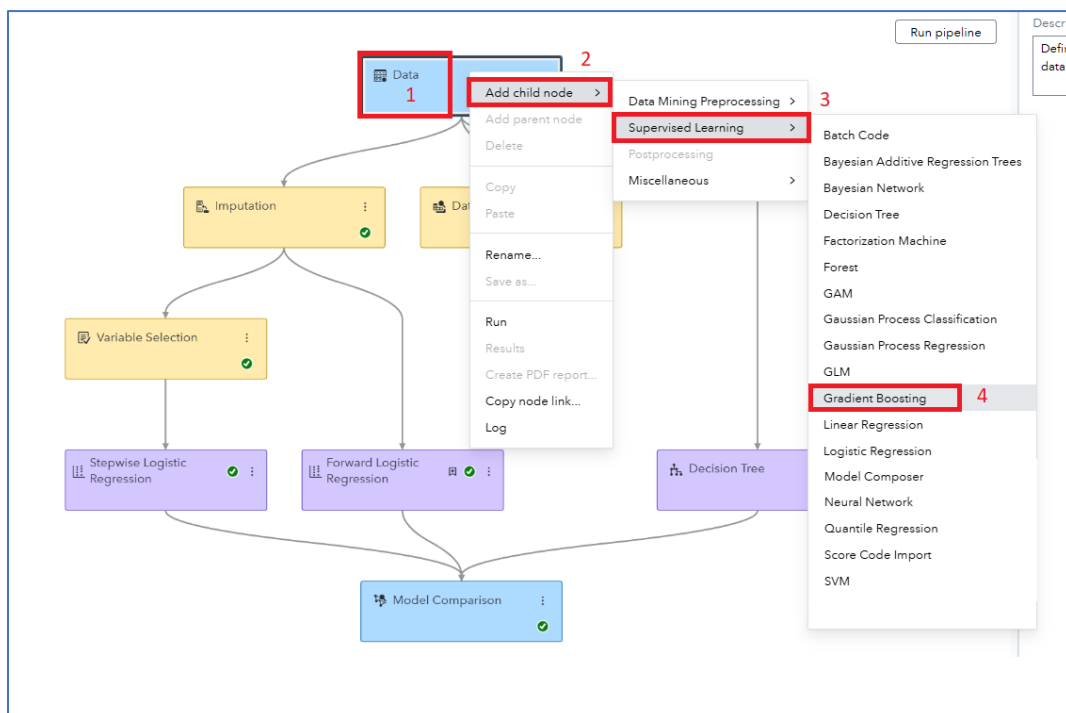
- With the **KS (Youden) statistic** as our default selection criteria for a classification target, we can see that the Forward Logistic Regression and Stepwise Regression models appear to select the same model, with a KS (Youden) statistic of 0.7250. The Decision tree performs far worse. You can also click on the **Assessment** tab to explore a variety of other model selection criteria:



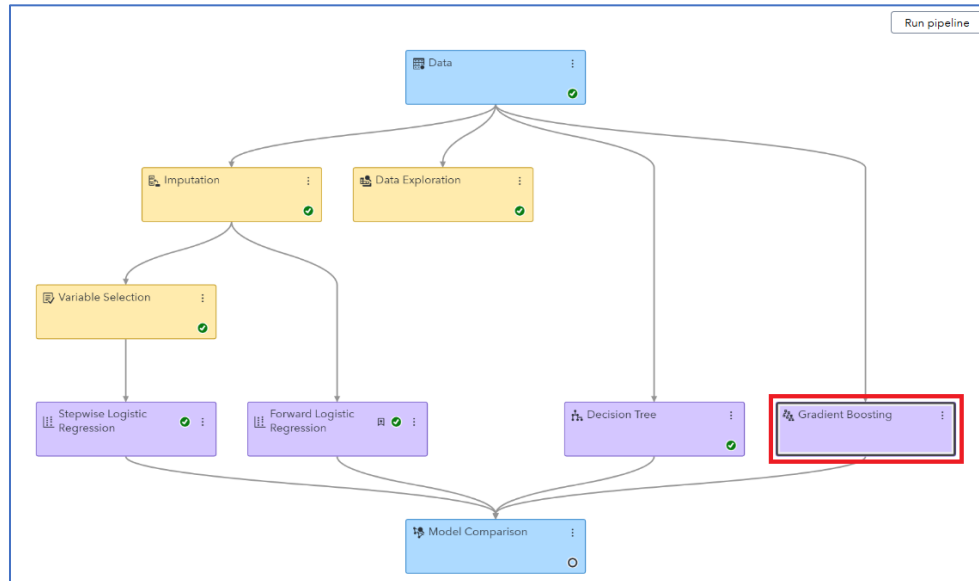
- Explore a bit, then hit **Close** when you're done.
- But... in my preamble didn't I promise machine learning models? I did. I sure did. So, let's add one true machine learning model to our pipeline. You can then tweak and add more models to your pipeline later.
- Go back to your original pipeline. Expand the **Supervised Learning** node to explore all the modeling options available to you:



- Yup – it’s a lot! Now I’ll show you another way to add a node to the pipeline... one that isn’t drag-and-drop (because that’s not everyone’s cup of tea). Find the **Data** node inside your pipeline. Right-click, select **Add child node >> Supervised Learning >> Gradient Boosting**. Like so:



- Gradient Boosting is a proper machine learning model. And shows up in the pipeline here:

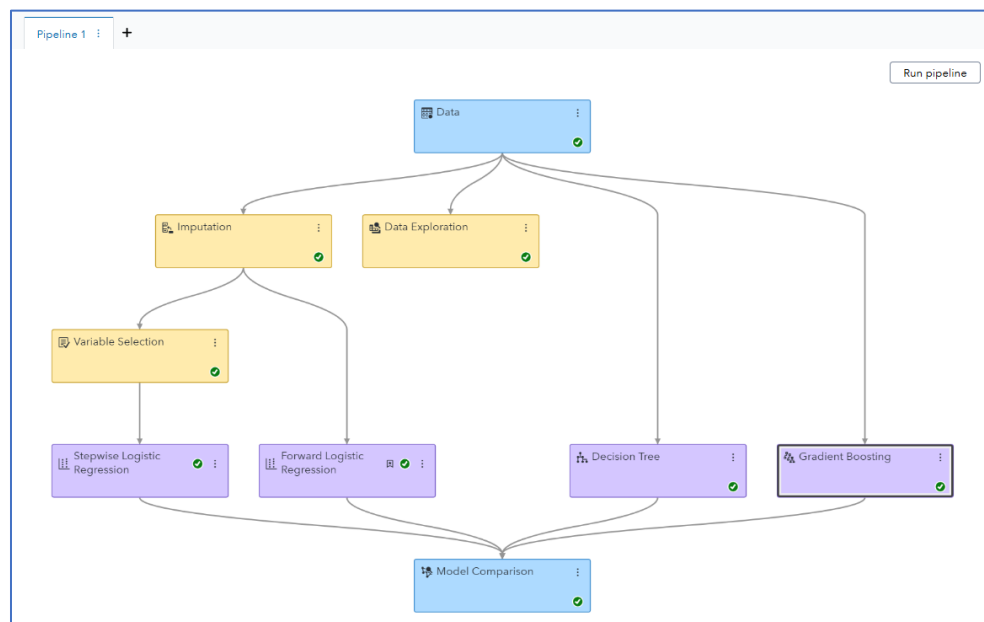


- Click **Run pipeline**. And then I'll see you in the next section. Great work!

### Crown a champion model

Modeling is fun. But we're not just doing modeling for modeling's sake. We're actually on a mission. And that's to choose the incoming class of 2025. Let's get on it!

- If you got lost, navigate back to *Pipeline 1*. You should have models to compare and a champion waiting under **Model Comparison**:



- A right-click and selecting **Results** on the **Model Comparison** node yields:

Cha...	Name	Algorithm Name	KS (Youden)	Accuracy	Averag...	Area U...	Cumula...	Cumula...	Cutoff	Data Role	Depth	F1 Score	False D...	False P
★	Forward Logistic Regression	Logistic Regression	0.7250	0.9100	0.0825	0.9138	4.5000	45	0.5000	TEST	10	0.7429	0.1333	0.02
	Stepwise Logistic Regression	Logistic Regression	0.7250	0.9100	0.0825	0.9138	4.5000	45	0.5000	TEST	10	0.7429	0.1333	0.02
	Decision Tree	Decision Tree	0.6000	0.8700	0.1089	0.7903	3.7500	37.5000	0.5000	TEST	10	0.6286	0.2667	0.05
	Gradient Boosting	Gradient Boosting	0.5875	0.8600	0.1053	0.8713	3.5000	35	0.5000	TEST	10	0.5625	0.2500	0.03

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)
selectionCriteriaInterval	Average squared error
selectionTable	Test
selectionDepth	10

- Whomp, womp. The Gradient Boosting model was, well, not good. Huh. But, perhaps, you can do better!
- Anyway, here is your homework for the rest of this section:
  - Add more models to the existing *Pipeline 1* template
  - Feeling brave? Add a **Text Mining** node as a child node to the **Data** node.
    - Does *Mission Statement* have any predictive power (hint, hint: it should).
  - Feeling even braver? Click the **Add new pipeline** button here:
 
    - Give the *Advanced Template for class target* a try and further refine those models!
- When you are finished, identify your champion model – i.e., the model with the best predictive power in your modeling.
  - Welcome champion!

Feel a bit cheated because I cut this short?

I bet you did. And we never actually predicted the incoming class of 2025. Well, you can get the rest of the story in our **sas-hackathon-boot-camp-2025 GitHub** repo. Search for the file: [SAS Guided Demo - Student Admissions + Ethical Data Analysis.pdf](#).