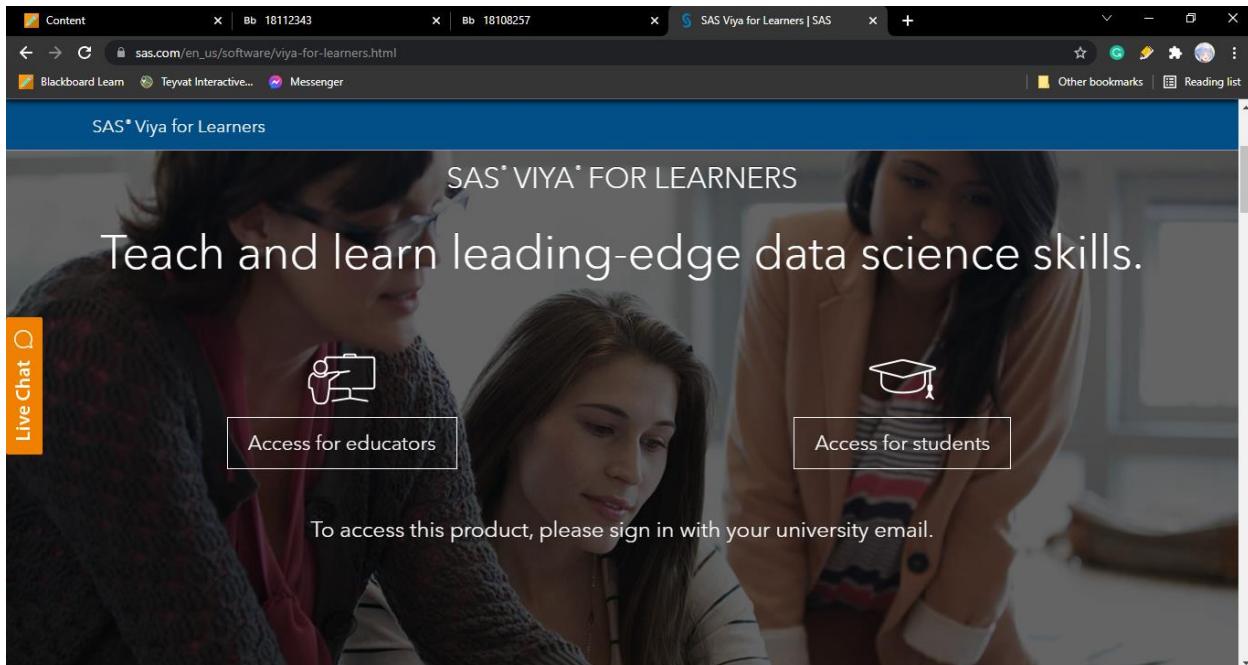


1. Read and download SAS Model Studio in SAS Viya for Learners - Quickstart Demos from Week 9.

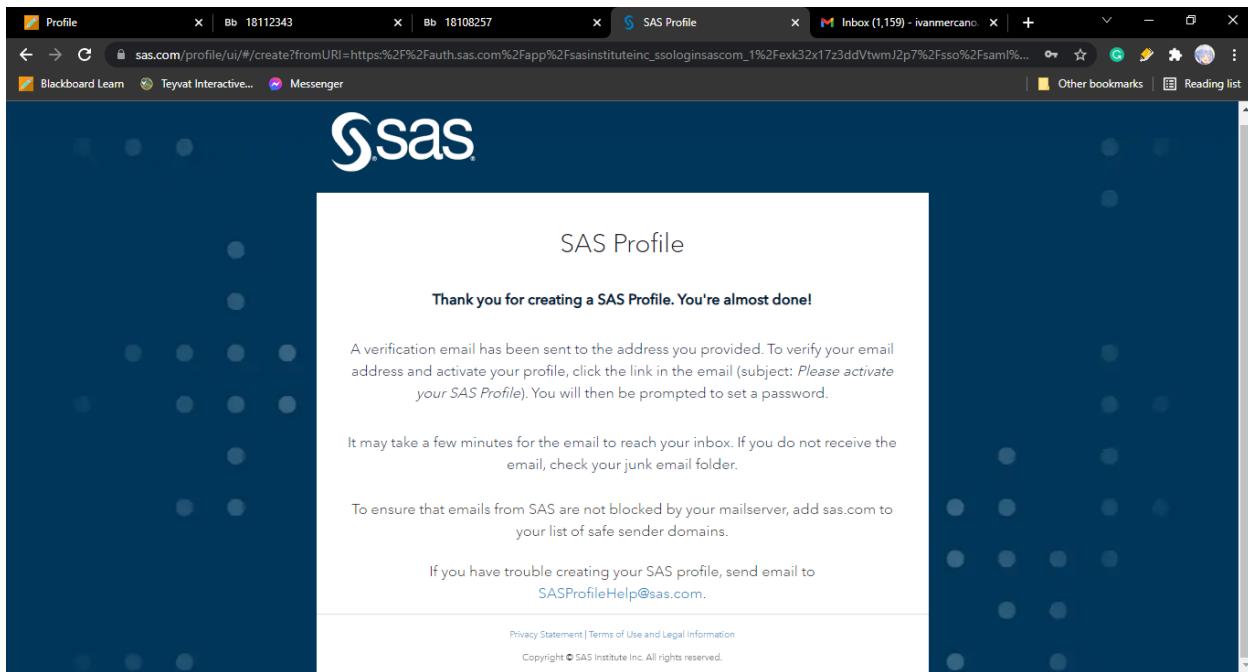
Accessing sas website to create an account



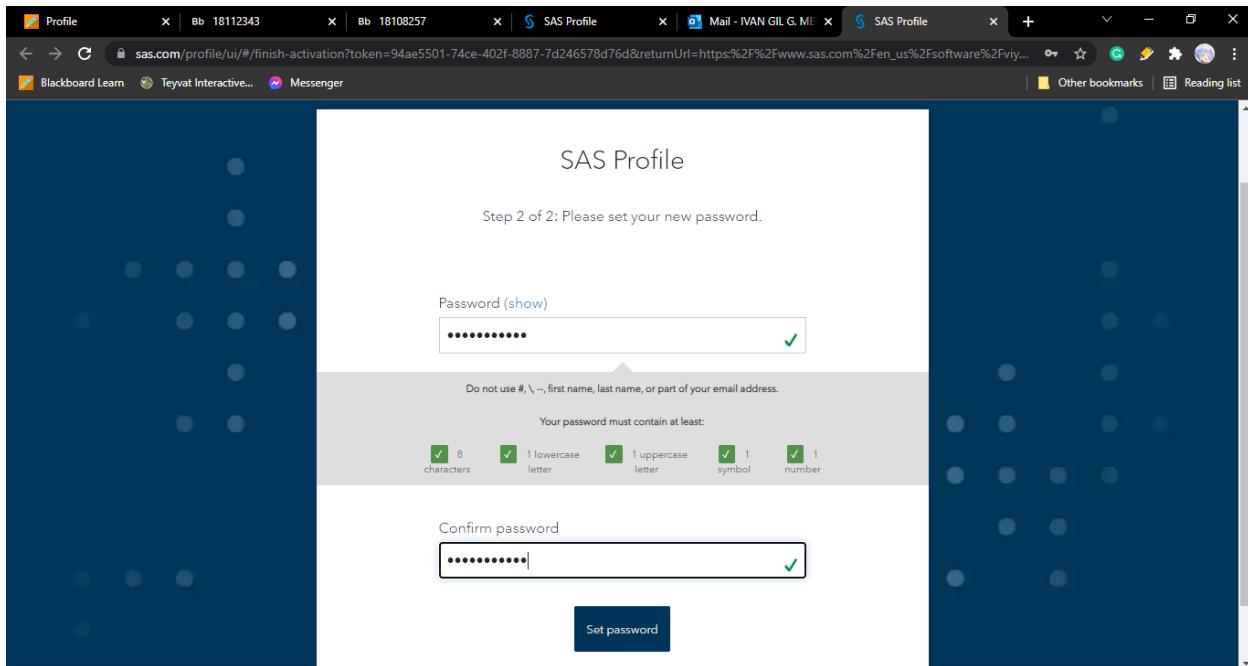
Filling out the form for the SAS profile

A screenshot of a web browser showing the 'SAS Profile' creation form. The URL in the address bar is sas.com/profile/ui/#/create?fromURI=https%3F%2Fauth.sas.com%2Fapp%2Fsasinsti... . The form is titled 'SAS Profile' and is labeled 'Step 1 of 2: Tell us about yourself.' It contains several input fields: 'Preferred Language' (dropdown menu showing 'English'), 'First Name *' (text input showing 'Ivan Gil'), 'Last Name *' (text input showing 'Mercano'), 'Email *' (text input showing 'ivanmercano.10@gmail.com'), and 'Country/Region *' (dropdown menu showing 'Philippines'). At the bottom, there's a link 'Affiliation With SAS *'.

Accepting terms and condition



Setting up password for the SAS account



Accessing the SAS account

The screenshot shows a web browser window with the following details:

- Title bar: Course: SAS Viya for Learners
- Address bar: vle.sas.com/course/view.php?id=7715
- Navigation: Dashboard / My courses / SAS Viya for Learners
- Content:
 - SAS logo and 'Virtual Learning Environment'
 - English (United States) (en_us)
 - 'Launch SAS Viya for Learners 3.5' button
 - 'More Information' link

Launch SAS Viya for Learners using the button below. Bookmark this page and return here each time you launch SAS Viya for Learners.

Launch SAS Viya
for Learners 3.5

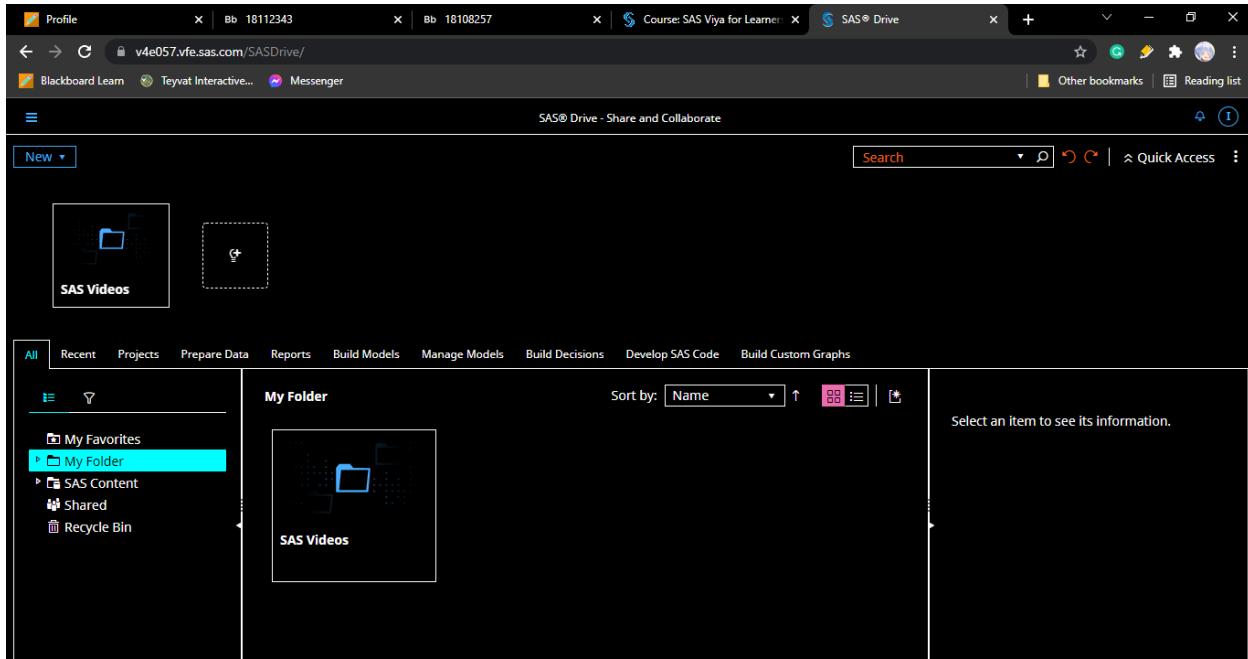
More Information

Launching the SAS Viya for Learners 3.5

The screenshot shows a web browser window with the following details:

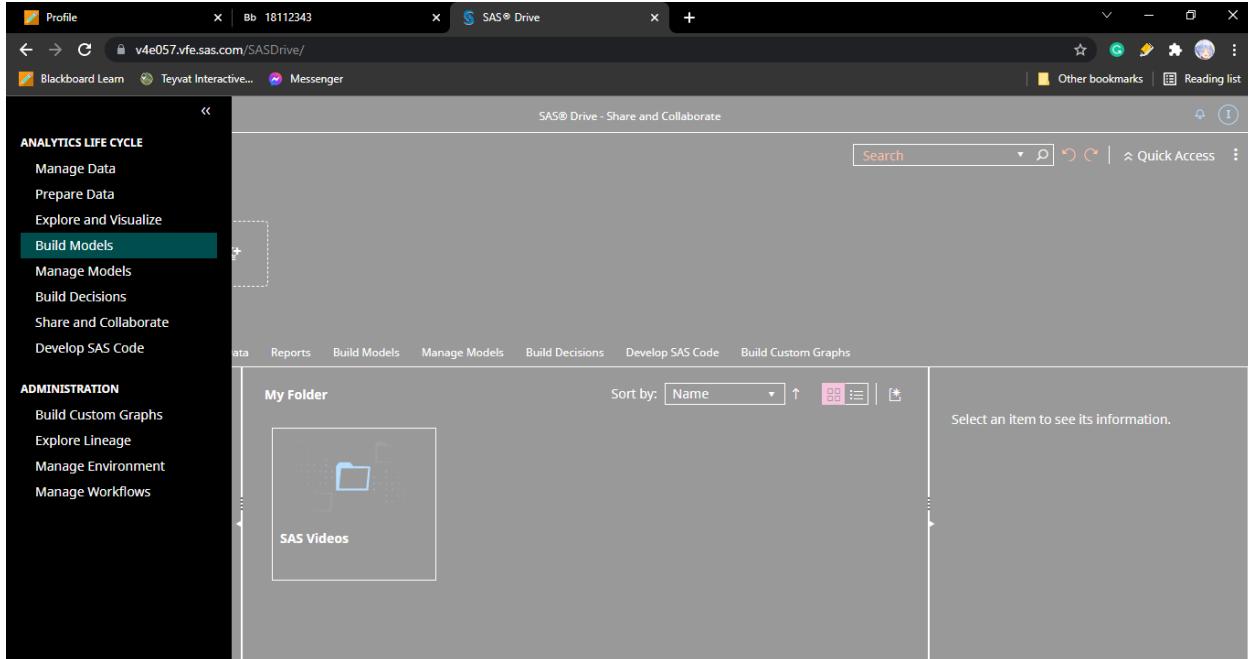
- Title bar: SAS Drive
- Address bar: v4e057.vfe.sas.com/SASDrive/
- Navigation: Blackboard Learn, Teyvat Interactive..., Messenger
- Content:
 - Welcome message: Welcome, iggmercane@mymail.mapua.edu.ph!
 - Organize and share content using SAS Drive.
 - Buttons: Watch Video, Let's Go!
 - SAS Drive - Share and Collaborate
 - Search bar
 - Build Decisions, Develop SAS Code, Build Custom Graphs
 - Sort by: Name
 - Select an item to see its information.

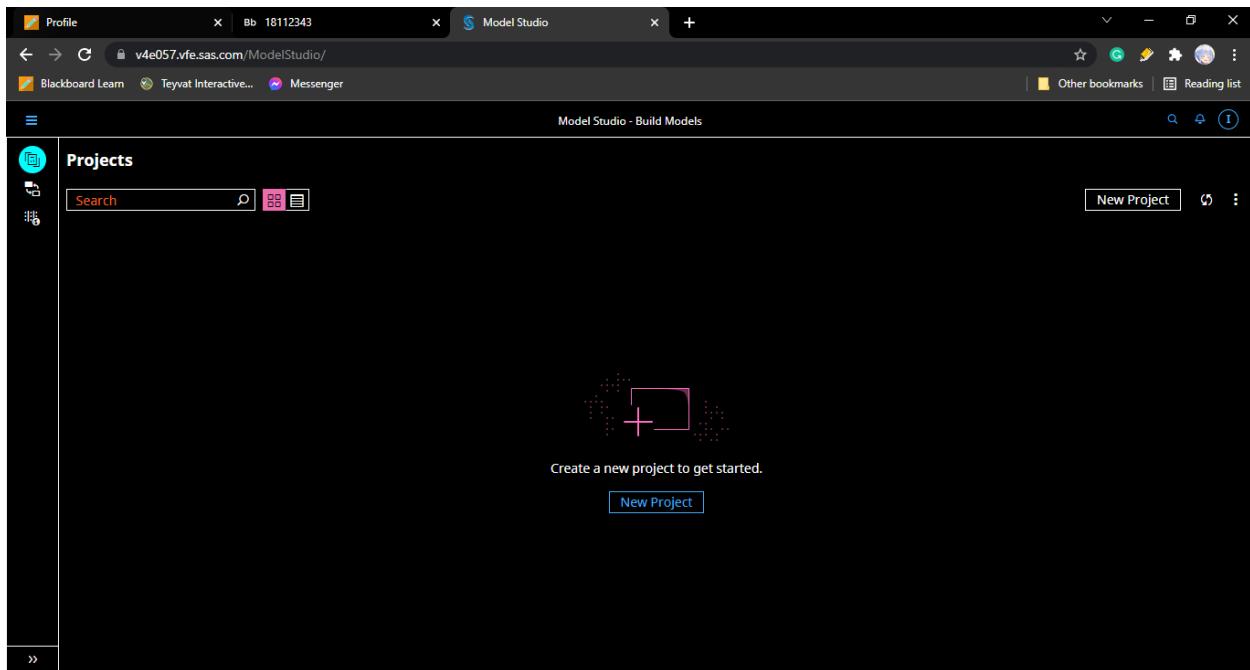
Exploring the Interface



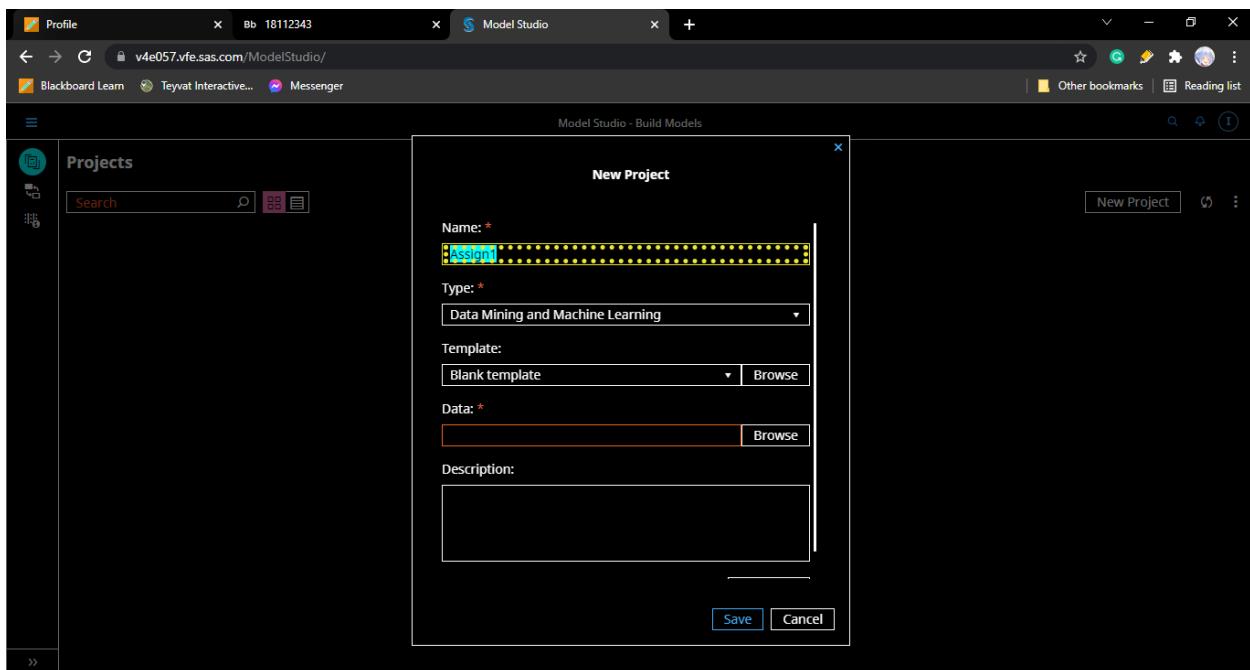
2. Do the sample activities from pages 12 to 45 and interpret the results.

Building new model inside SAS

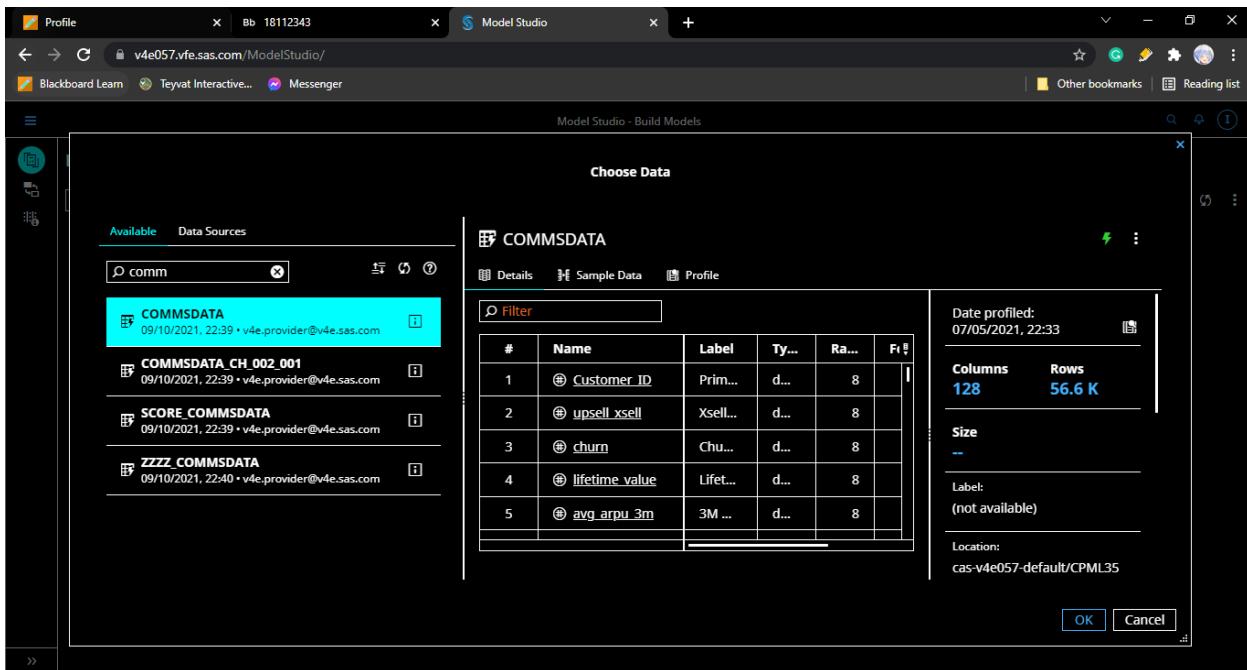




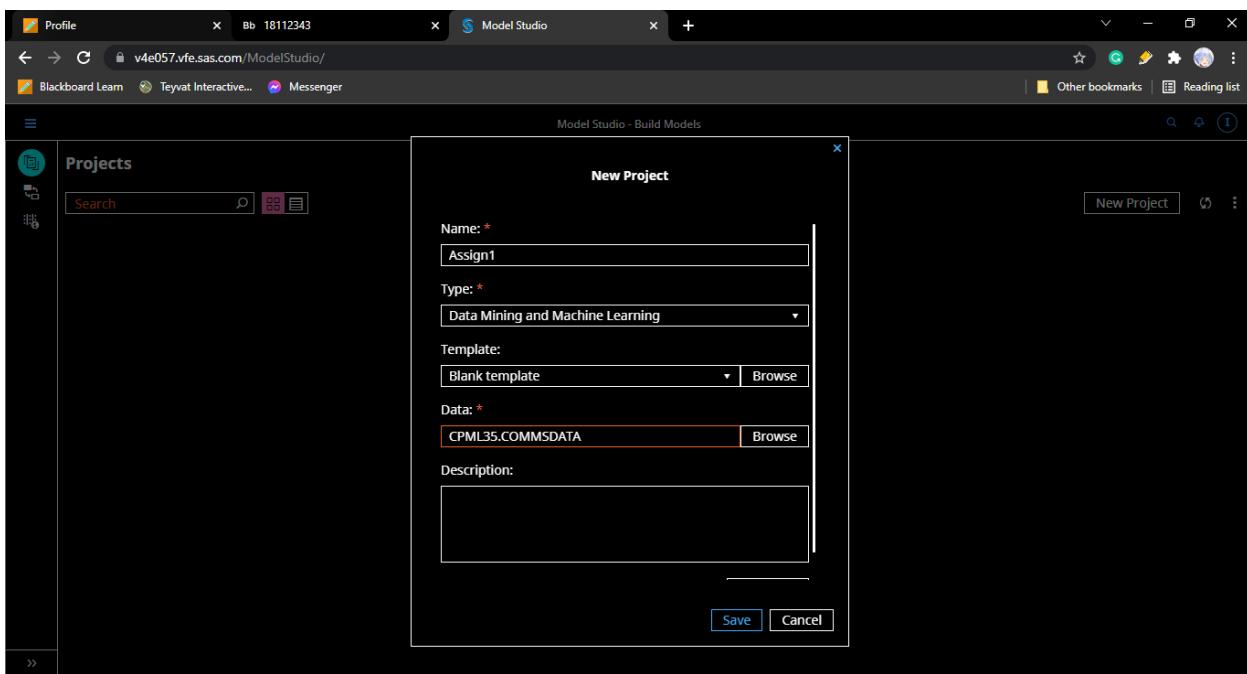
Naming the project 'Assign1'



Selecting dataset for the project

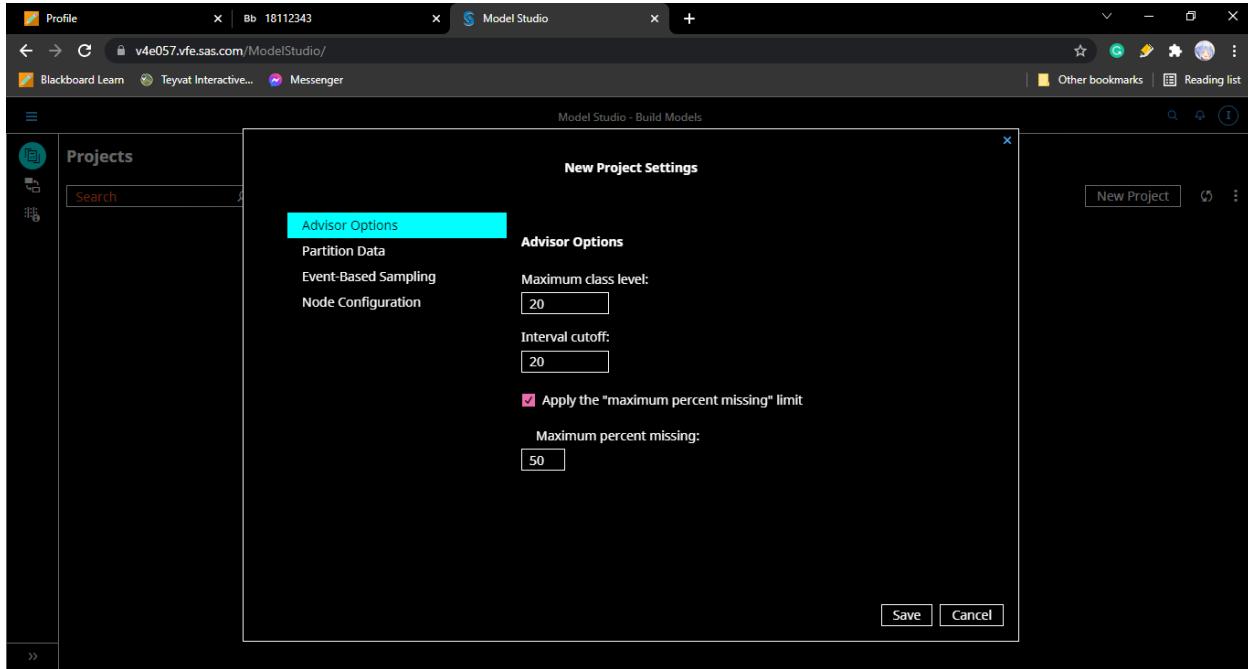


The screenshot shows the 'Choose Data' dialog in Model Studio. On the left, there's a sidebar with 'Available' and 'Data Sources' sections. Under 'Available', there are four entries: 'COMMSDATA', 'COMMSDATA_CH_002_001', 'SCORE COMMSDATA', and 'ZZZZ_COMMSDATA'. The 'COMMSDATA' entry is highlighted. On the right, the 'COMMSDATA' dataset is detailed. It has 128 columns and 56.6 K rows. The 'Label' field is '(not available)'. The 'Date profiled' is 07/05/2021, 22:33. There are tabs for 'Details', 'Sample Data', and 'Profile'. At the bottom right are 'OK' and 'Cancel' buttons.



The screenshot shows the 'New Project' dialog in Model Studio. The 'Name:' field contains 'Assign1'. The 'Type:' dropdown is set to 'Data Mining and Machine Learning'. The 'Template:' dropdown is set to 'Blank template'. The 'Data:' dropdown is set to 'CPML35.COMMSDATA'. The 'Description:' field is empty. At the bottom are 'Save' and 'Cancel' buttons. To the right of the dialog, there's a 'New Project' button.

Accessing the advisor options of the project, here we can access and adjust the settings for *Maximum Class Level, Interval cutoff, Maximum Percent Missing*.



Accessing the COMMS dataset

The screenshot shows the 'Model Studio - Build Models' interface with the 'Assign1' project selected. The 'Data' tab is active, displaying a table of variables. The table has columns: Variable Name, Label, Type, Role, and Level. The variables listed are: acct_age, avg_arpu_3m, avg_data_chrgs_3m, avg_data_prem_chrgs_3m, avg_days_susp, avg_overage_chrgs_3m, bill_data_usg_m03, bill_data_usg_m06, bill_data_usg_m09, bill_data_usg_tot, billing_cycle, and call_category_1. To the right of the table, a sidebar provides dataset metadata: Columns: 128, Rows: 56,557, Label: (not available), and Location: cas-v4e057-default/CPML35.

Assigning the target variable, for this project I assigned the variable *chum*.

The screenshot shows the Model Studio interface with the 'Data' tab selected. On the left, there's a sidebar with icons for Profile, Bb 18112343, Model Studio, and a video player for '1093 6 - The Drawbridge - You'. The main area displays a table of variables:

Variable Name	Label	Type	Role	Level
calls_out_pk	Calls Outgoing Peak	Numeric	Input	Ordinal
calls_total	Total Calls Curr	Numeric	Input	Ordinal
calls_TS_acct	Number Calls Tech Support	Numeric	Input	Interval
churn	Churn Flag	Numeric	Target	Binary
city	Account City	Character	ID	Nominal
city_lat	Account City Latitude	Numeric	Rejected	Interval
city_long	Account City Longitude	Numeric	Rejected	Ordinal
count_of_susensions_6m	Times Suspended Last 6M	Numeric	Input	Nominal
credit_class	Credit Class	Character	Input	Nominal
cs_afr_amer	Census Area African-American	Numeric	Input	Interval
cs_caucasian	Census Area Caucasian	Numeric	Input	Interval
cs_hispanic	Census Area Hispanic	Numeric	Input	Interval

To the right of the table, a detailed view for the 'churn' variable is shown, with 'Target' selected under 'Role'.

Specifying the target event level of the variable, for chum I select 12.13% as target event level.

The screenshot shows the 'Specify the Target Event Level' dialog box overlaid on the Data table. The dialog has a title 'Specify the Target Event Level' and a sub-section 'Target event level:' with a dropdown menu containing '1 - 6.862 (12.13%)'. At the bottom are 'Save' and 'Cancel' buttons.

The Data table shows the same list of variables as the previous screenshot, with the 'churn' row highlighted. To the right of the table, the detailed view for 'churn' remains the same, showing 'Target' selected under 'Role'.

MODIFYING DATA PARTITION

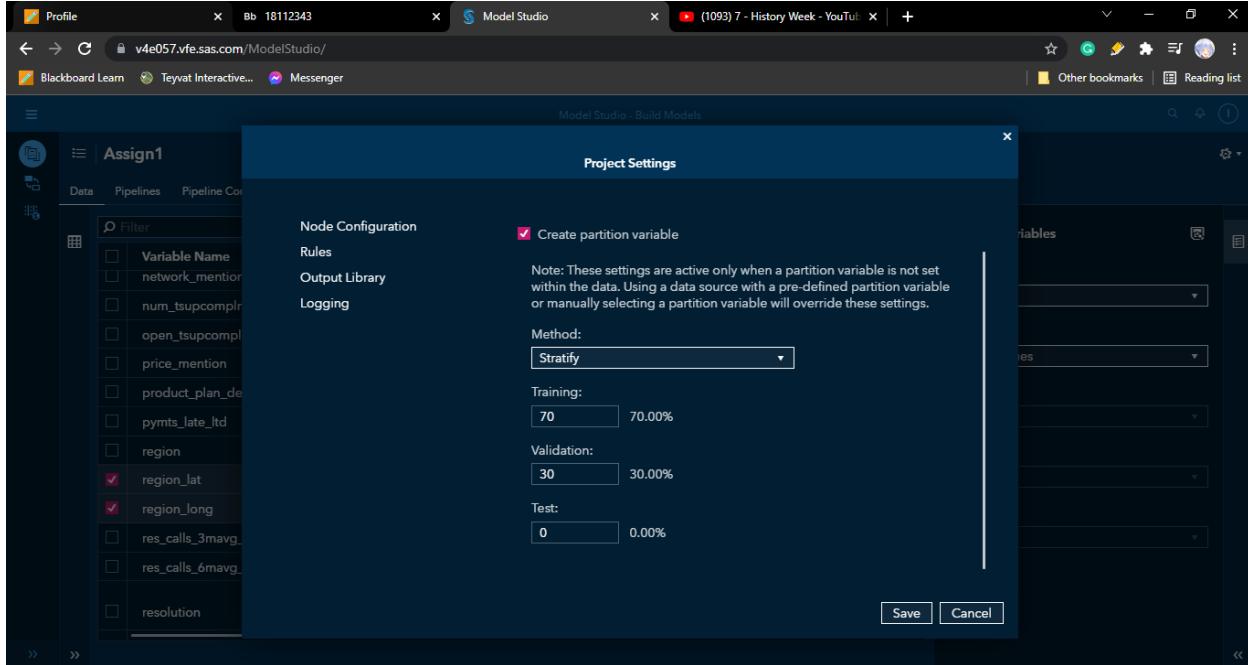
Selecting the 11 variable: *city, city_lat, city_long, data_usage_amt, mou_onnet_6m_normal, mour_roam_6m_normal, region_lat, region_long, state_lat, state_long, and tweedie_adjusted*.

Variable Name	Label	Type	Role	Level
network_mention	Network Issues Discussed	Numeric	Input	Binary
num_tsupcomplnts	Tech Support Complaints - LTD	Numeric	Input	Nominal
open_tsupcomplnts	Open Tech Support Complaints	Numeric	Input	Interval
price_mention	Price Issues Discussed	Numeric	Input	Nominal
product_plan_desc	Plan Name	Character	Input	Nominal
pymts_late_ltd	Total Late Payments Lifetime	Numeric	Input	Nominal
region	Account Region	Character	Input	Nominal
<input checked="" type="checkbox"/> region_lat	Account Region Latitude	Numeric	Rejected	Nominal
<input checked="" type="checkbox"/> region_long	Account Region Longitude	Numeric	Rejected	Ordinal
res_calls_3mavg_acct	Resolved Calls - 3Mo Average	Numeric	Input	Interval
res_calls_6mavg_acct	Resolved Calls - 6Mo Average	Numeric	Input	Interval
resolution	Final Resolution	Character	Text	Nominal

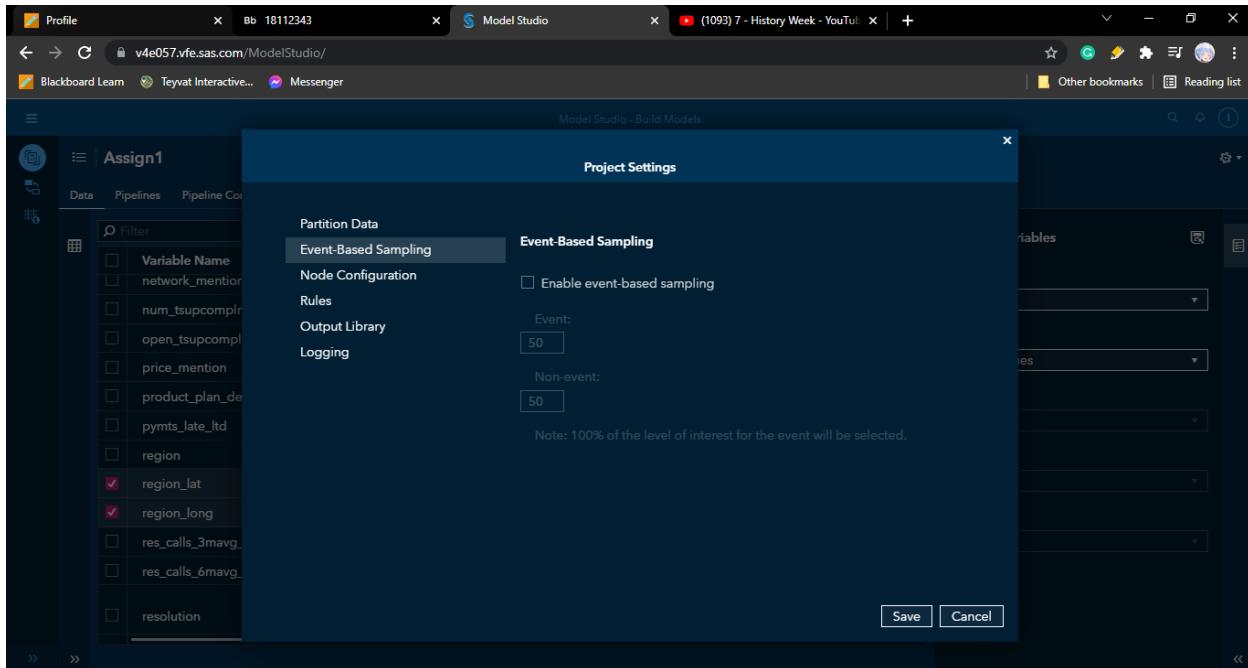
Changing the role of the selected variable to rejected

Variable Name	Label	Type	Role	Level
network_mention	Network Issues Discussed	Numeric	Input	Binary
num_tsupcomplnts	Tech Support Complaints - LTD	Numeric	Input	Nominal
open_tsupcomplnts	Open Tech Support Complaints	Numeric	Input	Interval
price_mention	Price Issues Discussed	Numeric	Input	Nominal
product_plan_desc	Plan Name	Character	Input	Nominal
pymts_late_ltd	Total Late Payments Lifetime	Numeric	Input	Nominal
region	Account Region	Character	Input	Nominal
<input checked="" type="checkbox"/> region_lat	Account Region Latitude	Numeric	Rejected	Nominal
<input checked="" type="checkbox"/> region_long	Account Region Longitude	Numeric	Rejected	Ordinal
res_calls_3mavg_acct	Resolved Calls - 3Mo Average	Numeric	Input	Interval
res_calls_6mavg_acct	Resolved Calls - 6Mo Average	Numeric	Input	Interval
resolution	Final Resolution	Character	Text	Nominal

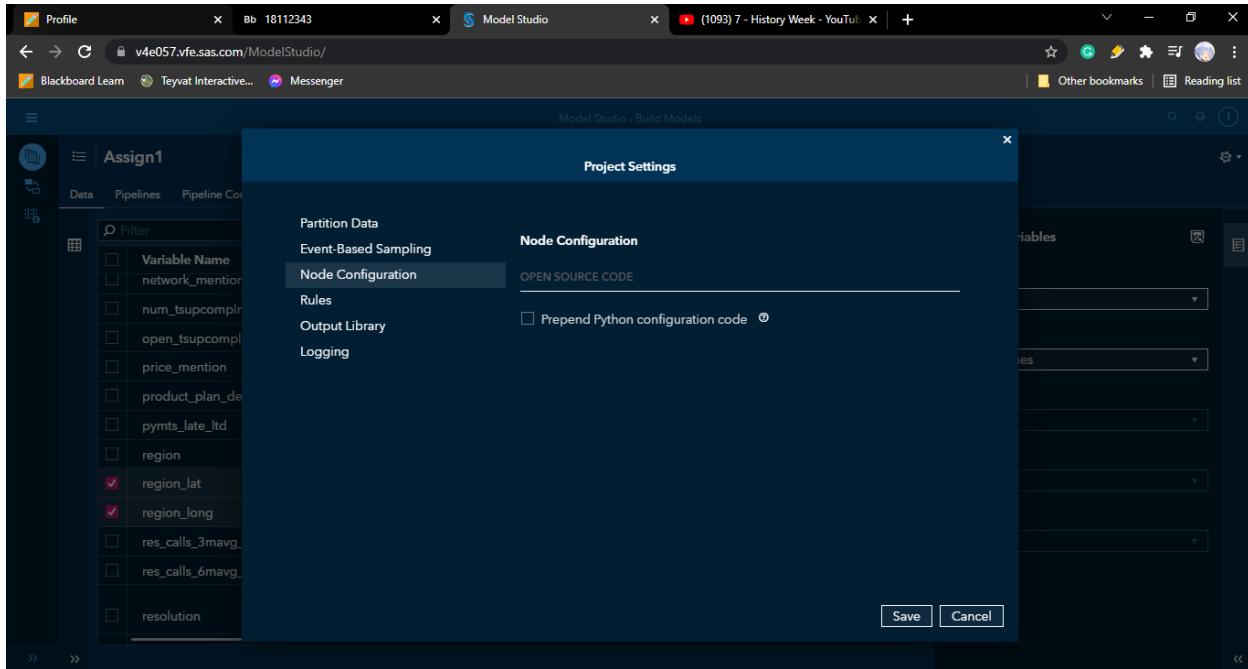
Adjusting the partition data in project settings by changing the training percentage to 70% and test percentage to 0%.



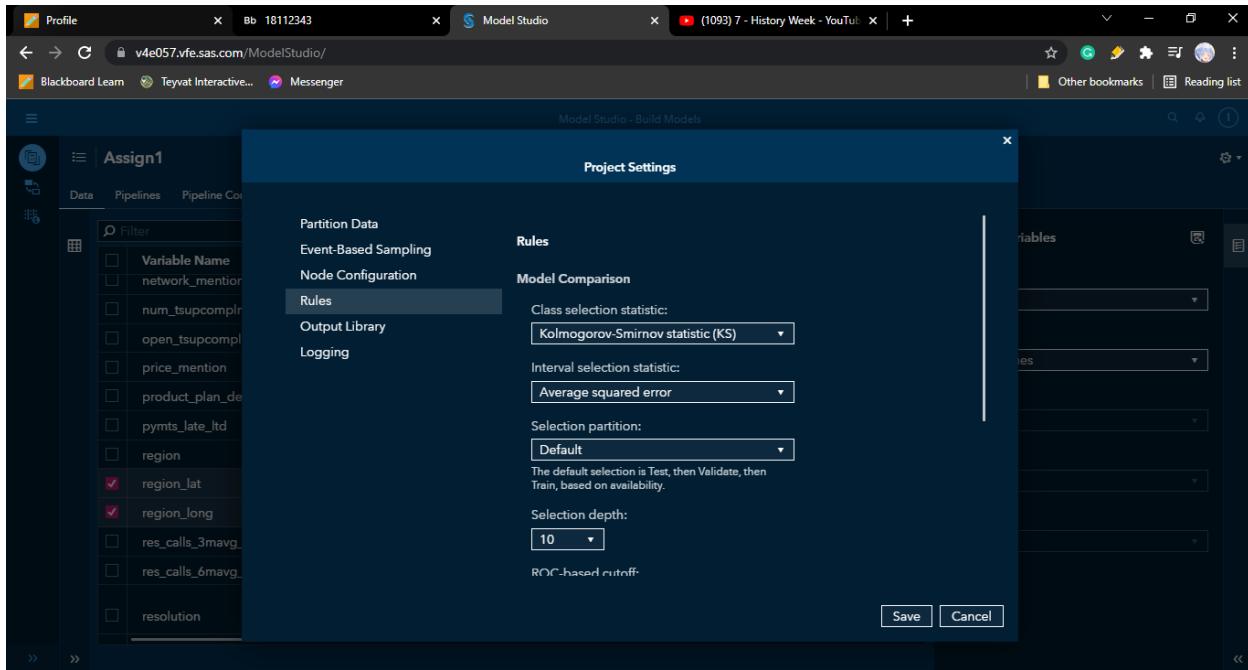
Leaving the event-based sampling in default



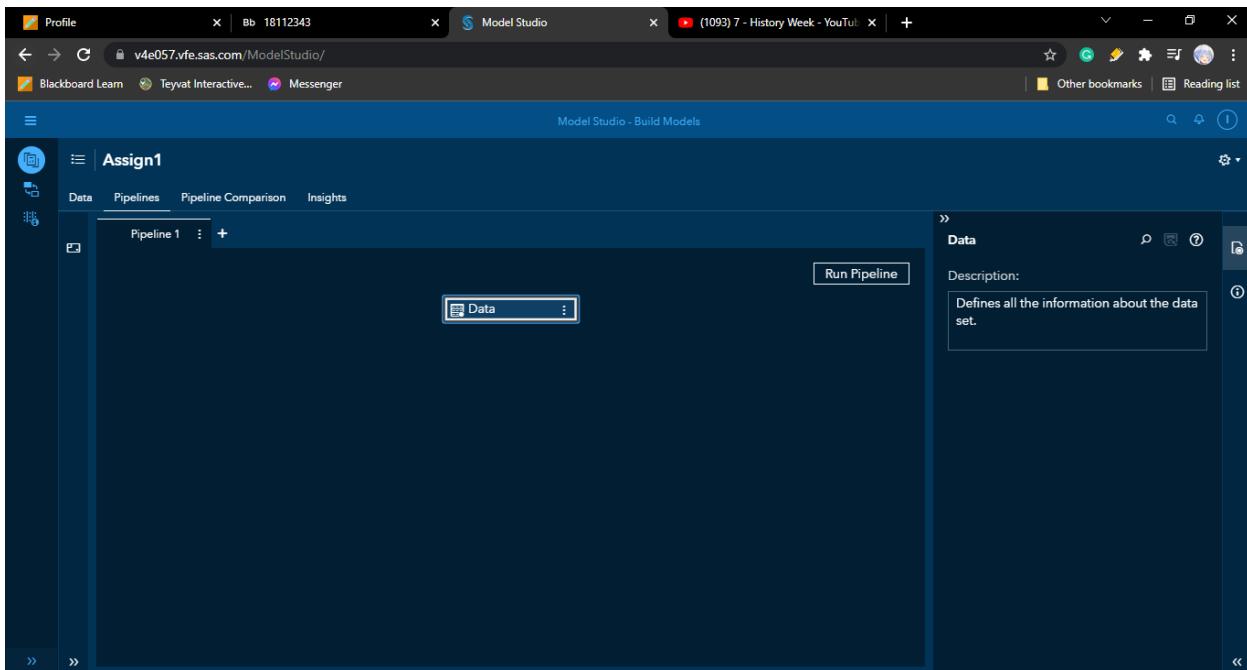
Leaving node configuration in default



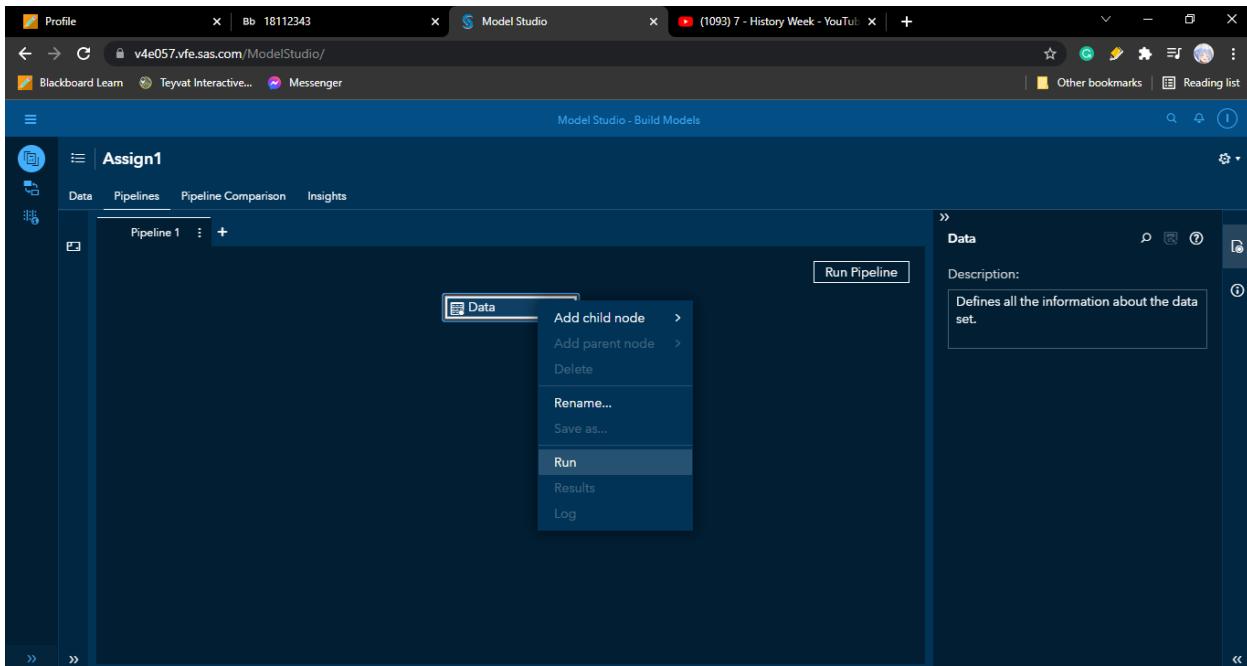
Leaving the rules in default



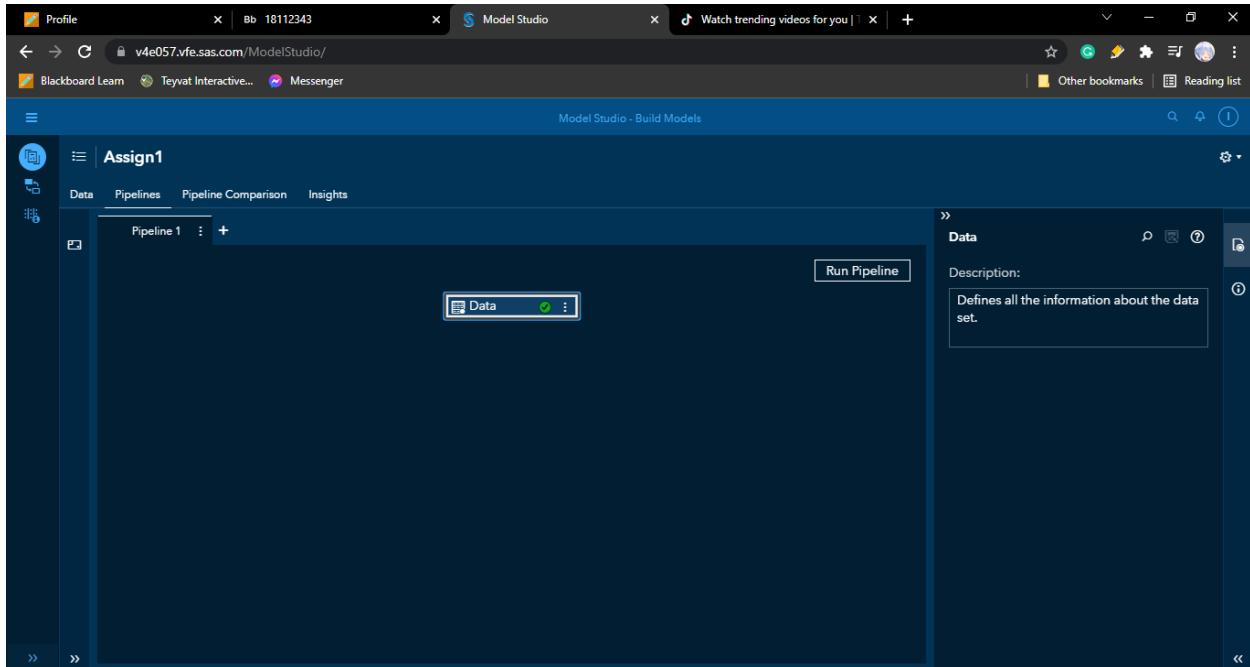
Accessing the pipeline



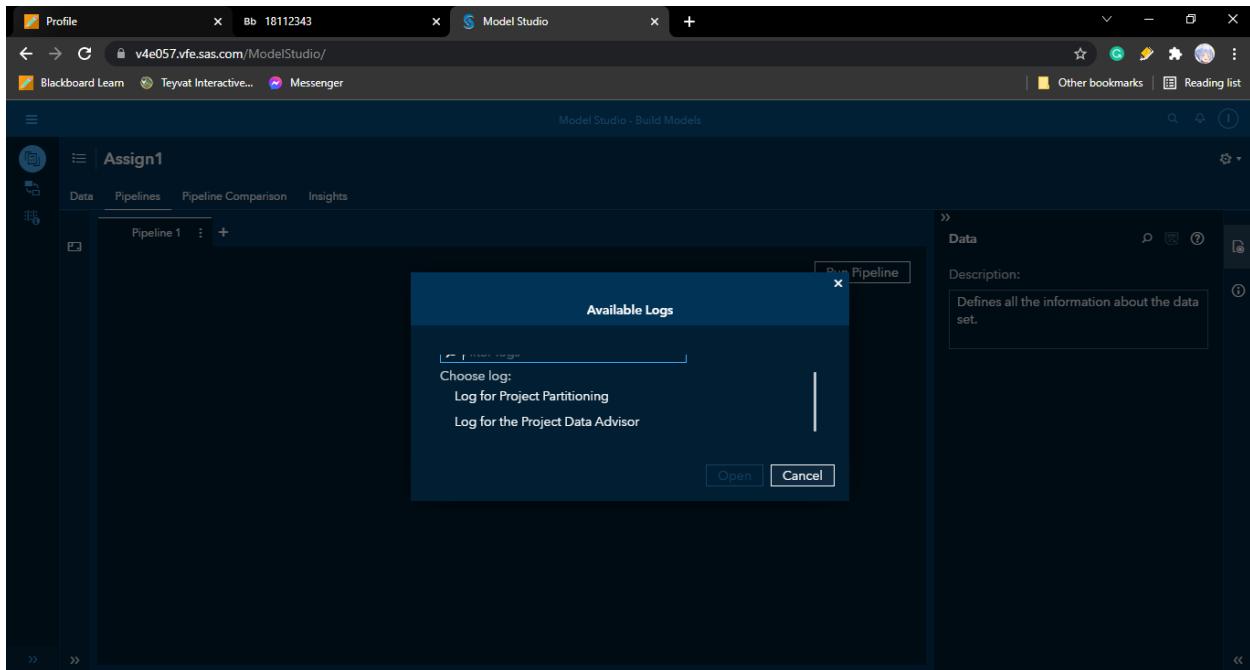
Running the data



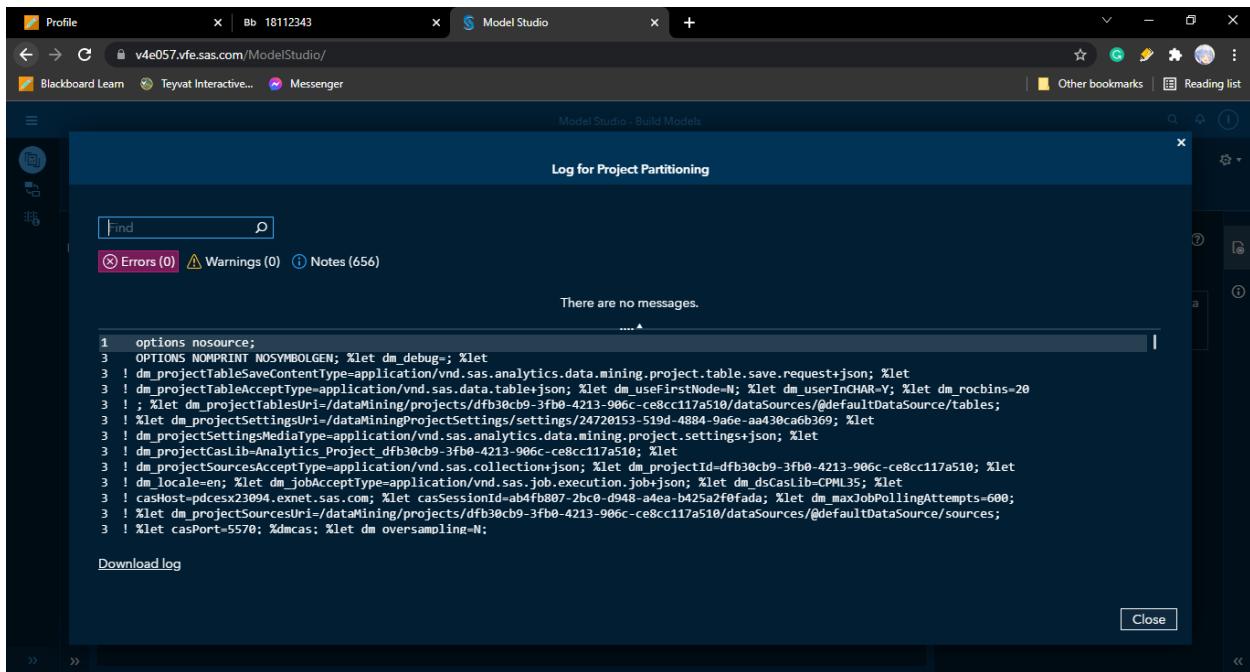
The green check beside the data means that the project run without any errors



Accessing the project logs



Here we can see or even download the logs of the project



The screenshot shows a browser window for Model Studio. The title bar says "Profile" and "Bb 18112343". The address bar shows "v4e057.vfe.sas.com/ModelStudio/". The main content area is titled "Log for Project Partitioning". It has a search bar and a status message "There are no messages.". Below this, there is a code block containing several lines of SAS log output. At the bottom right of the window is a "Close" button.

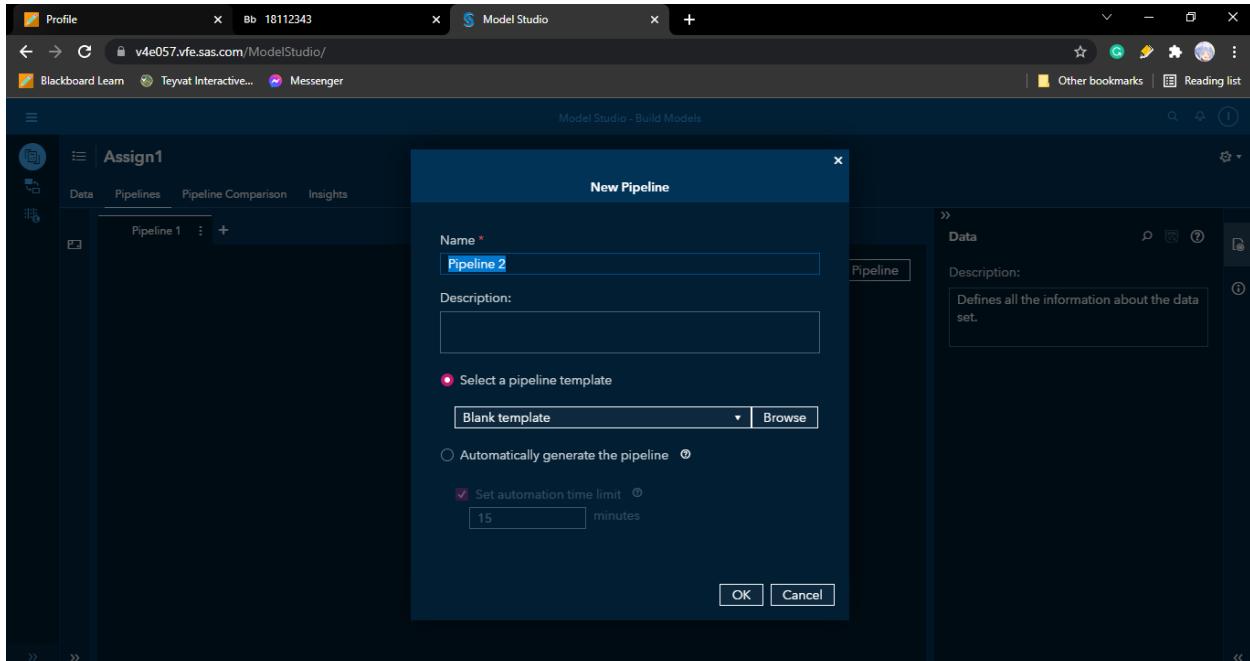
```
1 options nosource;
2 OPTIONS NOMPRINT NOSYMBOLGEN; %let dm_debug=; %let
3 ! dm_projectTableSaveContentType=application/vnd.sas.analytics.data.mining.project.table.save.request+json; %let
3 ! dm_projectTableAcceptType=application/vnd.sas.data.table+json; %let dm_useFirstNode=N; %let dm_userInCHAR=Y; %let dm_rochbins=20
3 ! ; %let dm_projectTablesUri=/dataMining/projects/dfb30cb9-3fb0-4213-906c-ce8cc117a510/dataSources/@defaultDataSource/tables;
3 ! %let dm_projectSettingsUri=/dataMiningProjectSettings/settings/24720153-519d-4884-9a6e-aa430cab6369; %let
3 ! dm_projectSettingsMediaType=application/vnd.sas.analytics.data.mining.project.settings+json; %let
3 ! dm_projectCasLib=Analytics_Project_dfb30cb9-3fb0-4213-906c-ce8cc117a510; %let
3 ! dm_projectSourcesAcceptType=application/vnd.sas.collection+json; %let dm_projectId=dfb30cb9-3fb0-4213-906c-ce8cc117a510; %let
3 ! dm_locale=en; %let dm_jobAcceptType=application/vnd.sas.job.execution.job+json; %let dm_dscCasLib=CPML3; %let
3 ! casHost=pcdesx23094.exnet.sas.com; %let casSessionId=ab4fb807-2b0-d948-a4ea-b425a2f0fada; %let dm_maxJobPollingAttempts=600;
3 ! %let dm_projectSourcesUri=/dataMining/projects/dfb30cb9-3fb0-4213-906c-ce8cc117a510/dataSources/@defaultDataSource/sources;
3 ! %let casPort=5570; %dmcas; %let dm_oversampling=N;
```

[Download log](#)

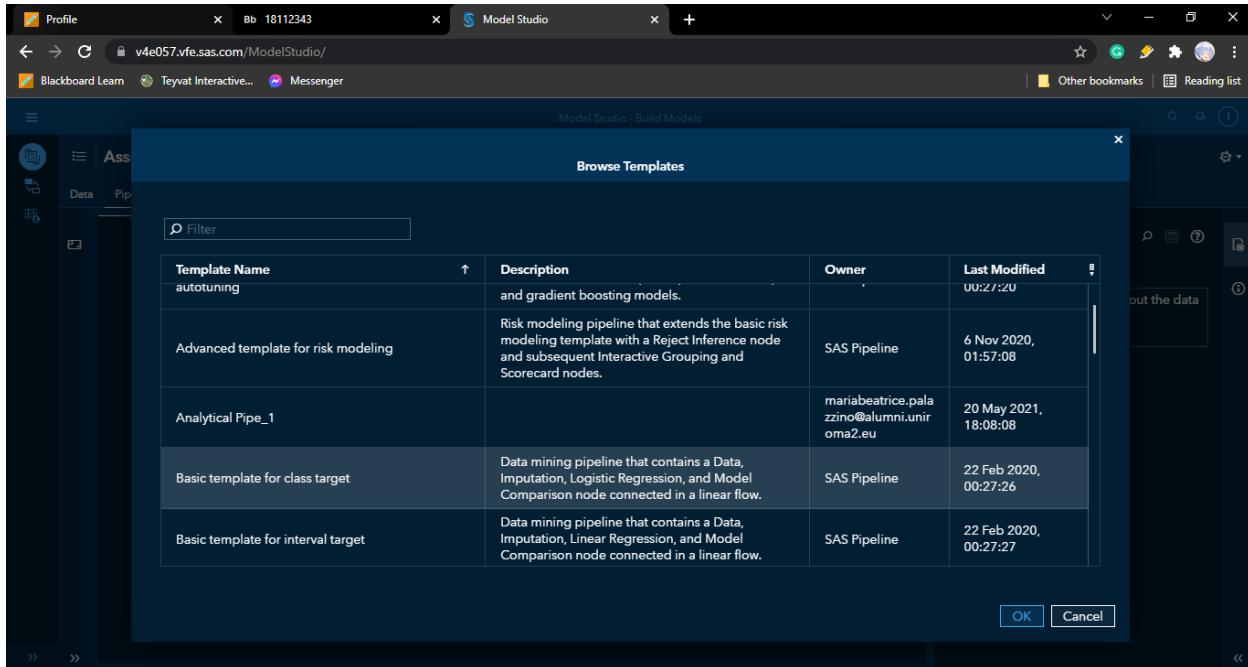
[Close](#)

BUILDING PIPELINE FROM A BASIC TEMPLATE

Creating new pipeline

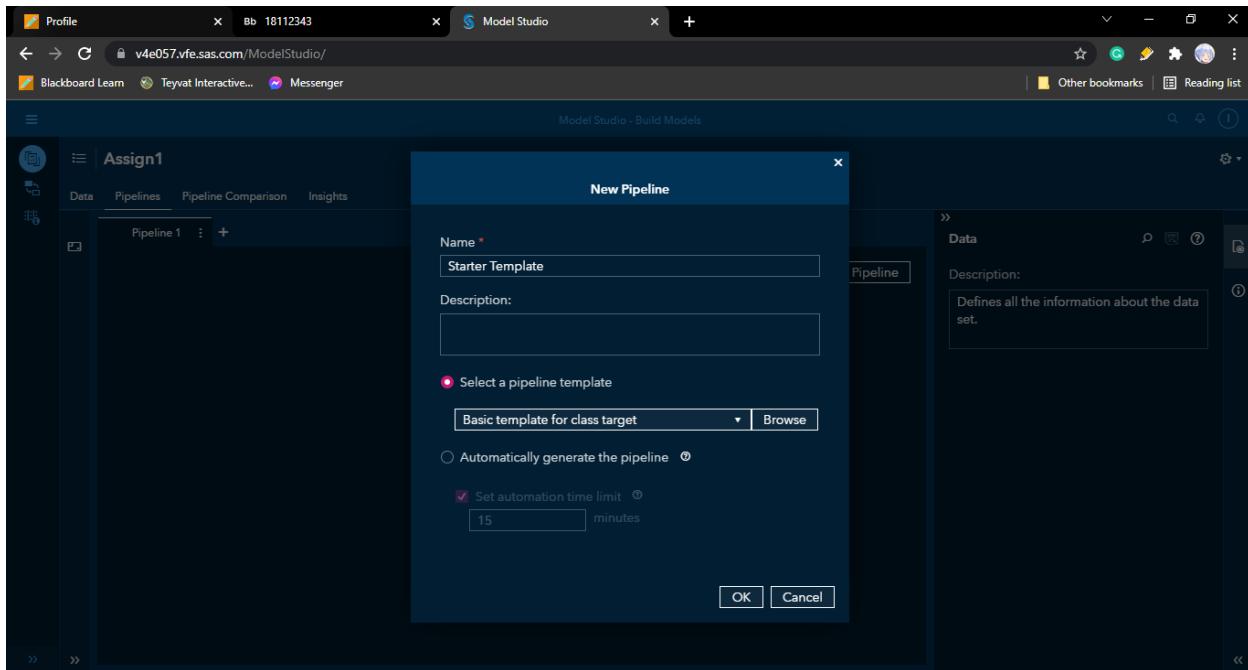


Selecting the template for the newly created pipeline



Template Name	Description	Owner	Last Modified
autotuning	and gradient boosting models.		00:27:20
Advanced template for risk modeling	Risk modeling pipeline that extends the basic risk modeling template with a Reject Inference node and subsequent Interactive Grouping and Scorecard nodes.	SAS Pipeline	6 Nov 2020, 01:57:08
Analytical Pipe_1		mariabeatrice.palazzino@alumni.uniroma2.eu	20 May 2021, 18:08:08
Basic template for class target	Data mining pipeline that contains a Data, Imputation, Logistic Regression, and Model Comparison node connected in a linear flow.	SAS Pipeline	22 Feb 2020, 00:27:26
Basic template for interval target	Data mining pipeline that contains a Data, Imputation, Linear Regression, and Model Comparison node connected in a linear flow.	SAS Pipeline	22 Feb 2020, 00:27:27

Naming the pipeline Starter Template



New Pipeline

Name:

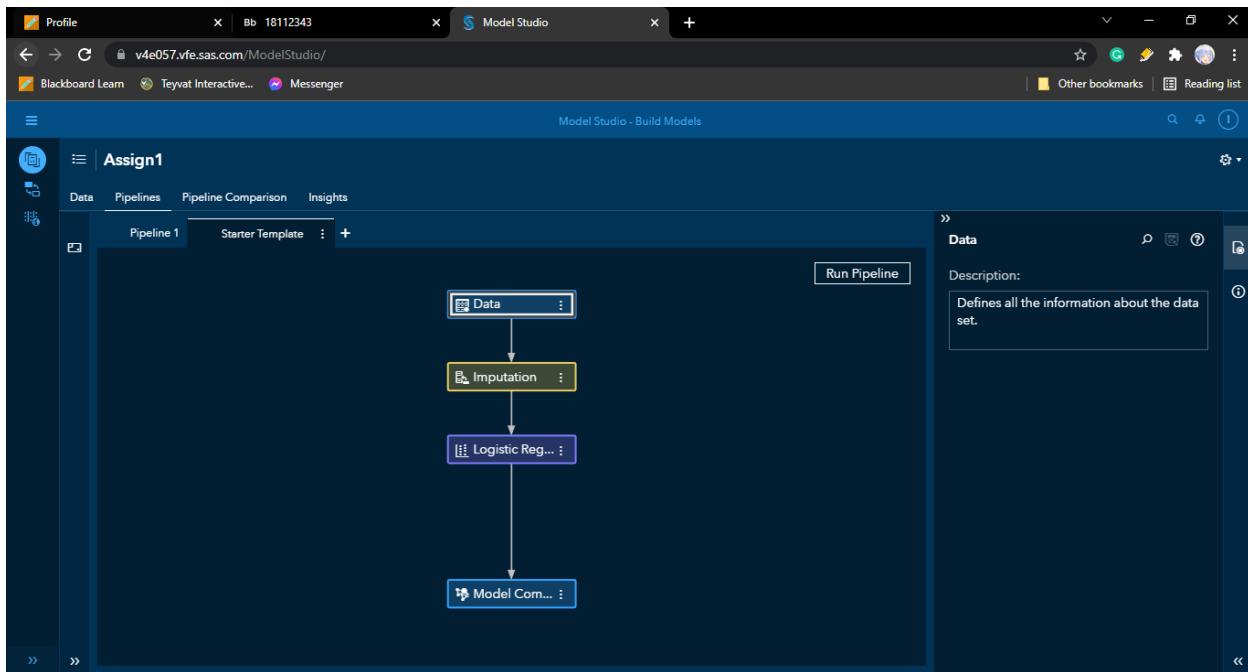
Description:

Select a pipeline template:

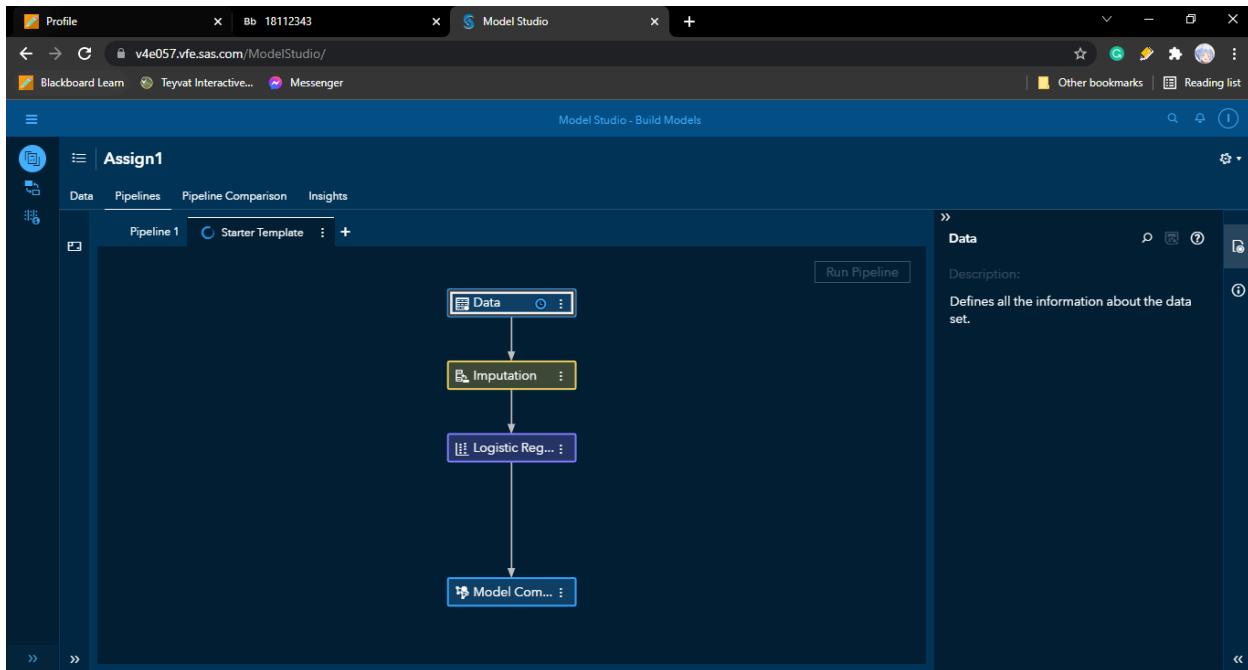
Automatically generate the pipeline:

Set automation time limit: minutes

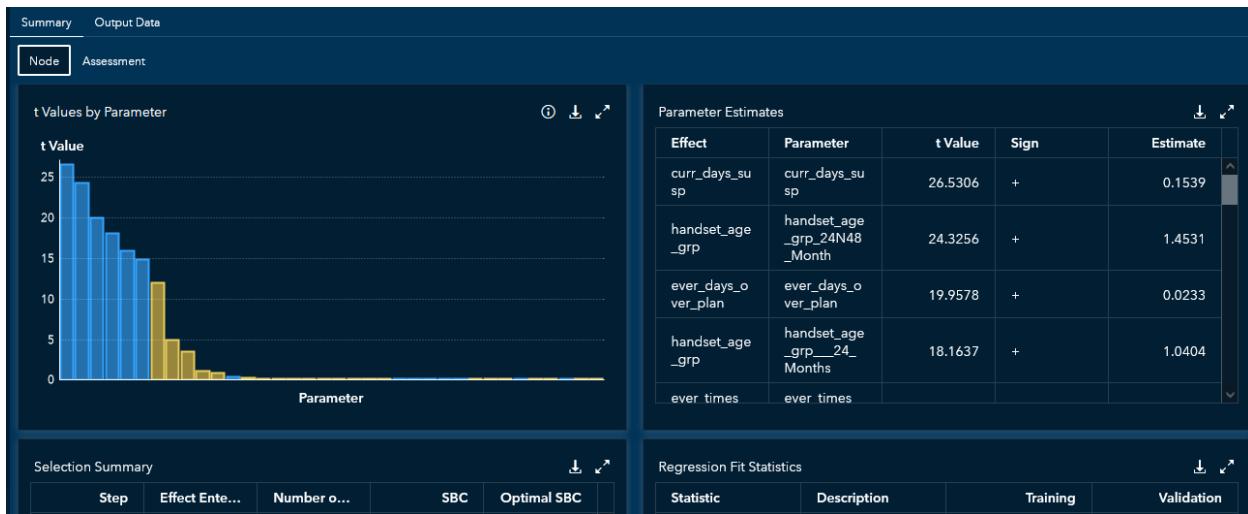
Accessing the starter template pipeline



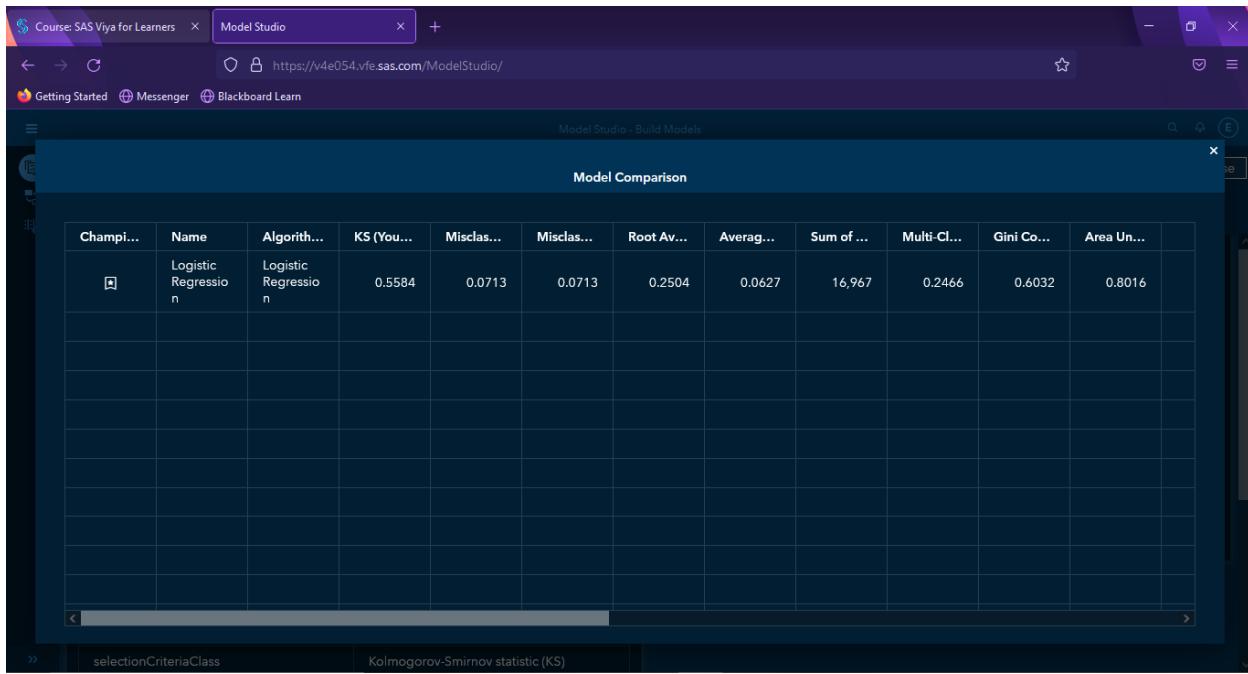
Running the pipeline



Results of the logistic regression

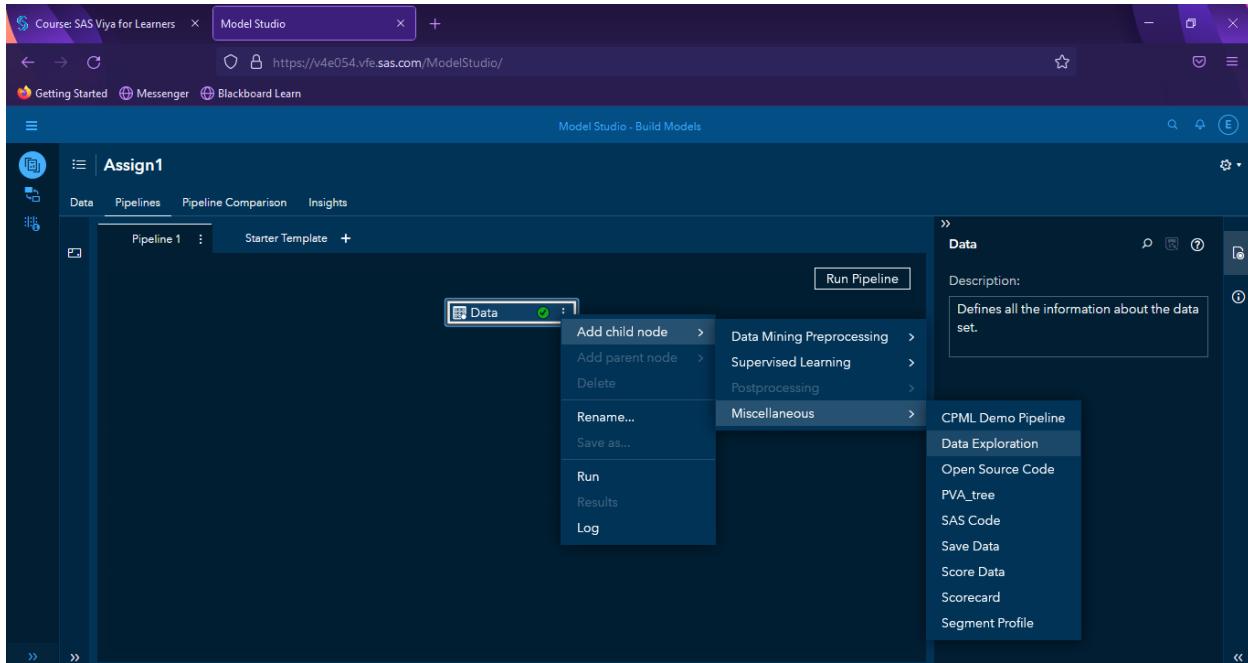


Results of the model comparison

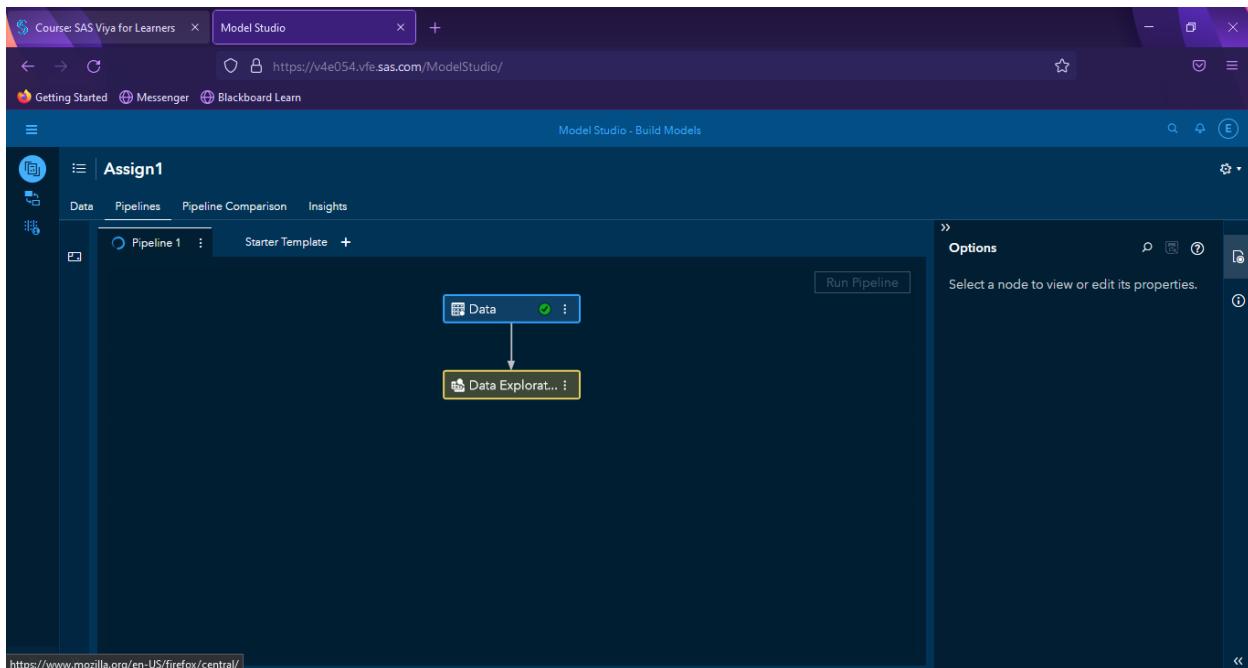


EXPLORING SOURCE DATA

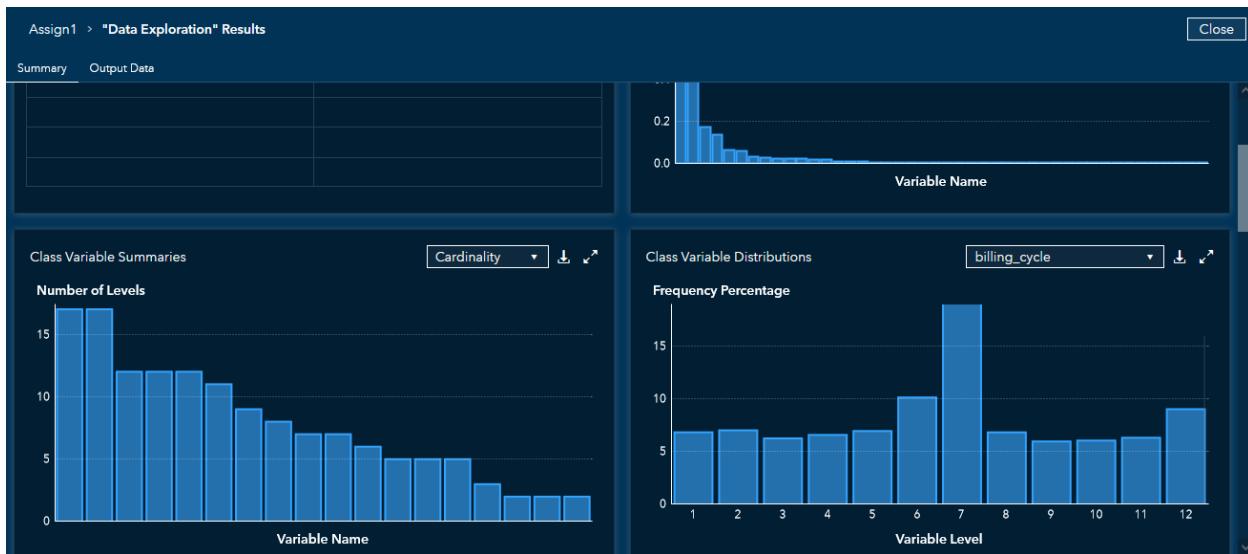
Adding childnode for the first pipeline



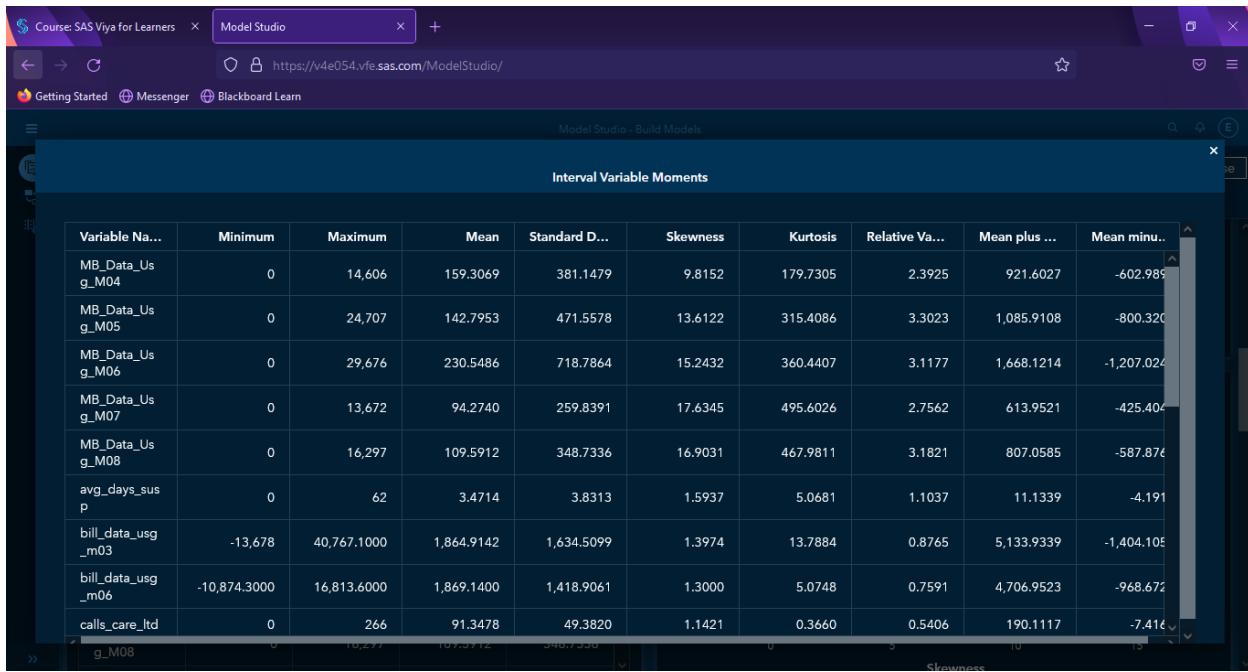
Running the pipeline



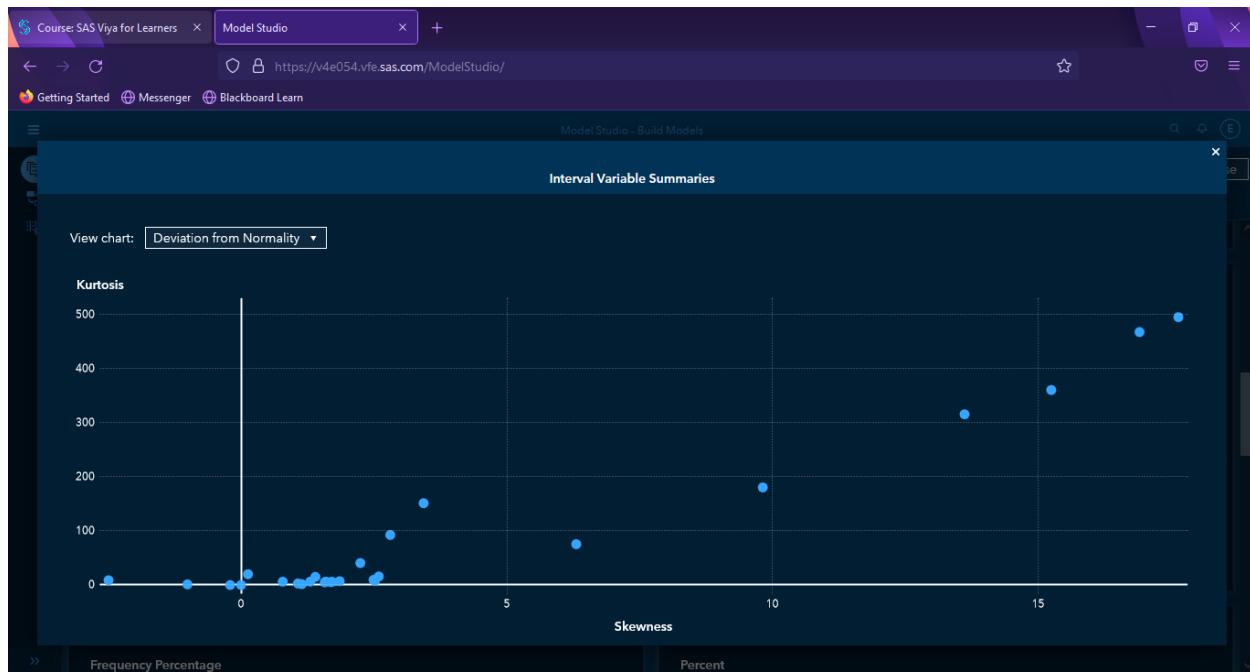
Checking the result of data exploration



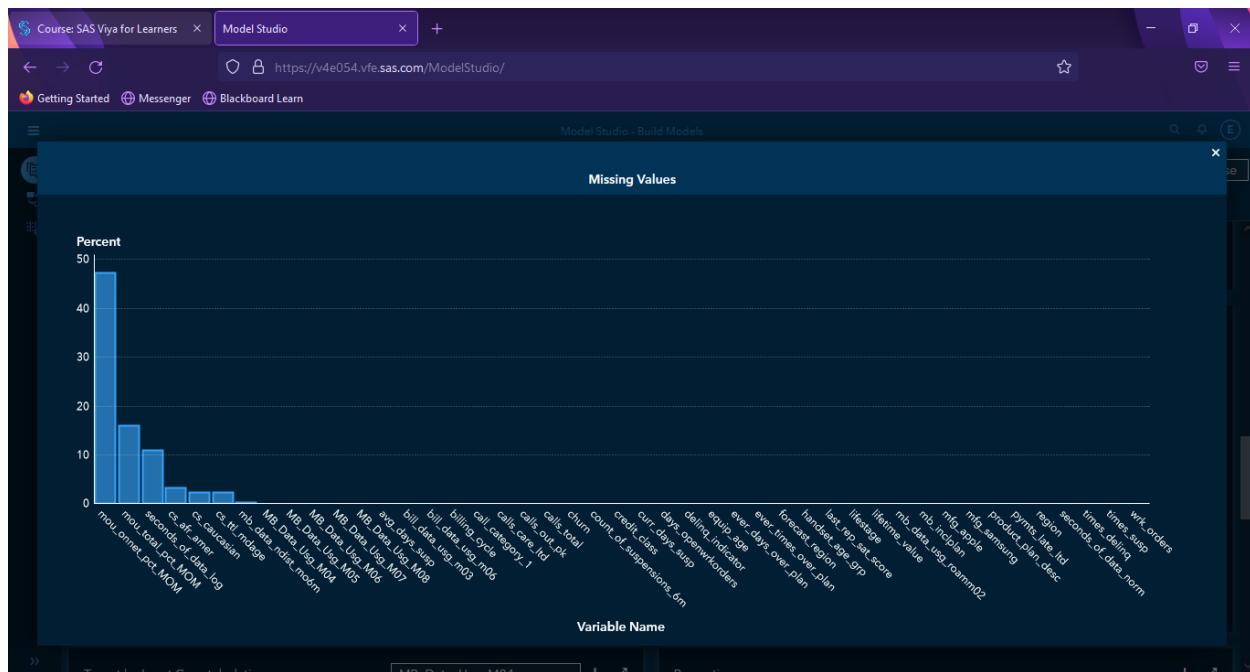
Checking the interval variable movement



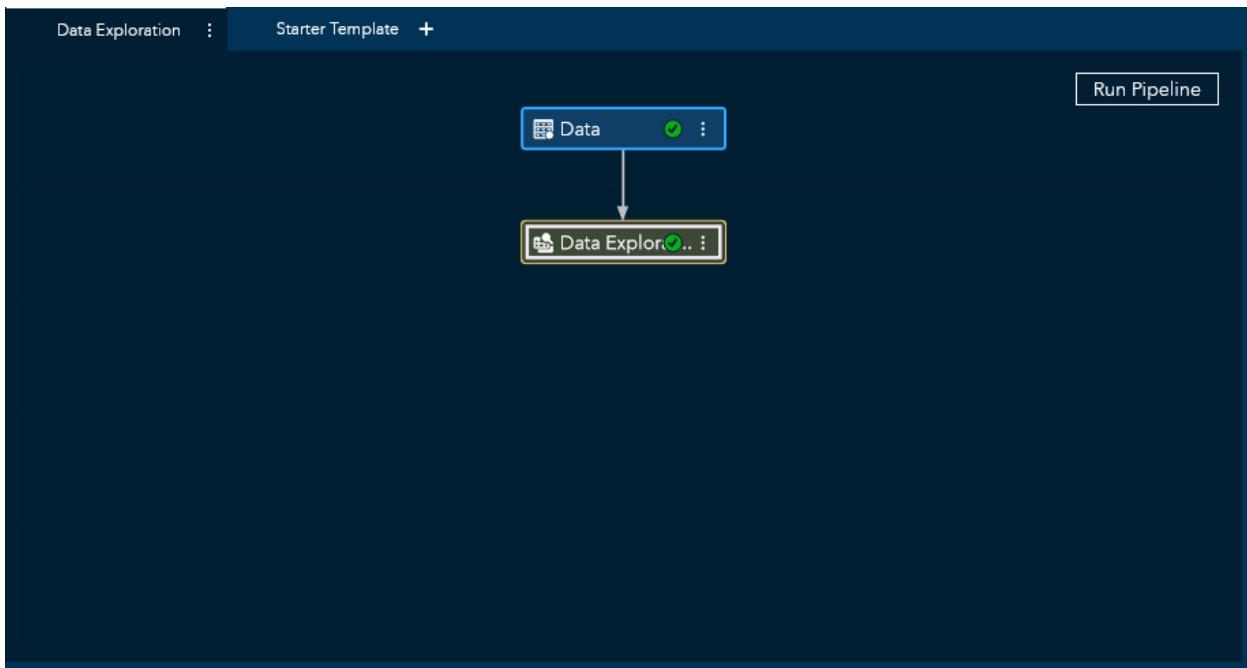
Checking the interval variable summary



Showing the missing values in the data exploration



Renaming the pipeline into *Data Exploration*



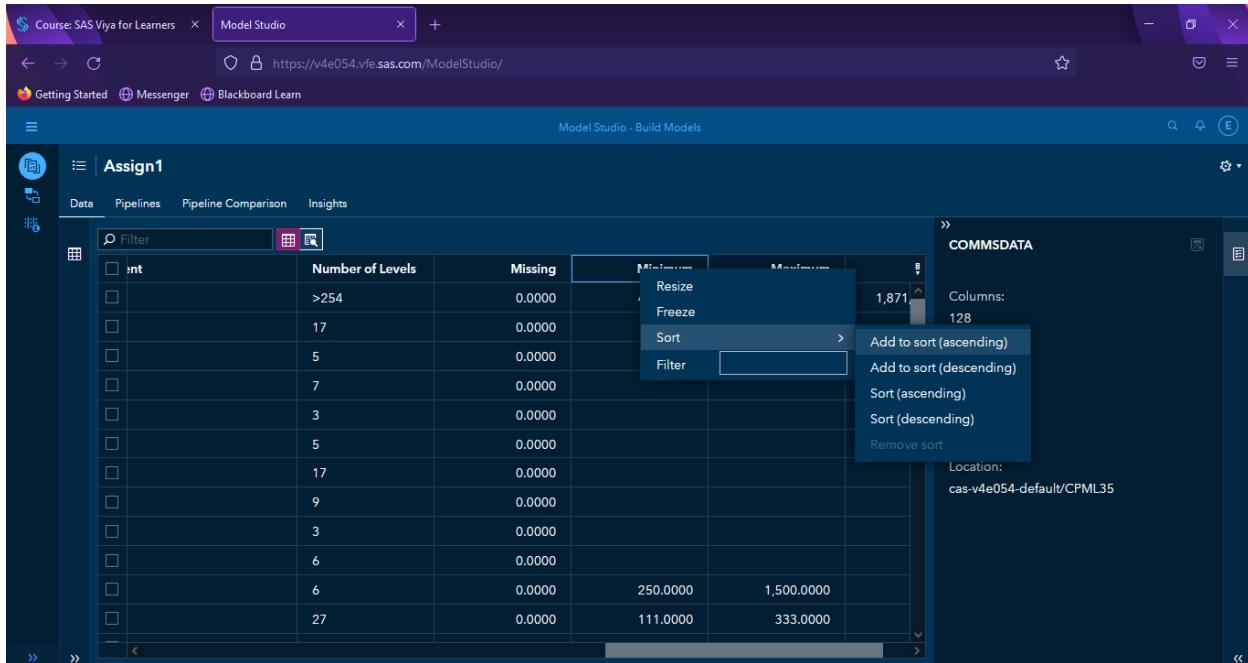
MODIFYING AND CORRECTING SOURCE DATA

Sorting the role in ascending format

The screenshot shows a web browser window for "Model Studio" with the URL "https://v4e054.vfe.sas.com/ModelStudio/". The page title is "Model Studio - Build Models". The main content area displays a data grid titled "Assign1". The grid has columns: "Variable Name", "Label", "Type", "Role", and "Interval". A context menu is open over the "Role" column for the first row, showing options: "Sort (ascending)", "Sort (descending)", and "Remove sort". The "Sort (ascending)" option is highlighted. The "Label" column for the first row is "acct_age" and the "Type" is "Numeric". The "Role" column for the first row is "Input". The "Interval" column for the first row is "Interv". The "Label" column for the second row is "Account Tenure".

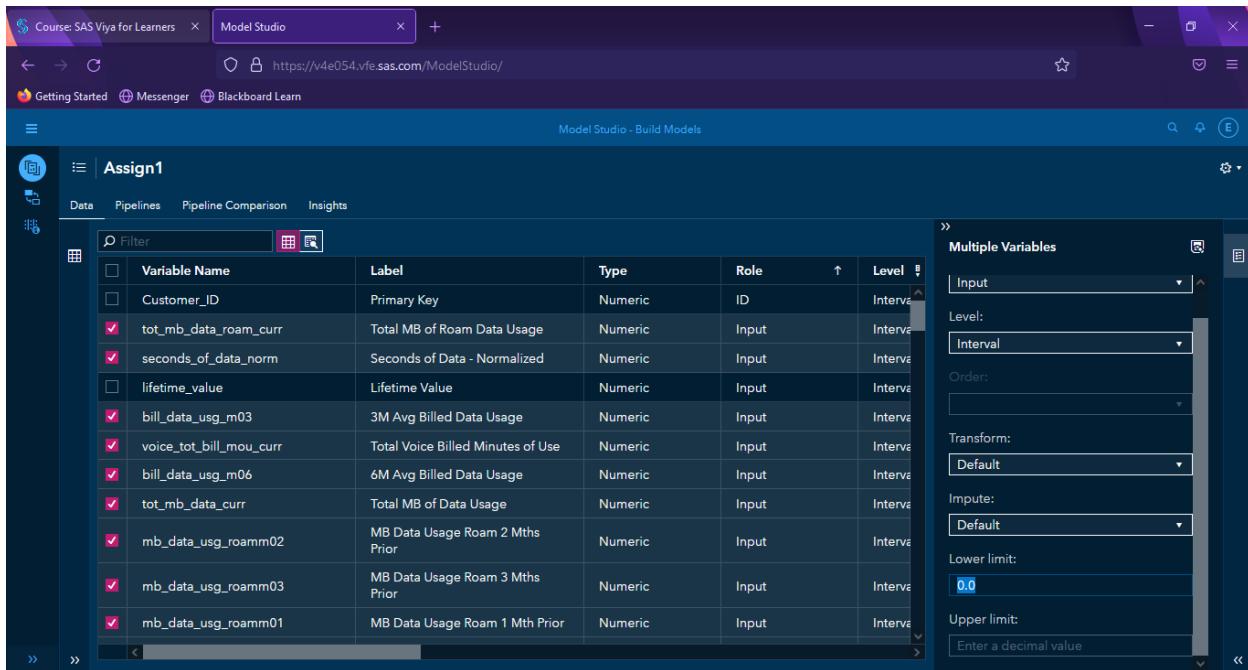
Variable Name	Label	Type	Role	Interval
acct_age	Account Tenure	Numeric	Input	Interv
avg_arpu_3m	3M Avg Revenue per User	Numeric	Input	Interv
avg_data_chrgs_3m	3M Avg Data Charges	Numeric	Input	Interv
avg_data_prem_chrgs_3m	3M Avg Premium Data Charges	Numeric	Input	Interv
avg_days_susp	Days Suspended Last 6M	Numeric	Input	Interv
avg_overage_chrgs_3m	3M Avg Overage Charges	Numeric	Input	Interv
bill_data_usg_m03	3M Avg Billed Data Usage	Numeric	Input	Interv
bill_data_usg_m06	6M Avg Billed Data Usage	Numeric	Input	Interv
bill_data_usg_m09	9M Avg Billed Data Usage	Numeric	Input	Interv
bill_data_usg_tot	Total Billed Data Usage	Numeric	Input	Interv
billing_cycle	Billing Cycle	Numeric	Input	Nomir
call_category_1	Call Center Category 1	Character	Input	Nomir

Sorting the minimum column



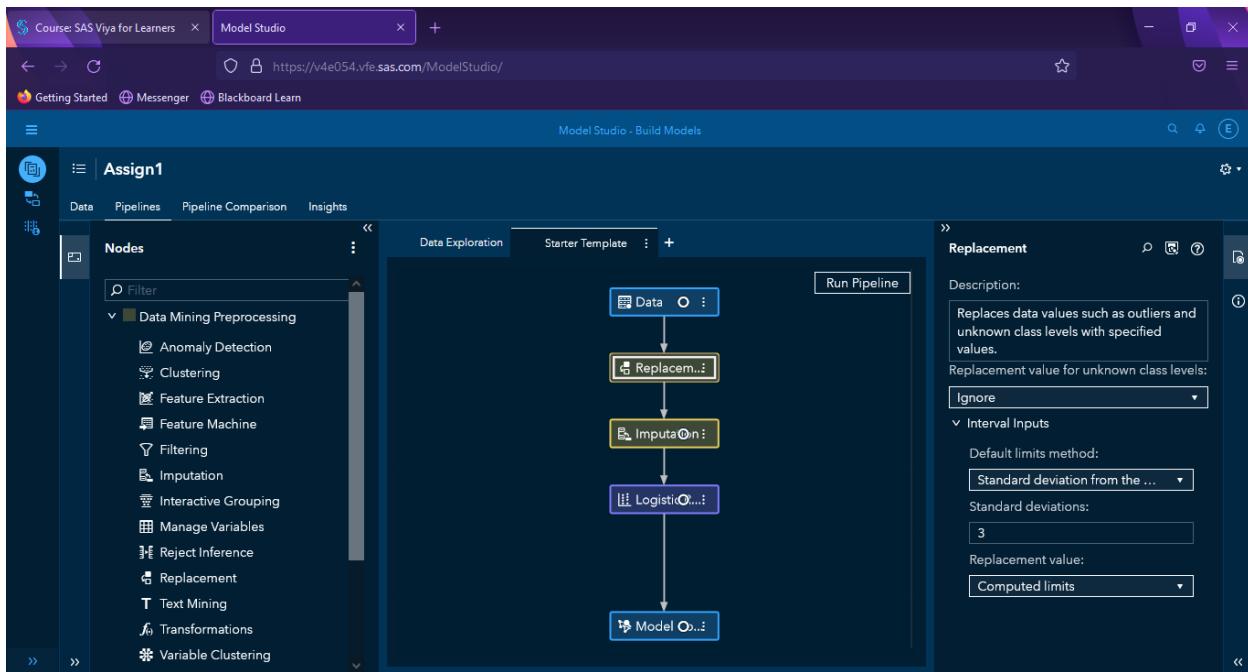
A screenshot of the SAS Viya Model Studio interface. The main area shows a data grid titled 'Assign1'. A context menu is open over the 'Minimum' column header, listing options: Resize, Freeze, Sort, Filter, Add to sort (ascending), Add to sort (descending), Sort (ascending), Sort (descending), and Remove sort. The 'Sort' option is highlighted. To the right of the grid, there is a panel titled 'COMMSDATA' showing 'Columns: 128' and a location path 'cas-v4e054-default/CPML35'.

Selecting the input variable and adjusting the lower limit to 0.0

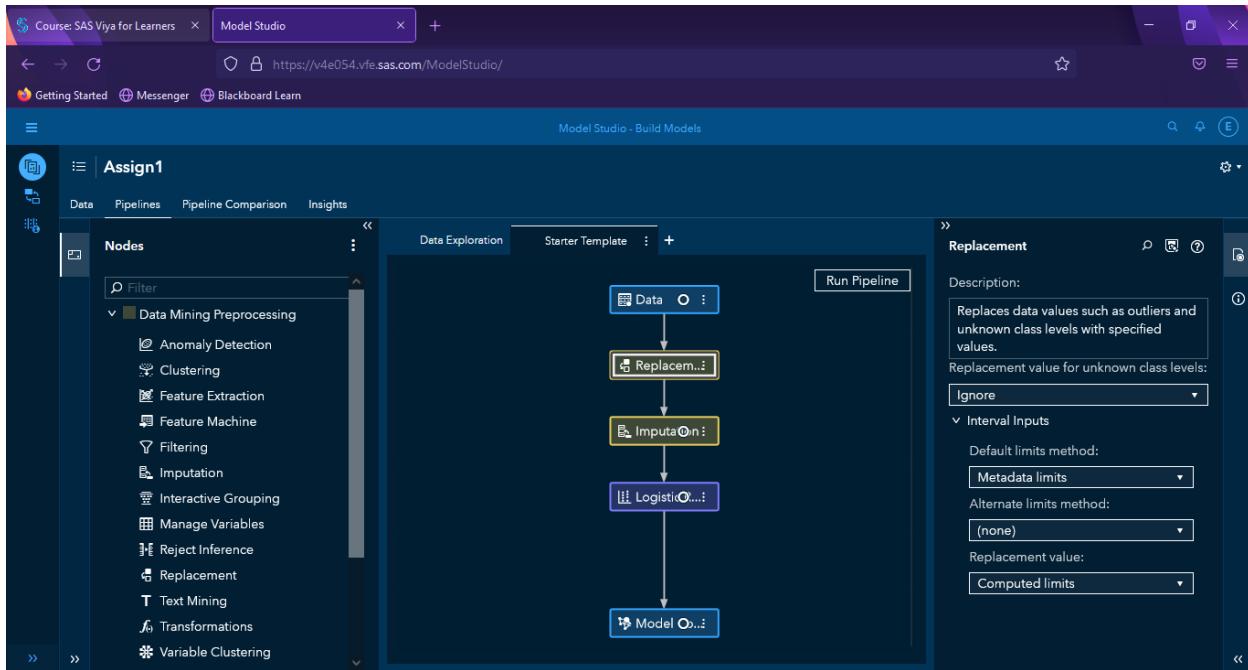


A screenshot of the SAS Viya Model Studio interface. The main area shows a data grid titled 'Assign1' with several variables selected. To the right, a panel titled 'Multiple Variables' is open, showing settings for 'Input' (Level: Interval, Order: , Transform: Default, Impute: Default). The 'Lower limit' field is set to '0.0' and the 'Upper limit' field is set to 'Enter a decimal value'.

Adding replacement in Starter template pipeline



Adjusting the interval input setting for the replacement node making the default limit method into metadata limit and alternate limit method into none



Running the replacement node

The screenshot shows the SAS Viya Model Studio interface. On the left, the 'Nodes' pane is open, displaying various data mining preprocessing nodes like Anomaly Detection, Clustering, Feature Extraction, etc. In the center, a pipeline diagram shows a flow from 'Data' to 'Replacement...', then to 'Imputation'. A context menu is open over the 'Replacement...' node, with 'Run' highlighted. To the right, the 'Replacement' node properties pane is visible, showing its description: 'Replaces data values such as outliers and unknown class levels with specified values.' It also includes settings for 'Replacement value for unknown class levels' (set to 'Ignore'), 'Default limits method' (set to 'Metadata limits'), and 'Alternate limits method' (set to '(none)').

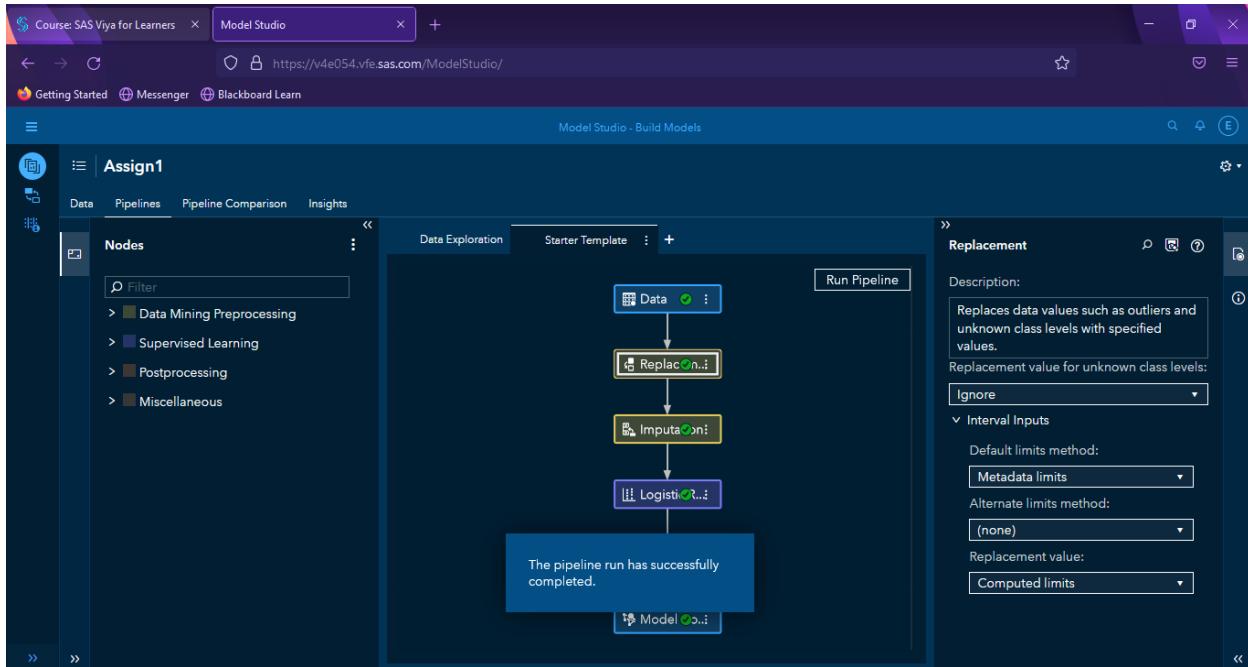
Checking the result of the replacement node

The screenshot shows the 'Assignment 1 > "Replacement" Results' page in Model Studio. It displays two main tables: 'Interval Variables' and 'Replacement Counts'. The 'Interval Variables' table lists four variables: BILL_DATA_USG_M03, BILL_DATA_USG_M06, CALLS_IN_O_FFPK, and CALLS_IN_P_K. The 'Replacement Counts' table lists the same variables with their respective replacement counts: 2,664, 1,891, 3,460, and 4,083. Below these tables are sections for 'Node Score Code' (containing code 1, 2, 3) and 'Properties' (with a single entry for 'unknownLevel' set to 'IGNORE').

Name	Variable L...	Replace V...	Limits Met...	Lower Limit
BILL_DATA_USG_M03	3M Avg Billed Data Usage	REP_BILL_D ATA_USG_M_03	METALIMIT	0
BILL_DATA_USG_M06	6M Avg Billed Data Usage	REP_BILL_D ATA_USG_M_06	METALIMIT	0
CALLS_IN_O_FFPK	Calls Incoming Off-Peak	REP_CALLS_IN_OFFPK	METALIMIT	0
CALLS_IN_PK	Calls Incoming Peak	REP_CALLS_IN_PK	METALIMIT	0

Name	Variable L...	Train	Role	Variable L...
BILL_DATA_USG_M03	3M Avg Billed Data Usage	2,664	INPUT	INTERVAL
BILL_DATA_USG_M06	6M Avg Billed Data Usage	1,891	INPUT	INTERVAL
CALLS_IN_O_FFPK	Calls Incoming Off-Peak	3,460	INPUT	INTERVAL
CALLS_IN_PK	Calls Incoming Peak	4,083	INPUT	INTERVAL

Running the whole pipeline



Checking the result for model comparison

The screenshot shows the 'Model Comparison' results table in the SAS Viya Model Studio. The table has columns: Champion, Name, Algorithm Name, KS (Youden), and Misclassification Rate. One row is present, showing Logistic Regression as the champion with an algorithm name of Logistic Regression, a KS (Youden) value of 0.5591, and a misclassification rate of 0.0705. Below the table, a properties panel shows the 'selectionCriteriaClass' property set to 'Kolmogorov-Smirnov statistic (KS)'.

Champion	Name	Algorithm Name	KS (Yoden)	Misclassification Rate
[Icon]	Logistic Regression	Logistic Regression	0.5591	0.0705

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)

TRANSFORMING INPUTS

Rerunning the data exploration

The screenshot shows the SAS Model Studio interface with a pipeline titled "Assign1". The pipeline consists of two steps: "Data" and "Data Explorer". A success message at the bottom states: "The pipeline run has successfully completed." On the right side, there is a panel titled "Data Exploration" with settings for variable selection and interval variables.

Description:
Displays summary statistics and plots for variables in your data.

Input data partition:
All data

Variable Selection
Variable selection criterion:
Importance

Maximum number of variables to select:
50

Screening Cutoffs

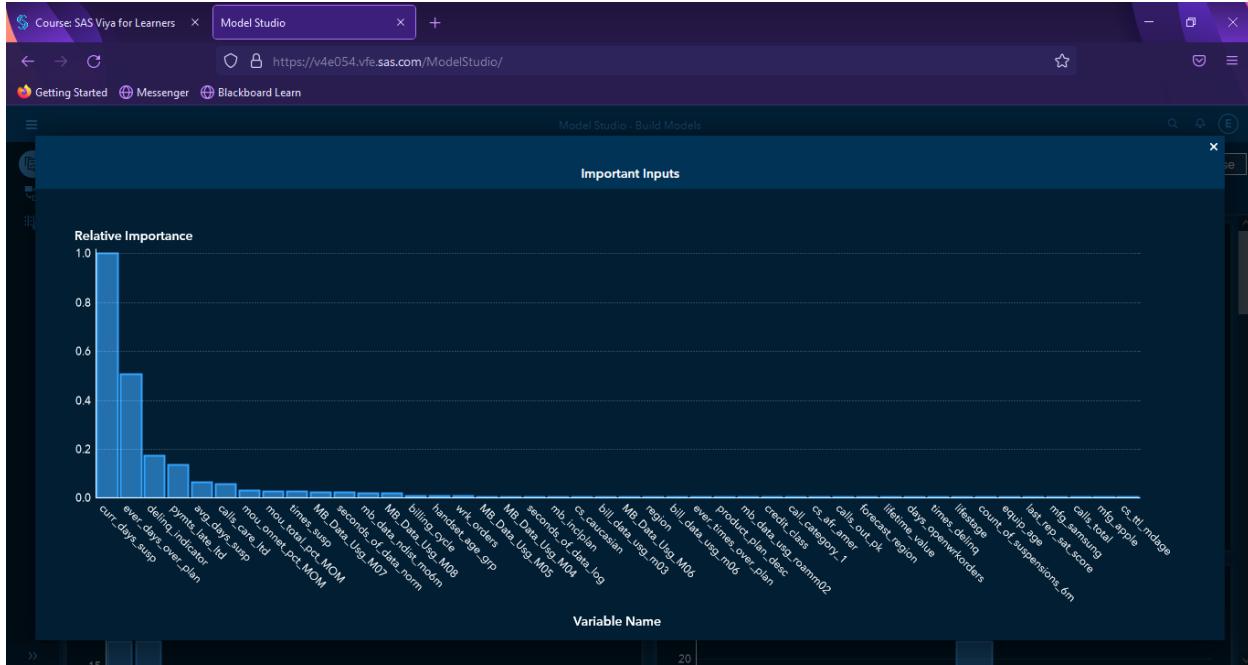
Interval Variables
Number of bins:

Checking the interval variable moments

The screenshot shows the SAS Model Studio interface displaying the "Interval Variable Moments" report. The report lists various variables along with their minimum, maximum, mean, standard deviation, skewness, kurtosis, relative variance, and mean plus/minus three standard deviations.

Variable Name	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis	Relative Variance	Mean plus 3SD	Mean minus 3SD
MB_Data_Us_g_M04	0	14,606	159.3069	381.1479	9.8152	179.7305	2.3925	921.6027	-602.985
MB_Data_Us_g_M05	0	24,707	142.7953	471.5578	13.6122	315.4086	3.3023	1,085.9108	-800.320
MB_Data_Us_g_M06	0	29,676	230.5486	718.7864	15.2432	360.4407	3.1177	1,668.1214	-1,207.024
MB_Data_Us_g_M07	0	13,672	94.2740	259.8391	17.6345	495.6026	2.7562	613.9521	-425.404
MB_Data_Us_g_M08	0	16,297	109.5912	348.7336	16.9031	467.9811	3.1821	807.0585	-587.876
avg_days_susp	0	62	3.4714	3.8313	1.5937	5.0681	1.1037	11.1339	-4.191
bill_data_usg_m03	-13,678	40,767.1000	1,864.9142	1,634.5099	1.3974	13.7884	0.8765	5,133.9339	-1,404.105
bill_data_usg_m06	-10,874.3000	16,813.6000	1,869.1400	1,418.9061	1.3000	5.0748	0.7591	4,706.9523	-968.672
calls_care_ltd	0	266	91.3478	49.3820	1.1421	0.3660	0.5406	190.1117	-7.416

Checking important input



Changing the variables into a transform log in the data tab

The screenshot shows the SAS Viya Model Studio interface for building models. The top navigation bar includes tabs for 'Getting Started', 'Messenger', 'Blackboard Learn', and 'Model Studio'. The main title is 'Model Studio - Build Models'. On the left, there's a sidebar with icons for Data, Pipelines, Pipeline Comparison, and Insights. The main area is titled 'Assign1' and contains a table with columns: Variable Name, Label, Type, Role, and Level. The table lists various variables like 'ever_times_over_plan', 'forecast_region', and multiple entries for 'MB_Data_Usg_M' (Month 4 to Month 9). To the right of the table is a panel titled 'Multiple Variables' with dropdown menus for Role (set to 'Input'), Level (set to 'Interval'), Order (empty), Transform (set to 'Log'), Impute (set to 'Default'), Lower limit ('Enter a decimal value'), and Upper limit ('Enter a decimal value').

Variable Name	Label	Type	Role	Level
ever_times_over_plan	Total Times Over Plan	Numeric	Input	Interval
forecast_region	Forecasted Region Key	Numeric	Input	Interval
MB_Data_Usg_M04	MB of Data Usage Month 4	Numeric	Input	Interval
MB_Data_Usg_M05	MB of Data Usage Month 5	Numeric	Input	Interval
MB_Data_Usg_M06	MB of Data Usage Month 6	Numeric	Input	Interval
MB_Data_Usg_M07	MB of Data Usage Month 7	Numeric	Input	Interval
MB_Data_Usg_M08	MB of Data Usage Month 8	Numeric	Input	Interval
MB_Data_Usg_M09	MB of Data Usage Month 9	Numeric	Input	Interval
mfg_apple	Own Apple	Numeric	Input	Binary
mfg_htc	Own HTC	Numeric	Input	Binary
mfg_lg	Own LG	Numeric	Input	Binary
mfg_motorola	Own Motorola	Numeric	Input	Binary

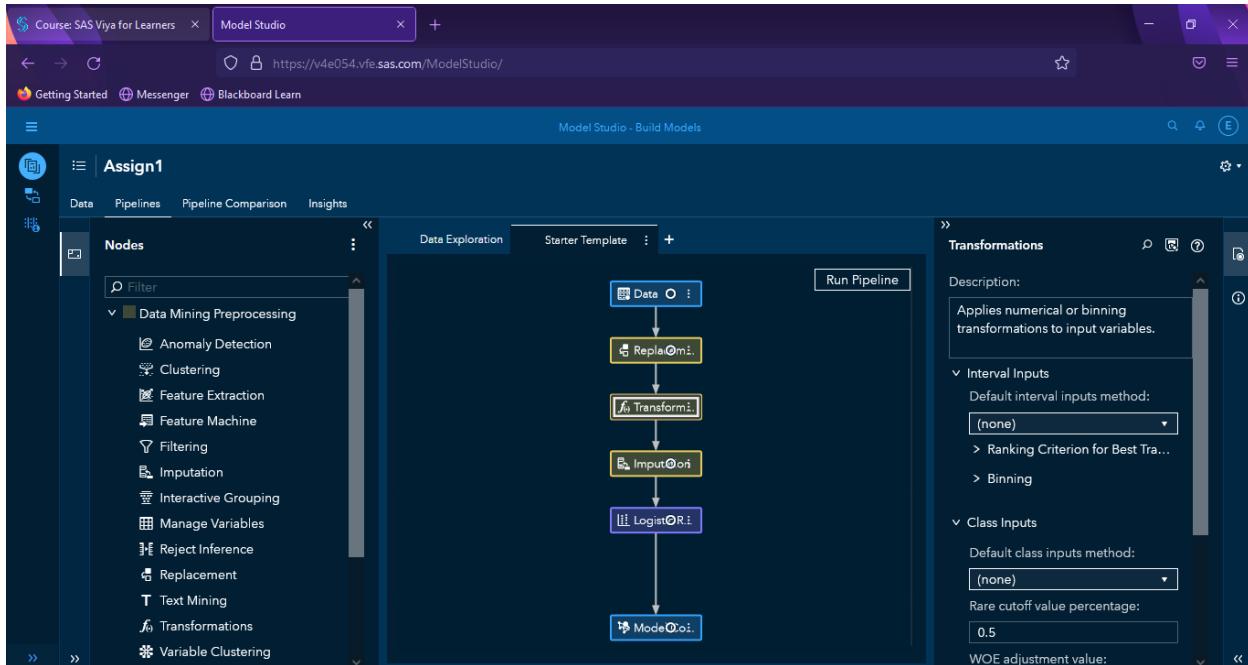
Changing the display of column using manage column

The screenshot shows the 'Manage Columns' dialog box in the SAS Viya Model Studio. The 'Displayed columns' section lists various statistical and metadata variables. The 'Transform' variable is currently selected. The 'Hidden columns' section lists other variables, including several starting with 'MB_Data_Usg_M04' through 'M09'. A tooltip for 'Transform' indicates it is a decimal value.

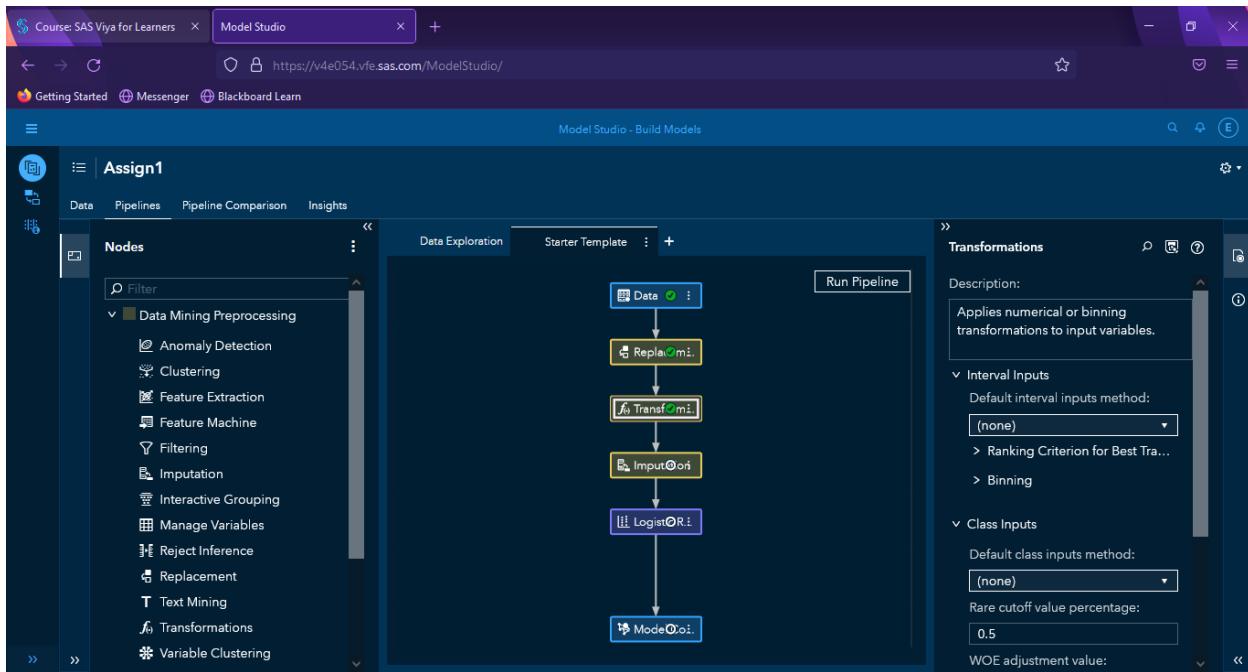
Checking if the configuration for the 6 variable is saved

The screenshot shows the 'Multiple Variables' configuration panel in the SAS Viya Model Studio. The table lists variables with their respective statistical properties and transformation types. The 'Transform' column includes entries like 'Default', 'Log', and 'Default'. The configuration panel on the right specifies 'Role: Input', 'Level: Interval', 'Transform: Log', and 'Impute: Default'.

Adding transformation node between replacement and imputation node



Running the transformation node



Checking the result of transformation node

The screenshot shows the 'Assign1 > "Transformations" Results' page in Model Studio. The 'Input Variable Statistics' table lists variables like acct_age, avg_arpu_3m, avg_data_chrgs_3m, avg_days_sp, and avg_data_usage, with their respective types, counts, and minimum values. The 'Transformed Variables Summary' table shows four LOG transformation nodes with formulas involving MB_Data_Usag_M04 through M07. The 'Properties' section shows a single entry for 'defaultMethod' set to 'NONE'. The 'Node Score Code' section contains a snippet of code related to the transformations.

Name	Variable L...	Number o...	Percent Mi...	Minimum
acct_age	INTERVAL	0	0	18
avg_arpu_3m	INTERVAL	807	2.0384	0
avg_data_chrgs_3m	INTERVAL	807	2.0384	0
avg_data_usage	INTERVAL	807	2.0384	0
avg_days_sp	INTERVAL	0	0	0

Transform...	Method	Input Vari...	Formula	Variable L...
LOG_MB_Data_Usag_M04	LOG	MB_Data_Usag_M04	log('MB_Data_Usag_M04'n + 1)	INTERVAL
LOG_MB_Data_Usag_M05	LOG	MB_Data_Usag_M05	log('MB_Data_Usag_M05'n + 1)	INTERVAL
LOG_MB_Data_Usag_M06	LOG	MB_Data_Usag_M06	log('MB_Data_Usag_M06'n + 1)	INTERVAL
LOG_MB_Data_Usag_M07	LOG	MB_Data_Usag_M07	log('MB_Data_Usag_M07'n + 1)	INTERVAL

Property Name	Property Value
defaultMethod	NONE

```

1
2 * Transformation Method = LOG ;
3 Label 'LOG_MB_Data_Usag_M04'n = 'Transformed MB of Data Usage'

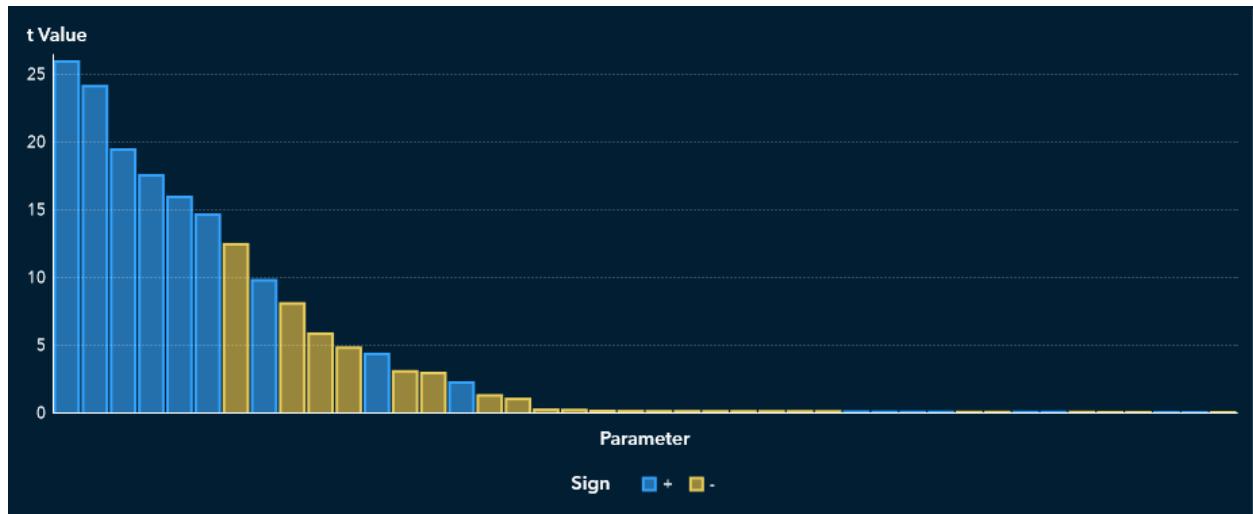
```

Running the whole pipeline

The screenshot shows the 'Assign1' pipeline in the Model Studio Pipeline Editor. The pipeline consists of a sequence of nodes: Data, Replace, Transform, Impute, Logistic R.E., and Model. The 'Transform' node is highlighted. The right panel displays the 'Transformations' node properties, including its description ('Applies numerical or binning transformations to input variables'), interval inputs settings (method: (none)), and class inputs settings (method: (none), rare cutoff value percentage: 0.5, WOE adjustment value: 0).

Checking the model comparison result

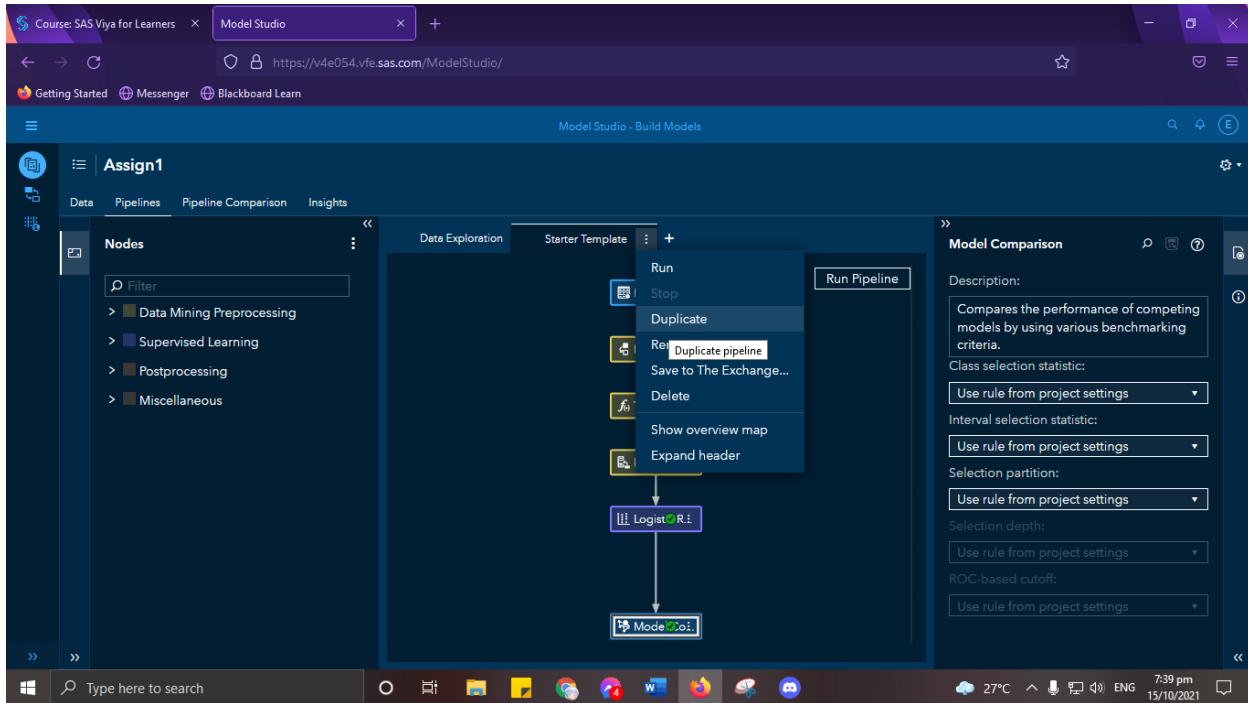
INTERPRETATION:



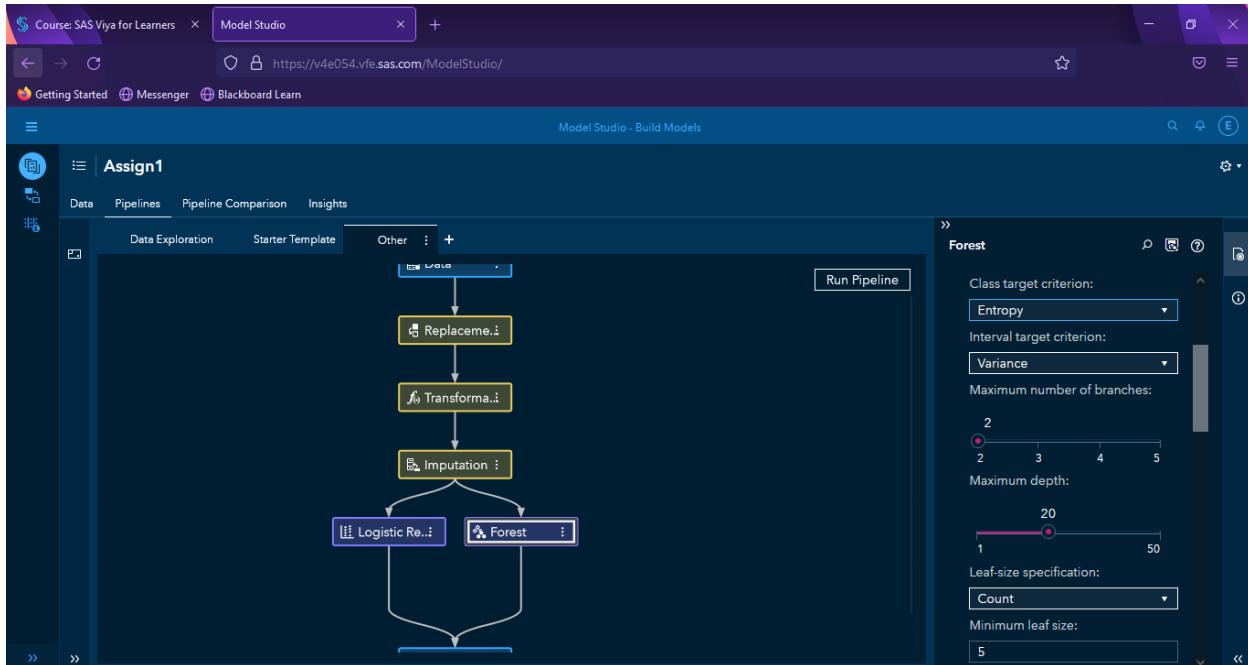
This table shows the absolute value of the t value of each parameter estimate in the logistic regression model. The larger the value is the more significant the value is. Out of all the variables, the curr_days_susp is the most significant having the t value of 25.97.

3. Try other classifiers (e.g., Decision Tree using Gini Index as Split Criterion)

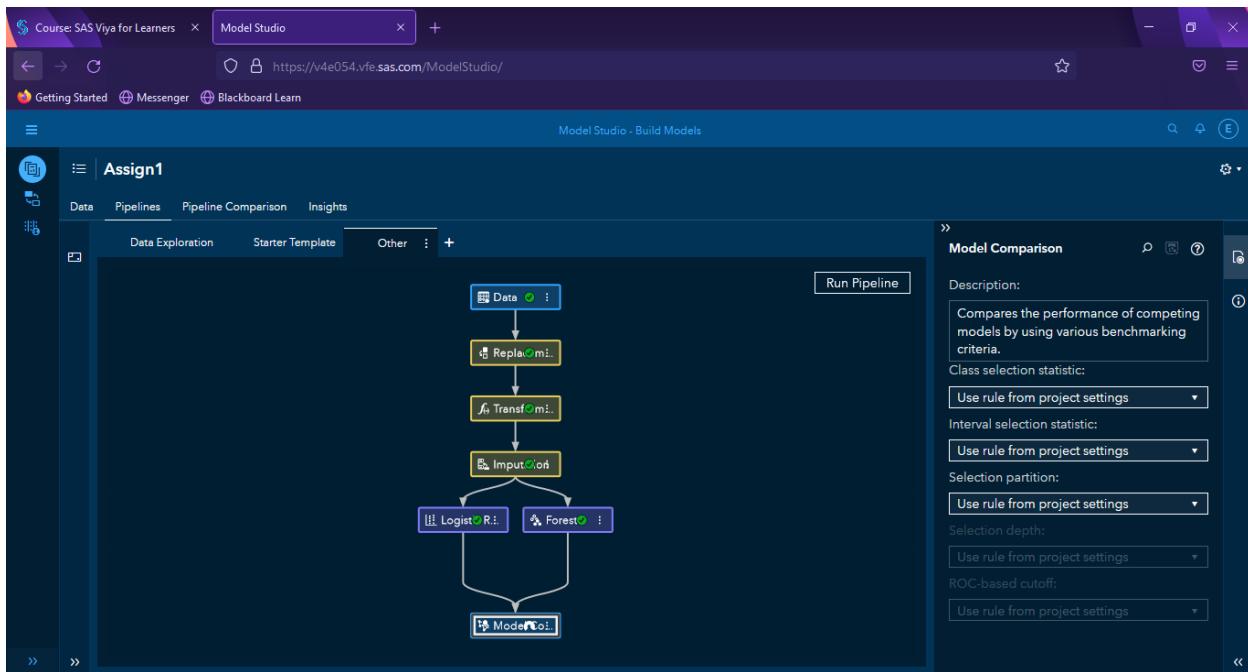
Duplicating the Starter Template and renaming it as *Other Classifiers*



Adding a random forest in the pipeline with the entropy as splitting criterion



Running the whole pipeline



Checking the model comparison results

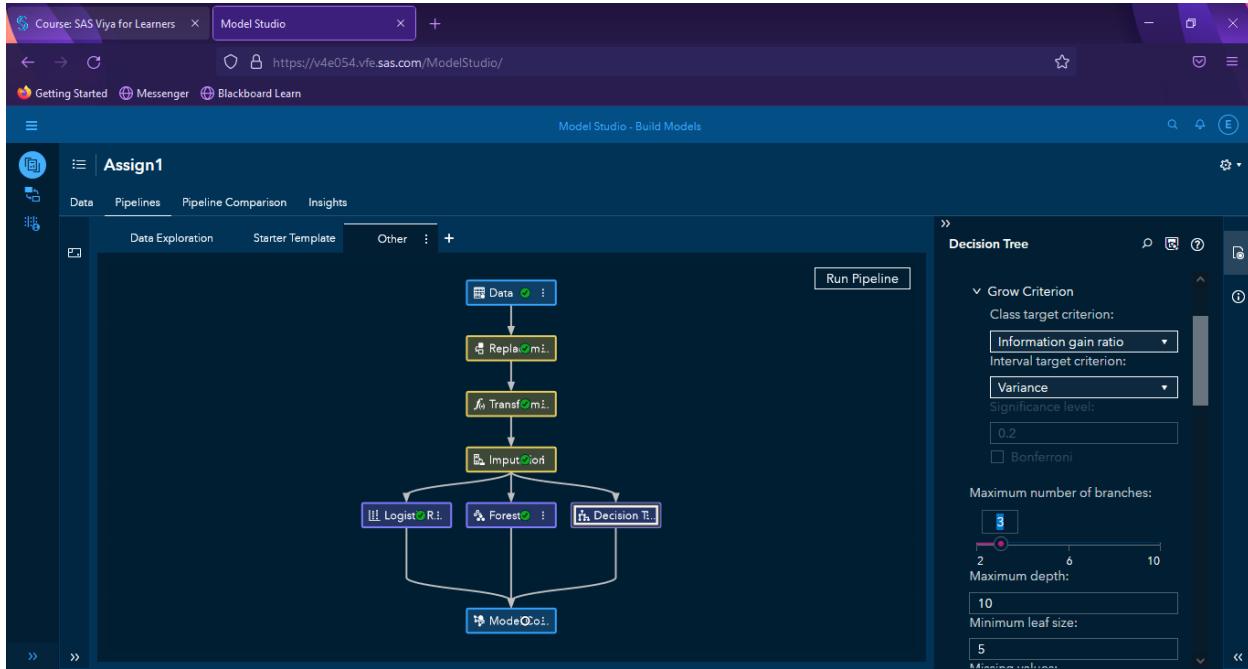
The screenshot shows the "Model Comparison" results table in the SAS Model Studio interface. The table compares two algorithms: Forest and Logistic Regression. The results are as follows:

Champion	Name	Algorithm Name	KS (Youden)	Misclassification Rate
Forest	Forest	Forest	0.5946	0.0593
	Logistic Regression	Logistic Regression	0.5608	0.0694

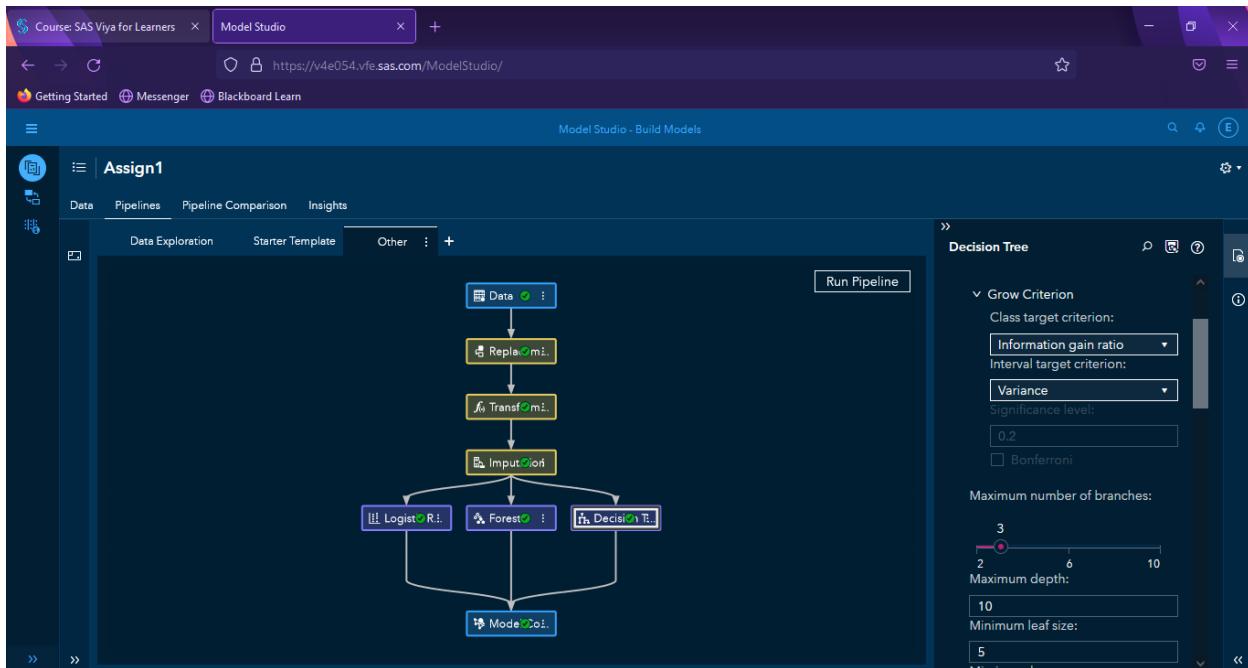
Below the table, a "Properties" section shows the following setting:

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)

Adding another classifier called decision tree with variance as interval target criterion and have a maximum branch of 3



Running the pipeline



Checking the model comparison

Champion	Name	Algorithm Name	KS (Younen)	Misclassification Rate
	Forest	Forest	0.5946	0.0593
	Logistic Regression	Logistic Regression	0.5608	0.0694
	Decision Tree	Decision Tree	0.4544	0.0822

Properties

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)

Adding another classifier in the pipeline called neural network

Assign1

Pipelines Pipeline Comparison Insights

Data Exploration Starter Template Other + Run Pipeline

Neural Network

Description: Fits a fully-connected neural network model.

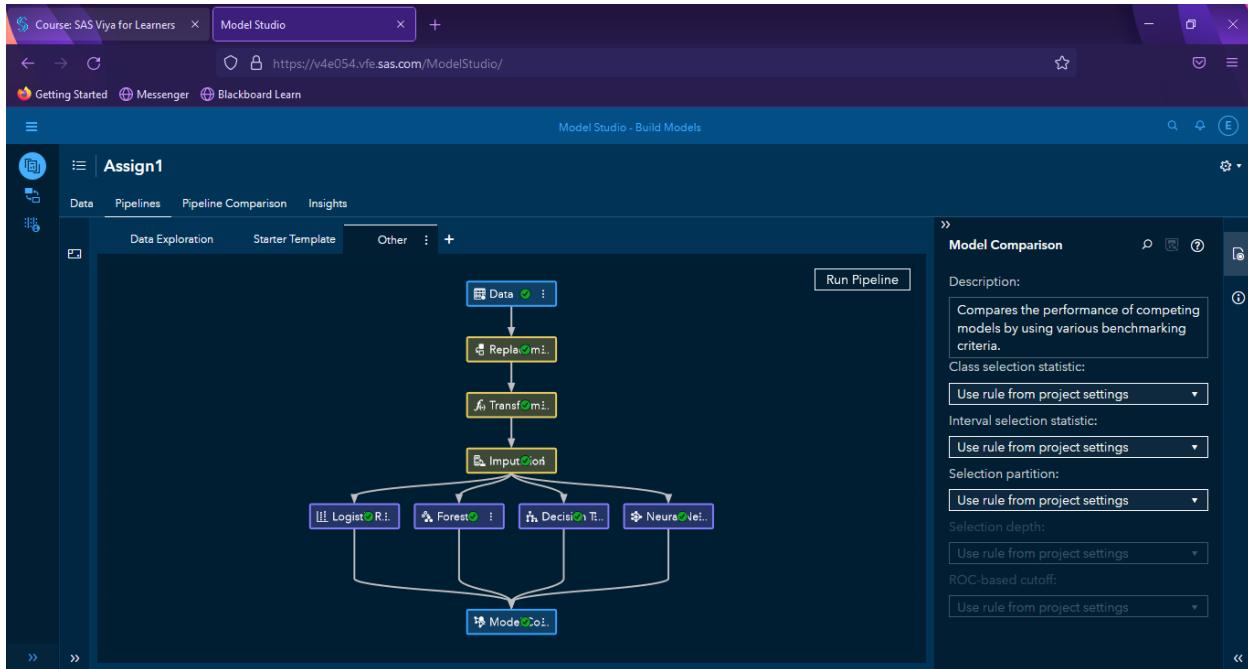
Include missing inputs

Input standardization: Midrange

Number of hidden layers: 1

Hidden Layer Options: Use same number of neurons in hidden layers (checked), Number of neurons per hidden layer: 50

Running the pipeline



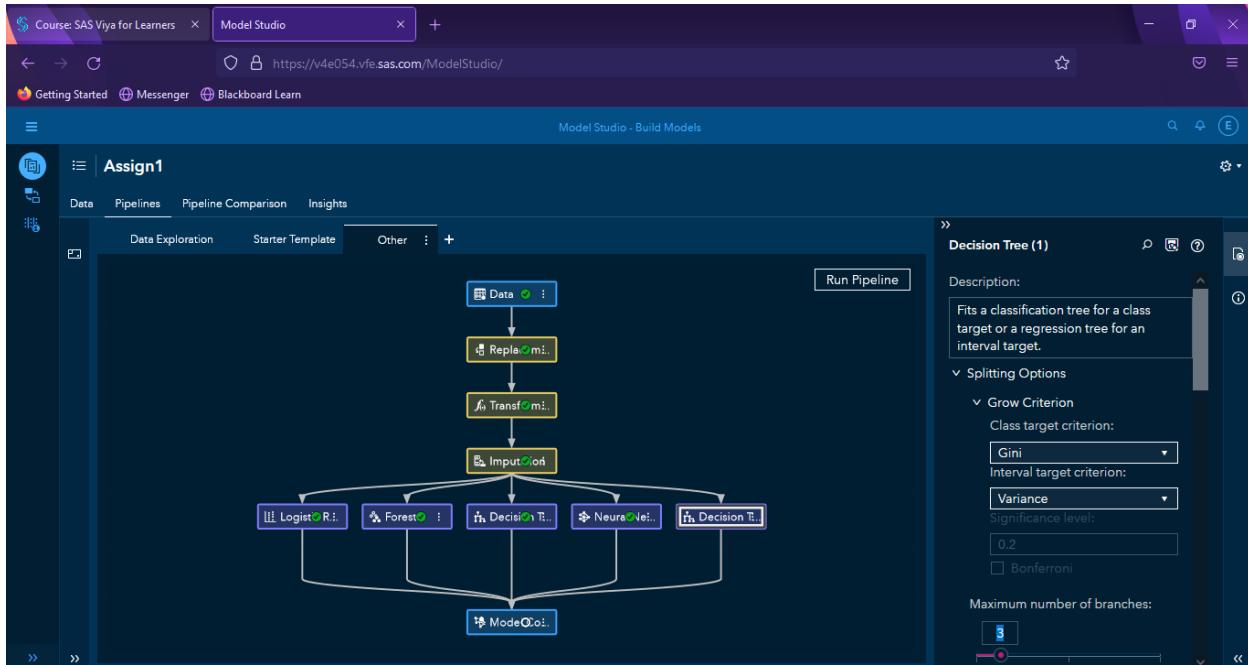
Checking the Model comparison

The screenshot shows the "Model Comparison" results for the "Assign1" pipeline. The results table is as follows:

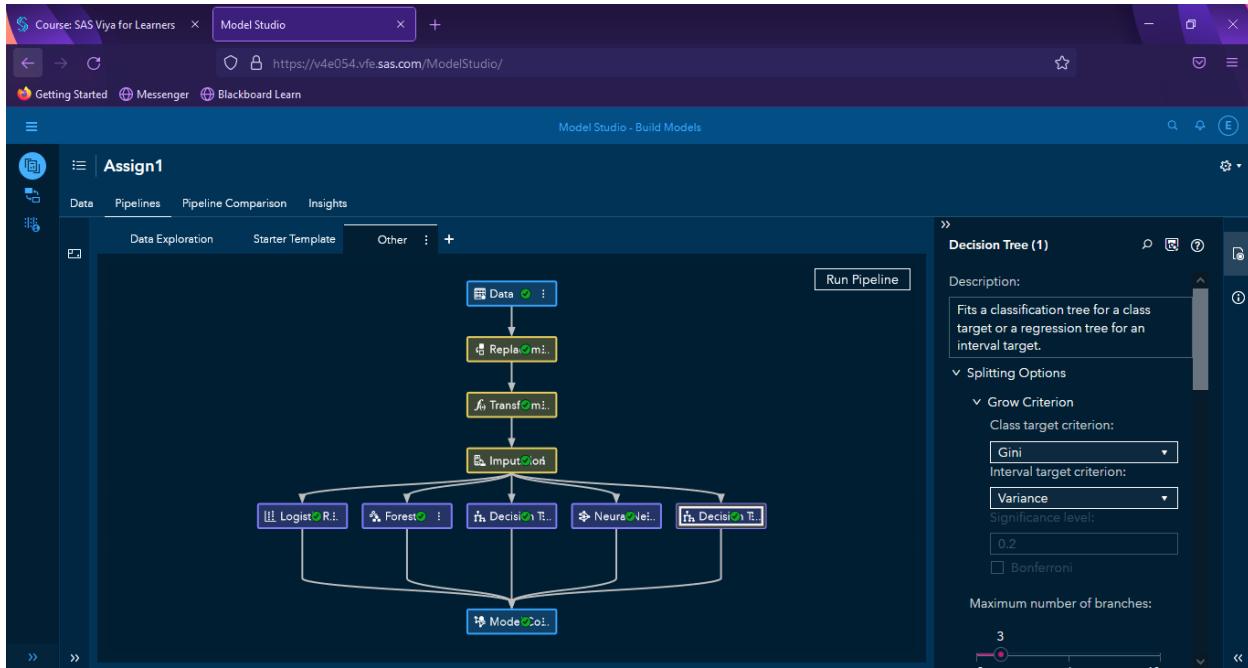
Champion	Name	Algorithm Name	KS (Yoden)	Misclassification Rate
Forest	Forest	Forest	0.5946	0.0593
Logistic Regression	Logistic Regression	Logistic Regression	0.5608	0.0694
Decision Tree	Decision Tree	Decision Tree	0.4544	0.0822
Neural Network	Neural Network	Neural Network	0.0278	0.1214

Below the table, there is a "Properties" section showing the value for "selectionCriteriaClass" as "Kolmogorov-Smirnov statistic (KS)".

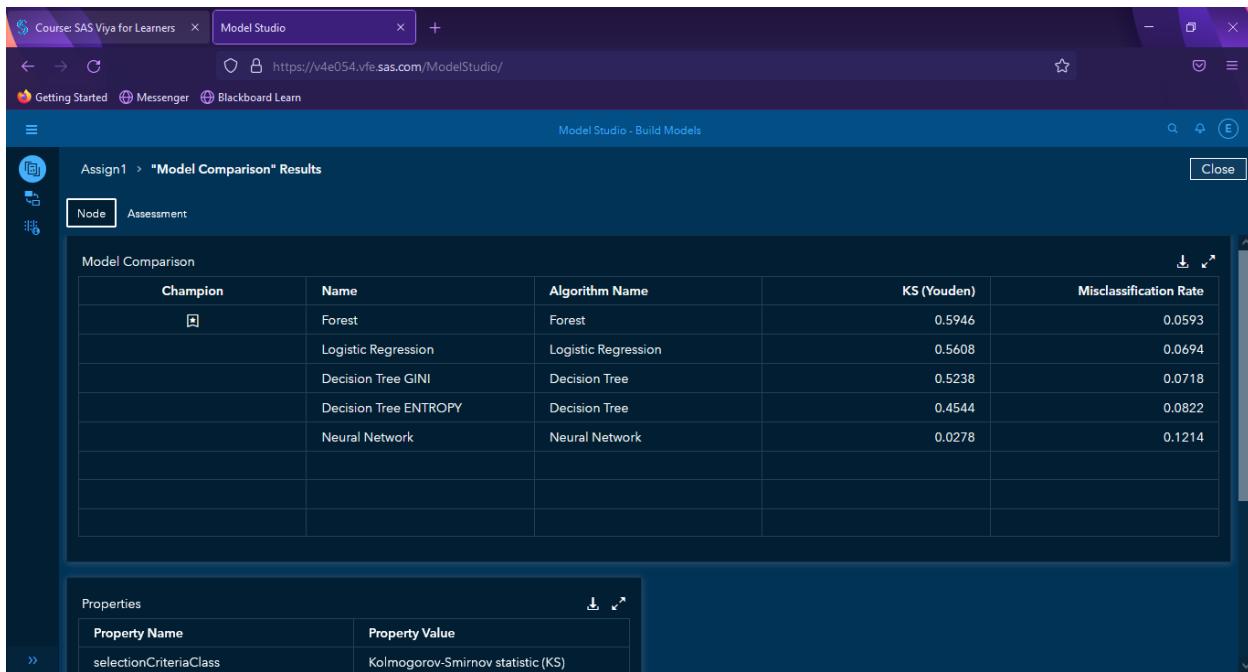
Adding another decision tree using GINI index as split criterion



Running the pipeline



Checking the model comparison



The screenshot shows the SAS Model Studio interface. The main window displays a 'Model Comparison' table titled 'Assign1 > "Model Comparison" Results'. The table lists five models: Forest, Logistic Regression, Decision Tree GINI, Decision Tree ENTROPY, and Neural Network. The columns represent the Champion, Name, Algorithm Name, KS (Younen), and Misclassification Rate. The 'Forest' model has the highest KS value (0.5946) and the lowest misclassification rate (0.0593). The 'Neural Network' model has the lowest KS value (0.0278) and the highest misclassification rate (0.1214). Below the table, a 'Properties' panel is visible, showing a single entry for 'selectionCriteriaClass' with the value 'Kolmogorov-Smirnov statistic (KS)'.

Champion	Name	Algorithm Name	KS (Younen)	Misclassification Rate
	Forest	Forest	0.5946	0.0593
	Logistic Regression	Logistic Regression	0.5608	0.0694
	Decision Tree GINI	Decision Tree	0.5238	0.0718
	Decision Tree ENTROPY	Decision Tree	0.4544	0.0822
	Neural Network	Neural Network	0.0278	0.1214

Property Name	Property Value
selectionCriteriaClass	Kolmogorov-Smirnov statistic (KS)

INTERPRETATION: In the model comparison, we can see that out of all five (forest, logistic regression, decision tree GINI, decision tree entropy, and neural network) the forest gave the highest result while Neural Network gave the lowest, what I think the reason is because Neural Network require a lot of data to train it effectively. On the other hand, the classifier that have the lowest misclassification rate is Neural Network while Decision Tree Entropy gave the highest.