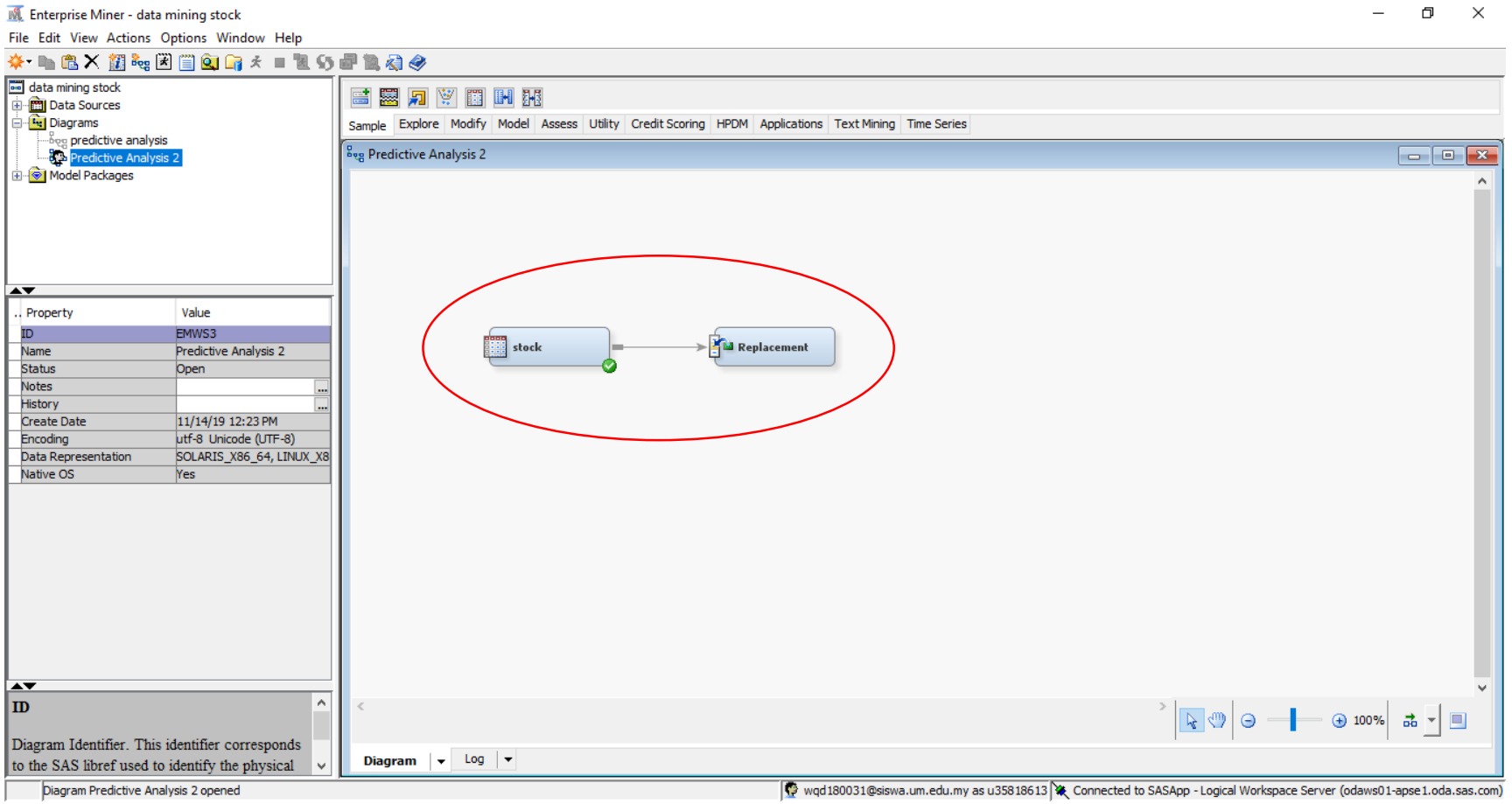
Milestone 4
Interpretation of Data

ZULKANAIN BIN HASAN

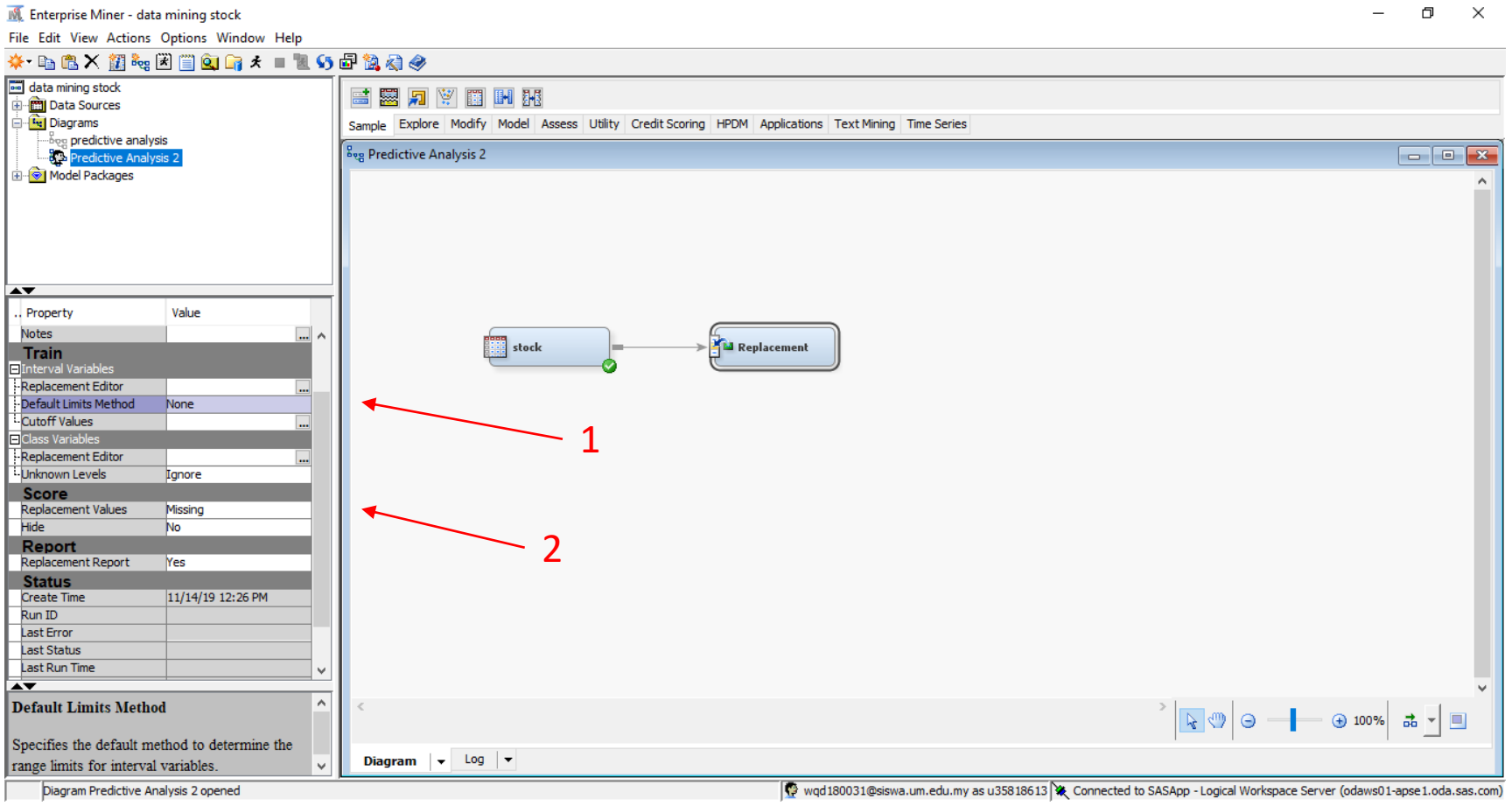
WQD180031

Import data source STOCK and REPLACEMENT node to diagram

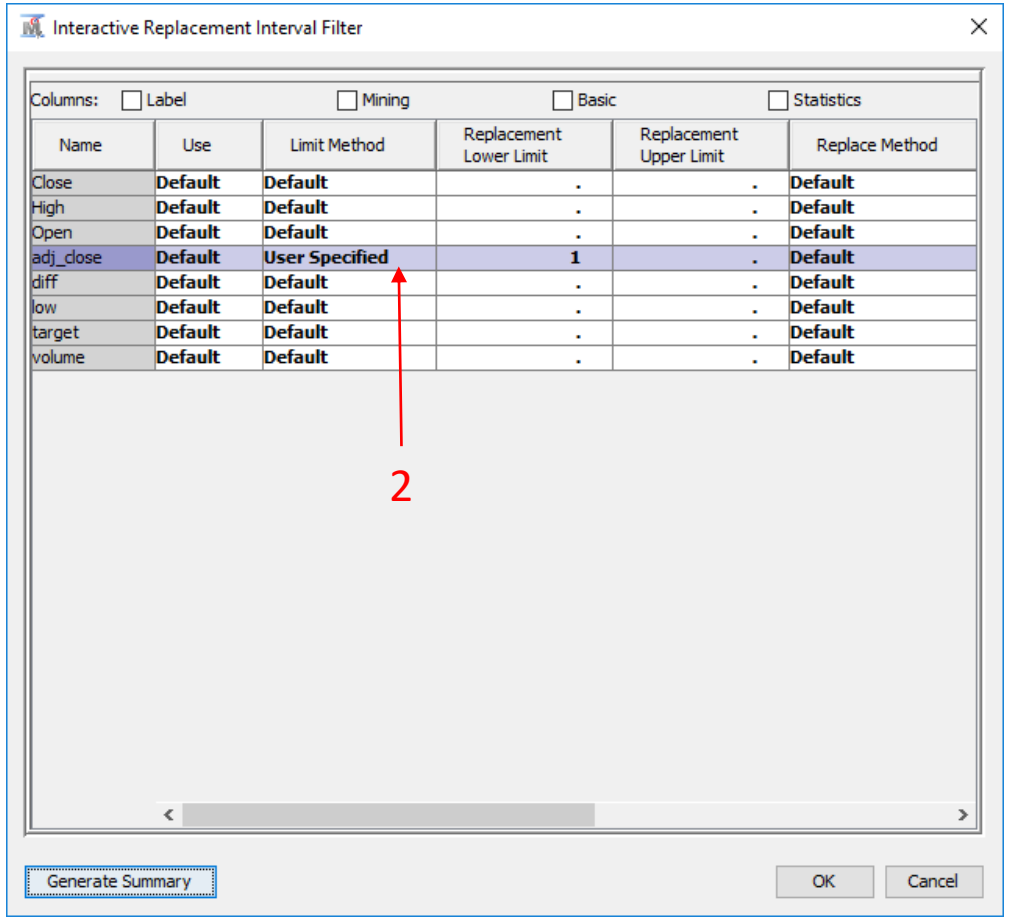


Change:

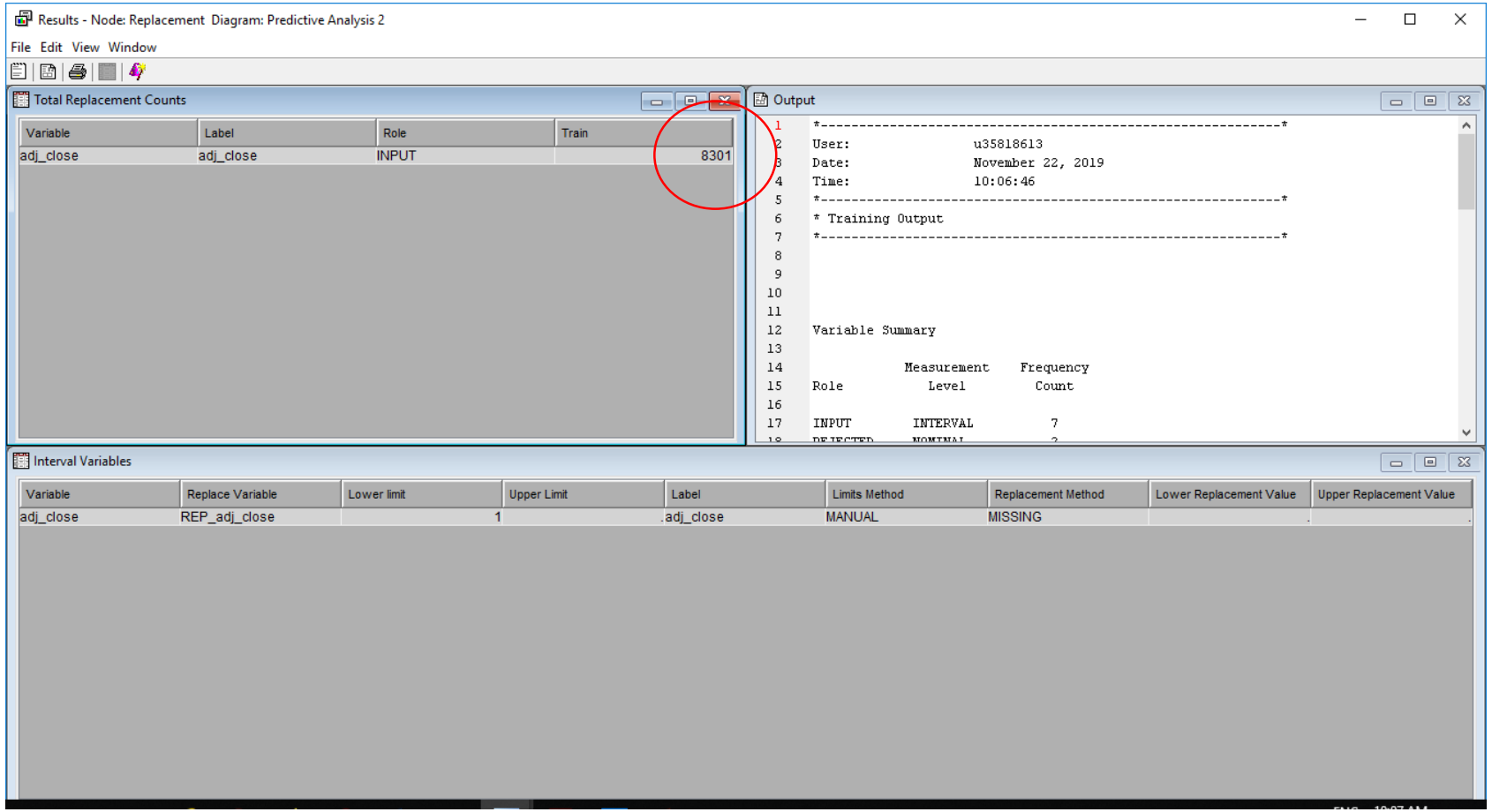
1. Default Limits Method to None
2. Replacement Values to Missing



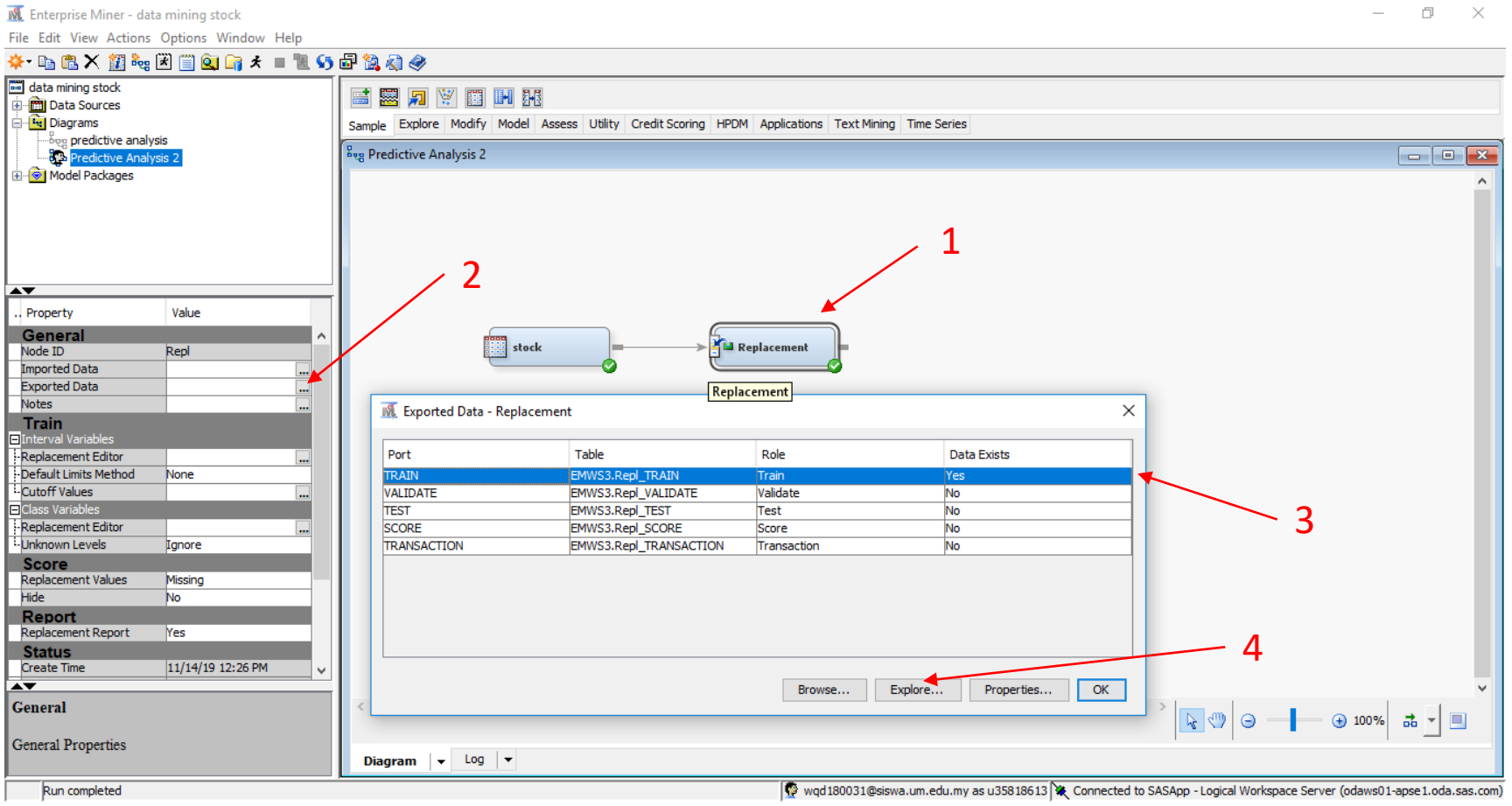
1. Open Replacement Editor for Interval Variables
2. Set adj_close to User Specified at set 1 for it Replacement Lower Limit



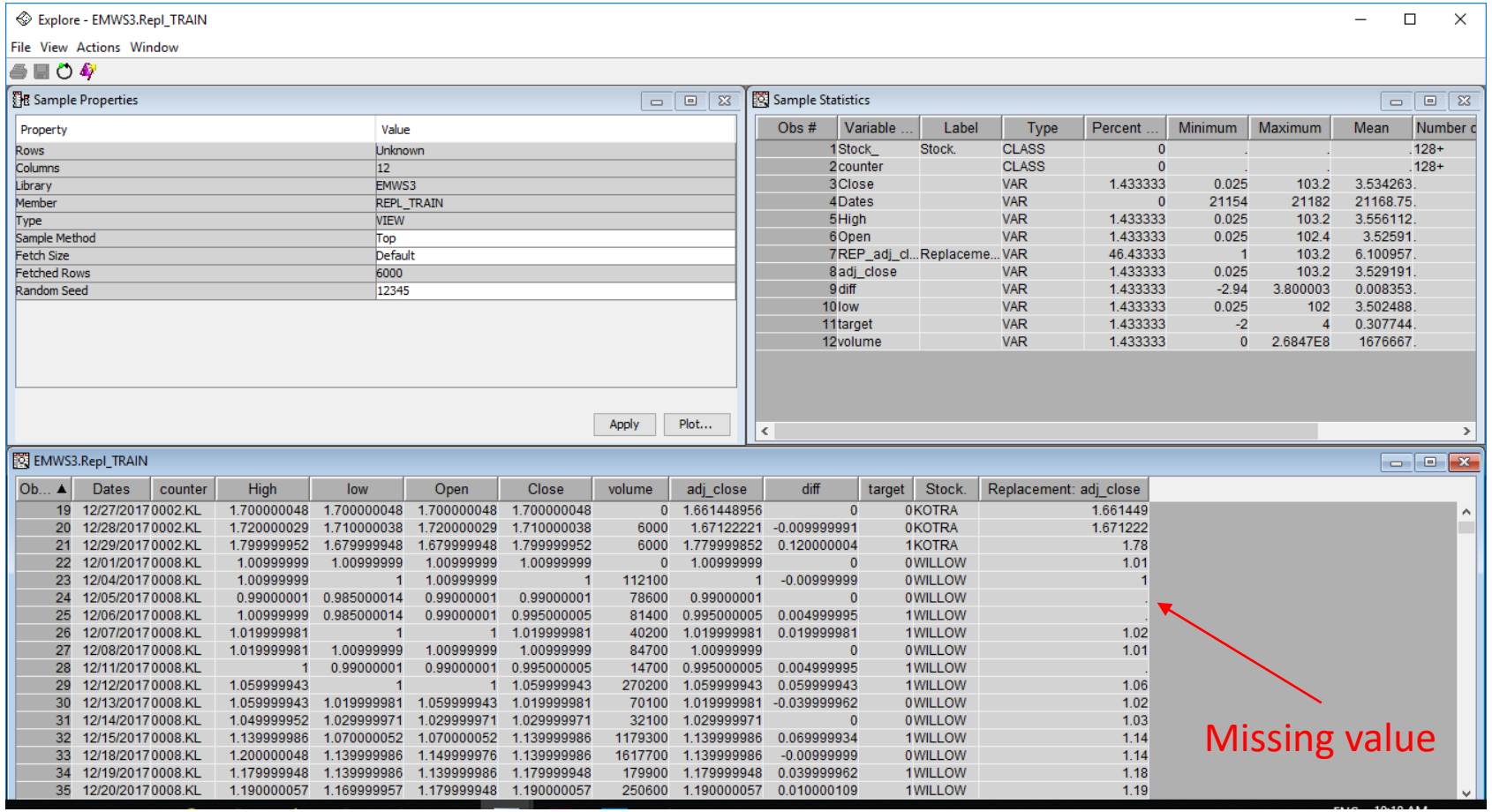
1. Run the REPLACEMENT node and view the results.
2. Result shows 8301 have been Train



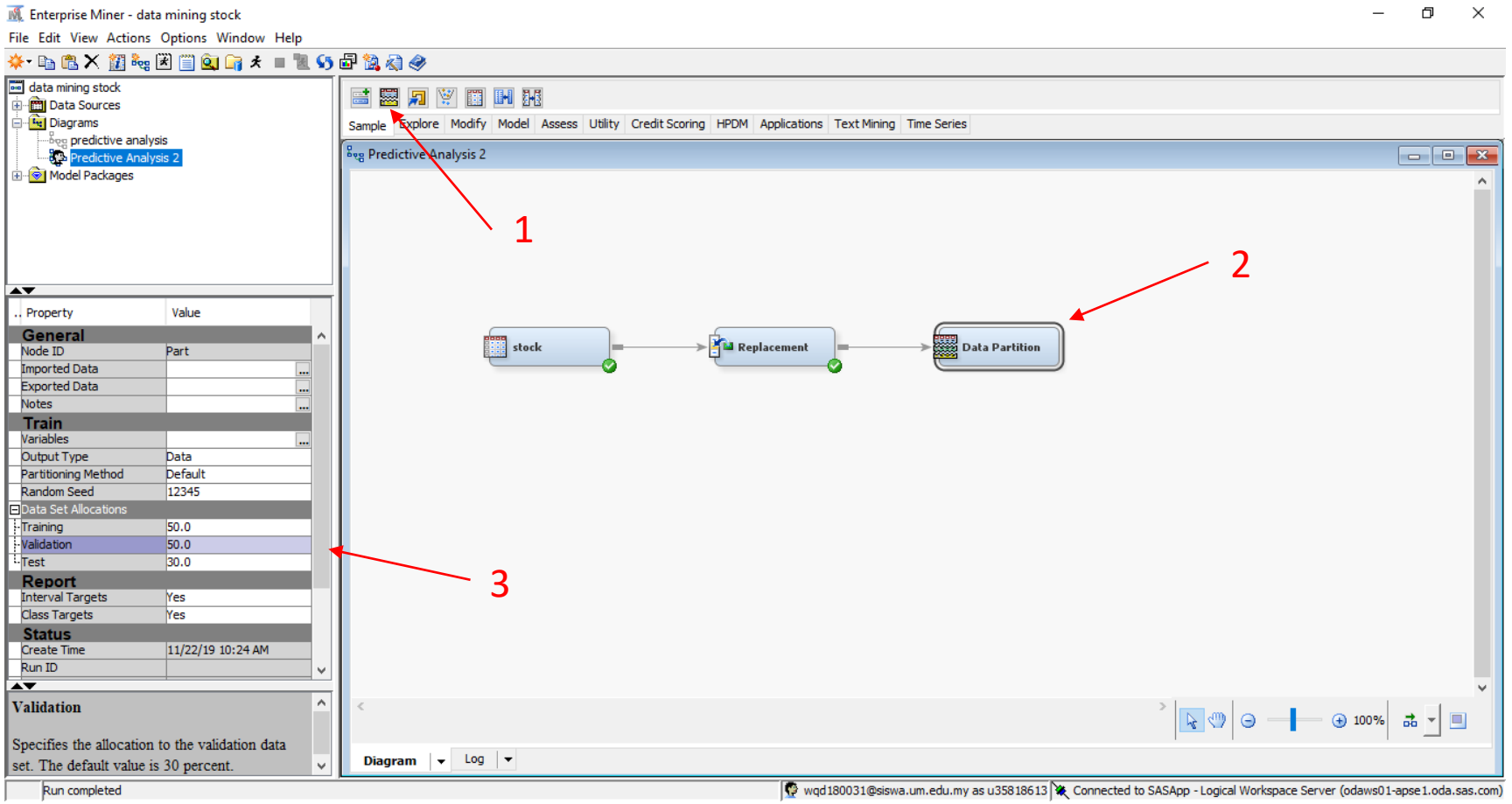
1. Go back to Diagram and select REPLACEMENT node
2. Open Exported Data in General
3. Select TRAIN and Explore



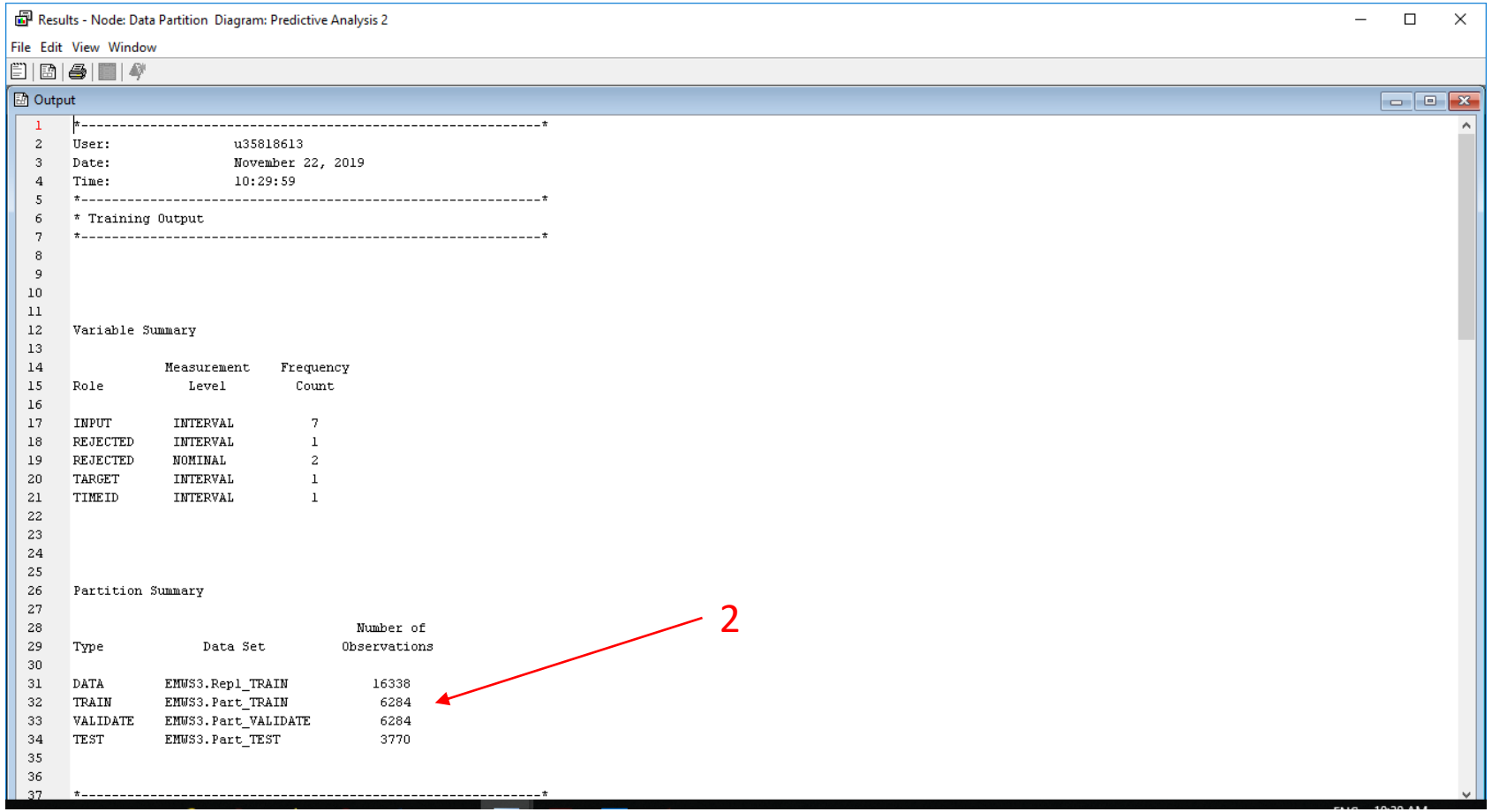
1. TRAIN table shows Replacement have been done to adj_close as per data table
2. Dot (.) indicates a missing value



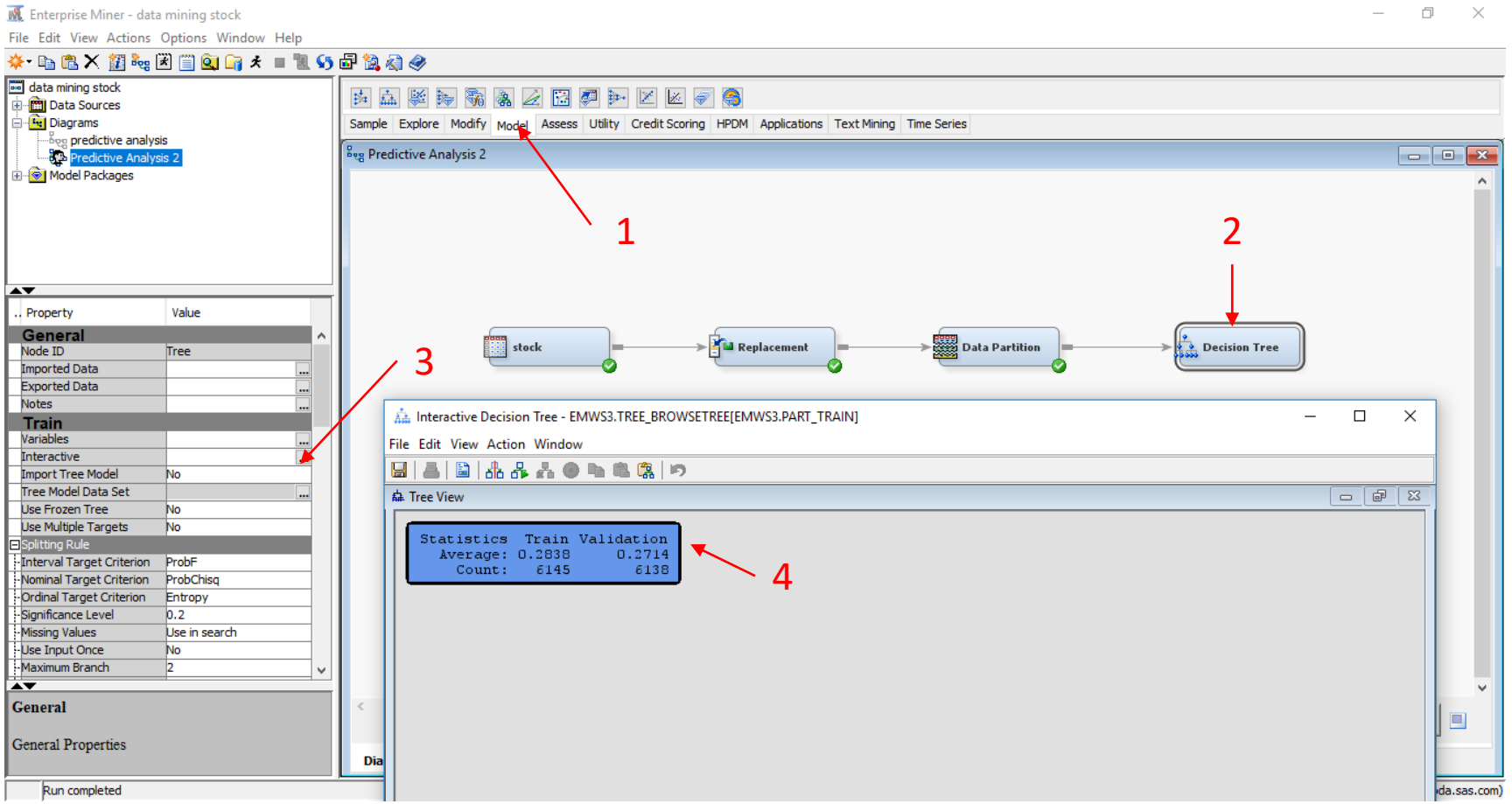
1. Open Sample and add DATA PARTITION node to diagram
2. Select DATA PARTITION node
3. Change Data Set Allocations Training to 50 and Validation to 50



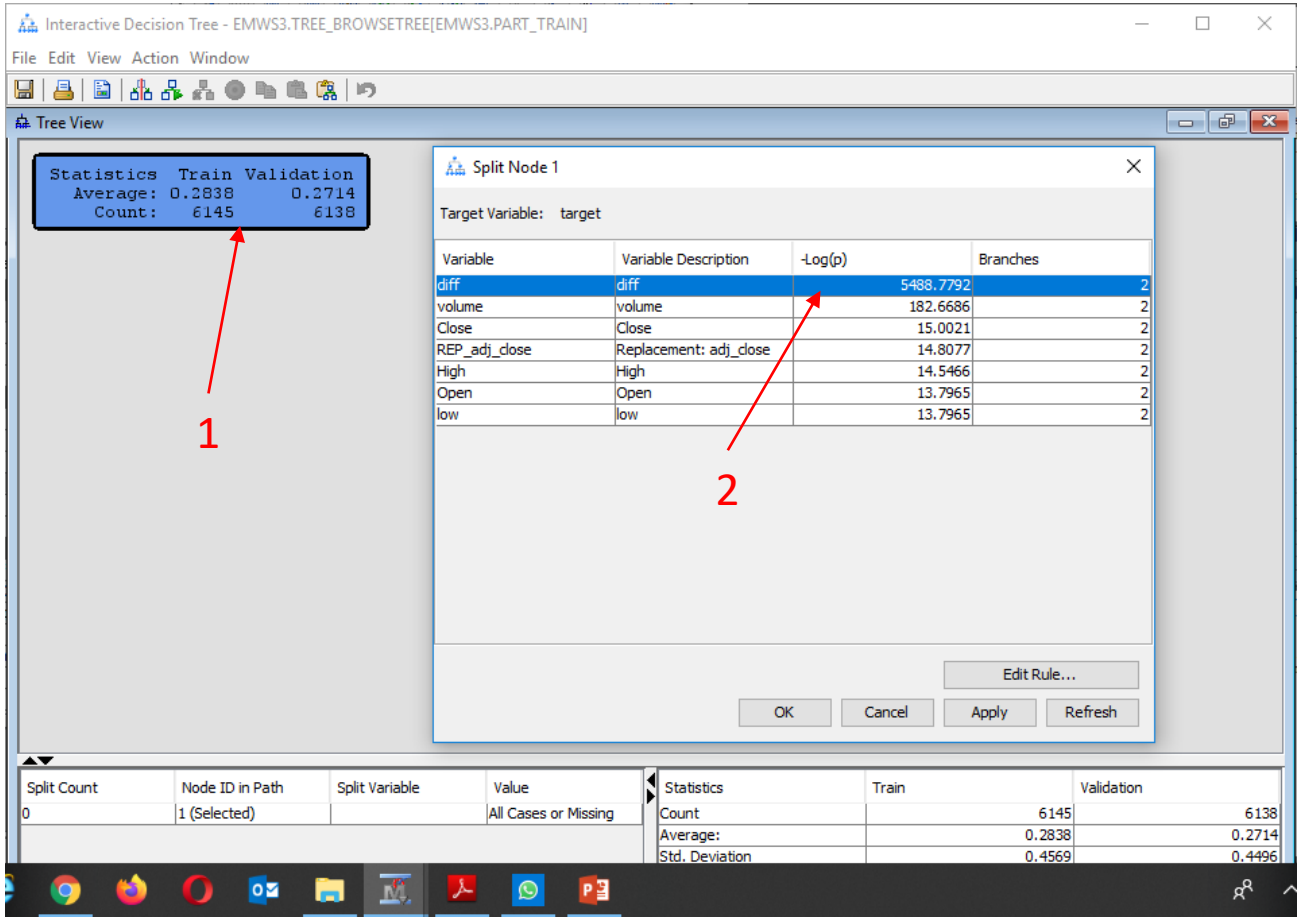
1. Run DATA PARTITION node and view the results
2. Results shows TRAIN and VALIDATE are divide equally 50:50



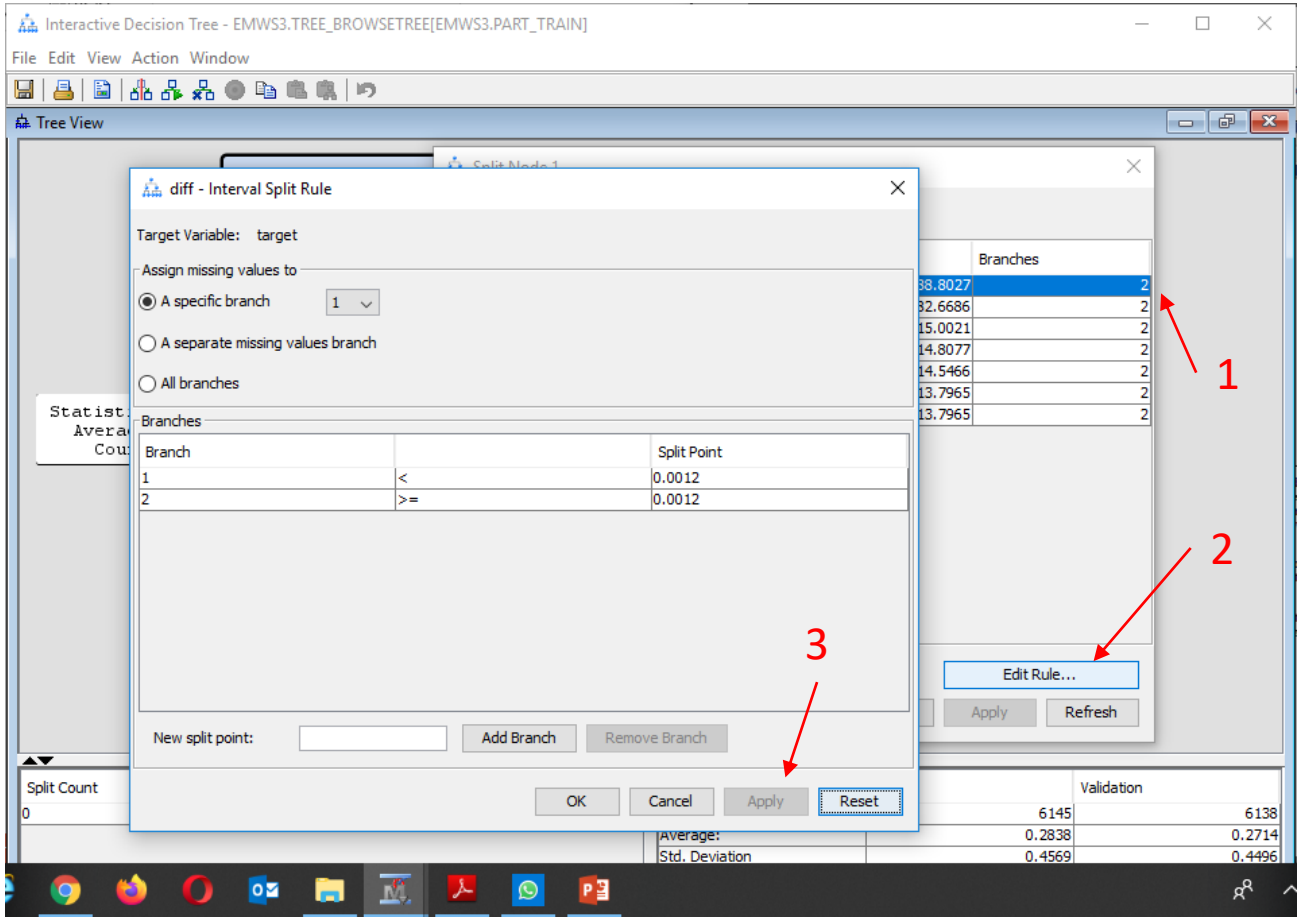
1. Open Model and add DECISION TREE node to diagram
2. Select DECISION TREE node
3. Open Interactive at Train
4. Interactive Decision Tree will shows it properties



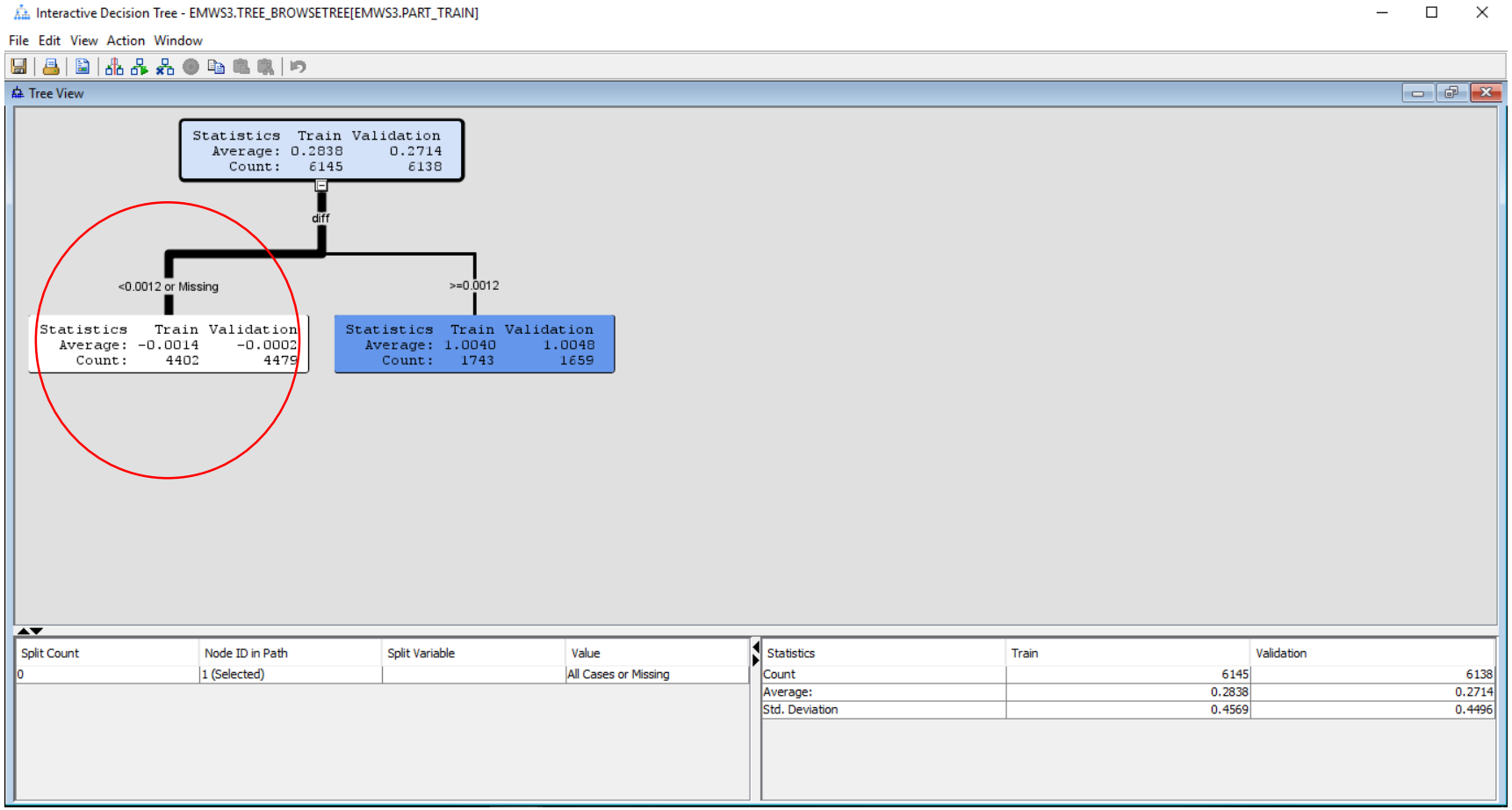
1. Open Split Node 1 by right-click purple box
2. Diff variable shows the highest logworth



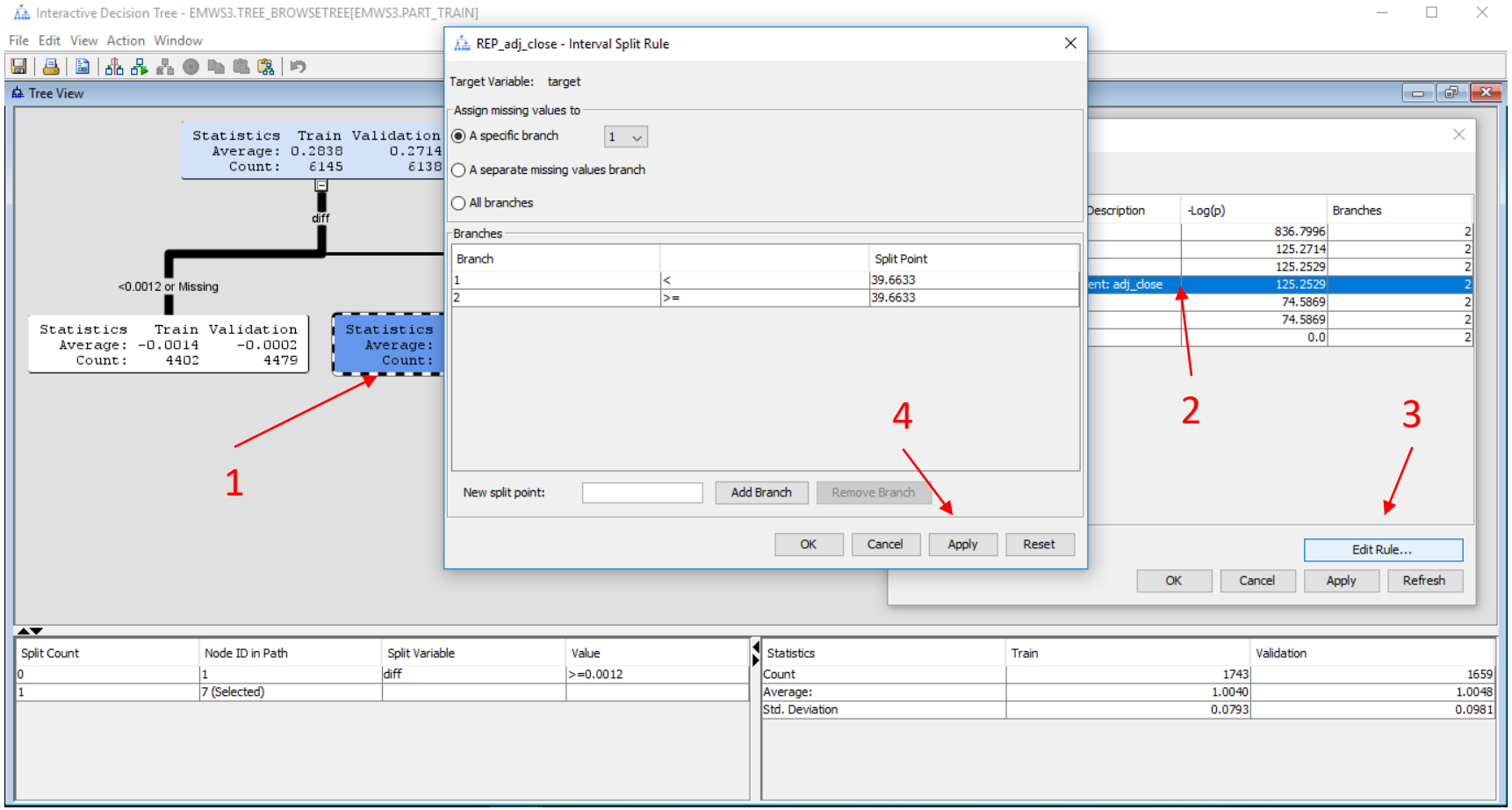
1. Select Input to split
2. Go to Edit Rule
3. Select Apply at Interval Split Rule



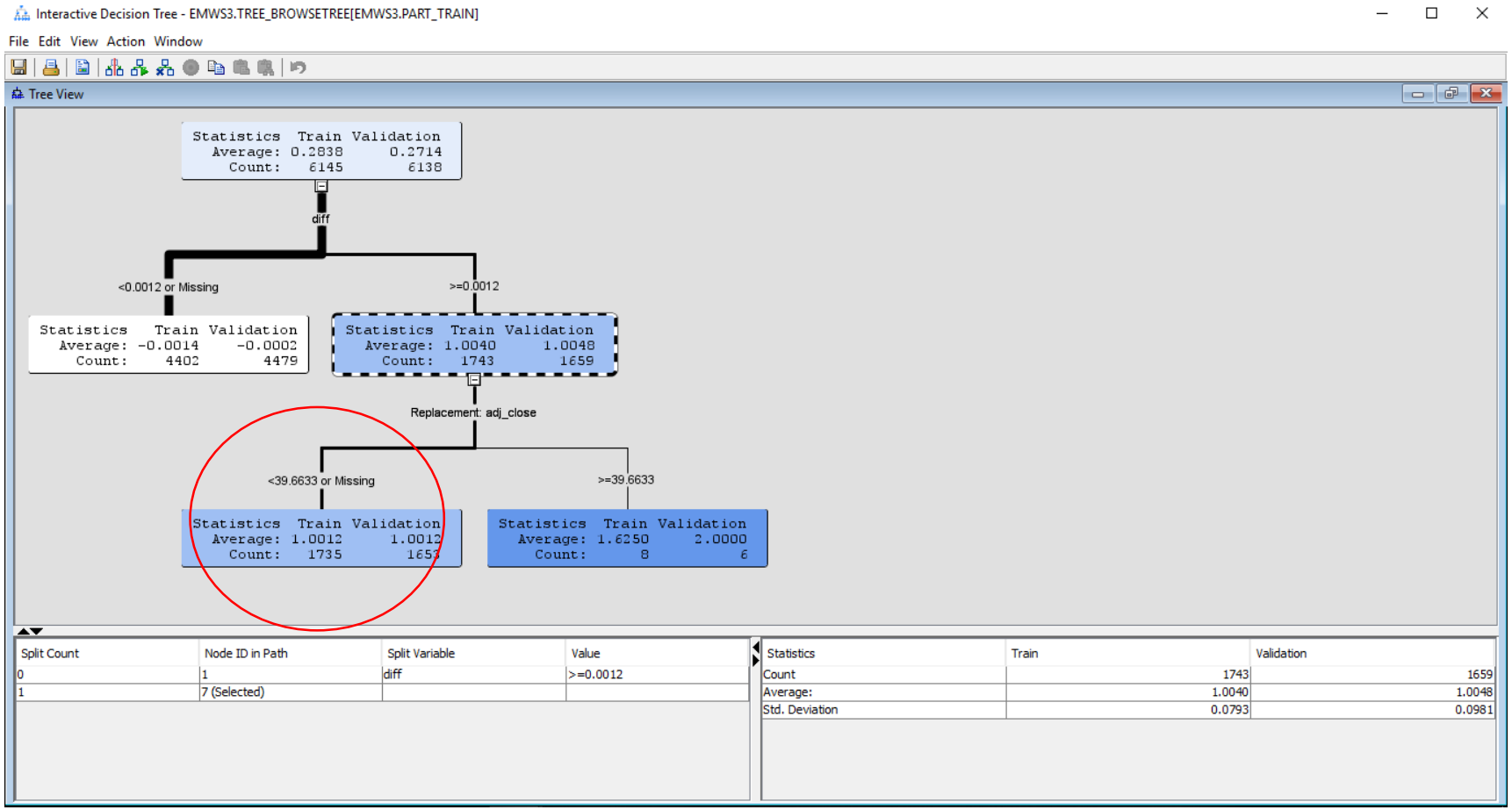
<0.0012 or Missing shows higher concentration



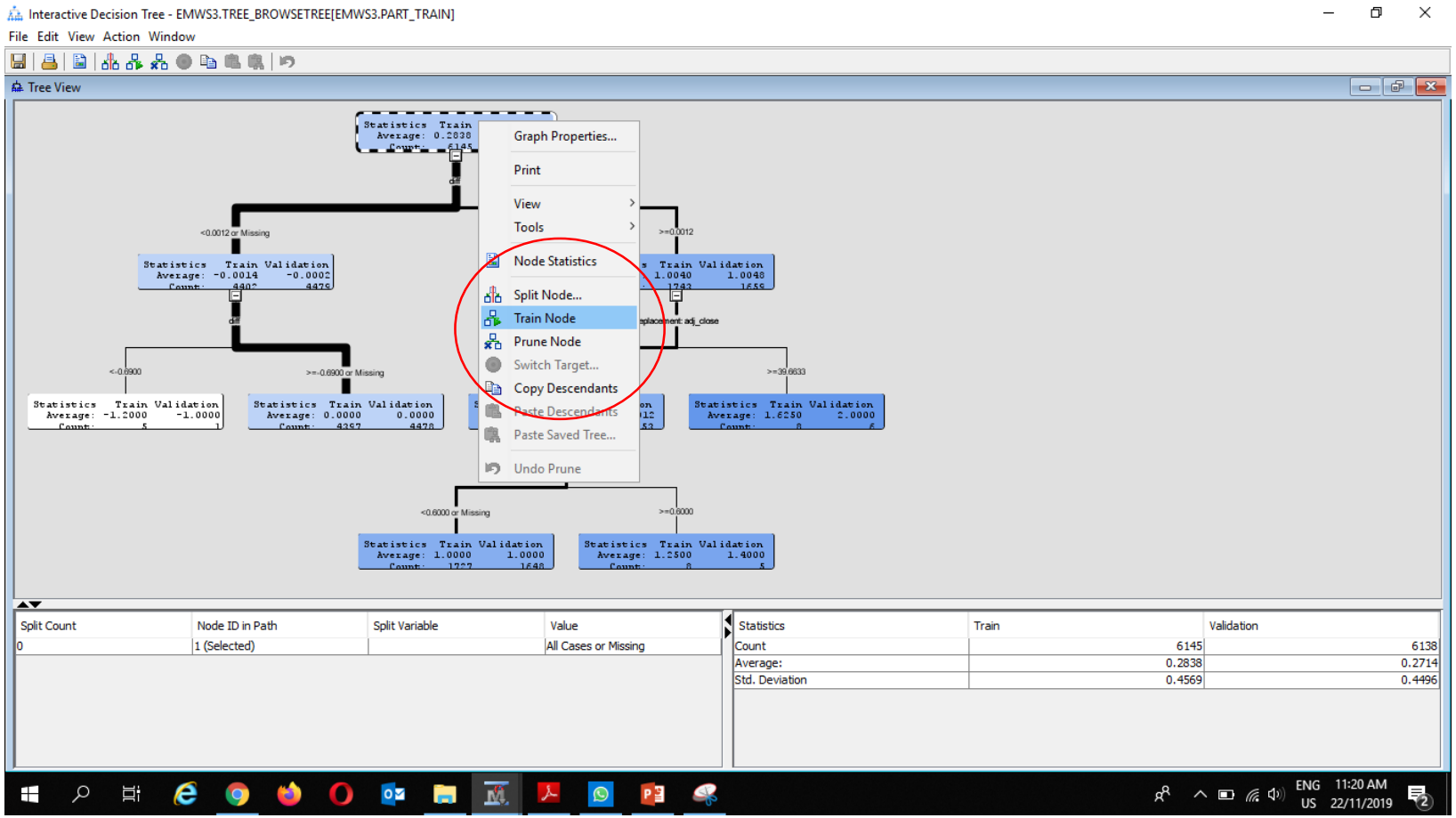
1. Split purple box
2. Select adj_close to Split
3. Go to Edit Rule
4. Apply the Interval Split Rule



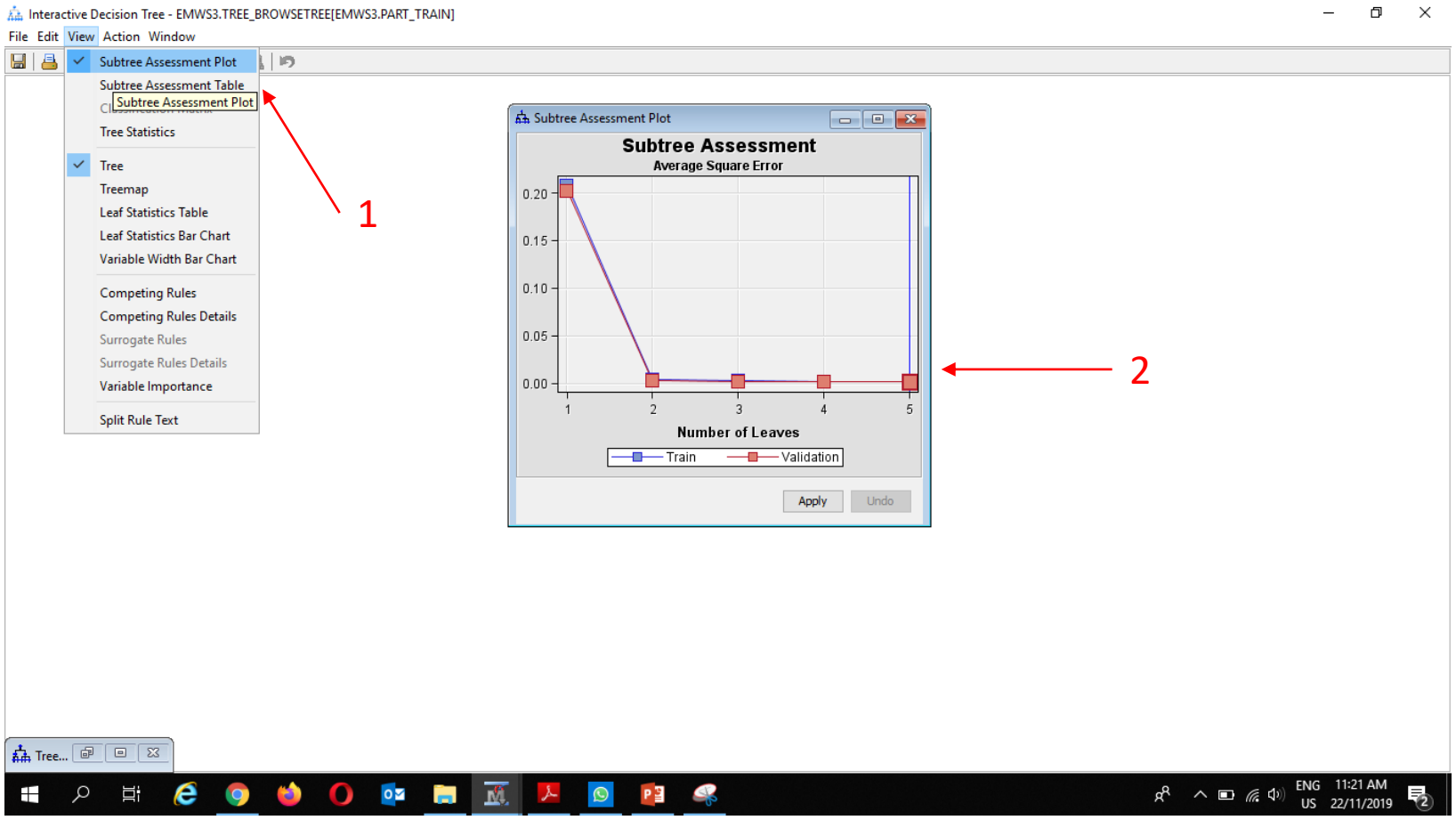
Split results shows …or Missing has higher concentration



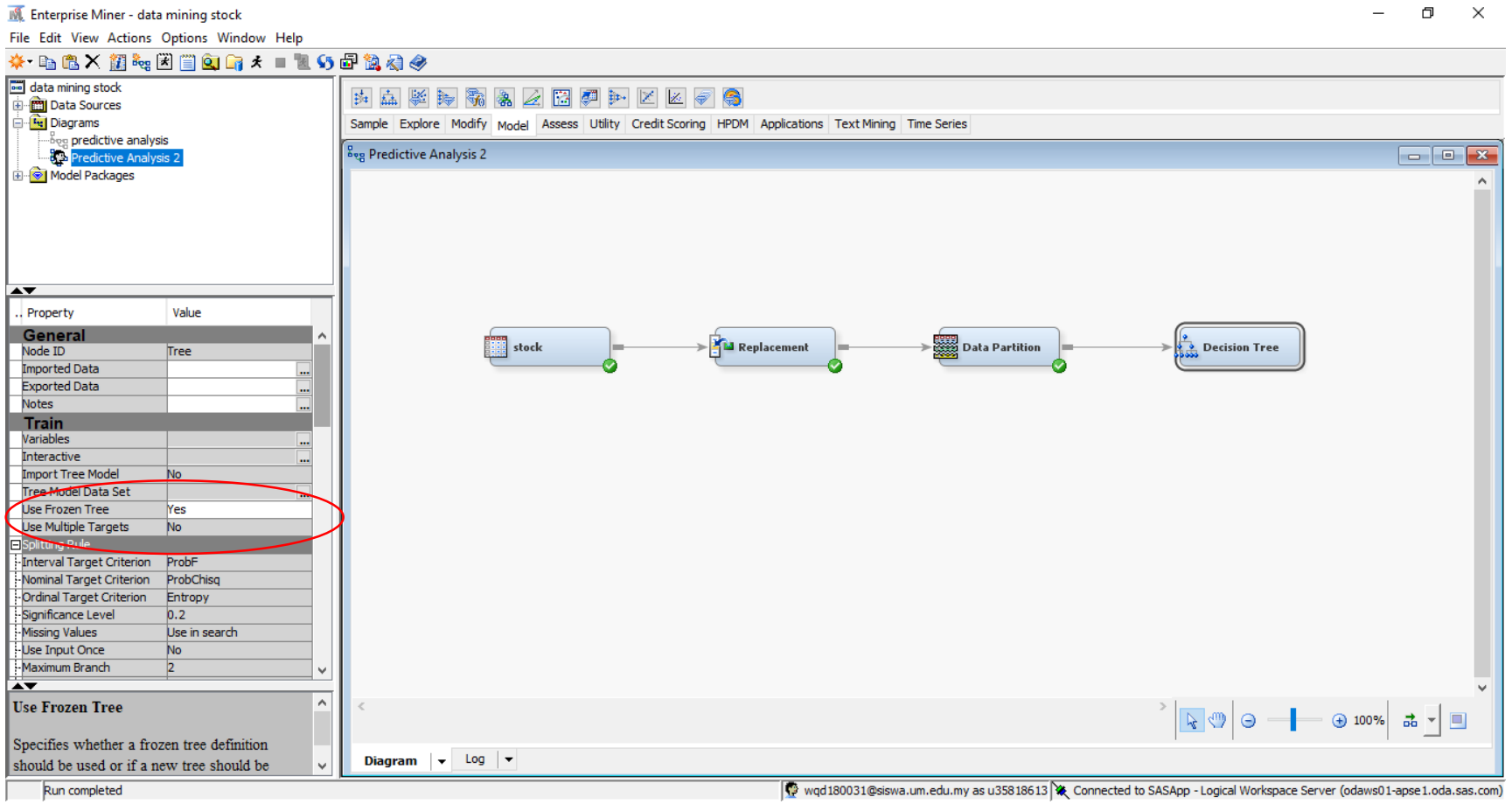
Select Train Node to create the Maximal Tree



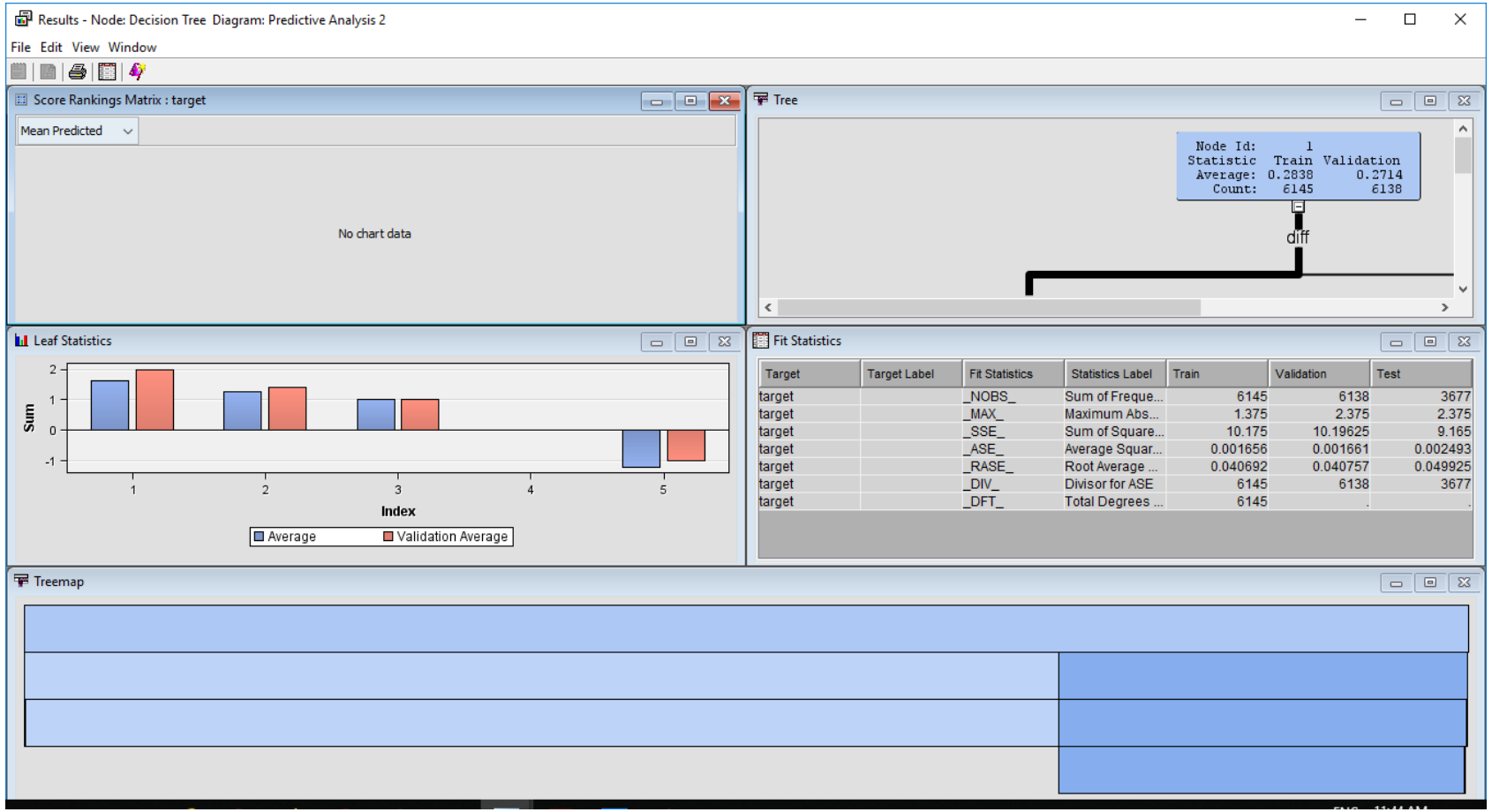
1. Select View and select Subtree Assessment Plot
2. 4 leaf-trees generate a lower misclassification rate



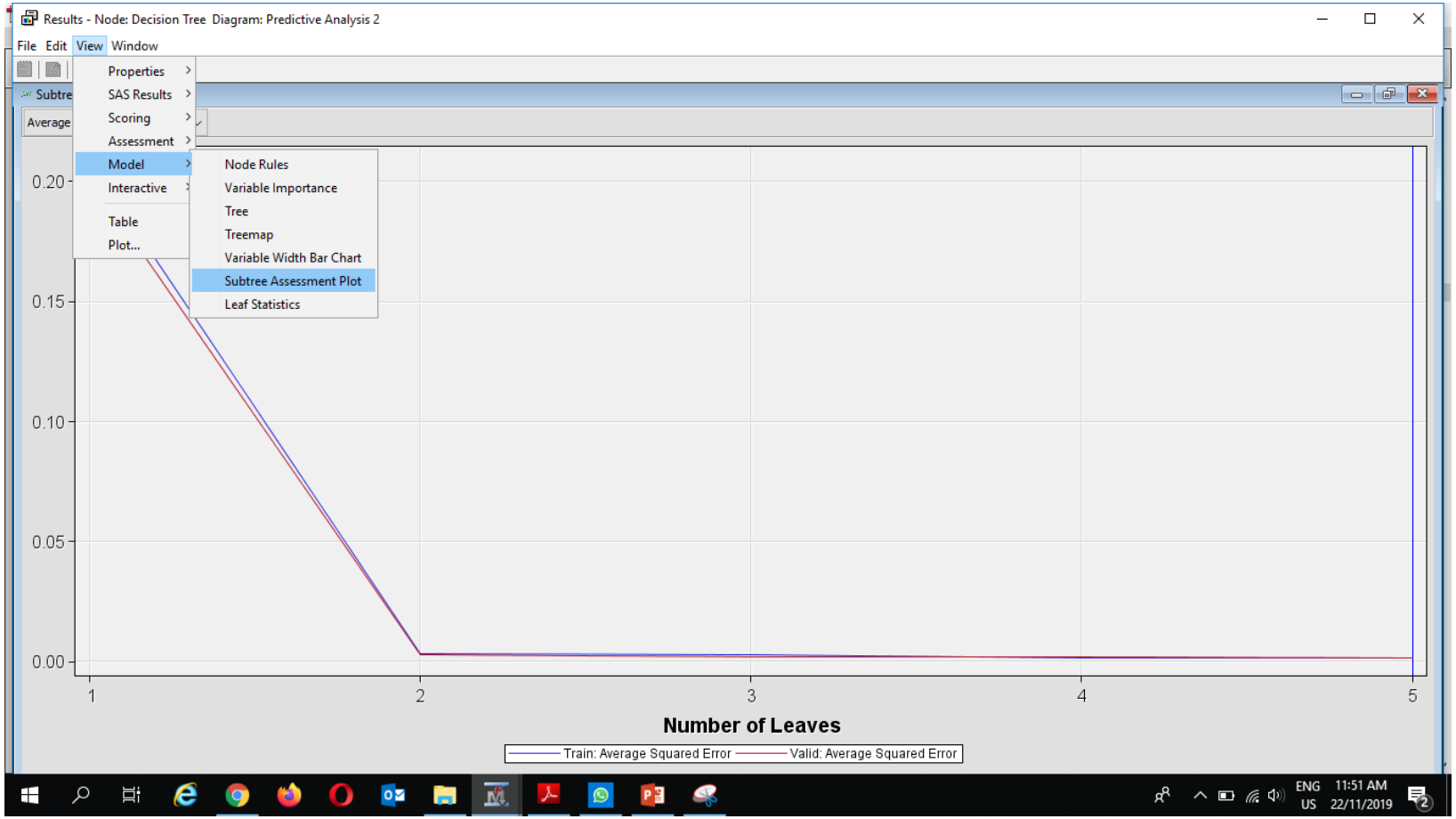
Change Use Frozen Tree in Train from No to Yes in order to assess Decision Tree



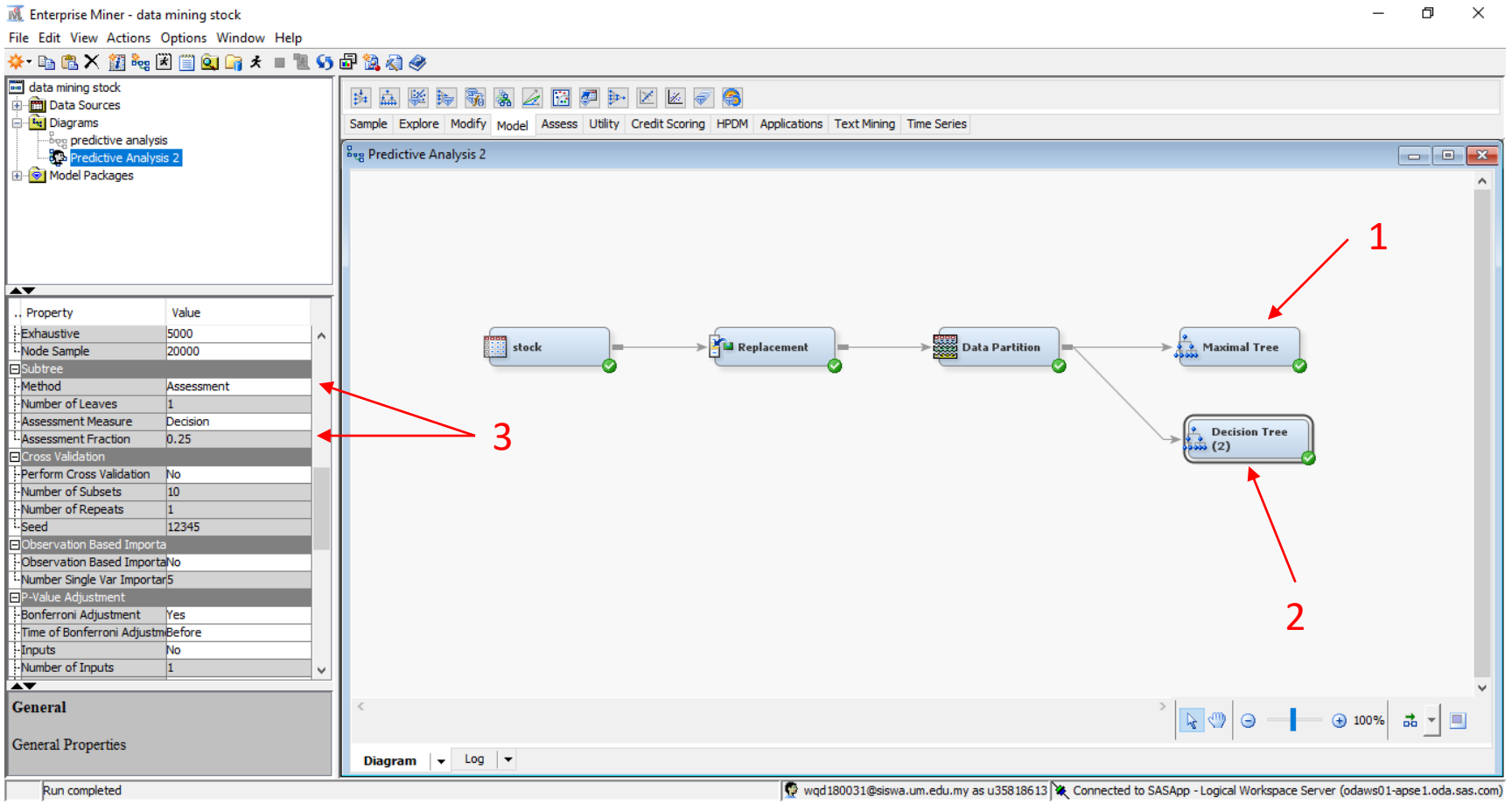
1. Run the DECISION TREE node and view the results
2. Results shows variety of diagnostic plots and tables



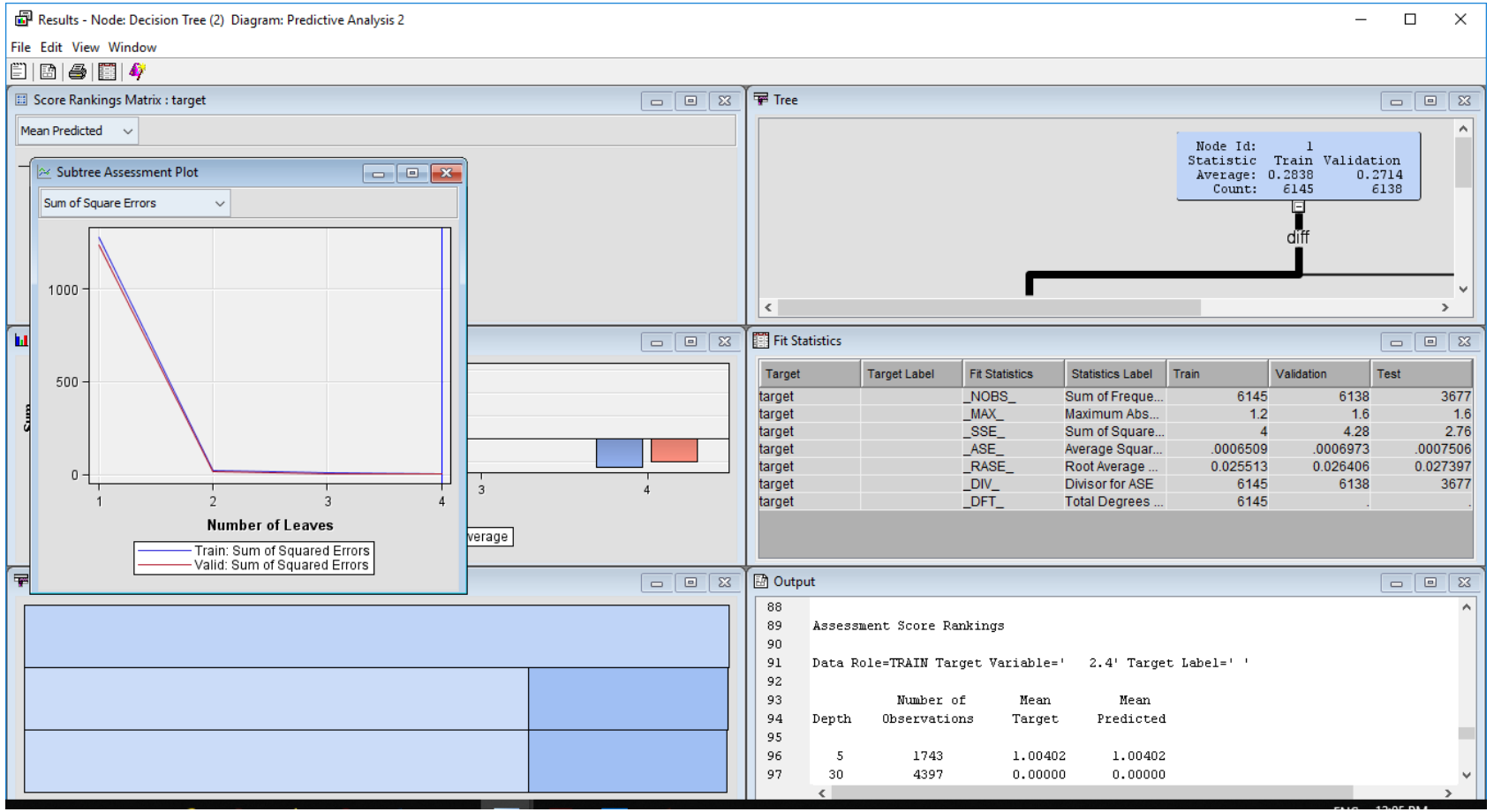
Here is how to assess Assessment Plot based on Average Square Error



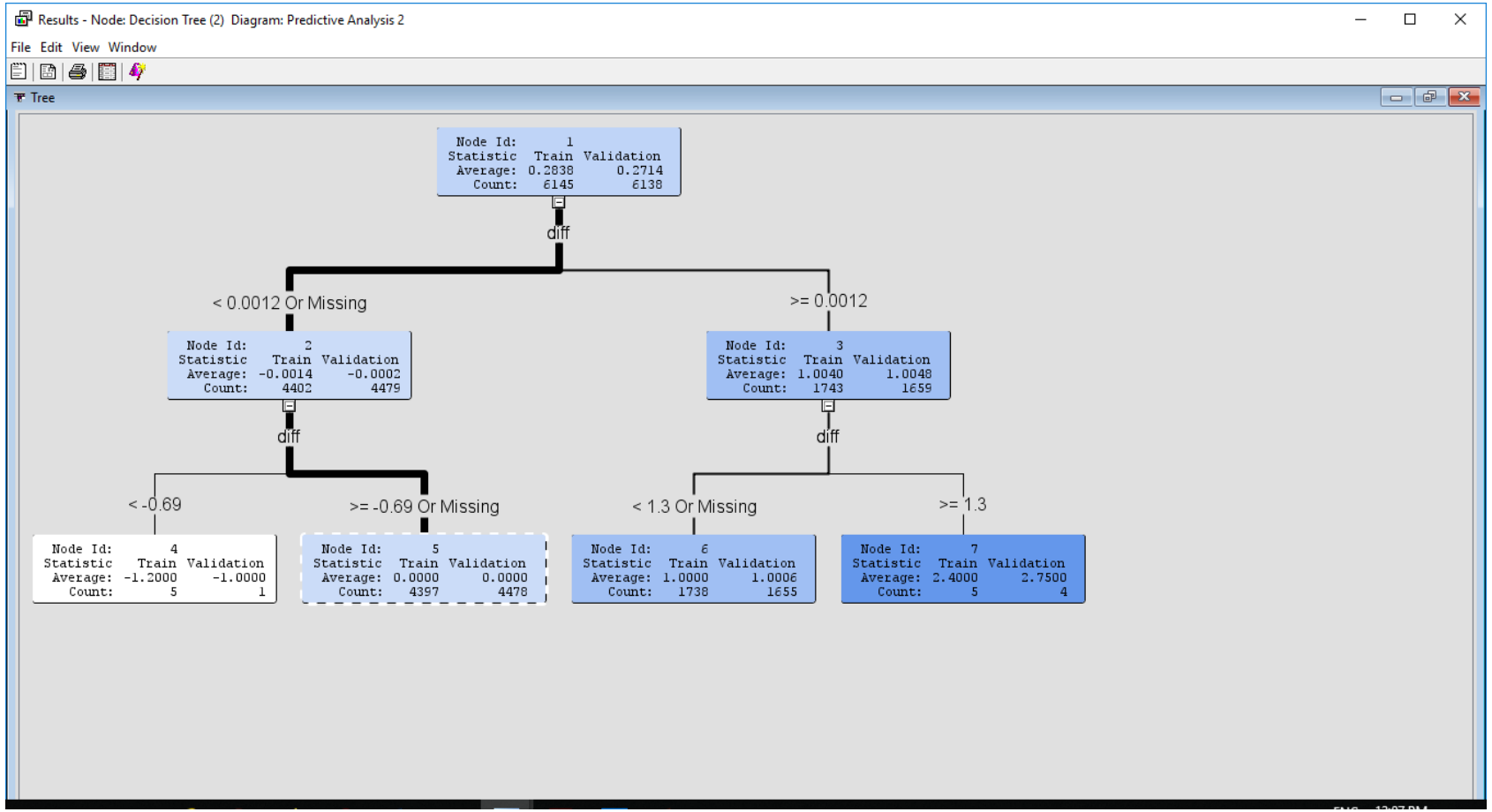
1. Back to diagram and rename DECISION TREE node to MAXIMAL TREE
2. Add another DECISION TREE node
3. Change DECISION TREE node properties under Subtree; Method to Assessment and Assessment Measure to Decision



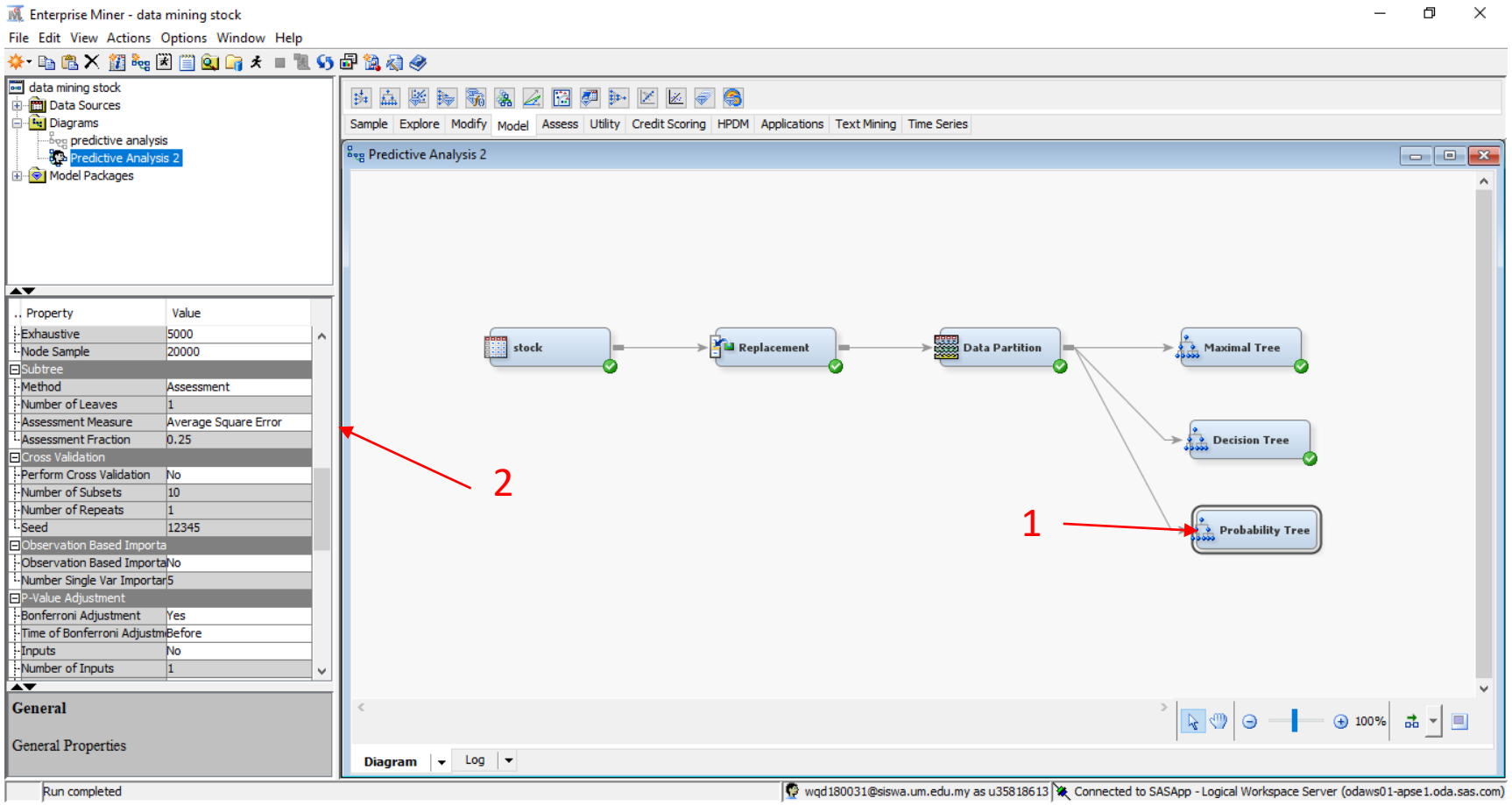
Results


Results - Node: Decision Tree (2) Diagram: Predictive Analysis 2
File Edit View Window
Score Rankings Matrix : target
Mean Predicted
Subtree Assessment Plot
Sum of Square Errors
1000
500
0
1
2
3
4
Number of Leaves
Train: Sum of Squared Errors
Valid: Sum of Squared Errors
Tree
Node Id: 1
Statistic Train Validation
Average: 0.2838 0.2714
Count: 6145 6138
diff
Fit Statistics
Target | Target Label | Fit Statistics | Statistics Label | Train | Validation | Test
target _NOBS_ Sum of Freque... 6145 6138 3677
target _MAX_ Maximum Abs... 1.2 1.6 1.6
target _SSE_ Sum of Square... 4 4.28 2.76
target _ASE_ Average Squar... .0006509 .0006973 .0007506
target _RASE_ Root Average ... 0.025513 0.026406 0.027397
target _DIV_ Divisor for ASE 6145 6138 3677
target _DFT_ Total Degrees ... 6145
Output
Assessment Score Rankings
Data Role=TRAIN Target Variable=' 2.4' Target Label=' '
Depth Number of Observations Mean Target Mean Predicted
5 1743 1.00402 1.00402
30 4397 0.00000 0.00000

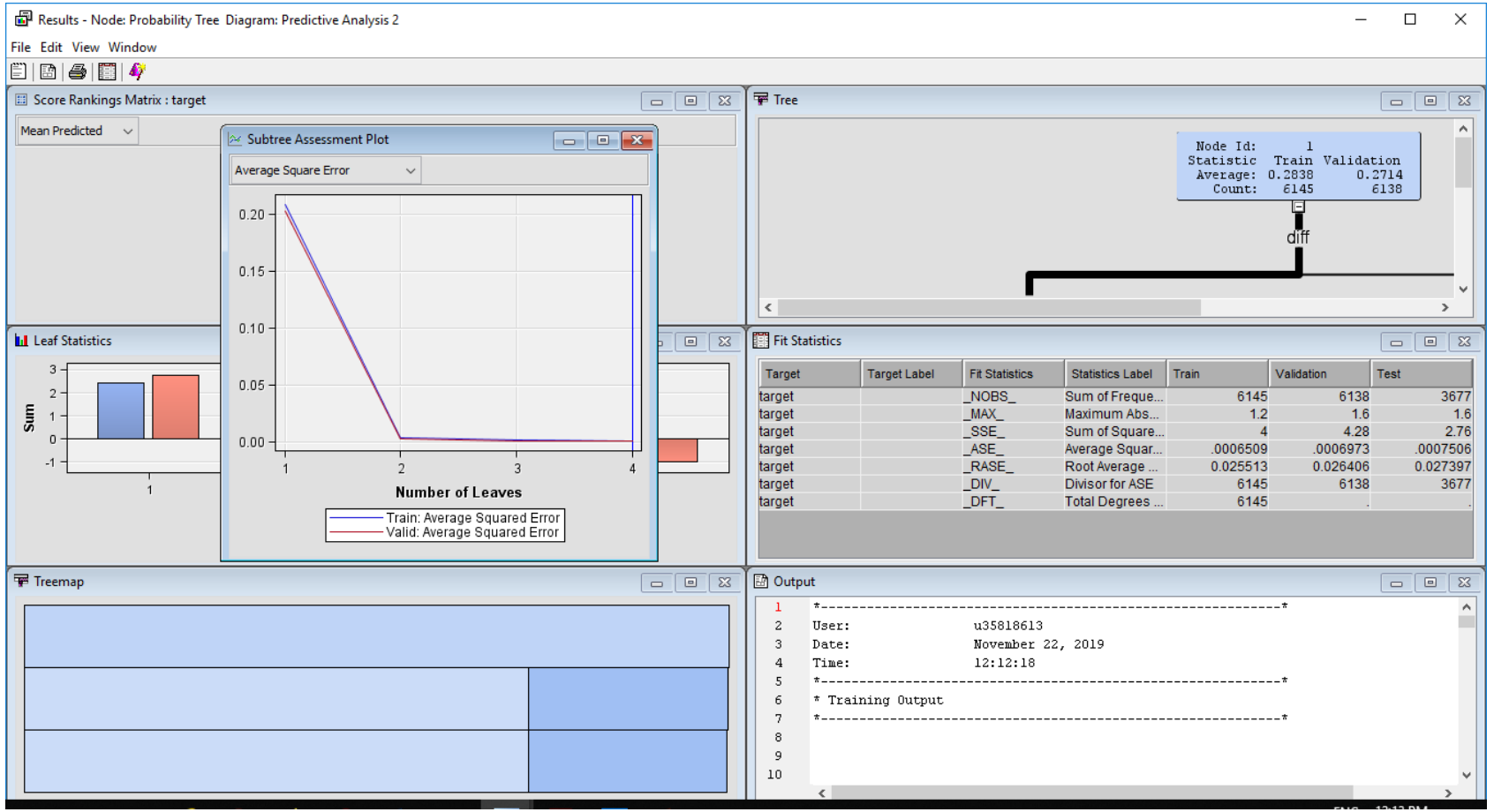
Optimal tree generated by DECISION TREE node



1. Create PROBABILITY TREE node by adding and renamed new DECISION TREE node
2. Change Assessment Measure in Subtree to Average Square Error



Run the PROBABILITY TREE node and view the results



Results of Probability Tree