1. **Ture or False.Decision Trees seek to form leaf nodes to maximize heterogeneity in each node.**

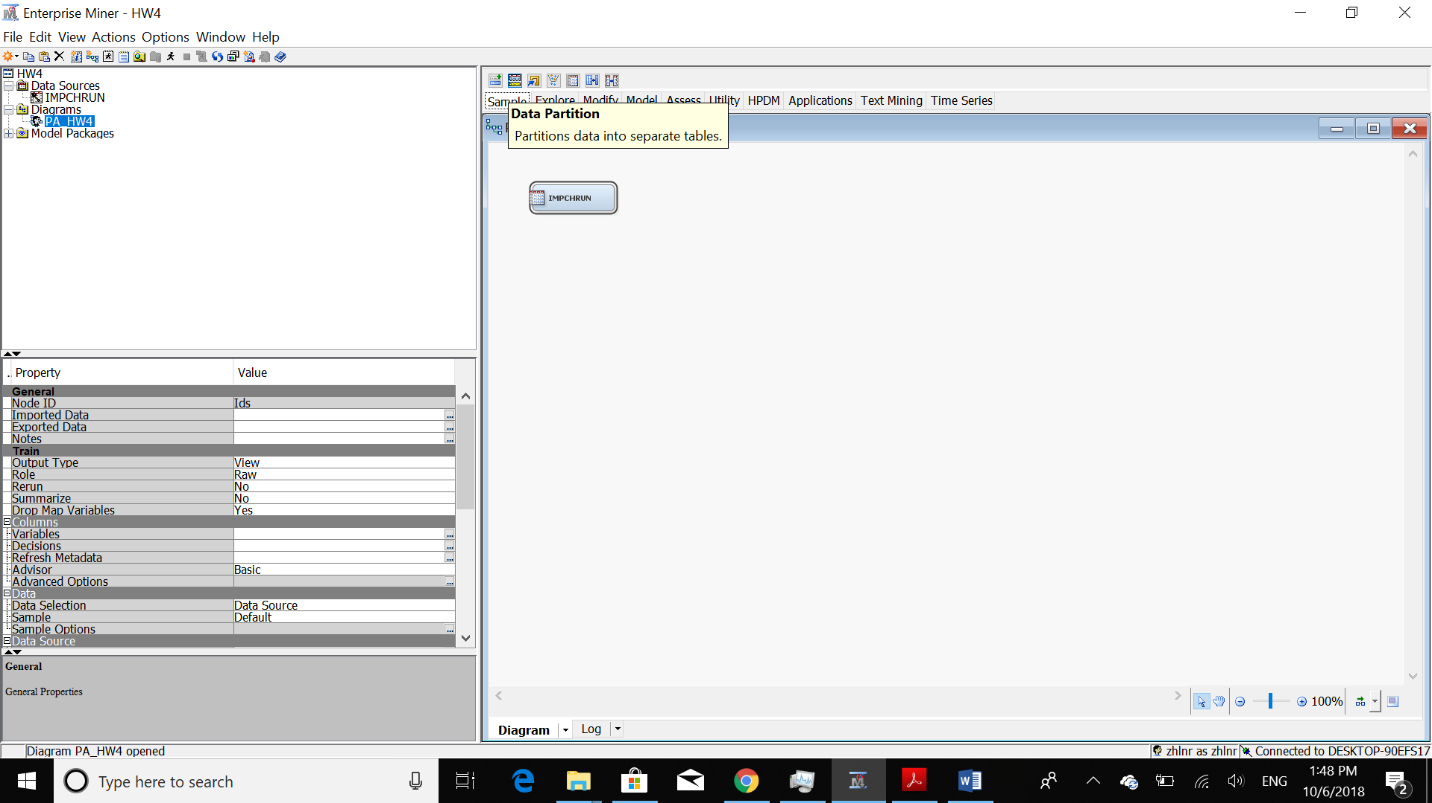
This statement is FALSE. As we have on page 167 of the textbook, Decision Trees seek to create a set of leaf nodes that are as “PURE” as possible. That is, where each of the records in a particular leaf node has the same classification. In this way, the decision tree may provide classification assignments with the highest measure of confidence available. To measure uniformity or conversely heterogeneity, we shall examine two of the many methods for measuring leaf node purity, which leads to the two leading algorithms for constructing decision trees:

1. Classification and regression trees (CART) algorithm
2. C4.5 Algorithm

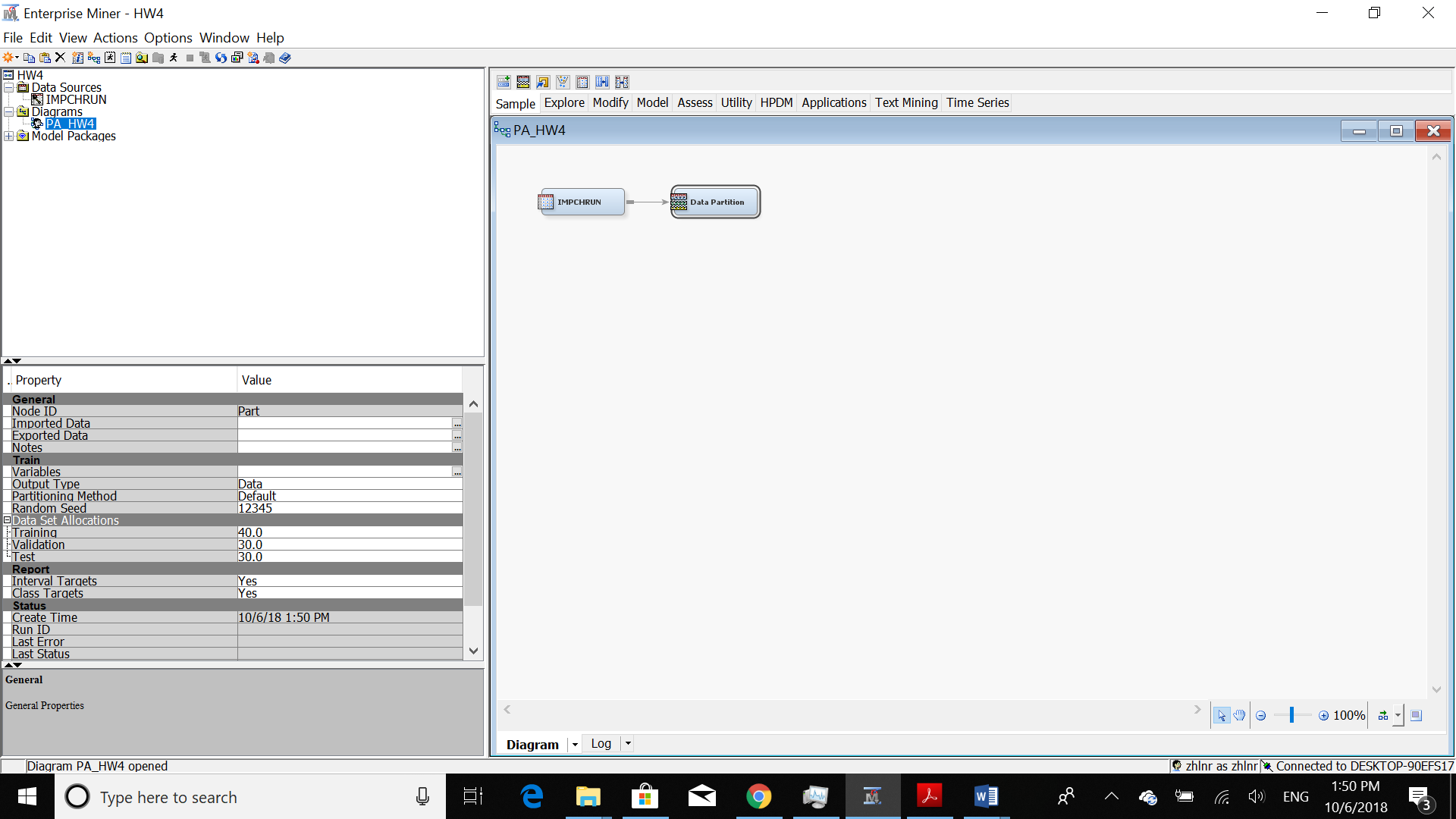
By using these two algorithms, we follow two different approaches to maximize the value calculated by these methods. In another word, we want to make the leaf nodes as pure as possible. Because we want to categorize the same values in the same nodes. So, maximizing the heterogeneity in each node contradicts with our goal. However, as we may have seen previously, not all leaf nodes are necessarily homogenous, which leads to a certain level of classification error.

1. **Generate the best decision tree based on the assessment measure of misclassification rate.**

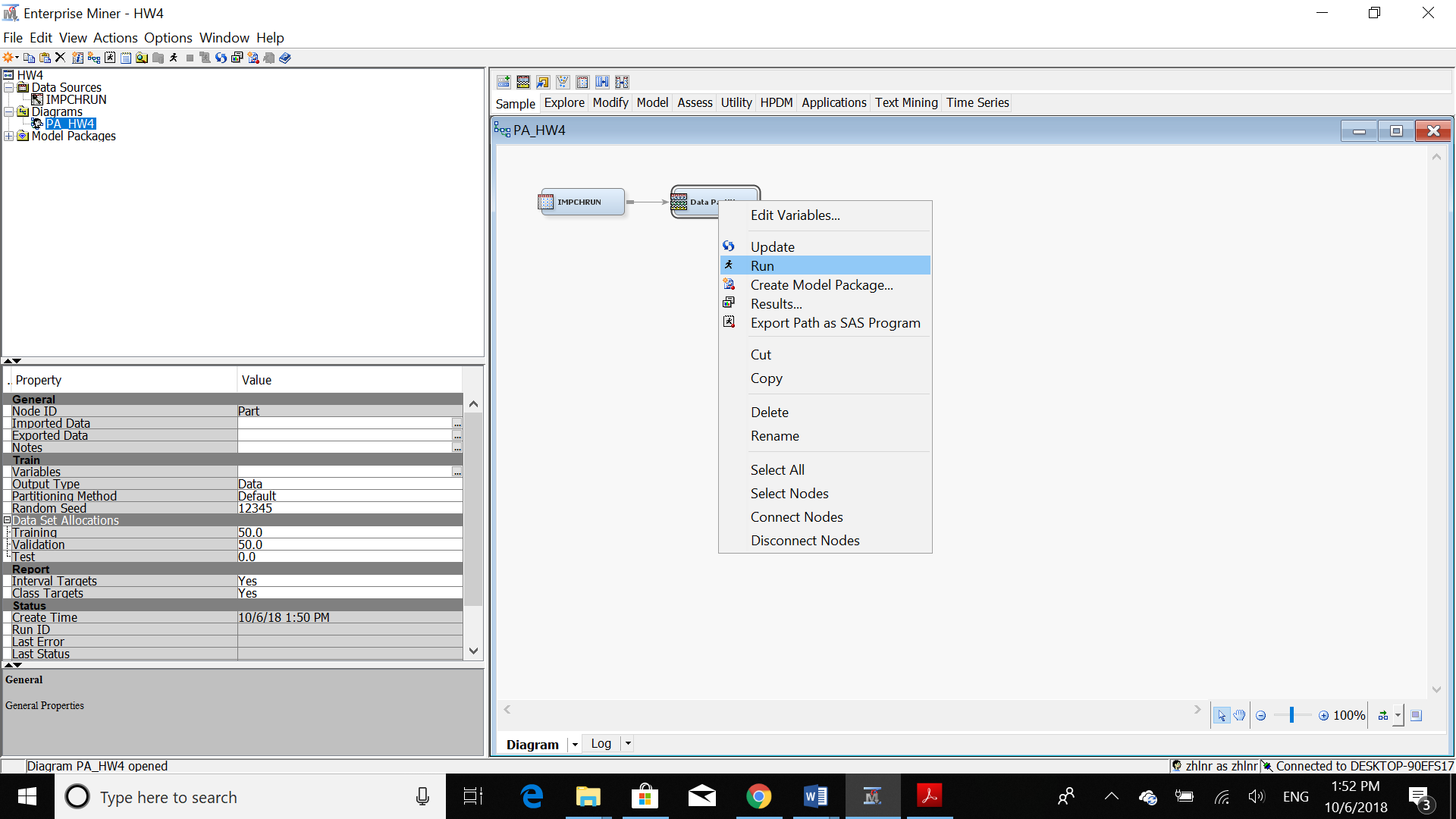
I followed the tutorial for answering this question. The first step in the tutorial after creating the data set was replacing the missing values. I did not add a replacement node. Because, as it was showed in previous homework, there is not any missing value for churn data set. So, because I know the data set as I’ve worked with it for answering previous home works, I went to the next step which is creating our decision tree. We can see these steps as follows:

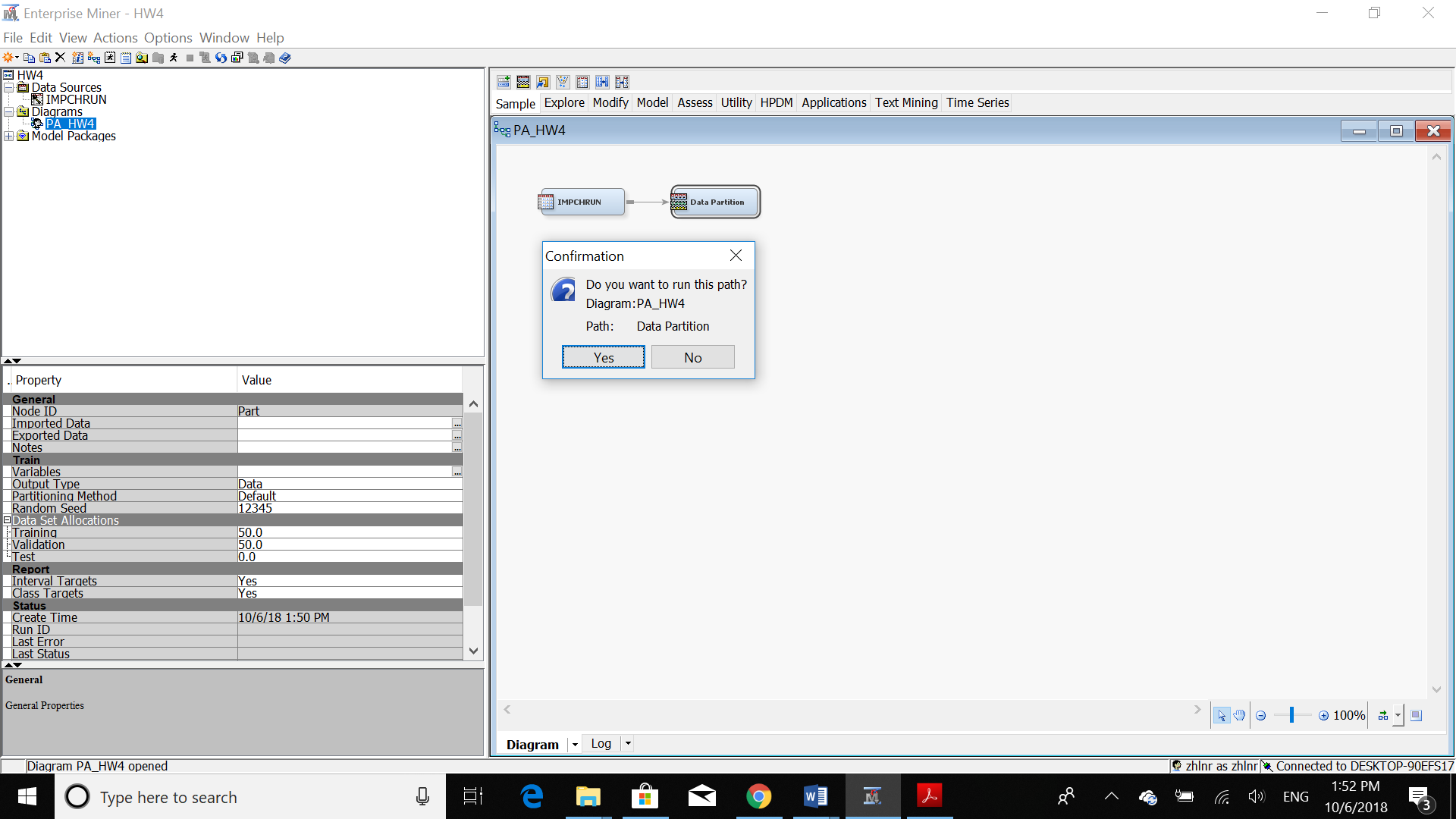


Next, we should add Data Partition node:

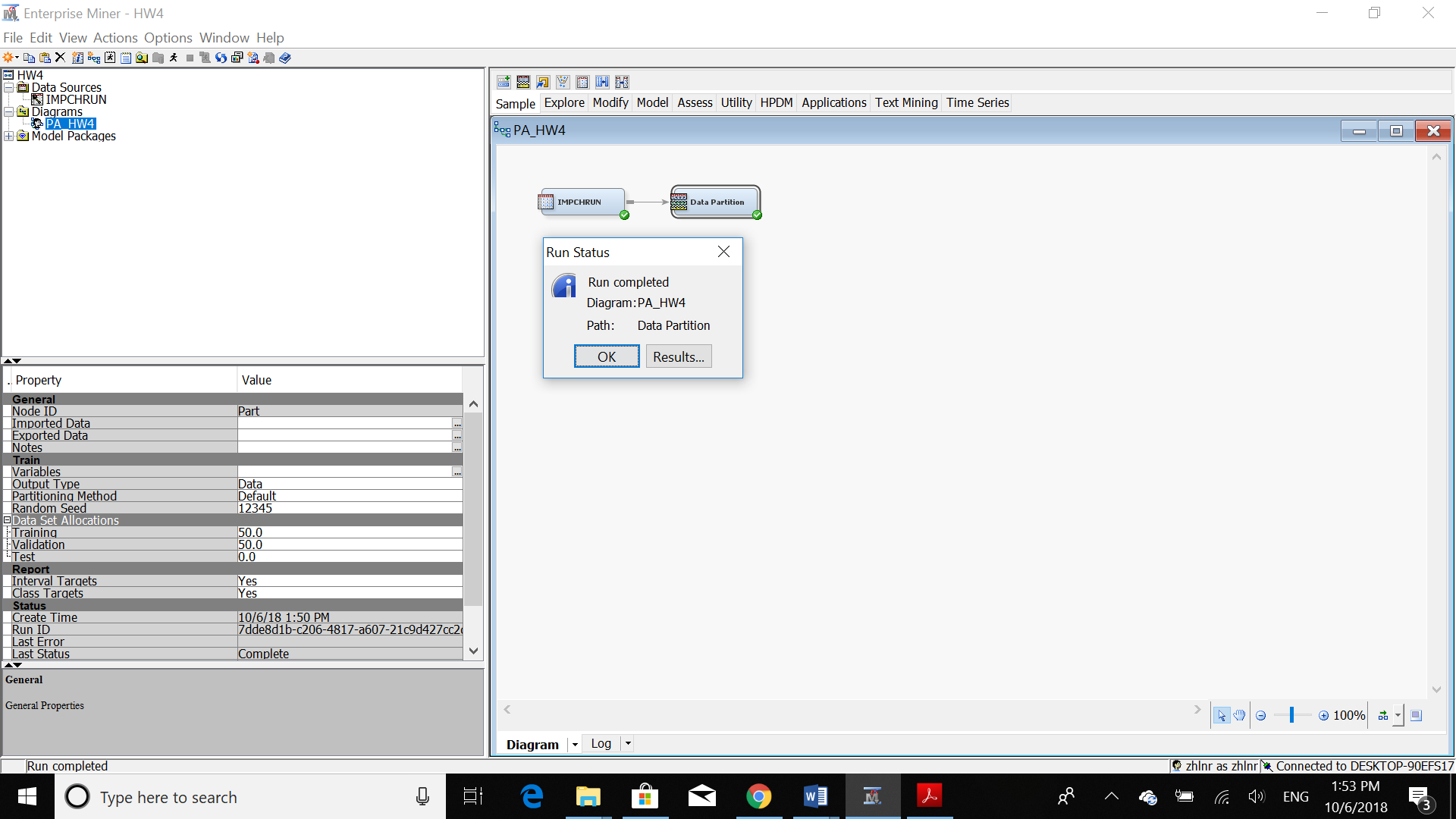


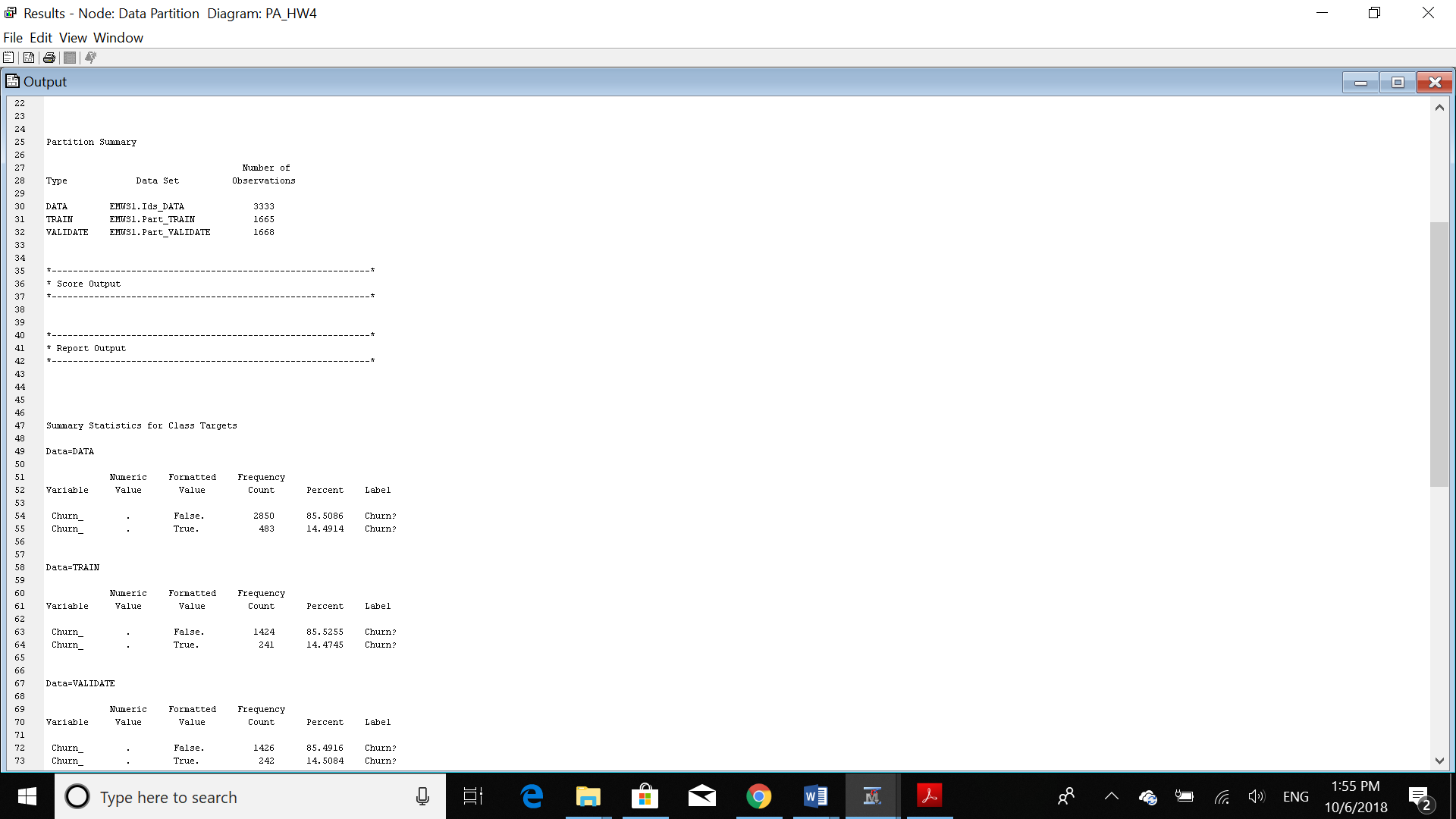
After adding, I run each node.



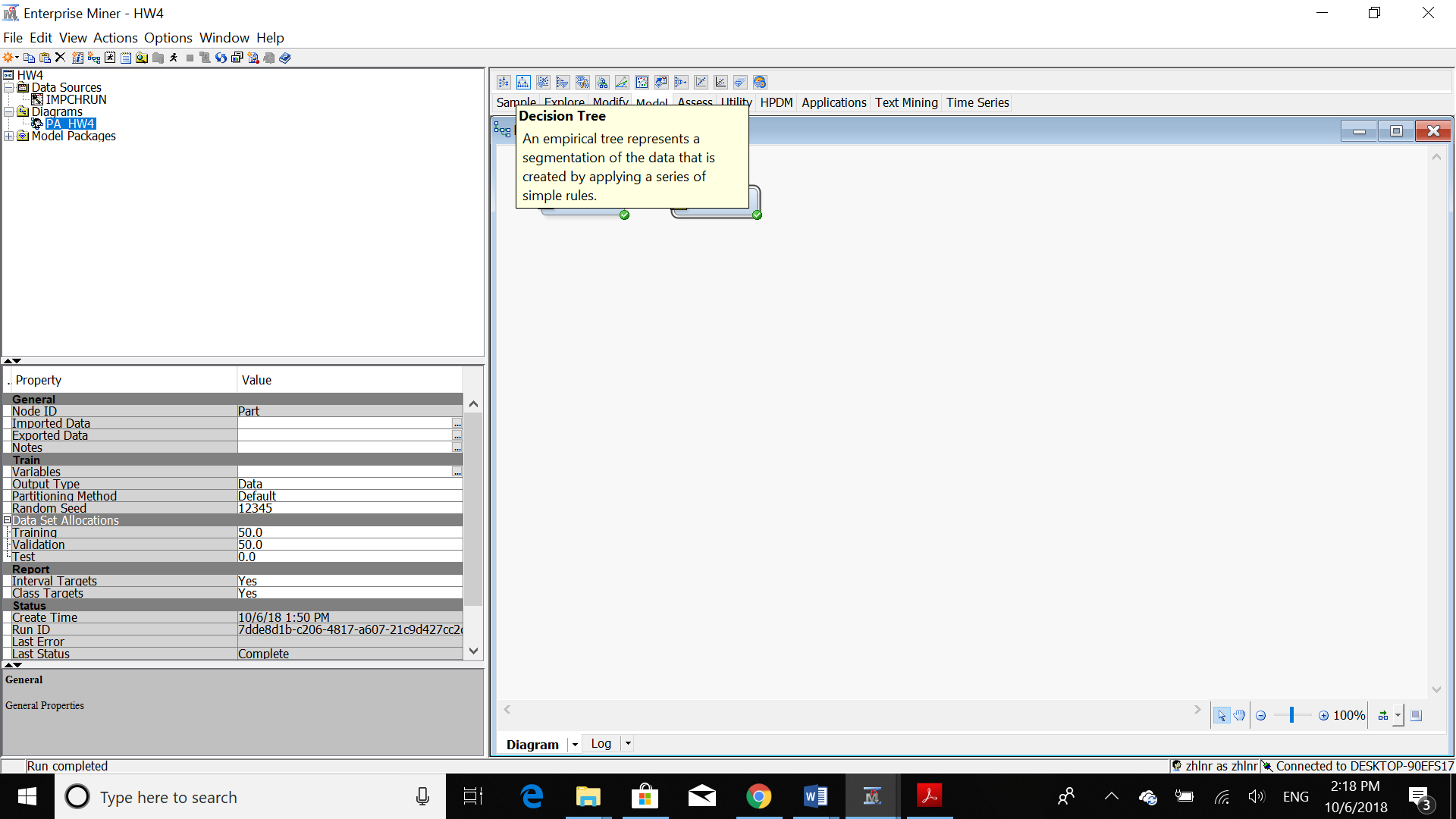


Now, we should choose” Results” in order to see it:

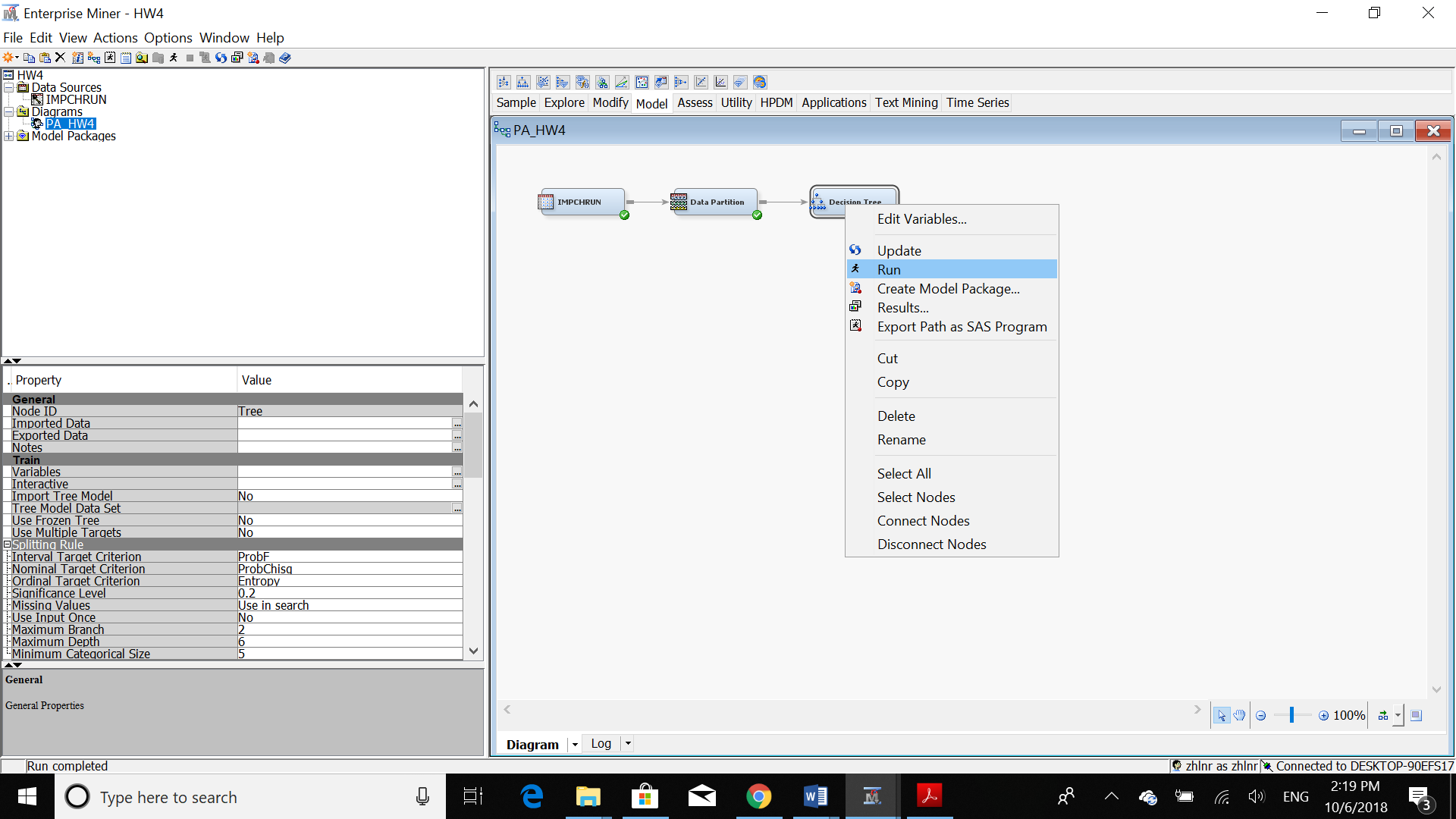




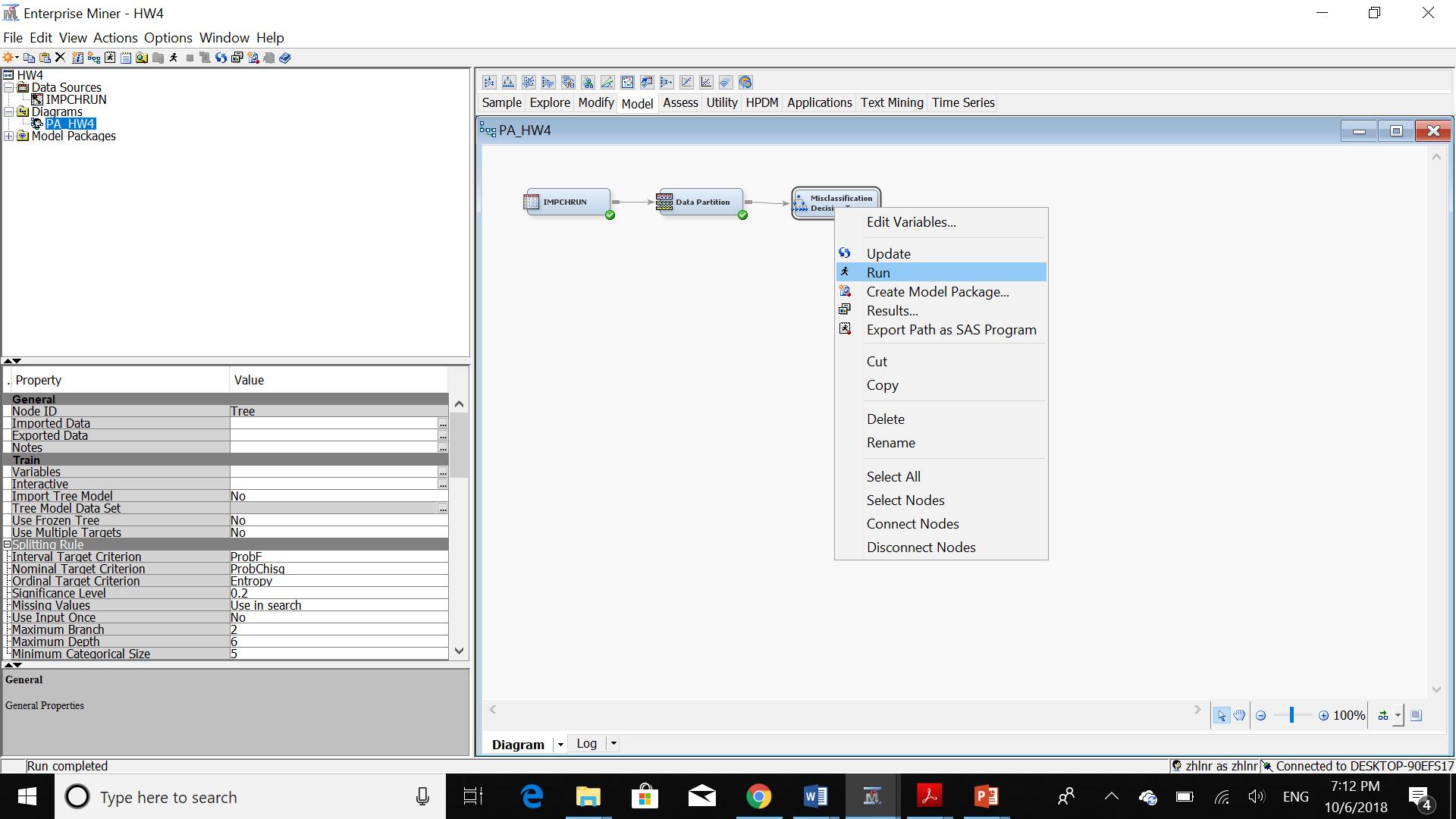
In this step, we should add “Decision Tree” node to our work station:

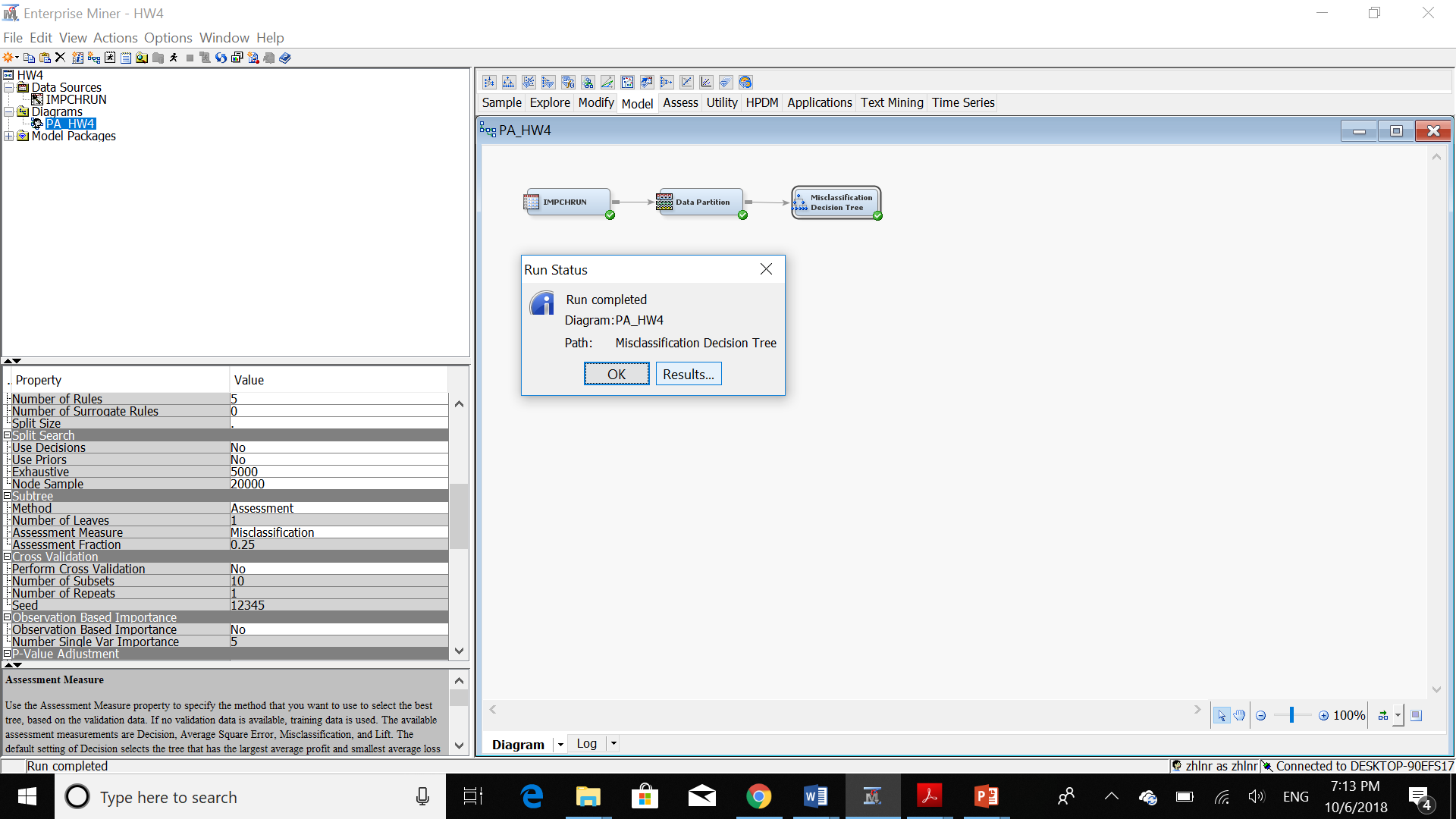


Then, we should run our decision tree in order to see the result:

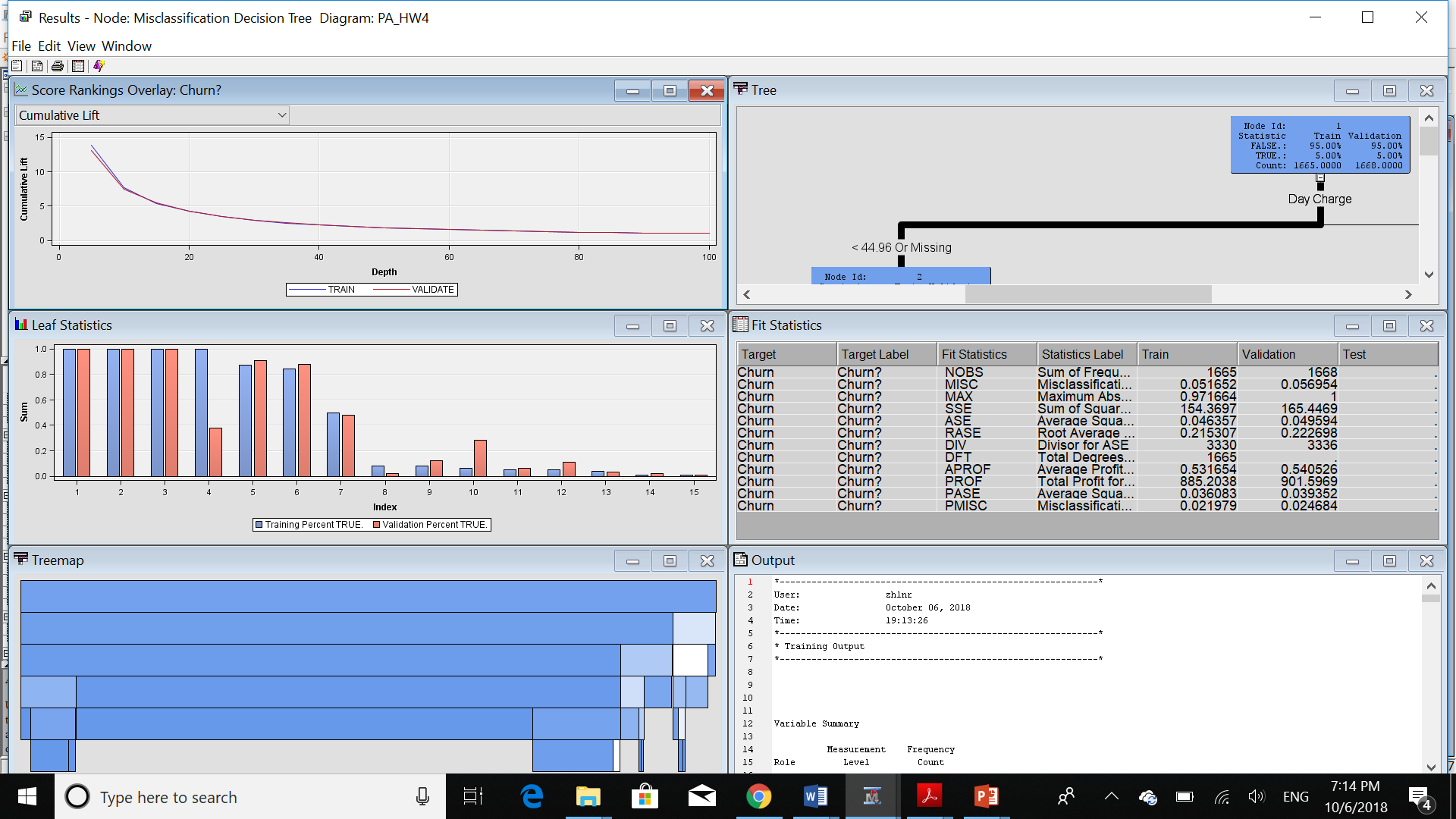


I renamed this node to Misclassification based Decision Tree:

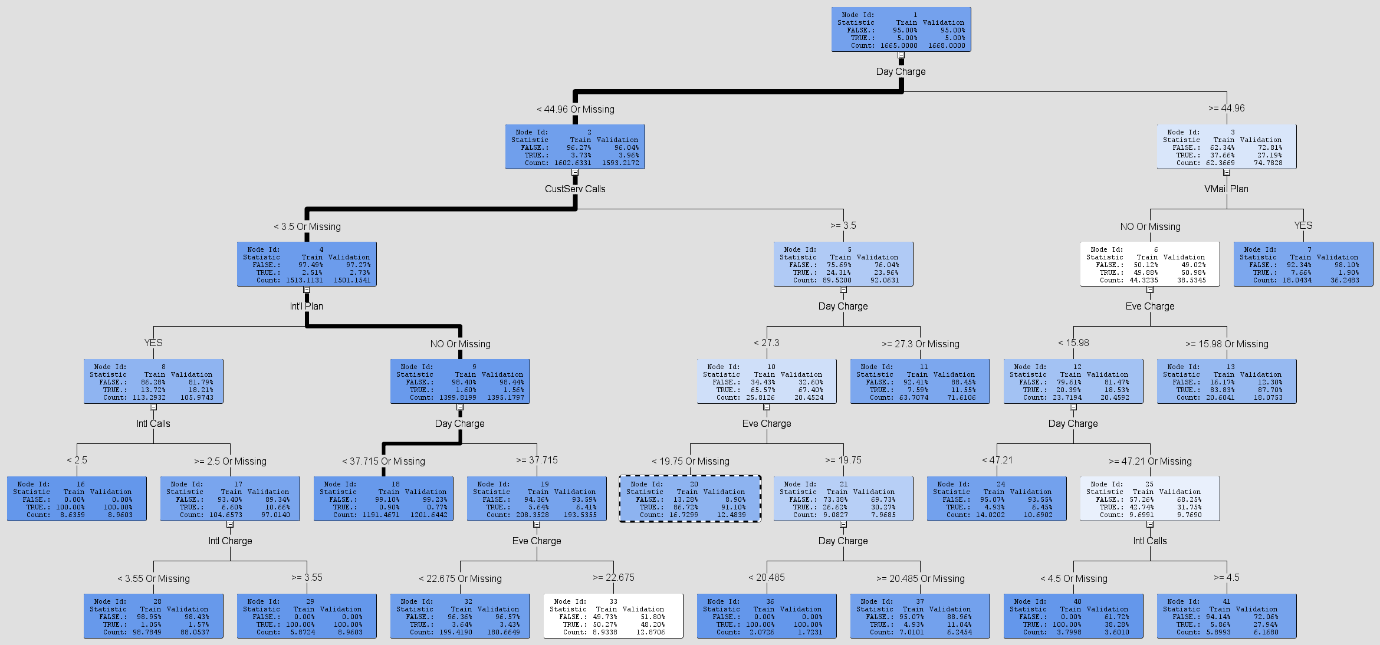




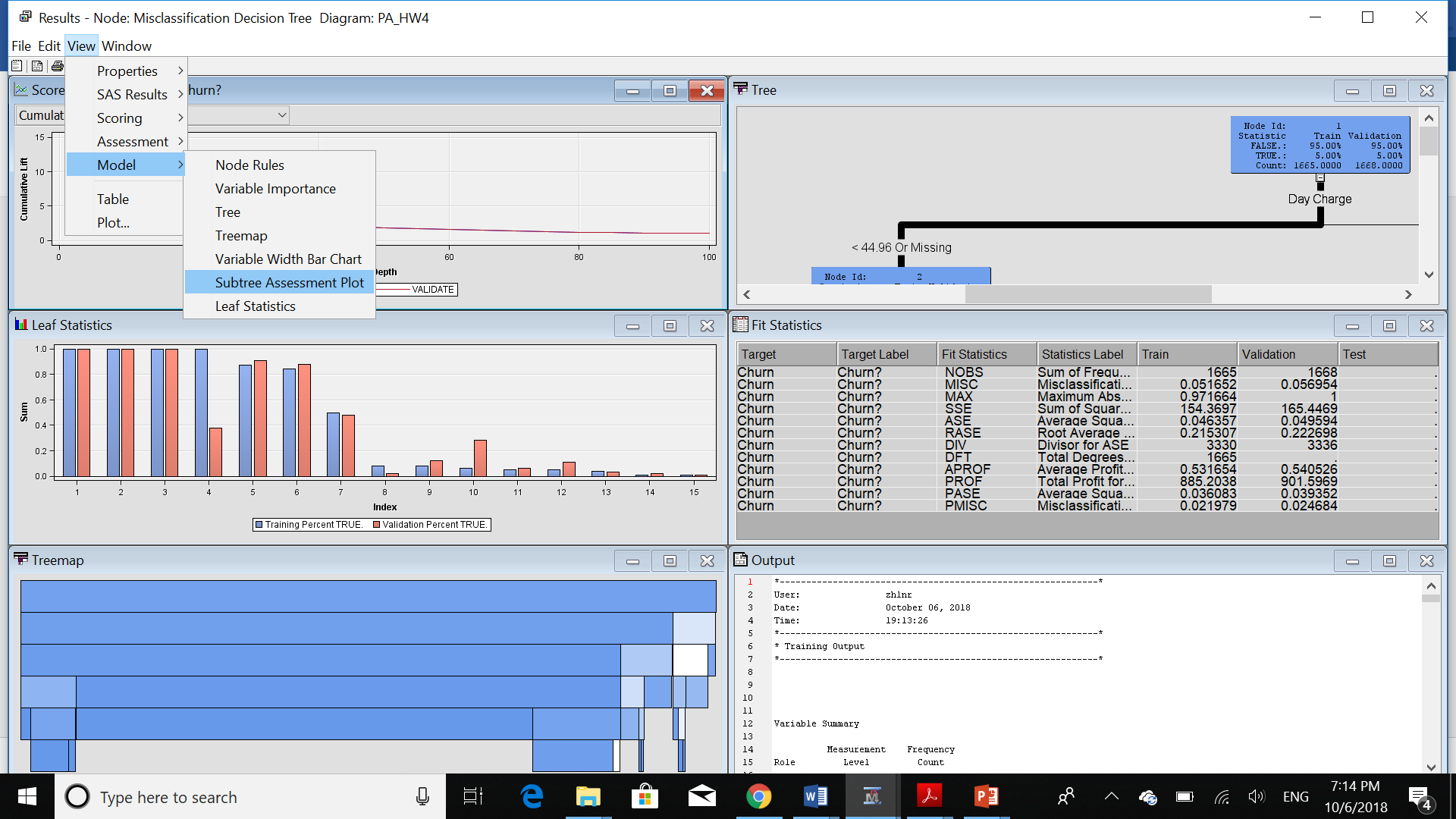
Here, we can see the result:



The Decision Tree for this question is:



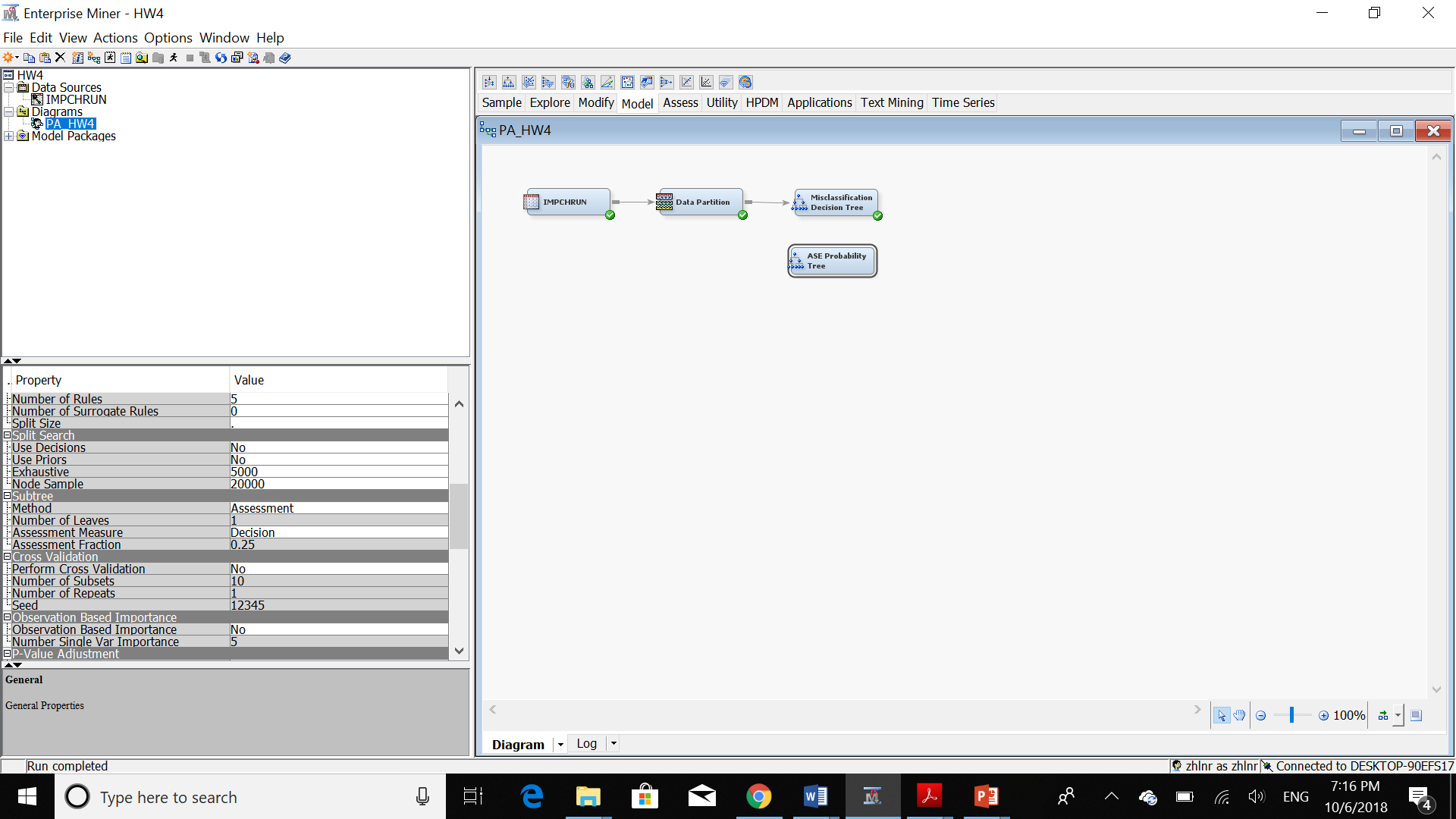
Now, let’s see the Subtree Assessment Plot:

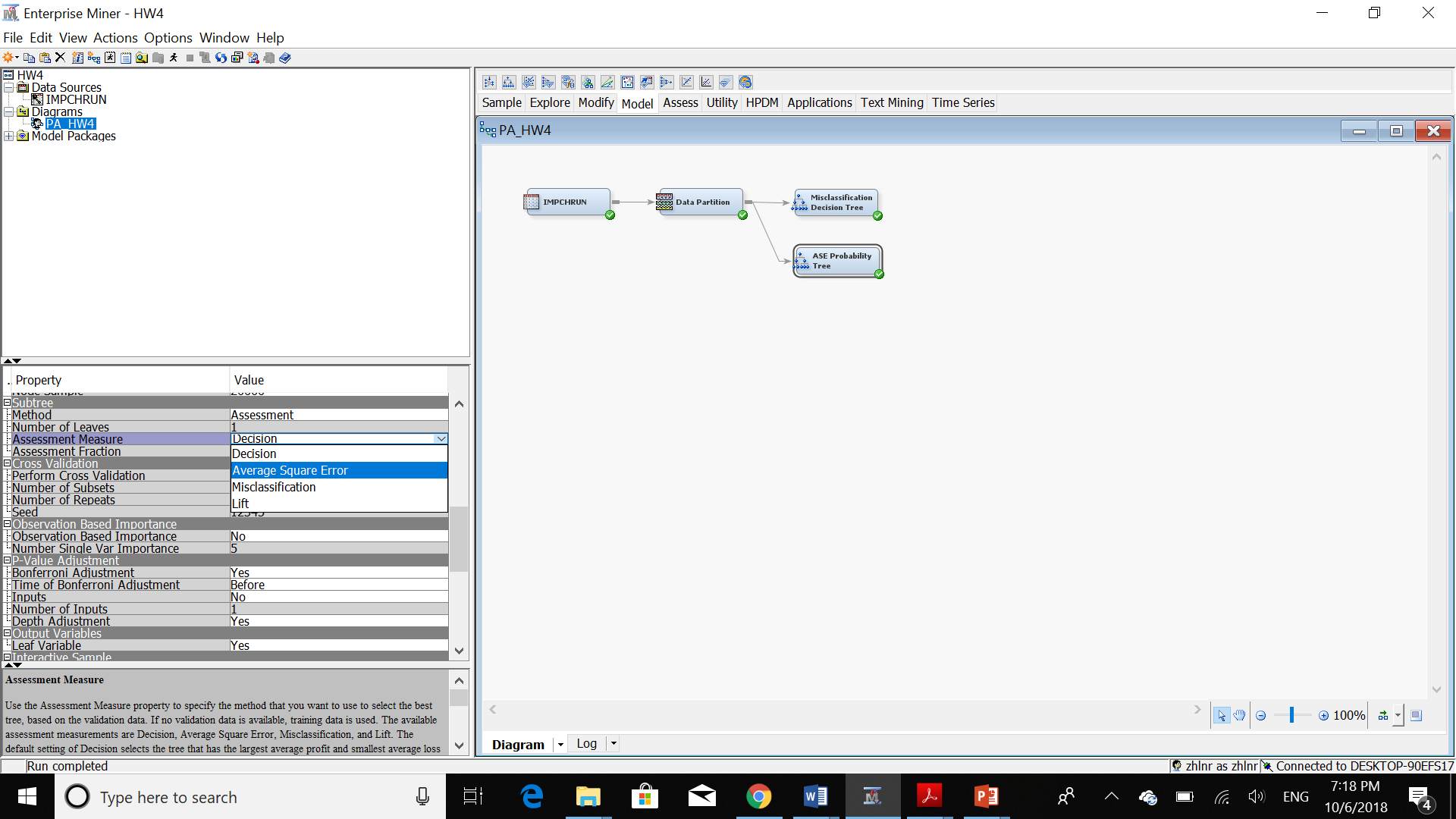




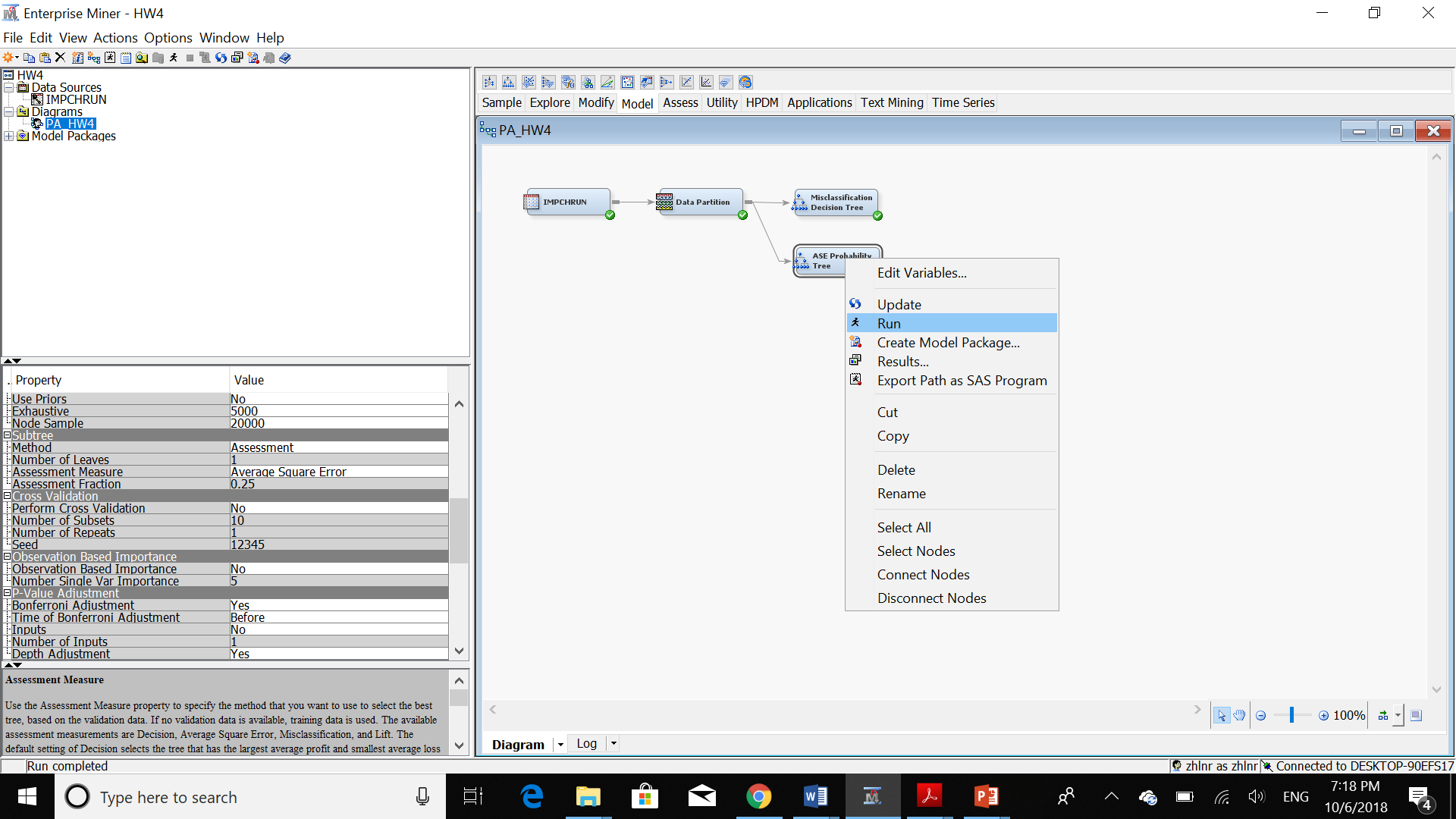
1. **Generate a best probability tree based on the assessment measure of mean(average) squared error.**

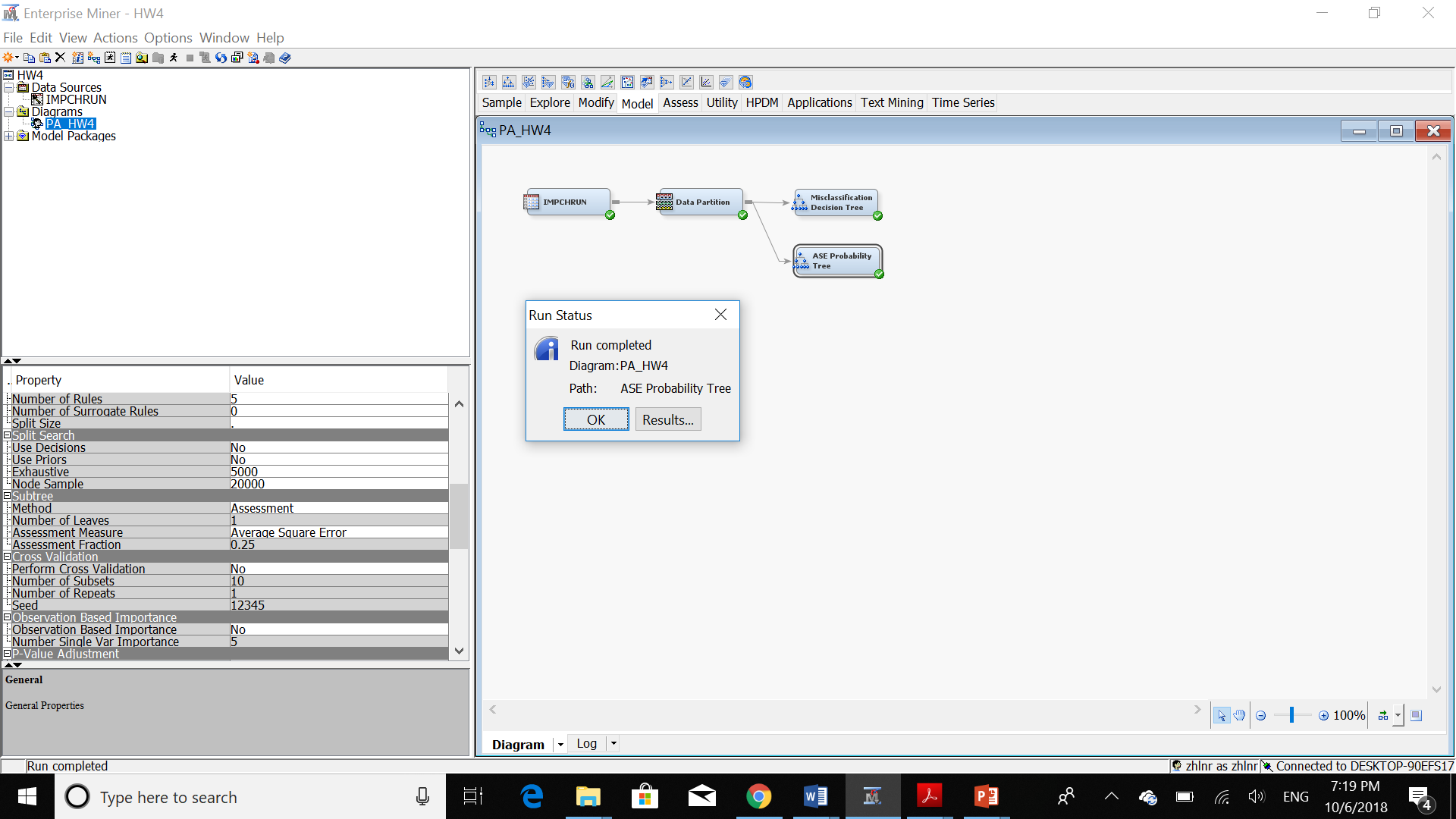
I created another node for this question:

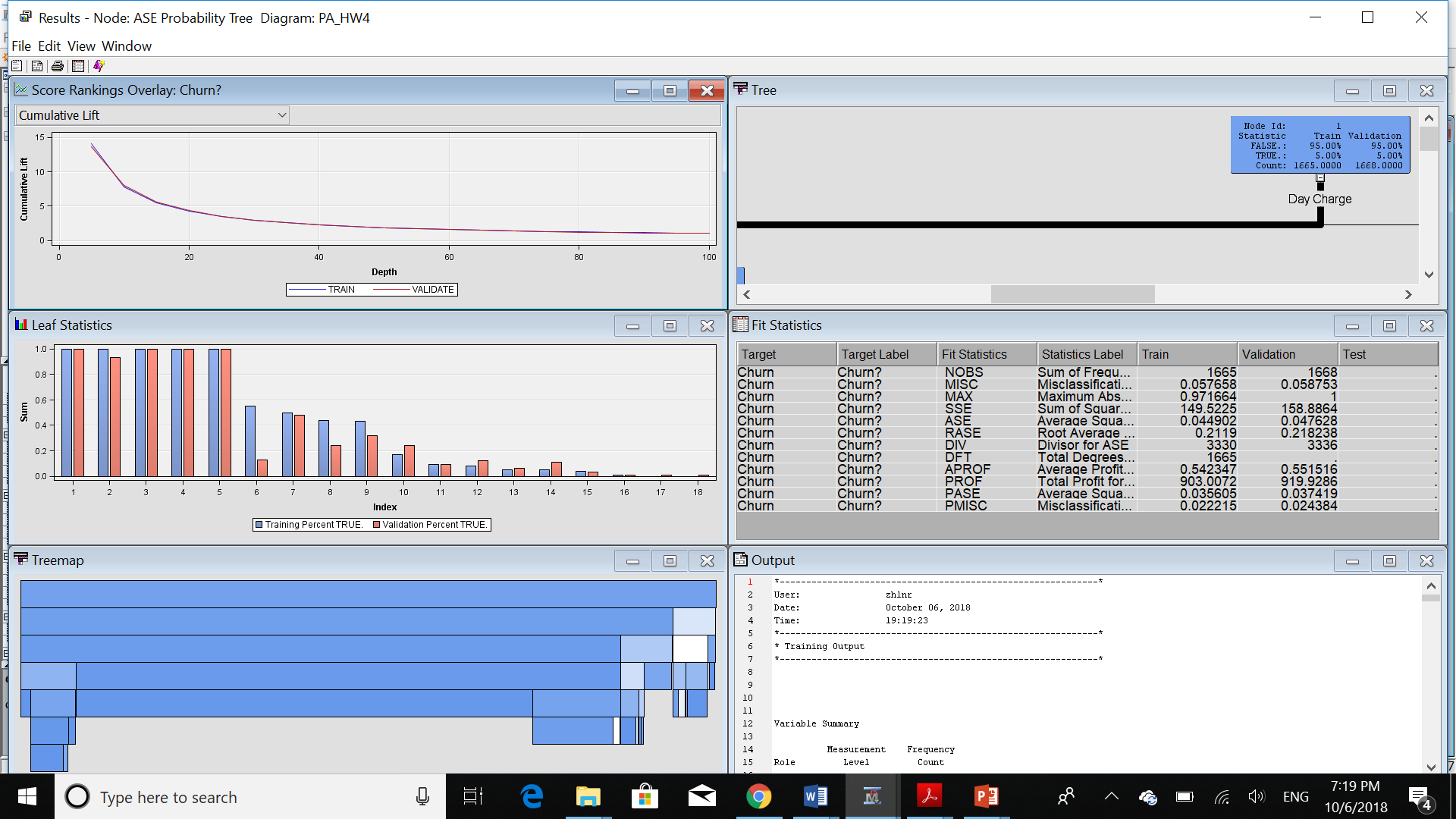




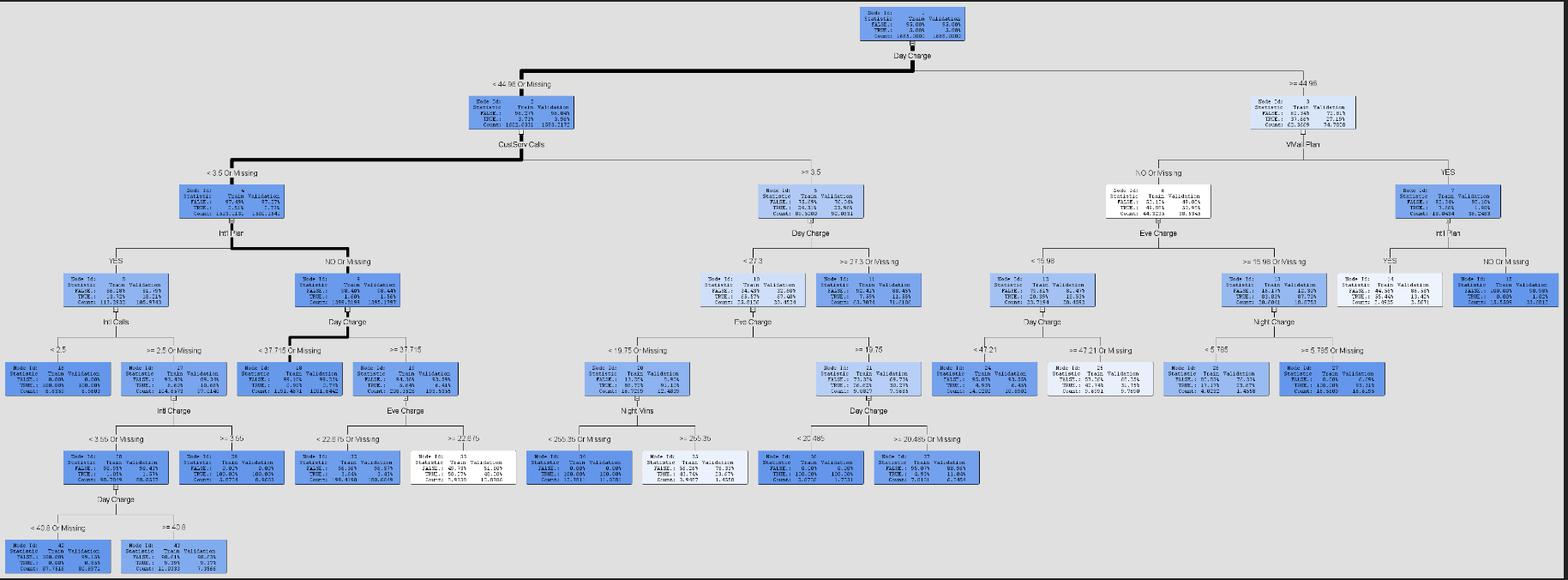
Then, we should run the new node:







Here is the tree:



Also, the subtree assessment:



For the next two questions: I considered the number of churn=TRUE as cases and the probability of these values as confidence. Also, I use the training set to answer these 2 questions. Furthermore, if the percentage of churn=TRUE is more than 50% I concluded that the consequent will be churning the customer and if the percentage of churn=FALSE is more than 50%, I concluded that the consequent will be the same as not churning the following customer.

1. **Generate a table of the full set of decision rules for the best decision tree.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Antecedent: If** | **Consequent: then** | **Cases** | **Confidence** |
| *Day Charge>=44.96 and VM Plan=Yes* | *Not Churning.* | 18.0434 | 7.66% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge >=27.3* | *Not Churning.* | 63.7074 | 7.59% |
| *Day Charge>=44.96 and VM Plan=No and Eve Charge>=15.98* | *Churning.* | 20.6041 | 83.83% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls<2.5* | *Churning.* | 8.6359 | 100% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’lPlan=No and Day Charge<37.715* | *Not Churning.* | 1191.4671 | 0.90% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge<27.3and Eve Charge<19.75* | *Churning.* | 16.7299 | 86.72% |
| *Day Charge>=44.96 and VMail Plan=No and Eve Charge<15.96 and Day Charge<47.21* | *Not Churning.* | 14.0202 | 4.93% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls>=2.5 and Int’l Charge<3.55* | *Not Churning.* | 98.7849 | 1.05% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls>=2.5 and Int’l Charge>=3.55* | *Churning.* | 5.8724 | 100% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=No and Day Charge>=37.715 and Eve Charge<22.675* | *Churning.* | 199.4190 | 3.64% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=No and Day Charge>=37.715 and Eve Charge>=22.675* | *Almost Equal.* | 8.9338 | 50.27% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge<27.3 and Eve Charge>=19.75 and Day Charge<20.485* | *Churning.* | 2.0726 | 100% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge<27.3 and Eve Charge>=19.75 and Day Charge>=20.485* | *Not Churning.* | 7.0101 | 4.93% |
| *Day Charge>=44.96 and VMail Plan=No and Eve Charge<15.98 and Day Charge>=47.21 and Int’l Calls<4.5* | *Churning.* | 3.7998 | 100% |
| *Day Charge>=44.96 and VMail Plan=No and Eve Charge<15.98 and Day Charge>=47.21 and Int’l Calls>=4.5* | *Not Churning.* | 5.8993 | 5.86% |

1. **Generate a table of the full set of decision rules for the best probability tree.**

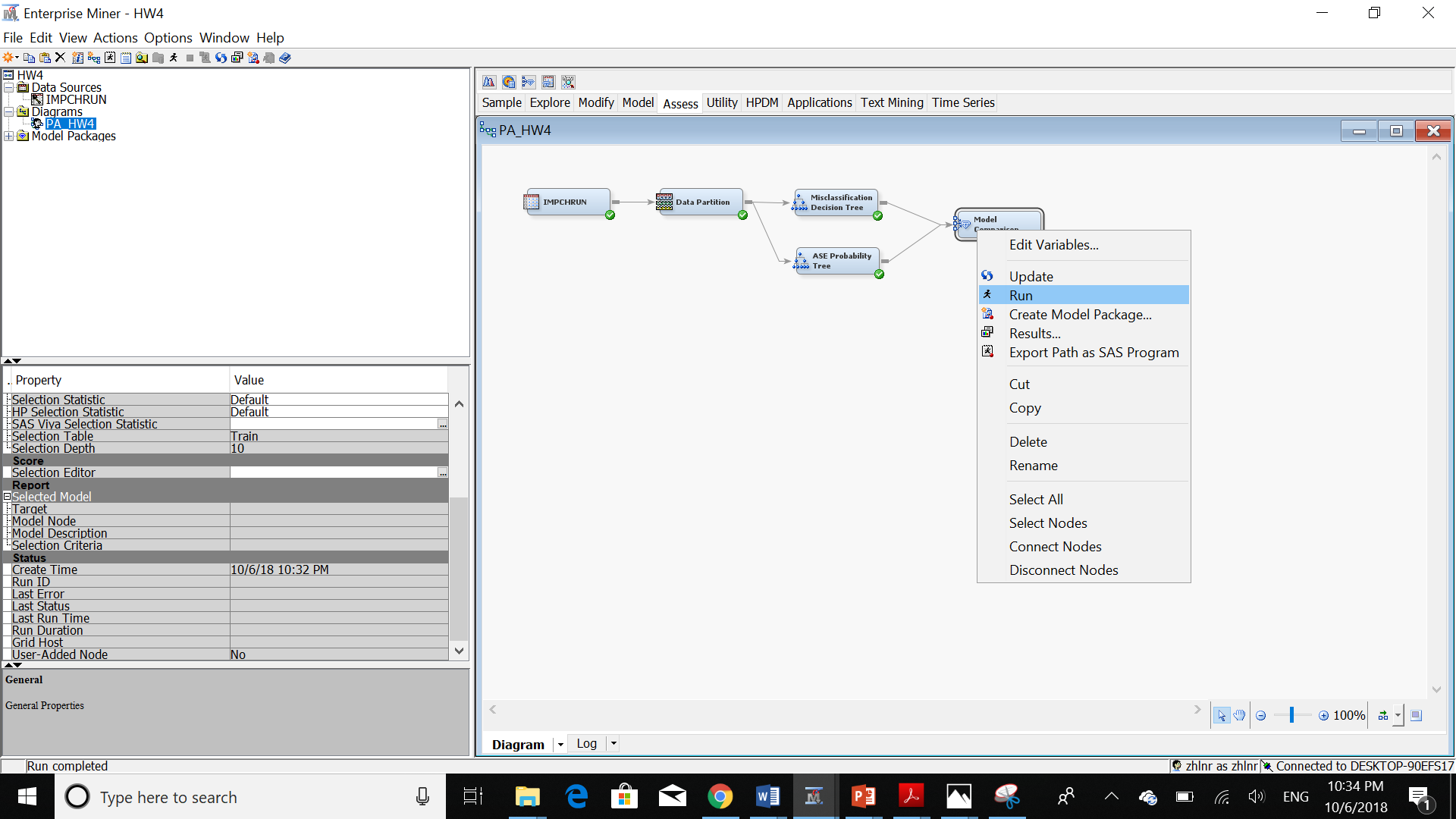
|  |  |  |  |
| --- | --- | --- | --- |
| **Antecedent: If** | **Consequent: then** | **Cases** | **Confidence** |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge>=27.3* | *Not Churning.* | 63.7074 | 7.59% |
| *Day Charge>=44.96 and VMail Plan=Yes and Int’l Plan=Yes* | *Almost Equal. Churning.* | 2.4925 | 55.44% |
| *Day Charge>=44.96 and VMail Plan=Yes and Int’l Plan=No* | *Not Churning.* | 15.5509 | 0% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls<2.5* | *Churning.* | 8.6359 | 100% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=No and Day Charge<37.715* | *Not Churning.* | 1191.4671 | 0.90% |
| *Day Charge>=44.96 and VMail Plan=No and Eve Charge<15.98 and Day Charge<47.21* | *Not Churning.* | 14.0202 | 4.93% |
| *Day Charge>=44.96 and VMail Plan=No and Eve Charge<15.98 and Day Charge>=47.21* | *Not Churning.* | 9.6991 | 42.74% |
| *Day Charge>44.96 and VMail Plan=No and Eve Charge>=15.98 and Night Charge<5.785* | *Not Churning.* | 4.0232 | 17.17% |
| *Day Charge>44.96 and VMail Plan=No and Eve Charge>=15.98 and Night Charge>=5.785* | *Churning.* | 16.5809 | 100% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls=2.5 and Int’l Charges=3.55* | *Churning.* | 5.8724 | 100% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=No and Day Charge>=37.715 and Eve Charge<22.675* | *Not Churning.* | 199.4190 | 3.64% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=No and Day Charge>=37.715 and Eve Charge>=22.675* | *Almost Equal. Churning.* | 8.9538 | 50.27% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge<27.3 and Eve Charge<19.75 and Night Mins<255.35* | *Churning.* | 12.7811 | 100% |
| *Day Charge<44.96 and CustServ Calls>=3.5 and Day Charge<27.3 and Eve Charge<19.75 and Night Mins>=255.35* | *Not Churning.* | 3.9487 | 43.74% |
| *Day Charge<44.96 and CustServe Calls>=3.5 and Day Charge<27.3 and Eve Charge>=19.75 and Day Charge<20.485* | *Churning.* | 2.0726 | 100% |
| *Day Charge<44.96 and CustServe Calls>=3.5 and Day Charge<27.3 and Eve Charge>=19.75 and Day Charge>=20.485* | *Not Churning.* | 7.0101 | 4.93% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls>=2.5 and Int’l Charge<3.55 and Day Charge<40.8* | *Not Churning* | 87.7516 | 0% |
| *Day Charge<44.96 and CustServ Calls<3.5 and Int’l Plan=Yes and Int’l Calls>=2.5 and Int’l Charge<3.55 and Day Charge>=40.8* | *Not Churning* | 11.0333 | 9.39% |

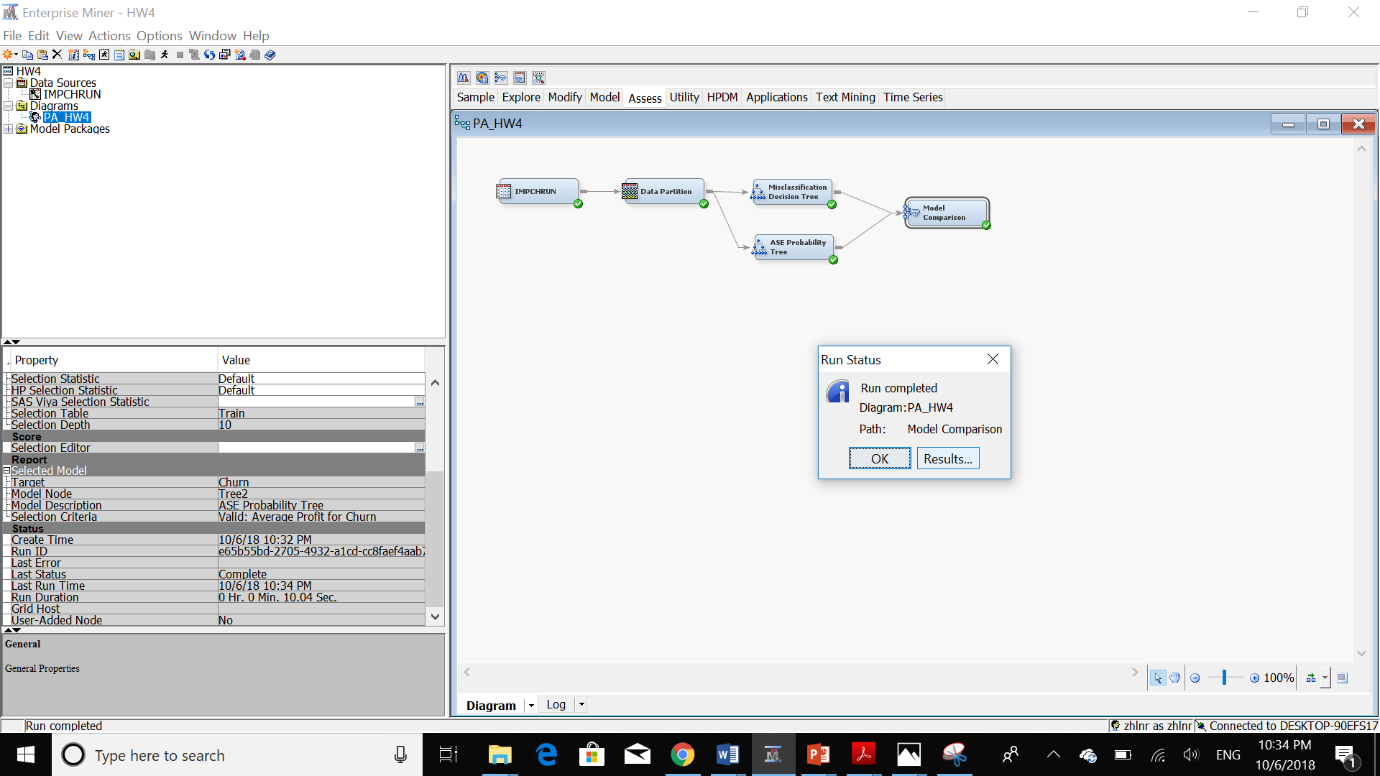
1. **Compare the two trees and discuss the benefits and drawbacks of each.**

I created a model comparison node in order to compare these two trees:

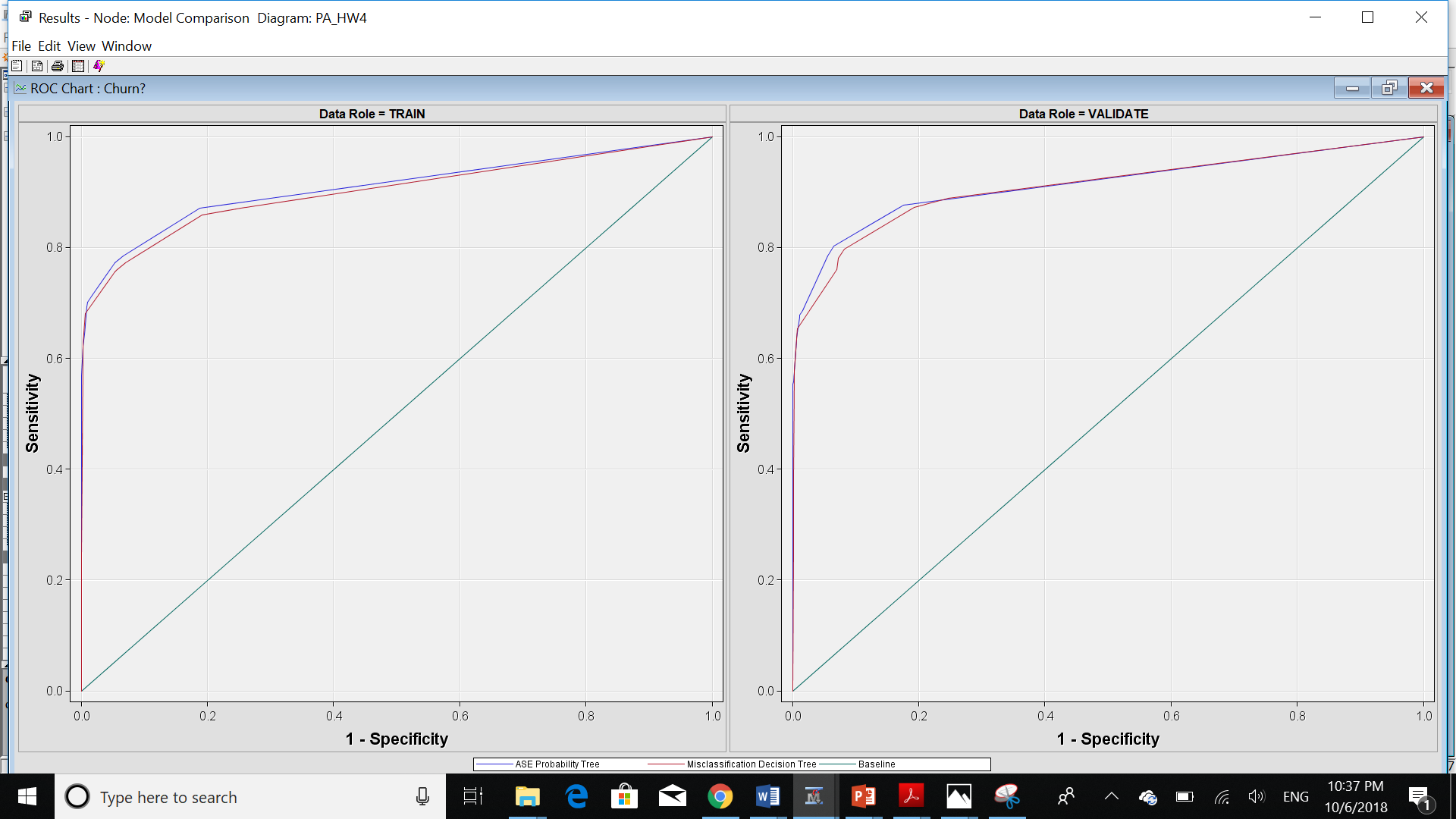


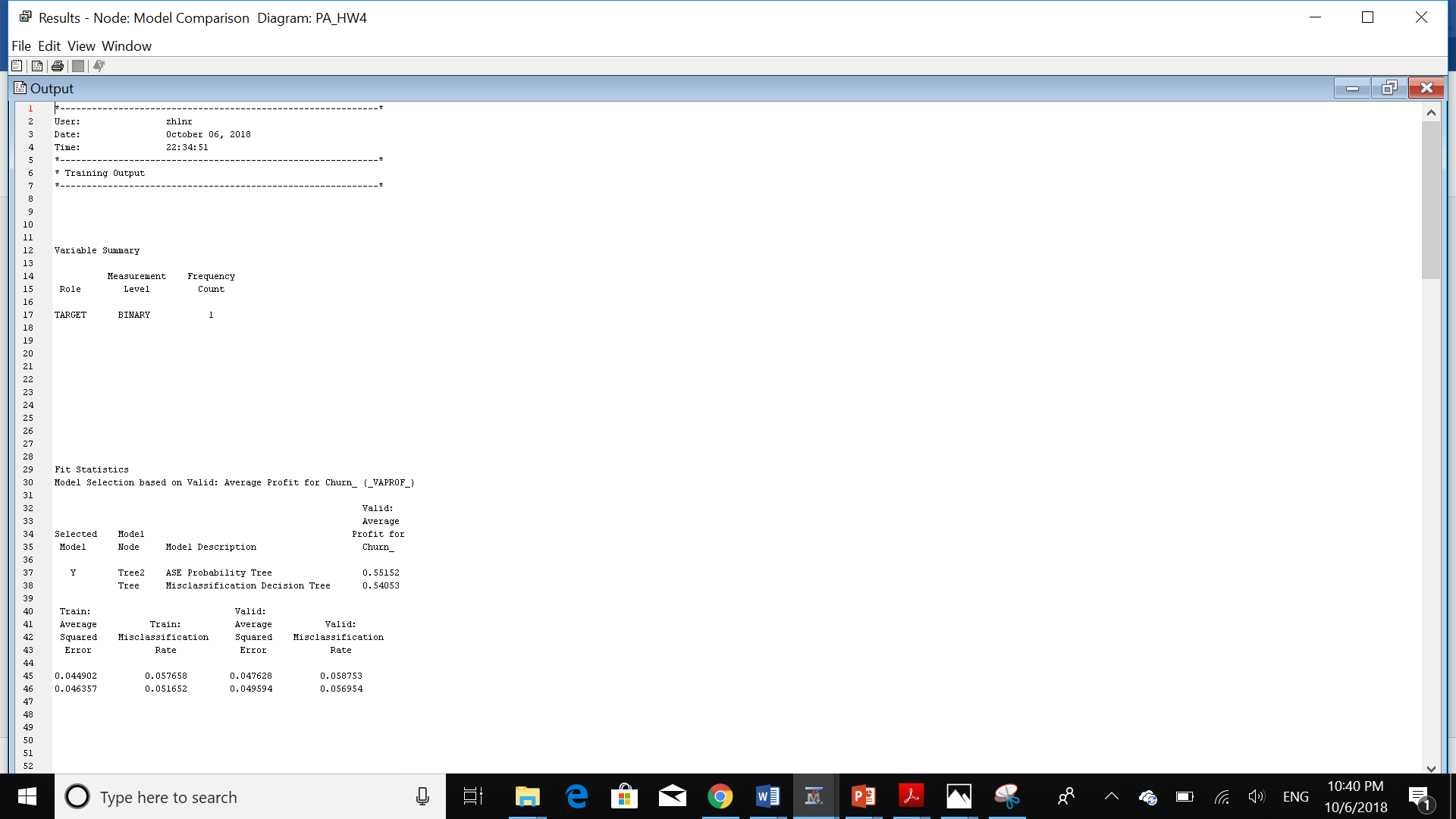
Then, we should run the new node:





We can see sensitivity and specificity of these two trees:

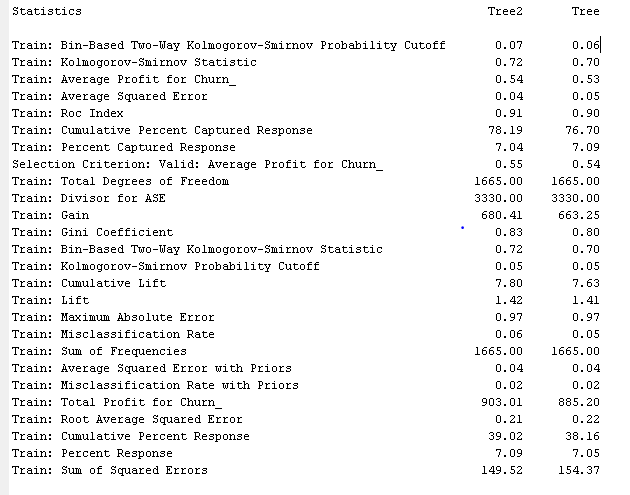




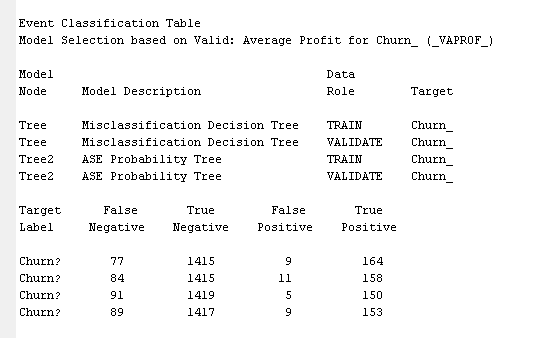
As we can see in the above picture, the Average Squared Error of ASE tree is less than the other one and the Misclassification rate of the decision tree base on the misclassification rate is lower than the ASE tree.

Also, the first tree has a height of 6 while the second one which is the ASE has a height of 7. So, we can conclude that it takes more time to parse the second tree and it’s easier to parse the first tree. So, in terms of the height of trees, the tree based on the misclassification rate is better. Furthermore, the ASE tree has 17 leaves while the other tree has 14 leaves. It means that to find the answer we should check for a greater number of leaves in the ASE tree. So, the tree based on the lower misclassification rate gives us an easier tree to parse. The other point which I would like to bring here is, in the first tree we have more leaves at lower levels. In other words, when we are parsing the first tree we can get to the final nodes, which are the leaves, sooner than in other trees. For example, as many leaves in the ASE tree are located at higher levels, we have to go through these nodes up to the highest level to find the answer. While we can find the answer easier in other trees.

For better comparing these two trees, let’s take a look at the statistic of the model comparison’ result:



On the other hand, the ROC index of the ASE tree is a little bit higher than the other tree’s ROC index which means that this tree is better in this matter. While the difference doesn’t make so much difference as it is just 0.1. But we can conclude that the ASE tree is better in terms of Sensitivity and Specificity values. Also, the sum of Squared Errors for ASE is less than the other tree as it is obvious from its name.



To compare these two trees, I copied the table from the result of model comparison. In this table, we can see the values of False Negative, True Negative, False Positive and True Positive for each tree, either for training set or validation set.

The first tree has the lower False Negative while the ASE tree has the lower False Positive value. Here, we know that the primarily important value for us is churn=true. So, the false positive means a customer churned while he/she did not. Also, false-negative means, a customer did not churn while he/she did. As we can see here the latter is more critical for us. Because we do not want to lose any customer. So, the lower the False Negative rate, the better. Here, the ASE tree has a higher False Negative rate which is worst. So, in terms of False Negative, the tree based on a lower misclassification rate is better. Also, if a customer churned, which means True positive, we should take a look at the True Positive column. The tree-based on lower misclassification rate has a higher true positive which is better than the other one.

Long story short, these two trees are not very different. But the ASE tree has lower Mean Squared Error and better ROC value. While the other tree has a lower misclassification rate and the other advantages as discussed before.