**Project Part 2**

Durgaprasad Koppala, Sri Sai Sathya Keshav Arigela, Sree Harshita Reddy Gangasani, Nikhil Vishnu Teja Madasu , Jyothi Naga Lakshmi Addala and Dileep Durgam

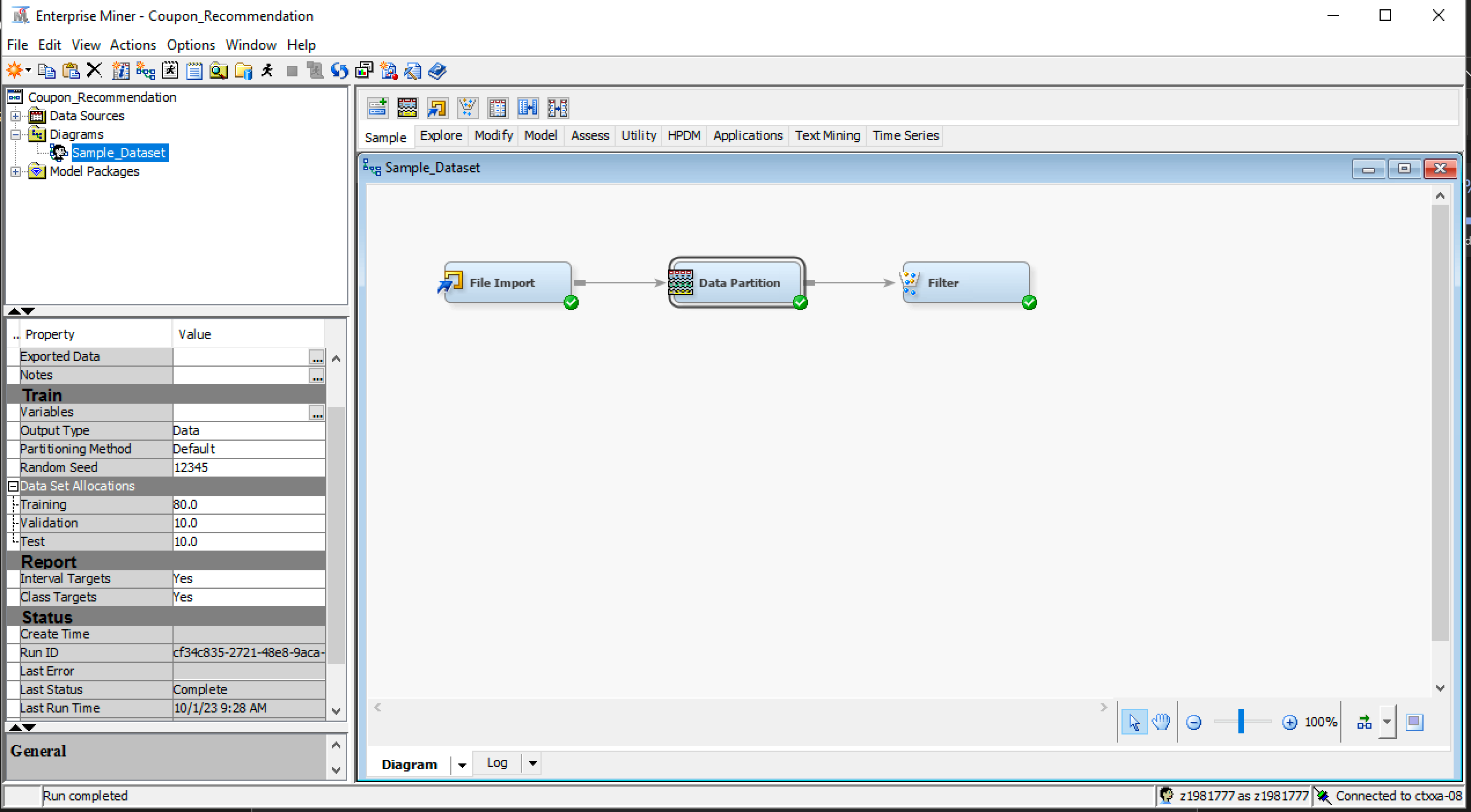
Northern Illinois University

OMIS 681: Predictive Bussiness Analytics

Prof. Karah Osterberg

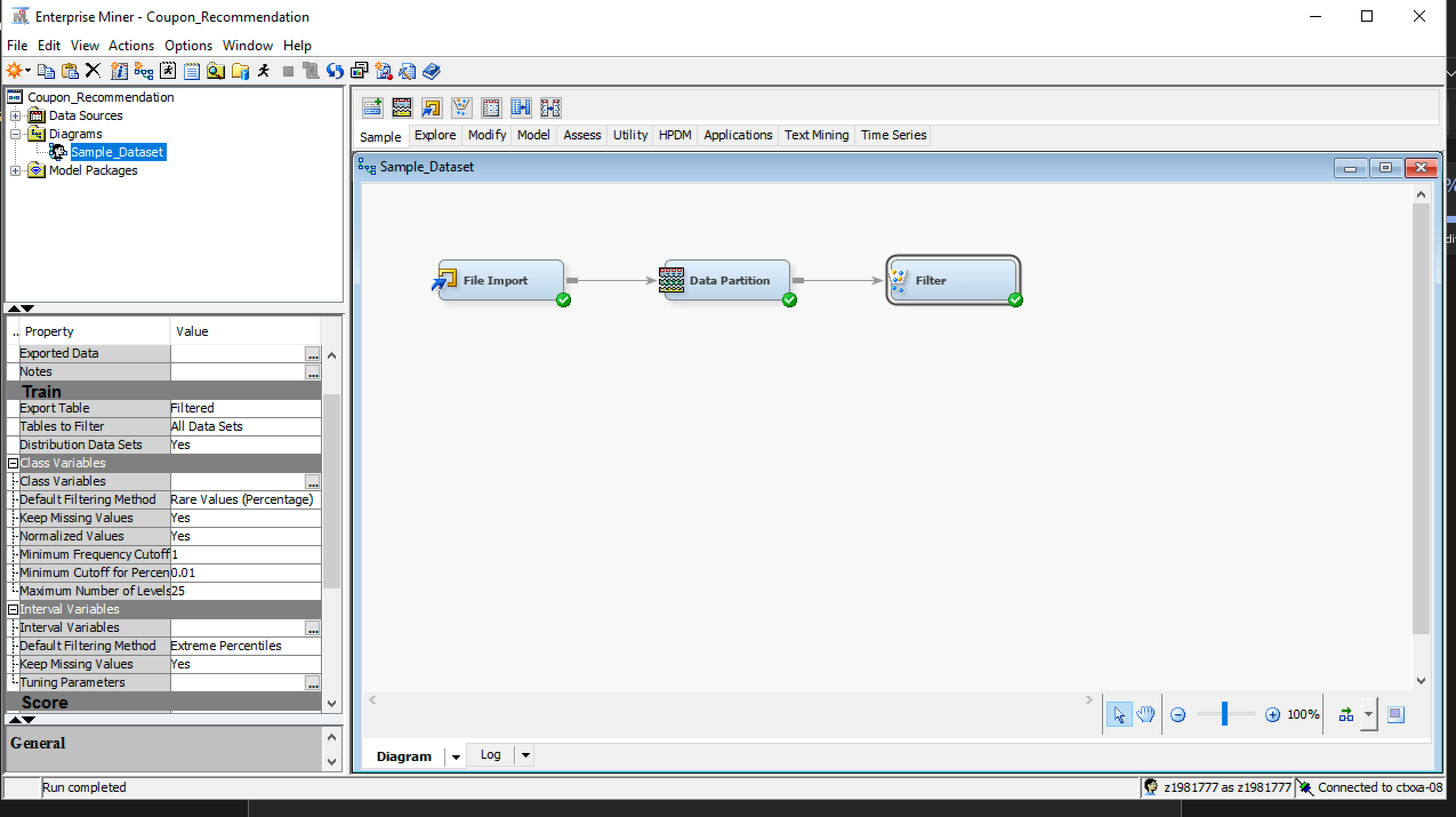
October 2, 2023

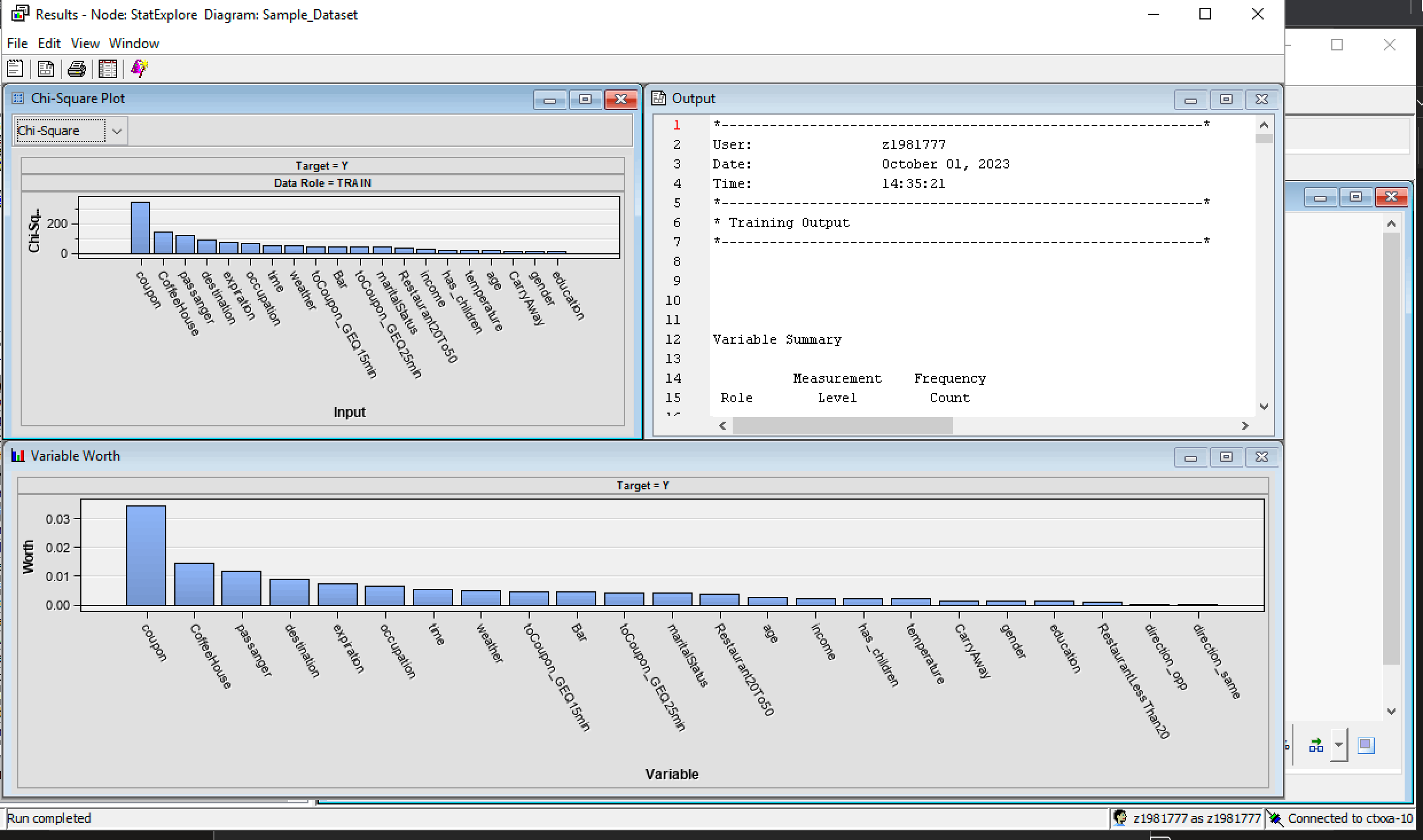
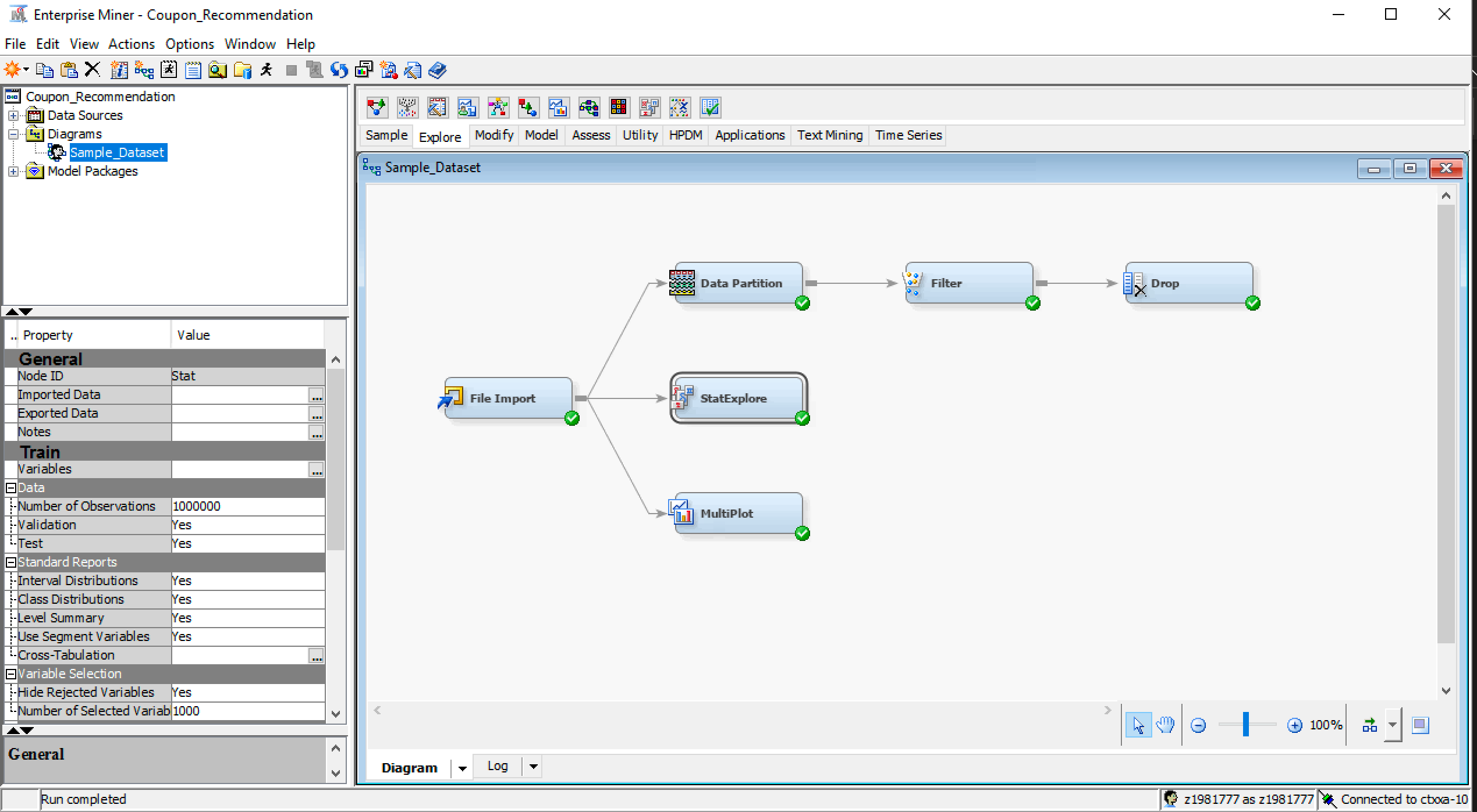
Data Partition: The data set is partitioned into 80 percent training data, 10 percent validation and the remaining 10 percent into testing data. Most of the data is assigned to training to get the better model.



File Import: The File Import node is used to import excel dataset file to SAS EM.

Filter: In the filter node, under its properties, the fault filtering method is set to ‘Extreme Percentiles’ and the table to filter is set to ‘All Data Set’. The reason for using the filter node is to filter out the outlier data present in the dataset that affects the model.



Stat Explore: 

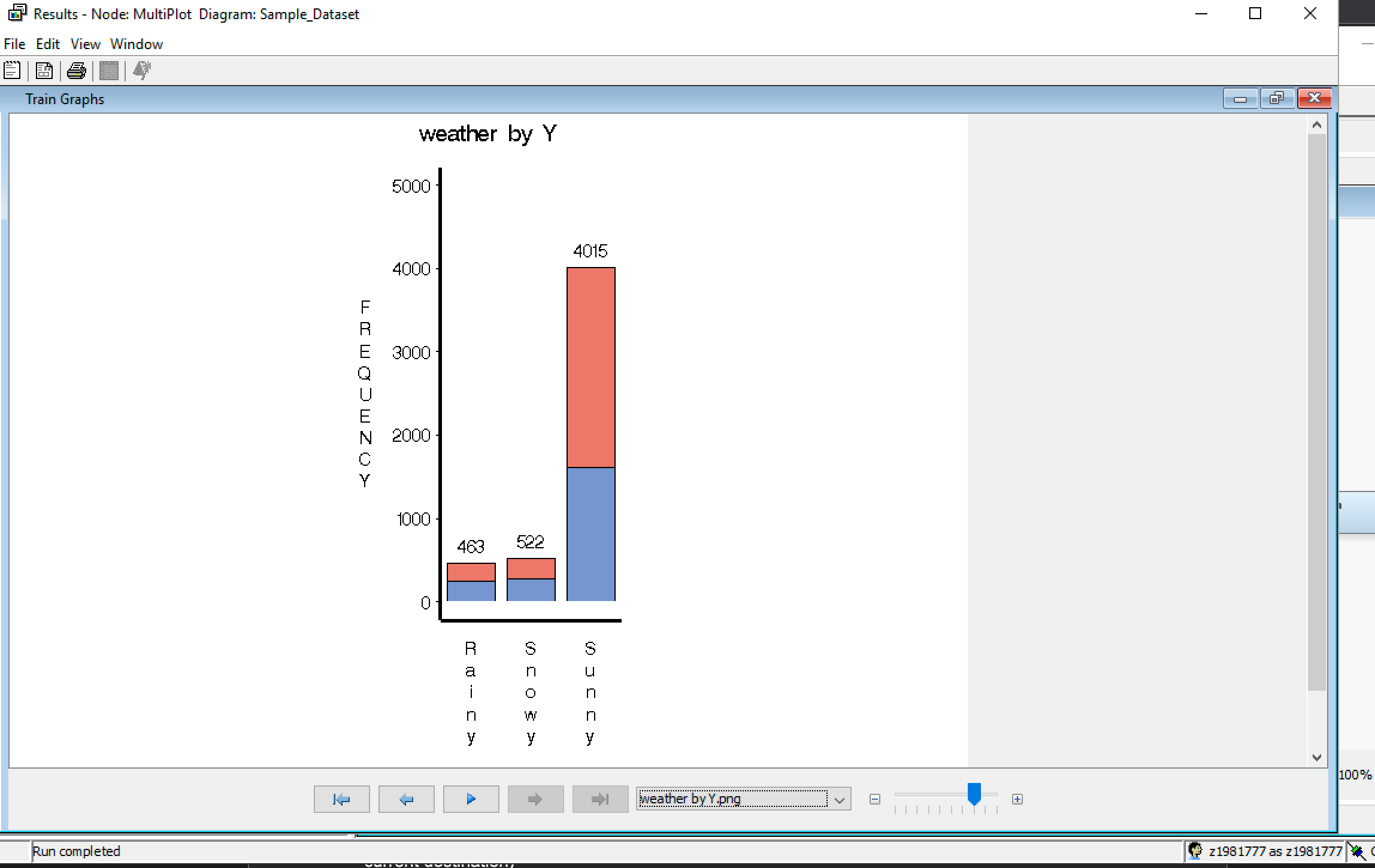
The Chi-square plot explains the relationship strength with the target variable ’Y’. The coupon variable has the highest correlation with the target variable.

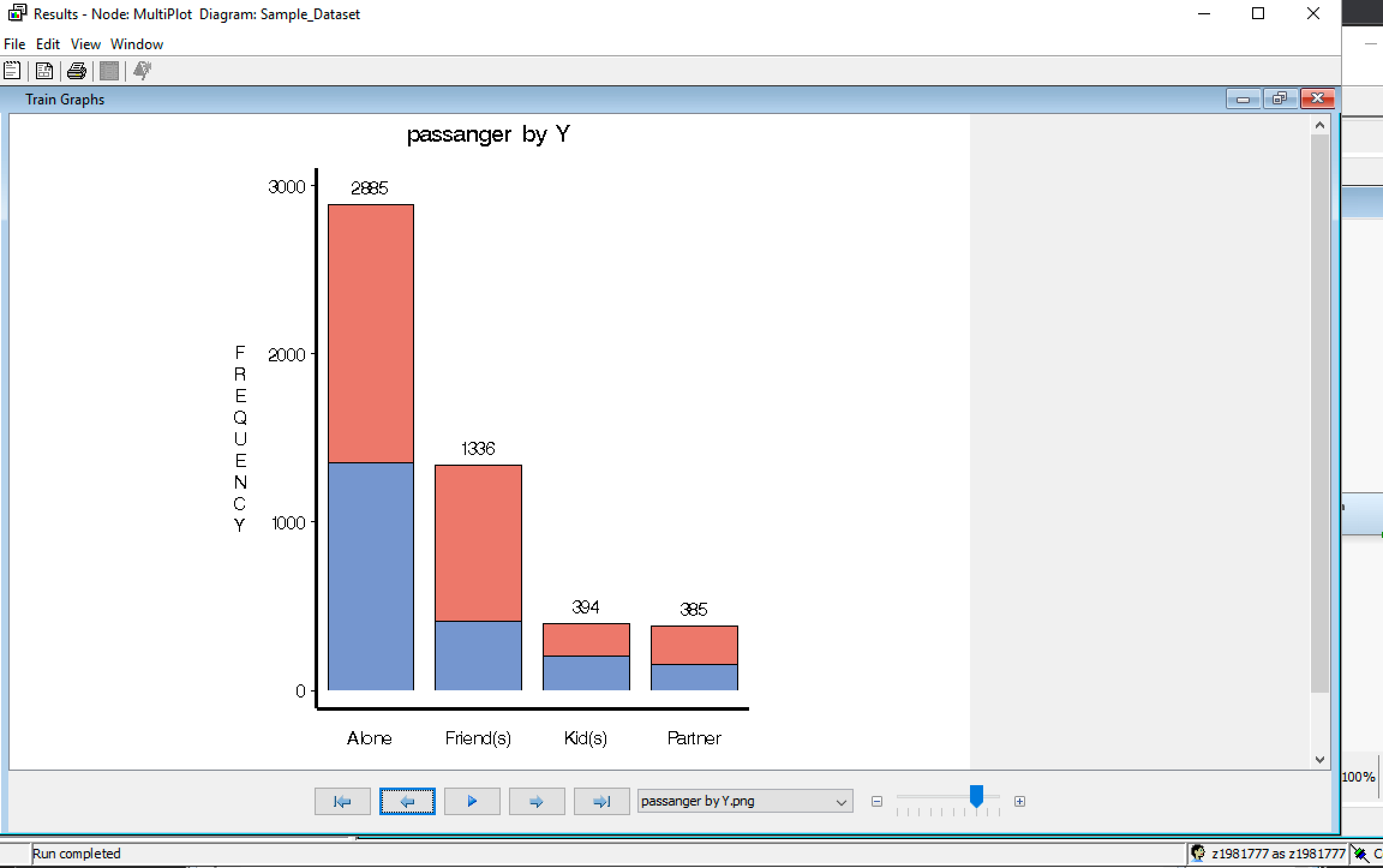
Variable worth is calculated using the P value of the Chi-square statistic. The coupon is the highly predicted target variable of ‘Y’ followed by CoffeeHouse, Passenger, and destination etc.

Properties: In the properties of Stat Explore, The Chi-Square and Interval variable are set to ‘yes’.

MultiPlot:

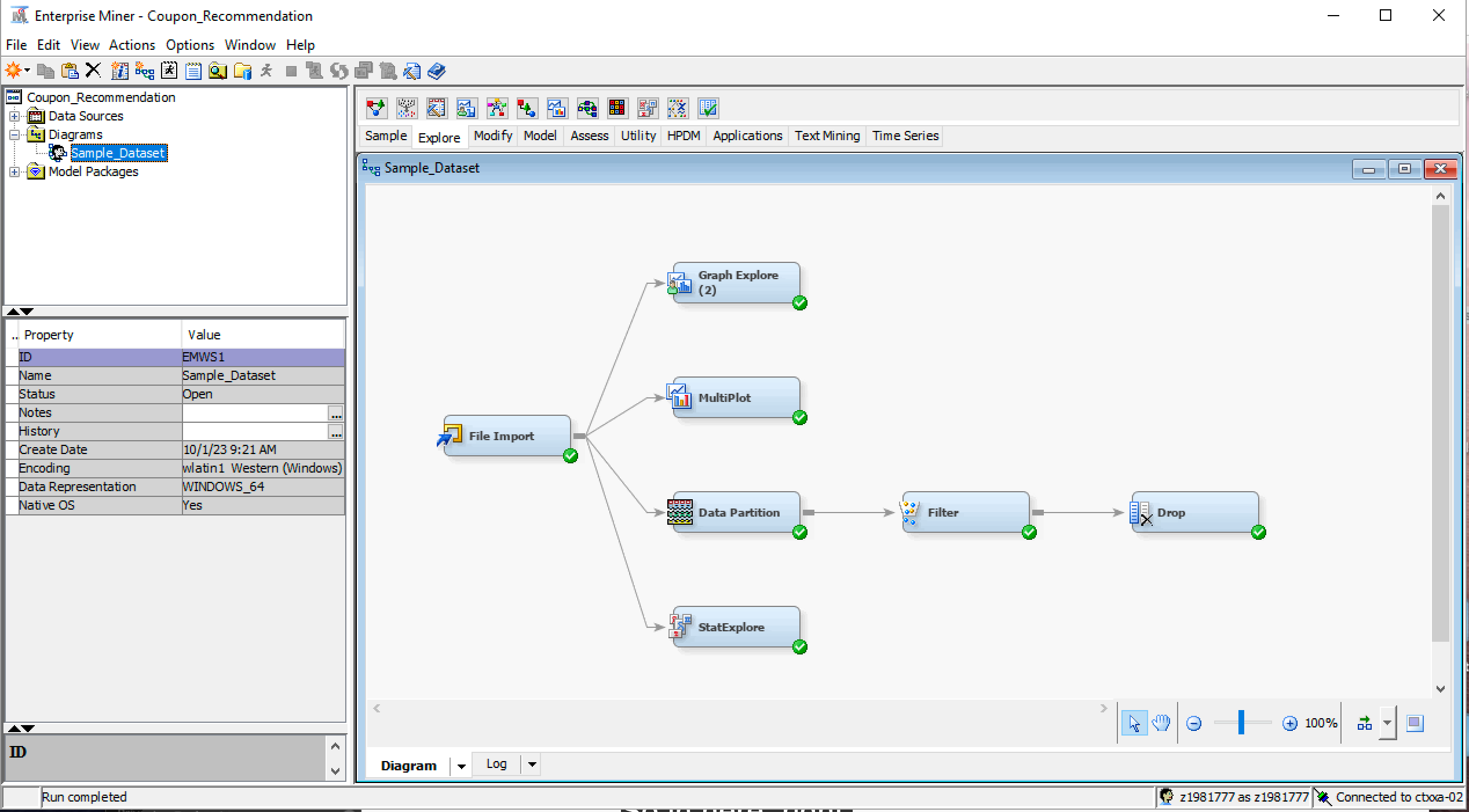
The MultiPlot node is used to analyze the individual variables vs Target variables plot. Some of the graph plots are explained below. If the weather is sunny, there is a high probability that the person will accept the coupon. If a driver has a friend in the passenger seat, then there is a high probability that the person will accept the coupon.

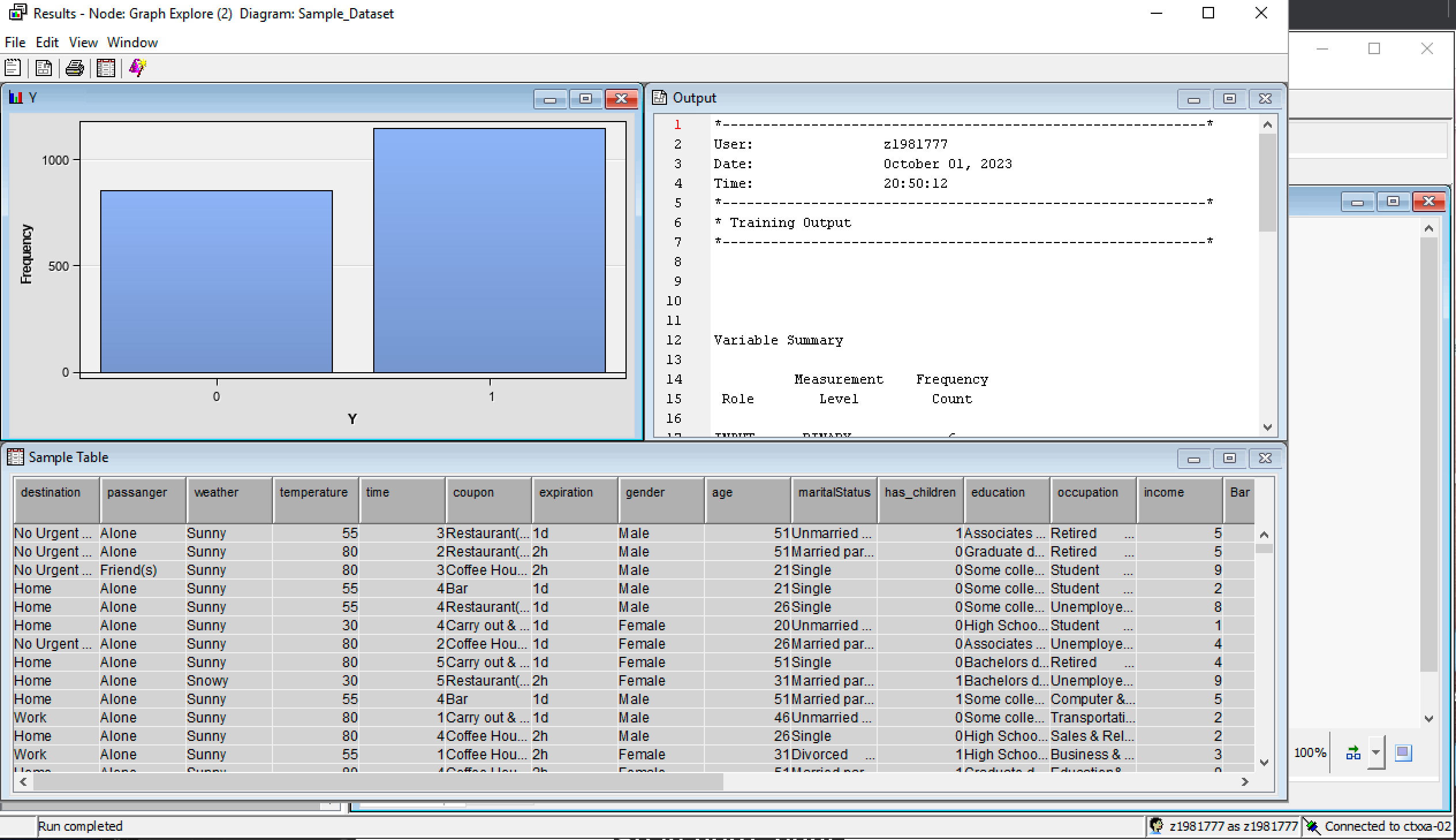




Graph Explore:

As per the results of the Graph Explore node, there is a higher frequency of one’s compared to zero’s. Meaning the acceptance ratio is higher than rejection ratio when a coupon is offered to a driving person.



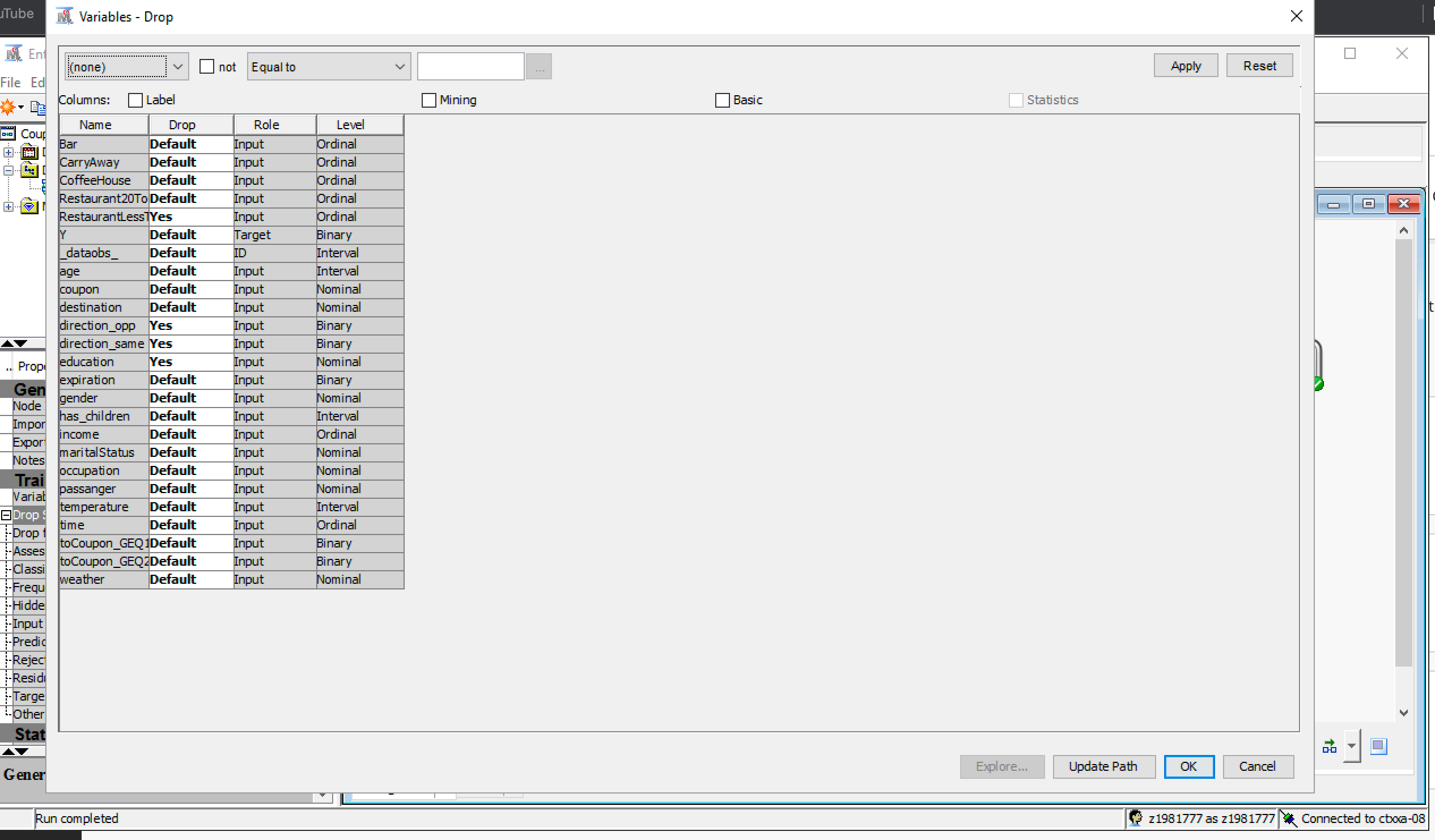


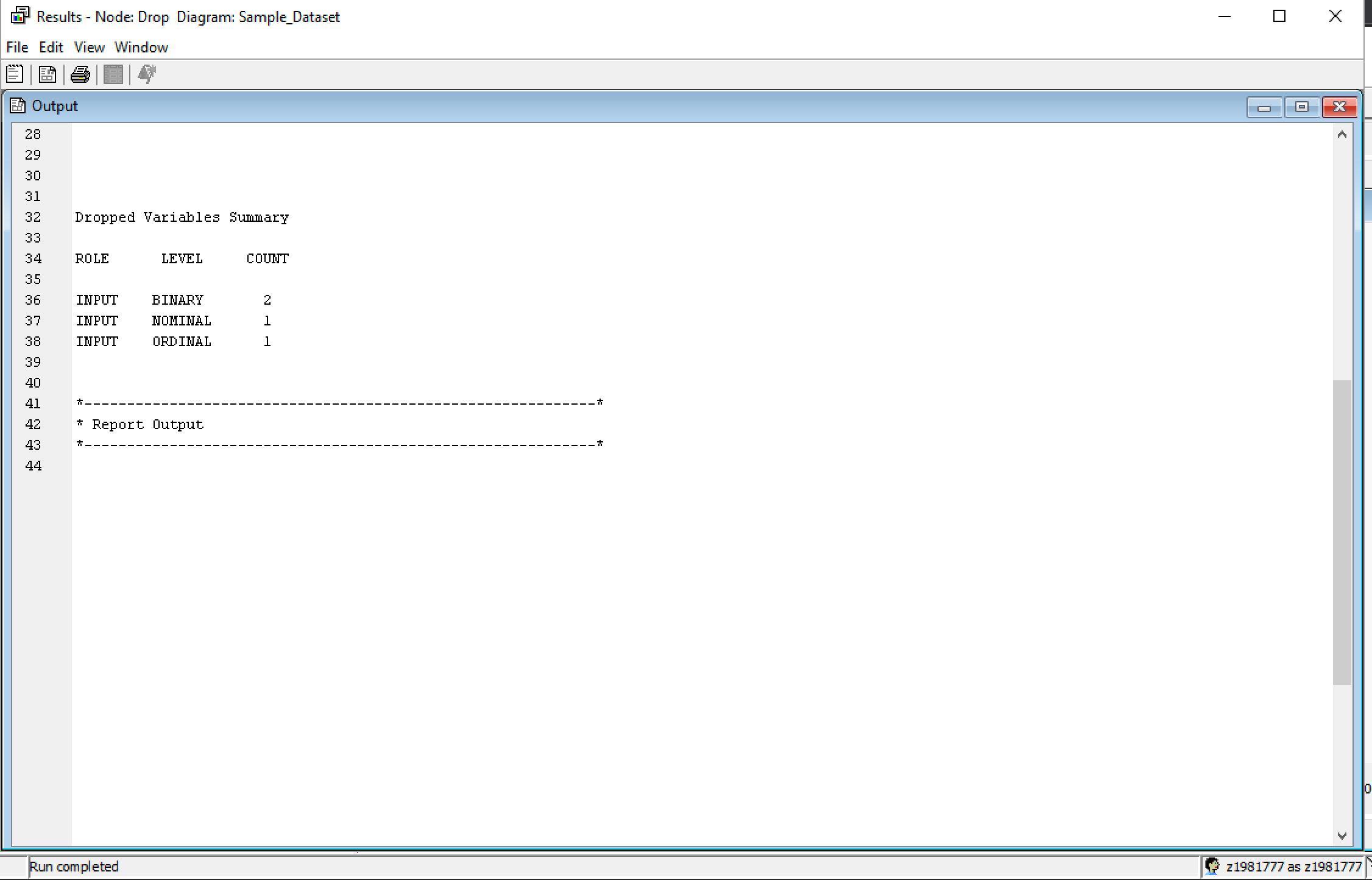
After changing the data options for ‘Gender’ variable, the acceptance rate for a male is slightly higher than female when a coupon is offered.



Drop:

As per the Stat Explore node, the variables ‘RestaurantLessThan20’, ‘direction\_opp’, ‘direction\_same’, and ‘education’ are less impactful in predicting the target variable. By using a Drop node, these impactful variables are dropped from the model.





Transformation Variable: The transformation node is used for variables transformation. The properties setting of Interval Input is changed to ‘Maximum Normal’ and class input is changed to ‘Dummy Indicators’. The reason for using the transformation node is to make the model fit better.

