School of Computer Application (DS & AI)

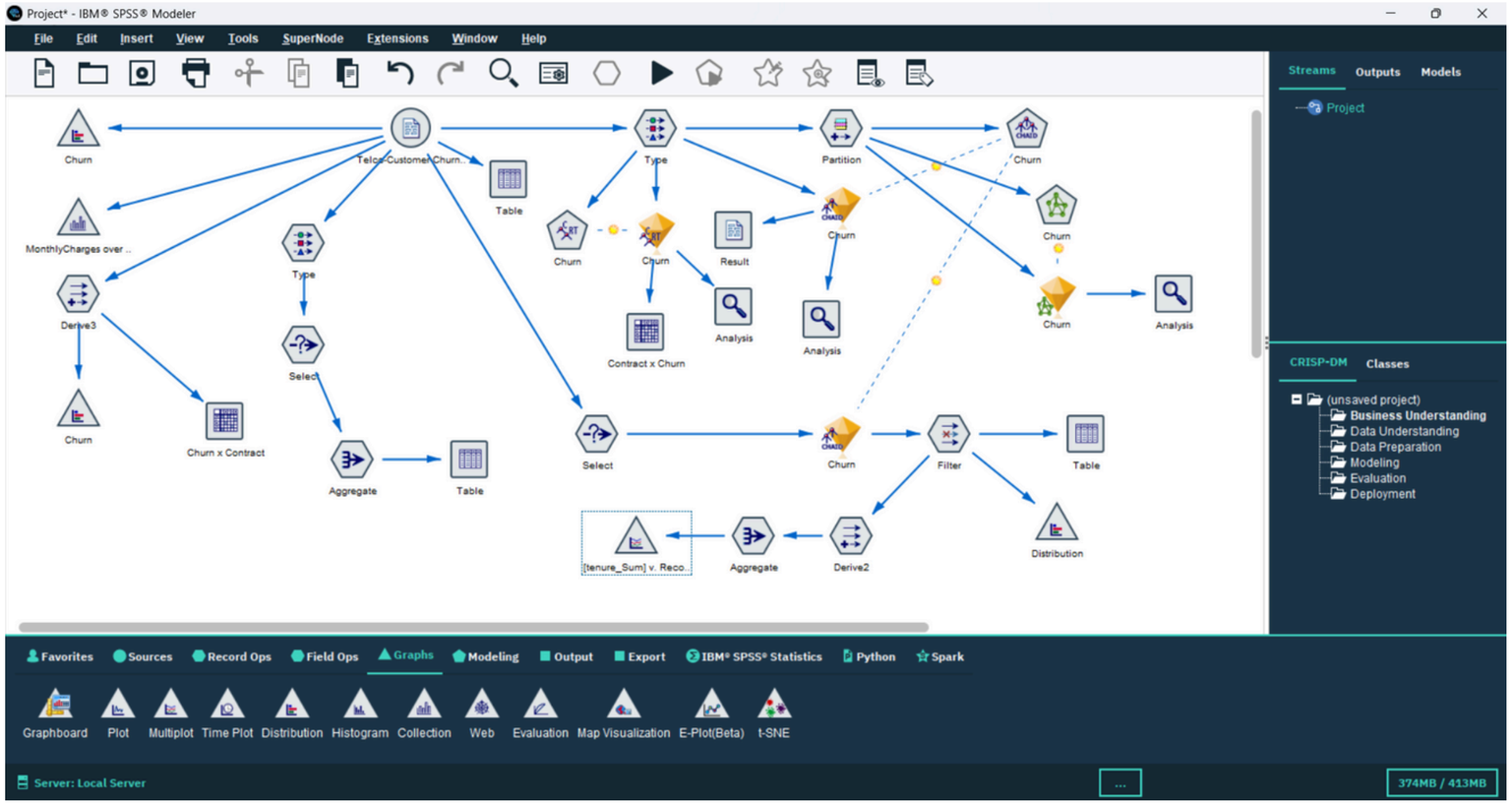
Predictive Analysis

Aditya Yadav Amit Yadav

Submitted to: Mr Vikash (IBM)

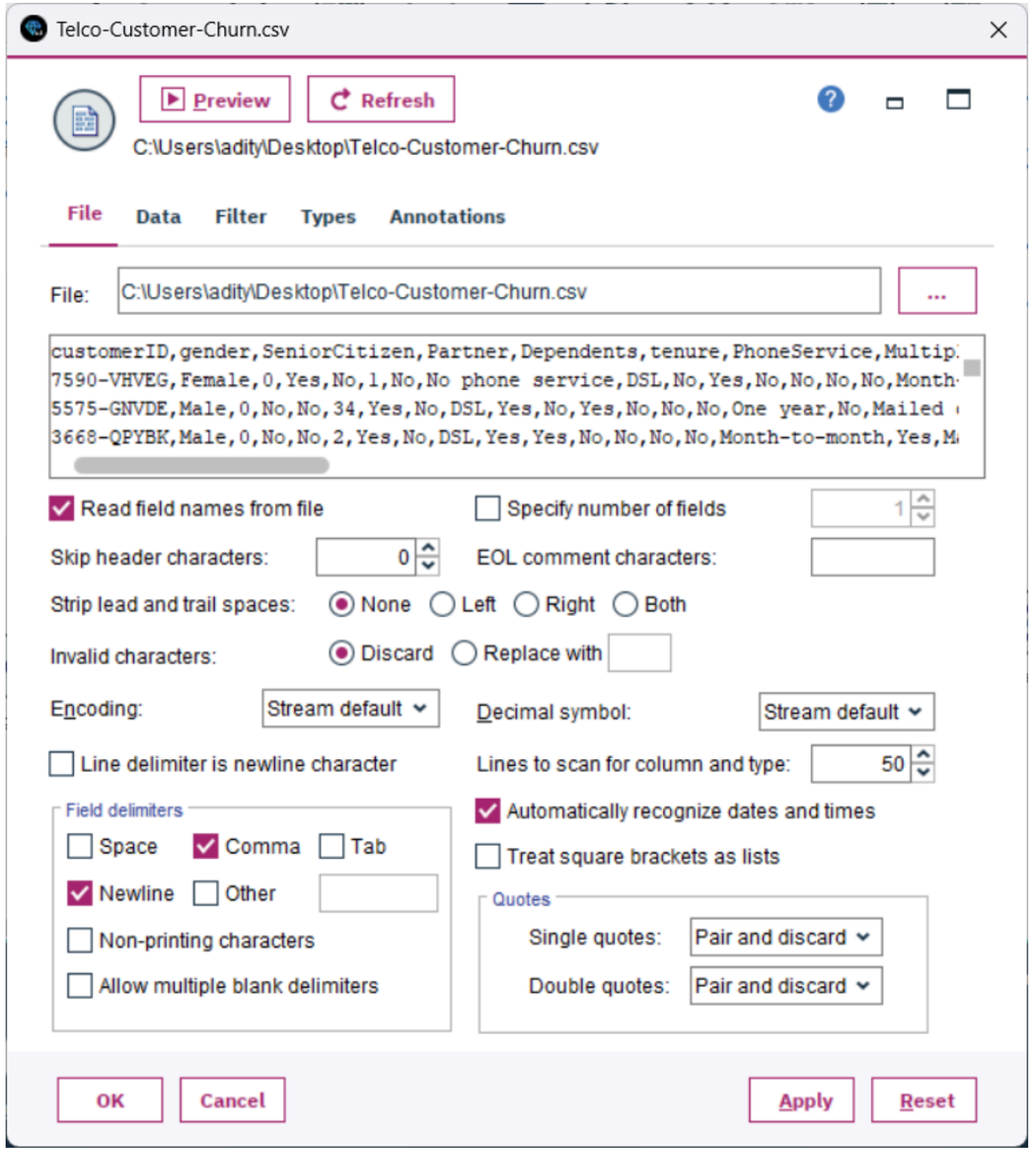
Churn analysis of telecom customers using IBM SPSS Modeler on an open source dataset

This is the Stream we made to analyze the telecom customer churn.

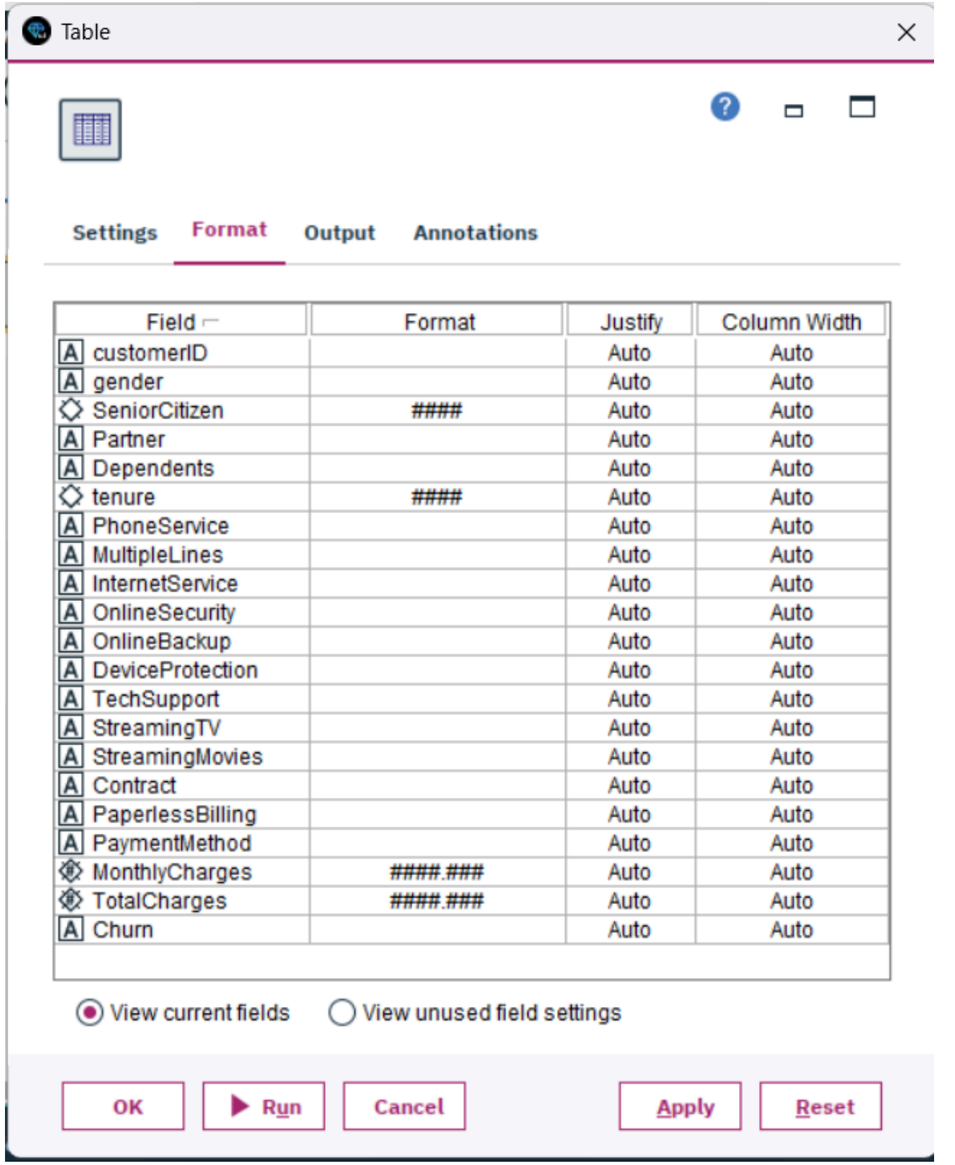


Now we will Understand the step-by-step procedure of creating this.

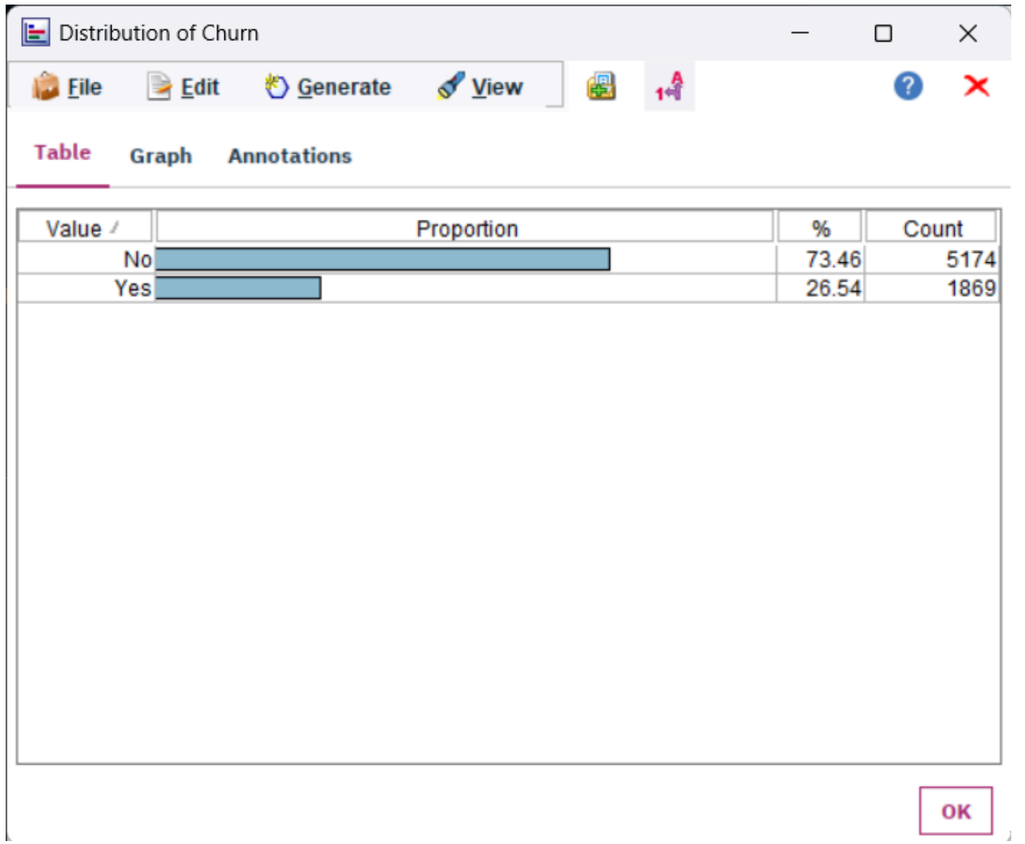
To Analyze the customer Churn we first need to have a proper dataset and upload it to IBM SPSS Modeler Using Var.file from Sourses.
The Data is in a CSV file format.
it contains 21 fields which are :-
customerID,gender,SeniorCitizen,Partner,
Dependents,tenure,PhoneService,MultipleLines,
InternetService,OnlineSecurity,OnlineBackup,
DeviceProtection,TechSupport,StreamingTV,
StreamingMovies,Contract,PaperlessBilling,
PaymentMethod,MonthlyCharges,TotalCharges,
Churn



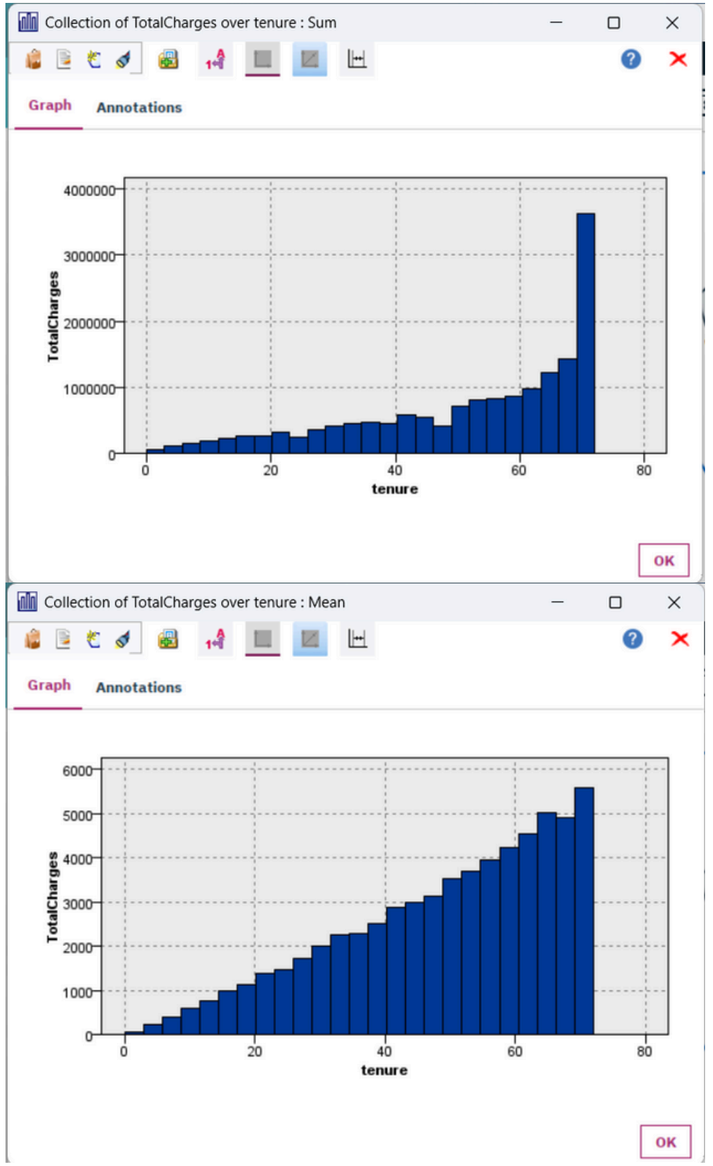
Now we can use the Table view from the Output to view and review the data and data types to make changes before the analysis begins. This also gives us an overview of the data.



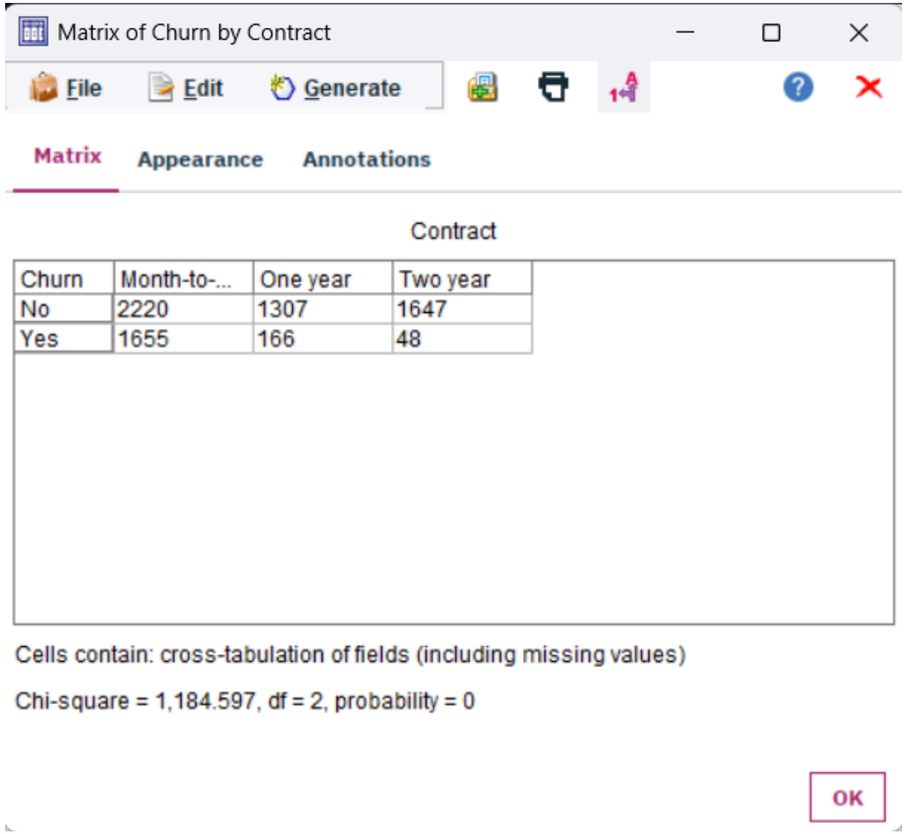
We now use Graph view to analyse the rate of churn of customers during a particular time period.



We can analyse the data with various factors to try to understand the relationship between different fields of data, like the example shown in the graph.



Here we use a matrix from the output to show the more complex relationships between data in our dataset, which shows that the rate of churn decreases inversely as the tenure of the contract increases and is high among short-term customers.



Now we need to understand the reason we want to do a churn analysis, so we compare the total revenue and revenue lost due to churn. We assume that we have 10% revenue growth due to new customers, which shows that we still have a net revenue loss of 7.8%, and it will be difficult to sustain the operations for long.

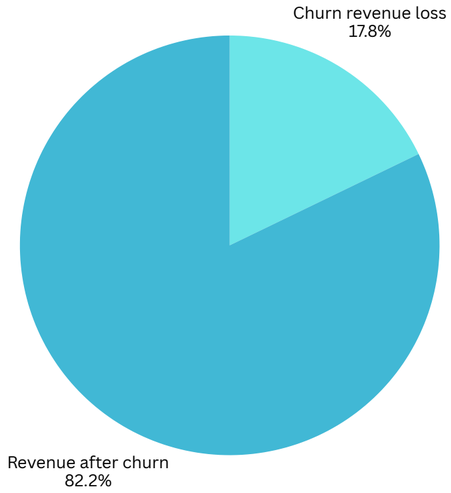


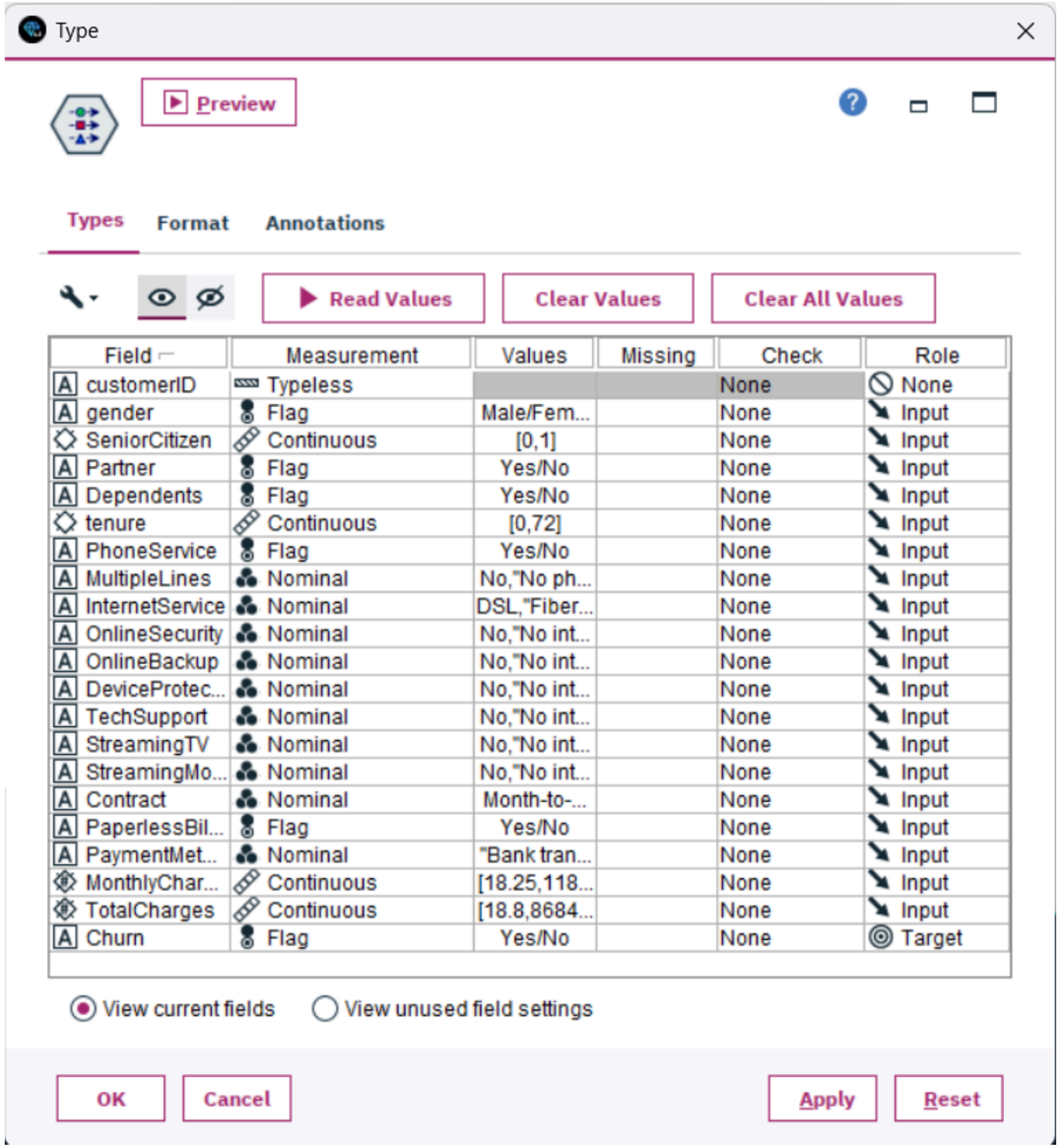
Table (3 fields, 1 records) #3

File Edit Generate

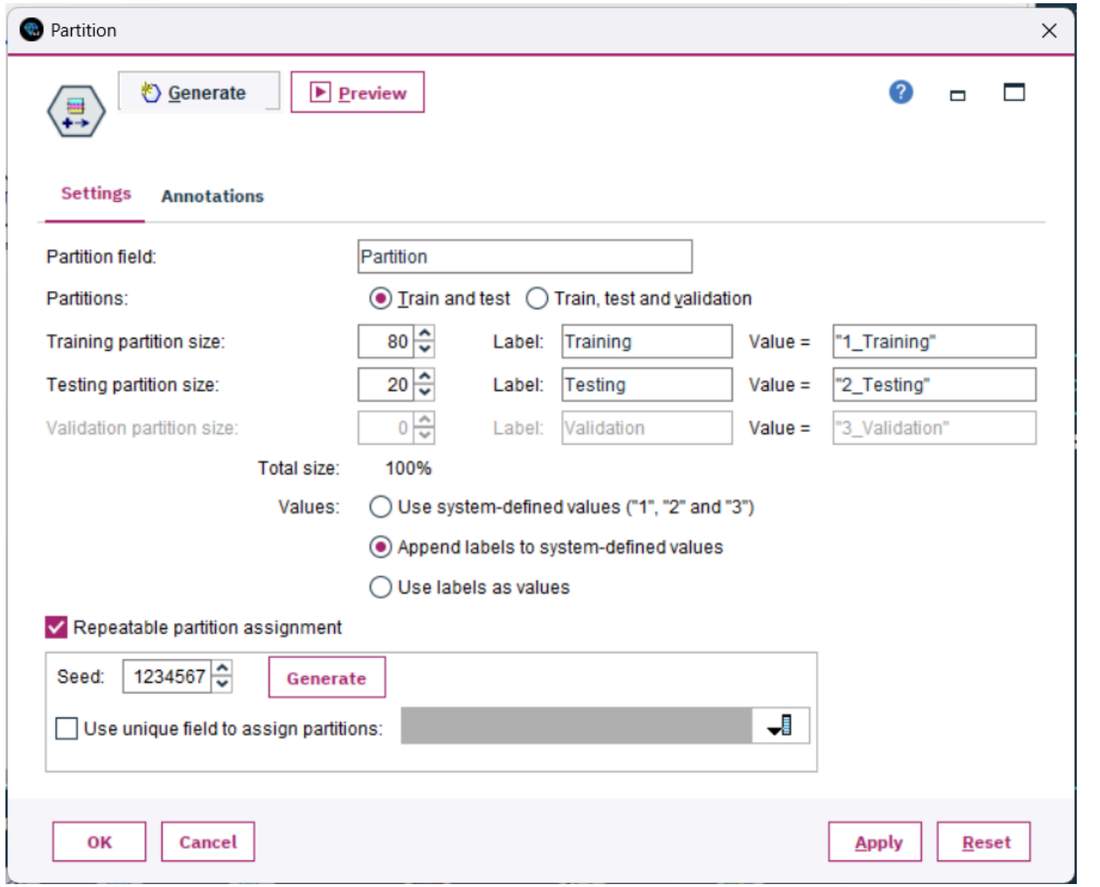
Table Annotations

	customerID_Count	TotalCharges_Sum	Record_Count
1	1869	2862926.900	1869

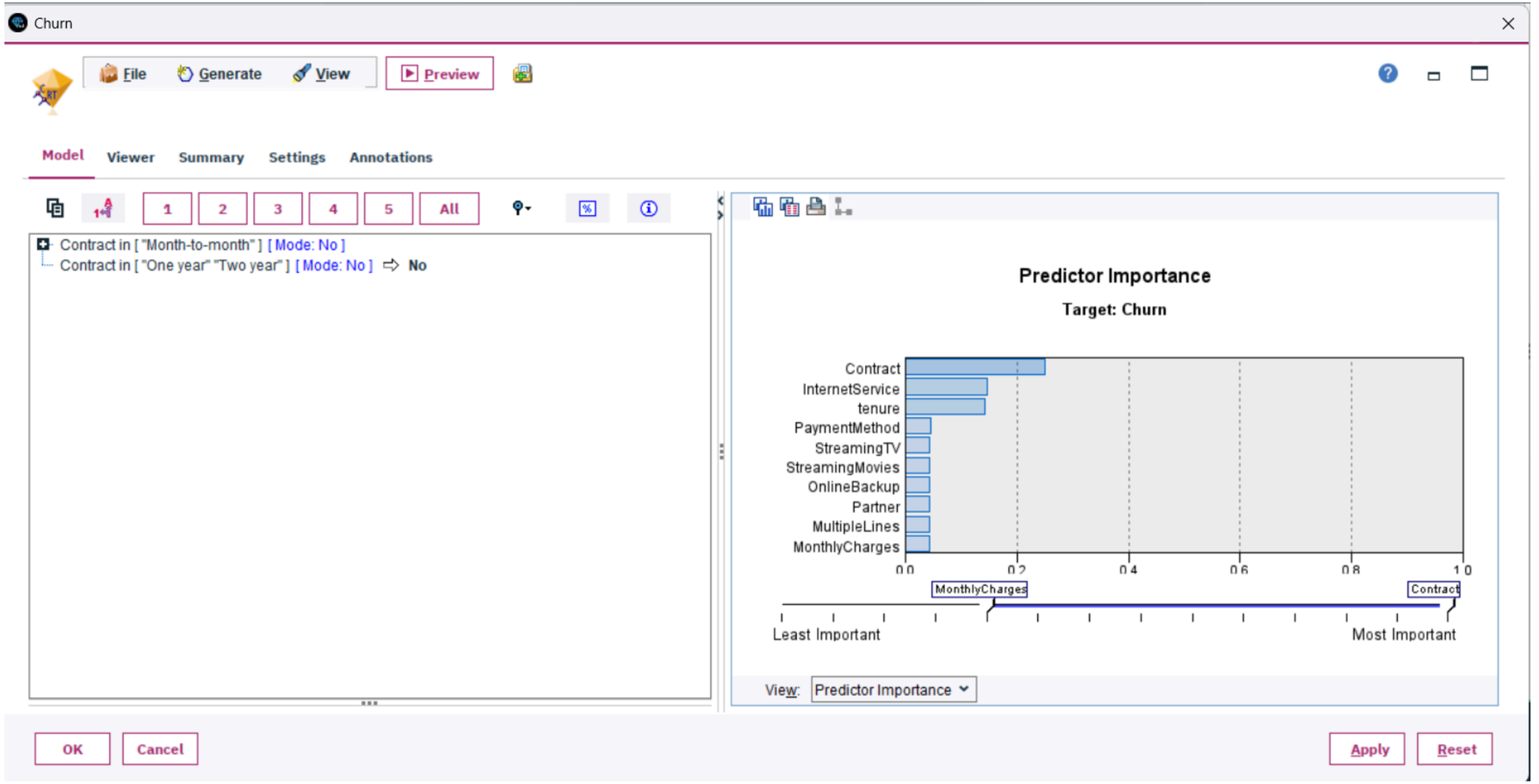
We will now use the type field to specify the various types of data fields into various categories such as input, none, and target.



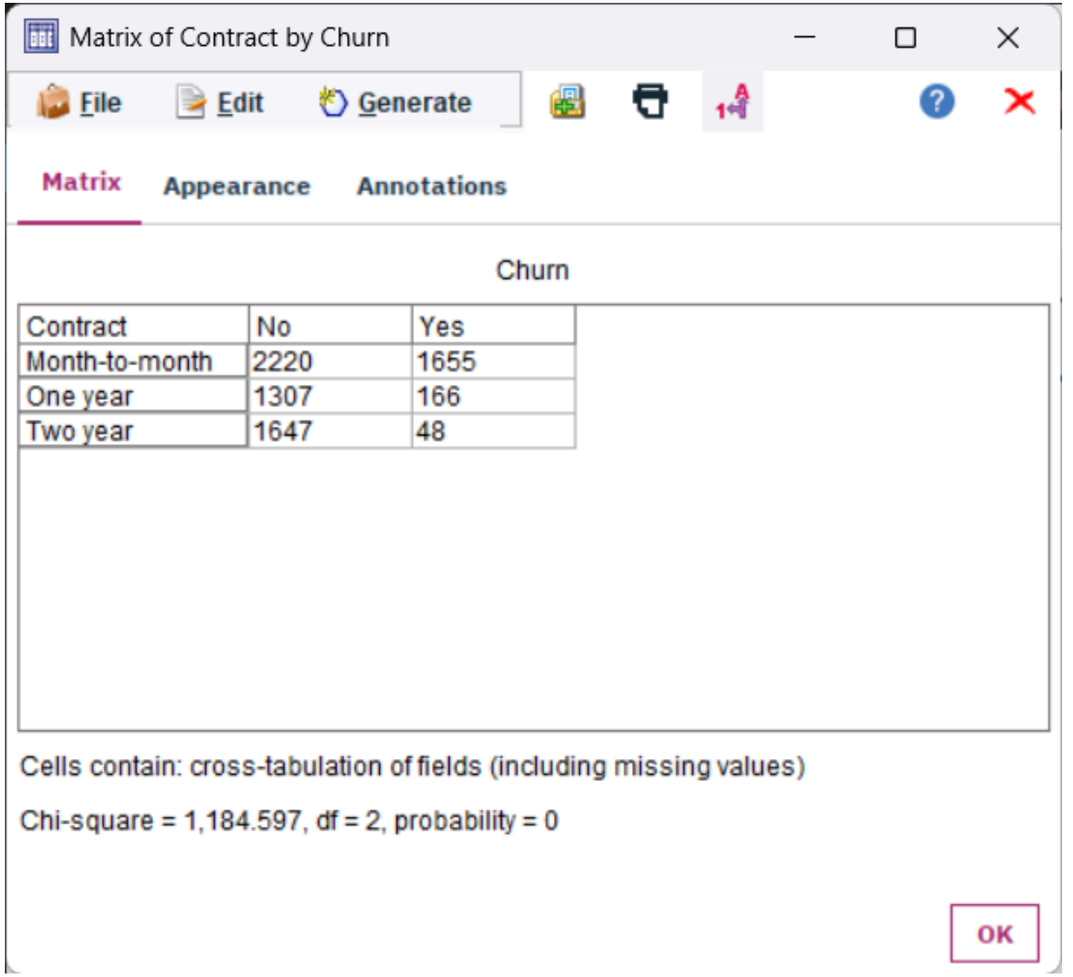
We can also use Partitioning to separate data into sets of training and testing data to test out the model accuracy.



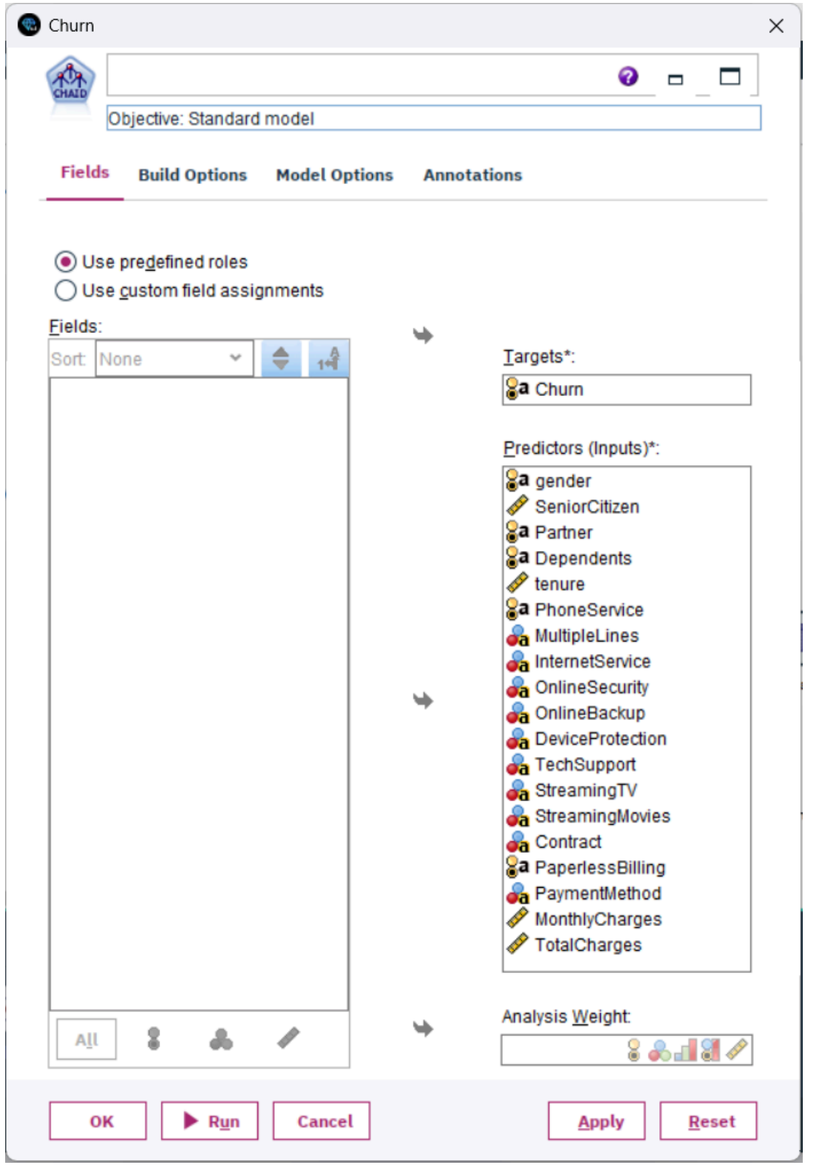
We use the CRT model to perform an analysis to find the factors affecting customer churn and make changes to improve these factors. For example, contract and internet service have a greater effect on customer churn than other fields such as monthly charges and payment method.



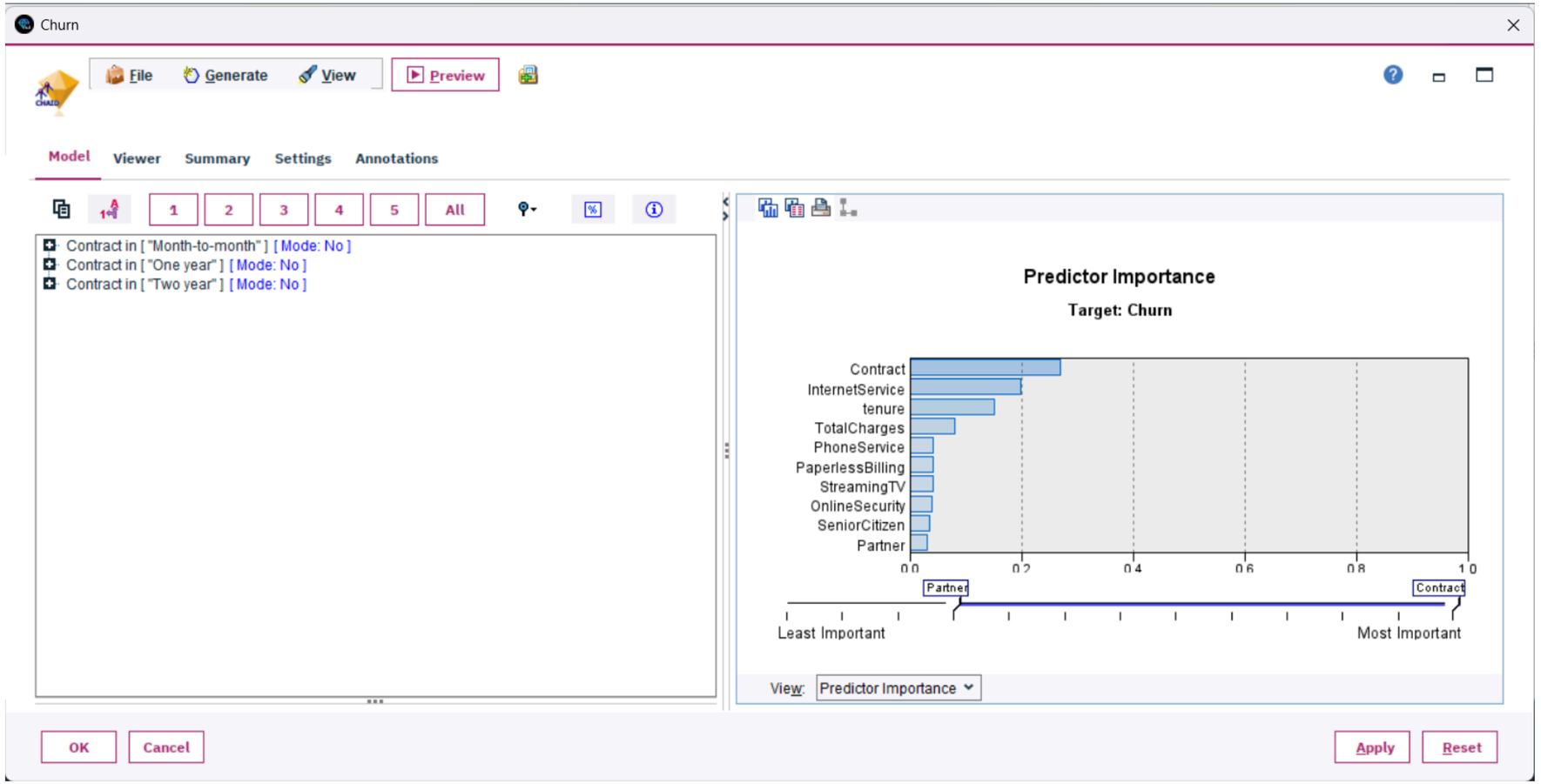
We now use an analysis node to analyse the output produced by the CRT model and analyse various fields.



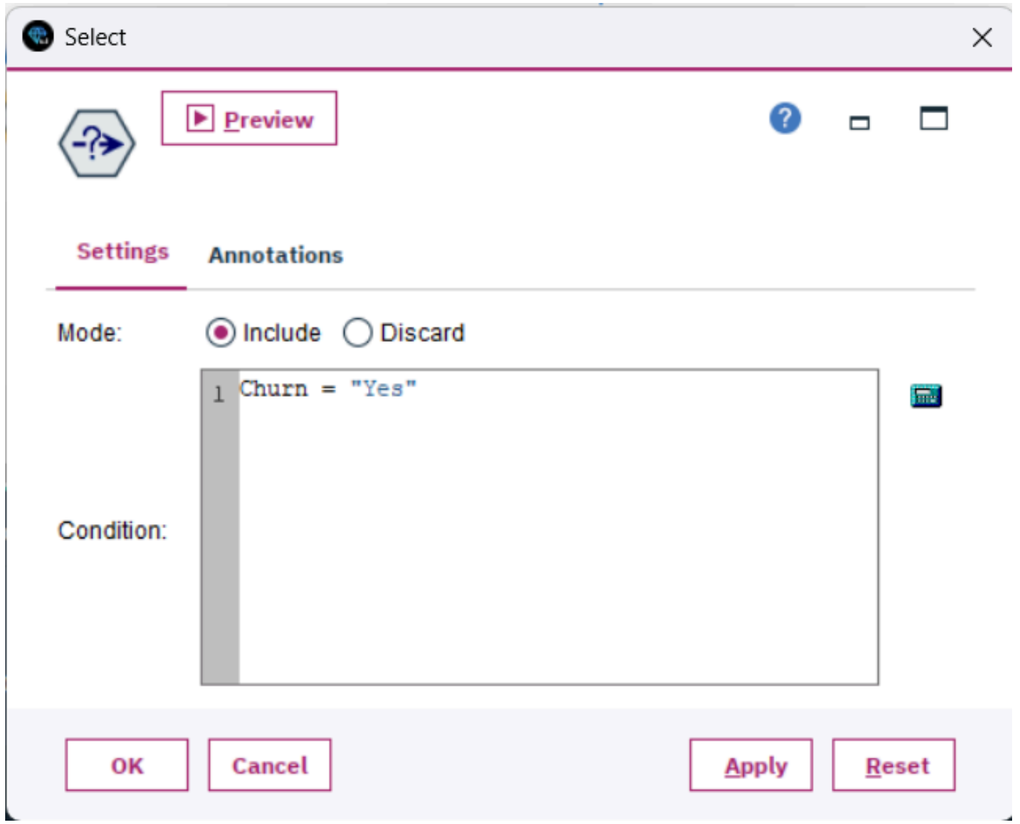
We will use the CHAID model to further improve our analysis, but first we have to specify the various parameters of the CHAID model in the setting of the model.



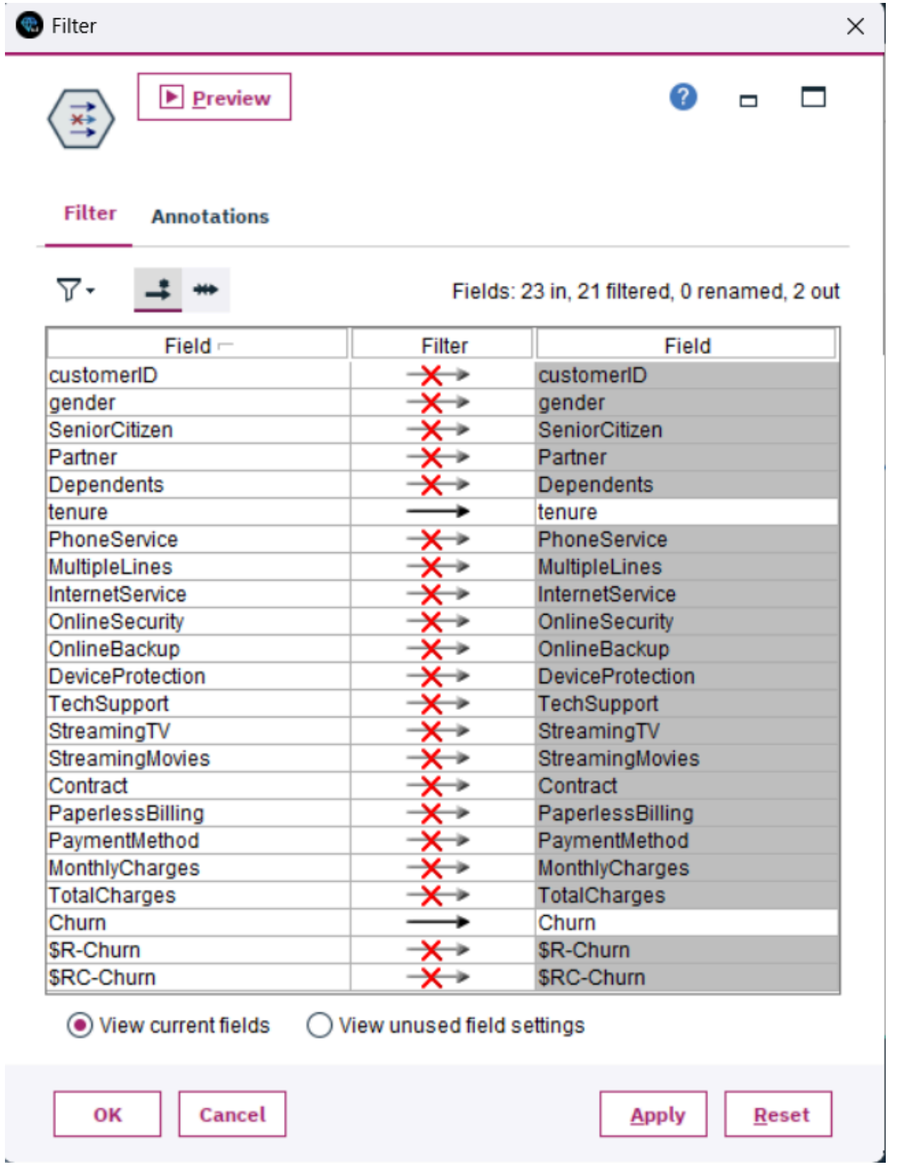
We use the CHAID model to perform an analysis to find the factors affecting customer churn and make changes to improve these factors. For example, contract and internet service have a greater effect on customer churn than other fields such as monthly charges and payment method.



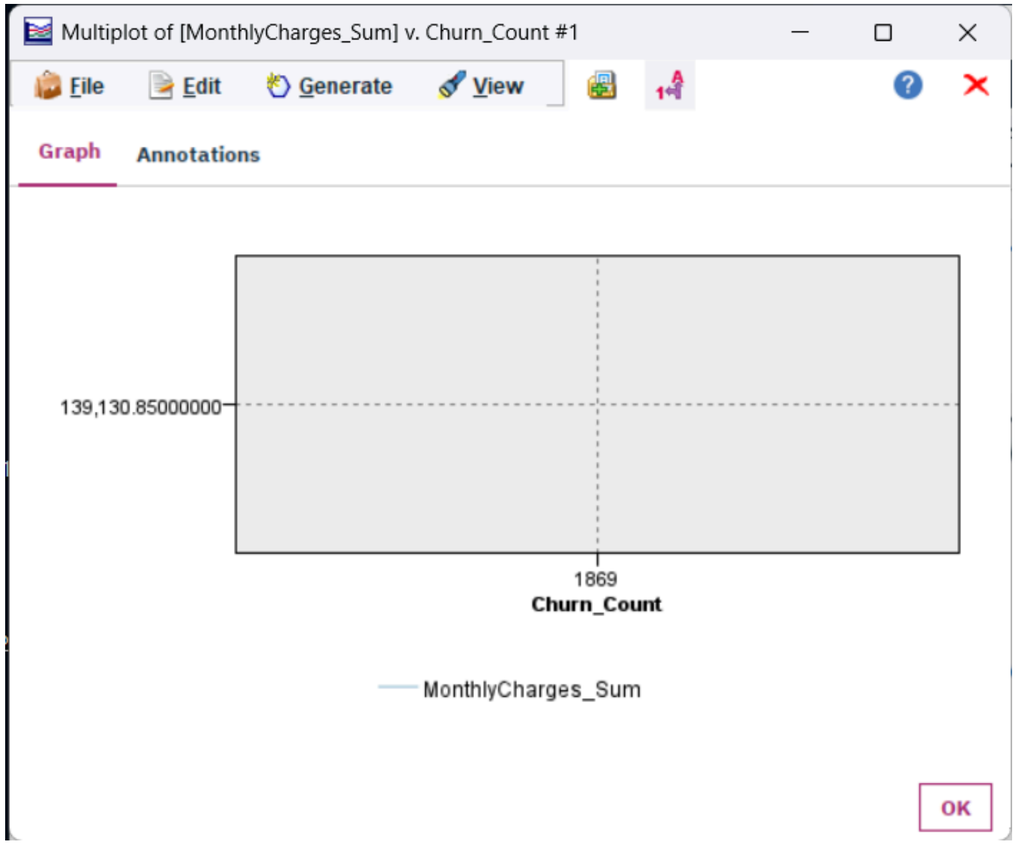
Add a select node to the output of the CHAID model and add various conditions to use the model output for further analysis of the model output and visualisation.



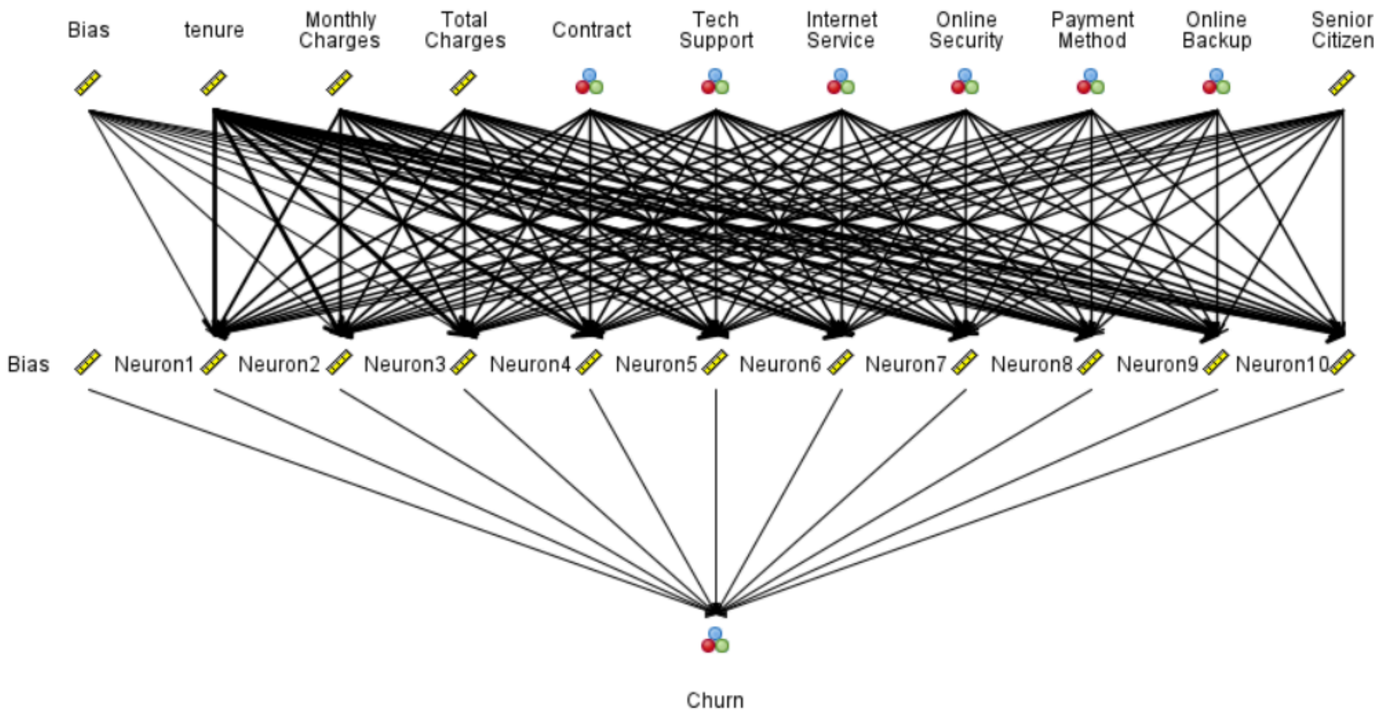
Use a filter node to select the fields you want to specify and analyse through various nodes such as visualisation.



Here we analyse the total monthly loss of revenue due to customer churn. We can change the parameters in the previous two nodes to adjust the fields and visualise it more thoroughly.



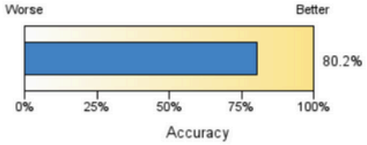
For future analysis of customer churn probability, we can add a neural network node to create a model and add a database node to predict customer churn probability as we add new customers and prepare a report and take actions to reduce churn.



Here are the model specifications for 80% accuracy, which can be improved with changes in data fields and the dataset.

Model Summary

Target	Churn
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	10



Classification for Churn

Overall Percent Correct = 80.3%

Observed	Predicted	
	No	Yes
No	90.2%	9.8%
Yes	46.2%	53.8%

Row Percent
100.00
80.00
60.00
40.00
20.00
0.00

Now we can produce an output of our churn prediction model analysis to our specified file.