BABU BANARASI DAS UNIVERSITY

School of Computer Applications (DS & AI)

Predictive Analysis

Aditya Yadav ( 1230258037 )

Amit Yadav (1230258036 )

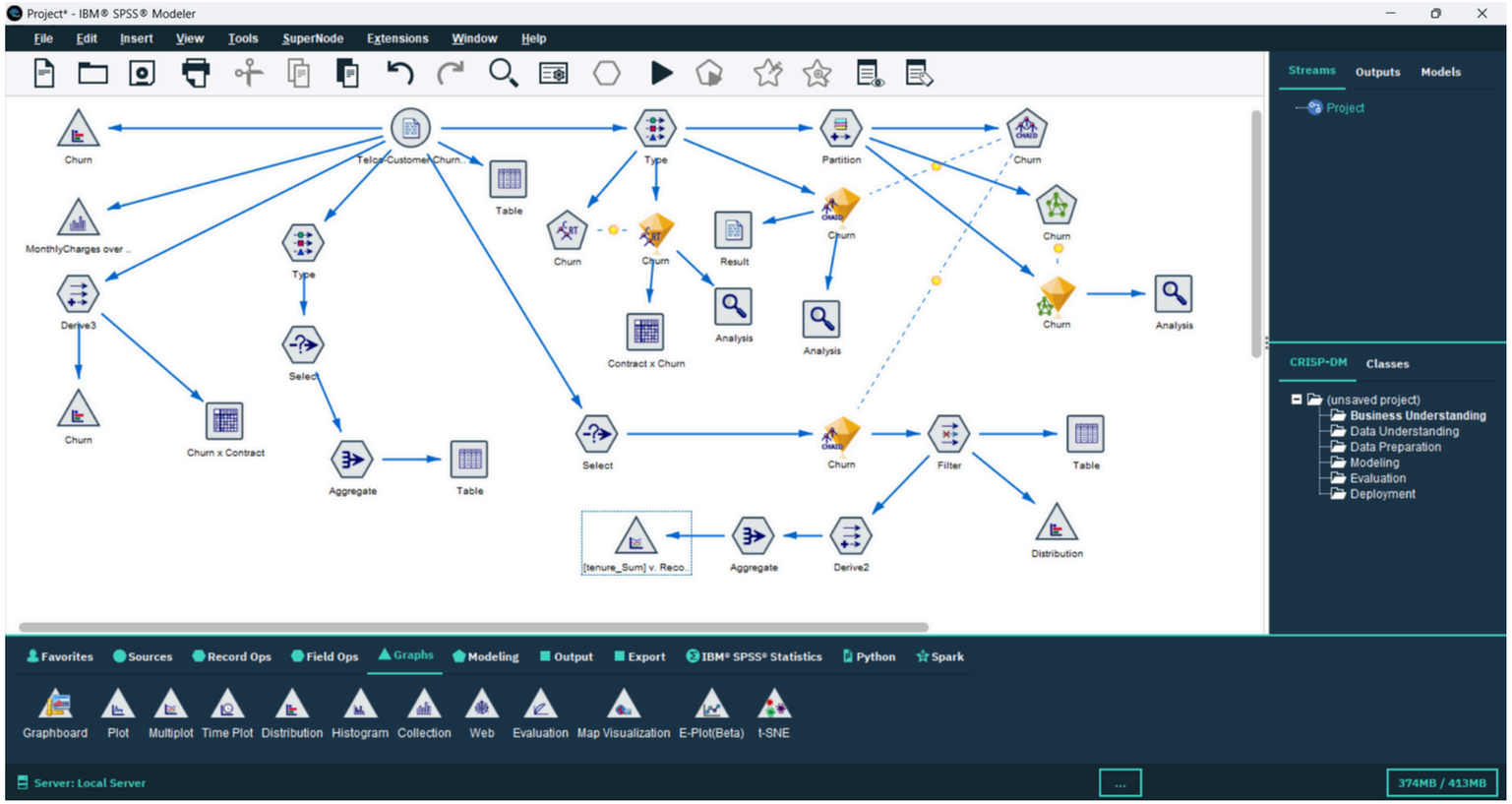
Aditya Srivastava ( 1230258065 )

Aditi Singh (1230258025 )

Submitted to:  Mr. Vikash (IBM)

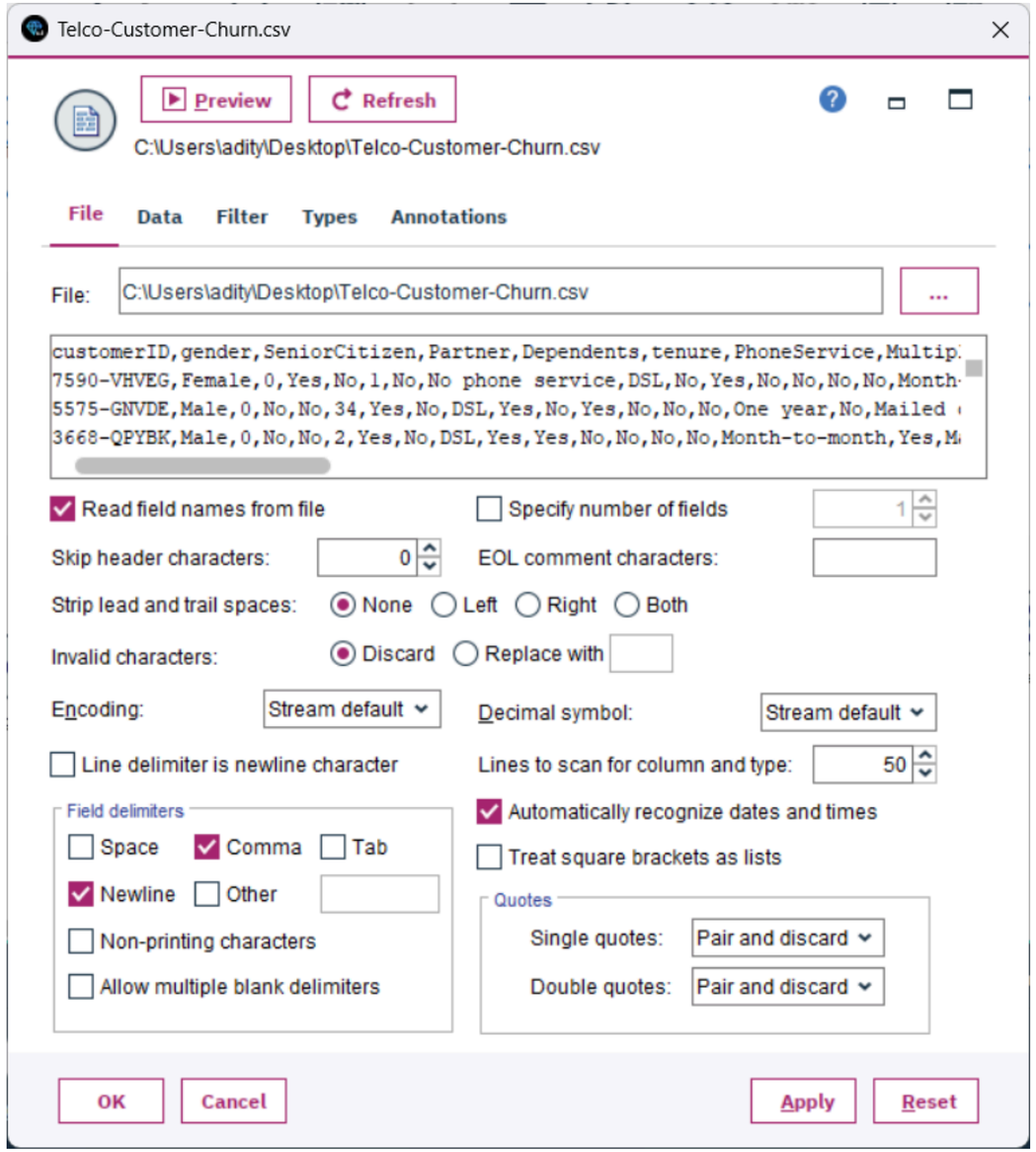
# Churn analysis of telecom customers using IBM SPSS Modeler on an open source dataset

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

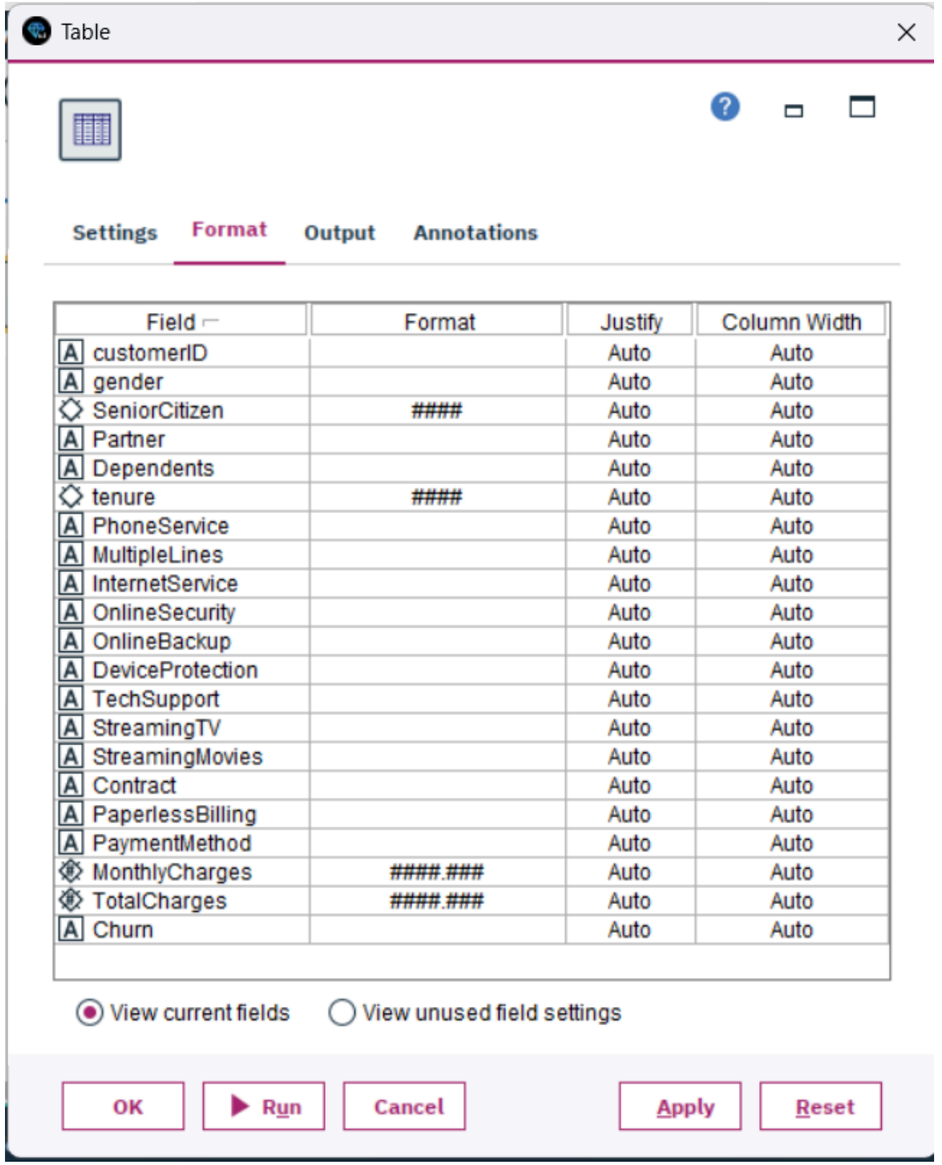



Now we will Understand the step-by-step procedure of creating this.

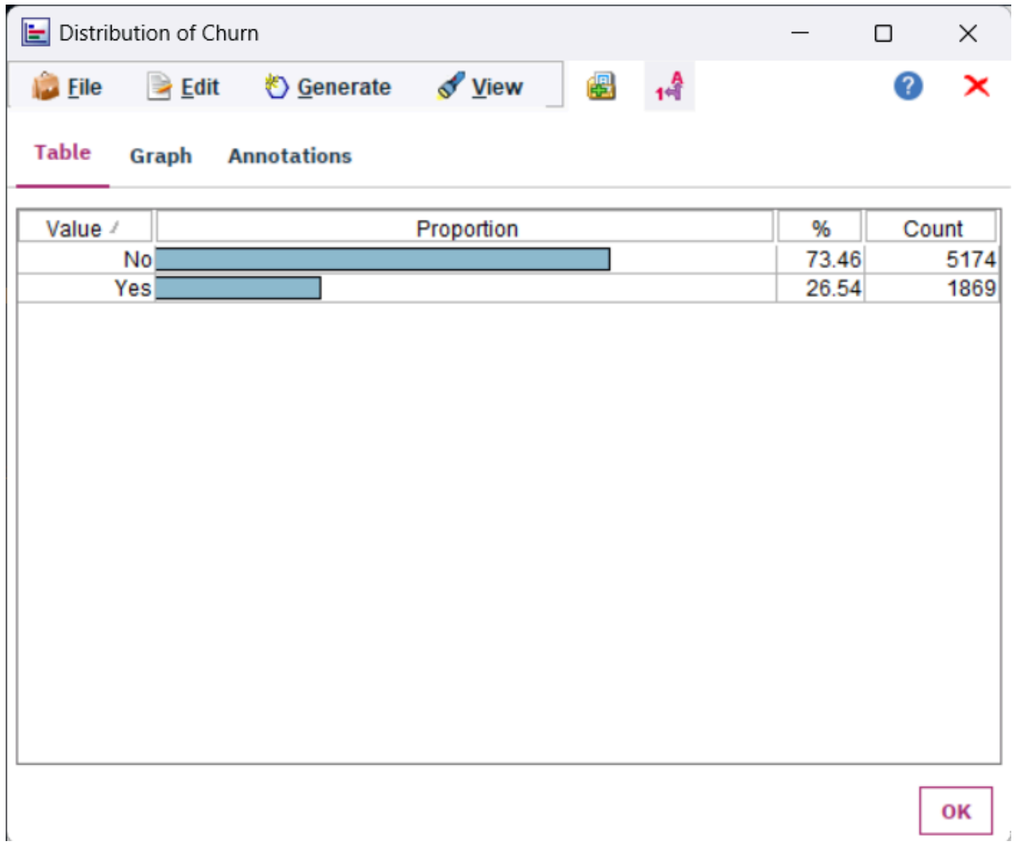
To analyse the customer churn, we first need to have a proper dataset and upload it to IBM SPSS Modeler using the Var.file from Sources. The data is in a CSV file format. It contains 21 fields which are: customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService,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



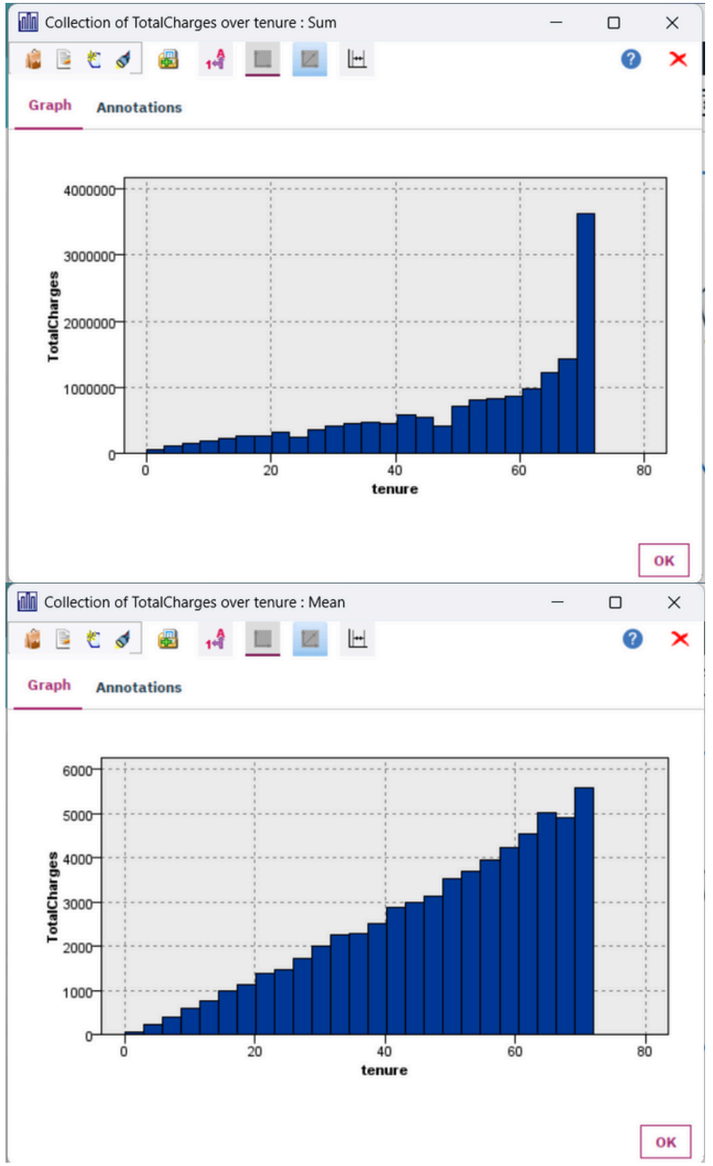
Now, we can utilize the Table view from the Output to examine and assess the data along with its types, allowing for any necessary adjustments before the analysis starts. This also provides us with a comprehensive overview of the data.



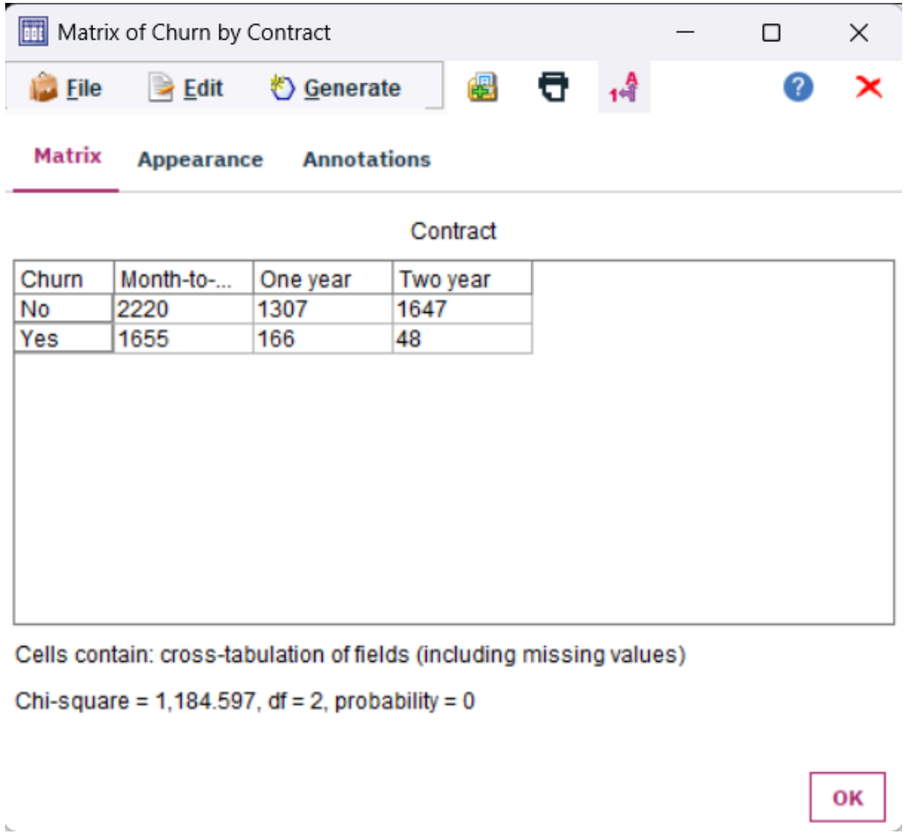
We now use Graph view to analyse the rate of churn of customers during a particular time period.



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 utilize a matrix derived from our output to illustrate the intricate relationships within our dataset. It reveals that the churn rate decreases inversely as contract tenure increases, with a notably high rate among short-term customers.



To effectively conduct a churn analysis, it's essential to grasp the purpose behind it: comparing total revenue with the revenue lost due to churn. We estimate a 10% revenue growth from new customers, yet we still face a net revenue loss of 7.8%. This situation indicates that sustaining operations in the long run will be challenging.

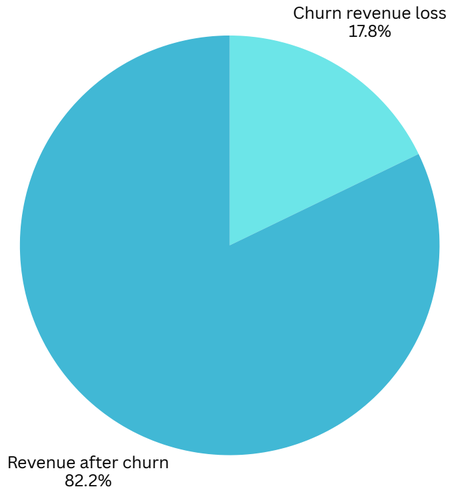


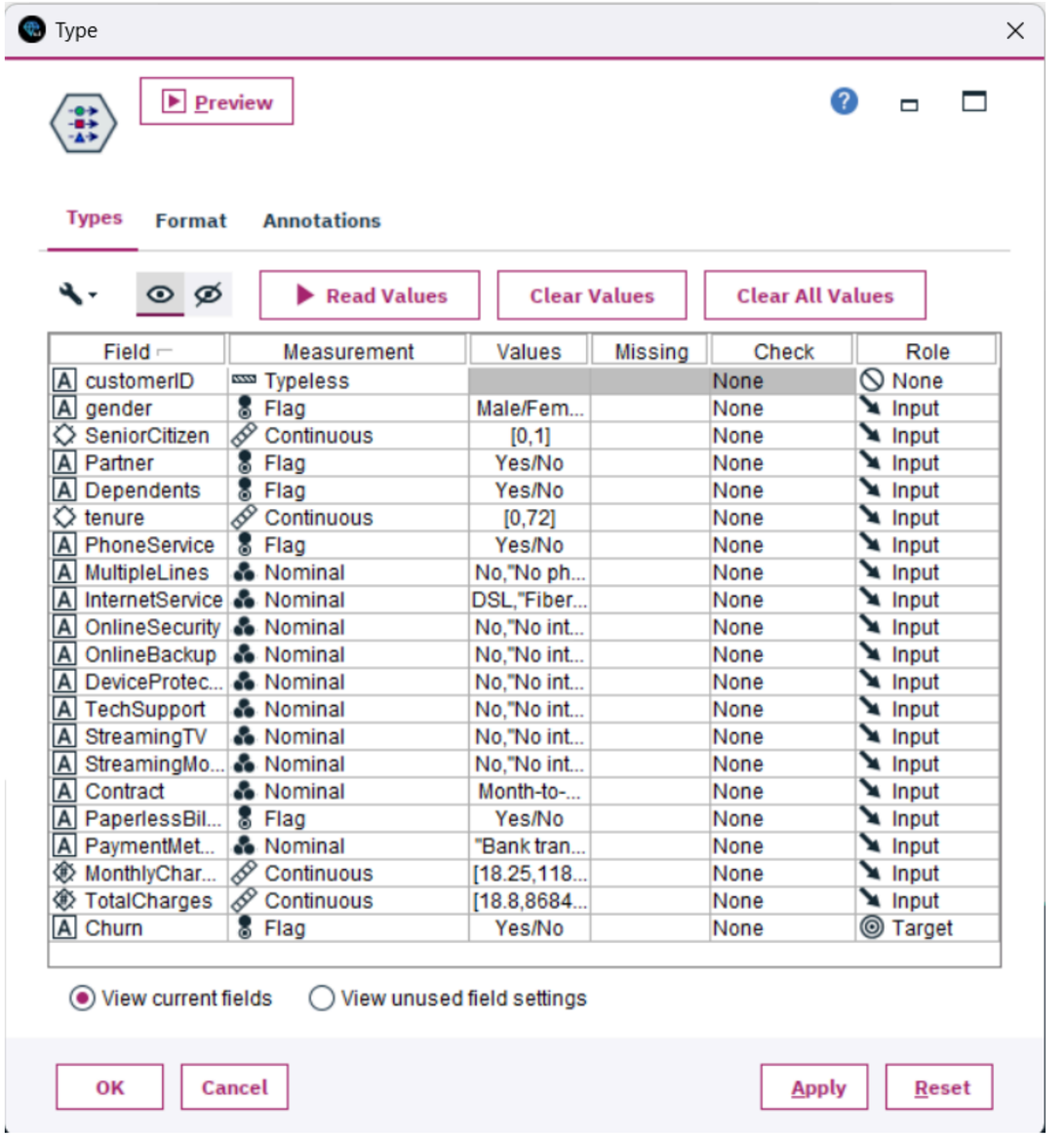

Table (3 fields, 1 records) #3

File Edit Generate

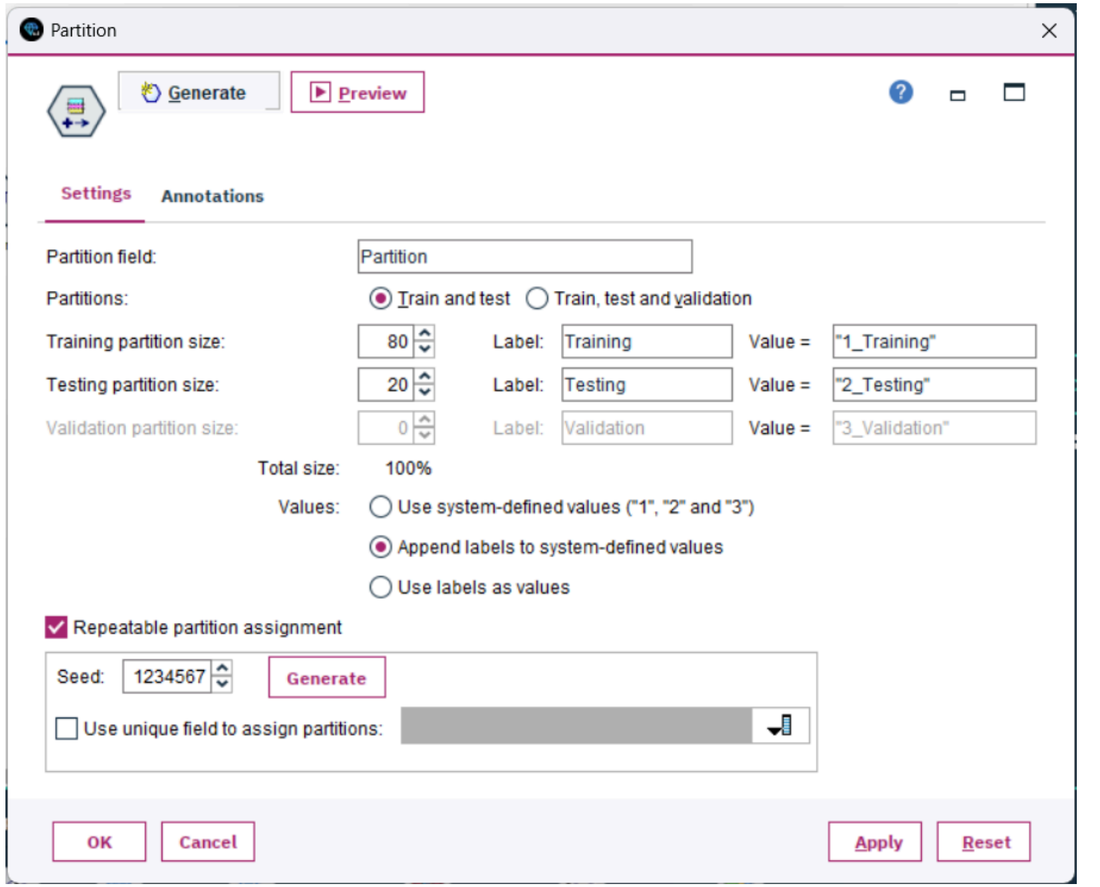
Table Annotations

| | customerID_Count | TotalCharges_Sum | Record_Count |
|---|---|---|---|
| 1 | 1869 | 2862926.900 | 1869 |

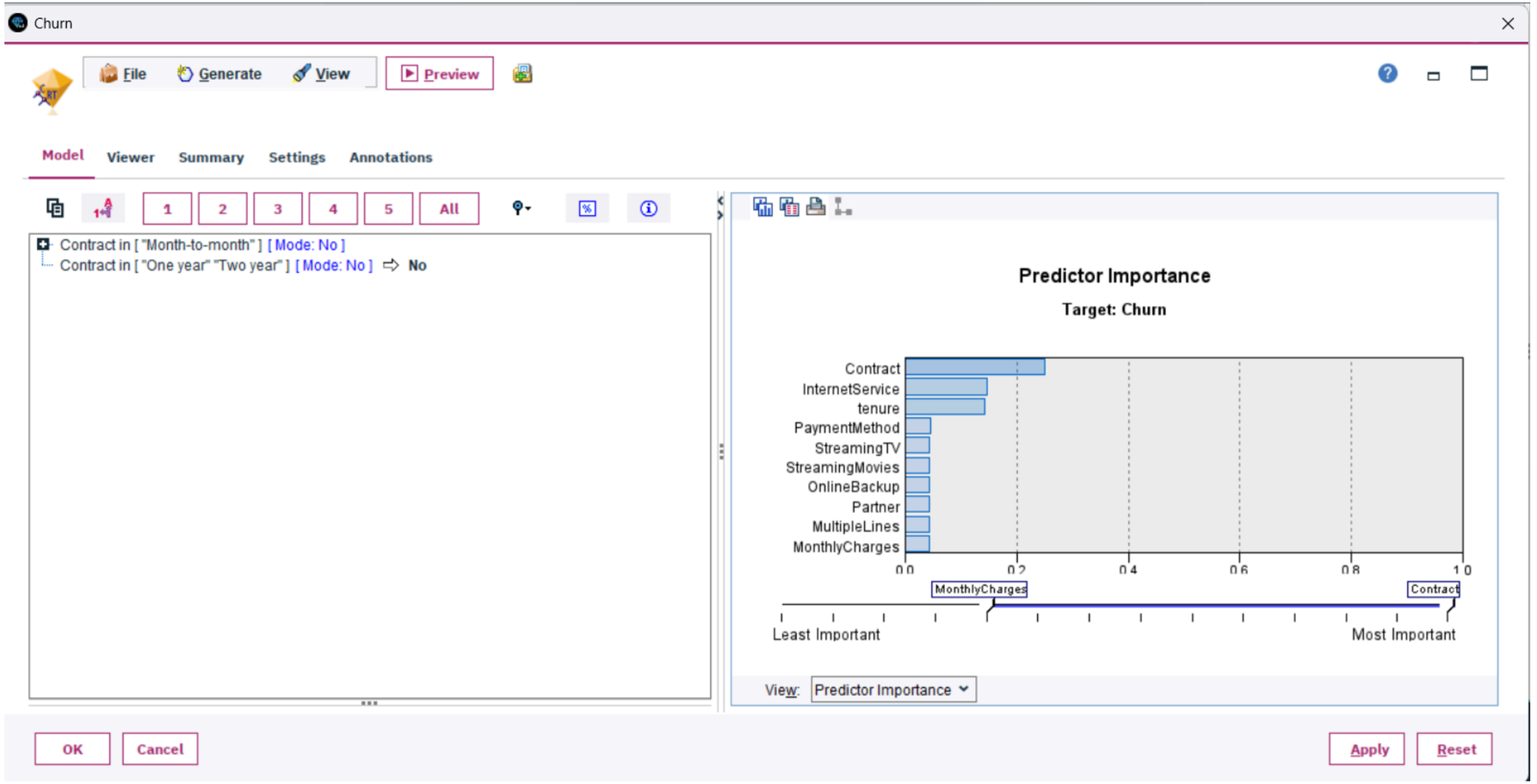
We will now utilize the type field to categorize the different types of data fields into various classifications, including input, none, and target.



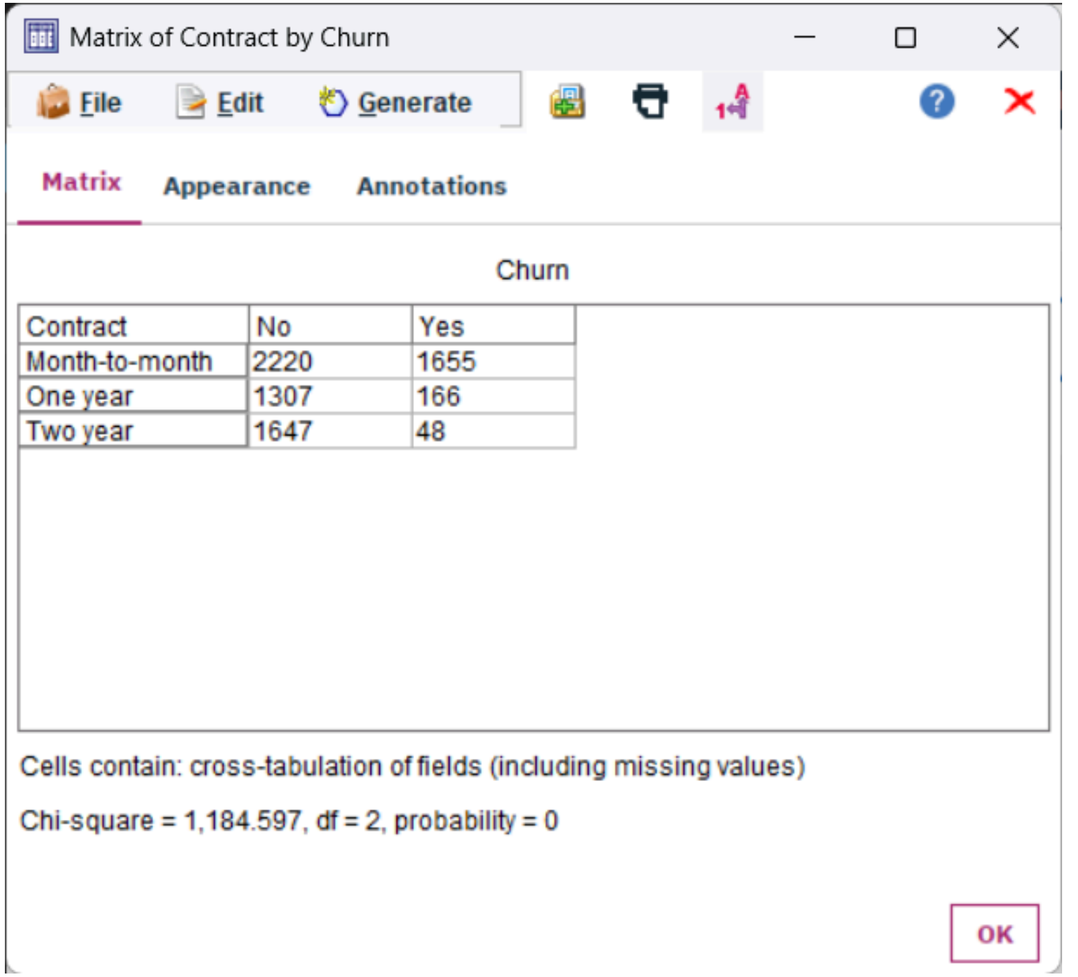
We can also use Partitioning to separate data into sets of training and testing data to test out the model accuracy.



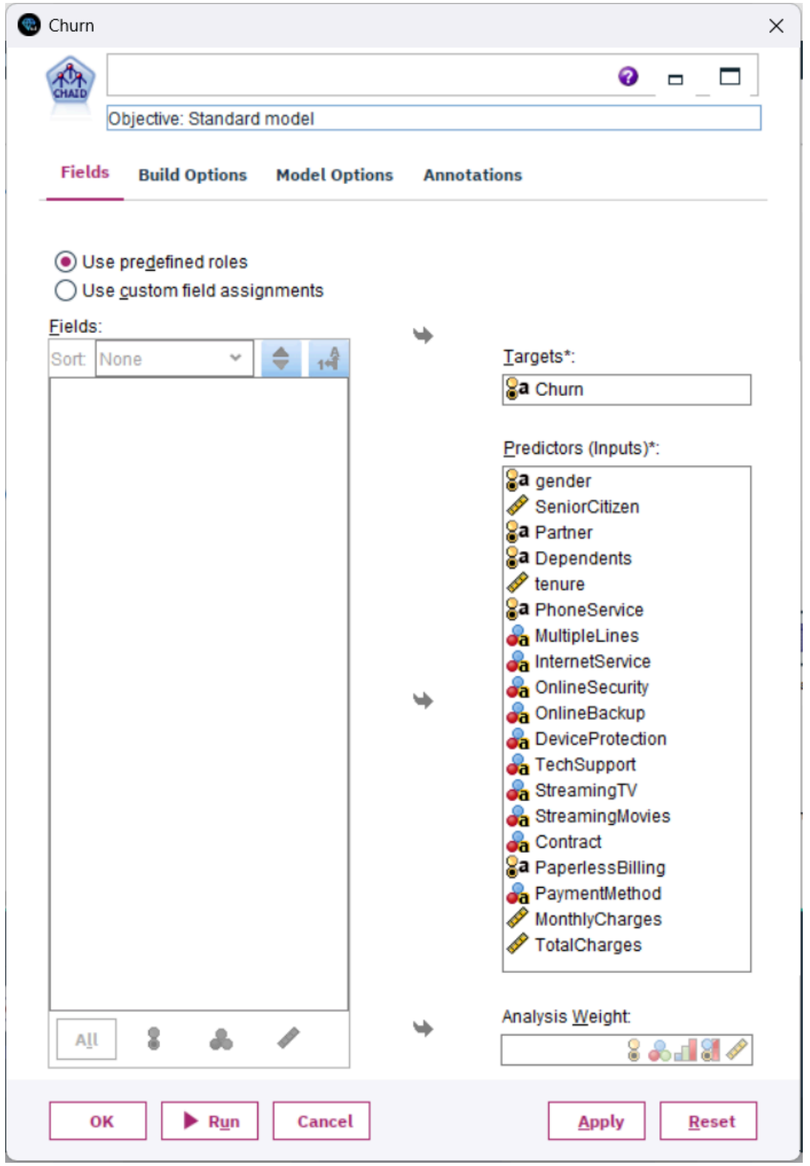
We utilize the CRT model to analyze the factors influencing customer churn and implement improvements accordingly. For instance, aspects like contract terms and internet service have a more significant impact on customer churn compared to other elements, such as monthly charges and payment methods.



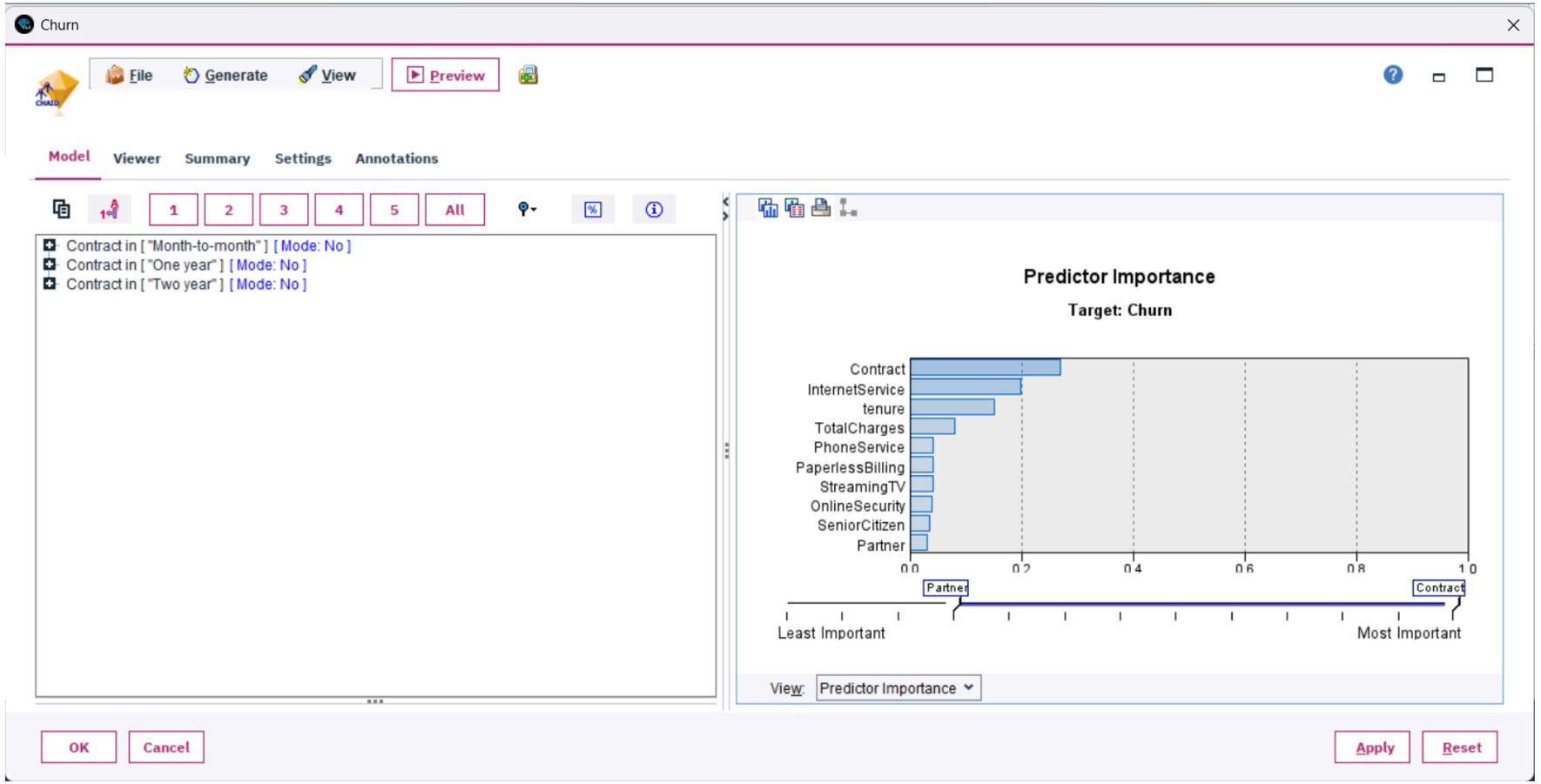
We now employ an analysis node to examine the output generated by the CRT model and evaluate various fields.



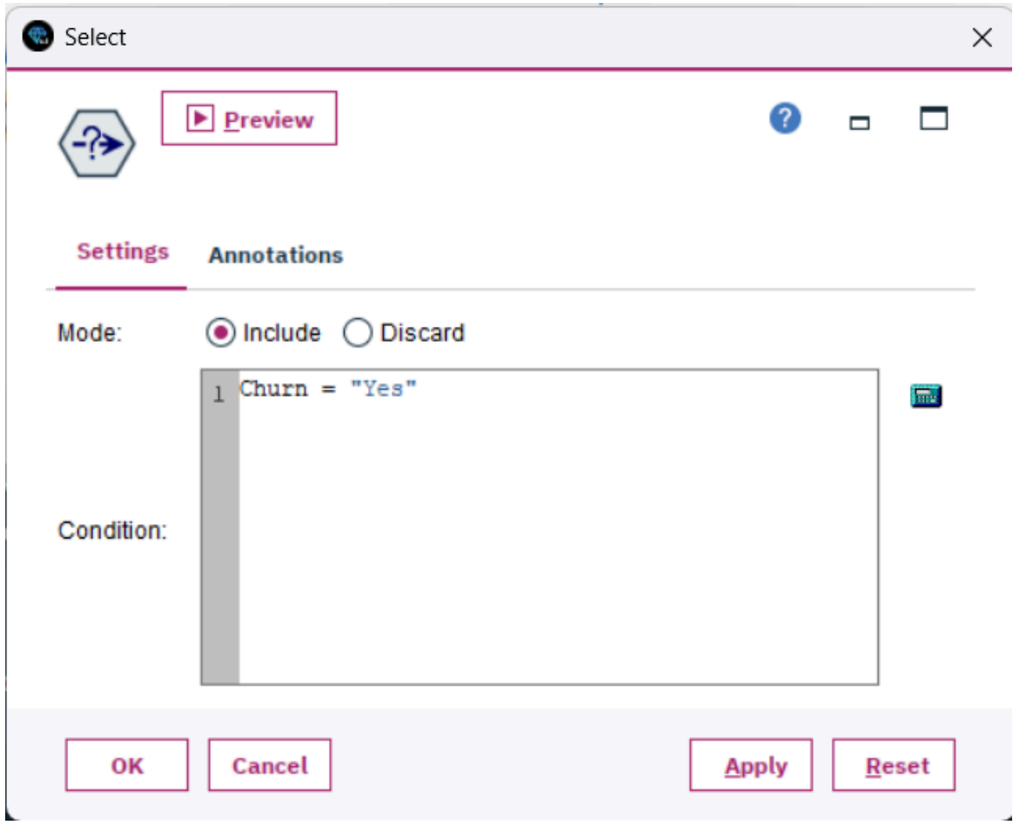
We will employ the CHAID model to enhance our analysis further; however, we first need to define the different parameters of the CHAID model within its context.



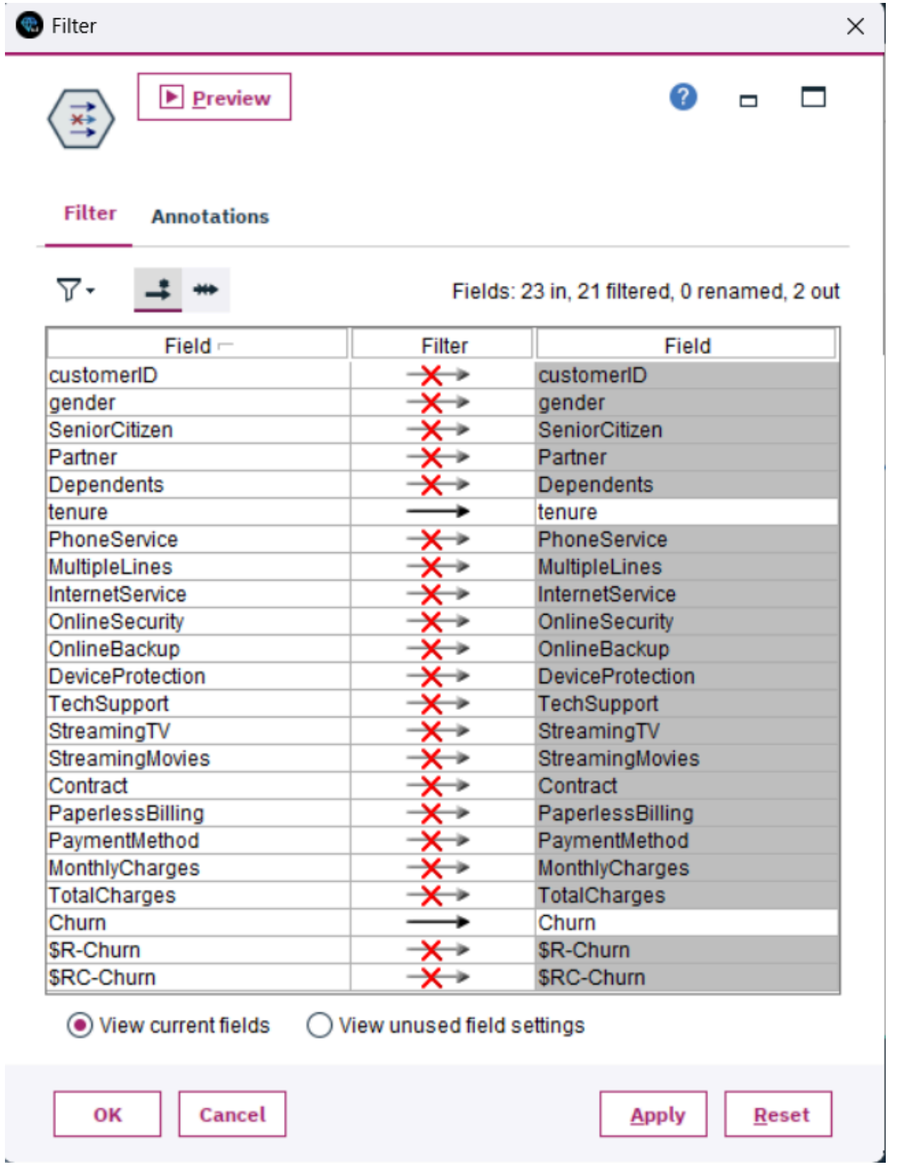
We use the CHAID model to perform an analysis to find the factors affecting customer churn and make changes to improve these factors. For example, contract and internet service have a greater effect on customer churn than other fields such as monthly charges and payment method.



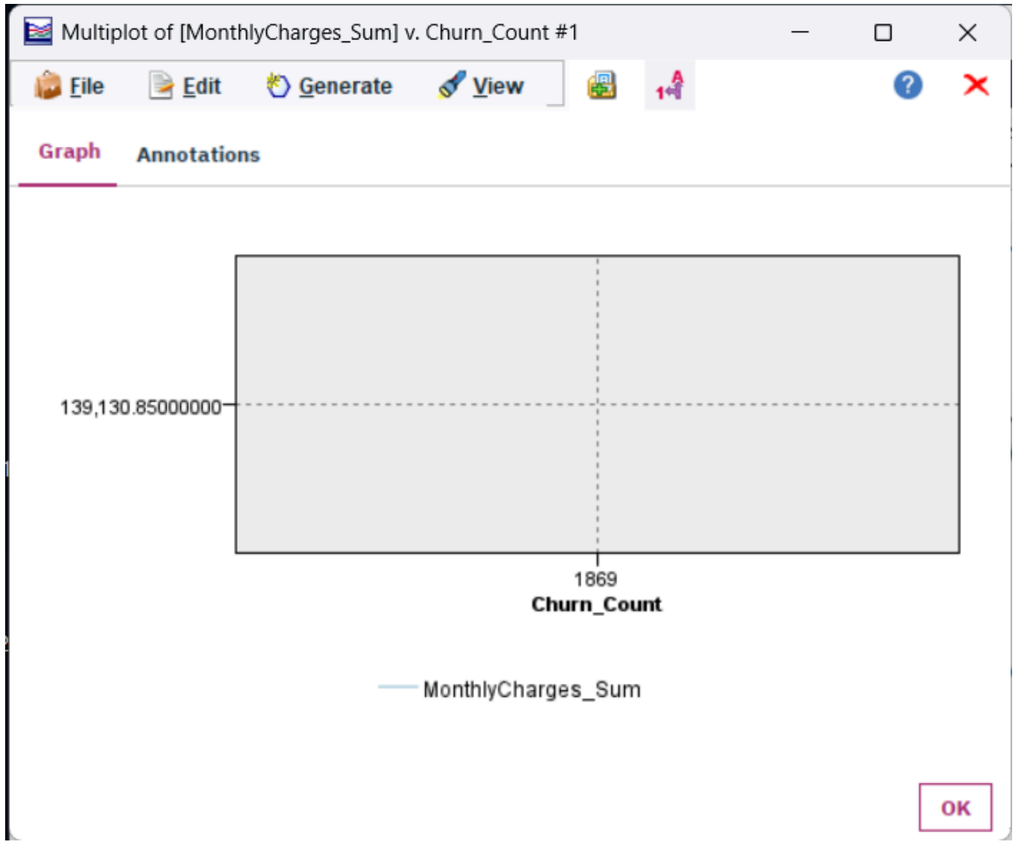
Add a select node to the output of the CHAID model and add various conditions to use the model output for further analysis of the model output and visualisation.



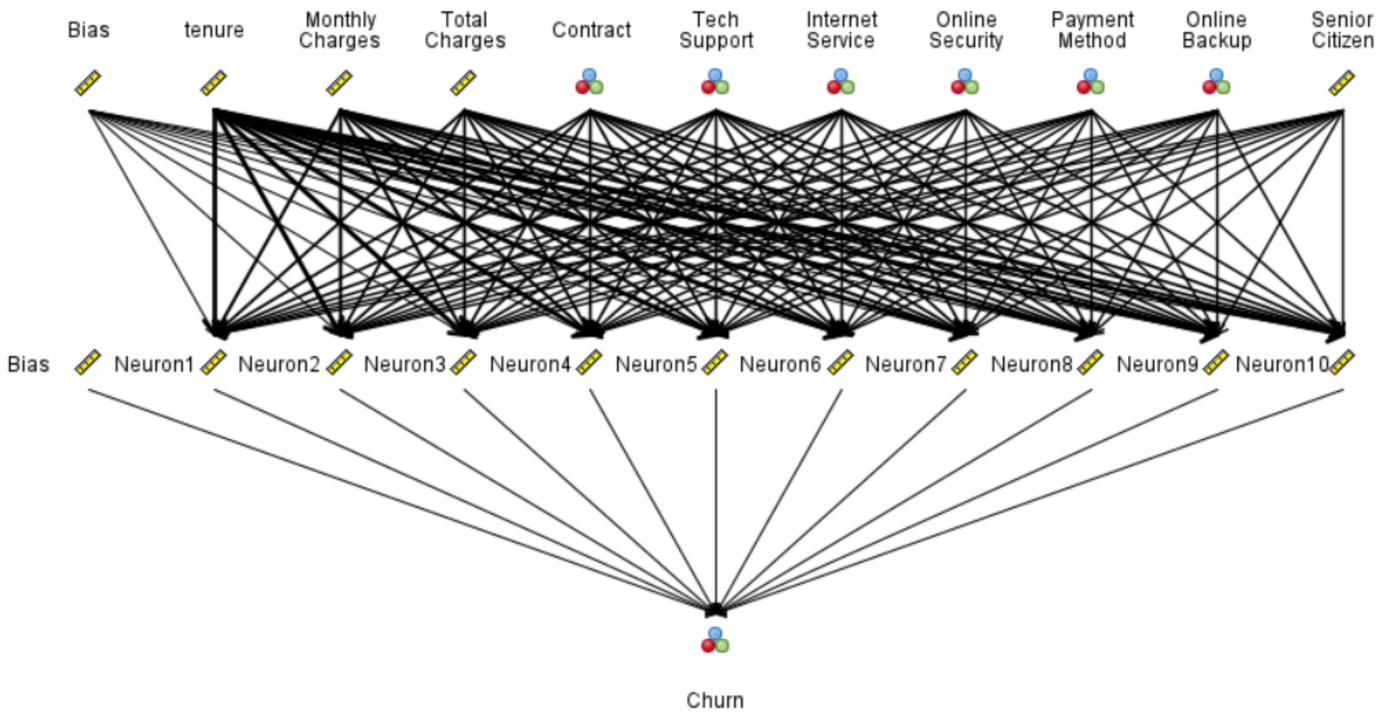
Use a filter node to select the fields you want to specify and analyse through various nodes such as visualisation.



Here, we examine the overall monthly revenue loss resulting from customer churn. By adjusting the parameters in the previous two nodes, we can modify the fields and gain a more comprehensive visualization.



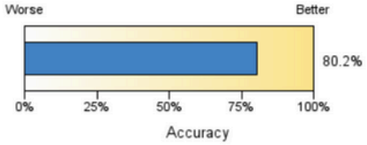
For future analysis of customer churn probability, we can add a neural network node to create a model and add a database node to predict customer churn probability as we add new customers and prepare a report and take actions to reduce churn.



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 generate the output from our churn prediction model analysis to a designated file. This will enable us to review and share insights with stakeholders, providing a clearer understanding of customer behavior. The findings from this analysis are crucial for identifying patterns and trends that can refine our strategies to enhance customer retention. By analyzing the data, we can identify the factors leading to churn and create targeted interventions to boost customer satisfaction and loyalty. Additionally, this output lays the groundwork for future iterations of the model, facilitating continuous improvement and more precise predictions.