



**School of Computer Application (DS & AI)**

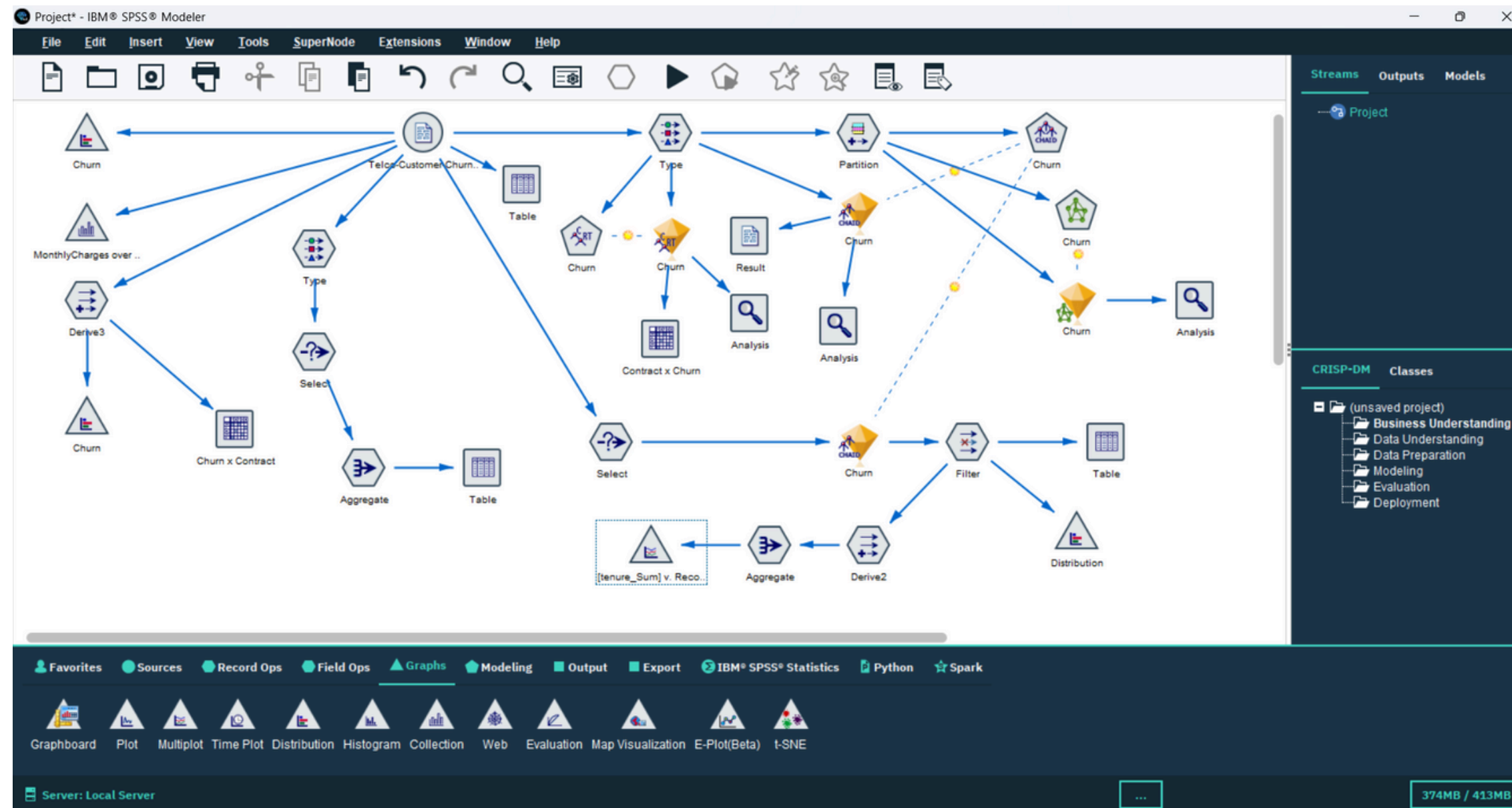
**Predictive Analysis**

**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 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,Churn

The screenshot shows the 'File' dialog box in IBM SPSS Modeler for the file 'Telco-Customer-Churn.csv'. The dialog is titled 'Telco-Customer-Churn.csv' and has a close button (X) in the top right corner. Below the title bar, there are icons for 'Preview' and 'Refresh', and a question mark icon. The file path is 'C:\Users\lady\Desktop\Telco-Customer-Churn.csv'. The 'File' tab is selected, showing a list of fields: 'customerID,gender,SeniorCitizen,Partner,Dependents,tenure,PhoneService,MultipleLines,InternetService,OnlineSecurity,OnlineBackup,DeviceProtection,TechSupport,StreamingTV,StreamingMovies,Contract,PaperlessBilling,PaymentMethod,MonthlyCharges,TotalCharges,Churn'. Below the field list, there are various options for file handling: 'Read field names from file' (checked), 'Specify number of fields' (1), 'Skip header characters' (0), 'EOL comment characters' (empty), 'Strip lead and trail spaces' (None selected), 'Invalid characters' (Discard selected), 'Encoding' (Stream default), 'Decimal symbol' (Stream default), 'Line delimiter is newline character' (unchecked), 'Lines to scan for column and type' (50), 'Field delimiters' (Comma checked, Newline checked), 'Automatically recognize dates and times' (checked), 'Treat square brackets as lists' (unchecked), and 'Quotes' (Single quotes: Pair and discard, Double quotes: Pair and discard). At the bottom, there are buttons for 'OK', 'Cancel', 'Apply', and 'Reset'.

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.

Table

?

SettingsFormatOutputAnnotations

Field	Format	Justify	Column Width
<input type="checkbox"/> customerID		Auto	Auto
<input type="checkbox"/> gender		Auto	Auto
<input checked="" type="checkbox"/> SeniorCitizen	####	Auto	Auto
<input type="checkbox"/> Partner		Auto	Auto
<input type="checkbox"/> Dependents		Auto	Auto
<input checked="" type="checkbox"/> tenure	####	Auto	Auto
<input type="checkbox"/> PhoneService		Auto	Auto
<input type="checkbox"/> MultipleLines		Auto	Auto
<input type="checkbox"/> InternetService		Auto	Auto
<input type="checkbox"/> OnlineSecurity		Auto	Auto
<input type="checkbox"/> OnlineBackup		Auto	Auto
<input type="checkbox"/> DeviceProtection		Auto	Auto
<input type="checkbox"/> TechSupport		Auto	Auto
<input type="checkbox"/> StreamingTV		Auto	Auto
<input type="checkbox"/> StreamingMovies		Auto	Auto
<input type="checkbox"/> Contract		Auto	Auto
<input type="checkbox"/> PaperlessBilling		Auto	Auto
<input type="checkbox"/> PaymentMethod		Auto	Auto
<input checked="" type="checkbox"/> MonthlyCharges	####.###	Auto	Auto
<input checked="" type="checkbox"/> TotalCharges	####.###	Auto	Auto
<input type="checkbox"/> Churn		Auto	Auto

☒ View current fields☐ View unused field settings

OK

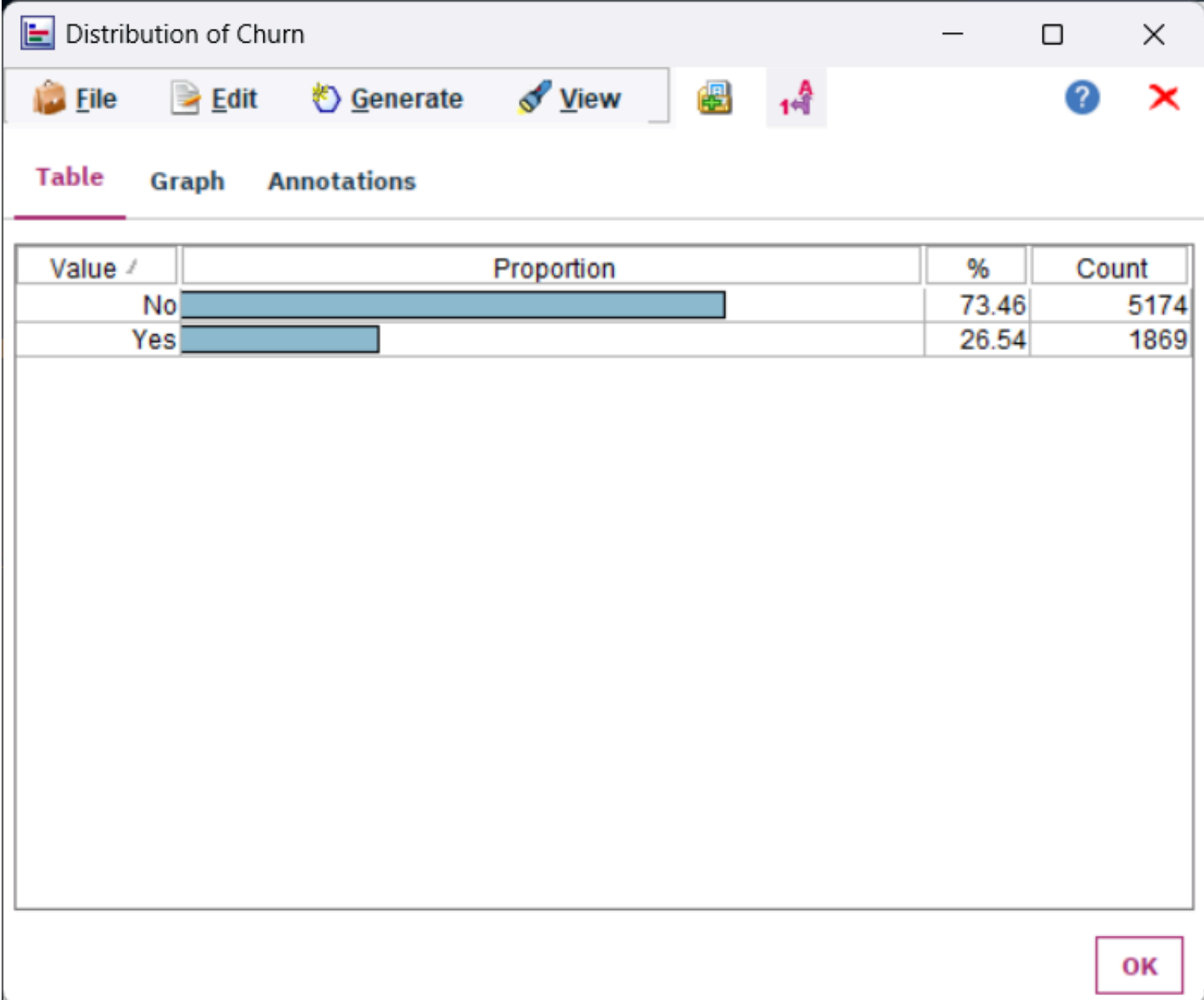
▶ Run

Cancel

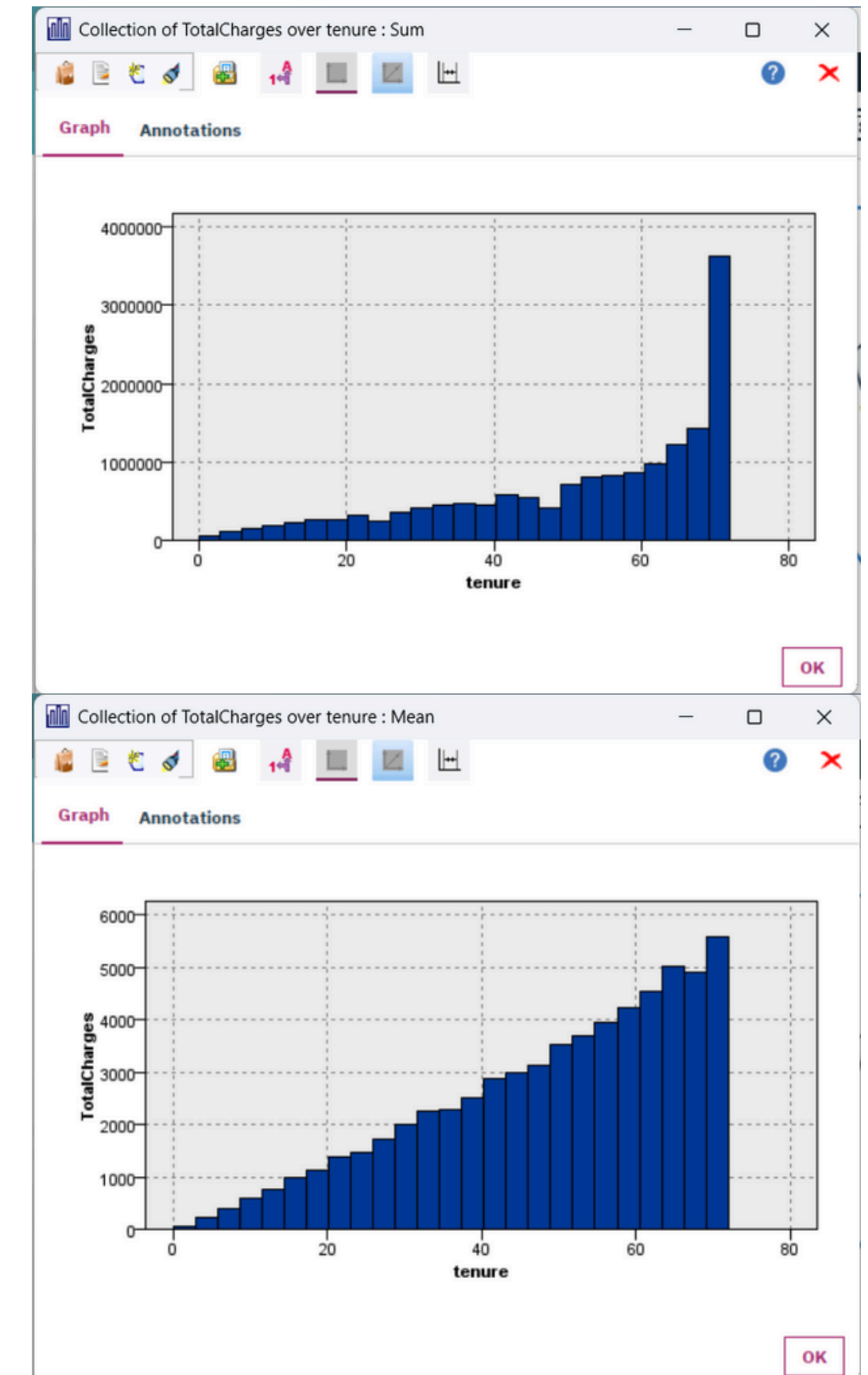
Apply

Reset

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

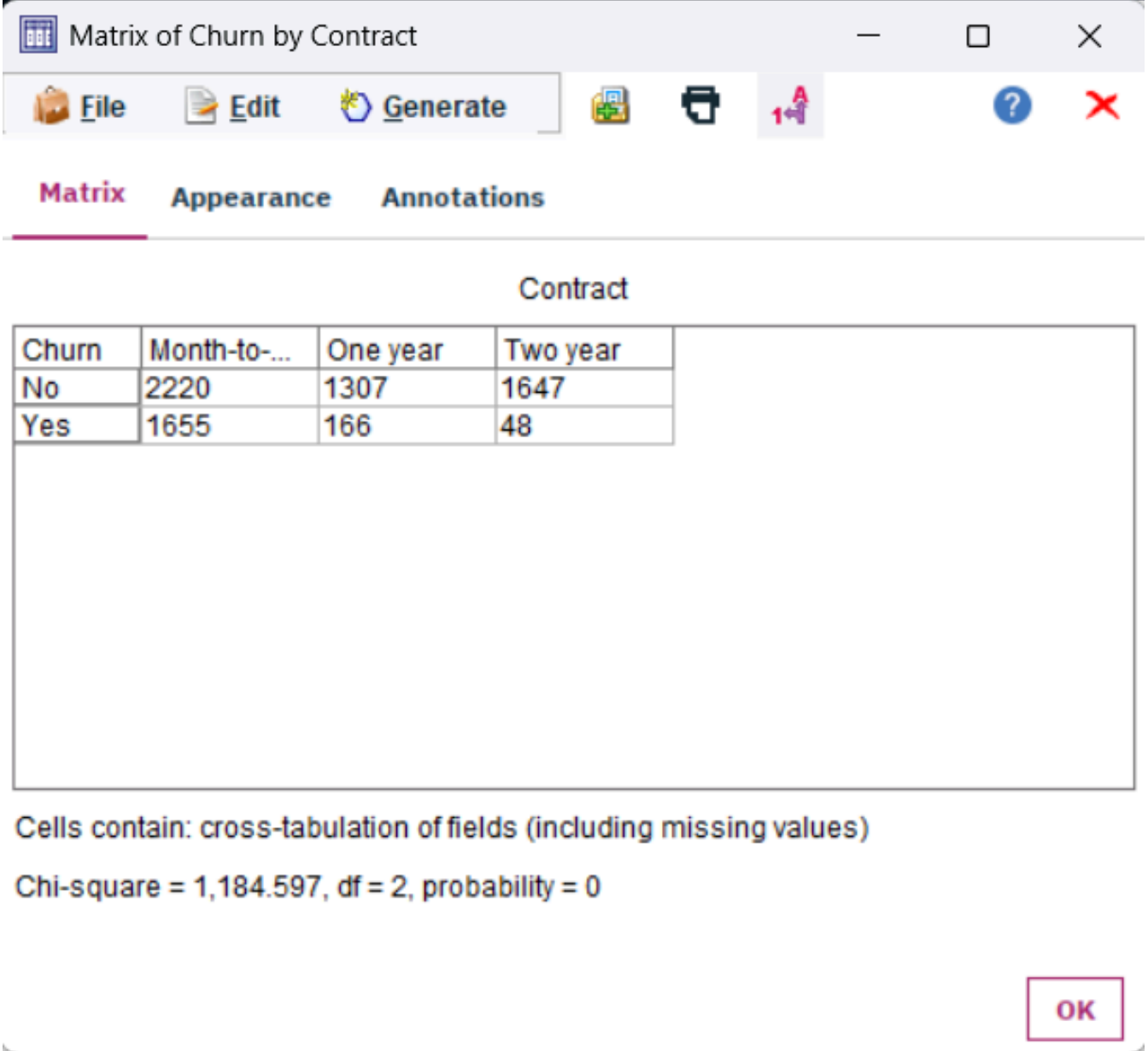


We can analyse the data with various factors to try to understand the relationship between different fields of data, like the example shown in the graph.





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.

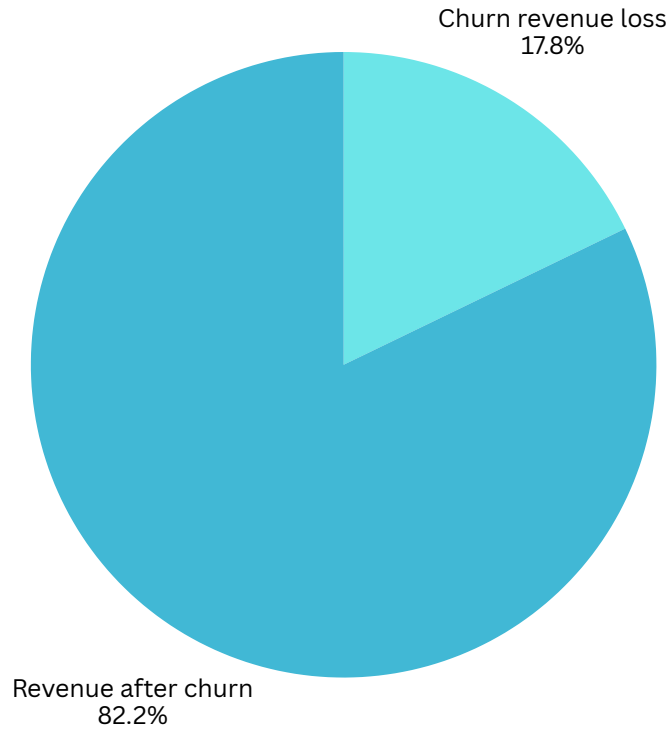


Table (3 fields, 1 records) #3			
File	Edit	Generate	
Table Annotations			
	customerID_Count	TotalCharges_Sum	Record_Count
1	1869	2862926.900	1869

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

Type

Preview

Types

Format

Annotations

Read Values

Clear Values

Clear All Values

Field	Measurement	Values	Missing	Check	Role
customerID	Typeless			None	None
gender	Flag	Male/Fem...		None	Input
SeniorCitizen	Continuous	[0, 1]		None	Input
Partner	Flag	Yes/No		None	Input
Dependents	Flag	Yes/No		None	Input
tenure	Continuous	[0, 72]		None	Input
PhoneService	Flag	Yes/No		None	Input
MultipleLines	Nominal	No,"No ph...		None	Input
InternetService	Nominal	DSL,"Fiber...		None	Input
OnlineSecurity	Nominal	No,"No int...		None	Input
OnlineBackup	Nominal	No,"No int...		None	Input
DeviceProtec...	Nominal	No,"No int...		None	Input
TechSupport	Nominal	No,"No int...		None	Input
StreamingTV	Nominal	No,"No int...		None	Input
StreamingMo...	Nominal	No,"No int...		None	Input
Contract	Nominal	Month-to-...		None	Input
PaperlessBil...	Flag	Yes/No		None	Input
PaymentMet...	Nominal	"Bank tran...		None	Input
MonthlyChar...	Continuous	[18.25,118...		None	Input
TotalCharges	Continuous	[18.8,8684...		None	Input
Churn	Flag	Yes/No		None	Target

☒ View current fields

☐ View unused field settings

OK

Cancel

Apply

Reset

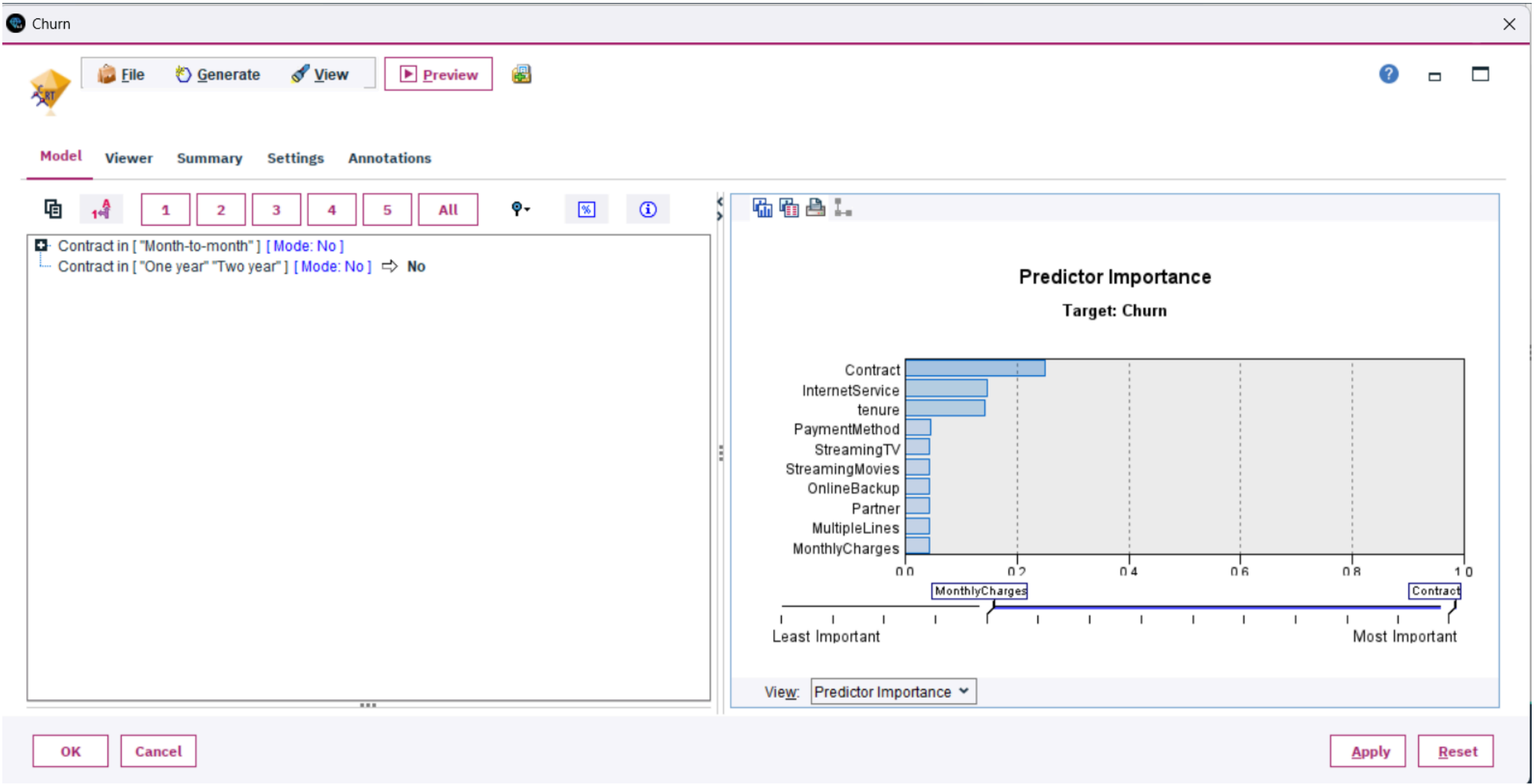
We can also use Partitioning to separate data into sets of training and testing data to test out the model accuracy.

The screenshot shows a 'Partition' dialog box with a title bar containing a close button. The interface includes a toolbar with a hexagonal icon, a 'Generate' button, and a 'Preview' button. Below the toolbar are two tabs: 'Settings' (active) and 'Annotations'. The 'Settings' tab contains the following controls:

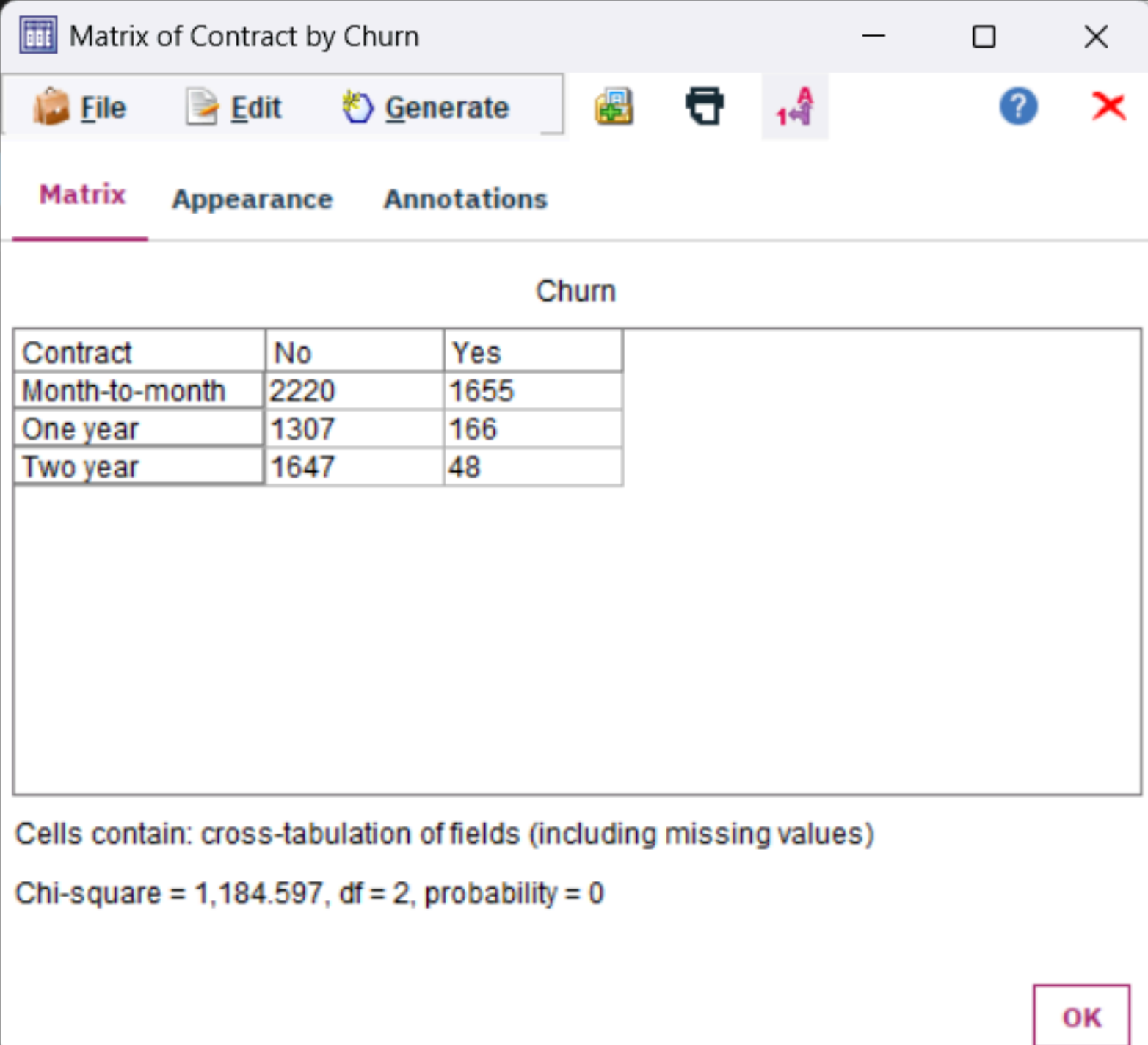
- Partition field:** A text box containing the word 'Partition'.
- Partitions:** Two radio buttons: 'Train and test' (selected) and 'Train, test and validation'.
- Training partition size:** A spinner box set to 80, a 'Label' text box with 'Training', and a 'Value =' text box with '1\_Training'.
- Testing partition size:** A spinner box set to 20, a 'Label' text box with 'Testing', and a 'Value =' text box with '2\_Testing'.
- Validation partition size:** A spinner box set to 0, a 'Label' text box with 'Validation', and a 'Value =' text box with '3\_Validation'.
- Total size:** A label 'Total size:' followed by '100%'.
- Values:** Three radio buttons: 'Use system-defined values ("1", "2" and "3")', 'Append labels to system-defined values' (selected), and 'Use labels as values'.
- Repeatability:** A checked checkbox labeled 'Repeatable partition assignment'.
- Seed:** A spinner box set to 1234567, a 'Generate' button, and a checkbox labeled 'Use unique field to assign partitions:' followed by a greyed-out text box and a dropdown arrow.

At the bottom of the dialog are four buttons: 'OK', 'Cancel', 'Apply', and 'Reset'.

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.



The screenshot shows a software window titled "Matrix of Contract by Churn". It has a menu bar with "File", "Edit", and "Generate". Below the menu bar are tabs for "Matrix", "Appearance", and "Annotations". The "Matrix" tab is selected, displaying a cross-tabulation table. The table has "Contract" as the row variable and "Churn" as the column variable. The data is as follows:

Contract	No	Yes
Month-to-month	2220	1655
One year	1307	166
Two year	1647	48

Below the table, the text reads: "Cells contain: cross-tabulation of fields (including missing values)" and "Chi-square = 1,184.597, df = 2, probability = 0". An "OK" button is located at the bottom right of the window.

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.

Churn

Objective: Standard model

Fields Build Options Model Options Annotations

☒ Use predefined roles  
☐ Use custom field assignments

Fields:

Sort: None

Targets\*:

Churn

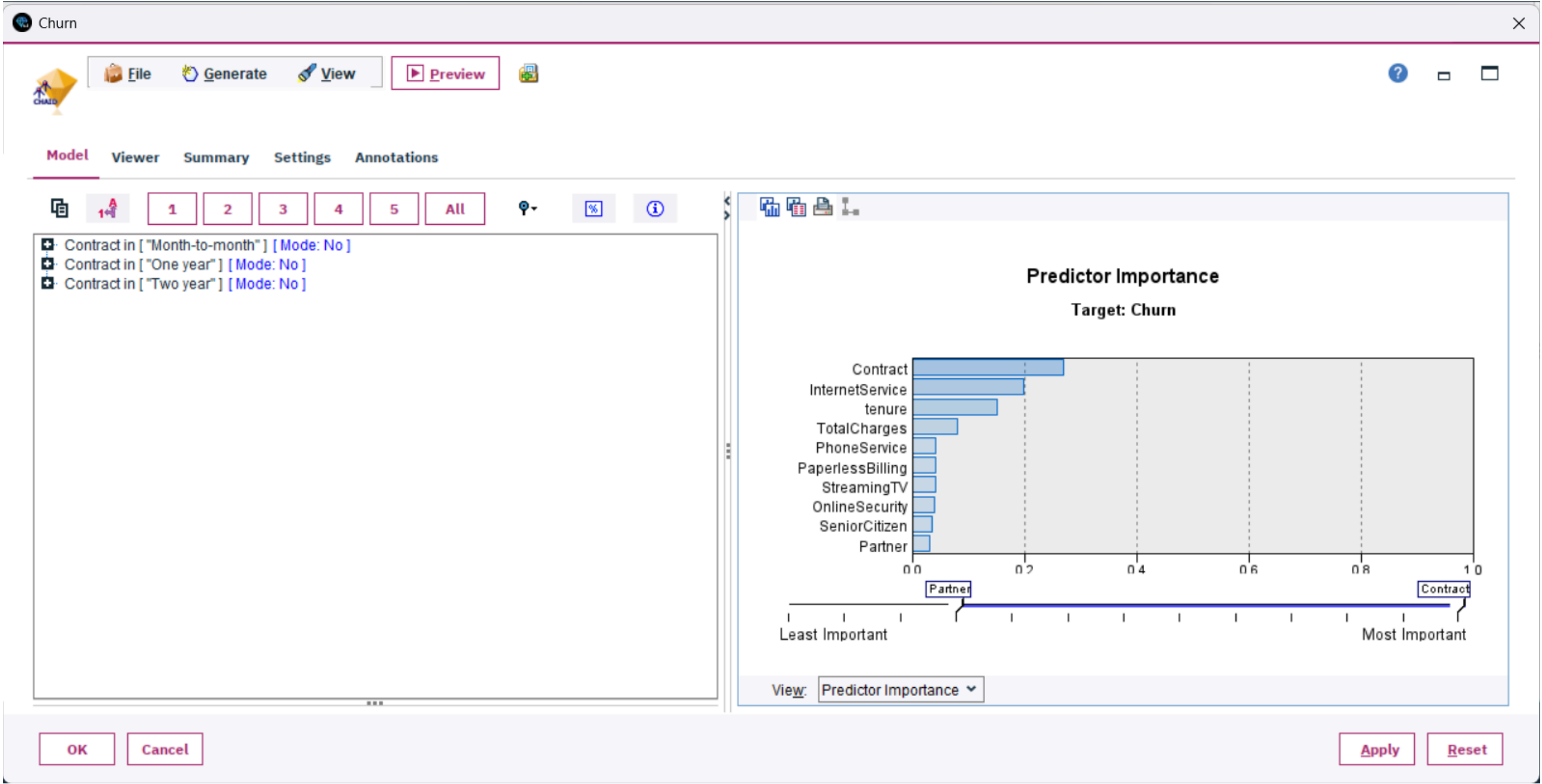
Predictors (Inputs)\*:

- gender
- SeniorCitizen
- Partner
- Dependents
- tenure
- PhoneService
- MultipleLines
- InternetService
- OnlineSecurity
- OnlineBackup
- DeviceProtection
- TechSupport
- StreamingTV
- StreamingMovies
- Contract
- PaperlessBilling
- PaymentMethod
- MonthlyCharges
- TotalCharges

Analysis Weight:

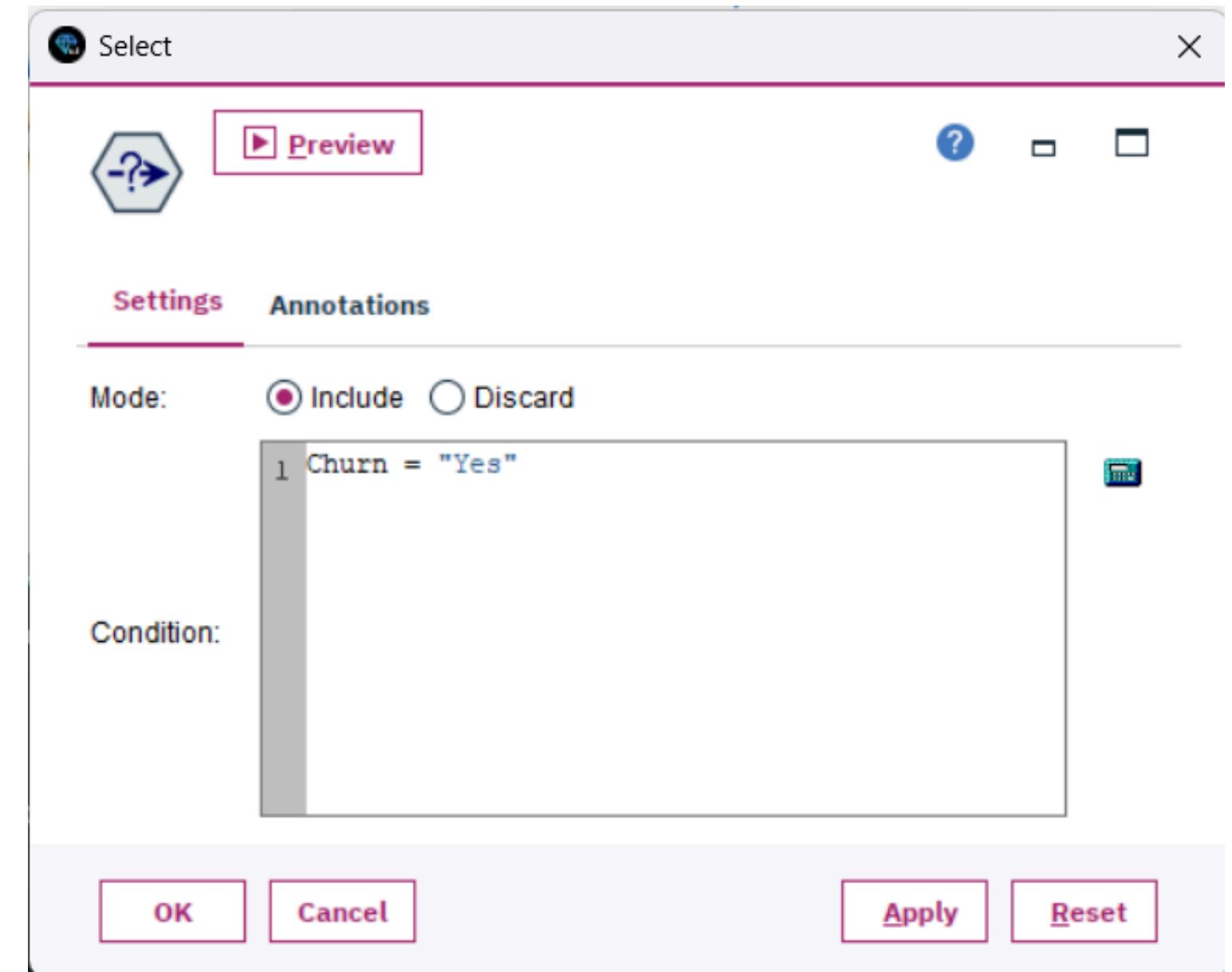
OK Run Cancel Apply Reset

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.



The screenshot shows a 'Select' dialog box with a title bar containing a close button (X). Inside the dialog, there is a toolbar with a hexagonal icon containing a question mark and arrow, a 'Preview' button, and a help icon (question mark). Below the toolbar, there are two tabs: 'Settings' (selected) and 'Annotations'. Under the 'Settings' tab, the 'Mode' is set to 'Include' (selected with a radio button) and 'Discard' (unselected). Below the mode selection, there is a 'Condition:' label and a text area containing the condition '1 Churn = "Yes"'. At the bottom of the dialog, there are four buttons: 'OK', 'Cancel', 'Apply', and 'Reset'.

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

Filter





Preview


?

Filter
























Annotations







Fields: 23 in, 21 filtered, 0 renamed, 2 out

Field	Filter	Field
customerID		customerID
gender		gender
SeniorCitizen		SeniorCitizen
Partner		Partner
Dependents		Dependents
tenure		tenure
PhoneService		PhoneService
MultipleLines		MultipleLines
InternetService		InternetService
OnlineSecurity		OnlineSecurity
OnlineBackup		OnlineBackup
DeviceProtection		DeviceProtection
TechSupport		TechSupport
StreamingTV		StreamingTV
StreamingMovies		StreamingMovies
Contract		Contract
PaperlessBilling		PaperlessBilling
PaymentMethod		PaymentMethod
MonthlyCharges		MonthlyCharges
TotalCharges		TotalCharges
Churn		Churn
\$R-Churn		\$R-Churn
\$RC-Churn		\$RC-Churn

☒ View current fields

☐ View unused field settings

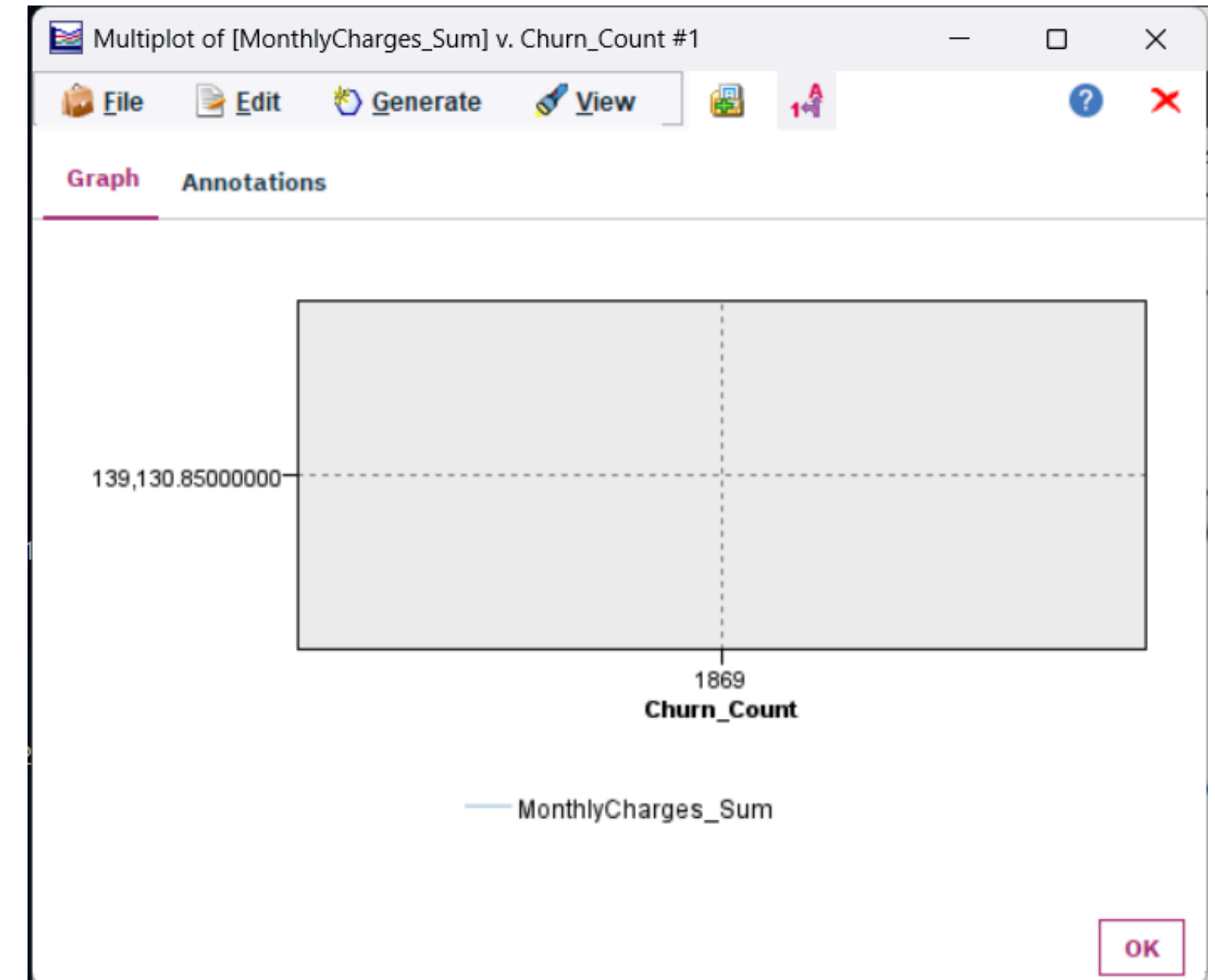
OK

Cancel

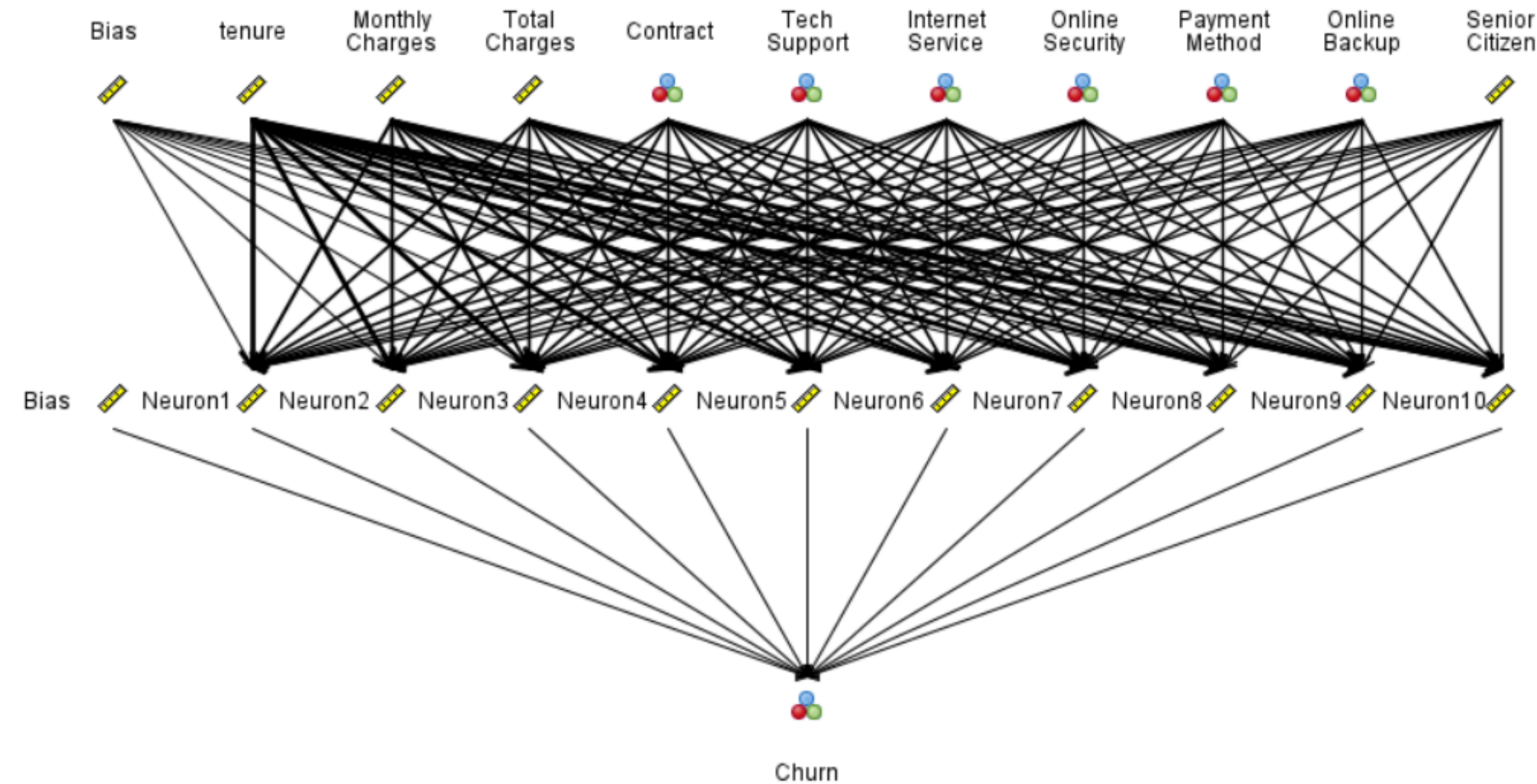
Apply

Reset

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.



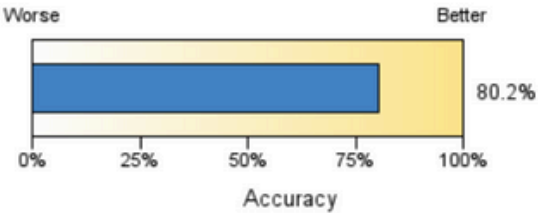
For future analysis of customer churn probability, we can add a neural network node to create a model and add a database node to predict customer churn probability as we add new customers and prepare a report and take actions to reduce churn.



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		Row Percent
	No	Yes	
No	90.2%	9.8%	
Yes	46.2%	53.8%	

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.

Result

?

Export

Publish

Annotations

Export file:

C:\Users\adity\Downloads\output

...

Write mode:

☒ Overwrite

☒ Include field names

☐ Append

☒ New line after each record

Field separator:

☒ Comma

☐ Tab

☐ Space

☐ Other:

Symbol quotes:

☐ None

☐ Single ( ' )

☒ Double ( " )

☐ Other:

Encoding:

Stream default

Decimal symbol:

Stream default

☐ Generate an import node for this data

OK

▶ Run

Cancel

Apply

Reset