



School of Computer Applications (DS & AI)
Predictive Analysis

Aditya Yadav (1230258037)

Amit Yadav (1230258036)

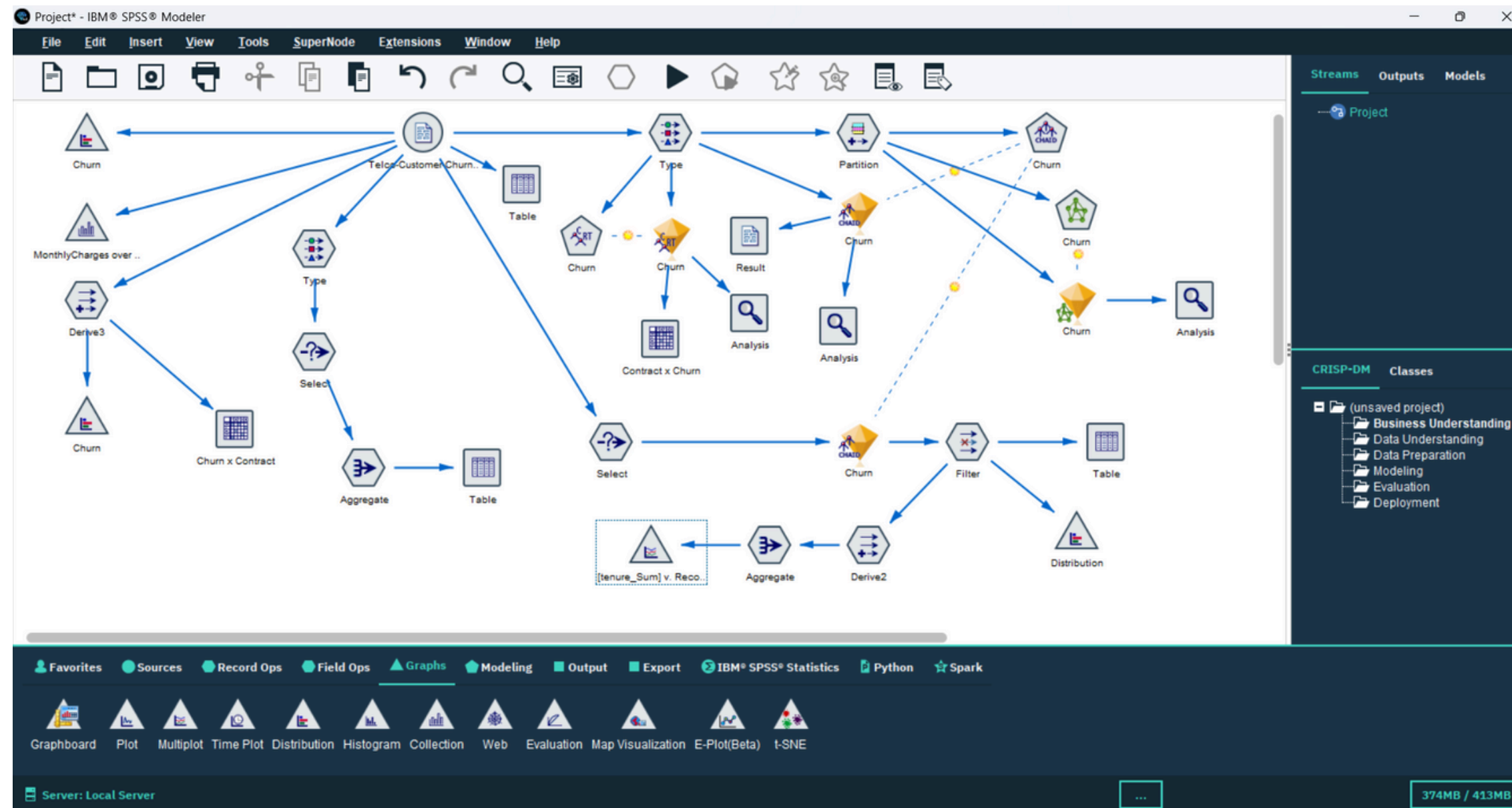
Aditya Srivastava (1230258065)

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.

The screenshot shows the 'Import' dialog box in IBM SPSS Modeler, titled 'Telco-Customer-Churn.csv'. The file path is 'C:\Users\adity\Desktop\Telco-Customer-Churn.csv'. The 'File' tab is selected, showing a preview of the CSV data. The preview shows three rows of data, with the first row being the header: 'customerID,gender,SeniorCitizen,Partner,Dependents,tenure,PhoneService,Multip.'. The dialog box has various options for importing the file, including 'Read field names from file' (checked), 'Specify number of fields' (set to 1), 'Skip header characters' (set to 0), 'EOL comment characters' (empty), 'Strip lead and trail spaces' (set to None), 'Invalid characters' (set to Discard), 'Encoding' (set to Stream default), 'Decimal symbol' (set to Stream default), 'Line delimiter is newline character' (unchecked), 'Lines to scan for column and type' (set to 50), 'Field delimiters' (Space, Comma, Tab, Newline, Other), 'Automatically recognize dates and times' (checked), 'Treat square brackets as lists' (unchecked), and 'Quotes' (Single quotes: Pair and discard, Double quotes: Pair and discard). The 'OK' button is highlighted.

Telco-Customer-Churn.csv

Preview Refresh

C:\Users\adity\Desktop\Telco-Customer-Churn.csv

File Data Filter Types Annotations

File: C:\Users\adity\Desktop\Telco-Customer-Churn.csv

customerID,gender,SeniorCitizen,Partner,Dependents,tenure,PhoneService,Multip.
7590-VHVEG,Female,0,Yes,No,1,No,No phone service,DSL,No,Yes,No,No,No,No,Month-
5575-GNVDE,Male,0,No,No,34,Yes,No,DSL,Yes,No,Yes,No,No,No,One year,No,Mailed
3668-QPYBK,Male,0,No,No,2,Yes,No,DSL,Yes,Yes,No,No,No,No,Month-to-month,Yes,M

☒ Read field names from file ☐ Specify number of fields 1

Skip header characters: 0 EOL comment characters:

Strip lead and trail spaces: ☒ None ☐ Left ☐ Right ☐ Both

Invalid characters: ☒ Discard ☐ Replace with

Encoding: Stream default Decimal symbol: Stream default

☐ Line delimiter is newline character Lines to scan for column and type: 50

Field delimiters
☐ Space ☒ Comma ☐ Tab
☒ Newline ☐ Other
☐ Non-printing characters
☐ Allow multiple blank delimiters

☒ Automatically recognize dates and times
☐ Treat square brackets as lists

Quotes
Single quotes: Pair and discard
Double quotes: Pair and discard

OK Cancel Apply Reset

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.

Table

?

SettingsFormatOutputAnnotations

Field	Format	Justify	Column Width
A customerID		Auto	Auto
A gender		Auto	Auto
SeniorCitizen	####	Auto	Auto
A Partner		Auto	Auto
A Dependents		Auto	Auto
tenure	####	Auto	Auto
A PhoneService		Auto	Auto
A MultipleLines		Auto	Auto
A InternetService		Auto	Auto
A OnlineSecurity		Auto	Auto
A OnlineBackup		Auto	Auto
A DeviceProtection		Auto	Auto
A TechSupport		Auto	Auto
A StreamingTV		Auto	Auto
A StreamingMovies		Auto	Auto
A Contract		Auto	Auto
A PaperlessBilling		Auto	Auto
A PaymentMethod		Auto	Auto
MonthlyCharges	####.###	Auto	Auto
TotalCharges	####.###	Auto	Auto
A Churn		Auto	Auto

☒ View current fields☐ View unused field settings

OK

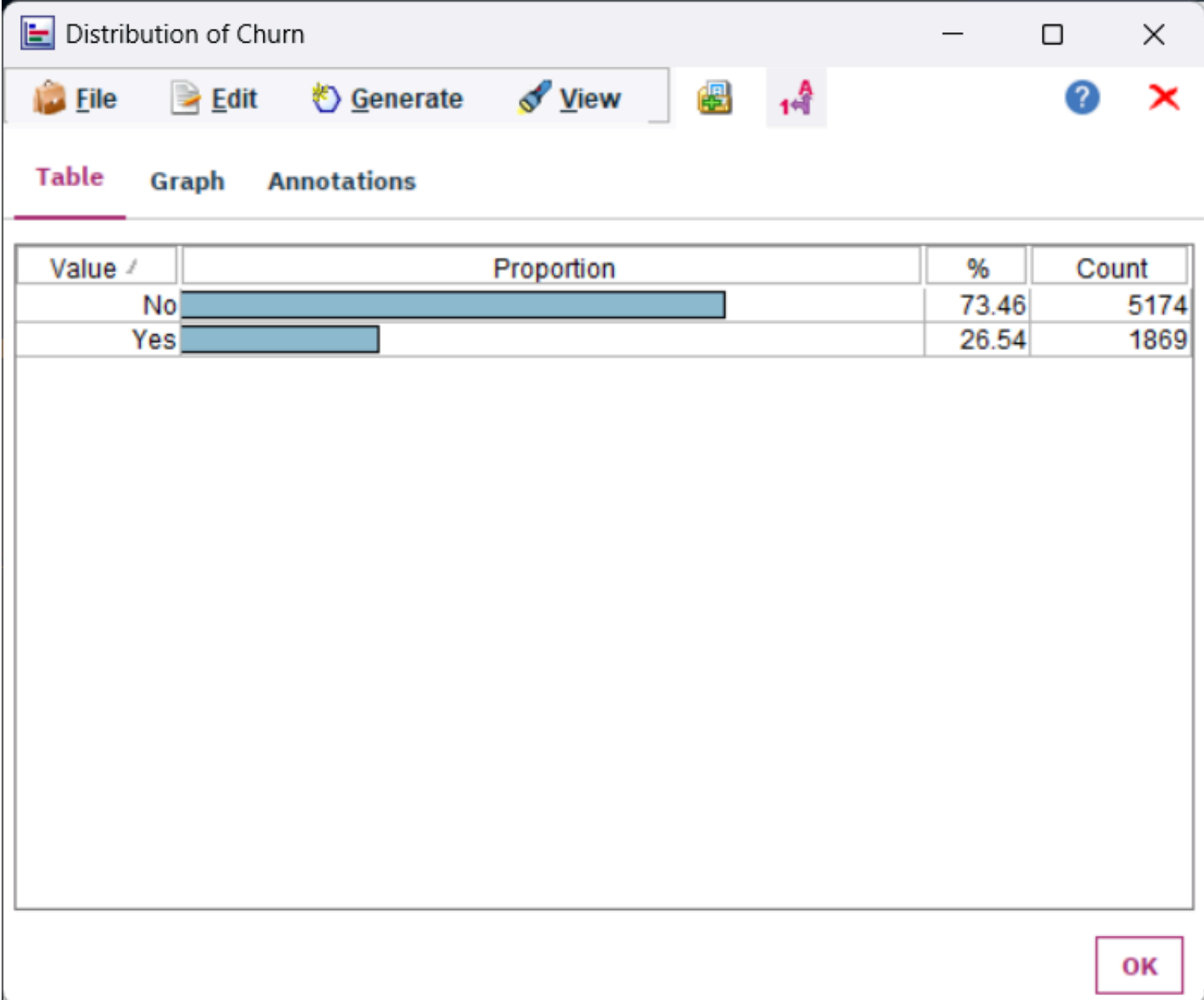
▶ Run

Cancel

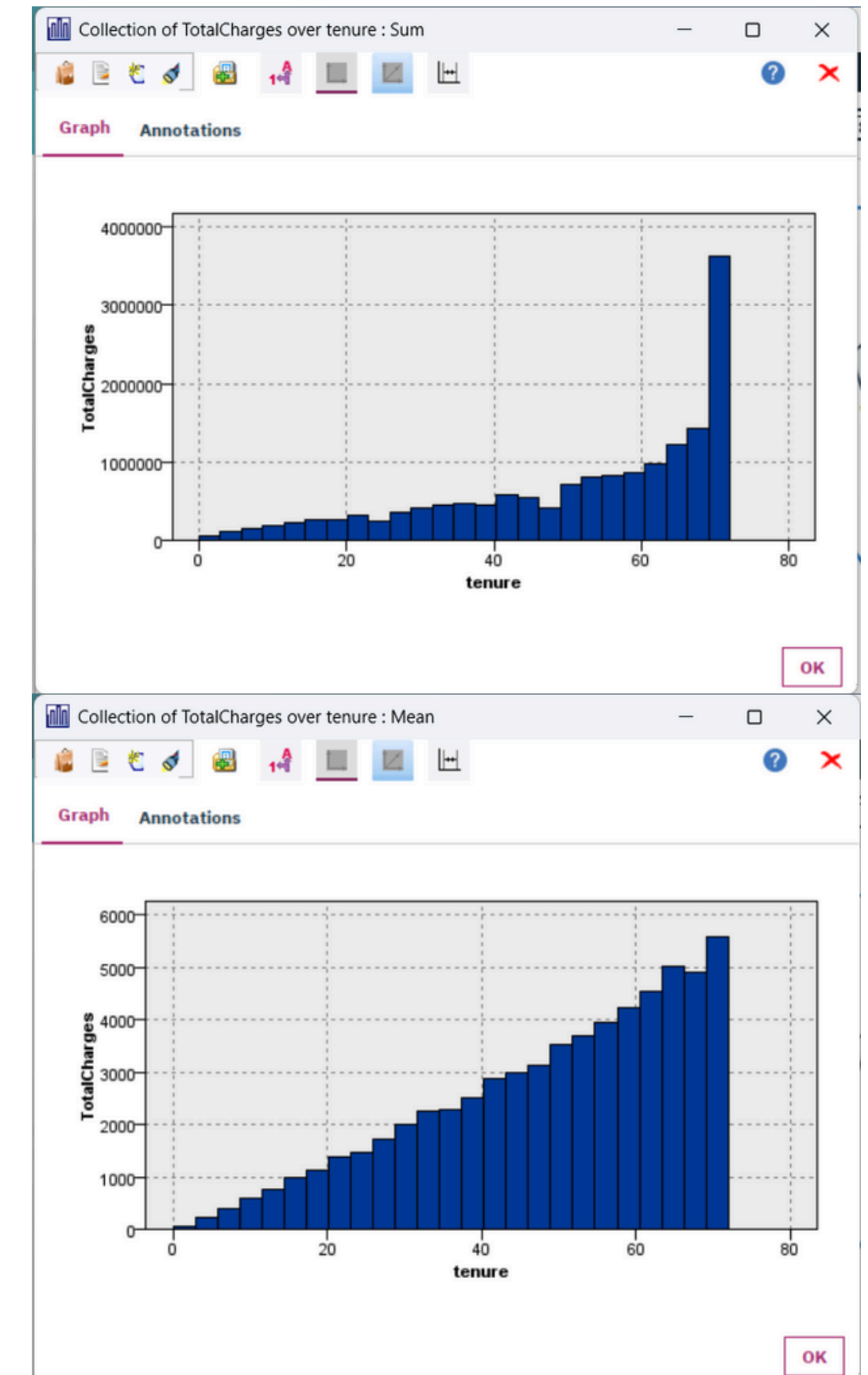
Apply

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 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.

Matrix of Churn by Contract

File Edit Generate

Matrix Appearance Annotations

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

Cells contain: cross-tabulation of fields (including missing values)

Chi-square = 1,184.597, df = 2, probability = 0

OK

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.

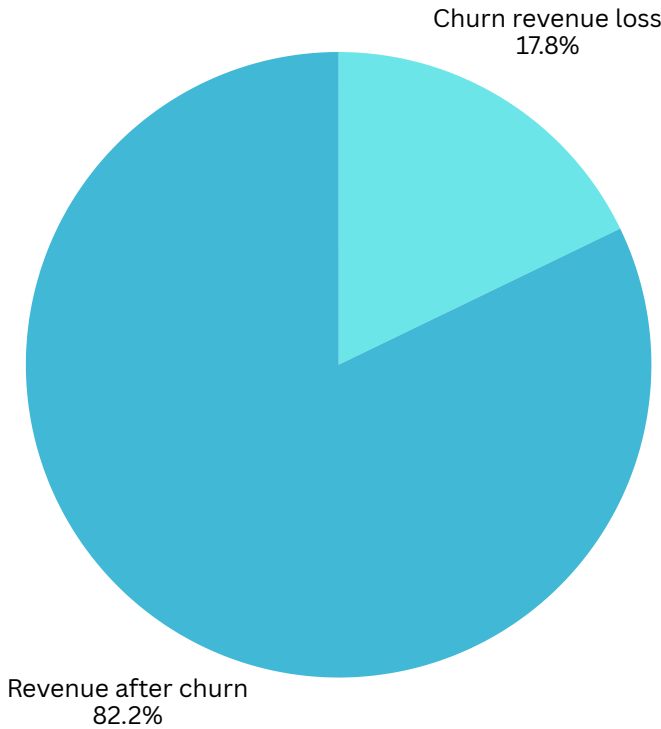


Table (3 fields, 1 records) #3

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.

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 (X) and a help icon (?). Below the title bar is a toolbar with a hexagonal icon, a 'Generate' button, and a 'Preview' button. The dialog is divided into two tabs: 'Settings' (active) and 'Annotations'. Under the 'Settings' tab, there are several configuration options: 'Partition field' is set to 'Partition'; 'Partitions' has two radio buttons, 'Train and test' (selected) and 'Train, test and validation'; 'Training partition size' is 80, 'Testing partition size' is 20, and 'Validation partition size' is 0, each with up/down arrows; 'Label' and 'Value' fields are provided for each partition size, with values '1_Training', '2_Testing', and '3_Validation' respectively; 'Total size' is 100%; 'Values' has three radio buttons: 'Use system-defined values ("1", "2" and "3")', 'Append labels to system-defined values' (selected), and 'Use labels as values'; 'Repeatable partition assignment' is checked; 'Seed' is 1234567 with a 'Generate' button; and an unchecked checkbox 'Use unique field to assign partitions:' with a dropdown menu. At the bottom are 'OK', 'Cancel', 'Apply', and 'Reset' buttons.

Partition

Generate Preview

Settings Annotations

Partition field: Partition

Partitions: ☒ Train and test ☐ Train, test and validation

Training partition size: 80 Label: Training Value = "1_Training"

Testing partition size: 20 Label: Testing Value = "2_Testing"

Validation partition size: 0 Label: Validation Value = "3_Validation"

Total size: 100%

Values: ☐ Use system-defined values ("1", "2" and "3")
☒ Append labels to system-defined values
☐ Use labels as values

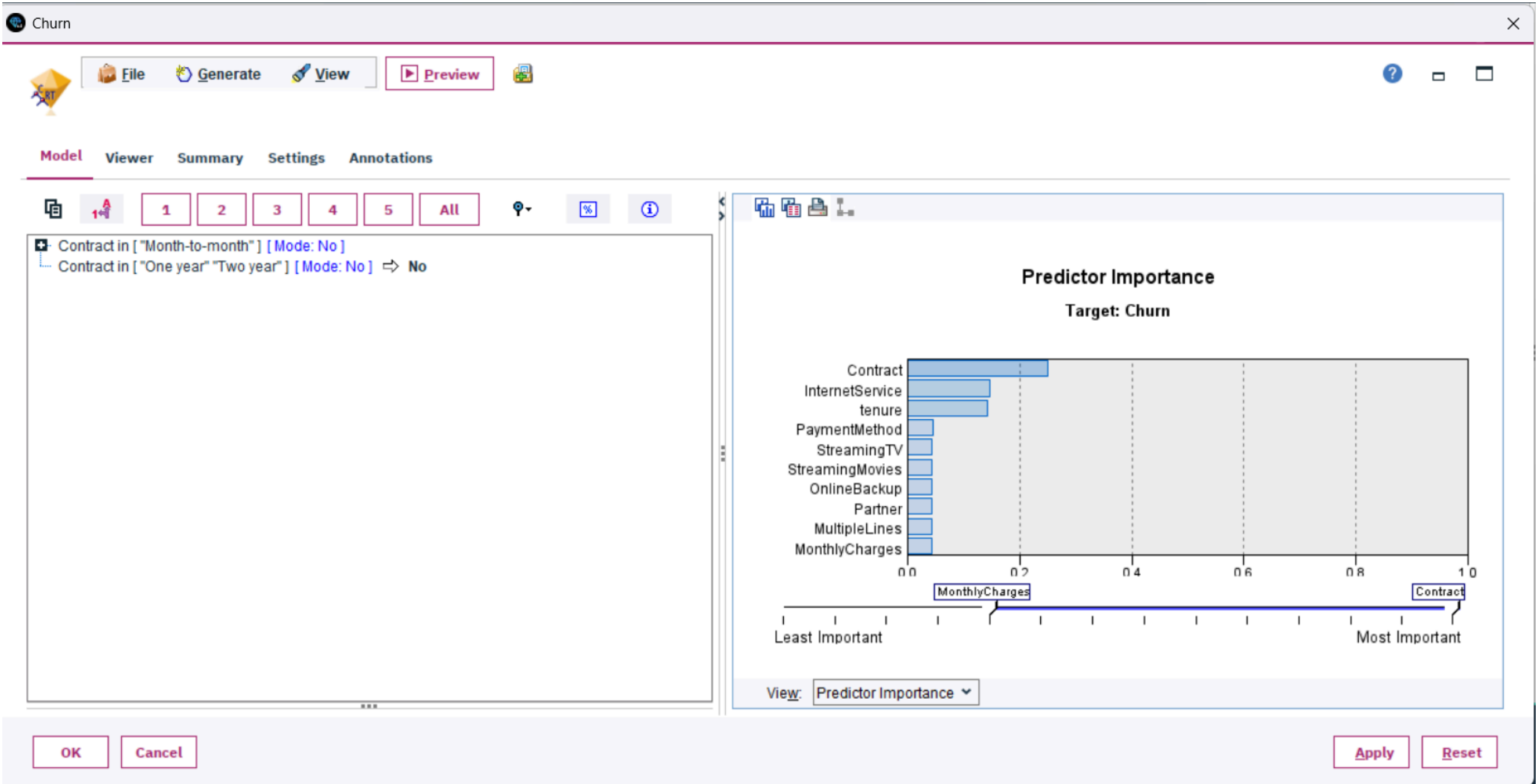
☒ Repeatable partition assignment

Seed: 1234567 Generate

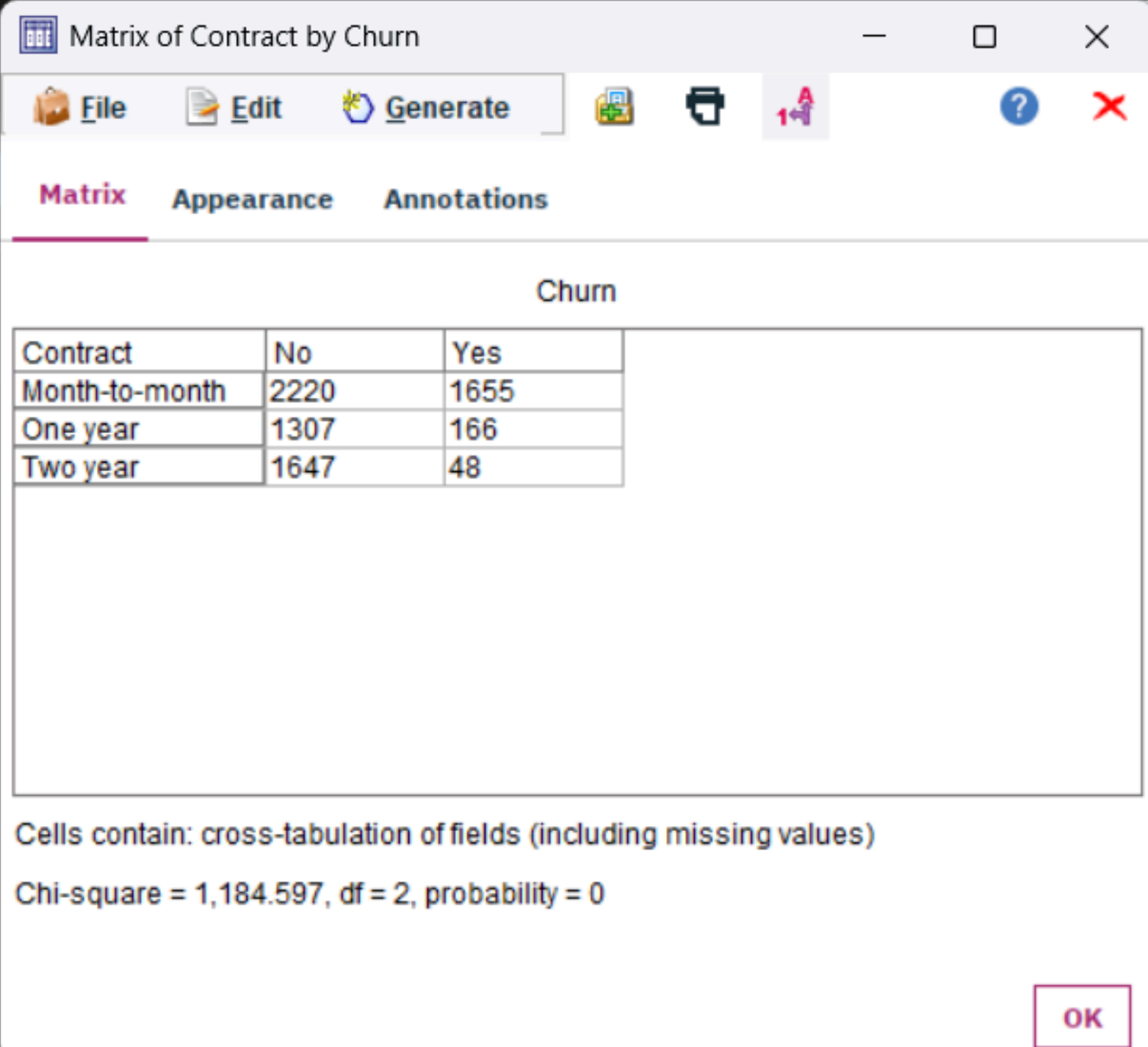
☐ Use unique field to assign partitions:

OK Cancel Apply Reset

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.



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 rows are "Month-to-month", "One year", and "Two year". The columns are "No" and "Yes". The table contains the following counts:

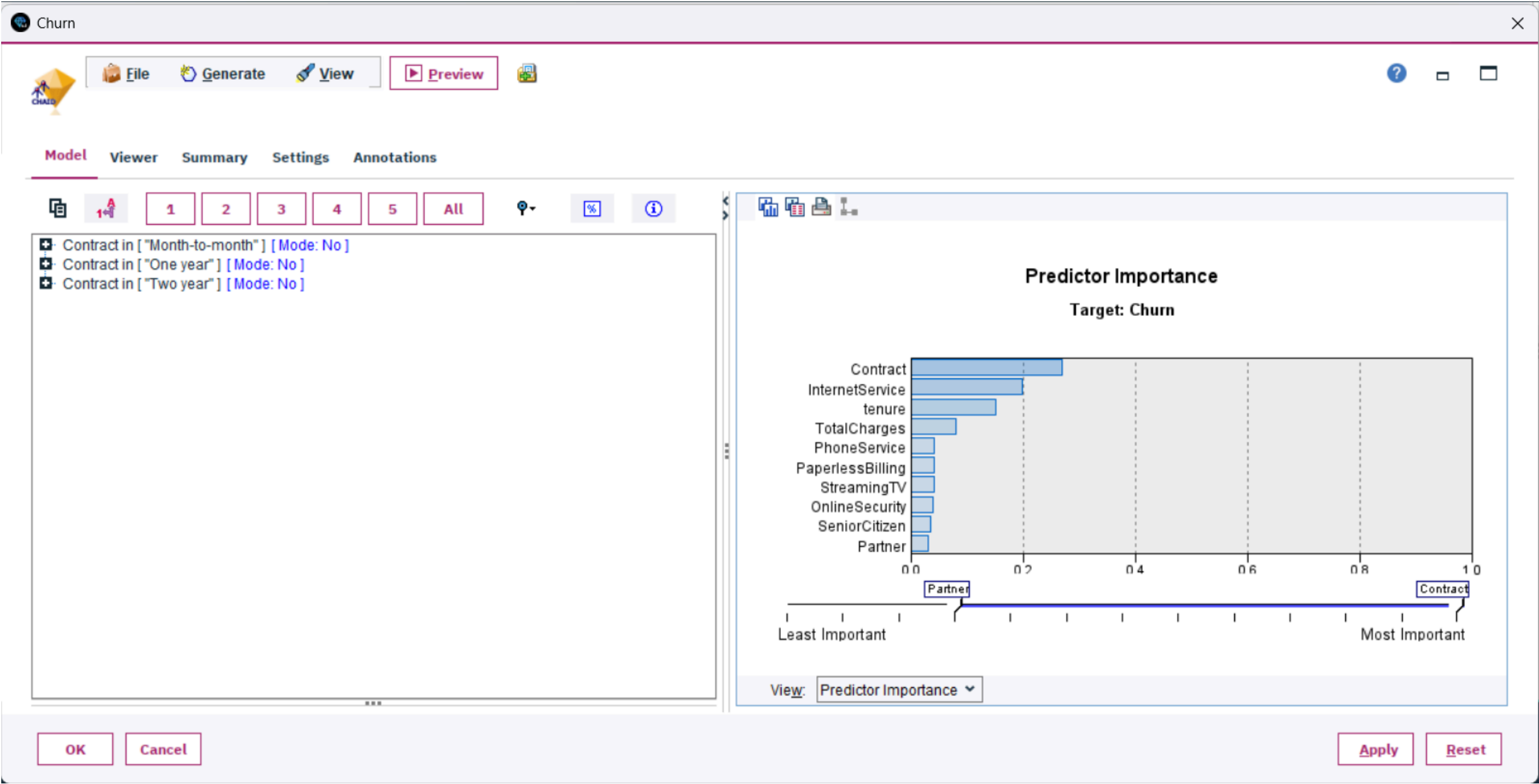
Contract	No	Yes
Month-to-month	2220	1655
One year	1307	166
Two year	1647	48

Below the table, the text "Cells contain: cross-tabulation of fields (including missing values)" is displayed. At the bottom, the statistical results are shown: "Chi-square = 1,184.597, df = 2, probability = 0". An "OK" button is located in the bottom right corner.

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.

The screenshot shows a software window titled "Churn" with a close button (X) in the top right corner. The window contains a CHAID logo and a text field labeled "Objective: Standard model". Below this, there are four tabs: "Fields", "Build Options", "Model Options", and "Annotations", with "Fields" being the active tab. In the "Fields" tab, there are two radio buttons: "Use predefined roles" (selected) and "Use custom field assignments". Below these is a "Fields:" section with a "Sort: None" dropdown and a list of fields. To the right of the "Fields:" list is a "Targets*" section with a dropdown menu showing "Churn". Below the "Targets*" section is a "Predictors (Inputs)*:" section with a list of predictors: gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, and TotalCharges. At the bottom right, there is an "Analysis Weight:" section with a dropdown menu. At the bottom of the window, there are five buttons: "OK", "Run", "Cancel", "Apply", and "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 text '1 Churn = "Yes"'. To the right of the text area is a small icon of a document with a checkmark. At the bottom of the dialog, there are four buttons: 'OK', 'Cancel', 'Apply', and 'Reset'.

Select

Preview

Settings Annotations

Mode: ☒ Include ☐ Discard

Condition: 1 Churn = "Yes"

OK Cancel Apply 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	<div><div></div></div>	customerID
gender	<div><div></div></div>	gender
SeniorCitizen	<div><div></div></div>	SeniorCitizen
Partner	<div><div></div></div>	Partner
Dependents	<div><div></div></div>	Dependents
tenure	<div><div></div></div>	tenure
PhoneService	<div><div></div></div>	PhoneService
MultipleLines	<div><div></div></div>	MultipleLines
InternetService	<div><div></div></div>	InternetService
OnlineSecurity	<div><div></div></div>	OnlineSecurity
OnlineBackup	<div><div></div></div>	OnlineBackup
DeviceProtection	<div><div></div></div>	DeviceProtection
TechSupport	<div><div></div></div>	TechSupport
StreamingTV	<div><div></div></div>	StreamingTV
StreamingMovies	<div><div></div></div>	StreamingMovies
Contract	<div><div></div></div>	Contract
PaperlessBilling	<div><div></div></div>	PaperlessBilling
PaymentMethod	<div><div></div></div>	PaymentMethod
MonthlyCharges	<div><div></div></div>	MonthlyCharges
TotalCharges	<div><div></div></div>	TotalCharges
Churn	<div><div></div></div>	Churn
\$R-Churn	<div><div></div></div>	\$R-Churn
\$RC-Churn	<div><div></div></div>	\$RC-Churn

View current fields

View unused field settings

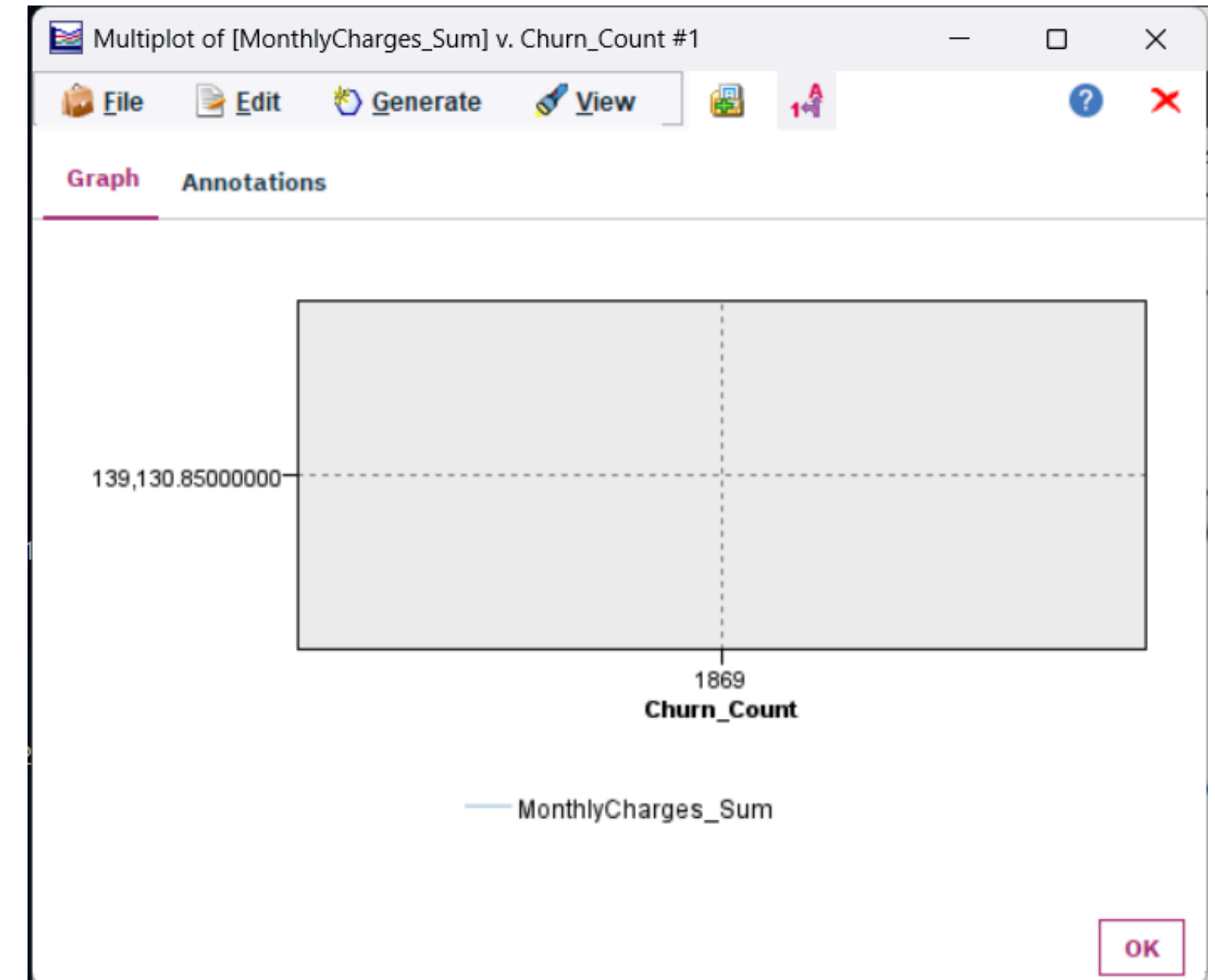
OK

Cancel

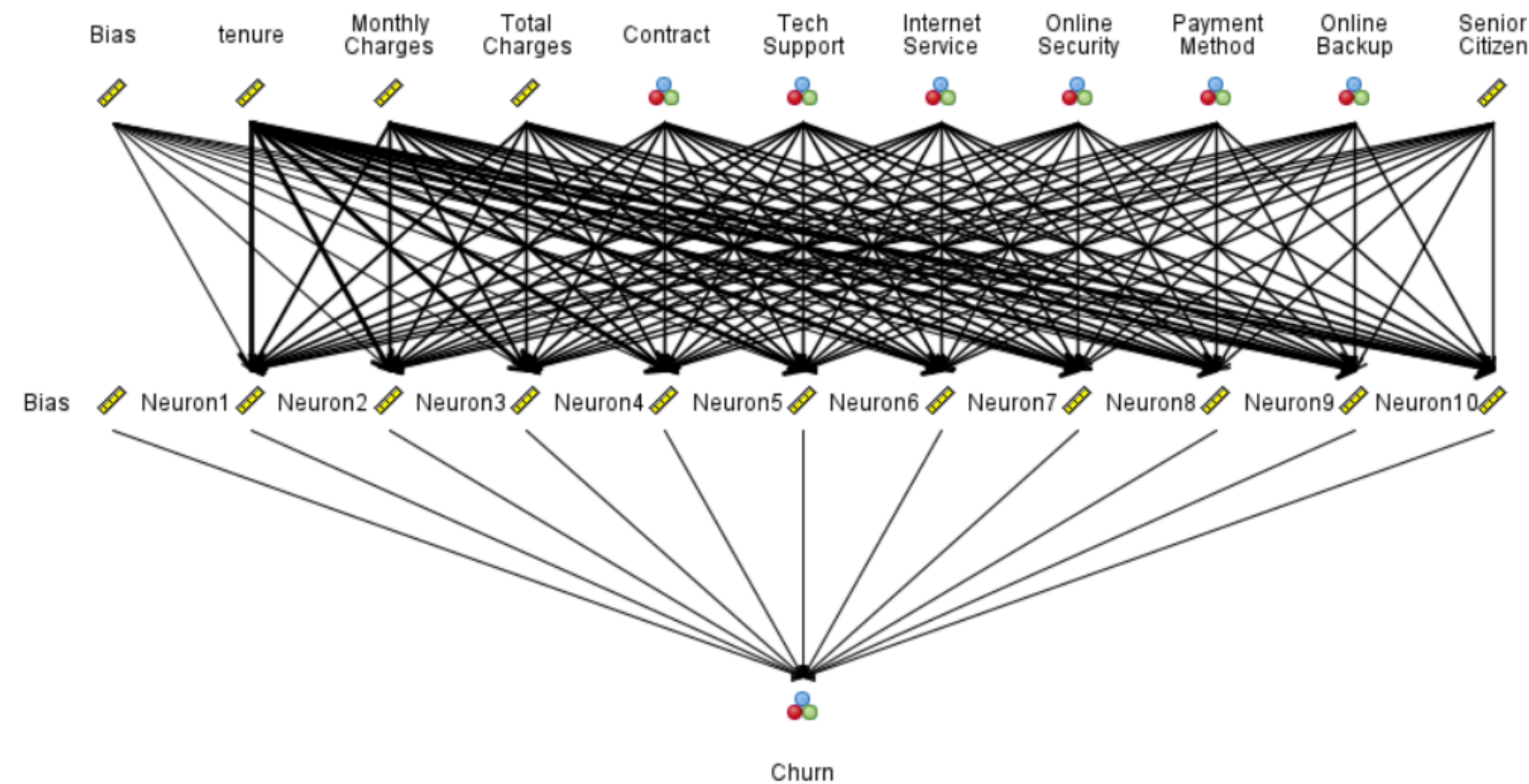
Apply

Reset

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.

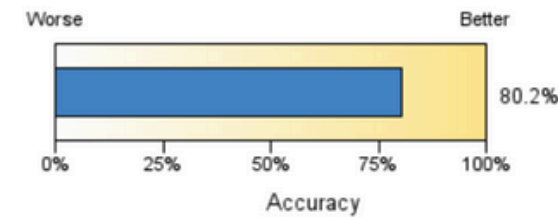


For future analysis of customer churn probability, we can add a neural network node to create a model and add a database node to predict customer churn probability as we add new customers and prepare a report and take actions to reduce churn.



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		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 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.

Result

?

Export

Publish

Annotations

Export file:

C:\Users\ladity\Downloads\output

...

Write mode:

Overwrite

Append

Include field names

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