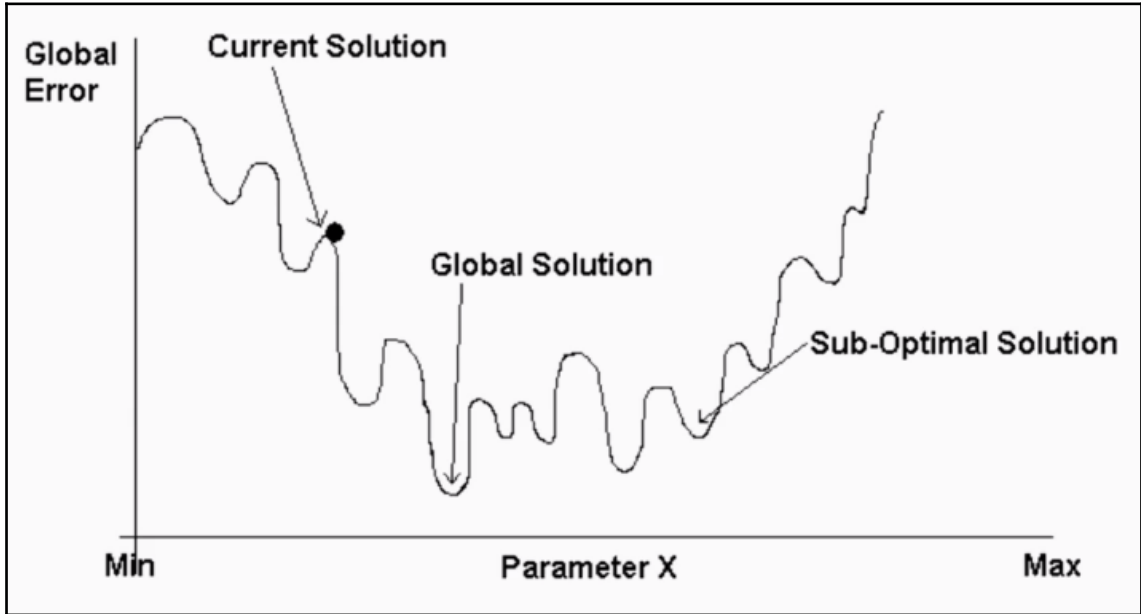
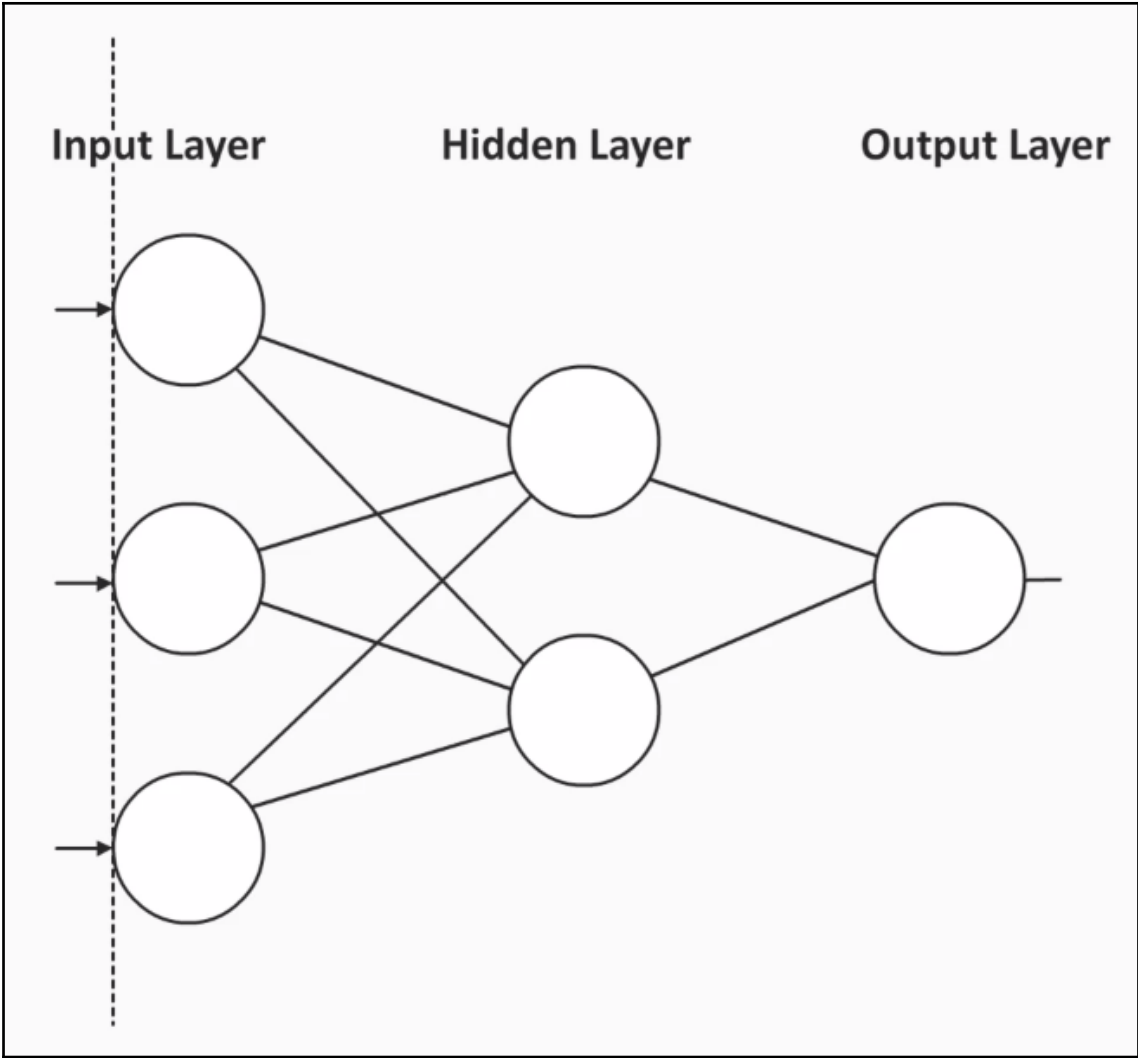
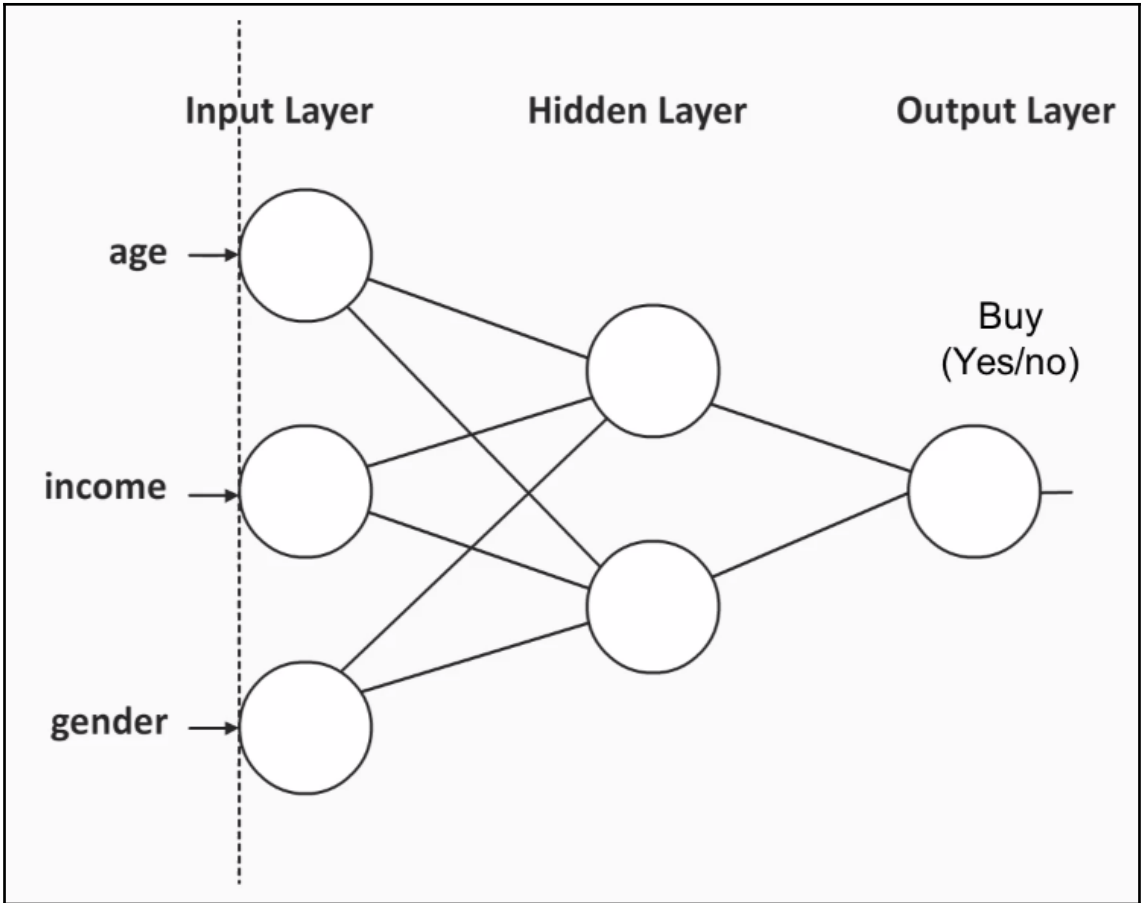
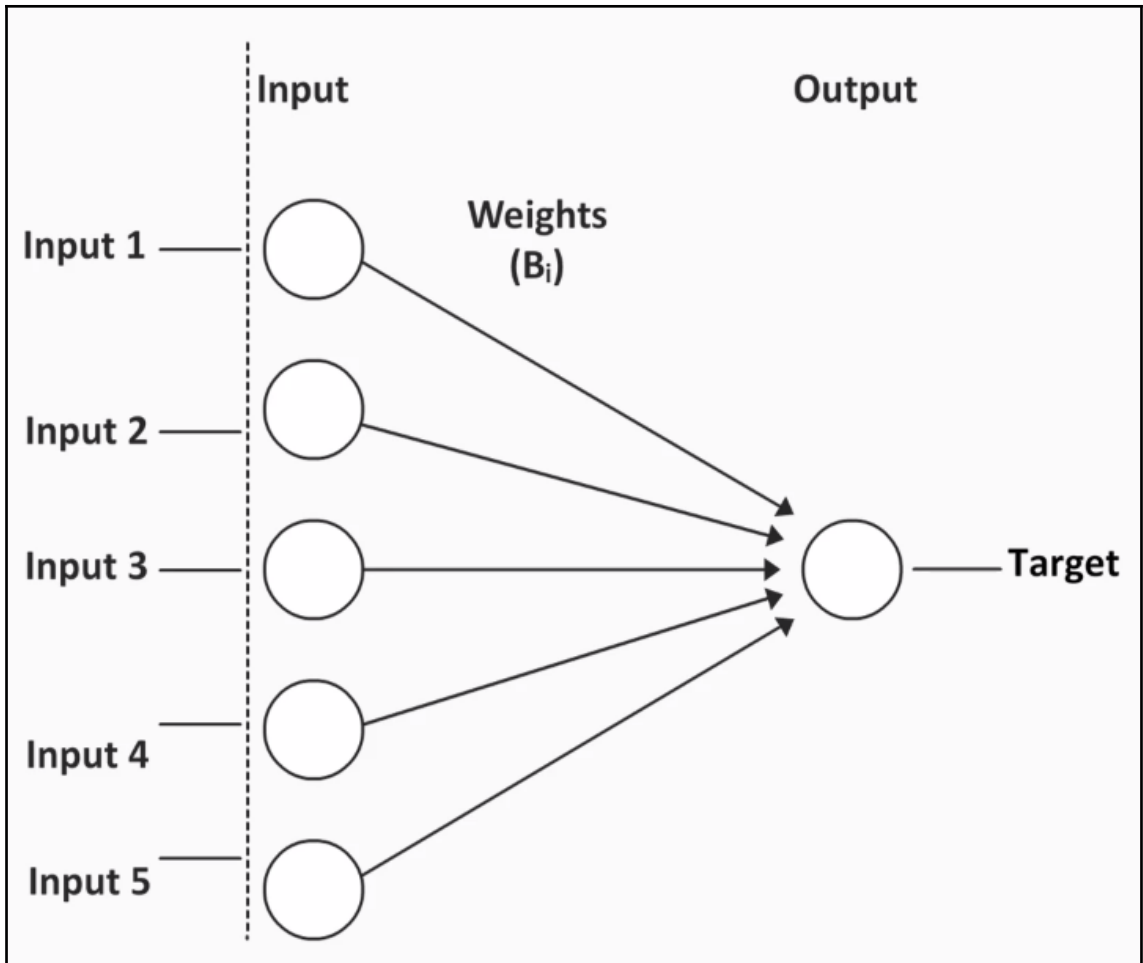


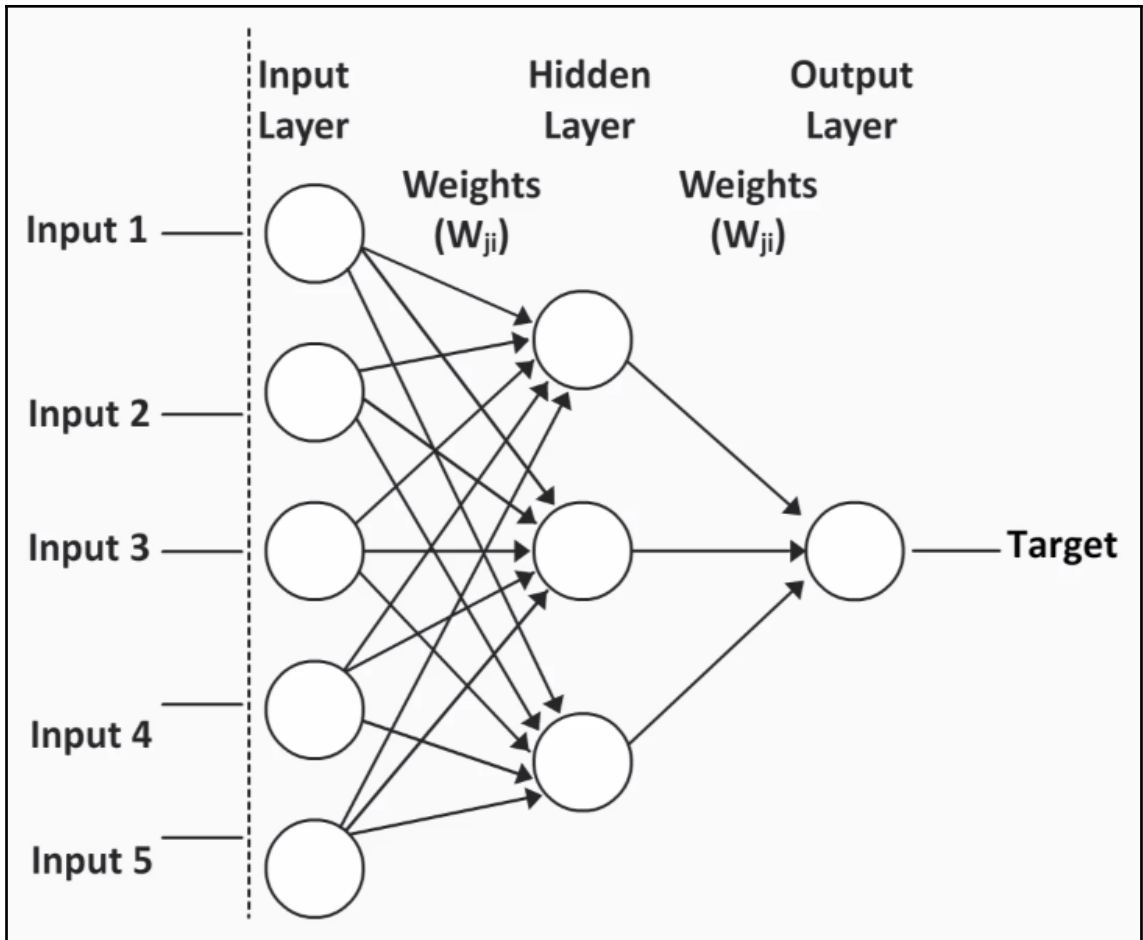
# Chapter 1: Introducing Machine Learning Predictive Models

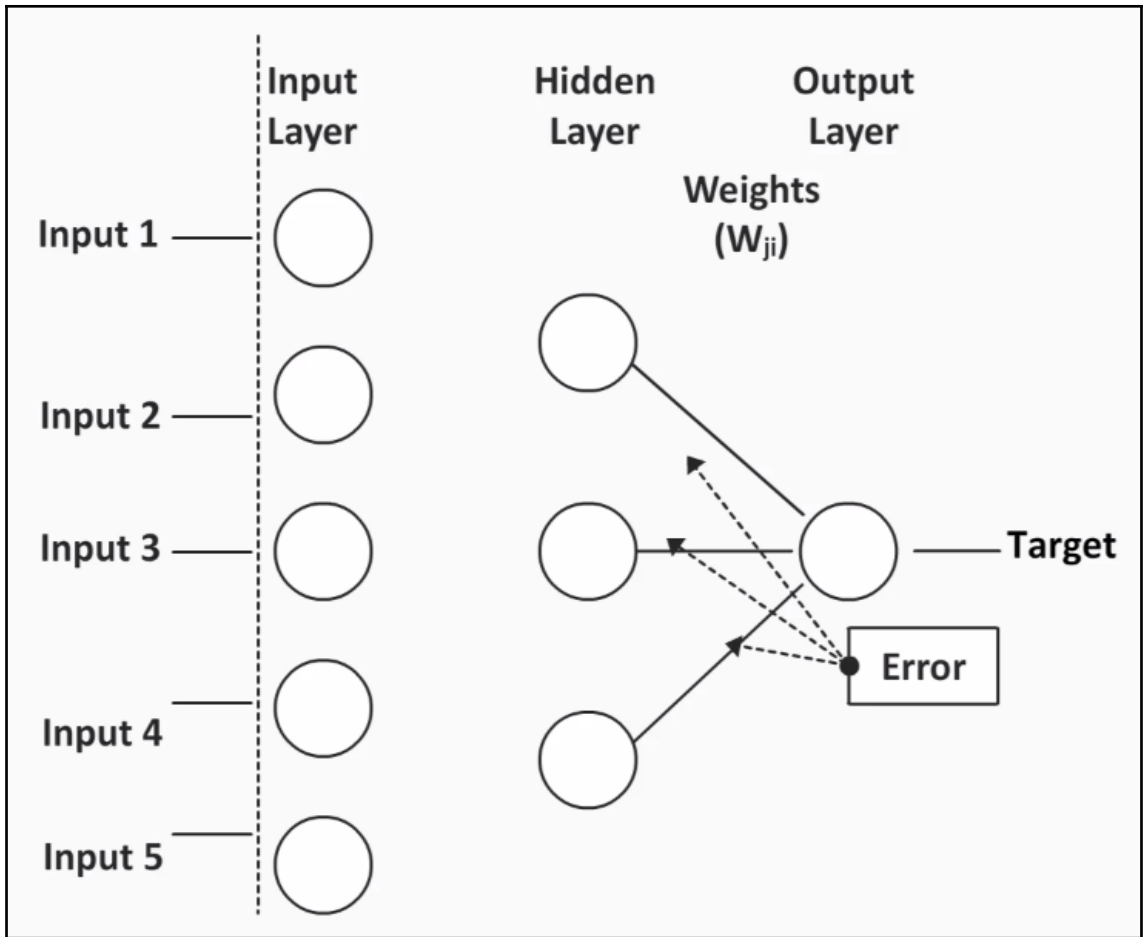


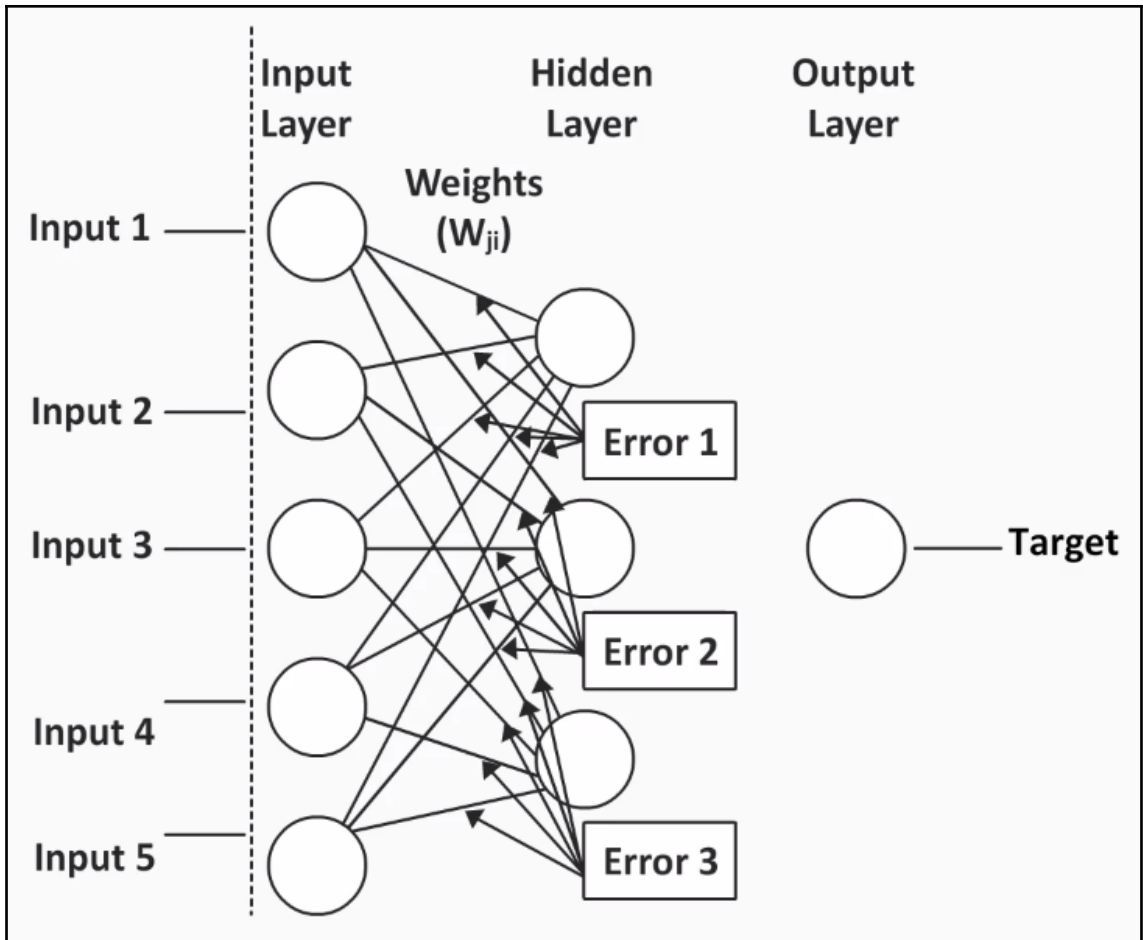




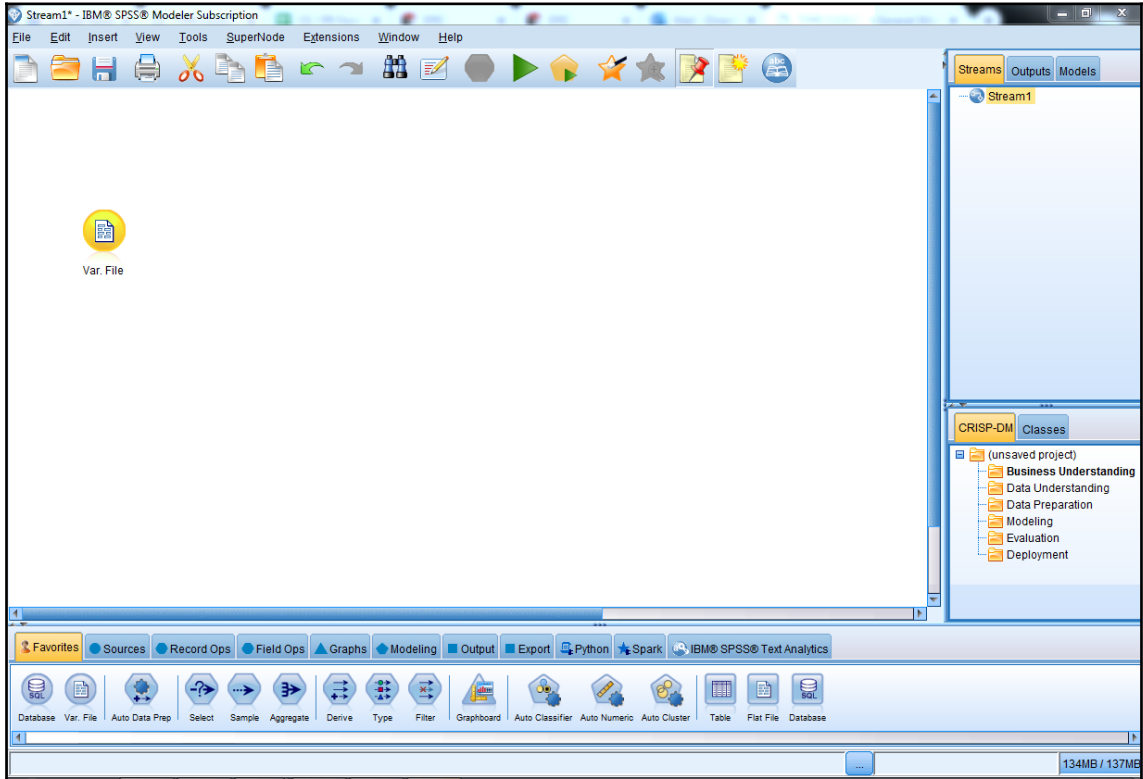







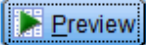
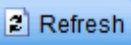





# Chapter 2: Getting Started with Machine Learning





Var. File X

(No current file selected)

**File** | Data | Filter | Types | Annotations

File:  ...

Read field names from file  Specify number of fields

Skip header characters:  EOL comment characters:

Strip lead and trail spaces:  None  Left  Right  Both

Invalid characters:  Discard  Replace with

Encoding:  Decimal symbol:

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

**Field delimiters**

Space  Comma  Tab

Newline  Other

Non-printing characters

Allow multiple blank delimiters

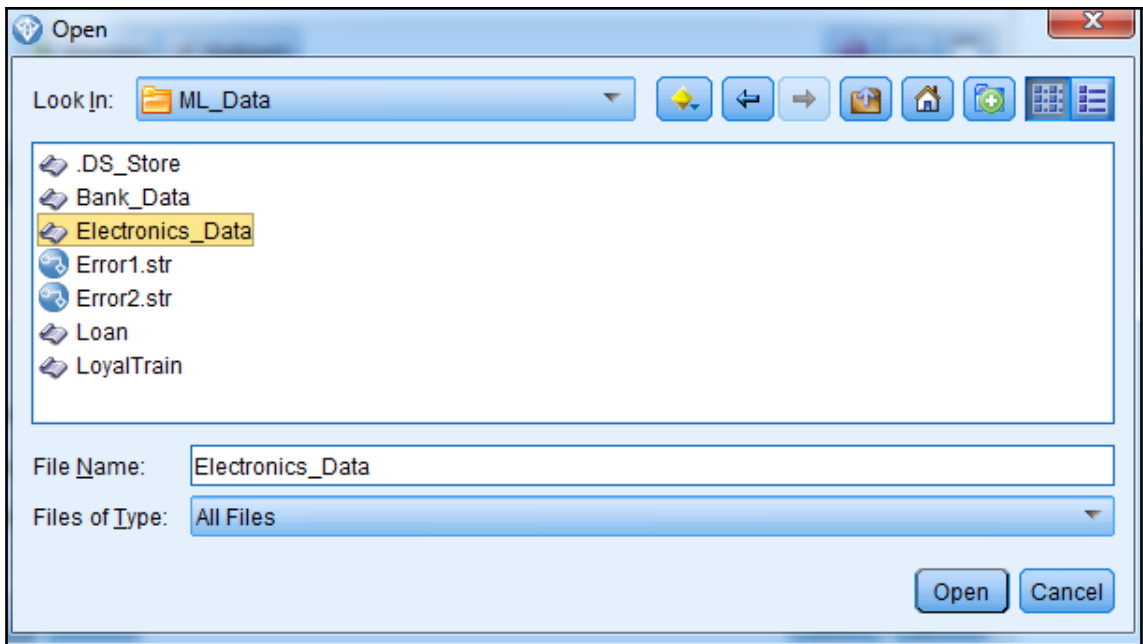
Automatically recognize dates and times

Treat square brackets as lists

**Quotes**

Single quotes:

Double quotes:



Var. File X

C:\Users\dineshchaudhary\Desktop\codes\Code\Code\ML\_Data\Electronics\_Data

Field	Measurement	Values	Missing	Check	Role
# ID	Continuous			None	Input
# Stereos	Continuous			None	Input
# TVs	Continuous			None	Input
# Speakers	Continuous			None	Input
# Delivery_Pro...	Continuous			None	Input
# Years_as_cu...	Continuous			None	Input
# Estimated_R...	Continuous			None	Input
⚠ Payment_Me...	Categorical			None	Input
⚠ Speaker_Dis...	Categorical			None	Input
⚠ Stereo_Disc...	Categorical			None	Input
⚠ Premier	Categorical			None	Input
⚠ Location	Categorical			None	Input
⚠ Status	Categorical			None	Input
# Number_Em...	Continuous			None	Input
# Number_Sto...	Continuous			None	Input
⚠ Manufacturer	Categorical			None	Input
⚠ Problems	Categorical			None	Input
⚠ TV_Categories	Categorical			None	Input
⚠ Potential_Risk	Categorical			None	Input

View current fields
  View unused field settings

Electronics\_Data

Preview Refresh

C:\Users\dineshchaudhary\Desktop\codes\Code\Code\ML\_Data\Electronics\_Data

File Data Filter **Types** Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
ID	Continuous	[1.0,13337...		None	Input
Stereos	Continuous	[0.0,35.0]		None	Input
TVs	Continuous	[0.0,15.0]		None	Input
Speakers	Continuous	[0.0,454.0]		None	Input
Delivery_Pro...	Continuous	[0.0,4.0]		None	Input
Years_as_cu...	Continuous	[2.0,11.0]		None	Input
Estimated_R...	Continuous	[11028.0,9...		None	Input
Payment_Me...	Nominal	"Auto Pay"...		None	Input
Speaker_Dis...	Flag	Regular/Di...		None	Input
Stereo_Disc...	Flag	Regular/Di...		None	Input
Premier	Flag	Yes/No		None	Input
Location	Flag	National/In...		None	Input
Status	Flag	Current/C...		None	Input
Number_Em...	Continuous	[102.0,100...		None	Input
Number_Sto...	Continuous	[1.0,10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categories	Nominal	Low,Mediu...		None	Input
Potential_Risk	Flag	Risk/Regu...		None	Input

View current fields  View unused field settings

OK Cancel Apply Reset

Electronics\_Data

Preview Refresh

\\MacHome\Desktop\Machine Learning\Electronics\_Data

File Data Filter **Types** Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
ID	Continuous	[1.0,13337...		None	Input
Stereos	<Default>	[0.0,35.0]		None	Input
TVs	Continuous	[0.0,15.0]		None	Input
Speakers	Categorical	[0.0,454.0]		None	Input
Delivery_Pro...	Flag	[0.0,4.0]		None	Input
Years_as_cu...	Flag	[2.0,11.0]		None	Input
Estimated_R...	Nominal	[11028.0,9...		None	Input
Payment_Me...	Ordinal	"Auto Pay",...		None	Input
Speaker_Dis...	Typeless	Regular/Di...		None	Input
Stereo_Disc...	Specify...	Regular/Di...		None	Input
Premier	Flag	Yes/No		None	Input
Location	Flag	National/In...		None	Input
Status	Flag	Current/C...		None	Input
Number_Em...	Continuous	[102.0,100...		None	Input
Number_Sto...	Continuous	[1.0,10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categories	Nominal	Low,Mediu...		None	Input
Potential_Risk	Flag	Risk/Regu...		None	Input

View current fields
  View unused field settings

OK Cancel Apply Reset

Electronics\_Data

Preview Refresh

\\MacHome\Desktop\Machine Learning\Electronics\_Data

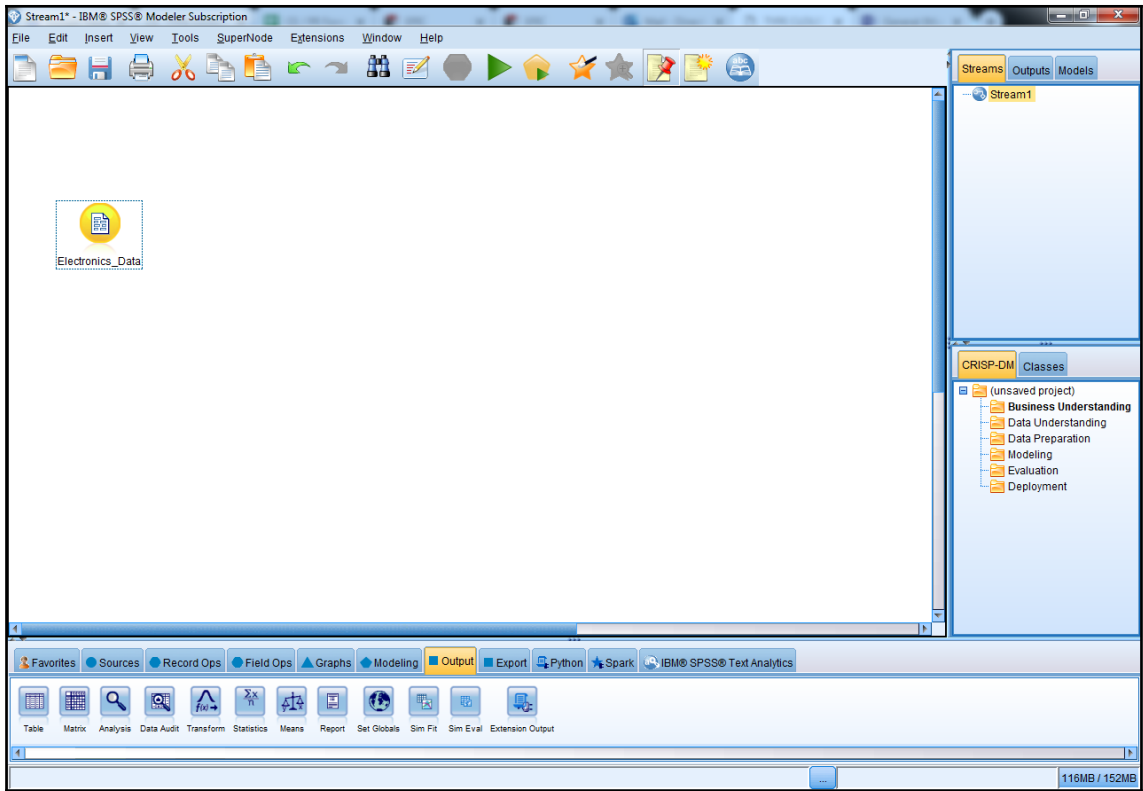
File Data Filter **Types** Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
ID	Typeless			None	None
Stereos	Continuous	[0.0,35.0]		None	Input
TVs	Continuous	[0.0,15.0]		None	Input
Speakers	Continuous	[0.0,454.0]		None	Input
Delivery_Pro...	Continuous	[0.0,4.0]		None	Input
Years_as_cu...	Continuous	[2.0,11.0]		None	Input
Estimated_R...	Continuous	[11028.0,9...		None	Input
Payment_Me...	Nominal	"Auto Pay",...		None	Input
Speaker_Dis...	Flag	Regular/Di...		None	Input
Stereo_Disc...	Flag	Regular/Di...		None	Input
Premier	Flag	Yes/No		None	Input
Location	Flag	National/In...		None	Input
Status	Flag	Current/C...		None	Input
Number_Em...	Continuous	[102.0,100...		None	Input
Number_Sto...	Continuous	[1.0,10.0]		None	Target
Manufacturer	Flag	Yes/No		None	Both
Problems	Flag	T/F		None	None
TV_Categories	Nominal	Low,Mediu...		None	Partition
Potential_Risk	Flag	Risk/Regu...		None	Split

View current fields View unused field settings

OK Cancel Apply Reset



The image displays the IBM SPSS Modeler software interface. The main workspace contains a workflow diagram with two nodes: 'Electronics\_Data' (represented by a document icon) and 'Table' (represented by a grid icon), connected by a blue arrow pointing from left to right.

On the right side, there is a 'Streams' panel with a tab labeled 'Stream1' and a 'CRISP-DM' panel with a 'Classes' tab. The CRISP-DM panel shows a project tree structure:

- (unsaved project)
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
  - Deployment

At the bottom, there is a toolbar with various icons and a row of tabs: 'Sources', 'Record Ops', 'Field Ops', 'Graphs', 'Modeling', 'Output', 'Export', 'Python', 'Spark', and 'IBM SPSS Text Analytics'. Below the tabs, the following icons are visible: Matrix, Analysis, Data Audit, Transform, Statistics, Means, Report, Set Globals, Sim Fit, Sim Eval, and Extension Output.



Table (19 fields, 5,003 records)

File Edit Generate

Table Annotations

	ID	Stereos	TVs	Speakers	Delivery_Problems	Years_as_customer	Estimated_Revenue	Payment_Method
1	1.0...	5.000	8.0...	86.000	0.000	6.000	2753530.000	Check
2	2.0...	0.000	0.0...	4.000	0.000	6.000	6463230.000	Credit Card
3	3.0...	6.000	0.0...	9.000	1.000	8.000	8100090.000	Credit Card
4	4.0...	14.000	6.0...	30.000	0.000	4.000	8746710.000	Credit Card
5	5.0...	14.000	3.0...	33.000	0.000	7.000	8322060.000	Credit Card
6	6.0...	0.000	0.0...	1.000	0.000	9.000	5029070.000	Credit Card
7	7.0...	0.000	0.0...	9.000	0.000	3.000	2085040.000	Check
8	8.0...	2.000	0.0...	20.000	0.000	6.000	8411260.000	Credit Card
9	9.0...	11.000	0.0...	34.000	0.000	10.000	377612.000	Credit Card
10	10....	0.000	0.0...	73.000	0.000	10.000	7386590.000	Auto Pay
11	11....	4.000	0.0...	19.000	0.000	8.000	3093360.000	Auto Pay
12	12....	20.000	0.0...	76.000	0.000	8.000	1230960.000	Credit Card
13	13....	10.000	2.0...	25.000	0.000	10.000	6986400.000	Check
14	14....	20.000	0.0...	103.000	0.000	10.000	9162060.000	Check
15	15....	26.000	1.0...	5.000	0.000	7.000	9650190.000	Credit Card
16	16....	26.000	0.0...	31.000	0.000	9.000	396854.000	Credit Card
17	17....	8.000	0.0...	11.000	0.000	6.000	1377420.000	Check
18	18....	3.000	0.0...	77.000	0.000	5.000	3942820.000	Check
19	19....	24.000	0.0...	63.000	0.000	4.000	498814.000	Credit Card
20	20....	9.000	4.0...	44.000	0.000	6.000	8575380.000	Check

OK

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Electronics\_Data Partition Table

Favorites Sources Record Ops **Field Ops** Graphs Modeling Output Export Python Spark IBM® SPSS® Text Analytics

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Partition

Generate Preview

Settings Annotations

Partition field: Partition

Partitions:  Train and test  Train, test and validation

Training partition size: 50 Label: Training Value = "1\_Training"

Testing partition size: 50 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

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

```
graph LR; Electronics_Data --> Partition; Partition --> Status; Electronics_Data --> Table;
```

Electronics\_Data Partition Status

Table

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export Python Spark IBM® SPSS® Text Analytics

All Supervised Association Segmentation

Auto Classifier Auto Numeric Auto Cluster Time Series TCM Random Trees Tree-AS C&R Tree Decision List Linear Linear-AS C5.0 Regression PCA/Factor Neural Net Feature Selection Discriminant

Status

Objective: Standard model

Fields Build Options Model Options Annotations

Use predefined roles  
 Use custom field assignments

Fields:

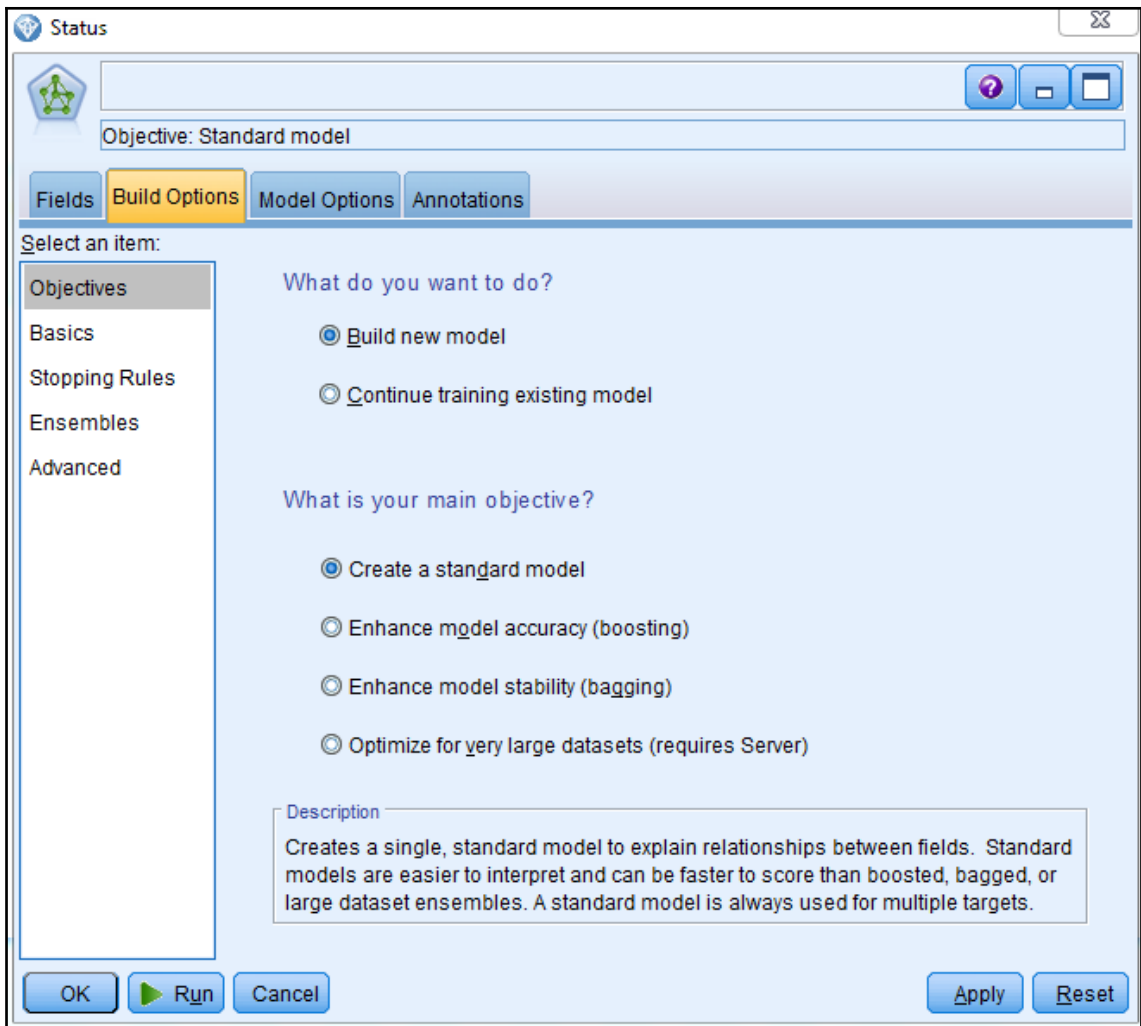
Sort: None

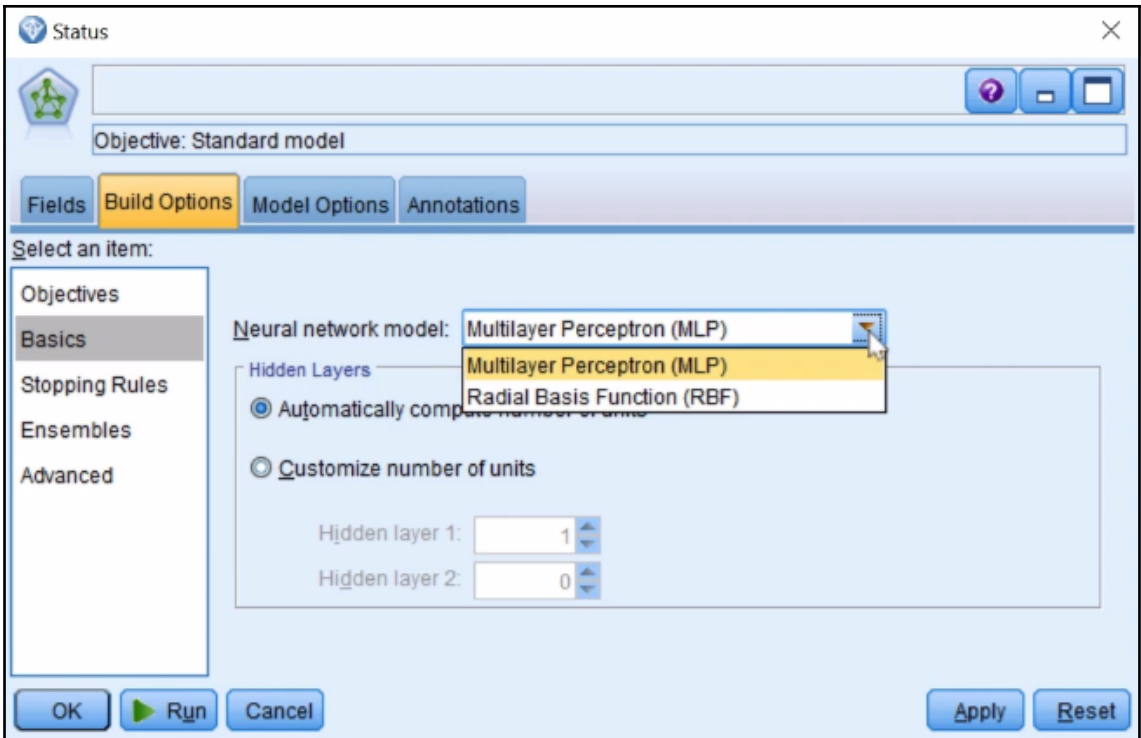
Targets\*:

Predictors (Inputs)\*:

- Stereos
- TVs
- Speakers
- Delivery\_Problems
- Years\_as\_customer
- Estimated\_Revenue
- Payment\_Method
- Speaker\_Discount
- Stereo\_Discount
- Premier

OK Run Cancel Apply Reset





Status

Objective: Standard model

Fields Build Options Model Options Annotations

Select an item:

- Objectives
- Basics
- Stopping Rules
- Ensembles
- Advanced

**i** Stopping rules are available for the Multilayer Perceptron model and apply to training the standard model (or component models if ensembles are in effect). Other stopping rules apply.

Use maximum training time (per component model)

Minutes:

Customize number of maximum training cycles

Maximum number of cycles:

Use minimum accuracy




Accuracy (%):

**i** The maximum training time does not include the time for the calculation of predictor importance.

OK Run Cancel Apply Reset



Status ×


  

Objective: Standard model

**Fields** Build Options Model Options Annotations

Select an item:

- Objectives
- Basics
- Stopping Rules
- Ensembles**
- Advanced

 These settings determine the behavior of ensembling that occurs when boosting, bagging, or very large datasets are requested in Objectives. Options that do not apply are ignored.

Bagging and Very Large Datasets

Default combining rule for categorical targets:

Default combining rule for continuous targets:

Boosting and Bagging

Number of component models for boosting and/or bagging:

Status

Objective: Standard model

Fields Build Options Model Options Annotations

Select an item:

- Objectives
- Basics
- Stopping Rules
- Ensembles
- Advanced

**i** Neural Network modeling internally separates records into a model building set and an overfit prevention set. Specify a percentage of records for the overfit prevention set.

Overfit prevention set (%):

Replicate Results

Random seed:

Missing values in predictors:

- Delete listwise
- Impute missing values

Stream1 - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

```
graph LR; Electronics_Data[Electronics_Data] --> Table[Table]; Electronics_Data --> Partition[Partition]; Partition --> Status1[Status]; Partition --> Status2[Status];
```

The screenshot displays the IBM SPSS Modeler Stream interface. The main workspace contains a workflow diagram with the following nodes and connections:

- Electronics\_Data** (Source node) connects to **Table** (Table node) and **Partition** (Partition node).
- Partition** connects to two **Status** (Status nodes).

On the right side, there is a **Streams** panel with a **Status** node. Below it is a **CRISP-DM** **Classes** panel showing a project structure:

- (unsaved project)
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
  - Deployment

At the bottom, there is a **Modeling** toolbar with various modeling techniques:

- Auto Classifier
- Auto Numeric
- Auto Cluster
- Time Series
- TCM
- Random Trees
- Tree-AS
- C&R Tree
- Decision List
- Linear
- Linear-AS
- C5.0
- Regression
- PCA-Factor
- Neural Net
- Feature Selection
- Discriminant
- Logistic
- GenLin
- GLMM
- GLE
- Co

128MB / 152MB

Status

File Generate Preview

Model Settings Annotations

### Model Summary

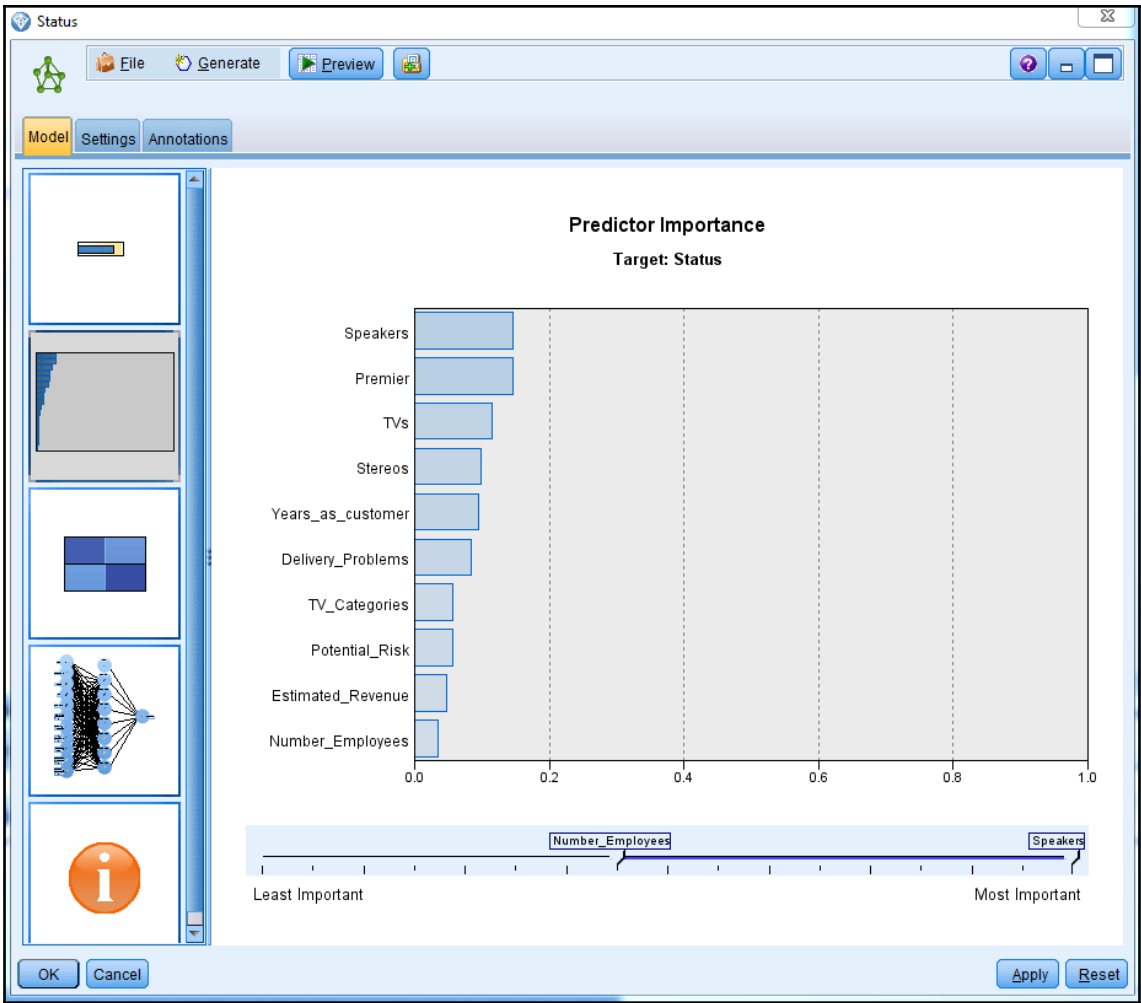
Target	Status
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	7

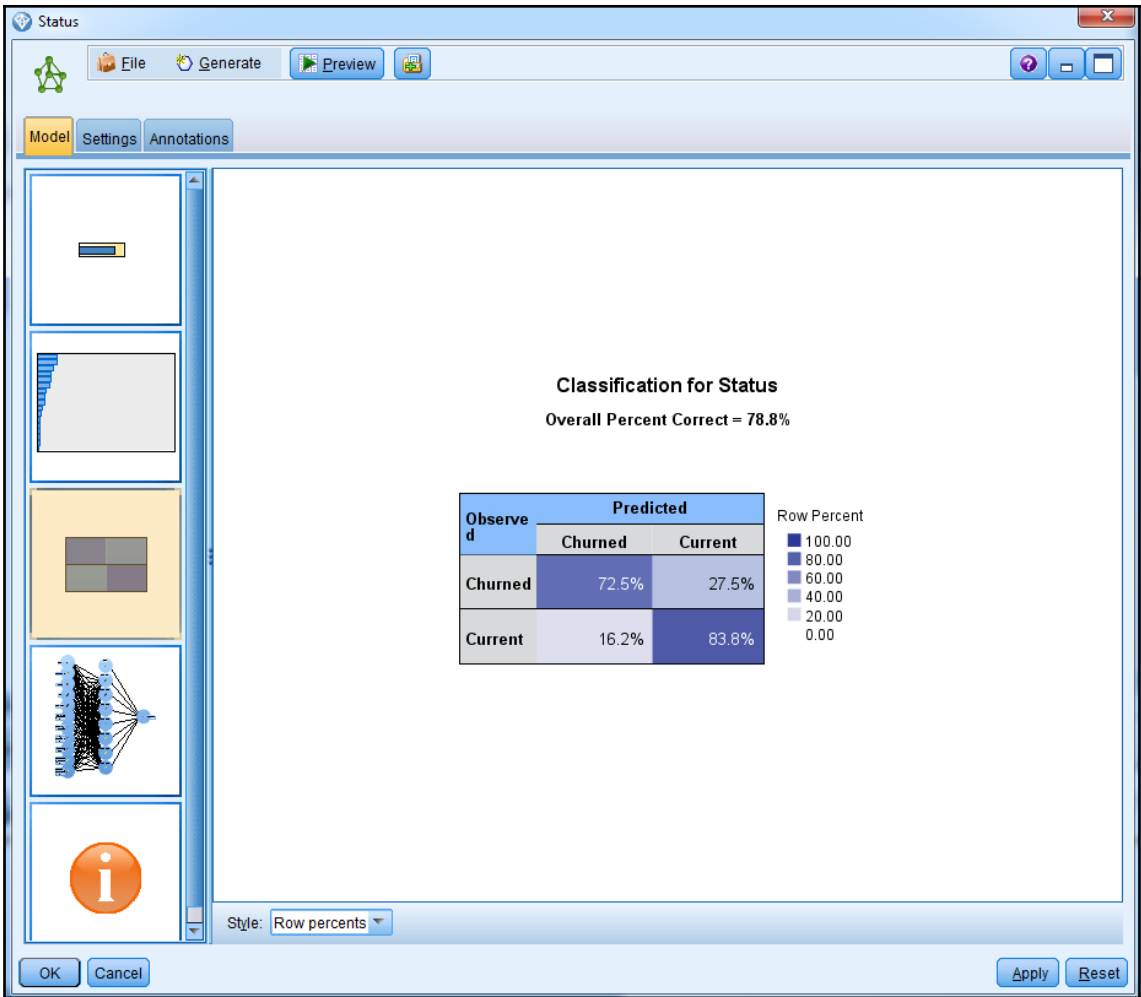
Worse Better

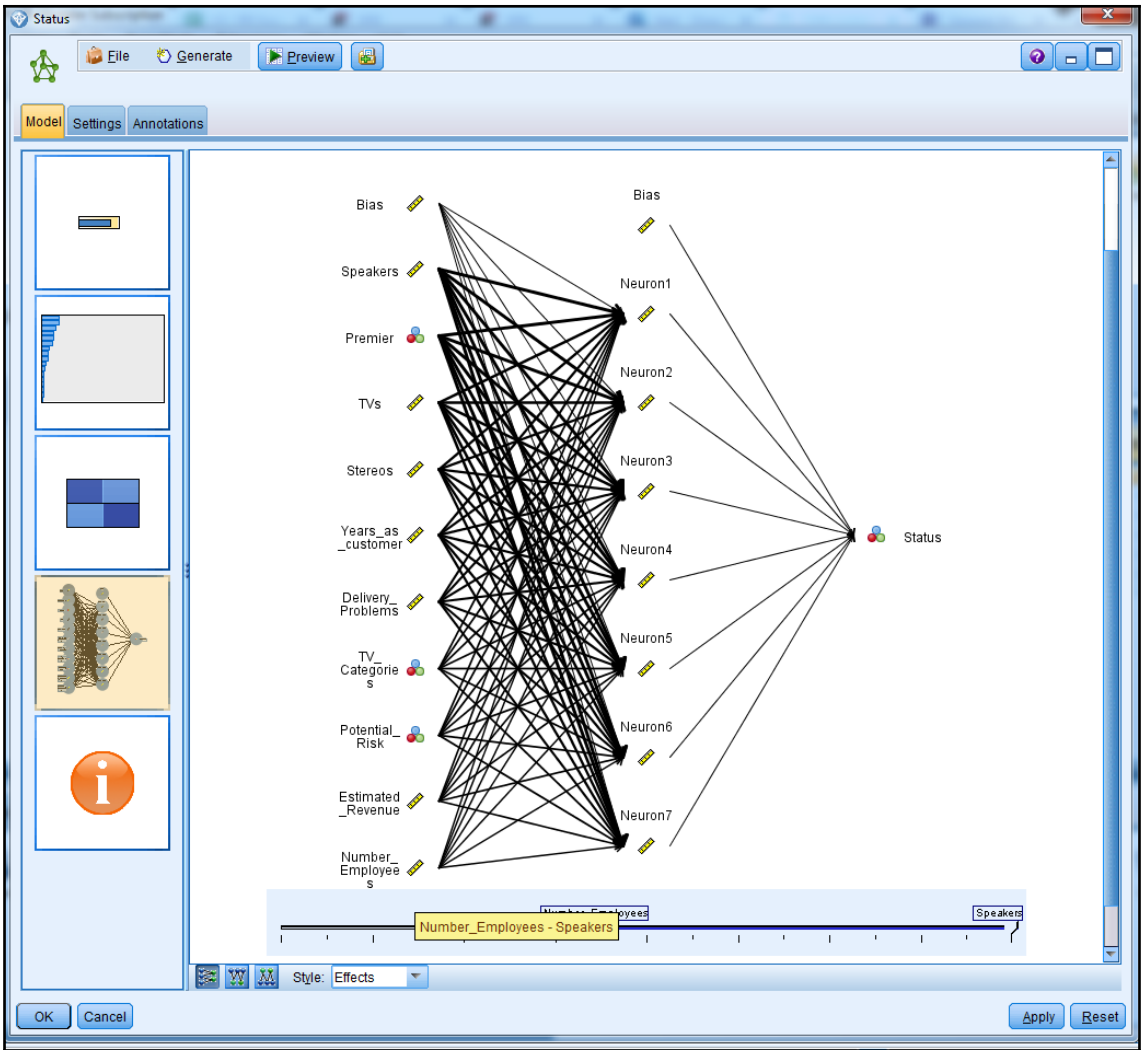
0% 25% 50% 75% 100% Accuracy

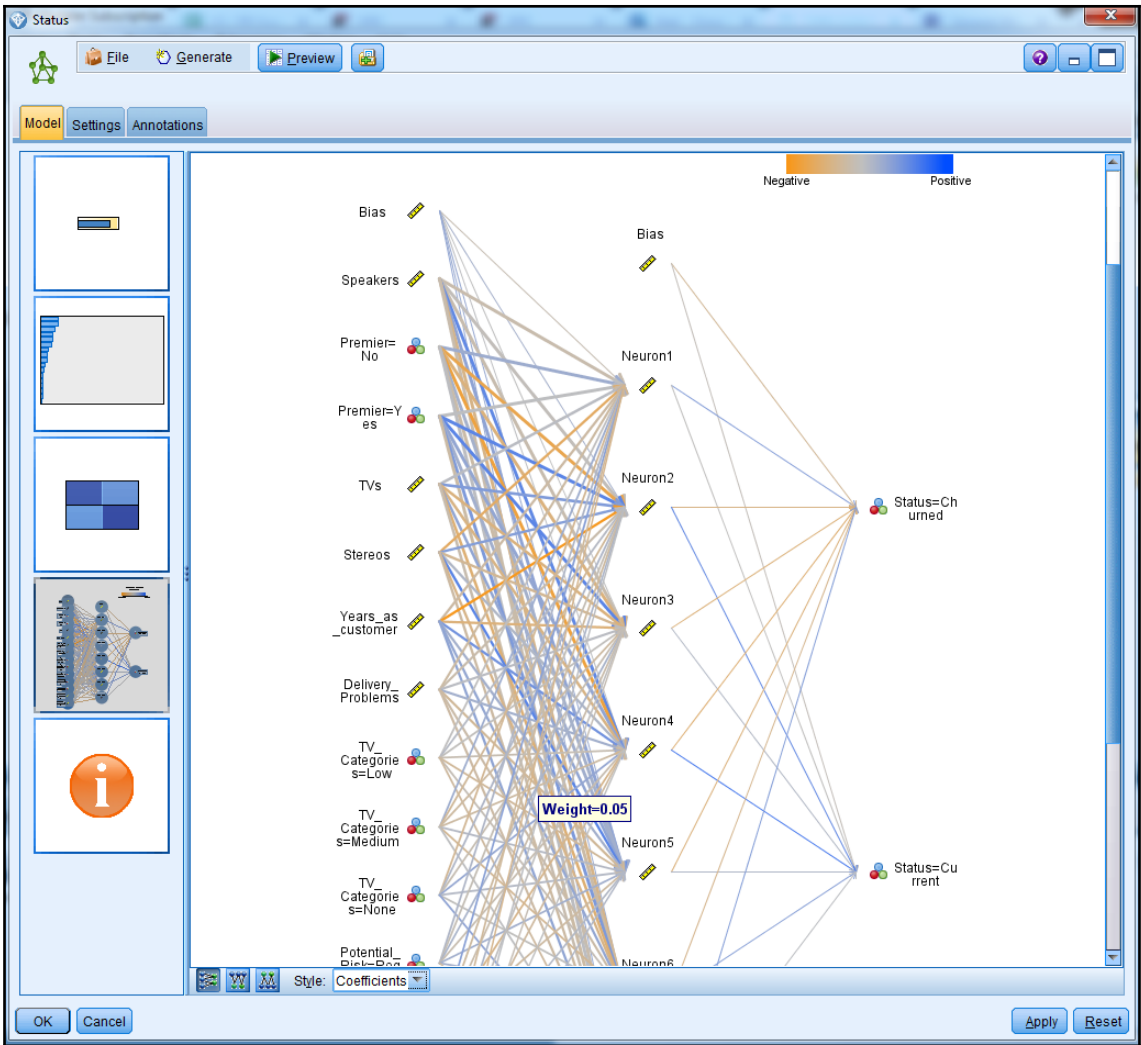
78.8%

OK Cancel Apply Reset







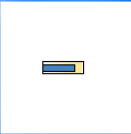

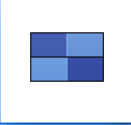






Status

File Generate Preview

Model Settings Annotations

**i** Neural Networks

Neural networks predict a continuous or categorical target based on one or more predictors by finding unknown and possibly complex patterns in the data.

Fields

- Target
  - Status
- Predictors (Inputs)
  - Use partitioned data: true
- Build Options
- Training Summary

OK Cancel Apply Reset

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Electronics\_Data Partition Status Status Table

Streams  
Outputs Models  
Status

CRISP-DM Classes  
[unsaved project]  
Business Und  
Data Understa  
Data Preparati  
Modeling  
Evaluation  
Deployment

Modeling Output Export Python Spark IBM® SPSS® Text Analytics

All Supervised Association Segmentation  
Auto Classifier Auto Numeric Auto Cluster Time Series TCM Random Trees Tree-AS CAR Tree Decision List Linear Linear-AS C5.0 Regression PCA-Factor Neural Net Feature Selection Discriminant Logistic GenLin GLMM GLE C

158MB / 170MB

Table (22 fields, 5,003 records)

File Edit Generate

Table Annotations

	_Stores	Manufacturer	Problems	TV_Categories	Potential_Risk	Partition	\$N-Status	\$NC-Status	
1	9.000	No	F	Medium	Regular	1_Training	Churned		0.929
2	3.000	Yes	F	None	Regular	1_Training	Churned		0.803
3	6.000	No	T	None	Regular	2_Testing	Churned		0.879
4	4.000	Yes	F	Medium	Regular	2_Testing	Current		0.735
5	9.000	No	F	Low	Regular	1_Training	Current		0.886
6	4.000	Yes	F	None	Regular	1_Training	Churned		0.830
7	3.000	Yes	F	None	Regular	2_Testing	Churned		0.617
8	8.000	No	F	None	Regular	1_Training	Current		0.736
9	9.000	Yes	F	None	Regular	1_Training	Churned		0.785
10	2.000	Yes	F	None	Regular	1_Training	Churned		0.770
11	1.000	No	F	None	Regular	2_Testing	Current		0.811
12	2.000	No	F	None	Regular	2_Testing	Current		0.577
13	5.000	Yes	F	Low	Regular	2_Testing	Current		0.740
14	5.000	No	F	None	Regular	1_Training	Churned		0.650
15	4.000	No	F	Low	Regular	2_Testing	Current		0.934
16	6.000	No	F	None	Regular	2_Testing	Current		0.991
17	3.000	No	F	None	Regular	1_Training	Churned		0.806
18	9.000	No	F	None	Regular	1_Training	Current		0.892
19	8.000	No	F	None	Regular	2_Testing	Current		0.568
20	2.000	No	F	Low	Regular	1_Training	Current		0.858

1

OK

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

The central workspace displays a workflow diagram with the following nodes and connections:

- Electronics\_Data** (Data Source) connects to **Partition** (Partitioning Node).
- Partition** connects to **Status** (Status Node).
- Status** connects to **Table** (Table Node).
- Table** connects to **Status** (Status Node).
- Status** connects to **Analysis** (Analysis Node).

Right-hand pane:





- Streams: Status
- CRISP-DM: Classes
  - (unsaved project)
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
  - Deployment

Bottom toolbar:

- Favorites Sources Record Ops Field Ops Graphs Modeling Output Export Python Spark IBM® SPSS® Text Analytics
- Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output

147MB / 156MB

Analysis X

 Analyze \$N-Status   

**Analysis** Output Annotations

Coincidence matrices (for symbolic targets)

Performance evaluation

Evaluation metric (AUC & Gini, binary classifiers only)

Confidence figures (if available)

Threshold for:  % correct

Improve accuracy:  fold

Find predicted/predictor fields using:



Model output field metadata

Field name format (for example, '\$<x>-<target field>')

Separate by partition

User defined analysis

Break down analysis by fields:

Analysis of [Status]

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

- Comparing \$N-Status with Status
 

'Partition'	1_Training		2_Testing	
Correct	1,936	78.83%	1,976	77.58%
Wrong	520	21.17%	571	22.42%
Total	2,456		2,547	
- Coincidence Matrix for \$N-Status (rows show actuals)
 

'Partition' = 1_Training		Churned	Current
Churned		790	299
Current		221	1,146
'Partition' = 2_Testing		Churned	Current
Churned		745	331
Current		240	1,231

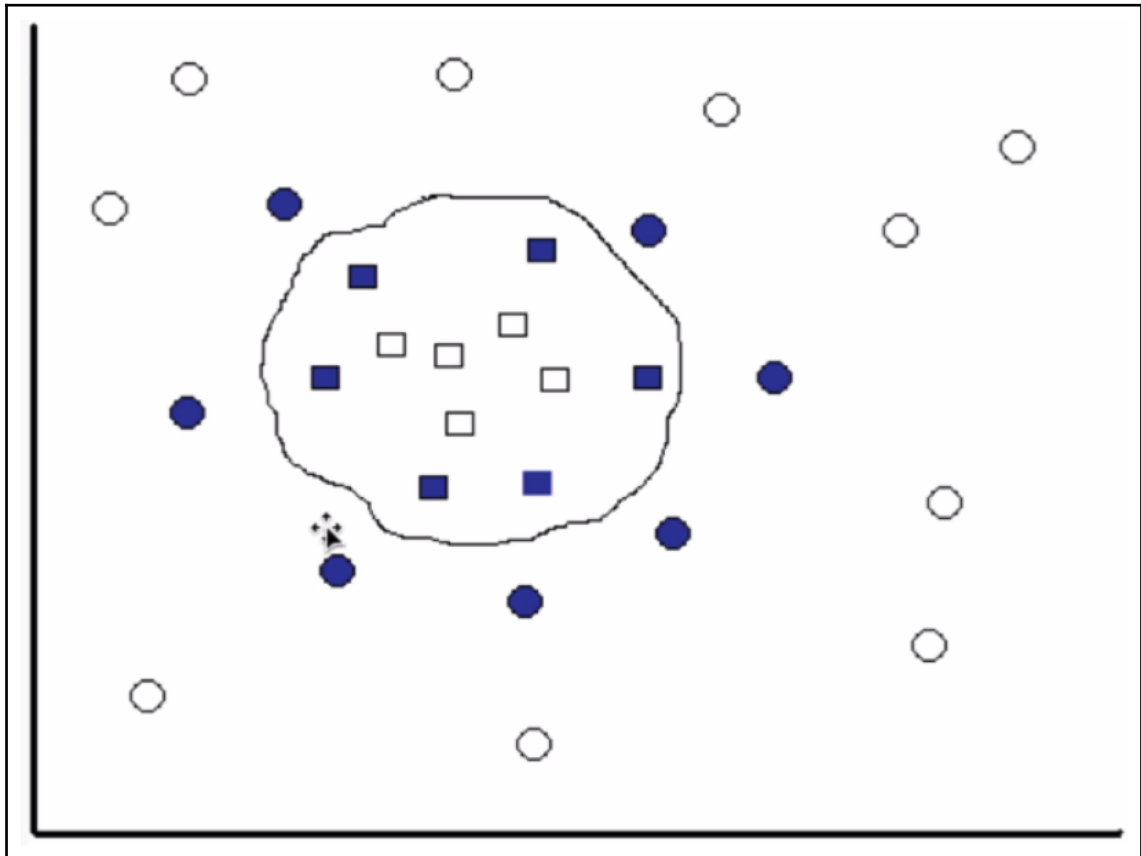
OK

Seed	Overall	Churned	Current	Consistent
229176228	77.58%	745 (68.4%)	1231 (83.7%)	Yes

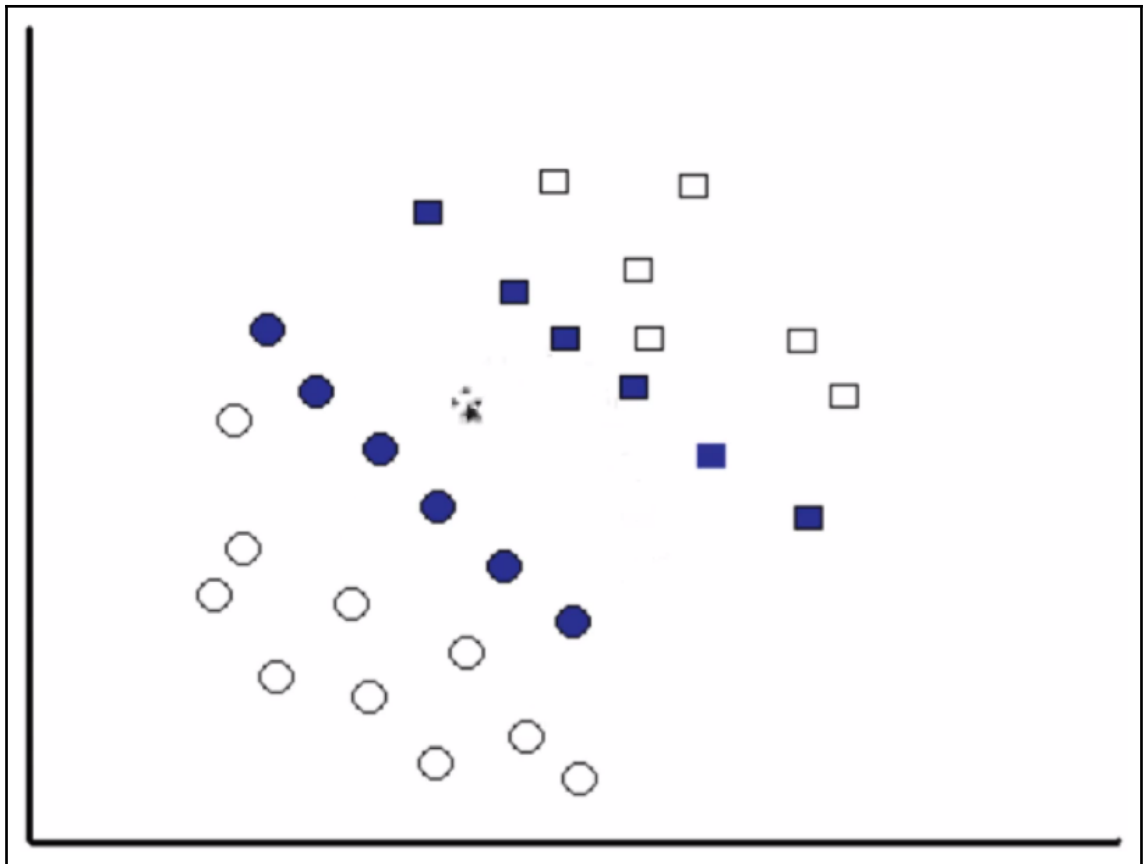
<b>229176228</b>
Speakers
Premier
TVs
Stereos
Years as customer
Delivery problems
TV categories
Potential risk
Estimated revenue
Number employees

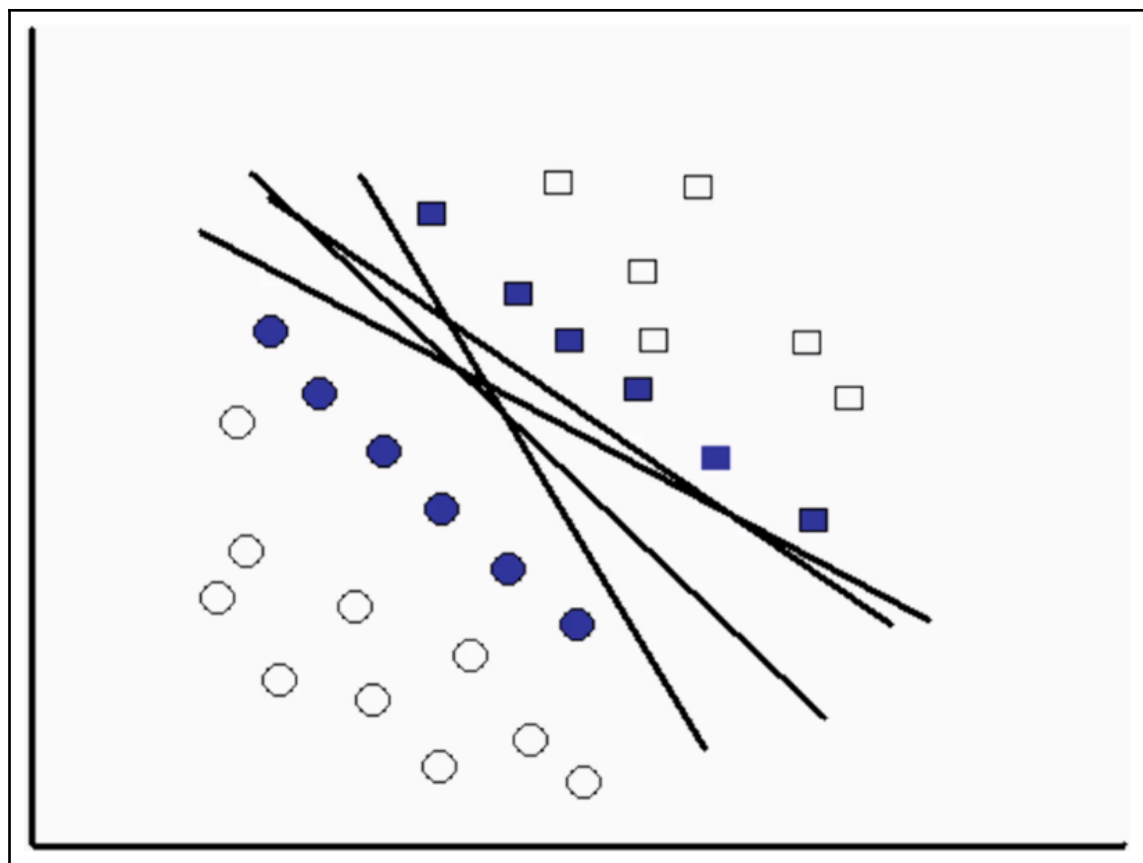
<b>Seed</b>	<b>Overall</b>	<b>Churned</b>	<b>Current</b>	<b>Consistent</b>
229176228	77.58%	745 (68.4%)	1231 (83.7%)	Yes
641835376	78.17%	779 (71.5%)	1212 (82.4%)	Yes
1	79.7%	837 (76.9%)	1193 (81.1%)	Yes
2552	78.41%	787 (72.3%)	1210 (82.3%)	Yes
5000	80.29%	877 (80.5%)	1168 (79.4%)	Yes

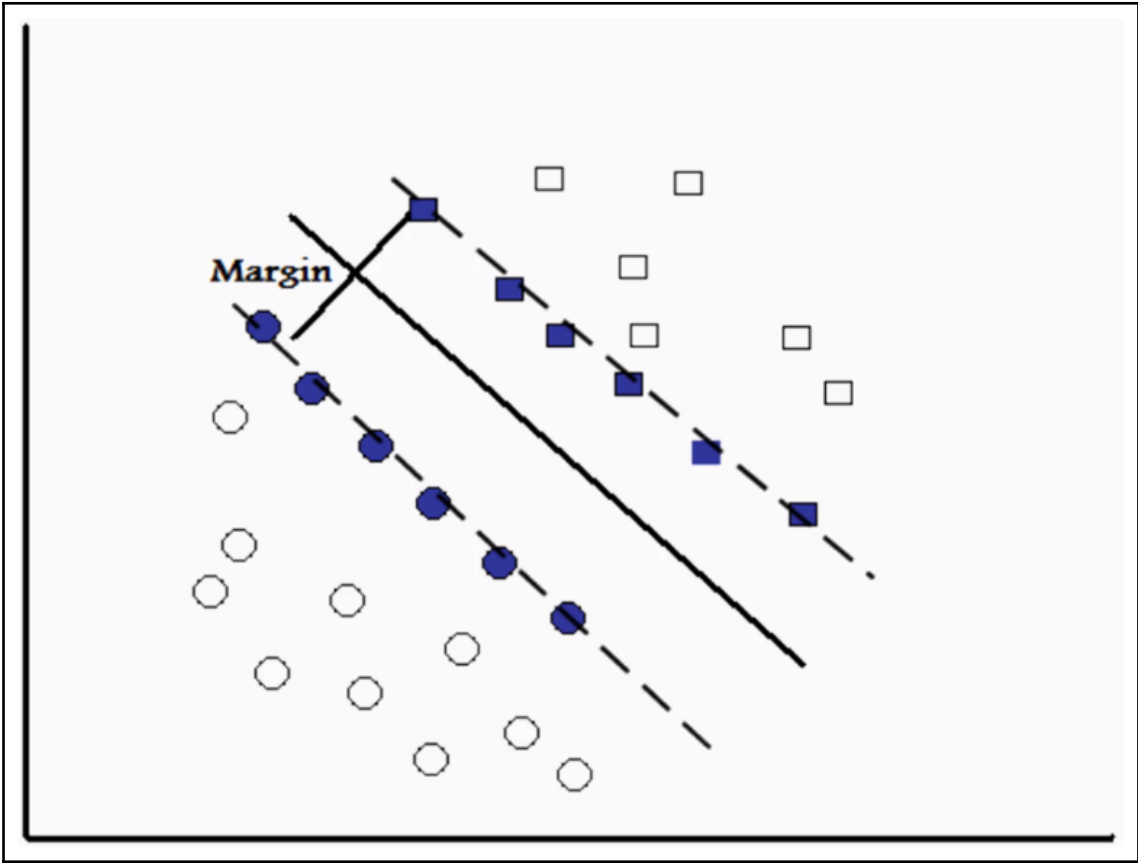
229176228	641835376	1	2552	5000
Speakers	Premier	Premier	Premier	Premier
Premier	Speakers	Years as customer	Stereos	Years as customer
TVs	Years as customer	Speakers	Speakers	Stereos
Stereos	Stereos	TV categories	TVs	TVs
Years as customer	TVs	Stereos	Years as customer	Speakers
Delivery problems	Delivery problems	TVs	Delivery problems	TV categories
TV categories	Estimated revenue	Estimated revenue	Potential risk	Number employees
Potential risk	TV categories	Payment method	TV categories	Estimated revenue
Estimated revenue	Problems	Number employees	Estimated revenue	Delivery problems
Number employees	Number stores	Delivery problems	Number employees	Problems











Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

The screenshot displays the IBM SPSS Modeler interface. The main workspace contains a workflow diagram with three nodes: 'Electronics\_Data' (source), 'Partition' (node), and 'Table' (target). The 'Partition' node is highlighted with a dashed border. The right-hand pane shows the 'Streams' tab with 'Stream1' and the 'CRISP-DM' tab with a 'Classes' section containing a project structure: (unsaved project), Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The bottom toolbar includes tabs for Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, Python, Spark, and IBM® SPSS® Text Analytics. The Modeling tab is active, showing a list of supervised association and segmentation algorithms: CART, Decision List, Linear, Linear-AS, C5.0, Regression, PCA Factor, Neural Net, Feature Selection, Discriminant, Logistic, GenLin, GLMM, GLE, Cox, SVM, LVM, Bayes Net, SLRM, Apriori, Cam, Sequence, K-Means, and K.

Streams Outputs Models

Stream1

CRISP-DM Classes

(unsaved project)

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

1

1

113MB / 156MB

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Stream1

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Electronics\_Data Partition Status

Table




Favorites Sources Record Ops Field Ops Graphs Modeling Output Export Python Spark IBM® SPSS® Text Analytics

All Supervised Association Segmentation

Decision Tree Decision List Linear Linear-AS C5.0 Regression PCA/Factor Neural Net Feature Selection Discriminant Logistic GenLin GLMM GLE Cox SVM LSVM Bayes Net SLRM Apriori Camm Sequence K-Means Ko

114MB / 156MB

Status ×

Fields Model **Expert** Analyze Annotations

Mode:  Simple  Expert

Append all probabilities (valid only for categorical targets)

Stopping criteria:



Regularization parameter (C):

Regression precision (epsilon):




Kernel type:

RBF gamma:  Bias:

Gamma:  Degree:

OK  Run Cancel Apply  Reset

Status X

Fields Model **Expert** Analyze Annotations

Mode:  Simple  Expert

Append all probabilities (valid only for categorical targets)

Stopping criteria: 1.0E-3



Regularization parameter (C): 5

Regression precision (epsilon): 0.1

Kernel type: Sigmoid

RBF gamma: 0.1 Bias: 0.0

Gamma: 1.0 Degree: 3

OK  Run Cancel Apply  Reset

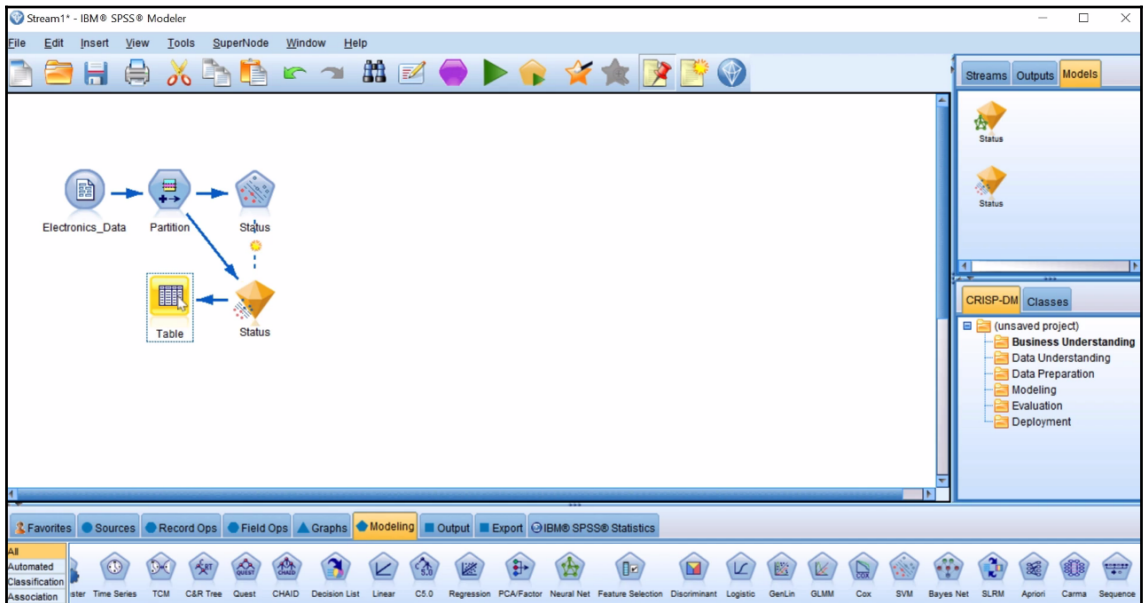
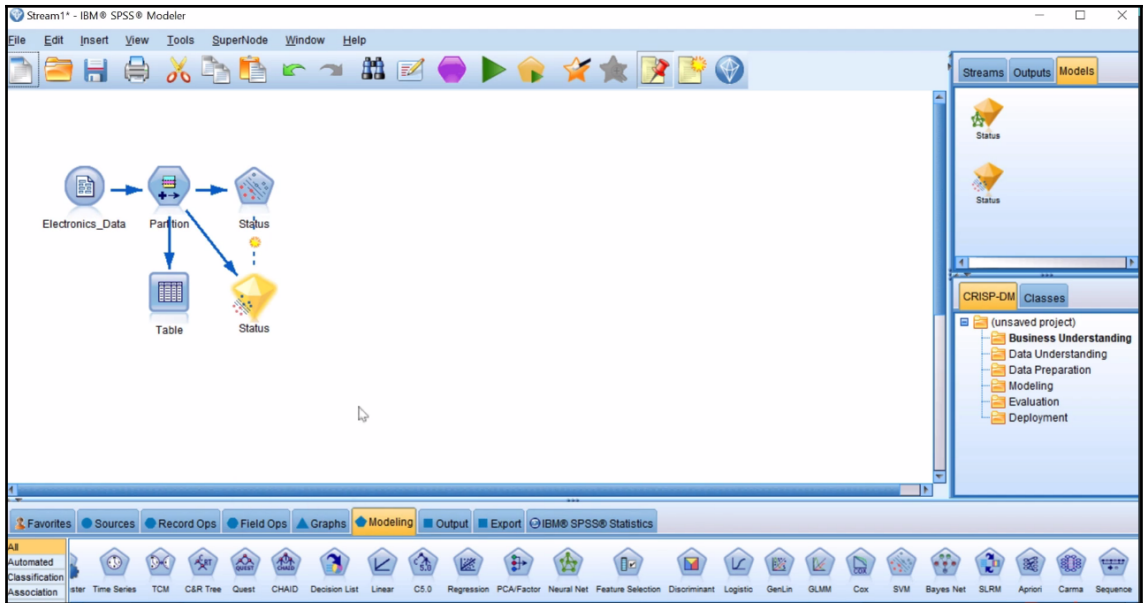




Table (22 fields, 5,003 records)

File Edit Generate

Table Annotations

	r_Stores	Manufacturer	Problems	TV_Categories	Potential_Risk	Partition	\$\$-Status	\$SP-Status	
1	9.000	No	F	Medium	Regular	1_Training	Churned	0.670	
2	3.000	Yes	F	None	Regular	1_Training	Churned	0.670	
3	6.000	No	T	None	Regular	2_Testing	Churned	0.670	
4	4.000	Yes	F	Medium	Regular	2_Testing	Current	0.803	
5	9.000	No	F	Low	Regular	1_Training	Current	0.803	
6	4.000	Yes	F	None	Regular	1_Training	Churned	0.670	
7	3.000	Yes	F	None	Regular	2_Testing	Churned	0.670	
8	8.000	No	F	None	Regular	1_Training	Current	0.803	
9	9.000	Yes	F	None	Regular	1_Training	Churned	0.670	
10	2.000	Yes	F	None	Regular	1_Training	Churned	0.670	
11	1.000	No	F	None	Regular	2_Testing	Current	0.803	
12	2.000	No	F	None	Regular	2_Testing	Churned	0.670	
13	5.000	Yes	F	Low	Regular	2_Testing	Current	0.803	
14	5.000	No	F	None	Regular	1_Training	Churned	0.670	
15	4.000	No	F	Low	Regular	2_Testing	Current	0.803	
16	6.000	No	F	None	Regular	2_Testing	Current	0.803	
17	3.000	No	F	None	Regular	1_Training	Churned	0.670	
18	9.000	No	F	None	Regular	1_Training	Current	0.803	
19	8.000	No	F	None	Regular	2_Testing	Churned	0.670	
20	2.000	No	F	Low	Regular	1_Training	Current	0.803	

1

OK

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Electronics\_Data Partition Status Status Analysis

Table

Streams

Outputs Models

Status

CRISP-DM Classes

- (unsaved project)
- Business Understar
- Data Preparation
- Modeling
- Evaluation
- Deployment

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globale Sim Fit Sim Eval Extension Output

160MB / 170MB

Analysis of [Status]

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

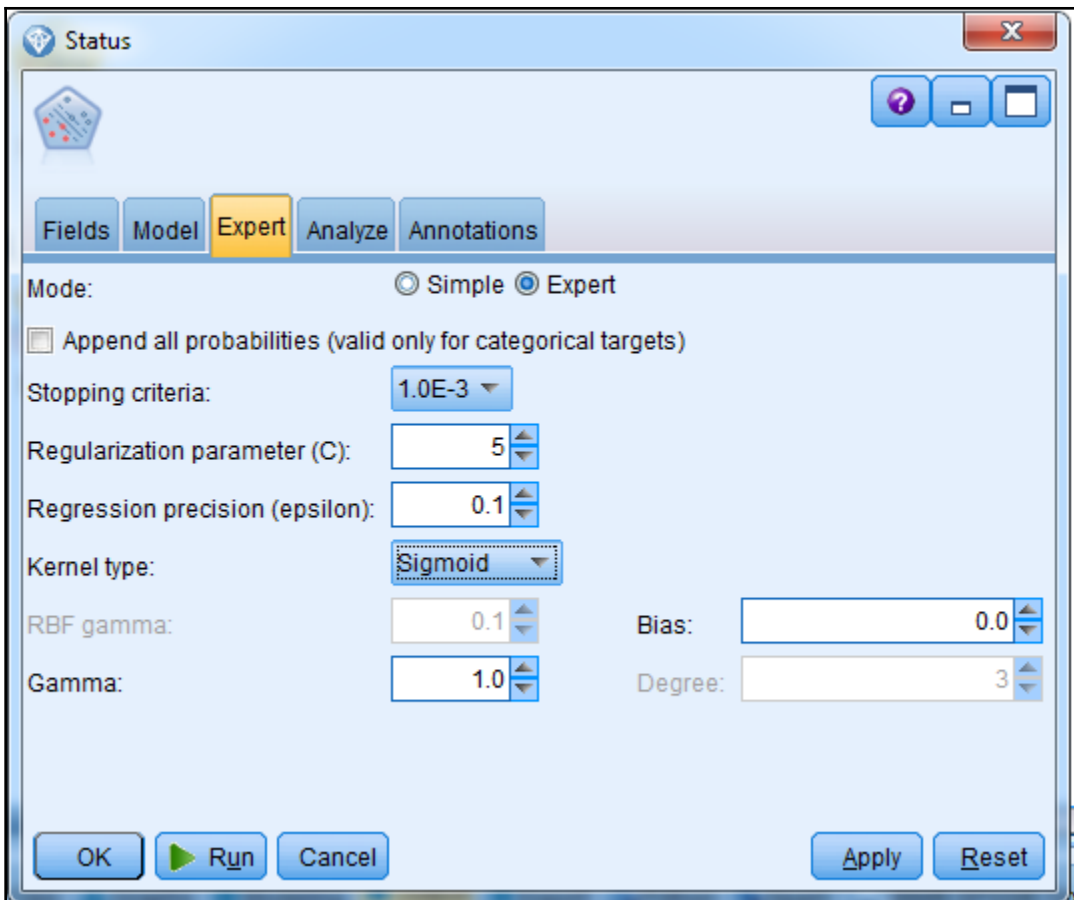
- Comparing \$\$-Status with Status
 

'Partition'	1_Training		2_Testing	
Correct	1,803	73.41%	1,870	73.42%
Wrong	653	26.59%	677	26.58%
Total	2,456		2,547	
- Coincidence Matrix for \$\$-Status (rows show actuals)
 

'Partition' = 1_Training	Churned	Current
Churned	858	231
Current	422	945
'Partition' = 2_Testing	Churned	Current
Churned	831	245
Current	432	1,039

OK

Model	Overall	Churned	Current	Consistent
Linear C=5	73.42%	831 (76.3%)	1039 (70.6%)	Yes



Analysis of [Status] #1

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Comparing \$\$-Status with Status

'Partition'	1_Training		2_Testing	
Correct	1,367	55.66%	1,471	57.75%
Wrong	1,089	44.34%	1,076	42.25%
Total	2,456		2,547	

Coincidence Matrix for \$\$-Status (rows show actuals)

'Partition' = 1_Training	Current
Churned	1,089
Current	1,367

'Partition' = 2_Testing	Current
Churned	1,076
Current	1,471

OK

Model	Overall	Churned	Current	Consistent
Linear C=5	73.42%	831 (76.3%)	1039 (70.6%)	Yes
Polynomial C=5; Degree = 2	80.17%	820 (75.3%)	1222 (83.1%)	Yes
RBF C=5; RBF gamma = .27	79.94%	842 (77.3%)	1194 (81.2%)	No
Sigmoid C=5; Gamma = 1	57.75%	0 (0%)	1471 (100%)	Yes

## Chapter 3: Understanding Models

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

Current Salary = 135.9 + 1.7(Beginning Salary) + 298  
(Education Level) – 59 (Age)

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

Current Salary = 135.9 + 1.7(Beginning Salary) + 298 (Education Level) – 59 (Age)

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

$$\text{Current Salary} = 135.9 + 1.7(\text{Beginning Salary}) + 298(\text{Education Level}) - 59(\text{Age})$$



Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

$$\text{Current Salary} = 135.9 + 1.7(\text{Beginning Salary}) + 298(\text{Education Level}) - 59(\text{Age})$$

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

$$\text{Current Salary} = 135.9 + 1.7(\text{Beginning Salary}) + 298(\text{Education Level}) - 59(\text{Age})$$

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	135.904	983.582		.138	.890
	salbeg	1.734	.059	.799	29.331	.000
	edlevel	298.049	67.220	.126	4.434	.000
	age	-58.950	12.734	-.102	-4.629	.000

Linear Regression Formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_ix_i$$

Current Salary = 135.9 + 1.7(Beginning Salary) + 298  
(Education Level) – 59 (Age)

Model Viewer Summary Settings Annotations

1 2 3 4 5 All

- Premier in ["No"] [Mode: Churned] (1,280)
  - Years\_as\_customer <= 3 [Mode: Current] (193)
    - Estimated\_Revenue <= 4,114,620 [Mode: Churned] (72)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (25; 1.0)
      - Location in ["National"] [Mode: Churned] ⇒ Churned (47; 0.66)
    - Estimated\_Revenue > 4,114,620 and Estimated\_Revenue <= 5,029,070 [Mode: Current] ⇒ Current (29; 0.586)
    - Estimated\_Revenue > 5,029,070 and Estimated\_Revenue <= 6,898,750 [Mode: Current] ⇒ Current (39; 0.923)
    - Estimated\_Revenue > 6,898,750 [Mode: Current] ⇒ Current (53; 0.736)
  - Years\_as\_customer > 3 and Years\_as\_customer <= 4 [Mode: Current] (156)
    - TV\_Categories in ["Low" "Medium"] [Mode: Current] (63)
      - Stereos <= 11 [Mode: Churned] ⇒ Churned (25; 0.8)
      - Stereos > 11 [Mode: Current] ⇒ Current (38; 0.763)
    - TV\_Categories in ["None"] [Mode: Current] (93)
      - Stereo\_Discount in ["Discount"] [Mode: Current] ⇒ Current (28; 1.0)
      - Stereo\_Discount in ["Regular"] [Mode: Current] ⇒ Current (65; 0.846)
  - Years\_as\_customer > 4 [Mode: Churned] (931)
    - Speakers <= 5 [Mode: Churned] ⇒ Churned (109; 0.982)
    - Speakers > 5 and Speakers <= 11 [Mode: Churned] (87)
      - Stereos <= 7 [Mode: Churned] ⇒ Churned (29; 0.931)
      - Stereos > 7 and Stereos <= 24 [Mode: Churned] ⇒ Churned (31; 0.516)
      - Stereos > 24 [Mode: Churned] ⇒ Churned (27; 0.815)
    - Speakers > 11 and Speakers <= 18 [Mode: Churned] ⇒ Churned (90; 0.878)
    - Speakers > 18 and Speakers <= 27 [Mode: Churned] (76)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (27; 0.778)
      - Location in ["National"] [Mode: Current] ⇒ Current (49; 0.531)
    - Speakers > 27 and Speakers <= 63 [Mode: Churned] (271)
      - Payment\_Method in ["Auto Pay"] [Mode: Churned] ⇒ Churned (43; 0.674)
      - Payment\_Method in ["Check" "Credit Card"] [Mode: Churned] (228)
        - TVs <= 3 [Mode: Churned] ⇒ Churned (184; 0.815)

Model Viewer Summary Settings Annotations

1 2 3 4 5 All

- Premier in ["No"] [Mode: Churned] (1,280)
  - Years\_as\_customer <= 3 [Mode: Current] (193)
    - Estimated\_Revenue <= 4,114,620 [Mode: Churned] (72)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (25; 1.0)
      - Location in ["National"] [Mode: Churned] ⇒ Churned (47; 0.66)
    - Estimated\_Revenue > 4,114,620 and Estimated\_Revenue <= 5,029,070 [Mode: Current] ⇒ Current (29; 0.586)
    - Estimated\_Revenue > 5,029,070 and Estimated\_Revenue <= 6,898,750 [Mode: Current] ⇒ Current (39; 0.923)
    - Estimated\_Revenue > 6,898,750 [Mode: Current] ⇒ Current (53; 0.736)
  - Years\_as\_customer > 3 and Years\_as\_customer <= 4 [Mode: Current] (156)
    - TV\_Categories in ["Low" "Medium"] [Mode: Current] (63)
      - Stereos <= 11 [Mode: Churned] ⇒ Churned (25; 0.8)
      - Stereos > 11 [Mode: Current] ⇒ Current (38; 0.763)
    - TV\_Categories in ["None"] [Mode: Current] (93)
      - Stereo\_Discount in ["Discount"] [Mode: Current] ⇒ Current (28; 1.0)
      - Stereo\_Discount in ["Regular"] [Mode: Current] ⇒ Current (65; 0.846)
  - Years\_as\_customer > 4 [Mode: Churned] (931)
    - Speakers <= 5 [Mode: Churned] ⇒ Churned (109; 0.982)
    - Speakers > 5 and Speakers <= 11 [Mode: Churned] (87)
      - Stereos <= 7 [Mode: Churned] ⇒ Churned (29; 0.931)
      - Stereos > 7 and Stereos <= 24 [Mode: Churned] ⇒ Churned (31; 0.516)
      - Stereos > 24 [Mode: Churned] ⇒ Churned (27; 0.815)
    - Speakers > 11 and Speakers <= 18 [Mode: Churned] ⇒ Churned (90; 0.878)
    - Speakers > 18 and Speakers <= 27 [Mode: Churned] (76)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (27; 0.778)
      - Location in ["National"] [Mode: Current] ⇒ Current (49; 0.531)
    - Speakers > 27 and Speakers <= 63 [Mode: Churned] (271)
      - Payment\_Method in ["Auto Pay"] [Mode: Churned] ⇒ Churned (43; 0.674)
      - Payment\_Method in ["Check" "Credit Card"] [Mode: Churned] (228)
        - TVs <= 3 [Mode: Churned] ⇒ Churned (184; 0.815)

Model Viewer Summary Settings Annotations

1 2 3 4 5 All

- Premier in ["No"] [Mode: Churned] (1,280)
  - Years\_as\_customer <= 3 [Mode: Current] (193)
    - Estimated\_Revenue <= 4,114,620 [Mode: Churned] (72)**
      - Location in ["International"] [Mode: Churned] ⇒ Churned (25; 1.0)
      - Location in ["National"] [Mode: Churned] ⇒ Churned (47; 0.66)
    - Estimated\_Revenue > 4,114,620 and Estimated\_Revenue <= 5,029,070 [Mode: Current] ⇒ Current (29; 0.586)
    - Estimated\_Revenue > 5,029,070 and Estimated\_Revenue <= 6,898,750 [Mode: Current] ⇒ Current (39; 0.923)
    - Estimated\_Revenue > 6,898,750 [Mode: Current] ⇒ Current (53; 0.736)
  - Years\_as\_customer > 3 and Years\_as\_customer <= 4 [Mode: Current] (156)
    - TV\_Categories in ["Low" "Medium"] [Mode: Current] (63)
      - Stereos <= 11 [Mode: Churned] ⇒ Churned (25; 0.8)
      - Stereos > 11 [Mode: Current] ⇒ Current (38; 0.763)
    - TV\_Categories in ["None"] [Mode: Current] (93)
      - Stereo\_Discount in ["Discount"] [Mode: Current] ⇒ Current (28; 1.0)
      - Stereo\_Discount in ["Regular"] [Mode: Current] ⇒ Current (65; 0.846)
  - Years\_as\_customer > 4 [Mode: Churned] (931)
    - Speakers <= 5 [Mode: Churned] ⇒ Churned (109; 0.982)
    - Speakers > 5 and Speakers <= 11 [Mode: Churned] (87)
      - Stereos <= 7 [Mode: Churned] ⇒ Churned (29; 0.931)
      - Stereos > 7 and Stereos <= 24 [Mode: Churned] ⇒ Churned (31; 0.516)
      - Stereos > 24 [Mode: Churned] ⇒ Churned (27; 0.815)
    - Speakers > 11 and Speakers <= 18 [Mode: Churned] ⇒ Churned (90; 0.878)
    - Speakers > 18 and Speakers <= 27 [Mode: Churned] (76)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (27; 0.778)
      - Location in ["National"] [Mode: Current] ⇒ Current (49; 0.531)
    - Speakers > 27 and Speakers <= 63 [Mode: Churned] (271)
      - Payment\_Method in ["Auto Pay"] [Mode: Churned] ⇒ Churned (43; 0.674)
      - Payment\_Method in ["Check" "Credit Card"] [Mode: Churned] (228)
        - TVs <= 3 [Mode: Churned] ⇒ Churned (184; 0.815)

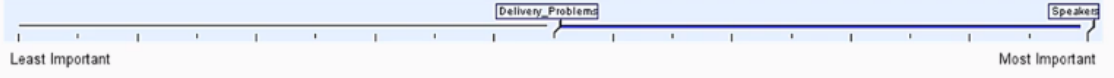
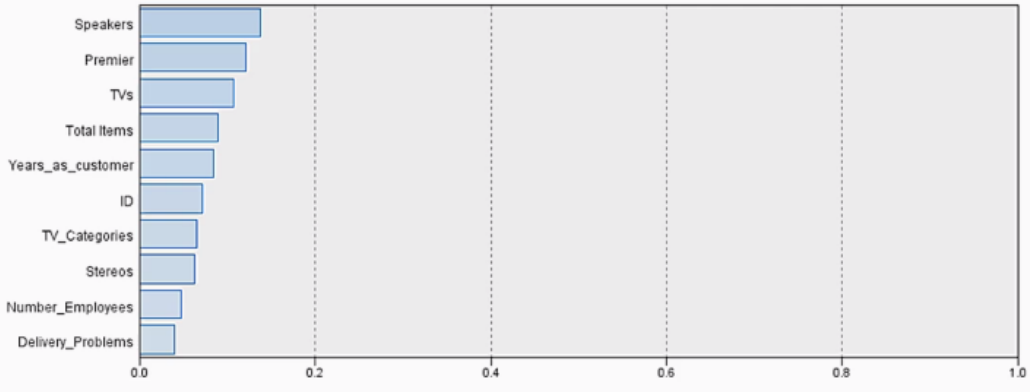
Model Viewer Summary Settings Annotations

1 2 3 4 5 All

- Premier in ["No"] [Mode: Churned] (1,280)
  - Years\_as\_customer <= 3 [Mode: Current] (193)
    - Estimated\_Revenue <= 4,114,620 [Mode: Churned] (72)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (25; 1.0)
      - Location in ["National"] [Mode: Churned] ⇒ Churned (47; 0.66)
    - Estimated\_Revenue > 4,114,620 and Estimated\_Revenue <= 5,029,070 [Mode: Current] ⇒ Current (29; 0.586)
    - Estimated\_Revenue > 5,029,070 and Estimated\_Revenue <= 6,898,750 [Mode: Current] ⇒ Current (39; 0.923)
    - Estimated\_Revenue > 6,898,750 [Mode: Current] ⇒ Current (53; 0.736)
  - Years\_as\_customer > 3 and Years\_as\_customer <= 4 [Mode: Current] (156)
    - TV\_Categories in ["Low" "Medium"] [Mode: Current] (63)
      - Stereos <= 11 [Mode: Churned] ⇒ Churned (25; 0.8)
      - Stereos > 11 [Mode: Current] ⇒ Current (38; 0.763)
    - TV\_Categories in ["None"] [Mode: Current] (93)
      - Stereo\_Discount in ["Discount"] [Mode: Current] ⇒ Current (28; 1.0)
      - Stereo\_Discount in ["Regular"] [Mode: Current] ⇒ Current (65; 0.846)
  - Years\_as\_customer > 4 [Mode: Churned] (931)
    - Speakers <= 5 [Mode: Churned] ⇒ Churned (109; 0.982)
    - Speakers > 5 and Speakers <= 11 [Mode: Churned] (87)
      - Stereos <= 7 [Mode: Churned] ⇒ Churned (29; 0.931)
      - Stereos > 7 and Stereos <= 24 [Mode: Churned] ⇒ Churned (31; 0.516)
      - Stereos > 24 [Mode: Churned] ⇒ Churned (27; 0.815)
    - Speakers > 11 and Speakers <= 18 [Mode: Churned] ⇒ Churned (90; 0.878)
    - Speakers > 18 and Speakers <= 27 [Mode: Churned] (76)
      - Location in ["International"] [Mode: Churned] ⇒ Churned (27; 0.778)
      - Location in ["National"] [Mode: Current] ⇒ Current (49; 0.531)
    - Speakers > 27 and Speakers <= 63 [Mode: Churned] (271)
      - Payment\_Method in ["Auto Pay"] [Mode: Churned] ⇒ Churned (43; 0.674)
      - Payment\_Method in ["Check" "Credit Card"] [Mode: Churned] (228)
        - TVs <= 3 [Mode: Churned] ⇒ Churned (184; 0.815)

### Predictor Importance

Target: Status





Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

```
graph LR; Electronics_Data --> Partition; Partition --> Status1[Status]; Partition --> Status2[Status]; Status1 --> Analysis;
```

Streams Outputs Models

Status

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export Python Spark IBM® SPSS® Text Analytics

Database Var. File Auto Data Prep Select Sample Aggregate Derive Type Filter Graphboard Auto Classifier Auto Numeric Auto Cluster Table Flat File Database

146MB / 182MB


Status

Objective: Standard model

Fields Build Options Model Options Annotations

Select an item:

- Objectives
- Basics
- Stopping Rules
- Ensembles
- Advanced

 Neural Network modeling internally separates records into a model building set and an overfit prevention set. Specify a percentage of records for the overfit prevention set.

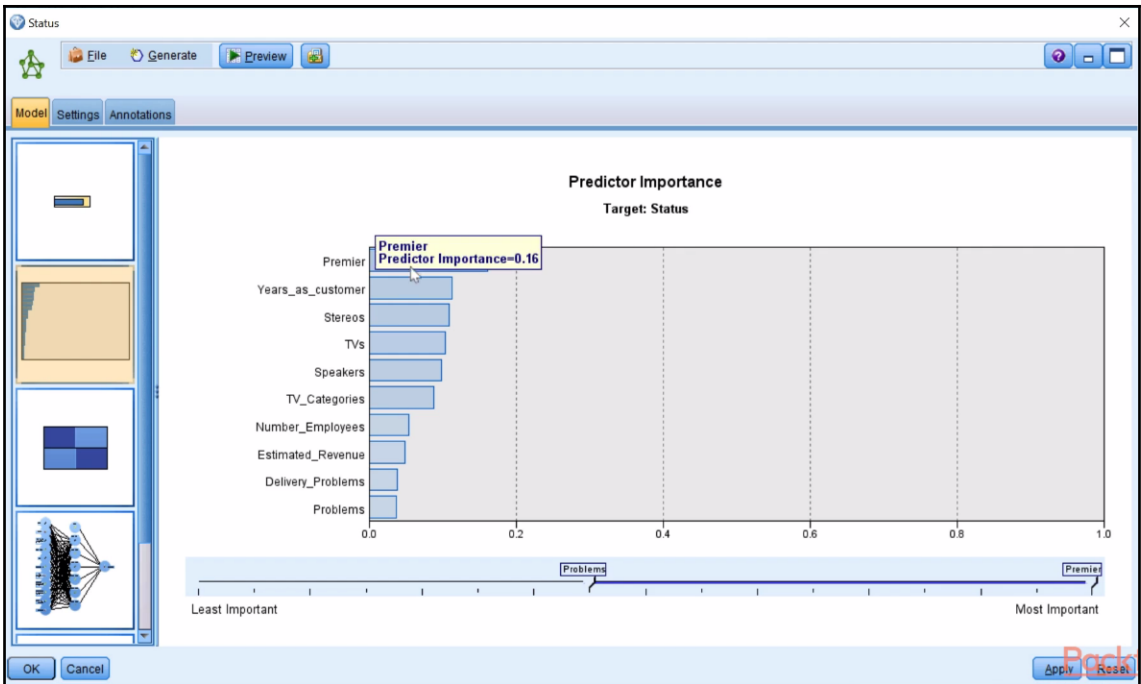
Overfit prevention set (%):

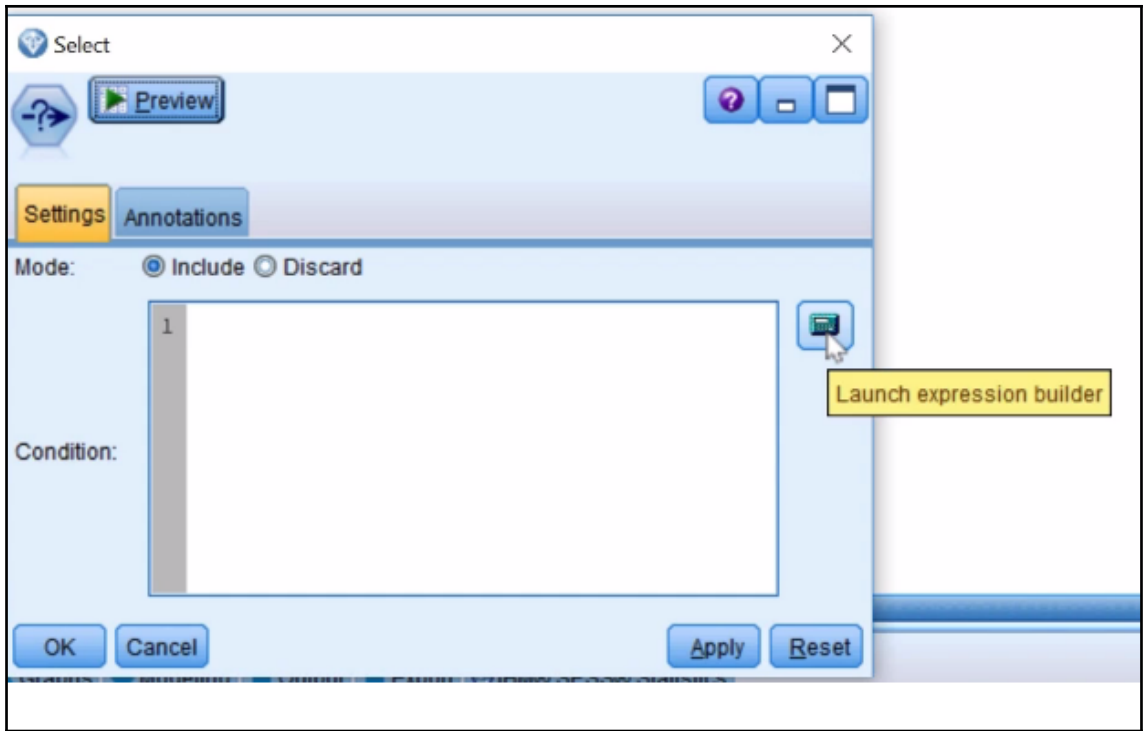
Replicate Results

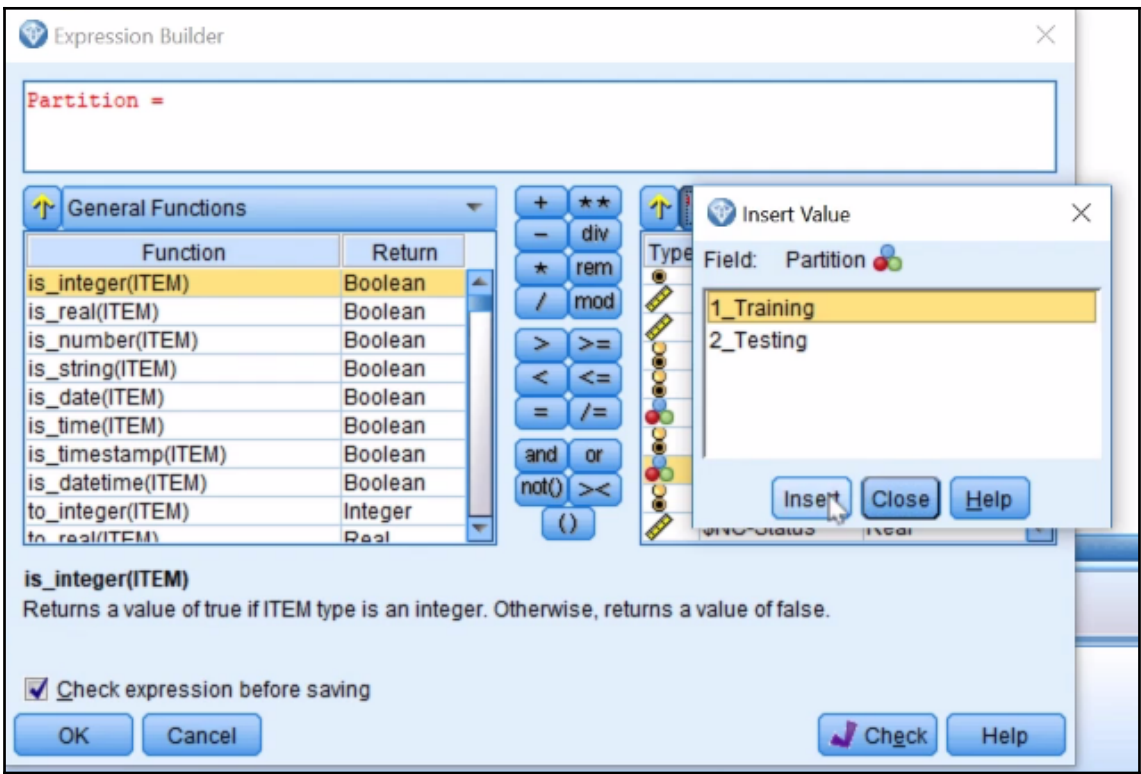
Random seed:

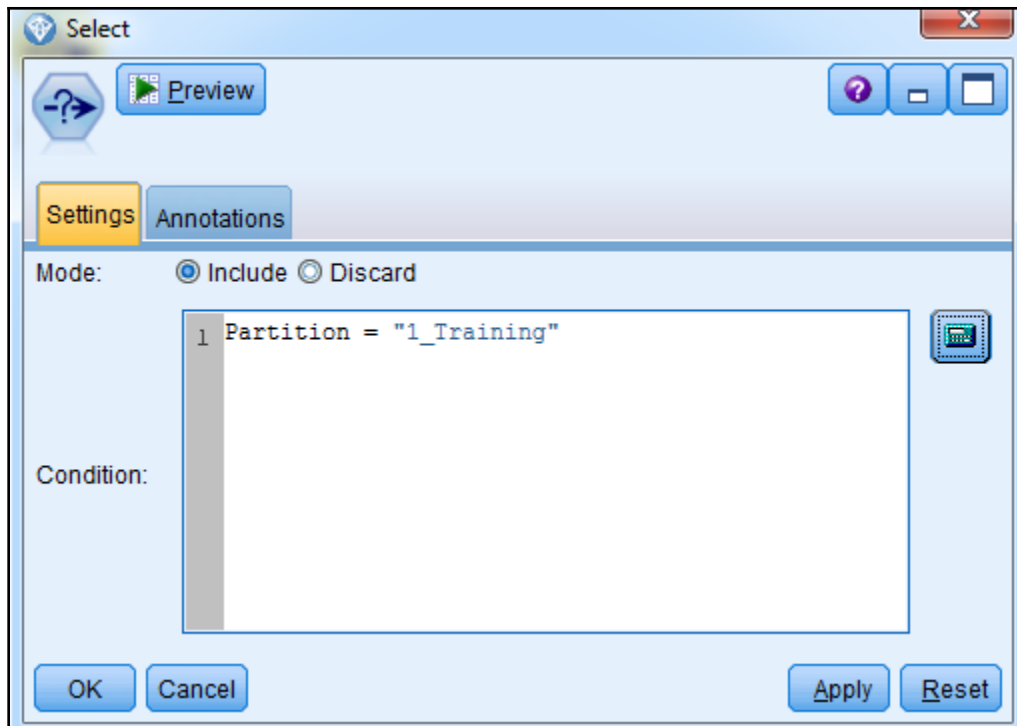
Missing values in predictors:

- Delete listwise
- Impute missing values









Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Streams

- Status

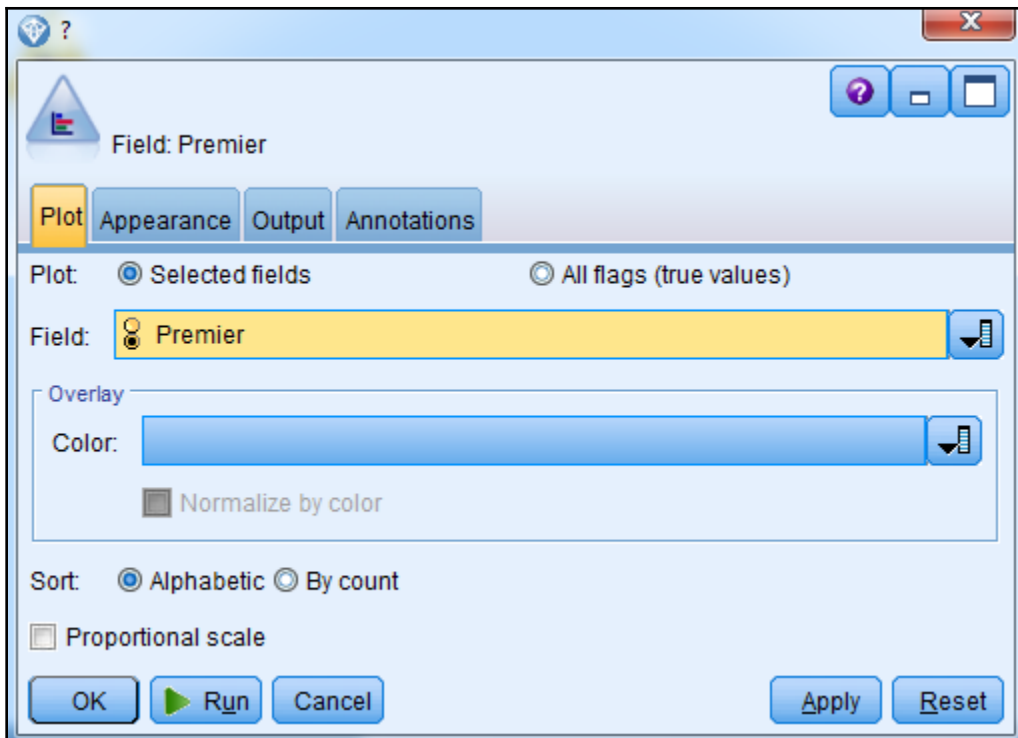
CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

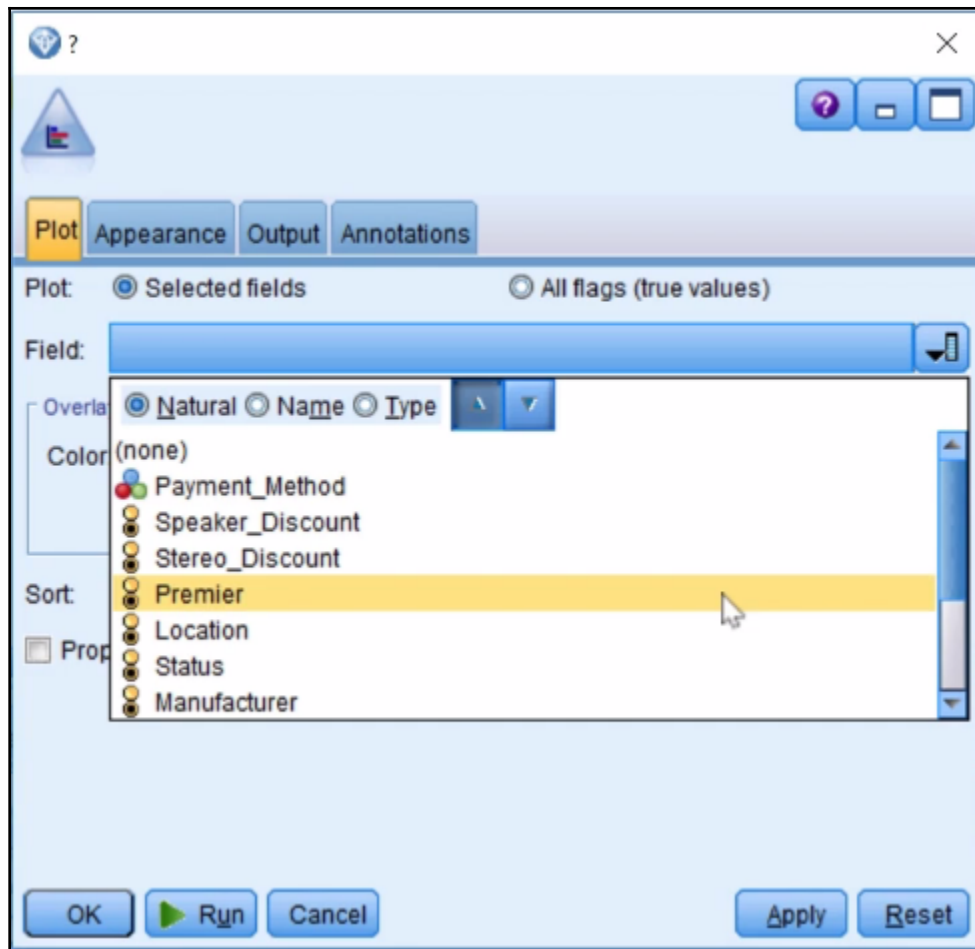
Analysis Partition Status Status Select ?

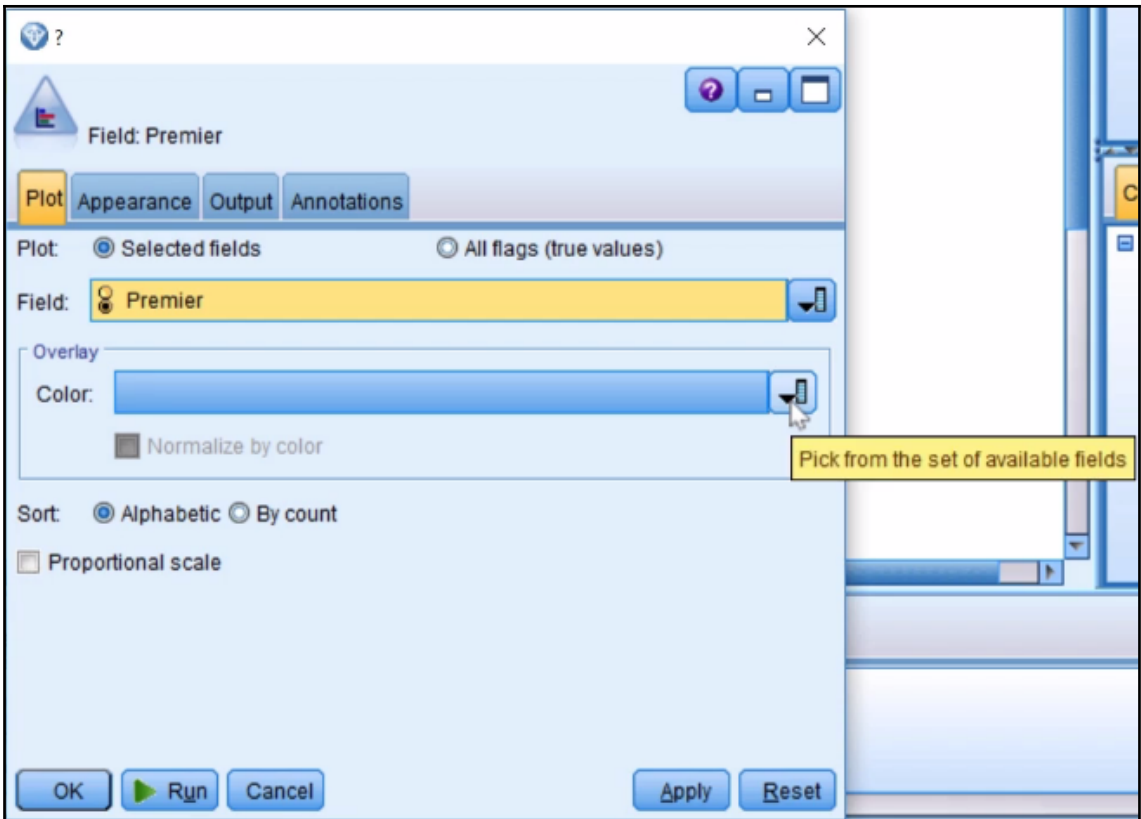
Graphboard Plot Multiplot Time Plot Distribution Histogram Collection Web Evaluation Map Visualization E-Plot(Beta) t-SNE

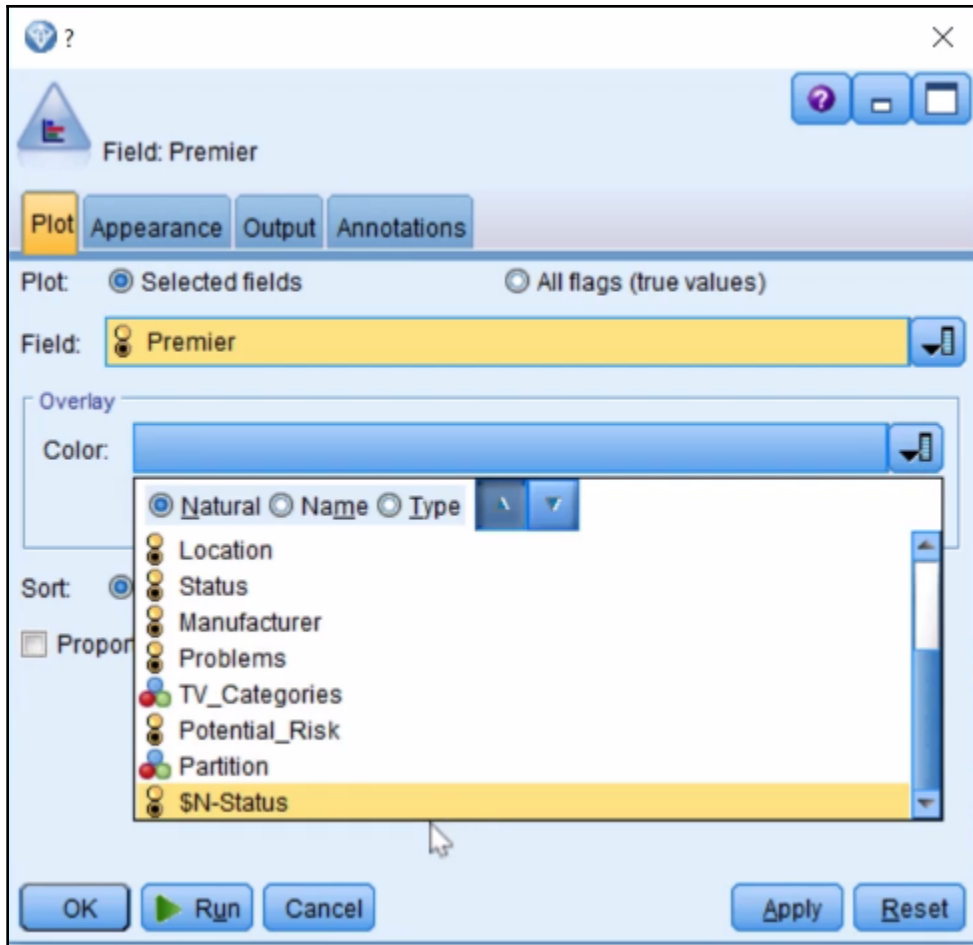
139MB / 198MB

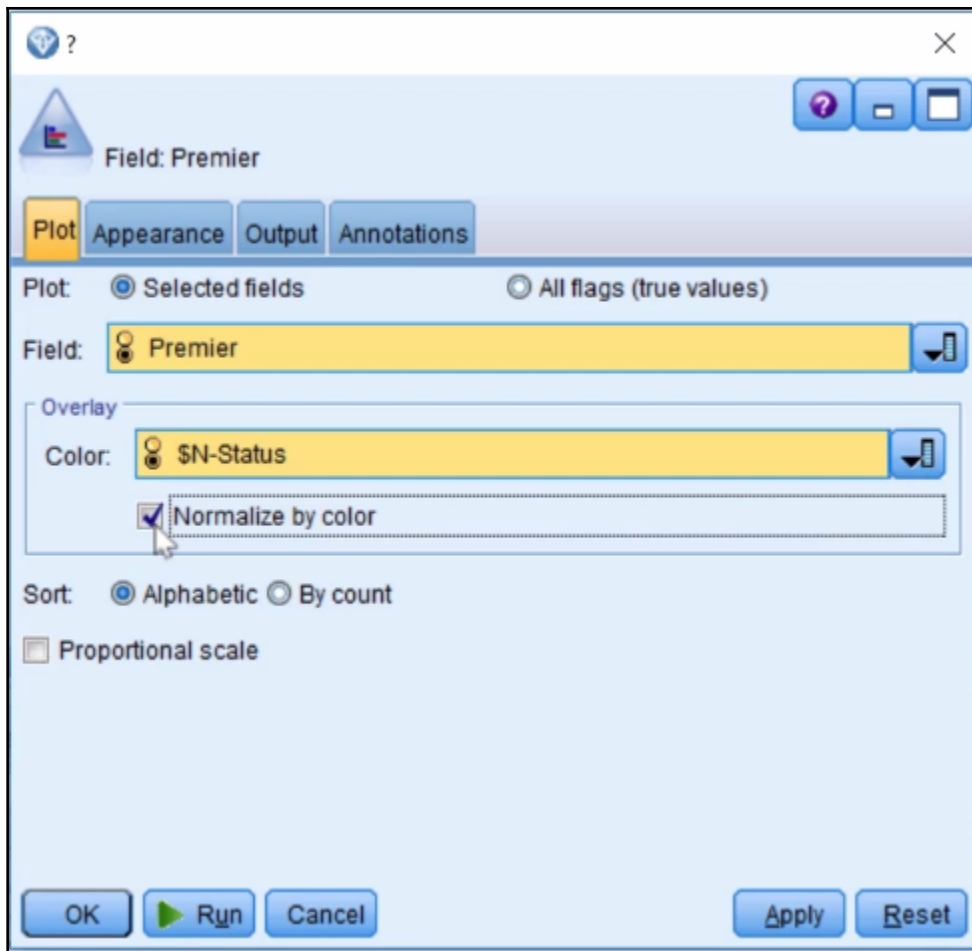


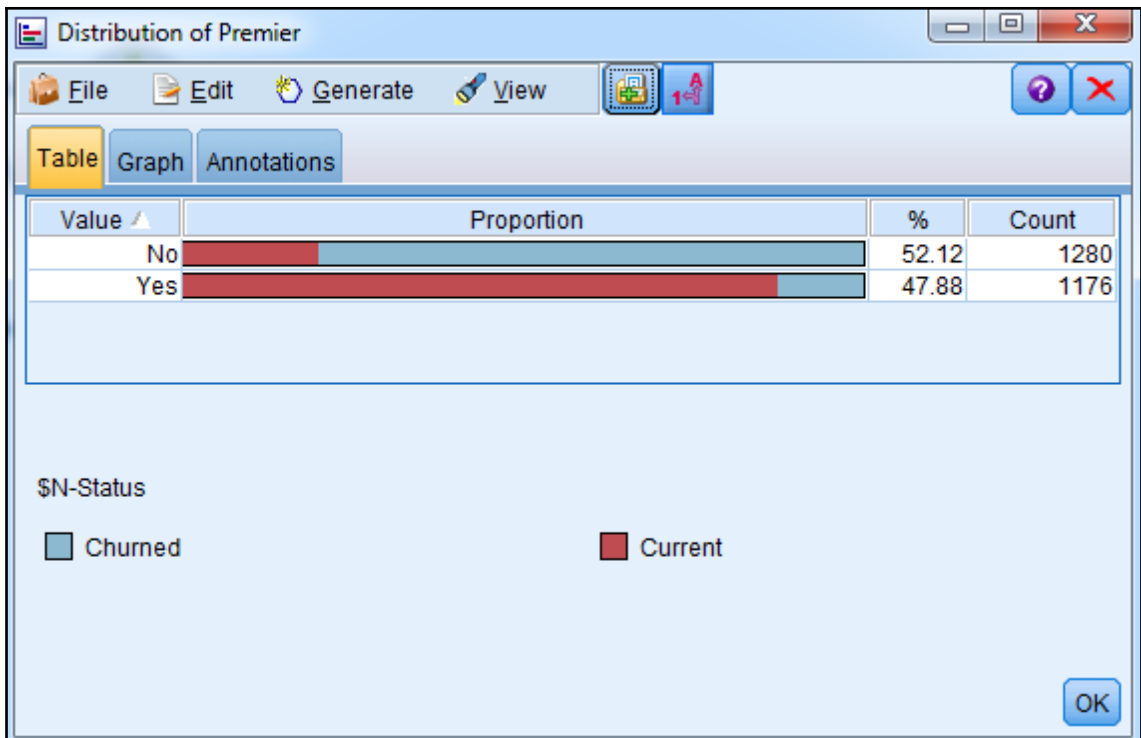












Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

The diagram in the main workspace consists of the following nodes and connections:

- Electronics\_Data** (Data source icon) connects to **Partition** (Partitioning icon).
- Partition** connects to **Status** (Status icon).
- Analysis** (Analysis icon) connects to the **Status** node.
- The **Status** node connects to **Select** (Selection icon).
- Select** connects to a **?** (Unknown icon).

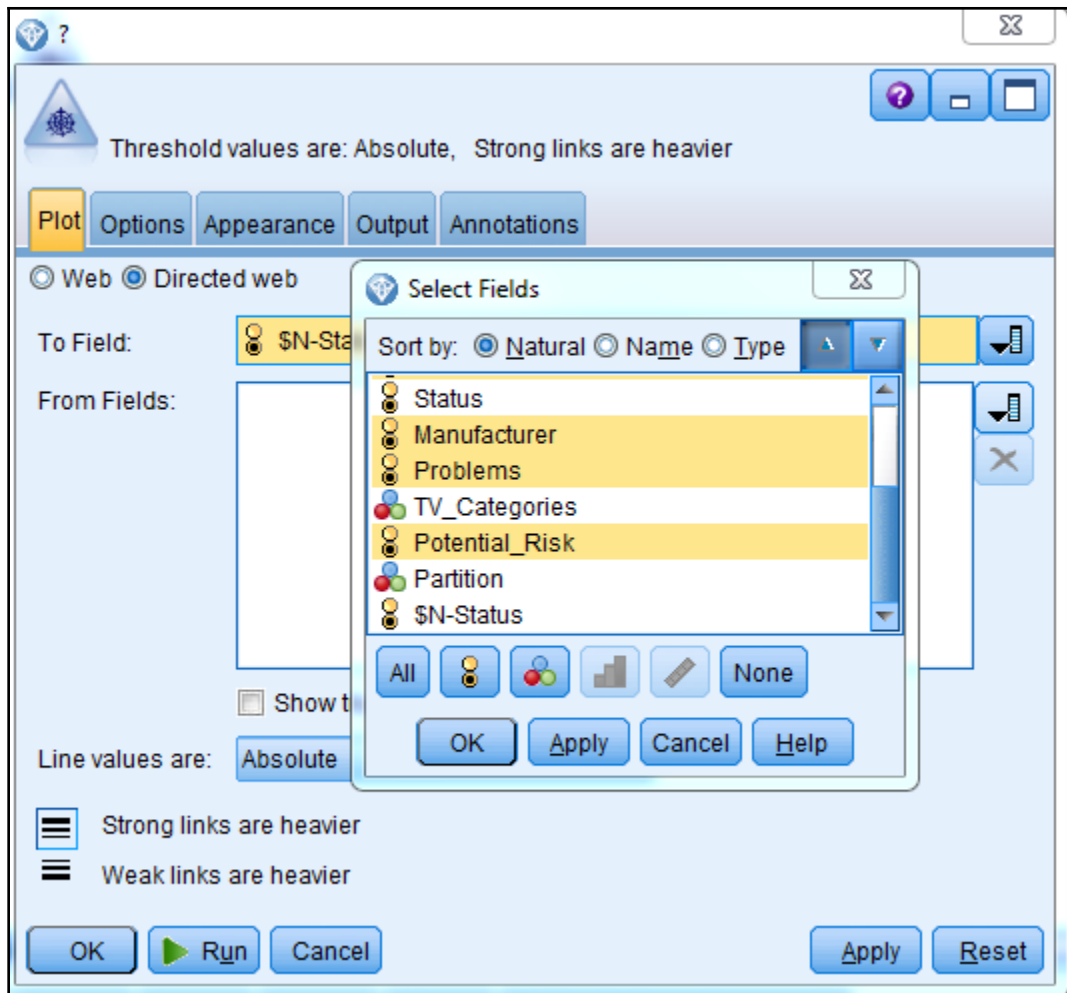
Right sidebar (CRISP-DM Classes):

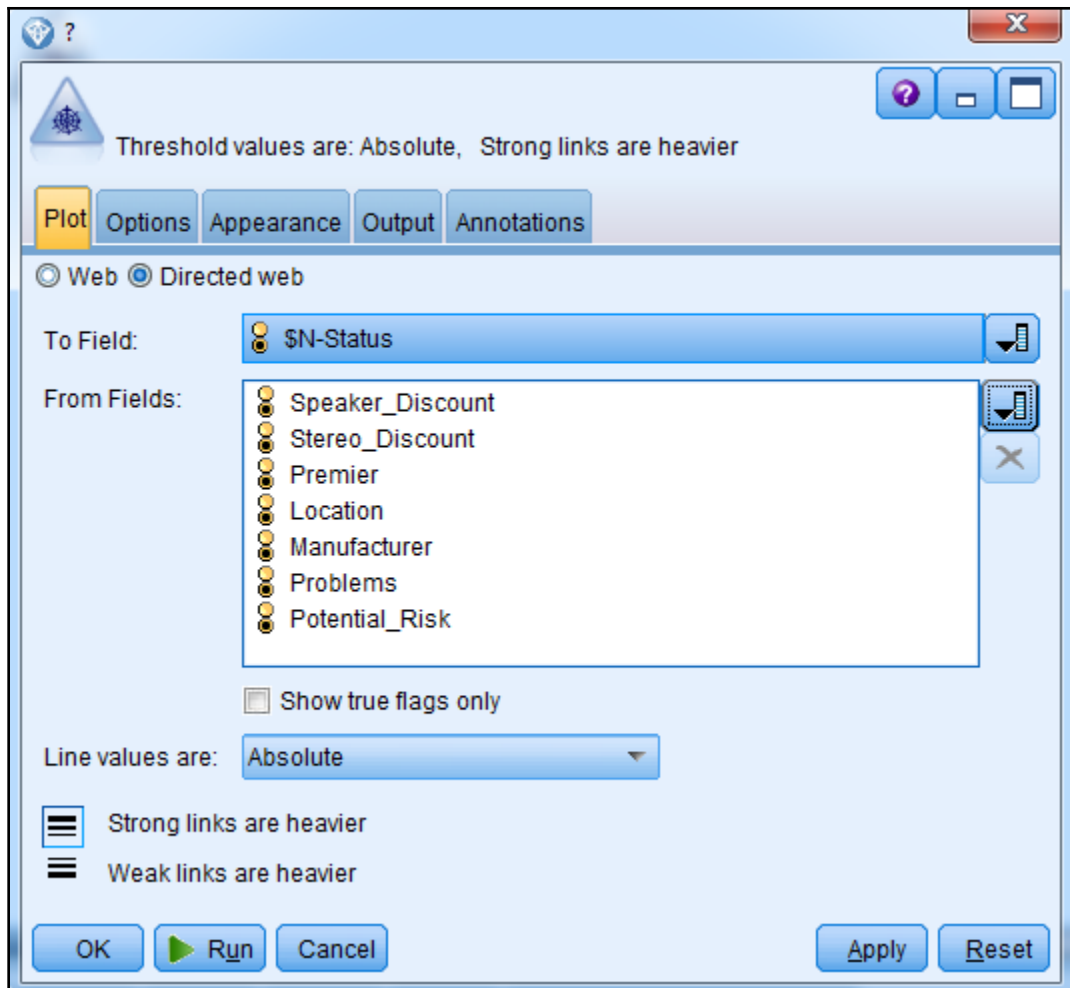
- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Bottom toolbar:

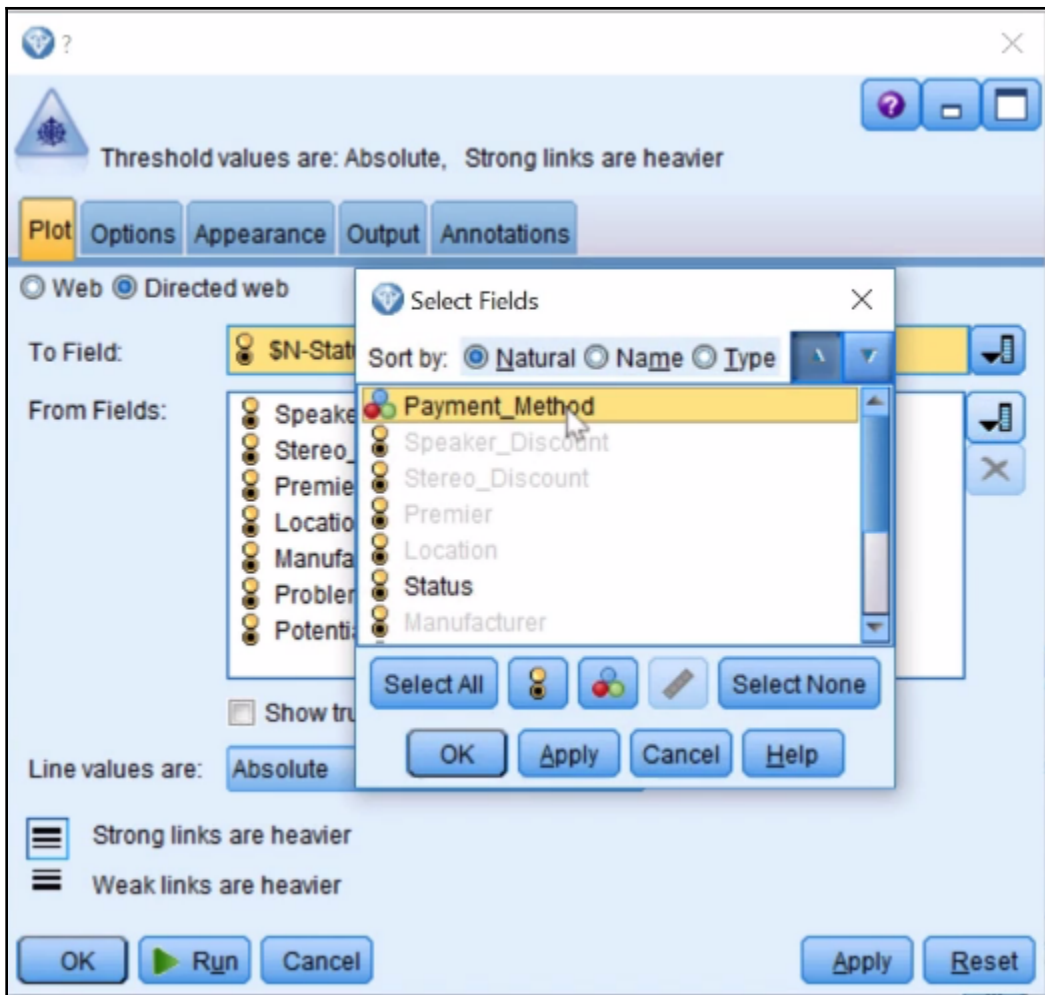
- Graphboard
- Plot
- Multiplot
- Time Plot
- Distribution
- Histogram
- Collection
- Web
- Evaluation
- Map Visualization
- E-Plot(Beta)
- t-SNE

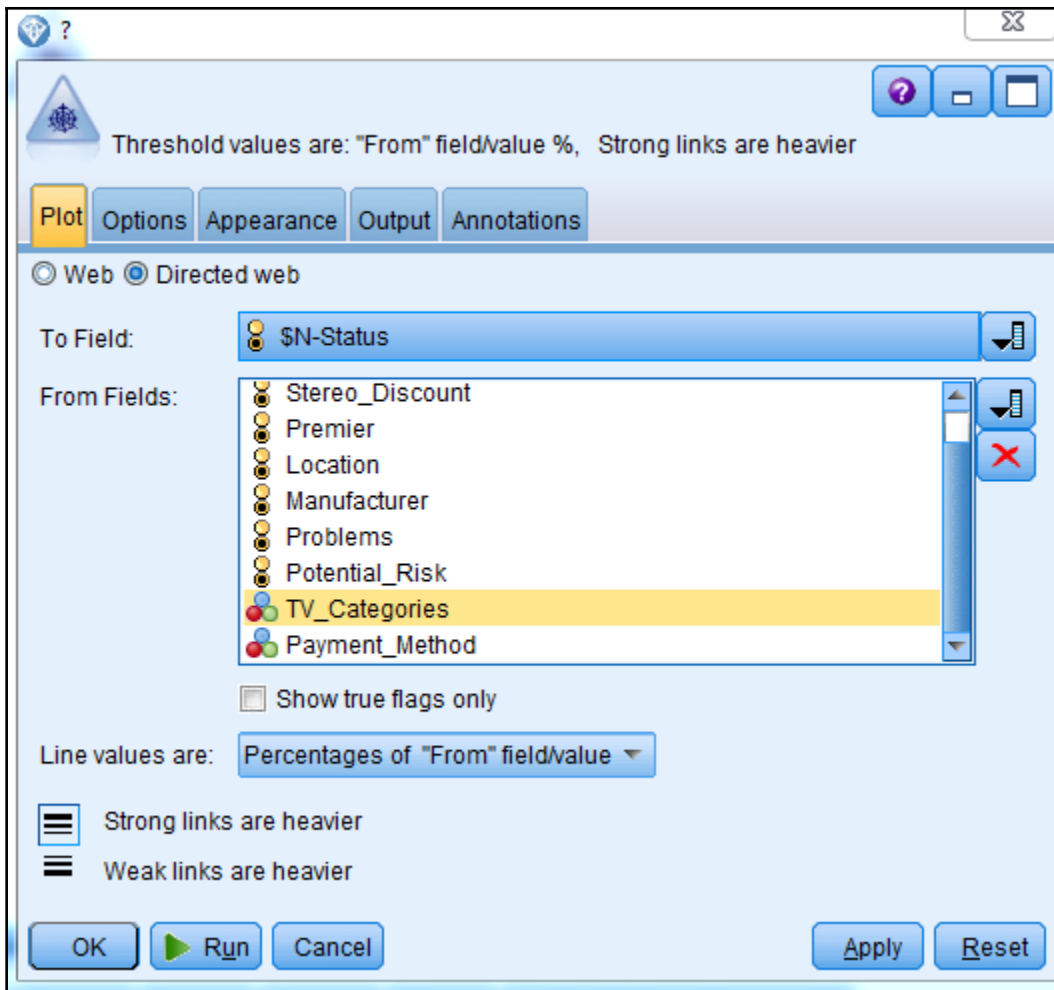
148MB / 198MB











Threshold values are: Absolute, Strong links are heavier

Plot Options Appearance Output Annotations

Web  Directed web

To Field:

From Fields:

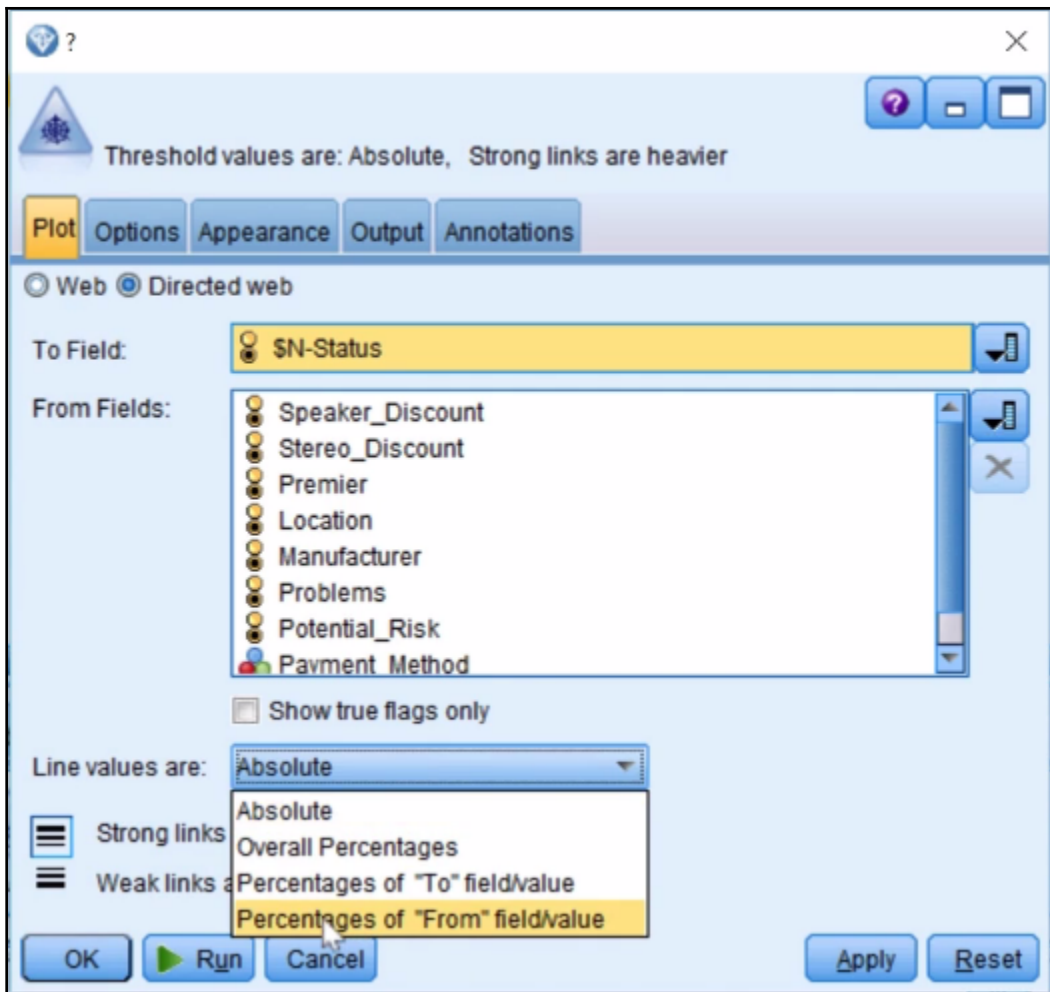
- Speaker\_Discount
- Stereo\_Discount
- Premier
- Location
- Manufacturer
- Problems
- Potential\_Risk
- Payment Method

Show true flags only

Line values are:

Strong links are heavier  
 Weak links are heavier

OK Run Cancel Apply Reset



Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Status

Electronics\_Data Partition Status

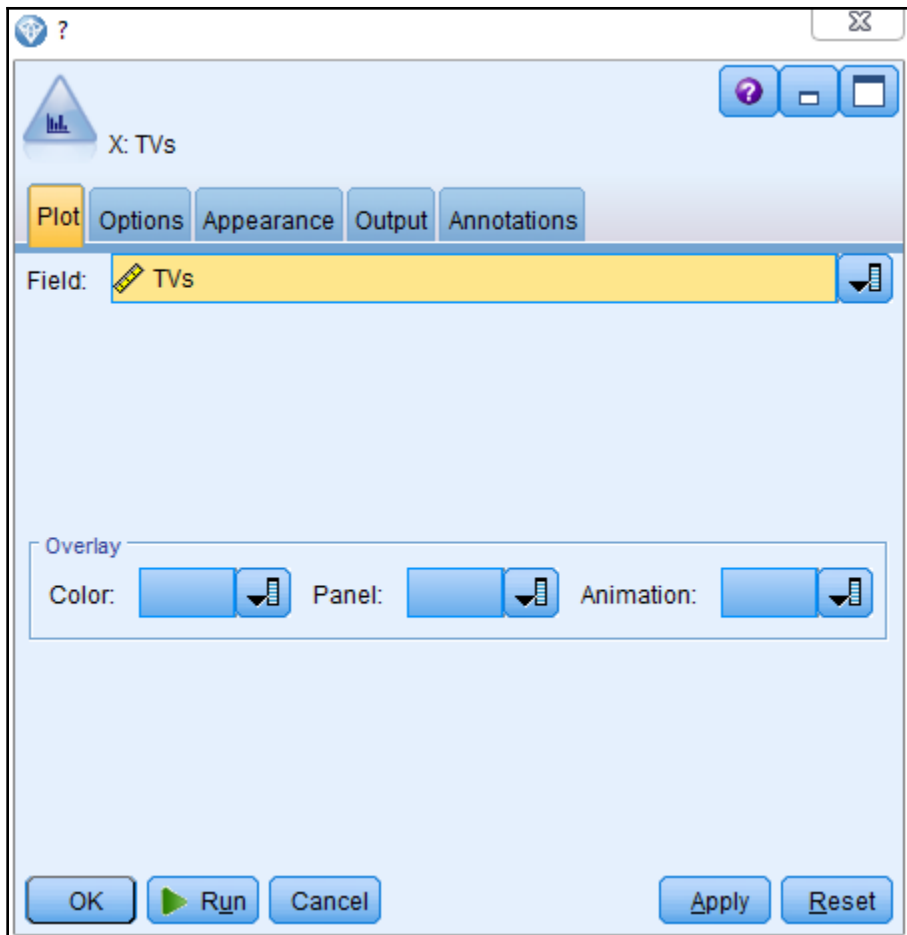
Analysis Status Select ?

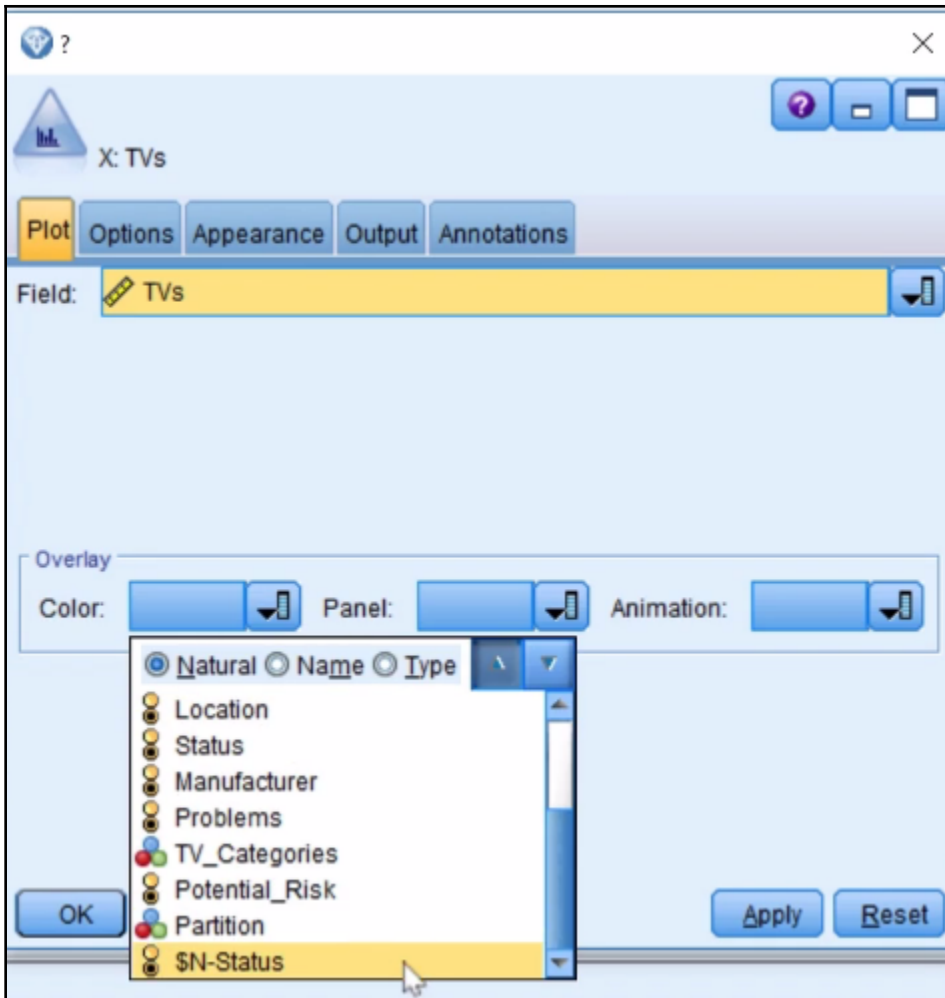
CRISP-DM Classes

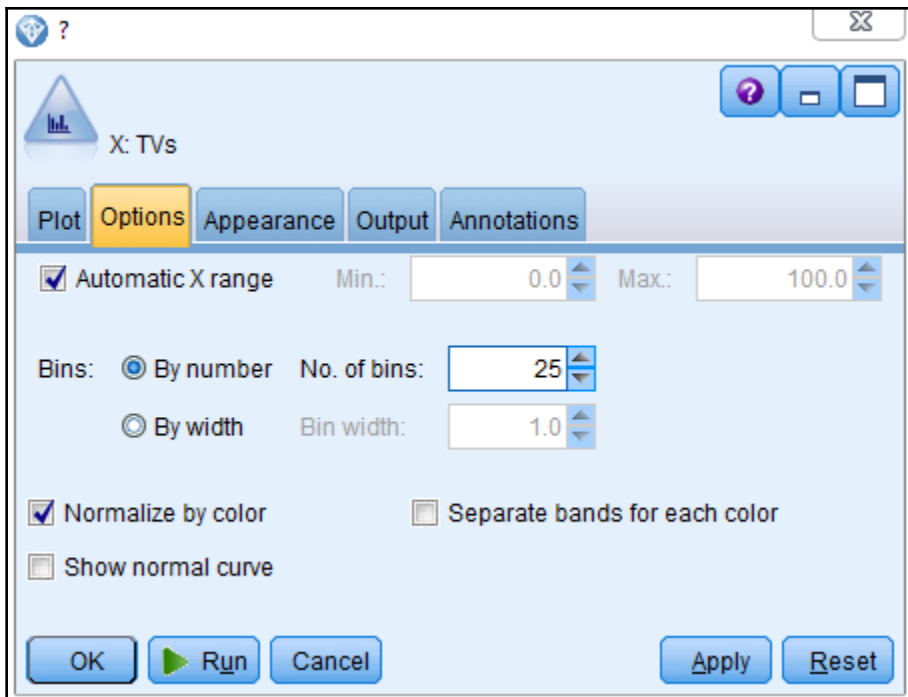
- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Graphboard Plot Multiplot Time Plot Distribution Histogram Collection Web Evaluation Map Visualization E-Plot(Beta) t-SNE

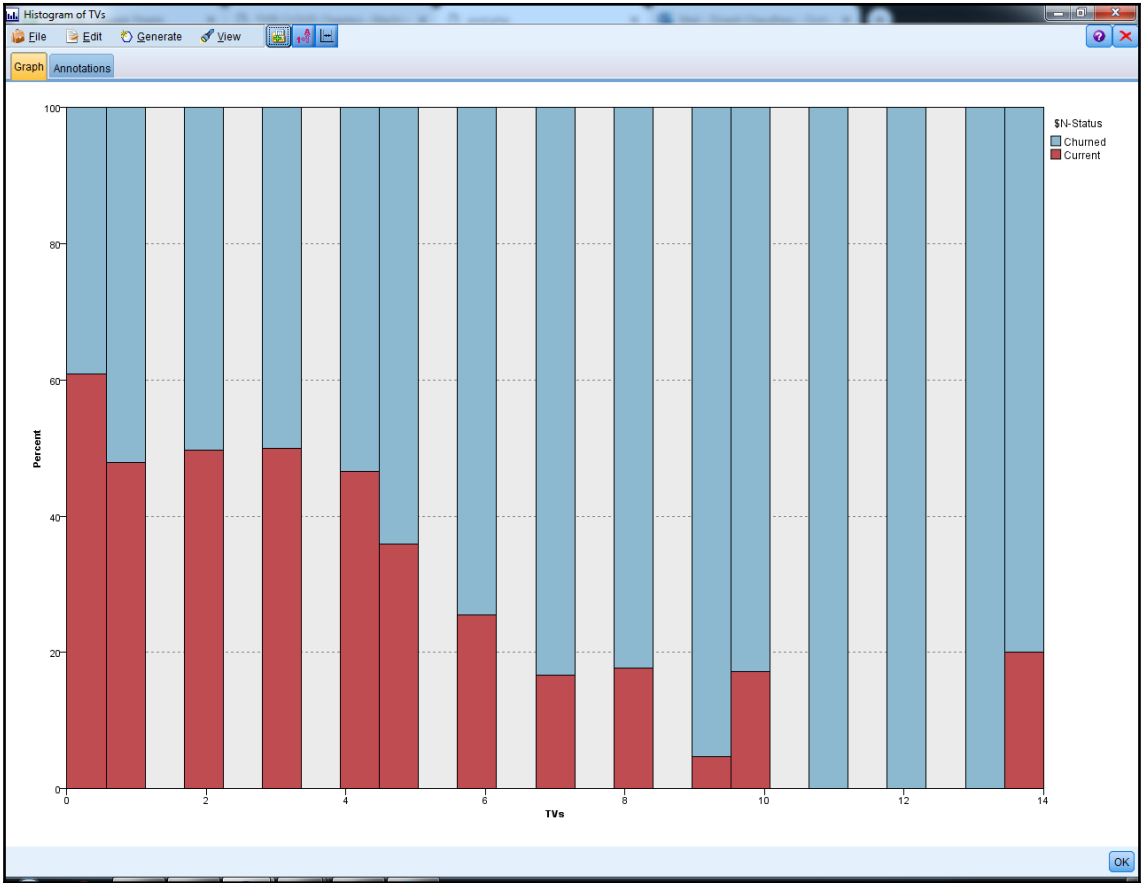
177MB / 224MB











Var. File X

Preview Refresh ? ▢ ▢

\\Mac\Home\Desktop\Machine Learning\Bank\_Data

**File** Data Filter Types Annotations

File: \\Mac\Home\Desktop\Machine Learning\Bank\_Data ...

```
id,salbeg,sex,time,age,salnow,edlevel,work,jobcat,minority,sexrace
628.000000,8400.000000,"Males",81.000000,28.500000,16080.000000,16.000000,0.2
630.000000,24000.000000,"Males",73.000000,40.330000,41400.000000,16.000000,12
632.000000,10200.000000,"Males",83.000000,31.080000,21960.000000,15.000000,4.
```

Read field names from file  Specify number of fields

Skip header characters:  EOL comment characters:

Strip lead and trail spaces:  None  Left  Right  Both

Invalid characters:  Discard  Replace with

Encoding:  Decimal symbol:

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

**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:

Double quotes:

OK Cancel Apply Reset

Bank\_Data

Preview Refresh

C:\Users\dineshchaudhary\Desktop\codes\Code\Code\ML\_Data\Bank\_Data

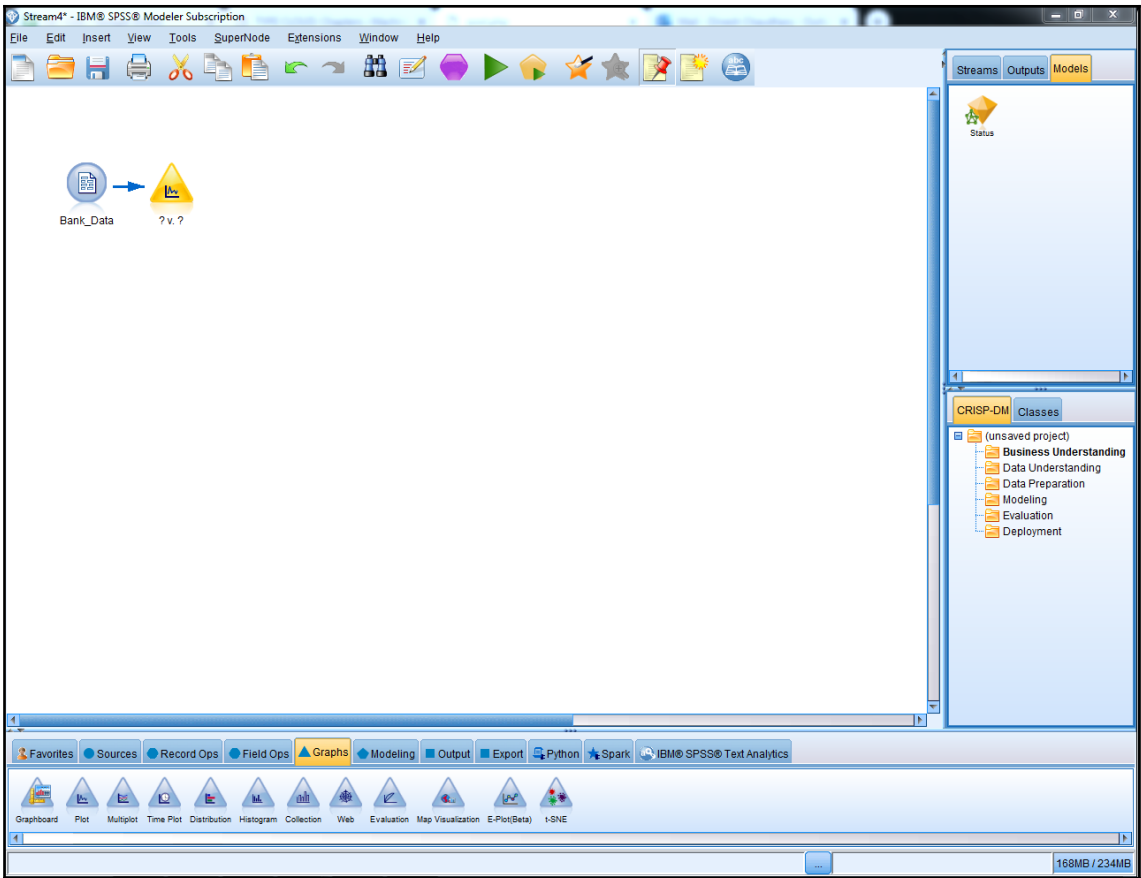
File Data Filter **Types** Annotations

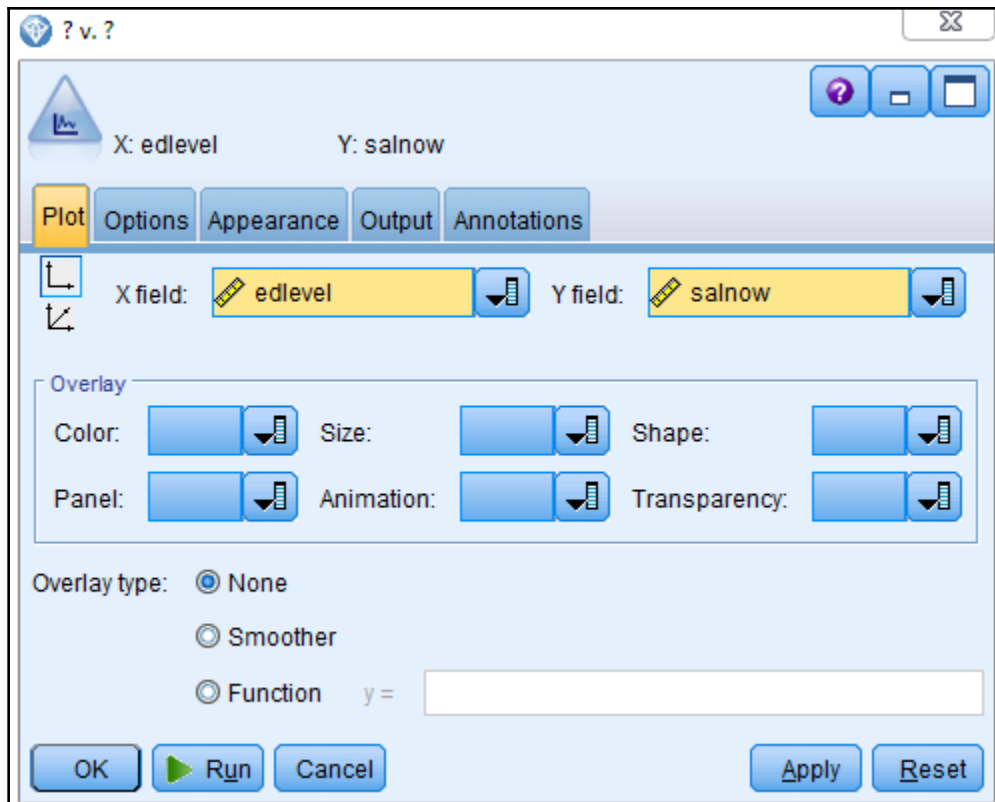
Read Values Clear Values Clear All Values

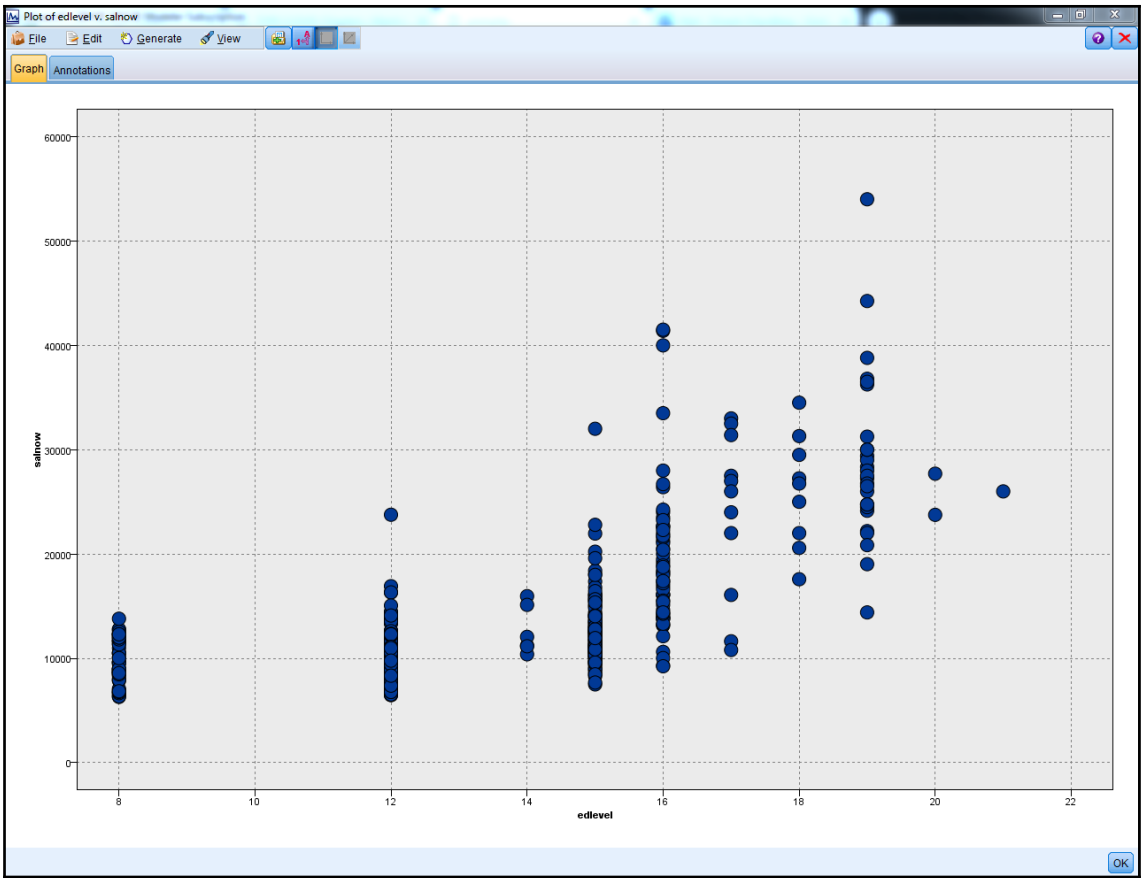
Field	Measurement	Values	Missing	Check	Role
id	Continuous	[626.0,112...		None	None
salbeg	Continuous	[3600.0,31...		None	Input
sex	Flag	Males/Fe...		None	Input
time	Continuous	[63.0,98.0]		None	Input
age	Continuous	[23.0,64.5]		None	Input
salnow	Continuous	[6300.0,54...		None	Target
edlevel	Continuous	[8.0,21.0]		None	Input
work	Continuous	[0.0,39.67]		None	Input
jobcat	Nominal	Clerical,"C...		None	Input
minority	Flag	White/Non...		None	Input
sexrace	Nominal	"Minority fe...		None	Input

View current fields
  View unused field settings

OK Cancel Apply Reset







Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

- Stream1
- Stream1
- Stream2
- Stream2
- Stream3
- Stream1
- Stream4

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment


Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output


185MB / 234MB

Matrix

Settings Appearance Output Annotations


Fields:  Selected  All flags (true values)  All Numerics

Rows:  

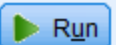

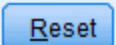
Columns:  

Include missing values

Cell contents:  Cross-tabulations  Function

Field:  

Function:  Mean  Sum  SDev  
 Max  Min

OK  Cancel  



Matrix

Settings Appearance Output Annotations

Rows and columns:

Unsorted  Ascending  Descending

Overlay:

Highlight top

Highlight bottom

Cross-tabulation cell contents:

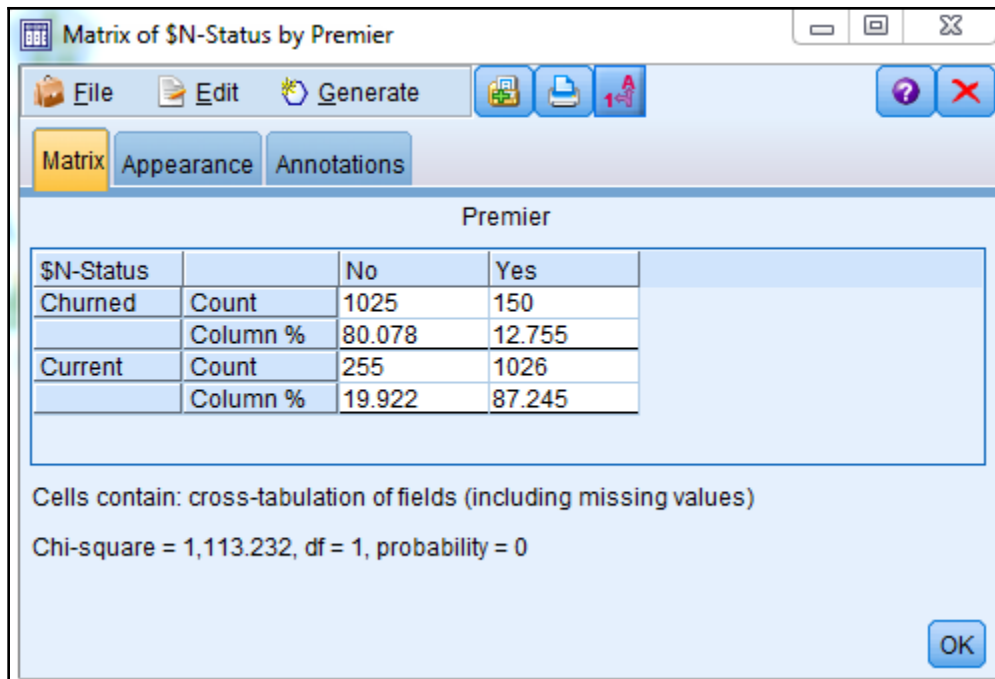
Counts  Percentage of row

Expected values  Percentage of column

Residuals  Percentage of total

Include row and column totals

OK Run Cancel Apply Reset



Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

- Stream1
- Stream1
- Stream2
- Stream2
- Stream3
- Stream1
- Stream4

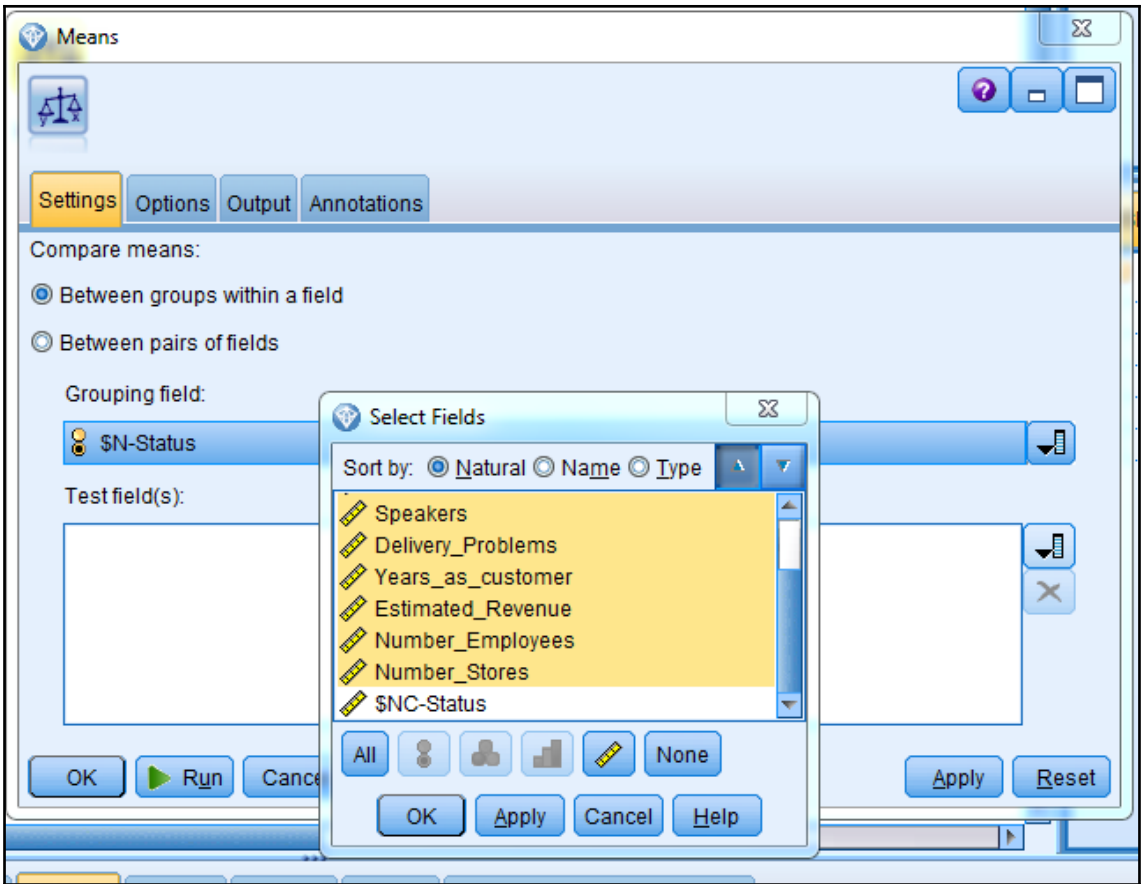
CRISP-DM Classes

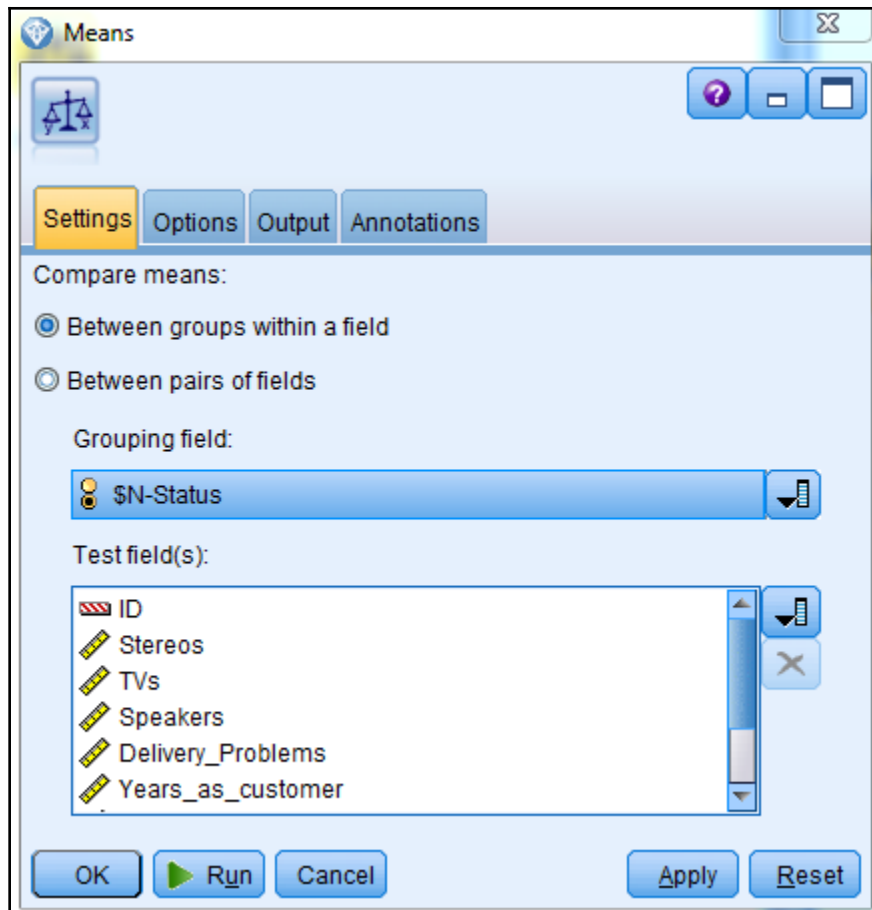
- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Analysis Status Select Means

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output

188MB / 234MB





Means of [\$N-Status][ID Stereos TVs Speakers Delivery\_P...]

File Edit

Means Annotations

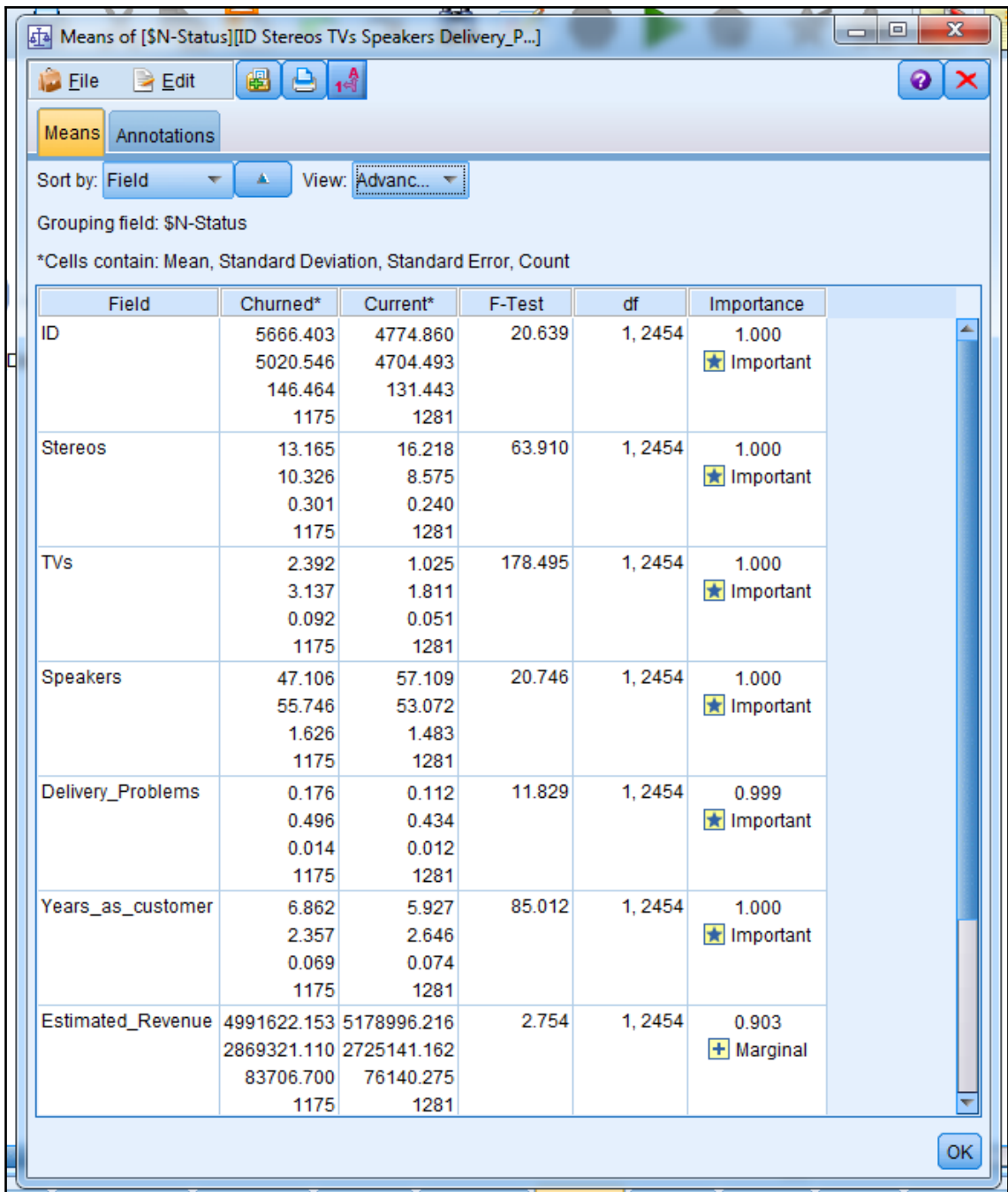
Sort by: Field View: Simple

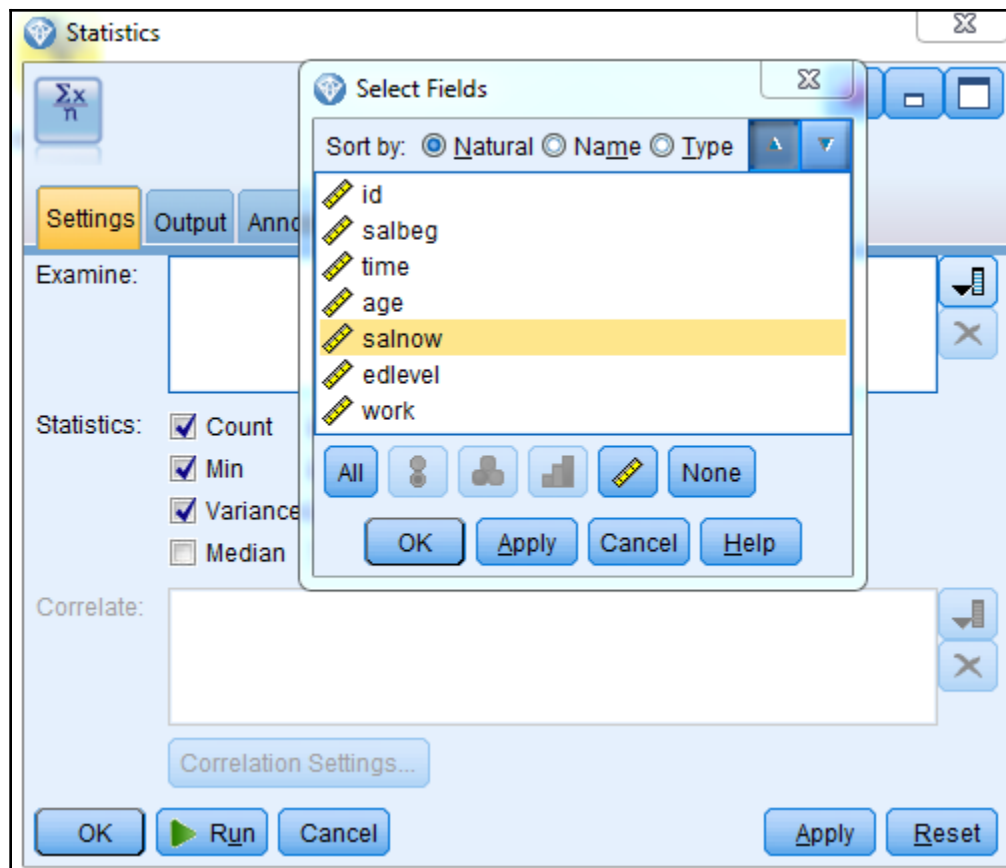
Grouping field: \$N-Status

\*Cells contain: Mean

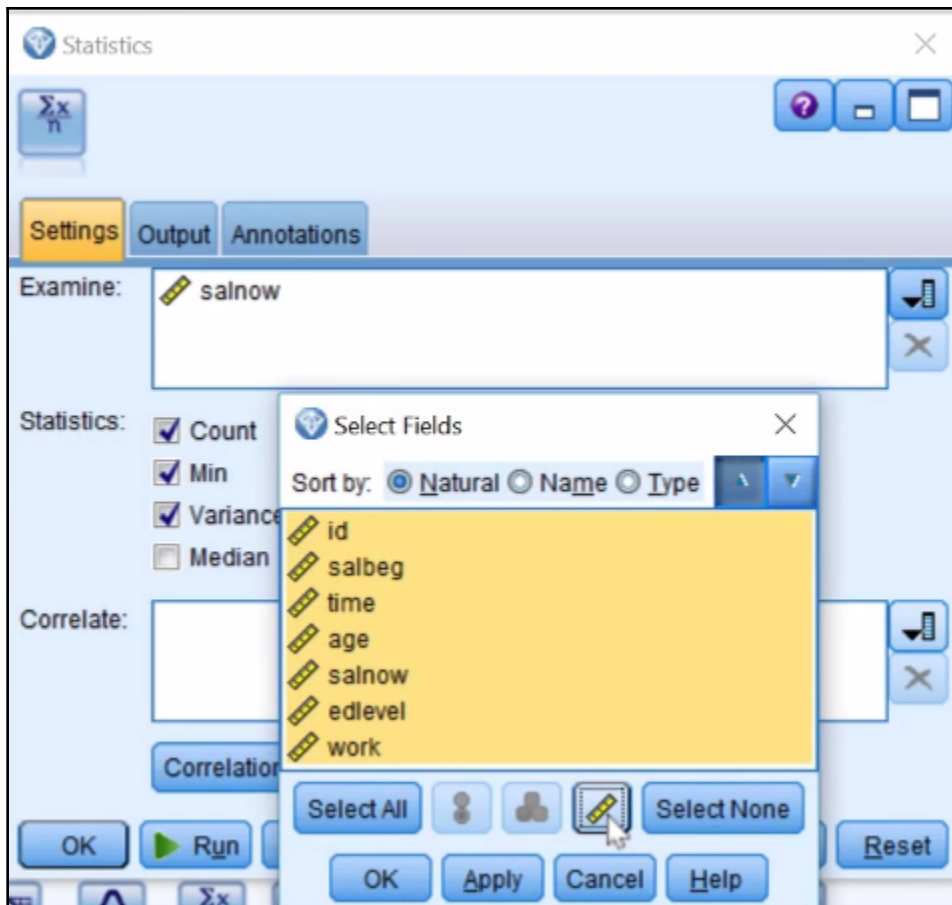
Field	Churned*	Current*	Importance
ID	5666.403	4774.860	1.000 ★ Important
Stereos	13.165	16.218	1.000 ★ Important
TVs	2.392	1.025	1.000 ★ Important
Speakers	47.106	57.109	1.000 ★ Important
Delivery_Problems	0.176	0.112	0.999 ★ Important
Years_as_customer	6.862	5.927	1.000 ★ Important
Estimated_Revenue	4991622.153	5178996.216	0.903 + Marginal
Number_Employees	5059.466	4994.847	0.427 □ Unimportant
Number_Stores	5.342	5.629	0.993 ★ Important

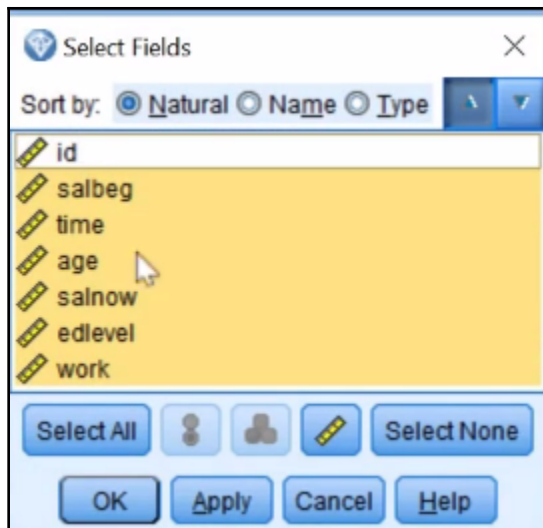
OK











Statistics of [salnow][6 fields]

File Edit Generate

Statistics Annotations

Collapse All Expand All

salnow

- Statistics

Count	474
Mean	13767.827
Min	6300.000
Max	54000.000
Range	47700.000
Variance	46652514.313
Standard Deviation	6830.265
Standard Error of Mean	313.724

- Pearson Correlations

salbeg	0.880	Strong
time	0.084	Medium
age	-0.146	Strong
edlevel	0.661	Strong
work	-0.097	Strong

OK

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

- Stream1
- Stream1
- Stream2
- Stream2
- Stream3
- Stream1
- Stream4

CRISP-DM Classes





- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Analysis Status Type




Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals




























210MB / 234MB

Type ✕

 Preview   

Types **Format** Annotations

   ▶ Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
 Premier	 Flag	Yes/No		None	 Input
 Location	 Flag	National/...		None	 Input
 Status	 Flag	Current/C...		None	 None
 Number_Em...	 Continuous	[102.0, 10...		None	 Input
 Number_Sto...	 Continuous	[1.0, 10.0]		None	 Input
 Manufacturer	 Flag	Yes/No		None	 Input
 Problems	 Flag	T/F		None	 Input
 TV_Categori...	 Nominal	Low, Medi...		None	 Input
 Potential Risk	 Flag	Risk/Req...		None	 Input

View current fields     View unused field settings

OK    Cancel    Apply    Reset

Type

Preview

Types Format Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
Number_Ent...	Continuous	[102.0, 10...		None	Input
Number_Sto...	Continuous	[1.0, 10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categori...	Nominal	Low, Medi...		None	Input
Potential_Risk	Flag	Risk/Reg...		None	Input
Partition	Nominal	*1_Trainin...		None	Partition
\$N-Status	Flag	Current/C...		None	Target
\$NC-Status	Continuous	[0.0, 1.0]		None	Input

View current fields
  View unused field settings

OK Cancel Apply Reset

Type ✕

Preview ? - □

Types Format Annotations

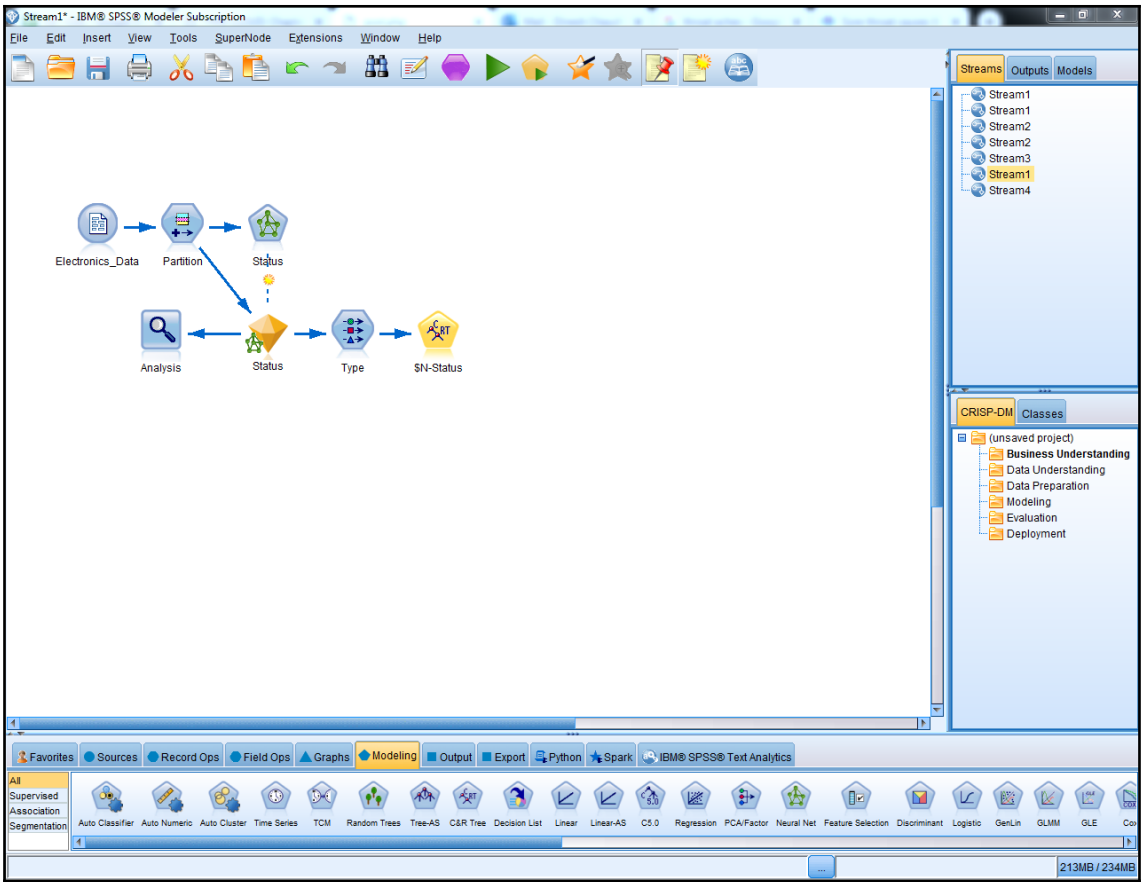
▶ Read Values
Clear Values
Clear All Values

Field	Measurement	Values	Missing	Check	Role
Number_Ent...	Continuous	[102.0, 10...		None	Input
Number_Sto...	Continuous	[1.0, 10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categori...	Nominal	Low, Medi...		None	Input
Potential_Risk	Flag	Risk/Reg...		None	Input
Partition	Nominal	"1_Trainin...		None	Partition
\$N-Status	Flag	Current/C...		None	Target
\$NC-Status	Continuous	[0.0, 1.0]		None	None

View current fields
  View unused field settings

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help



Streams Outputs Models

- Stream1
- Stream2
- Stream3
- Stream4

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Modeling

Auto Classifier Auto Numeric Auto Cluster Time Series TCM Random Trees Tree-AS C&R Tree Decision List Linear Linear-AS C5.0 Regression PCA-Factor Neural Net Feature Selection Discriminant Logistic GenLin GLM GLE Co...

213MB / 234MB



**\$N-Status** [Close]

Objective: Standard model

**Fields** | Build Options | Model Options | Annotations

Use predefined roles  
 Use custom field assignments

Fields:  
Sort: None

- Status
- \$NC-Status

Targets\*:  
\$N-Status

Predictors (Inputs)\*:

- Stereos
- TVs
- Speakers
- Delivery\_Problems
- Years\_as\_customer
- Estimated\_Revenue
- Payment\_Method
- Speaker\_Discount
- Stereo\_Discount
- Premier
- Location
- Number\_Employees

Analysis Weight:

OK Run Cancel Apply Reset

Stream1\* - IBM® SPSS® Modeler Subscription

File Edit Insert View Tools SuperNode Extensions Window Help

Streams Outputs Models

Streams

- Status
- SN-Status

Models

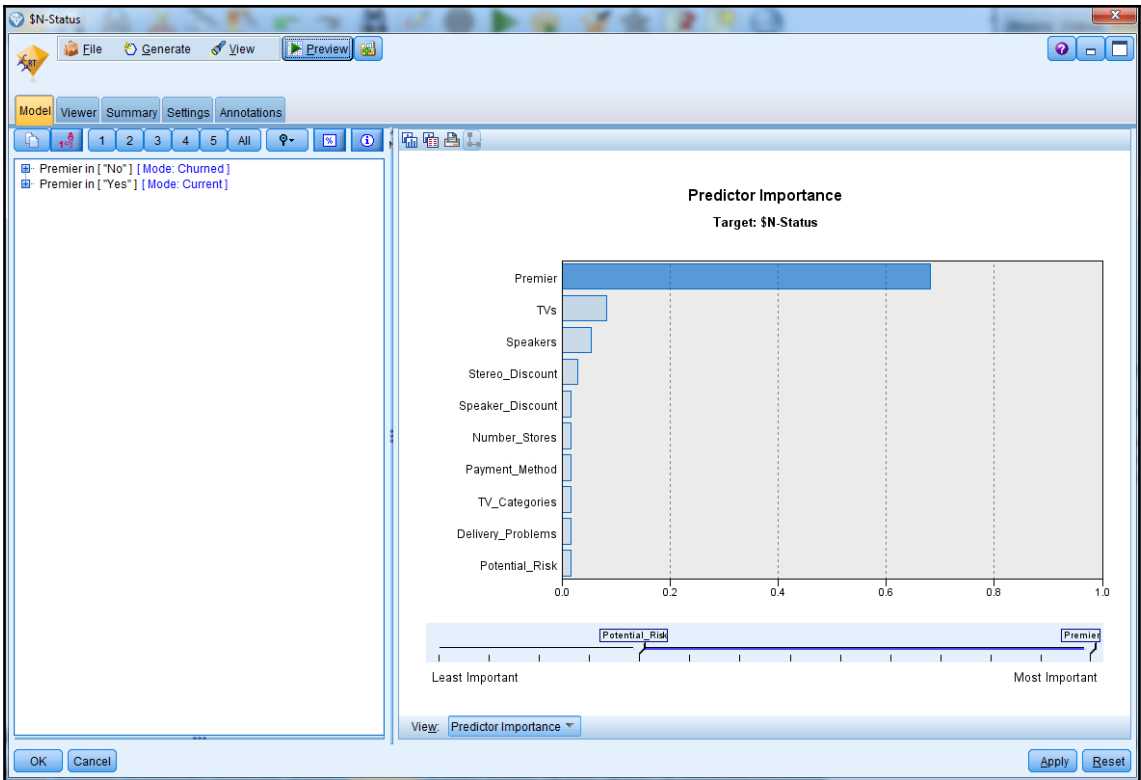
CRISP-DM | Classes

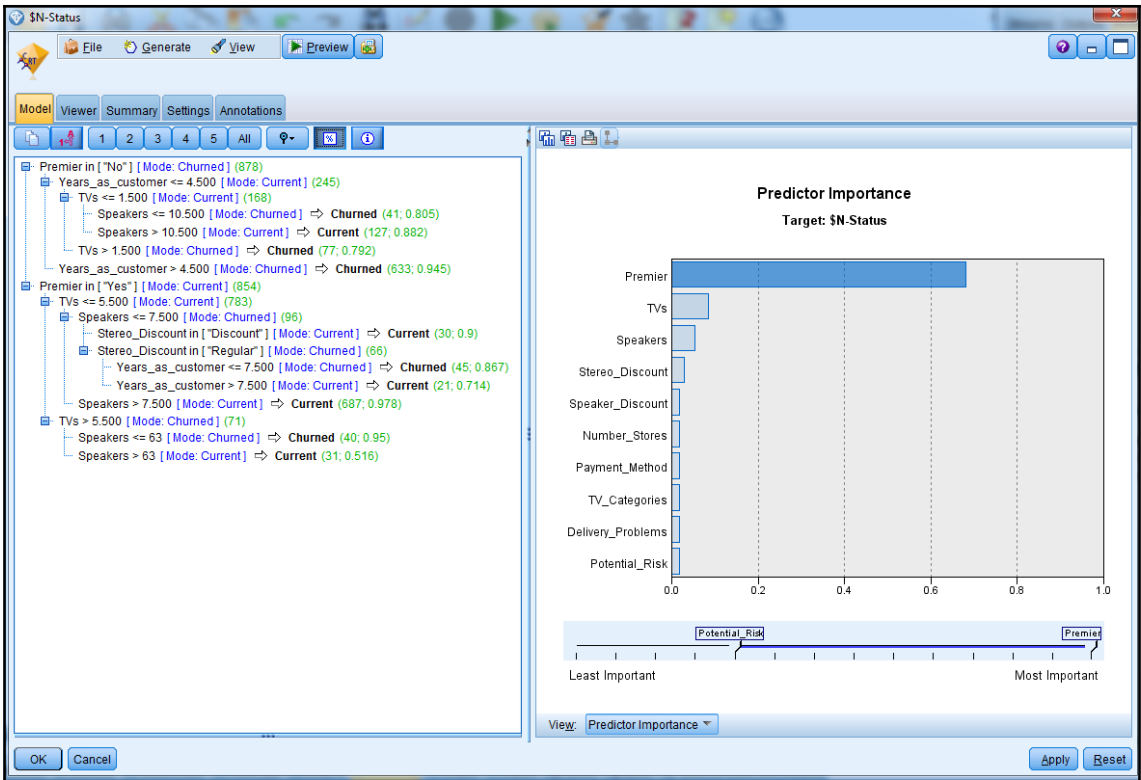
- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Modeling

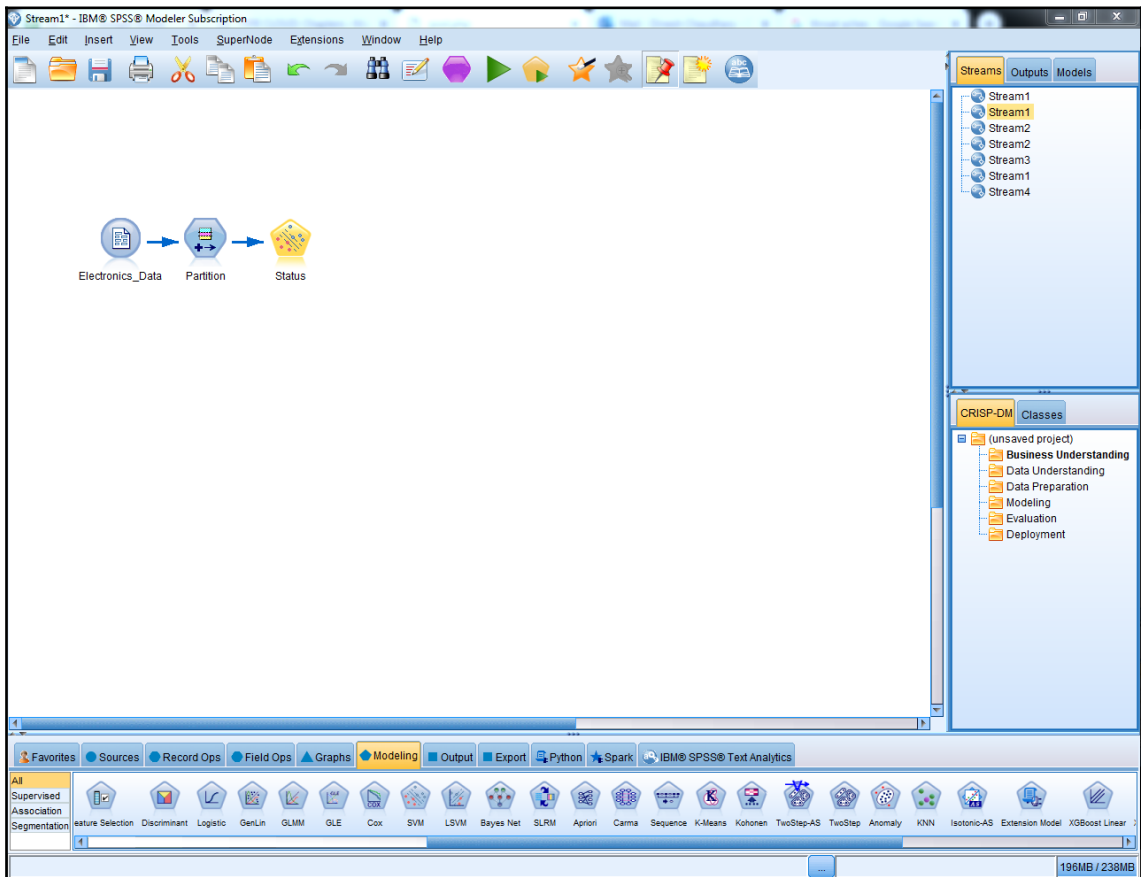
Auto Classifier Auto Numeric Auto Cluster Time Series TCM Random Trees Tree-AS C&R Tree Decision List Linear Linear-AS C5.0 Regression PCA-Factor Neural Net Feature Selection Discriminant Logistic GenLin GLM OLE Cov

230MB / 234MB





# Chapter 4: Improving Individual Models



Analysis of [Status]

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

- Comparing \$\$-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,379	96.86%	1,987	78.01%
Wrong	77	3.14%	560	21.99%
Total	2,456		2,547	

- Coincidence Matrix for \$\$-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	1,057	32
Current	45	1,322

'Partition' = 2_Testing	Churned	Current
Churned	819	257
Current	303	1,168

OK

Analysis of [Status] #1

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Comparing \$\$-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,078	84.61%	2,042	80.17%
Wrong	378	15.39%	505	19.83%
Total	2,456		2,547	

Coincidence Matrix for \$\$-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	905	184
Current	194	1,173

'Partition' = 2_Testing	Churned	Current
Churned	820	256
Current	249	1,222

OK

Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot shows the IBM SPSS Modeler Stream2 interface. The main workspace contains a workflow with three nodes: 'Electronics\_Data' (a document icon), 'Partition' (a cube icon), and 'Status' (a document icon with a green checkmark). Arrows indicate the flow from Electronics\_Data to Partition, and then to Status. The right-hand pane is divided into 'Streams', 'Outputs', and 'Models' tabs. The 'Models' tab is active, showing a tree view of 'CRISP-DM Classes' under '(unsaved project)'. The classes listed are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The bottom toolbar is divided into several sections: Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling (which is highlighted), Output, Export, and IBM® SPSS® Statistics. The Modeling section contains various icons for different modeling techniques, including Auto Classifier, Auto Numeric, Auto Cluster, Time Series, TCM, C&R Tree, Quest, CHAID, Decision List, Linear, CS.0, Regression, PCA/Factor, Neural Net, Feature Selection, Discriminant, Logistic, GenLin, QLM, Cox, SVM, and Bayes Net.

Streams Outputs Models

Electronics\_Data Partition Status

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

All Automated Classification Association

Auto Classifier Auto Numeric Auto Cluster Time Series TCM C&R Tree Quest CHAID Decision List Linear CS.0 Regression PCA/Factor Neural Net Feature Selection Discriminant Logistic GenLin QLM Cox SVM Bayes Net



Status X

Objective: Standard model

Fields Build Options Model Options Annotations

Select an item:

- Objectives
- Basics
- Stopping Rules
- Ensembles
- Advanced

**i** Neural Network modeling internally separates records into a model building set and an overfit prevention set. Specify a percentage of records for the overfit prevention set.

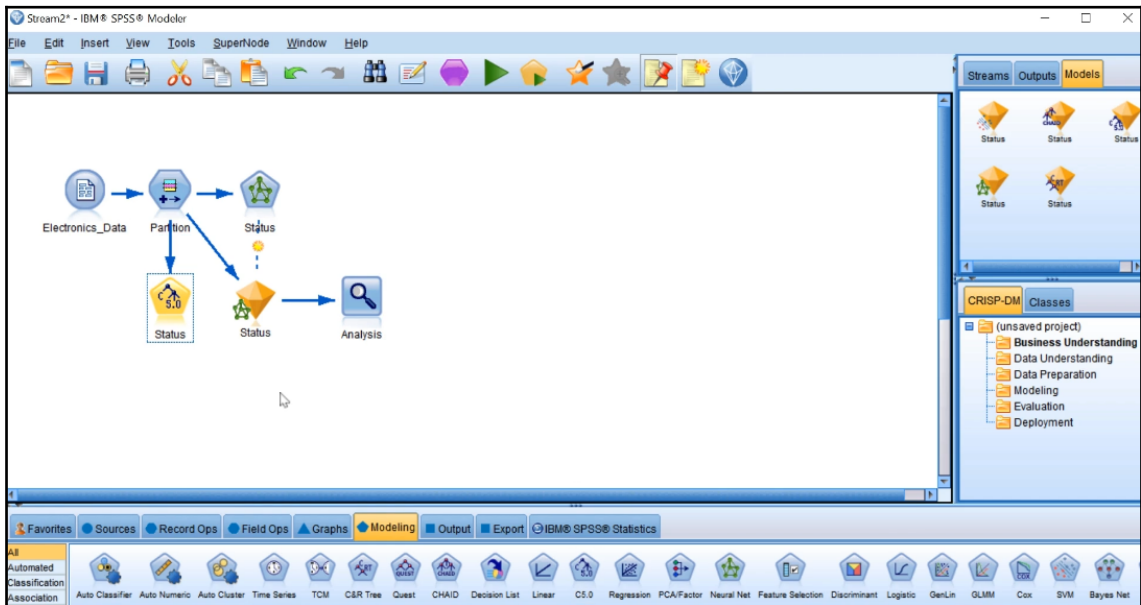
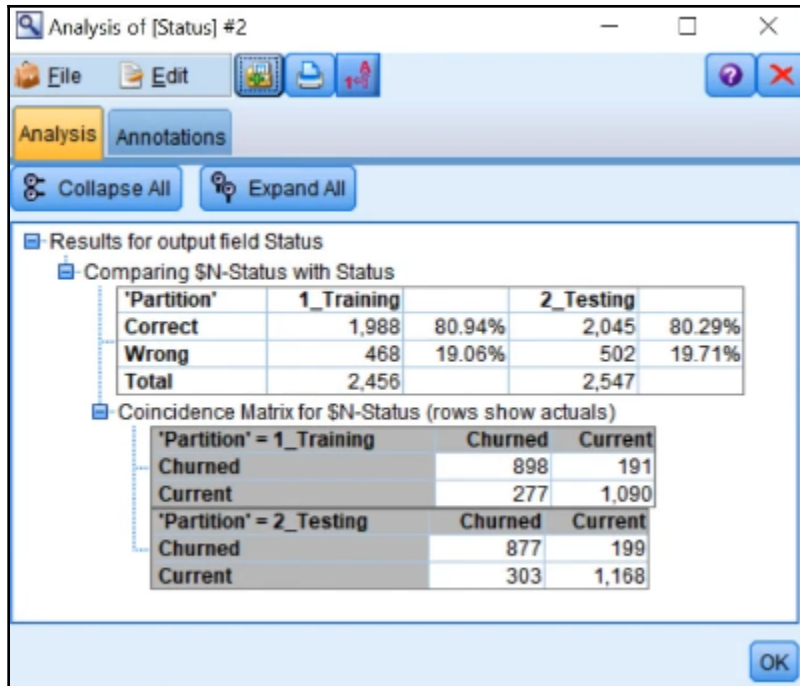
Overfit prevention set (%):

Replicate Results

Random seed:

Missing values in predictors:

- Delete listwise
- Impute missing values



Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot shows the IBM SPSS Modeler interface. The main workspace contains a workflow diagram with the following nodes and connections:

- Electronics\_Data** (Source node) connects to **Partition** (Partitioning node).
- Partition** connects to two **Status** (Monitoring node) nodes.
- The top **Status** node connects to **Analysis** (Analysis node).
- The bottom **Status** node is highlighted with a mouse cursor.

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

All Automated Classification Association

Auto Classifier Auto Numeric Auto Cluster Time Series TCM C&R Tree Quest CHAID Decision List Linear C5.0 Regression PCA/Factor Neural Net Feature Selection Discriminant Logistic GenLin GLMM Cox SVM Bayes Net

Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot shows the IBM SPSS Modeler interface. The main workspace contains a workflow diagram with the following nodes and connections:

- Electronics\_Data** (Source node) connects to **Partition** (Partitioning node).
- Partition** connects to two **Status** (Monitoring node) nodes.
- The top **Status** node connects to **Analysis** (Analysis node).
- The bottom **Status** node connects to **Analysis**.
- The **Analysis** node is highlighted with a mouse cursor.

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

All Automated Classification Association

Auto Classifier Auto Numeric Auto Cluster Time Series TCM C&R Tree Quest CHAID Decision List Linear C5.0 Regression PCA/Factor Neural Net Feature Selection Discriminant Logistic GenLin GLMM Cox SVM Bayes Net

Analysis of [Status] #6

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Individual Models

Comparing \$N-Status with Status

'Partition'	1_Training		2_Testing	
Correct	1,988	80.94%	2,045	80.29%
Wrong	468	19.06%	502	19.71%
Total	2,456		2,547	

Coincidence Matrix for \$N-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	898	191
Current	277	1,090

'Partition' = 2_Testing	Churned	Current
Churned	877	199
Current	303	1,168

Comparing \$C-Status with Status

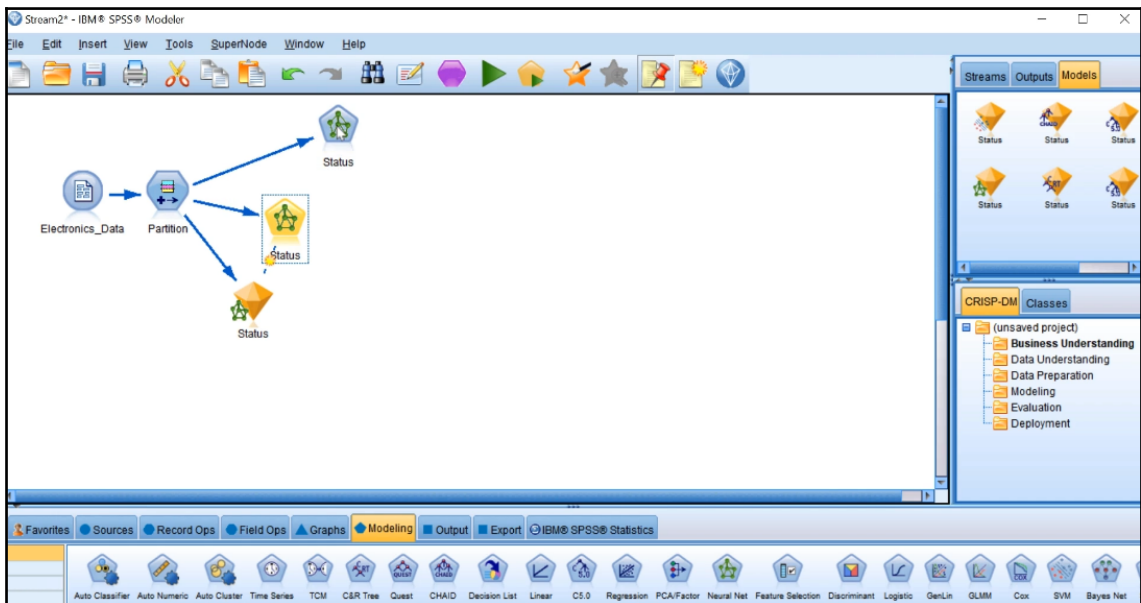
'Partition'	1_Training		2_Testing	
Correct	2,246	91.45%	2,171	85.24%
Wrong	210	8.55%	376	14.76%
Total	2,456		2,547	

Coincidence Matrix for \$C-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	994	95

OK

229176228	641835376	1	2552	5000
Speakers	Premier	Premier	Premier	Premier
Premier	Speakers	Years as customer	Stereos	Years as customer
TVs	Years as customer	Speakers	Speakers	Stereos
Stereos	Stereos	TV categories	TVs	TVs
Years as customer	TVs	Stereos	Years as customer	Speakers
Delivery problems	Delivery problems	TVs	Delivery problems	TV categories
TV categories	Estimated revenue	Estimated revenue	Potential Risk	Number employees
Potential risk	TV categories	Payment method	TV categories	Estimated revenue
Estimated revenue	Problems	Number employees	Estimated revenue	Delivery problems
Number employees	Number stores	Delivery problems	Number employees	Problems



Status

Objective: Standard model

Fields Build Options Model Options Annotations

Use predefined roles  
 Use custom field assignments

Fields:

Sort: None

- Speaker\_Discount
- Stereo\_Discount
- Location
- Manufacturer

Targets\*:

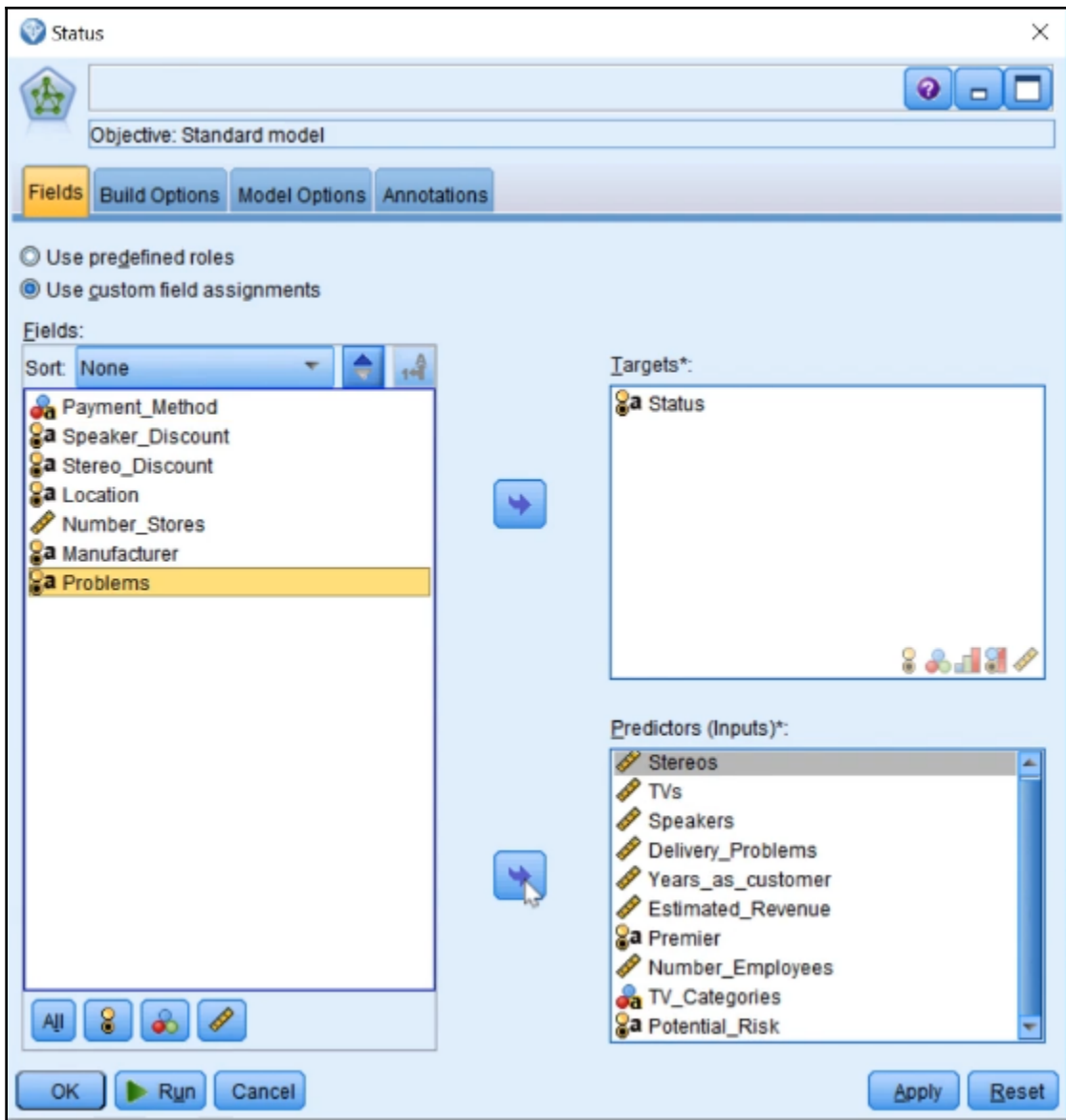
- Status

Predictors (Inputs)\*:

- Delivery\_Problems
- Years\_as\_customer
- Estimated\_Revenue
- Payment\_Method
- Premier
- Number\_Employees
- Number\_Stores
- Problems
- TV\_Categories
- Potential\_Risk

All [Icon] [Icon] [Icon]

OK Run Cancel Apply Reset



Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot displays the IBM SPSS Modeler Stream2\* interface. The main workspace contains a workflow diagram starting with 'Electronics\_Data' (a document icon) pointing to a 'Partition' node (a cube icon). From the 'Partition' node, three arrows lead to three 'Status' nodes (hexagonal icons with a green tree). A dashed arrow connects the top 'Status' node to a 'Status' node (a yellow cube icon). Below this, another 'Status' node (a green cube icon) points to a final 'Status' node (a yellow cube icon). The right-hand side features a 'Streams' tab with a grid of 'Status' nodes. Below that is a 'CRISP-DM' section with a 'Classes' tree showing a project structure: (unsaved project) -> Business Understanding -> Data Understanding -> Data Preparation -> Modeling -> Evaluation -> Deployment. The bottom toolbar includes tabs for Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, and IBM® SPSS® Statistics. Below these are icons for Table, Matrix, Analysis, Data Audit, Transform, Statistics, Means, Report, Set Globals, Sim Fit, R Output, and Sim Eval.

Streams Outputs Models

Electronics\_Data Partition Status Status Status Status Status Status

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit R Output Sim Eval



Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot displays the IBM SPSS Modeler Stream2\* interface. The main workspace contains a workflow diagram starting with 'Electronics\_Data' (document icon) pointing to 'Partition' (hexagon icon). From 'Partition', three arrows lead to three 'Status' (hexagon icon) nodes. The top 'Status' node has a dashed arrow pointing to another 'Status' node. The bottom 'Status' node has a solid arrow pointing to a third 'Status' node. This third 'Status' node has a solid arrow pointing to an 'Analysis' (magnifying glass icon) node. The right-hand side features a 'Streams' tab with a grid of 'Status' nodes. Below that is a 'CRISP-DM' section with a 'Classes' tree showing: (unsaved project), Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The bottom toolbar includes tabs for Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, and IBM® SPSS® Statistics. Below these are icons for Table, Matrix, Analysis, Data Audit, Transform, Statistics, Means, Report, Set Globals, Sim Fit, R Output, and Sim Eval.

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit R Output Sim Eval

Analysis of [Status] #9

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

- Individual Models
  - Comparing \$N-Status with Status
 

'Partition'	1_Training		2_Testing	
Correct	1,988	80.94%	2,045	80.29%
Wrong	468	19.06%	502	19.71%
Total	2,456		2,547	

    - Coincidence Matrix for \$N-Status (rows show actuals)
 

'Partition' = 1_Training	Churned	Current
Churned	898	191
Current	277	1,090

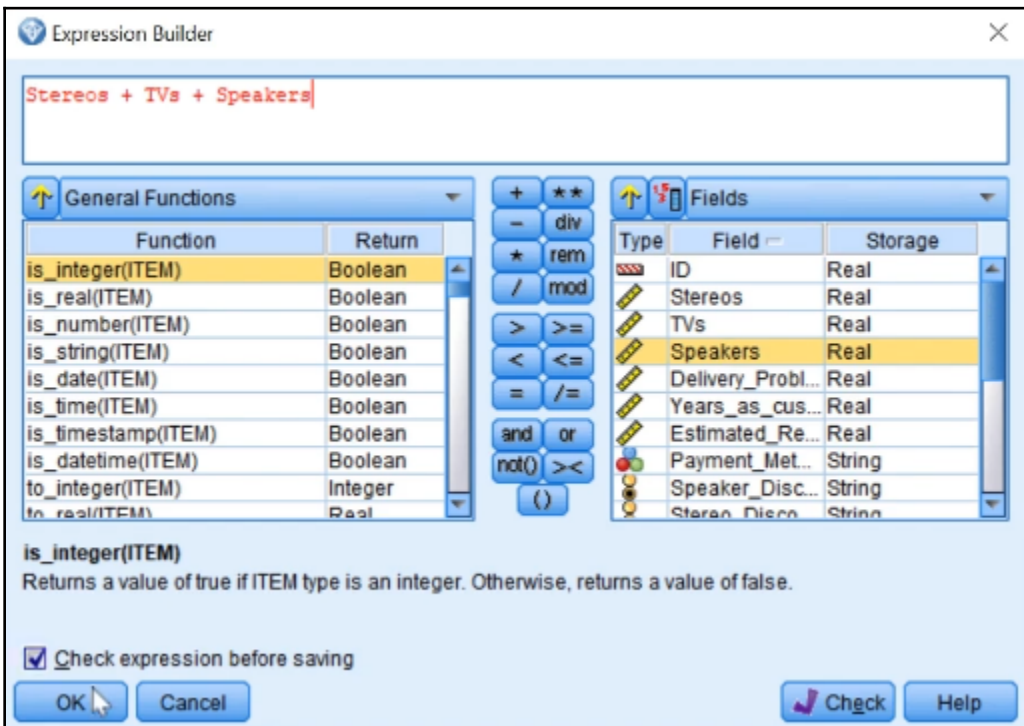
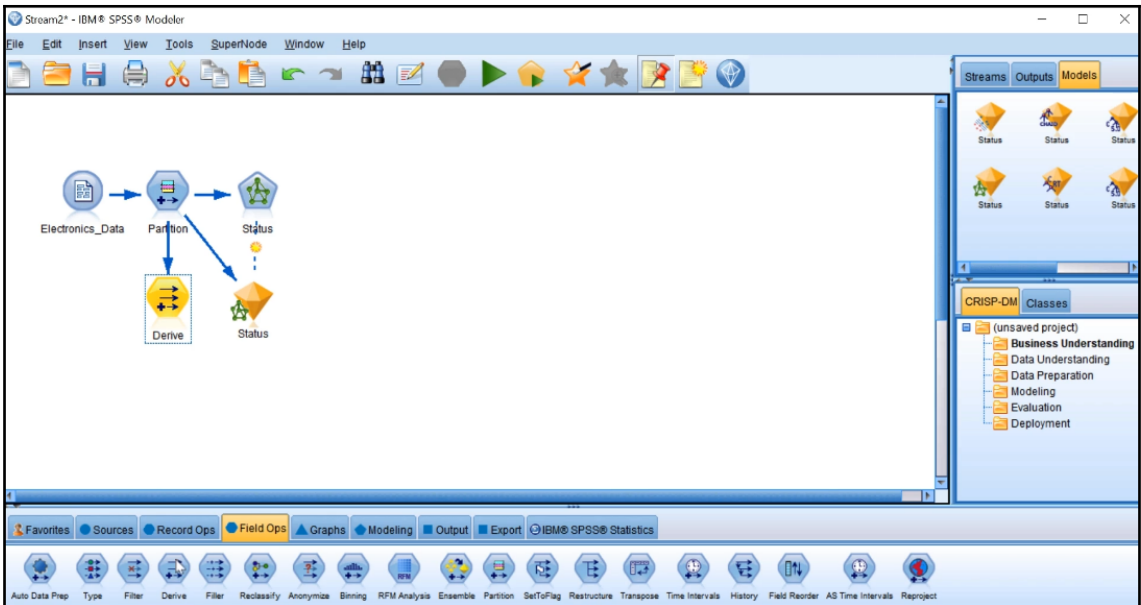
'Partition' = 2_Testing	Churned	Current
Churned	877	199
Current	303	1,168
    - Comparing \$N1-Status with Status
 

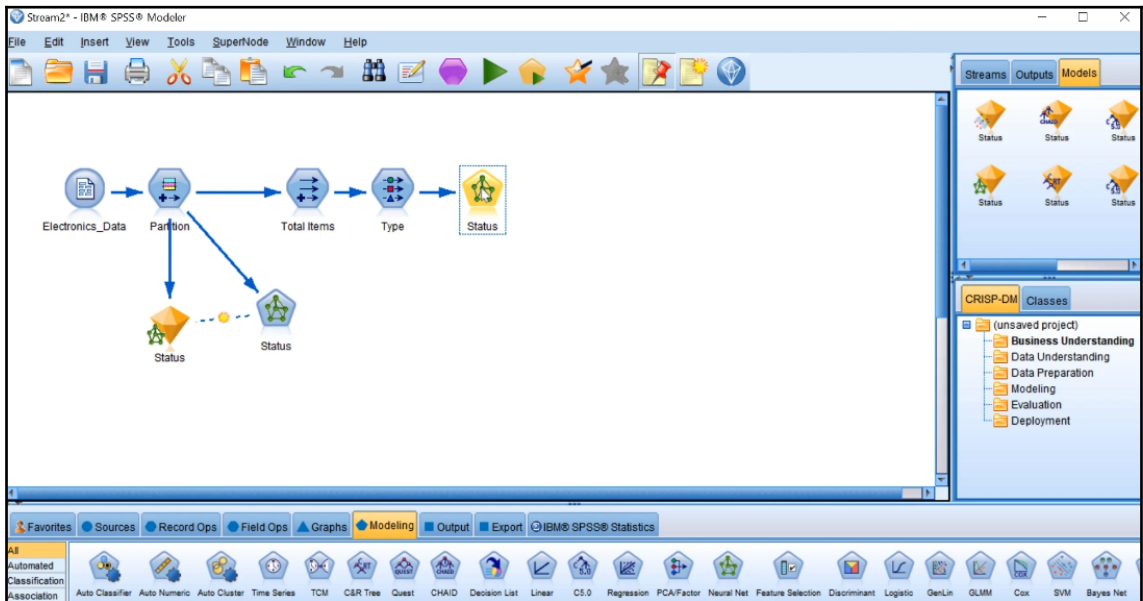
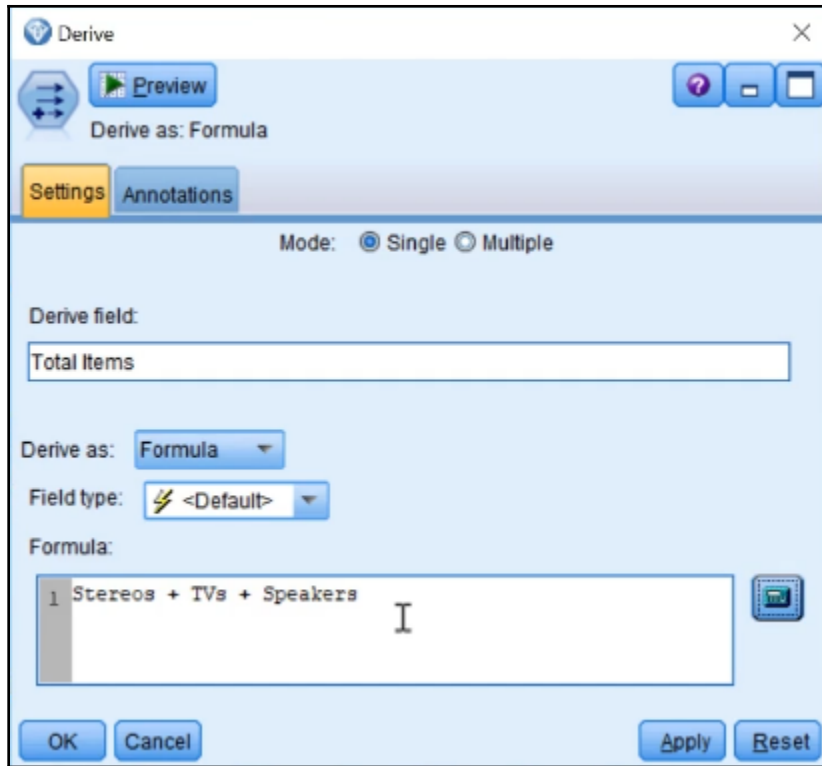
'Partition'	1_Training		2_Testing	
Correct	2,023	82.37%	2,098	82.37%
Wrong	433	17.63%	449	17.63%
Total	2,456		2,547	

      - Coincidence Matrix for \$N1-Status (rows show actuals)
 

'Partition' = 1_Training	Churned	Current
Churned	911	178

OK





Type

Preview

Types Format Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
Status	Flag	Current...		None	Target
Number_Em...	Continuous	[102.0,10...		None	Input
Number_Sto...	Continuous	[1.0,10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categori...	Nominal	Low,Medi...		None	Input
Potential_Risk	Flag	Risk/Reg...		None	Input
Partition	Nominal	"1_Trainin...		None	Partition
Total Items	Continuous	[0.0,488.0]		None	Input

View current fields View unused field settings

OK Cancel Apply Reset

Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

```

    graph LR
      Electronics_Data[Electronics_Data] --> Partition[Partition]
      Partition --> Total_Items[Total Items]
      Partition --> Status1[Status]
      Partition --> Status2[Status]
      Total_Items --> Type[Type]
  
```

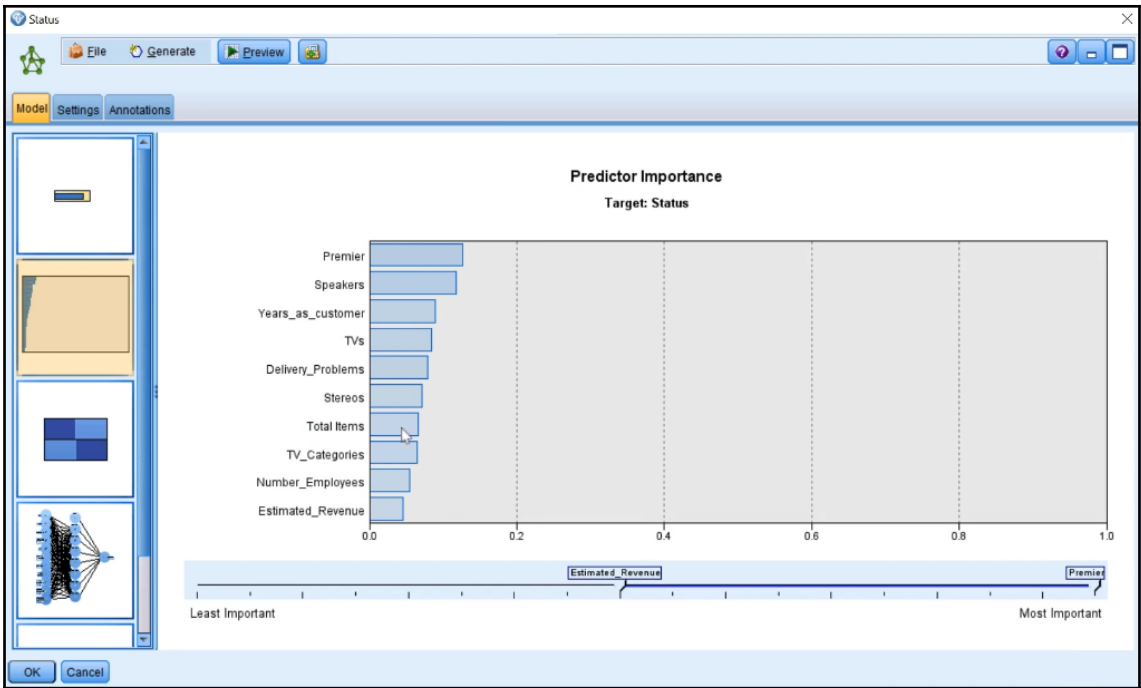
Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
  - Business Understanding
  - Data Understanding
  - Data Preparation
  - Modeling
  - Evaluation
  - Deployment

Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

Auto Data Prep Type Filter Derive Filter Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose Time Intervals History Field Reorder AS Time Intervals Reproject



Stream2<sup>®</sup> - IBM<sup>®</sup> SPSS<sup>®</sup> Modeler

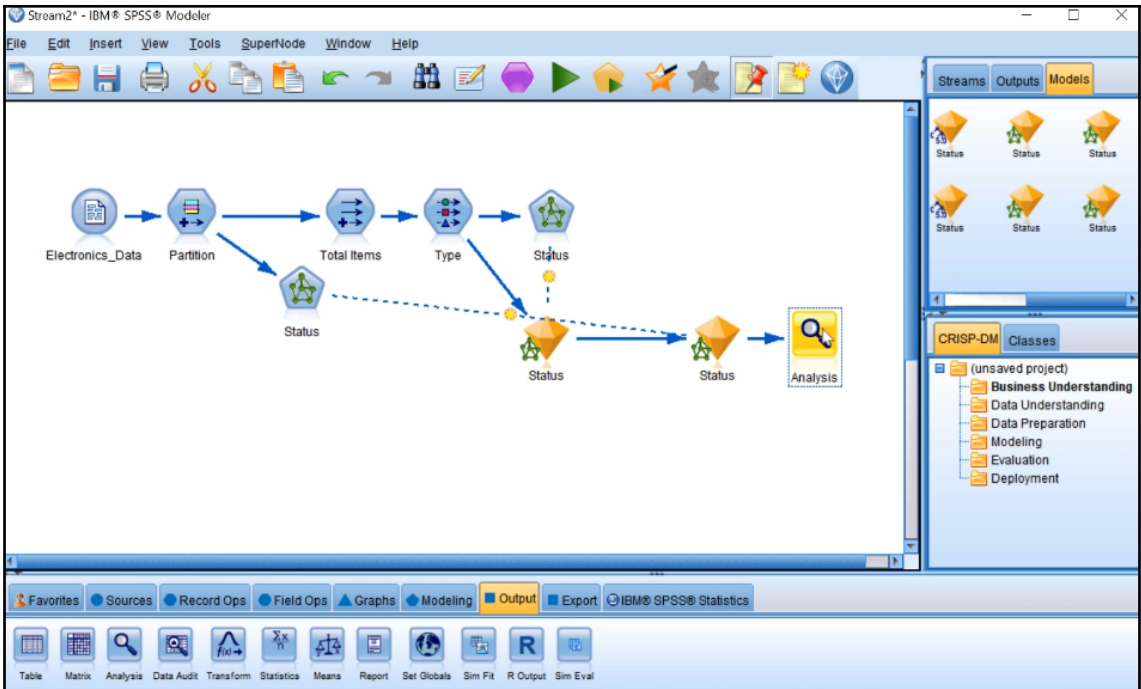
File Edit Insert View Tools SuperNode Window Help

Streams Outputs Models

CRISP-DM Classes

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit R Output Sim Eval





Analysis of [Status] #10

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Individual Models

Comparing \$N-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,002	81.51%	2,048	80.41%
Wrong	454	18.49%	499	19.59%
Total	2,456		2,547	

Coincidence Matrix for \$N-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	886	203
Current	251	1,116

'Partition' = 2_Testing	Churned	Current
Churned	850	226
Current	273	1,198

Comparing \$N1-Status with Status

'Partition'	1_Training		2_Testing	
Correct	1,988	80.94%	2,045	80.29%
Wrong	468	19.06%	502	19.71%
Total	2,456		2,547	

Coincidence Matrix for \$N1-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	898	191

OK

Var. File ×

\\MacHome\Desktop\Loan

File: \\MacHome\Desktop\Loan

```

age, income, gender, marital, numkids, numcards, howpaid, mortgage, storecar, risk, inc
44.000000, 59944.000000, "m", "married", 1.000000, 2.000000, "monthly", "y", 2.000000
35.000000, 59692.000000, "m", "married", 1.000000, 1.000000, "monthly", "y", 1.000000
34.000000, 59463.000000, "m", "married", 0.000000, 2.000000, "monthly", "y", 1.000000

```

Read field names from file
  Specify number of fields:

Skip header characters: 
 EOL comment characters:

Strip lead and trail spaces:
  None
  Left
  Right
  Both

Invalid characters:
  Discard
  Replace with

Encoding: 
 Decimal symbol:

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

Automatically recognize dates and times

Treat square brackets as lists

**Field delimiters**
 Space
  Comma
  Tab
  Newline
  Other

Non-printing characters
  Allow multiple blank delimiters

**Quotes**
 Single quotes: 
 Double quotes:

Var. File ×

Preview Refresh

W:\MacHome\Desktop\Loan

File Data Filter **Types** Annotations

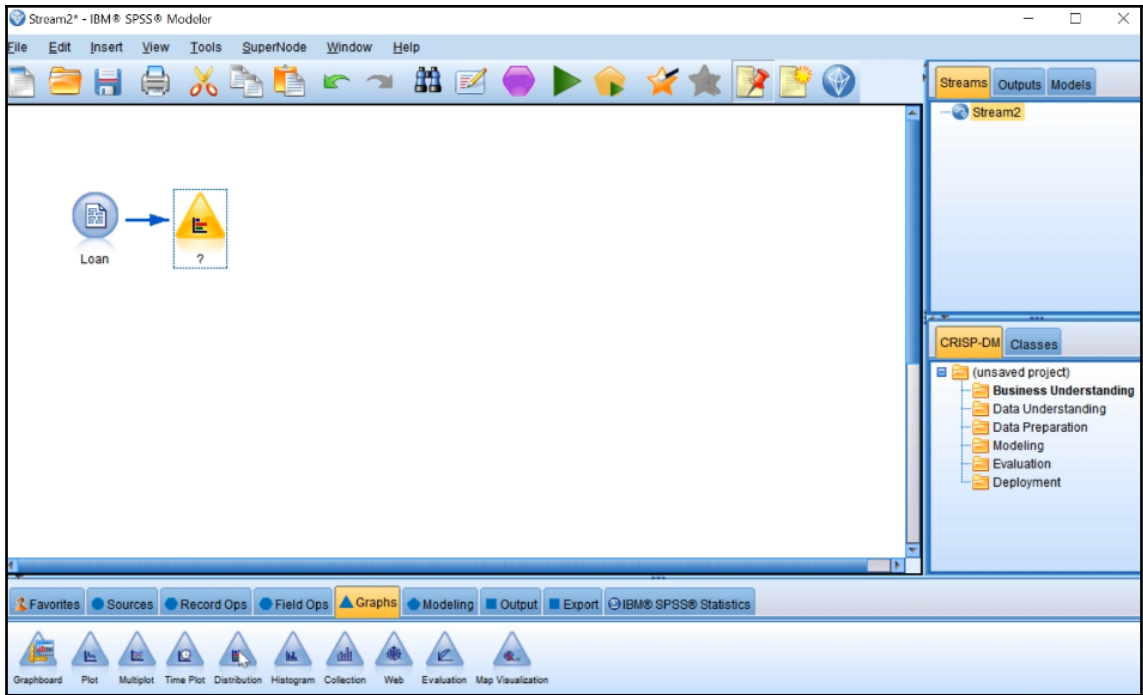
Read Values Clear Values Clear All Values

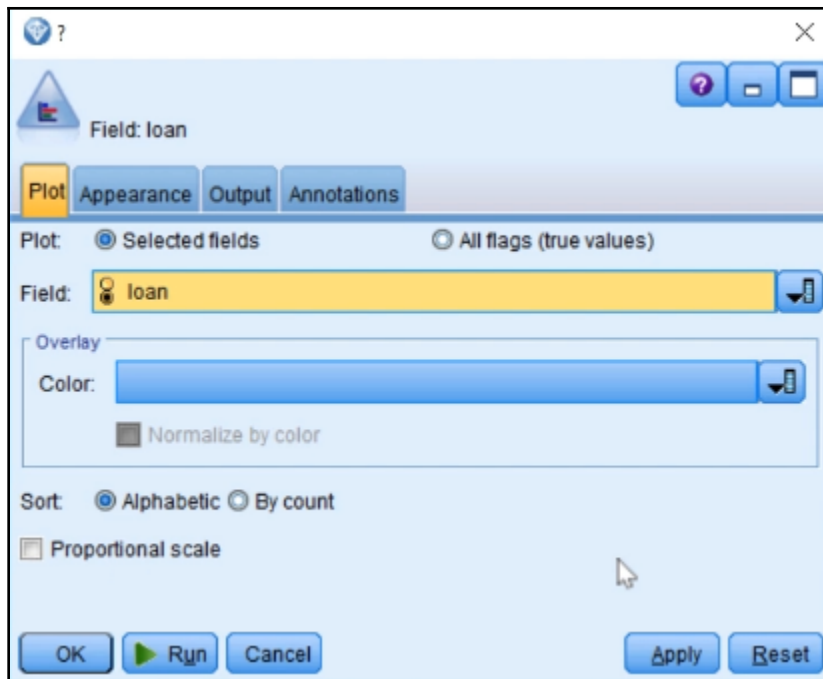
Field	Measurement	Values	Missing	Check	Role
age	Continuous			None	Input
income	Continuous			None	Input
gender	Categorical			None	Input
marital	Categorical			None	Input
numkids	Continuous			None	Input
numcards	Continuous			None	Input
howpaid	Categorical			None	Input
mortgage	Categorical			None	Input
storecar	Continuous			None	Input
risk	Categorical			None	Input
income1k	Continuous			None	Input
loan	Categorical			None	Input

- Input
- Target**
- Both
- None
- Partition
- Split
- Frequency
- Record ID

View current fields  View unused field settings

OK Cancel Apply Reset

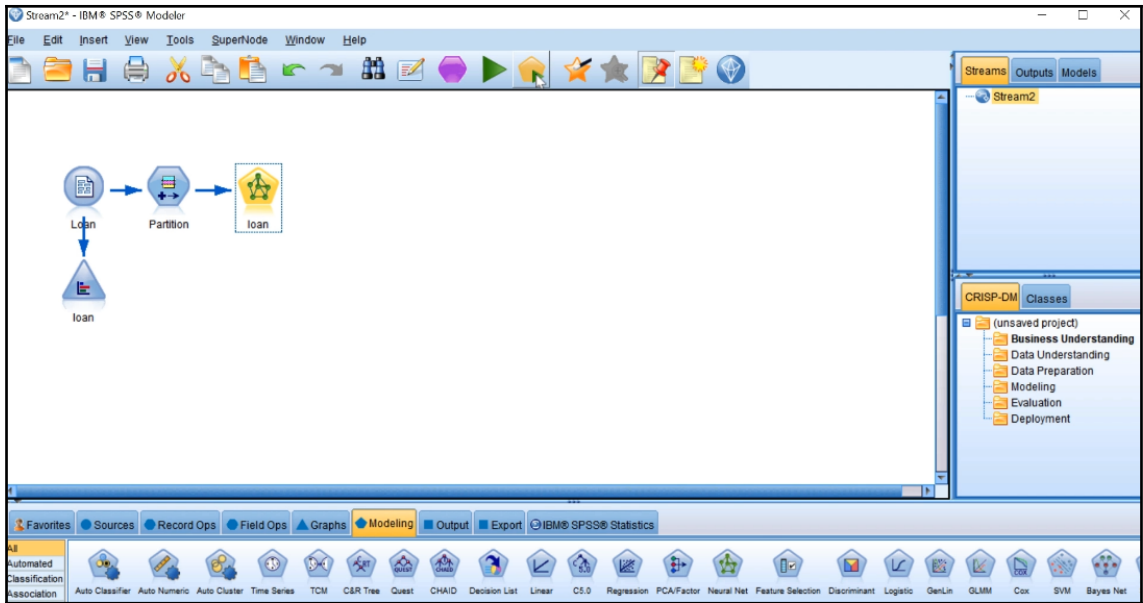




The window has a menu bar with 'File', 'Edit', 'Generate', and 'View'. Below the menu are 'Table', 'Graph', and 'Annotations' tabs. The table displays the distribution of the 'loan' field.

Value	Proportion	%	Count
No		13.65	335
Yes		86.35	2120

An 'OK' button is located at the bottom right of the window.



Analysis of [loan] #2

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field loan

Comparing \$N-loan with loan

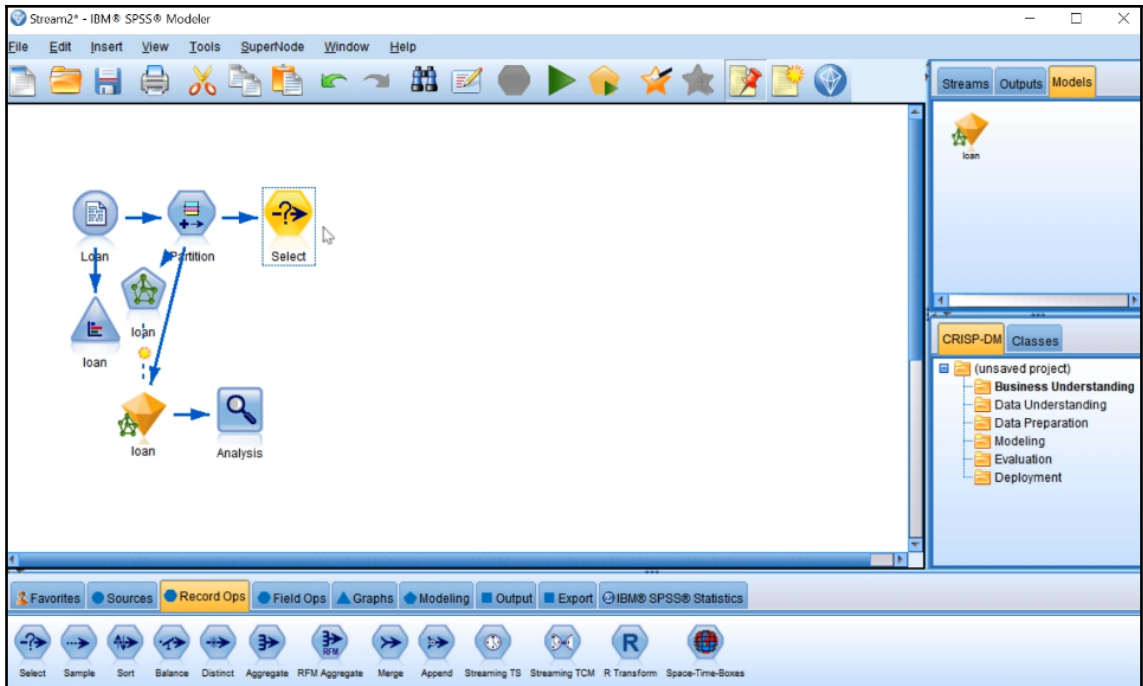
'Partition'	1_Training		2_Testing	
Correct	1,082	93.28%	1,213	93.67%
Wrong	78	6.72%	82	6.33%
Total	1,160		1,295	

Coincidence Matrix for \$N-loan (rows show actuals)

'Partition' = 1_Training		No	Yes
No		121	47
Yes		31	961

'Partition' = 2_Testing		No	Yes
No		124	43
Yes		39	1,089

OK



Expression Builder

Partition

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

Fields

Type	Field	Storage
Integer	total	String
Integer	numkids	Real
Integer	numcards	Real
Integer	howpaid	String
Integer	mortgage	String
Integer	storecar	Real
Integer	risk	String
Integer	income1k	Real
Integer	loan	String
Integer	Partition	String

**is\_integer(ITEM)**  
Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel Check Help



Expression Builder

Partition = |

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

Fields

Type	Field Name	Return Type
	numkids	Real
	numcards	Real
	howpaid	String
	mortgage	String
	storecar	Real
	risk	String
	income1k	Real
	loan	String
	Partition	String

**is\_integer(ITEM)**  
Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel Check Help

Expression Builder

Partition =

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

is\_integer(ITEM)  
Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel Check Help

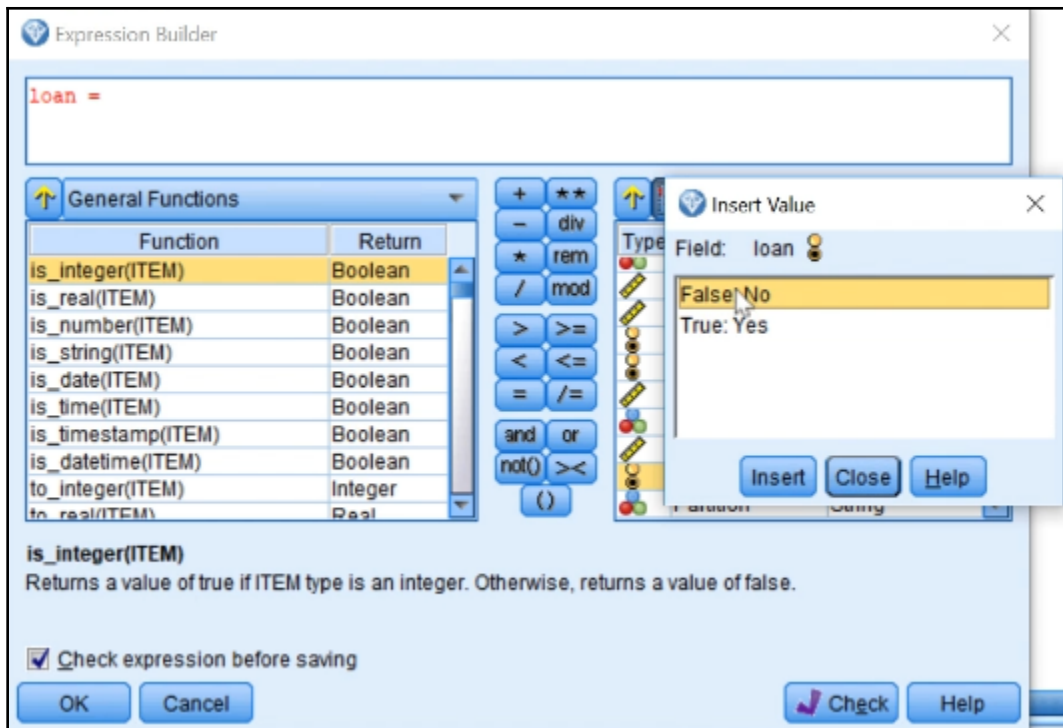
Insert Value

Field: Partition

1\_Training  
2\_Testing

Insert Close Help





Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot displays the IBM SPSS Modeler Stream2\* interface. The main workspace contains a workflow diagram with the following nodes and connections:

- A **Loan** source node (document icon) connects to a **Partition** node (grid icon).
- The **Partition** node connects to a **Select** node (filter icon).
- The **Select** node branches into two paths:
  - One path leads to another **Select** node (filter icon).
  - The other path leads to a **Select** node (filter icon) that is highlighted with a dashed border.
- A **Loan** source node (document icon) connects to an **Analysis** node (magnifying glass icon).
- The **Analysis** node connects to a **loan** target node (diamond icon).

The right-hand pane shows the **Streams** tab with a **loan** stream. Below it, the **CRISP-DM** classes are listed:

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

The bottom toolbar includes the following icons and labels:

- Select
- Sample
- Sort
- Balance
- Distinct
- Aggregate
- RFM Aggregate
- Merge
- Append
- Streaming TS
- Streaming TCM
- R Transform
- Space-Time-Boxes

Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot displays the IBM SPSS Modeler Stream2\* interface. The main workspace contains a workflow diagram with the following nodes and connections:

- Loan** (Source) connects to **Partition** (Record Ops).
- loan** (Source) connects to **Partition**.
- Partition** connects to **Select** (Record Ops).
- Select** connects to two other **Select** nodes.
- Select** (bottom) connects to **Sample** (Record Ops).
- loan** (Source) connects to **Analysis** (Analysis).

On the right side, the **Streams** panel shows a **loan** stream. Below it, the **CRISP-DM** **Classes** panel lists the following stages:

- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

At the bottom, the **Record Ops** palette is active, showing icons for various operations: Select, Sample, Sort, Balance, Distinct, Aggregate, RFM Aggregate, Merge, Append, Streaming TS, Streaming TCM, R Transform, and Space-Time-Boxes.

Sample ×

... ▶ Preview ? ▭ ▭

**Settings** Annotations

Sample method:  Simple  Complex

Simple Sample

Mode:  Include sample  Discard sample

Sample:  First  1-in-n  Random %

Use block-level sampling (in-database only)

Maximum sample size

Repeatable partition assignment Seed: 1162541 Generate

OK Cancel Apply Reset

Stream2\* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Window Help

The screenshot displays the IBM SPSS Modeler Stream2\* interface. The main workspace contains a workflow diagram with the following nodes and connections:

- Loan** (Source) connects to **Partition** (Partitioning).
- loan** (Source) connects to **Partition**.
- Partition** connects to **Select** (Filtering).
- loan** (Source) connects to **Analysis** (Analysis).
- Select** (Filtering) connects to two **Select** (Filtering) nodes.
- The top **Select** (Filtering) node connects to **Append** (Merging).
- The bottom **Select** (Filtering) node connects to a **33.0%** (Output) node.
- Append** (Merging) connects to the **33.0%** (Output) node.

The right-hand pane shows the **Streams** tab with a **loan** stream. Below it, the **CRISP-DM** class tree is visible, showing the following structure:





- CRISP-DM
- Classes
- (unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

The bottom toolbar includes the following icons and labels:

- Record Ops
- Field Ops
- Graphs
- Modeling
- Output
- Export
- IBM® SPSS® Statistics
- Select
- Sample
- Sort
- Balance
- Distinct
- Aggregate
- RFM Aggregate
- Merge
- Append
- Streaming TS
- Streaming TCM
- R Transform
- Space-Time-Boxes



Append ×
























 **Preview**   

Append 2 datasets

**Inputs** Append Annotations

Match fields by:  Position  Name  Match case

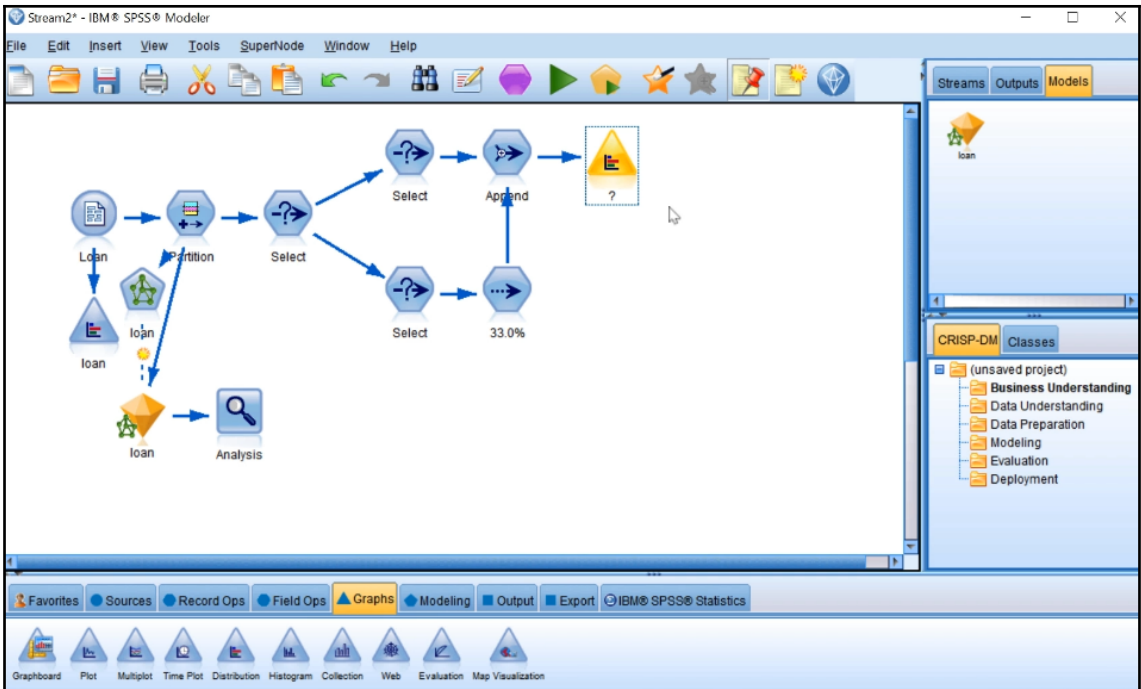
Preview of field matches and structure

Output Field	1[Loan:Select]	2[Loan:33.0%]
 age	 age	 age
 income	 income	 income
 gender	 gender	 gender
 marital	 marital	 marital
 numkids	 numkids	 numkids
 numcards	 numcards	 numcards
 howpaid	 howpaid	 howpaid
 mortgage	 mortgage	 mortgage

Include fields from:  Main dataset only  All datasets

Tag records by including source dataset in field

**OK** **Cancel** **Apply** **Reset**



Distribution of loan #4

File Edit Generate View

Table Graph Annotations

Value	Proportion	%	Count
No		33.6	168
Yes		66.4	332

OK

Analysis of [loan] #3

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field loan

- Individual Models
  - Comparing \$N-loan with loan
 

'Partition'	1_Training		2_Testing	
Correct	1,082	93.28%	1,213	93.67%
Wrong	78	6.72%	82	6.33%
Total	1,160		1,295	

    - Coincidence Matrix for \$N-loan (rows show actuals)
 

'Partition' = 1_Training		No	Yes
No		121	47
Yes		31	961

'Partition' = 2_Testing		No	Yes
No		124	43
Yes		39	1,089
    - Comparing \$N1-loan with loan
 

'Partition'	1_Training		2_Testing	
Correct	1,091	94.05%	1,211	93.51%
Wrong	69	5.95%	84	6.49%
Total	1,160		1,295	

      - Coincidence Matrix for \$N1-loan (rows show actuals)
 

'Partition' = 1_Training		No	Yes
No		167	1

OK

# Chapter 5: Advanced Ways of Improving Models

Model 1 Prediction	Model 1 Confidence	Model 2 Prediction	Model 2 Confidence	Model 3 Prediction	Model 3 Confidence	Combined Prediction	Final Confidence
Leave	.80	Leave	.70	Leave	.60	Leave	.70
Leave	.80	Leave	.80	Stay	.90	Leave	.53

Model 1 Prediction	Model 1 Confidence	Model 2 Prediction	Model 2 Confidence	Model 3 Prediction	Model 3 Confidence	Combined Prediction	Final Confidence
Leave	.80	Leave	.70	Leave	.60	Leave	.80
Leave	.80	Leave	.80	Stay	.90	Stay	.90

Table (24 fields, 5,003 records)

File Edit Generate

Table Annotations

	TV_Categories	Potential_Risk	Partition	\$N-Status	\$NC-Status	\$S-Status	\$SP-Status	
1	Medium	Regular	1_Training	Churned		0.840	Churned	0.967
2	None	Regular	1_Training	Churned		0.754	Churned	0.967
3	None	Regular	2_Testing	Churned		0.935	Churned	0.897
4	Medium	Regular	2_Testing	Current		0.524	Current	0.712
5	Low	Regular	1_Training	Current		0.879	Current	0.916
6	None	Regular	1_Training	Churned		0.934	Churned	0.972
7	None	Regular	2_Testing	Churned	None	0.774	Churned	0.780
8	None	Regular	1_Training	Current		0.733	Current	0.844
9	None	Regular	1_Training	Churned		0.914	Churned	0.949
10	None	Regular	1_Training	Churned		0.899	Churned	0.934
11	None	Regular	2_Testing	Current		0.794	Current	0.939
12	None	Regular	2_Testing	Churned		0.803	Current	0.550
13	Low	Regular	2_Testing	Current		0.962	Current	0.987
14	None	Regular	1_Training	Churned		0.646	Churned	0.661
15	Low	Regular	2_Testing	Current		0.950	Current	0.953
16	None	Regular	2_Testing	Current	None	0.986	Current	0.978
17	None	Regular	1_Training	Churned		0.832	Churned	0.899
18	None	Regular	1_Training	Current		0.955	Current	0.853
19	None	Regular	2_Testing	Current		0.746	Current	0.586
20	Low	Regular	1_Training	Current		0.891	Current	0.832

OK

Analysis of [Status]

File Edit

Analysis Annotations

Collapse All Expand All

Correct	1,988	80.94%	2,045	80.29%
Wrong	468	19.06%	502	19.71%
Total	2,456		2,547	

☐ Coincidence Matrix for \$N-Status (rows show actuals)

'Partition' = 1_Training		Churned	Current
Churned		898	191
Current		277	1,090

'Partition' = 2_Testing		Churned	Current
Churned		877	199
Current		303	1,168

☐ Comparing \$\$-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,078	84.61%	2,042	80.17%
Wrong	378	15.39%	505	19.83%
Total	2,456		2,547	


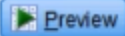



☐ Coincidence Matrix for \$\$-Status (rows show actuals)

'Partition' = 1_Training		Churned	Current
Churned		905	184
Current		194	1,173

'Partition' = 2_Testing		Churned	Current
Churned		820	256
Current		249	1,222


OK

Ensemble ×

Number of models included in ensemble: 2

**Settings** Annotations

Target field for ensemble:  

Filter out fields generated by ensembled models

Flag Target

Ensemble method:

If voting is tied, select value using:

Random selection  Highest confidence  Raw propensity  Adjusted propensity

Analysis of [Status] #1

File Edit

Analysis Annotations

Collapse All Expand All

	Churned	Current
'Partition' = 2_Testing	877	199
Churned	877	199
Current	303	1,168

Comparing \$\$-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,078	84.61%	2,042	80.17%
Wrong	378	15.39%	505	19.83%
Total	2,456		2,547	

Coincidence Matrix for \$\$-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	905	184
Current	194	1,173

'Partition' = 2_Testing	Churned	Current
Churned	820	256
Current	249	1,222

Comparing \$XF-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,068	84.2%	2,085	81.86%
Wrong	388	15.8%	462	18.14%
Total	2,456		2,547	

Coincidence Matrix for \$XF-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	918	171
Current	217	1,150

'Partition' = 2_Testing	Churned	Current
Churned	868	208
Current	254	1,217

Agreement between \$N-Status \$\$-Status \$XF-Status

'Partition'	1_Training		2_Testing	
Agree	2,170	88.36%	2,214	86.93%
Disagree	286	11.64%	333	13.07%
Total	2,456		2,547	

Comparing Agreement with Status

'Partition'	1_Training		2_Testing	
Correct	1,890	87.1%	1,877	84.78%
Wrong	280	12.9%	337	15.22%

Expression Builder

```
if ('$NC-Status' > '$SP-Status') then '$N-Status' elseif ('$SP-Status' > '$NC-Status') then '$S-Status' else '0' endif
```

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

**is\_integer(ITEM)**  
Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel

Fields

Type	Field	Storage
	number_stores	Real
	Manufacturer	String
	Problems	String
	TV_Categories	String
	Potential_Risk	String
	Partition	String
	\$N-Status	String
	\$NC-Status	Real
	\$S-Status	String
	\$SP-Status	Real

and or  
not() ><  
( )

Check Help



Table (25 fields, 5,003 records)

File Edit Generate

Table Annotations





	ial_Risk	Partition	\$N-Status	\$NC-Status	\$S-Status	\$SP-Status	Combined_prediction
1	ar	1_Training	Churned		0.840	Churned	0.967 Churned
2	ar	1_Training	Churned		0.754	Churned	0.967 Churned
3	ar	2_Testing	Churned		0.935	Churned	0.897 Churned
4	ar	2_Testing	Current		0.524	Current	0.712 Current
5	ar	1_Training	Current		0.879	Current	0.916 Current
6	ar	1_Training	Churned		0.934	Churned	0.972 Churned
7	ar	2_Testing	Churned		0.774	Churned	0.780 Churned
8	ar	1_Training	Current		0.733	Current	0.844 Current
9	ar	1_Training	Churned		0.914	Churned	0.949 Churned
10	ar	1_Training	Churned		0.899	Churned	0.934 Churned
11	ar	2_Testing	Current		0.794	Current	0.939 Current
12	ar	2_Testing	Churned		0.803	Current	0.550 Churned
13	ar	2_Testing	Current		0.962	Current	0.987 Current
14	ar	1_Training	Churned		0.646	Churned	0.661 Churned
15	ar	2_Testing	Current		0.950	Current	0.953 Current
16	ar	2_Testing	Current		0.986	Current	0.978 Current
17	ar	1_Training	Churned		0.832	Churned	0.899 Churned
18	ar	1_Training	Current		0.955	Current	0.853 Current
19	ar	2_Testing	Current		0.746	Current	0.586 Current
20	ar	1_Training	Current		0.891	Current	0.832 Current

OK

Prediction	Confidence	Propensity
Leave	.95	.95
Leave	.60	.60
Stay	.55	.45
Stay	.90	.10

- Propensity = Confidence (for group of interest)
- Propensity = 1 – Confidence (for the other group)

Status ×

Objective: Standard model

Fields Build Options **Model Options** Annotations

Model name:  Auto  Custom

Model Evaluation


Calculate predictor importance

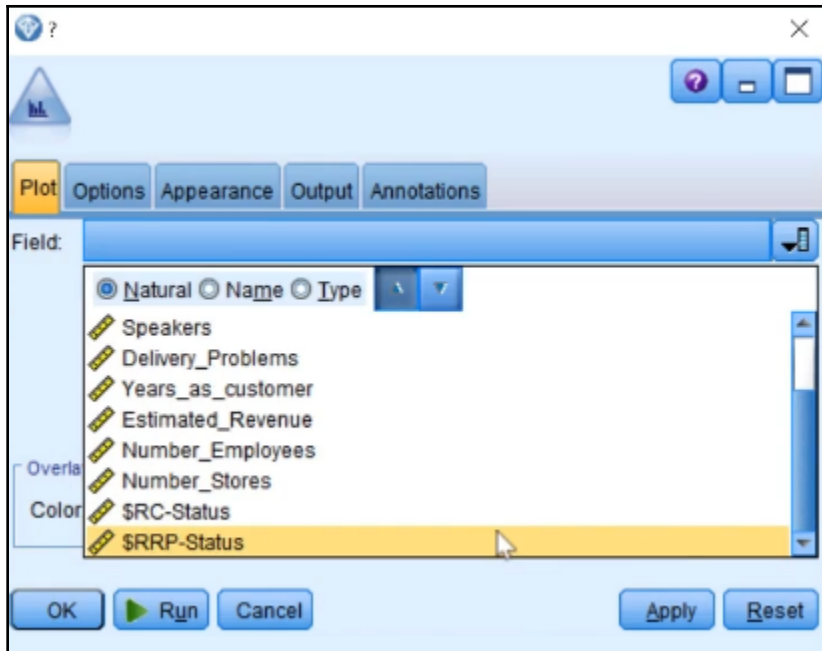
Propensity Scores (valid only for flag targets)

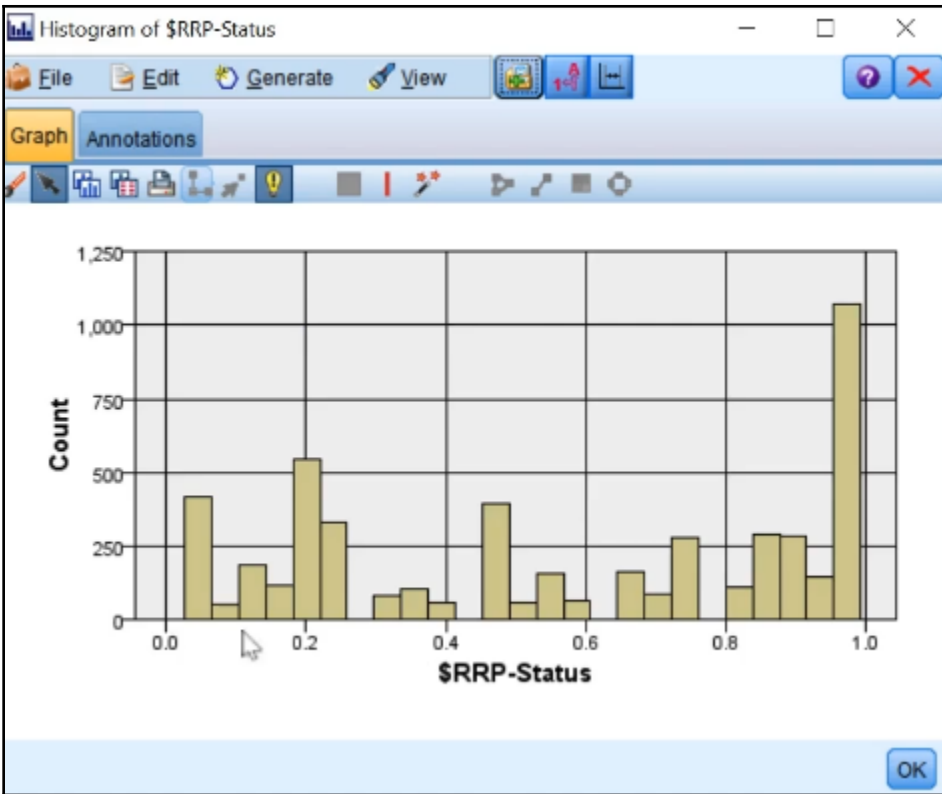
Calculate raw propensity scores

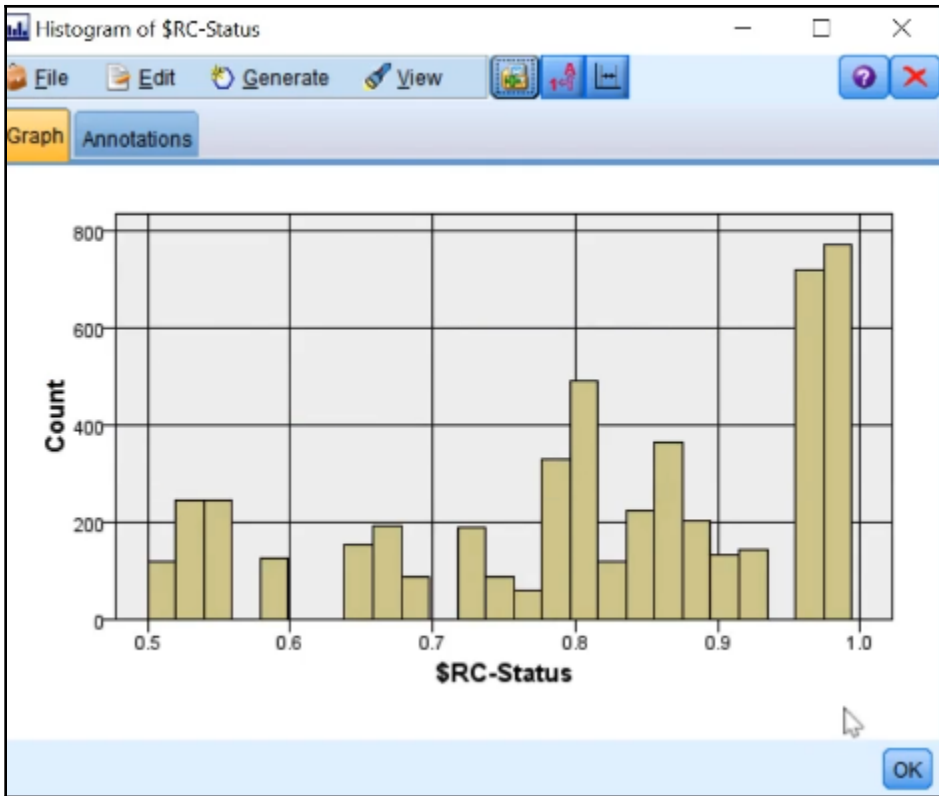
Calculate adjusted propensity scores

Based on:  Testing partition  Validation partition

OK  Cancel Apply Reset







Analysis of [Status] #2

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Individual Models

Comparing \$R-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,028	82.57%	2,038	80.02%
Wrong	428	17.43%	509	19.98%
Total	2,456		2,547	

Coincidence Matrix for \$R-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	921	168
Current	260	1,107

'Partition' = 2_Testing	Churned	Current
Churned	867	209
Current	300	1,171

Comparing \$N-Status with Status

'Partition'	1_Training		2_Testing	
Correct	1,988	80.94%	2,045	80.29%
Wrong	468	19.06%	502	19.71%
Total	2,456		2,547	

Coincidence Matrix for \$N-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	898	191

OK

Type

Preview

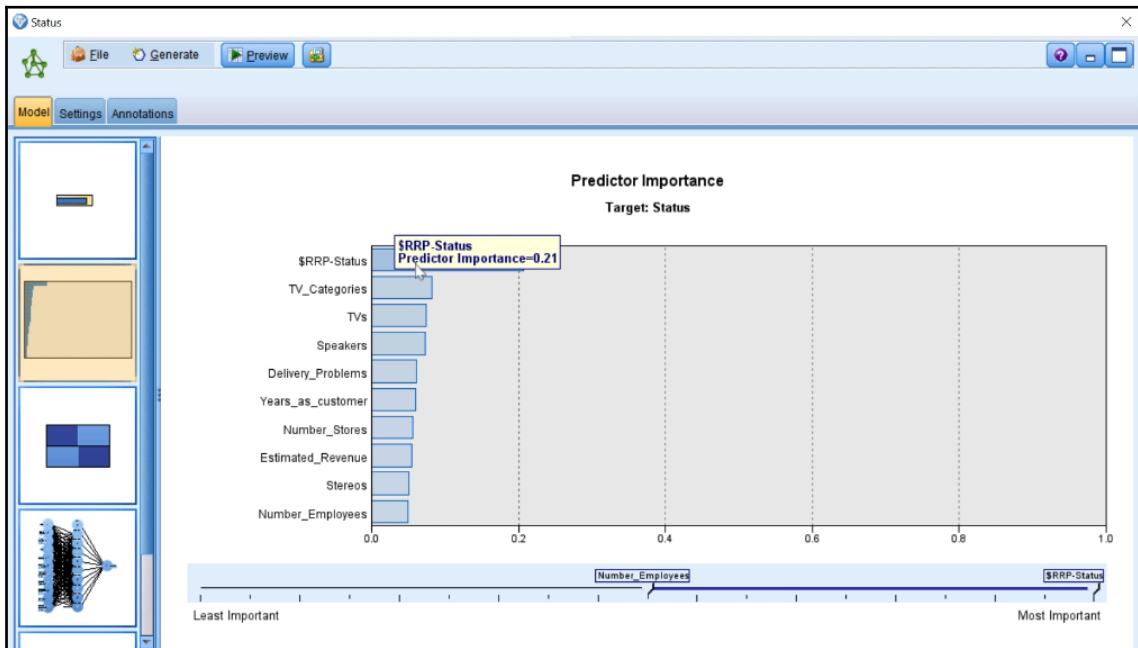
Types Format Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
number_sto...	Continuous	[1.0, 10.0]		None	Input
Manufacturer	Flag	Yes/No		None	Input
Problems	Flag	T/F		None	Input
TV_Categori...	Nominal	Low, Medi...		None	Input
Potential_Risk	Flag	Risk/Reg...		None	Input
Partition	Nominal	"1_Trainin...		None	Partition
\$R-Status	Flag	Current/C...		None	Input
\$RC-Status	Continuous	[0.0, 1.0]		None	None
\$RRP-Status	Continuous	[0.0, 1.0]		None	Input

View current fields View unused field settings

OK Cancel Apply Reset



Analysis of [Status] #3

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

Individual Models

Comparing \$R-Status with Status

'Partition'	1_Training		2_Testing	
Correct	2,028	82.57%	2,038	80.02%
Wrong	428	17.43%	509	19.98%
Total	2,456		2,547	

Coincidence Matrix for \$R-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	921	168
Current	260	1,107

'Partition' = 2_Testing	Churned	Current
Churned	867	209
Current	300	1,171

Comparing \$N-Status with Status

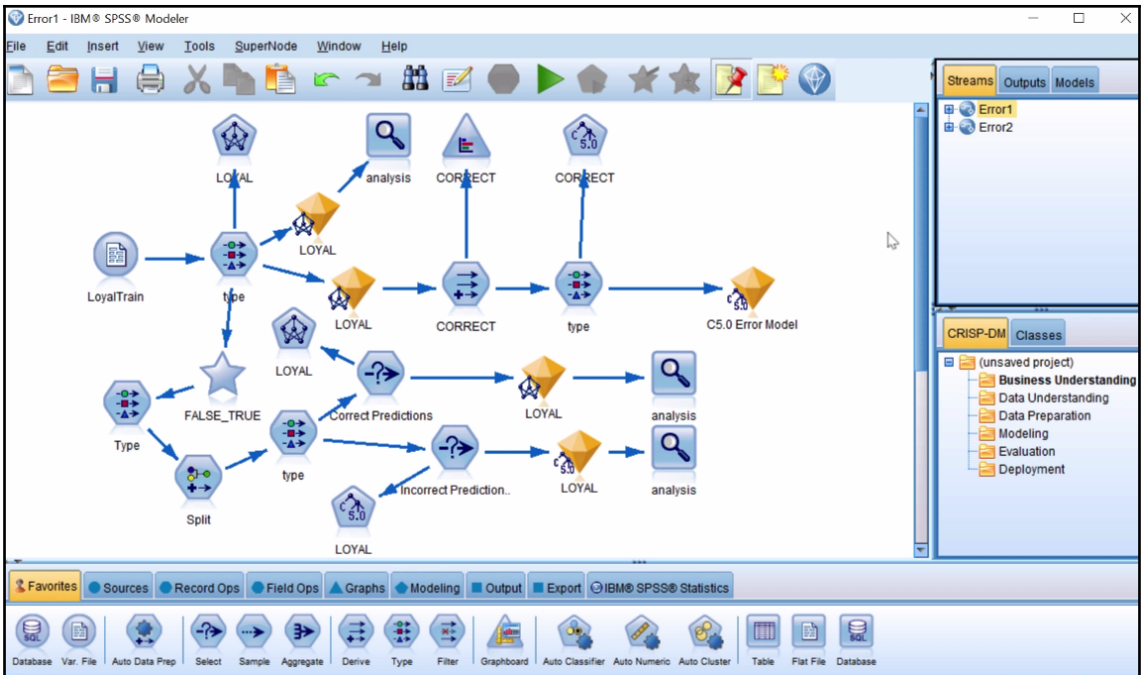
'Partition'	1_Training		2_Testing	
Correct	2,066	84.12%	2,064	81.04%
Wrong	390	15.88%	483	18.96%
Total	2,456		2,547	

Coincidence Matrix for \$N-Status (rows show actuals)

'Partition' = 1_Training	Churned	Current
Churned	911	178

OK





Analysis of [LOYAL] #3

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field LOYAL

- Comparing \$N-LOYAL with LOYAL
 

Correct	872	78.7%
Wrong	236	21.3%
Total	1,108	
- Coincidence Matrix for \$N-LOYAL (rows show actuals)
 


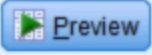
	Leave	Stay
Leave	394	98
Stay	138	478
- Performance Evaluation
 

Leave	0.512
Stay	0.401
- Confidence Values Report for \$NC-LOYAL
 

Range	0.001 - 0.993
Mean Correct	0.642
Mean Incorrect	0.418
Always Correct Above	0.966 (0.99% of cases)
Always Incorrect Below	0.009 (0.27% of cases)
90.03% Accuracy Above	0.59
2.0 Fold Correct Above	0.574 (89.38% of cases)

OK

CORRECT ×

  Preview

Derive as: Flag

**Settings** Annotations

Mode:  Single  Multiple

Derive field:

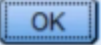
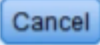
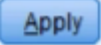
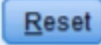
Derive as:

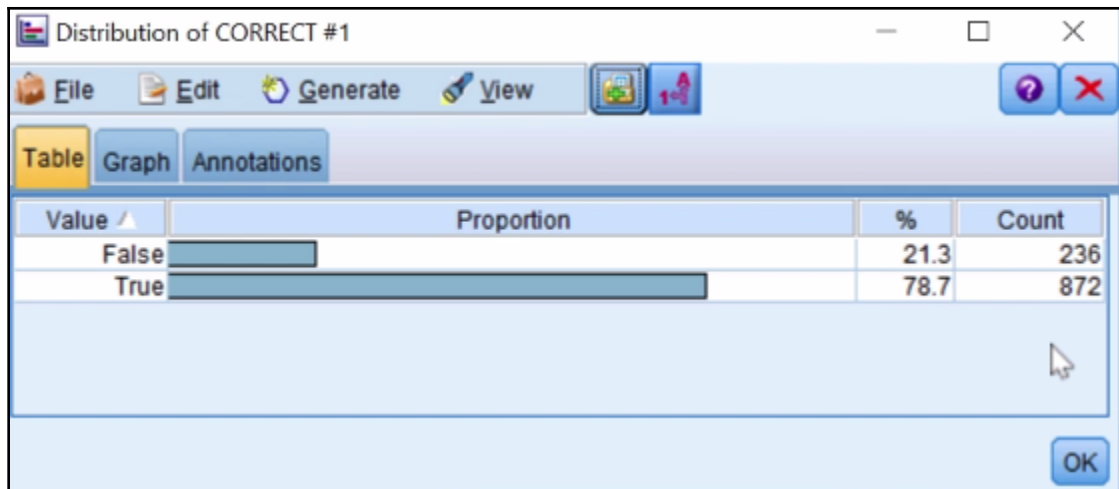
Field type:

True value:  False value:

True when:  

```
1 LOYAL == 'N-LOYAL'
```



C5.0 Error Model

File Generate Preview

Model Summary Settings Annotations

1 2 3 All %

- [-] Rules for False - contains 4 rule(s)
- [-] Rules for True - contains 10 rule(s)
  - Default: True

OK Cancel Apply Reset

C5.0 Error Model

File Generate Preview

Model Summary Settings Annotations

1 2 3 All %

Rules for False - contains 4 rule(s)

- Rule 1 for **False**
  - if SEX = M
  - and International <= 0.986
  - and LONGDIST <= 0.104
  - and STATUS = S
  - then **False**
- Rule 2 for **False**
  - if SEX = M
  - and International <= 0.986
  - and LONGDIST > 0.104
  - and AGE <= 29
  - and Est\_Income <= 40200.700
  - then **False**
- Rule 3 for **False**
  - if SEX = M
  - and International > 0.986
  - and AGE <= 62
  - then **False**
- Rule 4 for **False**
  - if SEX = F
  - and LONGDIST > 0.775
  - then **False**

OK Cancel Apply Reset

C5.0 Error Model

File Generate Preview

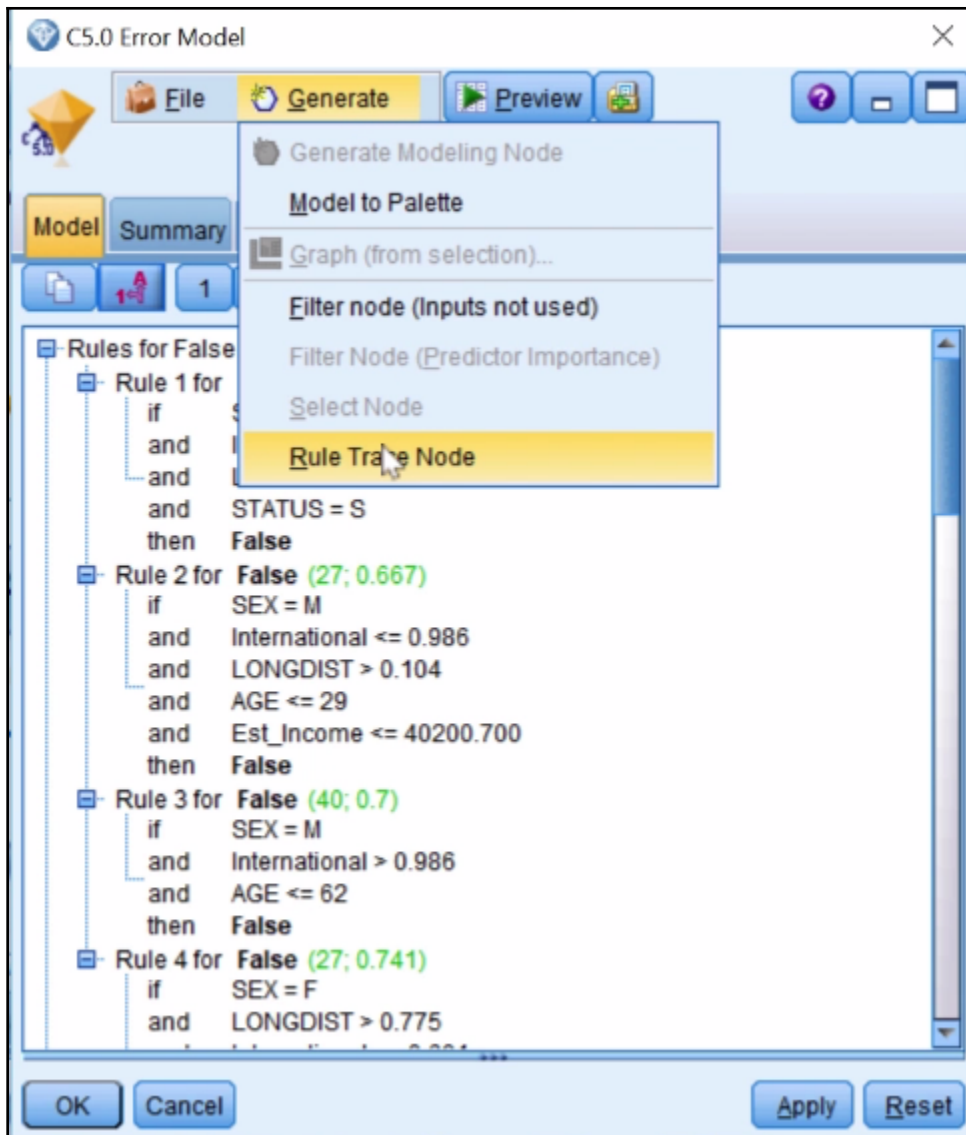
Model Summary Settings Annotations

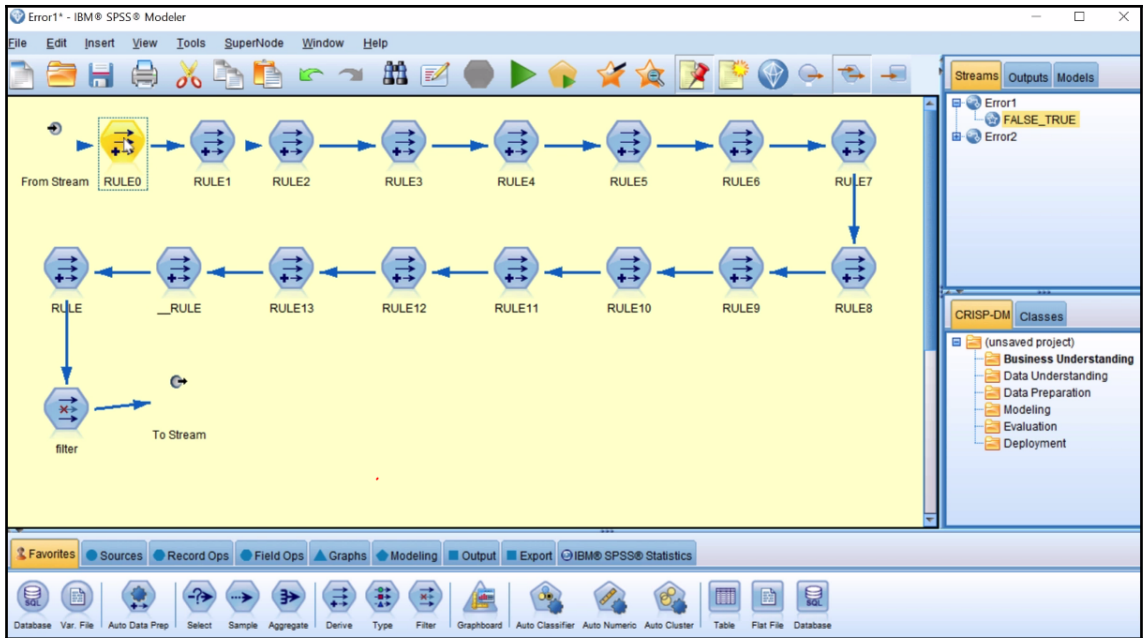
1 2 3 All %

Rules for False - contains 4 rule(s)

- Rule 1 for **False** (22; 0.818)
  - if SEX = M
  - and International <= 0.986
  - and LONGDIST <= 0.104
  - and STATUS = S
  - then **False**
- Rule 2 for **False** (27; 0.667)
  - if SEX = M
  - and International <= 0.986
  - and LONGDIST > 0.104
  - and AGE <= 29
  - and Est\_Income <= 40200.700
  - then **False**
- Rule 3 for **False** (40; 0.7)
  - if SEX = M
  - and International > 0.986
  - and AGE <= 62
  - then **False**
- Rule 4 for **False** (27; 0.741)
  - if SEX = F
  - and LONGDIST > 0.775

OK Cancel Apply Reset







RULE0

Derive as: Conditional

Settings Annotations

Mode:  Single  Multiple

Expression Builder

SEX = "M" and International <= 0.985628 and LONGDIST <= 0.10447 and STATUS = "S"

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

Fields


Type	Field	Storage
	ID	Integer
	LONGDIST	Real
	International	Real
	LOCAL	Real
	DROPPED	Integer
	PAY_MTHD	String
	LocalBillType	String
	LongDistance...	String
	AGE	Integer
	SEX	String

**is\_integer(ITEM)**  
Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel Check Help

Split ×



 Preview ? □ □

**Settings** Annotations

Mode:  Single  Multiple





Reclassify into:  New field  Existing field

Reclassify field:





 RULE 

New field name:

Reclassify values:

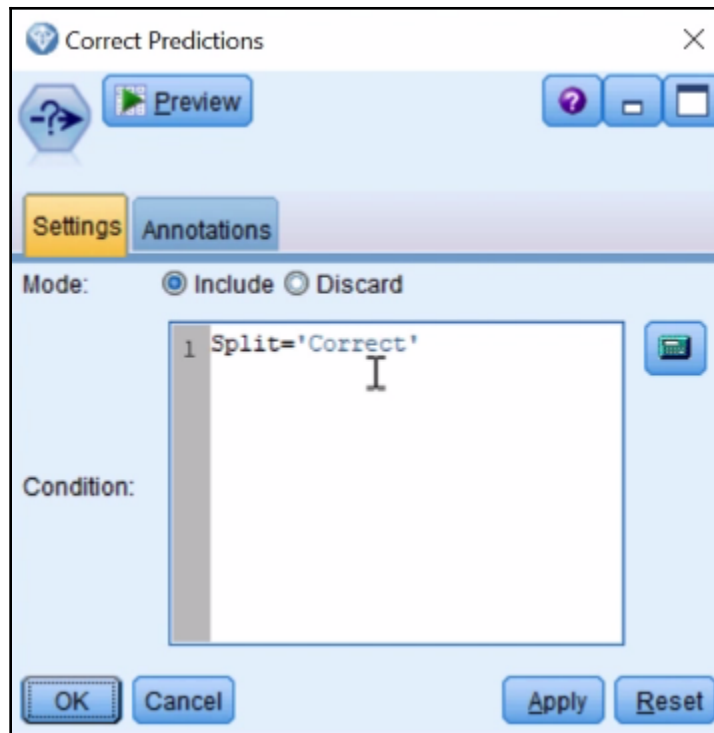
 Get
 Copy
 Clear new
 Auto...

Original value	New value
False_1(0.818)	Incorrect
False_2(0.667)	Incorrect
False_3(0.700)	Incorrect
False_4(0.741)	Incorrect
True_1(0.793)	Correct

For unspecified values use:  Original value  Default value

OK Cancel Apply Reset



Analysis of [LOYAL] #4

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field LOYAL

- Comparing \$N-LOYAL with LOYAL
 

Correct	832	83.87%
Wrong	160	16.13%
Total	992	
- Coincidence Matrix for \$N-LOYAL (rows show actuals)
 

	Leave	Stay
Leave	364	51
Stay	109	468

OK

Incorrect Prediction..

Preview

Settings Annotations

Mode:  Include  Discard

Condition:

```
1 Split='Incorrect'
```

Analysis of [LOYAL] #5

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field LOYAL

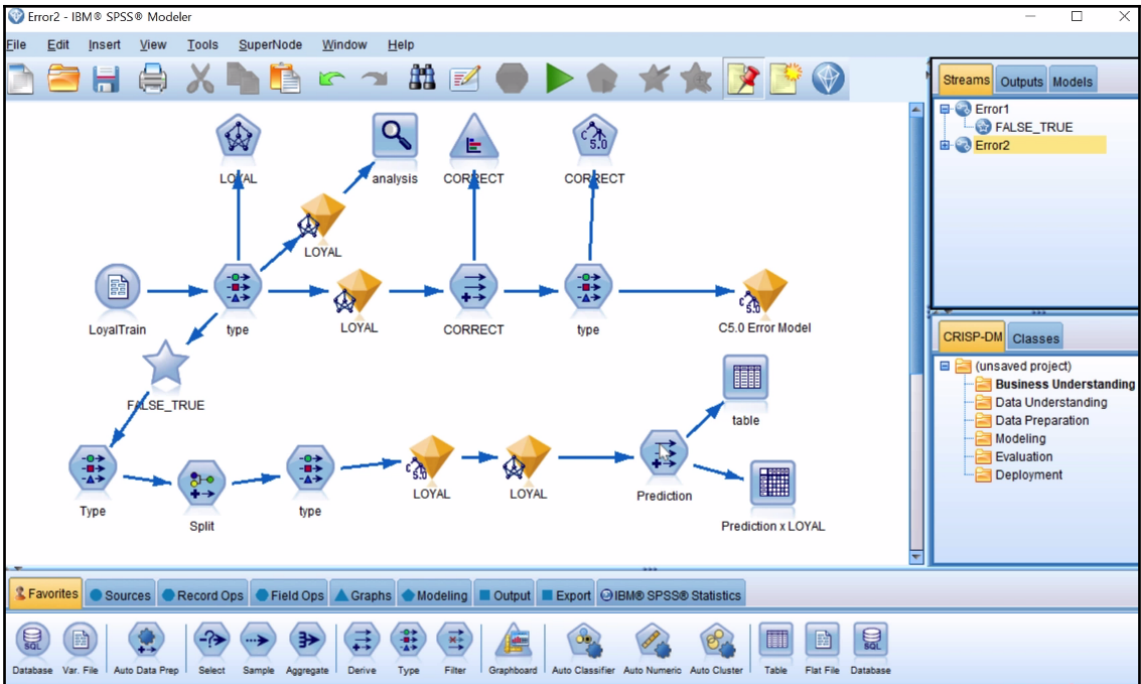
- Comparing \$C-LOYAL with LOYAL

<b>Correct</b>	103	88.79%
<b>Wrong</b>	13	11.21%
<b>Total</b>	116	


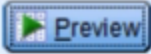



- Coincidence Matrix for \$C-LOYAL (rows show actuals)

	Leave	Stay
Leave	73	4
Stay	9	30

OK



Prediction X


Derive as: Conditional


**Settings** Annotations


Mode:  Single  Multiple


Derive field:

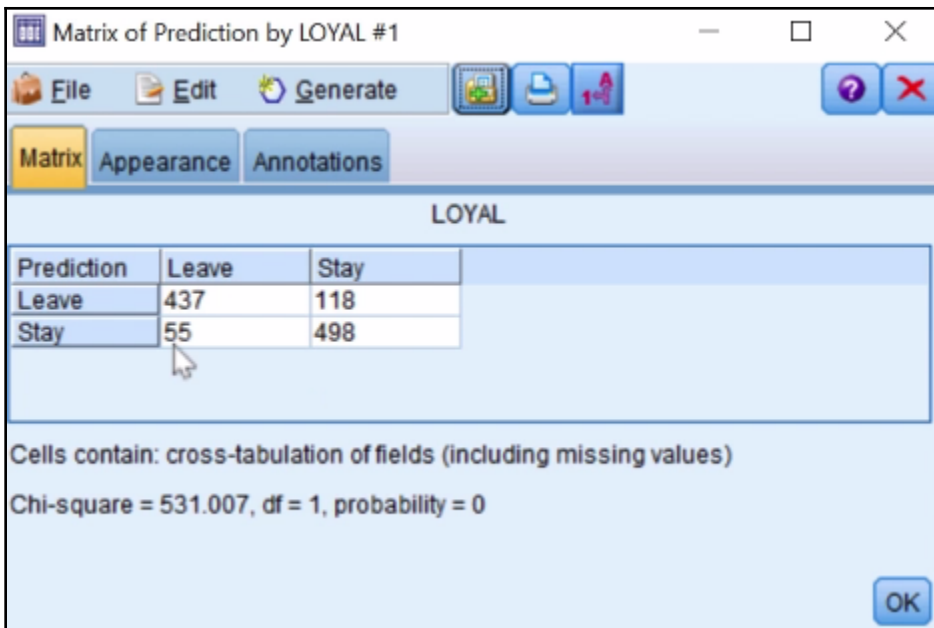
Derive as:

Field type: 

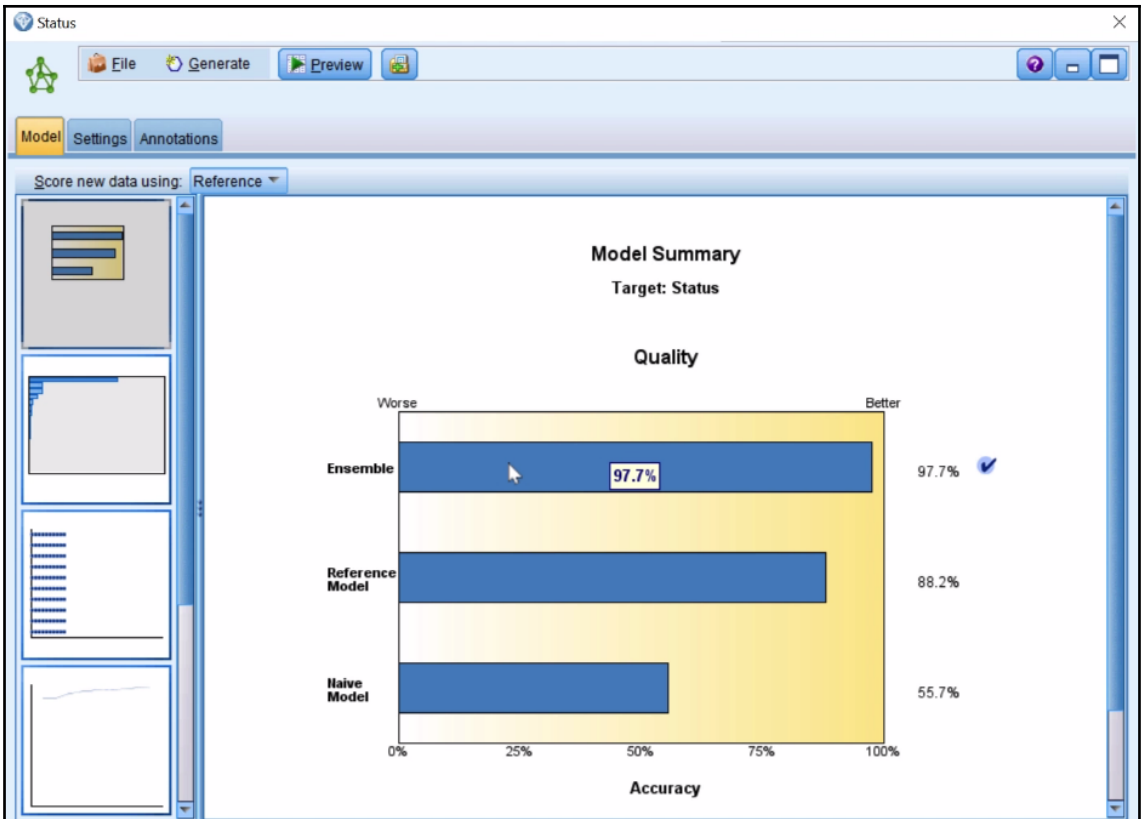
If:  
 

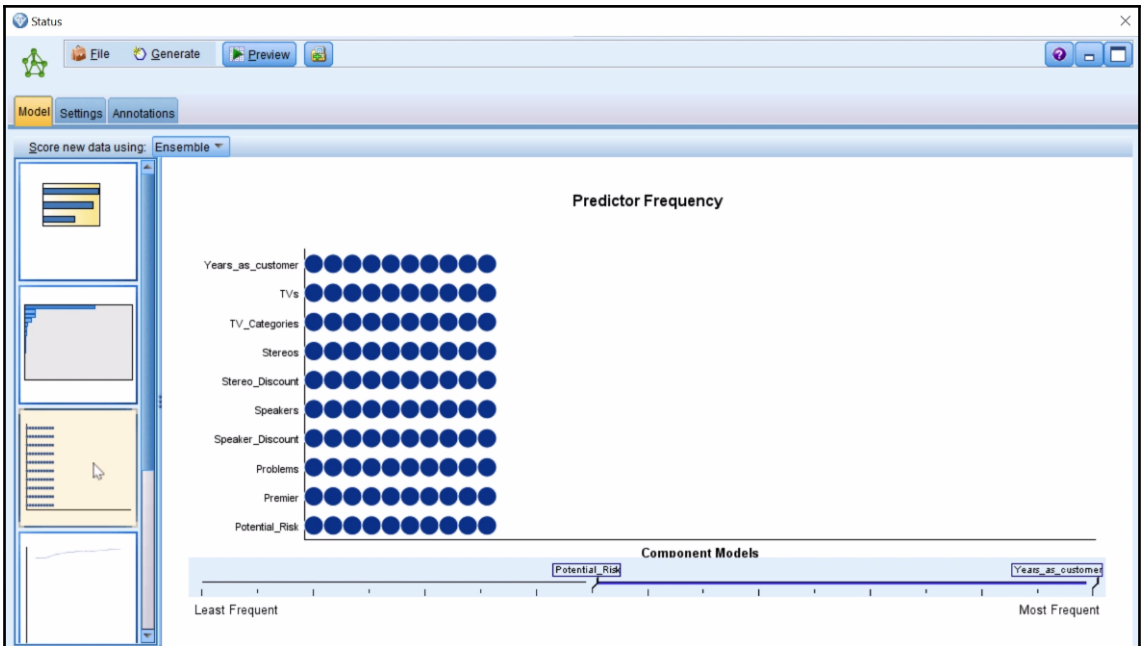
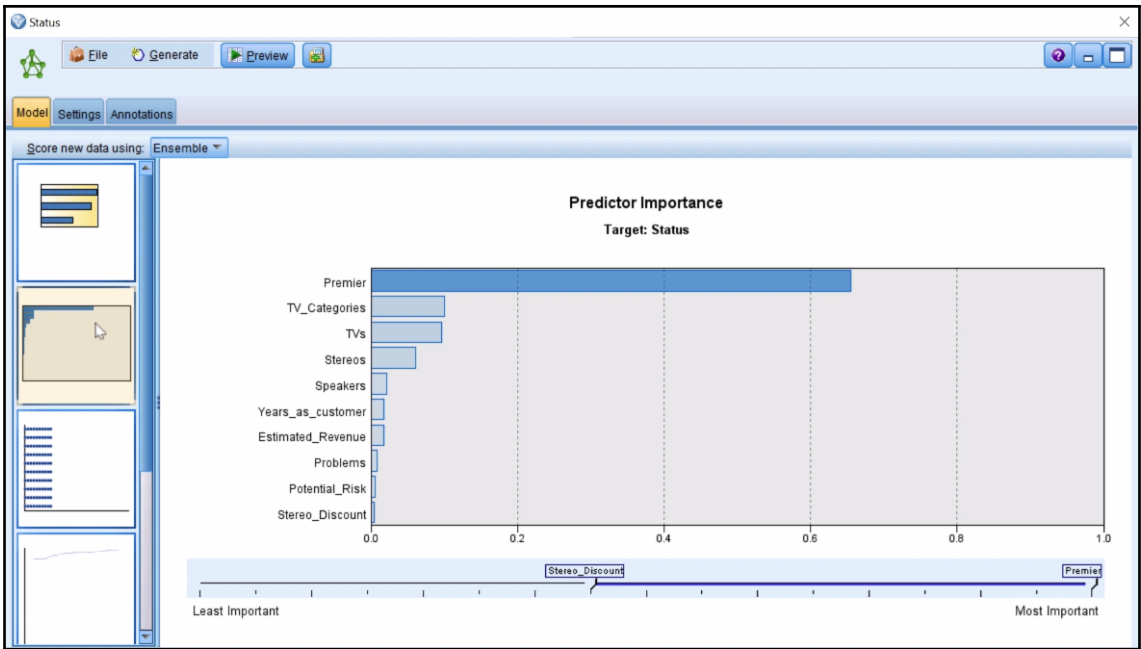
Then:  
 

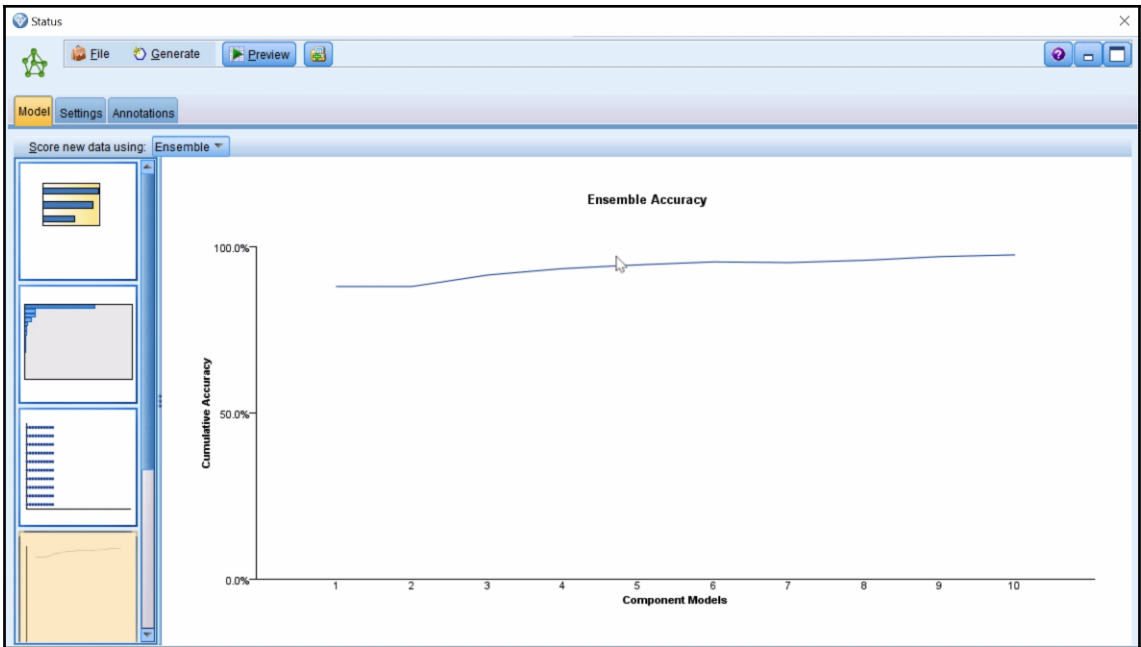
Else:  
 









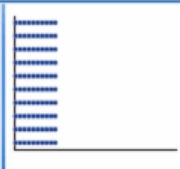
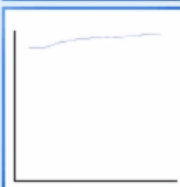


Status ×


File Generate Preview

Model Settings Annotations










Score new data using: Ensemble

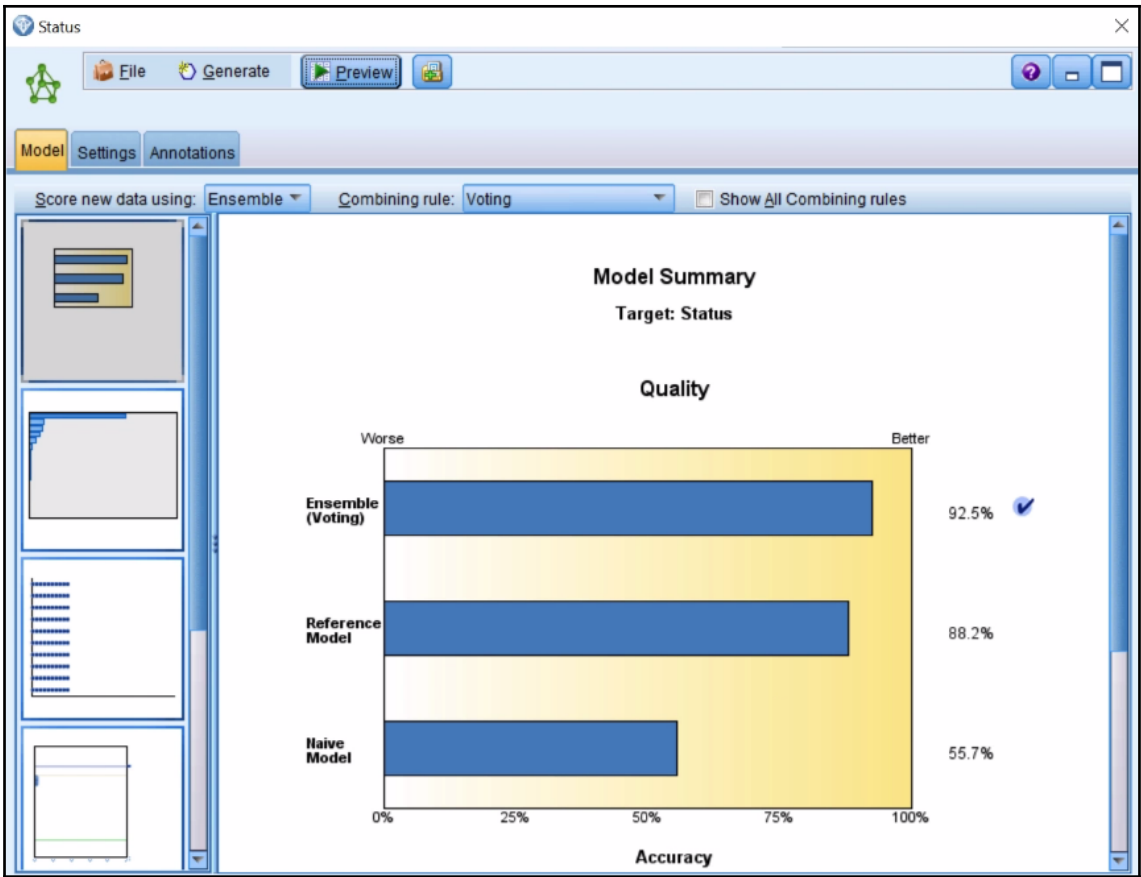



Model	Accuracy	Method	Predictors	Model Size (Synapses)	Records
1	88.2%	Ensemble	17	250	2,456
2	77.5%	Ensemble	17	281	2,456
3	75.1%	Ensemble	17	219	2,456
4	78.4%	Ensemble	17	281	2,456
5	61.8%	Ensemble	17	188	2,456
6	66.9%	Ensemble	17	157	2,456
7	62.5%	Ensemble	17	126	2,456
8	60.1%	Ensemble	17	126	2,456
9	70.8%	Ensemble	17	157	2,456



### Component Model Details

Model	Accuracy	Method	Predictors	Model Size (Synapses)	Records
1	88.2%		17	250	2,456
2	77.5%		17	281	2,456
3	75.1%		17	219	2,456
4	78.4%		17	281	2,456
5	61.8%		17	188	2,456
6	66.9%		17	157	2,456
7	62.5%		17	126	2,456
8	60.1%		17	126	2,456
9	70.8%		17	157	2,456



Analysis of [Status] #6

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field Status

- Comparing \$N-Status with Status
 

'Partition'	1_Training		2_Testing	
Correct	2,273	92.55%	2,147	84.3%
Wrong	183	7.45%	400	15.7%
Total	2,456		2,547	
- Coincidence Matrix for \$N-Status (rows show actuals)
 

'Partition' = 1_Training	Churned	Current
Churned	1,019	70
Current	113	1,254

'Partition' = 2_Testing	Churned	Current
Churned	909	167
Current	233	1,238

OK

Model 1 Prediction	Model 2 Prediction	Model 3 Prediction	Combined Prediction
7	9	8	8
12.5	12.5	12.5	12.5
10	9	8.5	9.167
2	3	10	5

Bank\_Data ×

Preview Refresh
?

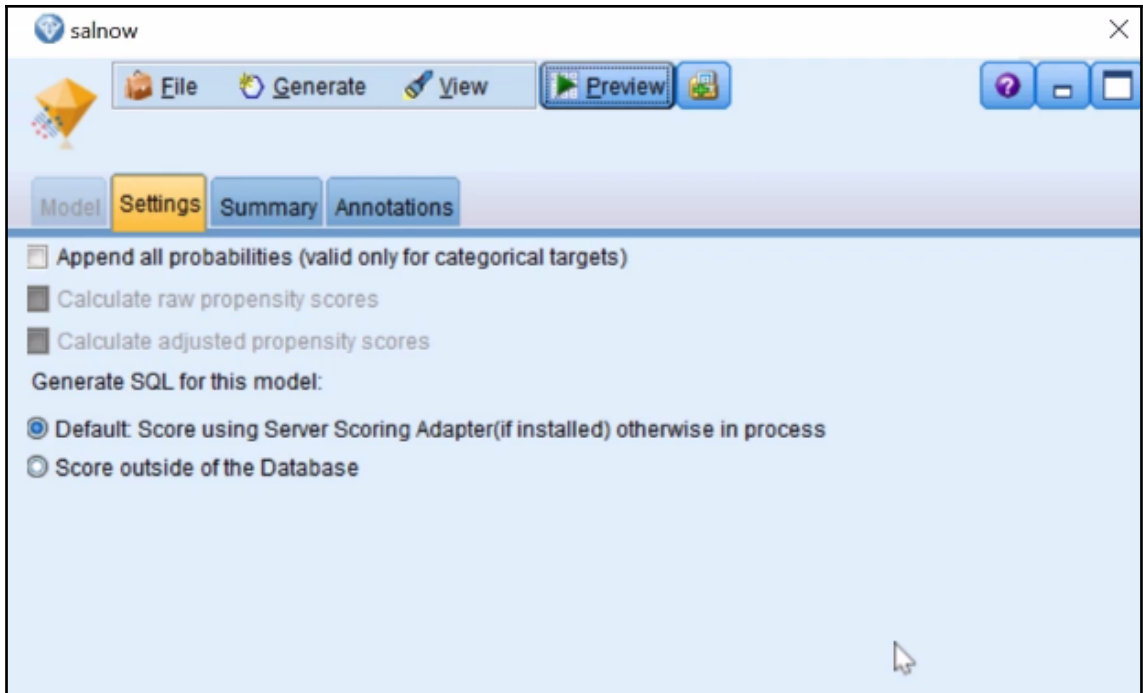
\\MacHome\Desktop\Machine Learning\Bank\_Data

File Data Filter Types Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
salbeg	Continuous	[3600.0,31...		None	Input
sex	Flag	Males/Fe...		None	Input
time	Continuous	[63.0,98.0]		None	Input
age	Continuous	[23.0,64.5]		None	Input
salnow	Continuous	[6300.0,54...		None	Target
edlevel	Continuous	[8.0,21.0]		None	Input
work	Continuous	[0.0,39.67]		None	Input
jobcat	Nominal	Clerical,"C...		None	Input
minority	Flag	White/Non...		None	Input
sexrace	Nominal	"Minority fe...		None	Input

View current fields
  View unused field settings





Analysis of [salnow] #7

File Edit

Analysis Annotations

Collapse All Expand All

Results for output field salnow

- Individual Models
  - Comparing \$N-salnow with salnow

'Partition'	1_Training	2_Testing
Minimum Error	-6585.348	-8455.256
Maximum Error	13967.272	18214.351
Mean Error	-118.414	19.35
Mean Absolute Error	1605.156	2135.244
Standard Deviation	2258.828	3322.512
Linear Correlation	0.931	0.897
Occurrences	238	236
  - Comparing \$S-salnow with salnow

'Partition'	1_Training	2_Testing
Minimum Error	-4047.638	-4890.045
Maximum Error	21305.415	31883.945
Mean Error	917.323	944.225
Mean Absolute Error	2374.65	2885.95
Standard Deviation	3858.52	5000.19
Linear Correlation	0.835	0.806
Occurrences	238	236
- Agreement between \$N-salnow \$S-salnow
  - Comparing Agreement with salnow

OK

Analysis of [salnow] #8

File Edit

Analysis Annotations

Collapse All Expand All

Standard Deviation	3858.52	5000.19
Linear Correlation	0.835	0.806
Occurrences	238	236

Comparing \$XR-salnow with salnow

'Partition'	1_Training	2_Testing
Minimum Error	-4179.669	-4929.77
Maximum Error	17210.738	25049.148
Mean Error	399.455	481.788
Mean Absolute Error	1796.208	2268.59
Standard Deviation	2761.154	3918.25
Linear Correlation	0.919	0.883
Occurrences	238	236

Agreement between \$N-salnow \$\$-salnow \$XR-salnow

Comparing Agreement with salnow

'Partition'	1_Training	2_Testing
Minimum Error	-4179.669	-4929.77
Maximum Error	17210.738	25049.148
Mean Error	399.455	481.788
Mean Absolute Error	1796.208	2268.59
Standard Deviation	2761.154	3918.25
Linear Correlation	0.919	0.883
Occurrences	238	236

OK