



# Introduction to Next Generation Business Intelligence with Automated Analytics Mode in SAP Predictive Analytics Case Study

**Product**

SAP Predictive Analytics

**Level**

Undergraduate  
Graduate  
Intermediate

**Focus**

Business Analytics

**Authors**

Michael Bliemel,  
Rowe School of Business,  
Dalhousie University

**Version**

1.1 for UWM Workshop

**Motivation**

This case study provides an introduction into four different use cases of Automated Analytics in SAP Predictive Analytics to demonstrate how business users can develop predictive models that can be published to target specific customers, assess risks or predict what products are of interest to which customers.

**Prerequisites**

A basic understanding of predictive analytics is recommended. Students should be familiar with classification, correlation and regression concepts.

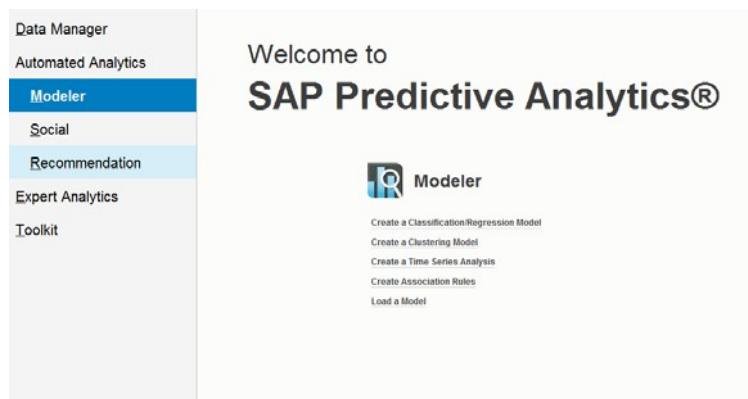
To complete these exercises you will require SAP Predictive Analytics and the accompanying data files which should be installed in the \07 Predictive Analytics Automated folder.

*Updated by Nancy Jones  
and Hossam Ali-Hassan for  
the SAP UA Analytics  
Summer Workshop 2016 @  
UWM.*

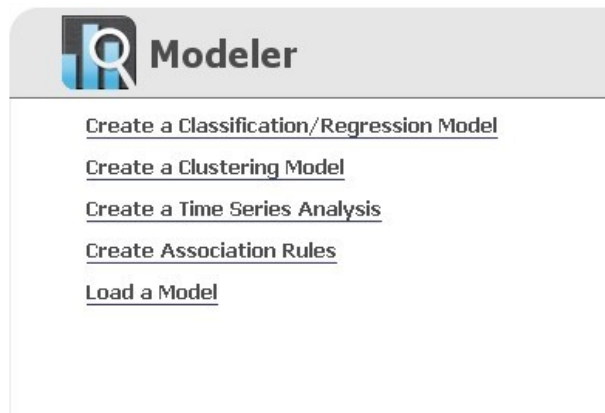
## Exercise 1: Auto Insurance Risk Analysis with SAP Predictive Analytics

In this exercise, an analyst of an Insurance company notices the number of accident claims has increased. He decides that it would be better to analyze the key factors that lead to claim and use them to evaluate the risk of claim for new contracts in the future. The analyst uses SAP Predictive Analytics to statistically analyze what factors lead to a claim. Using the pattern developed using past data, she is able to predict the risk of claim for new contracts considering the profile of the subscriber and the car insured.

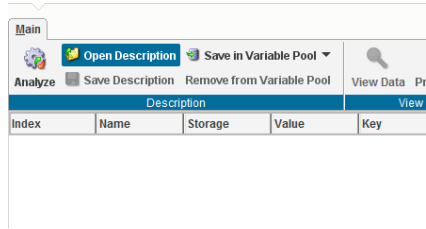
Start SAP Predictive Analytics by following the menu **Start > All Programs > SAP Business Intelligence > SAP Predictive Analytics Desktop > SAP Predictive Analytics**



Click on the option **Modeler**.



Field Label	Value	Description
		Click on the option <b>Create a Classification / Regression Model</b> .
Data Type	Text files	

Field Label	Value	Description
Folder	.\07 Predictive Analytics Automated	Click <b>Browse</b> for the <b>Data Set</b> option.
		Select the file <b>AutoInsuranceRisk_training.csv</b> .
		Click <b>OK</b> .
		Click <b>Next</b> .
		Click <b>Open Description</b> . <b>Data Description</b>
		
		Click <b>Browse</b> for the <b>Description</b> .
		Select the file <b>AutoInsuranceRisk_training_desc.csv</b> .
		Click <b>OK</b> .
		Click <b>OK</b> .

Main

Edition

Structures

Open Description

Save in Variable Pool

Save Description

Remove from Variable Pool

Description

View

Description: AutoInsuranceRisk\_training\_desc.csv

Index	Name	Storage	Value	Key	Order	Missing	Group	Description	Structure
1	ContractId	integer	continuous	1	0				
2	Age	integer	continuous	0	0				
3	Gender	string	nominal	0	0				
4	Children	string	nominal	0	0				
5	Profession	string	nominal	0	0				
6	Customer Type	string	nominal	0	0				
7	Multiple cars	string	nominal	0	0				
8	Driving Licenc...	integer	continuous	0	0				
9	Car category	string	nominal	0	0				
10	Annual Kilome...	integer	continuous	0	0				
11	Gearbox	string	nominal	0	0				
12	Fuel	string	nominal	0	0				
13	Claim	string	nominal	0	0				

Click **View Data** to see the actual records.

Take some time to understand what data you have available by looking at the actual records. It is important to understand the business problem and the data to make the right decisions based on using these tools.

Field Label	Value	Description
		Click <b>Close</b> to close the sample data.
		Click <b>Next</b> .

### Selecting Variables

The screenshot shows the 'Selecting Variables' interface in SAP Predictive Analytics. It features three main panels:

- Explanatory Variables Selected (11):** A list of variables including Age, Gender, Children, Profession, Customer Type, Multiple cars, Driving Licence Years, Car category, Annual Kilometers, Gearbox, and Fuel. Navigation arrows (right and left) are present between this list and the others.
- Target Variables (1):** A list containing the variable 'Claim'.
- Excluded Variables (1):** A list containing the variable 'Contractid'.

Additional features include a 'Weight Variable' field set to 0, and checkboxes for 'Alphabetic Sort' at the bottom of each variable list.

The model identifies a pattern in how one, few or all of the explanatory variables lead to a claim.

Note: The target variable is the phenomena we would like to explain, in this case the fact that the contract led to an accident claim. The explanatory variables are the potential variables that could explain the phenomena. Here we think that the driver age, its gender, the car category, etc. can potentially have an impact on the risk of accident. SAP Predictive Analytics is going to identify the most significant variables that contribute to explain the risk of accident. Note that the Contract Id will be excluded, as it may not have any impact on the risk of accident.

Check that the column **Contract ID** is in the list of the **Excluded Variables**. Click **Next**.

### Summary of Modeling Parameters

Model Name: Claim\_AutoInsuranceRisk\_training

Description:

**Kxen.RobustRegression**

Data to be Modeled: ExerciseData/AutoInsuranceRisk\_training

Cutting Strategy: Random without test

Compute Decision Tree: ☒

Enable Auto-selection: ☒

Autosave... Export KxShell Script... Advanced...

Cancel Previous Generate

Activate the option **Compute Decision Tree**. Click **Generate**.

Report Type: Model Overview

### Overview

Model: Claim_AutoInsuranceRisk_training	
Data Set:	AutoInsuranceRisk_training.csv
Initial Number of Variables:	14
Number of Selected Variables:	11
Number of Records:	8,221
Building Date:	2016-04-12 15:45:23
Learning Time:	1 s
Engine Name:	Kxen.RobustRegression
Author:	Nancy

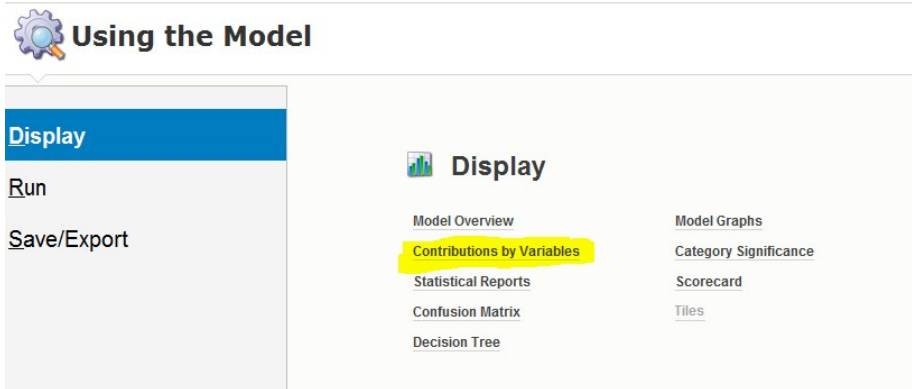
### Nominal Targets

Claim	
Target Key	Yes
No - Frequency	90.99%
Yes - Frequency	9.01%

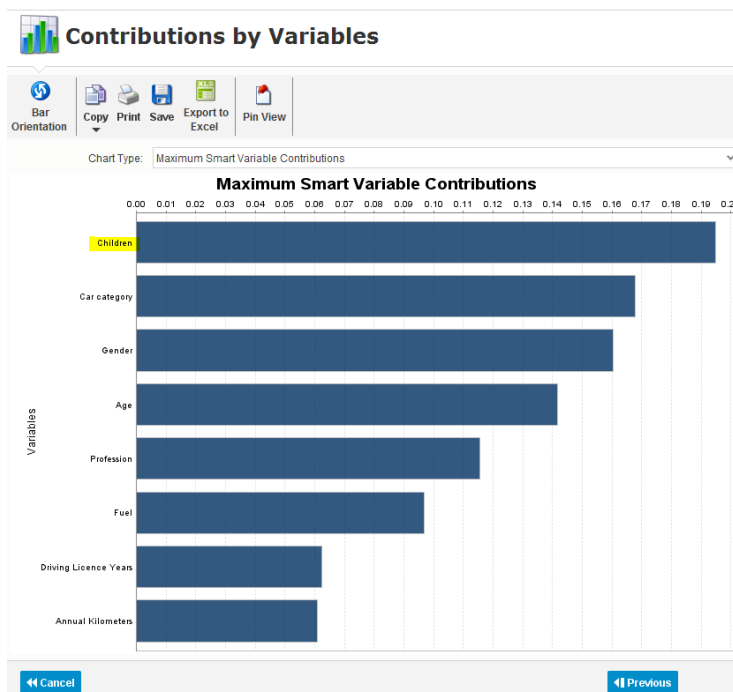
### Selection Process Selected Iteration

1	
Predictive Power (KI)	0.4784
Prediction Confidence (KR)	0.9768
Nb. Variables Kept	7

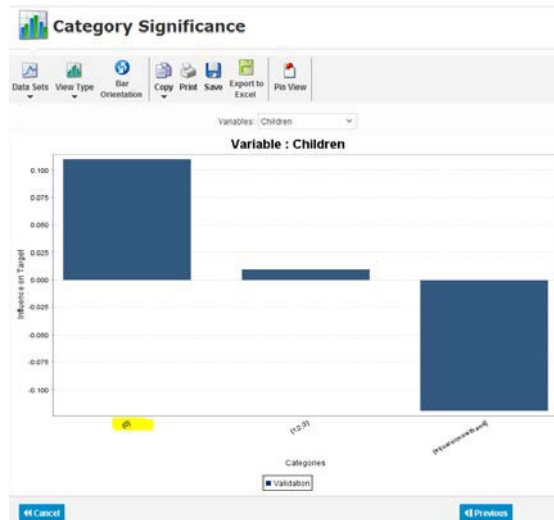
At the bottom you can see that SAP Predictive Analytics found 7 variables that are influencing the claims. Click **Next**.



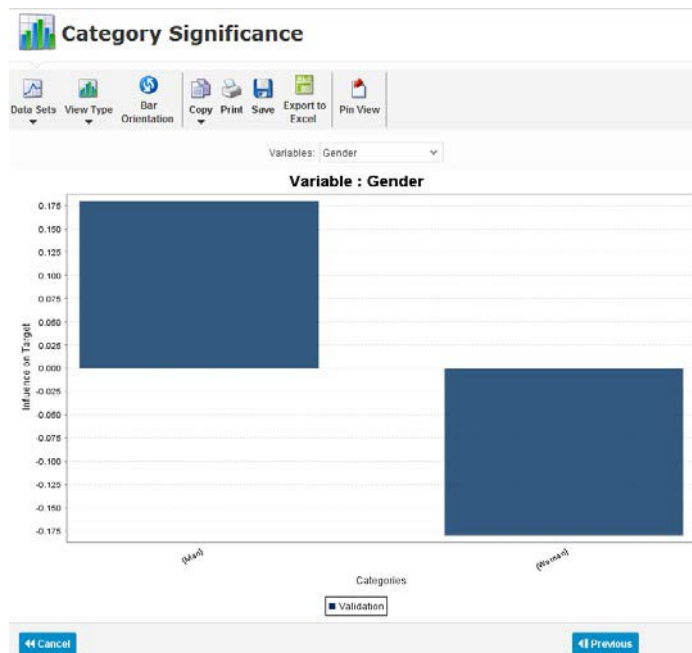
Click on **Contributions by Variables**. – You can change the direction of the Bars by clicking on the Bar **Orientation** Button on the top left.



As you can see the number of children is the most important factor. Double-click on the bar for **Children**.



You can see now that customers without children have a much higher propensity to have an accident compared to customers with 4 or more children. Another way to see this is that positive numbers for Influence on Target means that having no children has a greater likelihood of accidents, while the negative numbers for Influence on Target means that having 4 or more children has a smaller likelihood of claiming an automobile accident. Click on **Previous**. Double-click on the bar for **Gender**.

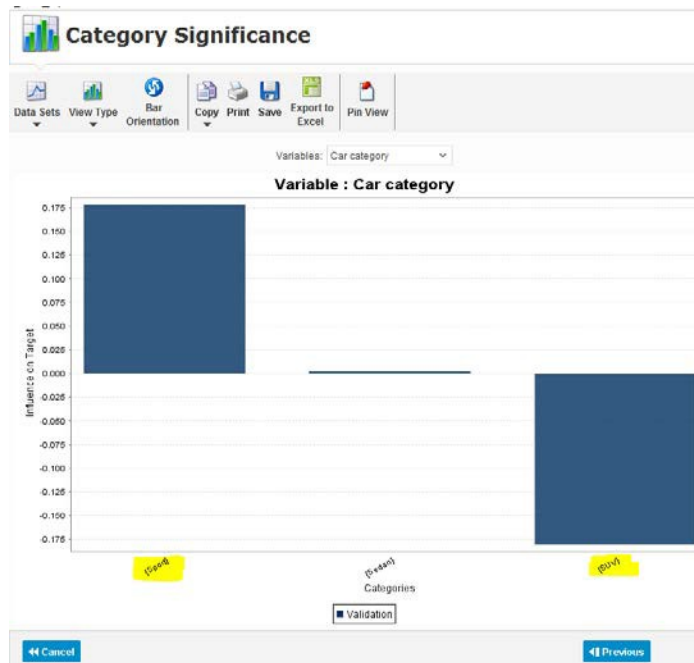


You see that Gender makes a difference.

Q1: Is a Man or Woman likely to have an accident?

1 mark

Field Label	Value	Description
		Click on <b>Previous</b> .
		Double-click on the bar for <b>Car Category</b> .



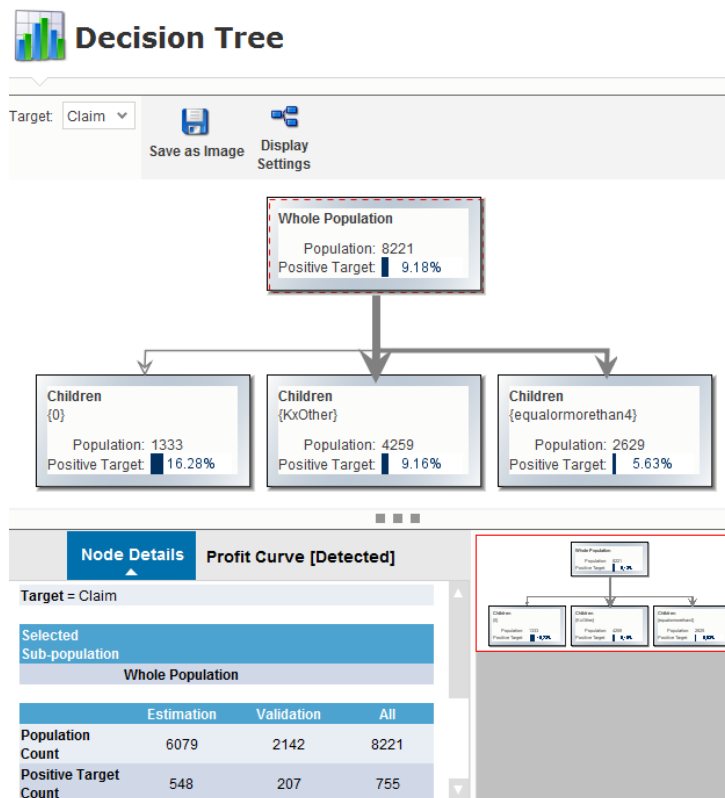
So a man in a sports car is a much higher risk than a woman in a SUV.

Field Label	Value	Description
		Click <b>Previous</b> .
		Click <b>Previous</b> .

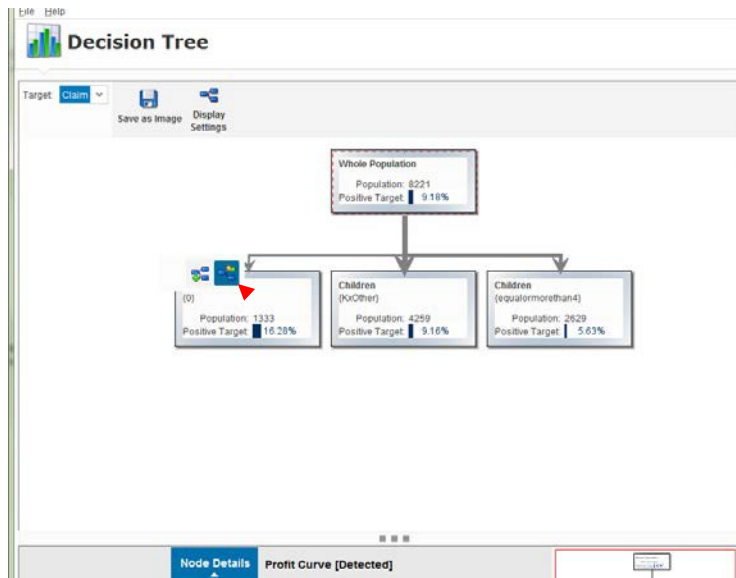
The figure shows the 'Using the Model' interface. On the left is a sidebar with a gear icon and a list of options: 'Display' (highlighted in blue), 'Run', and 'Save/Export'. The main area is titled 'Display' and contains a list of model components: 'Model Overview', 'Contributions by Variables', 'Statistical Reports', 'Confusion Matrix', 'Decision Tree' (highlighted in yellow), 'Model Graphs', 'Category Significance', 'Scorecard', and 'Tiles'.



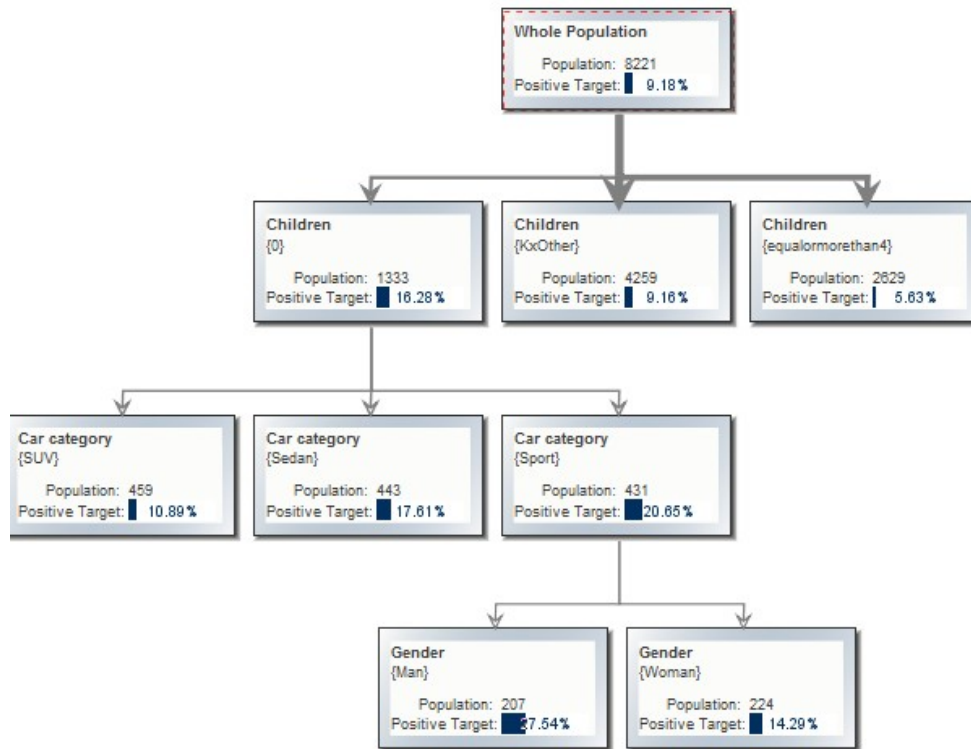
Click on **Decision Tree**.



The decision tree accompanied the model shows that comparing to overall claim rate of 9.18%, the customers who don't have children had a claim rate of 16.28%. On the contrary, those who have 4 children and more only had a claim rate of 5.63%.



Move the mouse cursor on top of the node of 0 **Children** and **expand the node**. Repeat this and expand the node for the **Car Category Sport**. (you may have to play about with one of the other boxes to view this)



So we can see that Man without Children driving Sport cars have a claim rate of 27.54% - a very high risk group.

Q2. What is the claim rate for women without children driving sports cars? 1 - Mark

Below the node, you can examine the details for Males without Children in Sport cars.

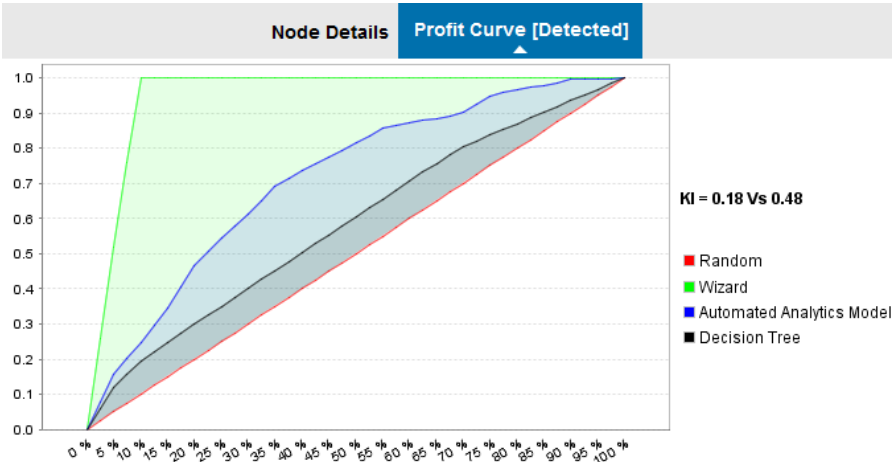
Node Details		Profit Curve [Detected]	
Target = Claim			
Selected Sub-population			
	Children is in {0}		
AND	Car category is in {Sport}		
AND	Gender is in {Man}		
	Estimation	Validation	All
Population Count	146	61	207
Positive Target Count	41	16	57
Positive Target Ratio	28.08%	26.23%	27.54%
Negative Target Count	105	45	150
Negative Target Ratio	71.92%	73.77%	72.46%
Variance	0	0	
Weighted Population	146.0	61.0	

Here you can see how accurate the model is at predicting claims based on a cutting strategy of estimating using 146 drivers and validating with 61 drivers.

Click on **Profit Curve [Detected]** to view the ROC (Receiver operating characteristic Curve) for the Predictive Analytics model compared to the Decision Tree and a Random selection.

The Red line would be the performance if we used no model.

The Green line would be the model performance if we created a theoretically perfect model. The Blue line is how well our current model performed.



Field Label	Value	Description
		Click <b>Previous</b> .
		Select the option <b>Run</b> .

**Using the Model**

Display

**Run**

Save/Export

**Run**

Analyze Deviations

**Simulation**

Apply Model

Select Variables

Select the option **Simulation**. Enter the following values:

Field Label	Value	Description
Children	0	
Car Category	SUV	
Gender	Man	
Age	40	



## Simulating the Model

Explanatory Variables
  
Sort by: Contribution of Claim

Names	Values	
Children	0	
Car category	SUV	
Gender	Man	
Age	40	
Profession		
Fuel		
Annual Kilometers		

Reset Run
  
  
Variable: Age  
Min: 18  
Max: 95

Click **Run**.

Results	
Output	Value
Score of Claim	0.4172
Proba. of (Claim = Yes)	0.2082

The probability for a claim of our 40 year old man driving a SUV without any children is 20.82%.

Note: this means the probability that this driver belongs to the group “Claim=Yes” is 20.82%, which is a subtle difference from saying that this driver has a 20.82% chance of having an accident. For the purposes of pricing Insurance premiums this subtlety is not a problem because one deals with aggregated customer groups who collectively will have 20.82% of their members having claims, which allows insurers to calculate the cost of insuring those customers.

Q. 3 Enter the following and Run:

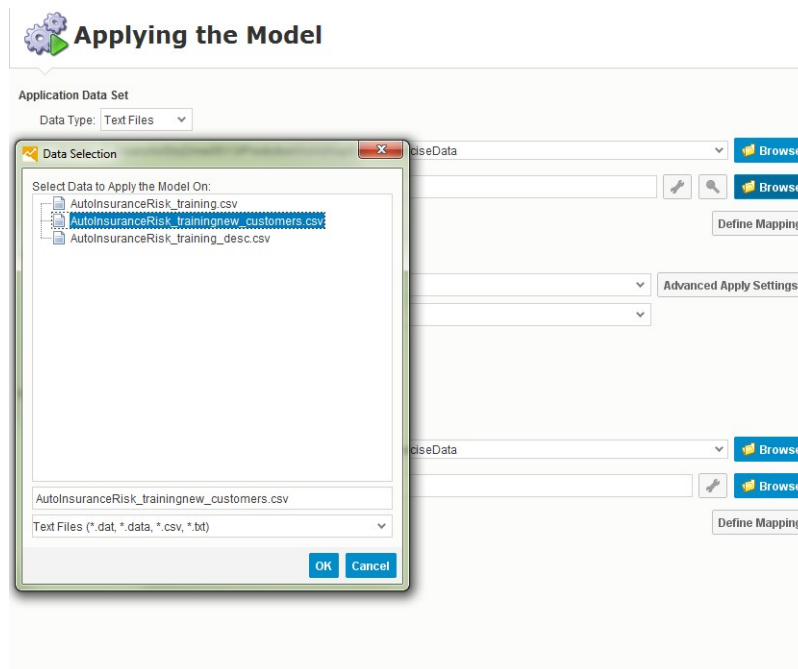
What is the probability for a claim in this instance? 1 - Mark

Field Label	Value	Description
Children	4	
Car Category	SUV	
Gender	Man	
Age	55	

Field Label	Value	Description
		Click <b>Previous</b> .
		Click <b>Apply Model</b> .

Now we are going to load a set of new customers and apply the probabilistic model to predict their risk. For this we load a data file with no claim information, apply the model and then output a new data file with the calculated risk scores. In a real-time business, one would have the model applied as new customer information is entered to generate an instant risk score instead of the process we are doing here.

Field Label	Value	Description
		Click <b>Previous</b> .
		Click <b>Apply Model</b> .



Click on **Browse** for the **Data** and select the file **AutoInsuranceRisk\_trainingnew\_customers.csv**.

Field Label	Value	Description
		Click <b>OK</b> .
		Click on <b>Advanced Apply Settings</b> .



## Model Advanced Apply Settings

Advanced Apply Settings

General Outputs

Gain Chart

Outputs for Target 'Claim'

☐ Copy Weight Variable
   
☒ Copy Variables
   
☒ All
   
☐ Individual

User Defined Constant Outputs

Visibility	Name	Storage	Value	Key
<input type="checkbox"/>	Model Name	string	Claim_Autoinsurance...	0
<input type="checkbox"/>	Build Date	datetime	2015-02-15 11:48:23	0
<input type="checkbox"/>	Apply Date	datetime	2015-02-15 12:22:16	0
<input type="checkbox"/>	Model Version	integer	1	0

Add

Delete

Cancel

Previous

OK

Field Label	Value	Description
		Activate the option <b>Copy Variables</b> .
		Activate the option <b>All</b> .
		Click <b>OK</b> .
		Click on <b>Browse</b> for the <b>Folder</b> in the Results Generated by the Model area.
		Check that the folder <b>/07 Predictive Analytics Automated</b> is set as location to store the file.
		Enter <b>YourNameExercise1.txt</b> for the field <b>Data</b> .
		Click <b>Apply</b> .

Be careful when entering Directory and filenames

## Applying the Model

Stop View Type Copy Print Save

✓ Your model was applied successfully

Claim\_AutoInsuranceRisk\_training  
 ExerciseData/AutoInsuranceRisk\_trainingnew\_customers.csv  
 (Kxen.FileStore)  
 ExerciseData/MichaelExercise1.txt (Kxen.FileStore)

View Output

Cancel Previous Next

Click **View Output**.

## Applying the Model

Stop View Type Copy Print Save

Output Name: JonesExercise1.txt

First Row Index: 1 Last Row Index: 100

Data Statistics Graph

	Gender	Children	Profession	Customer ...	Multiple c...	Driving Lic...	Car categ...	Annual Kil...	Gearbox	Fuel	Claim	rr_Claim
1	55 Woman	4+	Retired	Agency	No	14	Saloon	24074	Automatic	Diesel		0.0531151...
2	58 Man	1	Unemployed	Agency	No	13	Sport	19328	Manual	Diesel		0.3430313...
3	26 Woman	4+	Independant	Agency	No	3	Sport	21141	Automatic	Diesel		0.2650568...
4	47 Woman	1	Unemployed	Agency	No	22	Estate	41040	Manual	Petrol		-0.0835518...
5	54 Woman	2	Private Sec...	On-line	No	18	Estate	31620	Manual	Diesel		0.0510711...
6	78 Man	4+	Public Sect...	Agency	No	18	Estate	17837	Manual	Diesel		0.4682320...
7	39 Man	2	Independant	On-line	No	21	Saloon	20065	Automatic	Diesel		0.2138477...
8	43 Woman	2	Retired	Agency	No	11	Saloon	35688	Automatic	Diesel		0.0139257...
9	51 Woman	0	Private Sec...	Agency	Yes	20	Sport	24503	Manual	Diesel		0.3009015...
10	42 Woman	0	Public Sect...	Agency	Yes	18	Estate	24053	Manual	Diesel		0.2551527...
11	52 Woman	1	Student	Agency	Yes	18	Estate	27859	Automatic	Diesel		0.0495802...
12	38 Man	3	Public Sect...	Agency	No	20	Sport	21000	Automatic	Diesel		0.3242916...
13	20 Woman	2	Independant	Agency	No	2	Saloon	19898	Automatic	Diesel		0.0821736...
14	36 Man	0	Independant	Agency	No	18	Sport	22955	Manual	Diesel		0.5632096...
15	46 Man	2	Independant	On-line	No	7	Saloon	18803	Manual	Petrol		0.0538806...
16	43 Man	0	Private Sec...	Agency	Yes	13	Saloon	21278	Automatic	Diesel		0.2184297...

If you look at the rightmost column, you can see the prediction that was added by applying the model to the dataset. The first row is a 55 year old Woman in a Saloon (=Sedan in American English) with 4+ Children, while the second one is a 58 year old Man with one Child in a Sport car, who is clearly a higher insurance risk!

**Q4. Which Contract id has the highest Insurance Risk? (1 Mark)**

## Exercise 2: Long-Term Customer Relationships with SAP Predictive Analytics

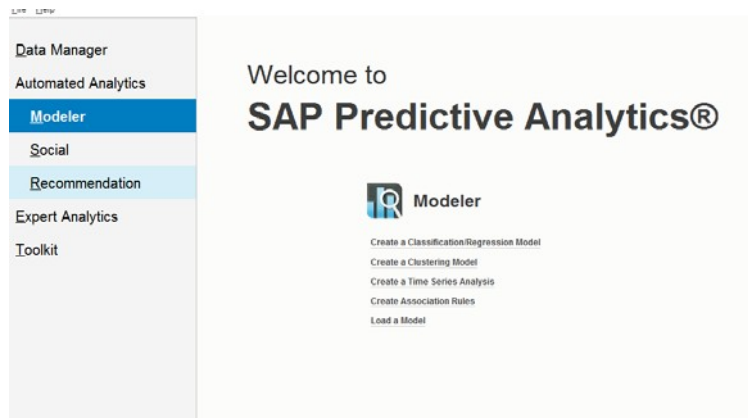
In an ever-competitive market, banks face challenges to keep customers, especially when they are highly profitable. One of many KPIs regularly monitored in the retail bank is customer attrition.

**When a highly profitable customer leaves, it hurts the bottom line.** Giving near real-time visibility to attrition helps executives to develop proactive strategies.

The head of sales for retail banking sees that the **attrition numbers are rising**. He is made aware of high-value customers at risk of attrition through his dashboard and then can get a 360-degree view of the customer, including suggested **offers to entice the customers to stay**. A comprehensive customer view means finding insights that only come from a complete picture.

Mary, an analyst at BestRun Bank, has been tasked to find the factors that contributed to customer leave.

Start SAP Predictive Analytics by following the menu **Start > All Programs > SAP Business Intelligence > SAP Predictive Analytics Desktop > SAP Predictive Analytics**



Click on the option **Modeler**, then **Create a Classification / Regression Model**. Ensure the option **Data Type** is set to **Text Files**. Enter the following to the option Folder: **\\07 Predictive Analytics Automated**.

## Select a Data Source

☒ Use a File or a Database Table
 ☐ Use Explorer

Data Type: **Text Files**

Folder:  **Browse**

Data Set:  **Browse**

**Cutting Strategy...**

**Metadata** Metadata are stored in the same place as data source.

**Cancel** **Previous** **Next**

Field Label	Value	Description
		Click <b>Browse</b> for the <b>Data Set</b> option.
		Select the file <b>Customer_Loyalty.csv</b> .
		Click <b>OK</b> .
		Click <b>Next</b> .
		Click <b>Analyze</b> .

Here the data is automatically described. We could edit and save these definitions if we wanted to here. Have a look to understand what values we will use in the model.

## Data Description

**Main** | Edition | Structures

**Analyze** | Open Description | Save in Variable Pool | View Data | Properties

Save Description | Remove from Variable Pool

Description | View

GuesSED Description

Index	Name	Storage	Value	Key	Order	Missing	Group	Description	Structure
1	Id	integer	continuous	0	0				
2	Gender	string	nominal	0	0				
3	Salary	number	continuous	0	0				
4	Marital	string	nominal	0	0				
5	CheckingAccount	string	nominal	0	0				
6	NumberOfSaving...	integer	ordinal	0	0				
7	PersonalAccount	string	nominal	0	0				
8	Mortgage	string	nominal	0	0				
9	PensionPlan	string	nominal	0	0				
10	Region	string	nominal	0	0				
11	Status	integer	nominal	0	0				
12	KIndex	integer	continuous	1	0			Automatically add...	

Click **View Data** to examine and understand what we are working with.

First Row Index:  Last Row Index:  

Data Statistics Graph												
	id	Gender	Salary	Marital	Checking...	NumberOf...	PersonalA...	Mortgage	PensionPl...	Region	Status	KxIndex
1	13101	Female	38962.599...	Single	No Checki...	0	Personal A...	Mortgage	No Pensio...	Land	0	1
2	13102	Male	39127.599...	Single	No Checki...	3	No Person...	Mortgage	Pension Pl...	Land	1	2
3	13103	Male	77196.800...	Married	No Checki...	0	Personal A...	No Mortgage	Pension Pl...	City	1	3
4	13104	Male	41508.599...	Single	No Checki...	0	No Person...	Mortgage	No Pensio...	City	0	4
5	13105	Female	27729.200...	Married	Checking A...	3	Personal A...	No Mortgage	No Pensio...	City	0	5
6	13106	Male	73198	Married	Checking A...	3	Personal A...	No Mortgage	Pension Pl...	Suburban	0	6
7	13107	Male	91712.199...	Single	Checking A...	1	Personal A...	No Mortgage	No Pensio...	Suburban	0	7
8	13108	Male	44724.599...	Single	Checking A...	0	Personal A...	No Mortgage	No Pensio...	Suburban	0	8
9	13109	Female	43968	Married	No Checki...	1	Personal A...	No Mortgage	Pension Pl...	Suburban	0	9
10	13110	Male	22146	Married	Checking A...	2	No Person...	No Mortgage	No Pensio...	Suburban	1	10

Field Label	Value	Description
		Click <b>Close</b> .
		Click <b>Next</b> .
Explanatory Variables Selected		Move the column <b>ID</b> to the <b>Excluded Variables</b> .
		Ensure the column <b>Status</b> is set in <b>Target Variables</b> .

Explanatory Variables Selected 9

Gender  
Salary  
Marital  
CheckingAccount  
NumberOfSavingsAccounts  
PersonalAccount  
Mortgage  
PensionPlan  
Region

☐ Alphabetic Sort

Target Variables 1

Status

☐ Alphabetic Sort

Weight Variable 0

Excluded Variables 2

KoIndex  
id

☐ Alphabetic Sort

Cancel

Previous

Next

© 2015 SAP SE or an SAP affiliate company. All rights reserved.

SAP Predictive Analytics is going to identify the most significant variables that contribute to explaining the leave. Note that the Id is excluded, as it may not have any impact on the risk of leave. In this dataset, positive target corresponds with records that have '0' value, which is customers who stay.

Field Label	Value	Description
		Click <b>Next</b> .

### Summary of Modeling Parameters

Model Name: Status\_Customer\_Loyalty

Description:

#### Kxen.RobustRegression

Data to be Modeled: ExerciseData/Customer\_Loyalty.csv

Cutting Strategy: Random without test

Target Variable: Status

Weight Variable (Optional): None

Compute Decision Tree: ☒

Enable Auto-selection: ☒

Autosave...
Export KxShell Script...
Advanced...

Cancel
Previous
Generate

Field Label	Value	Description
Compute Decision Tree		Set checkbox.
		Click <b>Generate</b> .
		Click <b>Next</b> .

Warning (New)
4/12/16 4:41:24 PM

Found significant deviation on data set Validation: model will be suspicious. The following variables have deviations: Gender. Consider changing the cutting strategy or excluding these variables from the model.

Close the error message and click Next.

## Using the Model

**Display**  
Run  
Save/Export

**Display**  

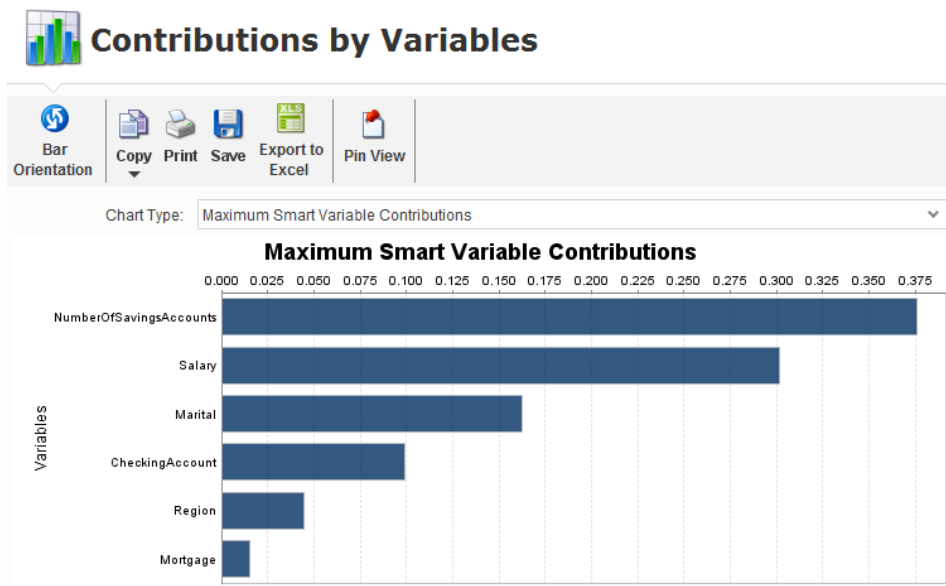
Model Overview  
Contributions by Variables  
Statistical Reports  
Confusion Matrix  
Decision Tree

Model Graphs  
Category Significance  
Scorecard  
Tiles

Cancel
Previous

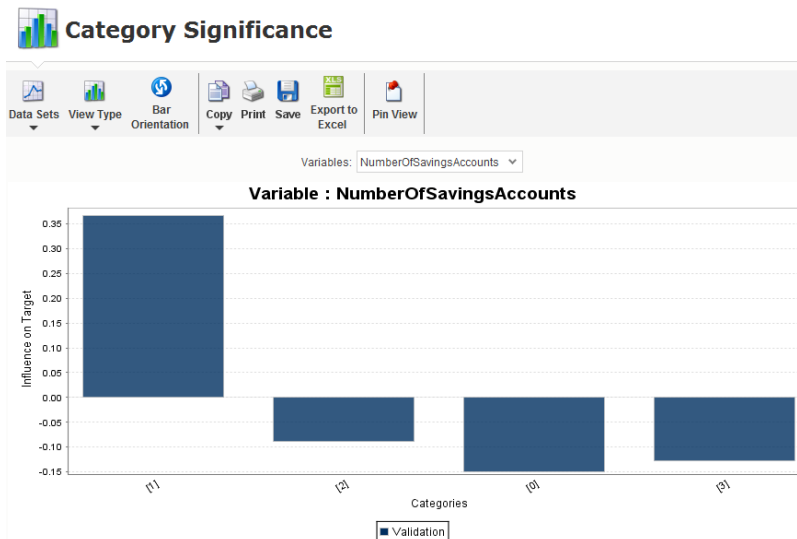
Field Label	Value	Description
		Click on <b>Contributions by Variables</b>
		Click on <b>Bar Orientation</b>

Now you should see the bars and text horizontally which is easier to read.



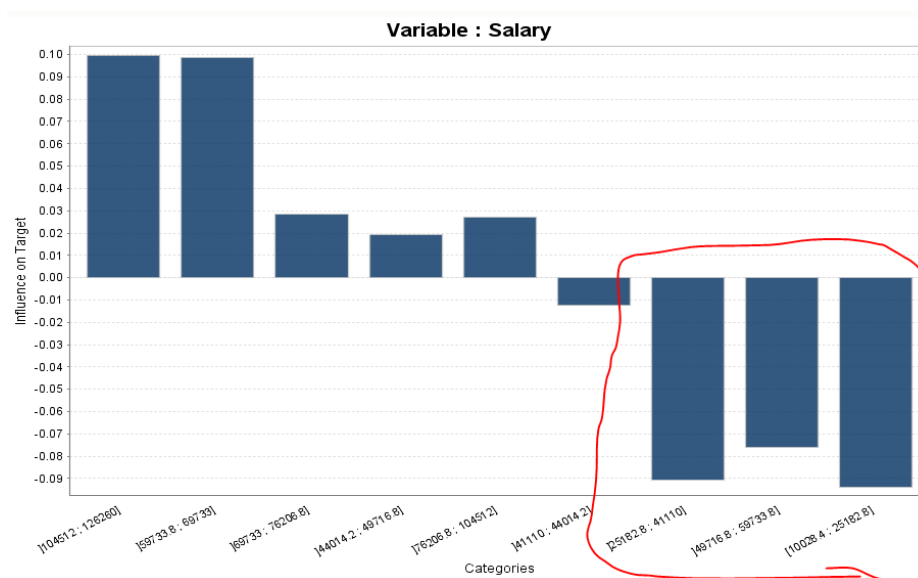
Double-Click on the bar for the **Number of Saving Accounts**.





Customers with no savings account or more than 1 saving accounts have a higher propensity to leave the bank than customers who have just 1 savings account.

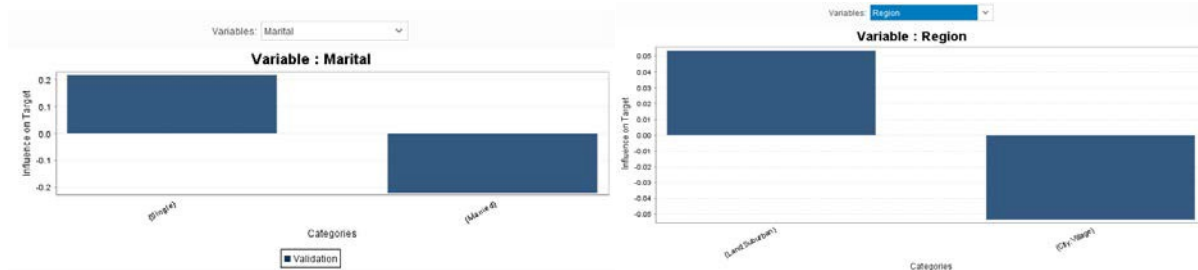
Field Label	Value	Description
		Click <b>Previous</b> .
		Double-click on the bar for <b>Salary</b> .



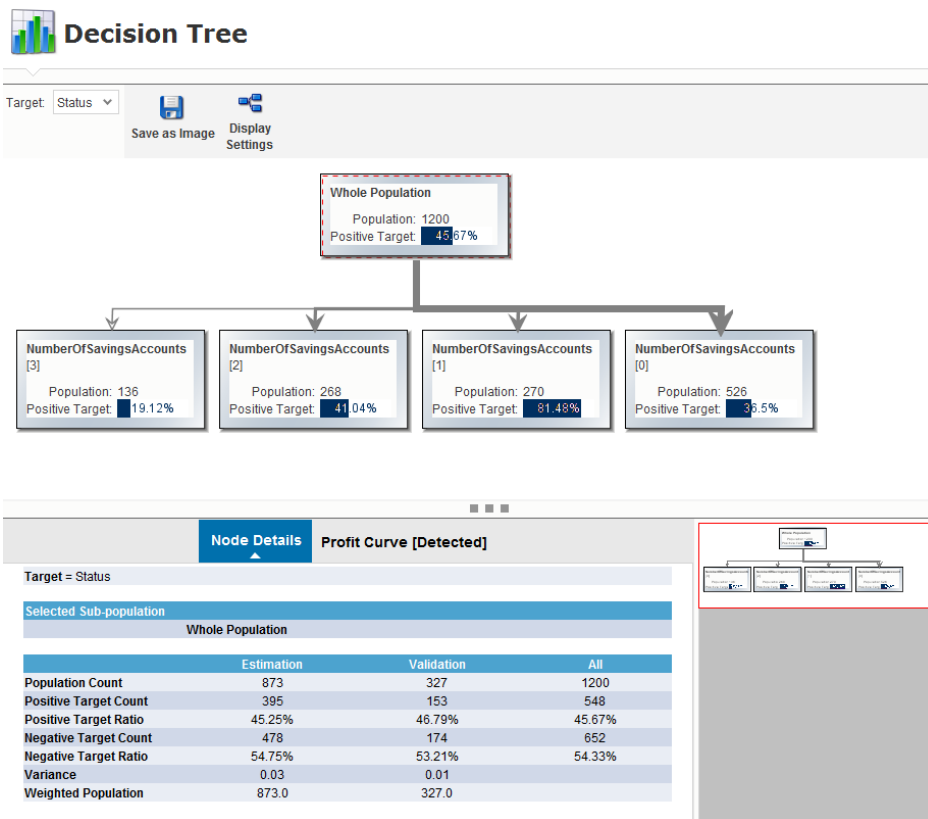
Customers with a lower income have a higher propensity to leave. Click **Previous**. Take a look at the details for the variables Marital and Region. You will notice:

- Married customers have a higher propensity to leave the bank.

- Customer from the city or from a village have a higher propensity.



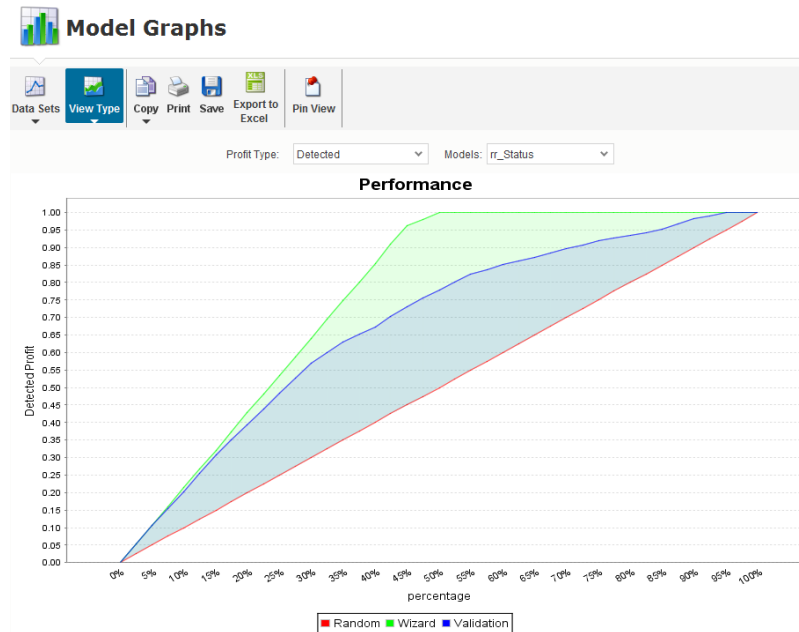
Field Label	Value	Description
		Click <b>Previous</b> .
		Click <b>Previous</b> .
		Click on <b>Decision Tree</b> .



We can see that 45.67 % of the population is not about to leave our bank. The population statistics also show that SAP Predictive Analytics used a cutting strategy with 873 customers to create the model and 327 to validate it.

By expanding the decision tree we can see that the percentage value of loyal customers goes up to 81 % for customers with 1 Savings account. Means we should focus on customers with more than 1 saving account or 0 saving account.


Field Label	Value	Description
		Click <b>Previous</b> .
		Click <b>Model Graphs</b> .



Ensure the **Profit Type** is set to **Detected** and the **Models** option is set to the value **rr\_Status**. The Red line would be the performance if we used no model. The Green line would be the model performance if we created a theoretically perfect model. The Blue line is how well our current model performed. Click **Previous**. Navigate to the **Run** area. Select the option **Simulation**.

Field Label	Value	Description
Salary value	90000	
Marital value	Married	Click <b>Run</b> .

The results of the simulation will appear in the Results section. You will obtain the Predicted value (score) of the observation described in the table of Explanatory variables, as well as the probability that this observation belongs to the target category of the target variable.



## Simulating the Model

Explanatory Variables

Sort by: Contribution of Status

Names	Values	
NumberOfSavingsAccounts	0	
Salary	90000	
Marital	Married	
CheckingAccount	Checking Account	
Region	City	
Mortgage	KxOther	

Results

Output	Value
Score of Status	-0.0539
Proba. of (Status = 0)	0.2903

The probability of our customers – Married with 90,000 \$ income – staying is 29%. Note that certain variables of the table of Explanatory variables were automatically completed upon execution of the simulation. In fact, the model automatically completed certain missing values that were essential to the simulation. In our example the model assumed 0 Savings Accounts, 1 Checking Account, and in the Region City.

Change the **NumberOfSavingsAccounts** to 1 and **Run** the simulation again. This time the probability that the customer is will be loyal jumped to 87%. Click **Previous**, then **Apply Model**.

**Q5. What is the probability of customers with an income of \$106,000, a mortgage, checking account, married, region city and one saving account? 1 – Mark**

We will apply the model to a new data file of customers to now predict their probability of leaving. For this we will load New\_Customers.csv which contains a list of 20 new customers.

Field Label	Value	Description
		Click <b>Browse</b> and select file <b>New_Customers.csv</b> .
		Click <b>OK</b> .

## Applying the Model

**Application Data Set**

Data Type: Text Files

Folder: ExerciseData Browse

Data: New\_Customers.csv Browse

Define Mapping

**Generation Options**

Generate: Advanced Apply Settings Advanced Apply Settings...

Mode: Apply

☐ Add Score Deviation

**Results Generated by the Model**

Data Type: Text Files

Folder: ExerciseData Browse

Data: MichaelExercise2 Browse

Define Mapping

☐ Use Direct Apply in the Database

Cancel Previous Apply

Click on **Advanced Apply Settings**.

File Help

## Model Advanced Apply Settings

**Advanced Apply Settings**

General Outputs

☐ Gain Chart

☒ Outputs for Target 'Status'

☐ Copy Weight Variable

☒ Copy Variables Individual

**User Defined Constant Outputs**

Visibility	Name	Storage	Value	Key
<input type="checkbox"/>	Model Name	string	Status_Customer_Lo...	0
<input type="checkbox"/>	Build Date	datetime	2015-02-17 13:33:56	0
<input type="checkbox"/>	Apply Date	datetime	2015-02-17 15:47:30	0
<input type="checkbox"/>	Model Version	integer	1	0


Add Delete

Cancel Previous OK






Field Label	Value	Description
Copy Variables		Set <b>checkbox</b> .
All		Set this option.
		Click <b>OK</b> .
		Click <b>Browse</b>

For the **Folder** option of **Results Generated by the Model** select the folder **\\07 Predictive Analytics Automated** if not automatically. For the **Results Generated by the model** enter a file name **YourNameExercise2** to store the output.


Field Label	Value	Description
		Click <b>Apply</b> .
		Click <b>View Output</b> .



## Applying the Model

Output Name: JonesExercise2.bt

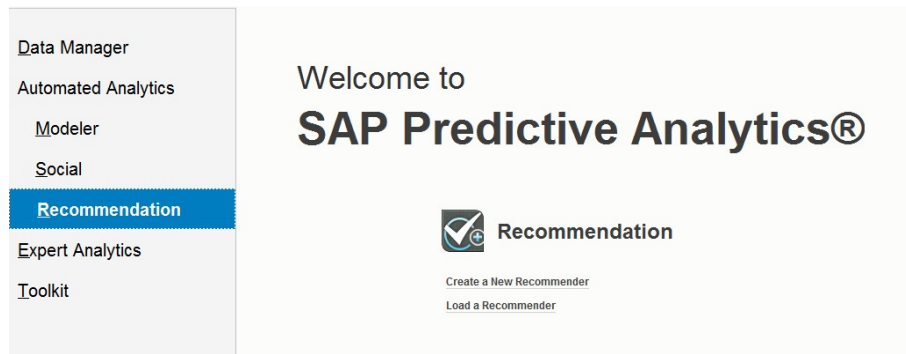
First Row Index:  Last Row Index:  

												Data	Statistics	Graph
	KxIndex	id	Gender	Salary	Marital	Checking...	NumberOf...	PersonalA...	Mortgage	PensionPL...	Region	Status	rr_Status	
1	1		Female	35092	Single	No Checki...	1	No Person...	No Mortgage	No Pensio...	City		0 0.4640692...	
2	2		Male	60170.199...	Married	No Checki...	3	Personal A...	Variable M...	Pension PL...	Village		1-0.1888638...	
3	3		Female	33150.800...	Married	Checking A...	0	Personal A...	No Mortgage	Pension PL...	City		1-0.2216367...	
4	4		Female	40750.800...	Married	No Checki...	3	Personal A...	No Mortgage	No Pensio...	Village		1-0.2218350...	
5	5		Female	101152.60...	Married	Checking A...	0	No Person...	No Mortgage	No Pensio...	Land		1 0.0182832...	
6	6		Female	75739.199...	Married	Checking A...	2	Personal A...	No Mortgage	No Pensio...	Village		0-0.0599577...	
7	7		Male	17754.139...	Single	No Checki...	0	Personal A...	No Mortgage	No Pensio...	Land		0-0.0209429...	
8	8		Male	49893.199...	Married	Checking A...	0	Personal A...	No Mortgage	Pension PL...	Village		1-0.2057600...	
9	9		Female	50608.599...	Married	No Checki...	2	No Person...	No Mortgage	Pension PL...	Suburban		1 0.0011107...	
10	10		Male	48424.199...	Married	Checking A...	2	Personal A...	No Mortgage	Pension PL...	Village		1-0.1717067...	
11	11		Female	119607.8	Married	Checking A...	0	Personal A...	No Mortgage	No Pensio...	Village		1 0.1685010...	
12	12		Female	53317.599...	Single	Checking A...	0	Personal A...	Mortgage	Pension PL...	City		1-0.1083619...	
13	13		Female	31471.599...	Married	Checking A...	1	Personal A...	Variable M...	No Pensio...	Village		0 0.1218232...	
14	14		Female	110409.39...	Married	Checking A...	1	Personal A...	Variable M...	Pension PL...	Village		0 0.5246688...	
15	15		Male	38949.199...	Married	Checking A...	0	Personal A...	Variable M...	No Pensio...	Land		1-0.0949828...	
16	16		Female	44684.199...	Married	Checking A...	0	Personal A...	Mortgage	Pension PL...	City		1-0.2061978...	
17	17		Female	35459.599...	Married	No Checki...	2	No Person...	Mortgage	No Pensio...	Village		1-0.0761856...	
18	18		Female	82032	Married	Checking A...	0	No Person...	Mortgage	No Pensio...	Suburban		1-0.0370178...	
19	19		Female	53818.400...	Married	Checking A...	0	No Person...	No Mortgage	No Pensio...	City		0-0.2511181...	
20	20		Male	45045.599...	Married	Checking A...	0	Personal A...	No Mortgage	Pension PL...	Village		1-0.1780536...	

**Q6.** Here you can see the scores for the new 20 customers. Which one is most likely to leave? **1 - Mark**

Click on the **rr\_Status** column to sort to see the smallest likelihood of being a loyal customer. This should make sense, given what we learned about our customers. The least likely customer to stay is Married, has a lower salary of 53,818, no Savings Accounts, and from a Region “City”. Click **Next**.

Navigate back to the Home Panel of SAP Predictive Analytics (using Previous or Cancel). Click **Recommendation**.



Click **Create a New Recommender**. Here we will load a file that contains events which trigger a recommendation.

### Transaction Data

Main

Analyze

Open Description

Save in Variable Pool

Save Description

Remove from Variable Pool

View Data

Properties

Description
View

Data Type: Text Files

Folder: ExerciseData

Events: Transaction.csv

Index	Name	Storage	Value	Key	Order	Missing	Group	Description

Cancel

You must analyze the data or open a description file.

Previous

Next

Field Label	Value	Description
Data Type	Text Files	
Folder	. \07 Predictive Analytics Automated	Click <b>Browse</b> for the <b>Events</b> option. Select the file <b>Transaction.csv</b> .
		Click <b>OK</b> .
		Click <b>Analyze</b> .

Field Label	Value	Description
		Click <b>View Data</b> to inspect what kind of data we are working with.

Data Set: Transaction.csv

First Row Index: 1 Last Row Index: 100

	KxVar1	KxVar2	KxIndex
1	12101	Fixed Mortg...	1
2	12599	Savings Ac...	2
3	12600	Rent Depo...	3
4	12600	Checking A...	4
5	12600	Savings Ac...	5
6	12601	Checking A...	6
7	12602	Variable M...	7
8	12602	Rent Depo...	8
9	12602	Checking A...	9
10	12602	Savings Ac...	10
11	12603	Checking A...	11
12	12603	Savings Ac...	12
13	12604	Checking A...	13
14	12604	Savings Ac...	14
15	12605	Checking A...	15
16	12607	Variable M...	16
17	12607	Checking A...	17
18	12607	Savings Ac...	18
19	12608	Variable M...	19
20	12608	Checking A...	20
21	12608	Savings Ac...	21

As you can see here we have a customer number and a transaction related to a specific bank product. So when a customer opens a new Checking Account, we will predict what else that customer might be interested in purchasing.

Field Label	Value	Description
		Click <b>Close</b> .
		Click <b>Next</b> .

## Transactions Description

Column Roles

User: KxVar1

Item: KxVar2

Weight:

Time Period


Date Column:

Start Date: 2015-02-17 12:19:56

End Date: 2015-02-17 12:19:56



Field Label	Value	Description
User	KxVar1	
Item	KxVar2	Click <b>Next</b> .
		Click <b>Next</b> .
		Click <b>Generate</b> .
		Click <b>Next</b> .
		Click on <b>Recommendation</b> in the <b>Display</b> area.



### Recommendation

Recommender: KxVar2 ☒ Do not recommend if already owned


☐ Include Best Sellers

KxVar1 12101 ☒ Keep Top N 5

[Get Recommendations](#)

You can now enter a customer ID number – for example **12101** – into the field **KxVar1**.

Click on **Get Recommendations** to receive the recommendations.



### Recommendation

Recommender: KxVar2 ☒ Do not recommend if already owned

☐ Include Best Sellers

KxVar1 12101 ☒ Keep Top N 5

[Get Recommendations](#)

Recommendation	Confidence
Savings Account	65.67%
Variable Mortgage	17.91%
Rent Deposit Account	6.53%

For the customer 12101, three products are recommended - savings account is the number one product that should be recommended to this customer based on the Transaction of her getting a Fixed Mortgage, which you can see in the transaction table from step 64.

If you look in the raw data excel file Customer\_Loyalty.csv you will see what her profile looks like.

200	13299	Male	29423.6	Married	Checking	0	No Personal	Variable	No Pension	City	1
201	13300	Male	53343.2	Single	No Checking	0	Personal	Variable	Pension	Village	0
202	12101	Female	35092	Single	No Checking	1	No Personal	No Mortgage	No Pension	City	0
203	12102	Male	60170.2	Married	No Checking	3	Personal	Variable	Pension	Village	1
204	12103	Female	33150.8	Married	Checking	0	Personal	No Mortgage	Pension	City	1

Q7. What is the recommendation for customer 12605? Only has Checking Account currently – 1 Mark

7 - Marks