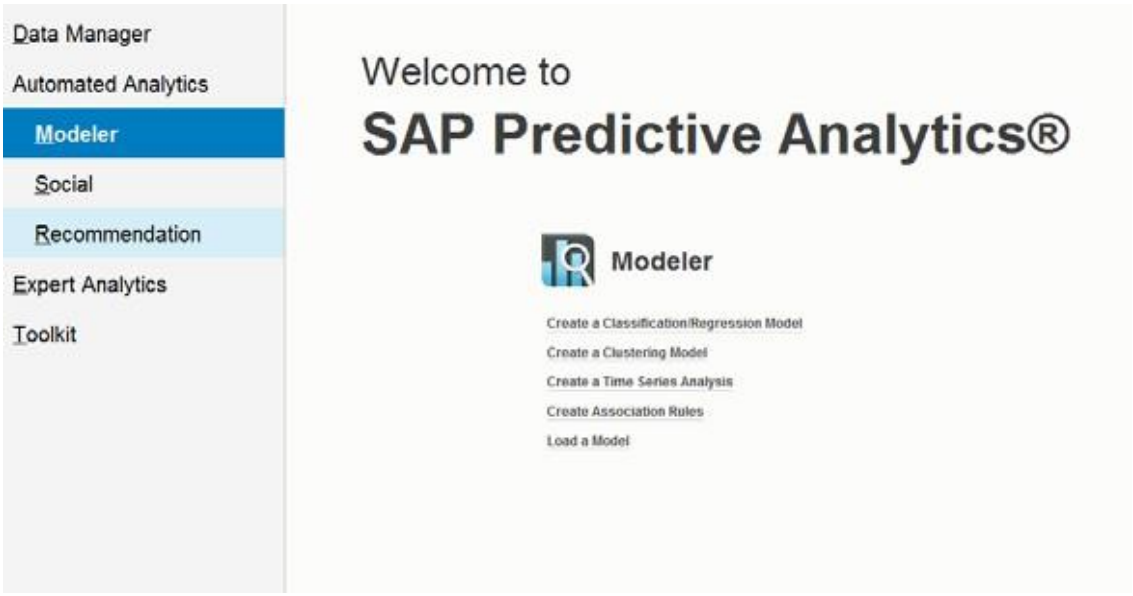
Teaching Material Sample

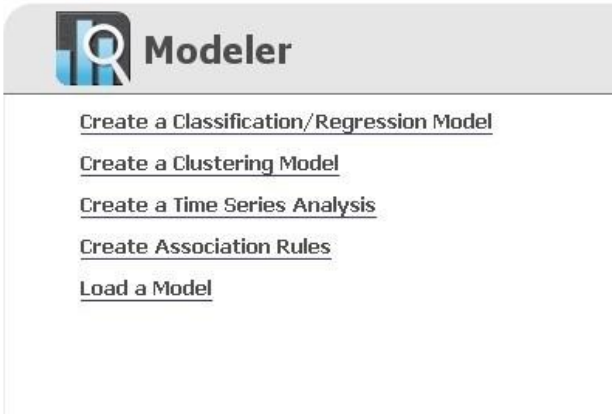
Exercise 1: Auto Insurance Customers/ CRM /Risk Analysis with SAP Predictive Analytics

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.

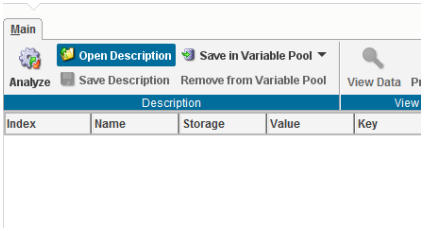
Start SAP Predictive Analytics by following the menu **Start > All Programs > SAP Business Intelligence > SAP Predictive Analytics Desktop > SAP Predictive Analytics**

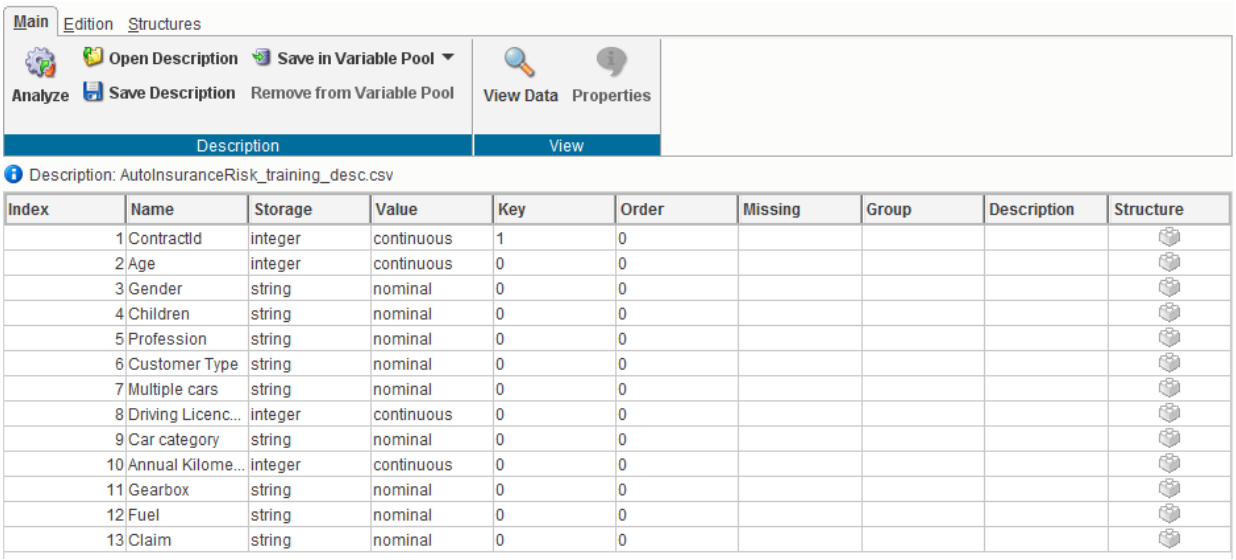


Click on the option **Modeler**.



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

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

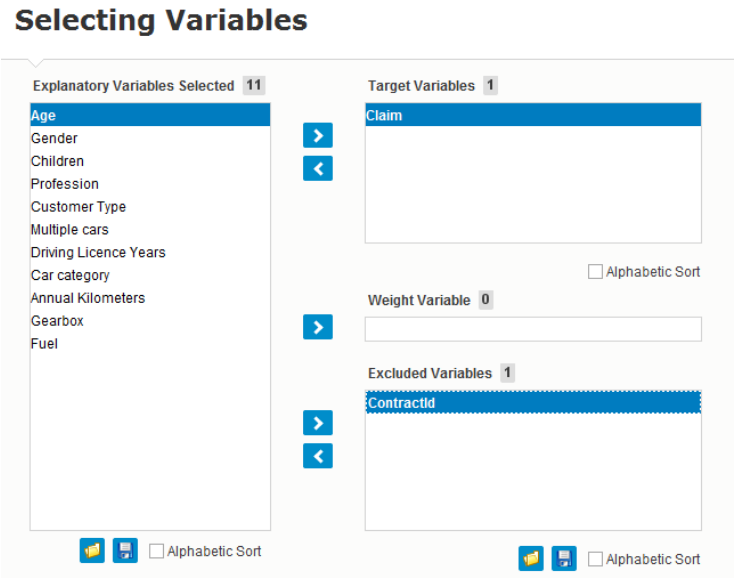


Index	Name	Storage	Value	Key	Order	Missing	Group	Description	Structure
1	ContractId	integer	continuous	1	0				
2	Age	integer	continuous	0	0				
3	Gender	string	nominal	0	0				
4	Children	string	nominal	0	0				
5	Profession	string	nominal	0	0				
6	Customer Type	string	nominal	0	0				
7	Multiple cars	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.

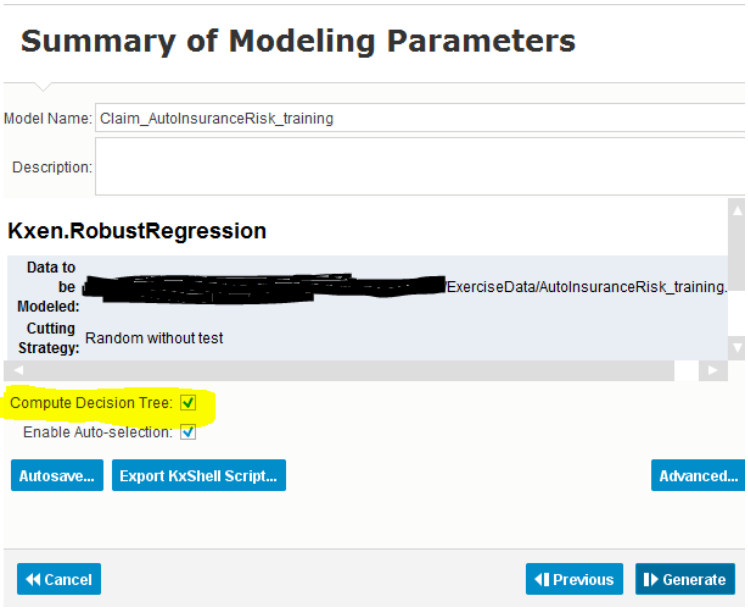
Field Label	Value	Description
		Click **Close** to close the sample data.
		Click **Next**.



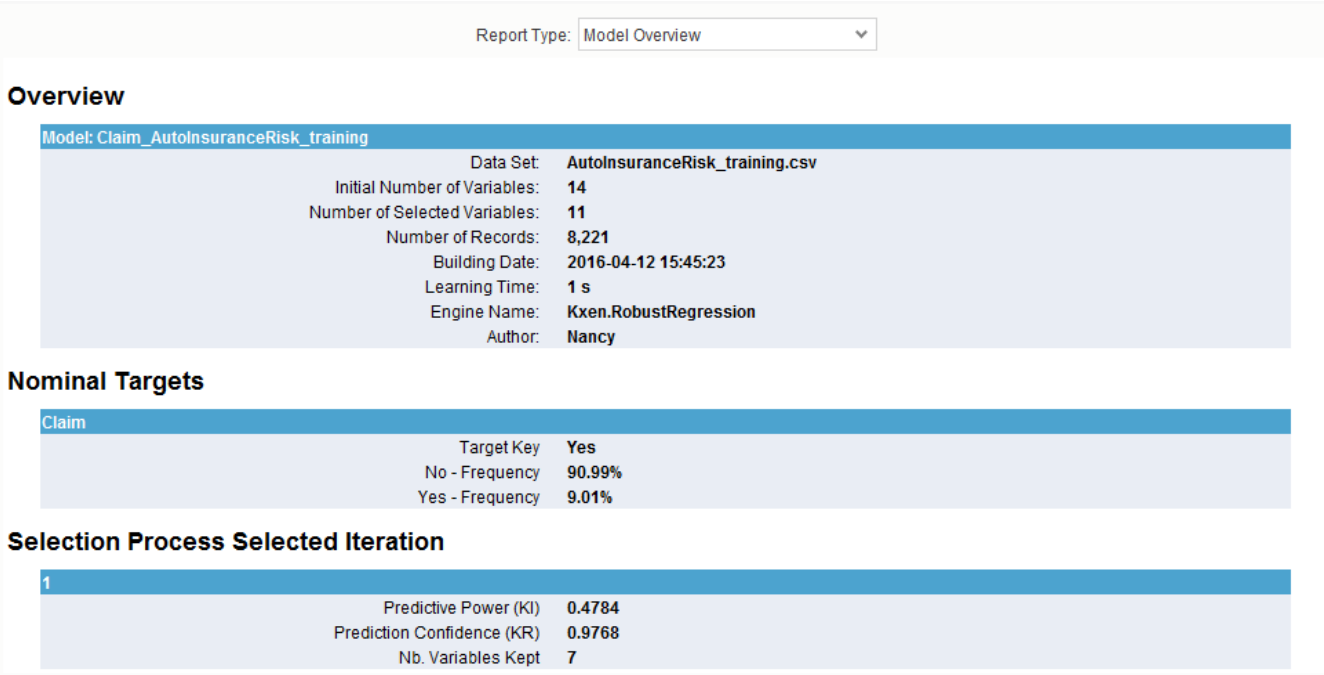
The model identifies a pattern in how one, few or all of the explanatory variables lead to a claim.

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.

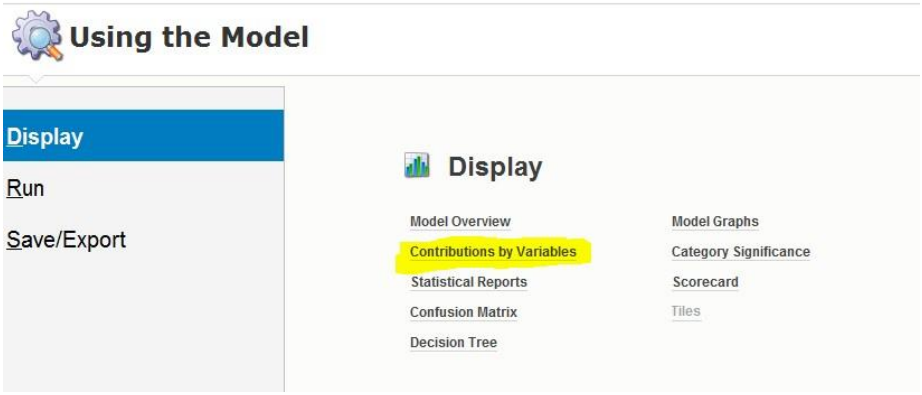
Check that the column **Contract ID** is in the list of the **Excluded Variables**. Click **Next**.



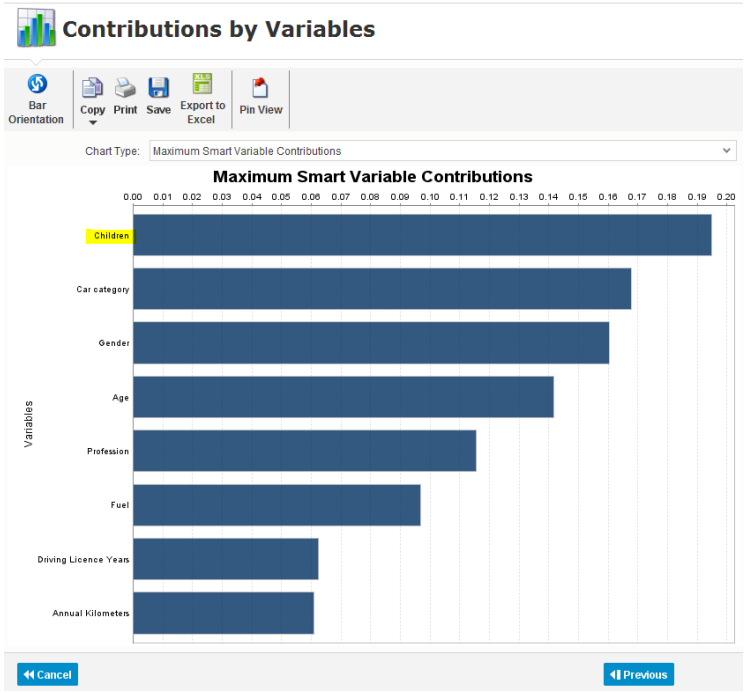
Activate the option **Compute Decision Tree**. Click **Generate**.



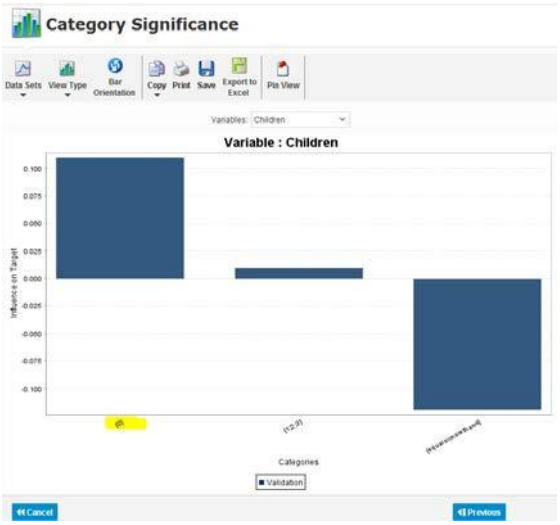
At the bottom you can see that SAP Predictive Analytics found 7 variables that are influencing the claims. Click **Next**.



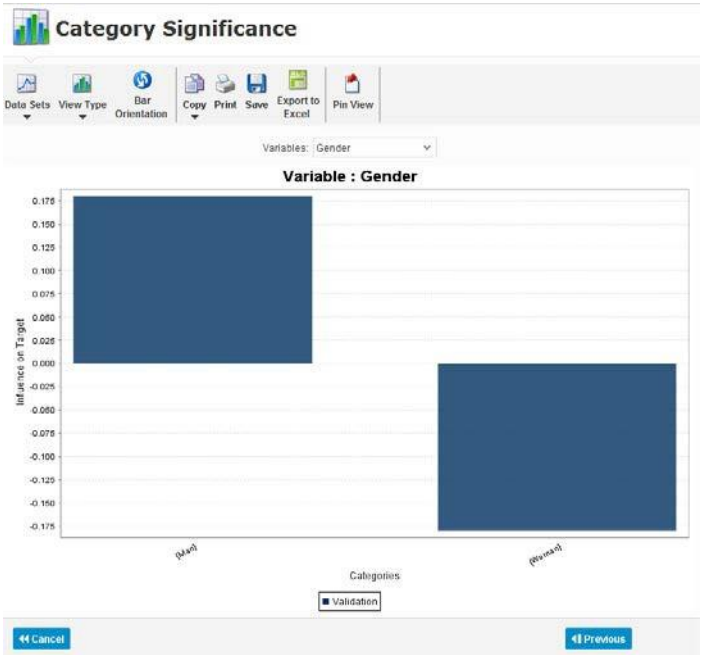
Click on **Contributions by Variables**. – You can change the direction of the Bars by clicking on the Bar **Orientation** Button on the top left.



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

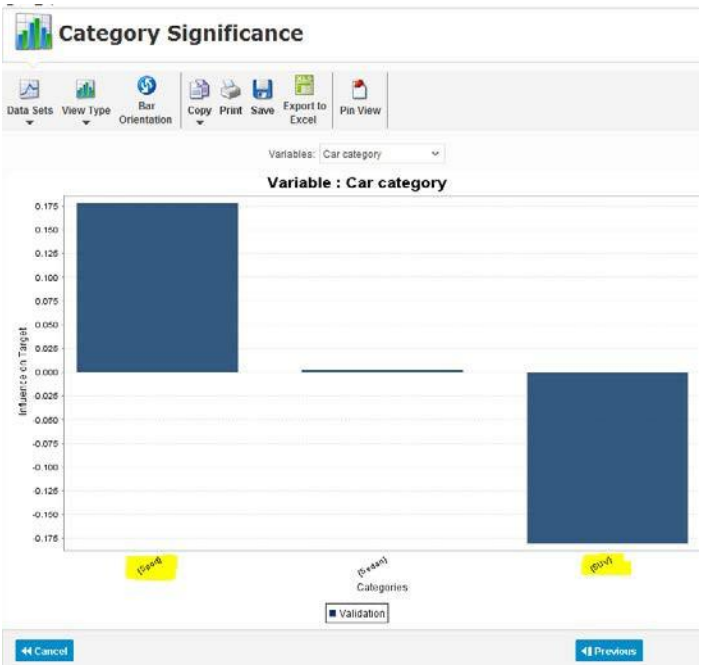


You can see now that customers without children have a much higher propensity to have an accident compared to customers with 4 or more children. Another way to see this is that positive numbers for Influence on Target means that having no children has a greater likelihood of accidents, while the negative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



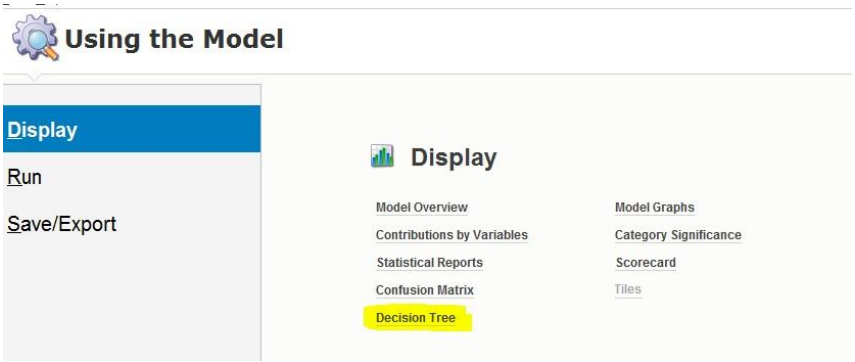
You see that Gender makes a difference.

Field Label	Value	Description
		Click on **Previous**.
		Double-click on the bar for **Car Category.**

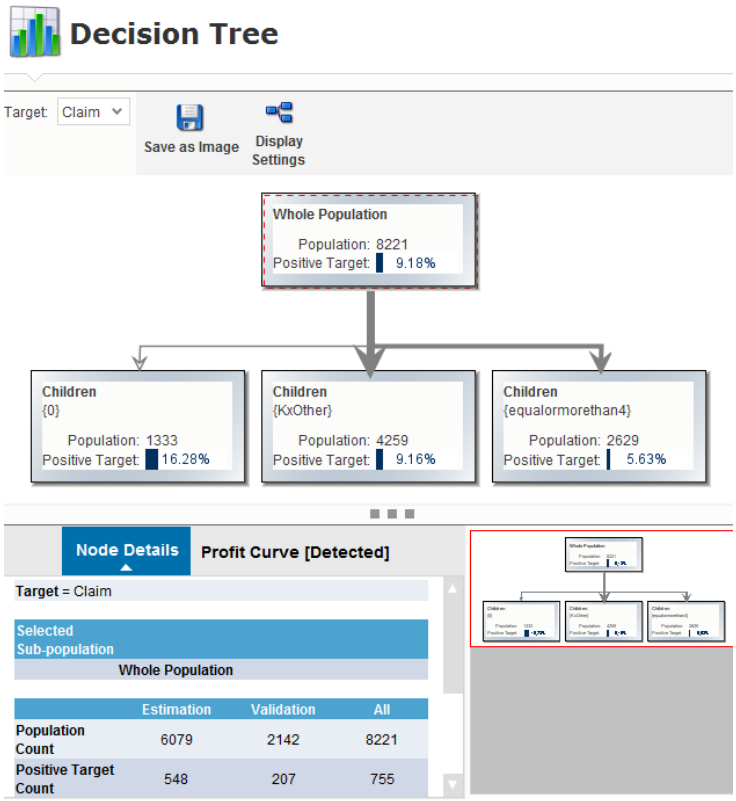


So a man in a sports car is a much higher risk than a woman in a SUV.

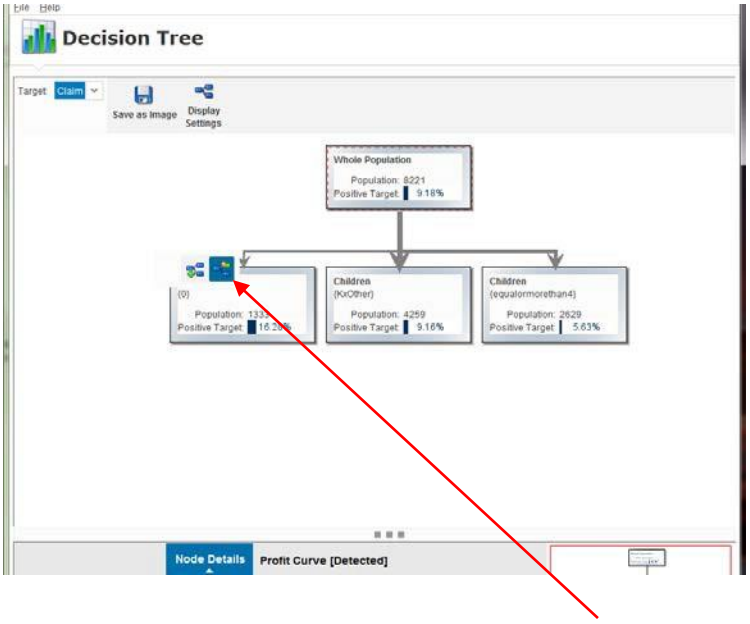
Field Label	Value	Description
		Click **Previous**.
		Click **Previous**.



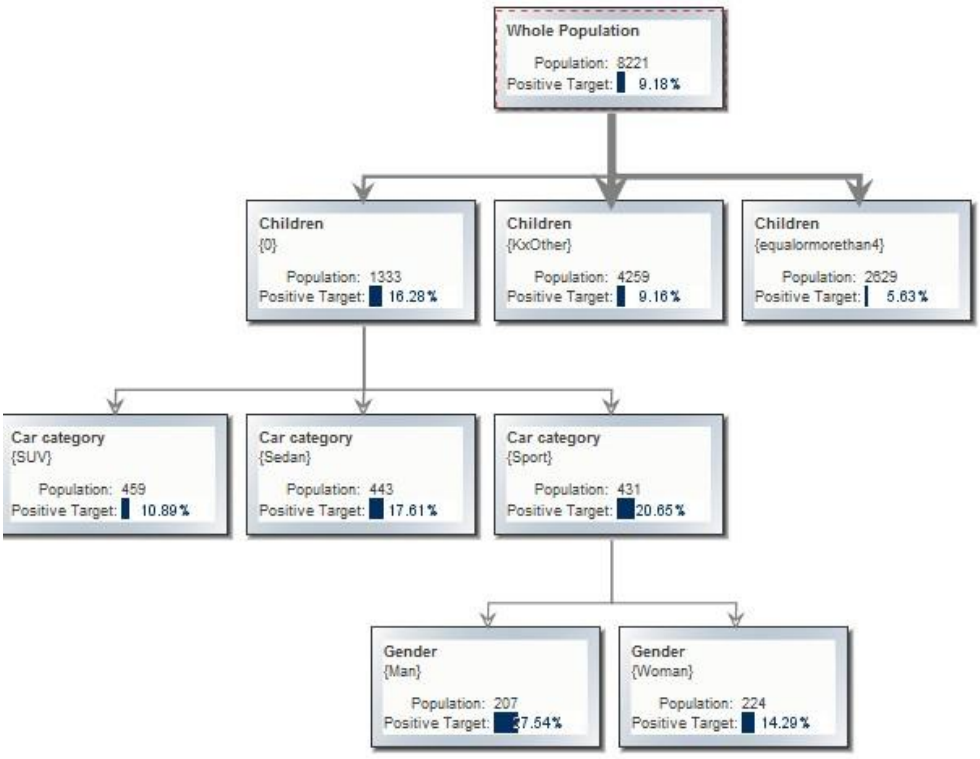
Click on **Decision Tree**.



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**.



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

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}**

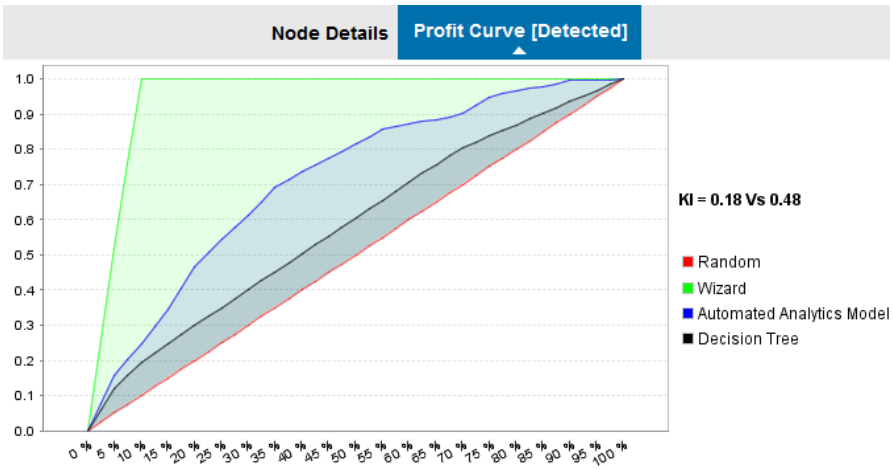
	Estimation	Validation	All
Population Count	146	61	207
Positive Target Count	41	16	57
Positive Target Ratio	28.08%	26.23%	27.54%
Negative Target Count	105	45	150
Negative Target Ratio	71.92%	73.77%	72.46%
Variance	0	0	
Weighted Population	146.0	61.0	

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.

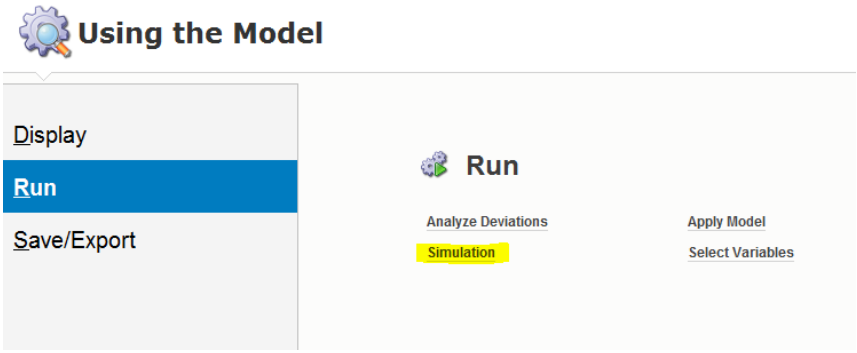
Click on **Profit Curve [Detected]** to view the ROC curve for the Predictive Analytics model compared to the Decision Tree and a Random selection.

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.

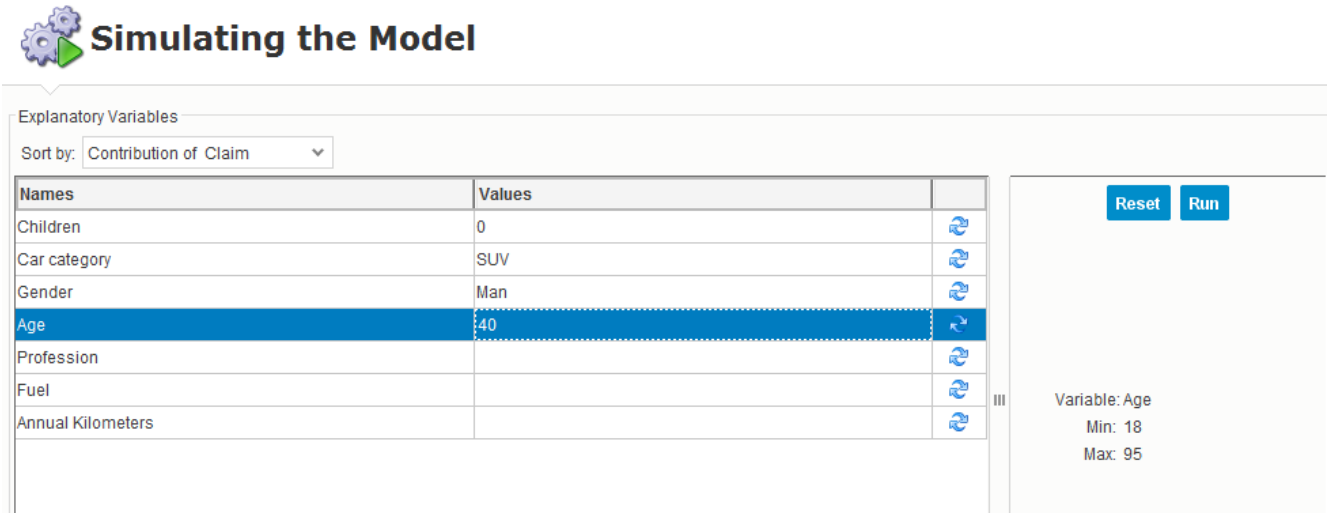


Field Label	Value	Description
		Click **Previous**.
		Select the option **Run**.

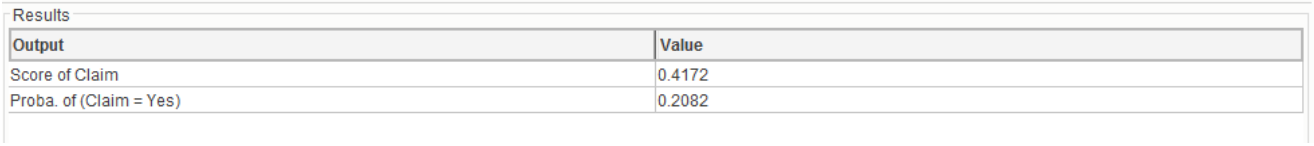


Select the option **Simulation**. Enter the following values:

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



Click **Run**.



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.