



# **Universität Mannheim**

## **DataFest 2015 Midnight Challenge**

### **SAP Predictive Analytics Instructions**

v3.3



## Table of Contents

---

Quick Start Guide.....	3
Welcome.....	4
Challenge Details.....	4
Prerequisites.....	5
Software Installation and Setup.....	5
Installing Predictive Analytics 2.0.....	6
Connecting Automated Analytics to SAP HANA .....	7
Creating a Model.....	13
Analyzing a Model.....	21
Applying a Model .....	28
Assignment Documentation .....	33
Assignment Submission.....	34
Additional Resources .....	35
Using Expert Analytics (Optional) .....	35
About SAP.....	39
About the SAP Big Data Analytics Team .....	39

## Quick Start Guide

---

*(Estimated time for this section: ~5 minutes)*



### Challenge

- Use [Deutsche Eishockey Liga](#) historical data (1994 through 2014) to create a model in SAP Predictive Analytics that allows you to predict which hockey players have the highest likelihood for the most wins in the current season (2014 through 2015)
- This is an individual challenge
- The challenge is 2 hours in length
- Materials are accessible in the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015))



### Prize

- Two winners will be chosen who will each receive an interview for an internship on the Big Data Analytics team at SAP



### Assignment Documentation

1. What was your methodology and process to create your model?
2. What is your rationale and understanding of your model?
3. What other interesting findings would you like to mention?
4. Who are the top 6 players who have the most likelihood for the most wins (players can be from any team and any position)?



### Checklist

- ☐ Install SAP Predictive Analytics 2.0
- ☐ Connect SAP Predictive Analytics (Automated Analytics) to SAP HANA
- ☐ Create a regression model with one of the 1994-2014 datasets
- ☐ Apply the model to the 2014-2015 dataset
- ☐ Document your findings in the PowerPoint template
- ☐ Submit your findings via email to Thorsten Henrichs ([t.henrichs@sap.com](mailto:t.henrichs@sap.com))

## Welcome

---

*(Estimated time for this section: ~1 minute)*

Greetings,

Thank you for taking part in the DataFest Midnight Challenge!

Over the next few hours you will be using SAP technology to predict which players in the Deutsche Eishockey Liga (DEL) are most likely to win the most games.

The following document will guide you through what you need to do in order to be successful.

We are glad you have chosen to participate and hope you enjoy the challenge!

Sincerely,

The SAP Big Data Analytics Team | SAP Enterprise Analytics

## Challenge Details

---

*(Estimated time for this section: ~5 minutes)*

Each participant will be given a set of historical data (1994 through 2014) that contains statistics for hockey players in the [Deutsche Eishockey Liga](#) (DEL).

Your challenge is to use this data to create a model that will allow you to predict which hockey players have the highest likelihood for the most wins in the current 2014-2015 season. You will select the top six players whom you find are likely to have the most wins. Players can be from any team and any position.

You will also be asked to document your answers to a series of questions that help explain the details within your model.

By using SAP Predictive Analytics you will analyze players' performance statistics (e.g. goals and assists), physical characteristics (e.g. height and weight), and other variables to find how they impact wins. Through a regression analysis you will individually score players in the current season according to likelihood for most wins.

SAP Predictive Analytics will help you complete the challenge through use of the Automated Analytics Modeler.

This is an individual challenge; final submissions must come from single students.

You have two hours to work on the challenge. If you cannot complete all steps, or answer all questions within the time allowed, please submit what results you do have at the end of the

challenge. An SAP expert and member of the Big Data Analytics team, Thorsten Henrichs, will be on-site during the challenge to moderate the event.

The two winners of the challenge will be chosen based upon completeness of answers, number of actual wins from the six players selected, documented work, thought process, and creativity.

The two winners will each receive an interview for an internship within the Big Data Analytics team in SAP.

## Prerequisites

---

*(Estimated time for this section: ~1 minute)*

The following knowledge, hardware, and tools will be needed to succeed in this event:

1. English language proficiency
2. Understanding of statistical concepts and data analysis
3. Computer running Windows
4. Internet connection
5. Around 10 GB of space
6. SAP Predictive Analytics 2.0 (provided)
7. Microsoft PowerPoint (or other compatible presentation software)

If a section is marked as **OPTIONAL**, it is not a required step. Supporting documentation or tools may be provided if you need additional information, are running into issues, or have the curiosity to explore something further. However, given the challenge must be completed in two hours, please plan your time accordingly.

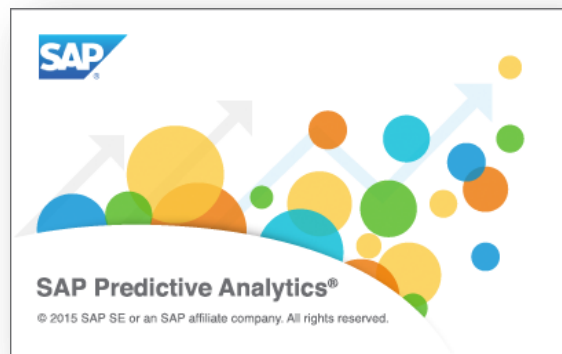
## Software Installation and Setup

---

*(Estimated time for this section: ~30 minutes)*

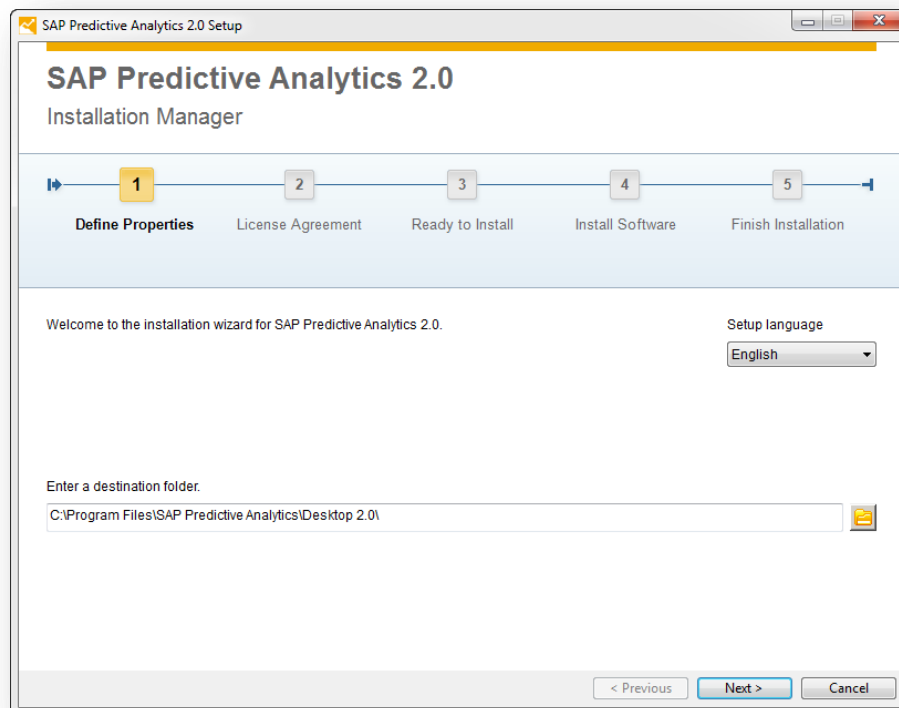
All materials needed to participate and succeed in the challenge can be found in a public [SAP Mobile Documents folder](#). You may utilize this shortened link to access: [spr.ly/datafest2015](http://spr.ly/datafest2015)

To get started, you'll first need to download and install SAP Predictive Analytics.

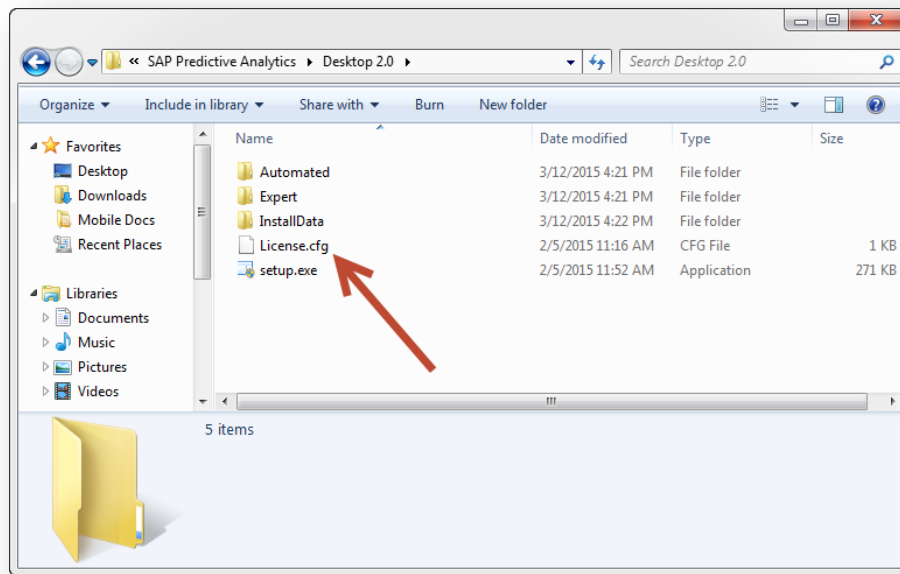


## Installing Predictive Analytics 2.0

1. Download the “PADESKTOP2000\_0.EXE” file from the SAP Mobile Documents folder and double click the icon to begin installation.
2. Follow the 5 steps as guided by the installation manager.
3. **OPTIONAL:** For detailed instructions please refer to the “Installation\_Guide” PDF within the SAP Mobile Documents folder.



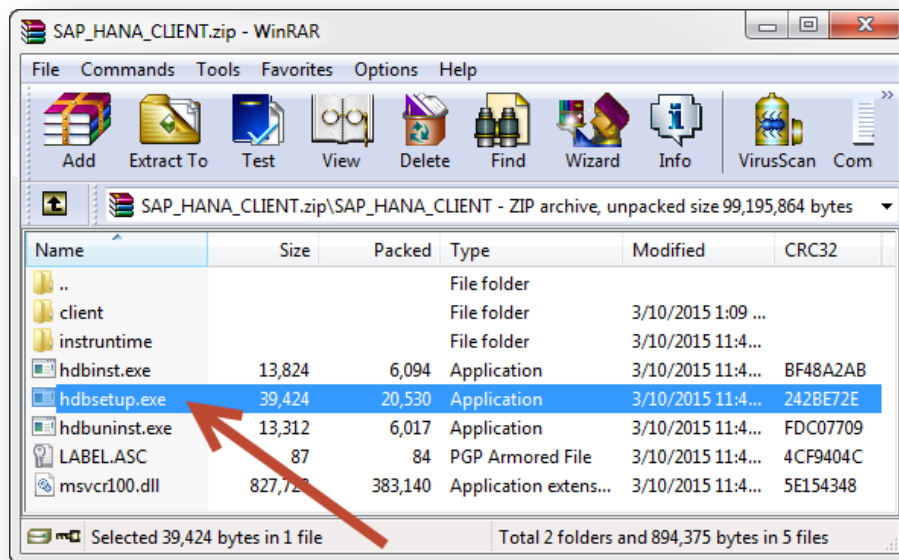
4. Make sure to close Predictive Analytics as you now need to update the license key. Navigate to the C:\Program Files\SAP Predictive Analytics\Desktop 2.0 folder. Copy the “License.cfg” file from the SAP Mobile Documents folder to your desktop. Then drag the “License.cfg” file over into the SAP Predictive Analytics\Desktop 2.0 folder and overwrite the existing file. **OPTIONAL:** Or you may change the keycode string in the “License.cfg” file to the following: DG20Z-R6YWH38-WK61WAB-0AYT279-X7



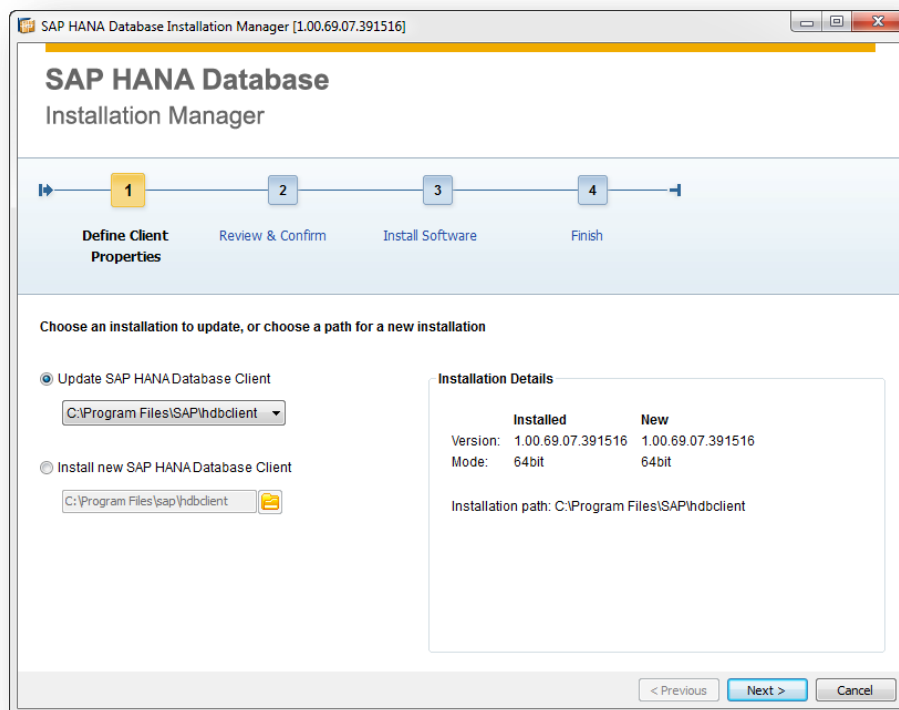
## Connecting Automated Analytics to SAP HANA

In order to use Automated Analytics on the DEL dataset, you need to connect to the SAP HANA database where the data resides.

1. Download the “SAP\_HANA\_CLIENT.zip” from the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015)) and run the “hdbsetup.exe” file. (You may need to tell your computer to trust this program.)

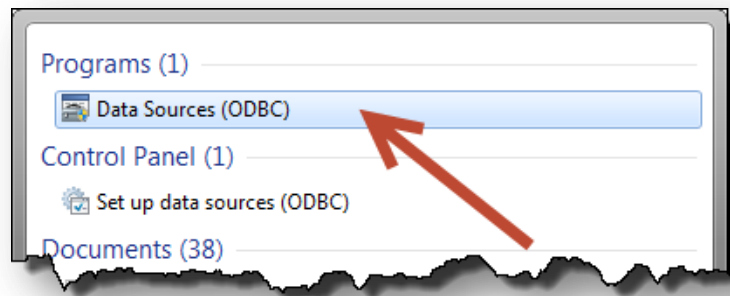


2. Follow the 4 steps as guided by the installation manager.

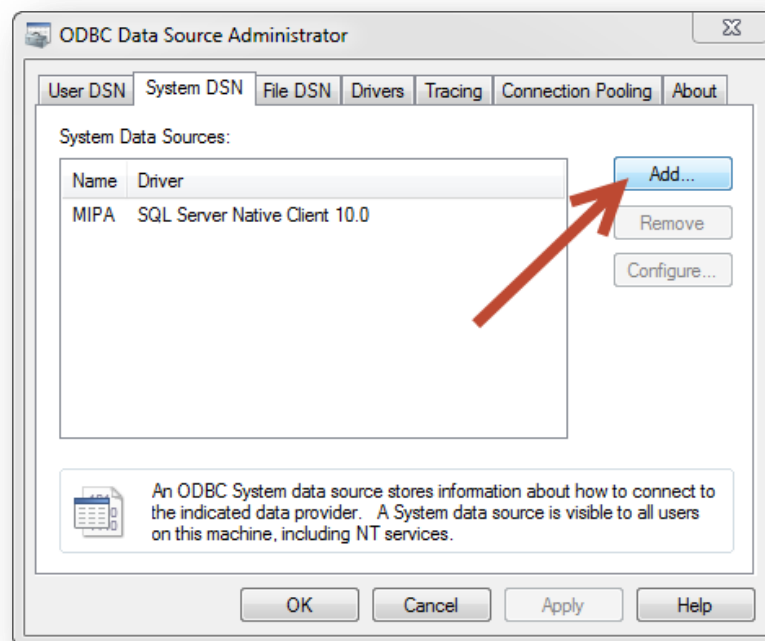


3. Open the Windows Start menu and type in "ODBC" and run the "Data Sources (ODBC)" program. (If running Windows 8, search for "ODBC" and select "ODBC Data Sources (64-bit)" or 32-bit according to your computer.)

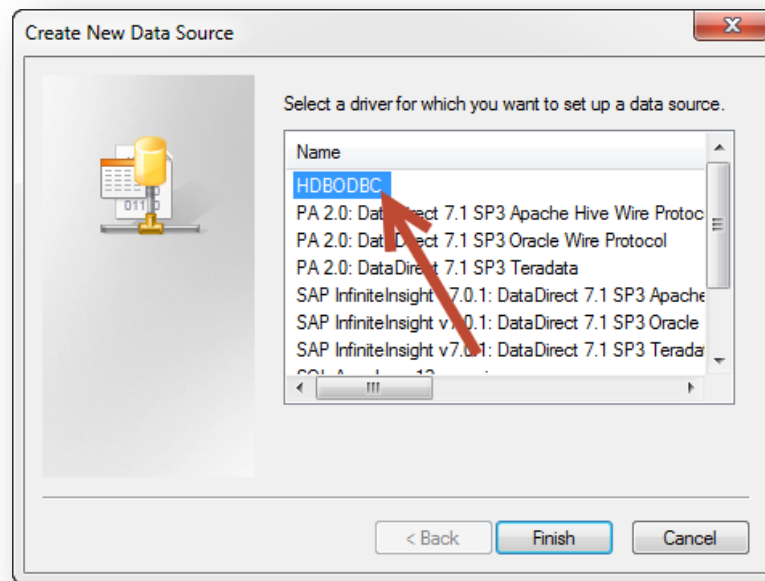




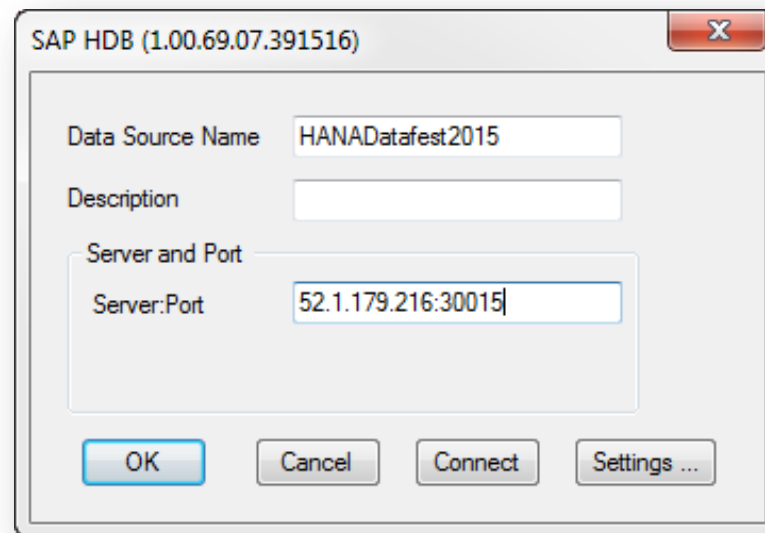
1. Navigate to the “System DSN” tab. Click “Add”.



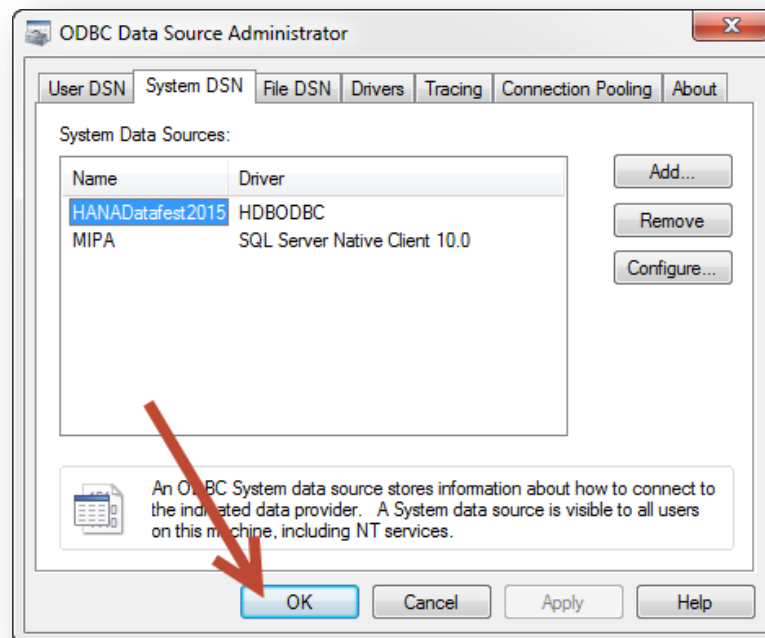
2. Select “HDBODBC” and click “Finish”.



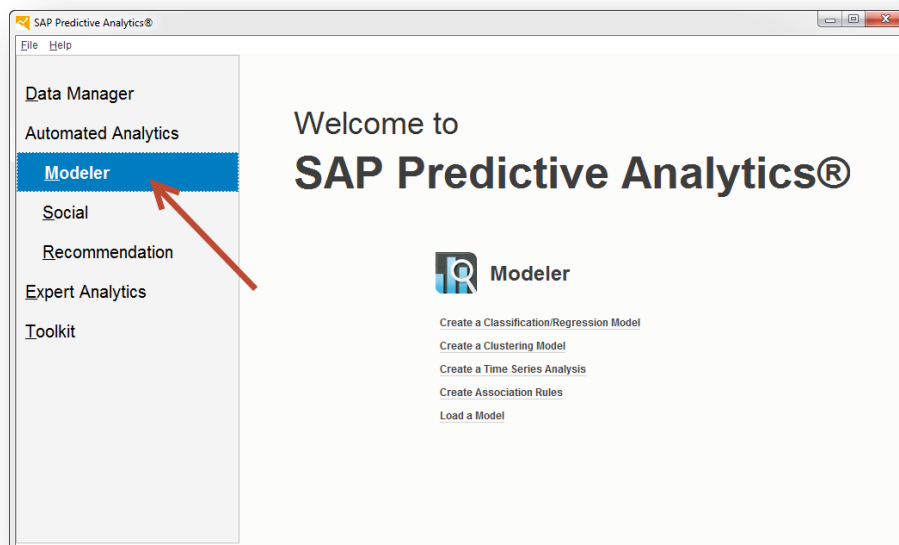
3. In the “Data Source Name” field type in HANADatafest2015. In the “Server:Port” field type in 52.1.179.216:30015 Click “OK”.



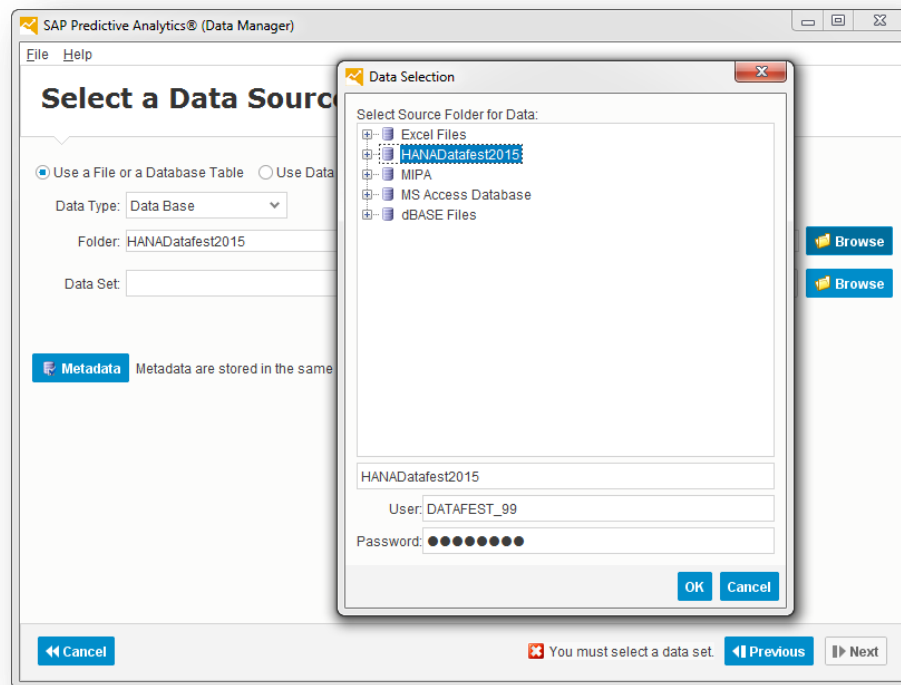
4. Back on the “ODBC Data Source Administrator” window, click “OK”.



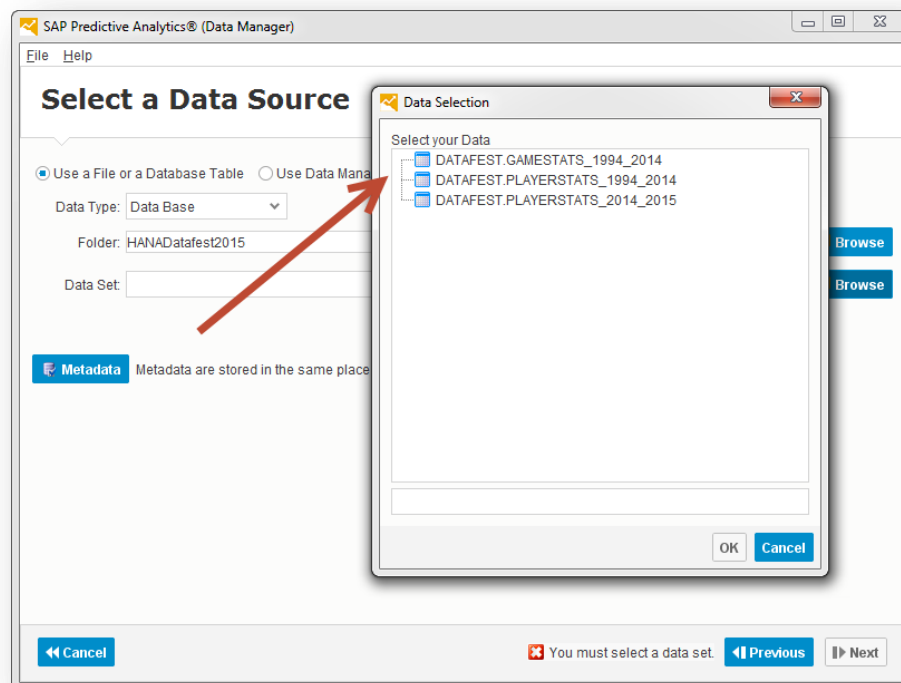
5. Open Predictive Analytics and navigate to Automated Analytics: Modeler. Click the “Create a Classification/Regression Model” hyperlink to launch the tool.



6. Set the “Data Type” to “Data Base”. Click on “Browse” next to the “Folder” line and select “HANADatafest2015”. In the “User” field type in your username (DATAFEST\_XX) and type in Data1234 in the “Password” field. Click “OK” to establish a connection. (Please see Thorsten Henrichs or other on-site SAP representative for your username.)



- Click on “Browse” next to the “Data Set” line and select either the “PLAYERSTATS\_1994\_2014” or “GAMESTATS\_1994\_2014” table to ensure the connection is working. Click “OK”.



To continue working in Automated Analytics, move ahead in this documentation to the section “Creating a Model”. If you desire to first visualize or further explore the data, move on to “Using Expert Analytics (Optional)” section under “Additional Resources” at the end of this document.

## Creating a Model

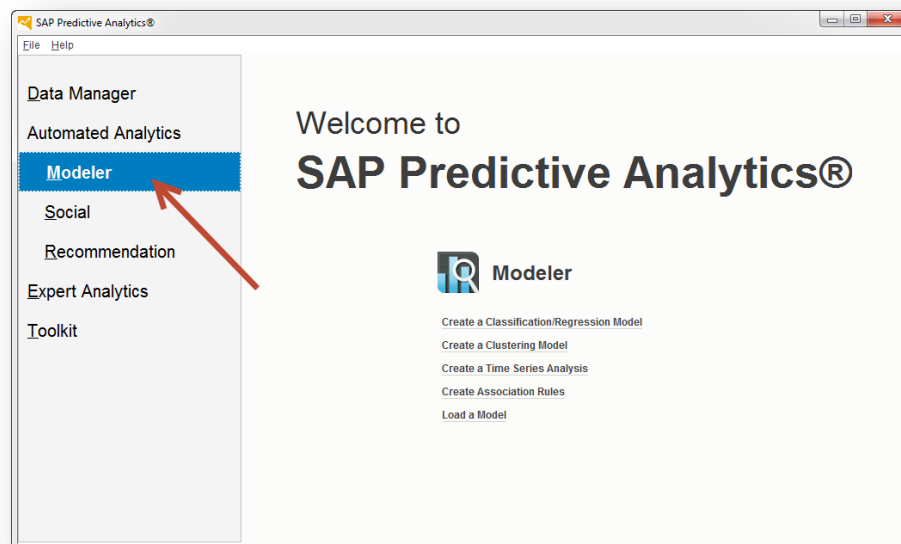
---

*(Estimated time for this section: ~15 minutes)*

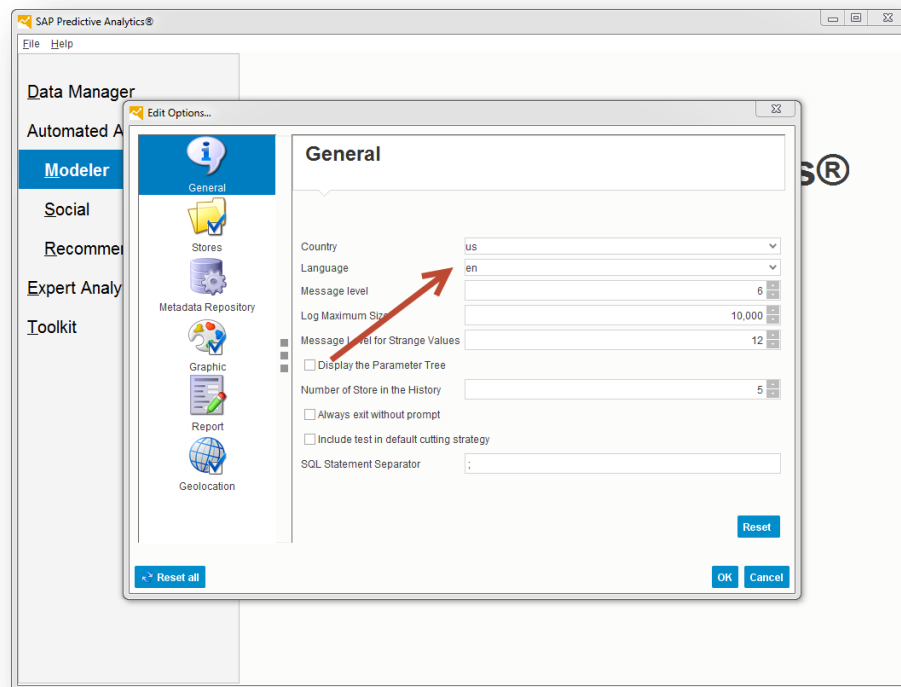
In order to select the players who are most likely to have the most wins, a regression analysis must be performed to generate a model. The following steps will give you an overview of how to create a model in SAP Predictive Analytics.

**OPTIONAL:** For more detailed instructions please refer to the “Regression\_Guide” PDF within the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015)).

1. If not there already, open Predictive Analytics and navigate to Automated Analytics: Modeler. Click on the “Create a Classification/Regression Model” hyperlink.



2. **OPTIONAL:** If default language is not set to English, and you would prefer it to be (in order to better follow these instructions), click on “File”, “Preferences”, and set “Language” (or the second dropdown list) to “en”.

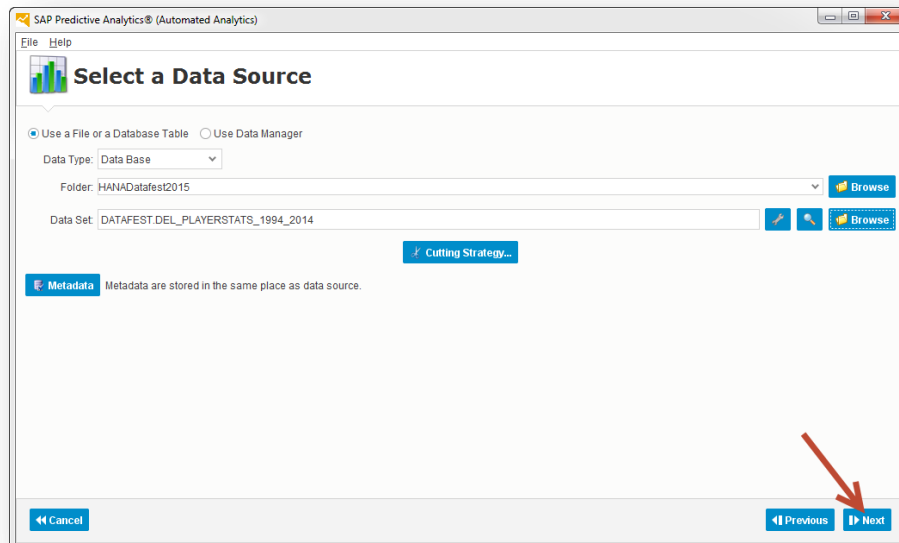


3. Click on the “Browse” button next to the “Data Set” line, you will select either the “PLAYERSTATS\_1994\_2014” or “GAMESTATS\_1994\_2014” table.

*As seen in the diagram below, you have the option of building a model off of a data set aggregated on the game level (and building a model with a nominal target) or aggregated on the player level (and building a model with a continuous target). Decide how you would like to proceed and click “OK” to continue.*

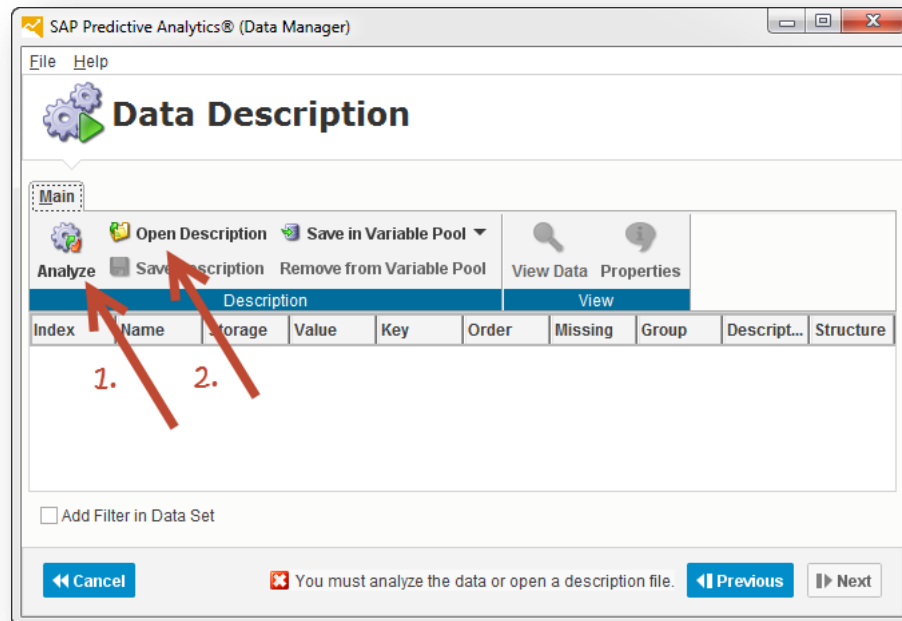
Used for	Table	Used by
<i>Model Development</i>	<i>for Binary Regression</i> <b>GAMESTATS_1994_2014</b> Contains <b>running averages</b> of the players historic performance as well as performance indicators of the last season for <b>all games</b> conducted between 1994 and 2014	<i>Participants</i>
	<i>for Metric Regression</i> <b>PLAYERSTATS_1994_2014</b> Contains <b>averages</b> of the players historic performance as well as performance indicators of the last season for <b>all players</b> being active between 1994 and 2014	
<i>Model Application</i>	<b>PLAYERSTATS_2014_2015</b> Contains <b>averages</b> of the players historic performance as well as performance indicators of the last season for <b>all players</b> being active between 1994 and 2014 <b>and</b> being active 2014/2015	<i>Participants</i>
<i>Model Validation</i>	<b>VALIDATION_DATASET</b> Results of season 2014/2015 year to date	<i>Judges</i>

4. Click “Next” to proceed.



5. Click “Analyze” and correctly classify each field of data. If you would like to use the same descriptions we did, download either the “GAMESTATS Data Description.xls” or “PLAYERSTATS Data Description.xls” file from the public SAP Mobile Documents folder to your computer. Click on the “Open Description” file icon, choose “Excel Files” under the “Data Type:” and point to the correct file. Click “OK” to continue.

- a. *Continuous variables are variables whose values are numerical, continuous and sortable. Arithmetic operations may be performed on these values, such as determination of their sum or their mean.*
- b. *Ordinal variables are variables with discrete values, that is, they belong to categories, and they are sortable. Ordinal variables may be numeric or textual.*
- c. *Nominal variables are variables whose values are discrete, that is, belong to categories, and are not sortable. Nominal variables may be numeric or textual.*



The data dictionary for the fields you may utilize within the challenge from the GAMESTATS\_1994\_2014 and PLAYERSTATS\_1994\_2014 tables are as follows.

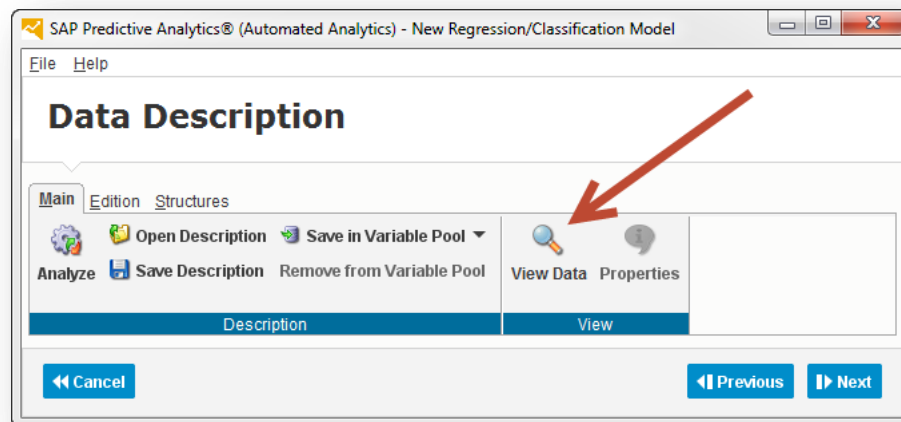
Variable Name	Description	GAMESTATS	PLAYERSTATS
PlayerID	Unique player identifier	x	x
AgeAsOf2015	Age of the player as of January 1st 2015	x	x
Height	Height of the player	x	x
Weight	Weight of the player	x	x
ShootsLeftRight	Indication whether the player shoots with his right / left hand	x	x
SeasonName	Hockey season	x	x



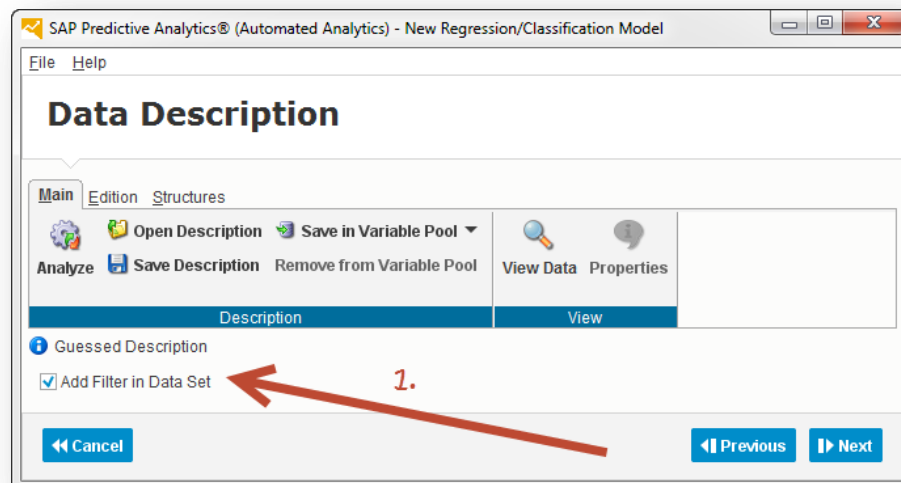
NoOfTeamsPlayed	# of teams the player has been played during his prior DEL career	x	x
NoOfSeasonsPlayed	# of seasons the player has been played during his prior DEL career	x	x
NoOfSeasonsPlayed_Descending	# of seasons the player will still be playing (descending order) until the beginning of season 2014/2015	x	x
NoOfPlayOffsPlayed	# of playoffs the player has been played during his prior DEL career	x	x
Win	Binary variable indicating whether the game has been won or not	x	
Wins	# of wins the player achieved in his DEL history at the beginning of the season 2014/2015		x
Position	Players position on the ice	x	x
Position_Category	Categorization into "Goalie", "Defense", "Offense", "multiple" of the players position on the ice	x	x
AvgGoalsPerGame	Average goals per game which have been shot by the player during his prior DEL career	x	x
AvgAssistsPerGame	Average assists per game provided by the player during his prior DEL career	x	x
AvgPointsPerGame	Average points (1 for goals, 1 for assists) per game achieved by the player during his prior DEL career	x	x
SharePowerPlayGoals	Share of power play goals shot by the player relative to all goals shot by the player during his prior DEL career. A team is said to be on a "power play" when at least one opposing player is serving a penalty, and the team has a numerical advantage on the ice (whenever both teams have the same number of players on the ice, there is no power play).	x	x
ShareShortHandedGoals	Share of short-handed goals (goals made while the opposing team more players on ice than the own team) shot by the player relative to all goals shot by the player during his prior DEL career	x	x
ShareUnassistedGoals	Share of unassisted goals shot by the player relative to all goals shot by the player during his prior DEL career	x	x
PercentGameWinningGoals	Percent of game-winning goals shoot by the player relative to all wins achieved by the team during his prior DEL career	x	x
FaceOffWinRate	Win rate of face offs achieved by the player during his prior DEL career	x	x
AvgPenaltyMinutesPerGame	Average penalty minutes per game attributed to the player during his prior DEL career	x	x
ShareMinorPenalty	Share of minor penalties (less than five minute duration) relative to all penalties attributed to the player during his prior DEL career	x	x
ShareMajorPenalty	Share of major penalties (at least five minutes duration) relative to all penalties attributed to the player during his prior DEL career	x	x
ShareTenMinutePenalty	Share of ten minute penalties (at least five minutes duration) relative to all penalties attributed to the player during his prior DEL career	x	x
ShareShotsOnGoal	Share of shots on goal relative to all shots done by the player during his prior DEL career	x	x
ShotEfficiency	Share of goals shot by the player relative to all shots done by the player during his prior DEL career	x	x
ShootoutEfficiency	Share of goals shot by the player relative to all shootouts the player participated during his prior DEL career	x	x
SaveRate	Share of saves of a goalie relative to all shots on goal during his prior DEL career	x	x
ShootoutSaveRate	Share of saves of a goalie relative to all shootout attempts during his prior DEL career	x	x
GamePlayed_LastSeason	# of games played in prior DEL season	x	x

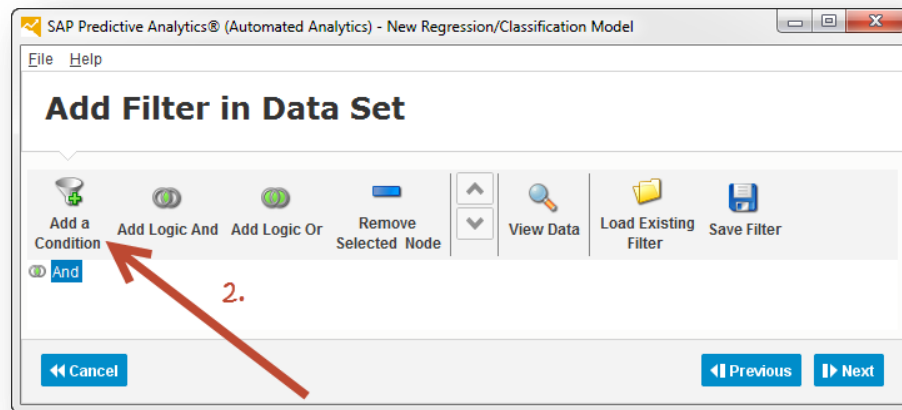
Wins_LastSeason	# of wins the player achieved in the prior DEL season	x	x
WinsHome_LastSeason	# of wins the player achieved in the prior DEL season in his home stadium	x	x

6. **OPTIONAL:** To facilitate accurate classification, click on “View Data” to see a sample of the records.



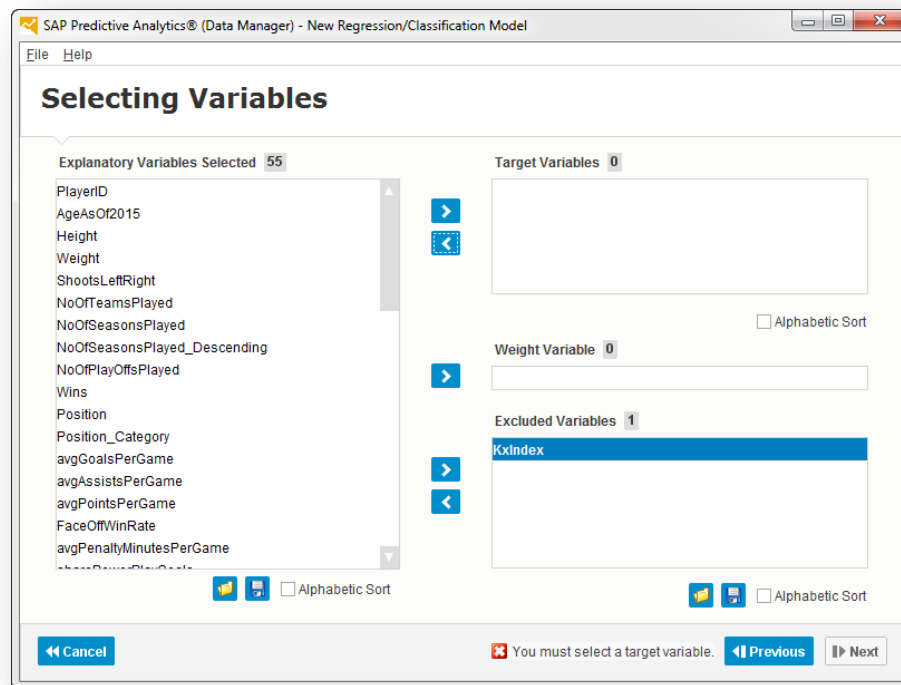
7. **OPTIONAL:** To filter the dataset and select only a subset of the records for the regression model, check the “Add Filter in Data Set” box. Click “Next” to continue and then click the “Add a Condition” button to create a filter condition. (For additional conditions, use the “Add Logic And” or “Add Logic Or” buttons.)



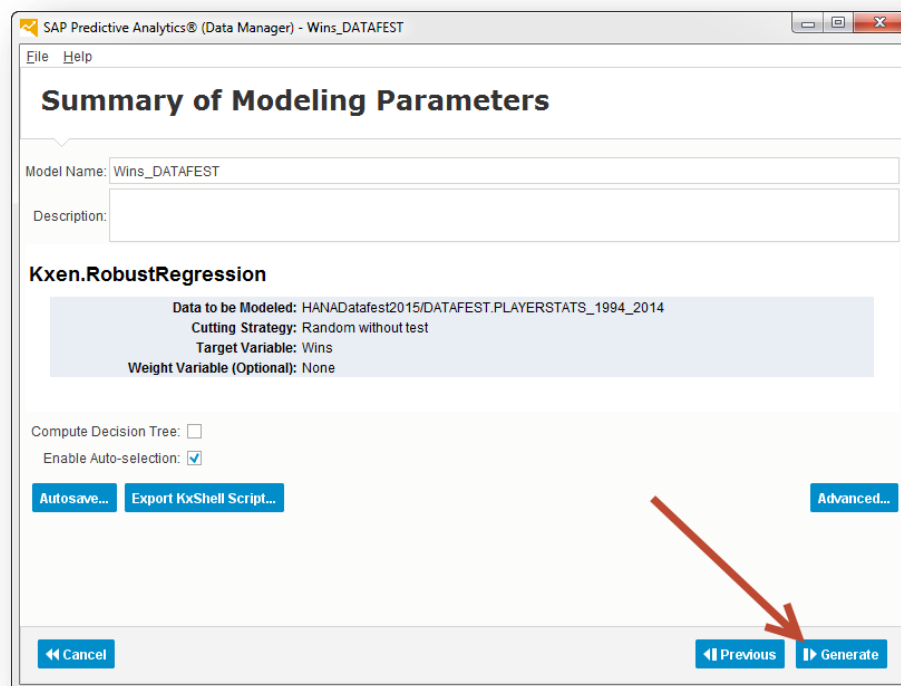


8. Click “Next” to continue.
9. Decide which variables will remain under the “Explanatory Variables Selected” section, meaning those that will serve as independent variables or predictors of the dependent variable. Move variables over to the “Excluded Variables” section to not use them in the model. Finally, move the variable which you are trying to predict or the dependent variable into the “Target Variables” section.

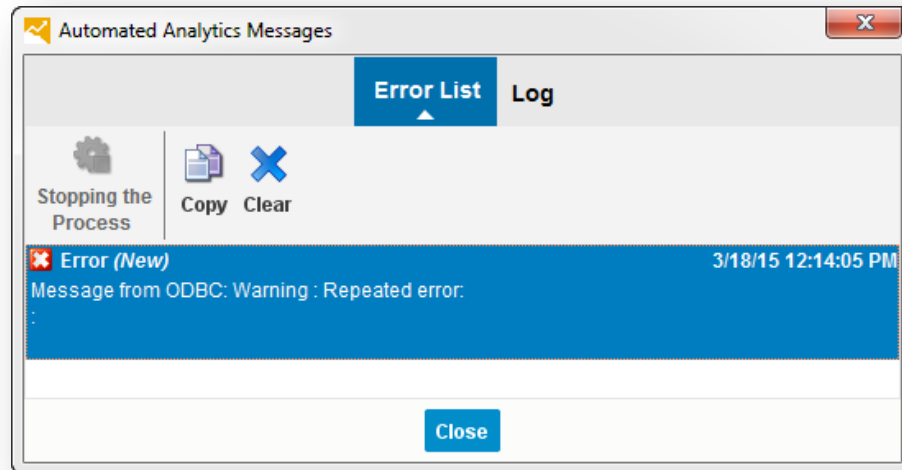
*For this challenge, select the “WINS” variable (which must be a continuous integer) if you are using the “PLAYERSTATS\_1994\_2014” table or the “WIN” variable (which must be a nominal integer, representing win-loss) if you are using the “GAMESTATS\_1994\_2014” table.*



10. Click “Generate” and give the tool a few minutes to train the model. (You can click on “View Type” and then “Log” while the model is being trained.)



11. **OPTIONAL:** If presented with the error message “Message from ODBC: Warning : Repeated error:” or additional warnings, press the “Close” button to continue.



## Analyzing a Model

---

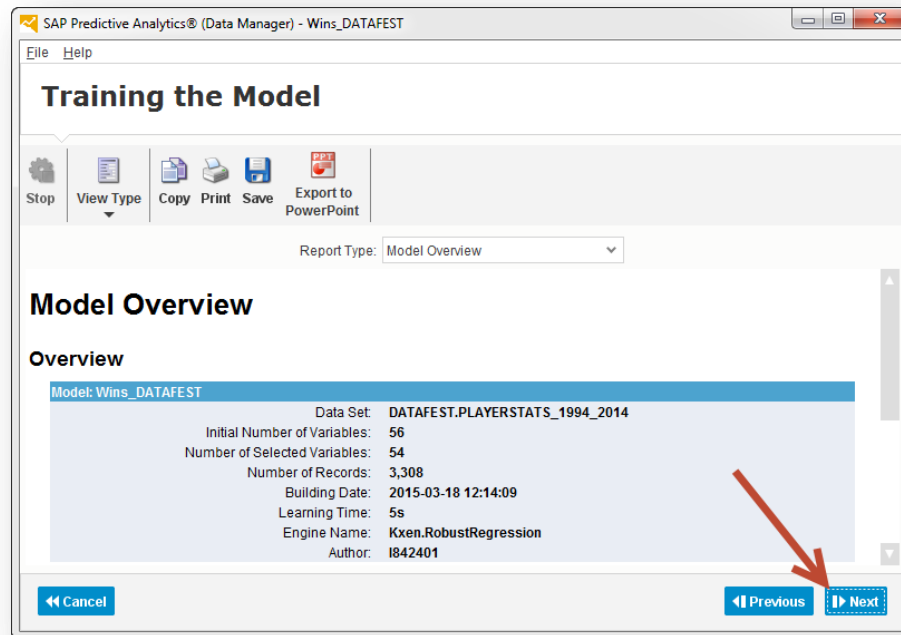
(Estimated time for this section: ~20 minutes)

When you have finished creating a model, you are brought to the “Model Overview”. This and the following screens will let you understand and explore the model in greater detail.

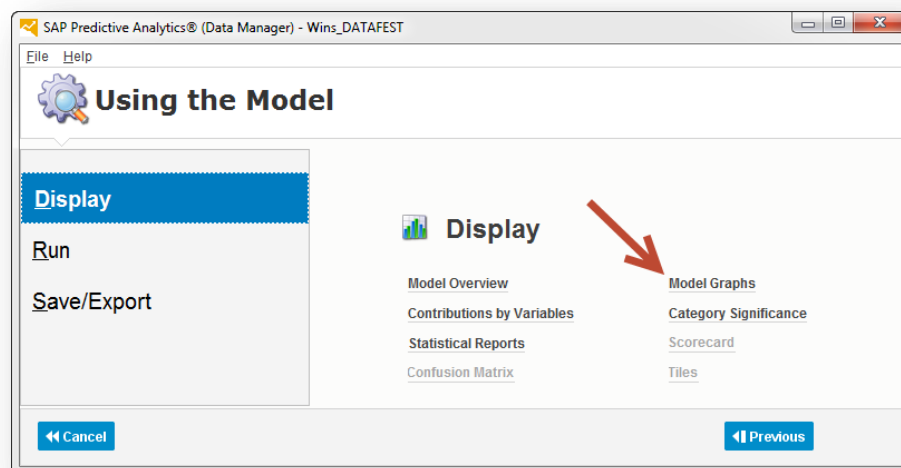
**OPTIONAL:** For more detailed instructions please refer to the “Regression\_Guide” PDF within the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015)).

1. Review the “Model Overview” statistics. Take note of the “Predictive Power (KI)” and the “Prediction Confidence (KR)”. When finished, click “Next” button.
  - a. *The predictive power allows you to evaluate the explanatory power of the model, that is, its capacity to explain the target variable when applied to the training dataset. A perfect model possesses a predictive power of 1 and a completely random model possesses a predictive power of 0. No minimum threshold is required for the predictive power of a model. This depends upon the context of your work, that is, your domain of application, the nature of your data and your business issue. A positive predictive power indicates that the model generated will perform better than a random model.*
  - b. *The prediction confidence defines the degree of robustness of the model, that is, its capacity to achieve the same explanatory power when applied to a new dataset. In other words, the degree of robustness corresponds to the predictive power of the model when applied to an application dataset. A model with a*

*prediction confidence inferior to 0.95 must be considered with caution. The performance of such a model is very likely to vary between the training dataset and the application datasets.*



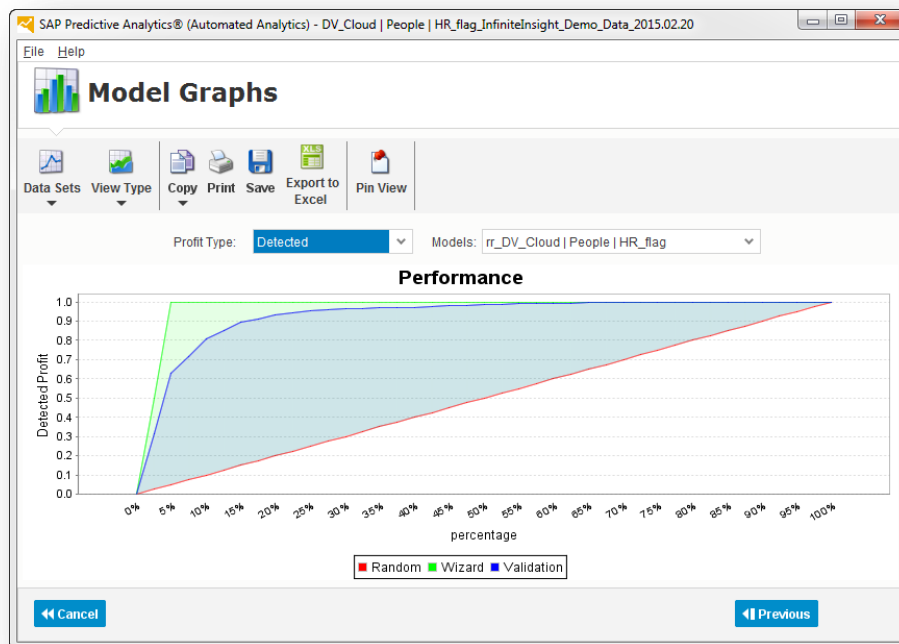
- Click on the “Model Graphs” hyperlink from the “Using the Model” screen.



- Take note of the “Detected Profit” graph or “Predicted Vs. Actual” graph depending upon if you are running a regression with a nominal target (with the GAMESTATS\_1994\_2014 table) or continuous target (with the PLAYERSTATS\_1994\_2014 table). When finished, click on the “Previous” button to return to “Using the Model” screen.

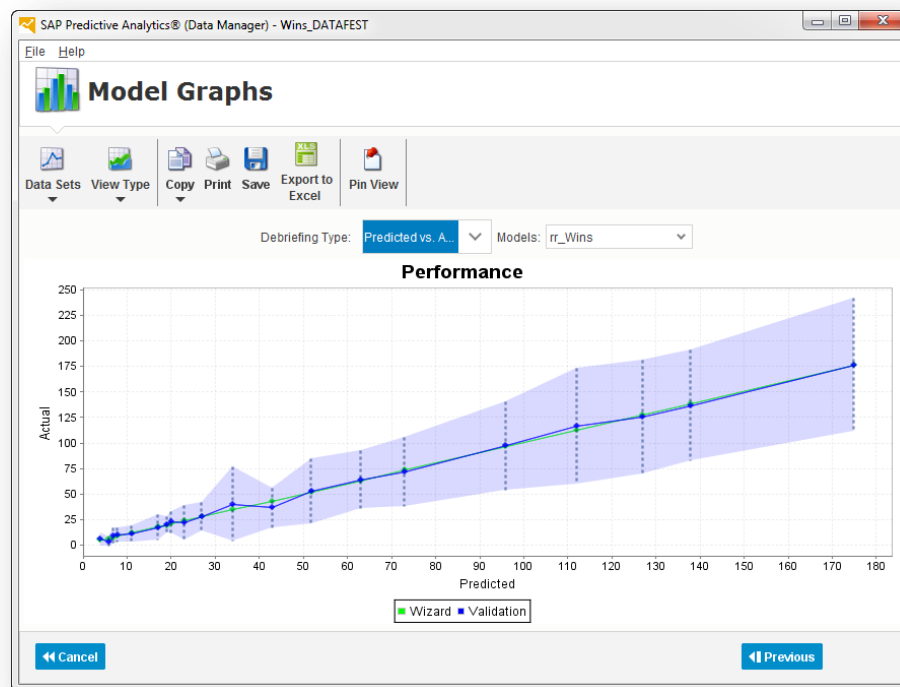
*For a model with a nominal target:*

- a. The default parameters display the profit curves corresponding to the Validation sub-set (blue line), the hypothetical perfect model (Wizard; green line) and a random model (Random; red line).
- b. On the plot, the curves for each type of model represent the profit that may be realized (Y axis), that is, the percentage of observations that belong to the target variable, in relation to the number of observations selected from the entire initial dataset (X axis). On the X axis, the observations are sorted in terms of decreasing "score", that is, the decreasing probability that they belong to the target category of the target variable.

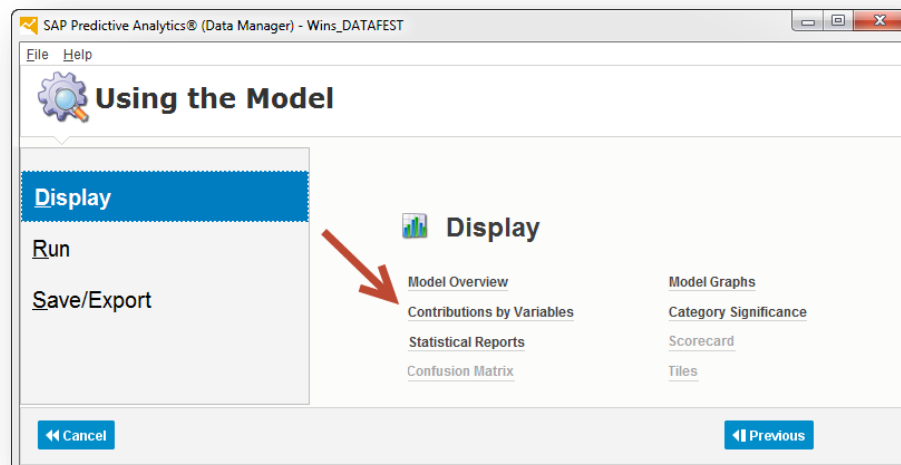


For a model with a continuous target:

- a. The default graphic displays the actual target values as a function of predicted target values. Two curves are displayed: one for the Validation sub-set (blue line) and another for the hypothetical perfect model (Wizard; green line). The Validation curve gives Actual Target value as a function of Predicted Target value.
- b. The blue area represents the expected deviation of the current model.

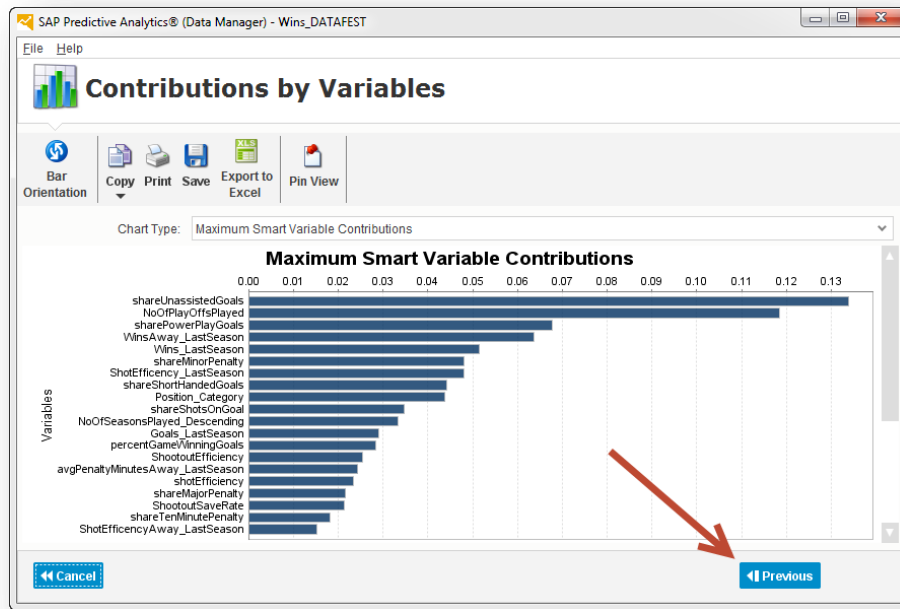


4. Click on the “Contributions by Variables” hyperlink from the “Using the Model” screen.

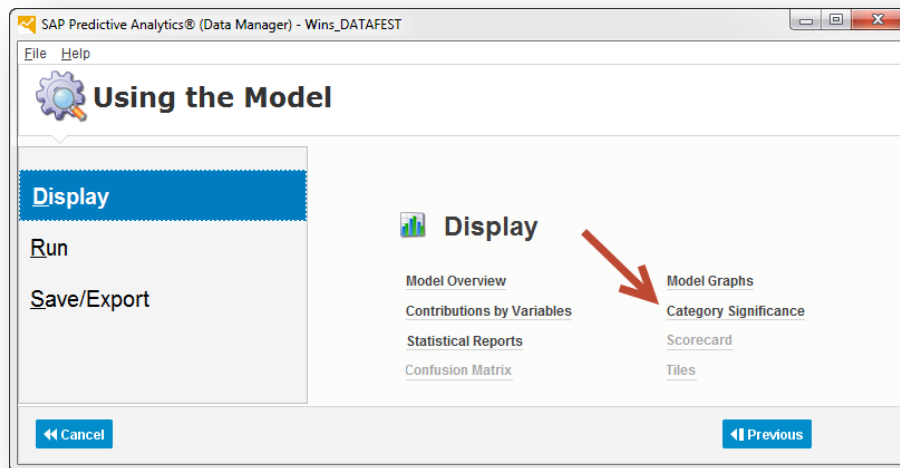


5. Take note of the “Maximum Smart Variable Contributions” chart. When finished, click on the “Previous” button to return to “Using the Model” screen.
  - a. *The Contributions by Variables option allows the user to examine the relative significance of each of the explanatory variables in relation to the target variable. This significance is relative, as the weight of each variable is pro-rated as a function of the significance of the other explanatory variables.*





6. Click on the “Category Significance” hyperlink from the “Using the Model” screen.

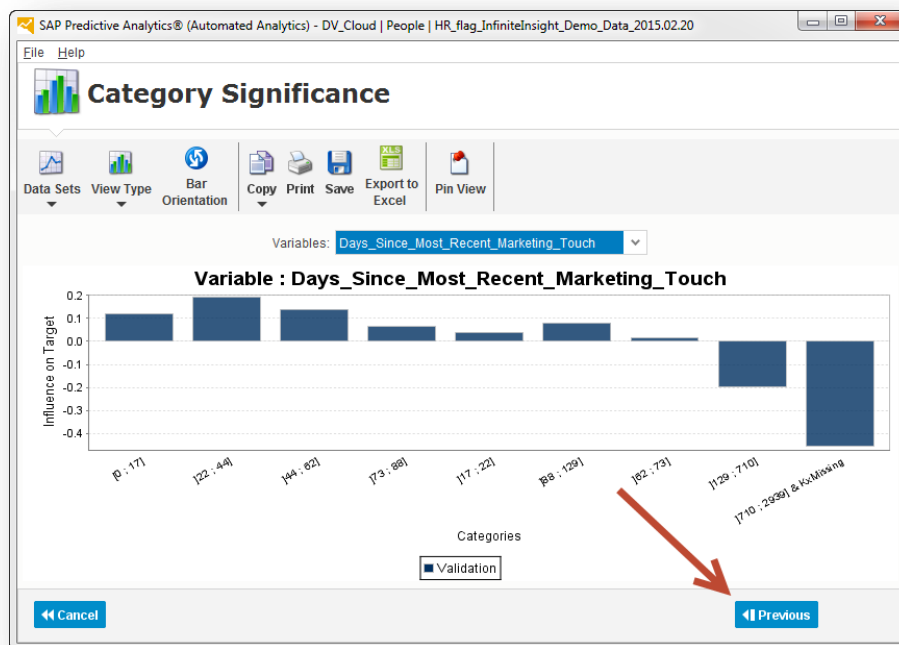


7. Take note of the category significance within each variable. When finished, click on the “Previous” button to return to “Using the Model” screen.

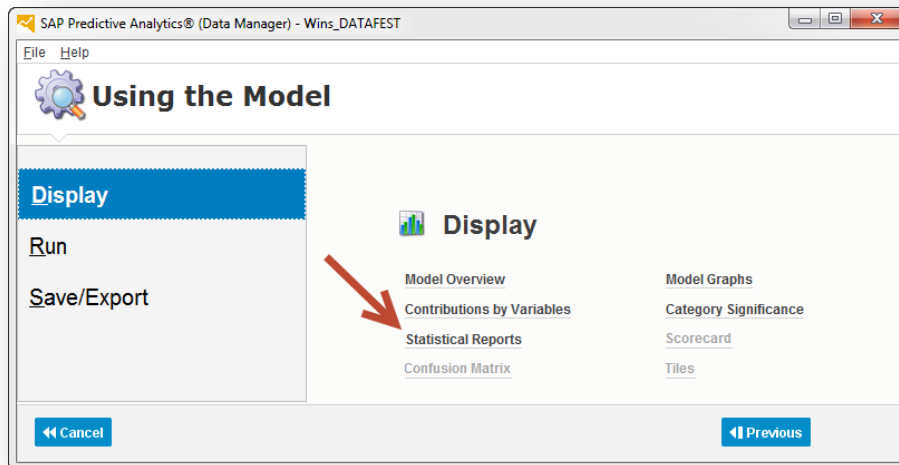
- a. *The Significance of Categories plot illustrates the relative significance of the different categories of a given variable with respect to the target variable.*
- b. *On this type of plot:*
  - i. *The higher on the screen one finds a category, the greater the positive effect on the target category (or hoped-for value) of the target variable. In other words, the higher a category appears on the screen, the more*

*representative that category is of the target category of the target variable.*

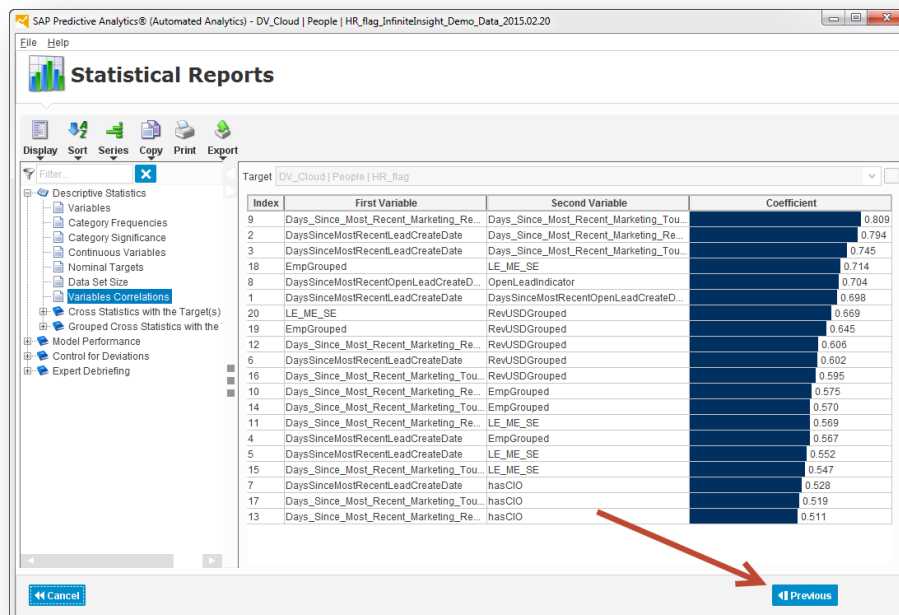
- ii. *The width and direction of the bar correspond to the profit contributed by that category. In other words, they correspond to the relationship of that category to the target variable and whether that category has more or less observations belonging to the target category of the target variable. For a given category, a positive bar (above 0.0) indicates that the category contains more observations belonging to the target category of the target variable than the mean (calculated on the entire dataset). A negative bar (below 0.0) indicates that the category contains a lower concentration of target category of the target variable than the mean.*



8. Click on the “Statistical Reports” hyperlink from the “Using the Model” screen.



9. Take note of the “Variable Correlations” chart. When finished, click on the “Previous” button to return to “Using the Model” screen.
  - a. *To say that variables are correlated implies a certain level of redundancy, that they each contribute some of the same information with respect to the target variable. Two variables said to be highly correlated would describe the same information, or the same concept, to an even greater degree.*



10. **OPTIONAL:** If time allows and depending upon your results you may try to re-run the model to improve its performance by adding or removing variables, running it on the

different dataset from what you selected (either by player or by game), or modifying filters you applied to the dataset to include more records.

## Applying a Model

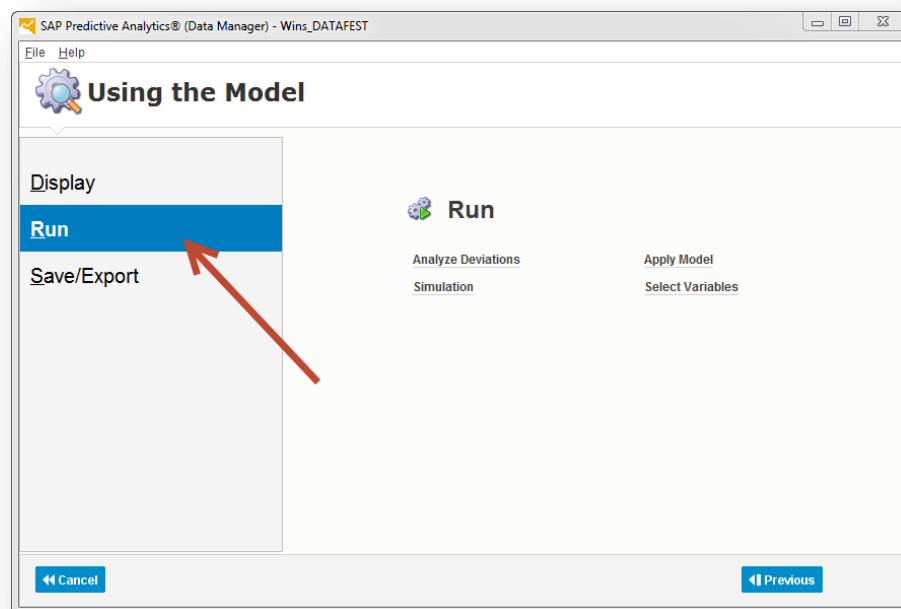
---

*(Estimated time for this section: ~10 minutes)*

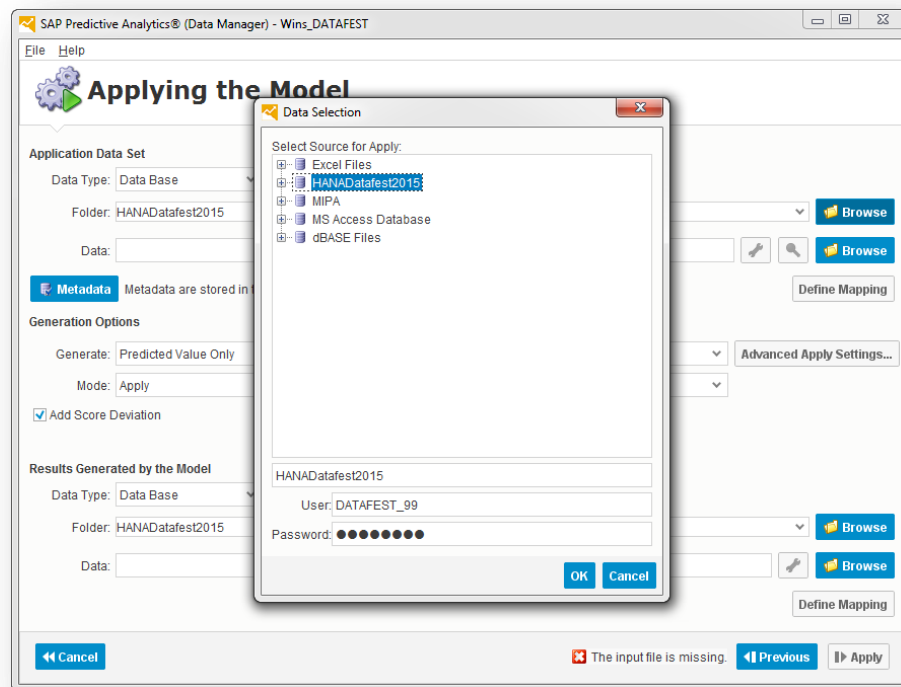
When you are comfortable with the model you have built, you must apply it to the 2014-2015 regular season. The following steps will give you an overview of how to apply the model in SAP Predictive Analytics.

**OPTIONAL:** For more detailed instructions please refer to the “Regression\_Guide” PDF within the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015)).

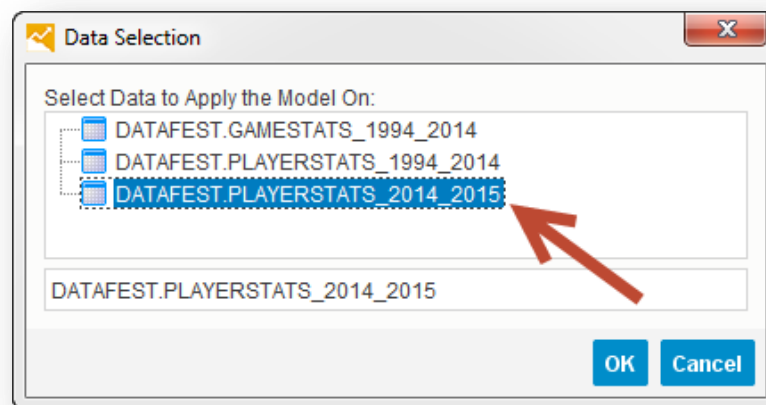
1. To apply the model to the 2014-2015 season, click on “Run” and then the “Apply Model” hyperlink.



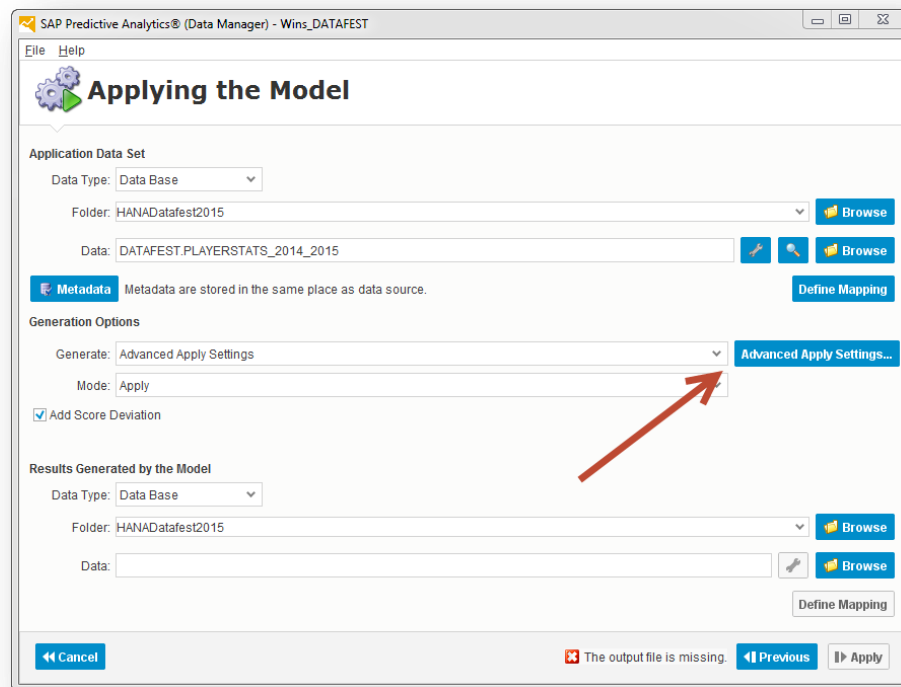
2. Under the “Application Data Set”, set the “Data Type” to “Data Base”. Click on “Browse” next to the “Folder” line and select “HANADatafest2015”. Your username and password will be the same as previously entered. Click “OK” to establish a connection.



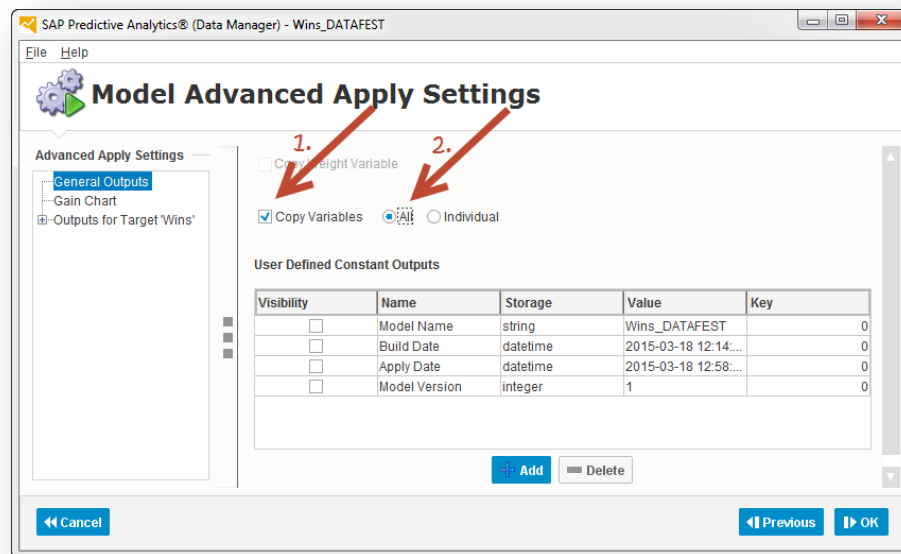
- Click on “Browse” next to the “Data Set” line and select the “PLAYERSTATS\_2014\_2015” table. Click “OK”.



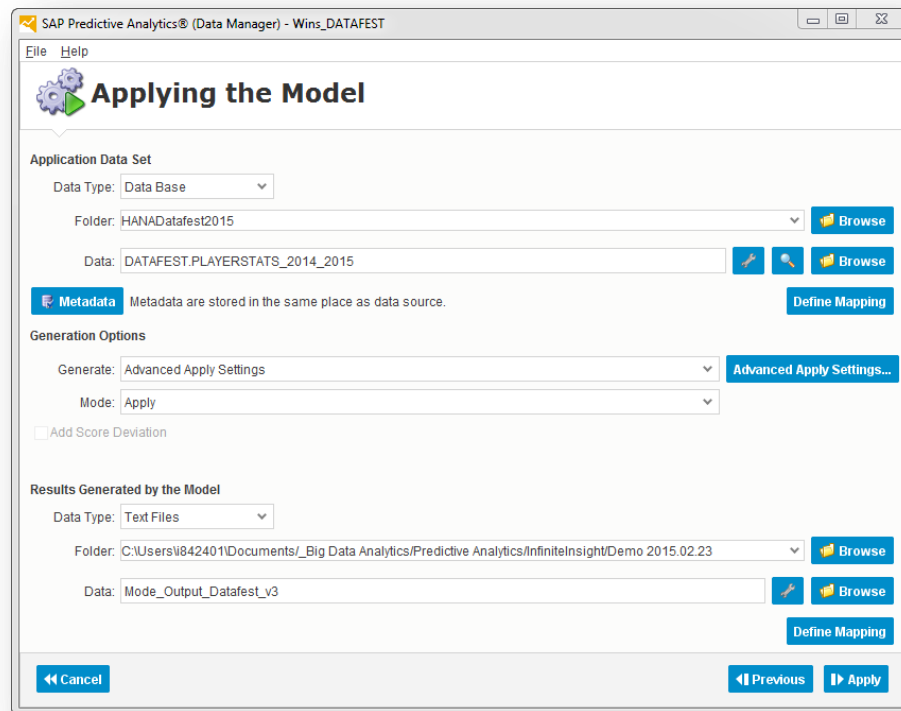
- Under the “Generation Options” section, set “Generate” to “Advanced Apply Settings”, and set “Mode” to “Apply”. Then click on “Advanced Apply Settings...”.



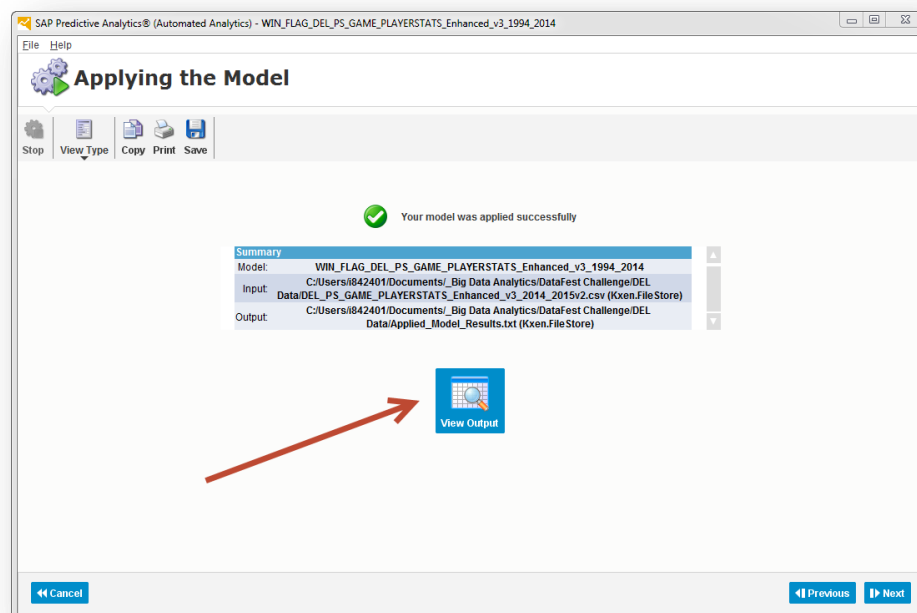
5. Check the box “Copy Variables” and click the “All” button. Press “OK” to continue.



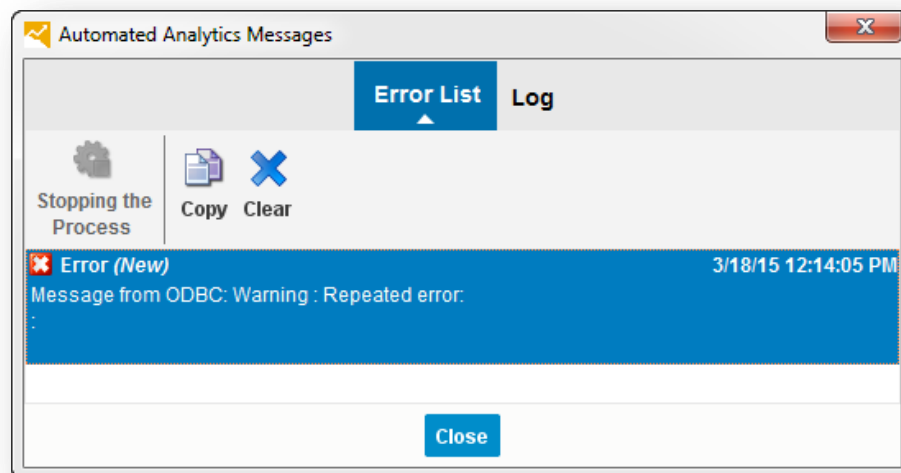
6. Under the “Results Generated by the Model” section, set “Date Type” to “Text Files”, and click on the “Browse” button next to the “Folder” line to select a folder within your computer. Next to the “Data” line, click on the “Browse” button and type in a name for your file and press “OK”.



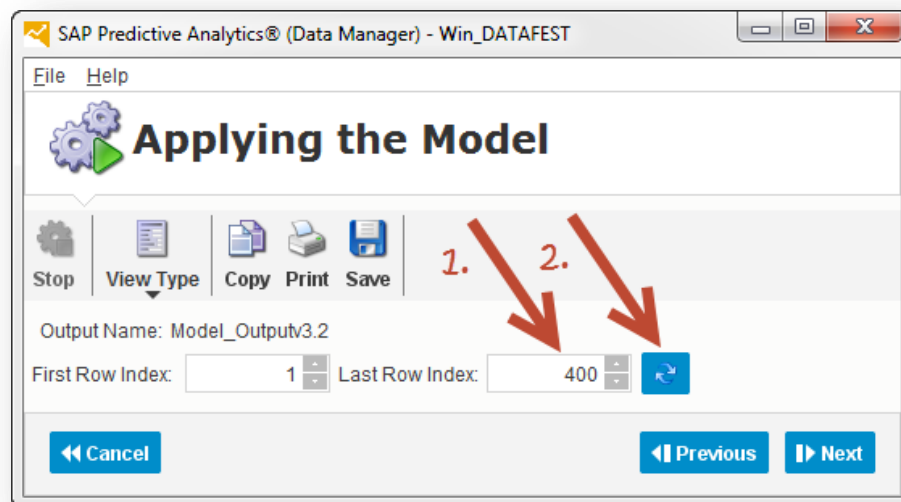
7. Click “Apply” and the model will apply to the new dataset. Then click on the “View Output” icon in the center of the screen.



12. **OPTIONAL:** If presented with the error message “Message from ODBC: Warning : Repeated error:” or additional warnings, press the “Close” button to continue.



8. Change the "Last Row Index" to be 400 records and click the refresh icon.



9. Scroll over to the "rr\_WIN" or "rr\_WINS" field (depending upon the type of regression you are running), and click the field header two times to sort the column descending.



	avgPenalt...	GameWin...	ShotEffice...	ShotEffice...	ShotEffice...	SaveRate...	SaveRate...	SaveRate...	GamePlay...	rr_Wins
1			1	1.14						212
2			3	2.61						204
3			4	2.41						199
4			3	2.29						185
5			6	2.37						182
6			3	1.44						180
7			1	1.29						179
8			1	2.23						171
9			3	1.27						168
10			0	0		0.93	0.93	0.92		166
11			1	1.78						164
12			2	2.15						162

10. Locate the “PLAYERID” field and note the top 6 players who have the most likelihood to win for your final submission.

## Assignment Documentation

(Estimated time for this section: ~15 minutes)

To facilitate comparable results among individuals, please fill out the provided PowerPoint template found in the [SAP Mobile Documents folder](#) (short link: [spr.ly/datafest2015](http://spr.ly/datafest2015)).

Please provide in 1 summary slide and save the file as “DATAFEST\_XX Midnight Challenge Submission” where XX is the number from your username:

1. Information about yourself such as:
  - a. DataFest Username
  - b. Name
  - c. Email address
  - d. Area of study
  - e. Expected graduation date
  - f. **OPTIONAL**: Other interesting facts or interests you would like to mention

2. Brief answers to the following questions (you may want to reference the predictive power, prediction confidence, most or least impactful variables, most or least impactful categories, most or least correlated variables, number model of iterations, etc.):
  - a. What was your methodology and process to create your model?
  - b. What is your rationale and understanding of your model?
  - c. **OPTIONAL**: What other interesting findings would you like to mention?
3. List the top 6 players, by PLAYERID, who have the most likelihood for the most wins. These players can be from any team (for example you could select 6 players from 6 different teams) and can be from any position (for example you could select 6 goaltenders or 6 defensemen).
4. **OPTIONAL**: Screenshots may be submitted on subsequent slides

The screenshot shows a presentation slide with a yellow header bar. The title is "Predicting DEL winning players with SAP Predictive Analytics" and the subtitle is "DataFest 2015 Midnight Challenge Results". Below the title, there is a form with the following fields:

- DataFest Username: <DATAFEST\_XX>
- Name and email address: <FirstName> <LastName>, <Email Address>
- Area of study, expected graduation date: <Area of Study>, <Expected Graduation Date>
- 1. What was your methodology and process to create your model? <Brief Explanation>
- 2. What is your rationale and understanding of your model? <Brief Explanation>
- 3. What other interesting findings would you like to mention? <Brief Explanation>

On the right side of the slide, there is a section titled "Top 6 players chosen:" with a list of 6 players, each with a placeholder <PLAYERID\_X>.

At the bottom right, there is a section titled "For judges only:" with the text "Sum of actual wins of these 6 players in the regular 2014-2015 season:".

At the bottom of the slide, there is a footer with the text "© 2015 SAP SE or an SAP affiliate company. All rights reserved." and "Internal 1".

## Assignment Submission

---

*(Estimated time for this section: ~5 minutes)*

Congratulations! You have worked towards creating, applying, and understanding how to select players from the Deutsche Eishockey Liga who are most likely to have the most wins with SAP Predictive Analytics.

You also have a chance at an internship with SAP on the Big Data Analytics team!

Please email your PowerPoint slide to Thorsten Henrichs ([t.henrichs@sap.com](mailto:t.henrichs@sap.com)) and title the email "DATAFEST\_XX Midnight Challenge Submission" where XX is the number from your username. **OPTIONAL**: Please let us know if you do not want to be considered for an internship. Otherwise, we will consider all submissions as eligible.

We hope to engage with you again soon, thank you for your participation!

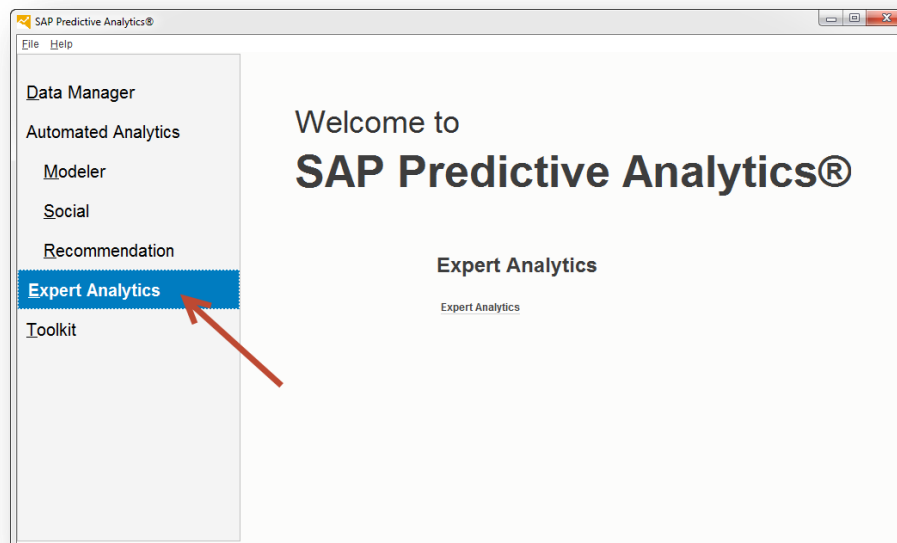
## Additional Resources

---

### Using Expert Analytics (Optional)

None of the questions you need to answer for the final submission require Expert Analytics. However, if you wish to use Expert Analytics to visualize or further explore the data, follow the steps below.

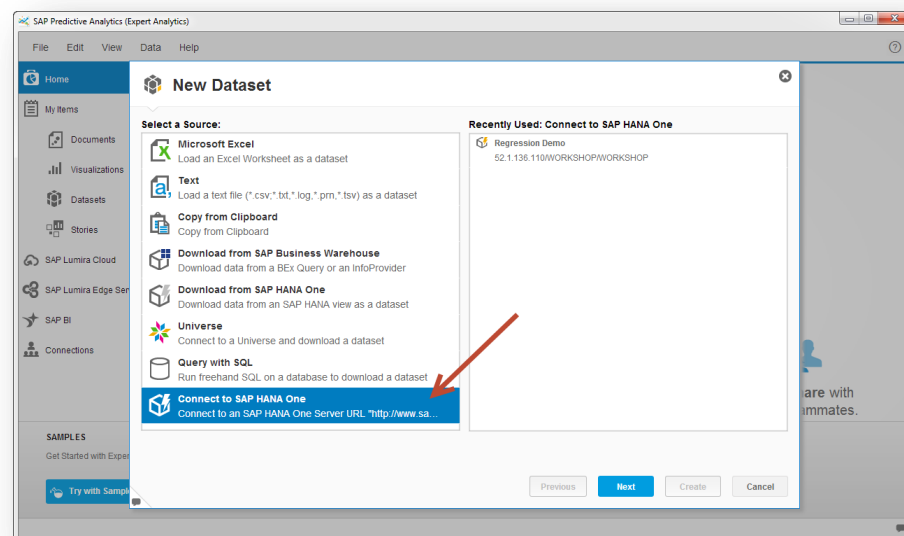
1. Open Predictive Analytics and navigate to Expert Analytics. Click on the “Expert Analytics” hyperlink to launch the tool. If prompted to enter a keycode, type in the following: DB20S-NVDWH38-P7N1WAB-0A0R1HB-EX



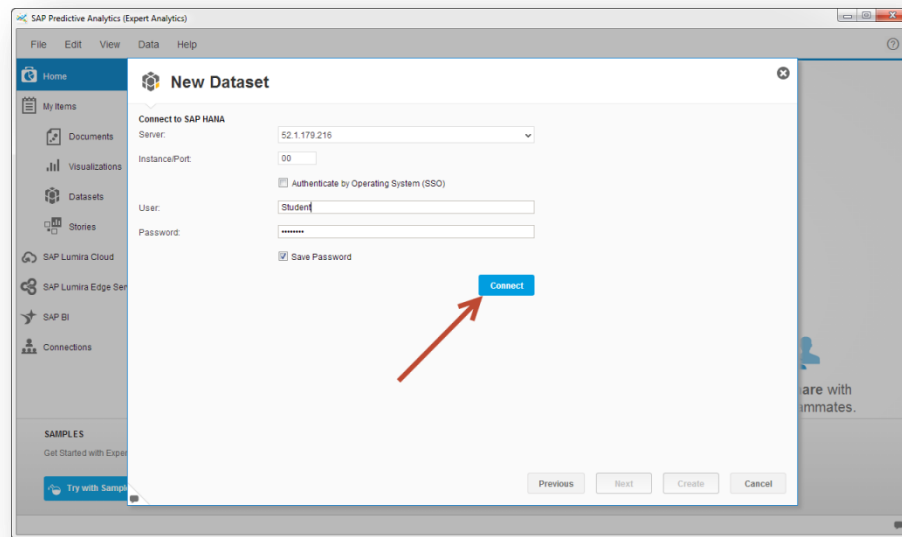
2. In order to use Expert Analytics on the DEL dataset, you will need to connect to the SAP HANA database where the data resides. Click on the blue, circular “Acquire Data” icon.



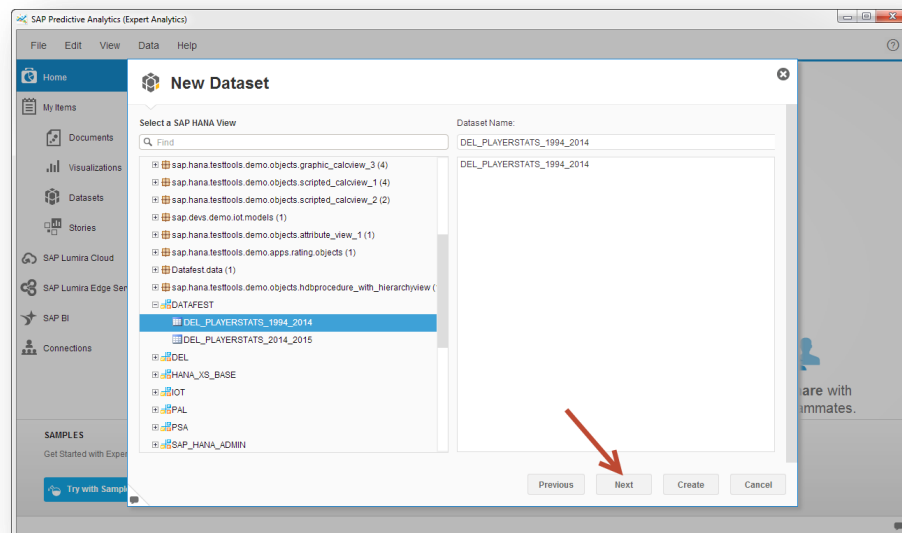
- On the “New Dataset” screen, click on the last item on the list “Connect to SAP HANA One”. Then click “Next” to continue.



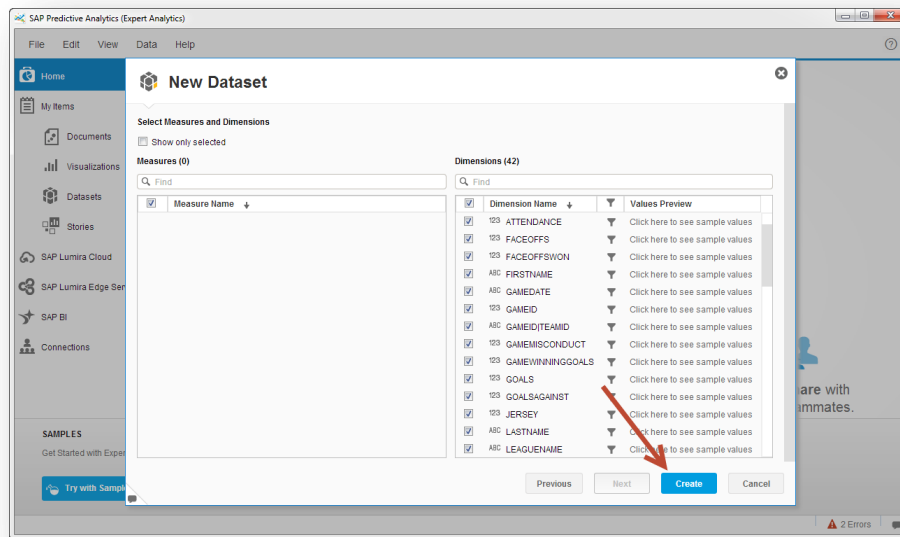
- In the “Server Name” field type in 52.1.179.216. In the “Instance/Port” field, keep the default 00. Do not check the “Authenticate by Operating System (SSO)”. In the “User” field type in DATAFEST\_XX and type in Data1234 in the “Password” field. Check the “Save Password” box. Click “Connect”. (Please see Thorsten Henrichs, the on-site SAP representative, for your username.)



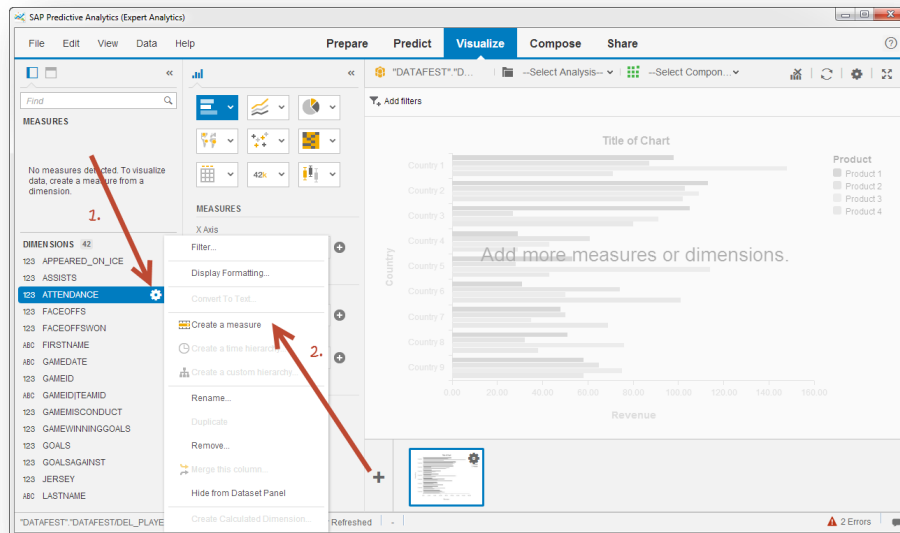
5. Locate the “DATAFEST” view. Click on the “PLAYERSTATS\_1994\_2014” or “GAMESTATS\_1994\_2014” table to ensure the connection is working. Click “Next” to continue. (For more information on the differences between tables, please see the “Creating a Model” section.)



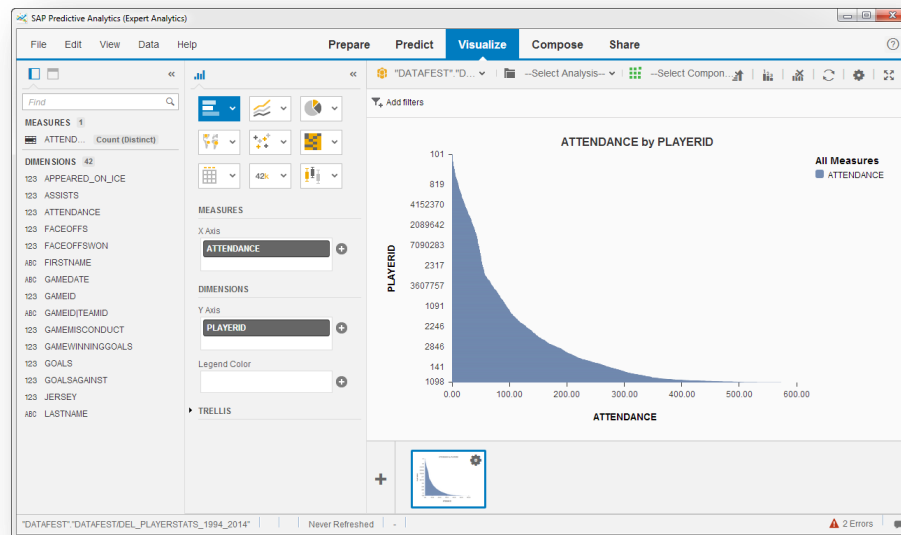
6. Click “Create” and wait a few moments for the data to be brought into Expert Analytics.



7. **OPTIONAL:** After a connection has been made, you will be brought to the “Visualize” tab within Expert Analytics. In order to graphically display some of the fields you may need to change them from being classified as DIMENSIONS to MEASURES. Hover over a field listed under the DIMENSIONS and click on the cog icon for Options. Click on “Create a measure”.



8. Select your visualization from the icons available, drag in appropriate MEASURE(S) and/or DIMENSION(S), and then apply filtering, sorting, or other formatting as desired.



## About SAP

SAP is at the center of today's technology revolution. The market leader in enterprise application software, SAP helps organizations fight the damaging effects of complexity, generate new opportunities for innovation and growth, and stay ahead of the competition.

Visit [our corporate site](#) to explore news, information for investors and analysts, career resources, SAP's history, and more.

## About the SAP Big Data Analytics Team

Our vision is to help the world unleash the potential of big data insights and to propel SAP into the market leader in predictive analytics powered by SAP HANA.

We are a diverse team who consist of not only data scientists but also program managers to ensure the right mix of science, art, operational rigor, and storytelling to drive adoption and advocacy.

We are part of Enterprise Analytics, an organization focused on revolutionizing the creation and consumption of inspiring analytical insights based on breakthrough innovations, thereby laying the foundation for steering excellence at SAP.

---

**www.sap.com**

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

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see <http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark> for additional trademark information and notices. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

