

# **Your First Hire in Predictive Analytics (Hint: It's Not a Data Scientist)**

**Keith McCormick**

The Modeling Agency, LLC

June 17, 2015

# Sponsor



# Speakers



**Keith McCormick**  
Senior Consultant,  
The Modeling Agency, LLC



**Pierre Leroux**  
Director of Product Marketing,  
SAP



THE  
MODELING  
AGENCY

# Your First Hire in Predictive Analytics

---

Hint: It's Not a Data Scientist

---

## **Keith McCormick**

Senior Consultant, TMA  
(281) 667-4200 x233

[keith@the-modeling-agency.com](mailto:keith@the-modeling-agency.com)



## **Pierre Leroux**

Director of Product Marketing, SAP

@pileroux

[pierre.leroux@sap.com](mailto:pierre.leroux@sap.com)

# Keith McCormick

**Keith McCormick** is a highly accomplished professional consultant, mentor, and trainer, having served as keynote and moderator at international conferences focused on analytic practitioners and leadership alike.

Keith has leveraged statistical software since 1990 along with deep expertise utilizing popular industry advanced analytics solutions such as IBM SPSS Statistics, IBM SPSS Modeler, AMOS, Answer Tree, popular open source and other tools involving text and big data analytics.

Mr. McCormick has guided organizations to establish highly effective analytical practices across industries, to include public sector, media, marketing, healthcare, retail, finance, manufacturing and higher education.

As a Senior Consultant with The Modeling Agency, Keith holds a very unique blend of tactical and strategic skill along with the business acumen to ensure superior project design, oversight and outcomes that align with organizational targets.







# What We'll Cover

- The Data Scientist
- The PA Project
- Project Roles
- The Solution





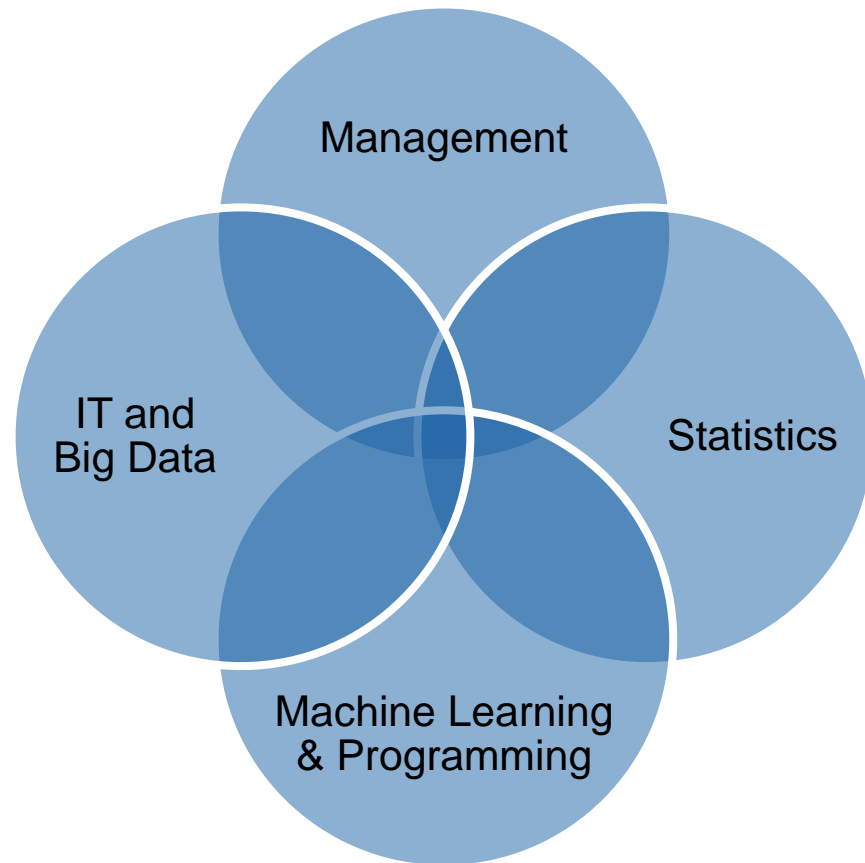
THE  
MODELING  
AGENCY

# Why 'Data Scientist'?

Who Are They, What Can They Do,  
And Where Do You Get One?



# Managing Data Scientist







# Where Do You Find Them?

## Curriculum

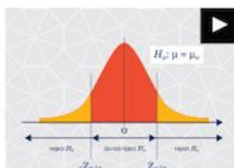
The Master of Information and Data Science is designed to be completed in 20 months, but other options are available to complete the program on an accelerated basis. You will complete 27 units of course work over three to five terms. Courses are divided into foundation courses (15 units), advanced courses (9 units), and a synthetic capstone (3 units). You will also complete an immersion at the UC Berkeley campus.

### Foundation Courses (15 units)



Research Design  
and Application for  
Data and Analysis

3 UNITS



Exploring and  
Analyzing Data

3 UNITS



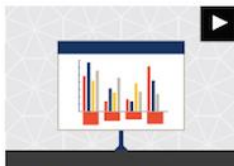
Storing and  
Retrieving Data

3 UNITS



Applied Machine  
Learning

3 UNITS



Visualizing and  
Communicating  
Data

3 UNITS

### Advanced Courses will include (9 units)



Experiments and  
Experimentation  
with Data

3 UNITS



Legal, Policy, and  
Ethical  
Considerations for  
Data Scientists

3 UNITS



Really Big Data:  
Scaling up and  
Parallelism

3 UNITS

### Capstone Course (3 units)



Synthetic  
Capstone Course

3 UNITS

Request more information about [datascience@berkeley](mailto:datascience@berkeley), or speak with an admissions counselor at 855-678-MIDS.

<sup>1</sup> Students may take a maximum of 9 units each term.<sup>†</sup>

Learn more about [datascience@berkeley](mailto:datascience@berkeley)



# Data Science 'Unicorn' vs. Teams

**Data Science tasks need a rare combination of statistics, hacking, database, business, presentation, and other skills, but finding versatile ("unicorn") data scientists with all the required skills is hard. Which approach is better: [304 votes total]**

Individual: Seek and Train versatile Data Scientists that have all (or most) of needed skills (135)

44.4%



Not sure (33)

10.9%

Team: Build a data science team where each member mainly focuses on one skill (136)

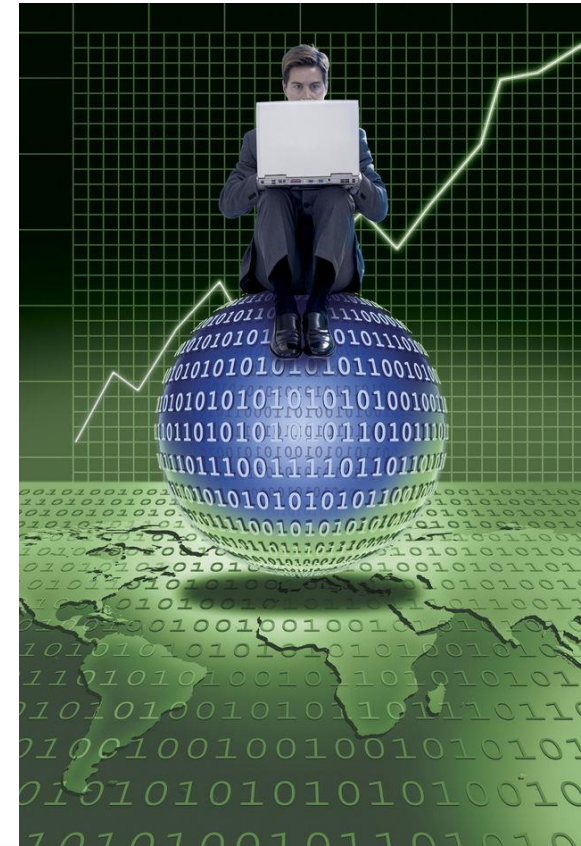
44.7%

<http://www.kdnuggets.com/2014/01/split-on-data-science-skills-individual-vs-team-approach.html>



# Data Science Salaries

- O'Reilly '14: \$98,000 Int'l Median Salary
- USA only: 105k Base and 144k Total
- 12% were over \$200,000
- Glassdoor: \$118,000 Average Salary
- Burtch Works: \$91,000 for 'junior level'
- \$250,000+ for management with 50k+ in bonuses







THE  
MODELING  
AGENCY

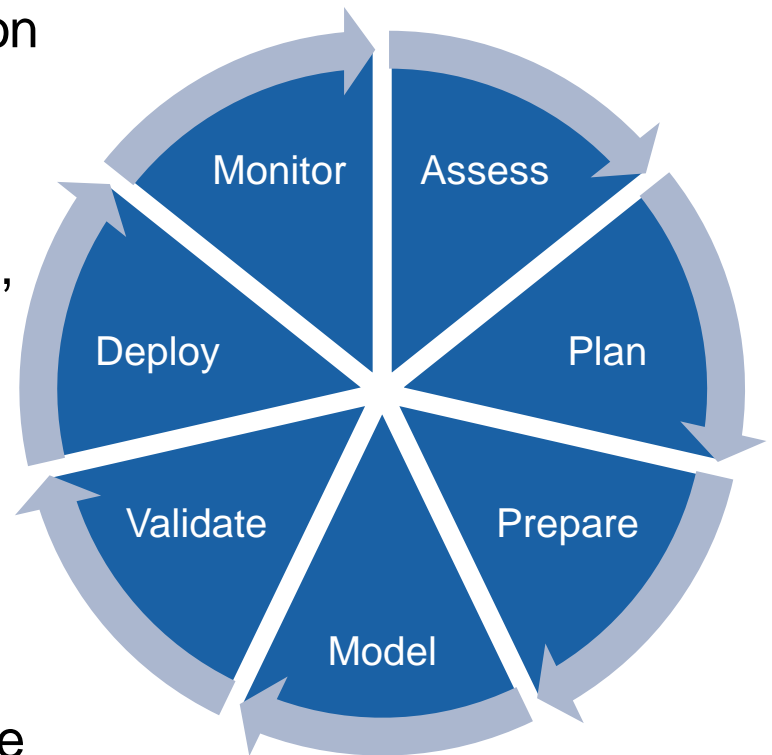
# What is Predictive Analytics?

What Does a PA Project Look Like?



# TMA's Modeling Practice Framework™

- **Assess:** Assemble team; gather information
- **Plan:** Analyze resources, situation and environment; define the project
- **Prepare:** Select, merge, manipulate, clean, construct and transform the data
- **Model:** Build Predictive Models
- **Validate:** Verify that the model solved the problem and will hold up in production
- **Deploy:** Put the solution into action
- **Monitor:** Evaluate performance and update process as needed

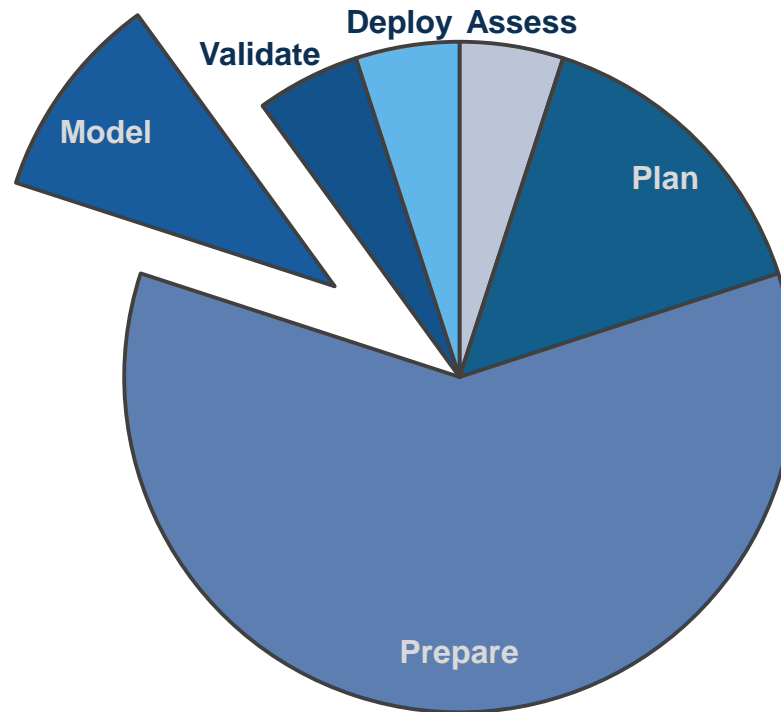


TMA Modeling  
Practice Framework™



# First Project Effort Allocation

*Modeling Is Only About 10%*





# Solo Sr. Data Scientist's First Year



- Managing
- Modeling
- Evangelizing
- Hiring
- Cleaning
- Learning



THE  
MODELING  
AGENCY

# Predictive Analytics Team Roles

What Will Your Eventual Team Look Like?





# Project Roles



# Project Role 'Hats'

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



# Project Roles You've Got Covered

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		

# Potential Existing Assets?

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



# Potential Existing Assets – Phase 2

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		

# Project Roles You Probably Need

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



THE  
MODELING  
AGENCY

# The Solution

The Cross Functional Team  
*Promote from Within*



# Phase 1

## Hire to Replace from Below

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



# Find a Trainer and Mentor

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



## Phase 2

### Promote and/or Hire

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



# Long Term Full Time Team

	Support	Core Team	Management
<b>Project Sponsorship</b>		Internal Customer	Senior Sponsor
<b>SME</b>	End User		Process Management
			Process Expert
<b>Data</b>	Data Steward	Data Prep	IT Management
<b>Modeling</b>		Modeler	Architect
			Analytics Management
<b>Deployment</b>	Interface Support	Deployment Lead	Interface Management
<b>Monitor</b>	Model Manager		



THE  
MODELING  
AGENCY

# Wrap Up





# Primary Takeaways

## Your First Hire in Predictive Analytics

- Predictive Analytics is always a team effort.
- Your existing team already has some of the necessary skills, and your existing team already understands your business.
- Data Scientists have valuable and rare skills – utilize them primarily for those skills. Don't make them a team of one. They won't stay or be effective.
- Your projects, especially in the first year, do not require the rare skills on a daily basis.
- Grow your team's skills by: Promoting from within, Replacing from below, Cross-training and Mentoring, and 'Rent' Data Science skills until you need them full time.

# Speak Up!...

## Discussion Session

**Enter your questions now...**

to obtain tailored advice from the experts!

---

## Attend Keith's TDWI Seminar

**Serious Play for Predictive Analytics**

*What Works, What Doesn't & Why*

**TDWI's "The Analytic Experience"**

*July 30, 2015 | Boston*



# SAP Predictive Analytics


Confidently anticipate and drive better business outcomes

June 17, 2015



Pierre Leroux  
Director of Product Marketing

# SAP Predictive Analytics Demo

 SAP Predictive Analytics®

File Help

Data Manager

Automated Analytics

**Modeler**

Social


Recommendation

Expert Analytics

Toolkit

Welcome to

**SAP Predictive Analytics®**

 **Modeler**

[Create a Classification/Regression Model](#)

[Create a Clustering Model](#)

[Create a Time Series Analysis](#)

[Create Association Rules](#)

[Load a Model](#)



SAP Predictive Analytics® (Automated Analytics) - New Regression/Classification Model

File Help

Select a Data Source

☒ Use a File or a Database Table ☐ Use Data Manager

Data Type: Text Files

Folder: C:\Users\i818552\Documents\Predictive\Demos\Demo Predictive

Data Set: Promo01.csv

Cutting Strategy...

Browse

Browse

Metadata

Metadata are stored in the same place as data source.

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

Public

33

SAP Predictive Analytics® (Automated Analytics) - New Regression/Classification Model

File Help

Selecting Variables

Explanatory Variables Selected 13

age  
workclass  
fnlwgt  
education  
education-num  
marital-status  
occupation  
relationship  
sex  
capital-gain  
capital-loss  
hours-per-week  
native-country

Target Variables 1

Risk\_Leaving


☐ Alphabetic Sort


Weight Variable 0


Excluded Variables 1


KxIndex


Training the Model



Stop


View Type


Copy


Print


Save


Export to PowerPoint

Report Type: Model Overview

Model Overview

Overview

Model: Risk-Leaving_Promo01		
	Data Set:	Promo01.csv
	Initial Number of Variables:	15
	Number of Selected Variables:	13
	Number of Records:	48,842
	Building Date:	2015-06-15 13:37:40
	Learning Time:	11s
	Engine Name:	Kxen.RobustRegression
	Author:	I818552

Nominal Targets

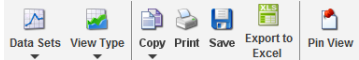
Risk-Leaving		
	Target Key	1
	0 - Frequency	76.05%
	1 - Frequency	23.95%

Selection Process Selected Iteration

3		
	Predictive Power (KI)	0.8085
	Prediction Confidence (KR)	0.9959
	Nb. Variables Kept	7

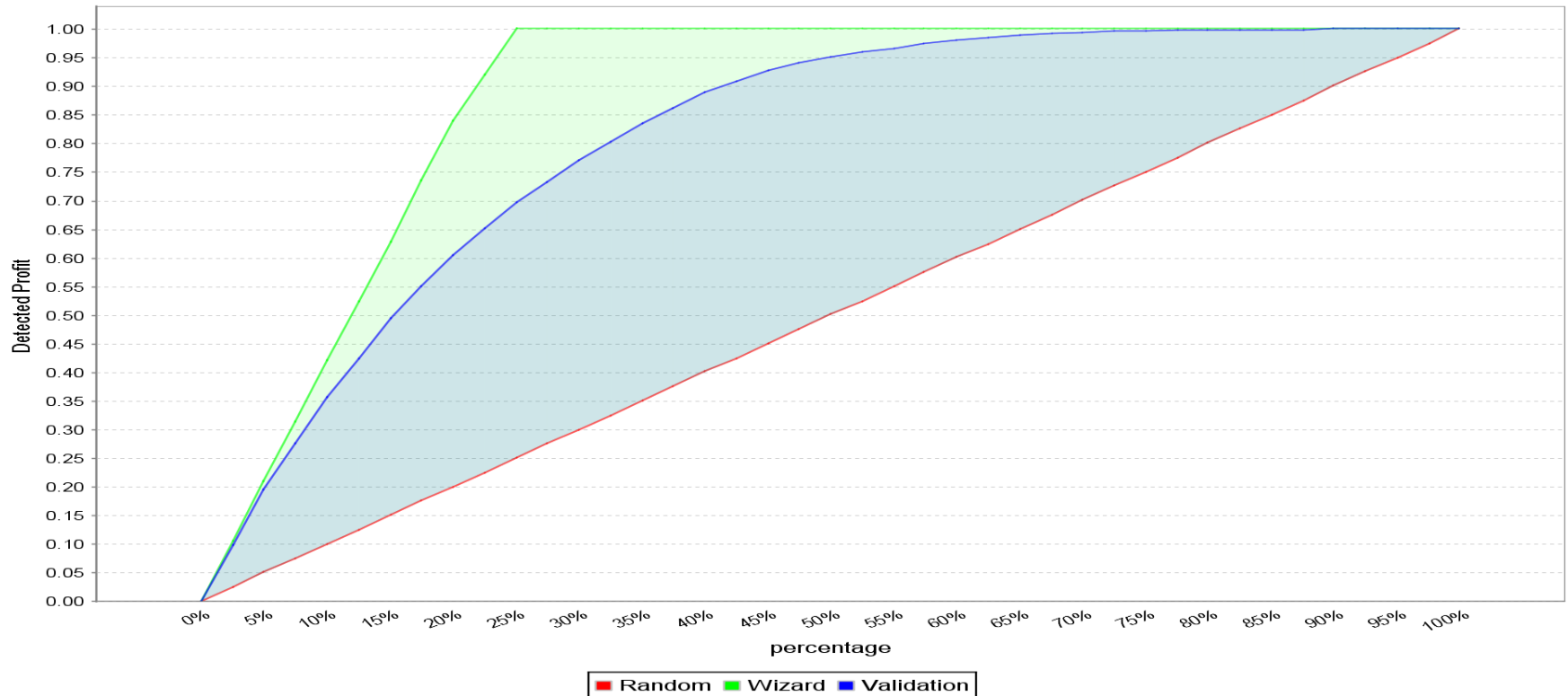


## Model Graphs

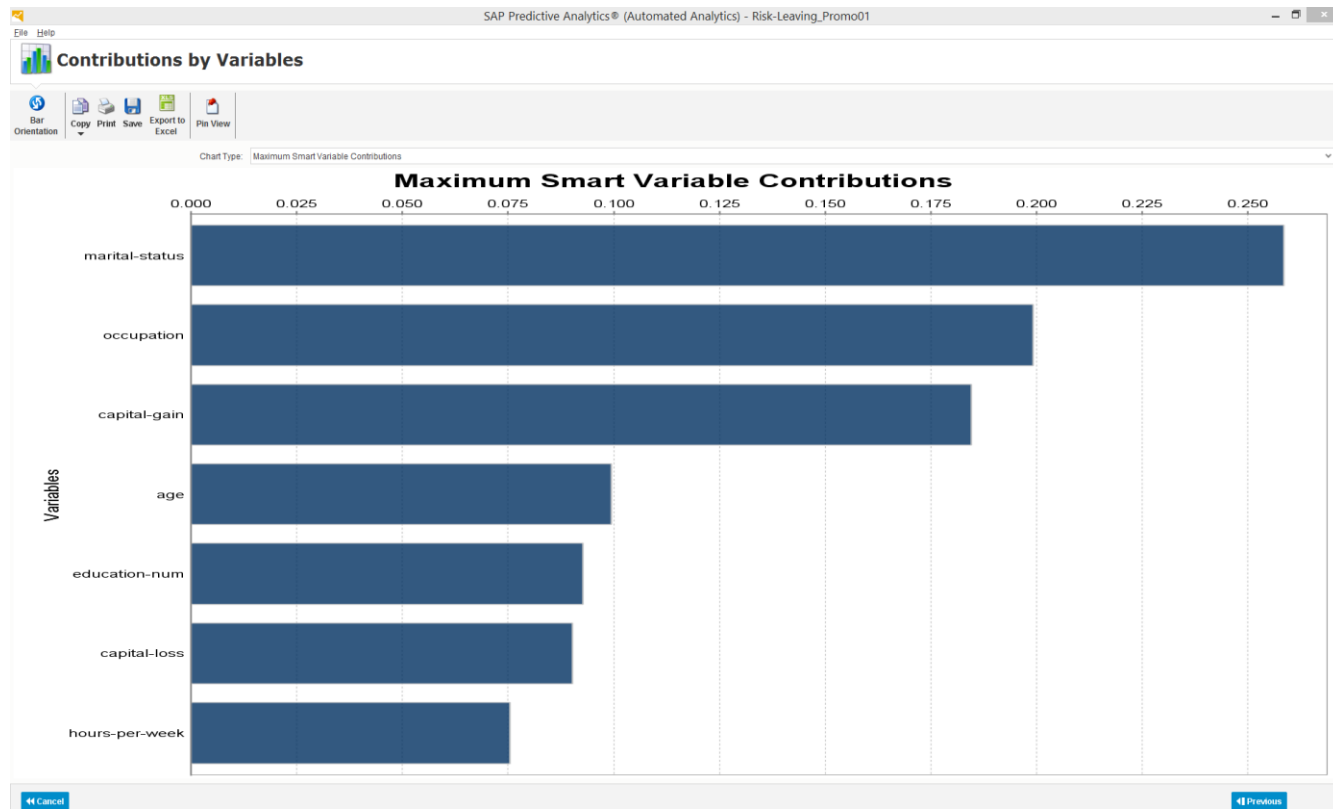


Profit Type: Detected Models: rr\_Risk-Leaving

## Performance








SAP Predictive Analytics® (Automated Analytics) - Risk-Leaving\_Promo01

File Help

 **Generating Code**

Standard Settings

Advanced Settings


Target to be used: Risk-Leaving Information to be generated: Score/Estimates


Code Settings

Code type: HTML(Javascript)

Note:  
The results are the same as when using Automated Predictive Kernel.

Output

Folder: C:\Users\l818552\Documents\Predictive\Demos\Demo Predictive 

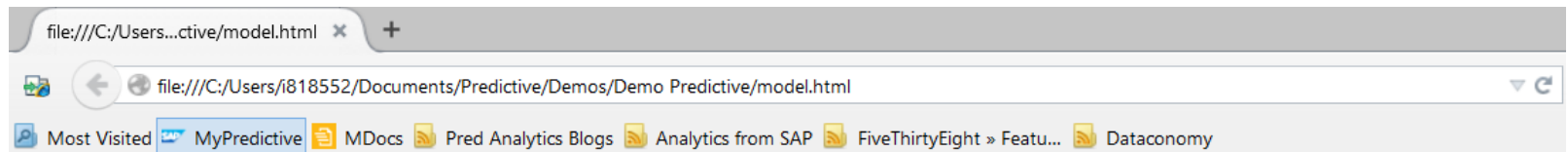
Generated File: .html 

☐ View Generated Code

Cancel

Previous

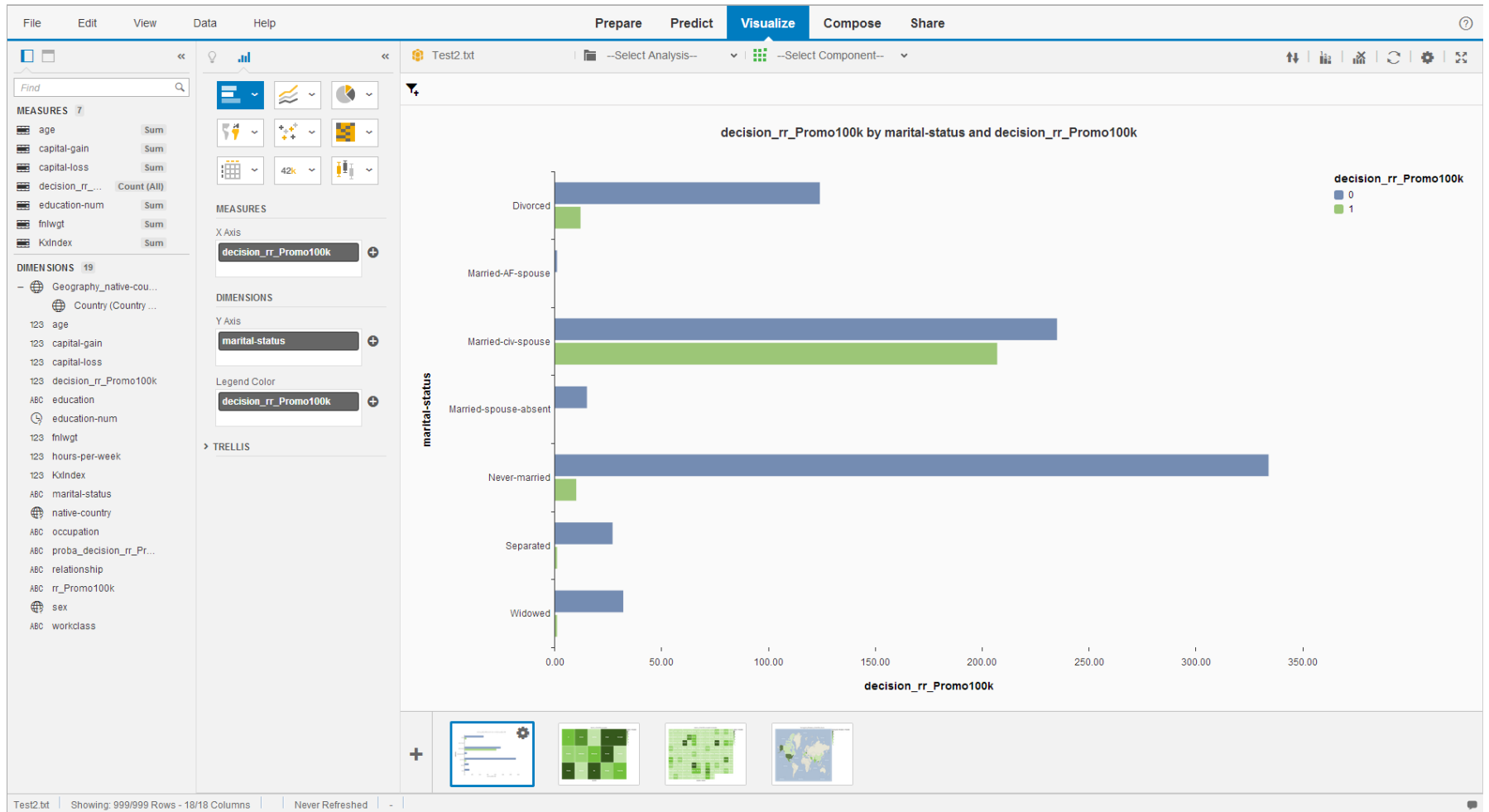
Generate



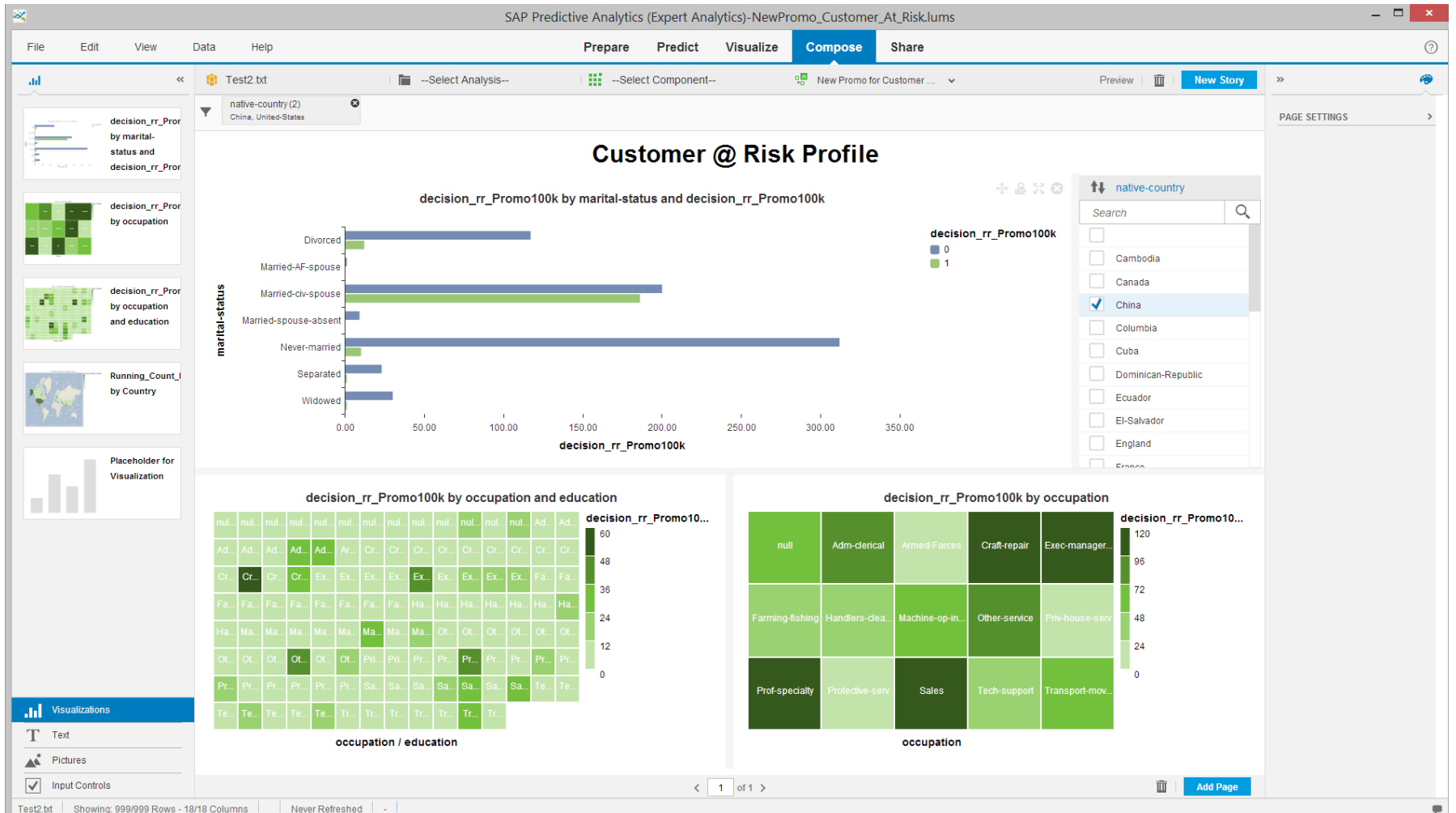
Model	Transform
Risk-Leaving_Promo01	Kxen.RobustRegression

Variable	Value	Score
age	45	0.02466372878336169
education-num	[12] ▾	0.07913405019338772
marital-status	Never-married ▾	-0.13378966182623014
occupation	KxOther ▾	-0.07673808968646559
capital-gain	15000	0.2877036426127835
capital-loss	0	-0.018870197925081364
hours-per-week	50	0.03655246370521886

Target	Value
<a href="#">Risk-Leaving</a>	0.19865593585697466







# Next Steps

- 1 Read customers case studies**  
[www.sap.com/predict-and-me](http://www.sap.com/predict-and-me)
- 2 Learn more and watch it in action**  
[www.sap.com/predictive](http://www.sap.com/predictive) & <http://scn.sap.com/docs/DOC-32651>
- 3 Try it!**  
[www.sap.com/trypredictive](http://www.sap.com/trypredictive)

# THANK YOU

**Pierre Leroux**

**@pileroux**



[www.sap.com/predictive](http://www.sap.com/predictive)



[scn.sap.com/community/predictive-analytics](http://scn.sap.com/community/predictive-analytics)



#sappredictive • @sapanalytics

# Questions?



# Contact Information

If you have further questions or comments:

Keith McCormick, The Modeling Agency  
[keith@the-modeling-agency.com](mailto:keith@the-modeling-agency.com)

Pierre Leroux, SAP  
[pierre.leroux@sap.com](mailto:pierre.leroux@sap.com)