SAS® FORUM
UNITED KINGDOM 2015

SAS Factory Miner and Next Generation SAS Forecasting

John Spooner

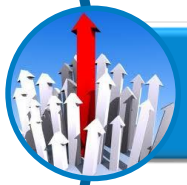
THE POWER TO KNOW.

Agenda

- Trends in the analytics space
- SAS Factory Miner
- SAS Forecast Server Client

Market challenges

Market opportunities & risks – volatile and fragmented



Time to market is crucial for the business

Digital economy, Internet of things are rising

Predictive Analytics challenges

Model proliferation
With targeted segmentation and micromarketing campaigns

Less time
to (re)act and differentiate on new market changes

Big Data
High volume, variety and velocity of big data

Examples & facts

CISCO

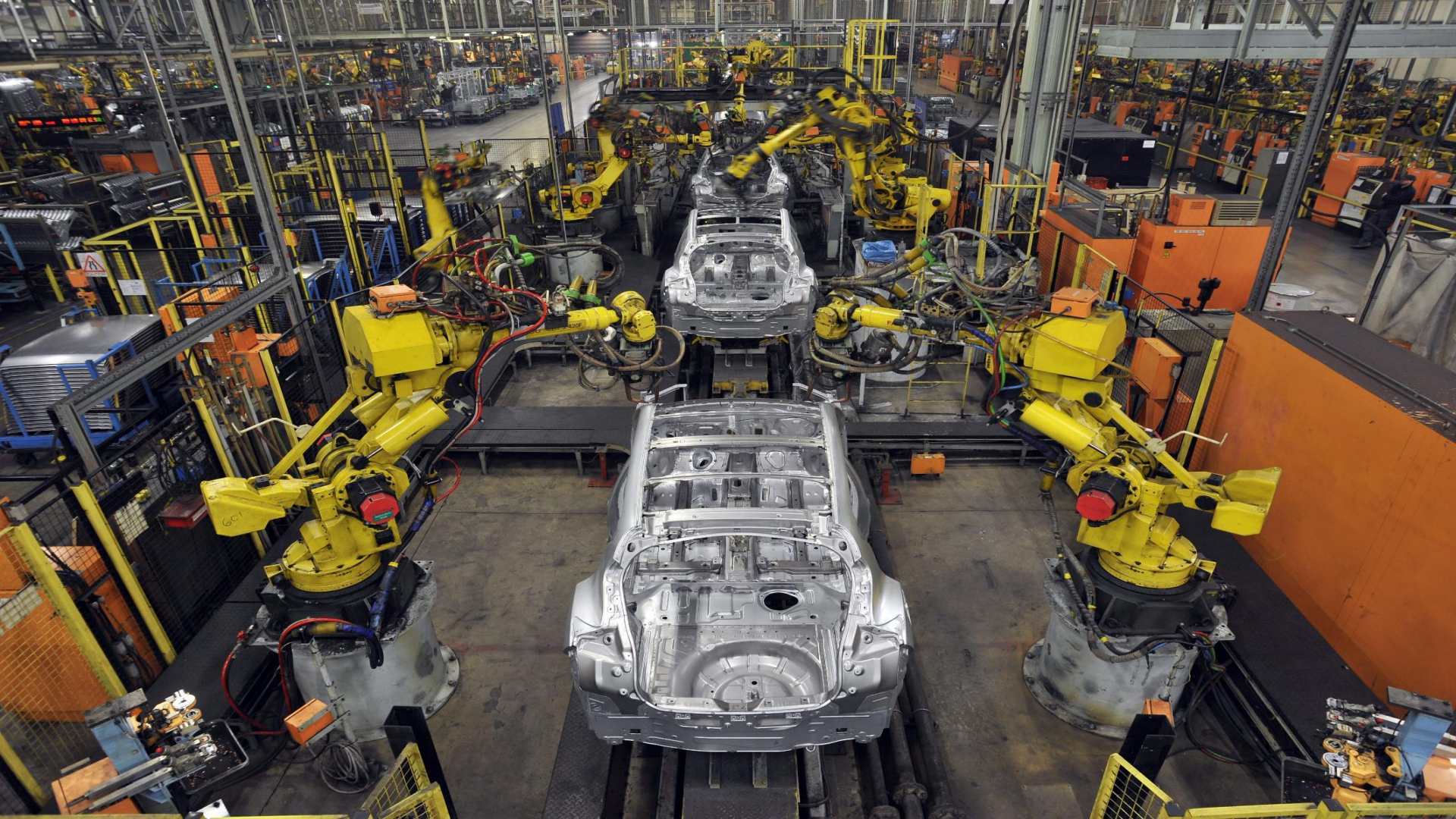
35,000 models for marketing campaigns to manage & monitor

HSBC

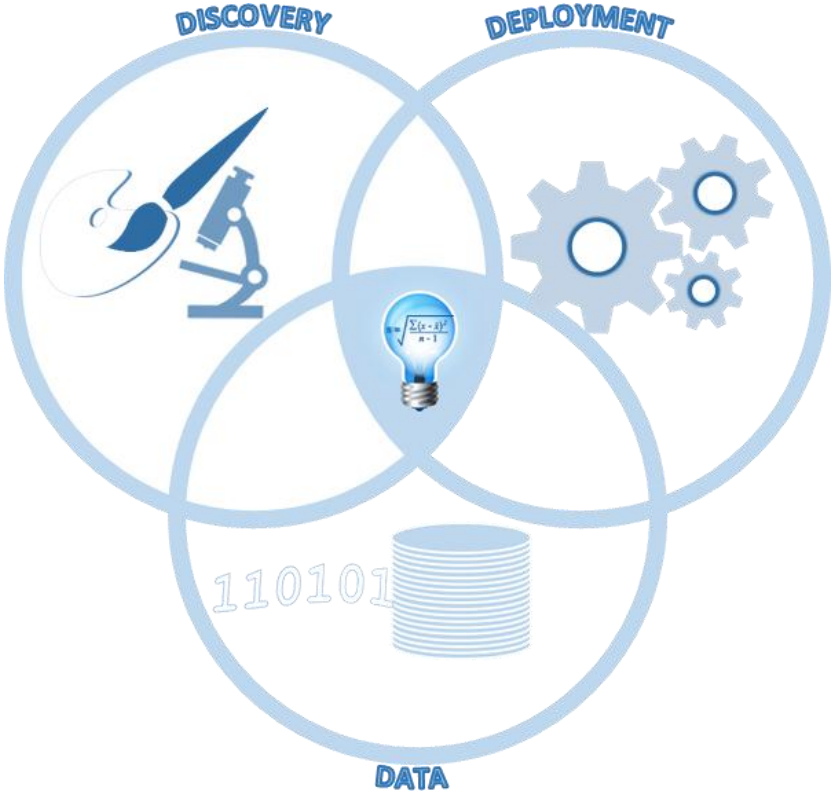
Over a year to select the best variables & deliver a robust & manageable model in production

Dun & Bradstr

Credit risk score on businesses cost millions of dollars due to volume & complexity of data



Modeling at Scale



Creating Even More Granular Models

DATA

Summed customer transactions

Customer type

Community type

Distance

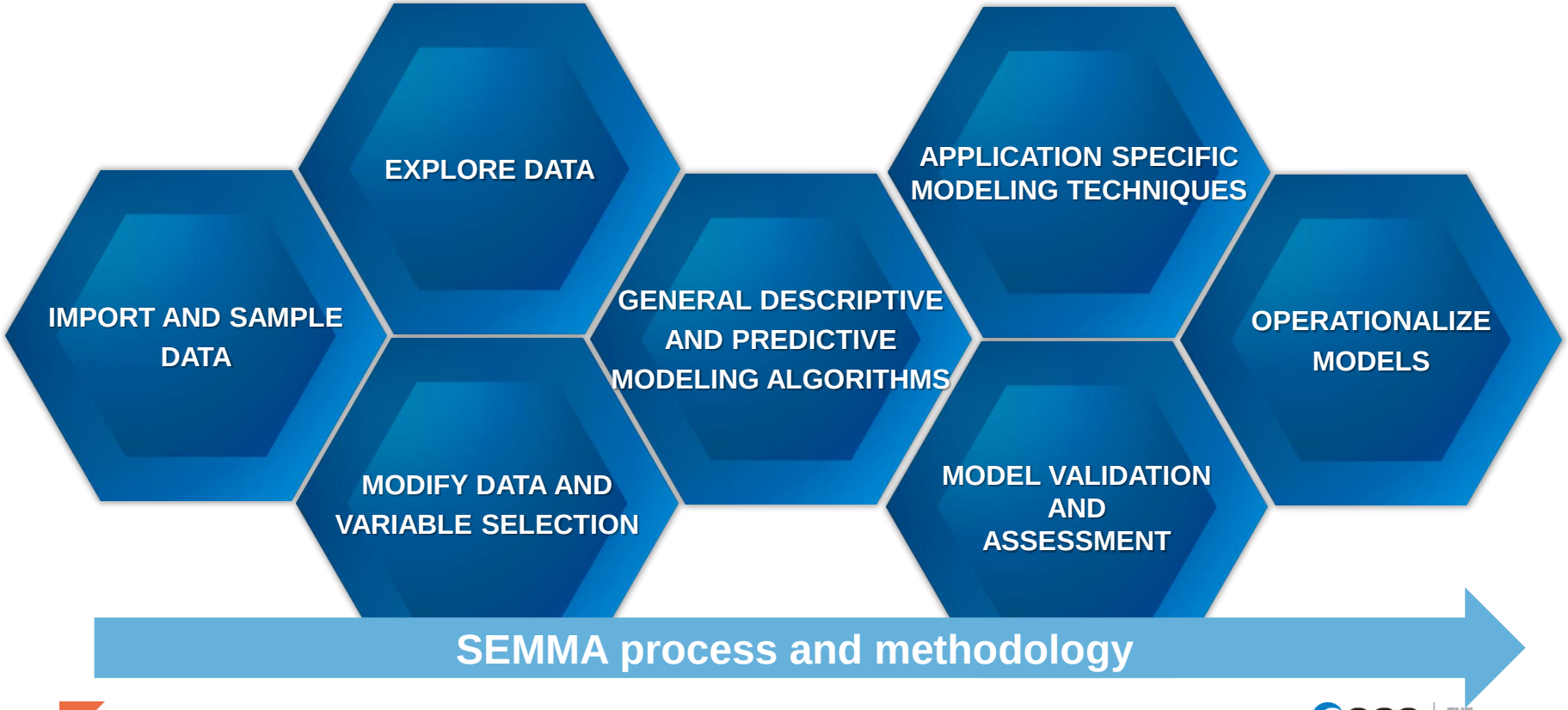
Spend frequency

Increasing refinement / granularity

Creating a **hierarchy** of models, choosing the “best” model at each level

Introducing SAS Factory Miner

Where we are now



SAS Factory Miner – Key features

- Usability
- Efficiency
- Model Completeness
- Model Variety
- Replicability and control

Demonstration

http://fcmrdemo.na.sas.com/SASLogon/login?service=http%3/

SAS® Logon Manager

Page Safety Tools

SAS®... SWW ... SAS ... Comp... Tran... Siti... OWA2... Port... Siti... Supp... audit Basi... Cred... Deci... Demo docu... ENI ... mode... New ... Pers...

Accesso a SAS®

ID utente:

itamcn

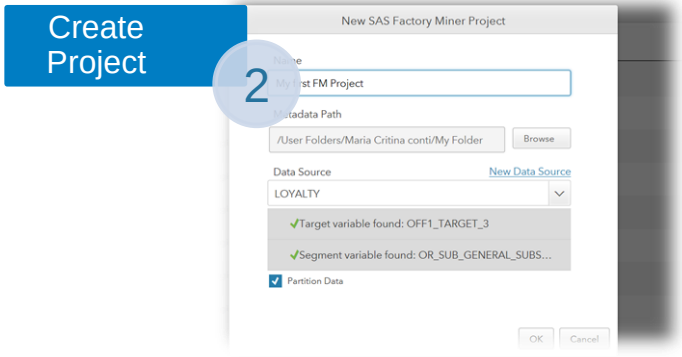
Password:

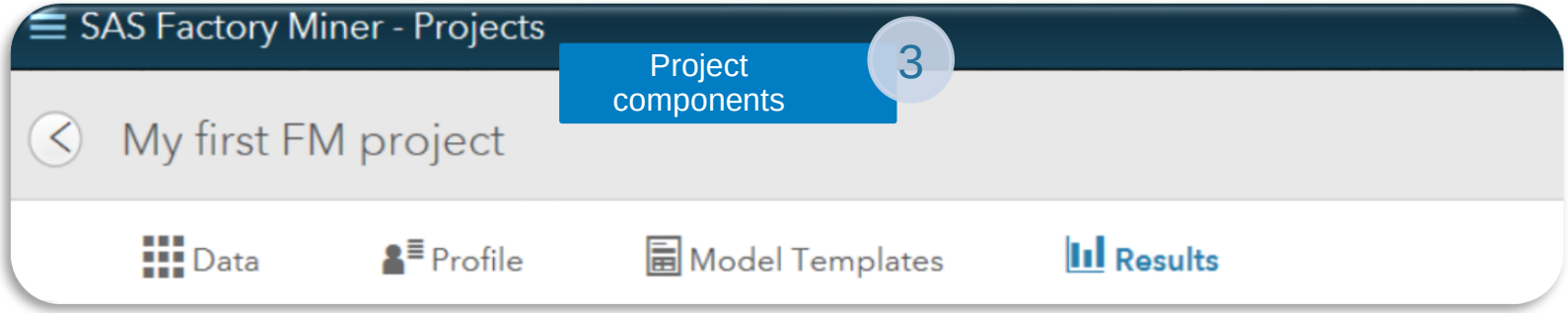
••••••••••

ACCEDI

Informazioni

Overview





Setting the scene: data and segments

4

Analyze inputs, segments and targets distributions

My first FM project

Data | Profile | Model Templates | Results

Build Profile

Variable	Label	Type	Role	Level	Order	Percent Missir	Number of Le	Min			Impute	
AIRPORT_C	Number...	Numeric	Input	Interval	None	0.0000					None	None
aquarius	Number...	Numeric	Input	Interval	None					44.0000	None	None
AVG_DAYS_BTN_S...	Average...	Numeric	Input	Interval				124.2368	3820.0000		None	None
BUS_PERCENT	Business...	Numeric	Input				0.0000	0.0731	1.0000		None	None

5

Generate profiles for each segment

Training: Event Rate Reset

0 | 1

Filter

Exclude | Include

Segment ID	Market	Member_Type	Training: Observati...	Training: Event Rate	Run
Segment 1	Midwest	BASE	5165	0.4391	Include
Segment 2	Midwest	GOLD	4973	0.1649	Include
Segment 3	Midwest	PLATINUM	2053	0.4233	Include
Segment 4	Midwest	SILVER	35935	0.0290	Exclude
Segment 5	Northeast	BASE	25790	0.4463	Include
Segment 6	Northeast	GOLD	24908	0.1944	Include

Model definition

6

Select list of templates to test selecting from available algorithms

7

Define characteristics and parameters of template

New Model Tem

Name

Model Template 2

Model Algorithm

Forest-Based Model

Make this template available everywhere

Mark as default

OK

Cancel

Model Template 3

Use

Drag variables here

Transform Variables

Filter

Variable Selection

Forest-Based Variable Selec...

Tree-Based Variable Selection

Impute

Forest-Based Model

Maximum Number of Trees: 100

Proportion of obs in each sample: 0.6

Maximum Depth: 15

Missing Values: Use In Search

Method to Determine Leaf Size.: Default

Leaf Size.: 1

Save and Run

Close

Model execution, evaluation and selection

9

Summary of models, results and performance by segment

Data | Profile | Model Templates | Results

Status

ew All

mpleted (46)

nning (7)

heduled (12)

led (0)

Canceled (0)

Pending Changes (0)

Excluded (15)

servations Reset

Filter

Actions

Segme...	Market	Member_Type	Champion	Valid: Kolmo	Ov...	Validation: Ob							
Segme...	Northeast	PLATINUM			No	4028	3	1	0	0	0	0	0
Segme...	Northeast	SILVER	Baseline T...	0.0000	No	77090	1	0	0	0	0	0	3
Segme...	Northwest	BASE	Regressio...	0.3423	No	597	4	0	0	0	0	0	0
Segme...	Northwest	GOLD			No	659	2	0	2	0	0	0	0
Segme...	Northwest	PLATINUM			No	160	2	0	2	0	0	0	0
Segme...	Northwest	SILVER	Baseline T...	0.0000	No	5081	1	0	0	0	0	0	3
Segme...	Southeast	BASE			N	310	3	0	1	0	0	0	0

8

Define settings for selection and execution

Edit Default Settings

Partition Data

Rules

Rules

Minimum Observations

100

Minimum Event Rate

10

Champion Criterion

Kolmogorov-Smirnov

10

Drill down on individual segment

Add | Actions

		Valid: Misclassification Rate	Valid: Kolmogorov-Smirn...	Status
		0.3306	0.3421	Completed
		0.3567	0.2483	Completed
Baseline Template	Baseline	0.4449	0.0000	Completed
Forest based template	Forest-Based Model	0.2980	0.3882	Completed

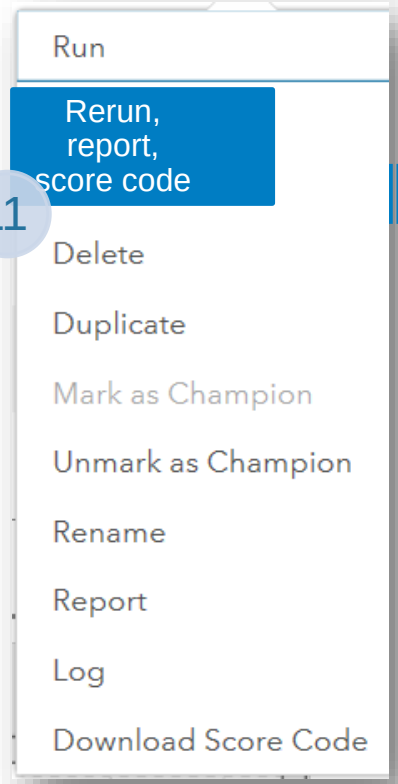
Model Graphs

Lift and Cumulative Lift

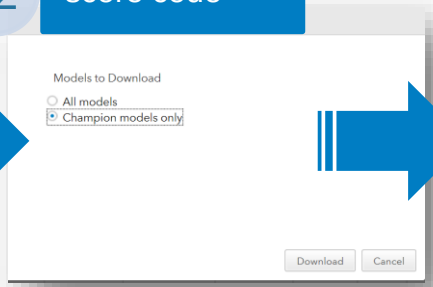
T-Values

Deployment, report and retraining

11 Rerun, report, score code

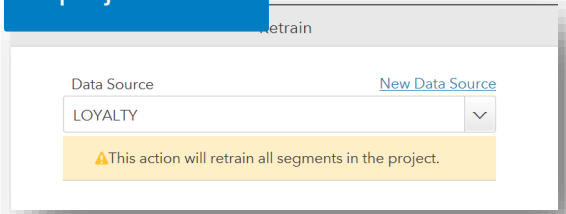


12 Export score code



```
*-------------------------------------------------------------*;
*-------------------------------------------------------------* ;
* Segment:          8aa59de94bbd5770014bdf6f076a565e;
* SegmentServerId:  1;
*-------------------------------------------------------------* ;
if strip(put(Market, $24.0)) ='Midwest' and strip(put(Member_Type, $15.0)) ='BASE'  then do;
*-------------------------------------------------------------* ;
* NODE:  HPImp;
*-------------------------------------------------------------* ;
Label IMP_PROPLOYALTY = 'Imputed Brand Loyalty                                  ';
Format IMP_PROPLOYALTY COMMA5.3;
if missing("PROPLOYALTY"n) then IMP_PROPLOYALTY = 0.6284682254;
else IMP_PROPLOYALTY ="PROPLOYALTY"n;
*-------------------------------------------------------------* ;
* NODE:  HPReg;
*-------------------------------------------------------------* ;
****************************************;
** SAS Scoring Code for PROC Hplogistic;
****************************************;
```

13 Retrain project



Forecasting challenges

Model proliferation
With targeted segmentation and micromarketing campaigns

Less time
to (re)act and differentiate on new market changes

Big Data
High volume, variety and velocity of big data

Examples & facts

- Different demand profiles requiring different modelling strategies
- Over 40 million forecasts generated every day
- Tracking of accuracy across forecasting iterations

Large scale automatic modelling

Scalable	Create a very large number of forecasts quickly
Manageable	Create statistical forecasts with limited resources
Reliable	Create sound forecasts that follow best practices

SAS Forecast Server client

- Approachable analytical workflow (automation)
- Productivity environment for statistical modeler
- Capitalizing on existing products while adding value
- Modeling by exception

SAS Forecast Server client

Extensibility allows users to define and reuse their own strategies:

- Segmentation
- Modeling
- Exceptions

Demonstration

LARGE SCALE AUTOMATIC FORECASTING

DEMO: FORECAST SERVER CLIENT

SAS® FORUM
UNITED KINGDOM 2015

sas | THE POWER TO KNOW.

www.SAS.com