



Business Analytics

IBM
Software
Solutions
Group

IBM SPSS Predictive Analytics Overview

with IBM SPSS Modeler

Rebecca Young– Data Scientist, Advanced Analytics, IBM Canada

Soha Moosavi, Business Analyst, BCNet

April 26, 2017



Presenters



Rebecca Young
Data Scientist, IBM Canada



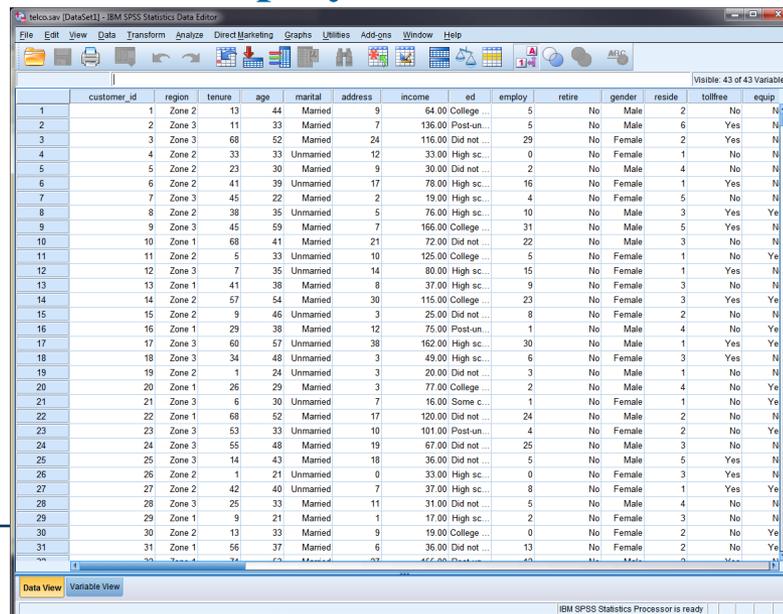
Soha Moosavi
Business Analyst, BCNET

SPSS Products and Components – Statistics

SPSS Statistics

Statistics Desktop: rich desktop application for statistical analysis and hypothesis testing. Note that Statistics Desktop is self-sufficient; it does not require a connection to the Server. It is not strictly a “client” in the usual sense.

Statistics Server: server component used for statistical analysis and for executing in batch mode. Statistics Desktop can connect to this server. Statistics Server is also capable of running scheduled processes, invoked by SPSS Collaboration and Deployment Services or any other job scheduler.



IBM SPSS Statistics Desktop Data Editor window showing a data table with 43 variables and 31 rows of data. The table includes columns for customer_id, region, tenure, age, marital, address, income, ed, employ, retire, gender, reside, tollfree, and equip.

	customer_id	region	tenure	age	marital	address	income	ed	employ	retire	gender	reside	tollfree	equip
1	1	Zone 2	13	44	Married	9	64.00	College	5	No	Male	2	No	No
2	2	Zone 3	11	33	Married	7	136.00	Post-un	5	No	Male	6	Yes	No
3	3	Zone 3	68	52	Married	24	116.00	Did not	29	No	Female	2	Yes	No
4	4	Zone 2	33	33	Unmarried	12	33.00	High sc	0	No	Female	1	No	No
5	5	Zone 2	23	30	Married	9	30.00	Did not	2	No	Male	4	No	No
6	6	Zone 2	41	39	Unmarried	17	78.00	High sc	16	No	Female	1	Yes	No
7	7	Zone 3	45	22	Married	2	19.00	High sc	4	No	Female	5	No	No
8	8	Zone 2	38	35	Unmarried	5	76.00	High sc	10	No	Male	3	Yes	Yes
9	9	Zone 3	45	59	Married	7	166.00	College	31	No	Male	5	Yes	No
10	10	Zone 1	68	41	Married	21	72.00	Did not	22	No	Male	3	No	No
11	11	Zone 2	5	33	Unmarried	10	125.00	College	5	No	Female	1	No	Yes
12	12	Zone 3	7	35	Unmarried	14	80.00	High sc	15	No	Female	1	Yes	No
13	13	Zone 1	41	38	Married	8	37.00	High sc	9	No	Female	3	No	No
14	14	Zone 2	57	54	Married	30	115.00	College	23	No	Female	3	Yes	Yes
15	15	Zone 2	9	46	Unmarried	3	25.00	Did not	8	No	Female	2	No	No
16	16	Zone 1	29	38	Married	12	75.00	Post-un	1	No	Male	4	No	Yes
17	17	Zone 3	60	57	Unmarried	38	162.00	High sc	30	No	Male	1	Yes	Yes
18	18	Zone 3	34	48	Unmarried	3	49.00	High sc	6	No	Female	3	Yes	No
19	19	Zone 2	1	24	Unmarried	3	20.00	Did not	3	No	Male	1	No	No
20	20	Zone 1	26	29	Married	3	77.00	College	2	No	Male	4	No	Yes
21	21	Zone 3	6	30	Unmarried	7	16.00	Some c	1	No	Female	1	No	Yes
22	22	Zone 1	68	52	Married	17	120.00	Did not	24	No	Male	2	No	No
23	23	Zone 3	53	33	Unmarried	10	101.00	Post-un	4	No	Female	2	No	Yes
24	24	Zone 3	55	48	Married	19	67.00	Did not	25	No	Male	3	No	No
25	25	Zone 3	14	43	Married	18	36.00	Did not	5	No	Male	5	Yes	No
26	26	Zone 2	1	21	Unmarried	0	33.00	High sc	0	No	Female	3	Yes	No
27	27	Zone 2	42	40	Unmarried	7	37.00	High sc	8	No	Female	1	Yes	Yes
28	28	Zone 3	25	33	Married	11	31.00	Did not	5	No	Male	4	No	No
29	29	Zone 1	9	21	Married	1	17.00	High sc	2	No	Female	3	No	No
30	30	Zone 2	13	33	Married	9	19.00	College	0	No	Female	2	No	Yes
31	31	Zone 1	56	37	Married	6	36.00	Did not	13	No	Female	2	No	Yes

IBM SPSS Statistics Desktop:

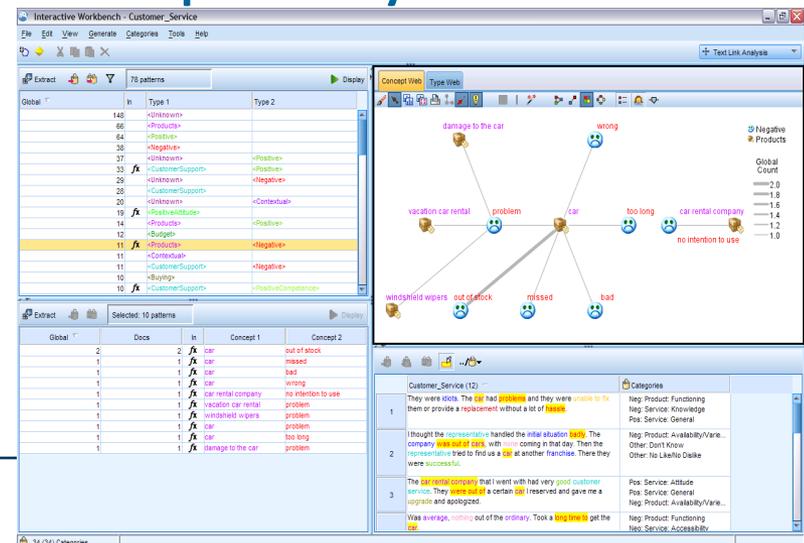
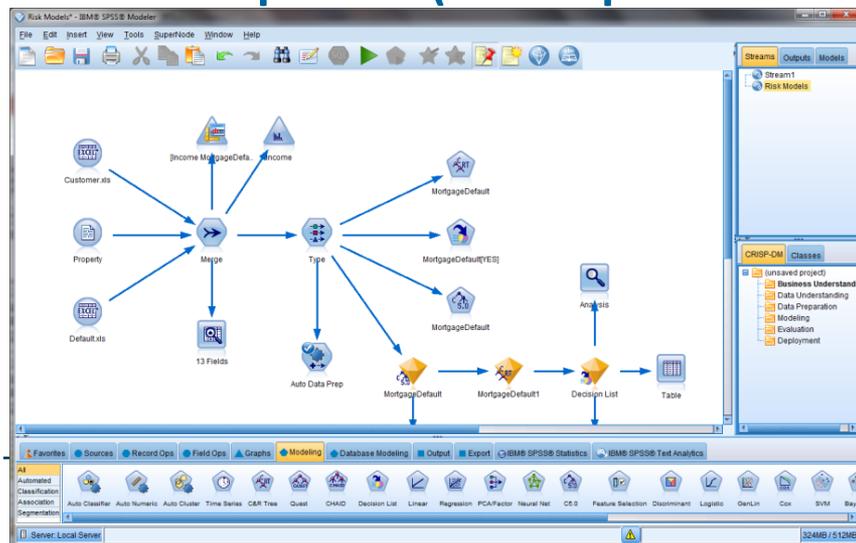
SPSS Products and Components – Modeler

Modeler

Modeler Desktop: a rich desktop application for developing models. Note that Modeler Desktop is self-sufficient; it does not require a connection to the Server. It is not strictly a “client” in the usual sense.

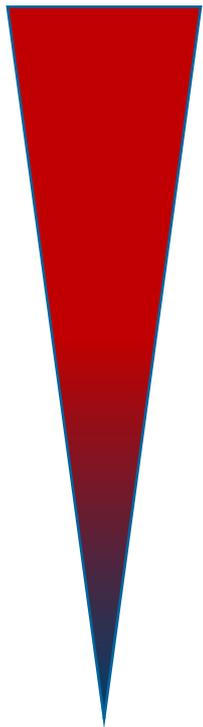
Modeler Server: server component used for building models and for executing models in batch mode. Modeler Desktop can connect to this server. Modeler Server is also capable of performing batch scoring, Analytical Decision Management (together with C&DS), and other features.

Each component (Desktop and Server) can run independently of the other.



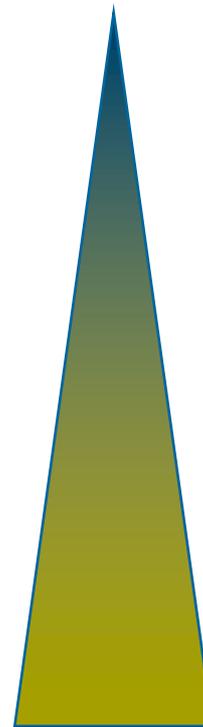
Statistical Analysis and Modeling Both Drive Predictive Analytics

Statistics Approach



- **Statistical approach involves:**
 - **forming a theory about a possible relationship**
 - **converting the theory to a hypothesis**
 - **testing that hypothesis, using statistical methods**
- **It is a manual, user-driven, top-down approach to data analysis**
- **Used to test hypotheses**

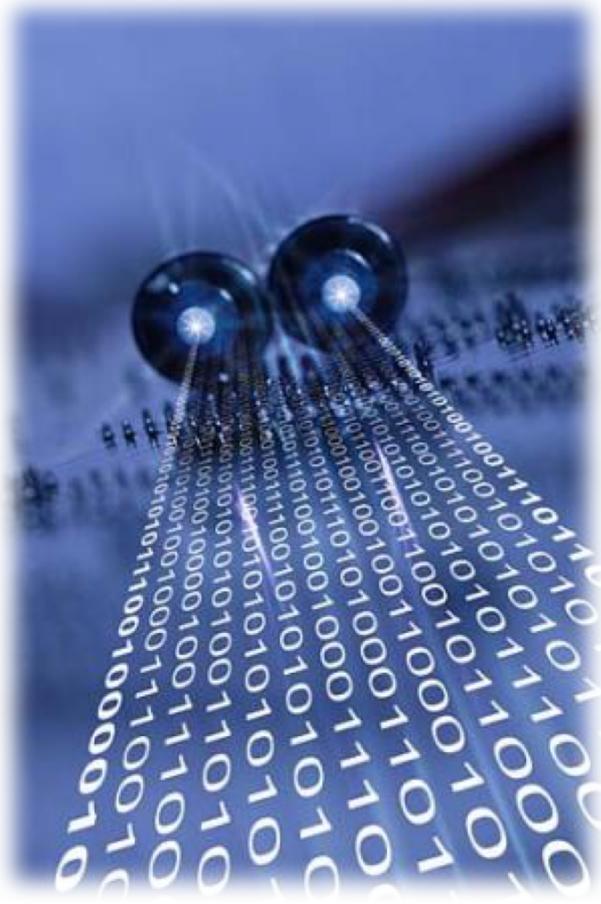
Modeling Approach



- **Data mining involves:**
 - **the interrogation of the data**
 - **determined by the method and goal, rather than by the user**
 - **listening to the voice of the data**
- **It is a data-driven, self-organizing, bottom-up approach to data analysis**
- **Used to generate hypotheses**

Note that Both Approaches Drive Predictive Analytics

What is predictive analytics?

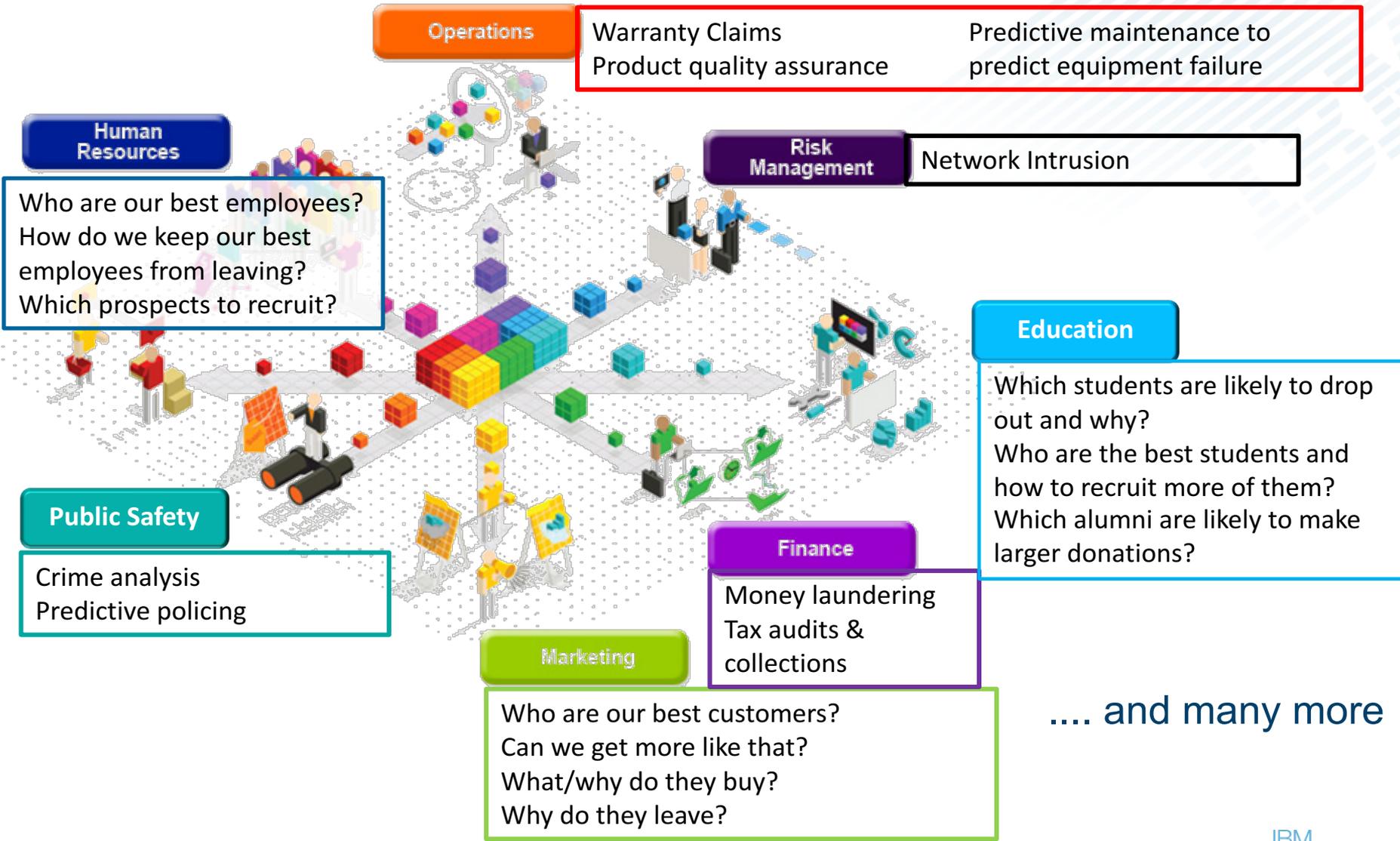


Predictive Analytics helps connect **data** to **effective action** by drawing reliable conclusions about current conditions and future events

Gareth Herschel, Research Director, Gartner Group

Enabling businesses to use **predictive models** to exploit patterns found in historical data to **identify** potential **risks** and **opportunities** before they occur.

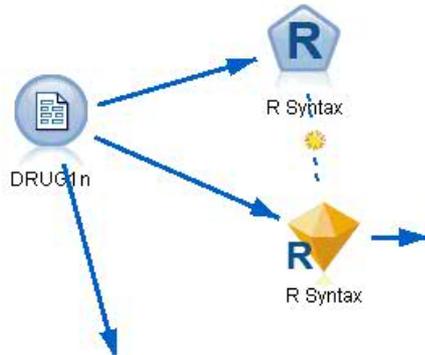
Areas for Predictive Analytics



Modeling Techniques in IBM SPSS Modeler

Technique	Usage	Algorithms
Classification (or prediction)	<ul style="list-style-type: none">• Used to predict group membership (e.g., will this employee leave?) or a number (e.g., how many widgets will I sell?)	<ul style="list-style-type: none">• Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc.
Segmentation	<ul style="list-style-type: none">• Used to classify data points into groups that are internally homogenous and externally heterogeneous.• Identify cases that are unusual	<ul style="list-style-type: none">• Auto Clustering, K-means, etc.• Anomaly detection
Association	<ul style="list-style-type: none">• Used to find events that occur together or in a sequence (e.g., market basket)	<ul style="list-style-type: none">• APRIORI, Carma, Sequence

Extend Capabilities through Open Source (R, Python)



R Integration

R Build/Score, Process and Output node support

Scale R execution by leveraging database vendor provided R engines

The image shows two screenshots from the IBM Business Analytics interface. The left screenshot displays the 'Custom Dialog Builder' window, which is used to create custom dialog boxes. It features a 'Tools' palette on the left with various controls like Source List, Target List, Check Box, Combo Box, List Box, Text control, Number control, Static Text, Item Group, Radio Group, Check Box Group, File Browser, and Sub-dialog Button. The main workspace shows a 'Variables' list with Variable1, Variable2, and Variable3, and a 'Target List' with a 'Variable' control. Below the workspace is a 'Target Variable List Properties' table.

Property	Value
Identifier	pred_var
Title	Predictors List
ToolTip	
Target list type	Multiple item list
Mnemonic Key	
Required for execution	True

The right screenshot shows the 'Syntax Template' dialog box. It contains instructions on how to use syntax templates to generate syntax for a dialog. It includes an example line: `!VARIABLES %%var_list%%`. A tip explains that users can select from a list of available control identifiers to insert into the template. The main area shows a list of control identifiers with their corresponding syntax:

- 1 # model building
- 2 attach(modelerData)
- 3 model<-lm(%%stg_var%% ~ %%pred_var%%)
- 4 model
- 5
- 6
- 7
- 8 # scoring
- 9 attach(modelerData)
- 10 output<-predict(model)
- 11 output
- 12

 The dialog has 'OK', 'Cancel', and 'Help' buttons at the bottom.

Custom Dialog Builder for R or Python

Provides the ability to create new Modeler Algorithm nodes and dialogs that run R processes

Makes R usable for non-programmers

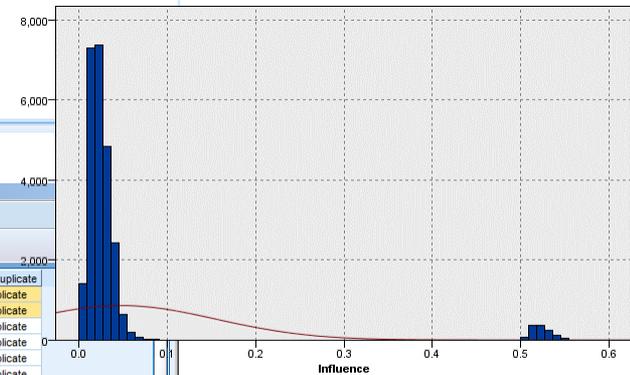
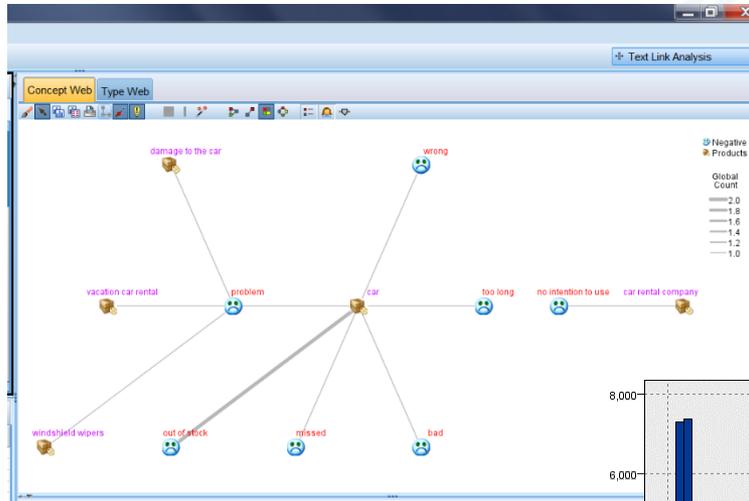
Uncovering Patterns in Unstructured Data

- Text Analytics

- Natural Language
- Sentiment Analysis

- Social Network Analysis

- Uncover relationships
- Find leaders and followers



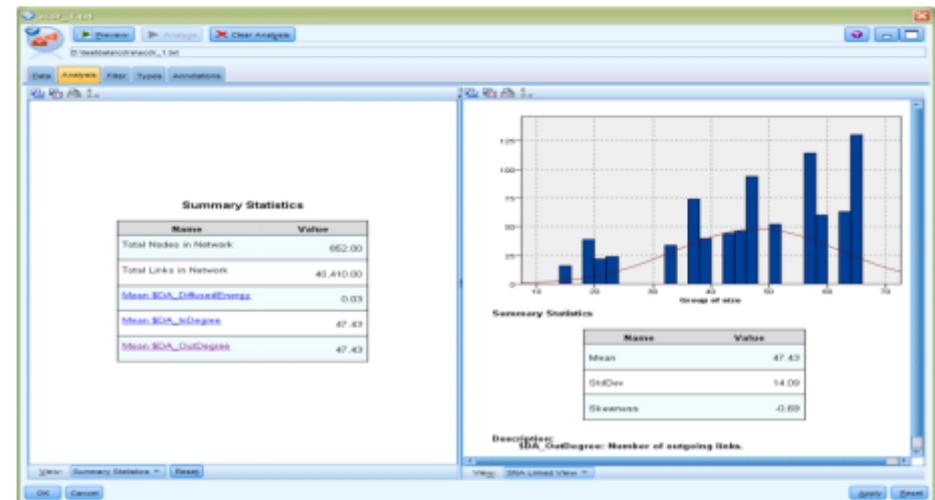
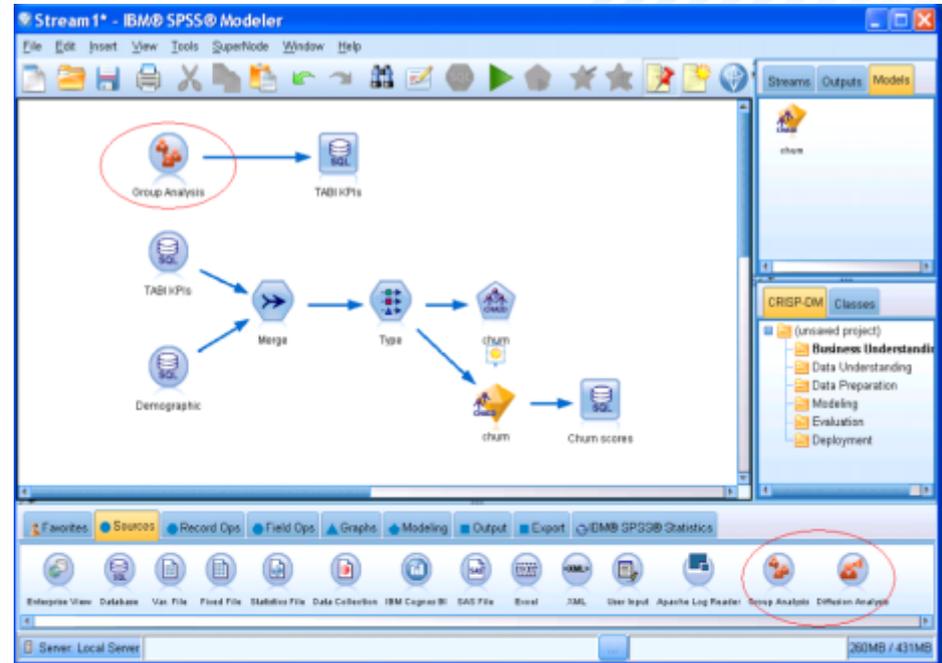
Summary Statistics

Table (13 fields, 850 records) #1

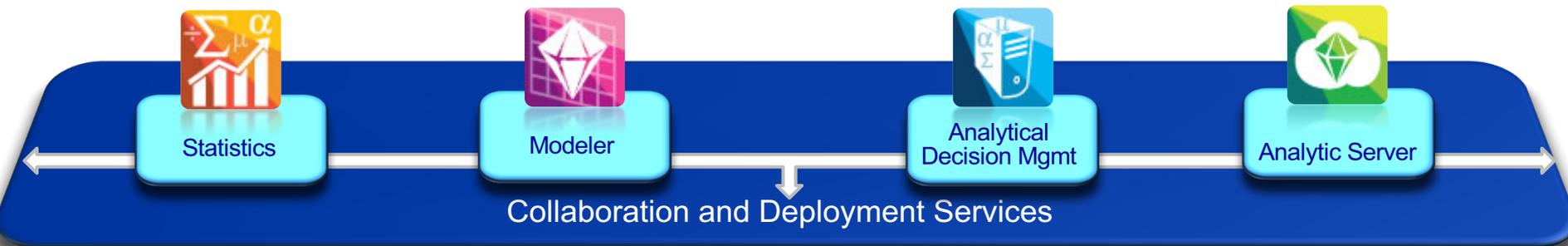
SEA-ID	SEA-SRC	key	SEA-SC	SEA-RULE	SEA-DOB_DATE	SEA-GENDER	SEA-NAME_SUR_NAME	SEA-NAME_GIVEN_NAME	SEA-PASSPORT_ID_NUM	SEA-PHONE_PHONE_NUM	SEA-SSN_ID_NUM	IsDuplicate
11	273 TEST	REF00000273	0.000	SnuIS	1926-04-03	M	MARTINEZ	KEVIN	317451	SnuIS	320-19-4	Duplicate
12	273 TEST	REF00000728	10.000	SF1_PNAME_CFF_CSTAB	1926-04-03	M	MARTINEZ	EUGENE	317451	SnuIS	320-19-4	Duplicate
13	334 TEST	REF00000334	0.000	SnuIS	1995-02-13	M	BUTLER	JOSHUA	692453	SnuIS	510-98-6	Duplicate
14	334 TEST	REF00000520	10.000	SF1_PNAME_CFF_CSTAB	1995-02-13	M	BUTLER	TODD	692453	SnuIS	510-98-6	Duplicate
15	342 TEST	REF00000517	8.000	SF1_PNAME_CFF_DSTAB	1967-12-02	F	HUGHES	DEBRA	989511	SnuIS	156-04-7	Duplicate
16	342 TEST	REF00000857	0.000	SnuIS	1970-12-18	F	HUGHES	DEBRA	989511	SnuIS	156-04-7	Duplicate
17	377 TEST	REF00000377	0.000	SnuIS	1950-09-28	M	WHITE	DANIEL	844897	SnuIS	389-32-9	Duplicate
18	377 TEST	REF00000780	8.000	SF1_PNAME_CFF_DSTAB	1949-12-01	M	WHITE	DAN	844897	SnuIS	389-32-9	Duplicate
19	388 TEST	REF00000388	0.000	SnuIS	1937-03-26	M	HILL	RUSSELL	104791	SnuIS	551-95-8	Duplicate
20	388 TEST	REF00000628	10.000	SF1_PNAME_CFF_CSTAB	1937-03-26	M	HILL	RUSSELL	104791	SnuIS	551-95-8	Duplicate
21	437 TEST	REF00000573	9.000	SF1_PNAME_CSTAB	1937-09-16	F	JENKINS	DORIS	462623	SnuIS	688-19-5	Duplicate
22	437 TEST	REF00000437	0.000	SnuIS	1937-09-16	F	JENKINS	DORIS	462623	SnuIS	688-19-5	Duplicate
23	501 TEST	REF00000501	0.000	SnuIS	1978-04-03	F	GRIFFIN	RUTH	571208	SnuIS	516-52-9	Duplicate
24	501 TEST	REF00000621	10.000	SF1_PNAME_CFF_CSTAB	1978-04-03	F	GRIFFIN	RUTH	571208	SnuIS	516-52-9	Duplicate
25	1 TEST	REF00000001	0.000	SnuIS	1938-11-16	M	BROOKS	JOE	147882	SnuIS	338-14-3	Unique
26	2 TEST	REF00000002	0.000	SnuIS	1989-04-05	F	HALL	ANNIE	554947	SnuIS	413-31-8	Unique
27	3 TEST	REF00000003	0.000	SnuIS	1931-06-25	M	BROWN	AARON	856779	SnuIS	997-89-0	Unique
28	4 TEST	REF00000004	0.000	SnuIS	1990-01-02	M	WASHINGTON	WAYNE	642815	SnuIS	485-80-9	Unique
29	5 TEST	REF00000005	0.000	SnuIS	1983-01-21	F	ADAMS	KIMBERLY	762208	SnuIS	440-90-5	Unique
30	6 TEST	REF00000006	0.000	SnuIS	1946-05-25	M	SMITH	CLARENCE	530356	SnuIS	833-82-1	Unique
31	7 TEST	REF00000007	0.000	SnuIS	1974-08-12	M	BROWN	TERRY	213824	SnuIS	274-76-1	Unique

Social Network Analysis

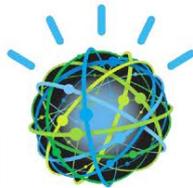
- Processes CDR (Call Data Record) data companies to produce social analysis
- Focuses around identifying groups, leaders and probabilities that others will churn based on influence
- Enhances existing churn predictions of Modeler
 - Expressed as two new nodes in the Sources Palette
 - Group Analysis – what are the groups in my data and who are the leaders
 - Diffusion Analysis – uses existing churn information to determine who else that churning is likely to influence to leave



IBM SPSS Modeler Within the Business Analytics Ecosystem



Cognos
software
InfoSphere



ASSETS MANAGEMENT **IBM. maximo**



NETEZZA
an IBM® Company

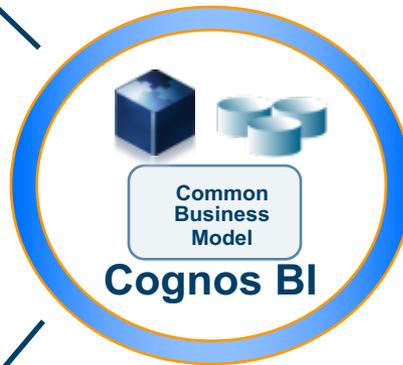
IBM Research
Etc...

WebSphere software

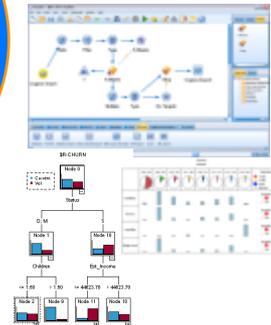
Integration with IBM Cognos



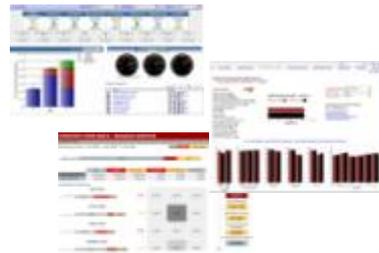
1) Leveraging BI, identify problem or situation needing attention



2) SPSS predictive analytics feed results back into the BI layer



3) Results widely distributed via BI for consumption by business Users



Overview takeaways

Easy to use, visual interface

- Short timeframe to be productive with actionable results
- Does not require knowledge of programming language
- No proprietary data formats
- Open architecture

Business results focused

- Leverages the investments already made in technology
- Cost effective solution that delivers powerful results across organization
- Full range of algorithms for your business problems
- Big Data enabled (Hadoop, SQL Pushback)

End-to-end solution

- Data preparation through real time interactions
- Use structured, unstructured and semi-structured data
- Integrated portfolio for business analytics
- Scales from a single desktop to an enterprise deployments

SPSS Agreement with BCNET

- Members pay the same fee as before, for three years, without the 6% annual increase
- After year three members can exit the service if they wish
- After three years, members only pay the maintenance fee component and a small BCNET administration fee for their seats (less than 25% of current fees)
- How you use the licenses will not change
- Any renewals or new SPSS purchases will be completed through BCNET