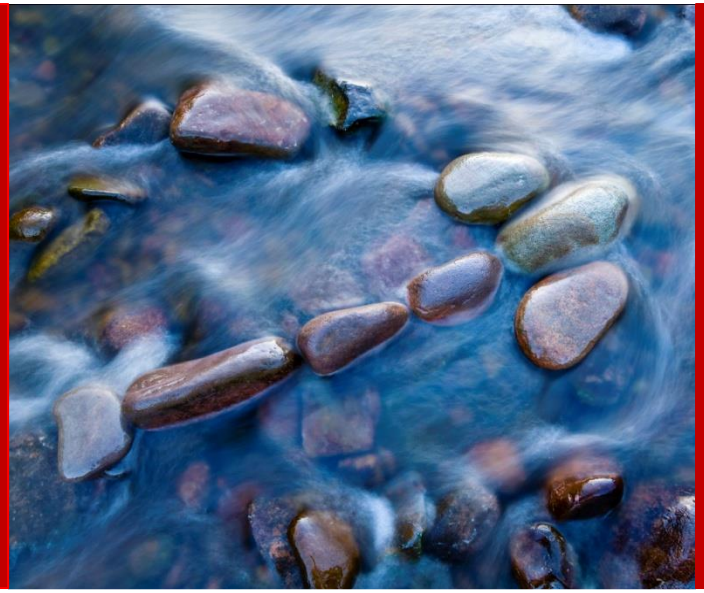


PREDICTIVE ANALYTICS REVELATION

MINDSTREAM® ANALYTICS

02.27.2014

Agenda

- Introduction
- Who we are!
- How to build Predictive Models?
- Demonstration: IBM SPSS
- Success Stories
- Questions and Answers

Why MindStream Analytics

CIO DataAnalytics Consulting 2013

20 Most Promising Data Analytics Consulting Companies

Windstream was awarded the 2014 ProformaTECH Award for Excellence, for the Most Effective Adaption of Technology by a Finance Organization.

Putting a smile on executives face!

MindStream helps guide executives; integrating data from different sources and extracting value from it.

Leveraging our Analytics technology partnership with IBM reduces our customers' risk, while providing high value, award winning projects that take a direct line to bottom line value.

Last webinar

What is Predictive Analytics?

- Empirically-derived models used for predicting future outcomes

Analyze	Forecast	Plan	Collaborate	Simulate
Survey	Govern	Discover	Model	Predict
Mine	Report	Score	Visualize	Decide

Better predict customer behavior, increasing profits and revenue

MINDSTREAM

5

Goals of Predictive Analytics

- Bring key business insights into our decision-making processes
- Solution to our biggest challenges with data mining
- Integration of predictive analytics with data driven decision making
- Positive ROI and superior outcomes

MINDSTREAM

6

Goals of Predictive Analytics

Customer

Banking
- Increase account profitability

Insurance
- Retain policy holders with better service & marketing

Retail
- Understand sales patterns

Telecommunications
- Reduce churn with custom retention offers

Finance

Government
- Effective budget management

Retail
- Develop dynamic merchandise plans

Industrial
- Plan and forecast sales & operations

Operations

Industrial
- Predict maintenance issues before they occur

Retail
- Improve store performance with P&L reports

Telecommunications
- Understand & manage network traffic

Banking
- Measure branch performance

Insurance
- Streamline claims process

Government
- Reduce fraud and waste

Risk

Banking
- Align risk strategy and financial planning
- Improve compliance & regulatory response

Insurance
- Improve compliance & regulatory response

MINDSTREAM

Who needs Predictive Analytics?

- Companies that need to:

1. **secure** their competitiveness
2. **increase** sales and grow customer base
3. **manage** fraud
4. **meet** customer's expectations
5. **improve** core business capacity

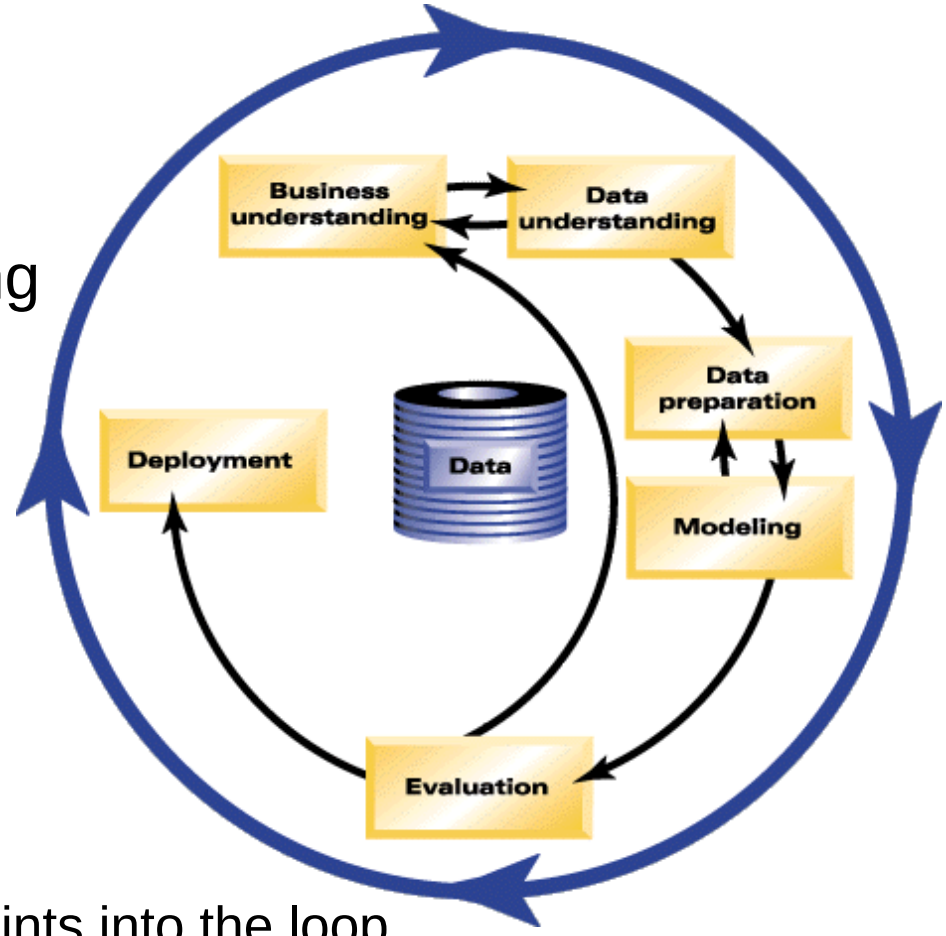
MINDSTREAM

How to build Predictive Models

CRISP-DM

6 Phases

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment



- Not strictly ordered
  - Several possible entry points into the loop
- Reflects iterative nature of data mining

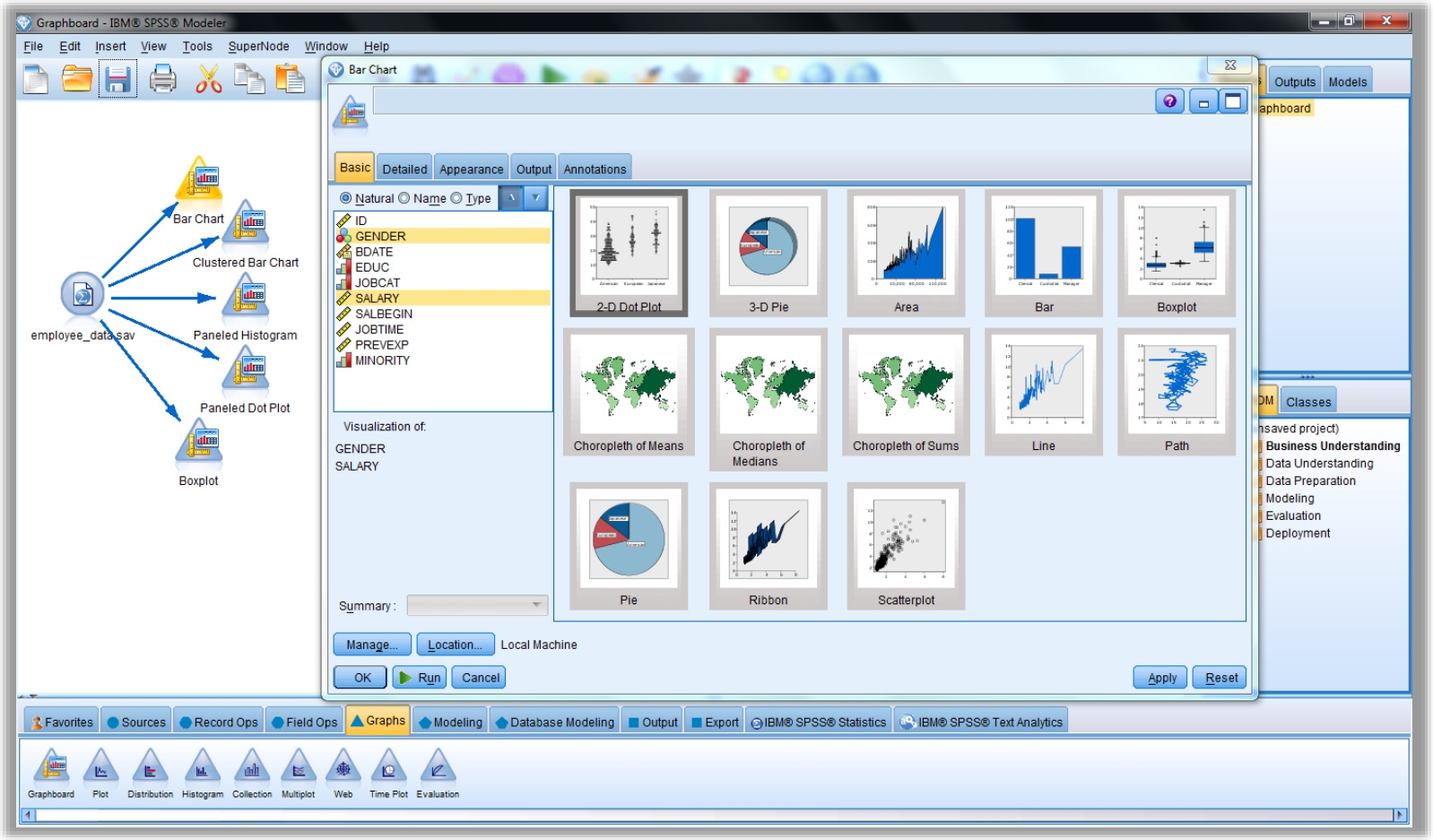
SPSS Modeler

- High-performance data mining and text analytics
- Utilizes structured and unstructured data
- Creates predictive analytics for data driven decision making
- Enables superior outcomes and positive ROI

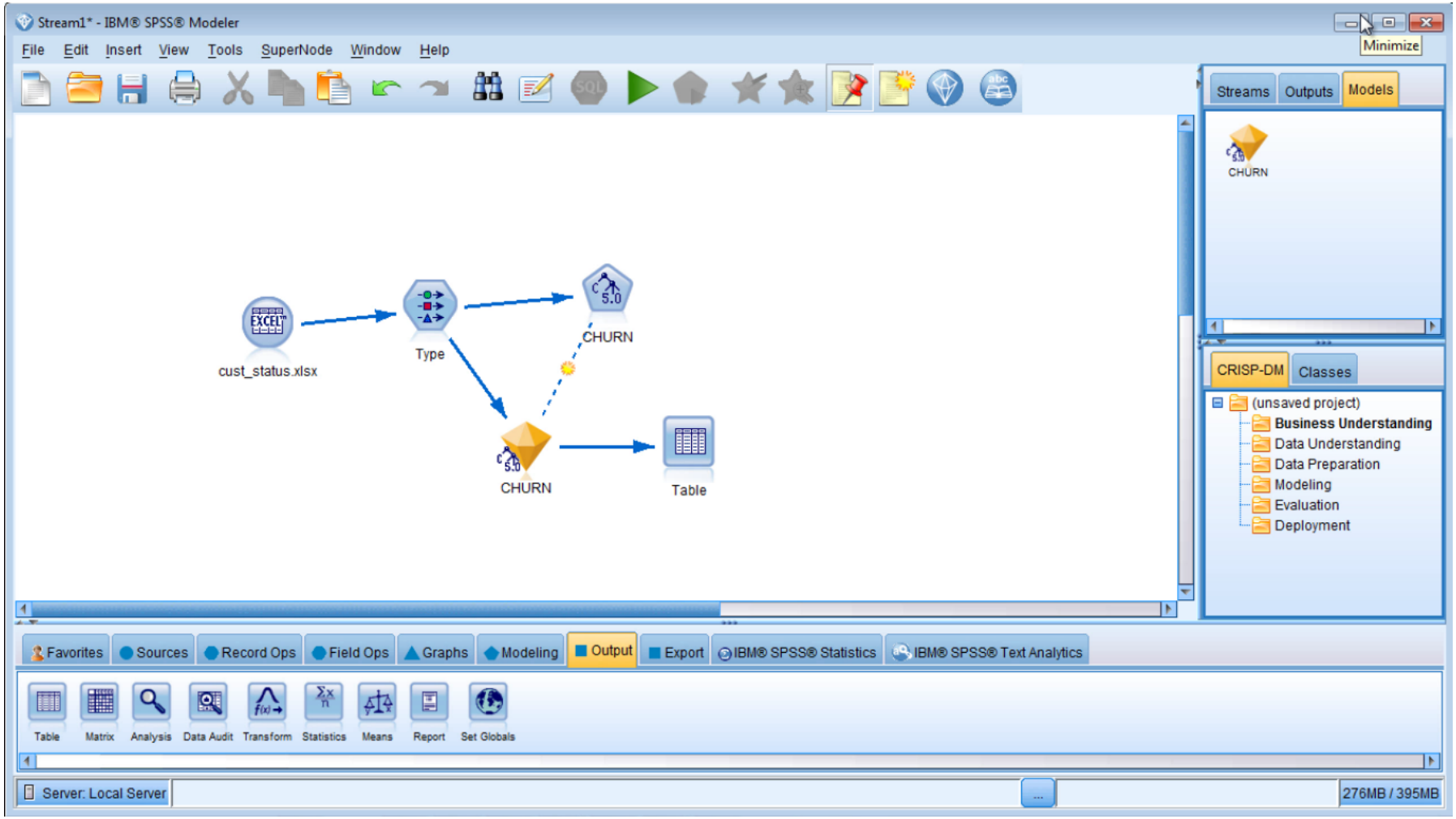
SPSS Modeler

- Easy-to-use, interactive interface without the need for programming
- Automated modeling and data preparation capabilities
- Access ALL data – structured and unstructured – from disparate sources
- Natural Language Processing (NLP) to extract concepts and sentiments in text
- Entity Analytics ensures the quality of the data and results in more accurate models

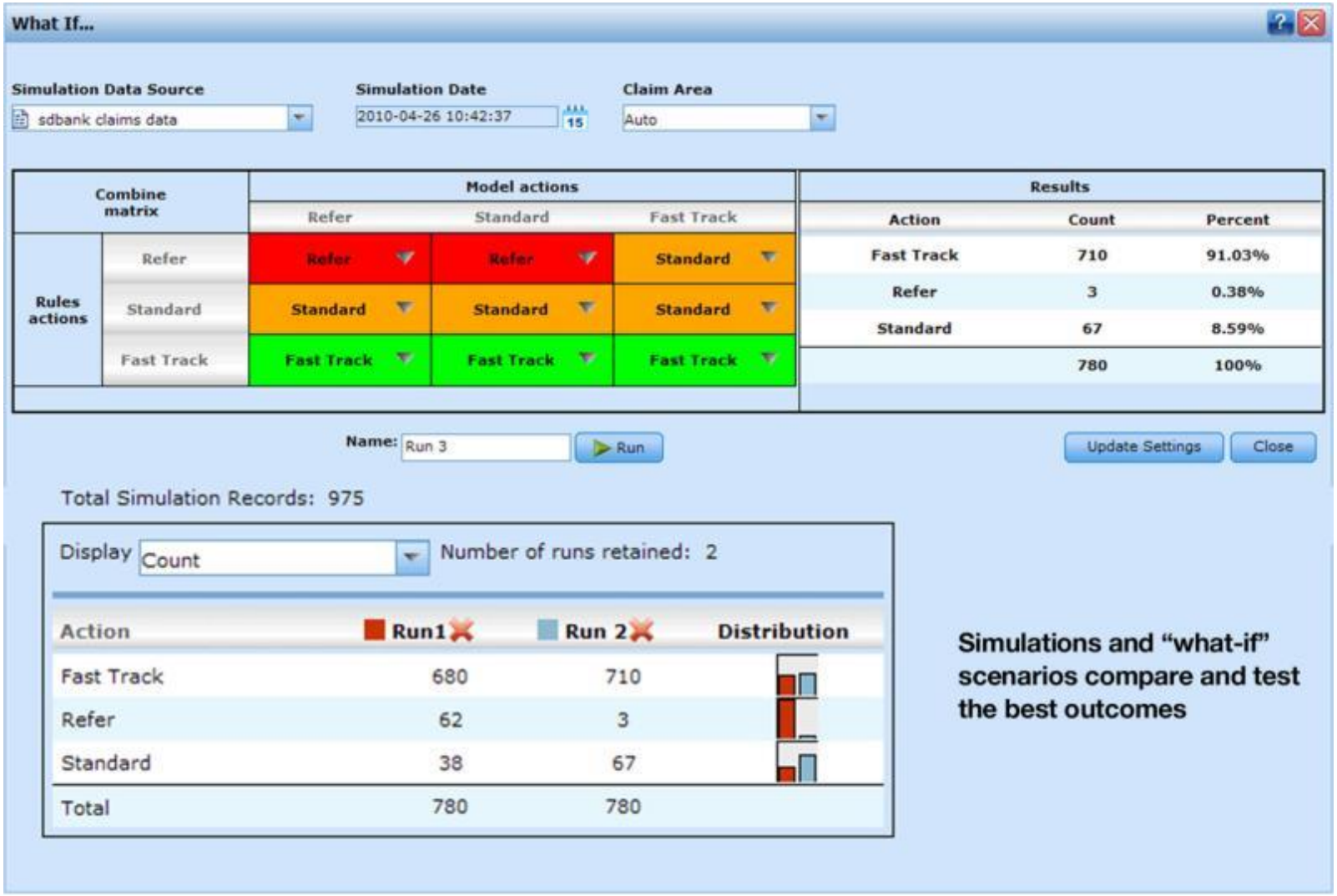
SPSS Modeler



SPSS Modeler



SPSS Decision Management



Simulations and "what-if" scenarios compare and test the best outcomes

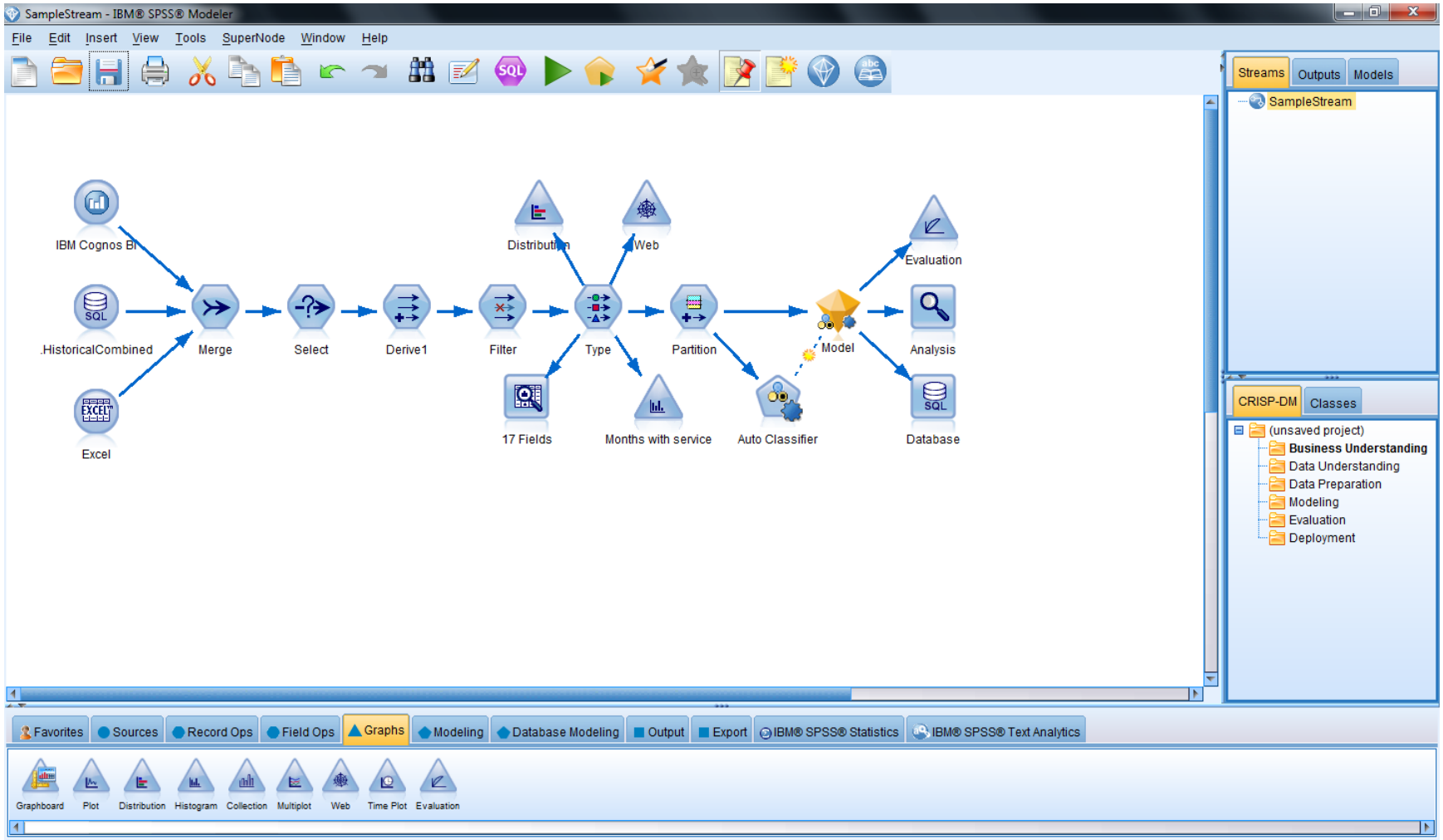
Data on Airlines in SFO

- Collected monthly via customer interviews held at all airport terminals and boarding areas from July 2005 through March 2011
- Interviews were done using a stratified random sample of flights selected by airport staff
- The questionnaires were available in English, Japanese, Chinese, and Spanish.

Data on Airlines in SFO

- **Understanding the data**: determining which fields in the data to use as predictors and which ones to discard.
- **Partitioning the file**
- **Training the model:** we will use one of the partitioned data to train the model.
- **Scoring the model:** we will use the other partitioned data to score the model.

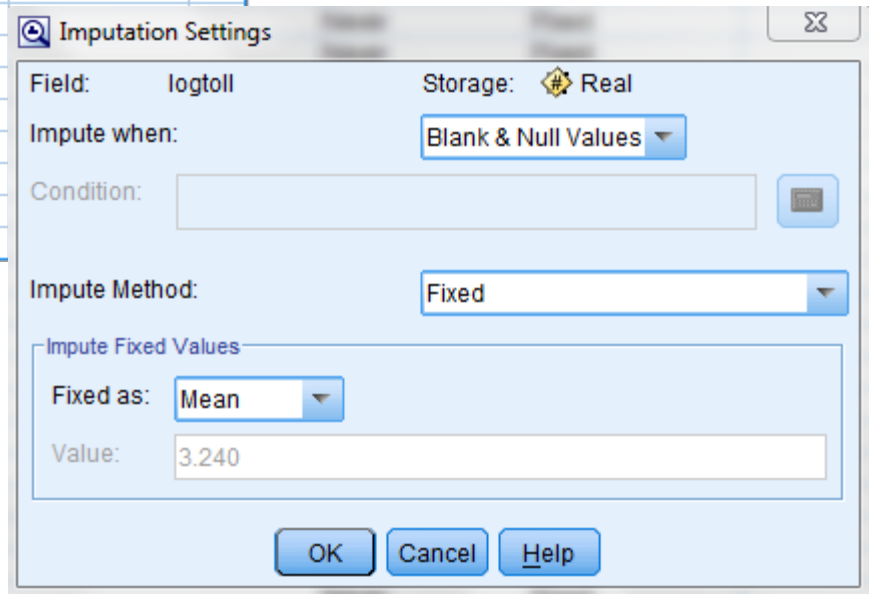
Airlines in SFO - Model



Airlines in SFO
Handling Missing Values and Outliers

Measurement	Outliers	Extremes	Action	Impute Missing	Method	%
Nominal	--	--	--	Never	Fixed	
Continuous	0	0	None	Never	Fixed	
Continuous	0	0	None	Never	Fixed	
Flag	--	--	--	Blank & Null Values	Fixed	
Continuous	12	0	None	Never		
Continuous	9	6	None	Never		
Ordinal	--	--	--	Never		
Continuous	8	0	None	Never		
Nominal	--	--	--	Never		
Nominal	--	--	--	Never		
Ordinal	--	--	--	Never	Fixed	
Flag	--	--	--	Never		
Flag	--	--	--	Never		
Flag	--	--	--	Never		
Flag	--	--	--	Never		
Continuous	18	4	None	Never		
Continuous	9	1	None	Never		
Continuous	2	0	None	Never		
Continuous	11	3	None	Never		

Fixed
Random
Expression...
Algorithm
Specify...



Airlines in SFO – Bar Chart

- Passengers by Origin or Destination

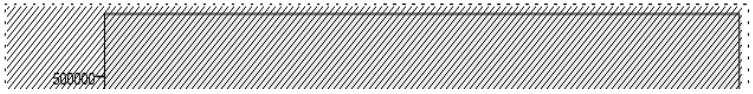
*Output1 [Document1] - IBM SPSS Stati

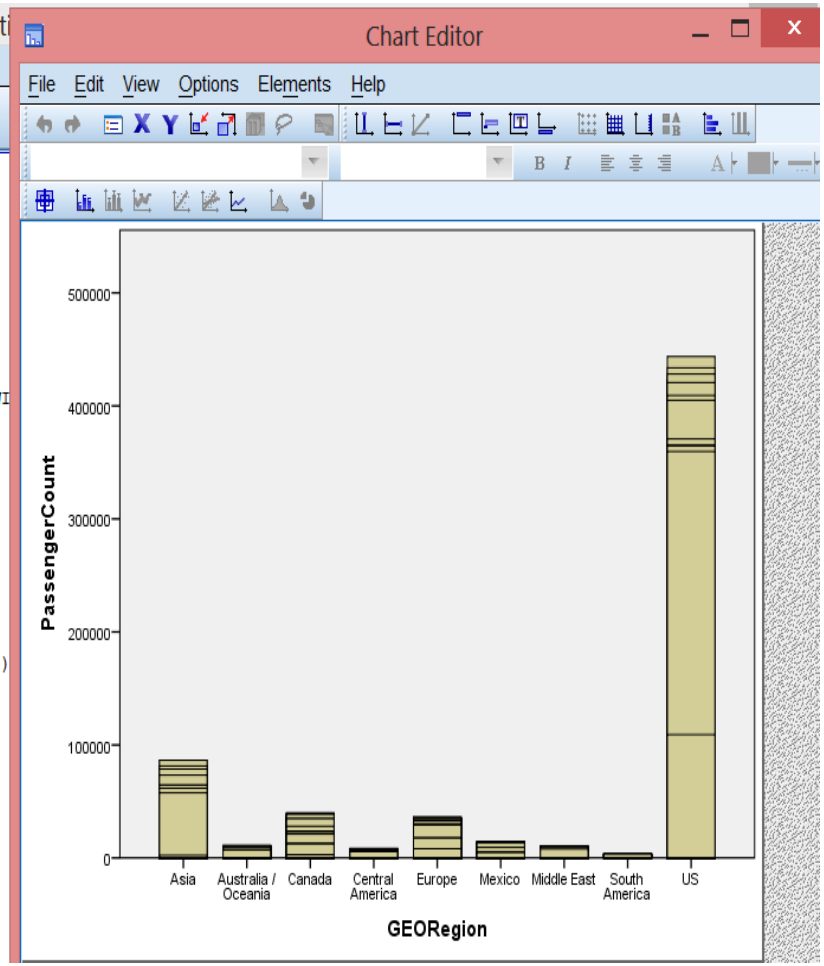
form Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

```
* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=GEORegion PassengerCount MISSING=LISTWI
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: GEORegion=col(source(s), name("GEORegion"), unit.category())
  DATA: PassengerCount=col(source(s), name("PassengerCount"))
  GUIDE: axis(dim(1), label("GEORegion"))
  GUIDE: axis(dim(2), label("PassengerCount"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: interval(position(GEORegion*PassengerCount), shape.interior(shape.square))
END GPL.
```

GGraph

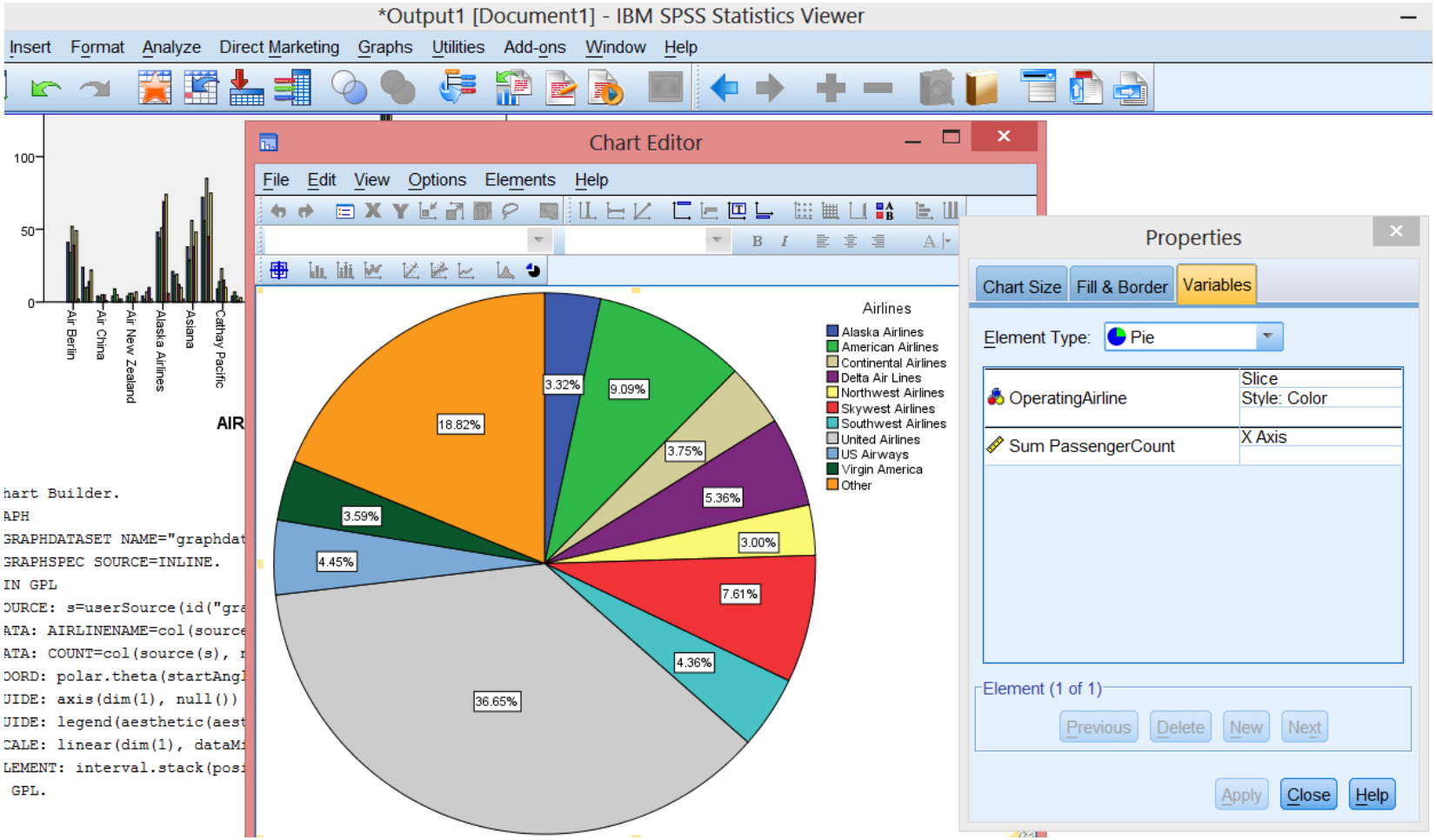
[DataSet3]





Airlines in SFO – Pie Chart

- Passengers by Airlines



Airlines in SFO - Crosstab

- Cross Tabulation of Destinations and Incomes

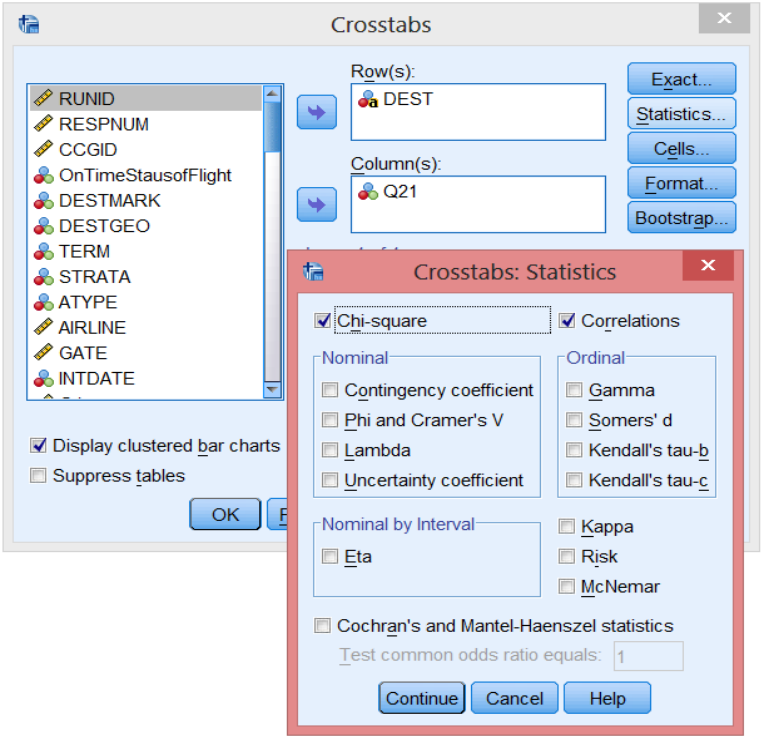
*Output1 [Document1] - IBM SPSS Statistics Viewer

orm Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

```
CROSSTABS
  /TABLES=DEST BY Q21
  /FORMAT=DVALUE TABLES
  /STATISTICS=CHISQ CORR
  /CELLS=COUNT COLUMN
  /COUNT ROUND CELL
  /BARCHART.
```

DEST * Q21 Crosstabulation

		Q21						
		0	1	2	3	4	5	Total
DEST	TOKYO-NRT	10	8	20	20	13	5	76
		1.3%	1.2%	2.1%	3.1%	1.6%	16.7%	2.0%
	SEATTLE	26	32	46	29	35	2	170
		3.5%	4.7%	4.8%	4.5%	4.4%	6.7%	4.4%
	PHILADELPHIA	26	13	28	18	36	2	123
		3.5%	1.9%	2.9%	2.8%	4.5%	6.7%	3.2%
	NEW YORK-JFK	45	25	53	33	64	2	222
		6.0%	3.7%	5.5%	5.1%	8.0%	6.7%	5.7%
	LOS ANGELES	39	53	72	48	32	1	245
		5.2%	7.8%	7.5%	7.4%	4.0%	3.3%	6.3%
	LONG BEACH	5	16	20	9	11	1	62
		0.7%	2.3%	2.1%	1.4%	1.4%	3.3%	1.6%
	LONDON-HEATH	21	14	19	12	17	2	85
		2.8%	2.0%	2.0%	1.8%	2.1%	6.7%	2.2%
	LAS VEGAS	33	34	35	25	23	2	152
		4.4%	5.0%	3.6%	3.8%	2.9%	6.7%	3.9%
	KLAMATH FALL	5	3	5	0	1	0	14
		0.7%	0.4%	0.5%	0.0%	0.1%	0.0%	0.4%
	KAHULUI	5	3	8	6	4	0	26
		0.7%	0.4%	0.8%	0.9%	0.5%	0.0%	0.7%
	HOUSTON	15	20	23	19	24	0	101
		2.0%	2.9%	2.4%	2.9%	3.0%	0.0%	2.6%
	HONOLULU	24	26	24	7	16	0	97
		3.2%	3.8%	2.5%	1.1%	2.0%	0.0%	2.5%



- Q21_Income Group column: 1=Under 50,000, 2=$50,000 - $100,000, 3=$100,001 - $150,000, 4=Over $150,000, 5=Over $500,000

Airlines in SFO – Crosstab (cont.)

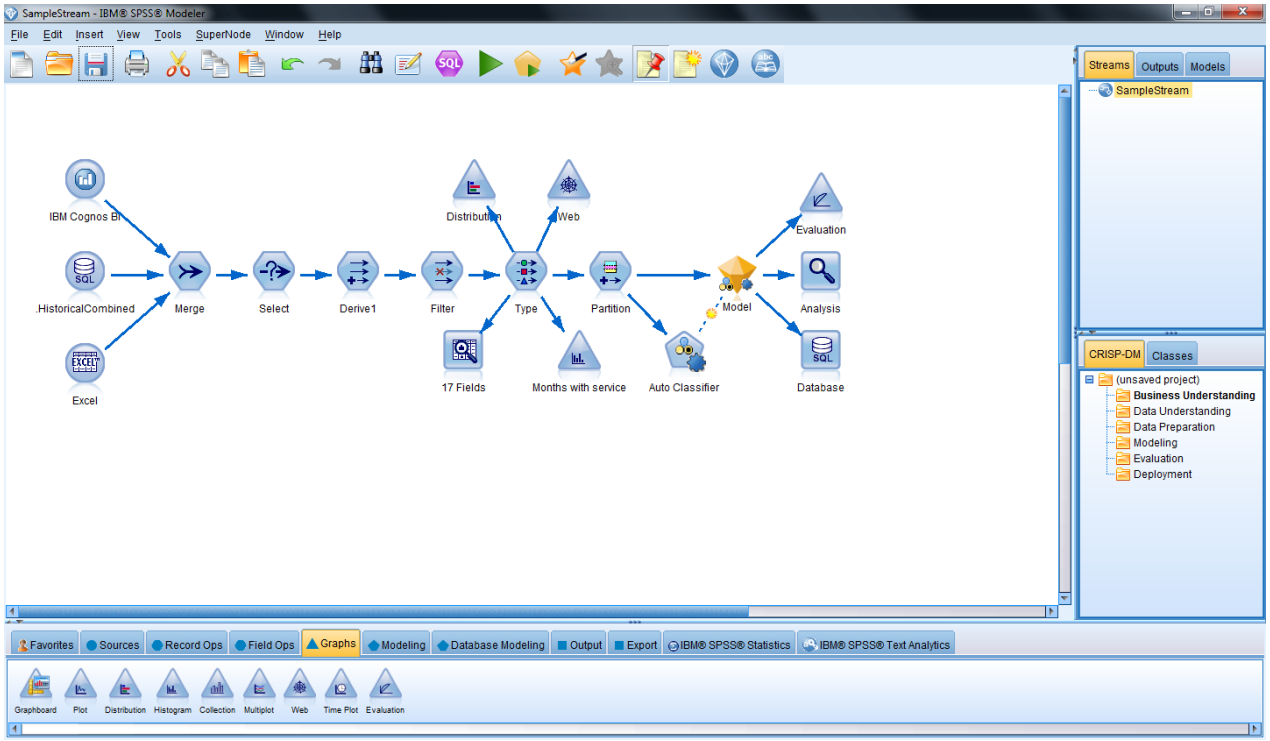
DEST * Q21 Crosstabulation

	Q21						
	0	1	2	3	4	5	Total
TOKYO-NRT	10	8	20	20	13	5	76
	1.3%	1.2%	2.1%	3.1%	1.6%	16.7%	2.0%
SEATTLE	26	32	46	29	35	2	170
	3.5%	4.7%	4.8%	4.5%	4.4%	6.7%	4.4%
PHILADELPHIA	26	13	28	18	36	2	123
	3.5%	1.9%	2.9%	2.8%	4.5%	6.7%	3.2%
NEW YORK-JFK	45	25	53	33	64	2	222
	6.0%	3.7%	5.5%	5.1%	8.0%	6.7%	5.7%
LOS ANGELES	39	53	72	48	32	1	245
	5.2%	7.8%	7.5%	7.4%	4.0%	3.3%	6.3%
LONG BEACH	5	16	20	9	11	1	62
	0.7%	2.3%	2.1%	1.4%	1.4%	3.3%	1.6%
LONDON-HEATH	21	14	19	12	17	2	85
	2.8%	2.0%	2.0%	1.8%	2.1%	6.7%	2.2%
LAS VEGAS	33	34	35	25	23	2	152
	4.4%	5.0%	3.6%	3.8%	2.9%	6.7%	3.9%
KLAMATH FALL	5	3	5	0	1	0	14
	0.7%	0.4%	0.5%	0.0%	0.1%	0.0%	0.4%
KAHULUI	5	3	8	6	4	0	26
	0.7%	0.4%	0.8%	0.9%	0.5%	0.0%	0.7%
HOUSTON	15	20	23	19	24	0	101
	2.0%	2.9%	2.4%	2.9%	3.0%	0.0%	2.6%
HONOLULU	24	26	24	7	16	0	97
	3.2%	3.8%	2.5%	1.1%	2.0%	0.0%	2.5%

- Q21_Income Group column: 1=Under 50,000, 2=$50,000 - $100,000, 3=$100,001 - $150,000, 4=Over $150,000, 5=Over $500,000

- It is obvious that Airlines should provide First Class and Business Class on the top 8 routes
- On the remaining routes it is not necessary

Airlines in SFO – Scoring Model



'Partition'	1_Training		2_Testing		3_Validation	
Correct	33,145	96.33%	9,402	95.82%	4,814	96.22%
Wrong	1,261	3.67%	410	4.18%	189	3.78%
Total	34,406		9,812		5,003	

HCDE

Challenges

- Increase its high school graduation rate
- Intervene with at-risk students early enough to prevent them from dropping out
- Analyze 23,000 text-based surveys and other data
- Provide the right data to make daily decisions that will help their students achieve a brighter future

10%

the increase in graduation rates

25 hours

cut off the workload of creating each report

HCDE

MINDSTREAM ANALYTICS

Benefits from using Predictive Analytics

- Teachers and administrators can now identify at-risk students and constructively intervene with personalized assistance
- Predict which intervention activities will have the optimal impact on students
- Produces year-over-year improvements in behavior and attendance

10%

the increase in graduation rates

25 hours

cut off the workload of creating each report

XO Communications

Challenges

- Telecommunications companies need to control churn
- Numerous small or mid sized customers to manage
- Higher propensity to churn than large customers
- Inefficient to reach out to each customer
- No clear reliable means to identify customers at risk

376%

return on investment

5 Months

The time it took to pay back the investment

$3M

Average annual benefit

XO Communications

Results using Predictive Analytics

- Customers are scored on likelihood to churn per month
- Through the BI web interface, client service managers access predictive data and customer profiles based on territory and prioritize customer outreach

376%
return on investment

5 Months
The time it took to pay back the investment

$3M
Average annual benefit

Centerstone Research Institute

MINDSTREAM ANALYTICS

CRI CENTERSTONE RESEARCH INSTITUTE

Challenges

- Want to ensure that patients actually receive the benefits of new breakthroughs
- Patients receive correct diagnoses and treatment less than 50 percent of the time on first pass through system
- Want to use emergent analytics technologies to connect researchers and healthcare providers

42%

improvement in patient outcomes

58%

anticipated reduction in cost per unit of outcome change

Centerstone Research Institute

Benefits from Predictive Analytics

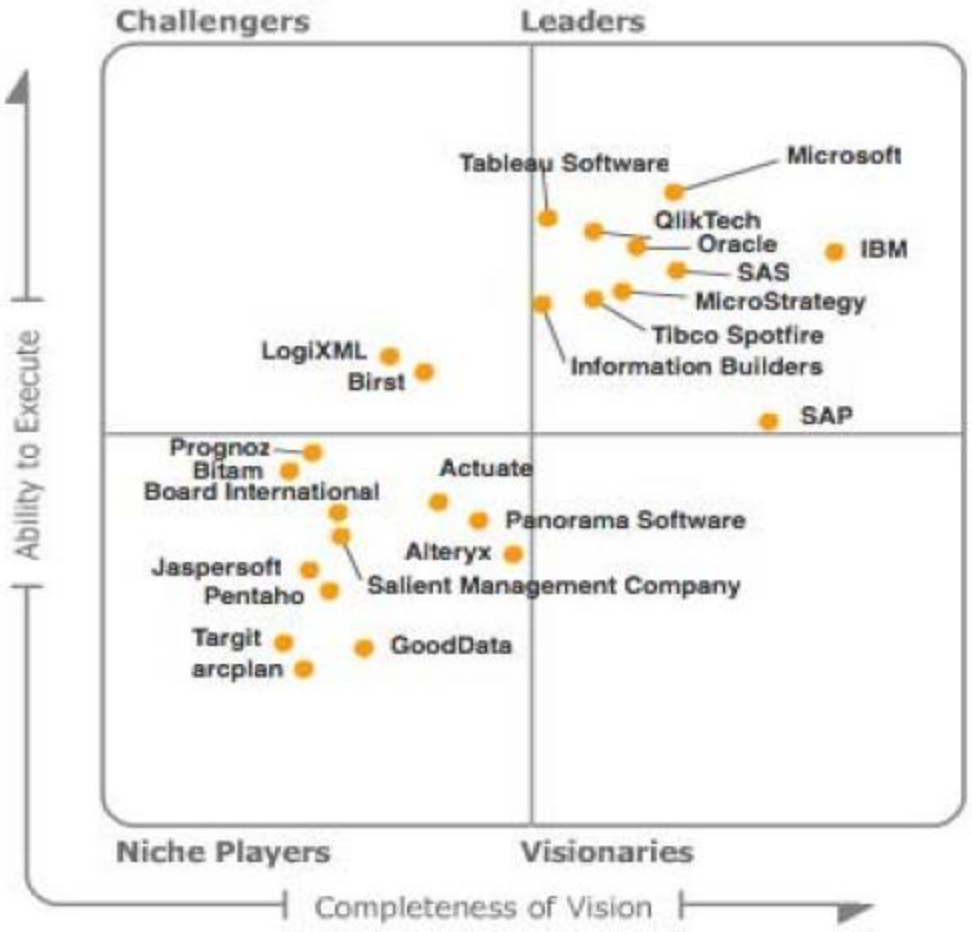
- Created predictive models to assess the effectiveness of various treatment options based on thousands of patients
- Foresight into how treatment would work over time
- Improved operating costs, productivity within clinics and insurance reimbursements

42%
improvement in patient outcomes

58%
anticipated reduction in cost per unit of outcome change

Conclusion

Magic Quadrant for Business Intelligence & Analytics Platforms



Recent Clients

MINDSTREAM ANALYTICS

Pharmaceuticals	Media / Telco	Financial Services	Consumer	Technology
Bayer HealthCare Consumer Care	verizon	vantiv	Stanley Black & Decker	IBM
gsk	CABLEVISION	AgFirst Farm Credit Bank	Chiquita	salesforce
Celgene	windstream communications	ING DIRECT	AVON	brightcove
Bayer HealthCare Pharmaceuticals	Time Warner Cable	FirstMarblehead	CHANEL	OPERA software
Bayer HealthCare Animal Health	Associated Press	VISA		

Energy	Industrial	Education	Healthcare	Real Estate / REIT
ATCO	tyco	Penn University of Pennsylvania	Boston Scientific	SIMON PROPERTY GROUP, INC.
HESS	CNH	University of Phoenix	Catholic Health Partners	Taubman
noble energy	BARRICK	Chicago Public Schools CPS	Cigna	Iron Mountain
KCP&L energizing life	EXAL	Edison Schools	Brookdale Senior Living	HEALTHCARE REIT
	EXIDE TECHNOLOGIES		HEALTHSOUTH	

Thank You !

Del Rogers
VP of Sales
MindStream Analytics
Cell: 214.417.4613
drogers@mindstreamanalytics.com
Twitter: @deljr

Cuong Nguyen Tien
Business Intelligence Consultant
MindStream Analytics
917-294-3417
ctien@mindstreamanalytics.com
Twitter: @cuongcz