



La piedra angular del Análisis Predictivo

José Ignacio Marín Alberdi
IBM SPSS Client Technical Professional



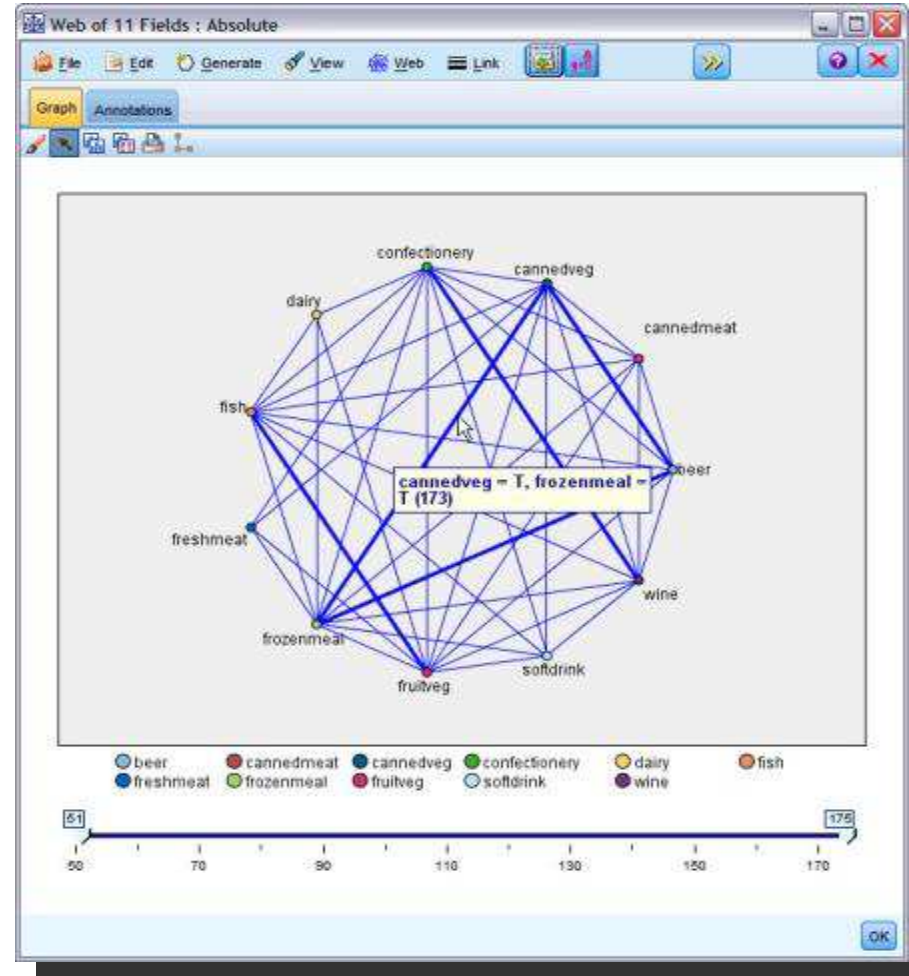
Objetivos de la sesión

- En la sesión de hoy cubriremos los siguientes puntos:
 - Una definición operativa del término “análisis predictivo”
 - Niveles de aplicación del análisis predictivo: desde la simple expansión de las capacidades analíticas a procesos de transformación a nivel de departamento u organización
 - Dos enfoques de análisis predictivo: de arriba abajo y de abajo arriba
 - Cómo IBM SPSS Statistics e IBM SPSS Modeler forman la piedra angular del análisis predictivo

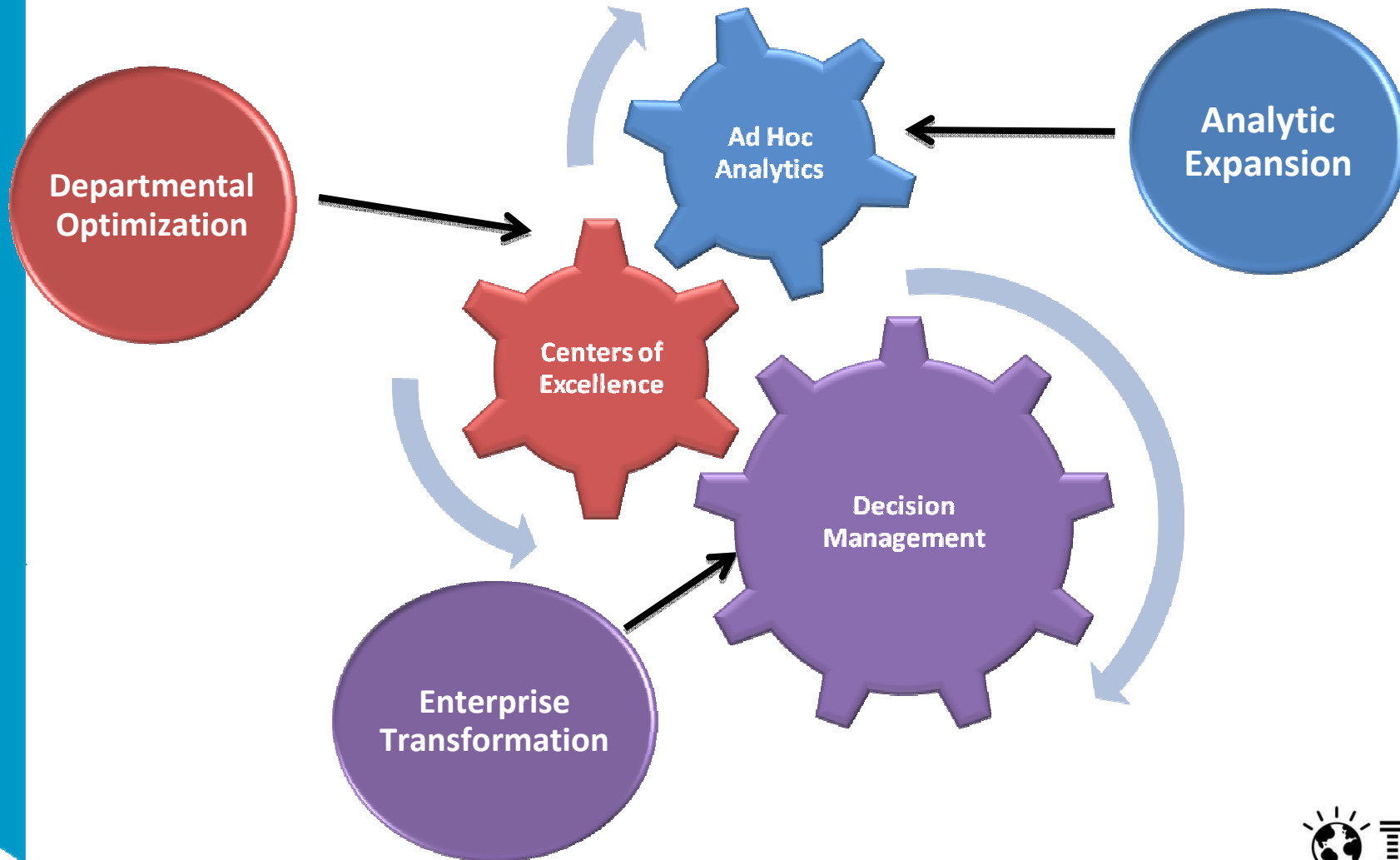
¿Por qué Análisis Predictivo?

- El Análisis Predictivo ayuda a conectar los **datos** con **acciones** eficaces extrayendo conclusiones acerca de las circunstancias actuales y eventos futuros.

• *Gareth Herschel,*
Research Director,
Gartner Group



Las necesidades de negocio dictan el enfoque de Análisis Predictivo más necesario





El ámbito del problema determina el alcance de la solución

Alcance Necesidad	Pequeño	Medio	Grande
Expansión capacidades analíticas	Banco de Trabajo de Analítica Avanzada	Banco de Trabajo de Minería de Datos y Texto	Entorno avanzado de Análisis Predictivo
Optimización departamental	Centro de Excelencia de Analítica Avanzada	Centro de Excelencia de Análisis Predictivo	Centro de Excelencia a nivel de toda la organización
Transformación organizacional	Aplicación Predictiva departamental	Entorno de Soluciones configurables	Despliegue en tiempo real con alta criticidad

El ámbito del problema determina el alcance de la solución

Necesidad	Pequeño	Medio	Grande
	Comprender una muestra	Predecir para una población	Gobernanza analítica y Big Data
Expansión capacidades analíticas	Banco de Trabajo de Analítica Avanzada	Banco de Trabajo de Minería de Datos y Texto	Entorno avanzado de Análisis Predictivo
Optimización departamental	Centro de Excelencia de Analítica Avanzada	Centro de Excelencia de Análisis Predictivo	Centro de Excelencia a nivel de toda la organización
Transformación organizacional	Aplicación Predictiva departamental	Entorno de Soluciones configurables	Despliegue en tiempo real con alta criticidad

El ámbito del problema determina el alcance de la solución

Alcance	Pequeño	Mediano	Grande	
Necesidad	Analistas estadísticos con pocas exigencias en tratamiento de datos	Analistas de negocio con exigencias de tratamiento de datos	Colaboración entre diversos tipos de analistas. Retos en el tratamiento de datos	Entorno de trabajo
Expansión de capacidades analíticas	Analítica Avanzada	Minería de Datos y Texto	Análisis Predictivo	
Optimización departamental	Centro de Excelencia de Analítica Avanzada	Centro de Excelencia de Análisis Predictivo	Centro de Excelencia a nivel de toda la organización	
Transformación organizacional	Aplicación Predictiva departamental	Entorno de Soluciones configurables	Despliegue en tiempo real con alta criticidad	

El ámbito del problema determina el alcance de la solución

Alcance	Pequeño	Medio	Grande
Una solución pequeña a nivel de departamento. La solución se basa en una plantilla predefinida	Una infraestructura para optimizar aplicaciones de toma de decisión de forma configurable.	Aplicativos de máximo nivel de exigencia en tiempos de ejecución o criticidad	
capacidades analíticas	Trabajo de	Trabajo de	avanzado de
	Analítica Avanzada	Minería de Datos y Texto	Análisis Predictivo
Optimización departamental	Centro de Excelencia de Analítica Avanzada	Centro de Excelencia de Análisis Predictivo	Centro de Excelencia a nivel de toda la organización
Transformación organizacional	Aplicación Predictiva departamental	Entorno de Soluciones configurables	Despliegue en tiempo real con alta criticidad

Los retos pueden categorizarse en: Clientes, Operaciones, Riesgos/Fraude

Clientes

Adquirir
Desarrollar
Fidelizar



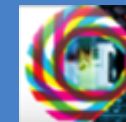
Operaciones

Gestionar
Mantener
Maximizar



Riesgos / Fraude

Monitorizar
Detectar
Controlar



Data Collection

CRM

Modeler

Statistics

Decision Management

Collaboration and Deployment Services

Cognos.
software



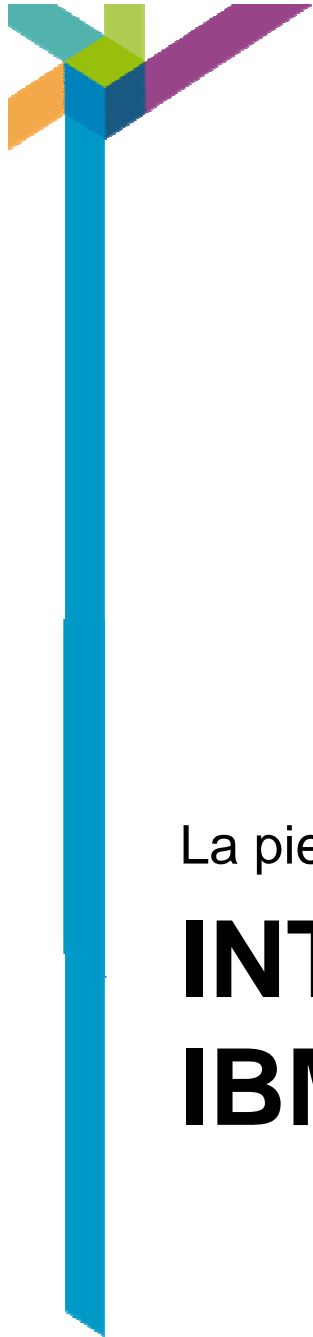
I B M



JRules (ILOG)

IBM Research





La piedra angular del Análisis Predictivo

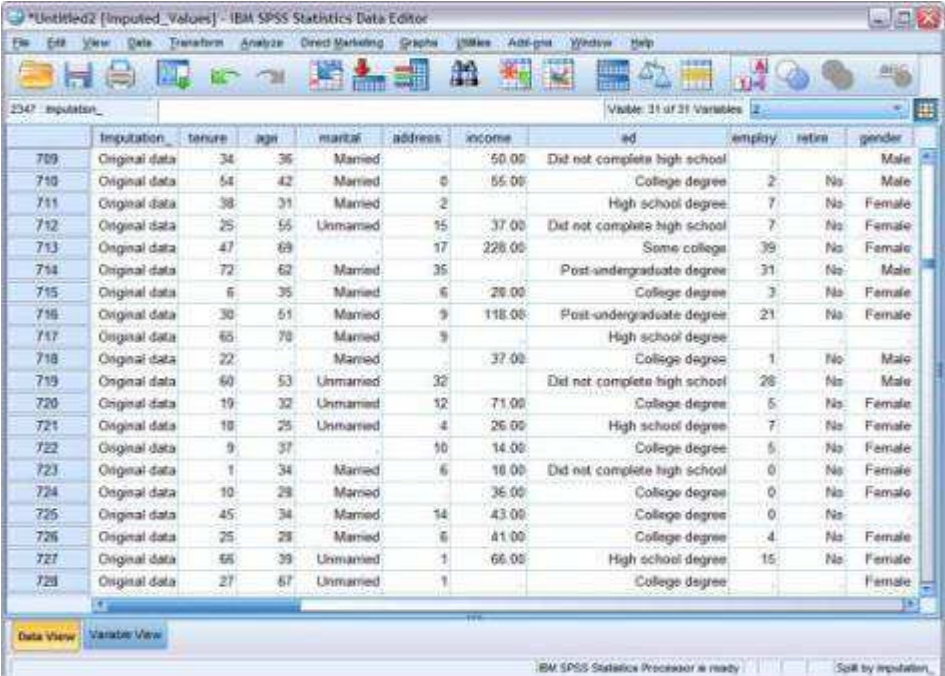
INTRODUCCIÓN A IBM SPSS STATISTICS



IBM SPSS Statistics: contraste de Hipótesis, Análisis Estadístico, enfoque analítico de arriba a abajo

- Modo de trabajo tipo hoja de cálculo
- Herramienta de análisis estadístico de propósito general
- Especialmente indicado para el contraste de hipótesis

- Análisis estadístico descriptivo
- Funciones estadísticas avanzadas
- Muchos tipos de regresión
- Análisis y salidas tabulares



The screenshot displays the IBM SPSS Statistics Data Editor interface. The main window shows a dataset with 21 variables and 2347 observations. The variables are: ingratation, tenure, age, marital, address, income, ad, employ, retire, and gender. The data is presented in a table format with columns for each variable and rows for individual observations. The 'ad' variable shows different education levels, and the 'gender' variable shows 'Male' and 'Female'.

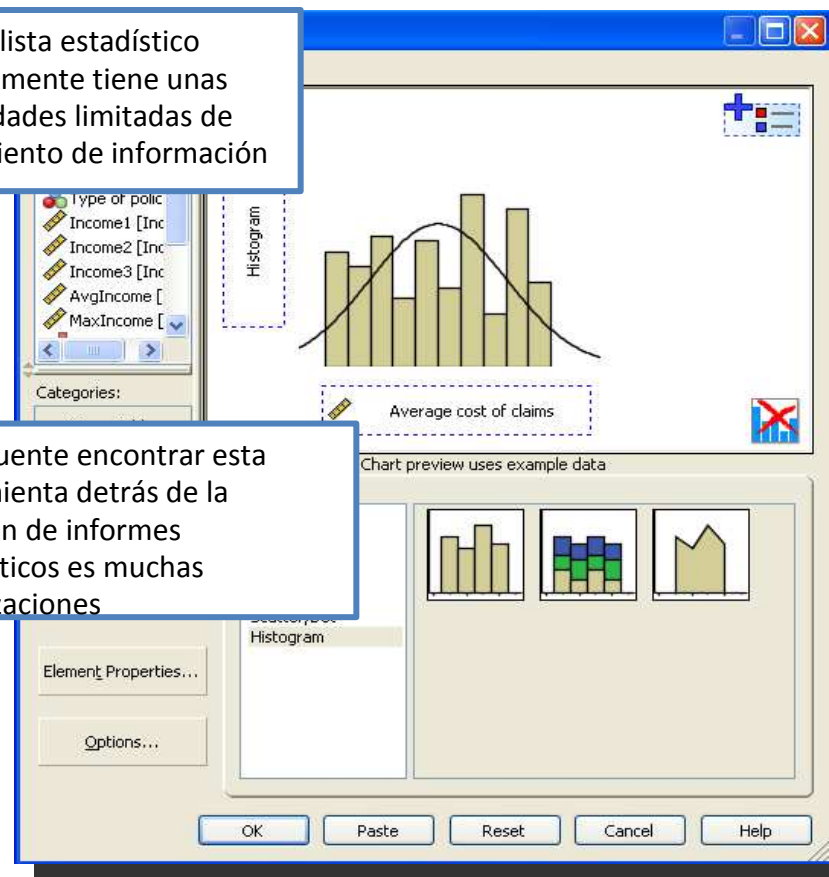
	ingratation	tenure	age	marital	address	income	ad	employ	retire	gender
709	Original data	34	36	Married		50.00	Did not complete high school			Male
710	Original data	54	42	Married	0	55.00	College degree	2	No	Male
711	Original data	38	31	Married	2		High school degree	7	No	Female
712	Original data	25	55	Unmarried	15	37.00	Did not complete high school	7	No	Female
713	Original data	47	69		17	228.00	Some college	39	No	Female
714	Original data	72	62	Married	35		Post-graduate degree	31	No	Male
715	Original data	6	35	Married	6	26.00	College degree	3	No	Female
716	Original data	30	51	Married	9	118.00	Post-graduate degree	21	No	Female
717	Original data	65	78	Married	9		High school degree			
718	Original data	22		Married		37.00	College degree	1	No	Male
719	Original data	60	53	Unmarried	32		Did not complete high school	28	No	Male
720	Original data	19	32	Unmarried	12	71.00	College degree	5	No	Female
721	Original data	18	25	Unmarried	4	26.00	High school degree	7	No	Female
722	Original data	9	37		10	14.00	College degree	5	No	Female
723	Original data	1	34	Married	6	18.00	Did not complete high school	0	No	Female
724	Original data	10	28	Married		36.00	College degree	0	No	Female
725	Original data	45	34	Married	14	43.00	College degree	0	No	
726	Original data	25	28	Married	6	41.00	College degree	4	No	Female
727	Original data	66	39	Unmarried	1	66.00	High school degree	15	No	Female
728	Original data	27	67	Unmarried	1		College degree			Female

IBM SPSS Statistics como plataforma analítica

- Acceder a los datos en cualquier formato
- Funciones de manipulación de la información
- Describir y explorar los datos de forma sencilla
- Análisis factorial y de conglomerados
- Modelos de regresión
- Presentación de los resultados en forma de tablas y de gráficos
- Resultados exportables a otras aplicaciones

Un analista estadístico normalmente tiene unas necesidades limitadas de tratamiento de información

Es frecuente encontrar esta herramienta detrás de la creación de informes estadísticos es muchas organizaciones



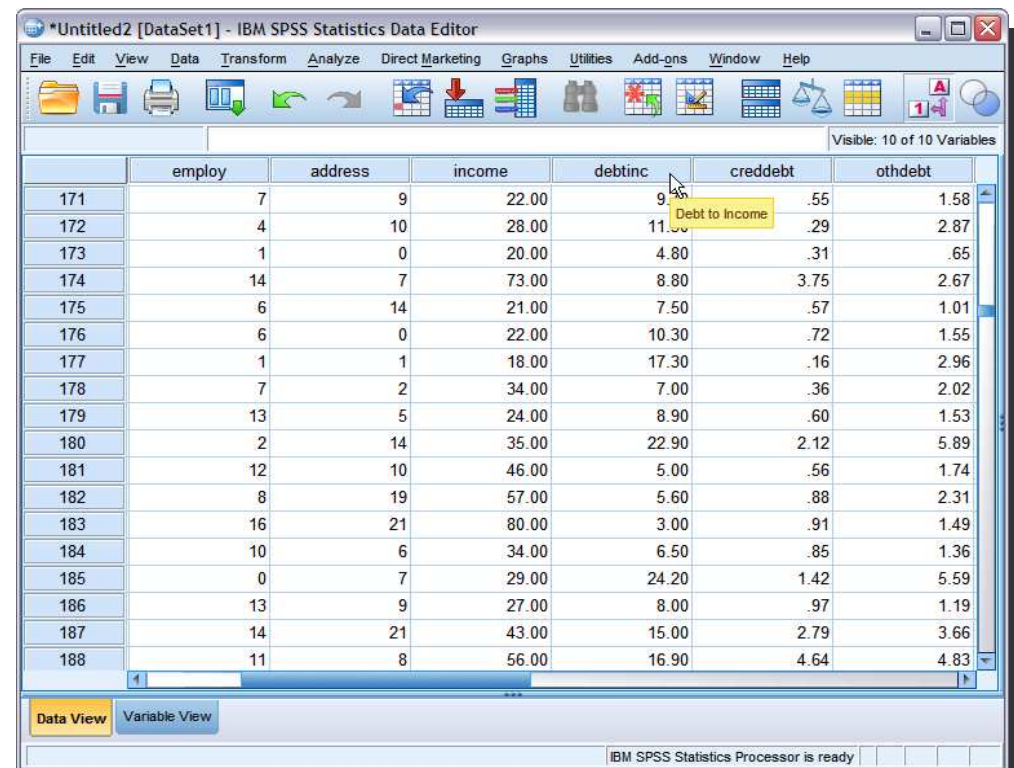
Ejemplo: comprendiendo los datos y analizando tendencias en 5 minutos

- **Reto de negocio**

- Basándome en datos **de mi oficina**, es la edad o el nivel de renta de mis clientes un posible indicador de riesgo a la hora de conceder créditos? Si no, ¿cuáles son los 2 o 3 atributos relevantes de mis clientes que me permiten entender los niveles de riesgo mencionados?

- **Datos disponibles**

- Edad
- Nivel educativo
- Años en empleo actual
- Años en dirección actual
- Nivel de ingresos
- Ratio Deuda / Ingresos
- Deuda en tarjeta de crédito
- Otros créditos



Visible: 10 of 10 Variables

	employ	address	income	debtinc	creddebt	othdebt
171	7	9	22.00	9.55	.55	1.58
172	4	10	28.00	11.00	.29	2.87
173	1	0	20.00	4.80	.31	.65
174	14	7	73.00	8.80	3.75	2.67
175	6	14	21.00	7.50	.57	1.01
176	6	0	22.00	10.30	.72	1.55
177	1	1	18.00	17.30	.16	2.96
178	7	2	34.00	7.00	.36	2.02
179	13	5	24.00	8.90	.60	1.53
180	2	14	35.00	22.90	2.12	5.89
181	12	10	46.00	5.00	.56	1.74
182	8	19	57.00	5.60	.88	2.31
183	16	21	80.00	3.00	.91	1.49
184	10	6	34.00	6.50	.85	1.36
185	0	7	29.00	24.20	1.42	5.59
186	13	9	27.00	8.00	.97	1.19
187	14	21	43.00	15.00	2.79	3.66
188	11	8	56.00	16.90	4.64	4.83

Data View Variable View

IBM SPSS Statistics Processor is ready

Resultado: comprender el comportamiento de mis clientes me hace tomar mejores decisiones

- Los indicadores de riesgo de impago basados en nuestros datos y por nivel de importancia son:
 - Ratio Deuda / Ingresos
 - Años en empleo actual
 - Deuda en tarjeta de crédito
 - Años en la dirección actual

Precisión analítica: 80%

- Podemos usar estas puntuaciones de riesgo para gestionar los préstamos que concedemos

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	debtinc	.121	.017	52.676	1	.000	1.129
	Constant	-2.476	.230	116.315	1	.000	.084
Step 2 ^b	employ	-.140	.023	38.158	1	.000	.869
	debtinc	.134	.018	54.659	1	.000	1.143
	Constant	-1.621	.259	39.038	1	.000	.198
	employ	-.244	.033	54.676	1	.000	.783
Step 3 ^c	debtinc	.069	.022	9.809	1	.002	1.072
	creddebt	.506	.101	25.127	1	.000	1.658
	Constant	-1.058	.280	14.249	1	.000	.347
	employ	-.247	.034	51.826	1	.000	.781
Step 4 ^d	address	-.089	.023	15.109	1	.000	.915
	debtinc	.072	.023	10.040	1	.002	1.074
	creddebt	.602	.111	29.606	1	.000	1.826
	Constant	-.605	.301	4.034	1	.045	.546

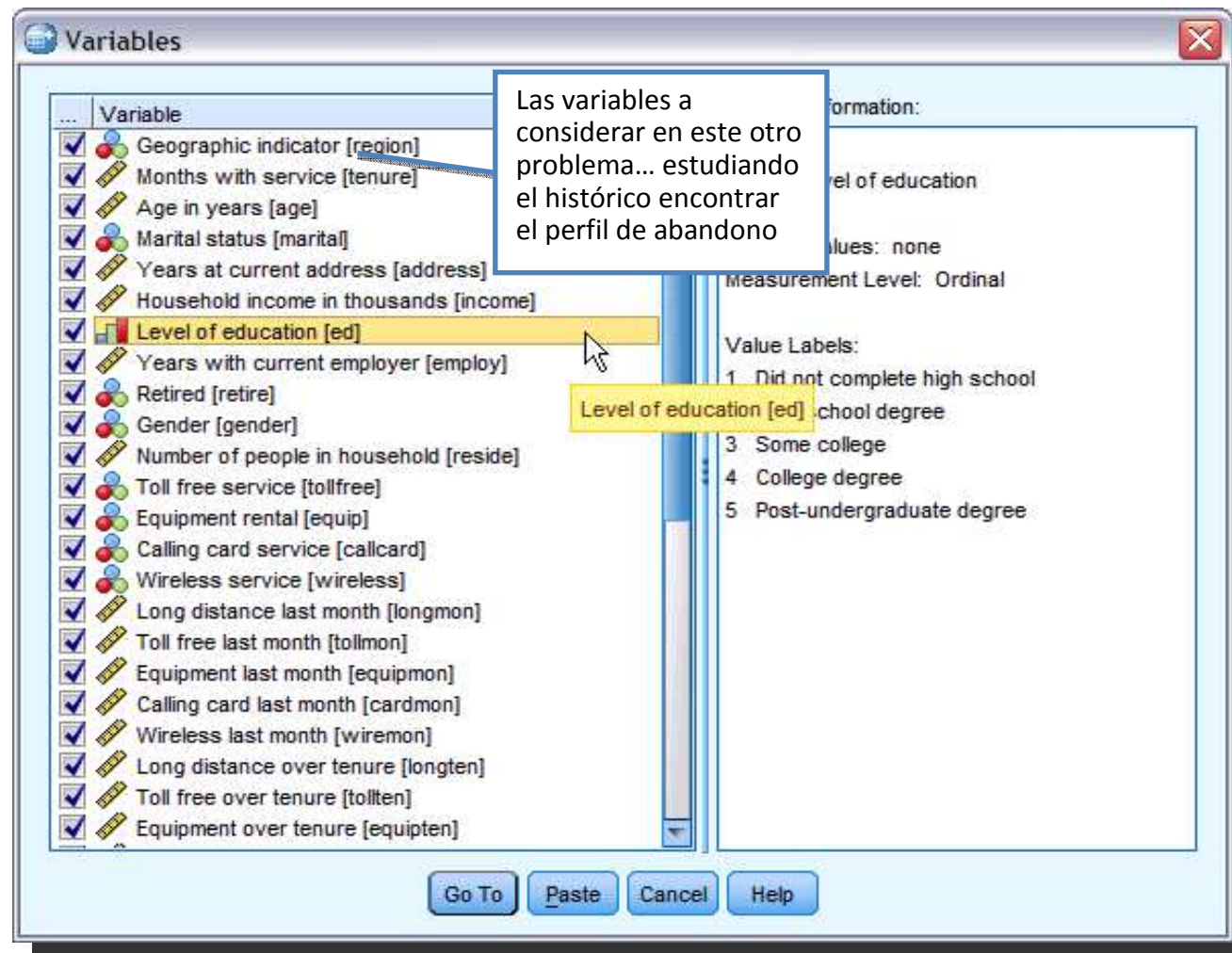
a. Variable(s) entered on step 1: debtinc.
 b. Variable(s) entered on step 2: employ.
 c. Variable(s) entered on step 3: creddebt.
 d. Variable(s) entered on step 4: address.

Classification Table^c

	Observed	Predicted					
		Selected Cases ^a			Unselected Cases ^b		
		default		Percentage Correct	default		Percentage Correct
0	1	0	1				
Step 1	default 0	361	14	96.3	137	5	96.5
	1	100	24	19.4	45	14	23.7
	Overall Percentage			77.2			75.1
Step 2	default 0	351	24	93.6	136	6	95.8
	1	80	44	35.5	36	23	39.0
	Overall Percentage			79.2			79.1
Step 3	default 0	348	27	92.8	135	7	95.1
	1	72	52	41.9	28	31	52.5
	Overall Percentage			80.2			82.6
Step 4	default 0	352	23	93.9	130	12	91.5
	1	67	57	46.0	27	33	54.2
	Overall Percentage			82.0			80.6

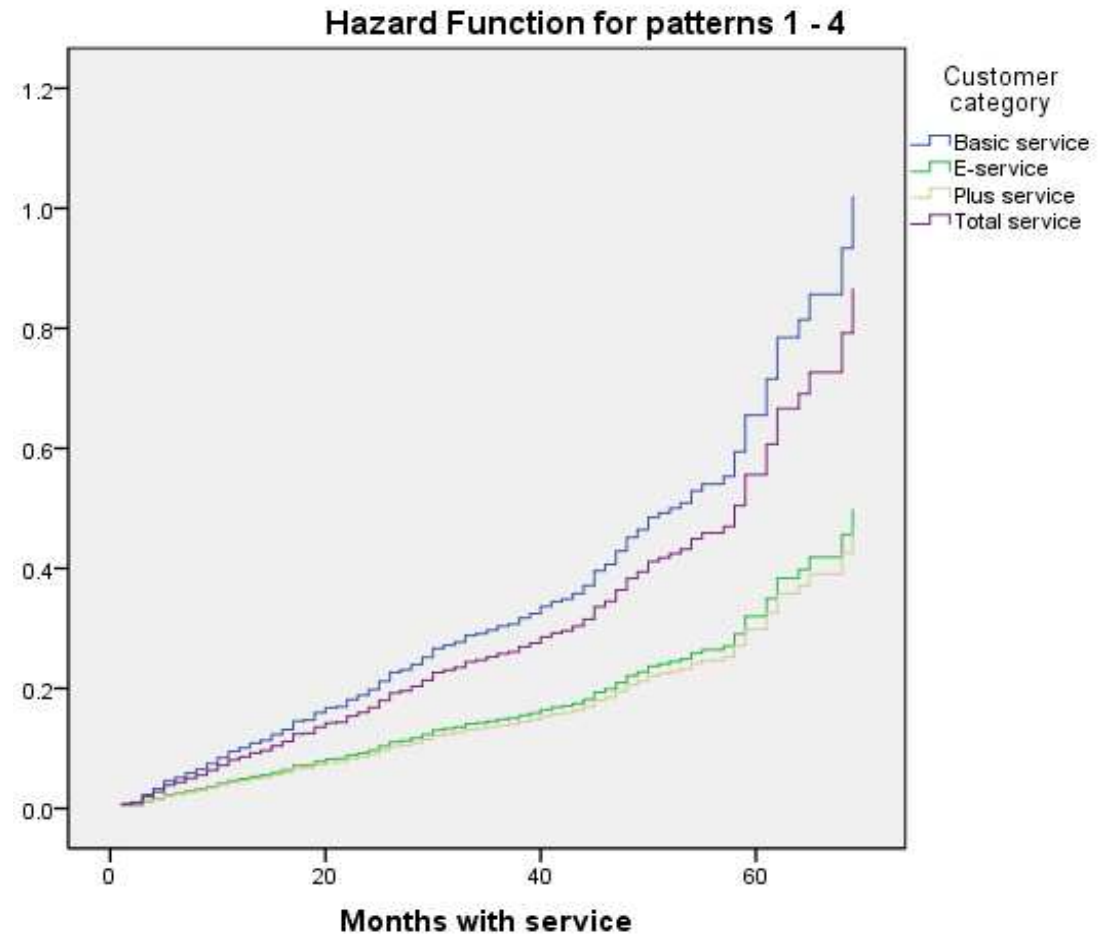
a. Selected cases validate EQ 1
 b. Unselected cases validate NE 1
 c. The cut value is .500

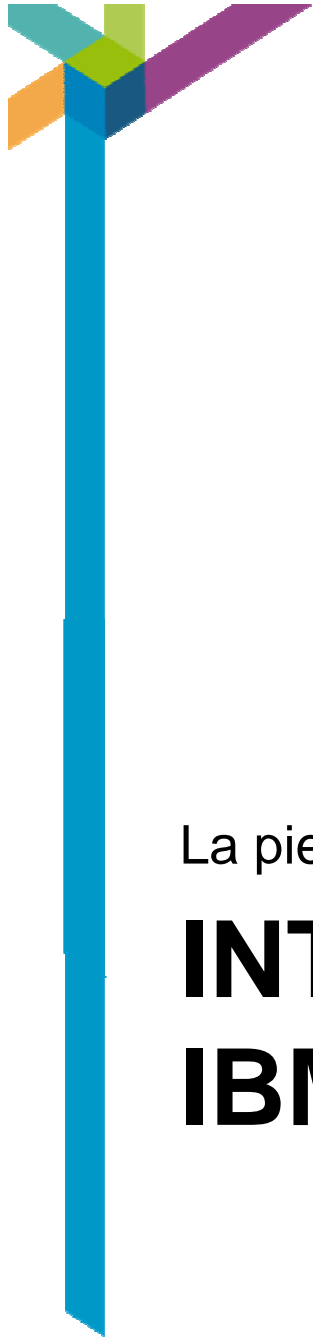
Otro ejemplo: análisis de supervivencia para encontrar perfiles de abandono



Conclusión: necesitamos mejores esfuerzos en determinado segmento los primeros meses

- Los clientes de los servicios “Basic Service” y “Total Service” son más propensos a abandonar y lo hacen más tempranamente
- La tendencia se marca más a medida que pasan los meses, la probabilidad de abandono casi se dobla al superar la marca de los 2 años según los datos





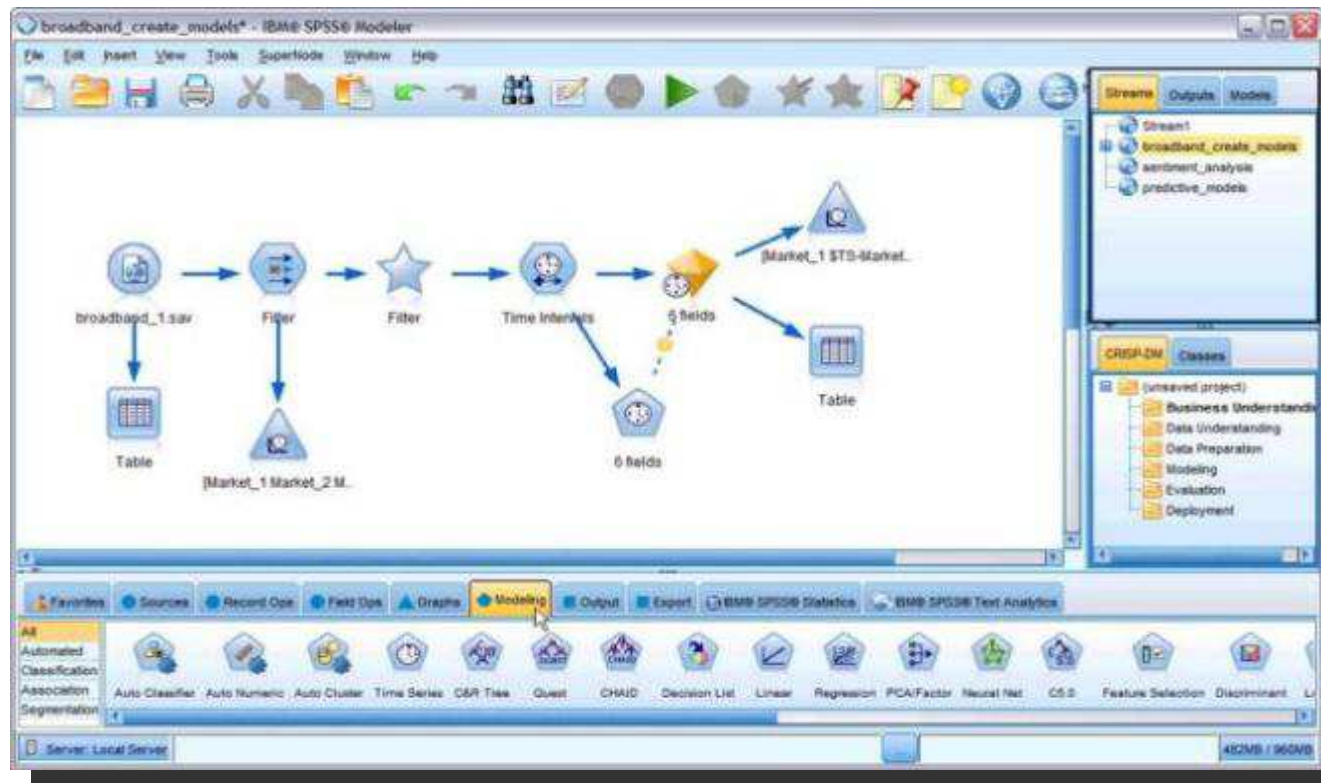
La piedra angular del Análisis Predictivo

INTRODUCCIÓN A IBM SPSS MODELER

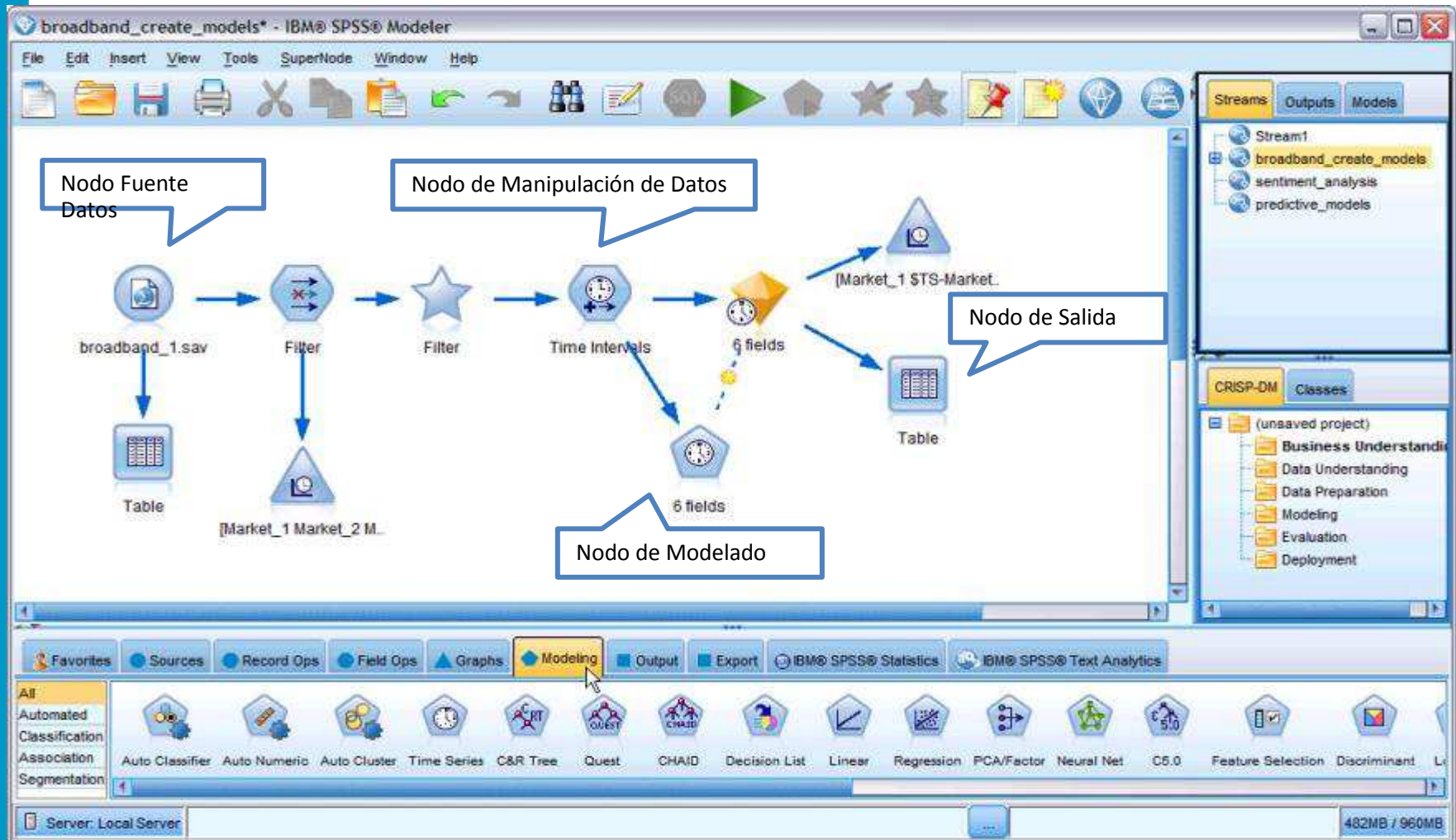


IBM SPSS Modeler: minería de datos y de texto, descubrir patrones, enfoque analítico de abajo a arriba

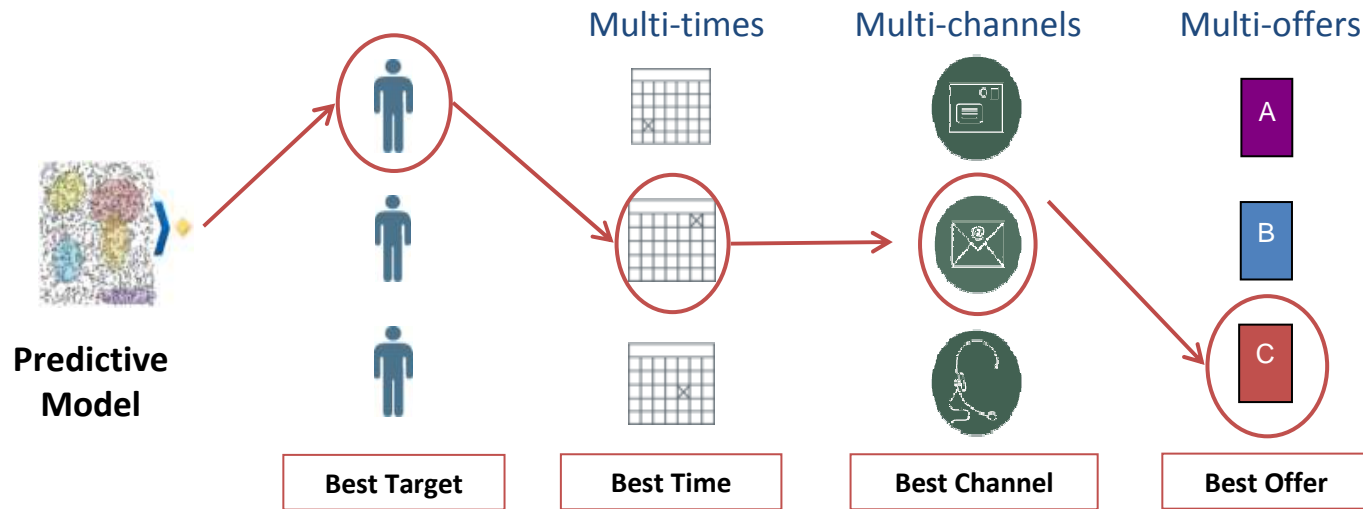
- IBM SPSS Modeler es una herramienta de minería de datos y texto de propósito general que pone el máximo foco en la sencillez de uso y el paradigma de la programación visual y asistida de cara al usuario de negocio



IBM SPSS Modeler como plataforma analítica



Un ejemplo: First Tennessee Bank



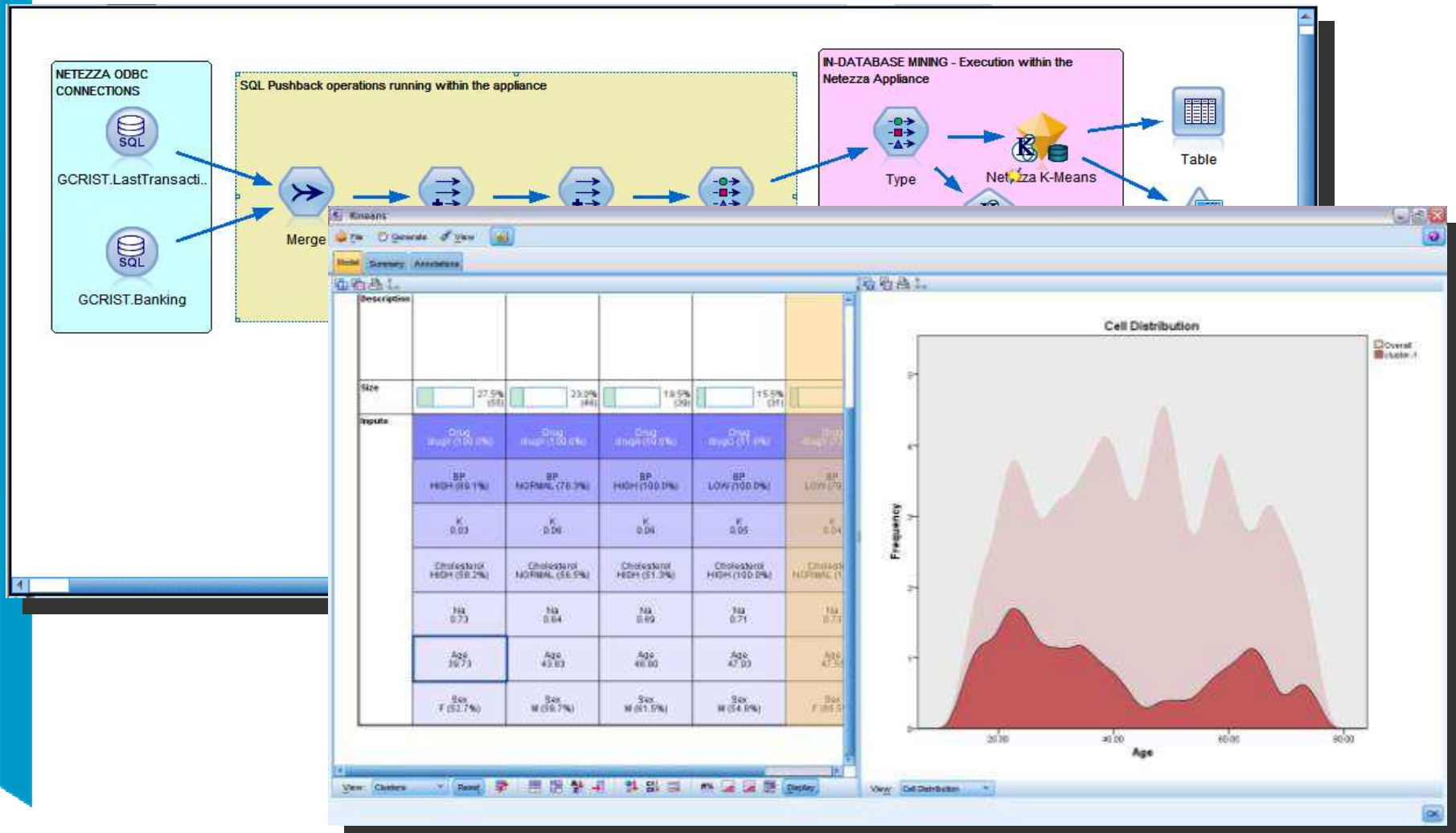
Potential Actions	Business Rules	Margin	Response Probability	Expected Value
A	✗			
B	✓	90	54%	49
C	✓	200	32%	64

Foco en el analista de negocio, preparación automatizada de datos y modelos

The screenshot displays the 'Auto Data Prep' software interface. The main window, titled 'Customer_Status', shows a list of generated models with their performance metrics. The 'Model' column includes icons for each model type. A tooltip 'Double click to view model details' is visible over the 'C5 1' model row.

Us...	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in	Lift{...}	Overall Accuracy (%)	No. Fields Used	Area Under Curve
<input checked="" type="checkbox"/>		Bayesian Network 1	< 1	70	50	1.609	61.667	39	0.742
<input checked="" type="checkbox"/>		CHAID 1	< 1	70	37	1.609	75	6	0.777
<input checked="" type="checkbox"/>		C&R Tree 1	< 1	75	45	1.567	76.667	11	0.806
<input checked="" type="checkbox"/>		C5 1	< 1				71.667	10	0.737
<input checked="" type="checkbox"/>		Decision List 1	< 1	40	44	1.353	65	1	0.648
<input checked="" type="checkbox"/>		Discriminant 1	< 1	55	56	1.149	61.667	2	0.675
<input checked="" type="checkbox"/>		Logistic regression 1	< 1	0	1	1	0	39	0.5

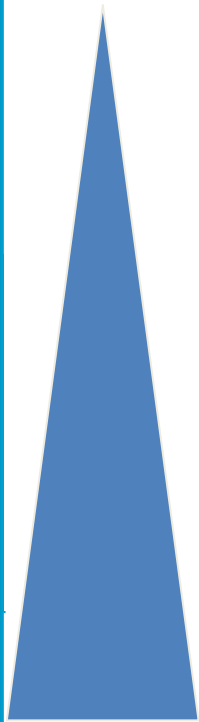
La filosofía de plataformas flexibles permite alcanzar un rápido ROI explotando los datos de la organización





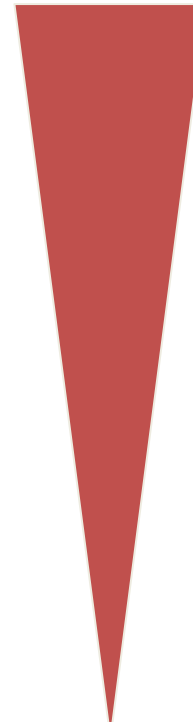
El Análisis Predictivo recoge ambos enfoques

Enfoque arriba a abajo



- El Análisis Estadístico requiere
 - Tener una teoría que explique las relaciones propuestas
 - Derivar una hipótesis de la teoría
 - Contrastar esa hipótesis con Análisis Estadístico
- Es un enfoque para analizar los datos manual y guiado por el usuario

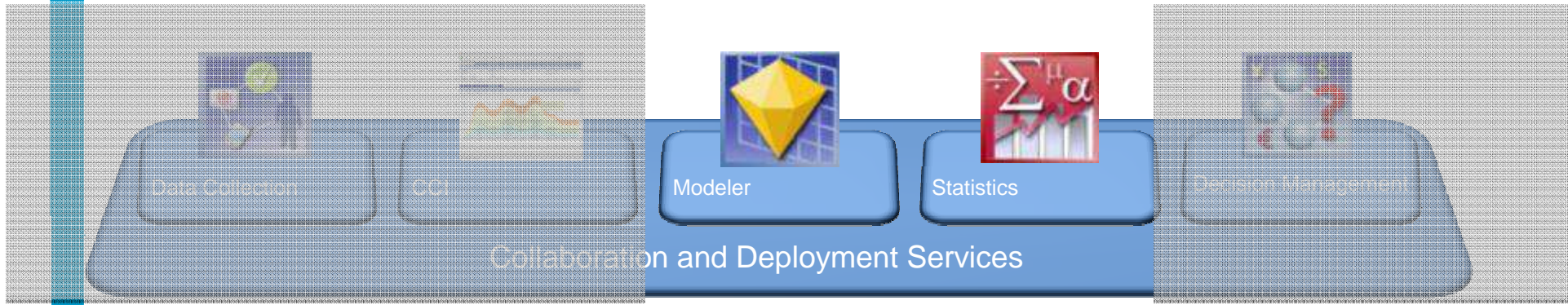
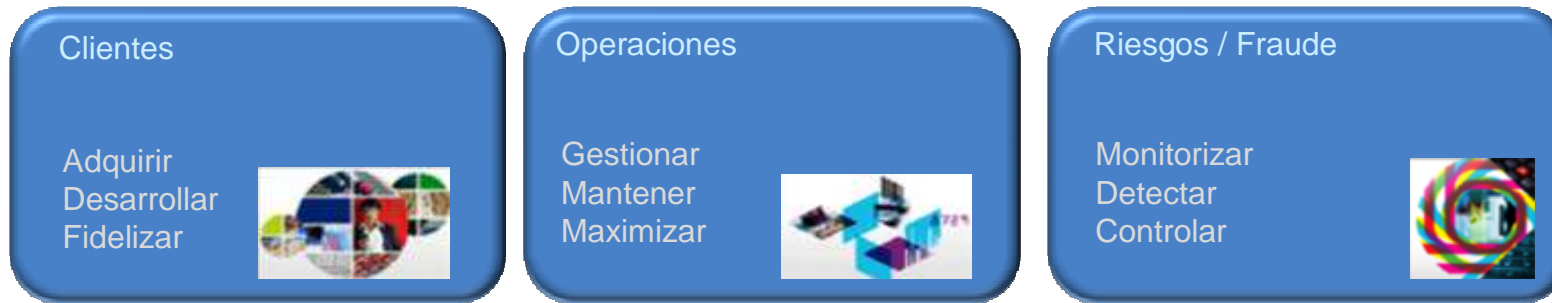
Enfoque abajo arriba



- La minería de datos extrae patrones de los datos mismos en un proceso guiado por el algoritmo y no el usuario
- Es un enfoque centrado en los datos, auto-organizado, de abajo arriba empleado en datos potencialmente masivos

La combinación de ambos enfoques es la fuerza del Análisis Predictivo

La piedra angular del Análisis Predictivo





Thank You





IBM®

