



Sobre un caso de Detección de Irregularidades

Estanislao Martin Irigoyen

Director – Intelgentia SRL

emirigoyen@intelligentia.com.ar

Agenda



- Definición de Analytics.
- El problema inicial.
- **IVEMIS-*ia***.
- Un caso de aplicación en Detección de Fraudes, enmarcado en la metodología **IVEMIS-*ia***.

¿Que es Analytics? Definición Funcional

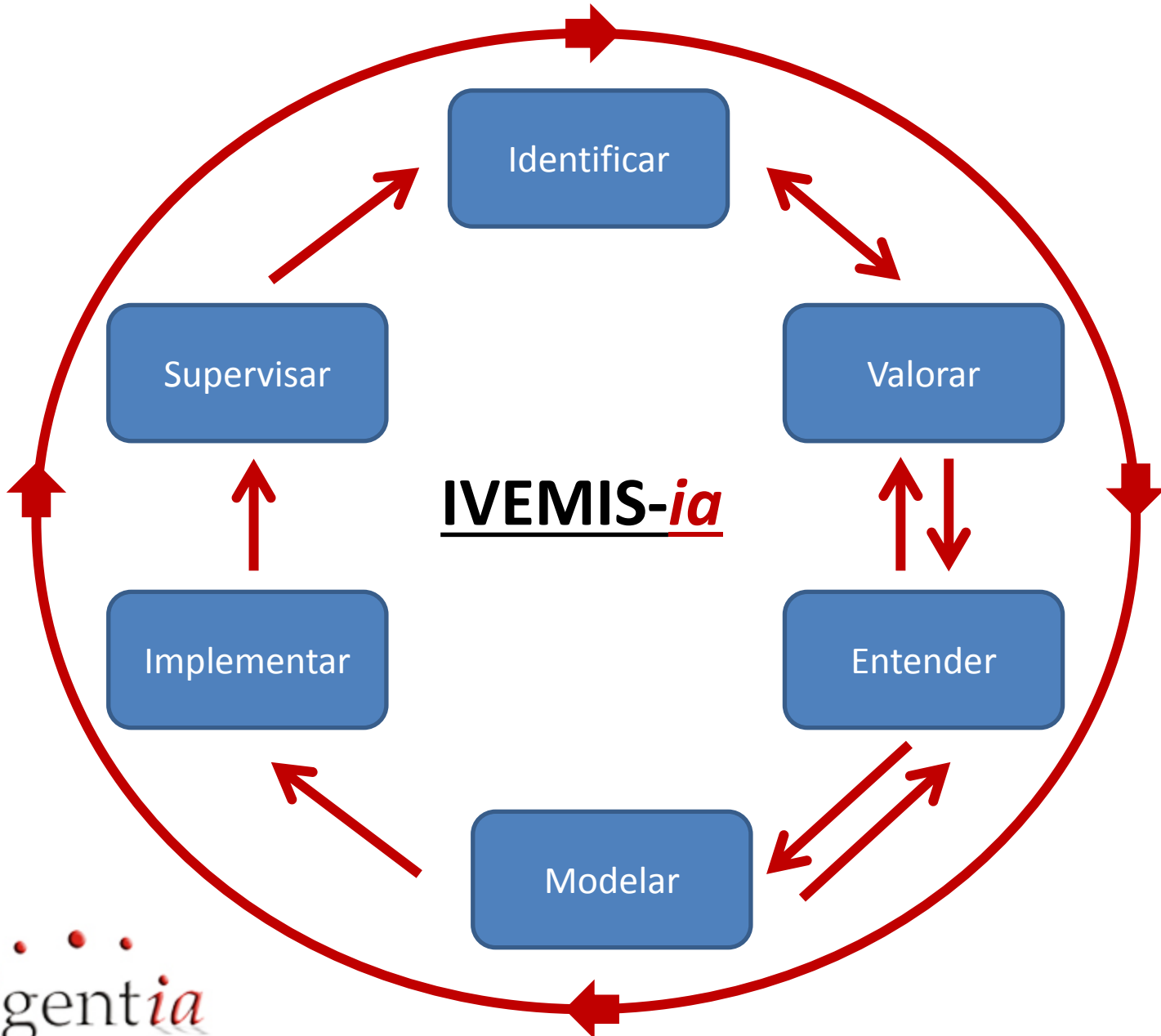
Analytics es una disciplina que tiene como fin “**explotar**” los datos almacenados en los sistemas de la empresa, “**extraer**” patrones estables allí ocultos, “**expresarlos**” en **modelos legibles, operativos y accionables** para luego “**transformarlos**” en conocimientos del negocio que mejoren la rentabilidad del mismo.

Analytics: El problema inicial.



Analytics es una disciplina compleja:

- Alto grado de «Multidisciplinaredad».
 - Profundos conocimientos del negocio.
 - Profundos conocimientos técnicos.
- La «Espectacularidad» de las soluciones conlleva un «vuelo hacia la estratosfera» por parte de los recursos No Técnicos.
 - La fijación de expectativas puede situarse en lugares inalcanzables!.
- No debe separarse al negocio, del proceso de desarrollo de la solución.
- Debe adoptarse una metodología de trabajo integral y funcional, que integre a todas las disciplinas.



Analytics: IVEMIS-*ia*

- Identificar: Detectar un problema del negocio.
 - Valorar: Cuantificar el impacto de la resolución del problema.
 - Entender: Comprender y Contextualizar el problema.
 - Modelar: Desarrollar la Solución Analítica.
 - Implementar: Poner en Producción la Solución desarrollada.
 - Supervisar: Controlar el correcto funcionamiento de la Solución Implementada.

Analytics: *IVEMIS-ia*

Identificar	Valorar	Entender	Modelar	Implementar	Supervisar
Detectar y/o Presumir la Existencia de un Problema.	Cuantificar el Impacto actual del Problema.	Determinar el GAP existente.	Realizar las Integraciones, Procesos y Transformaciones sobre los Datos.	Poner en Produccion el Modelo Obtenido.	Realizar un Control Continuo del modelo implementado.
Definir el Problema de Negocios con Claridad.	Estimar el valor aportado por una solucion (Fijar Expectativas).	Evaluar Riesgos y Mitigantes.	Ajustar los Modelos utilizando el/los Algoritmos seleccionados.	Informar el Conocimiento Adquirido.	Controlar: Estabilidad, Discriminacion, Precision, Exactitud...
	Estimar ROI (Relacion Costo/Beneficio).	Contextualizar el problema.	Evaluar el Modelo en cuanto a: Ajuste, Coherencia, Precision, Exactitud, Poder de Discriminacion, Estabilidad, Interpretabilidad.		
	Decidir si se encarara o no la solucion.	Comprender Reglas, Limitaciones, Exclusiones, Caracteristicas, Informacion Disponible (Fuentes de Datos, Variables).	Evaluar contra Resultados Pasados (Que hubiese pasado si?)		
	CRITERIO de SUBORDINACION: 1-Facturacion -> 2-Costo -> 3-ROI	Determinar Estrategia de Modelado.			
		Determinar "TAREA ANALITICA" a realizar.			
		Preseleccionar las Tecnicas (Algoritmos).			
		Definir el Objetivo Tecnico.			
		Explorar los Datos disponibles.			
		Definir un plan de accion para el Procesamiento de Datos.			
		Determinar Metricas de Evaluacion.			
		Definir Criterios de Éxito.			
		Determinar el Volumen de Datos Necesario para Modelar.			
		Determinar la Necesidad y/o La Pertinencia del Muestreo.			

Entonces..... Había una vez.....



IDENTIFICAR

- Detectar y/o Presumir la Existencia de un Problema.
- Definir el Problema de Negocios con Claridad.

« Tenemos la presunción de la existencia de abusos y/o irregularidades en el consumo de medicamentos por determinados grupos y/o bandas que se aprovechan de los servicios que prestamos para hacer negocios espurios, generar consumos para no afiliados y otras tantas posibilidades mas.....

Deseamos poder:

- Caracterizar los comportamientos de los usuarios.
- Detectar estos comportamientos.
- Identificar a dichos afiliados, a fin de poder accionarlos.»



VALORAR

- Cuantificar el Impacto actual del Problema.
- Estimar el valor aportado por una solución (Fijar Expectativas).
- Estimar ROI (Relación Costo/Beneficio).
- Decidir si se encarara o no la solución.

Grupo	Costo x Usuario	Lift
Base Completa	\$1.000,00	1,00
Según Pareto	\$4.000,00	4,00
TOP 1%	\$11.500,00	11,50

El top 1% gasta en promedio 11 veces y media lo que gasta el promedio, implicando esto un gasto de \$xxx.xxx.xxx,00, si se lograsen reducir los costos de este TOP 1% en un 10% el ahorro seria de aproximadamente \$xx.xxx.xxx,00 Mensuales, lo cual da las pautas de que **resulta una buena decisión encarar una solución a este problema.**

ENTENDER



- **Determinar el GAP existente:**
 - El Cliente no posee practica continua en Analytics, pero realiza análisis de información.
 - Tiene Motores de Bases de Datos, pero no Herramientas de Data Mining.
 - Tiene algunos conocimiento In House y desea mejorarlo.
 - Hoy en día, aplica reglas simples para la caracterización de los Abusos y con dichas reglas, selecciona casos que son enviados a una empresa que realiza la auditoria correspondiente y devuelve resultados marcando a los casos como Irregulares o Desestimados.
- **Evaluar Riesgos y Mitigantes.**
- **Contextualizar el problema.**
- **Comprender Reglas, Limitaciones, Exclusiones, Características, Información Disponible (Fuentes de Datos, Variables).**

Se analizaran usuarios que consuman medicamentos. El análisis de limitara a una Sucursal. Se Excluirán las Internaciones Domiciliarias y en la selección se quitaran a los afiliados con discapacidad y a los empleados de la empresa.

Los Usuarios pueden consumir productos con tres tipos de descuentos: 100% para PMI, 70% para usuarios con CRONICIDADES DECLARADAS y 40% al consumo general con Receta Medica.

Se dispone de la Información de los Afiliados, de sus características Filiatorias y de sus transacciones. Se utilizara una ventana histórica de 6 meses.

ENTENDER



- **Determinar Estrategia de Modelado.**

Se trabajara con una estrategia tipo «DIVIDE y VENCERAS» por considerarla la mas adecuada al caso, separando comportamiento Transaccional e Intensidad de Consumos.

- **Determinar "TAREA ANALITICA" a realizar.**

Se trabajara con una TAREA de SEGMENTACION.

- **Preseleccionar las Técnicas (Algoritmos).**

Se utilizara el Algoritmo de K-Medias para agrupar a los afiliados.

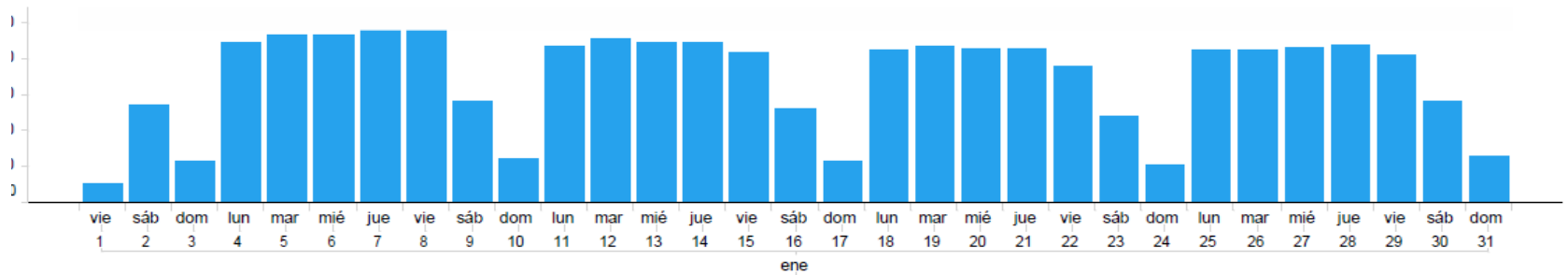
- **Definir el Objetivo Técnico.**

Marcar a los usuarios según la pertenencia (o no!) a determinados grupos.

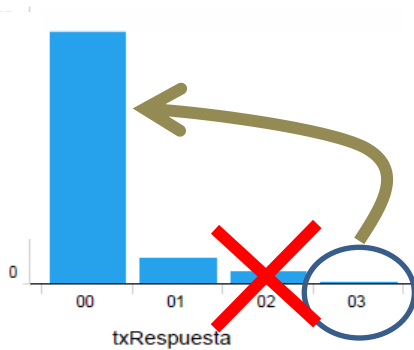
ENTENDER

- Explorar los Datos Disponibles.

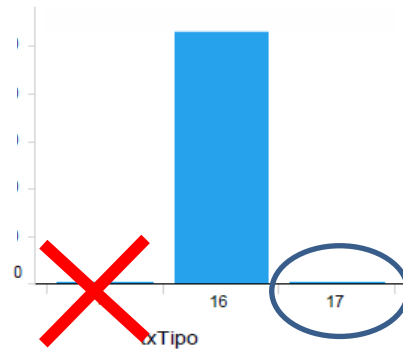
Transacciones x Fecha



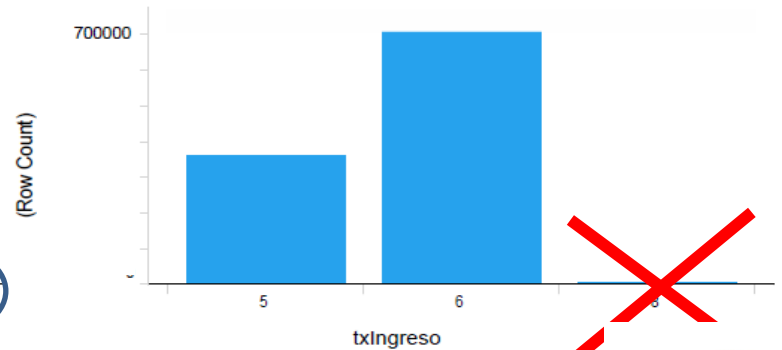
Transacciones x Respuesta



Transacciones x Tipo



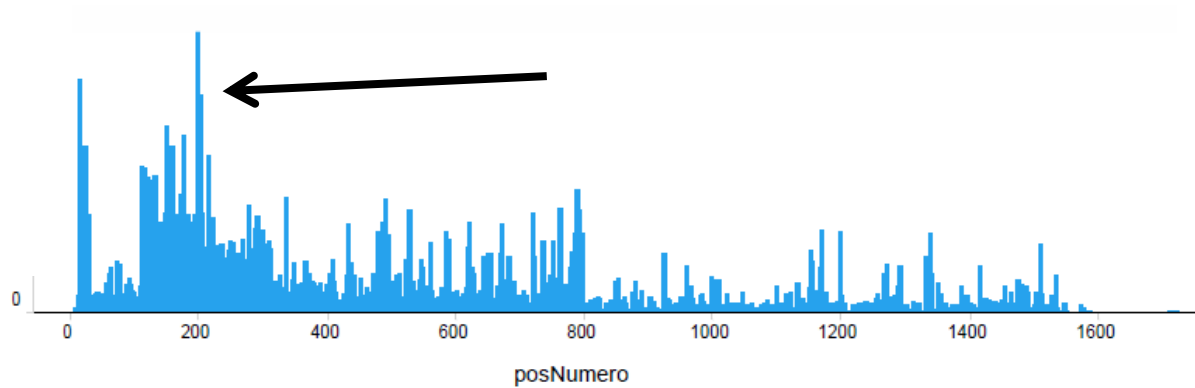
Transacciones x Ingreso



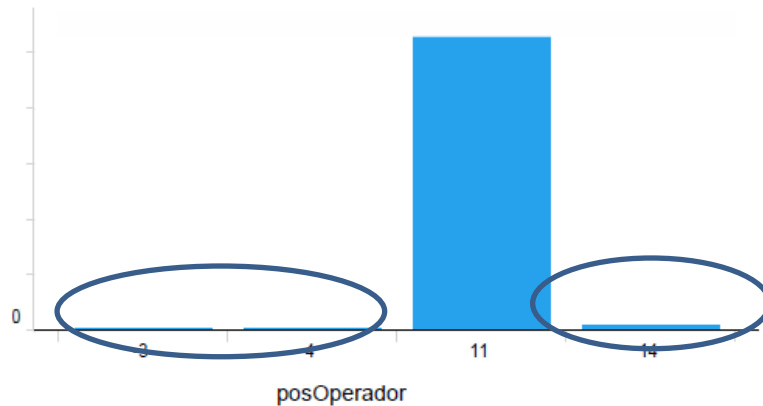
ENTENDER

- Explorar los Datos Disponibles.

TXs x Numero POS



TXs x Operador



TXs x Filial POS

posFilial	Casos
38
24	
2	
32	
14	
27	
25	
30	
11	
28	
37	1115
15	792
23	757

Casos por Prestador

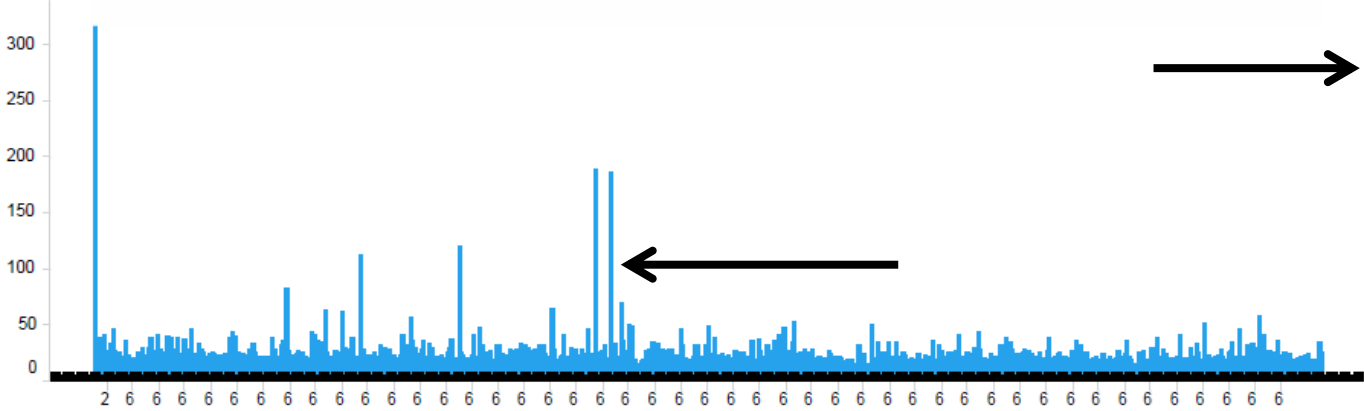
prestadorNombre	Ca
FARMCITY S.A.	
FARMACIA PHARMAKION S.C.S	
SALUD Y BELLEZA S.A	
SELMA JORGE OMAR	
FARMACIA SAN PEDRO S.C.S.	
SUPERFARMA S.A.	
WENT S.A.	
FARMACIA DANESA S.C.S.	
LAS MAGNOLIAS S.C.S.	
FARMACIA COSTANERA S.C.S.	
TEKIEL ISIDORO	
SANTA FE 2000 S.C.S.	
FARMACIA S.A.	
FARMACEUTICA CENTAURO S.A.	
FARMACIA NUEVA SAN AGUSTIN S.R.L.	
FARMACIA VARELA S.A.	
ANTIG. FARM. ROSSI DE CASTELAR S.C.S.	
AVENIDA CERVIÑO 4716 S.R.L.	
FARMACIA SOCIAL RAMOS MEJIA S.C.S	
ABC S.A	
FARMACIA CENTRAL MORON S.R.L.	
DANERI S.A	
RIVADAVIA 14166 S.C.S	
CENTRAL OESTE PARATI SCS	38
CENTENARIO 300 SCS	38
PARADIÑEIRO VICTOR HUGO	
NESTORE HONS.S.C.S.	

ENTENDER



- Explorar los Datos Disponibles.

Transacciones x Socio



Casos x Filial de Servicio

filialServicio	Casos
60	
0	
35	
14	
25	
33	
37	
22	
23	
24	
26	
3	
36	
31	
18	
11	
13	
32	
19	
2	
30	
12	
42	
15	
21	439
6	393
17	
5	

Casos per cronico

cronico	Casos
0	
1	
Grand to...	

Casos per pmí

pmí	Casos
0	
1	
Grand to...	

Casos per prescriptorTipo

prescriptorTipo	Casos
MW	
MN	
ML	
MJ	
MY	
MH	
MZ	
MG	140
MP	130

ENTENDER



- Definir un plan de acción para el Procesamiento de Datos.

Se procesaran los datos teniendo en cuenta la construcción de métricas de tipo RFM. Se calcularan los desvíos en los Consumos, contrastando al consumo individual contra la media. Se separaran los tipos de consumo.

- Determinar Métricas de Evaluación.

Se utilizaran como métricas de Evaluación, el LIFT (Elevación en el monto promedio por encima de la media) y la estabilidad de los segmentos en el tiempo.

- Definir Criterios de Éxito.

Austeramente, se espera un LIFT que supere 3. Cualquier resultado igual o mayor sera considerado exitoso.

- Determinar el Volumen de Datos Necesario para Modelar.

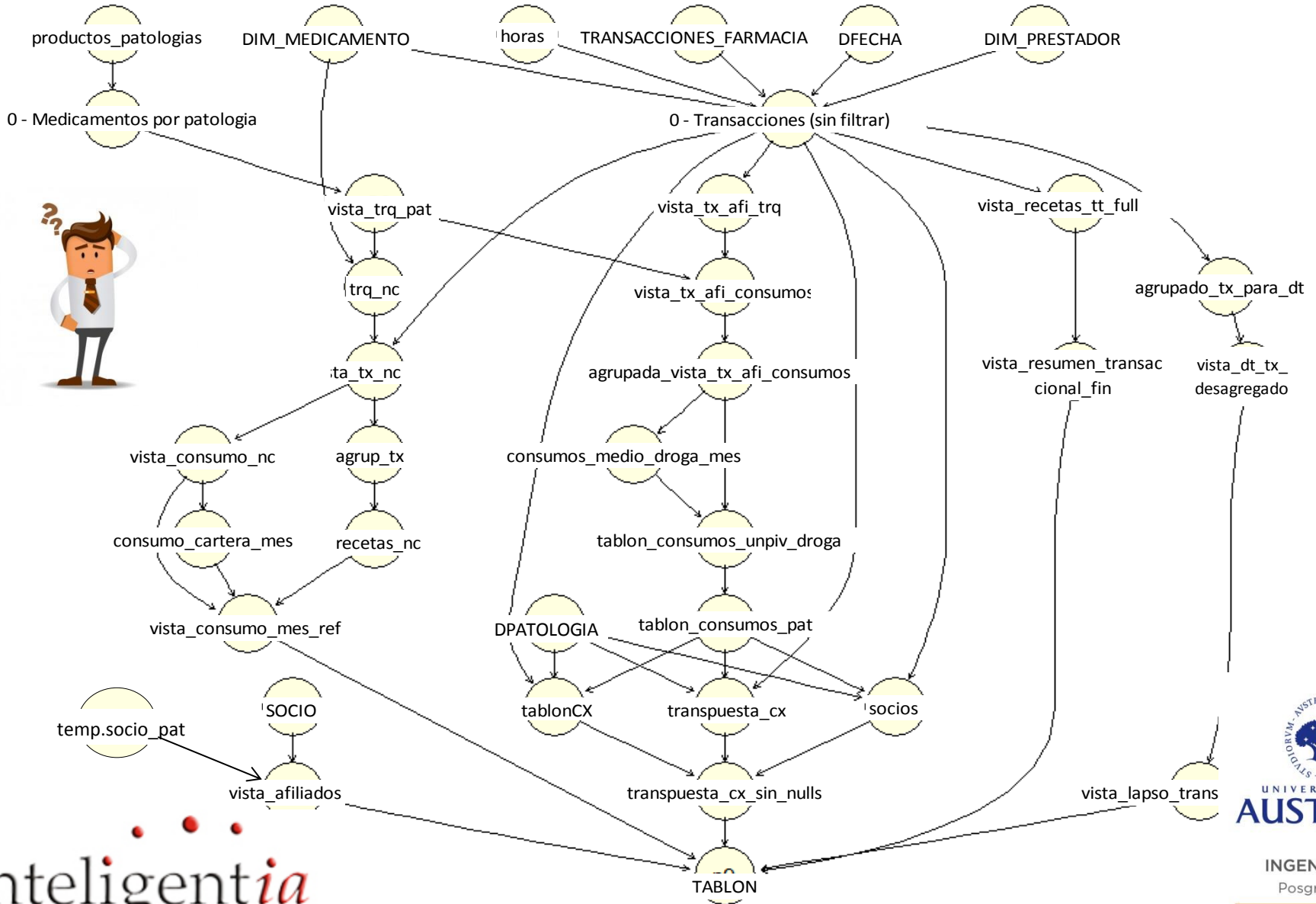
Se tomaran todos los consumos mensuales no limitando el volumen de datos a ser utilizado.

- Determinar la Necesidad y/o La Pertinencia del Muestreo.

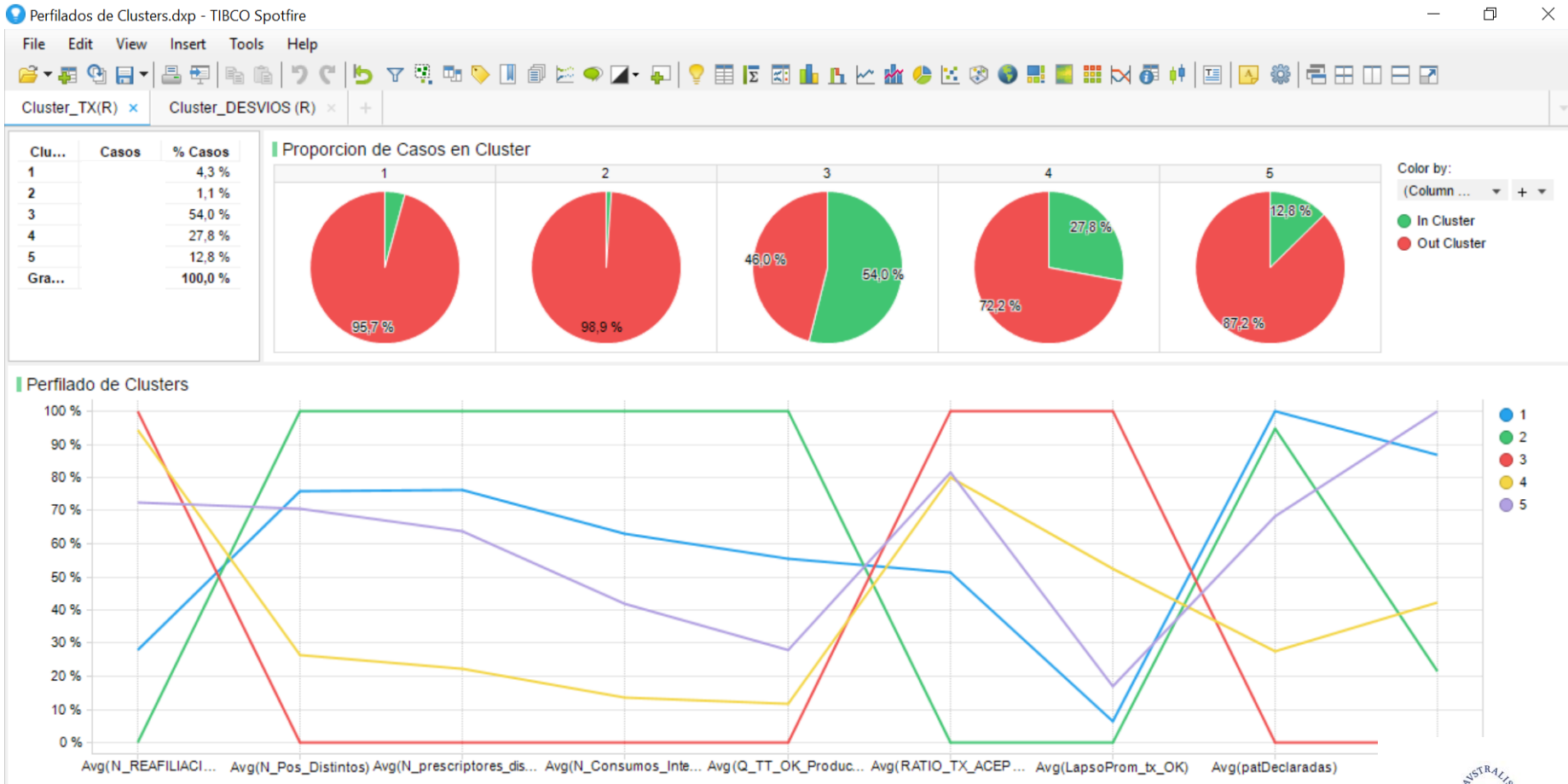
Al buscarse anomalías, no se recomienda realizar muestreos, ya que si el mismo no es bueno podrían perderse características interesantes a ser detectadas.

MODELAR

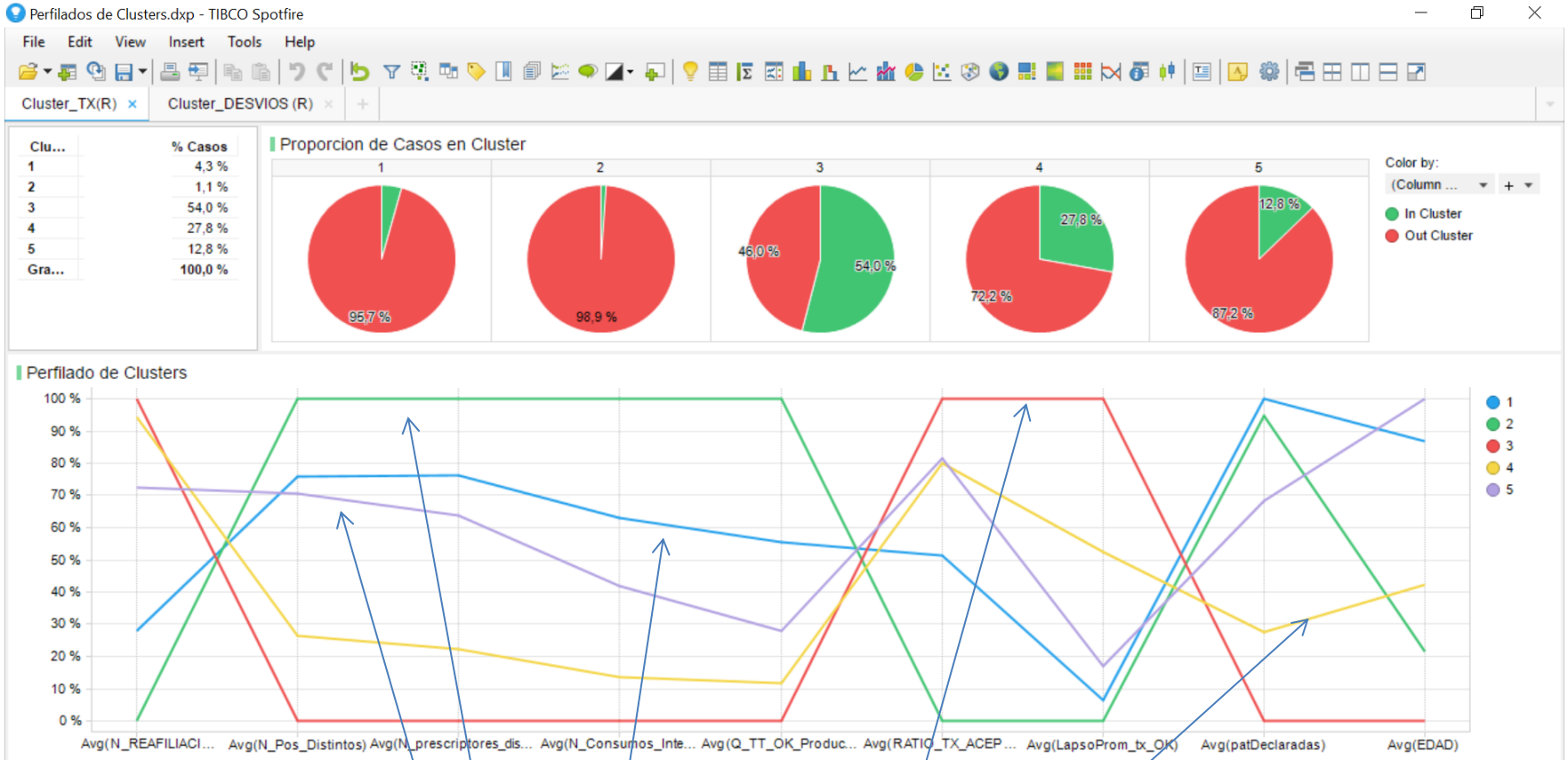
- Realizar las Integraciones, Procesos y Transformaciones sobre los Datos.



Modelo 1: Comportamental



Modelo 1: Comportamental



- Cluster 1: Professional
- Cluster 2: Bad Boys
- Cluster 3: Good Boys
- Cluster 4: Cronics Boys
- Cluster 5: Problematics

MODELAR

- Ajustar los Modelos utilizando el/los Algoritmos seleccionados.

Se presume que los desvíos en los consumos de un afiliado, impactaran generalmente sobre todos sus consumos.

- Variables de Desvíos Correlacionadas.
 - Un Modelo de Clusters no funciona correctamente cuando las variables de entrada están correlacionadas.
 - Se aplicara un modelo de Componentes Principales previo a la corrida del Clustering.
 - No se eliminaran componentes principales de manera tal de no perder información.
 - Luego, los Clusters se interpretaran en función de las variables Originales.

Campo	Lo Peor	→			Lo Mejor
Importe_Total	5	4	1	2	3
Desvio_Cr_Cr	1	5	4	3	2
Desvio_Cr_NoCr	4	2	5	1	3
Desvio_NC	5	4	1	3	2
Q_TT_OK_PRD_MES	5	4	1	2	3

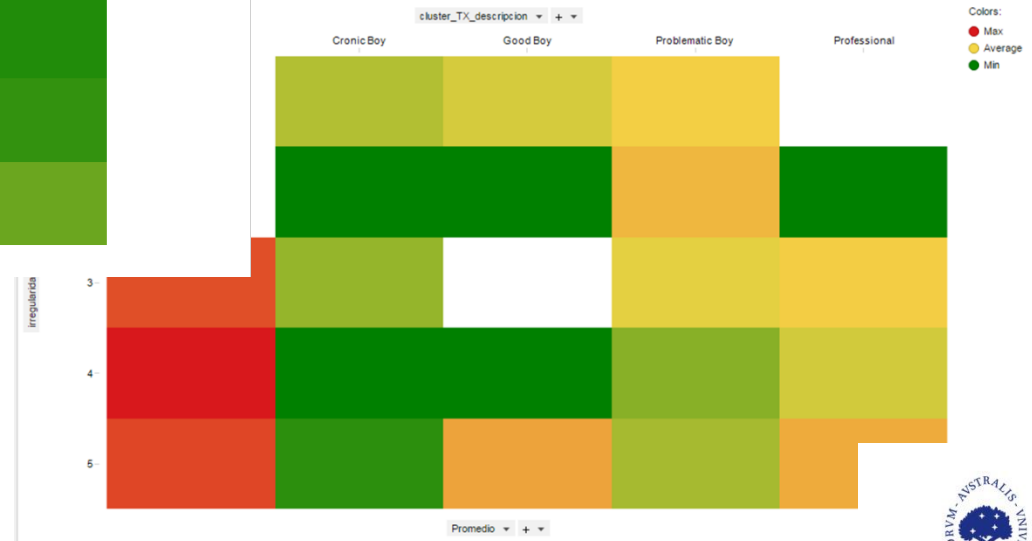
ORDEN INTERPRETADO Cluster Desvios	5	4	1	2	3
IRREGULARIDAD_DESVIOS	5	4	3	2	1

MODELAR

- Evaluar el Modelo en cuanto a: Ajuste, Coherencia, Precisión, Exactitud, Poder de Discriminación, Estabilidad, Interoperabilidad.
- Evaluar contra Resultados Pasados (Que hubiese pasado si?)

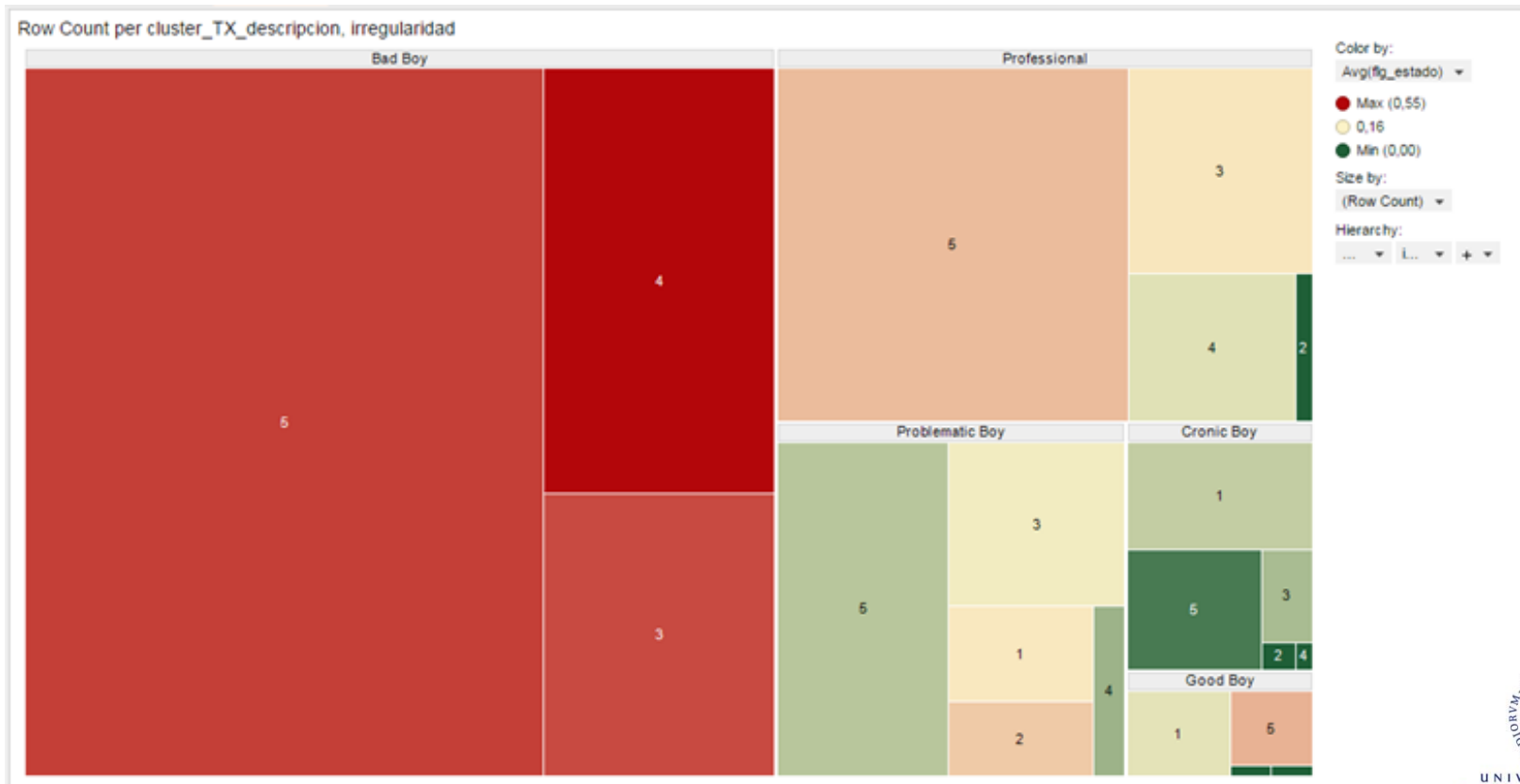


Muy Pocos Casos



MODELAR

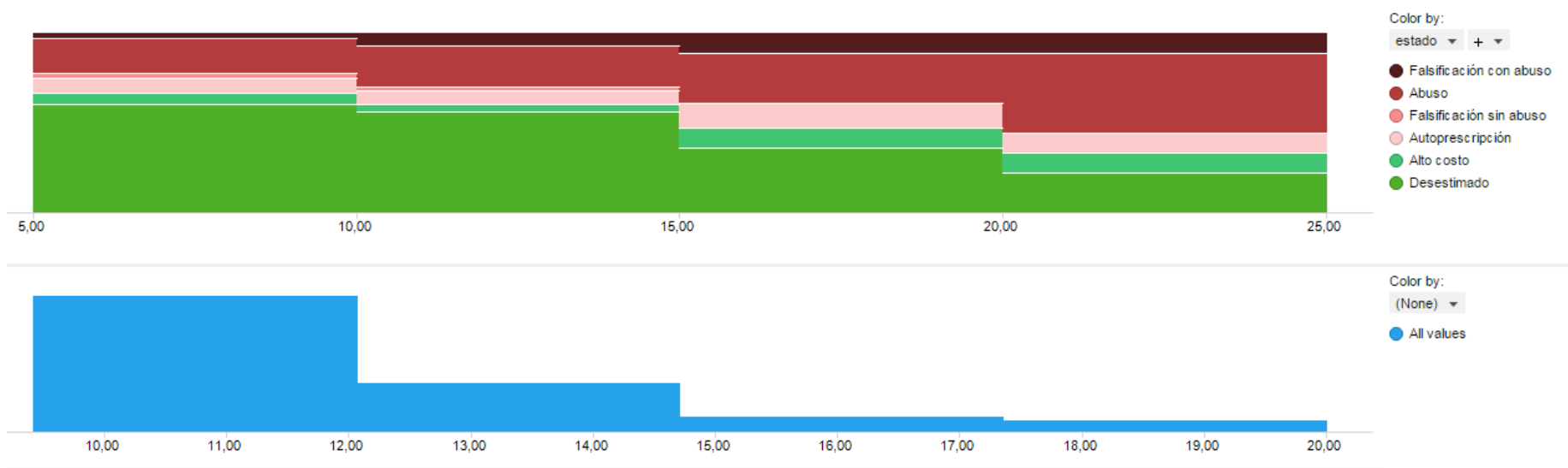
- Evaluar el Modelo en cuanto a: Ajuste, Coherencia, Precisión, Exactitud, Poder de Discriminación, Estabilidad, Interoperabilidad.
- Evaluar contra Resultados Pasados (Que hubiese pasado si?)



IMPLEMENTAR y SEGUIR



- El Modelo se pone en Producción.
 - Se realiza una Selección. (Inicialmente, distribuida en distintos grupos para realizar una evaluación POST del modelo).
 - Estos datos seleccionados son enviado a Auditoria desde donde vuelven Clasificados.
- Se realiza el seguimiento, cruzando la Devolución de Auditoria contra los resultados del Modelo.



El Grupo BAD BOYS – 5, contiene aproximadamente un 60% de Casos Fraudulentos

PREGUNTAS?

Muchas Gracias

Estanislao M. Irigoyen
INTELIGENTIA SRL

emirigoyen@inteligentia.com.ar

Metodologías de Trabajo Existentes:

- No son completas.
- Se olvidan del carácter multidisciplinario del problema.
 - **SEMMA:** Sample-Explore-Modify-Model-Assest.
 - Disocia al Negocio del análisis.
 - Se enfoca puramente en lo técnico.
 - Contiene errores conceptuales, los que pueden llevar a un incorrecto encare de la solución.
 - Puede llevar a NO ENCONTRAR resultados útiles.
- Disocian las etapas de trabajo, generando disociación y a veces «Presiones» sobre el equipo de trabajo.
 - Introducción de Errores en el trabajo debido a la necesidad de «MOSTRAR RESULTADOS» o «PRODUCIR ENTREGABLES».
- **CRISP-DM:**
 - Separa a la «PREPARACION DE DATOS» del «MODELADO».
 - La PREPARACION DE DATOS es la parte MAS IMPORTANTE del MODELADO.
- Ninguna de las anteriores se presenta como un Circulo de Mejora Continua ya que se olvidan de la SUPERVISION permanente de los resultados.