BIG DATA for BUSINESS

2.6 Casos de uso analíticos

Conecta Empleo

Contenido desarrollado por Synergic Partners



Índice del módulo

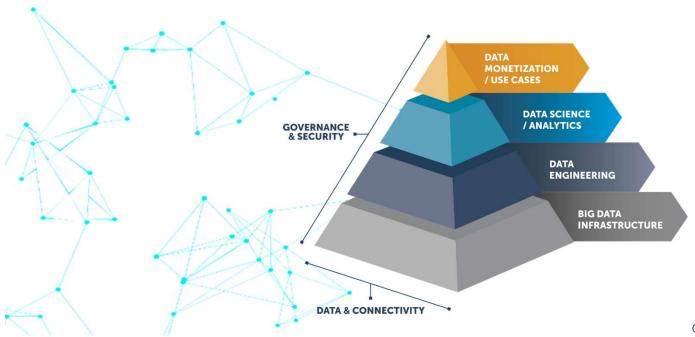
2.6 CASOS DE USO ANALÍTICOS

- Introducción
- Casos de Uso en una Telco
- Casos de Uso Reales
- Taller



Introducción

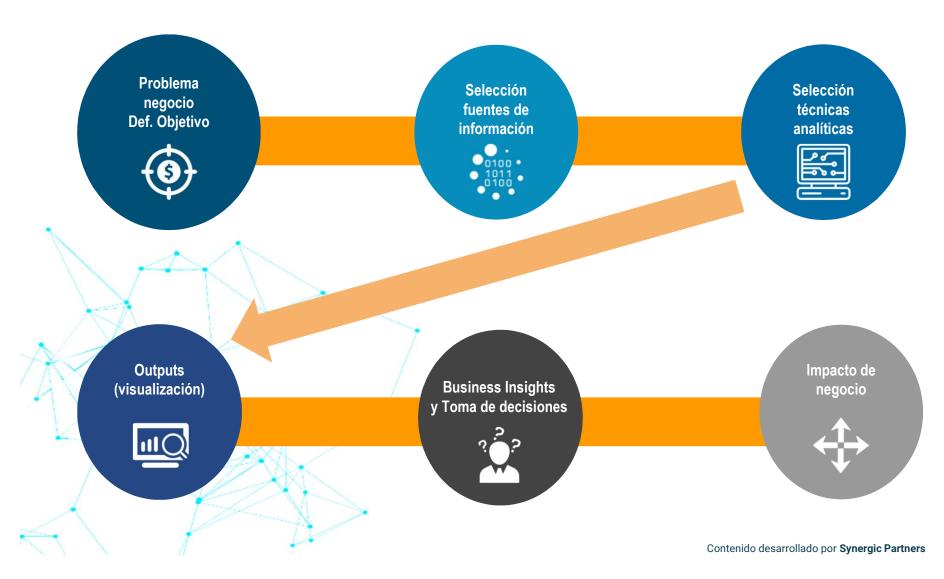
El fin último de la transformación de una compañía hacia una **organización orientada al dato** no es otro que satisfacer a los **principales retos y necesidades del negocio**. La forma más eficiente de aterrizar las tecnologías **Big Data** y la disciplina **Data Science** en una organización es a través del desarrollo de **casos de uso** que den respuesta a las preguntas más prioritarias del negocio, es decir, casos reales que dan solución a problemas reales de las compañías.



Contenido desarrollado por Synergic Partners

Introducción

Fases de un caso de uso



Glosario de Términos

DPI: Deep Packet Inspection. Datos del uso de internet de un cliente agrupados por paquetes (visitas web, social apps...)

CDR: Call Detail Record. Conjunto de datos producidos por una llamada de teléfono o cualquier otra transacción entre dispositivos de telecomunicación.

Billing: Datos de facturación

Revenue: Datos sobre la gestión de ingresos

Customer Insights: Objetivo de conocer al cliente internamente, sus gustos, preferencias, etc.

ARPU: Average Revenue per User. Ingresos medios por usuario.

KPI: Key Performance Indicators. Métricas para cuantificar los resultados de una determinada acción, es decir, permite medir el éxito de una acción.

CRM: Customer Relationship Management. Software para gestionar la relación con el cliente.



Casos de Uso en una Telco



Imaginemos una gran compañía internacional de telecomunicaciones, que abarque el sector de la telefonía, la televisión y la innovación. A continuación analizaremos las fuentes, la metodología y los posibles casos de uso.

Casos de Uso en una Telco

Fuentes de datos disponibles en una Telco



Datos de red

- CDR
- Datos de red
- Datos llamadas
- DPI



Datos clientes

- Billing / Revenue
- Información socioeconómica del usuario
- Uso de los productos
- Tipos de tarifa
- Tipo de dispositivo



Datos comportamiento clientes

- Localización
- Uso de las aplicaciones
- Audiencia canales TV
- Datos navegación Web
- Datos consumidos por las apps: YouTube, FB, Whatsapp, etc.



Fuentes externas

- Open Data: INE, catastro, ayuntamientos, etc...
- Redes sociales, LinkedIn, Twitter, etc.
- Datos compañías externas
- Estudios & Reports:
 - o Gartner, Forrester, Gfk, TNS, etc.



Casos de Uso en una Telco

Posibles casos de uso de la telco:



Casos de Uso en una Telco

Recomendador de paquetes de TV:

Herramienta que proporcione una recomendación de productos comerciales de video para cada cliente.



Fuentes de datos	KPIs	Analítica

Casos de Uso en una Telco

Recomendador de paquetes de TV:

Herramienta que proporcione una **recomendación de productos comerciales de video para cada cliente.** Se realiza en base a su histórico de compra, al uso de los productos y a los contenidos de TV preferidos. Como ejemplo del uso de este tipo de herramienta tenemos Netflix.

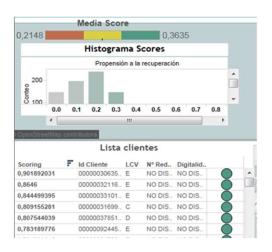


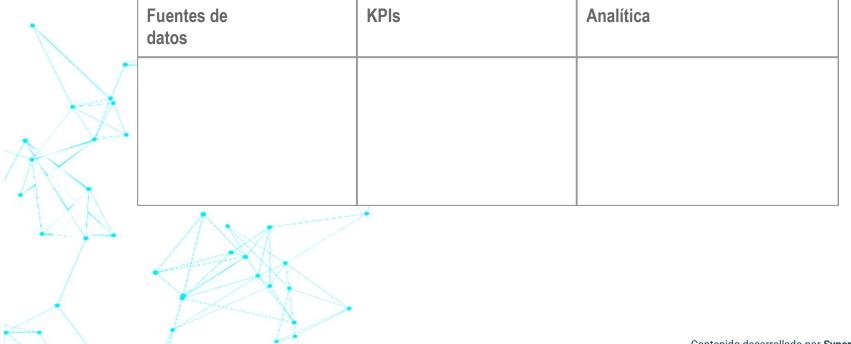


Casos de Uso en una Telco

Reactivación de clientes inactivos:

Conseguir recuperar a aquellos clientes considerados inactivos, es decir sin productos contratados o con baja actividad con la entidad.

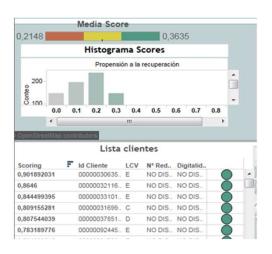




Casos de Uso en una Telco

Reactivación de clientes inactivos:

Conseguir recuperar a aquellos clientes considerados inactivos, es decir sin productos contratados o con baja actividad con la entidad. Para ello se enriquecerá la información disponible de los clientes utilizando para ello redes sociales y otras fuentes de datos externas junto con los datos bancarios que se tengan de dichos cliente.



1	Fuentes de datos	KPIs	Analítica
	Datos de clientes Fuentes externas	Patrones de rutas detectadas, número de clientes para cada patrón, Incremento del número de variables correlativas, tasa incremento margen medio por cliente	Preprocesamiento de la información (Natural Language processing, named entity recognition), Clasificación y Clustering

Casos de Uso en una Telco

Detección de Fraude / Smart Digits:

Smart Digits es un servicio de datos de localización de clientes que se ofrece a las entidades bancarias para que, previo consentimiento expreso, permita validar las extracciones de dinero en los cajeros automáticos.

Las entidades bancarias, por su parte, pueden ofrecer un servicio de seguridad a sus clientes para permitir la aprobación de las extracciones de dinero en los cajeros siempre y cuando el teléfono móvil del cliente se encuentre cercano al cajero. Este servicio se ofrece al cliente final para reducir el fraude ante el robo de tarjetas de crédito.



N	Fuentes de datos	KPIs	Analítica

Casos de Uso en una Telco

Detección de Fraude / Smart Digits:

Smart Digits es un servicio de datos de localización de clientes que se ofrece a las entidades bancarias para que, previo consentimiento expreso, permita validar las extracciones de dinero en los cajeros automáticos.

Las entidades bancarias, por su parte, pueden ofrecer un servicio de seguridad a sus clientes para permitir la aprobación de las extracciones de dinero en los cajeros siempre y cuando el teléfono móvil del cliente se encuentre cercano al cajero. Este servicio se ofrece al cliente final para reducir el fraude ante el robo de tarjetas de crédito.



Fuentes de datos	KPIs	Analítica
Fuentes externas Localización Datos de red	Evitar el fraude. Detectar robos. Dar seguridad al cliente.	Transformación de coordenadas, cálculos de distancias. Clasificación (nivel del riesgo de fraude)





Caso 1: Detección de Anomalías en Descuentos/Promociones



El **objetivo** de este caso de uso es la **detección de anomalías en descuentos y promociones** aplicadas a los clientes en campañas de marketing o por el área de Atención al Cliente.



Este caso de uso es solicitado por el departamento de **Business & Marketing Intelligence**. Tras detectar el área de Billing de la compañía una pérdida de facturación significativa a la hora de aplicar estos descuentos.



La detección de anomalías en descuentos y promociones genera una serie de pérdidas mensuales hasta ahora irrecuperables. Gracias a la detección de las mismas, se producirán ahorros directos para el área de Billing y mejorará la estructura de los descuentos futuros que se apliquen desde el área de Marketing y Atención al cliente.



- → **Valor empresa:** Ahorros derivados de aplicar los descuentos en el tiempo y forma preestablecido, evitando anomalías en su ejecución.
- → Valor Cliente: El cliente evitará cargos asociados a la mala ejecución de los descuentos concedidos





USE CASE:
DETECCIÓN ANOMALÍAS EN DESCUENTOS

Caso 1: Detección de Anomalías en Descuentos/Promociones

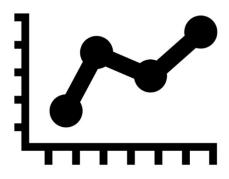
Descripción

Proyecto analítico de detección de anomalías en descuentos y promociones aplicadas a los clientes de una Telco.

Analizando los datos en los sistemas de información prestando especial atención a las estructuras de los descuentos, con el **objetivo de reducir posibles pérdidas causadas por errores en la asignación de descuentos** a determinados clientes o su continuidad en otros.

Por ejemplo, un descuento de 12 meses aplicado durante un tiempo superior o un descuento propio de la campaña de Navidad aplicado fuera de la misma.

Las anomalías detectadas podrían indicar posibles problemas en Billing o combinaciones de descuentos inapropiadas, originadas por campañas de marketing o por atención al cliente.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Valor de Negocio



Optimización de servicios

- Reducción número de llamadas Call Center por reclamación en descuentos mal aplicados.
- Reducción de costes de gestión de reclamaciones por parte del área de atención al cliente
- Adelantarse a las peticiones de los clientes, logrando aumentar el compromiso de los mismos.



Campañas de descuento

- Prevención de fallos en la asignación de los descuentos en las futuras campañas
- Resolución de anomalías identificadas con menor tiempo de respuesta



Revenue

- Incrementar el revenue assurance mediante la identificación de aquellos descuentos que se han aplicado de forma incorrecta.
- Incremento de la facturación derivados del aumento del número de anomalías detectadas que provocan problemas de billing.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Fuentes de datos

¿Qué fuentes de datos utilizarían para la realización del caso de uso?



Caso 1: Detección de Anomalías en Descuentos/Promociones

Fuentes de datos

ANÁLISIS



FUENTES INTERNAS

- CRM: tipo de contrato, beneficios asociados al cliente, producto. familia del producto, edad clientes, canal.
- Datos de facturación Account ID, Billable (Y/N), Billing Cycle, SIM State, Service type (datos, voz, SMS), Date/Time Received, Total Usage (raw), Total Usage (rounded), Bytes up/down, APN, Operator Network, Session Duration, Session Close Cause.
- **Datos red:** username, dirección IP, preferencia de aplicaciones.

器

FUENTES EXTERNAS

 Datos campañas descuentos: datos de campañas de descuentos y promociones realizados con anterioridad para detectar las anomalías más comunes a la hora de ejecutar los descuentos. (tiempos de descuento, no aplicación de descuento ofrecido...)



Caso 1: Detección de Anomalías en Descuentos/Promociones

Fuentes de datos

¿Otras fuentes?

En futuros proyectos los datos pueden ser alimentados por más fuentes.

Un ejemplo sería utilizar las redes sociales como fuentes externas:

- Twitter: sentiment analysis en base a los tweets de los clientes que han escrito un tweet en el que se menciona la campaña de descuentos realizada.
- Facebook: Información demográfica de los clientes que pueda extraerse de esta red social (edad, sexo, lugar de residencia). Sentiment analysis en base a los comentarios de los clientes sobre la campaña de descuento.

En este caso de uso no se usaron ya que la legislación del país en el que se encontraba la Telco prohíbe extraer los datos personales de los clientes.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Tratamiento de datos

El primer paso realizado con el conjunto de datos fue un **análisis exploratorio** con el objetivo de conocer el tipo de variable, número de nulos y valores distintos.

El output del modelo debía escribirse sobre una tabla de resultados con un esquema ya definido (fct_dad_ctr_score), de modo que hubo que realizar un **pre-procesamiento manual** de los datos de entrada (fct_dad_ctr_benefits). Este consistió en:

 Cambiar los nombres de las variables ya que sus nombres en la tabla de salida no coincidían con los nombres en la tabla de entrada

fct_dad_ctr_benefits	fct_dad_ctr_score
benefit_action_code	ben_action_code
benefit_value_type_desc	ben_value_type_desc
benefit_promotion_type_desc	ben_promo_type_desc
benefit_sl_channel_sub_group	ben_sl_chl_sub_gp
sales_channel_sub_group	sl_chl_sub_gp
product_family_desc_prev	prd_fam_prev
product_family_desc_currrent	prd_fam_curr
framework_contract_type	fw_ctr_type



Caso 1: Detección de Anomalías en Descuentos/Promociones

Tratamiento de datos

 Se crearon variables dummies para las variables categóricas, es decir, se creó una nueva variable con cada valor categóricos de la misma. Por ejemplo podemos ver cómo se hizo para los días de la semana, de este modo si antes un registro tenía una variable ordert_dt_weekday y este completo con monday, pasa a tener 7 variables donde la correspondiente a monday estará completada con un 1 y el resto con valores 0.

	fct_dad_ctr_benefits		fct_dad_ctr_score
	Variables	Categorical Values	Variables
	order_dt_weekday	Friday	order_dt_weekday_rs_friday
		Monday	order_dt_weekday_rs_monday
		Saturday	order_dt_weekday_rs_saturday
		Tuesday	order_dt_weekday_rs_tuesday
4		Sunday	order_dt_weekday_rs_sunday
		Thursday order_dt_weekday_rs_thu	order_dt_weekday_rs_thursday
		Wednesday	order_dt_weekday_rs_wednesday
/	11/	i e e e e e e e e e e e e e e e e e e e	

Caso 1: Detección de Anomalías en Descuentos/Promociones

Tratamiento de datos

• En algunos casos, algunos de los valores de las variables categóricas de la tabla de entrada no aparecían como valores residuales en la tabla de salida. Este conjunto de valores se agruparon en una variable denominada others.

	fct_dad_ctr_benefits		fct_dad_ctr_score
-	Variables	Categorical Values	Variables
	benefit_sl_channel_sub_group	Flächenmarkt Other, Other, Business Distribution	ben_sl_chl_sub_gp_rs_other
	product_family_desc_prev	Blau Postpaid, o2 Genion, o2 Genion L, o2 Genion L Aktion, o2 Genion M, o2 Genion Profi,o2 Genion XL, o2 Mobile Flat, o2 New Genion, o2 New Genion Profi,o2 nxt, o2 on, o2 on Upgrade,o2 Select, o2 Select Profi, o2 Starter	prd_fam_prev_rs_other
	product_family_desc_current	Blau Postpaid, o2 Genion, o2 Genion L, o2 Genion M, o2 New Genion, o2 nxt, o2 on","o2 on Upgrade	prd_fam_curr_rs_other



Caso 1: Detección de Anomalías en Descuentos/Promociones

Análitica

¿Cómo lo resolvemos?

Se estudiaron diferentes enfoques analíticos a la hora de abordar este problema. Estos pueden agruparse en 6 categorías:

- 1. **Estadístico.** Se basa en la construcción de un modelo de distribución de probabilidad con los datos y se hace un ranking con las muestras según se aproximen a la distribución del modelo.
- 2. Basado en proximidad. La idea es que un elemento que se encuentra alejado del resto, es una anomalía. Es computacionalmente costoso y no es práctico para grande volúmenes de datos.
- 3. Basado de densidad. Considera anomalías aquellos objetos que se encuentran en la zona de baja densidad.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Análitica

¿Cómo lo resolvemos? (cont.)

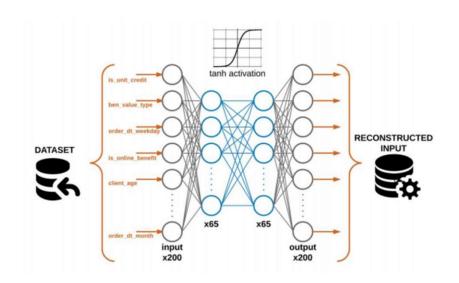
- **4. Basado en clustering**. Después de aplicar un modelo de clustering, se marcan todos los clusters que son más pequeños que un mínimo o que están muy alejados respecto al resto, como anomalías.
- **5. Deep learning**. Basado en un autoencoder, esta aproximación construye un modelo que comprime la colección de datos a una mínima dimensión e intenta reproducir los datos de entrada. Las anomalías serán datos pobremente replicados.
- **6. Teoría de la información**. Se basa en la medida de la cantidad de información que un punto proporciona, suponiendo que los datos anómalos proporcionarán la mayoría de la información.

A continuación veremos algunos de estos ejemplos.

Caso 1: Detección de Anomalías en Descuentos/Promociones

Análitica

Autoencoder



Analítica	Algoritmo	Funcionamiento
Redes Neuronales	Autoencoder.	Se introducen los datos de facturación y descuentos de clientes. El autoencoder aprende los patrones de estos datos de entrada, por ejemplo, que un tipo de descuento siempre tiene una duración de ciertos meses. Aquellos descuentos que no cumplan esta duración supondrán un patrón extraño, y no podrá reconstruirlo a la salida. Los casos cuya reconstrucción no sea correcta serán estudiados por expertos del departamento, ya que serán potenciales anomalías.

Caso 1: Detección de Anomalías en Descuentos/Promociones

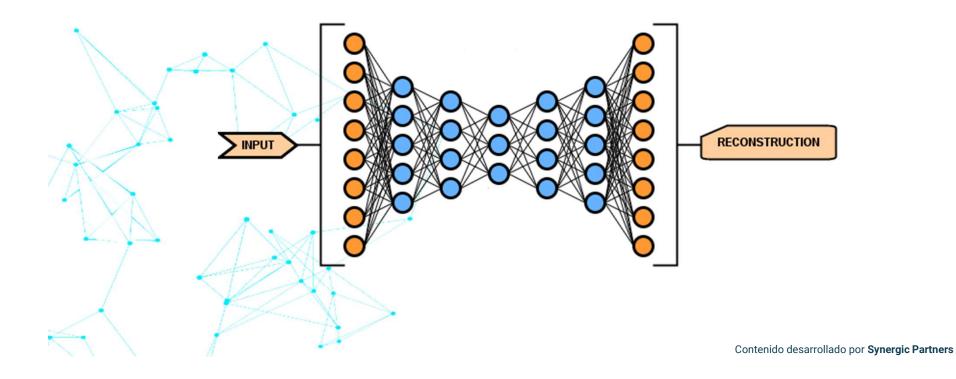
Análitica K-means 0.8 0.7 0.6 0.5 0.4 0.3 0.2 **Analítica** Algoritmo **Funcionamiento** Se agrupan los datos que comparten los mismos patrones de Clustering K-Means descuentos. Aquellos que no pertenezcan a ningún cluster, outliers, serán posibles anomalías.

Caso 1: Detección de Anomalías en Descuentos/Promociones

Análitica

Modelo Final

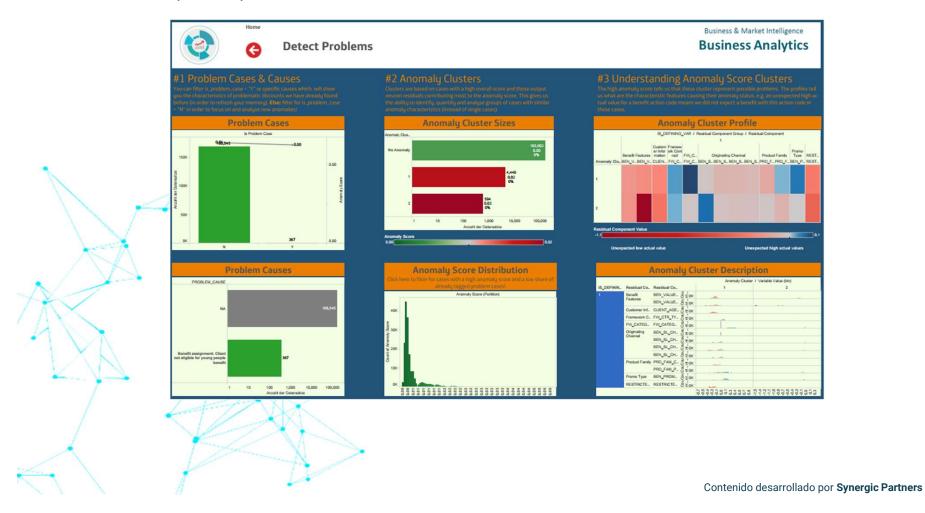
Después de estudiar las diferentes aproximaciones, se optó por un Autoencoder para resolver el problema, por requerimientos del cliente.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Output

Mediante las técnicas analíticas avanzadas, se genera un dashboard en el que se identifican las anomalías para su posterior resolución.



Caso 1: Detección de Anomalías en Descuentos/Promociones

Resumen del proceso

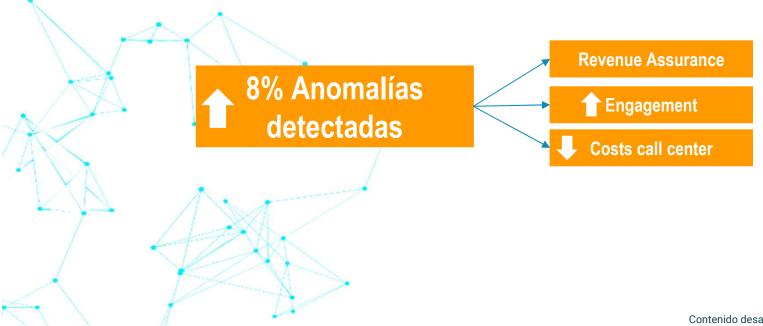


Caso 1: Detección de Anomalías en Descuentos/Promociones

Resultados

Como resultado este caso de uso supuso la detección del 8% de las anomalías.

El conjunto de anomalías se agrupó mediante clustering donde podemos destacar **dos** de ellos: uno de ellos proviene de errores en el sistema CRM formado por descuentos que fueron transferidos de una tarifa previa y reactivados tras ocurrir un cambio en la cuenta. Por otro lado el otro grupo estaba formado por clientes que amenazaban con rescindir su contrato por lo que se les ofrecían altos descuentos. Este segundo caso no es considerado anomalía.





Caso 2: Detección de puntos de venta de valor



El **objetivo** de este caso de uso es la **identificación de puntos de venta de valor tanto actuales como futuros:** zona geográfica, puntos de ventas y sus características.



Este caso de uso es solicitado por el departamento de **Business Insights** al detectar la posibilidad de realizar **campañas de marketing optimizadas** aumentar las ventas totales de sus puntos de venta mejorando los mismo en base a las tiendas con mayor índices de ventas totales y los potenciales puntos de venta de valor **aumentando así los ingresos de los puntos de venta**.



La detección de puntos de venta de valor ayuda a la compañía, por un lado, a identificar cuáles son las características que llevan a un punto de venta a obtener un elevado número de ventas de sus productos y, por otra parte, a detectar los puntos de venta con mayor potencial para así reducir el coste de las campañas comerciales de captación de nuevos clientes.



- → Valor empresa: Detección potenciales puntos de venta de valor. Optimización puntos de ventas actuales con posibilidad de mejora. Optimización campañas comerciales.
- → Valor Cliente: Implementar mejoras que lleven a incrementar el número de ventas.





USE CASE: <u>DETECCIÓN PUNTOS VEN</u>TA DE VALOR

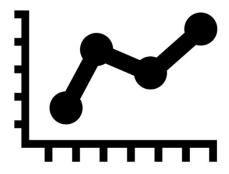
Caso 2: Detección de puntos de venta de valor

Descripción

Este caso de uso tiene dos objetivos: por un lado la detección de puntos de venta de valor y la localización de establecimientos "no clientes" con gran potencial de venta. Por otro lado segmentar los puntos de venta según variables externas.

Mediante el estudio de la zona geográfica seleccionada por el cliente, se realiza un estudio analítico de cuáles son los puntos de venta óptimos para la venta de los productos seleccionados y qué posibilidades de compras potenciales se detectan en los clientes actuales.

Una vez realizado el caso de uso se puede determinar que clientes son más rentables para atacar en las campañas comerciales del cliente.





Caso 2: Detección de puntos de venta de valor

Valor de Negocio

Clientes actuales:



Puntos de venta con mayor potencial de los clientes actuales de la compañía

- → Campañas comerciales con un mayor nivel de eficiencia.
- → Potenciación venta de los productos del cliente.
- → Información privilegiada para su óptima localización establecimientos propios

Clientes futuros:



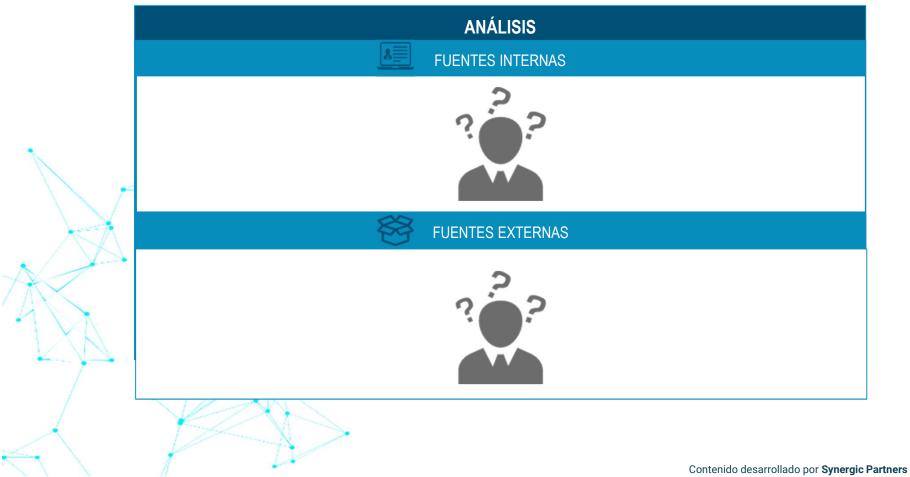
Detección clientes potenciales en base a las potenciales ventas del establecimiento

- → Captación de nuevos clientes, ya que, se podrá demostrar nivel de ventas de los productos en clientes actuales en establecimientos que poseen las mismas características.
- Área comercial: Mejor gestión de los esfuerzos para la captación de nuevos clientes de valor.

Caso 2: Detección de puntos de venta de valor

Fuentes de datos

¿Qué fuentes de datos utilizarían para la realización del caso de uso?



Caso 2: Detección de puntos de venta de valor

Fuentes de datos

ANÁLISIS



FUENTES INTERNAS

- **Número de SKUs:** (Número de referencia) es un identificador usado en el comercio con el objeto de permitir el seguimiento continuo de los productos y servicios ofrecidos a los clientes. Cada SKU se asocia con un objeto, producto, marca, servicio, cargos, etc.
- Informe Nielsen: este informe recoge las ventas totales en algunas tiendas de la marca
- **Tipo de establecimiento:** esta fuente de datos es necesaria para clasificar los diferentes establecimientos y conocer las posibles ventas potenciales de los mismos
- Informe ventas internas

容

FUENTES EXTERNAS

- Smart Steps: herramienta estadística que analiza movimientos de grupos de personas.
- Open street maps: puntos de interés; punto de ubicación específica que alguien puede encontrar útil o interesante, en este caso, se seleccionaron los puntos de interés identificados dentro de la zona geográfica seleccionada para el proyecto..
- Censo: esta fuente es de especial interés para determinar la población total, por género y por edad, censada en el área geográfica seleccionada para la ejecución del proyecto.
- Catastro: Censo estadístico por área geográfica que incluye número de viviendas totales, principales, secundarias y vacías.



Caso 2: Detección de puntos de venta de valor

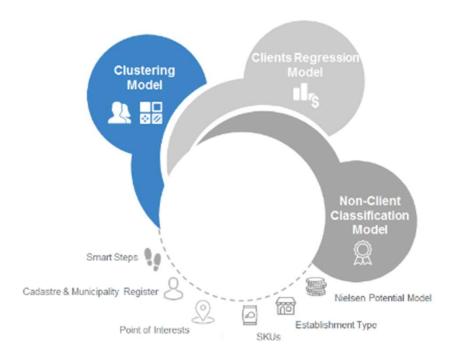
Analítica

¿Cómo lo resolvemos?

Se realizaron tres tipos de analíticas:

- Modelo predictivo de ventas totales de clientes
- Modelo de clasificación de no clientes en función de sus ventas
- Clustering para segmentar los puntos de venta clientes y no clientes en base a su entorno.

A continuación exploramos la metodología y las tres analíticas realizadas



Caso 2: Detección de puntos de venta de valor

Analítica

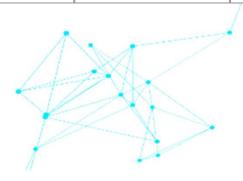
Modelo Predictivo

Se utilizaron variables internas (informe Nielsen, tipo de establecimiento, SKUs, ventas internas) y variables externas (Smart Steps, Puntos de interés, Catastro y censo).

REGRESSION



Analítica	Algoritmo	Funcionamiento		
Regresión	Random Forest	En esta analítica se busca predecir las ventas totales de un punto de venta. Para ello el algoritmo aprende de un valor objetivo de ventas totales (valor numérico) y devuelve la cantidad de ventas totales para ese punto.		



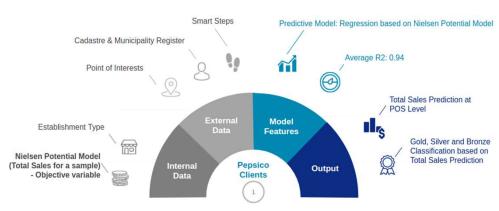
Caso 2: Detección de puntos de venta de valor

Analítica

Modelo de Clasificación

Se utilizaron variables internas (informe Nielsen, tipo de establecimiento) y variables externas (Smart Steps, Puntos de interés, Catastro y censo).

CLASSIFICATION



Analítica	Algoritmo	Funcionamiento
Clasificación	Random Forest	Esta analítica busca clasificar los tipos de venta binariamente, segúr sea un punto de alto valor de ventas (Oro/ Plata) o de bajo valor (Bronce). En este caso las variables son categóricas y el algoritmo devolverá si un punto de venta es Oro/Plata o Bronce.

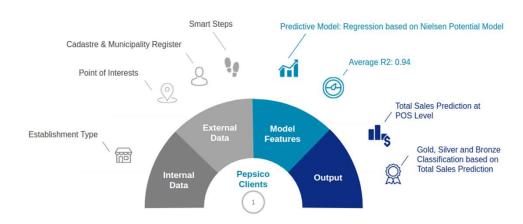
Caso 2: Detección de puntos de venta de valor

Analítica

Clustering

Se utilizaron variables internas (tipo de establecimiento) y variables externas (Smart Steps, Puntos de interés, Catastro y censo).

CLUSTERING



À	Analítica	Algoritmo	Funcionamiento
< X	Clustering	GMM	El objetivo de esta analítica es segmentar los puntos de venta en función de lo que sucede en su entorno. Por ejemplo, si es un lugar donde sólo compran residentes, o donde compran turistas, en qué momento del día se concentran el máximo de ventas para el establecimiento, etc.

Caso 2: Detección de puntos de venta de valor

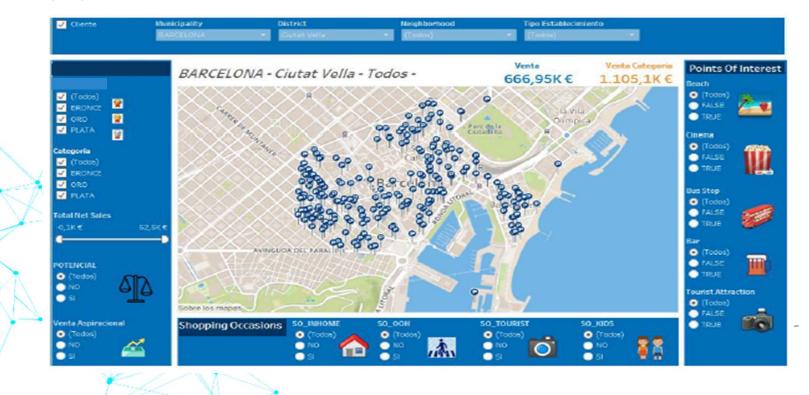
Resumen del proceso



Caso 2: Detección de puntos de venta de valor

Output

Gracias al diseño de un dashboard interactivo el cliente puede **identificar fácilmente los puntos de venta de alto valor**, tanto de clientes actuales como de posibles futuros clientes. El dashboard también aporta información relativa a **tipos de oportunidades de venta y puntos de interés identificados** en la zona geográfica estudiada.



Caso 2: Detección de puntos de venta de valor

Resultados

Identificando puntos de venta de alto valor pueden buscarse puntos de venta de la misma zona con igual características pero con menores ventas. Siguiendo como modelo el punto con mayores ventas de la zona puede usarse como referencia para aplicar sus mejoras en el resto de puntos y elevar las ventas.



Esto lleva a aumentar un 10% las ventas en puntos Silver&Bronze.





Caso 3: Activación de clientes dormidos



El **objetivo** de este caso de uso es la **recuperación de los clientes** del Banco que se encuentran en estado **"dormido"**, esto es, clientes que no presentan ninguna actividad en su cuenta en el último semestre.



Este caso de uso es solicitado por el departamento de **Business Intelligence** al detectar un **gran número de clientes inactivos en el CRM** de la compañía y con el fin de ejecutar una **acción comercial** en base a la **propensión de recuperación** de los clientes detectados como inactivos.



La existencia de una elevado número de clientes dormidos o inactivos supone un **riesgo** reseñable para el Banco, ya que, los **clientes inactivos tienen una gran tendencia a convertirse en clientes fugados.**



- → Valor empresa: Recuperación de clientes dormidos a clientes activos. valor recuperación ~ 90€.
- → Valor Cliente: El cliente evitará costes adicionales asociados al mantenimiento de una cuenta inactiva.





USE CASE: ACTIVACIÓN CLIENTE DORMIDOS