

La presente obra se encuentra protegida por el Derecho de Autor, fue elaborada por personal de IDATA S.A.S. quien ostenta la titularidad de los derechos patrimoniales de autor. Se prohíbe su reproducción total o parcial y se entrega a los estudiantes mediante una licencia Creative Commons:



## Línea de Formación: Científico de Datos

### CONTENIDO DEL CASO:

Incorporación de la analítica avanzada a la realidad del negocio

- Empresa: CottonPlus Inc.
- Industria: Retail de Moda

### CONTEXTO:

La compañía ha evolucionado, ahora han entendido el alto poder que tienen los datos y les ha permitido tomar decisiones estratégicas de negocio basados en el conocimiento de la propia operación y el conocimiento del comportamiento de sus clientes, sin embargo, se han dado cuenta que para poder hacer estrategias de impacto real deben dejar de mirar hacia el pasado y



empezar a mirar hacia el futuro. La estrategia se ha empezado a hacer preguntas como: ¿cuánto voy a vender?, ¿cuáles van a ser mis principales clientes?, ¿cuáles clientes voy a perder en los próximos meses?, entre otras preguntas. Ya intentaron acercarse al área de tecnología dado que son ellos quienes resguardan los datos en algunos servidores en nube, pero en el área consideran que no tienen perfiles para poder hacer predicciones, incluso, no entienden las preguntas que el negocio les está haciendo acerca de comportamientos futuros desde la dirección estratégica. Han empezado a preocuparse acerca de las nuevas capacidades que se requieren, no saben a quién contratar para poder entender los datos más desde la estadística y analizar posibles comportamientos futuros.

La gerencia ha participado en varios eventos de tecnología donde ha escuchado acerca de aplicaciones que utilizan inteligencia artificial, pero esto es algo que ve muy lejano a la realidad de la empresa, sin embargo, quiere encontrar qué tipo de soluciones de inteligencia artificial pudiese, en un futuro cercano, implementar en la compañía de tal forma que se pueda diferenciar de sus principales competidores.

## OBJETIVO:

Implementar una arquitectura analítica en nube que tenga los componentes necesarios para poder abordar iniciativas de análisis de datos a nivel estadístico, modelos que permitan hacer segmentación de clientes basado en comportamientos, predicción de eventos futuros y de interés para el negocio proyecciones financieras y proyecciones de ventas en general, e introducir algunas capas o componentes que le permitan realizar análisis de datos no estructurados, y soluciones basadas en inteligencia artificial.

Insumos: Bases de datos y componentes de análisis de datos en nube.

## NIVEL BÁSICO

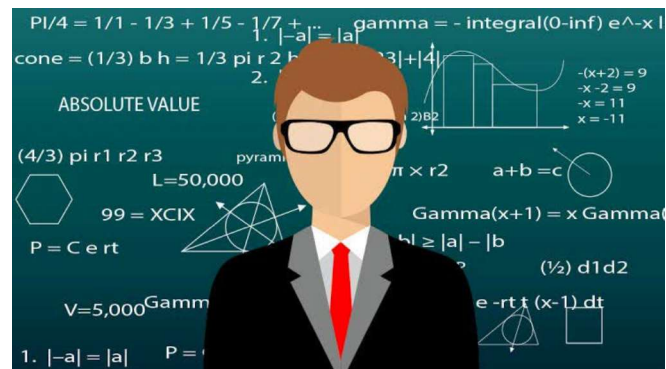
### Convirtiendo los datos en conocimiento del cliente

#### CAPÍTULO 1: Planear el uso de herramientas y datos estadísticos para explorar el comportamiento del cliente

**Insumos:** Tabla de datos de clientes

**Personajes:** Andrés Ortiz, director de tecnología David Roa (Gerente General Juliana Peláez, directora de mercadeo)

Ante el interés que ha tenido la Gerencia General de implementar algunas herramientas de analítica avanzada con enfoque predictivo, la dirección de TI en cabeza de Andrés Ortiz ha estado revisando qué tipo de software permite desarrollar ese tipo de iniciativas. Ha explorado ya software licenciado y software



gratuito, pero está un poco preocupado dado que entiende que cada herramienta se opera desde lenguajes de programación distintos, e incluso son estudiadas por profesionales egresados de diferentes carreras universitarias; no sabe cuál de ellas es mejor y mucho menos cuál de ellas es más adecuada para la realidad de la empresa. Es necesario realizar en este punto, un análisis acerca de los primeros pasos desde donde debería iniciar la empresa en cuanto a herramientas y capacidades. Lo que sí tienen claro, es que le gustaría probar herramientas sin tener que sacrificar costos importantes, incluso, ha escuchado recientemente acerca de algunas herramientas gratuitas que le puede permitir probar la tecnología para luego incorporarlas a las arquitecturas en nube. Algunos analistas que recientemente ha contratado la empresa, sugieren que se trabaje sobre R, y algunos otros sugieren que se trabaje sobre Python. Andrés quiere poner a prueba ambas herramientas buscando primero solucionar algunas deudas que tiene con la dirección estratégica, que son generar tableros de control basados en indicadores estadísticos.

### Objetivo:

- Instalar la herramienta Python en los equipos, y partiendo de la base de datos transaccional generar análisis descriptivo con análisis de percentiles, promedios, medidas de dispersión y análisis de probabilidades, entre otros indicadores que generen un conocimiento más profundo de los datos.

## CAPÍTULO 2. Análisis del comportamiento histórico en una base de datos con algunos problemas

**Insumos:** Tabla de datos de clientes

Algunas personas en la empresa que tienen gusto por los datos y el análisis estadístico, han iniciado con la exploración de los datos, el cálculo de indicadores, y cálculo de algunas probabilidades para generar información de valor para la estrategia. Sin embargo, rápidamente se han topado con que en la mayoría de los casos la información no coincide entre los diferentes analistas de datos. Se cree que



algunos de los errores están en las fuentes de información, sin embargo, también se han dado cuenta de que, en algunos casos, las fuentes de información son las mismas, pero los valores no responden a la realidad de la empresa. Han encontrado grandes falencias en la calidad de los datos.

#### **Objetivos:**

- Hacer un diagnóstico general de la calidad de los datos basado en datos duplicados, campos vacíos, y datos por fuera de parámetros normales del negocio.
- Establecer las diferencias básicas entre modelos probabilísticos y determinísticos

### **CAPÍTULO 3. Transformaciones y minería para mejorar el entendimiento de la base de datos**

**Insumos:** Tabla de datos de clientes

David Roa está altamente preocupado porque el esfuerzo que han venido haciendo capturando datos, durante tanto tiempo, están logrando pocas cosas, los errores en la calidad de datos son muy visibles, y le pide al equipo técnico encontrar algunas alternativas. Es por esto que, el equipo ha decidido hacer mejoramiento en los datos para alcanzar un mejor nivel de calidad. Así mismo, requieren aumentar el entendimiento que tiene la dirección acerca de las



bases de datos, para lo que han iniciado algunas transformaciones y minería de datos que les permita tener informes con mayor conocimiento del cliente; algunos de estos indicadores son lo que se ha llamado la metodología RFM.

#### **Objetivos:**

- Realizar una depuración de calidad sobre la base de datos y construir los indicadores RFM, entre algunos otros para mejorar el proceso de análisis de información.
- Procesamiento de matrices y vectores en ML
- Explorar modelos más utilizados en sector financiero como referencia a Retail
- Identificar el funcionamiento y el uso de los modelos de fuga de clientes (churn)

## CAPÍTULO 4. Modelando escenarios para segmentar clientes

Una vez fueron depurados los datos por algunos analistas de la compañía, David Roa ha empezado a obtener valor a partir de los datos pero ahora, ha encontrado un universo gigante de clientes donde todos se comportan de diferentes maneras y para poder diseñar estrategias efectivas cree que ahora es necesario segmentarlos. Las segmentaciones tradicionales que ha escuchado son basadas en análisis demográfico, es decir, en variables como la edad, el nivel socioeconómico, el género, entre otras variables; pero se han



dado cuenta que en la empresa hay personas de diferentes características demográficas que tienen comportamientos similares a nivel transaccional. Juliana Peláez desde el área de mercadeo, se pregunta si existen metodologías de segmentación basadas en comportamientos.

### Objetivos:

- Utilizar modelos de clasificación de datos para hallar diferentes segmentos de clientes y representarlos gráficamente.
- Explorar el uso de modelos más utilizados en la industria de seguros
- Analizar pasos requeridos para la industrialización de modelos