

Proyecto de Calidad y Preprocesamiento de Datos

Guillermo Gerardo Andrés Urbano
Luis Octavio Sánchez Hatadi
Edgar Callejas Hernández

I Escoger un área de negocio dentro del sector Salud, Finanzas, Educación, etc., ya sea global o en México

El negocio pertenece a sector de bienes y servicios enfocado a la educación. Se enfoca a la venta de productos escolares y de oficina desde pregrado hasta grado, además ofrece algunos básicos de consulta y pagos de servicios.

II. Problemática y entendimiento del negocio

1. Identificar una problemática que implique mala toma de decisiones derivada de datos de baja calidad en términos de las dimensiones de calidad vistas en clase.

Derivado a gran variedad de productos que se ofertan y que más de dos personas trabajan en el negocio, se ha vuelto complicado llevar un conteo preciso de los productos faltantes cada semana y se desconoce la utilidad.

Se cuenta con una base de datos inicial la cual se compone de una tabla llamada *inventario* y una llamada *ventas*. La primera es obtenida mediante un programa de punto de venta lo cual minimiza los errores; sin embargo, el llenado inicial es a mano y existen valores faltantes en los costos de los productos y errores ortográficos en la descripción del producto. En la tabla *ventas* solo se cuenta con la fecha de la venta, el ID del producto y la cantidad vendida.

El negocio implementó este almacenamiento de datos recientemente, por lo que solo se cuenta con las ventas de abril y mayo.

Sumado a esto, no se sabe con exactitud cuáles son los productos estrella ni los productos de temporada. Y se quiere saber si existen productos con una cierta regla de asociación. (Si se compra uno, entonces normalmente se compra el otro).

Además, no se sabe analíticamente cuáles son los productos que se pueden invertir en grandes cantidades al inicio de las fechas escolares.

Finalmente, se quiere realizar un proceso de optimización, es decir, se busca eliminar productos cuya utilidad no compense los costos de compra y almacenamiento, reducir la cantidad de productos en inventario e incrementar aquellos que se vendan más para evitar desabastos.

2. Identificar 10 preguntas de negocio que impliquen toma de decisiones a partir de un análisis de datos.
 - (a) ¿Cuál es el producto con mayor margen de ganancia?
 - (b) ¿Cuál es el top 5 de los más vendidos en un periodo de tiempo determinado?
 - (c) ¿Cuál es el top 5 de los menos vendidos en un periodo de tiempo determinado?
 - (d) ¿Cuál es el día de la semana con mayor venta?
 - (e) Se quiere proyectar las ventas para las siguientes semanas.
 - (f) Sé saber cuáles son los productos que se venden juntos.
 - (g) Saber si existe un incremento sostenido de las ventas por cada categoría.
 - (h) Dado los productos que más se venden, saber cuál es la marca que menos se vende.
 - (i) Se desea saber la preferencia del color de los productos.
 - (j) Saber cuál es la categoría con más ventas.

III Elaborar propuesta de solución

Se planea realizar una limpieza de los datos eliminando los errores y llenando los datos de acuerdo a las reglas del negocio. Igualmente, se hará un perfilado de datos para conocer en términos generales la proporción de datos faltantes, la consistencia y tener una idea de la cantidad de información que se puede extraer.

Posteriormente, se completará la tabla **ventas** con ayuda de la tabla **inventario** para poder responder a las preguntas de negocio. Sumado a esto, se aplicarán algunas técnicas estadísticas y de análisis de datos para realizar las proyecciones y estimaciones.

Al final se entregará un reporte sobre el análisis de datos, así como de las soluciones a las preguntas del negocio de una manera clara y concisa que le permita a los dueños del negocio tomar decisiones.