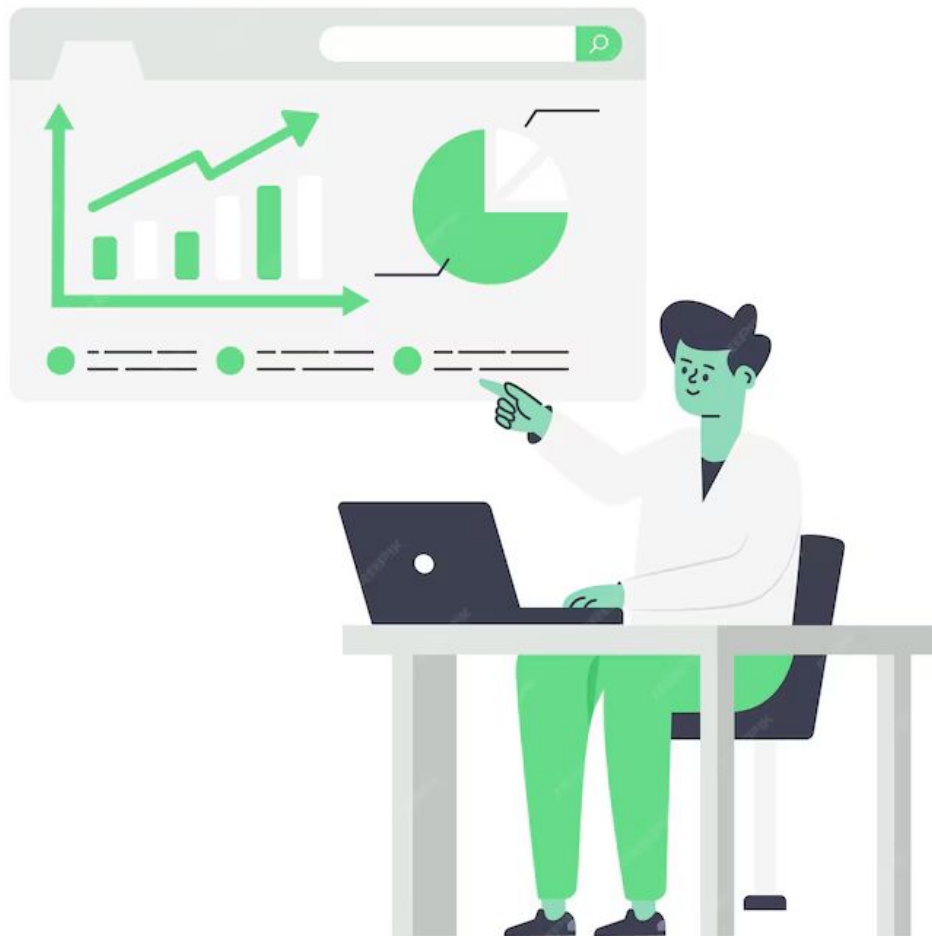




escuela  
británica de  
artes creativas  
y tecnología

# Profesión Científico de Datos





# Índice

01 La empresa

02 Briefing

03 Entregables

04 ¿Por qué participar?

05 Feedback

06 ¿Cómo participar?



# La empresa

Reckitt México, es una empresa líder en el mercado en productos de **higiene del hogar**, **salud** y **nutrición**, con marcas como Vanish, Lysol, Tempra, Sico, Durex y Chocomilk, entre muchas otras.

Para saber más sobre nosotros síguenos en nuestras redes sociales:

<https://www.instagram.com/thisisreckitt/>

<https://www.facebook.com/ThisIsReckitt/>

<https://www.linkedin.com/company/reckitt/>

<https://www.youtube.com/c/reckitt>





# Briefing

Para este reto utilizarás la información de una categoría en específico: Cloros y Blanqueadores de Ropa, la cual es una de las categorías más importantes para Reckitt con la marca Vanish. Sin embargo, no somos lo suficientemente fuertes en todos los segmentos de la categoría (Blanqueadores en Polvo, Liquido, Barra, etc), por lo que necesitamos detectar, dónde estamos perdiendo presencia y quiénes son nuestros principales competidores y qué acciones podemos implementar para mejorar.

En esta categoría también tenemos el producto Lysol Sanitizer y realmente no sabemos si nuestra estrategia ha sido la correcta para ganar participación en el segmento de Sanitizantes de Ropa.

# Briefing

Actualmente, la industria de Retail o Consumo Masivo es una de las que más se beneficia del uso de la Ciencia de Datos. En este sector, se procesan grandes volúmenes de datos para identificar patrones de comportamiento del consumidor, permitiendo anticipar tendencias y optimizar estrategias de marketing, posicionamiento, precios, entre otros. Esto ayuda a las empresas a diferenciarse de la competencia y aumentar la rentabilidad, al ofrecer soluciones más innovadoras y personalizadas para el cliente final, basadas en un análisis predictivo y descriptivo riguroso.

Dentro de esta industria hay empresas de clase mundial como Coca Cola, Pepsi, Nestle, P&G, Reckitt. Y que día a día tratan de usar al máximo las herramientas de Analítica y Ciencia de Datos para mantener su liderazgo, innovación y poder tener esa ventaja competitiva.



# Proyecto

Para resolver el desafío deberás analizar las siguientes preguntas y a partir de estas generar algunas recomendaciones:

- ¿Nuestra marca Vanish, en general está teniendo un buen desempeño?
- ¿Quién(es) es(n) el(los) competidor(es) clave(s) en cada uno de los segmentos?
- ¿Cuáles son los principales productos que están ganando o perdiendo margen de mercado?
- ¿Dónde están las oportunidades clave para hacer crecer el mercado y las ventas?
- ¿Estamos teniendo un buen desempeño en productos Sanitizantes?
- ¿Qué área presenta más oportunidades, y qué recomendación darías para mejorar?
- ¿Cuál es mi pronóstico de ventas para Vanish y Lysol para los siguientes 3 meses, qué margen de error tiene este pronóstico?

# Proyecto

Para responder a las pregunta anteriores, deberás desarrollar lo siguiente:

- El primer paso es cargar las diferentes tablas de datos en Python utilizando Pandas. Luego, deberás realizar la limpieza de datos, asegurando que no haya valores nulos, duplicados o inconsistencias. Posteriormente, realizarás la unión de las tablas relevantes para obtener un conjunto de datos consolidado. Finalmente, se aplicarán transformaciones necesarias para que los datos estén en un formato adecuado para el análisis posterior.
- Antes de construir los modelos, es importante visualizar los datos para entender mejor su distribución y las relaciones entre diferentes variables. Usando Python, se crearán visualizaciones que resalten las tendencias, patrones y posibles outliers en los datos. Esta visualización también servirá para identificar áreas clave para el análisis más profundo.
- Los datos originales se cargarán en una base de datos SQL, lo que permitirá realizar consultas eficientes y análisis relacional. Este paso es crucial para organizar y manejar los datos de manera estructurada dentro de la empresa. Tendrás que crear una base de datos y cargar cada tabla además de realizar algunas consultas y uniones para probar la base.

# Proyecto

- Necesitarás aplicar técnicas de clustering, como K-Means, para segmentar los productos o regiones según características específicas. Este análisis permitirá identificar segmentos clave en los que la marca Vanish tiene presencia y dónde podría mejorar su estrategia.
- Con los insights obtenidos deberás crear un dashboard completo en Power BI que permita a los stakeholders visualizar los resultados de manera clara y concisa. Este dashboard debe ser interactivo, permitiendo a los usuarios explorar diferentes segmentos, productos y escenarios predictivos.
- Por último construirás un modelo predictivo utilizando Python para prever las ventas futuras de productos clave como Vanish y Lysol. Este modelo debe ser entrenado y validado utilizando técnicas adecuadas para asegurar la precisión y fiabilidad de las predicciones. Puedes usar técnicas de regresión o series de tiempo. Deberás incrustar el modelo predictivo de ventas directamente en Power BI (opcional).

\*Recuerda que la información se encuentra almacenada en diferentes tablas, los datos están divididos por áreas; para tu análisis puedes considerar el Total México.



# Entrega final

Para la presentación final de este proyecto considera los siguientes puntos:

- Explica brevemente el objetivo del proyecto y su importancia para Reckitt México, mencionando la relevancia de Vanish y Lysol en la categoría analizada.
- Muestra un resumen del proceso de carga, limpieza y transformación de datos, destacando los pasos clave realizados para consolidar la información.
- Incluye visualizaciones clave que revelen tendencias y patrones en los datos, con breves notas que expliquen los principales insights.

# Entrega final

- Describe cómo se organizaron los datos en la base de datos SQL, incluyendo capturas de pantalla o diagramas que ilustren las consultas realizadas.
- Presenta el análisis de clustering realizado, mostrando los segmentos identificados y cómo estos insights pueden mejorar la estrategia de Vanish.
- Incluye capturas de pantalla del dashboard en Power BI, destacando su interactividad y cómo permite a los usuarios explorar los datos y predicciones.
- Presenta los resultados del modelo predictivo, con gráficos que comparen las predicciones de ventas con los datos históricos, e indica si fue integrado en Power BI.
- Resume los hallazgos clave y proporciona recomendaciones estratégicas, destacando los próximos pasos sugeridos para Reckitt México.

# Valora que tu proyecto tenga:

- Aplicabilidad real del proyecto
- Nivel de complejidad en el proyecto
- Uso correcto de las herramientas para ciencia de datos
- Innovación en los productos finales
- Documentación y entrega de la presentación final.