

# ProSales Analytics

*“Donde la Data Impulsa  
Ventas y  
Recomendaciones con  
Precisión Proactiva”*



Proyecto de Data Science  
para Predicción de  
Ventas y  
Recomendaciones de  
Productos.

Presentado por:  
Lada Corina  
Hellen Lanza

## ProSales Analytics

*“Donde la Data Impulsa Ventas y Recomendaciones con Precisión Proactiva”*

Nuestro proyecto para el bootcamp de Data Science se trata de la Predicción de Ventas y Recomendaciones Personalizadas de Productos. Para explicar nuestro proyecto vamos a seguir el siguiente orden:

- La estructura del Proyecto.
- Beneficiario, valor y utilidad del proyecto.
- Proceso del Proyecto
- Impacto y resultado
- Líneas de continuidad
- Conclusiones

### La estructura de proyecto

- Descripción de la empresa analizada.
- La identificación del problema y objetivos, justificación de la importancia del proyecto.
- Limpieza de datos y preprocesamiento.
- Análisis exploratorio de datos dividido en tres parte: de ventas, de productos y de clientes.
- Predicción de ventas para la próxima semana: implementación de un modelo de Serie Temporal con Random Forest.
- Categorización de Clientes mediante K-Means
- Recomendación personalizada de productos: implementación del modelo KNNBasic de la Librería Surprise.
- Conclusiones y líneas de continuidad.
- Anexos.

### Beneficiarios, valor y utilidad del proyecto

La empresa objeto de análisis se ve confrontada con desafíos significativos en la gestión de stock, los cuales se manifiestan a través de fluctuaciones en las cantidades solicitadas, problemas asociados a la caducidad de productos y, como consecuencia, la pérdida de ingresos. La predicción de ventas podría ser una solución porque:

- Ayuda a mantener niveles óptimos de inventario, evitando tanto excesos que conducen a costos de almacenamiento innecesarios como escasez que resulta en pérdida de ventas.
- La empresa puede minimizar la posibilidad de caducidad de productos, esto reduce las pérdidas asociadas con productos vencidos.
- Contribuye a una gestión más eficiente de la cadena de suministro y de las operaciones logísticas.
- Mantener niveles adecuados de stock asegura que los productos estén disponibles cuando los clientes los requieran.

La empresa tiene como objetivo incrementar sus ingresos, y como estrategia proponemos la implementación de un sistema de recomendación personalizada de productos para cada cliente. La importancia de un sistema de recomendación personalizada:

- Mejora la experiencia de compra, demostrando que la empresa comprende las preferencias individuales de cada cliente y ellos se sienten más conectados y valorados.
- Al sugerir productos relevantes basados en el historial de compras y preferencias, se aumenta la probabilidad de ventas por impulso.
- Puede ayudar a la empresa a destacar productos estratégicos, contribuyendo a la optimización de la estrategia de precios.
- Puede ser una ventaja competitiva, diferenciando a la empresa y posicionándola como una marca que brinda experiencias de compra únicas.

En conclusión, la implementación de un sistema de recomendación personalizada no solo se trata de aumentar las ventas de manera inmediata, sino de construir relaciones más sólidas con los clientes y mejorar la estrategia general de la empresa.

## Proceso del Proyecto.

### Limpieza

Dado que se trataba de datos reales, ambos dataframes contenían varias incongruencias que constituían un obstáculo para avanzar. Al cargar los archivos CSV en Python (utilizando pandas), comenzamos a analizar cada una de las columnas y las filas (en total 232253 en el CSV de ventas y 3992 en el CSV de productos). Creamos dos funciones para limpiar los dataframes, luego unirlos y crear un solo dataframe para continuar con las siguientes fases del proyecto.

### Análisis exploratorio de datos – EDA.

Al conocer la estructura del dataframe, tomamos la decisión de dividir nuestro análisis en ventas, productos y clientes. Tomando como base estos tres elementos, buscábamos observar cómo influyen las diversas columnas. Las herramientas que utilizamos para crear los gráficos fueron las bibliotecas de: Matplotlib, Seaborn y Plotly. Por otra parte, cabe destacar que el análisis de clientes solo se realizó para los del almacén.

### Creación de Modelos de Machine Learning.

Para dar respuesta a los problemas de la empresa, creamos dos enfoques. Por una parte, la predicción de ventas de productos, y por otra parte, la recomendación de productos a clientes.

#### Predicción de Ventas

Para predecir las ventas de productos, organizamos nuestro dataframe por semanas (en total, 20 semanas). Inicialmente, creamos las primeras iteraciones de diferentes modelos para un solo producto, pero el error era muy alto. Luego continuamos haciendo iteraciones

de modelos utilizando el dataframe completo y luego agregando variable, pero el error continuaba siendo muy alto.

Para nuestro modelo final, optamos por utilizar Random Forest Regressor para predecir el siguiente valor en la serie de tiempo semanal de cantidad vendida (One-Step prediction). Utilizamos los primeros cuatro meses como conjunto de entrenamiento y el mes de noviembre como conjunto de evaluación. Aunque este último modelo ha demostrado ser el más efectivo, el error sigue siendo alto. No obstante, somos conscientes de que con más datos, este modelo puede mejorar exponencialmente.

### **Recomendación de Productos a Clientes**

Creamos un modelo de recomendación de productos basado en el filtrado colaborativo. El filtrado colaborativo basado en clientes sugiere productos a un usuario en función de las preferencias y comportamientos de usuarios similares. En este contexto, utilizamos los clientes del almacén, los únicos diferenciados en la base de datos inicial. Para este modelo, utilizamos la librería Surprise y el modelo KNN. Surprise ayuda a analizar el comportamiento pasado de los usuarios y hace recomendaciones basadas en la similitud entre usuarios o elementos. Utilizamos el modelo KNN Basic encontrado en la librería Surprise porque puede proporcionar recomendaciones basadas en similitud, incluso cuando la información histórica es limitada.

### **Segmentación de Clientes**

Finalmente, realizamos una segmentación de los clientes en grupos homogéneos según comportamientos de compra y preferencias con el modelo de aprendizaje no supervisado K-means. Esto permite a la empresa personalizar estrategias de marketing para cada segmento. Los elementos recomendados pueden basarse en las preferencias del grupo al que pertenece el cliente.

## **Impacto y Resultado**

Tomando como base que los datos reales de la empresa abarcan solo 5 meses, los resultados de cada uno de los modelos no son los óptimos. La escasez de datos nos dificulta la creación de modelos con una precisión más alta.

En nuestro modelo de predicción de ventas, el error en el conjunto de entrenamiento es aproximadamente 0.7966, mientras que en el conjunto de prueba es alrededor de 0.7144. Ambos errores son relativamente bajos, lo que indica que el modelo está aprendiendo bien de los datos de entrenamiento y generalizando adecuadamente a nuevos datos. El  $R^2$  en los datos de entrenamiento es 0.3024, y en los datos de prueba es 0.3195. Estos valores sugieren que el modelo explica alrededor del 30% al 32% de la variabilidad en los datos, lo cual es relativamente bajo pero no inusual para problemas de predicción de series temporales. Hay factores no capturados por el modelo que contribuyen a la variabilidad en las ventas.

Por otra parte, el modelo de recomendación de productos muestra un error RMSE de 24.0376 y un  $R^2$  de 0.3177. Estos valores sugieren que el modelo explica alrededor del 31% de la variabilidad en los datos.

## Líneas de continuidad

Si el proyecto continúa en el futuro y se obtienen más datos, hay varias áreas clave en las que podríamos centrarnos para mejorar y refinar nuestros modelos como:

1. Recolectar datos adicionales a lo largo del tiempo para extender el historial de ventas. Un conjunto de datos más extenso permite capturar patrones estacionales y tendencias a largo plazo, mejorando la capacidad predictiva del modelo.
2. Con un conjunto de datos más amplio, podríamos analizar tendencias a largo plazo en el comportamiento de ventas. Esto podría incluir cambios estacionales anuales, patrones de crecimiento a lo largo de varios años y otros factores de influencia a largo plazo.
3. Identificar y agregar nuevas características relevantes que puedan influir en las ventas.
4. Ya con más datos se puede considerar modelos más avanzados y complejos para capturar relaciones no lineales y comportamientos más sofisticados en los datos.

En resumen, con un enfoque futuro en la expansión del conjunto de datos y la aplicación de técnicas más avanzadas, podríamos mejorar significativamente la capacidad del modelo para prever las ventas en un entorno más complejo y dinámico.

## Conclusiones

Por falta de datos, el proyecto no obtuvo los resultados más óptimos, a pesar de ello creamos dos modelos que sirven como base para mejorar.

Para solucionar el problema de predicción de ventas el modelo utilizamos las series temporales y el modelo de Random Regression con un error  $R^2$  de 0.3195.

Para solucionar el problema de recomendación de productos utilizamos un enfoque colaborativo utilizando la librería surprise y el modelo KNN Basic con un error  $R^2$  de 0.3177

Los resultados del Proyecto ProSales Analytics ayuda a la empresa a comprender la importancia de los datos y cómo estos pueden ayudarlo a tomar decisiones estratégicas.

El Proyecto ProSales Analytics proporciona una base para la toma de decisiones informada en la gestión de inventario, una mayor comprensión de los patrones de ventas, comprensión de clientes para poder aplicar estrategias de marketing personalizadas y saber que productos son más útiles para los clientes.

El proyecto sienta las bases para futuras mejoras con la incorporación de más datos y técnicas avanzadas. La implementación efectiva de estos resultados puede llevar a la optimización continua y al crecimiento sostenible de la empresa.