Informe Final del Proyecto de Análisis de Datos

\*\*Estudiante:\*\* Luis Alfredo Denis Garcia

\*\*Proyecto:\*\* Análisis Exploratorio y Predicción de Ventas

# 1. Introducción

Este informe describe el desarrollo completo del proyecto de análisis de datos aplicando técnicas de ciencia de datos y estadística descriptiva sobre un conjunto de datos de ventas. El objetivo principal fue identificar productos y locales más rentables, explorar tendencias de consumo, y aplicar un modelo predictivo para anticipar la demanda futura.

# 2. Objetivos del Proyecto

- Identificar los productos con mayor volumen de ventas e ingresos.  
- Determinar los productos más y menos rentables en términos de margen.  
- Analizar la distribución geográfica de las ventas.  
- Detectar patrones temporales y estacionalidad.  
- Aplicar un modelo de series de tiempo para predecir las ventas futuras.

# 3. Metodología

# 3.1 Recopilación de Datos

Se trabajó con un archivo CSV proporcionado por la organización, conteniendo información detallada de pedidos, productos, tiendas, precios y fechas.

# 3.2 Importación de Librerías

Se utilizaron librerías especializadas como pandas, numpy, matplotlib, seaborn, scipy, statsmodels, entre otras.

# 3.3 Limpieza de Datos

- Inspección de la estructura del dataset.  
- Tratamiento de valores faltantes y tipos de datos.  
- Cálculo del margen de ganancia (%) como nueva columna.

# 3.4 Análisis Exploratorio

- Análisis univariado y bivariado.  
- Identificación de productos y ubicaciones clave.  
- Visualizaciones con histogramas, boxplots, heatmaps.

# 3.5 Análisis Temporal

- Agrupación mensual y anual.  
- Identificación de estacionalidad en meses clave.  
- Detección de caídas post-2019.

# 3.6 Modelado Predictivo

- Modelo SARIMA (1,1,1)x(1,1,1,12).  
- Validación con prueba ADF.  
- Predicción a 12 meses con intervalo de confianza del 95%.

# 4. Principales Hallazgos

- PCs de escritorio son los más vendidos.  
- DVD binders tienen alto margen pero bajo volumen.  
- Refrigeradores y lavadoras generan pérdidas.  
- EE.UU., Online y UK dominan el volumen.  
- Diciembre, enero y febrero son meses pico.

# 5. Recomendaciones Estratégicas

- Asegurar stock de productos más vendidos.  
- Promocionar productos con alto margen.  
- Revisar o eliminar productos con pérdidas.  
- Invertir en campañas de temporada alta.  
- Optimizar el canal Online.

# 6. Conclusión

Este proyecto demuestra la utilidad de la ciencia de datos para extraer información valiosa de datos de ventas. Las herramientas estadísticas y predictivas respaldan decisiones estratégicas informadas.

# 7. Próximos Pasos

- Desarrollar dashboard interactivo.  
- Automatizar reportes.  
- Aplicar segmentación de productos o clientes.  
- Explorar modelos avanzados de machine learning.