# **Informe Final del Proyecto de Análisis Exploratorio de Datos**

**Estudiante:** Luis Alfredo Denis Garcia  
 **Proyecto:** Análisis Exploratorio y Predicción de Ventas

## **1. Introducción**

Este informe describe el desarrollo del proyecto de análisis de datos aplicado sobre un conjunto de datos de ventas. El objetivo fue identificar productos y ubicaciones con mayor rentabilidad y volumen de ventas, detectar patrones temporales y aplicar un modelo predictivo de series de tiempo para anticipar tendencias futuras.

## **2. Descripción del Problema**

La empresa cuenta con un gran volumen de datos de ventas, pero carece de un análisis sistemático que le permita tomar decisiones estratégicas fundamentadas. Se busca resolver preguntas clave como:

* ¿Cuáles son los productos más vendidos y rentables?
* ¿Qué países y estados tienen mayor volumen de ventas?
* ¿Cuáles son los períodos de mayor demanda?
* ¿Cómo se pueden anticipar tendencias futuras de ventas?

El análisis debe permitir tomar decisiones sobre inventario, campañas de marketing y rentabilidad por producto.

## **3. Datos Seleccionados**

Se utilizó un archivo CSV proporcionado por la empresa, que contiene información detallada de:

* Productos vendidos (nombre, precio, costo)
* Cantidades y fechas de venta
* Países, estados y canales de venta
* Variables calculadas como ingresos, ganancia y margen (%)

Este conjunto de datos incluye varios años de información y permite analizar ventas a nivel mensual y anual.

## **4. Metodología Aplicada**

### **4.1 Recopilación de Datos**

El dataset fue cargado en Python y procesado con herramientas de análisis de datos.

### **4.2 Importación de Librerías**

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

### **4.3 Limpieza y Transformación**

* Conversión de fechas de texto a datetime
* Los valores nulos de delivery date pase a los valores de order date
* Creación de la columna de Margen (%)
* Agrupación y resumen por producto y ubicación
* Deje como valor 0 cero a squere meters porque analice las ventas online.

### **4.4 Análisis Exploratorio**

* Análisis univariado y bivariado
* Visualización de variables clave
* Matriz de correlación

### **4.5 Análisis Temporal**

* Ventas agrupadas por mes y año
* Análisis de estacionalidad mensual

### **4.6 Modelado Predictivo**

* Modelo SARIMA para prever ventas futuras
* Validación del modelo (ADF, MAE, residuos)
* Gráfico de predicción con intervalo de confianza

## **5. Hallazgos Clave**

* PCs de escritorio son los productos más vendidos e ingresan más dinero
* Los organizadores de DVD tienen el mayor margen (%)
* Refrigeradores y lavadoras presentan márgenes negativos
* Los principales mercados son EE.UU., Online y Reino Unido
* Hay estacionalidad clara en diciembre, enero y febrero

## **6. Conclusiones**

Este análisis permitió responder preguntas clave sobre los productos más rentables, los períodos de mayor demanda y las ubicaciones más relevantes. Además, se demostró que aplicar modelos de series de tiempo como SARIMA permite anticipar la demanda futura con un nivel razonable de certeza.

Estas conclusiones pueden ser aplicadas directamente en la gestión del inventario, promociones, abastecimiento y asignación de recursos.

## **7. Recomendaciones**

* Asegurar stock de productos de alta rotación
* Promocionar productos con alto margen y baja venta
* Eliminar o revisar productos con pérdidas
* Focalizar campañas en los meses pico
* Reforzar el canal Online y automatizar reportes