



UNIVERSIDAD TECNOLÓGICA DE PANAMÁ
VICERRECTORÍA DE INVESTIGACIÓN, POSTGRADO Y
EXTENSIÓN FACULTAD DE INGENIERÍA DE SISTEMAS
COMPUTACIONALES



MAESTRÍA EN ANALÍTICA DE DATOS
MODELOS PREDICTIVOS

PROYECTO FINAL

Exploración de Modelos Predictivos para Ventas en AdventureWorks

ELABORADO POR:

CARRERA TELLO, IGNACIO

8-905-1860

PROFESOR:

JUAN MARCOS CASTILLO. PhD

GRUPO:

1AN-215

Introducción

En un entorno empresarial altamente competitivo, la capacidad de predecir tendencias de ventas y comportamiento de clientes es clave para la toma de decisiones. La aplicación de modelos predictivos en bases de datos empresariales permite optimizar estrategias de marketing, mejorar la gestión de inventarios y aumentar la rentabilidad. Este estudio se enfocará en la aplicación de técnicas de machine learning para analizar datos de ventas y predecir patrones de compra en la base de datos AdventureWorks.

1. Justificación La selección de AdventureWorks se basa en los siguientes criterios:

- **Disponibilidad y accesibilidad:** Es una base de datos pública y bien documentada.
- **Estructura empresarial realista:** Contiene datos representativos de un negocio de ventas minoristas.
- **Variedad de atributos:** Incluye información clave para análisis predictivos, como fechas de compra, productos vendidos, montos de venta y segmentos de clientes.
- **Aplicabilidad en modelos predictivos:** Se pueden construir modelos para predecir ventas futuras, segmentación de clientes y optimización del inventario.

2. Antecedentes AdventureWorks es una base de datos de muestra proporcionada por Microsoft, diseñada para entornos de aprendizaje y pruebas de aplicaciones empresariales. Contiene información estructurada en diversas tablas que representan distintos aspectos de un negocio de ventas minoristas en línea. Algunas de sus principales tablas incluyen:

- **SalesOrderHeader y SalesOrderDetail:** Información detallada de cada venta realizada.
- **Customer:** Datos demográficos y comerciales de los clientes.
- **Product y ProductCategory:** Catálogo de productos y sus clasificaciones.
- **SalesTerritory:** Información sobre ubicaciones geográficas de las ventas.

3. Definición del Problema El reto de muchas empresas es predecir las ventas futuras con precisión para optimizar inventarios y mejorar estrategias de marketing. En este estudio, se busca aplicar modelos de machine learning para analizar datos de ventas de AdventureWorks y predecir tendencias de compra, permitiendo tomar decisiones más informadas y reducir costos operativos.

4. Avance de Análisis Predictivo Se emplearán técnicas como regresión lineal, árboles de decisión y redes neuronales para identificar patrones de compra, estacionalidad de ventas y predicción de ingresos. Además, se evaluará el impacto de las características de los clientes en el comportamiento de compra y la fidelización.

5. Gráficas de Visualización Se incluirán gráficas que muestren:

- Distribución de ventas por categoría de producto.
- Tendencias de ventas mensuales y anuales.
- Segmentación de clientes según patrones de compra.
- Comparación de resultados entre distintos modelos predictivos.

✓

```
1 df.describe().T
```

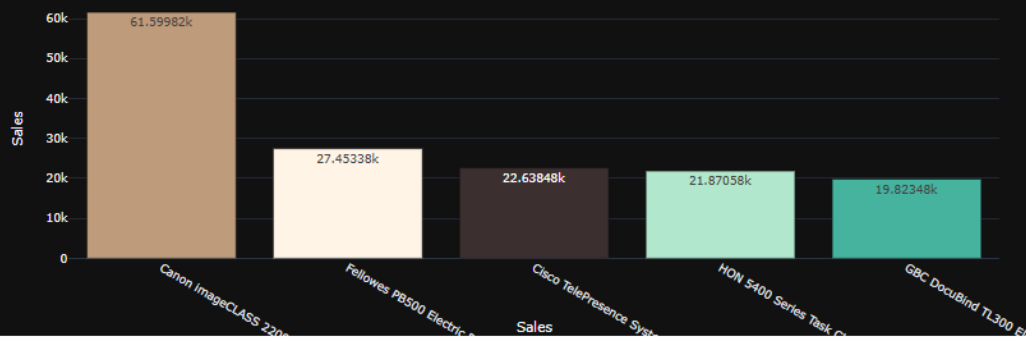
[5]

✓ <1 s - Command executed in 260 ms by IGNACIO TELLO on 2:26:44 PM, 4/08/25

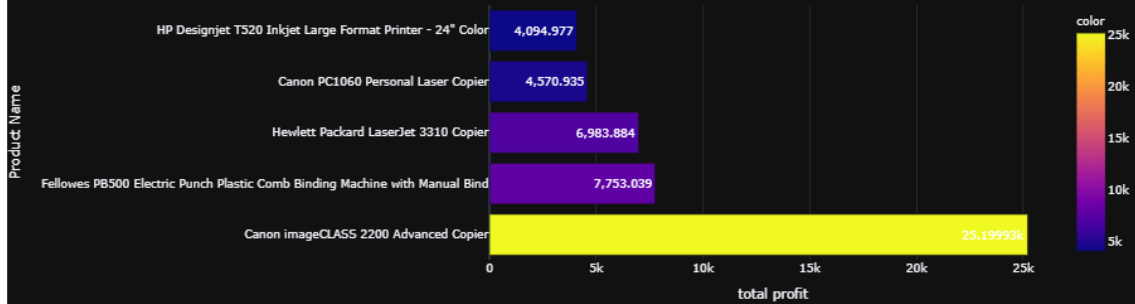
...

	count	mean	std	min	25%	50%	75%	max
Row ID	9994.0	4997.500000	2885.163629	1.000	2499.25000	4997.5000	7495.750	9994.000
Postal Code	9994.0	55190.379428	32063.693350	1040.000	23223.00000	56430.5000	90008.000	99301.000
Sales	9994.0	229.858001	623.245101	0.444	17.28000	54.4900	209.940	22638.480
Quantity	9994.0	3.789574	2.225110	1.000	2.00000	3.0000	5.000	14.000
Discount	9994.0	0.156203	0.206452	0.000	0.00000	0.2000	0.200	0.800
Profit	9994.0	28.656896	234.260108	-6599.978	1.72875	8.6665	29.364	8399.976

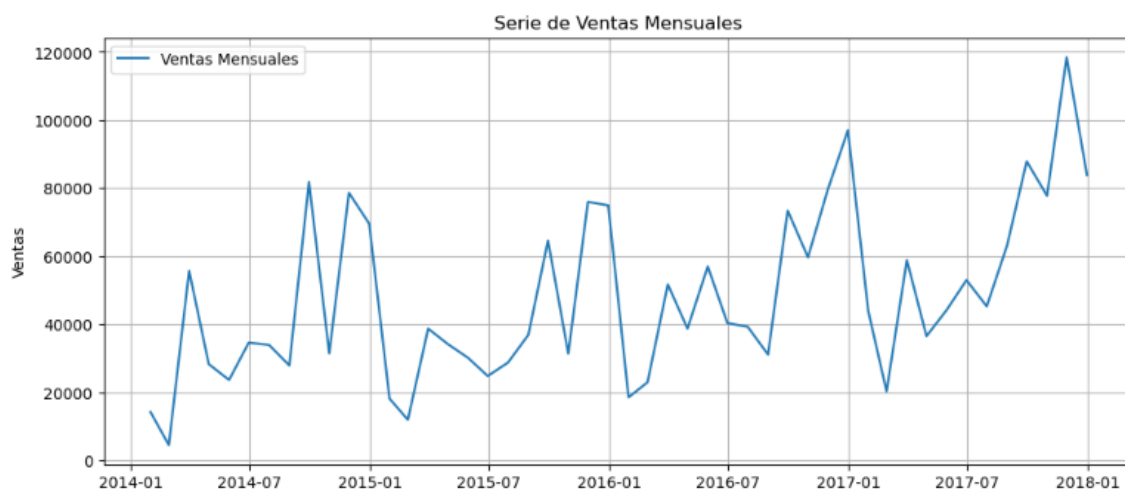
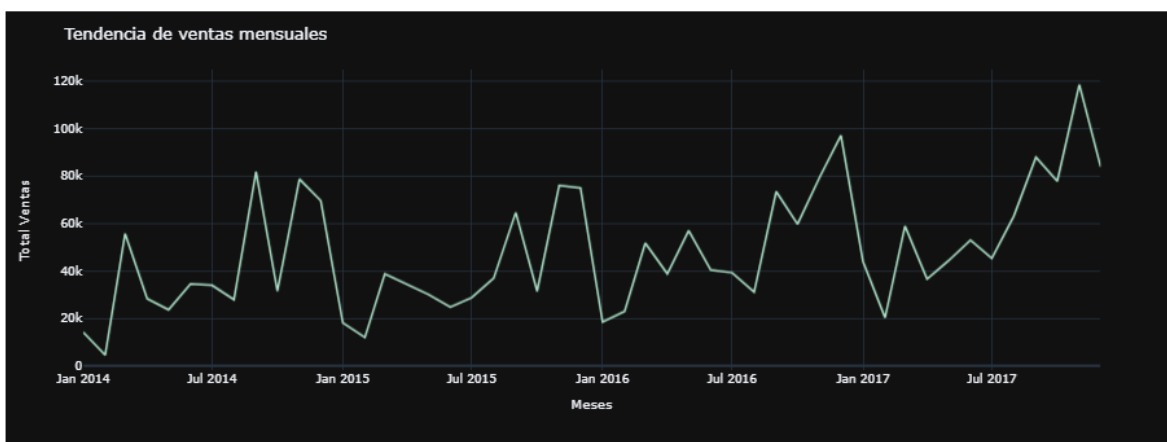
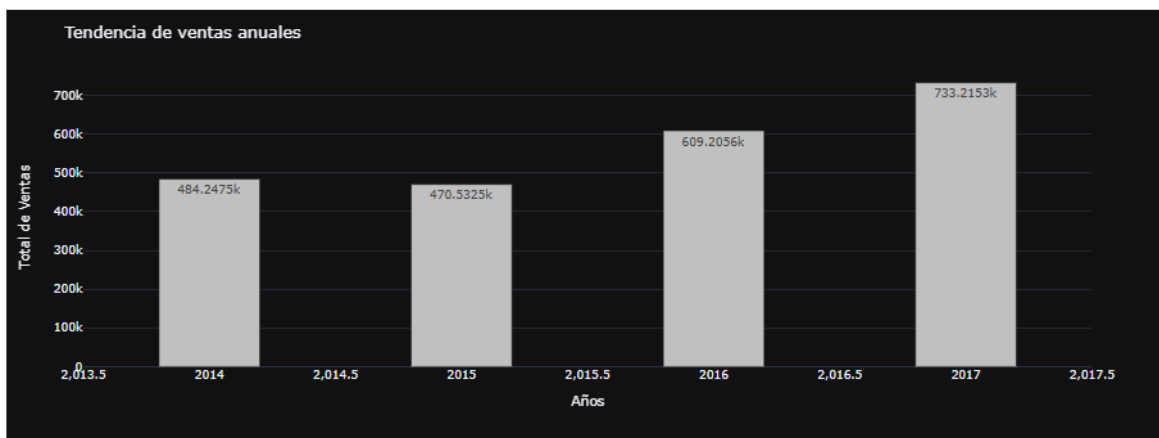
LOS 5 PRODUCTOS MÁS VENDIDOS

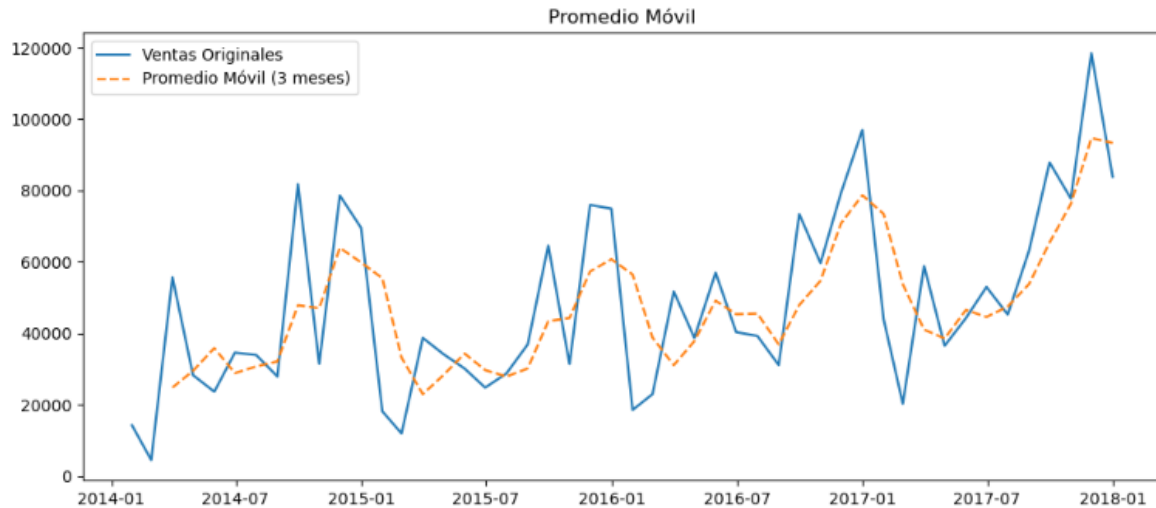


LOS 5 PRODUCTOS MÁS RENTABLES



¿Cuál es la tendencia de ventas a lo largo del tiempo (mensual, anual)?



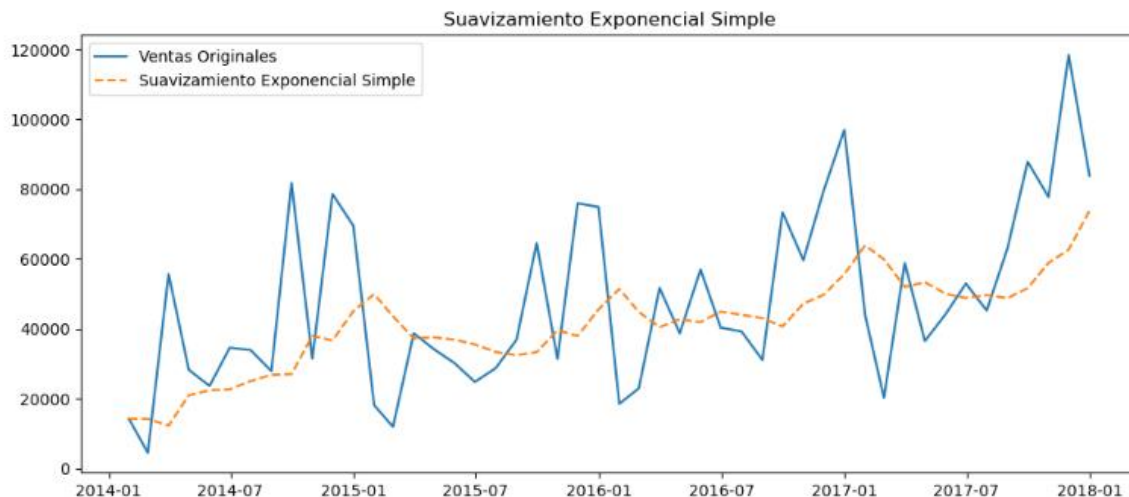


Promedio Móvil **Ventajas: Muy simple, útil para suavizar ruido.

Desventajas: No predice hacia el futuro (solo suaviza lo pasado), y responde lento a cambios.

Cuando usarlo: Si solo necesitas entender la tendencia general.

No es bueno para predicción real, solo para visualización de tendencia. **

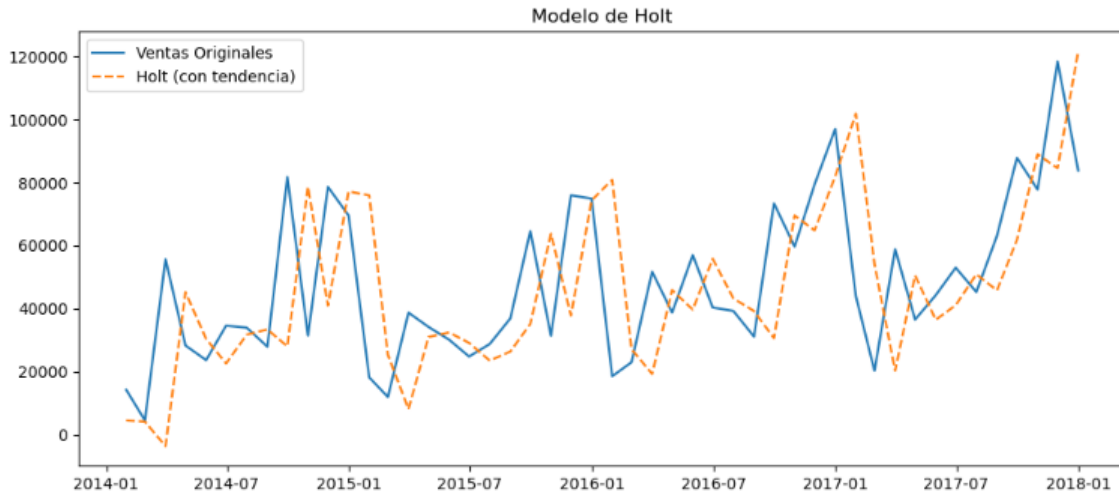


Suavizamiento Exponencial Simple (SES) Ventajas: Se adapta mejor que el promedio móvil.

Desventajas: No captura tendencias ni estacionalidad.

Cuando usarlo: Si los datos son estables sin patrón estacional ni tendencia clara.

En este caso no sería el más adecuado, porque las ventas tienen una tendencia creciente y posible estacionalidad anual.

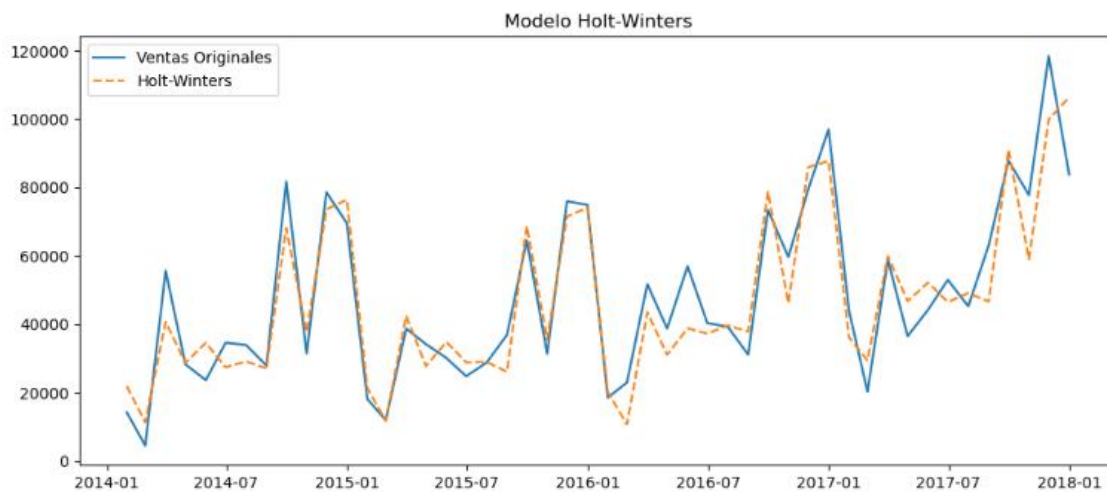


Modelo de Holt (tendencia lineal) Ventajas: Capta tendencia creciente o decreciente.

Desventajas: No capta estacionalidad.

Cuando usarlo: Datos con tendencia, pero sin patrón repetitivo estacional.

Funciona mejor que SES, pero no ideal si hay estacionalidad, como puede ser en ventas (ej: aumentan en fin de año).



Holt-Winters (tendencia + estacionalidad) Ventajas: Capta tendencia y patrones estacionales.

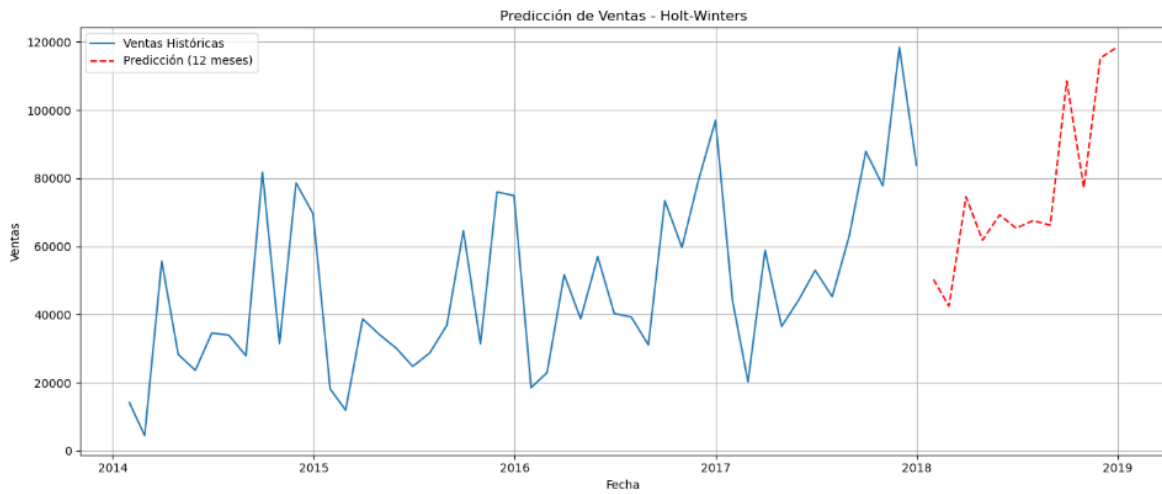
Desventajas: Puede sobreajustarse si no hay estacionalidad real.

Cuando usarlo: Cuando hay estacionalidad clara (mensual, anual) y tendencia.

Es el modelo más completo porque:

- Las ventas muestran una tendencia general creciente.

- Es probable que haya estacionalidad anual (ej: meses con más ventas).



Predicción para los próximos 12 meses:

2018-01-31	50321.382449
2018-02-28	42480.259261
2018-03-31	74560.388209
2018-04-30	61838.886811
2018-05-31	69236.677656
2018-06-30	65358.588777
2018-07-31	67568.990552
2018-08-31	66212.613835
2018-09-30	108524.429772
2018-10-31	77246.201781
2018-11-30	115291.869551
2018-12-31	118522.543554

Freq: M, dtype: float64

Línea de Tiempo de Investigación:

Fase	Actividad	Duración
Semana 1-2	Revisión de la base de datos, limpieza y selección de variables clave	2 semanas
Semana 3-4	Análisis exploratorio de datos (EDA), identificación de tendencias y patrones	2 semanas
Semana 5-6	Desarrollo y entrenamiento de modelos predictivos (Regresión, Árboles de Decisión, Redes Neuronales)	2 semanas
Semana 7	Evaluación de resultados, ajuste de hiperparámetros y comparación de modelos	1 semana
Semana 8	Documentación, generación de informes y presentación de hallazgos	1 semana

Este cronograma permitirá una investigación estructurada y efectiva para evaluar el potencial de los modelos predictivos en la mejora de la toma de decisiones empresariales. Además, se analizará la interpretabilidad de los modelos y su viabilidad en entornos empresariales reales.

Conclusiones

Con base en el análisis exploratorio y el modelado predictivo realizado utilizando la serie de ventas de la base de datos analizada, se pueden extraer varias conclusiones clave:

- **Holt-Winters fue el modelo más preciso**, al capturar tanto la tendencia como la estacionalidad del comportamiento de ventas. Su desempeño fue superior en términos de MAD (desviación absoluta media) y MAPE (error porcentual medio absoluto) en comparación con el Promedio Móvil, Suavizamiento Exponencial Simple y el modelo de Holt.
- **Las ventas muestran una clara estacionalidad anual**. Los picos tienden a repetirse en los mismos meses cada año, lo que hace que los modelos con componente estacional, como Holt-Winters, sean más adecuados.
- El análisis descriptivo reveló que **los productos más vendidos no siempre son los más rentables**, y que los descuentos tienen un impacto negativo en las ganancias cuando superan el 20%.
- **Las categorías de productos y regiones geográficas** juegan un papel importante en la rentabilidad y volumen de ventas, lo que sugiere que variables demográficas y de ubicación deben ser consideradas en futuros modelos predictivos.
- En cuanto al preprocesamiento, se confirmó que los datos no contenían valores nulos significativos ni duplicados, y se establecieron correctamente las fechas para los modelos temporales.
- **La validación del modelo Holt-Winters** con un conjunto de prueba (últimos 12 meses) mostró predicciones coherentes y estables, lo que respalda su aplicabilidad en un entorno empresarial.

Bibliografía

- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Microsoft Docs. (2024). *AdventureWorks Database Documentation*. Disponible en: <https://docs.microsoft.com/en-us/sql/samples/adventureworks>
- Witten, I., Frank, E., Hall, M., & Pal, C. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

ANEXOS

Total de ventas por año:

100 %

Results Messages		
	Año	TotalVentas
1	2014	22419498.3157
2	2013	48965887.9632
3	2012	37675700.312
4	2011	14155699.525

Top 5 productos más vendidos

Results Messages		
	Name	CantidadVendida
1	AWC LOGO CAP	8311
2	WATER BOTTLE - 30 OZ.	6815
3	SPORT-100 HELMET, BLUE	6743
4	LONG-SLEEVE LOGO JERSEY, L	6592
5	SPORT-100 HELMET, BLACK	6532

Clientes con mayor volumen de compras

Results Messages			
	CustomerID	CompanyName	TotalCompras
1	29818	Brakes and Gears	989184.082
2	29715	Excellent Riding Supplies	961675.8596
3	29722	Vigorous Exercise Company	954021.9235
4	30117	Totes & Baskets Company	919801.8188
5	29614	Retail Mall	901346.856

Selección de Variables

Results Messages				
	PromedioVentas	PromedioImpuestos	PromedioEnvio	PromedioSub Total
1	3915.9951	323.7557	101.1736	3491.0656

```

48 -- Verificar si el tipo de cliente influye en las ventas
49 SELECT SO.Description as CustomerType, COUNT(SH.SalesOrderID) AS NumeroVentas, SUM(SH.TotalDue) AS TotalVentas
50 FROM Sales.Customer C
51 Inner JOIN Sales.SalesOrderHeader SH ON C.CustomerID = SH.CustomerID
52 Inner JOIN Sales.SalesOrderDetail SD ON SH.SalesOrderID = SD.SalesOrderID
53 Inner JOIN Sales.SpecialOffer SO ON SD.SpecialOfferID = SO.SpecialOfferID
54 GROUP BY SO.Description
55 ORDER BY TotalVentas DESC;

```

Results Messages			
	CustomerType	NumeroVentas	TotalVentas
1	No Discount	115884	2728648489.9237
2	Volume Discount 11 to 14	3428	91608616.0965
3	Volume Discount 15 to 24	606	40820134.1373
4	Touring-3000 Promotion	524	25715167.5485
5	Touring-1000 Promotion	244	11041307.6498
6	Mountain-500 Silver Clearance Sale	169	7497361.0031
7	Volume Discount 25 to 40	80	5540046.7703
8	Sport Helmet Discount-2002	98	5142021.4538
9	Sport Helmet Discount-2003	84	4405695.0115
10	Road-650 Overstock	61	3424403.0756
11	Mountain-100 Clearance Sale	137	2959401.8322
12	Volume Discount 41 to 60	2	167479.5391

```

56
57 --e. Selección de Modelos
58 SELECT
59     YEAR(SH.OrderDate) AS Año,
60     MONTH(SH.OrderDate) AS Mes,
61     C.CustomerID,
62     D.Name as CompanyName,
63     P.Name AS Producto,
64     SD.OrderQty AS CantidadVendida,
65     SH.TotalDue AS MontoVenta,
66     SH.Freight AS CostoEnvio

```

Results Messages								
	Año	Mes	CustomerID	CompanyName	Producto	CantidadVendida	MontoVenta	CostoEnvio
1	2011	5	29825	Better Bike Shop	MOUNTAIN-100 BLACK, 42	1	23153.2339	616.0984
2	2011	5	29825	Better Bike Shop	MOUNTAIN-100 BLACK, 44	3	23153.2339	616.0984
3	2011	5	29825	Better Bike Shop	MOUNTAIN-100 BLACK, 48	1	23153.2339	616.0984
4	2011	5	29825	Better Bike Shop	MOUNTAIN-100 SILVER, 38	1	23153.2339	616.0984
5	2011	5	29825	Better Bike Shop	MOUNTAIN-100 SILVER, 42	1	23153.2339	616.0984
6	2011	5	29825	Better Bike Shop	MOUNTAIN-100 SILVER, 44	2	23153.2339	616.0984
7	2011	5	29825	Better Bike Shop	MOUNTAIN-100 SILVER, 48	1	23153.2339	616.0984
8	2011	5	29825	Better Bike Shop	LONG-SLEEVE LOGO JERSEY, M	3	23153.2339	616.0984
9	2011	5	29825	Better Bike Shop	LONG-SLEEVE LOGO JERSEY, XL	1	23153.2339	616.0984
10	2011	5	29825	Better Bike Shop	MOUNTAIN BIKE SOCKS, M	6	23153.2339	616.0984
11	2011	5	29825	Better Bike Shop	AWC LOGO CAP	2	23153.2339	616.0984
12	2011	5	29825	Better Bike Shop	SPORT-100 HELMET, BLUE	4	23153.2339	616.0984
13	2011	5	29672	Pedals Warehouse	ROAD-650 RED, 44	1	1457.3288	38.8276
14	2011	5	29672	Pedals Warehouse	ROAD-450 RED, 52	1	1457.3288	38.8276