

# Reporte Completo – Punto 2 (EDA Ventas Funcionales)

## 1. Contexto del Problema

Una cadena de retail desea analizar el comportamiento mensual de las ventas de **100 tiendas durante 24 meses** con el fin de **identificar patrones de transaccionalidad**. Esto servirá para segmentar tiendas en grupos comparables y luego estimar la curva de ventas de una nueva tienda (Tienda\_101).

**Base entregada:** `ventas_funcionales.csv` → ventas mensuales de las 100 tiendas.

**Objetivo del EDA:** Explorar calidad de datos, distribución, outliers, tendencias y preparar las distintas vistas que servirán como insumo para el análisis de clusterización funcional.

---

## 2. Calidad de Datos

- **Observaciones:** 24 meses × 100 tiendas (2.400 puntos).
  - **Nulos:** ninguno.
  - **Negativos:** ninguno.
  - **Zeros:** ninguno.
  - **Periodos:** secuenciales del 1 al 24.
  - **Conclusión:** base limpia, no requiere correcciones. Se genera copia estándar en `data/clean/ventas_funcionales_clean.csv`.
- 

## 3. Estadísticas Globales

- Media global: **1.756** ventas mensuales.
  - Mediana: **1.467** (menor que la media → **sesgo positivo**).
  - Desviación estándar: **1.258** (muy alta → dispersión marcada).
  - Rango: **107 – 10.946**.
  - Implicación: existen **outliers estructurales** con ventas extraordinarias.
- 

## 4. Outliers y Segmentación Preliminar

- Tiendas con medias más altas: **Tienda\_43 (~7.168)**, **Tienda\_99 (~6.210)**, **Tienda\_41 (~4.599)**, **Tienda\_44 (~4.044)**.
- Operan en otra escala → **hipermercados/premium**.
- Segmentos naturales observados:
- **Normales:** medias ~900–2200.

- **Media-alta:** medias 2500–3500.
  - **Hiper:** medias >4000.
- 

## 5. Tendencias Globales

- La **curva promedio** de todas las tiendas es **creciente y sostenida** en 24 meses.
  - La **banda intercuartílica (Q25–Q75)** crece ligeramente con el tiempo → diferencias entre tiendas se amplían.
  - Implica que aunque la tendencia global es positiva, la **heterogeneidad aumenta**.
- 

## 6. Distribución por Segmento

- **Normales:** homogéneos, estables, IQR angosto.
  - **Media-alta:** más heterogéneos en nivel, pero consistentes en forma (IQR ancho, estable).
  - **Hiper:** muy dispersos entre sí, IQR enorme desde el inicio. Confirma que no deben mezclarse con el resto sin estandarización.
- 

## 7. Crecimiento y Consistencia por Tienda

- **Crecimiento porcentual promedio:** identifica “emergentes” (ej. **Tienda\_55: 11% mensual**).
- **Correlación tiempo-ventas:** mide estabilidad de tendencia. Valores 0.99 → lineales y consistentes (ej. Tienda\_27). Valores 0.94–0.95 → más fluctuantes (ej. Tienda\_76).

### Insights:

- Existen tiendas pequeñas con gran crecimiento relativo (emergentes).
  - Hipermercados crecen menos en % pero con magnitudes absolutas muy altas.
  - Algunas curvas son mucho más estables que otras.
- 

## 8. Curvas Representativas

Selección automática de 4 casos: - **Típica y estable:** Tienda\_77 (media ~2.018, growth ~6%, corr=0.99).

- **Emergente:** Tienda\_55 (media ~375, growth ~11%, corr=0.95).
- **Hiper estable:** Tienda\_99 (media ~6.210, growth ~4%, corr=0.98).
- **Hiper volátil:** Tienda\_44 (media ~4.044, growth ~5%, corr=0.95).

### Comparación Normal vs Hiper:

- Normales llegan a ~2.000 al mes 24.
  - Hipers superan los ~8.000.
  - La brecha se amplía con el tiempo → magnitud es determinante.
-

## 9. Preparación de Datos para Clustering

Dado que el clustering funcional puede variar según cómo se mida la similitud entre curvas, se generan **múltiples vistas**:

1. **Escala original:** `ventas_funcionales_clean.csv` → para capturar diferencias de magnitud.
2. **Z-score por tienda:** `ventas_funcionales_zscore.csv` → centra y escala → compara **formas**.
3. **Normalización por media:** `ventas_funcionales_normalized.csv` → elimina nivel absoluto.
4. **Min-max (0-1):** `ventas_funcionales_minmax01.csv` → útil para comparar trayectorias relativas.

Adicionalmente: - **Metadata de tiendas** (`ventas_funcionales_metadata_tienda.csv`): media, std, CV, crecimiento %, correlación tiempo, segmento por magnitud, flag outlier.

- **Tiendas representativas** (`ventas_funcionales_tiendas_representativas.json`).

- **Formato largo (tidy)** para visualización y librerías de clustering:

- `ventas_funcionales_long.csv`
- `ventas_funcionales_long_zscore.csv`
- `ventas_funcionales_long_normalized.csv`
- `ventas_funcionales_long_minmax01.csv`

---

## 10. Conclusiones Finales del EDA

1. **Datos consistentes:** no requieren limpieza, solo estandarización y diferentes vistas.
2. **Estructura clara de segmentos:** normales, media-alta y hiper.
3. **Outliers estructurales:** hipermercados dominan magnitud, pero algunos son estables y otros volátiles.
4. **Emergentes:** tiendas pequeñas con fuerte crecimiento relativo pueden formar clusters aparte.
5. **Preparación de datos múltiple:** necesaria para comparar formas y niveles en el clustering.

👉 Con este EDA y las bases preparadas en `./data/clean/`, el siguiente paso es realizar el **análisis de clusterización funcional**, generando el dendrograma (Ward + distancia euclidiana en curvas z-score), seleccionar el número óptimo de clusters y describir cada grupo resultante.