

# Parcial – Ciencia de Datos Aplicada (DEFI IV)

---

Profesor: Gustavo Venegas

Duración: 1 semana

Fecha de entrega: jueves 4 (11:59 p.m.)

Formato de entrega: Un archivo en Google Sheets que debe incluir el número de documento del estudiante y nombre. Cada punto debe estar en una hoja diferente del archivo.

## Punto 1. Estimación de Ventas en el Mes 24 (33.3%)

Contexto:

Una cadena de retail desea estimar las ventas en el mes 24 de sus tiendas a partir de información geográfica, sociodemográfica y de competencia. El objetivo es identificar qué factores impulsan las ventas y poder predecir el comportamiento de nuevas tiendas.

Bases de datos entregadas:

- Punto1.tiendas\_100.csv → Contiene 100 tiendas con todas las covariables y la variable target ventas\_m24.
- tiendas\_10\_no\_target.csv → Contiene 10 tiendas adicionales (Tienda\_101 a Tienda\_110) sin la variable target.

Diccionario de covariables:

lat, lon, store\_cat, pop\_100m, pop\_300m, pop\_500m, commerces, gas\_stations, malls, foot\_traffic, car\_traffic, socio\_level, viviendas\_100m, oficinas\_100m, viviendas\_pobreza, competencia, tiendas\_peq.

Tareas a realizar:

1. Construir un modelo predictivo para ventas\_m24.
2. Evaluar el modelo usando el  $R^2$  y mape.
3. Identificar las variables más importantes e interpretar sus efectos sobre las ventas.
4. Con el modelo entrenado, estimar las ventas de las 10 tiendas nuevas.

Entregable en Sheets:

- Hoja llamada 'Punto 1' con:
  - Estimaciones de ventas\_m24 para las 10 nuevas tiendas.
  - Tabla de importancia de variables.
  - Explicación breve del modelo y su interpretación.

## Punto 2. Clusterización en Datos Funcionales (33.3%)

Contexto:

La empresa desea analizar el comportamiento mensual de las ventas de las 100 tiendas para identificar patrones de transaccionalidad. Esto servirá para diseñar estrategias de crecimiento y segmentación.

Base entregada: ventas\_funcionales.csv — Ventas mensuales de las 100 tiendas durante 24 meses.

Tareas a realizar:

1. Realizar un análisis de clusterización en datos funcionales.
2. Mostrar el dendrograma y justificar el número de clusters seleccionados.
3. Describir las características de cada cluster encontrado.
4. Estimar las ventas de una nueva tienda (Tienda\_101) para sus primeros 24 meses.

Entregable en Sheets:

- Hoja llamada 'Punto 2' con:
  - Dendrograma con explicación.
  - Número de clusters seleccionados y justificación.
  - Descripción de los clusters.
  - Tabla con la estimación de ventas (mes 1 a 24) para la Tienda\_101.

## Punto 3. Estimación de Transacciones de un Producto Financiero (33.3%)

Contexto:

Una oficina financiera desea pronosticar el uso del dinero (valor total) de un producto financiero. Para ello se dispone de la información de los meses de julio a octubre 2022 y se busca estimar los valores de noviembre 2022.

Bases entregadas:

- Punto3.SerieUsodeldinero — Serie diaria de julio a octubre.

Tareas a realizar:

1. Construir un modelo de pronóstico (AutoARIMA o Prophet).
2. Explicar el modelo elegido.
3. Usar los datos de julio a octubre como train y testeo.
4. Pronosticar el uso del dinero para el mes de noviembre.

Entregable en Sheets:

- Hoja llamada 'Punto 3' con:
  - Gráficas de la serie (train y test).
  - Cálculo del MAPE.
- Explicación del modelo y de los resultados.

- Predicciones para noviembre.

### **Criterios de Evaluación**

- Cada punto tiene el mismo peso: 33.3%.
- Dentro de cada punto:
  - MAPE (80%): precisión del modelo según la métrica indicada.
  - Calidad del entregable (20%): claridad de tablas, gráficas y explicación del modelo.