

---

**Entrega:** enviar por email a [lm.pozo@upm.es](mailto:lm.pozo@upm.es), **antes del 12 de julio a las 23:59 horas**, un documento PDF en el que se incluya, para cada pregunta:

1. El enunciado de la pregunta.
2. Los comandos de R utilizados para responder a la pregunta.
3. Los resultados mostrados por R como respuesta a cada comando.
4. Copia, en su caso, de las gráficas generadas por R.
5. Cuando se pida, el análisis de los resultados obtenidos.

**Valoración:** 100% de la nota final.

---

**Datos:** Los datos se han obtenido de <https://data.world> y de <https://www.kaggle.com>

El fichero `Customer_Data.csv` contiene diversos datos descriptivos del uso de tarjeta de crédito de más de 8000 clientes individuales. El fichero `DOC_Thames.1883.2014.monthly.csv` contiene los datos mensuales, desde 1883 hasta 2014, de concentración de carbono orgánico disuelto en el río Támesis a su desembocadura.

---

## 1. Clustering

- (a) Carga los datos del fichero `Customer_Data.csv` en un data frame.
- (b) Quédate solo con los datos correspondientes a las columnas `balance`, `purchases`, `cash_advance`, `purchases_frequency`, `cash_advance_frequency`, `credit_limit`, `prc_full_payment` y `tenure`.
- (c) Si hay alguna fila en la que falten datos (NA), elimínala.
- (d) Haz un clustering de los clientes. Considera si puedes ponderar o escalar los datos de alguna manera razonable. Usa el método elbow para determinar el valor más apropiado de  $k$ .
- (e) Estudia las características de los clusteres obtenidos e intenta hacer una descripción cualitativa de los mismos.

## 2. Series temporales

- (a) Carga los datos del fichero `DOC_Thames.1883.2014.monthly.csv` en un data frame.
- (b) Comprueba que la serie de datos es continua con todos sus datos consecutivos, sin datos perdidos (NA) intermedios. De no ser así, escoge el periodo más largo posible con datos ininterrumpidos.
- (c) Crea una serie temporal con los datos. Reserva los 5 últimos datos para validar el modelo.

- (d) Ajusta a la serie un modelo ARIMA posiblemente estacional (¿con qué periodo?). Razona los parámetros escogidos ( $d, D$ , mediante la gráfica de la serie y sus diferencias, y las varianzas;  $p, q$  y  $P, Q$ , usando las gráficas de ACF y PACF). Opcionalmente, compara con el resultado con el de `auto.arima`.
- (e) Estudia el ACF, PACF y la normalidad de los residuos del modelo. Evalúa si el modelo es razonablemente apropiado, en función de tu estudio.
- (f) Utiliza el modelo para realizar la predicción en los siguientes 5 meses, representa gráficamente las predicciones, sus intervalos de confianza, y compara con los datos reales que has reservado como conjunto de test. Comenta los resultados.