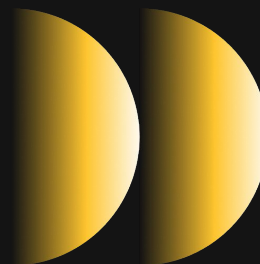


## Entregable 2

Supervised Machine Learning



# Contexto





## Contexto de negocio

A la compañía de telecomunicaciones MobileNOW le preocupa el descenso de número de clientes activos tras sufrir numerosas bajas en los últimos trimestres. Para combatirlo deciden impulsar una campaña de retención de clientes donde centrarán parte de los recursos de telemarketing en evitar que los clientes se den de baja del servicio ofreciéndoles unas condiciones especiales si decidieran renovar por un año más.



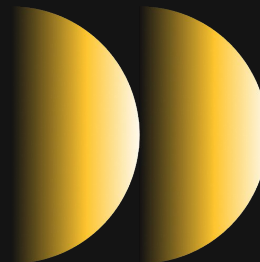
# Tu misión

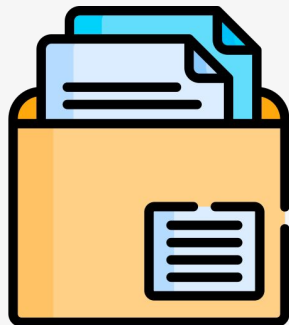
Como Data Scientist, se te encarga lo siguiente:

1. Modelizar la probabilidad de abandono (abandono:  $\text{churn}=1$ ) y calcular para 1.500 clientes preseleccionados la probabilidad de abandonar el próximo mes. Así se podrá centrar la campaña de retención en clientes más propensos a abandonar.
2. Entender qué factores tienen mayor impacto en la decisión del cliente de abandonar, para mejorar el servicio y combatir el abandono a largo plazo.



# Instrucciones





# Archivos entregados

Los archivos están cargados en la carpeta de recursos de la última sesión. Encontrarás:

- Un histórico de datos para entrenar
- Un dataset a predecir con 1.500 observaciones
- Un diccionario con la descripción de las variables del dataset.
- Un notebook vacío con la trama a seguir



# Instrucciones de entrega

- Trabajar el ejercicio por equipos (del TFM o no)
- Crear una carpeta de trabajo en Colab titulada "DSC{nº edicion} – Entregable 2 - {1ºApellido de cada miembro}" (darme permiso de acceso: [mateodebrab@gmail.com](mailto:mateodebrab@gmail.com))
- Guardar el notebook ejecutado con los outputs de las celdas visibles .
- Dejar un log de las transformaciones que aplicamos en el dataset. Ejemplo: `print("Se han eliminado las siguientes columnas: ...")`
- Escribir un código limpio que se pueda ejecutar al completo sin errores dándole al botón "Ejecutar todas las celdas".
- Entregar por Slack un csv con las 1.500 predicciones con el nombre correcto de las columnas, en minúscula.
- Confirmar Nombre y Apellido de los participantes por Slack
- Subir igualmente en entregable a Nuclio
- Se evaluará las predicciones sobre el F1\_SCORE



# Criterios de evaluación

	PROCESOS	PUNTOS	COMO SACAR UN 10
Preprocessing	Data Cleaning y EDA	0,5	Analizo la distribución de las variables
	Tratamiento de nulos	1	Trato de forma coherente los nulos del data set a entrenar y dataset a predecir
	Tratamiento de variables categóricas	1	Trato de forma óptima las variables categóricas
	Alta correlacion y baja varianza	0,5	Elimino variables altamente correlacionadas y de mínima varianza
Modelo	Rebalanceo	0,5	Pruebo rebalancear el train
	Estandarización	0,5	Estandarizo los datasets correctamente para la competición
	Competición de modelos	0,5	Realizo una competición de modelo con mínimo 2 modelos lineales y 2 modelos de reglas (árboles de decisión) sobre la métrica pedida (f1_score)
	Hiperparametrización	0,5	Parametrizo el modelo ganador de la competición (se valora el código, no si consigue mejorar los parametros por defecto del modelo)
Evaluación	Evaluación	0,5	Evaluo las métricas en el dataset de test, siendo el dataset de test lo más parecido al dataset original posible
Interpretabilidad	Interpretabilidad	1	Gráfico las variables más importantes y en qué sentido Saco los coeficientes de un modelo lineal Comento en una celda de texto el impacto positivo/negativo de las top 5 features en el churn.
Predicción	Datacleaning	0,5	Garantizo que se hace la predicción sobre un dataset en las mismas condiciones que el dataset usado para entrenar
	Predicción	3	Entrego un predicción para las 1.500 observaciones en un csv con el nombre correcto de las variables (0 puntos en caso contrario) Nota según <b>F1_SCORE</b> : > 0,85: 3 puntos > 0,8: 2 puntos > 0,7: 1 punto < 0,7: 0 puntos
	TOTAL	10	un código limpio no suma puntos pero uno sucio resta