

# Reporte de Avance

## Proyecto Churn - Predicción de bajas

## Objetivo del modelo

Predecir la probabilidad de baja de clientes dentro de los próximos 3 meses

## Primera etapa de integración

Para la integración se usan las tablas

- Interacciones
- Pagos
- Clientes
- Pólizas

Se adjunta lista de campos que se usan por cada una de ellas en un documento aparte

## Procesamiento de los datos

- **Pólizas:** Se agrupan por CIF\_ID para determinar las bajas y generar una tabla de churn. En base a esa tabla se realizan todas las integraciones por lo que el resto de las tablas debe agruparse por CIF\_ID.

## Procesamiento de los datos

- **Interacciones y Pagos:** Se reagrupan por CIF\_ID y fechas para hacer un análisis temporal
- **Clientes:** Se necesitan algunas transformaciones en los datos pero no es necesaria ninguna reagrupación

## Dudas con respecto a la información

- En el diccionario de datos faltan algunas definiciones de los valores de los campos [COD\_COBRO], [COD\_SECC] y [COD\_SITUACION] para la tabla de pagos. Lo mismo con en campo [GESTION] en la tabla interacciones.

## Dudas con respecto a la información

- Faltarían las interacciones del 2019, aunque sea hasta julio. Es necesario para entrenar al modelo con los datos más recientes posibles. Se puede usar un archivo aparte para no hacerlo tan pesado.
- Pregunta: ¿Puede haber pólizas sin interacciones o todas deben tener al menos una?

## Próximos pasos

- Finalizar la transformación de las tablas que requieran estar ordenadas por CIF\_ID.
- Integrar todos los datos en una sola tabla.
- A partir de los datos integrados obtener un modelo base para refinar iterativamente.



# Voolkia

\_software & services.



\_Como tenemos historia,  
tenemos futuro

Voolkia es una marca registrada de Staffing IT S.A.