## Trabajo fin de Máster: Customer Relationship Management y Software (CRM) Análisis descriptivo y pronóstico de abandono

## Título: MODELO DE DESCRIPTIVO Y PRONÓSTICO DE ABANDONO

# INTRODUCCIÓN

En los últimos años ha surgido la banca electrónica que se caracteriza por la oferta de servicios online, que secundariamente han generado grandes bases de datos cuyo análisis puede ofrecer estrategias de mercado de gran interés. Entre ellas destacamos el modelo predictivo de abandono que analiza el comportamiento de los clientes, e intenta predecir futuros desertores buscando optimizar costes.

# OBJETIVOS

**PRIMARIO:**

1. Realizar un análisis descriptivo de los diferentes canales de contratación para seleccionar aquellos con mejor comportamiento (contratación por canal, rentabilidad)
2. Diseñar un modelo predictivo de abandono de clientes

# DISEÑO DEL estudio

Se extraen datos de Weborama y salesforce entre Noviembre del 2017 hasta Junio de 2019 para el diseño del modelo predictivo. Para obtener el modelo predictivo de abandono se utilizó una regresión logística y randomforest

Los datos utilizados para el análisis descriptivo de los canales de contratación sólo corresponder a los meses entre Octubre del 2018 y Enero de 2019. Para dicho análisis se seleccionaron las siguientes variables:

1. Site\_Offer: Canales de generación de Leads.
2. ID: identificador del cliente
3. Conversion\_date\_hour: Registro de entradas
4. Conversion\_label: Proceso de contración

Para las variables continuas, se calcularon medias; Para las categóricas, se calcularon los porcentajes. Para el análisis de los datos se utilizó Jupyter notebook. Python

# I.ANÁLISIS DE LOS CANALES DE CONTRATACIÓN. RESULTADOS Y DISCUSIÓN

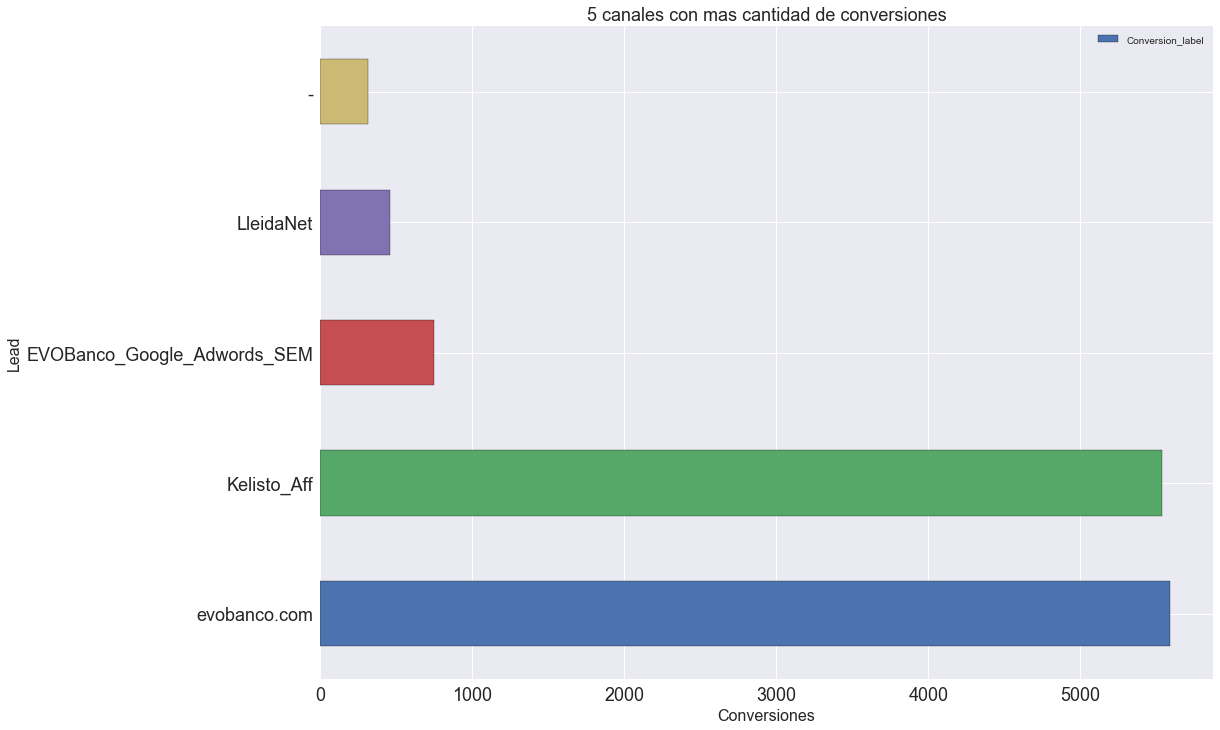
Se parte de una base de datos de 745632 registros.

Canales de contacto con el usuario:

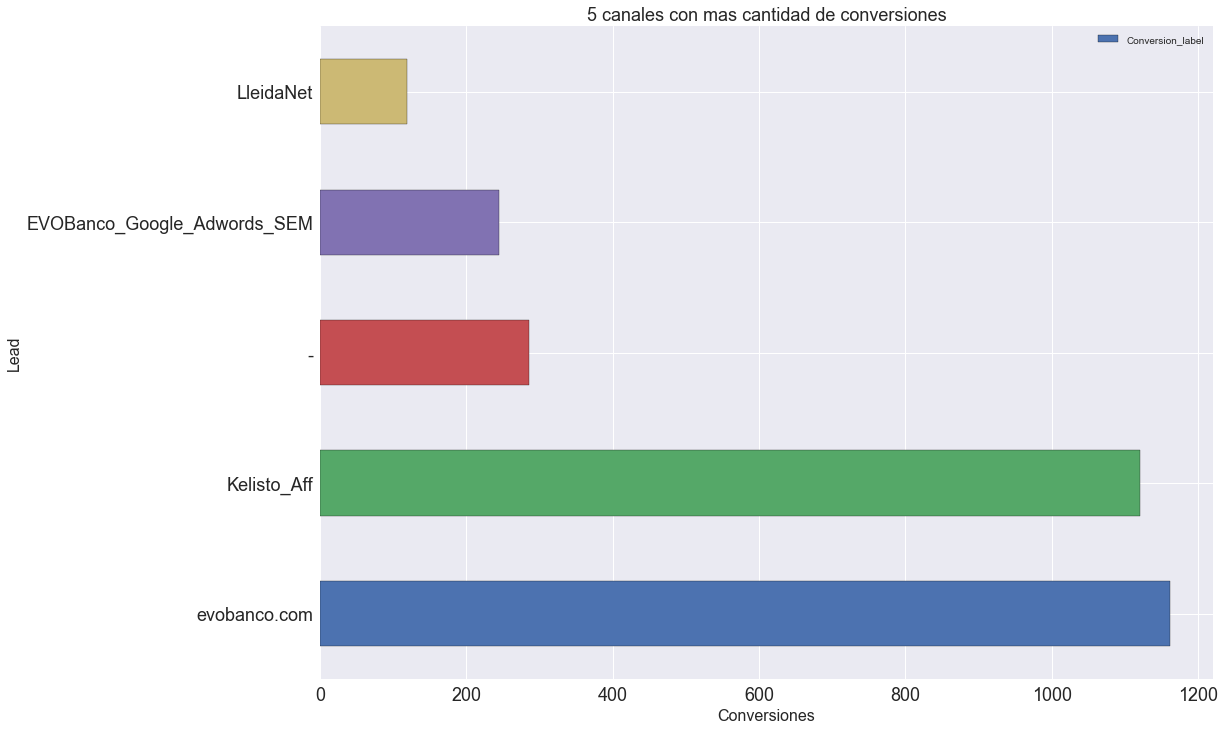
1\_Email\_movil es la primera forma del contacto del usuario con la plataforma en número absolutos. Le sigue 3\_Datos\_personales

**Resultado neto de conversión por canales:**

**Los cinco canales con posible conversión (bruto):**

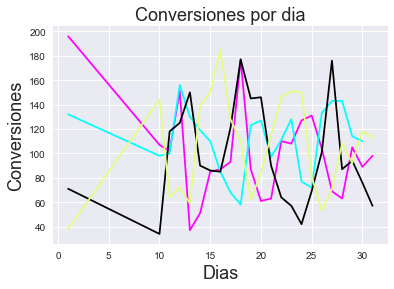


**Los cinco canales con mayor conversión (neto):**

****

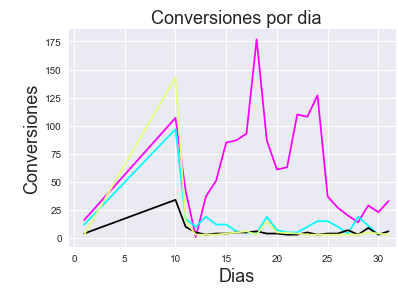
Se observa que la matriz (“evobanco.com”), es el canal con mejor conversión. El segundo canal es canalKelisto\_Aff.

**Tasa de conversión durante los días del mes**



*Nota pie de figura: Rosa (octubre), azul (noviembre), negro (diciembre), amarillo (enero).*

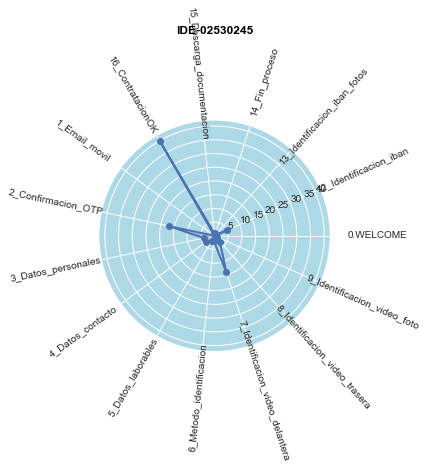
Del análisis se observa que en los meses de Octubre, Noviembre y Diciembre hay una mayor conversión en los primeros días del mes, que posteriormente presenta una disminución y después tres picos de conversión entre los días 10 y 25. En enero destaca un pico de conversión entre los días 5-11 por una campaña publicitaria de préstamos personales (solo para clientes)



**Franja horaria de mayor conversión**

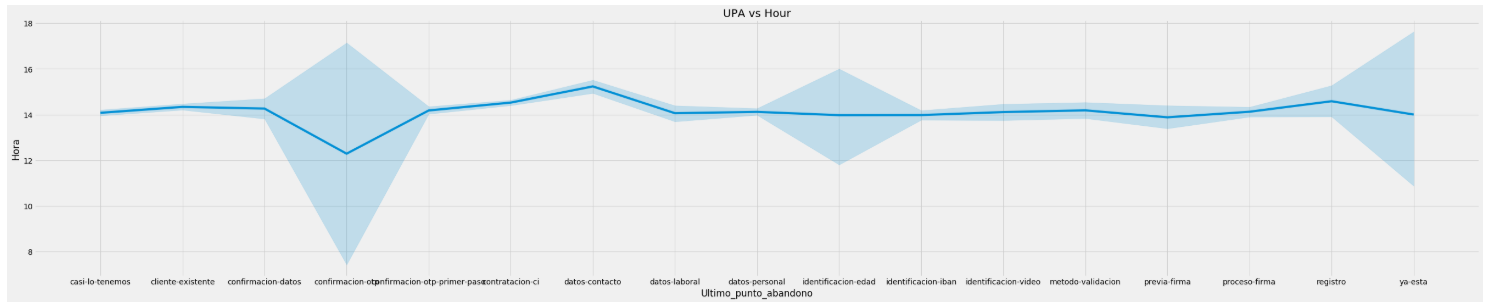
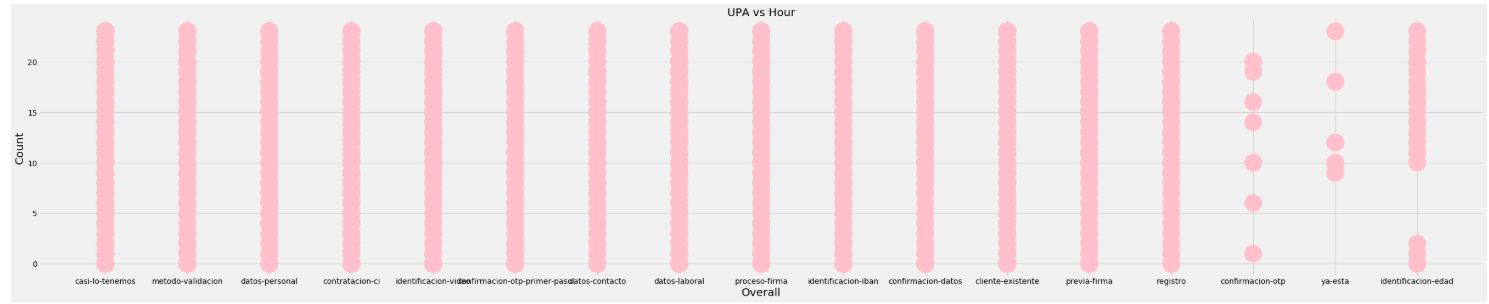
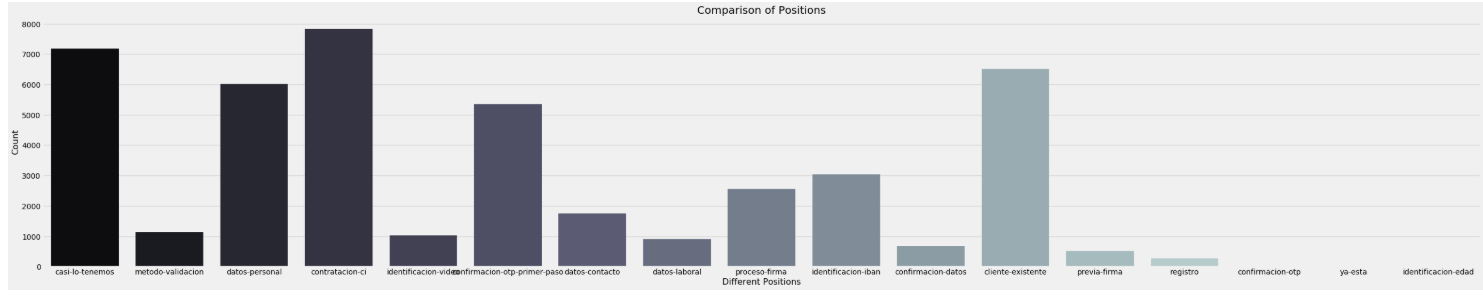
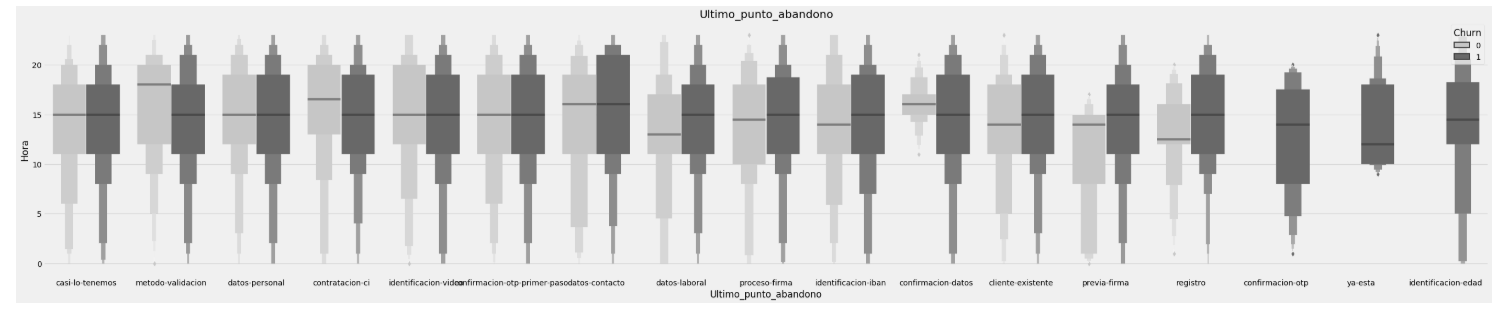
****

Mayor conversión en horario diurno y vespertino (de 10.00 am a 20.00 pm). Como acción adicional se produce un proceso de nurturing a través de SMS para reimpáctar a los clientes que ya están dentro del funnel. Esa acción se realiza sobre las 10:30 de la mañana, ya que dentro del proceso de firma los horarios de mayor conversión son desde las 10:30 a 11:30

**Ejemplo de pasos de clientes:**

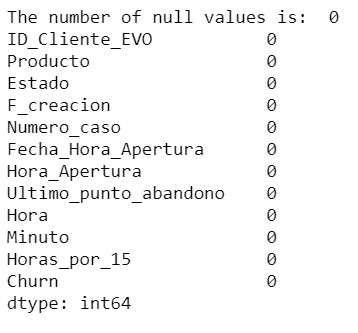
Se puede observar por todos los pasos que tiene que pasar un cliente para poder llegar a finalizar el proceso. Todo esto condicionado por el tiempo, ya sea porque refrescan la página o lo abandonan por tener que hacer otras funciones. Incluso también condicionado de si llama a atención al cliente para que le ayuden en el proceso.

# ii.GrAPHICAL ANALYSIS



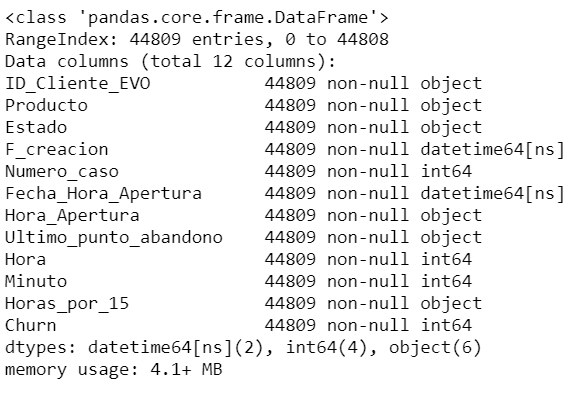
# IIi.MODELO PREDICTIVO DE ABANDONO. RESULTADOS Y DISCUSIÓN

1. Recopilación y limpieza de datos.



Se comprueba que no existen valores nulos

1. Seleccionamos que variables queremos incluir en el modelo. .

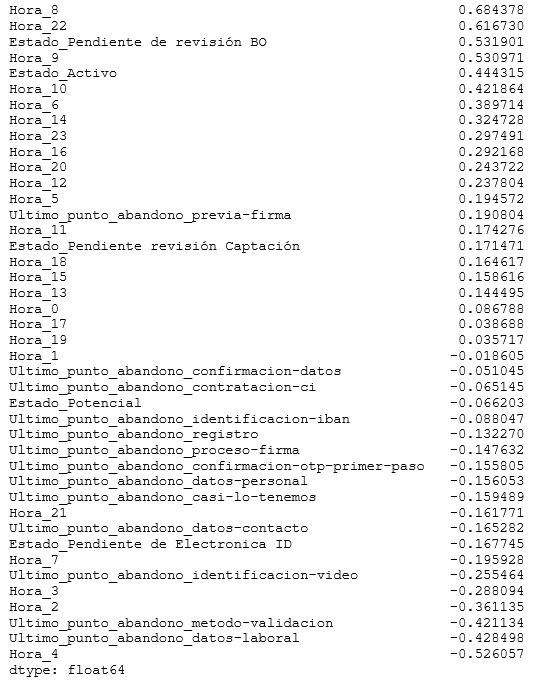


Se excluyen las variables no relevantes (ID\_cliente\_EVO, F\_Creacion, Numero\_Caso, Producto, Fecha\_Apertura, Hora\_Apertura, Minuto, Hora\_por\_15)

Las varibles categóricas incluidas se transformaron y agruparon en un formato numérico, ya que nuestro modelo de aprendizaje automático solo puede funcionar con datos númericos.

1. Prueba y entrenamiento del modelo:

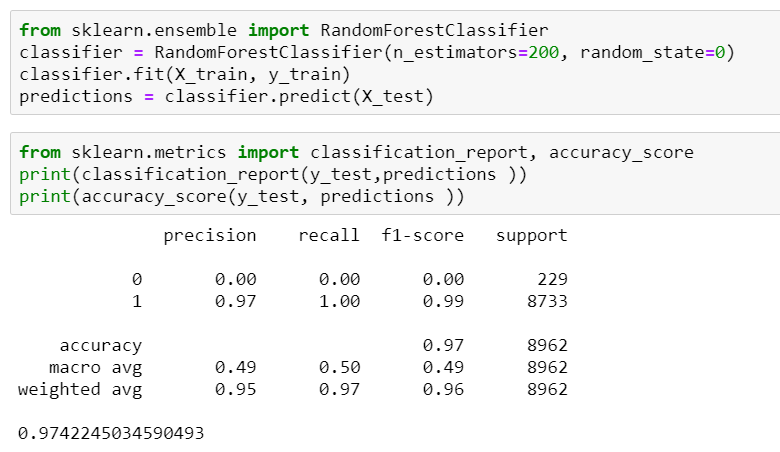
Regresión logística



Se obtiene un modelo con alta precisión (Accuracy 0,97)

En la regresión logística las variables con valor positivo se relacionan con el evento (“abandono”). Observamos que las que mejor predicen el abandono son haber iniciado el proceso web de contratación a las 8, 9, 22 horas y encontrarse en “estado pendiente de revisión”. Por el contrario, los procesos finales de la contratación web como la “validación e introducción de los datos laborables”, predicen que es poco probable que estos clientes abandonen.

Se realiza un análisis random forest, comprobando un similar grado de precisión



# CONCLUSIONES

1. El análisis de los datos generados a partir del registro de clientes en portales de banca online es útil para conocer los canales con mayor rentabilidad, además de los días y franjas horarias con mayor conversión. En este caso se observa que los canales con mejor conversión son EVO, Kelisto , SEM, - y Lleida.
2. Los canales con peor resultado son Rastreator\_Aff, Nurturing\_SMS y Direct Access.
3. En los meses analizados el mayor pico de conversión es a principios de mes, excepto Enero que presenta un comportamiento atípico debido a estrategias estacionales de marketing.
4. La franja horaria de mayor conversión es la comprendida entre las 11 a 1 de la mañana, debido al departamento de abandono que se encarga de volver a contactar con el cliente ya sea por email o SMS. Y como último recurso, se contacta con el cliente por el SmartCenter.
5. En el modelo predictivo las variables que mejor predicen el abandono son iniciar el proceso de contratación web 8:00, 9:00, 22:00 horas y el “estado pendiente de revisión” con alta precisión. Deberían diseñarse estrategias de marketing específicas para estas situaciones.