

# Modelo de Deserción en Clientes

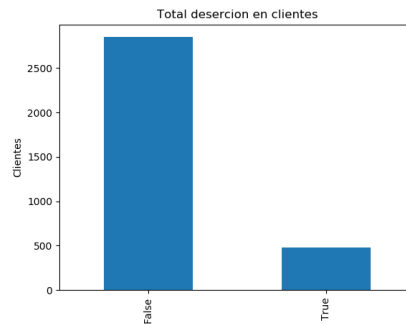
Ricardo Arango - Big Bang Data

March 20, 2019

La fuga de clientes se define como el *Movimiento de un cliente de un proveedor de servicio a otro*, mientras que la gestión de fuga de clientes describe el proceso por el cual el operador del servicio intenta evitar la fuga del cliente. En consecuencia, la utilización de modelos predictivos que permitan **identificar clientes en situación de riesgo de fuga**, permite al operador dirigir esfuerzos en prevenir la salida del cliente.

## 1 Los Datos

La variable *target* corresponde a la variable objetivo que se quiere predecir. En este caso, el abandono voluntario de clientes, esta variable es de carácter booleano, **FALSE: cliente permanece con sus servicios ó TRUE: clientes los abandona voluntariamente.**



El  $Odd_{desercion} = 0.17$ , es decir: por cada cliente que abandona los servicios, 6 clientes no lo hacen.

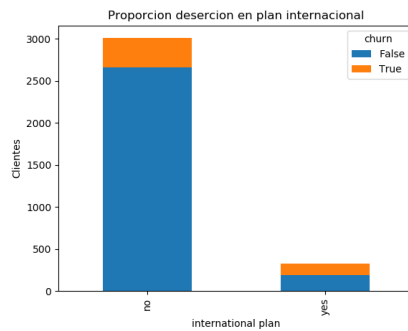
$$Odd_{desercion} \approx \frac{1}{6}$$

La meta es reducir dicho número lo máximo posible aumentando, considerablemente, la cantidad de usuarios que permanecen con sus servicios.

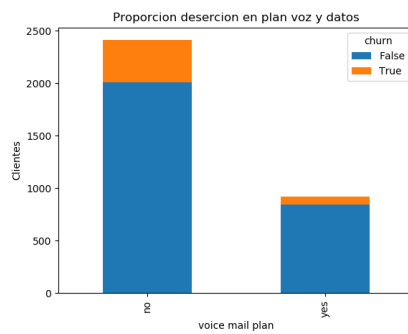
A continuación enumeraremos los rasgos más característicos presentes (históricamente) en los clientes que deciden abandonar los servicios. Esto con el fin de identificar posibles alarmas y evitar la fuga de cli

La medida del *ratio* entre las variables de deserción y plan internacional es:

$$ratio_{desercion,p-internacional} = 5.67$$



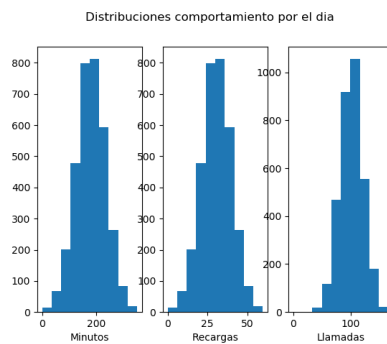
De dicho número podemos inferir que la relación entre estas dos variables es fuerte, es decir: **es muy probable que alguien con un plan internacional deserte.**



La medida del *ratio* entre las variables de deserción y plan de datos es:

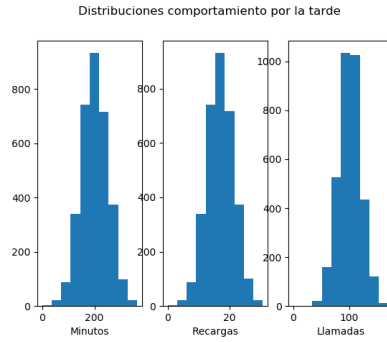
$$ratio_{desercion,p-datos} = 0.47$$

De dicho número podemos inferir que la relación entre estas dos variables es fuerte, es decir: **es muy probable que alguien con un plan de datos no deserte.**



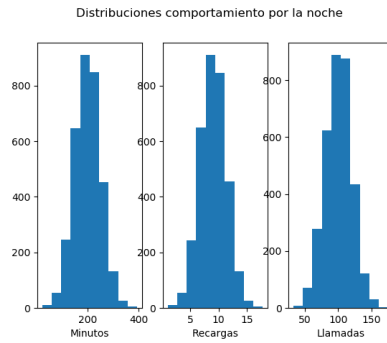
De estas variables podemos inferir que el comportamiento **promedio** de un desertor en las horas de la mañana es:

$$\text{minutos} = 206 \quad \text{recargas} = 35 \quad \text{llamadas} = 101$$



De estas variables podemos inferir que el comportamiento **promedio** de un desertor en las horas de la tarde es:

$$\text{minutos} = 212 \quad \text{recargas} = 18 \quad \text{llamadas} = 100$$



De estas variables podemos inferir que el comportamiento **promedio** de un desertor en las horas de la noche es:

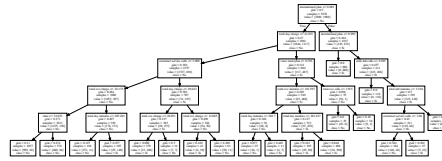
$$\text{minutos} = 205 \quad \text{recargas} = 9 \quad \text{llamadas} = 100$$

## 2 El Modelo

El problema de fuga de clientes ha sido objeto de una nutrida área de investigación, tanto desde el punto de vista del marketing, enfocándose en el comportamiento del consumidor y los componentes emocionales que la gatillan y desde el punto de vista de la minería de datos, enfocándose en la aplicación de algoritmos de machine learning para predecir el abandono de clientes. Del mismo modo, se han comparado diversos algoritmos para elaborar un puntaje de "propensión a la fuga" de clientes de telefonía celular. Los árboles de decisión y las regresiones logísticas son los algoritmos que ofrecen la mejor precisión. La desventaja de este último radica en que no es posible observar la cadena de condiciones que se deben cumplir para que un cliente tenga el potencial de abandonar el servicio. Dicho de otra forma, se trata de modelos de caja negra, en los cuales se hace difícil entender qué es lo que gatilla o motiva una fuga.

### 2.1 Árboles de Decisión

Los árboles de decisión se definen como un procedimiento recursivo, en el cual un número 'N' de registros se dividen progresivamente en grupos, de acuerdo a una regla de división que permita maximizar la homogeneidad o pureza de la variable *target*.



Para clasificar una instancia desconocida se recorre el árbol de arriba hacia abajo de acuerdo a los valores de los atributos probados en cada nodo y cuando se llega a una hoja, la instancia se clasifica con la clase indicada por esa hoja.

De forma general, el 93.1% de los datos el modelo los clasificará correctamente.

$$accuracy = 0.9310$$

además, el modelo tiene una precisión del 87% al momento de predecir a un usuario como desertor.

$$precision = 0.87$$