
Modelo Logístico con Datos Funcionales Aplicado al Churn en un Operador Movil.

Binomial Logistic Model with Functional Data Applied to Churn in a Virtual Mobile Operator that Offers Pre-Paid Telecommunications Services.

Maria Fernanda Vargas^a
mariavargas@usantotomas.edu.co

Wilmer Pineda^b
wilmerpineda@usantotomas.edu.co

Resumen

En el presente documento se pretende ilustrar el uso de datos funcionales multivariados con el análisis descriptivo, la implementación cluster jerárquicos y el modelo logístico funcional múltiple aplicado a la predicción de Churn de una compañía de telecomunicaciones. Los datos que se utilizaron para el desarrollo del trabajo son reales, tienen un histórico de 12 semanas y contemplan la información por usuario de antigüedad en la compañía, días de inactividad, uso de datos, uso de voz, recargas y la variable estado que indica si el usuario se encuentra activo o en estado churn.

Palabras clave: Datos funcionales multivariados, Cluster Jerarquicos, Modelo Logistico Funcional, Churn..

Abstract

This document is intended to illustrate the use of multivariate functional data with descriptive analysis, hierarchical cluster implementation and the multiple functional logistic model applied to the Churn prediction of a telecommunications company. The data that were used for the development of the work are real, have a history of 12 weeks and contemplate the information per user of seniority in the company, days of inactivity, data use, voice use, recharges and the state variable that indicates if the user is active or in churn state.

Keywords: Functional Data Multivariate, Cluster, Functional Logistic Model, Churn..

^aEstudiante. Facultad de Estadística, Universidad Santo Tomás

^bDocente. Facultad de Estadística, Universidad Santo Tomás

1. Introducción

En Colombia el diseño de la política sectorial de telecomunicaciones ha estado enfocada en aumentar la cobertura de los servicios de telecomunicaciones a nivel nacional, generando un clima de inversión para que los empresarios se motiven a invertir ampliando la infraestructura y asimismo, a diversificar y modernizar la oferta de servicios. Con esta política el gobierno nacional busca impulsar el desarrollo económico y social del país y a su vez a desarrollar acciones enfocadas a generalizar el acceso de todos los habitantes a servicios básicos de telecomunicaciones, garantizando calidad y eficiencia en su prestación a todos los habitantes del territorio nacional.(TIC, 2001)

Al posicionarse como un sector competitivo y en constante crecimiento las empresas han pasado de tener como objetivo principal la obtención de nuevos clientes a enfocarse en la retención de usuarios y en prevenir el fenómeno de abandono de servicio, debido a que con esto conduce a una mejora en ventas y una reducción de costos de marketing en comparación con los nuevos clientes. Estos hechos han dado lugar a que algunos indicadores relacionados tengan mayor relevancia con el paso del tiempo, entre los cuales esta el Churn Rate o Tasa de Cancelación de clientes, la cual es una métrica que mide el número de clientes y suscriptores que abandonaron la una compañía (o han comenzado a seguirla) en un período de tiempo.

Para poder reducir la tasa de cancelación de clientes, es necesario tener en cuenta varios aspectos a la hora de realizar acciones de mercadeo, para esto la predicción de abandonos de clientes es una parte indispensable en el proceso de toma de decisiones estratégicas y planificación del sector de la telecomunicación.

En las últimas décadas, las organizaciones se centran cada vez más en las relaciones a largo plazo con sus clientes y observan el comportamiento de un cliente en determinados momentos. Sin embargo es indispensable hacer un monitoreo continuo de su comportamiento con el fin de generar alertas de indicios de abandono del servicio, para esto se han implementado técnicas de descubrimiento de conocimiento aplicado en técnicas de base de datos para extraer relaciones ocultas entre diferentes entidades y atributos en una avalancha de bancos de datos. Otros métodos usados han sido las predicciones elaboradas con métodos estadísticos.

Entre los modelos más usados esta la regresión lineal, arboles de decisión, maquinas de soporte vectorial, entre otros modelos que permiten obtener la probabilidad de que un usuario deje de usar el servicio, para su construcción se han postulado variables medidas en un solo momento, por ejemplo; cantidad de llamadas en determinado mes, cantidad de MB usadas en el mes, valor de recargas acompañado de otras variables estructurales como la edad, ciudad, genero, etc, las cuales no permiten ver el comportamiento a lo largo del tiempo de un cliente, lo cual no genera un entendimiento óptimo de como fueron las últimas semanas de consumo del usuario en la compañía, por tal razón el presente trabajo pretende usar variables de tipo funcional, de tal manera que se midan las variables con su respectiva historia, es decir si se decide utilizar la variable "Cantidad de llamadas" se mida en doce semanas y se tenga en cuenta si su comportamiento es constante, creciente o decreciente.

2. Sector de Telecomunicaciones

A partir del año 1998 el sector ha venido presentando importantes avances en materia de comunicaciones en pro de incrementar la productividad y la competitividad de las industrias nacionales, modernizar las instituciones gubernamentales y socializar el acceso a la información. Según el plan estratégico de tecnologías de Información y comunicación sectorial del año 2017-2020, el objetivo principal es que al año 2019 todos los ciudadanos puedan acceder a los servicios de telecomunicaciones y puedan insertarse en la sociedad del conocimiento. El gobierno continúa con el mismo enfoque de invertir en la ampliación de la cobertura en infraestructura (internet o telecentros comunitarios), ampliación de las redes de radio, televisión, telefonía fija e inalámbrica (hospitales, guarniciones, alcaldías, poblaciones rurales e indígenas)

y continuar la adaptación de centros de acceso a telefonía para personas con discapacidad auditiva y de centros de acceso a internet para invidentes.(TIC, 2001)

Para lograr alcanzar la inclusión digital y masificar la banda ancha existen tres estrategias fundamentales:

- Reducción de costos en acceso a internet y a telefonía móvil para promover uso en el sector publico y privado.
- Uso pleno de la capacidad instalada de las redes de telecomunicaciones de fibra óptica y de operadores de televisión por cable, así como la reventa de servicios y la promoción de tecnologías inalámbricas de tercera generación.
- Incrementar el uso de las tecnologías de la Información a partir del fortalecimiento de los dos ejes anteriores.

Así mismo, en el desarrollo de Planes de Masificación de servicios de telecomunicaciones se contempla la posibilidad de otorgar subsidios a grupos específicos de población de menores ingresos o en condiciones socioeconómicas y geográficas menos favorables, para el suministro de los servicios de telecomunicaciones, los equipos terminales, los paneles solares, las aplicaciones y los servicios de capacitación para la apropiación de dicha tecnología. Dentro de este programa de masificación, en poblaciones con menos de 50.000 habitantes, el Fontic también podrá subsidiar la prestación del servicio de acceso a internet a nuevos usuarios que reúnan las mismas características, para lo cual los operadores de televisión comunitaria podrán prestar servicios de Internet y telefonía sin afectar su licencia de televisión, previo cumplimiento de inscripción en el Registro de TIC.

Como resultado a la política sectorial del país y al plan estratégico de tecnologías de Información y comunicación sectorial elaborado cada año, Colombia se encuentra bajo uno de los fenómenos más característicos de la evolución del sector de las telecomunicaciones de los últimos años, lo que se podría denominar como “Expansión del Sector”. Detrás de este concepto se encuentra la mayoría de las causas que están motivando los cambios recientes, y originando que en la actualidad el sector constituya uno de los soportes básicos de la sociedad futura. El concepto de expansión del sector surge por un proceso de convergencia de múltiples factores económicos, tecnológicos, institucionales y sociales que, sin embargo, se pueden concretar en siete fuerzas básicas:

- Amenaza de nuevos competidores.
- Poder negociador de los usuarios.
- Poder negociador de la industria de equipos y los proveedores de servicios.
- Amenaza de productos o servicios alternativos.
- Grado de diversidad y competencia interna.
- Presión reguladora.
- Características globales.

Con estas fuerzas se puede resumir un cambio drástico en el mercado con aumento en la oferta de servicios y las intenciones adquisitivas de los consumidores, lo cual crea una gran disputa de usuarios en las empresas de telecomunicaciones, debido a que el sector esta obligado por la Comisión de Regulación de Comunicaciones (CRC) a ofrecer libertad de migrar de una empresa a otra. Según la CRC la libre elección de servicio, de plan y de operador es un derecho que tienen los usuarios, esta elección es importante para que coincidan las necesidades con las decisiones de consumo, sobre el operador que presta el servicio y el

valor a pagar. Cabe resaltar que la suspensión o cambio de contrato del servicio de telecomunicaciones se puede suspender en cualquier momento sin pagar penalidad, multa o sanción alguna.

Informe de las TICs para el cuarto trimestre del año 2009

De acuerdo con el reporte realizado por los proveedores de redes y servicios de telecomunicaciones, si se tiene en cuenta el total de suscripciones fijas y móviles de acceso a Internet, se observa que a Diciembre de 2009 Colombia alcanzó una penetración de suscriptores de este servicio equivalente al 7,07 %, lo cual se evidencia en la Tabla 1 en donde se presenta el número de abonados que tienen las empresas del sector y su respectiva participación para el cuarto trimestre de 2009, relacionando el tráfico asociado por demanda (prepago y postpago) y por suscripción para cada proveedor. Adicionalmente, en la Tabla 2 se muestra el comportamiento para el cuarto trimestre del 2018.

Tabla 1: Participación por Proveedor en el Sector para el Año 2009. Fuente: Informe TIC 2009

PROVEEDOR	ABONADOS EN SERVICIO	PREPAGO	POSTPAGO	PARTICIPACIÓN EN EL MERCADO %
Avantel S.A.S.	15.911			2 %
Colombia Móvil	306.256	456.914	171.001	33 %
Comcel	339.285	1.167.061	1.023.644	37 %
Telefónica Móviles	253.828	382.33	168.803	28 %
TOTAL NACIONAL	915.28	839.244	339.804	100 %

Tabla 2: Participación por Proveedor en el Sector para el Año 2018. Fuente: Informe TIC 2018

PROVEEDOR	ABONADOS EN SERVICIO	PREPAGO	POSTPAGO	LINEAS RETIRADAS	PARTICIPACIÓN %
ALMACENES EXITO INVERSIONES S.A.S.	1.350.267	1.350.267	0	279.259	2 %
AVANTEL S.A.S.	2.221.618	1.580.942	640.676	227.919	3 %
COLOMBIA MOVIL S.A. E.S.P.	11.812.492	10.160.441	1.652.051	2.322.019	18 %
COLOMBIA TELECOMUNICACIONES S.A. ESP	15.716.216	11.880.734	3.835.482	1.541.173	24 %
COMUNICACION CELULAR S A COMCEL S A	29.680.853	22.744.160	6.936.693	4.661.702	46 %
EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP.	486.397	287.046	199.351	142.032	1 %
LOGISTICA FLASH COLOMBIA S.A.S	555.937	555.937	0	67.567	1 %
SUMA MOVIL SAS	2.460	2.460	0	23	0 %
UFF MOVIL SAS					0 %
UNE EPM TELECOMUNICACIONES S.A.					0 %
VIRGIN MOBILE COLOMBIA S.A.S.	2.687.737	2.687.737	0	366.148	4 %
TOTAL NACIONAL	64.513.977	51.249.724	13.264.253	9.607.842	100 %

2.1. Churn

Cuando se habla de un cliente en estado churn, se hace referencia a un usuario que lleva de forma consecutiva 90 días de inactividad, es decir, 90 días en los cuales el cliente con su línea no ha tenido ningún tipo de tráfico (llamadas salientes, llamadas entrantes, llamadas a buzón, mensajes de texto, uso de datos), por lo cual al cumplir estos 90 días sin actividad la compañía cataloga a este cliente como churn.

Las posibles categorías del estado del cliente son las siguientes:

- Activo: El usuario ha tenido actividad del en las últimas 24 horas.
- Inactivo: El usuario ha tenido actividad en los últimos 90 días, pero no en el último día.
- Churn: A partir del día 91 de inactividad el usuario se clasifica como churn, este estado no tiene implicaciones en la plataforma por lo cual el usuario en cualquier momento puede usar la línea y cambiar de estado.



Figura 1: Transición de estado. Fuente: Elaboración propia.

Razones de abandono o cancelación de servicio:

- Voluntario: Aquellos clientes que quieren dejar de usar el servicio y pasar a otro proveedor.
- No voluntario: Se presenta cuando una empresa interrumpe el servicio a un cliente.
- Silencioso: Aquellos clientes que dejan de usar el servicio sin previo aviso a la compañía.

3. Análisis de datos funcionales

3.1. Datos Funcionales

El análisis de datos funcionales es una rama de las matemáticas y más concretamente de la estadística, que estudia y analiza la información contenida en curvas, superficies o cualquier elemento que varíe sobre un espacio continuo.

El conjunto de datos más simple encontrado en FDA es de la forma:

$$x_n(t_{j,n}) \in R, t_{j,n} \in [T_1, T_2], n = 1, 2, \dots, N, j = 1, 2, \dots, J_n \quad (1)$$

Con lo cual se asume que se observan N curvas en un intervalo común $[T_1, T_2]$.

3.2. Espacio $L^2[T]$

Sea $L^2[T]$, con $T = [a, b] \in R$ el espacio de las funciones cuadrado integrables (Espacio de Hilbert)

$$L^2[T] = f : R : R \mid \int_a^b f(t)^2 dt \quad (2)$$

Con producto interno

$$\langle a, b \rangle = \int_a^b f(t)g(t) dt \quad (3)$$

3.3. Base

Una base es un conjunto de funciones conocidas e independientes $\{\phi_k\}_k \in N$ tales que cualquier función puede ser aproximada, también como se quiera, mediante una combinación lineal de K de ellas con K suficientemente grande. De esta forma, la observación funcional puede aproximarse como:

$$X(t) \approx \sum_{k=1}^K C_k \phi_k(t). \quad (4)$$

3.4. La base B-spline para funciones spline

Es necesario especificar un sistema de funciones básicas (t) , y estas tendrán las siguientes propiedades esenciales:

- Cada función base $\phi_k(t)$ es en sí misma una función de spline definida por un orden m y una secuencia de nodos τ .
- Dado que un múltiplo de una función de spline es una función de spline, y dado que las sumas y las diferencias de splines también son splines, cualquier combinación lineal de estas funciones básicas es una función de spline.
- Cualquier función de spline definida por m y τ puede expresarse como una combinación lineal de estas funciones básicas.

Este comportamiento de límite de las funciones básicas de B-spline se logra colocando, de hecho, m nodos en los límites. Es decir, cuando B-splines se calculan realmente, la secuencia de nodos τ se extiende en cada extremo para agregar una réplica $m - 1$ adicional del valor del nodo límite. Como notamos antes, hay algunas aplicaciones donde no queremos derivadas continuas $m - 2$ en ciertos puntos fijos en el interior del intervalo. Esto se puede acomodar fácilmente mediante B-splines. Colocamos un nodo en dichos puntos fijos, y luego, para cada reducción en la diferenciabilidad, también se coloca un nodo adicional en esa ubicación. Por ejemplo, si estuviéramos trabajando con el orden cuatro splines, y quisiéramos que la derivada pudiera cambiar abruptamente a un cierto valor de t pero aún quisiéramos que la función ajustada fuera continua, colocaríamos tres nodos en ese valor.

La notación $B_k(t, \tau)$ se usa a menudo para indicar el valor en t de la función de base B-spline definida por la secuencia de punto de interrupción τ . Aquí k se refiere al número del mayor nodo de valor t . Los nodos $m - 1$ agregados al punto de ruptura inicial también se cuentan en este esquema, y esto es consistente con el hecho de que las primeras funciones básicas de m B-spline tienen apoyos, todos comenzando en el límite izquierdo. Esta notación nos da funciones básicas $m + L - 1$, como se requiere en el caso habitual donde todos los nudos interiores son discretos. De acuerdo con esta notación, una función de spline $S(t)$ con nodos interiores discretos se define como

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau) \quad (5)$$

(Ramsay Silverman, Functional Data Analysis, 2005)

3.5. Modelo de regresión logística funcional

Formulación modelo

El objetivo del modelo FLR es explicar una variable de respuesta binaria en términos de variables explicativas de tiempo continuo (predictor funcional).

De tal manera se considera una magnitud cuya evolución en el tiempo se modela como un proceso estocástico de segundo orden $\{X(t), t \in T\}$, y $x_1(t), \dots, x_n(t)$ una muestra de observaciones de tal proceso que se supone que tienen funciones integrables cuadradas ($\int_T x_i^2 dt < \infty$).

Entonces el modelo se formula como:

$$y_i = \pi_i + \xi_i, i = 1, \dots, n \quad (6)$$

donde

$$y_i \sim Ber(\pi) \quad (7)$$

y la probabilidad $\pi_i = P\{Y = 1 | X(t) = x_i(t)\}$:

$$\ln\left(\frac{\pi}{1-\pi}\right) = Z_i^t \gamma + \sum_{j=i}^p x_{ij}(t) \beta_j(t) dt \quad (8)$$

Con Z como una variable escalar y con γ como su parámetro, $\beta(t)$ una función de parámetros de las variables funcionales x_{ij} .

4. Metodología

En primera instancia es importante conocer el comportamiento de los datos, para esto se utilizan la media y la desviación estándar, pero antes de implementar estos estadísticos se realiza una segmentación de los usuarios activos, con el fin de agrupar perfiles de usuarios similares entre si y controlar la varianza.

Para realizar el proceso de segmentación se usa como referencia el método de cluster jerárquicos implementado en el trabajo de Venagas (2017) "Método de Cluster Jerárquicos para Datos Funcionales Multivariados: Una Aplicación en Mercadeo"

Proceso Análisis Clúster Jerárquico

Paso 1: Selección de las variables. Se recomienda que las variables sean del mismo tipo.

Las variables que se usaran son las que resumen el uso del usuario:

- Cantidad de megas usadas
- Cantidad de llamadas salientes
- Valor de recargas

Paso 2: Detección de valores atípicos El análisis clúster es muy sensible a la presencia de objetos muy diferentes del resto (valores atípicos). Con el fin de poder detectar los datos atípicos de las variables se plantea el diagrama de caja HDR que resume la descomposición del componente principal funcional.

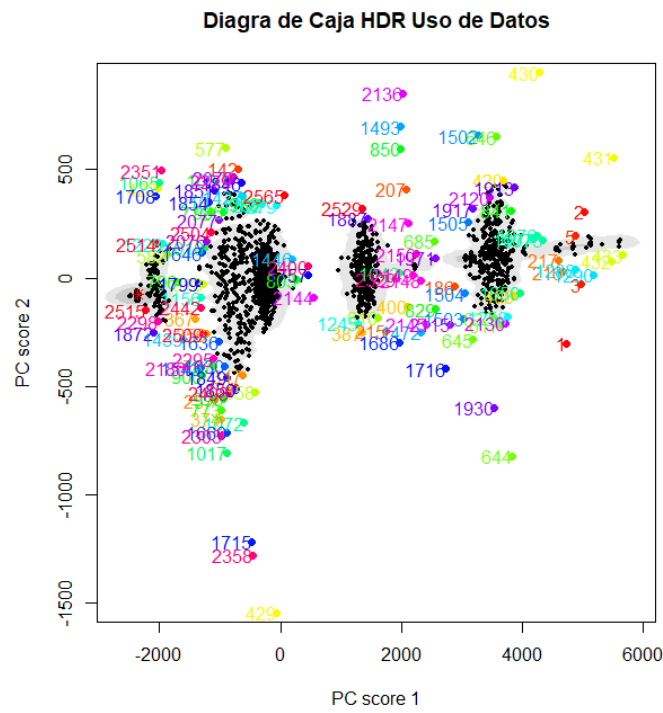


Figura 2: Diagrama de caja HDR uso de datos. Fuente: Elaboración propia.

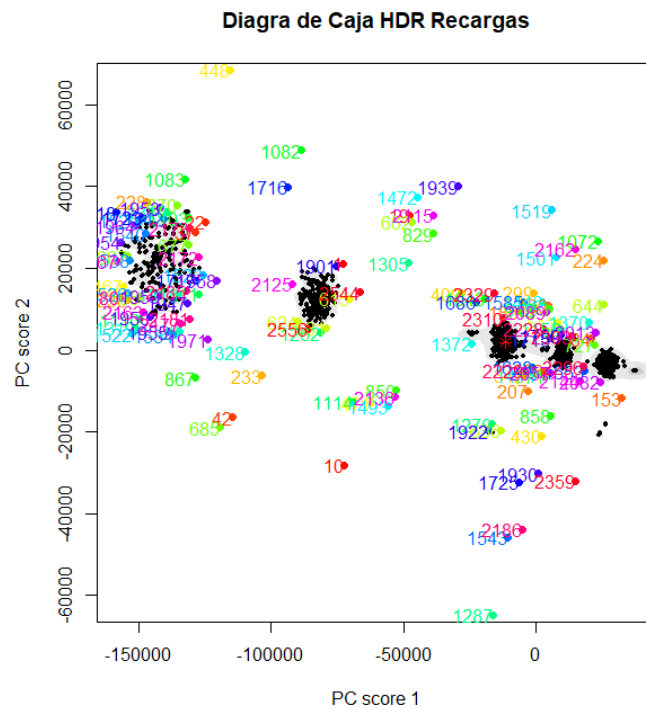


Figura 3: Diagrama de caja HDR recargas. Fuente: Elaboración propia.

Se logra observar que a pesar de que existe concentración en algunos intervalos de las variables, también existe una cantidad considerable de datos atípicos, lo cual es normal debido a la naturaleza de las variables y del sector.

Paso 3: Elección de una medida de similitud. entre objetos y obtención de la matriz de distancias. Mediante estas medidas se determinan los clusters iniciales.

Sea $f(t) = \sum_{i=1}^k a_i \phi_i(t)$ y $g(t) = \sum_{i=1}^k b_i \phi_i(t)$ datos funcionales en $t \in [a, b]$, en un intervalo, entonces la distancia entre dos funciones $D^2(f, g) = \sum_i (a_i - b_i)^2$

Para cada variable funcional se construye una matriz de similitud denotada como S_i

Paso 4: Buscar los clusters más similares. Apartir de la matriz de similitud S se aplica el algoritmo para métodos de aglomeración:

Utilizamos el método de Ward para contruir el dendograma,

$$W = \sum_g \sum_{i \in G} (X_{ig} - \hat{X}_g)' (X_{ig} - \hat{X}_g) \tag{9}$$

Se proponen 4 cluster obteniendo el siguiente dendograma:

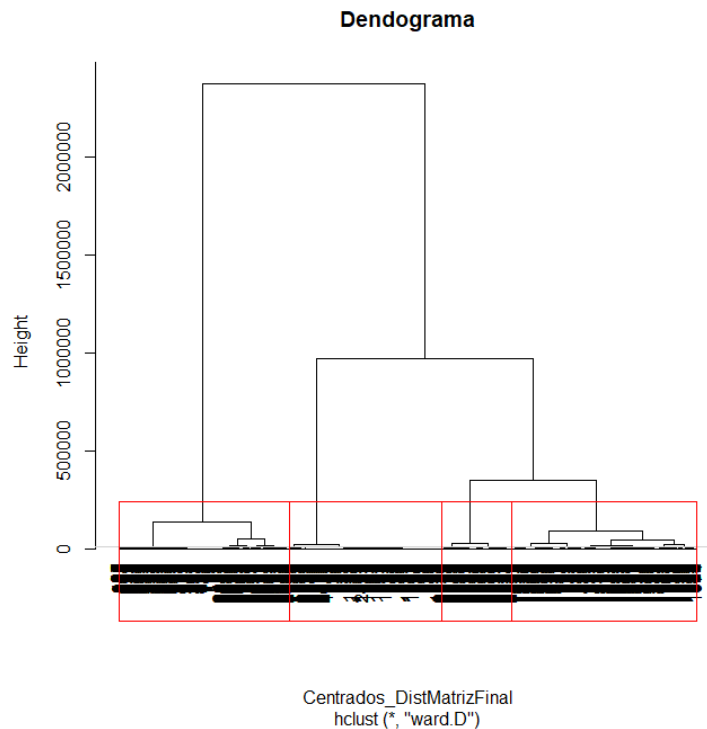


Figura 4: Dendograma. Fuente: Elaboración propia.

El proceso de agrupación jerárquico se puede resumir gráficamente mediante una representación gráfica en forma de árbol que recibe el nombre de Dendograma. Los objetos similares se enlazan y su posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los objetos.

Procedemos a graficar los cluster, para poder tener una noción más exacta de su comportamiento.

Se logra ver que el cluster 1 acoge a varios individuos que se encuentran alejados de su concentración. Se realiza el análisis de estos individuos en particular para poder conocer en cuál de las tres variables son datos atípicos.

Individuos: 2,3,9,6,4,16,1,21,29,22,15,17,2396,8,7

Teniendo los cluster construidos se procede a generar la media y la desviación estándar de cada uno para poder describirlos.

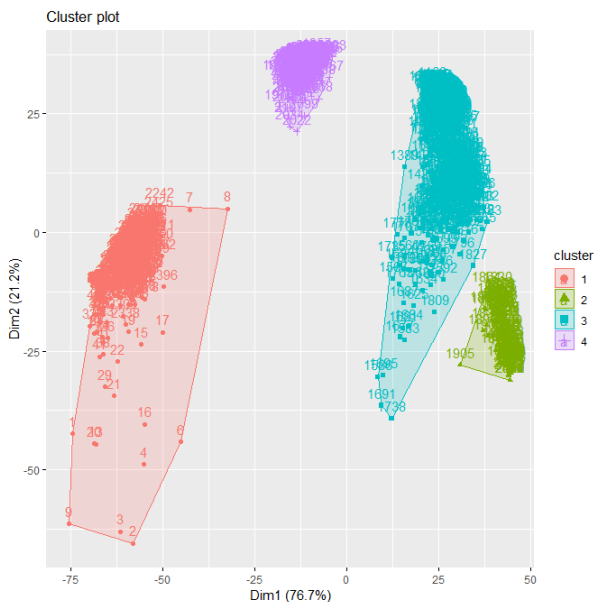


Figura 5: Cluster plot. Fuente: Elaboración propia.

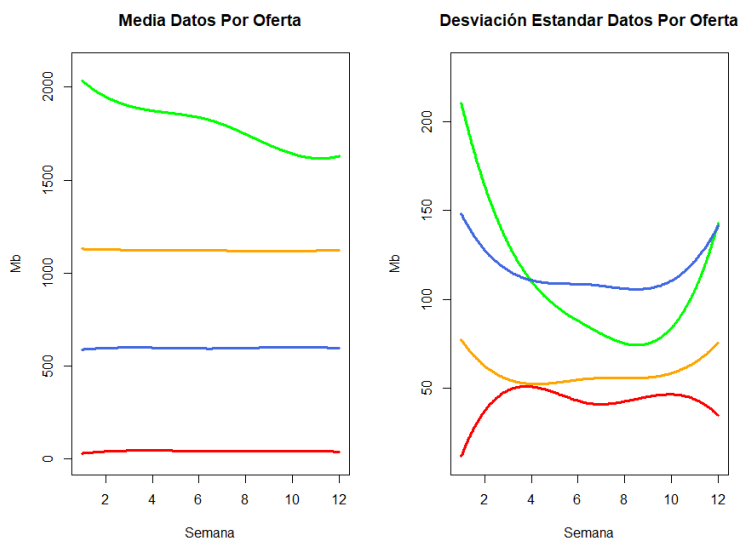


Figura 6: Media datos por oferta / Desviación estándar datos por oferta. Fuente: Elaboración propia.

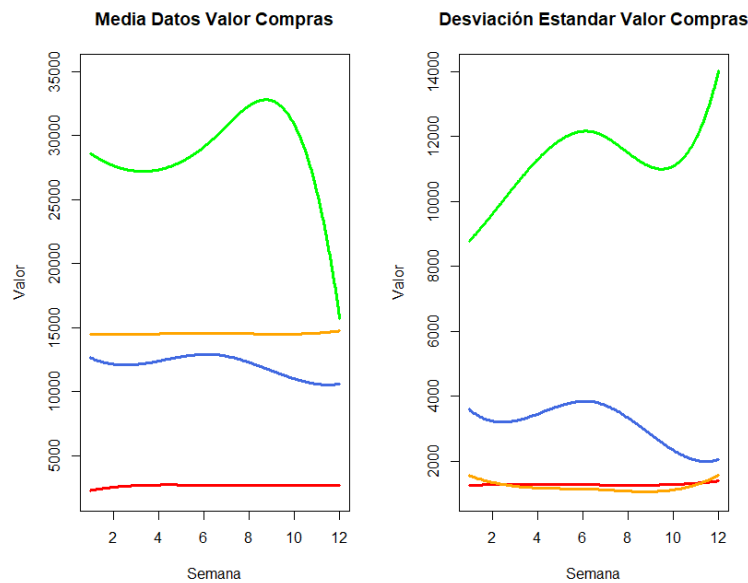


Figura 7: Media datos valor compras / Desviación estándar valor compras. Fuente: Elaboración propia.

Cluster 1: Usuarios de alto valor:

Usuarios que usan más de 3200 mb por semana, realizan más de 22 llamadas y su promedio de recarga esta sobre los \$45.000.

Cluster 2: Usuarios de medio valor:

Usuarios que usan de 2300 a 2600 mb, en promedio por semana hacen, 17 llamadas y su promedio de recarga pasó de estar en \$25.000 en la semana 10 a \$12.000 en la última semana.

Cluster 3: Usuarios de valor medio - bajo:

Usuarios que usan en promedio 1600 mb por semana, realizan entre 5 y 7 llamadas por semana y su promedio de recarga ha sido constante en el tiempo con \$15.000.

Cluster 4: Usuarios de bajo valor:

Usuarios que usan de 400 a 600 mb por semana, hacen entre 3 y 6 llamadas, y su promedio de recarga es constante en el tiempo con \$8.0000.

Cluster 5: Usuarios de bajo valor:

Usuarios que no generan uso de datos, sin embargo, uso de voz es elevado debido a que realizan más de 10 llamadas en la semana, su promedio de recarga es constante y es de \$6000.

Modelo de regresión logística funcional múltiple

Sea Y una variable dependiente binaria con dos posibles valores: 0 y 1 (activo o churn) y un conjunto de variables independientes, (x_1, x_2, \dots, x_n) , observadas con el fin de predecir o explicar el valor de Y .

La variable que se quiere modelar es el churn. Se plantea como variable respuesta del modelo, de tal modo que:

$$Y_i = \begin{cases} 0 & \text{si el usuario esta en estado Activo} \\ 1 & \text{si el usuario esta en estado Churn} \end{cases}$$

Vemos y_i como la realización de una variable aleatoria Y_i que puede tomar los valores uno y cero con probabilidades π y $1 - \pi$, respectivamente. La distribución de Y_i se llama distribución de Bernoulli con el parámetro π

Las variables significativas son:

- Antigüedad del usuario en la compañía.
- Cantidad de megas usadas.
- Cantidad de llamadas salientes.

Los β de las variables funcionales se comportan de la siguiente manera:

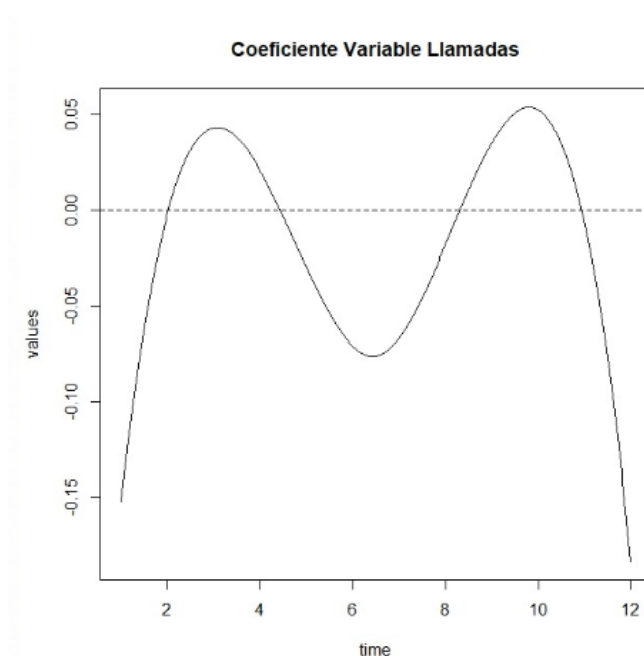


Figura 8: Coeficiente variable llamadas. Fuente: Elaboración propia.

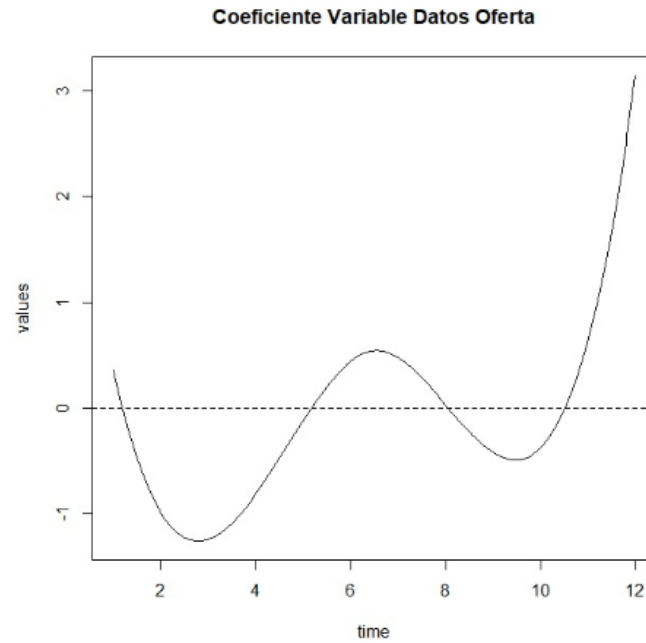


Figura 9: Coeficiente variable datos oferta. Fuente: Elaboración propia.

Tabla 3: Coeficientes Variable Llamadas. Fuente: Elaboración propia

Semana	1	2	3	4	5	6	7	8	9	10	11	12
Beta	-0,15	-0,025	0,04	0,025	-0,025	-0,075	-0,075	-0,01	0,025	0,06	0,025	0,2
Exp(Beta)	0,86	0,77	1,04	1,28	0,77	0,47	0,47	0,99	0,97	0,94	0,97	0,81

Tabla 4: Coeficientes Variable Datos. Fuente: Elaboración propia

Semana	1	2	3	4	5	6	7	8	9	10	11	12
Beta	-0,01	-1,1	-1,3	-0,8	-0,3	0,5	0,6	0,01	-0,04	-0,03	0,02	3
Exp(Beta)	0,99	0,33	0,27	0,44	0,74	1,64	1,82	1,01	0,96	0,97	1,02	20

$$Sensibilidad = \frac{VP}{P} = \frac{1894}{2000} = 0,947 \tag{10}$$

$$Especificidad = \frac{VN}{N} = \frac{1967}{2000} = 0,983 \tag{11}$$

$$Exactitud = \frac{VP}{P} = \frac{1894 + 1967}{4000} = 0,965 \tag{12}$$

Tabla 5: Matriz de confusión. Fuente: Elaboración propia.

Observaciones / Predicciones	Activo	Churn
Activo	1967	33
Churn	106	1894

Referencias

- Bande, M. F. (2008). A PRESENT OVERVIEW ON FUNCTIONAL DATA ANALYSIS. Universidad de Santiago de Compostela, 6.
- Ferraty, F., Vieu, P. (2006). Nonparametric Functional Data Analysis. STAPH.
- Franke, R. (s.f.). Smooth Interpolation of Scattered Data by Local Thin Plate Splines. Gran Bretaña: Computer and Mathematics with Applications.
- Ramsay, J., Silverman, B. (. (2005). Functional Data Analysis. Springer.
- Ramsay, J., Hooker, G., Graves, S. (2009). Functional Data Analysis with R and MATLAB. London: Springer.
- TIC. (2001). El sector de las telecomunicaciones. Bogotá: CRT.
- Bande, M. (s.f.). Functional Data Analysis. Artículos de Estadística.
, M. S. (2011). Predicción de picos de ponen mediante regresión logística funcional. Universidad de Granada.
- Ramirez, A., Garatejo, O., Pineda, W. (s.f.). Análisis multivariado de datos funcionales aplicado a curvas de encefalogramas. Repositorio USTA.
- Venegas, G. (2017). Método de Cluster Jerárquicos para Datos Funcionales Multivariados: Una Aplicación en Mercadeo. Repositorio USTA.
- Comisión de regulación de telecomunicaciones. (1998). El sector de las telecomunicaciones en Colombia 1998-2001. CRT.