



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DESARROLLO DE UN MODELO PREDICTIVO PARA DETECTAR CASOS
DE FRAUDE INTERNO EN UNA INSTITUCIÓN BANCARIA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DIEGO IGNACIO GARCÍA JURADO

PROFESOR GUÍA:
LUIS ABURTO LAFOURCADE

MIEMBROS DE LA COMISIÓN:
SEBASTIÁN RÍOS PÉREZ
VICENTE PEIROTÉN ÁLVAREZ

SANTIAGO DE CHILE
2016

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE: Ingeniero Civil Industrial

POR: Diego Ignacio García Jurado

Fecha: 11/11/2016

PROFESOR GUÍA: Luis Aburto Lafourcade

DESARROLLO DE UN MODELO PREDICTIVO PARA DETECTAR CASOS DE FRAUDE INTERNO EN UNA INSTITUCIÓN BANCARIA

El presente trabajo desarrolla metodologías de minería de datos, a partir de datos transaccionales de una institución bancaria nacional, generar un modelo predictivo que sea capaz de detectar sospechosos de la comisión de fraude interno. Es decir, fraude cometido por empleados de la misma institución.

El banco posee 11.723 empleados, cerca de 3,6 millones de clientes y más de 500 sucursales, lo que genera alrededor de 21 millones de transacciones diarias. Debido a que se cuenta exclusivamente con 5 registros de fraude, se opta por abordar el problema desde una óptica de modelos no supervisados, que permiten extraer conocimiento de los datos sin tener información a priori de ellos.

Se utilizan tres modelos para generar con cada uno una lista de sospechosos de haber cometido fraude interno, ya sea por semejanza con los fraudes, o por presentar un comportamiento que se desvía del comportamiento común del resto de los datos (outlier). Primero se utilizan medidas de distancia para encontrar los vecinos más cercanos a cada uno de los registros de fraude, luego se implementa el algoritmo Local Outlier Factor (LOF) que es capaz de identificar outliers a partir de la búsqueda de diferencias significativas entre la densidad de un dato y la de sus vecinos. Posteriormente se usa Análisis de Componentes Principales (PCA), que sirve para reducir la dimensionalidad de los datos generando combinaciones lineales de las variables, para ver la ubicación topológica de los registros de fraudes y seleccionar sospechosos que se encuentren en su entorno. Finalmente se consolida una lista con los sospechosos entregados por los tres criterios, por medio de majority voting, considerando exclusivamente los que son considerados sospechosos por al menos 2. Este conjunto de metodologías genera un modelo con un lift de 61,18.

Así se obtiene una lista de 74 usuarios sospechosos que presentan una media de transacciones diarias muy por debajo del resto de los datos, especialmente en las variables “consulta de datos básicos de la cuenta” y “localización de personas. Además de presentar un tiempo promedio entre transacciones inferior al resto de los datos.

AGRADECIMIENTOS

En éste, mi último paso en la vida estudiantil, la culminación de mi recorrido en esta Universidad, no puedo dejar de mencionar y agradecer a tantas personas que fueron claves en este caminar;

En primer lugar, a mis padres Patricio y María Eugenia, por el amor y la dedicación, por el apoyo incondicional entregado a lo largo de mi vida;

A mi hermana Daniela, mi amiga, confidente y Psicóloga, importante respaldo en la visión humanista de la vida que podemos compartir;

A mis amigos de los Colegios Patrocinio de San José y Antupirén, compañeros de esparcimiento, de maravillosos y alegres momentos que lograban sacarme del estrés cotidiano;

A Nacho, mi colega en la música, mi gran amigo, aún ahora en la distancia por su residencia en París;

A mis amigos de la Universidad, por todos los años compartidos, por su tiempo y colaboración en largas jornadas de estudio y otras tantas de diversión;

A mis profesores de la sección del Trabajo de Título, por la enseñanza y la ayuda prestada;

Especial gratitud a mi Profesor Guía, señor Luis Aburto, por su apoyo, su mirada experta y aliento cuando el camino se tornó difícil;

A mi Universidad y al Departamento de Industrias que fuera mi casa por todos estos años, por los conocimientos y la formación entregados.

A Everis por la cálida acogida, la confianza depositada en mí y el apoyo a lo largo del desarrollo de esta memoria;

A todos ellos mi gratitud, por forjar en gran parte mi futuro tanto en lo profesional como en lo humano.

TABLA DE CONTENIDO

1.	INTRODUCCIÓN.....	7
1.1.	ANTECEDENTES GENERALES.....	7
1.2.	DEFINICIÓN Y JUSTIFICACIÓN DEL PROYECTO	11
1.3.	OBJETIVOS	18
1.3.1.	GENERAL.....	18
1.3.2.	ESPECÍFICOS	18
1.4.	ALCANCES	18
1.5.	RESULTADOS ESPERADOS.....	19
2.	MARCO CONCEPTUAL	20
2.1.	DEFINICIÓN DE FRAUDE	20
2.2.	TIPOS DE FRAUDE	20
2.3.	DEFINICIÓN DE FRAUDE INTERNO.....	22
3.	MARCO TEÓRICO	23
3.1.	EL PROCESO KDD	23
3.1.1.	RECOLECCIÓN Y PREPARACIÓN DE LOS DATOS	25
3.1.2.	PRE-PROCESAMIENTO DE LOS DATOS.....	25
3.1.3.	TRANSFORMACIÓN DE LOS DATOS	26
3.1.4.	DATA MINING	27
3.1.4.1.	MODELOS SUPERVISADOS.....	28
3.1.4.2.	MODELOS NO SUPERVISADOS	29
3.1.4.2.1.	ANÁLISIS DE VECINOS MÁS CERCANOS	30
3.1.4.2.2.	ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)	31
3.1.4.2.3.	LOCAL OUTLIER FACTOR (LOF)	32
3.1.5.	ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	33
3.2.	HERRAMIENTAS TECNOLÓGICAS	34
4.	APLICACIÓN DE TÉCNICAS DE DATA MINING	35
4.1.	RECOLECCIÓN Y PREPARACIÓN DE LOS DATOS	35
4.1.1.	CARACTERIZACIÓN Y ESTRUCTURA DE LOS DATOS	35
4.1.2.	CARACTERIZACIÓN DE LOS REGISTROS DE FRAUDE	36
4.1.3.	SELECCIÓN DE LOS DATOS	38

4.2.	PRE-PROCESAMIENTO DE LOS DATOS.....	39
4.2.1.	LIMPIEZA DE LOS DATOS	39
4.3.	TRANSFORMACIÓN DE LOS DATOS	40
4.3.1.	GENERACIÓN Y SELECCIÓN DE VARIABLES	40
4.3.2.	ANÁLISIS PRELIMINAR DE LAS VARIABLES	42
4.3.3.	NORMALIZACIÓN DE VARIABLES	50
4.4.	DATA MINING	50
4.4.1.	ANÁLISIS DE VECINOS MÁS CERCANOS	51
4.4.2.	LOCAL OUTLIER FACTOR (LOF)	55
4.4.3.	ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)	61
4.5.	ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	62
5.	CONCLUSIONES Y TRABAJOS FUTUROS	66
6.	BIBLIOGRAFÍA.....	68
7.	ANEXOS.....	70
	ANEXO A: EVOLUCIÓN DE INDICADORES DE PROFUNDIDAD FINANCIERA Y COMPARACIÓN CON OTROS PAÍSES.....	70
	ANEXO B: CORRELACIÓN ENTRE INGRESO PER-CÁPITA Y PROFUNDIDAD BANCARIA	71
	ANEXO C: EVOLUCIÓN DEL INGRESO PER-CÁPITA EN CHILE.....	72
	ANEXO D: EVOLUCIÓN DE LA PENETRACIÓN DE TARJETAS DE CRÉDITO, DÉBITO Y CUENTAS CORRIENTES.....	73
	ANEXO E: PENETRACIÓN DE LAS TARJETAS DE CRÉDITO POR PAÍS	74
	ANEXO F: PAÍSES EN QUE EVERIS TIENE PRESENCIA.....	75
	ANEXO G: FACTURACIÓN POR OFICINA	76
	ANEXO H: FACTURACIÓN POR ACTIVIDAD	77
	ANEXO I: ARQUITECTURA BIG DATA DE EVERIS	78
	ANEXO J: FUNCIONES DE LA ARQUITECTURA BIG DATA EVERIS.....	79
	ANEXO K: MUESTRA DE LA BASE DEL LOG DE TECLEO	80
	ANEXO L.1: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON818.....	81
	ANEXO L.2: 30 VECINOS MÁS CERCANOS AL FRAUDE JTAPIAL	82
	ANEXO L.3: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON243 (1)	83
	ANEXO L.4: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON643.....	84
	ANEXO L.5: 30 VECINOS MÁS CERCANOS FRAUDE CCON243	85

ANEXO M.1: DISTANCIA DE VECINOS MÁS CERCANOS A CCON818 A MEDIA DE FRAUDES	86
ANEXO M.2: DISTANCIA DE VECINOS MÁS CERCANOS A JTAPIAL A MEDIA DE FRAUDES	87
ANEXO M.3: DISTANCIA DE VECINOS MÁS CERCANOS A CCON243 (1) A MEDIA LOS FRAUDES	88
ANEXO M.4: DISTANCIA DE VECINOS MÁS CERCANOS A CCON643 A MEDIA DE FRAUDES	89
ANEXO M.5: DISTANCIA DE VECINOS MÁS CERCANOS A CCON243 A MEDIA DE FRAUDES	90
ANEXO N: USUARIOS CLASIFICADOS COMO OUTLIERS POR LOF	91
ANEXO O: DISTANCIA DE LOS SOSPECHOSOS LOF A LA MEDIA DE LOS DATOS	92
ANEXO P: RESULTADO ANÁLISIS DE LAS COMPONENTES PRINCIPALES .	93
ANEXO Q: LISTA COMPLETA DE SOSPECHOSOS ENCONTRADOS POR LOS TRES MODELOS	98
ANEXO S: DISTANCIA DE SOSPECHOSOS DE LA LISTA FINAL A LA MEDIA DE LOS DATOS	104

1. INTRODUCCIÓN

1.1. ANTECEDENTES GENERALES

El conjunto de instituciones financieras en Chile se divide entre entidades bancarias y no bancarias. En el país, al año 2016, se encuentran en operación 24 bancos, de los cuales 19 corresponden a bancos nacionales, 4 sucursales de bancos extranjeros y 1 banco estatal. Por su parte las entidades no bancarias se constituyen, principalmente por Cooperativas de Ahorro y Crédito (CAC's), entidades emisoras y operadoras de tarjetas de crédito, y otros oferentes de créditos masivos.

Al año 2015, el sistema en su conjunto, poseía activos cercanos a los 294 mil millones de dólares, 97% correspondiente al sector banca, lo que equivale a 1,23 veces el PIB del país¹ [1].

En la siguiente tabla se puede observar la distribución de los activos por tipo de institución:²

Entidades	Número	Activos (millones de dólares)
Bancos	24	287.851
Filiales de bancos supervisadas por la SBIF	16	1.127
Filiales de bancos que tienen supervisión conjunta con la SVS	41	6.294
Filiales extranjeras de bancos locales	2	17.184
Sociedades de apoyo al giro de bancos y CACs	22	1.274
Sociedades de apoyo al giro extranjeras de bancos locales	1	0
Sucursales de bancos locales en el exterior	3	9.009
Oficinas de representación de bancos extranjeros en Chile	22	
Cooperativas de ahorro y crédito (CACs)	7	2.287
Emisores y operadores de tarjetas de crédito	17	3.502
Operadores de tarjetas de débito	1	2
Sociedades evaluadoras de riesgo	4	
Auditores externos	10	
Instituciones de garantía recíproca (IGR)	16	
Sociedades evaluadoras de IGR	4	
Almacenes generales de depósito (AGD)	3	
Firmas evaluadores de AGD	2	
Total	195	294.216

Tabla 1: Entidades bajo la fiscalización de la SBIF a Diciembre de 2015

Fuente: SBIF

¹ Con respecto al PIB de Chile del año 2015.

² Los valores en blanco corresponden a entidades que, por diversos motivos, no requieren enviar información financiera a la Superintendencia de Bancos e Instituciones Financieras (SBIF).

El mercado financiero chileno ha mostrado una tendencia a la bancarización en las últimas décadas que se ha mantenido acorde al crecimiento del Producto Interno Bruto (PIB) del país. Esta tendencia se ve reflejada tanto en cobertura, como en profundidad e intensidad de uso de los distintos productos que ofrece el sector bancario.

La siguiente tabla introduce las dimensiones de la bancarización. Profundidad, cobertura e intensidad [2]:

Profundidad	<i>Dimensión asociada a la importancia relativa agregada del sistema financiero sobre la economía. Los indicadores más comunes de profundidad son total de depósitos y captaciones sobre PIB y total de activos bancarios sobre PIB.</i>
Cobertura	<i>Dimensión asociada a la distribución de los servicios financieros entre los distintos grupos de usuarios. En este ámbito, los indicadores más comunes son medidas tales como; número de medios de pago sobre la población, número de ATMs y/o sucursales por cada 100 mil habitantes, entre otros.</i> <i>En esta dimensión, se puede destacar también la problemática del acceso al financiamiento de las Mipymes, de los grupos de consumidores de menores ingresos y de las zonas geográficamente apartadas.</i>
Intensidad de uso	<i>Dimensión vinculada a la cantidad de transacciones bancarias realizadas por una población de referencia. Algunos ejemplos de tales indicadores son: el número de créditos (o captaciones) por cada 100 mil habitantes, el número de transacciones realizadas con medios de pago bancarios distintos de cheques y efectivos por habitante, el número de créditos sobre la fuerza de trabajo, entre otros.</i>

Tabla 2: Dimensiones de la Bancarización
Fuente: Morales y Yáñez, SBIF

Ya en 2005 se puede observar que existía una tendencia creciente en cuanto a la profundidad, tanto en colocaciones³ sobre el PIB, como en depósitos sobre el PIB. Así mismo estos indicadores de profundidad, para Chile, se encuentran bastante por sobre los niveles de los otros países de la región [3]⁴.

Estos valores además poseen una correlación directa con el ingreso per-cápita del país⁵, que sabemos ha ido creciendo con el transcurso de los años⁶.

³ Venta de activos financieros a instituciones o personas naturales.

⁴ Detalle de la evolución de los indicadores de profundidad financiera, y la comparación de estos valores con otros países, se puede encontrar en el Anexo A.

⁵ Correlación entre ingreso per-cápita e indicadores de profundidad se puede observar en el Anexo B.

⁶ En el Anexo C se grafica el crecimiento del ingreso per-cápita de Chile, a valor dólar de Agosto 2016, en los últimos años.

De la misma forma, se observa un aumento sostenido en cobertura, particularmente en número de cuentas corrientes, números de tarjetas de crédito y de débito⁷. Mas, el nivel de penetración de las tarjetas de crédito en Chile, comparado con otros países de la región, resulta aún baja⁸.

La intensidad de uso, por su parte, muestra tendencias similares de crecimiento para las tarjetas de crédito y débito. En desmedro del uso de cheques, que ha mostrado últimamente una clara tendencia a la baja.

Todas estas tendencias han gatillado que el sistema bancario se constituya como uno de los grandes actores y sostenes de la economía, generando una utilidad acumulada de 1.050.084 millones de pesos a Junio de 2016 [4].

En 2013, el sistema contaba con más de 4 millones de clientes con crédito, casi 3,5 millones de personas con cuenta corriente, 800 mil personas con depósitos a plazo y 280 millones de transferencias electrónicas de fondos. Se estima, además, que la contribución al PIB del sector bancario para ese año fue de unos 10.900 millones de dólares, lo que equivale a un 3,9% del PIB del país para ese año. Porcentaje que ha mostrado una tendencia creciente a lo largo del último tiempo, a pesar de haber sufrido una contracción entre 2009 y 2011 [5]:



Figura 1: Participación de la Banca en el PIB total (porcentaje)
Fuente: ABIF. Con datos de la SBIF y el Banco Central

⁷ El Anexo D muestra el número de tarjetas de crédito, débito y cuentas corrientes en el sistema bancario chileno a través de los años (en millones).

⁸ La comparación de la penetración de las tarjetas de crédito en Chile, con respecto a los otros países se encuentra en el Anexo E (Nº de tarjetas cada 1.000 habitantes).

El fenómeno descrito, adicionalmente, conlleva un mayor tráfico operacional para las instituciones bancarias que tienen que lidiar con una mayor cantidad de clientes, mayor cantidad de cuentas, más transacciones diarias y, por ende, una mayor complejidad a la hora de realizar un seguimiento a todas las transacciones que se generan.

También ha hecho del sector bancario un mercado atractivo para nuevos entrantes, lo que se ha visto reflejado en los últimos años con la inclusión de nuevos actores. Muchos de ellos pertenecientes a empresas que se desenvuelven en otros rubros y que buscan diversificar sus negocios ingresando al mercado de la banca. Pero, a pesar de que existe una gran cantidad de entidades bancarias presentes en el país, el mercado se encuentra bastante concentrado en las tres mayores instituciones bancarias que poseen más del 50% de la participación de mercado. Banco Santander con un 19,4%, Banco de Chile con un 18,4% de participación y Banco Estado con un 14,4% en el año 2016 [6].

En la gráfica, a continuación, se puede observar las participaciones de mercado en la banca y cómo ellas han ido evolucionando durante el tiempo:

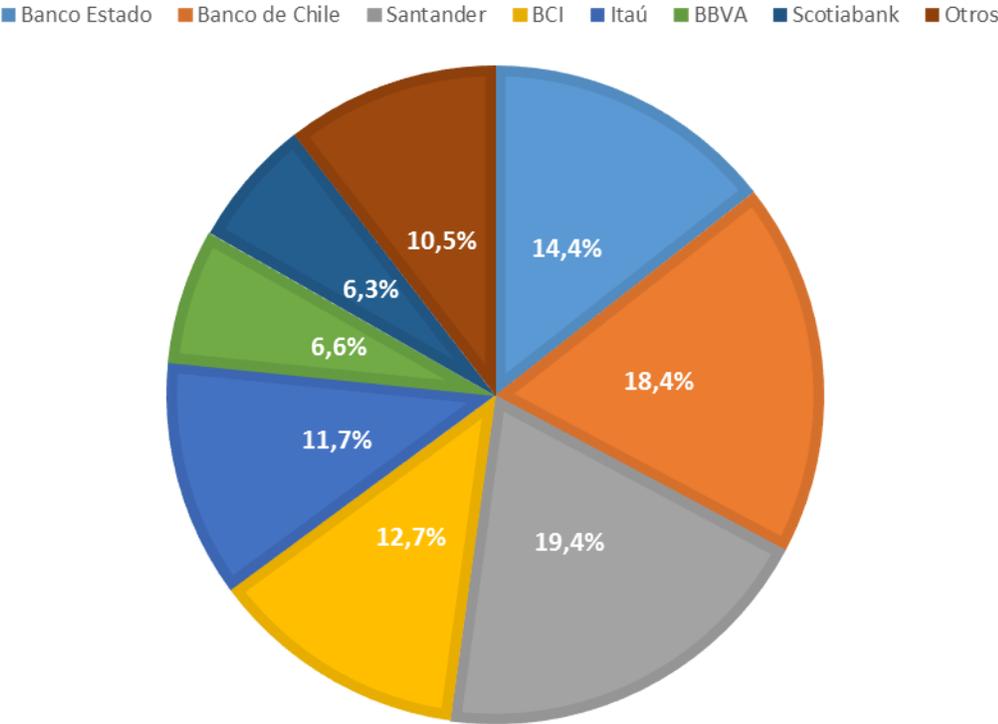


Figura 2.1: Participaciones de mercado a Junio 2016 (Colocaciones/Colocaciones Sistema)

Fuente: Elaboración propia, con datos de la SBIF

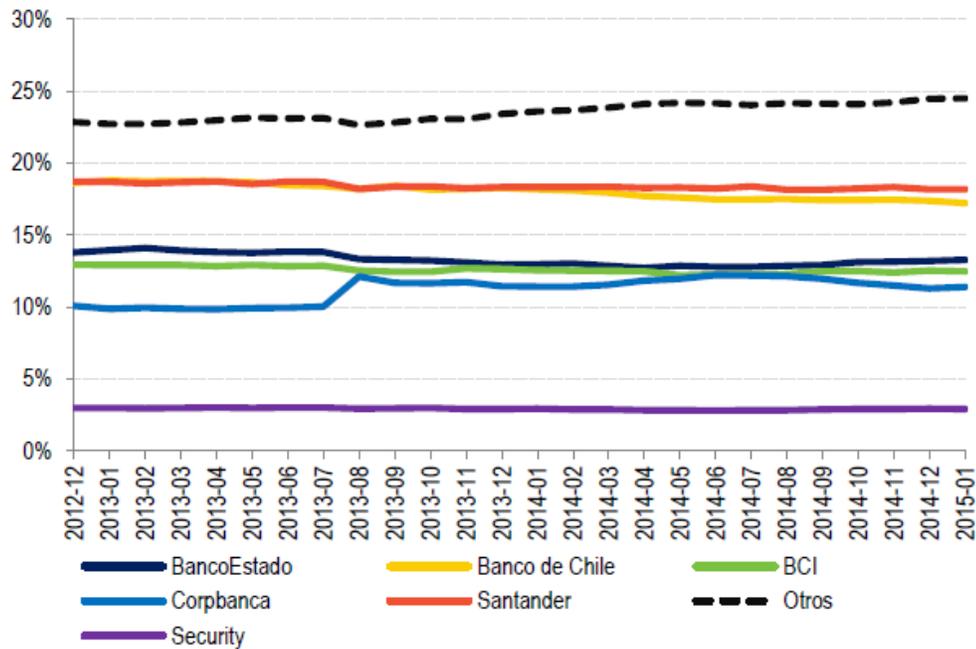


Figura 2.2: Evolución de participaciones de mercado (Colocaciones/Colocaciones Sistema)

Fuente: CorpResearch, con datos de la SBIF

1.2. DEFINICIÓN Y JUSTIFICACIÓN DEL PROYECTO

Uno de los principales afanes de toda institución bancaria es proyectar una imagen de probidad, responsabilidad y seguridad para que sus clientes sientan la confianza de depositar sus activos financieros en la institución. Por esto resulta sumamente importante coartar cualquier intento de fraude⁹ que generase un perjuicio económico, o pueda poner en riesgo los activos del banco y sus clientes.

Bajo esta premisa es que una de las instituciones bancarias de mayor tamaño con presencia en Chile solicita a la empresa española de consultoría Everis, en el marco de una prueba de concepto, desarrollar un modelo predictivo que les

⁹ Dentro del Marco Conceptual se puede encontrar la definición de fraude, y los tipos de fraudes existentes

permita detectar la ocurrencia de fraudes internos llevados a cabo por cualquier empleado de la institución, con el fin revelarlos, y a futuro evitarlos.

Everis es una consultora de origen español fundada en 1996 y que actualmente pertenece al grupo japonés NTT Data, la sexta compañía de servicios IT más grande del mundo. Cuenta con una plana de más de 11.000 profesionales en las 22 oficinas que posee alrededor del mundo (En Argentina, Bélgica, Brasil, Chile, Colombia, Italia, México, Perú, Portugal, Reino Unido, USA y 11 en España)¹⁰, en el año 2013 tuvo una facturación del orden de los 608 millones de euros a nivel global (41 millones de euros en Chile)¹¹, y se enfoca principalmente en entregar soluciones de negocio, estrategia, desarrollo y mantenimiento de aplicaciones tecnológicas, y outsourcing [7]. Siendo los mayores focos de recaudación los servicios y mantenimiento de sistemas, junto con el desarrollo de software¹².

Se desarrolla en diversos sectores económicos como lo son Salud, Telecomunicaciones, Utilities (Servicios Básicos), Industria, Banking (Entidades Financieras), Sector Público, etc.

La empresa compite directamente en Chile con otras consultoras como Sonda, Accenture, Indra, Experian, Coasin, entre otras. También se pueden considerar competencia empresas como IBM y HP como proveedor de servicios tecnológicos.

La entidad bancaria por su parte, es una de las tres que posee mayor participación de mercado en Chile, cuenta con 11.723 empleados, cerca de 3,6 millones de clientes, 1.536 cajeros automáticos, y más de 470 sucursales a lo largo de Chile, según la memoria anual del banco del año 2015.

En el gráfico presentado a continuación se puede ver la evolución de empleados y sucursales de la institución bancaria en los últimos años:

¹⁰ Detalle de los países en que tiene presencia Everis y NTT DATA se encuentra en el Anexo F.

¹¹ Detalle de la facturación de Everis se encuentra en el Anexo G.

¹² Detalle de facturación de Everis por tipo de servicio se encuentra en el Anexo H.

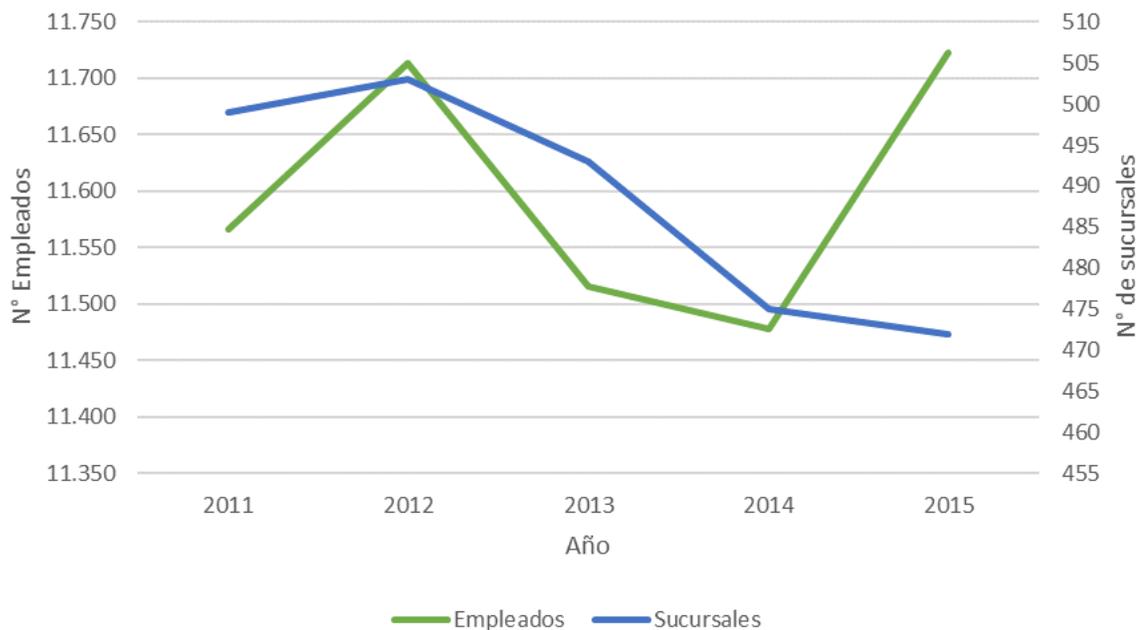


Figura 3: Evolución de cantidad de empleados y sucursales del banco
Fuente: Elaboración propia, a partir de los informes anuales del banco

Mientras que el número de clientes ha ascendido desde 3,4 millones en el año 2011 a más de 3,6 millones en 2015.

Esta cantidad de clientes, empleados y sucursales, generan aproximadamente 20 millones de transacciones¹³ diarias (más de 7.600 millones al año) que complican en demasía el monitoreo del comportamiento de los empleados. Por lo que, hacer seguimiento y control manual para discernir si el conjunto de transacciones de un usuario es constitutivo de fraude o no, resulta prácticamente imposible.

Todas estas transacciones asociadas a los distintos usuarios del sistema quedan registradas diariamente en el Log de Tecleo del banco. Existen más de 2.000 tipos de transacciones, y una transacción por sí sola no puede constituir fraude. Un fraude se explica por un conjunto de transacciones sobre una cuenta particular, que luego se concretan en el cobro de un documento, ya

¹³ Transacciones hace referencia a cualquier acción que lleve a cabo un usuario (empleado) de la institución bancaria en el sistema computacional. Todas estas transacciones quedan registradas en el Log de Tecleo del Banco. Algunas de ellas son: Consultas de saldo de una cuenta, mantenimiento de órdenes de no pago, consulta de últimos movimientos, entre otras.

sea cheque o vale vista. Esto complica aún más la detección de casos de fraude, ya que se debe revisar el conjunto de transacciones realizado por un usuario sobre una cuenta. Adicionalmente estas transacciones pueden no ser secuenciales, ya que el usuario puede realizar las transacciones en distintos momentos del día o realizar, en el intertanto, transacciones sobre otras cuentas.

La importancia de la detección y prevención del fraude radica no sólo en el menoscabo que genera a la imagen de probidad, seguridad y confianza que debe proyectar un banco para que sus clientes confíen su dinero a esta institución, y en la pérdida de activos por parte del banco o el cliente. Sino también en que los fraudes no son un problema aislado y poco común en las instituciones chilenas.

No se cuenta con datos que muestren la cantidad de fraudes que se generan en la institución bancaria en cuestión, ni la cuantía de estos. Sin embargo, un estudio realizado en el año 2009, revela que un 41% de las empresas chilenas declaraba haber sufrido algún tipo de fraude [8]. Esta cifra, a pesar de ser menores a los niveles registrados en otros países de la región, no deja de ser considerable:

Pais	Fraudes
Malasia	83%
México	77%
Estados Unidos	74%
Sudáfrica	72%
Alemania	71%
Brasil	69%
Austria	57%
Australia	45%
Chile	41%
Bélgica	37%
Arabia Saudita	32%
Kuwait	32%
Omán	32%
Qatar	32%

Tabla 3: Porcentaje de empresas que declaran haber sufrido fraude por país
Fuente: KPMG Chile

Si nos remitimos sólo a los casos de fraude interno¹⁴, es decir los fraudes cometidos por empleados de la misma compañía, el porcentaje presente en empresas chilenas es de un 29%. Aún por debajo de los otros países de la región:

País	Fraudes Internos
Brasil	58%
México	46%
Estados Unidos	37%
Argentina	33%
Chile	29%

Tabla 4: Porcentaje de empresas que declaran haber sufrido fraude interno por país
Fuente: KPMG Chile

Dentro de los fraudes internos presentes en las empresas chilenas, las tres categorías que concentran un mayor porcentaje del universo son la colusión entre miembros de la organización con un 36% del universo, fraudes entre empleados y terceras partes con un 15%, y el fraude cibernético con un 11%. Por otro lado, los fraudes de caja y bancarios constituyen un 15% del total de fraudes internos presentes en Chile:

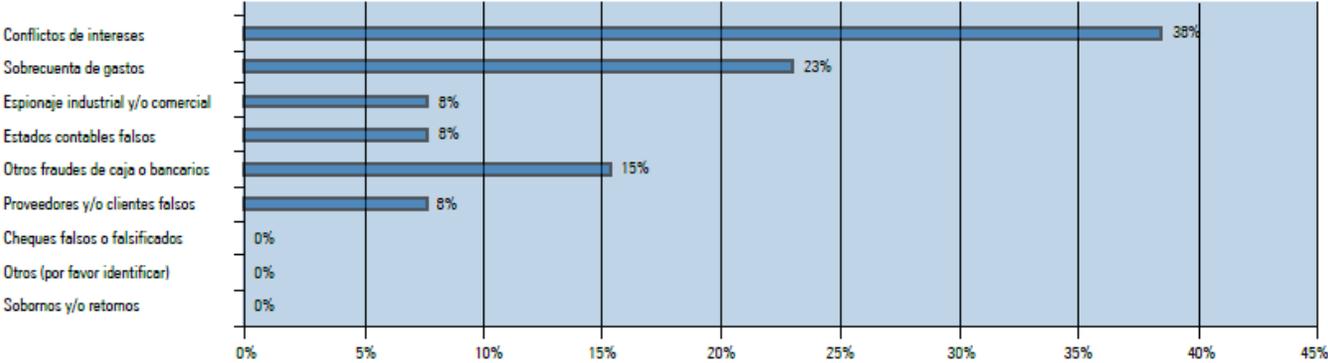


Figura 4: Distribución de los tipos de fraude interno en las empresas de Chile
Fuente: KPMG Chile

Adicionalmente, se indica que un 91% de los casos el defraudador comete más de un ilícito.

¹⁴ Dentro del Marco Conceptual se puede encontrar una definición más acabada de fraude interno.

Se estima que una organización pierde típicamente el 5% de sus ingresos debido a fraudes cada año [9]. Considerando los ingresos declarados por el banco en su memoria anual del año 2015, el monto perdido debido a fraudes podría ascender a más 62.760 millones de pesos.

Si consideramos que el 62% de los fraudes corresponden a colusión entre miembros de la organización y fraude entre empleados y terceras partes, ambos considerados casos de fraude interno [8], el monto en juego es de más de 38.911 millones de pesos. Por otra parte, un 15% de estos fraudes internos corresponden a fraudes de caja o bancarios, que significan alrededor de 5.800 millones de pesos.

Finalmente, si el modelo propuesto pudiese evitar la ocurrencia de un 5% de estos fraudes significaría reducir la pérdida en más de 290 millones de pesos anuales.

Por todo esto resulta sumamente relevante detectar estas faltas o delitos, no sólo con el fin de prevenir la comisión de nuevos ilícitos, sino que para generar conocimiento a partir de ellos. Todo esto, comprendiendo que el fraude es, según Van Vlasselaer et al. (2015): Un crimen poco común, bien planeado, imperceptible, que evoluciona en el tiempo y muchas veces cuidadosamente planeado, que puede presentarse de muchas formas distintas [10].

Debido al impacto económico que generan los fraudes en las corporaciones, y a las nuevas posibilidades que ofrecen el desarrollo tecnológico y las herramientas de analítica para detectarlos, es que cobra relevancia el estudio de Human Resources Analytics y Fraud Analytics.

El primero hace referencia al uso de modelos predictivos para el análisis del comportamiento de los empleados de una institución, así como para administrar de la mejor forma posible a su capital humano. Algunas de las áreas de acción que maneja esta disciplina son:

- Reducir la rotación de personal
- Anticipar el desempeño de un empleado para un cargo particular

- Definir la asignación de personal para maximizar la eficiencia
- Definición de políticas de compensación
- Mejorar la satisfacción de los empleados
- Reclutamiento de personal
- Identificar desempeños deficientes o anómalos

Por otro lado, Fraud Analytics hace referencia al uso de modelos predictivos que sean capaces de identificar patrones en los datos que se desvían del comportamiento normal, y que pueden constituir fraudes. El fin último de esta disciplina es, a partir del conocimiento extraído de los datos y los modelos disponibles, establecer reglas para identificar fraudes, caracterizarlos y finalmente implementar algoritmos capaces de reconocer estos patrones a tiempo para evitar la comisión de fraudes antes de que sucedan.

Algunos de los tipos de fraude que aborda esta disciplina son el fraude de tarjeta de crédito, fraude de seguro, corrupción, falsificación, fraude de telecomunicación, lavado de dinero, robo de identidad, evasión de impuestos, plagio, entre otros.

El desarrollo del presente trabajo se enmarca dentro de ambas, ya que pretende estudiar el comportamiento de empleados, con el fin de encontrar patrones anómalos que pudieran constituir casos de fraude interno.

1.3. OBJETIVOS

1.3.1. GENERAL

Desarrollar un modelo predictivo que sea capaz de detectar, a partir de data transaccional, usuarios sospechosos de haber cometido fraude interno.

1.3.2. ESPECÍFICOS

- Detectar patrones anómalos dentro de los datos que puedan constituir indicios de fraude interno.
- Definir criterios, por detección de outliers y comparación con los registros de fraude, que permitan seleccionar usuarios sospechosos de haber cometido fraude interno.
- Establecer qué variables y comportamientos resultan buenos indicadores de la ocurrencia de fraude interno.
- Entregar un listado de usuarios sospechosos, a partir de los criterios empleados con anterioridad, y caracterizarlos.

1.4. ALCANCES

- Se analizarán exclusivamente casos de fraude interno, es decir, fraudes o faltas llevadas a cabo por empleados de la institución bancaria.
- Exclusivamente se analizarán los tipos de fraudes presentes en los registros entregados indicados como sospechosos por el banco, y no otros tipos de fraude interno que se puedan perpetrar. Por lo que el modelo no resultará generalizable a otros tipos de fraude.
- Se considerará exclusivamente las transacciones realizadas por los usuarios del banco el día 02 de Abril de 2015, que son los datos que el banco puso a disposición.

- No se desarrollará un modelo que sea aplicable a tiempo real, debido a que los datos proporcionados no lo permiten.
- Se estudiarán solamente los tipos de transacciones presentes en los casos de fraude proporcionados por la entidad bancaria, por lo que no se estudiarán fraudes asociados a otro tipo de transacciones.
- Debido a la disponibilidad de datos se analizará exclusivamente el comportamiento transaccional de los empleados, y no se realizarán análisis de secuencialidad en las transacciones.

1.5. RESULTADOS ESPERADOS

Con el presente trabajo se espera llegar a un modelo que permita establecer una cantidad razonable de usuarios sospechosos de haber cometido fraudes internos similares a los presentes en los registros de fraude proporcionados por el banco.

Además de identificar las similitudes de estos con los casos de fraudes, y variables que puedan resultar en buenos indicadores de la presencia de este tipo de fraudes.

Finalmente se pretende entregar una lista de usuarios sospechosos, indicando cuáles de ellos presentan un patrón de transacciones que tienen mayor probabilidad de constituir casos de fraude interno similares a los analizados, detallando además su comportamiento y similitudes con dichos casos.

2. MARCO CONCEPTUAL

2.1. DEFINICIÓN DE FRAUDE

Existen diversas definiciones de fraude, todas ellas, indicándolo como un actuar tendencioso y deliberado para obtener beneficio personal, perjudicando a un tercero.

La Real Academia Española (RAE) describe fraude como una “acción contraria a la verdad y a la rectitud, que perjudica a la persona contra quien se comete”, o como un “acto tendente a eludir una disposición legal en perjuicio del Estado o de terceros”.

Otra definición que caracteriza de mejor forma el fenómeno del fraude es la entregada por el Oxford Dictionary que lo describe como un “engaño criminal o ilegal realizado con la intención de obtener beneficios financieros o personales”.

Una definición que ahonda más en las cualidades del e intenciones de un fraude es la entregada por Van Vlasselaer et al. (2015), que lo define como “Un crimen poco común, bien planeado, imperceptible, que evoluciona en el tiempo y muchas veces cuidadosamente planeado, que puede presentarse de muchas formas distintas”.

Es por estas características que resulta sumamente difícil reconocer y detectar casos de fraude, ya que el defraudador trata de mimetizarse y pasar desapercibido entre los registros no fraudulentos. Y aprende, evoluciona y perfecciona nuevos métodos para defraudar cuando un tipo de fraude particular ya fue detectado.

2.2. TIPOS DE FRAUDE

Los fraudes se dan en muchos ámbitos, mercados e instituciones. Y si bien, cada fraude se constituye como un caso particular y único, existen fraudes que resultan más recurrentes y se pueden agrupar bajo ciertas clasificaciones.

Algunos de ellos son [9]:

- Fraude de tarjeta de crédito: Fraudes asociados al uso de tarjetas de créditos hurtadas, robadas o perdidas. También puede ser por la adquisición de tarjetas de crédito por medio de identificaciones falsas, u otros tipos.
- Fraude de seguro: Cualquier tipo de fraude asociado a un seguro, ya sea por el lado del comprador o vendedor.
- Corrupción: Es el mal uso o abuso de una situación de poder para generar ganancia personal.
- Falsificación: Una imitación, generalmente de algún producto de valor, que pretende hacerse pasar por algo genuino. Pueden ser identificaciones, documentos, dinero, tarjetas de crédito, productos, etc.
- Fraude de telecomunicación: Comprende el robo de servicios de telecomunicación, la clonación de números telefónicos, suplantación de identidad en correos y otras cuentas personales, entre otras.
- Lavado de dinero: Consiste en hacer que activos provenientes de actividades ilícitas circulen de forma normal por el sistema financiero.
- Robo de identidad: Es el crimen consistente en obtener la información personal o financiera de otra persona, con el fin de asumir la identidad de esa persona para obtener beneficios económicos.
- Evasión de impuestos: Falsear información con el fin de evadir los deberes tributarios que se tienen.
- Plagio: Robar y hacer pasar por propias las ideas, o propiedad intelectual de otra persona.

2.3. DEFINICIÓN DE FRAUDE INTERNO

Los fraudes también pueden clasificarse según el origen de la persona que los comete. En particular, si esta persona pertenece o no a la institución en que se comete el fraude. Esto resulta relevante porque el defraudador que pertenece a la organización en que se realiza el fraude se encuentra en una situación de poder y ventaja frente al defraudado, ya que posee acceso a información que puede resultar provechosa para generar ganancias personales y perjuicio en la otra persona.

Así se divide el fraude entre externo, realizado por una persona ajena a la institución en que se realiza el fraude, e interno, que lo realiza un empleado de la institución en que se desarrolla el caso de fraude.

El fraude externo se define como las "Pérdidas derivadas de algún tipo de actuación orientada a cometer fraude, apropiarse de bienes o vulnerar la legislación por parte de terceros (ajenos al banco), por ejemplo, delitos informáticos y robos en sucursales."

Mientras que el fraude interno es definido por la SBIF como las "Pérdidas derivadas de algún tipo de actuación intencional cometida por personal del banco orientada a cometer fraude, incurriendo en prácticas o procedimientos que no son legales o no están autorizados, anteponiendo sus intereses personales a los de la organización." [11]

A su vez el fraude interno se puede subdividir en dos categorías [12]:

- Actividades no autorizadas: Operaciones no reveladas intencionalmente, operaciones no autorizadas con pérdidas monetarias y valoración errónea intencional de posiciones.
- Hurto y fraude: Fraude, fraude crediticio, depósitos sin valor, hurto, extorsión, malversación, robo, apropiación indebida de activos, destrucción dolosa de activos, falsificación, utilización de cheques sin fondo, contrabando, apropiación de cuentas de identidad, etc.

3. MARCO TEÓRICO

3.1. EL PROCESO KDD

El proceso KDD o “Knowledge Discovery in Databases” propuesto por Fayyad, Piatetsky-Shapiro y Smyth en 1996 es definido como [13]:

“El proceso no trivial de identificar patrones en los datos que resulten válidos, originales, potencialmente útiles y finalmente comprensibles.”

Básicamente es una serie de etapas y pasos para descubrir patrones en los datos que puedan aportar inteligencia para la resolución de un problema de negocio.

Los pasos que definen a grandes rasgos este proceso son la recolección y preparación de los datos, el pre-procesamiento de los datos, transformación de los datos, minería de datos y evaluación e interpretación de resultados.

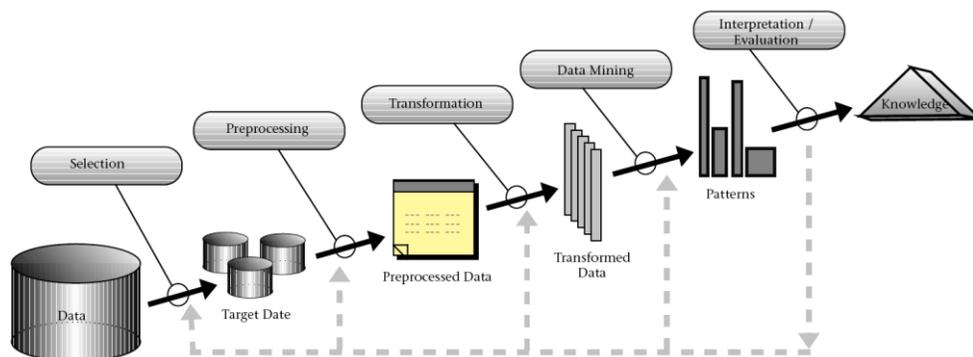


Figura 5: Etapas del proceso KDD
Fuente: Fayyad et al.

Mas esta subdivisión se puede ampliar a una serie más acabada tareas como las que se exponen a continuación:

- Entendimiento del dominio de aplicación, del conocimiento previo requerido e identificar el objetivo del proceso KDD desde el punto de vista del cliente.
- Selección de la data o el subconjunto de datos y variables sobre los que se generará el descubrimiento de información.
- Limpieza de los datos y pre-procesamiento: Operaciones básicas como remover datos que constituyan "ruido", recolectar información necesaria para el modelo y decisión de estrategias para manejar datos faltantes.
- Reducción de la data: Encontrar formas apropiadas de representar los datos, acorde al objetivo de la tarea a realizar. Usando reducción de dimensionalidad o métodos de transformación para reducir el número de variables.
- Alinear los objetivos del proceso KDD con un método de data mining en particular. Por ejemplo: Reglas de clasificación, árboles de decisión, regresiones, clustering, etc.
- Elección del algoritmo de data mining: Selección del método a utilizar para encontrar los patrones en los datos. Esto incluye también la elección de parámetros para el modelo.
- Minería de datos: Búsqueda de patrones de interés en los datos a través del modelo seleccionado.
- Interpretación de los patrones encontrados incluyendo, tal vez, visualización de estos patrones y de los datos.
- Consolidación del conocimiento descubierto: Incorporación del conocimiento a otros sistemas para acciones posteriores. O simplemente documentar este conocimiento.

3.1.1. RECOLECCIÓN Y PREPARACIÓN DE LOS DATOS

Las fuentes de datos no suelen ser pulcras y óptimas para aplicar directamente los algoritmos de data mining. Existen muchos datos que no aportan a generar el conocimiento que se espera lograr. Por ello, en esta etapa se unen fuentes de datos, se seleccionan los datos y las variables que resultan de interés para realizar los posteriores análisis, usualmente con la guía de un experto del negocio. La idea es contar con un subconjunto de datos que permita extraer la mayor cantidad de información minimizando los tiempos de procesamiento.

3.1.2. PRE-PROCESAMIENTO DE LOS DATOS

Esta etapa consiste en la limpieza y preparación de los datos para su posterior uso. El objetivo es eliminar cualquier tipo de incongruencia en los datos, manejar datos faltantes o fuera de rango. Con el fin de obtener un conjunto de datos consistente para la transformación y posterior aplicación de algoritmos de data mining.

Entre las técnicas más utilizadas para tratar datos faltantes se encuentran [14]:

- **Análisis de casos completos:** Se consideran exclusivamente los registros que no tienen datos faltantes para ninguna de las variables.
- **Análisis de casos disponibles:** Se estiman los valores de los datos faltantes a partir los datos disponibles tanto de la variable en cuestión, como de otras variables y se utiliza el coeficiente de correlación.
- **Imputación:** Con el afán de preservar la cantidad de registros originales, se rellenan los valores faltantes con estadísticos descriptivos calculados a partir de los datos disponibles. Por ejemplo, la media de la variable.

Otros métodos utilizados es la imputación a partir de la simulación de la distribución de los registros disponibles para la variable. O la maximización de la verosimilitud a través del algoritmo EM (Expectation-Maximization).

3.1.3. TRANSFORMACIÓN DE LOS DATOS

Consiste en el tratamiento preliminar de los datos. Muchas veces los datos vienen como registros transaccionales que deben ser procesados y transformados para ser ingresados a un modelo. En esta etapa se generan nuevas variables, con una estructura de datos apropiada que permita la aplicación de modelos.

Entre las técnicas más utilizadas para la normalización y transformación de variables se encuentran [9]:

- Estandarización Min-Max: $x_{nuevo} = \frac{x - \min(x)}{\max(x) - \min(x)} (nuevomax - nuevomin) - nuevomin$
Donde nuevomin y nuevomax son los rangos mínimos y máximos impuestos para la variable.
- Estandarización z-score: $z_i = \frac{x_i - \mu}{\sigma}$ con μ la media de la variable y σ la desviación estándar.
- Escalamiento decimal: $x_{nuevo} = \frac{x}{10^n}$ con n el número de dígitos del valor absoluto máximo.
- Binarización de variables categóricas: Los modelos de data mining, en general, trabajan exclusivamente con variables numéricas. Por esto, cuando tenemos una variable categórica es necesario transformarla generando n variables binarias, cada una representativa de alguna de las n categorías, que tendrán valor 1 si la variable pertenece a la categoría o 0 si esta no pertenece a dicha categoría. Es decir:

$$x_i = \begin{cases} 1 & \text{si la variable pertenece a la categoría } i \\ 0 & \text{sino} \end{cases}$$

3.1.4. DATA MINING

Corresponde a la fase de modelamiento, donde algoritmos y métodos de inteligencia son aplicados para descubrir los patrones subyacentes en los datos que resulten útiles, que fuesen desconocidos y comprensibles con el fin de extraer el mayor conocimiento posible.

Los modelos se dividen en tres categorías [15]:

- Modelos supervisados: Modelos que requieren que los datos vengan pre-clasificados como normales o anormales. Los modelos ajustan parámetros para maximizar el ajuste con los datos de entrenamiento y posteriormente su validez puede ser contrastada con los datos de testeo.
- Modelos no supervisados: Clasifican los datos sin un conocimiento previo de ellos. No requieren de una variable objetivo con valores predeterminados, generan conocimiento a partir de los datos disponibles, reconociendo la distribución subyacente de los datos e identificando los puntos que se alejan de esa distribución.
- Modelos semi-supervisados: Esta es una categoría algo nueva, ya que los dos anteriores son los enfoques más tradicionales para afrontar problemas de data mining. Requieren que una pequeña porción de los datos esté pre-clasificada y hace uso tanto de la data clasificada, como de la no clasificada.

Tradicionalmente los dos primeros enfoques son los más utilizados en problemas de detección de fraude. Por ello, a continuación, se analizarán algunos de los modelos más utilizados para afrontar un problema de detección de fraude.

3.1.4.1. MODELOS SUPERVISADOS

Entre los modelos supervisados más utilizados para la detección de fraude se encuentran:

- Regresión Logística [16]: Es un tipo de análisis de regresión para analizar el comportamiento de una variable categórica dicotómica (que puede tomar dos valores), en función de las variables independientes. Es una adaptación de una regresión simple, aplicando una función que fuerce a la variable dependiente a estar entre 0 y 1.

- Árboles de decisión [17]: Los árboles de decisión son modelos que clasifican registros a partir de los valores de una serie de variables predictivas, en iteraciones escalonadas. Cada nodo del árbol representa una instancia de decisión para clasificar un registro a partir del valor de una variable particular. Y cada rama representa el valor que puede tomar dicha variable. Los registros parten por el nodo inicial, y van siendo derivados por las ramas del árbol de acuerdo a cómo se compara el valor de sus variables con la regla de decisión.

- Support Vector Machine (SVM): La idea de un Support Vector Machine es generar un hiperplano óptimo, entre dos clases separables (por ejemplo, una variable objetivo con valores 0 o 1), que maximiza la distancia entre el plano y elemento más cercano de ambas clases. Así los datos que quedan a un costado del hiperplano son categorizados con un valor 1, mientras que los que quedan del otro lado se les asigna el valor 0.

- Redes Neuronales Artificiales: Las redes neuronales artificiales son modelos que replican el funcionamiento del sistema nervioso humano. Consisten, básicamente, en una capa con nodos de entrada en que los "inputs" son multiplicados por un vector de pesos que simula la intensidad de la sinapsis neuronal y luego una función matemática determina la activación de la neurona que permite el paso de las salidas. Otra función calcula el "output" de la neurona. Los pesos que ligan las neuronas se van modificando de modo, aprendiendo de los datos, para maximizar el ajuste del modelo.

Como estos modelos requieren que una gran cantidad de datos vengan pre-clasificados como normales o anormales, que no es el caso particular de este trabajo, un enfoque no supervisado resulta más apropiado y de mayor utilidad.

3.1.4.2. MODELOS NO SUPERVISADOS

Los mecanismos de detección que se basan en modelos no supervisados o analítica descriptiva, son una gran herramienta para encontrar registros que se desvían del comportamiento del resto de los datos, es decir, son una buena herramienta para encontrar outliers.

Un outlier es un registro que parece desviarse marcadamente de los otros miembros de la muestra en que se presenta.

Otra definición más profunda explica que un outlier es una observación o conjunto de observaciones que parece ser inconsistente con el conjunto restante de datos [18].

Las técnicas de detección de outlier poseen gran valor y permiten encontrar una fracción significativa de casos fraudulentos [9].

Es por esto, y por la poca cantidad de datos categorizados como fraudes con que se dispone que se ha optado por estos métodos para desarrollar el presente trabajo.

Probablemente uno de los modelos no supervisados más utilizados sean los métodos de clustering o segmentación. Estos agrupan los datos en distintos segmentos tratando de maximizar el parecido de los elementos de un mismo segmento y maximizar, a la vez, la diferencia entre elementos de distintos segmentos.

Algunos de los algoritmos de clustering más utilizados son [19]:

- Clustering Jerárquico Aglomerativo (Single Linkage): Parte con tantos clusters como datos hay. Es decir, cada dato es un cluster. Luego mide las distancias mínimas entre los objetos de todos los clusters y une los dos clusters cuya distancia mínima es menor. Se define un criterio para detener el algoritmo de acuerdo a cuantos clusters se desea obtener.
- Clustering Jerárquico Aglomerativo (Complete Linkage): Se mantienen los mismos criterios de inicio y detención del algoritmo que en el método anterior. Pero el criterio para juntar dos clusters consiste en calcular las distancias máximas entre los elementos de los clusters, y unir los dos clusters cuyas distancias máximas son mínimas.
- K-means: Este algoritmo requiere que se defina a priori el número de clusters que resultarán. El algoritmo parte con una configuración inicial y luego asigna cada punto al cluster cuyo centro (media) es más cercano. El algoritmo se detiene cuando entre dos iteraciones no hay movimientos de puntos entre clusters.
- Fuzzy-Logic: Este algoritmo de clusterización genera clusters difusos, en que cada punto posee un factor de pertenencia a cada cluster. Así todos los puntos pertenecen en alguna medida (puede ser cero) a todos los clusters.

Las metodologías utilizadas en esta memoria se enmarcan dentro de los modelos no supervisados y como métodos de detección de outliers.

3.1.4.2.1. ANÁLISIS DE VECINOS MÁS CERCANOS

Para data de alta dimensionalidad, la distancia euclidiana proporciona una buena medida de similitud entre dos puntos. Siempre y cuando las variables compartan escalas similares para que las variables que poseen mayores medias no se vean sobrestimadas al momento del cálculo de la distancia [18].

La distancia euclidiana se define por:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Donde X_i es la componente i del vector X e Y_i es la componente i del vector Y . Ambos vectores poseen n elementos.

Básicamente calcula la suma de las distancias absolutas de cada variable para un par de registros, lo que otorga una medida de similitud.

3.1.4.2.2. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El análisis de componentes principales posee como objetivos extraer información de los datos, reducir la dimensionalidad de los datos manteniendo la información relevante, simplificar la descripción de los datos y analizar la estructura de ellos [20].

Este análisis genera nuevas variables, llamadas componentes principales, que resultan de combinaciones lineales de las variables anteriores.

La primera componente requiere tener la mayor varianza posible, luego la segunda componente se fuerza a que sea ortogonal a la primera, y así sucesivamente.

Este análisis permite visualizar la distribución de datos que poseen alta dimensionalidad, lo que resulta una buena herramienta para visualizar outliers.

3.1.4.2.3. LOCAL OUTLIER FACTOR (LOF)

Es un algoritmo que permite encontrar outliers midiendo las desviaciones de la densidad de un punto con respecto a la densidad de los puntos dentro de su vecindad. A cada punto le asigna un factor que, mientras mayor, indica una mayor desviación de la densidad del punto con respecto a sus vecinos más cercanos. Si este valor es cercano a 1 quiere decir que la densidad del punto es muy similar a la de sus vecinos. Un valor significativamente menor a 1 indica una región más densa, y un valor significativamente mayor a 1 indica que el punto posee una densidad bastante menor a sus vecinos, y por tanto sería un outlier [21].

Una de las ventajas de utilizar este método es que es capaz de detectar outliers locales que pueden no resultar outliers al hacer un análisis sobre todo el conjunto de datos o sobre otro conjunto de datos distintos.

La desventaja de utilizar este método es que el factor entrega un número que no posee una escala única para determinar si un dato es outlier. Los rangos que puede tomar el factor variarán entre dos conjuntos de datos distintos, por lo que se debe definir un threshold (punto de corte) a partir del cual los datos serán considerados outliers.

El factor LOF se calcula de la siguiente forma:

Se define la *distancia* – $k(A)$ como la distancia entre a y su k -vecino más cercano. Es decir, la distancia que contendrá a los k -vecinos más cercanos de A .

Se define la *reachability* – $distance_k(A, B)$ como el máximo entre la distancia de los k -vecinos más cercanos a B y la distancia entre A y B :

$$reachability - distance_k(A, B) = \text{Max}\{distancia - k(B), d(A, B)\}$$

Se calcula también la "local reachability distance" que resulta ser el inverso del promedio de las reachability-distance entra A y sus vecinos.

$$lrd(A) = 1 / \left(\frac{\sum_{B \in N_k(A)} reachability - distance_k(A, B)}{|N_k(A)|} \right)$$

Con $N_k(A)$ el conjunto de los k vecinos más cercanos a. Posteriormente, comparando este valor con el de los k -vecinos más cercanos de A se obtiene el factor LOF como sigue:

$$LOF_k(A) = \frac{\sum_{B \in N_k(A)} \frac{lrd(B)}{lrd(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} lrd(B)}{|N_k(A)|} / lrd(A)$$

3.1.5. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En esta etapa se identifican los patrones obtenidos, se analizan los desempeños de los modelos, se interpreta la información, todo con el fin de obtener insights que puedan ayudar a resolver el problema de negocio que se quiere solucionar.

Con el fin de unificar los resultados arrojados por los tres modelos y generar una lista única de sospechosos, se utilizan criterios de Ensemble Classification para obtener un output final.

Ensemble Classification hace referencia al conjunto de métodos por el cual se combinan los resultados o predicciones obtenidas por dos o más modelos predictivos. Existen dos enfoques principales para afrontar este tipo de tarea: Los métodos Meta-learning y los métodos Weighting [22].

Los métodos Meta-learning son métodos de aprendizaje sucesivo en que la predicción de un modelo sirve como input o data de entrenamiento para el siguiente. Resultan útiles para los casos en que ciertos modelos clasifican consistentemente mal alguna instancia. Dentro de estas técnicas se encuentra el stacking, arbiter trees, combiner trees y grading.

En los métodos Weighting los resultados de los distintos modelos son combinados usando pesos o ponderaciones que se asignan a cada uno. El peso indica el efecto que tendrá en la clasificación final. Estos métodos resultan útiles cuando los modelos realizan la misma tarea de clasificación y poseen un desempeño similar. Entre estos métodos se encuentran: majority voting, performance weighting y bayesian combination.

Debido a que la mayoría de ellos, como performance weighting y bayesian combination, requieren conocer el ajuste del modelo, y por tanto tener los datos pre-clasificados (tener data de testeo), se utiliza majority voting.

Majority voting clasifica una instancia de acuerdo a la clase que obtiene mayor cantidad de votos, es decir que resulta más frecuente.

Por ejemplo, si tenemos diez modelos predictivos que clasifican una instancia entre A, B y C, con los siguientes resultados:

Modelo	Puntaje A	Puntaje B	Puntaje C	Clase
1	0,2	0,7	0,1	B
2	0,1	0,1	0,8	C
3	0,2	0,3	0,5	C
4	0,1	0,8	0,1	B
5	0,2	0,6	0,2	B
6	0,6	0,3	0,1	A
7	0,25	0,65	0,1	B
8	0,2	0,7	0,1	B
9	0,2	0,2	0,8	C
10	0,4	0,3	0,3	A

Tabla 5.1: Ilustración de Majority voting (Resultados de los modelos)
Fuente: Elaboración propia

	A	B	C
Votos	2	5	3

Tabla 5.2: Ilustración de Majority voting (Votos por clase)
Fuente: Elaboración propia

Como la clase que más se repite es B para la instancia analizada esta será clasificada como B.

3.2. HERRAMIENTAS TECNOLÓGICAS

Para el desarrollo de este proyecto se utilizará el software R, una herramienta open source basada, principalmente en lenguaje C, que posee un gran potencial para el manejo de vectores, matrices y aplicación de algoritmos. Una

de las mayores ventajas de R es que al ser un lenguaje open source, los usuarios son libres de crear sus propios paquetes con funciones que luego son dispuestos públicamente para ser usados por todos los usuarios. Así, gran cantidad de modelos estadísticos y de data mining ya se encuentran programados, y están disponibles en algún paquete.

Por otra parte, los datos se encuentran disponibles en un entorno Big Data Hadoop, por lo que se utiliza la interfaz Hue para la visualización de los datos. Además de los intérpretes SQL, Hive e Impala, que posee Hadoop para realizar consultas SQL sobre datos en entornos Big Data. Lo que permite la extracción y manipulación de los datos requeridos. Todo esto sobre la arquitectura Big Data proporcionada por Everis¹⁵.

4. APLICACIÓN DE TÉCNICAS DE DATA MINING

4.1. RECOLECCIÓN Y PREPARACIÓN DE LOS DATOS

4.1.1. CARACTERIZACIÓN Y ESTRUCTURA DE LOS DATOS

La institución bancaria puso a disposición de Everis una base de datos que contiene todas las transacciones realizadas el día 02 de Abril de 2015 por los empleados del banco en el sistema computacional interno de la institución. Todas estas transacciones quedan registradas en el Log de Tecleo y son almacenadas en un ambiente Big Data Hadoop.

Corresponden a más de 60 gb de datos que comprenden 21.470.112 transacciones realizadas por 10.057 usuarios distintos sobre 4.117.940 cuentas de clientes distintas.

Cada registro corresponde a una transacción realizada por un usuario sobre una cuenta particular, con una serie de datos sobre el usuario que las realizó, el tipo y características de la transacción, entre otros.

¹⁵ En el Anexo I se puede observar los softwares y plataformas de la arquitectura Big Data y Analytics de Everis. El Anexo J muestra las funciones asociadas que cumple cada parte de esta arquitectura. En rojo se encuentran las herramientas utilizadas.

Entre las variables disponibles se encuentran¹⁶:

- Fecha de la transacción: La fecha en que se llevó a cabo la transacción. Sólo se dispone de los datos de transacciones para el día 02 de Abril de 2015.
- Hora de la transacción: Hora del día en que se realizó la transacción con hora y minuto.
- Tipo de transacción: Código asociado a un tipo de transacción particular.
- Usuario: Código de usuario del empleado que realizó la transacción.
- Perfil del usuario: Asociado al cargo que posee el usuario. Entre estos está Ejecutivo de Cuentas, Cajeros, Analista, entre otros 108 perfiles.
- Variables asociadas a la sesión: Una serie de variables asociadas a la conexión del usuario al sistema.

Si bien existen una serie de otras variables como el terminal desde donde se realiza la transacción, estas no se pudieron utilizar debido a que no se contaba con el apoyo de expertos del banco que pudieran explicar su significado y valores.

4.1.2. CARACTERIZACIÓN DE LOS REGISTROS DE FRAUDE

La institución bancaria proporcionó exclusivamente 5 registros que son considerados, según expertos del banco, sospechosos de haber cometido algún tipo de fraude interno, que no se detalló de mayor forma. No se tiene mayor información de qué es lo que hace sospechoso al conjunto de transacciones que constituyen los fraudes proporcionados.

Cada registro de fraude corresponde a una serie de transacciones llevadas a cabo por un usuario, dentro del mismo día, sobre una cuenta en específico,

¹⁶ En el Anexo K se puede observar una captura de pantalla de la base proporcionada por el banco desplegada en la plataforma Hue a través de una consulta SQL con el intérprete Hive.

que a ojos de los expertos del banco resultan sospechosas. Y que se concretan, además, con un cobro de cheque, generalmente realizado por otro usuario.

Cabe recalcar que una transacción por sí sola no es constitutiva de fraude. Es un conjunto de transacciones sobre la cuenta que presentan un patrón sospechoso.

Los registros de fraudes son llevados a cabo por 4 usuarios, ya que dos casos de fraude los realizó la misma persona, y no se corresponden con la fecha de la base proporcionada del Log de Tecleo. Estos registros de fraude se llevaron a cabo todos en distintas fechas comprendidas entre el 27 de Mayo y el 15 de Junio de 2015¹⁷.

Las variables presentes en los registros de fraude proporcionados por el banco son:

- Fecha de la transacción: Fecha en que se llevó a cabo la transacción sobre la cuenta.
- Hora de la transacción: Hora en que se realizó la transacción.
- Oficina: Código único de la oficina en que se realizó la transacción.
- Nombre oficina: El nombre de la oficina asociada al código.
- Terminal: No se posee información sobre esta variable.
- Operador: Código único del empleado para registrar sus transacciones en el sistema.
- Rut operador: Rut del usuario.
- Nombre operador: El nombre del operador que llevó a cabo las transacciones.

¹⁷ En esta memoria se muestran exclusivamente datos agregados del comportamiento de los usuarios fraudulentos, y no los registros detallados del fraude, para no exponer información que puede resultar sensible para el banco.

- Transacción: Código único del tipo de transacción llevada a cabo
- Glosa Transacción: Nombre de la transacción. Por ejemplo, consulta de saldos.

4.1.3. SELECCIÓN DE LOS DATOS

Como los datos se encuentran en un ambiente Big Data de Hadoop, como tipo de archivo HDFS, debieron ser exportados por medio de los intérpretes SQL Hive e Impala, para así poder trabajar con bases de dato de tipo relacional (RDBMS) en R.

Debido a que los tipos de transacciones presentes en los casos de fraude son limitados, y resultaría inútil hacer análisis sobre otro tipo de transacciones debido a que no se podría comparar, se consideraron exclusivamente las transacciones presentes en dichos casos.

Los códigos de estas transacciones y la descripción asociada son la siguiente:

- B413: Consulta de datos básicos de una cuenta.
- KC01: Consulta de saldos.
- Z401: Consulta de últimos movimientos.
- PE47: Localización de Personas.
- B431: Consulta de talonarios de una cuenta.
- B412: Cabecera estándar de una cuenta.
- KP06: Producción y subida dada la operación.
- B433: Consulta de cheques de talonario.
- B741: Mantenimiento de órdenes de no pago.

Si bien, algunas de estas descripciones se explican por sí mismas para caracterizar el tipo de transacción, no se contó con la información por parte del banco que caracterice de mejor forma estas transacciones.

Para los registros que realizaban las transacciones presentadas anteriormente se exportaron los datos considerando las variables hora de la transacción, código de la transacción, perfil del usuario, código del usuario y el número de cuenta sobre el que se realizó la transacción. Llegando a una base de datos con 3.017.229 registros de transacciones.

4.2. PRE-PROCESAMIENTO DE LOS DATOS

4.2.1. LIMPIEZA DE LOS DATOS

Con respecto a la limpieza de los datos la base generada presenta dos principales problemas con los cuales se debe lidiar: Outliers y datos faltantes de perfil.

Con respecto a los outliers existen usuarios que presentan más de 100.000 transacciones diarios, y 8.780 poseen sólo 1 transacción. Probablemente el primer caso se deba a transacciones automáticas llevadas a cabo por robots que, por supuesto, no se quieren considerar en el análisis.

Quitar outliers, cuando se utilizan metodologías de outliers detection, resulta bastante sensible, ya que no se quieren remover datos que potencialmente serían sospechosos de fraude. Pero, tampoco se desea indicar como sospechosos datos que en realidad son sólo una anomalía, o se deben a procesos diferentes. Esto porque después repercutiría en un esfuerzo posterior de investigación de los casos sospechosos.

Por esto, se optó por quitar el 5% de los usuarios con mayor número de transacciones (aquellos registros que están por sobre el percentil 95) y el 5 % con menor número de transacciones. De un total de 8.780 usuarios se quitaron 441 de la cola superior de la distribución y 476 de la cola inferior. Los primeros porque poseían más de 283 transacciones diarias, y los segundos por tener menos de 2.

Por otra parte, existen registros que no poseen asociado un perfil de usuario. Este campo viene en blanco. Esto según fue explicado por expertos del banco se debe a que son procesos automáticos llevados a cabo por máquinas. Por lo anterior, se consideró sólo los registros que poseían valores en todas sus variables (Análisis de casos completos). Con esto se removieron 87.337 transacciones o registros.

Así, finalmente, la base quedó constituida por 630.226 registros de transacciones, de 7.541 usuarios distintos, sobre 80.897 cuentas distintas. Esto sin considerar aún los registros de fraude.

4.3. TRANSFORMACIÓN DE LOS DATOS

4.3.1. GENERACIÓN Y SELECCIÓN DE VARIABLES

Como el objetivo del trabajo es determinar usuarios sospechosos de estar cometiendo fraude interno, y la data viene dado como registro de transacciones, se deben generar variables que representen el comportamiento del usuario durante el día.

Para ello lo primero fue establecer como entidad de la base a los usuarios del banco, en particular el código único de usuario. Luego se generaron variables de frecuencia para cada una de los tipos de transacciones analizadas, contabilizando el número de veces que un usuario realiza cada tipo de transacción en el estudiado. Así se obtienen 9 variables de frecuencia, correspondientes a los 9 tipos de transacciones.

Otra variable generada, a partir de la observación de los registros de fraudes, es el tiempo promedio transcurrido entre las transacciones que realiza el usuario durante el día. Esto debido, a que las transacciones presentes en los casos de fraude se suceden con intervalos de tiempo muy pequeños (menores a 2 minutos), lo que hace presumir que estos pueden tener un tiempo promedio entre transacciones menor al del resto de los usuarios. Por lo anterior, se genera una variable que representa el tiempo promedio en minutos transcurrido entre las transacciones que genera el usuario en el día.

A continuación, se muestran los datos agregados de las transacciones diarias llevadas a cabo por los usuarios que cometieron fraude:

Variable	User id				
	fraude CCON818	fraude JTAPIAL	fraude CCON243(1)	fraude CCON643	fraude CCON243
Datos básicos	1	1	2	4	24
Saldos	0	0	0	1	0
Ult. Movimientos	1	0	1	1	1
Loc. Personas	3	5	3	4	24
Talones	0	2	0	1	0
Cabecera Cuenta	0	1	0	1	0
Prod. y Sub. de OP.	1	0	2	0	0
Cheques	0	1	0	0	0
Ordenes no pago	0	1	0	0	0
Tiemp trans.	1	5,211	0	1	8,99
Perfil	AUDIOM	AUDIOM	AUDIOM	AUDIOM	AUDIOM

Tabla 6: Valores de las variables para los 5 registros de fraude
Fuente: Elaboración propia

Debido a que los modelos a utilizar no pueden manejar variables categóricas, estas se transformaron en 11 variables binarias, correspondientes a los 10 perfiles que concentran más usuarios y otra variable para el resto, que toman el valor 1 si el usuario pertenece a la categoría de la variable y 0 sino.

$$perfil_{i,j} = \begin{cases} 1 & \text{si el usuario } i \text{ pertenece al perfil } j \\ 0 & \text{sino} \end{cases}$$

Perfil	N° de usuarios
EJECTAS	2.013
CAJASTL	1.074
AGENTES	561
AUDIOM	459
ATPUBS	429
SUPCATL	405
TESOR	368
BFEJECOM	308
GSENIOR	203
GRCEJE	187
Resto	1.539

Tabla 7: 10 perfiles más frecuentes y número de usuarios
Fuente: Elaboración propia

Pero, estas variables de perfil posteriormente no serían utilizadas en los modelos, para dar paso a un análisis posterior del perfil como se explica en la sección 4.4.

4.3.2. ANÁLISIS PRELIMINAR DE LAS VARIABLES

Para ver qué tan relevante resultarán las variables generadas a la hora de implementar los modelos y discriminar si un usuario tiene un comportamiento fraudulento, resulta interesante contrastar el comportamiento de ellas entre los usuarios defraudadores y el resto.

En los histogramas mostrados a continuación se expone la densidad del N° de transacciones para las distintas variables, el valor que tiene cada registro de fraude, la media de los fraudes y la media de los datos sin clasificar (datos del Log de Tecleo). El mismo ejercicio se repite con la variable que mide tiempo promedio entre transacciones.

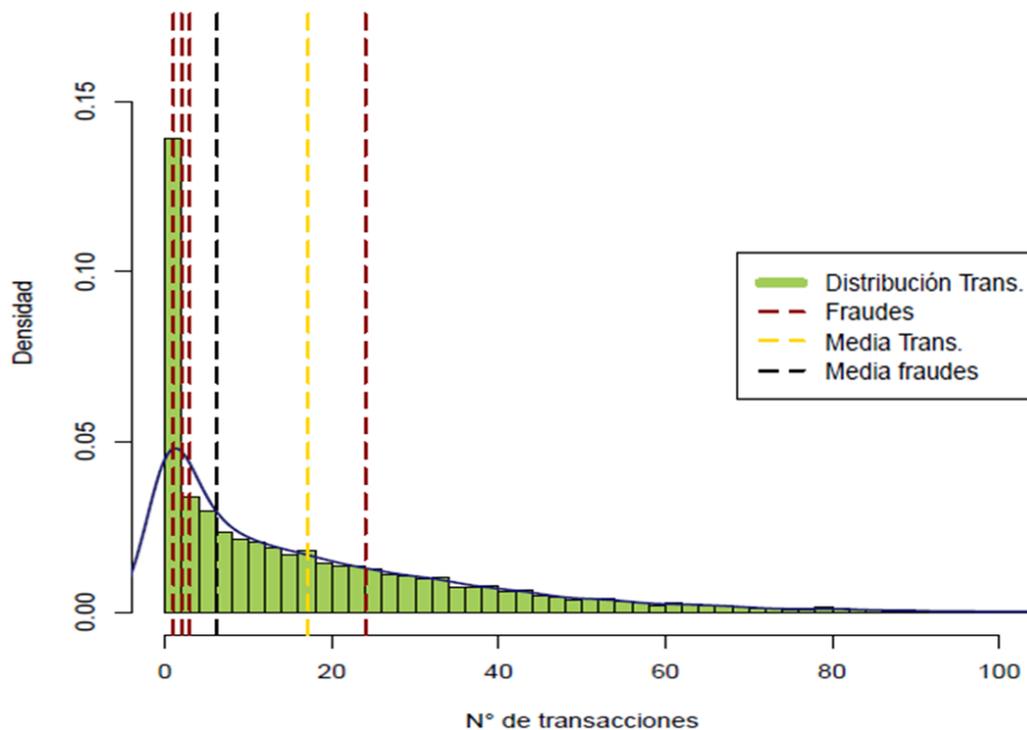


Figura 6.1: Densidad Consulta de datos básicos
Fuente: Elaboración propia

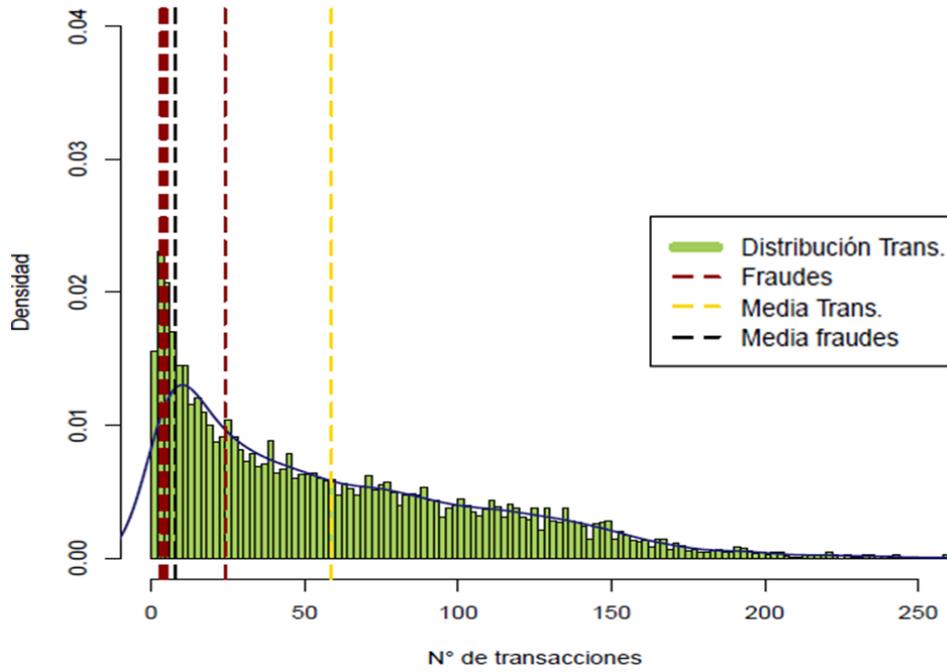


Figura 6.2: Densidad Localización de personas
 Fuente: Elaboración propia

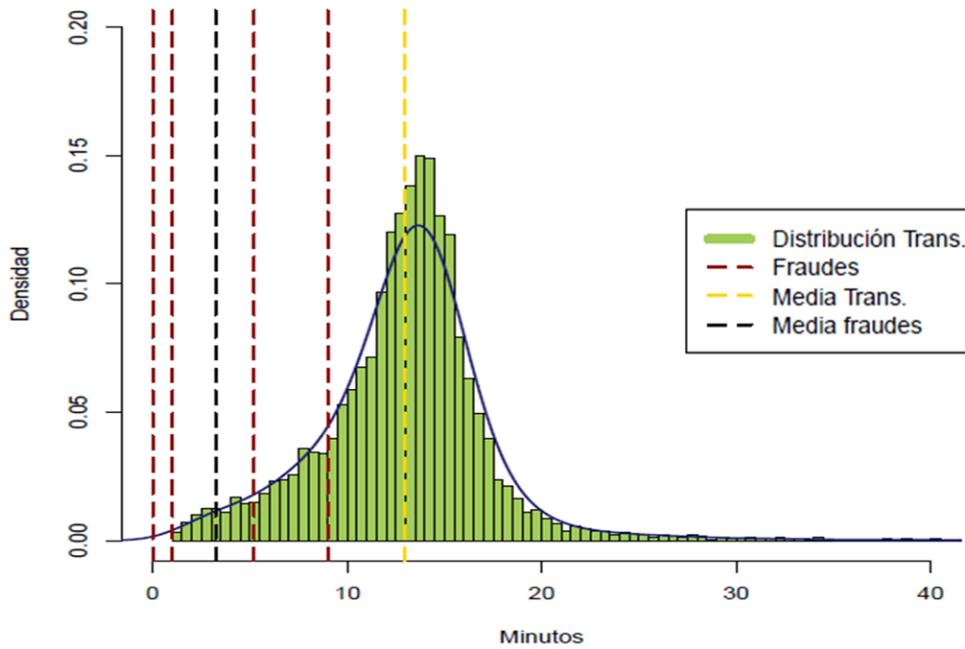


Figura 6.3: Densidad Tiempo promedio entre transacciones
 Fuente: Elaboración propia

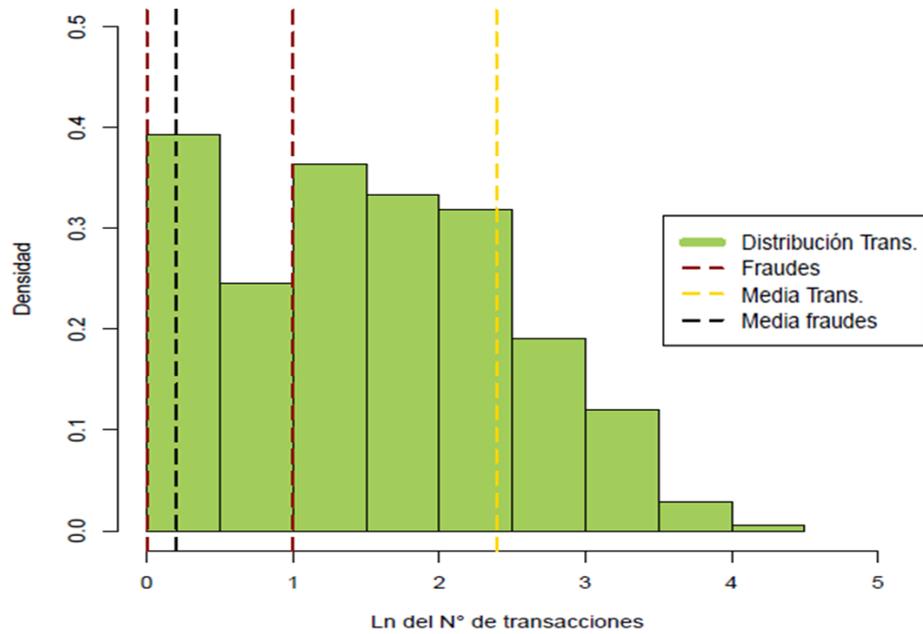


Figura 6.4: Densidad de Consulta de Saldos (Se aplicó Logaritmo natural porque los datos estaban muy concentrados)
Fuente: Elaboración propia

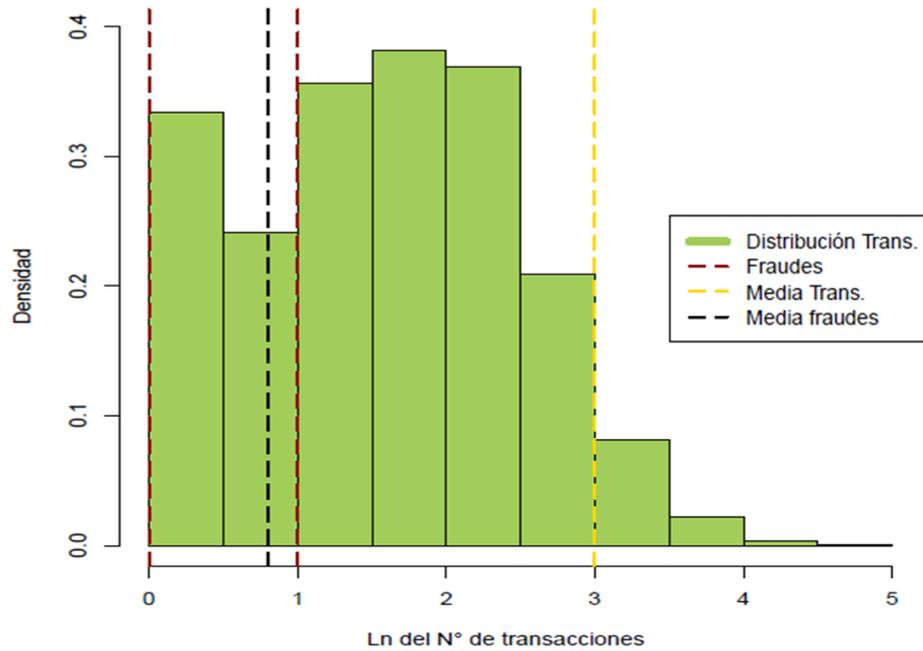


Figura 6.5: Densidad de Consulta de últimos movimientos (Se aplicó Logaritmo natural porque los datos estaban muy concentrados)
Fuente: Elaboración propia

Se puede observar que, para todas las transacciones expuestas en los gráficos anteriores, los usuarios fraudulentos realizan menos transacciones que el resto. Exceptuando un caso de fraude en que se realizan más consultas de datos básicos de la cuenta que la media del resto de los usuarios. Esto puede deberse a que los usuarios cuando comenten fraude pretenden dejar el menor registro de actividad con el fin de no ser fiscalizados.

En los siguientes gráficos de distribución con respecto a pares de variables se puede corroborar que en general los fraudes presentan medias menores que el resto de los datos:

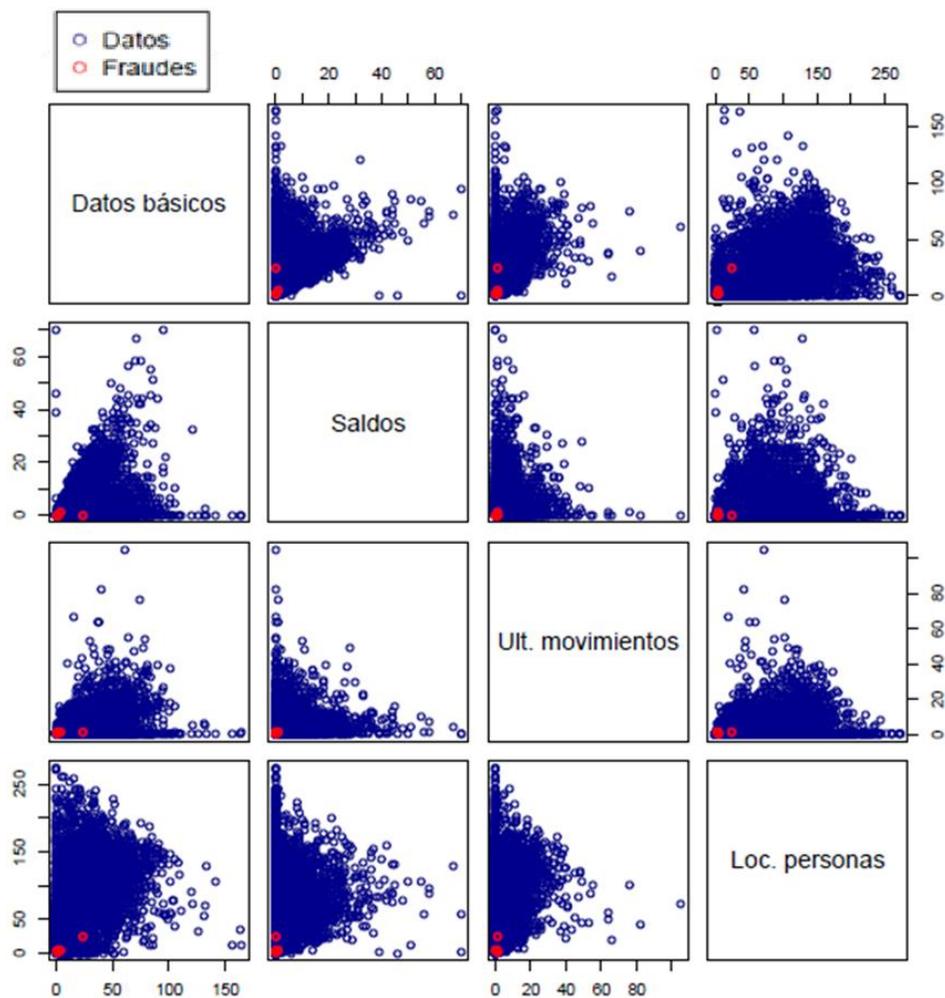


Figura 7.1: Distribución de las frecuencias de transacciones
Fuente: Elaboración propia

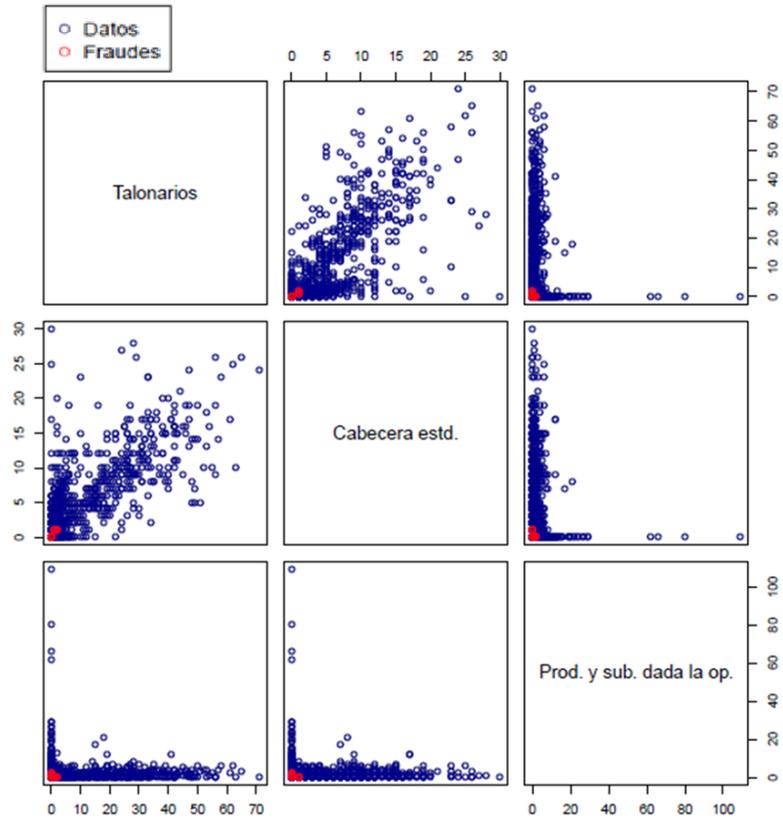


Figura 7.2: Distribución de las frecuencias de transacciones
Fuente: Elaboración propia

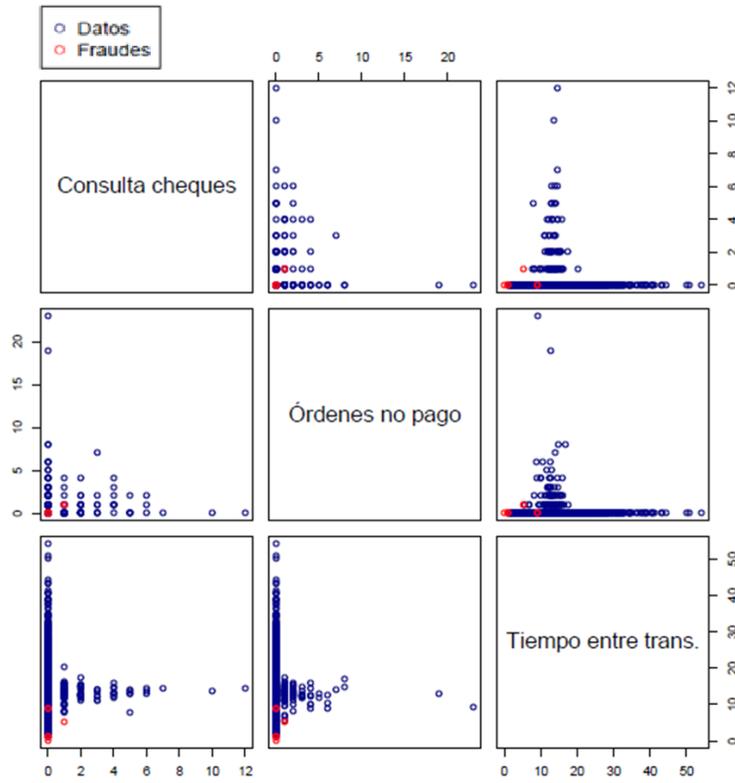


Figura 7.3: Distribución de las frecuencias de transacciones
Fuente: Elaboración propia

Otra explicación podría ser que se está comparando a tipos de usuarios distintos. Ya que todos los usuarios que cometieron fraude poseen perfil AUDIOM, y este comportamiento de tener menos transacciones puede ser propio de ese perfil. Por tanto, se estaría comparando con otros perfiles que pueden tener un promedio de transacciones más altas durante el día.

Para descartar este efecto se grafican las densidades exclusivamente para los usuarios con perfil AUDIOM:

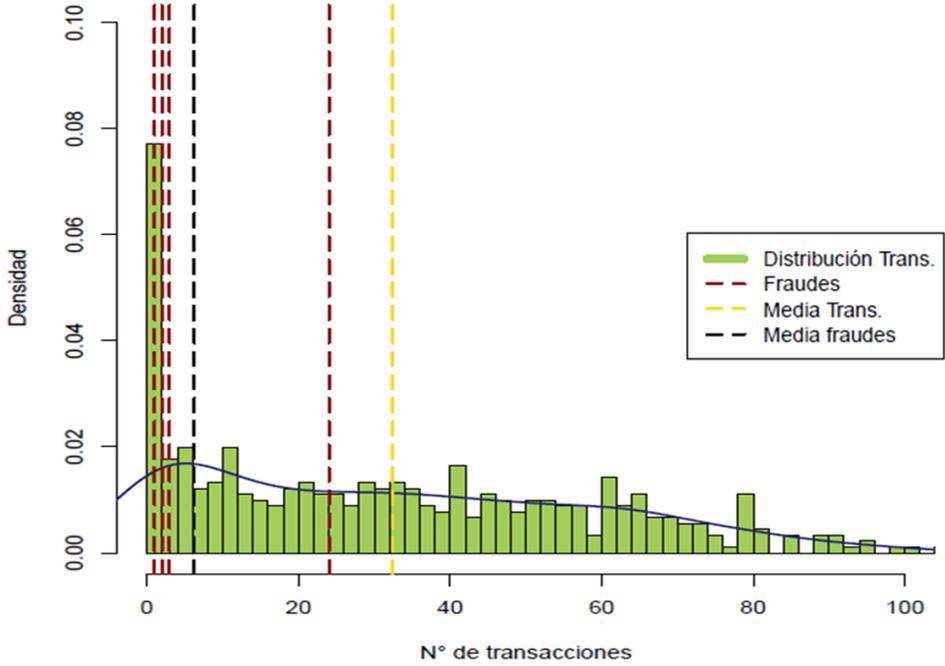


Figura 8.1: Densidad Consulta de datos básicos (Usuarios AUDIOM)
Fuente: Elaboración propia

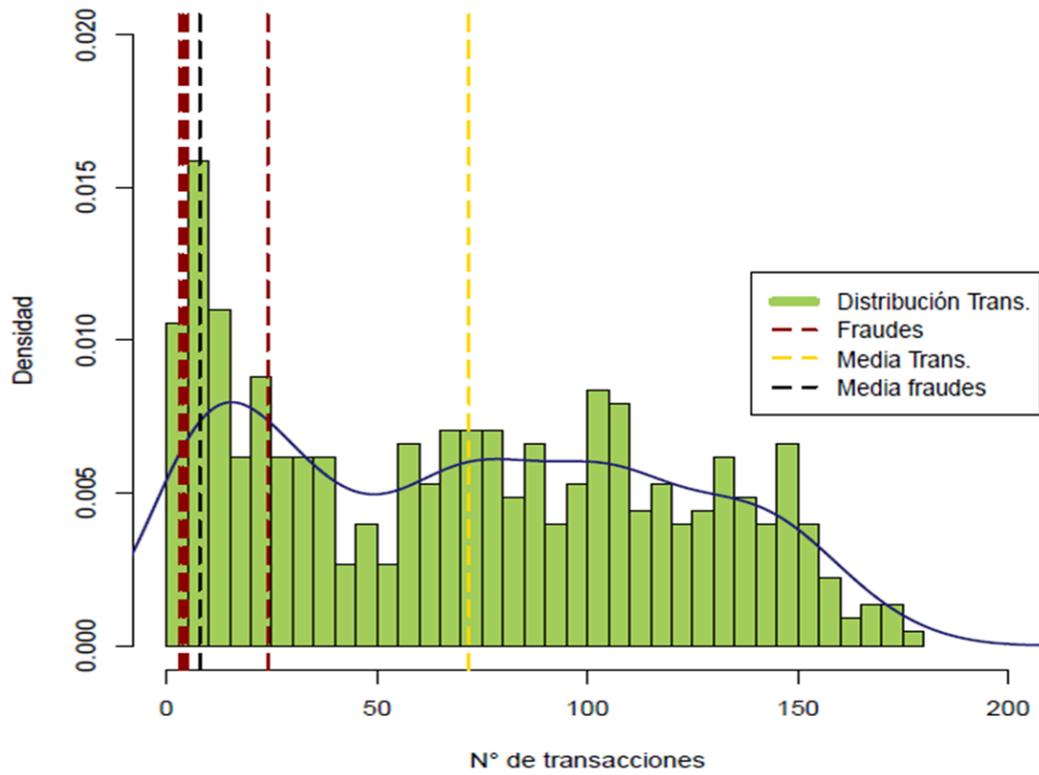


Figura 8.2: Densidad Localización de personas (Usuarios AUDIOM)
Fuente: Elaboración propia

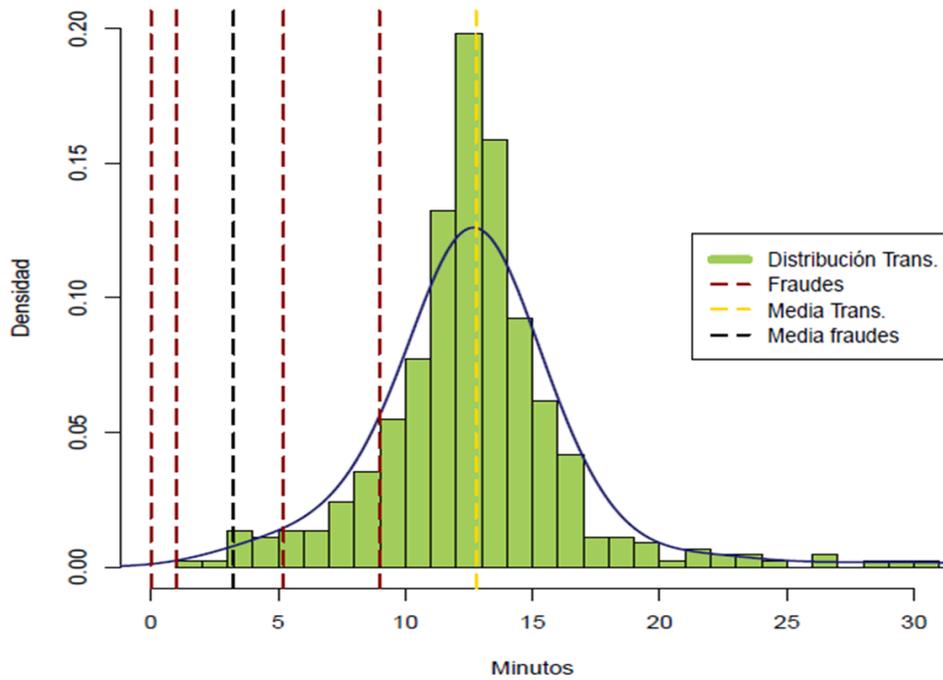


Figura 8.3: Densidad Tiempo promedio entre transacciones (Usuarios AUDIOM)
Fuente: Elaboración propia

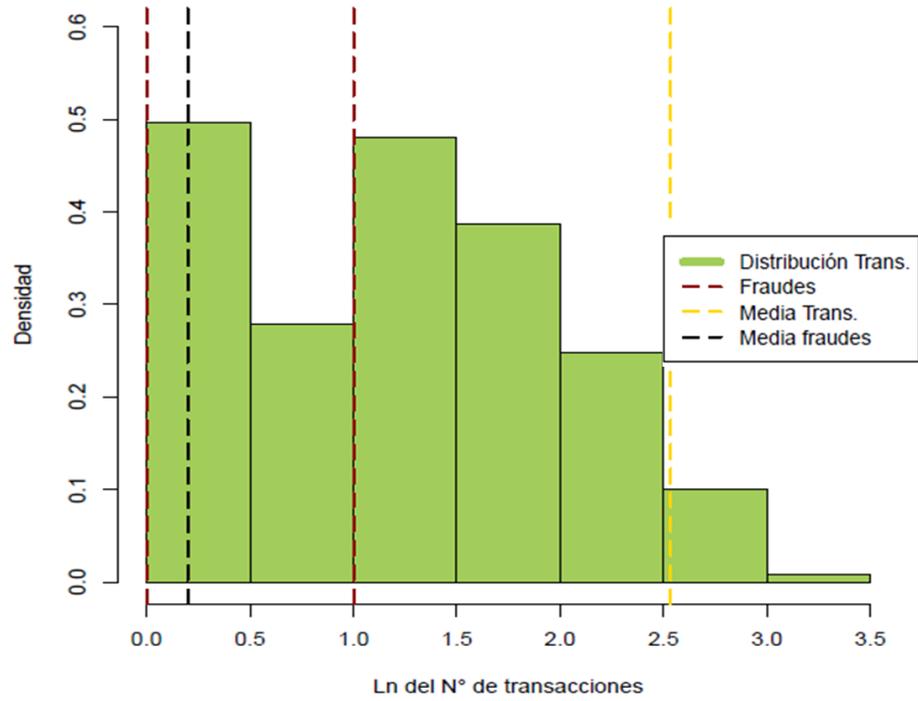


Figura 8.4: Densidad de Ln de Consulta de Saldos (Usuarios AUDIOM)
Fuente: Elaboración propia

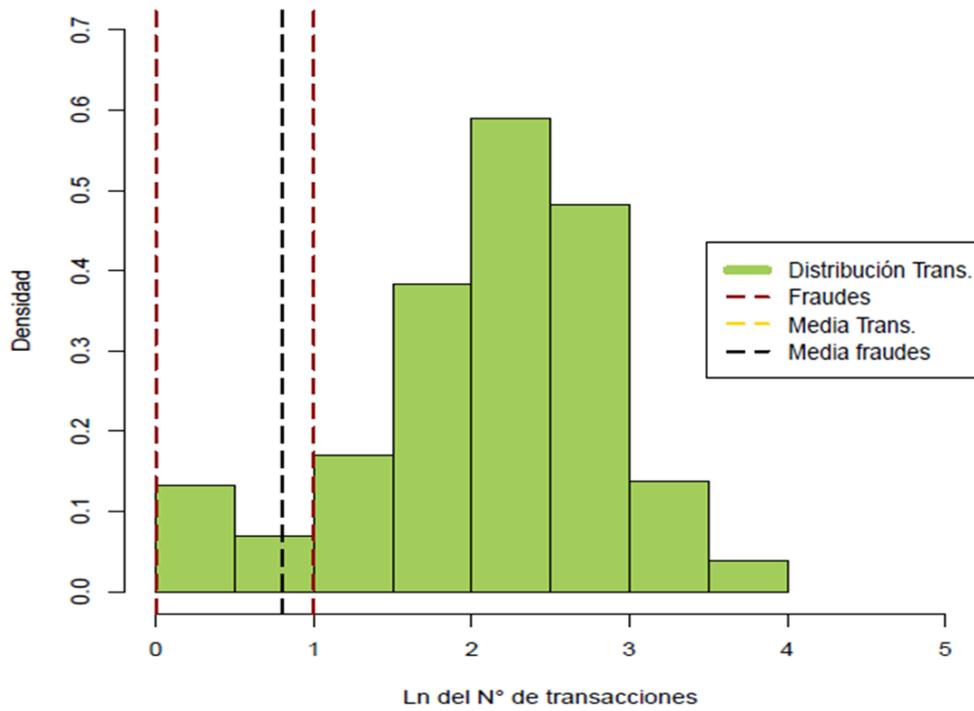


Figura 8.5: Densidad de Ln de Consulta de últimos movimientos (Usuarios AUDIOM)
Fuente: Elaboración propia

Vemos que se sigue manteniendo la tendencia con respecto a la comparación con todos los usuarios. Los registros de fraudes poseen menos transacciones diarias que el resto de los usuarios que tienen su mismo perfil.

4.3.3. NORMALIZACIÓN DE VARIABLES

Como se utilizarán métodos de detección de outliers que trabajan con medidas de distancia y densidad, resulta recomendable escalar las variables de forma que aquellas que poseen medias mayores al resto, no se vean sobrestimadas al momento de calcular la distancia. De no escalar las variables, los resultados de cercanía, al medir distancia euclidiana, por ejemplo, se verían muy influenciados por las variables que poseen más varianza y magnitud.

Para no generar este problema se escalan las variables con la transformación Min-Max, de forma que queden todas con rangos entre 0 y 1. Esta transformación además asegura que se mantenga la forma original de la distribución de la variable. Así las nuevas variables vienen dadas por:

$$x_{nuevo} = \frac{x - \min(x)}{\max(x) - \min(x)} (nuevomax - nuevomin) - nuevomin$$

Donde nuevomin (el nuevo valor mínimo) sería 0 y nuevomax (el nuevo valor máximo) sería 1.

4.4. DATA MINING

En este apartado se optó por usar distintos modelos no supervisados, pues estos no requieren que los datos vengan categorizados como anómalos o no. Ya que sólo se cuenta con 5 casos que se sabe son fraudulentos, este enfoque entrega una herramienta para lidiar con el problema de detección de otros posibles fraudes sin conocimiento a priori.

Por otra parte, estos modelos suelen desempeñarse bien para descubrir datos que se desvían de la media o comportamiento promedio de los demás, además de permitir un análisis más descriptivo.

Finalmente, a partir de los tres modelos, escogidos se consolidará una lista de individuos sospechosos.

4.4.1. ANÁLISIS DE VECINOS MÁS CERCANOS

El primer enfoque para detectar usuarios que tengan patrones de comportamiento sospechosos y similares a los fraudes, es el análisis de los vecinos más cercanos. Este permite establecer los puntos que poseen mayor similitud con estos, a partir de una medida de distancia.

La medida de distancia a utilizar será la distancia euclidiana que se define como:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Con X e Y vectores de n variables.

Para escalar las variables a rangos similares, como se mencionó anteriormente, se utilizó la transformación Min-Max con el fin de que todas las variables quedaran acotadas entre 0 y 1. De forma que la relevancia que tuvieran a la hora del cálculo de la distancia euclidiana no se viera afectada por la magnitud de la variable.

En un comienzo se incluirían en la implementación del algoritmo las variables binarias generadas para los perfiles de usuario. Pero, como estas sólo toman valores 0 y 1, en general aportan en 1 a la distancia euclidiana. Mientras que las otras variables, al estar escaladas poseen valores decimales y se ven subestimadas frente a estas.

Por ello al correr el algoritmo considerando las variables binarias de perfil el algoritmo arroja como resultado que todos los vecinos más cercanos a los fraudes corresponden a perfiles AUDIOM. Es decir, sobrestima la variable. Lo

que no proyecta, realmente, una similitud entre el comportamiento de los usuarios.

Debido a lo anterior, es que se decidió prescindir de estas variables de perfil y correr el algoritmo exclusivamente con las variables de frecuencia de las transacciones, y el tiempo promedio entre transacciones. Posteriormente se desarrollará un análisis para ver si el comportamiento de los usuarios sospechosos en realidad es propio de su perfil.

Para cada usuario fraudulento, utilizando las variables anteriormente mencionadas se calcularon las distancias euclidianas a todos los usuarios, y se extrajeron los 30 usuarios que poseen una menor distancia (vecinos más cercanos). Se escogió este número con el fin de tener un número de sospechosos suficientes para cruzar posteriormente con los sospechosos arrojados por los otros modelos utilizados.

Así tenemos 5 listas de 30 sospechosos cada una¹⁸.

En las siguientes tablas podemos observar las medias y desviaciones estándar para los grupos de sospechosos, con respecto a los datos y los fraudes (sin considerar para las medias el caso CCON243 que posee una clara desviación del comportamiento de los demás fraudes):

Grupo	Datos básicos		SalDOS		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	0,63	0,96	0,00	0,00	0,23	0,50	3,43	1,45	0,00	0,00	0,00	0,00	0,20	0,55	0,00	0,00	0,00	0,00	1,63	0,40

Tabla 8.1: Medias y Desv. Std. de los vecinos más cercanos al fraude CCON818

Fuente: Elaboración propia

Grupo	Datos básicos		SalDOS		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	0,37	0,67	0,00	0,00	0,03	0,18	4,87	1,61	0,00	0,00	0,03	0,18	0,03	0,18	0,00	0,00	0,00	0,00	5,15	0,43

Tabla 8.2: Medias y Desv. Std. de los vecinos más cercanos al fraude JTAPIAL

Fuente: Elaboración propia

¹⁸ Los Anexos L.1, L.2, L.3, L.4 y L.5 contienen las listas de sospechosos obtenidas por el método de los vecinos más cercanos.

Grupo	Datos básicos		SalDOS		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	0,83	1,18	0,00	0,00	0,23	0,50	3,40	1,71	0,00	0,00	0,00	0,00	0,27	0,69	0,00	0,00	0,00	0,00	1,64	0,26

Tabla 8.3: Medias y Desv. Std. de los vecinos más cercanos al fraude CCON243 (1)

Fuente: Elaboración propia

Grupo	Datos básicos		SalDOS		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	2,50	2,05	0,07	0,25	0,43	0,73	3,40	2,43	0,10	0,40	0,10	0,40	0,07	0,25	0,00	0,00	0,00	0,00	1,75	0,37

Tabla 8.4: Medias y Desv. Std. de los vecinos más cercanos al fraude CCON643

Fuente: Elaboración propia

En general se puede observar que los grupos de sospechosos poseen un mayor parecido con el grupo de fraudes en las variables "Tiempo entre transacciones" y "Localización de personas". Mientras que difieren en las variables "Cabecera estándar de una cuenta", "Consulta de cheques de talonario" y "Mantenimiento de órdenes de no pago"¹⁹.

A partir de este análisis surgieron usuarios que aparecen como vecino más cercano de más de un caso de fraude. Estos usuarios resultan más sospechosos y vale la pena analizarlos.

La siguiente tabla posee los usuarios que aparecen como vecino más cercano de dos o más casos de fraude:

¹⁹ En los Anexos M.1, M.2, M.3, M.4 y M.5 se pueden observar las distancias de cada uno de los sospechosos a la media de los fraudes.

User_jd	Datos básicos		Saldos		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo trans.	Perfil	N° Apariciones
CRIVERP	3	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2,00	EJECTAS	3	
HSEPULT	2	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1,62	GSENIOR	3	
DMARGAA	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1,33	GSENIOR	3	
EMUNOZG	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,56	GSENIOR	3	
FLYONAU	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1,75	NEGINTE	3	
VT18093	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1,67	FUERVEN	3	
FVTA232	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1,80	FUERVEN	3	
CBAHAMO	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1,94	GSENIOR	3	
TE01299	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1,13	TESOR	3	
AG05297	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2,00	AGENTES	3	
TE01213	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2,00	TESOR	3	
AG02032	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1,37	AGENTES	3	
AG02157	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1,40	AGENTES	3	
EJ01866	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1,42	GRCEJE	3	
PDEMEDP	4	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1,72	ANALSRIE	2	
EJOU052	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,83	AUDIOM	2	
EJ17170	0	0	0	3	0	0	0	0	0	1	0	0	0	0	0	0	0	1,38	GSENIOR	2	
DLOPEZV	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1,47	FMLOCAL	2	
N596659	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1,50	TESOR	2	
NREYE01	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1,46	SCLBTR	2	
PALVARM	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1,46	EJECTAS	2	
JREYESI	0	0	0	3	0	0	0	0	2	0	0	0	0	0	0	0	0	1,87	ATPUBS	2	
TE02198	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1,53	TESOR	2	
AC01004	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	1,88	EJECTAS	2	
CB03860	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1,74	GRCCAS	2	
RPARRAP	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1,82	EJECTAS	2	
EFIGUEC	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1,81	EJECTAS	2	

Tabla 9: Usuarios que son vecinos cercanos de 2 o más fraudes
Fuente: Elaboración propia

Se debe poner especial atención en aquellos usuarios que poseen varias apariciones, como vecinos de más de un caso de fraude, y que poseen usuario AUDIOM.

Grupo	Datos básicos		Saldos		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	0,89	1,22	0,00	0,00	0,22	0,51	3,22	1,63	0,00	0,00	0,00	0,00	0,15	0,46	0,00	0,00	0,00	0,00	1,65	0,24

Tabla 10: Medias y Desv. Std. de los vecinos más cercanos a 2 o más fraudes
Fuente: Elaboración propia

Al igual que en los casos individuales la lista de sospechosos resultante del cruce anterior posee similitudes con los fraudes principalmente en las variables "Tiempo entre transacciones" y "Localización de personas".

4.4.2. LOCAL OUTLIER FACTOR (LOF)

El segundo modelo utilizado para identificar sospechosos de haber cometido fraude interno es local outlier factor. Un modelo para detectar outliers que identifica puntos que poseen una desviación significativa en su densidad con respecto a su vecindad.

Como input del algoritmo se usaron las variables de frecuencia de cada una de las transacciones y los tiempos promedio entre transacciones, ya escaladas con la transformación Min-Max para acotar sus valores entre 0 y 1.

El algoritmo entrega un factor para cada uno de los 7.546 usuarios (incluidos los fraudes) que, para este caso, posee valores entre 0,867 y 7,428. Este último valor corresponde justamente a uno de los registros de fraude, JTAPIAL.

Ahora bien, a partir de este factor entregado por el algoritmo se debe definir un punto de corte (threshold), a partir del cual los usuarios que poseen un factor LOF mayor al punto de corte son considerados outliers locales.

Para definir este threshold recurrimos, en este caso, a maximizar el lift del modelo.

El lift es una medida del ajuste de un modelo de clasificación, que consiste en medir cuanto mejor es la respuesta del modelo al clasificar, fraudes en este caso, con respecto a la distribución existente en la población. Es, básicamente, la razón entre los casos de fraudes captados sobre los clasificados como sospechosos, divididos por la proporción de fraudes presentes en la población.

Para este caso correspondería a:

$$lift = \frac{N^{\circ} \text{ fraudes clasificados como outliers por LOF} / N^{\circ} \text{ de outliers que entrega LOF}}{N^{\circ} \text{ de usuarios fraudulentos en la base} / N^{\circ} \text{ de usuarios totales}}$$

La fracción de razón del denominador es un número fijo (5/7.546), mientras que la razón que se ubica en el numerador variará dependiendo el threshold utilizado.

Se definieron varios escenarios de número de outliers clasificados por el algoritmo, mediante la variación del threshold, con el fin de encontrar un escenario óptimo. Que posea un gran lift, y detecte como outliers una cantidad razonable de fraudes.

En la tabla que se muestra a continuación se disponen los distintos escenarios, con el número de outliers clasificados, el número de fraudes clasificados como outliers, el punto de corte para el factor LOF, el lift y el porcentaje de outliers sobre el total de datos.

Corte	Outliers encontrados	Fraudes encontrados	Fraudes/Outliers	Lift	%Outliers datos
3,429	10	1	10,00%	150,92	0,133%
2,784	20	1	5,00%	75,46	0,27%
2,598	30	1	3,33%	50,31	0,40%
2,385	40	1	2,50%	37,73	0,53%
2,319	50	1	2,00%	30,18	0,66%
2,238	60	1	1,67%	25,15	0,80%
2,161	70	2	2,86%	43,12	0,93%
2,118	80	3	3,75%	56,60	1,06%
2,062	90	3	3,33%	50,31	1,19%
2,019	100	3	3,00%	45,28	1,33%
1,955	125	3	2,40%	36,22	1,66%
1,888	150	3	2,00%	30,18	1,99%
1,821	175	3	1,71%	25,87	2,32%
1,785	200	3	1,50%	22,64	2,65%
1,72	225	3	1,33%	20,12	2,98%
1,685	250	3	1,20%	18,11	3,31%
1,656	275	3	1,09%	16,46	3,64%
1,632	300	3	1,00%	15,09	3,98%
1,604	325	4	1,23%	18,57	4,31%
1,576	350	4	1,14%	17,25	4,64%

Tabla 11: Escenarios lift-threshold para Local Outlier Factor (LOF)
Fuente: Elaboración propia

Finalmente se escoge un escenario intermedio entre el que clasifica 70 y el que clasifica 80 outliers, ya que es ahí donde el algoritmo pasa de clasificar 2 a 3 fraudes como outliers. Lo que eleva el lift, e implica clasificar un número razonable de fraudes como outliers.

Escenario elegido					
Corte	Outliers encontrados	Fraudes encontrados	Fraudes/Outliers	Lift	%Outliers datos
2,131	75	3	4,00%	60,37	0,994%

Tabla 12: Escenario escogido para Local Outlier Factor (LOF)
Fuente: Elaboración propia

Así, clasificando como outliers a aquellos usuarios que obtuvieron un valor del factor LOF mayor a 2,131, se obtienen 75 outliers, 3 de ellos son casos de fraudes, que corresponden a cerca del 1% del total de usuarios²⁰.

Para poder disponer gráficamente los resultados se realizó un análisis de componentes principales (PCA), dado que captan la mayor cantidad de varianza de los datos en las primeras componentes, se obtiene la mayor información al graficar las primeras dos y tres componentes, para dos y tres dimensiones respectivamente.

Los pesos asociados a las componentes para cada variable son los siguientes:

	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talones	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiempo trans.
Comp.1	-0,437	-0,198	-0,357	-0,349	-0,436	-0,473	-0,109	-0,232	-0,187	-0,077
Comp.2	0,384	0,476	0,243	0,314	-0,417	-0,379	0,004	-0,320	-0,178	0,130
Comp.3	-0,075	-0,208	-0,279	0,300	0,095	0,005	-0,023	0,024	-0,230	0,849
Comp.4	-0,100	0,047	-0,003	-0,066	0,090	-0,004	0,922	-0,048	-0,346	-0,053
Comp.5	0,074	-0,034	0,090	-0,002	0,269	0,064	-0,347	0,049	-0,852	-0,244
Comp.6	-0,005	-0,440	0,204	0,147	0,112	0,213	-0,001	-0,823	0,061	-0,067
Comp.7	-0,112	0,576	-0,048	-0,565	0,233	0,250	-0,106	-0,347	-0,005	0,293
Comp.8	0,108	0,258	-0,786	0,369	0,137	0,114	-0,001	-0,190	0,046	-0,314
Comp.9	0,785	-0,310	-0,249	-0,457	-0,008	-0,085	0,070	-0,012	-0,004	0,076
Comp.10	0,024	-0,022	-0,063	-0,027	-0,679	0,707	0,012	0,069	-0,169	0,010

Tabla 13: Pesos asociados a cada variable de las componentes principales
Fuente: Elaboración propia

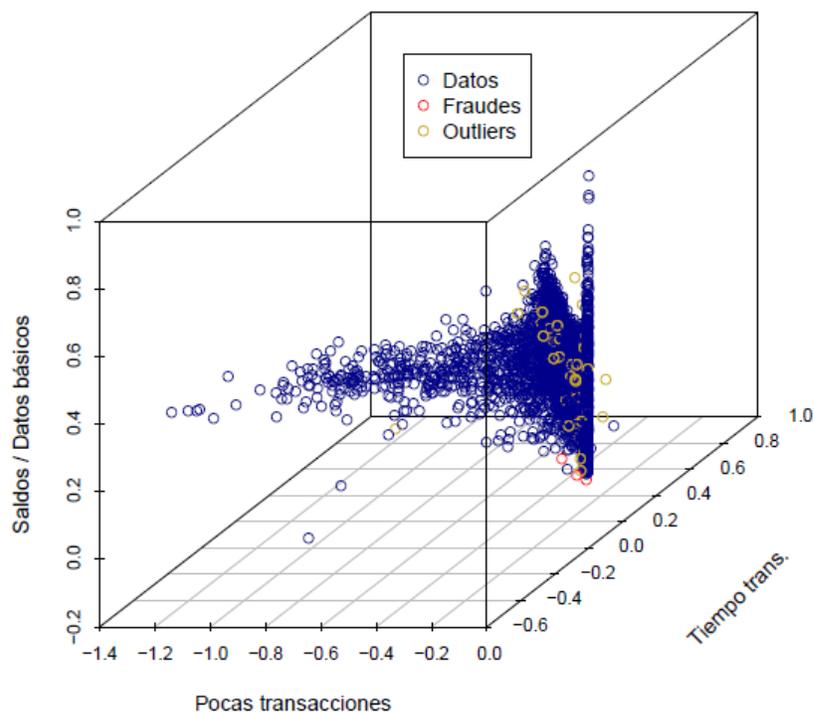
A partir de la distribución de estos pesos, se les puede dar nombre a las componentes de acuerdo a la importancia que asignan a cada variable.

²⁰ El anexo N muestra los usuarios clasificados como outliers por LOF para el escenario elegido.

Componente	Nombre
Comp.1	Pocas transacciones
Comp.2	Saldos/datos básicos
Comp.3	Tiempo trans.
Comp.4	Prod. y Sub. dada la Op.
Comp.5	Talonarios
Comp.6	Cabecera Cuenta/Ult. Movimientos
Comp.7	Saldos
Comp.8	Loc. Personas
Comp.9	Datos básicos cuenta
Comp.10	Cabecera Cuenta

Tabla 14: Nombres asociados a las componentes principales
Fuente: Elaboración propia

Así, utilizando las dos primeras y tres primeras componentes principales, se generaron gráficos en dos y tres dimensiones que detallan la distribución de los outliers arrojados por LOF y los fraudes.



These three components explain 53.41 % of the point variability.

Figura 9.1: Distribución outliers LOF según las 3 primeras componentes principales
Fuente: Elaboración propia

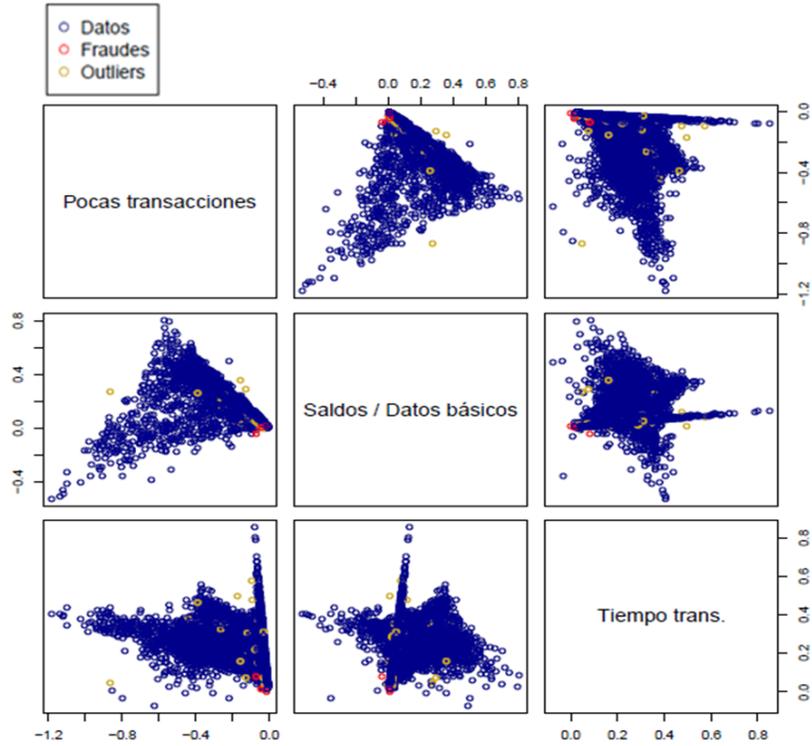


Figura 9.2: Gráficos en 2D de los resultados LOF según las 3 primeras componentes
Fuente: Elaboración propia

En el gráfico anterior se puede apreciar que los fraudes se muy cerca del origen del gráfico de las componentes "Pocas transacciones" y "Tiempo Trans.", lo que motiva la siguiente metodología.

Una vista más de cerca permite notar que, en general, los fraudes poseen pocas transacciones y sus tiempos promedios entre transacciones son pequeños.

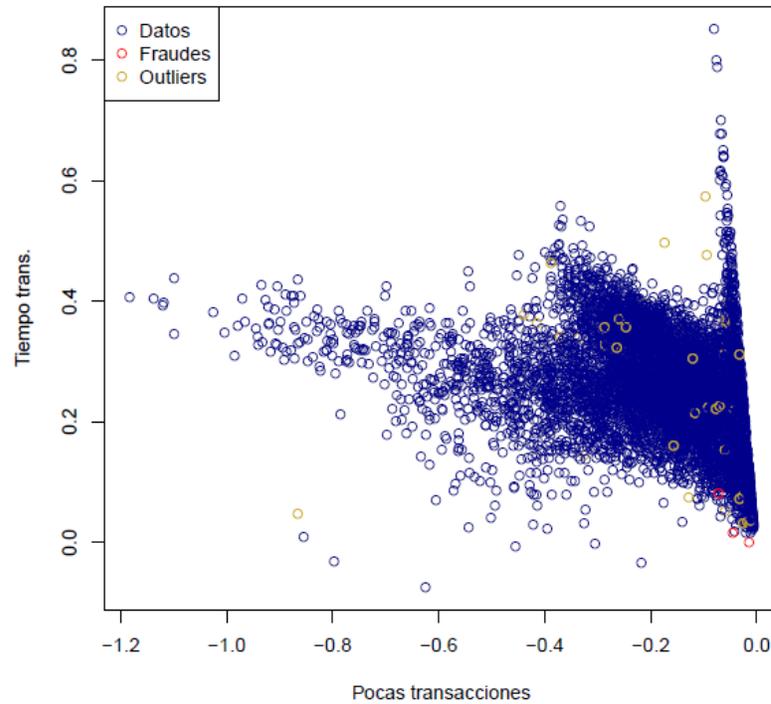


Figura 10: Resultados LOF con respecto a las componentes “Pocas transacciones” y “Tiempo trans.”
Fuente: Elaboración propia

A continuación, se muestran las medias y desviaciones estándar para los datos, los fraudes (sin considerar para las medias el caso CCON243 que posee una clara desviación del comportamiento de los demás fraudes) y los sospechosos arrojados por LOF:

Grupo	Datos básicos		Saldos		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Órdenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	9,32	12,46	1,83	7,11	2,60	9,25	69,56	72,53	1,08	2,58	1,13	3,35	1,25	4,46	0,03	0,17	0,39	2,75	12,58	6,77

Tabla 14: Medias y Desv. Std. de los sospechosos arrojados por LOF
Fuente: Elaboración propia

En este caso los sospechosos poseen un mayor parecido con los fraudes en las variables “Órdenes de no pago” y “Prod. y sub. dada la operación”, mientras que se comportan como outliers con respecto a los datos principalmente en las variables “Consulta de datos básicos” y “Localización de personas”²¹.

²¹ En el Anexo O se puede observar la distancia de cada registro sospechoso arrojado por LOF a la media de los datos.

4.4.3. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El análisis de componentes principales (PCA) es una técnica estadística que permite reducir la dimensionalidad de un conjunto de datos, por medio de la generación de variables que resultan de combinaciones lineales de las variables originales. Estas nuevas variables se llaman componentes principales.

A partir de los análisis desarrollados en el apartado anterior, se pudo encontrar que los registros de fraude se encuentran en su mayoría muy cercanos al origen de la primera y la tercera componente principal.

El fin entonces es identificar los 350 usuarios más cercanos al origen de ambos componentes. Se eligió ese número porque es un número lo suficientemente grande para asegurar un número de sospechosos que posteriormente pueda ser contrastado con los otros criterios. Corresponde a aproximadamente un 20% de la base de datos.

Para lograr lo anterior se calculó la distancia euclidiana de cada usuario de la base al origen (0,0), obviamente calculando los valores de las componentes para cada usuario con anterioridad. Recordemos que la distancia euclidiana viene dada por:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Luego estos datos se ordenaron de menor a mayor y se extrajeron los 350 primeros usuarios (los más cercanos al origen)²².

Todos estos registros comparten la característica de poseer pocas transacciones diarias en comparación con el resto de los usuarios, y tener tiempos promedios entre transacciones bajo el promedio.

²² El Anexo P incluye el listado completo de los 350 usuarios más cercanos al origen de las componentes "Pocas transacciones" y "Tiempo trans."

Grupo	Datos básicos		SalDOS		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pago		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	2,02	3,06	0,23	1,16	0,45	1,29	5,64	4,07	0,02	0,17	0,02	0,15	0,06	0,37	0,00	0,00	0,00	0,00	3,52	1,13

Tabla 15: Medias y Desv. Std. de los sospechosos cercanos al origen

Fuente: Elaboración propia

Se puede observar que este criterio arroja sospechosos que poseen gran parecido con los casos de fraude, con valores muy similares para las medias de las variables "Consulta de datos básicos", Consulta de saldos", Consulta de últimos movimientos" y "Localización de personas".

4.5. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

Con los tres modelos desarrollados se genera una lista única con los sospechosos que son encontrados por todos los modelos, que totalizan 448 usuarios²³. Un 6% aproximadamente de la base de datos total.

Por otra parte, se genera una lista, utilizando el criterio de Majority Voting. Es decir, aquellos usuarios que son catalogados como sospechosos en 2 o más de los modelos utilizados. En total son 74 usuarios que poseen 2 o más apariciones, 4 de ellos posee 3 y el resto 2. Estos registros corresponden a un 1% aproximadamente de la base de datos, lo que sería mucho más factible de fiscalizar.²⁴ Esto corresponde a un lift del modelo final de 61,18.

Dos de los usuarios poseen perfil AUDIOM, al igual que los fraudes, además de presentar un patrón de comportamiento similar. Estos deben concentrar los mayores esfuerzos de fiscalización.

²³ El Anexo Q contiene la lista completa de usuarios sospechosos encontrados mediante los tres modelos

²⁴ El anexo R contiene los usuarios que surgen como sospechosos en más de un modelo.

Comparando las medias con la de los fraudes, se puede observar que al aumentar el número de sospechosos esto aumenta la desviación estándar de los datos, pero se mantiene la tendencia de cercanía entre las medias de los fraudes y los sospechosos para las variables “Consulta de datos básicos de la cuenta”, “Localización de personas” y “Tiempo entre transacciones”, que resultan notablemente más bajas que la del resto de los datos²⁵.

Grupo	Datos básicos		Saldos		Ult. Movimientos		Loc. Personas		Talonarios		Cabecera Cuenta		Prod. y Sub. de Op.		Cheques		Ordenes no pague		Tiempo entre trans.	
	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd	Media	Sd
Todos	17,16	19,40	2,39	5,80	3,00	6,09	58,73	50,15	1,26	5,89	0,65	2,51	0,43	2,42	0,03	0,34	0,03	0,46	12,96	4,33
Fraudes	2,00	1,41	0,25	0,50	0,75	0,50	3,75	0,96	0,75	0,96	0,50	0,58	0,75	0,96	0,25	0,50	0,25	0,50	1,80	2,32
Sospechosos	1,41	2,63	0,07	0,30	0,19	0,51	4,16	2,20	0,04	0,26	0,05	0,28	0,12	0,47	0,00	0,00	0,00	0,00	2,75	1,41

Tabla 16: Medias y Desv. Std. de los sospechosos lista final
Fuente: Elaboración propia

Podemos ver que la gran mayoría de los usuarios de este conjunto se constituyen como outlier principalmente en las variables localización de personas, consulta de datos básicos de una cuenta (que son las dos variables con mayores medias de los datos. Y los registros de fraudes, como se ha nombrado anteriormente, poseen un menor número de transacciones totales que el resto de los datos), consulta de saldos, consulta de últimos movimientos y tiempo promedio entre transacciones.

Esta tendencia a generar menos transacciones diarias con un tiempo en promedio entre ellas menor, es un claro ejemplo de esfuerzos por minimizar el registro de actividad en ese día por parte del usuario a fin de no levantar sospecha.

Ahora bien, estos comportamientos podrían resultar propio de alguno de los perfiles, que los haga lucir como outliers y ser categorizados como sospechosos por los algoritmos.

Para descartar esta posibilidad se analizarán los usuarios de los 5 perfiles que concentran mayor cantidad de usuarios y se verá si poseen un comportamiento que se desvía de su media. Se considerarán sólo los perfiles que poseen más de un usuario en la lista.

²⁵ El Anexo S muestra la distancia entre los sospechosos de la lista final y la media del resto de los datos.

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
AC01004	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	3
EJ03327	Yellow	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
CRIVERP	Yellow	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
N328981	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
PALVARM	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
RPARRAP	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
EFIGUEC	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2
EJ01322	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	EJECTAS	2
EJ01316	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	EJECTAS	2



Figura 11.1: Mapa de calor registros outlier con perfil EJECTAS
Fuente: Elaboración propia

El perfil CAJASTL que resulta ser el segundo con mayor cantidad de usuarios, no presenta usuarios sospechosos en el listado final.

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
AG01238	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
CFUERTH	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
AG05297	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
AG02032	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
AG02157	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
LBALMDE	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
N431695	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
FBUSTAM	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2
AG03442	Red	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AGENTES	2



Figura 11.2: Mapa de calor registros outlier con perfil AGENTES
Fuente: Elaboración propia

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
EJOU052	Yellow	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AUDIOM	2
EJOU040	Red	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Red	AUDIOM	2



Figura 11.3: Mapa de calor registros outlier con perfil AUDIOM
Fuente: Elaboración propia

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
TE02248											TESOR	3
TE01292											TESOR	2
TE01101											TESOR	2
TE02043											TESOR	2
TE01299											TESOR	2
TE01213											TESOR	2
N596659											TESOR	2
TE02198											TESOR	2



Figura 11.4: Mapa de calor registros outlier con perfil TESOR
Fuente: Elaboración propia

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
GGALASS											GSENIOR	3
GPOZOAR											GSENIOR	2
BSILV01											GSENIOR	2
HSEPULT											GSENIOR	2
DMARGAA											GSENIOR	2
EMUNOZG											GSENIOR	2
CBAHAMO											GSENIOR	2
EJ17170											GSENIOR	2



Figura 11.5: Mapa de calor registros outlier con perfil GSENIOR
Fuente: Elaboración propia

Podemos ver que todos los usuarios de los perfiles analizados poseen desviaciones significativas con respecto a la media de las variables para usuarios de su mismo perfil. Particularmente, se observa que la mayor desviación se concentra en las variables localización de personas, últimos movimientos, datos básicos de la cuenta y tiempo promedio entre transacciones.

Es decir, se mantienen prácticamente los mismos patrones que al comparar con las medias generales de todos los usuarios. Por ende, el comportamiento de outlier no responde a un patrón de comportamiento común a estos perfiles.

Los usuarios indicados como sospechosos presentan comportamientos anómalos, incluso dentro de su mismo perfil.

5. CONCLUSIONES Y TRABAJOS FUTUROS

Ante la imposibilidad de desarrollar metodologías supervisadas, debido a la poca cantidad de registros categorizados como fraudulentos, los métodos no supervisados se alzan como una gran herramienta para generar conocimiento a partir de los datos, sin tener un conocimiento a priori.

Considerando dos criterios de búsqueda, cercanía a los fraudes y búsqueda de outliers; por medio de los tres modelos utilizados, vecinos más cercanos a los fraudes, análisis de componentes principales y local outlier factor, unidos a través de majority voting; se pudo encontrar usuarios sospechosos, que poseen comportamientos similares a los casos de fraudes, permitiendo entregar un listado de 74 usuarios sospechosos cuyo comportamiento debe ser monitoreado.

Estos modelos poseen un buen desempeño al momento de determinar qué puntos se desvían de los patrones de comportamiento del resto de los datos, llegando el modelo final a tener un lift de 61,18. Y dado que los comportamientos que se alejan significativamente de la media del resto de los datos se consideran anómalos y potenciales candidatos a fraude, las técnicas de outlier detection resultan muy útiles a la hora de predecir fraude.

Los usuarios fraudulentos demostraron tener una tendencia a reducir el número de transacciones diarias, generando en promedio 72 transacciones diarias menos que el resto de los datos. Probablemente con el fin de minimizar la actividad registrada en el Log de Tecleo por el usuario, y así llamar menos la atención para evadir posibles fiscalizaciones. También se demostró que el tiempo promedio entre transacciones resulta una buena variable predictiva para estos casos particulares de fraude, ya que los usuarios defraudadores poseen valores menores para esta variable que el resto de los usuarios.

Adicionalmente los fraudes se comportan como outliers de manera marcada en las variables "Consultad de datos básicos de una cuenta" y "Localización de personas", generando en promedio 15 y 54 transacciones diarias menos para esas variables. Por lo que estas variables pueden constituir un buen predictor de casos de fraude interno similares a los analizados.

Sin embargo, el no contar con el asesoramiento del banco y el poseer registros tan limitados en cantidad e información, conllevó la imposibilidad de lograr una comprensión cabal del funcionamiento de negocio y del problema en sí.

No fue posible lograr una comprensión profunda de los registros de fraude, particularmente, de la metodología y secuencialidad utilizada por los usuarios defraudadores para generar el ilícito.

Por todo lo anterior, los modelos propuestos no resultan del todo generalizables a otros tipos de fraude, lo que no permite su implementación como modelo para detectar en tiempo real y prevenir eventualmente nuevos casos de fraude.

Para que esto se pudiera llevar a cabo, debieran incorporarse más registros de fraudes para que los modelos pudieran aprender de ellos, ya que resulta de suma importancia ampliar el universo de tipos de fraudes que considera el modelo. También podría ser recomendable incorporar modelos supervisados para generar una clasificación más exacta y determinante de los datos que pueda ser implementada en tiempo real para poder prevenir futuros ilícitos. Pero, para poder implementar un modelo así se requiere de mayor cantidad de información y conocimiento del negocio que debe ser proporcionado por expertos de la institución bancaria.

Por otra parte, resultaría recomendable realizar experimentos de fiscalización para analizar el comportamiento de los defraudadores en periodos más largos de tiempo y ante escenarios en que el usuario está al tanto de que se está haciendo un seguimiento, y un escenario en que no lo está. Estas situaciones pueden variar el comportamiento del usuario fraudulento y se constituye como una oportunidad para aprender más del comportamiento transaccional de los usuarios fraudulentos y de la evolución de los fraudes.

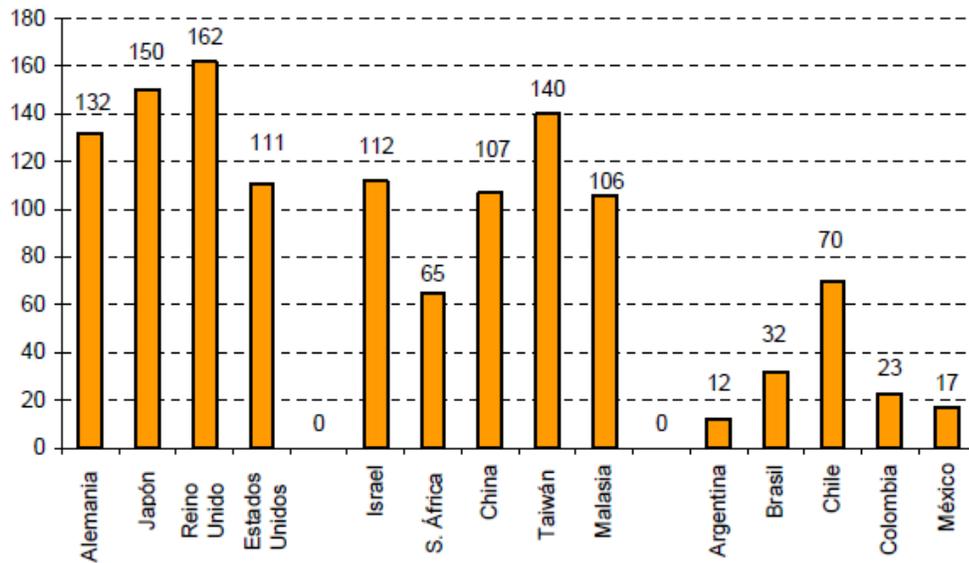
6. BIBLIOGRAFÍA

- [1] Superintendencia de Bancos e Instituciones Financieras. 2015. *Informe Anual SBIF 2015* (pp. 15-16). Santiago.
- [2] Morales, L. & Yañez, Á. 2006. *La Bancarización en Chile: Concepto y Medición*. Santiago: Superintendencia de Bancos e Instituciones Financieras.
- [3] Zahler, R. 2008. *Bancarización Privada en Chile*. Santiago: Naciones Unidas.
- [4] Mackenney, M. & Urbina, M. 2016. *Reporte de Monitoreo Mensual (Estados Financieros Junio 2016) Industria Bancaria*. Santiago: ICRCHILE.
- [5] ABIF. El PIB del sector bancario alcanzó a US\$ 10.900 millones en 2013. 2014. *Abifinforma*, (27).
- [6] CorpResearch. (2015). *Resultados Enero 2015*. Santiago.
- [7] Everis S.A. 2015. *Memoria Anual Everis 2014*. Madrid.
- [8] KPMG Chile. 2009. *Encuesta de Fraude y Corrupción en Chile 2009*. Santiago.
- [9] Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. 2015. *Gotcha! Network-based Fraud Detection for Social Security Fraud* (1st ed., p. 2). Management Science.
- [10] Baesens, B., Van Vlasselaer, V., & Verbeke, W. 2015. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. John Wiley & Sons.
- [11] Superintendencia de Bancos e Instituciones Financieras (SBIF). 2016. *Circular N° 3.601 de Bancos*. Santiago.
- [12] Pacheco, D. 2009. *Riesgo Operacional: Conceptos y Mediciones*. Santiago: Superintendencia de Bancos e Instituciones Financieras.
- [13] Fayyad, U., Piatestky-Shapiro, G., Smyth, P. 1996. *From data mining to knowledge discovery in databases*. American association for artificial intelligence 0738-4602, p. 37-54.

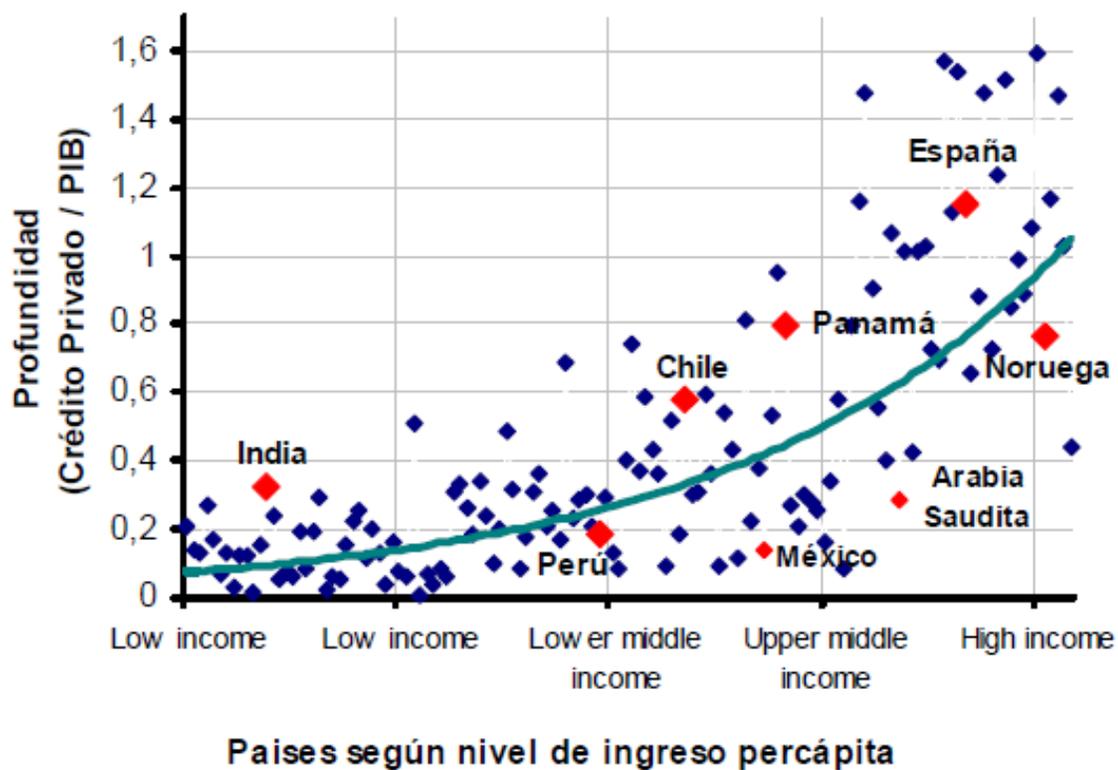
- [14] Pigott, T. 2001. A Review of Methods for Missing Data. *Educational Research and Evaluation*, 7(4), 353-383. <http://dx.doi.org/10.1076/edre.7.4.353.8937>
- [15] Hodge, V. & Austin, J. 2004. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2), 85-126. <http://dx.doi.org/10.1023/b:aire.0000045502.10941.a9>
- [16] So, Y. 1995. A Tutorial on Logistic Regression. *SAS Institute Inc.*
- [17] Kotsiantis, S. 2007. Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249-268.
- [18] Barnett, V. & Lewis, T. 1994. *Outliers in statistical data*. Chichester: Wiley.
- [19] Xu, R. & WunschII, D. 2005. Survey of Clustering Algorithms. *IEEE Trans. Neural Netw.*, 16(3), 645-678. <http://dx.doi.org/10.1109/tnn.2005.845141>
- [20] Breunig, M., Kriegel, H., Ng, R., & Sander, J. 2000. LOF. *ACM SIGMOD Record*, 29(2), 93-104. <http://dx.doi.org/10.1145/335191.335388>
- [21] Abdi, H. & Williams, L. 2010. Principal component analysis. *Wires Comp Stat*, 2(4), 433-459. <http://dx.doi.org/10.1002/wics.101>
- [22] L. Rokach. 2010. Pattern classification using ensemble methods, World Scientific
- [23] Patiño, V. 2014. Modelo de detección de fraude en clients del servicio de agua potable de una empresa sanitaria. Memoria para optar al título de Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [24] Castellón, P. 2012. Caracterización y detección de contribuyentes que presentan facturas falsas al SII mediante técnicas de Data Mining. Tesis para optar al grado de Magíster en Gestión de Operaciones. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

7. ANEXOS

ANEXO A: EVOLUCIÓN DE INDICADORES DE PROFUNDIDAD FINANCIERA Y COMPARACIÓN CON OTROS PAÍSES



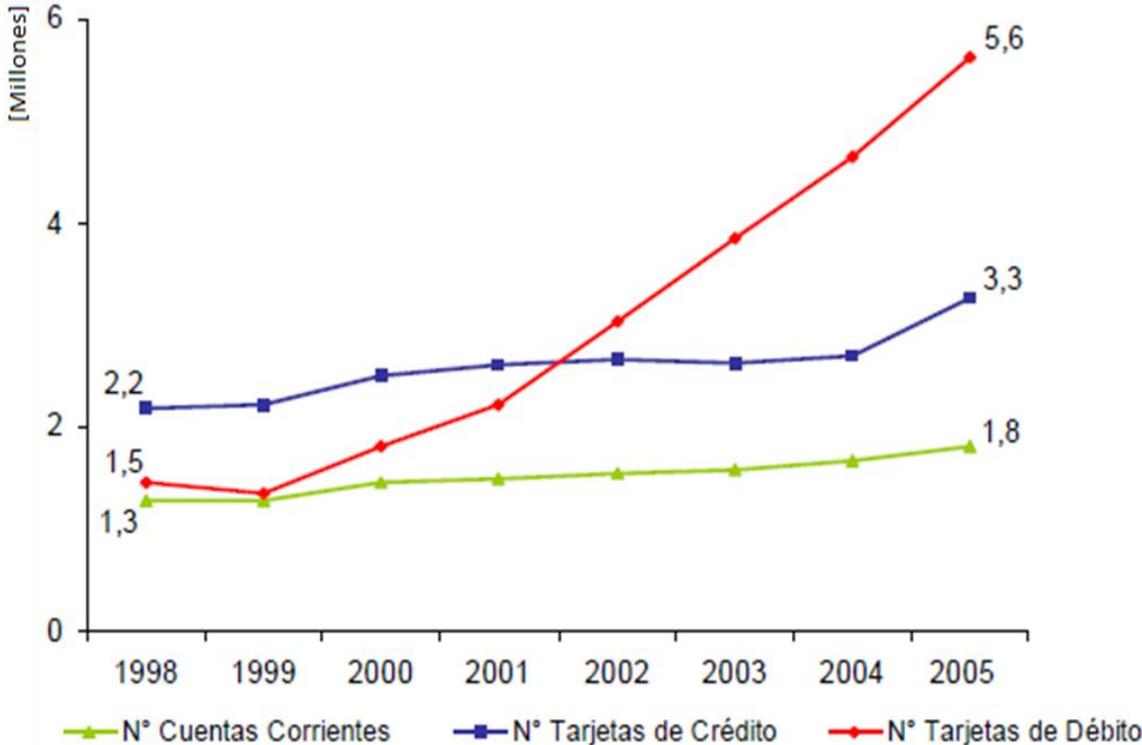
ANEXO B: CORRELACIÓN ENTRE INGRESO PER-CÁPITA Y PROFUNDIDAD BANCARIA



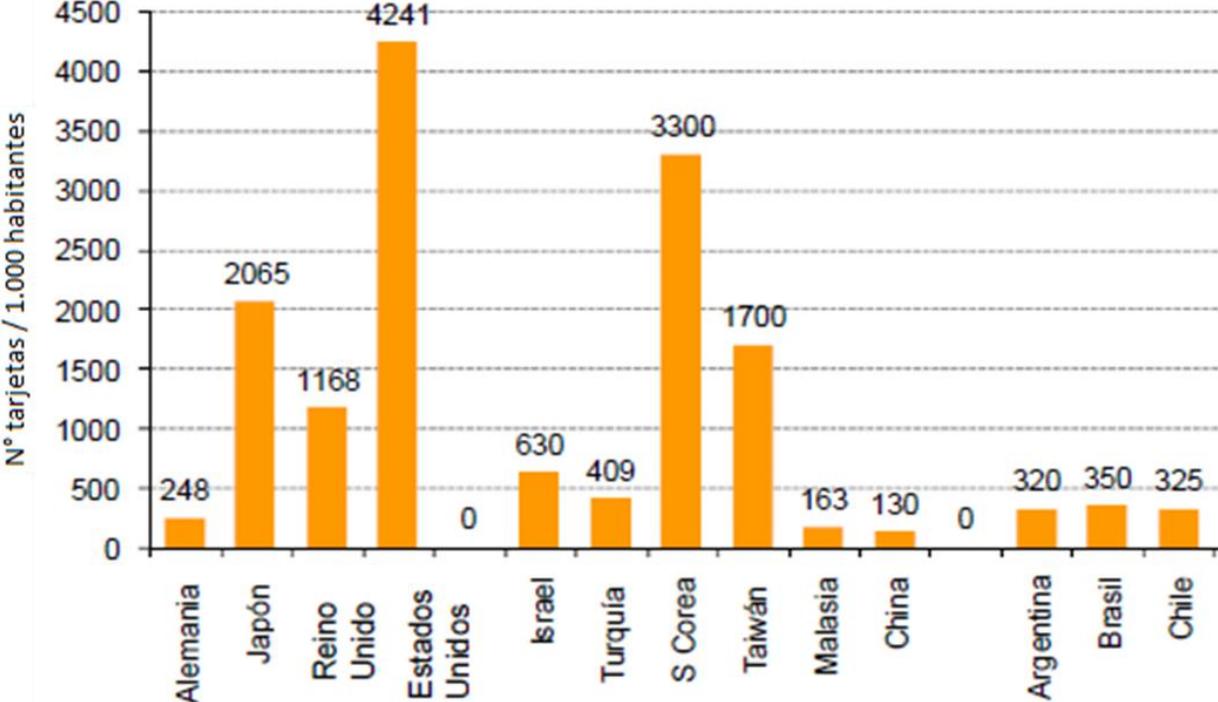
ANEXO C: EVOLUCIÓN DEL INGRESO PER-CÁPITA EN CHILE



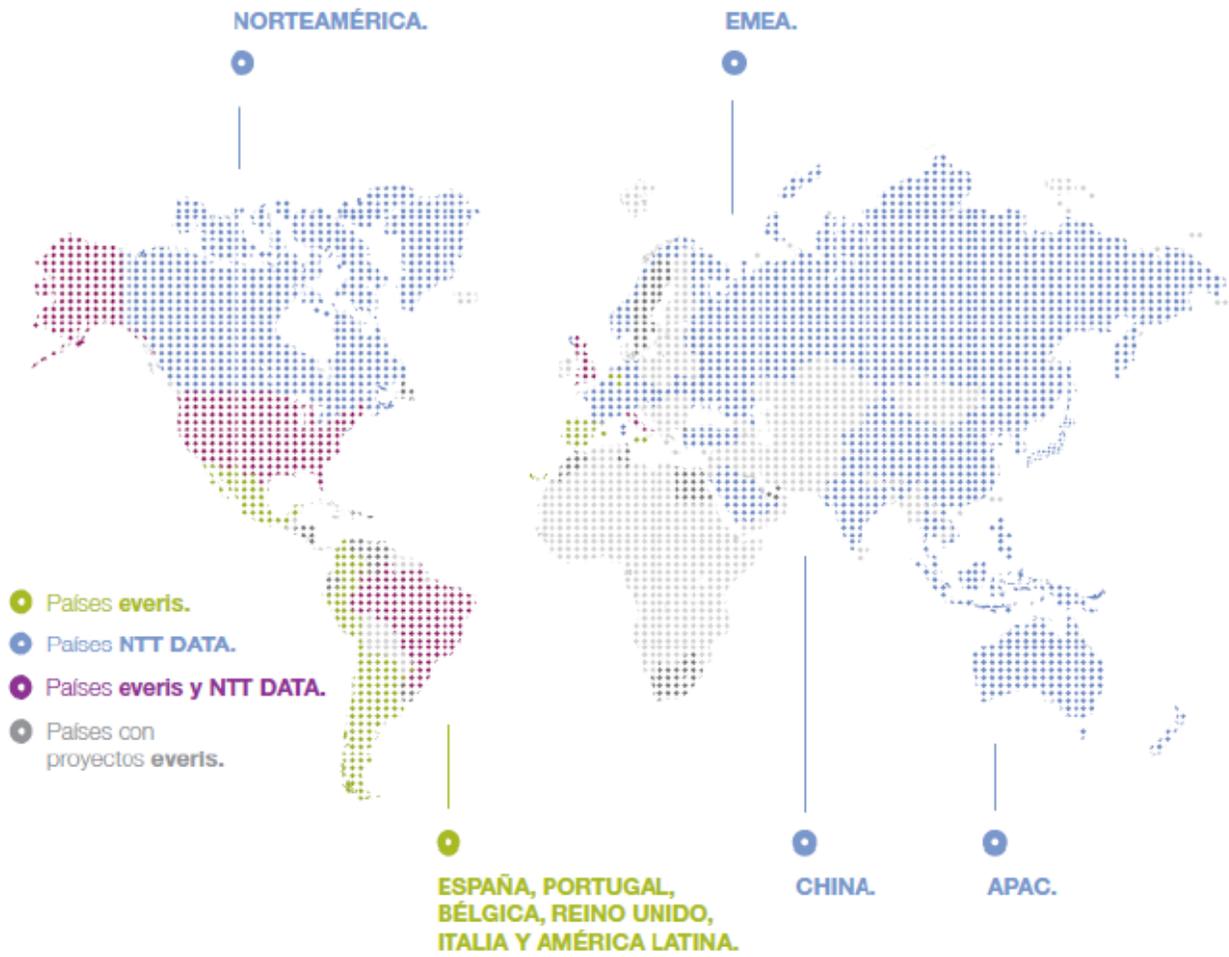
ANEXO D: EVOLUCIÓN DE LA PENETRACIÓN DE TARJETAS DE CRÉDITO, DÉBITO Y CUENTAS CORRIENTES



ANEXO E: PENETRACIÓN DE LAS TARJETAS DE CRÉDITO POR PAÍS

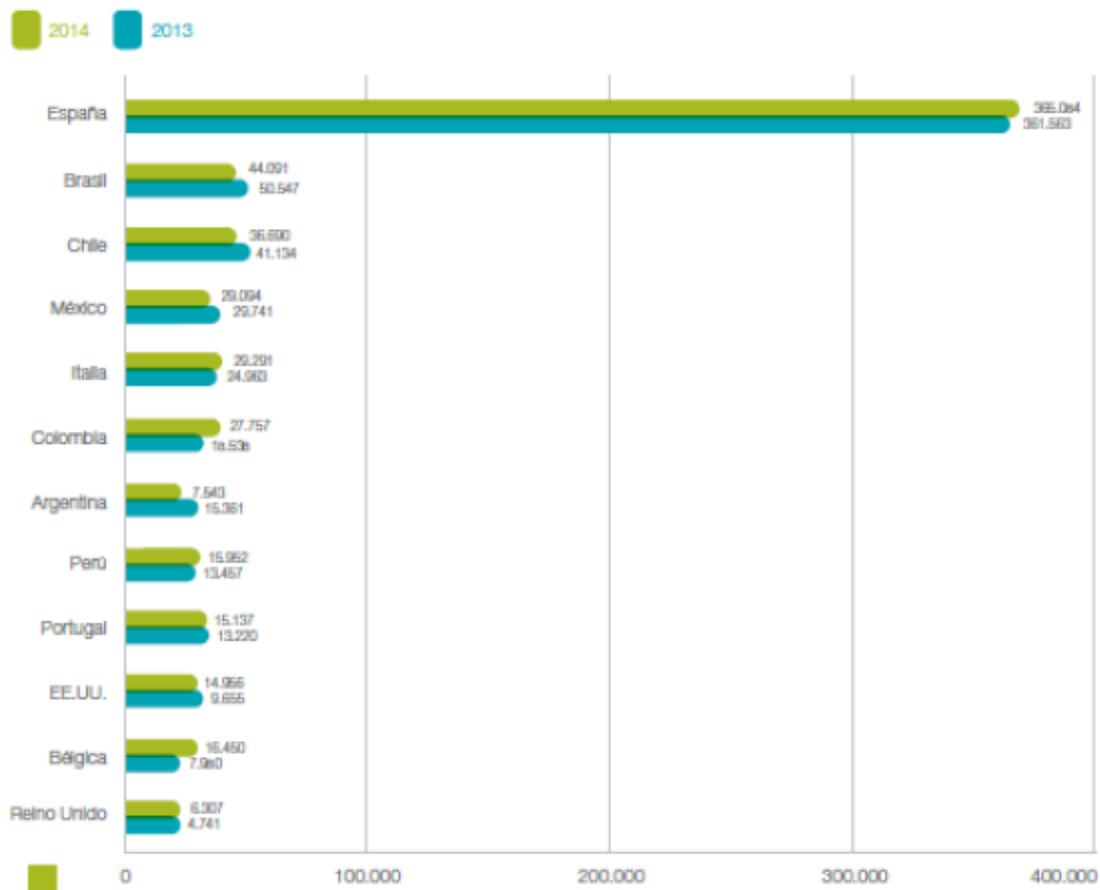


ANEXO F: PAÍSES EN QUE EVERIS TIENE PRESENCIA



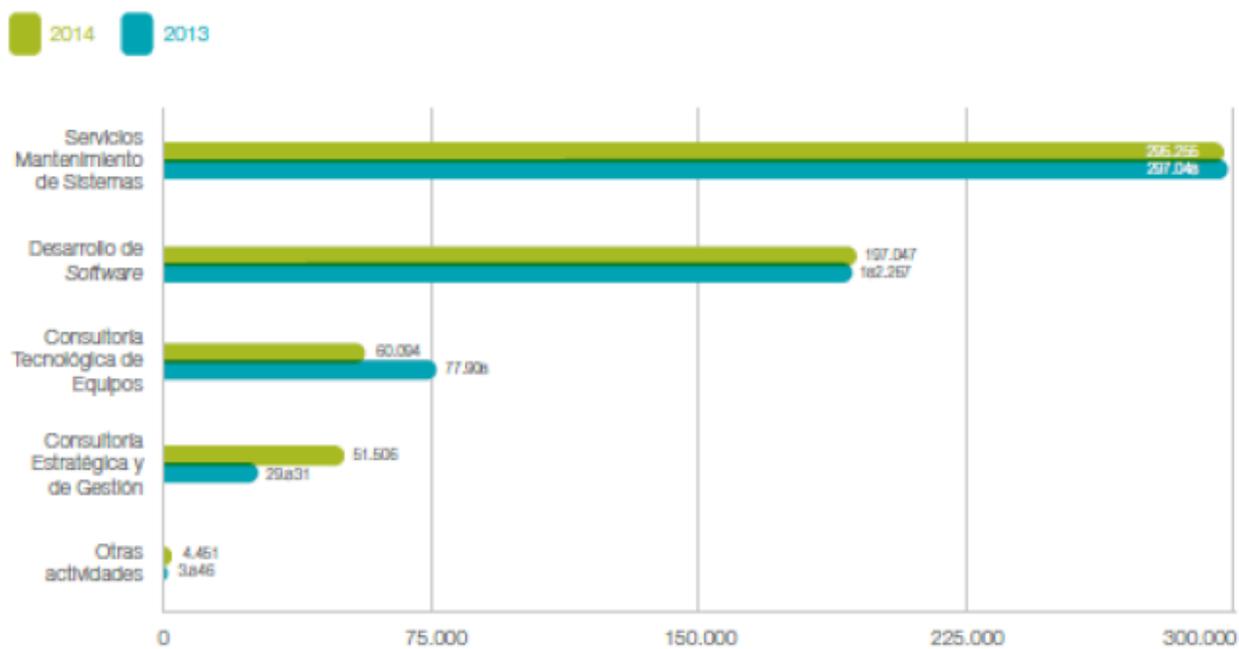
ANEXO G: FACTURACIÓN POR OFICINA

Facturación por países (en miles de €).

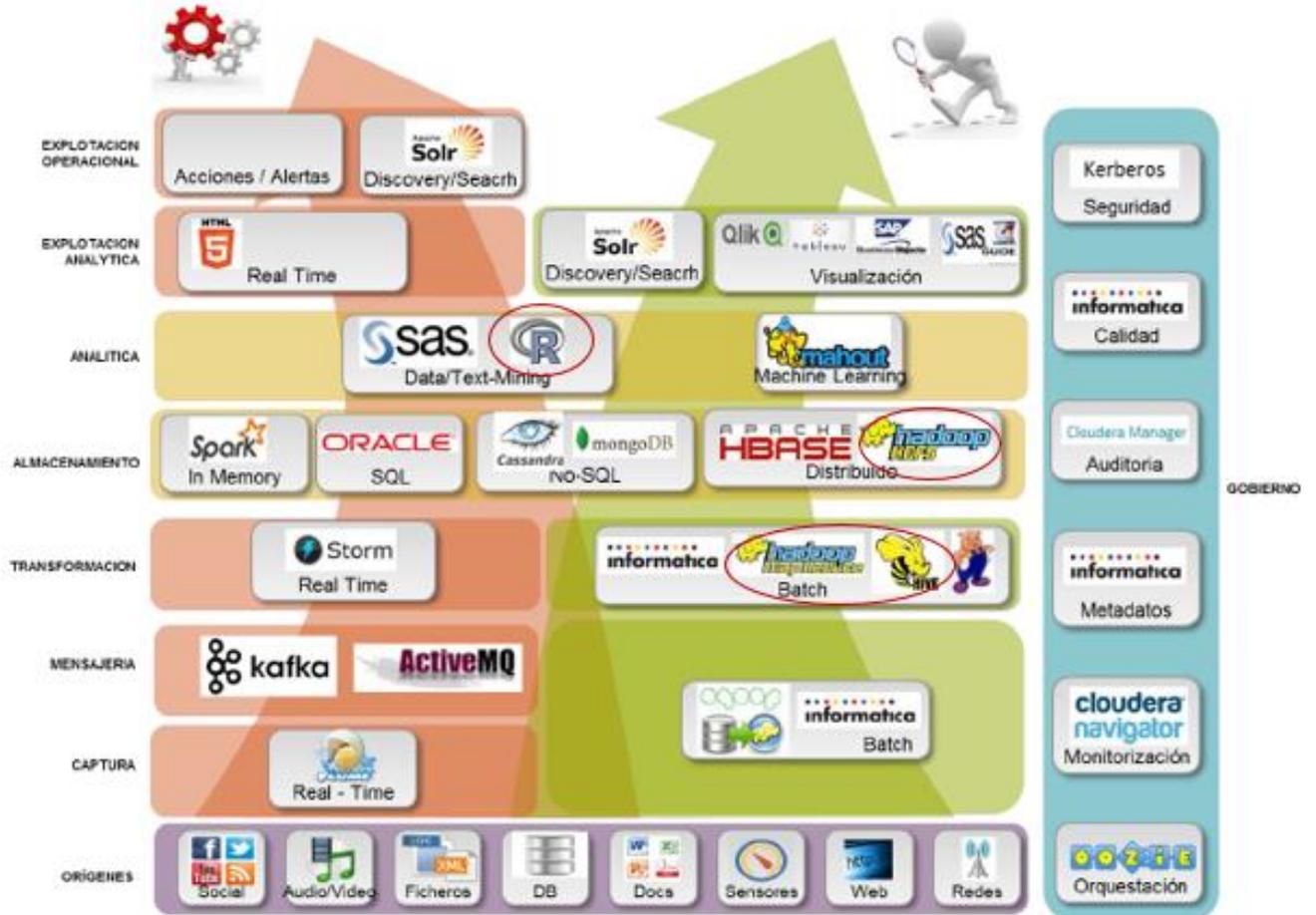


ANEXO H: FACTURACIÓN POR ACTIVIDAD

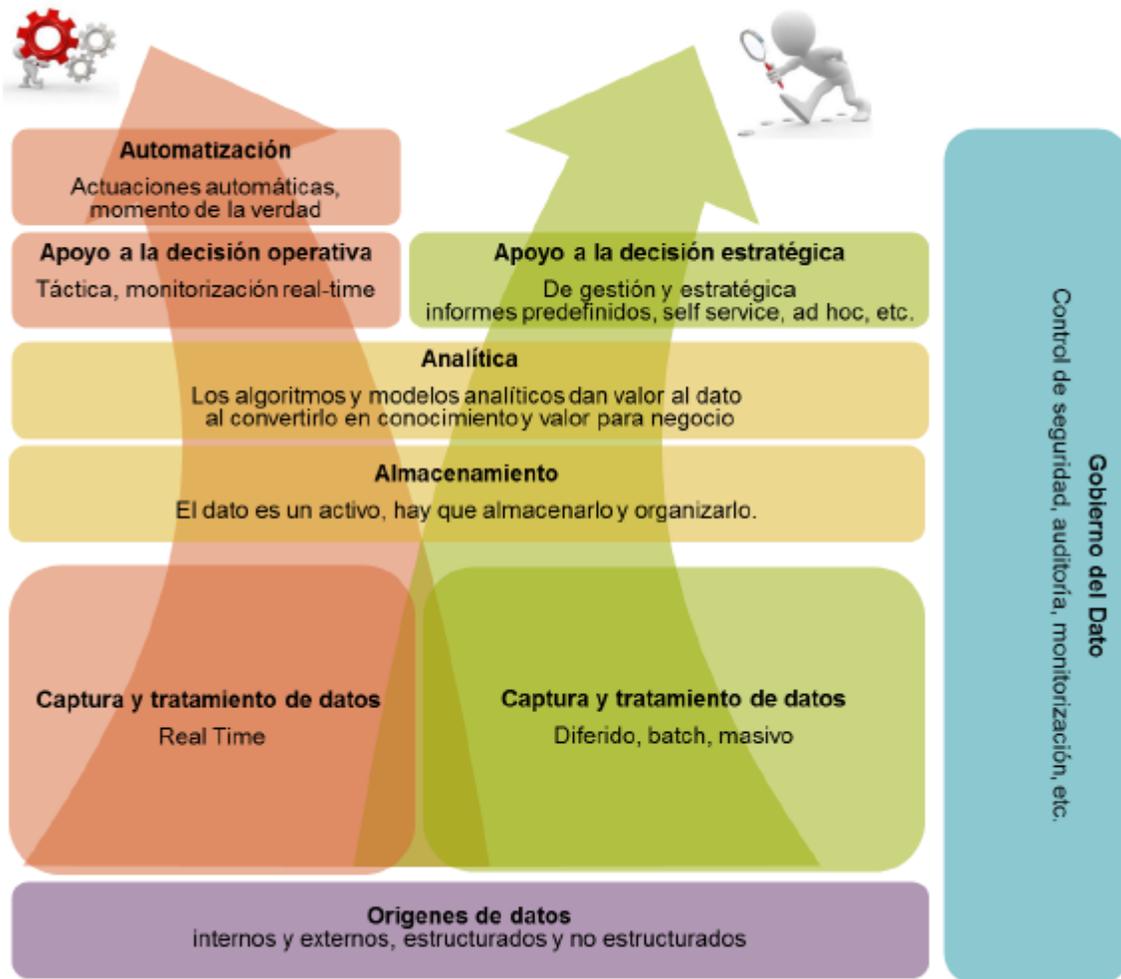
Facturación por actividad (en miles de €).



ANEXO I: ARQUITECTURA BIG DATA DE EVERIS



ANEXO J: FUNCIONES DE LA ARQUITECTURA BIG DATA EVERIS



ANEXO K: MUESTRA DE LA BASE DEL LOG DE TECLEO

	tcL_timestamp	tcL_transaccion	tcL_perfil_usuario	tcL_userid	tcL_cuentacte
0	2015-04-02-12.47.56.416081	MPV3		GHOBP	0035005680
1	2015-04-02-12.47.56.433882	PE62		GHOBP	R 0016
2	2015-04-02-12.47.56.461372	B900	EJECTAS	EJ08170	0170CLP2065
3	2015-04-02-12.47.56.462838	BU89		GOFBP	5000
4	2015-04-02-12.47.56.467984	K443		GHOBP	89100CLP
5	2015-04-02-12.47.56.476495	PE62		GHOBP	R 0015
6	2015-04-02-12.47.56.492692	BU89		GOFBP	5000
7	2015-04-02-12.47.56.498099	KP12		GOFBP	7190917 004 0
8	2015-04-02-12.47.56.511520	KP12		GOFBP	9990007 004 0
9	2015-04-02-12.47.56.521149	KT58		GSERCLI	837320GBP000000000
10	2015-04-02-12.47.56.535616	XP52	EJECTAS	EJ16183	R 00761637207
11	2015-04-02-12.47.56.562667	KP21		GHOBP	0166234212 A0003004
12	2015-04-02-12.47.56.566565	MPV3		MOGOFBP	0035040180
13	2015-04-02-12.47.56.580474	KA01		MOGHOBP	35639CLP
14	2015-04-02-12.47.56.632883	BU80		GHOBP	CLP00350371008271
15	2015-04-02-12.47.56.671893	PE62		GHOBP	R 0010
16	2015-04-02-12.47.56.675256	K443		GHOBP	75455CLP
17	2015-04-02-12.47.56.686124	BU82		GHOBP	
18	2015-04-02-12.47.56.690882	YP80	EJECTAS	EJ14183	YP00350183

ANEXO L.1: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON818

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON818	1	0	1	3	0	0	1	0	0	1	AUDIOM
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR
EJ17170	0	0	0	3	0	0	1	0	0	1,38	GSENIOR
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR
AG02157	0	0	0	3	0	0	0	0	0	1,40	AGENTES
EJ01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE
N596659	0	0	0	3	0	0	0	0	0	1,50	TESOR
DLOPEZV	0	0	0	4	0	0	0	0	0	1,47	FMLOCAL
AG02032	0	0	0	5	0	0	0	0	0	1,37	AGENTES
NREYE01	0	0	0	5	0	0	0	0	0	1,46	SCLBTR
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVEN
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR
PALVARM	0	0	0	6	0	0	0	0	0	1,46	EJECTAS
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVEN
CB03860	0	0	0	4	0	0	0	0	0	1,74	GRCCAS
TE02198	0	0	0	6	0	0	0	0	0	1,53	TESOR
RPARRAP	0	0	0	3	0	0	0	0	0	1,82	EJECTAS
AG05297	1	0	1	2	0	0	0	0	0	2	AGENTES
TE01213	1	0	1	2	0	0	0	0	0	2	TESOR
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NEGINTE
fraude CCON243(1)	2	0	1	3	0	0	2	0	0	0	AUDIOM
JREYESI	0	0	0	3	0	0	2	0	0	1,87	ATPUBS
EFIGUEC	0	0	0	5	0	0	0	0	0	1,81	EJECTAS
C919003	0	0	0	4	0	0	0	0	0	1,95	BFAJERO
EJ01316	0	0	0	3	0	0	0	0	0	1,99	EJECTAS
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR
CB1A345	0	0	0	4	0	0	0	0	0	2	GRCCAS
AI01332	0	0	0	5	0	0	0	0	0	1,96	CAPTADO
CRIVERP	3	0	1	4	0	0	0	0	0	2	EJECTAS

ANEXO L.2: 30 VECINOS MÁS CERCANOS AL FRAUDE JTAPIAL

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talones	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude JTAPIAL	1	0	0	5	2	1	0	1	1	5,21052632	AUDIOM
GGALASS	1	0	1	5	0	1	0	0	0	3,79	GSENIOR
N451798	0	0	0	6	0	0	0	0	0	5,27	TARPAMPN
C929001	0	0	0	6	0	0	0	0	0	5,12	BFCAJERO
MCANETS	0	0	0	6	0	0	0	0	0	5,30	BFCONSUL
C004031	1	0	0	7	0	0	0	0	0	5,29	CAJASTL
EJ01322	0	0	0	6	0	0	0	0	0	5,07	EJECTAS
SM01179	0	0	0	4	0	0	0	0	0	4,91	FUERVEN
N373190	0	0	0	4	0	0	0	0	0	4,90	BFEJECOM
LBALMDE	0	0	0	4	0	0	0	0	0	4,86	AGENTES
RROMERN	2	0	0	5	0	0	0	0	0	5,62	GSENIOR
TE03193	0	0	0	7	0	0	0	0	0	5,33	TESOR
C386001	1	0	0	4	0	0	0	0	0	5,71	SUPCATL
SM02186	0	0	0	3	0	0	0	0	0	5,38	FUERVEN
N431695	2	0	0	5	0	0	0	0	0	4,76	AGENTES
AV10966	0	0	0	3	0	0	0	0	0	5,00	BFCONSUL
C043001	1	0	0	7	0	0	0	0	0	5,60	SUPCATL
AV60999	0	0	0	3	0	0	0	0	0	4,94	BFCONSUL
TE01930	0	0	0	5	0	0	0	0	0	4,72256944	BFTESSUC
FBUSTAM	0	0	0	4	0	0	0	0	0	4,75953704	AGENTES
EJOU040	1	0	0	6	0	0	0	0	0	4,65	AUDIOM
CF11936	0	0	0	3	0	0	0	0	0	4,87444444	GRCCAS
LSEPU04	0	0	0	3	0	0	0	0	0	5,56	ANALSRIE
C207001	0	0	0	5	0	0	0	0	0	5,75	SUPCATL
AV01966	0	0	0	3	0	0	0	0	0	5,63	BFCONSUL
N450601	0	0	0	4	0	0	0	0	0	4,67	JEFEOPE
AG03442	2	0	0	2	0	0	0	0	0	5,22	AGENTES
AV14929	0	0	0	8	0	0	0	0	0	5,14	BFCONSUL
JS03220	0	0	0	4	0	0	1	0	0	5,5125	JEFEOPE
N484117	0	0	0	6	0	0	0	0	0	5,80	BFJOFUSC
EJ02970	0	0	0	8	0	0	0	0	0	5,40900794	BFEJECOM

ANEXO L.3: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON243 (1)

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON243(1)	2	0	1	3	0	0	2	0	0	0	AUDIOM
fraude CCON818	1	0	1	3	0	0	1	0	0	1,00	AUDIOM
EJ17170	0	0	0	3	0	0	1	0	0	1,38	GSENIOR
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR
AGO2157	0	0	0	3	0	0	0	0	0	1,40	AGENTES
EJ01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE
AGO2032	0	0	0	5	0	0	0	0	0	1,37	AGENTES
DLOPEZV	0	0	0	4	0	0	0	0	0	1,47	FMLOCAL
N596659	0	0	0	3	0	0	0	0	0	1,50	TESOR
NREYE01	0	0	0	5	0	0	0	0	0	1,46	SCLBTR
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR
PALVARM	0	0	0	6	0	0	0	0	0	1,46	EJECTAS
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVEN
JREYESI	0	0	0	3	0	0	2	0	0	1,87	ATPUBS
TE02198	0	0	0	6	0	0	0	0	0	1,53	TESOR
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NEGINTE
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS
PDEMEDP	4	0	0	3	0	0	0	0	0	1,72	ANALSRIE
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVEN
CB03860	0	0	0	4	0	0	0	0	0	1,74	GRCCAS
RPARRAP	0	0	0	3	0	0	0	0	0	1,82	EJECTAS
EJOU052	3	0	0	0	0	0	0	0	0	1,83	AUDIOM
JS01300	0	0	0	5	0	0	3	0	0	2,00	JEFOPE
EFIGUEC	0	0	0	5	0	0	0	0	0	1,81	EJECTAS
AG05297	1	0	1	2	0	0	0	0	0	2,00	AGENTES
CRIVERP	3	0	1	4	0	0	0	0	0	2,00	EJECTAS
TE01213	1	0	1	2	0	0	0	0	0	2	TESOR
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR
CF04913	0	0	0	7	0	0	0	0	0	1,77	GRCCAS

ANEXO L.4: 30 VECINOS MÁS CERCANOS AL FRAUDE CCON643

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talones	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON643	4	1	1	4	1	1	0	0	0	1	AUDIOM
FLARA	0	0	0	6	0	1	0	0	0	1,61	SCLBTR
PDEMEDP	4	0	0	3	0	0	0	0	0	1,72	ANALSRIE
CRIVERP	3	0	1	4	0	0	0	0	0	2,00	EJECTAS
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR
fraude CCON818	1	0	1	3	0	0	1	0	0	1,00	AUDIOM
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR
EJ23300	5	0	0	3	0	0	0	0	0	2,02	GRCEJE
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NETINTE
EJOU052	3	0	0	0	0	0	0	0	0	1,83	AUDIOM
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVEN
DP61002	1	1	0	4	0	0	0	0	0	2,03	DADMCPPO
C178001	3	0	0	1	0	0	0	0	0	2,04	SUPCATL
AR31300	6	0	0	0	0	0	0	0	0	1,71	ANALSRIE
CFUERTH	3	0	0	8	0	0	0	0	0	1,91	AGENTES
CMARIAN	7	0	0	0	0	0	0	0	0	1,12	JANALRIE
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVEN
CSUAZOD	5	0	1	8	0	0	0	0	0	2,08	OFCOMEX
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR
TE02043	6	1	3	9	0	0	0	0	0	1,35	TESOR
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR
JS01046	2	0	1	3	1	0	0	0	0	2,46	JEFEPOE
AG05297	1	0	1	2	0	0	0	0	0	2,00	AGENTES
TE01213	1	0	1	2	0	0	0	0	0	2,00	TESOR
CB16879	6	0	0	3	0	0	0	0	0	2,19	GRCCAS
AG02032	0	0	0	5	0	0	0	0	0	1,37	AGENTES
TE02248	3	0	1	7	2	2	1	0	0	2,22	TESOR
AG02157	0	0	0	3	0	0	0	0	0	1,40	AGENTES
N328981	3	0	0	4	0	0	0	0	0	2,35	EJECTAS
EJ01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE

ANEXO L.5: 30 VECINOS MÁS CERCANOS FRAUDE CCON243

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON243	24	0	1	24	0	0	0	0	0	8,99	AUDIOM
C600001	20	0	0	25	0	0	0	0	0	9,46	SUPCATL
EJOU070	21	0	0	24	0	0	0	0	0	7,43	AUDIOM
N568224	30	0	0	26	0	0	0	0	0	10,08	AGENTES
AG03393	17	0	1	26	0	0	0	0	0	9,23	AGENTES
EJ01183	18	0	0	28	0	0	0	0	0	7,96	EJECTAS
C626001	21	0	0	15	0	0	0	0	0	7,76	SUPCATL
AR29300	18	0	0	30	0	0	0	0	0	9,73	ANALSRIE
AG03067	17	1	1	26	0	0	0	0	0	8,56	AGENTES
AG01002	25	3	1	20	0	0	0	0	0	9,81	AGENTES
AR38300	18	0	0	16	0	0	0	0	0	9,36	ANALSRIE
AG01247	17	0	0	18	0	0	0	0	0	8,51	AGENTES
N431443	27	0	0	14	0	0	0	0	0	10,45	ANALSRIE
N722241	22	0	0	36	0	0	0	0	0	9,98	AGENTES
AG02399	17	0	3	22	0	0	0	0	0	9,98	AGENTES
EJ02877	21	0	0	11	0	0	0	0	0	8,77	GRCEJE
AR11234	18	2	0	23	0	0	0	0	0	10,25	ANALSRIE
EJ05020	25	3	0	26	0	0	3	0	0	9,63	EJECTAS
C077003	29	0	0	33	0	0	0	0	0	10,71	CAJASTL
EJ32345	25	0	0	33	0	0	0	0	0	11,44	GRCEJE
AG02162	18	0	1	24	0	0	0	0	0	11,46	AGENTES
PGALVEM	16	0	0	27	0	0	0	0	0	10,64	ANALISTA
KPRADOD	27	0	4	36	0	0	0	0	0	10,19	AUDIOM
AG01861	28	0	0	10	0	0	0	0	0	9,88	GRCEJE
AG01467	18	0	0	22	0	0	0	0	0	11,47	AGENTES
AG01207	18	0	3	18	0	0	0	0	0	11,04	AGENTES
EJ05252	26	2	0	34	0	0	0	0	0	10,91	EJECTAS
EJ02873	21	0	0	11	0	0	0	0	0	10,72	GRCEJE
C055001	26	0	0	36	0	0	0	0	0	11,12	SUPCATL
EJOUT56	20	0	0	30	0	0	0	0	0	11,74	AUDIOM
AG01269	24	0	3	40	0	0	0	0	0	9,18	AGENTES

ANEXO M.1: DISTANCIA DE VECINOS MÁS CERCANOS A CCON818 A MEDIA DE FRAUDES

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
DMARGAA											GSENIOR
EJ17170											GSENIOR
TE01299											TESOR
AG02157											AGENTES
EJ01866											GRCEJE
N596659											TESOR
DLOPEZV											FMLOCAL
AG02032											AGENTES
NREYE01											SCLBTR
VT18093											FUERVEN
HSEPULT											GSENIOR
PALVARM											EJECTAS
AC01004											EJECTAS
FVTA232											FUERVEN
CB03860											GRCCAS
TE02198											TESOR
RPARRAP											EJECTAS
AG05297											AGENTES
TE01213											TESOR
FLYONAU											NEGINTE
fraude CCON243(1)											AUDIOM
JREYESI											ATPUBS
EFIGUEC											EJECTAS
C919003											BFCAJERO
EJ01316											EJECTAS
EMUNOZG											GSENIOR
CBAHAMO											GSENIOR
CB1A345											GRCCAS
AI01332											CAPTADO
CRIVERP											EJECTAS



ANEXO M.2: DISTANCIA DE VECINOS MÁS CERCANOS A JTAPIAL A MEDIA DE FRAUDES

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
GGALASS											GSENIOR
N451798											TARPAMPN
C929001											BFCAJERO
MCANETS											BFCONSUL
C004031											CAJASTL
EJ01322											EJECTAS
SM01179											FUERVEN
N373190											BFEJECOM
LBALMDE											AGENTES
RROMERN											GSENIOR
TE03193											TESOR
C386001											SUPCATL
SM02186											FUERVEN
N431695											AGENTES
AV10966											BFCONSUL
C043001											SUPCATL
AV60999											BFCONSUL
TE01930											BFTESSUC
FBUSTAM											AGENTES
EJOU040											AUDIOM
CF11936											GRCCAS
LSEPU04											ANALSRIE
C207001											SUPCATL
AV01966											BFCONSUL
N450601											JEFEOPE
AG03442											AGENTES
AV14929											BFCONSUL
JS03220											JEFEOPE
N484117											BFJOFUSUC
EJ02970											BFEJECOM



ANEXO M.3: DISTANCIA DE VECINOS MÁS CERCANOS A CCON243 (1) A MEDIA LOS FRAUDES

User_id	Datos básicos	SalDOS	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON818											AUDIOM
EJ17170											GSENIOR
DMARGAA											GSENIOR
TE01299											TESOR
AG02157											AGENTES
EJ01866											GRCEJE
AG02032											AGENTES
DLOPEZV											FMLOCAL
N596659											TESOR
NREYE01											SCLBTR
HSEPULT											GSENIOR
EMUNOZG											GSENIOR
PALVARM											EJECTAS
VT18093											FUERVEN
JREYESI											ATPUBS
TE02198											TESOR
FLYONAU											NEGINTE
AC01004											EJECTAS
PDEMEDP											ANALSRIE
FVTA232											FUERVEN
CB03860											GRCCAS
RPARRAP											EJECTAS
EJOU052											AUDIOM
JS01300											JEFEOPE
EFIGUEC											EJECTAS
AG05297											AGENTES
CRIVERP											EJECTAS
TE01213											TESOR
CBAHAMO											GSENIOR
CF04913											GRCCAS



ANEXO M.4: DISTANCIA DE VECINOS MÁS CERCANOS A CCON643 A MEDIA DE FRAUDES

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
FLARA											SCLBTR
PDEMEDP											ANALSRIE
CRIVERP											EJECTAS
HSEPULT											GSENIOR
DMARGAA											GSENIOR
fraude CCON818											AUDIOM
EMUNOZG											GSENIOR
EJ23300											GRCEJE
FLYONAU											NEGINTE
EJOU052											AUDIOM
VT18093											FUERVEN
DP61002											DADMCPO
C178001											SUPCATL
AR31300											ANALSRIE
CFUERTH											AGENTES
CMARIAN											JANALRIE
FVTA232											FUERVEN
CSUAZOD											OFCOMEX
CBAHAMO											GSENIOR
TE02043											TESOR
TE01299											TESOR
JS01046											JEFOPE
AG05297											AGENTES
TE01213											TESOR
CB16879											GRCCAS
AG02032											AGENTES
TE02248											TESOR
AG02157											AGENTES
N328981											EJECTAS
EJ01866											GRCEJE



ANEXO M.5: DISTANCIA DE VECINOS MÁS CERCANOS A CCON243 A MEDIA DE FRAUDES

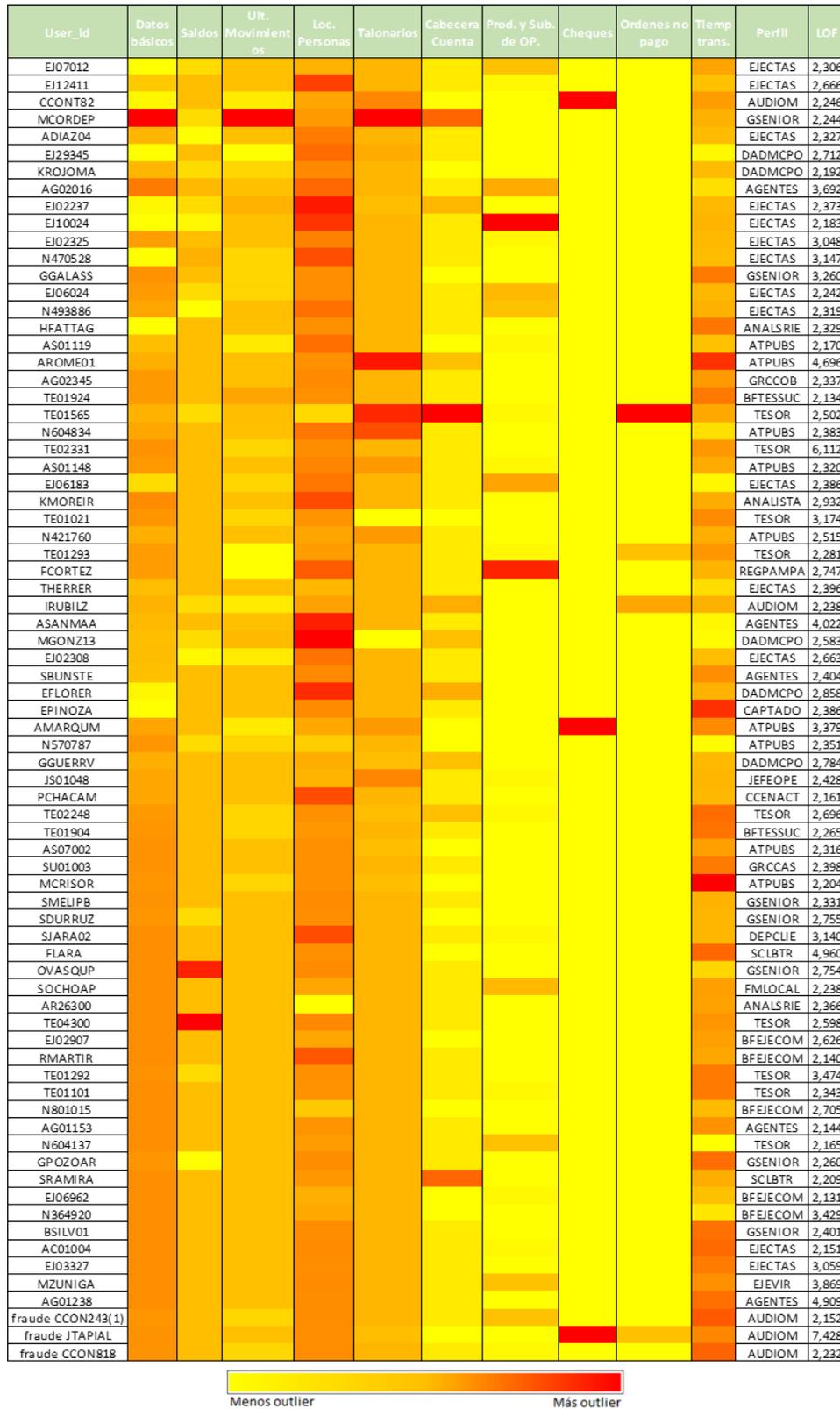
User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
C600001											SUPCATL
EJOU070											AUDIOM
N568224											AGENTES
AG03393											AGENTES
EJ01183											EJECTAS
C626001											SUPCATL
AR29300											ANALSRIE
AG03067											AGENTES
AG01002											AGENTES
AR38300											ANALSRIE
AG01247											AGENTES
N431443											ANALSRIE
N722241											AGENTES
AG02399											AGENTES
EJ02877											GRCEJE
AR11234											ANALSRIE
EJ05020											EJECTAS
C077003											CAJASTL
EJ32345											GRCEJE
AG02162											AGENTES
PGALVEM											ANALISTA
KPRADOD											AUDIOM
AG01861											GRCEJE
AG01467											AGENTES
AG01207											AGENTES
EJ05252											EJECTAS
EJ02873											GRCEJE
C055001											SUPCATL
EJOUT56											AUDIOM
AG01269											AGENTES



ANEXO N: USUARIOS CLASIFICADOS COMO OUTLIERS POR LOF

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	LOF
EJ07012	18	1	6	80	0	0	2	0	0	17,44	EJECTAS	2,306
EJ12411	20	5	0	186	0	0	1	0	0	14,10	EJECTAS	2,666
CCONT82	16	0	2	27	5	1	0	1	0	7,68	AUDIOM	2,246
MCORDEP	74	1	76	102	12	12	0	0	0	9,98	GSENIOR	2,244
ADIAZ04	11	2	0	130	0	0	0	0	0	11,22	EJECTAS	2,327
EJ29345	18	0	3	145	3	0	0	0	0	13,27	DADMCPO	2,712
KROJOMA	11	1	1	117	0	1	0	0	0	14,58	DADMCPO	2,192
AG02016	40	6	0	148	0	0	5	0	0	13,65	AGENTES	3,692
EJ02237	16	1	11	221	2	3	0	0	0	10,93	EJECTAS	2,373
EJ10024	18	3	6	196	0	0	29	0	0	10,47	EJECTAS	2,183
EJ02325	30	5	0	125	0	0	1	0	0	14,91	EJECTAS	3,048
N470528	18	8	1	172	0	0	0	0	0	11,44	EJECTAS	3,147
GGALASS	1	0	1	5	0	1	0	0	0	3,79	GSENIOR	3,260
EJ06024	31	1	5	113	0	0	3	0	0	15,05	EJECTAS	2,242
N493886	28	2	7	141	0	0	2	0	0	15,87	EJECTAS	2,319
HFATTAG	18	0	0	6	0	0	0	0	0	3,39	ANALSRIE	2,329
AS01119	21	0	4	142	0	1	1	0	0	14,10	ATPUBS	2,170
AROME01	9	0	0	6	11	2	0	0	0	30,54	ATPUBS	4,696
AG02345	3	0	0	117	0	0	0	0	0	7,49	GRCCOB	2,337
TE01924	3	0	16	6	0	0	0	0	0	22,45	BFTESSUC	2,134
TE01565	24	1	7	64	10	23	1	0	23	9,11	TESOR	2,502
N604834	7	0	0	135	8	0	0	0	0	13,65	ATPUBS	2,383
TE02331	1	0	1	3	0	0	0	0	0	7	TESOR	6,112
AS01148	3	0	0	122	4	0	1	0	0	16,55	ATPUBS	2,320
EJ06183	15	0	1	135	0	0	6	0	0	12,61	EJECTAS	2,386
KMOREIR	35	0	0	175	0	0	0	0	0	16,42	ANALISTA	2,932
TE01021	2	0	1	9	1	1	0	0	0	20,17	TESOR	3,174
N421760	9	0	0	25	4	0	0	0	0	16,20	ATPUBS	2,515
TE01293	4	0	3	18	0	0	1	0	1	6,89	TESOR	2,281
FCORTEZ	4	0	3	160	0	0	24	0	0	10,33	REGPAMPA	2,747
THERRER	13	0	0	76	0	0	0	0	0	13,68	EJECTAS	2,396
IRUBILZ	10	1	4	22	0	4	0	0	4	10,02	AUDIOM	2,238
ASANMAA	12	0	0	216	0	0	0	0	0	13,29	AGENTES	4,022
MGNONZ13	21	1	8	243	1	2	0	0	0	13,24	DADMCPO	2,583
EJ02308	13	3	2	138	0	0	0	0	0	14,41	EJECTAS	2,663
SBUNSTE	14	0	0	0	0	0	0	0	0	19,78	AGENTES	2,404
EFLORER	16	0	0	204	0	4	0	0	0	15,68	DADMCPO	2,858
EPINOZA	18	0	0	2	0	0	0	0	0	30,70	CAPTADO	2,386
AMARQUM	6	0	2	29	4	1	0	1	0	20,20	ATPUBS	3,379
NS70787	2	1	1	52	0	1	0	0	0	13,24	ATPUBS	2,351
GGUERRV	9	0	0	84	2	2	0	0	0	14,87	DADMCPO	2,784
JS01048	7	0	0	40	5	0	1	0	0	10,57	JEFEPOE	2,428
PCHACAM	7	0	6	174	0	0	0	0	0	15,01	CCENACT	2,161
TE02248	3	0	1	7	2	2	1	0	0	2,22	TESOR	2,696
TE01904	2	0	1	12	0	0	0	0	0	2,99	BFTESSUC	2,265
AS07002	1	0	0	6	2	1	0	0	0	17,94	ATPUBS	2,316
SU01003	1	0	0	6	0	0	0	0	0	3,68	GRCCAS	2,398
MCRISOR	2	0	1	7	2	1	0	0	0	36,15	ATPUBS	2,204
SMELIPB	1	0	6	2	0	0	0	0	0	15,63	GSENIOR	2,331
SDURRUZ	2	1	0	4	0	1	0	0	0	10,57	GSENIOR	2,755
SJARA02	0	0	0	173	0	0	1	0	0	15,32	DEPCLUE	3,140
FLARA	0	0	0	6	0	1	0	0	0	1,61	SCLBTR	4,960
OVASQUP	0	39	0	1	0	0	0	0	0	12,11	GSENIOR	2,754
SOCHOAP	0	0	0	27	0	0	3	0	0	17,95	FMLOCAL	2,238
AR26300	0	0	0	58	0	0	0	0	0	8,33	ANALSRIE	2,366
TE04300	0	46	0	0	0	0	0	0	0	18,96	TESOR	2,598
EJ02907	0	0	0	26	0	1	0	0	0	7,90	BFEJECOM	2,626
RMARTIR	0	0	0	165	0	0	0	0	0	17,26	BFEJECOM	2,140
TE01292	1	1	0	6	0	0	0	0	0	3,75	TESOR	3,474
TE01101	0	0	0	8	0	0	1	0	0	3,83	TESOR	2,343
N801015	0	0	0	66	0	1	0	0	0	14,77	BFEJECOM	2,705
AG01153	0	0	0	9	0	0	0	0	0	19,45	AGENTES	2,144
N604137	0	0	0	17	0	0	2	0	0	12,73	TESOR	2,165
GPOZOAR	2	2	0	3	0	0	0	0	0	2,42	GSENIOR	2,260
SRAMIRA	0	0	0	12	0	12	0	0	0	9,65	SCLBTR	2,209
EJ06962	0	0	0	34	0	1	1	0	0	11,72	BFEJECOM	2,131
N364920	0	0	0	29	0	1	0	0	0	12,36	BFEJECOM	3,429
BSILV01	0	0	0	3	0	0	0	0	0	2,66	GSENIOR	2,401
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS	2,151
EJ03327	0	0	0	3	0	0	0	0	0	3,89	EJECTAS	3,059
MZUNIGA	0	0	0	2	0	0	2	0	0	19,67	EJEVIR	3,869
AG01238	0	0	0	3	0	0	0	0	0	2,57	AGENTES	4,909
fraude CCON243(1)	2	0	1	3	0	0	2	0	0	0	AUDIOM	2,152
fraude JTAPIAL	1	0	0	5	2	1	0	1	1	5,21	AUDIOM	7,428
fraude CCON818	1	0	1	3	0	0	1	0	0	1	AUDIOM	2,232

ANEXO O: DISTANCIA DE LOS SOSPECHOSOS LOF A LA MEDIA DE LOS DATOS



ANEXO P: RESULTADO ANÁLISIS DE LAS COMPONENTES PRINCIPALES

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil
fraude CCON	2	0	1	3	0	0	2	0	0	0,00	AUDIOM
CMARIAN	7	0	0	0	0	0	0	0	0	1,12	JANALRIE
TE02043	6	1	3	9	0	0	0	0	0	1,35	TESOR
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR
fraude CCON	1	0	1	3	0	0	1	0	0	1,00	AUDIOM
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR
AR31300	6	0	0	0	0	0	0	0	0	1,71	ANALSRIE
EJ17170	0	0	0	3	0	0	1	0	0	1,38	GSENIOR
fraude CCON	4	1	1	4	1	1	0	0	0	1,00	AUDIOM
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR
AG02157	0	0	0	3	0	0	0	0	0	1,40	AGENTES
EJ01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE
N596659	0	0	0	3	0	0	0	0	0	1,50	TESOR
AG02032	0	0	0	5	0	0	0	0	0	1,37	AGENTES
DLOPEZV	0	0	0	4	0	0	0	0	0	1,47	FMLOCAL
EJOU052	3	0	0	0	0	0	0	0	0	1,83	AUDIOM
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NEGINTE
NREYE01	0	0	0	5	0	0	0	0	0	1,46	SCLBTR
PDEMEDP	4	0	0	3	0	0	0	0	0	1,72	ANALSRIE
PALVARM	0	0	0	6	0	0	0	0	0	1,46	EJECTAS
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVEN
N484075	4	1	5	4	0	0	0	0	0	2,65	EJECTAS
AG05297	1	0	1	2	0	0	0	0	0	2,00	AGENTES
TE01213	1	0	1	2	0	0	0	0	0	2,00	TESOR
TE02198	0	0	0	6	0	0	0	0	0	1,53	TESOR
JINZUNZ	6	0	0	0	0	0	0	0	0	2,09	GRCCOB
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS
CB03860	0	0	0	4	0	0	0	0	0	1,74	GRCCAS
RPARRAP	0	0	0	3	0	0	0	0	0	1,82	EJECTAS
CRIVERP	3	0	1	4	0	0	0	0	0	2,00	EJECTAS
C178001	3	0	0	1	0	0	0	0	0	2,04	SUPCATL
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVEN
JREYESI	0	0	0	3	0	0	2	0	0	1,87	ATPUBS
FLARA	0	0	0	6	0	1	0	0	0	1,61	SCLBTR
DP61002	1	1	0	4	0	0	0	0	0	2,03	DADMCPPO
EFIGUEC	0	0	0	5	0	0	0	0	0	1,81	EJECTAS
EJ23300	5	0	0	3	0	0	0	0	0	2,02	GRCEJE
EJ01316	0	0	0	3	0	0	0	0	0	1,99	EJECTAS
C040001	6	0	0	1	0	0	0	0	0	2,25	SUPCATL
GPOZOAR	2	2	0	3	0	0	0	0	0	2,42	GSENIOR
C919003	0	0	0	4	0	0	0	0	0	1,95	BFCAJERO
CF04913	0	0	0	7	0	0	0	0	0	1,77	GRCCAS
AP02186	0	0	0	3	0	0	0	0	0	2,06	DADMCPPO
CB1A345	0	0	0	4	0	0	0	0	0	2,00	GRCCAS
N484327	1	0	1	2	0	0	0	0	0	2,33	TESOR
AS01231	1	0	2	2	0	0	0	0	0	2,50	ATPUBS
CB16879	6	0	0	3	0	0	0	0	0	2,19	GRCCAS
AC02234	1	0	1	4	0	0	0	0	0	2,22	EJECTAS
A101332	0	0	0	5	0	0	0	0	0	1,96	CAPTADO
JS01300	0	0	0	5	0	0	3	0	0	2,00	JEFEPOE
N449286	0	0	0	4	0	0	0	0	0	2,08	BFEJECOM
CGUTIEL	2	0	5	4	0	0	0	0	0	2,93	GSENIOR
CSUAZOD	5	0	1	8	0	0	0	0	0	2,08	OFCOMEX
FVTA167	2	0	0	3	0	0	0	0	0	2,23	FUERVEN
EJ02319	0	0	0	4	0	0	0	0	0	2,11	EJECTAS
TE02087	0	0	0	8	0	0	0	0	0	1,85	TESOR
CB71345	2	0	0	1	0	0	0	0	0	2,39	GRCCAS
SU02856	1	0	0	3	0	0	0	0	0	2,25	GRCCAS
JRIVE12	2	0	2	8	0	0	0	0	0	2,23	GSENIOR
MMIRA01	3	0	0	0	0	0	0	0	0	2,56	EJECTAS
CFUERTH	3	0	0	8	0	0	0	0	0	1,91	AGENTES
PTAMAYO	0	0	0	4	0	0	0	0	0	2,22	DADMCPPO
CF11913	0	0	0	5	0	0	0	0	0	2,15	GRCCAS
MBALTO1	0	0	0	4	0	0	0	0	0	2,22	SCLBTR
EJ45345	9	0	0	2	0	0	0	0	0	2,45	GRCEJE
EJ04013	3	0	0	2	0	0	0	0	0	2,48	GRCEJE
JS01046	2	0	1	3	1	0	0	0	0	2,46	JEFEPOE
RJ01288	0	0	0	5	0	0	0	0	0	2,21	JANALRIE
N328981	3	0	0	4	0	0	0	0	0	2,35	EJECTAS
FOTONDO	0	0	0	4	0	0	0	0	0	2,31	JANALRIE
DLEPE	3	0	0	0	0	0	0	0	0	2,67	GSENIOR
CSAEZGO	1	0	1	2	0	0	0	0	0	2,67	PROCENT
AV13966	0	0	0	6	0	0	0	0	0	2,19	BFCONSUL

JSEIN	2	0	0	3	0	0	0	0	0	2,46	ANALISTA
TE01427	6	0	4	10	0	0	0	0	0	2,63	TESOR
TE01992	0	0	0	4	0	0	0	0	0	2,35	BFTESSUC
AG02248	1	0	0	6	0	0	0	0	0	2,20	AGENTES
CH02297	0	0	0	6	0	0	0	0	0	2,22	EJECTAS
TE01647	0	0	0	3	0	0	0	0	0	2,44	TESOR
MRICCB	2	0	0	2	0	0	0	0	0	2,57	ANALSRIE
EJ11874	1	0	0	2	0	0	0	0	0	2,56	GRCEJE
JS01903	0	0	0	3	0	0	0	0	0	2,50	BFJOPSUC
N364581	0	0	0	3	0	0	0	0	0	2,51	BFCAJERO
RECO142	0	0	0	3	0	0	0	0	0	2,53	AUDIOM
MESPINP	0	0	0	3	0	0	0	0	0	2,53	EJECTAS
MLLANOC	0	0	0	5	0	0	0	0	0	2,40	NEGINTE
EJ21288	0	0	0	3	0	0	0	0	0	2,54	EJECTAS
TE01032	0	0	0	3	0	0	0	0	0	2,57	TESOR
JS29300	4	0	1	8	0	0	1	0	0	2,45	JEFEPOE
AG01238	0	0	0	3	0	0	0	0	0	2,57	AGENTES
AG01445	2	0	1	5	0	0	0	0	0	2,64	AGENTES
AC03297	0	0	0	4	0	0	0	0	0	2,52	EJECTAS
FVTA115	0	0	0	4	0	0	0	0	0	2,55	FUERVEN
AV04905	0	0	0	4	0	0	0	0	0	2,55	BFCONSUL
CH06300	0	0	0	3	0	0	0	0	0	2,64	EJECTAS
C045001	2	0	0	5	0	0	0	0	0	2,56	SUPCATL
BSILV01	0	0	0	3	0	0	0	0	0	2,66	GSENIOR
AOYARCC	0	0	0	4	0	0	0	0	0	2,61	BFCONSUL
LROMERS	0	0	0	4	0	0	0	0	0	2,61	GSENIOR
JTORREA	1	0	0	3	0	0	0	0	0	2,72	GSENIOR
NGONZAD	5	0	0	0	0	0	0	0	0	3,03	GSENIOR
TE02248	3	0	1	7	2	2	1	0	0	2,22	TESOR
CF22843	8	0	0	0	0	0	0	0	0	3,07	GRCCAS
CB01870	2	0	0	3	0	0	0	0	0	2,81	GRCCAS
EJ11402	0	0	0	6	0	0	0	0	0	2,57	EJECTAS
CT66345	0	0	0	5	0	0	0	0	0	2,65	GRCCAS
AP01045	0	0	0	5	0	0	0	0	0	2,65	ADMCPPO
C031006	0	0	0	7	0	0	0	0	0	2,52	CAJASTL
C182001	3	0	0	3	0	0	0	0	0	2,88	SUPCATL
MBILBAB	0	0	0	6	0	0	0	0	0	2,63	MERCAP
MMIQU01	0	0	0	3	0	0	0	0	0	2,86	BFCONSUL
TE01195	6	0	1	10	0	0	0	0	0	2,64	TESOR
GNEIRO1	0	0	0	5	0	0	0	0	0	2,77	EJECTAS
VDUQUME	0	0	0	6	0	0	0	0	0	2,71	EJECTAS
AG01443	3	2	3	7	0	0	1	0	0	3,44	AGENTES
CF16823	0	0	0	3	0	0	0	0	0	2,95	GRCCAS
C174002	3	0	0	4	0	0	0	0	0	2,96	SUPCATL
C024001	3	0	0	6	0	0	0	0	0	2,86	SUPCATL
TE01192	1	0	1	2	0	0	0	0	0	3,33	TESOR
C284001	0	0	0	4	0	0	0	0	0	3,01	SUPCATL
RJ04012	0	0	0	6	0	0	0	0	0	2,86	JANALRIE
EJ03856	2	0	0	7	0	0	0	0	0	2,83	GRCEJE
EJ61345	3	0	0	0	0	0	0	0	0	3,39	GRCEJE
JRIVERZ	1	0	1	4	0	0	0	0	0	3,23	JEFEPOE
ARETE03	1	0	1	4	0	0	0	0	0	3,24	ATPUBS
C268006	10	0	0	1	0	0	0	0	0	3,44	CAJASTL
N346118	0	0	0	10	0	0	0	0	0	2,65	ADMGTS
AV01927	0	0	0	4	0	0	0	0	0	3,09	BFCONSUL
EJOU089	1	1	0	1	0	0	0	0	0	3,50	AUDIOM
EJ05297	0	0	0	6	0	0	0	0	0	2,96	EJECTAS
EJ20186	1	0	0	2	0	0	0	0	0	3,31	EJECTAS
JSTARKA	3	0	0	0	0	0	0	0	0	3,50	AGENTES
DDAROCG	0	0	0	12	0	0	0	0	0	2,52	AGENTES
EJOU112	0	0	0	4	0	0	0	0	0	3,18	AUDIOM
AG03874	5	0	0	0	0	0	0	0	0	3,57	GRCCOB
JNORAMV	1	0	0	4	0	0	0	0	0	3,21	GSENIOR
AR03288	0	0	0	3	0	0	0	0	0	3,27	ANALSRIE
JS01051	1	0	1	4	0	0	0	0	0	3,39	JEFEPOE
AC07300	3	0	0	2	0	0	0	0	0	3,41	GSENIOR
SM03017	3	0	0	1	0	0	0	0	0	3,55	FUERVEN
EV01911	0	0	0	10	0	0	0	0	0	2,85	BFEVASUC
MAGUILA	0	0	0	4	0	0	0	0	0	3,30	BFANAADM
EJ05068	0	0	0	11	0	0	0	0	0	2,81	EJECTAS
CTRULO1	0	0	0	4	0	0	0	0	0	3,31	AUDIOM
TE01623	6	0	1	11	0	0	1	0	0	3,07	TESOR
BBRAVOH	0	0	0	6	0	0	0	0	0	3,19	GSENIOR
FVTA017	0	0	0	5	0	0	0	0	0	3,27	FUERVEN
JLEOND	0	0	0	3	0	0	0	0	0	3,42	GSENIOR

CCONTRC	0	0	0	5	0	0	0	0	0	3,30	NEGINTE
TE01904	2	0	1	12	0	0	0	0	0	2,99	BFTESSUC
EJ04252	3	0	0	8	0	0	0	0	0	3,14	EJECTAS
TE01912	0	0	0	5	0	0	0	0	0	3,34	BFTESSUC
TE01273	0	0	0	10	0	0	0	0	0	3,00	TESOR
EJ05905	0	0	0	9	0	0	0	0	0	3,08	BFEJECOM
RBARRAP	2	0	0	7	0	0	0	0	0	3,27	ANALSRIE
EJ02050	3	1	0	12	0	0	0	0	0	3,08	EJECTAS
FVTA056	0	0	0	5	0	0	0	0	0	3,39	FUERVEN
MLEIVAF	0	0	0	7	0	0	0	0	0	3,28	MERCAP
TE01455	1	0	2	2	0	0	0	0	0	4,00	TESOR
EVALENZ	0	0	0	5	0	0	0	0	0	3,46	TALLSYG
CB05860	0	0	0	5	0	0	0	0	0	3,50	GRCCAS
TE01247	4	0	3	7	0	0	0	0	0	3,98	TESOR
JS01226	5	4	1	14	0	0	0	0	0	3,74	JEFEOPE
RSLAUGH	0	0	0	4	0	0	0	0	0	3,64	JANALRIE
SS12093	3	0	0	1	0	0	0	0	0	3,91	AGENTES
EJ02350	8	8	0	12	0	0	0	0	0	4,28	EJECTAS
TE01366	2	1	2	2	0	0	0	0	0	4,33	TESOR
EJ10288	0	0	0	10	0	0	0	0	0	3,26	EJECTAS
EJ05088	5	0	1	12	0	0	0	0	0	3,37	EJECTAS
TE01292	1	1	0	6	0	0	0	0	0	3,75	TESOR
JS01939	0	0	0	8	0	0	0	0	0	3,43	BFJOPSUC
AG01873	4	0	0	2	0	0	0	0	0	3,93	GRCEJE
MFUENTM	0	0	0	5	0	0	0	0	0	3,65	ADMGT5
GGALASS	1	0	1	5	0	1	0	0	0	3,79	GSENIOR
JS02914	0	0	0	3	0	0	0	0	0	3,82	BFJOPSUC
AV06999	0	0	0	7	0	0	0	0	0	3,53	BFCONSUL
RVICARI	3	0	0	3	0	0	0	0	0	3,89	GSENIOR
SM02021	5	0	0	0	0	0	0	0	0	4,16	FUERVEN
AG01015	8	2	0	3	0	0	0	0	0	4,19	AGENTES
MCARVAD	3	0	0	0	0	0	0	0	0	4,14	UAIPERSO
MCALDEA	6	0	0	14	0	0	0	0	0	3,13	CCENACT
N573151	2	0	1	2	0	0	0	0	0	4,15	BFEJECOM
SU01003	1	0	0	6	0	0	0	0	0	3,68	GRCCAS
AR01300	8	0	0	0	0	0	0	0	0	4,20	ANALSRIE
EJ03327	0	0	0	3	0	0	0	0	0	3,89	EJECTAS
DLEPPES	0	0	0	4	0	0	0	0	0	3,83	AGENTES
JSOTOU	6	0	0	1	0	0	0	0	0	4,11	EJECTAS
C070001	0	0	0	3	0	0	0	0	0	3,92	SUPCATL
AP01003	8	0	2	12	0	0	0	0	0	3,67	PREGIO
ABURGOS	1	3	1	2	0	0	0	0	0	4,68	OFCOMEX
NFERRAN	5	0	0	0	0	0	0	0	0	4,25	GSENIOR
EJ01868	0	0	0	5	0	0	0	0	0	3,81	GRCEJE
HFATTAG	18	0	0	6	0	0	0	0	0	3,39	ANALSRIE
AALBORN	5	0	1	10	0	0	0	0	0	3,70	SUPERVIS
GRIVERO	0	0	0	3	0	0	0	0	0	3,98	DELIVERY
EQUIROZ	0	0	0	7	0	0	0	0	0	3,70	DADMCP0
AR01011	9	0	0	0	0	0	0	0	0	4,30	ANALSRIE
MCARES	0	0	0	3	0	0	0	0	0	4,00	ADMPAMPA
FVTA065	0	0	0	3	0	0	0	0	0	4,00	FUERVEN
AG01214	0	0	0	10	0	0	0	0	0	3,47	AGENTES
GGONZAF	0	0	0	6	0	0	0	0	0	3,80	AUDIOM
EJOU109	6	0	0	10	0	0	0	0	0	3,55	AUDIOM
JS01908	0	0	0	7	0	0	0	0	0	3,75	BFJOPSUC
AG21300	0	0	0	5	0	0	0	0	0	3,88	AGENTES
BK01919	0	0	0	4	0	0	0	0	0	3,97	BFJOPSUC
CB04856	4	0	0	13	0	0	0	0	0	3,39	GRCCAS
EJ07297	1	0	1	3	0	0	0	0	0	4,24	EJECTAS
CF04916	11	0	0	0	0	0	0	0	0	4,45	GRCCAS
AI01392	0	0	0	8	0	0	0	0	0	3,73	CAPTADO
DGATI11	0	0	0	3	0	0	0	0	0	4,09	ADMGT5
EJ01297	0	0	0	14	0	0	0	0	0	3,30	EJECTAS
N730635	0	0	0	6	0	0	0	0	0	3,88	TESOR
PGALA01	0	0	0	12	0	0	0	0	0	3,44	DADMCP0
OX01288	3	0	1	6	0	0	0	0	0	4,09	NEGINTE
AV68999	0	0	0	3	0	0	0	0	0	4,12	BFCONSUL
SM01172	0	0	0	13	0	0	0	0	0	3,41	FUERVEN
EJ05822	0	0	0	5	0	0	0	0	0	3,99	BFEJECOM
EJ49300	1	0	1	7	0	0	0	0	0	4,04	EJECTAS
EJ01243	0	0	0	5	0	0	0	0	0	4,02	EJECTAS
ASELAME	3	1	2	8	0	0	0	0	0	4,34	ATPUBS
TE01101	0	0	0	8	0	0	1	0	0	3,83	TESOR
TPASSIM	4	3	0	3	0	0	0	0	0	4,69	AUDIOM
JS02966	0	0	0	5	0	0	0	0	0	4,04	BFJOPSUC

EJ06411	0	0	0	9	0	0	0	0	0	3,76	EJECTAS
DHERNAA	0	0	0	12	0	0	0	0	0	3,56	BFEJEDEM
PJERIAM	7	0	4	11	0	0	0	0	0	4,35	AGENTES
CFREDO1	0	0	0	8	0	0	0	0	0	3,86	JANALRIE
FVTA061	0	0	0	5	0	0	0	0	0	4,10	FUERVEN
EJ01875	2	0	0	3	0	0	0	0	0	4,28	GRCEJE
AV05936	0	0	0	3	0	0	0	0	0	4,25	BFCONSUL
ATOLEDO	4	0	0	2	0	0	0	0	0	4,39	CLASIRIE
MVERGAM	1	0	0	11	0	0	0	0	0	3,59	JANALRIE
FVTA011	0	0	0	5	0	0	0	0	0	4,13	FUERVEN
SATRIA	0	0	0	6	0	0	0	0	0	4,06	GSENIOR
AC02322	2	0	5	10	0	0	0	0	0	4,59	EJECTAS
SCORRET	0	0	0	3	0	0	0	0	0	4,31	GRCCAS
EJ10241	7	0	0	0	0	0	0	0	0	4,62	GRCEJE
TE01291	1	0	1	7	0	0	0	0	0	4,23	TESOR
EJ80345	3	0	0	0	0	0	0	0	0	4,61	GRCEJE
EJ02252	6	2	0	3	0	0	0	0	0	4,76	EJECTAS
EJ08879	5	0	0	0	0	0	0	0	0	4,65	GRCEJE
SCID	5	3	2	8	0	0	0	0	0	4,89	ATPUBS
EJ59345	0	0	0	8	0	0	0	0	0	4,00	GRCEJE
EJ01237	1	0	0	17	0	0	0	0	0	3,35	GRCEJE
CF04930	0	0	0	6	0	0	0	0	0	4,17	GRCCAS
MFUCHSL	4	2	0	8	0	0	0	0	0	4,43	EJECTAS
C933001	0	0	0	3	0	0	0	0	0	4,39	BFCAJERO
TE01488	11	0	4	15	0	0	0	0	0	4,23	TESOR
TE01088	3	0	2	8	0	0	0	0	0	4,42	TESOR
EJ08288	0	0	0	4	0	0	0	0	0	4,34	EJECTAS
SM01222	0	0	0	8	0	0	0	0	0	4,06	FUERVEN
TE01604	0	0	0	19	0	0	4	0	0	3,23	TESOR
RLARREA	0	0	0	6	0	0	0	0	0	4,25	EJECTAS
JS04234	0	0	0	8	0	0	0	0	0	4,13	JEFEPOE
CGONZAE	2	0	0	2	0	0	0	0	0	4,62	CCENACT
AV16966	0	0	0	7	0	0	0	0	0	4,23	BFCONSUL
FVTA043	2	0	0	1	0	0	0	0	0	4,71	FUERVEN
TE01825	1	0	1	5	0	0	0	0	0	4,59	BFTESSUC
CCORONA	1	0	0	6	0	0	0	0	0	4,35	ANALSRIE
TE01180	1	0	1	16	0	0	0	0	0	3,79	TESOR
SGONZAT	0	0	0	5	0	0	0	0	0	4,44	DADMCPO
GSOTO	0	0	0	6	0	0	0	0	0	4,39	GSENIOR
PGALAZF	0	0	0	5	0	0	0	0	0	4,47	ADMCPPO
N562291	0	0	0	5	0	0	0	0	0	4,48	BFEJECOM
FVTA109	3	0	0	5	0	0	0	0	0	4,52	FUERVEN
C831001	0	0	0	6	0	0	0	0	0	4,42	BFCAJERO
SM01243	0	0	0	10	0	0	0	0	0	4,14	FUERVEN
PTORR12	3	0	2	6	0	0	0	0	0	4,82	PREGIO
JS01288	0	0	0	13	0	0	0	0	0	3,93	JEFEPOE
CF20843	0	0	0	3	0	0	0	0	0	4,66	GRCCAS
MCHAPRU	18	15	6	18	0	0	0	0	0	5,18	EJECTAS
FVTA008	8	0	0	2	0	0	0	0	0	4,81	FUERVEN
VCHAVEG	18	0	7	24	0	0	0	0	0	3,90	AUDIOM
CCON176	4	0	3	11	0	0	0	0	0	4,61	AUDIOM
AAGUILE	0	6	0	0	0	0	0	0	0	5,91	GSENIOR
PT01915	0	0	0	7	0	0	0	0	0	4,40	BFEJECOM
N359455	0	0	0	9	0	0	0	0	0	4,26	EJECTAS
CF01904	0	0	0	5	0	0	0	0	0	4,57	GRCCAS
TE01368	8	0	8	15	0	0	0	0	0	4,96	TESOR
CVARG02	5	0	1	13	0	0	0	0	0	4,23	NEGINTE
N450601	0	0	0	4	0	0	0	0	0	4,67	JEFEPOE
EJ02561	4	2	2	14	0	0	0	0	0	4,45	EJECTAS
TE01389	1	0	1	11	0	0	0	0	0	4,34	TESOR
N422347	2	0	0	7	0	0	0	0	0	4,46	AGENTES
CB01886	0	0	0	10	0	0	0	0	0	4,27	GRCCAS
JMUNO16	0	0	0	6	0	0	0	0	0	4,55	DIRERIES
SPIZARG	10	5	0	11	0	0	0	0	0	4,91	EJECTAS
AG07288	0	0	0	9	0	0	0	0	0	4,39	AGENTES
TE01910	0	0	0	10	0	0	0	0	0	4,33	BFTESSUC
EJOU040	1	0	0	6	0	0	0	0	0	4,65	AUDIOM
FBUSTAM	0	0	0	4	0	0	0	0	0	4,76	AGENTES
TE01930	0	0	0	5	0	0	0	0	0	4,72	BFTESSUC
MRATTOD	0	0	0	11	0	0	0	0	0	4,30	MERCAP
CF11936	0	0	0	3	0	0	0	0	0	4,87	GRCCAS
AI01297	0	0	0	11	0	0	0	0	0	4,30	CAPTADO
N431695	2	0	0	5	0	0	0	0	0	4,76	AGENTES
EJ03881	6	0	0	0	0	0	0	0	0	5,18	GRCEJE
TE02180	0	0	0	12	0	0	0	0	0	4,27	TESOR

C332001	11	0	0	0	0	0	0	0	0	5,21	SUPCATL
LBALMDE	0	0	0	4	0	0	0	0	0	4,86	AGENTES
AV60999	0	0	0	3	0	0	0	0	0	4,94	BFCONSUL
CF01936	0	0	0	7	0	0	0	0	0	4,67	GRCCAS
MSALAMA	2	1	1	2	0	0	0	0	0	5,40	NEGINTE
JS01365	0	0	0	8	0	0	0	0	0	4,61	JEFEOPE
N373190	0	0	0	4	0	0	0	0	0	4,90	BFEJECOM
AR02234	13	0	0	0	0	0	0	0	0	5,22	GRCEJE
SM01179	0	0	0	4	0	0	0	0	0	4,91	FUERVEN
SHENRIV	3	0	0	8	0	0	0	0	0	4,64	ANALSRIE
AALLENS	0	0	0	10	0	0	0	0	0	4,48	DADMCPO
AV10966	0	0	0	3	0	0	0	0	0	5,00	BFCONSUL
TE01363	7	0	1	8	0	0	0	0	0	4,93	TESOR
N571300	0	0	0	9	0	0	1	0	0	4,60	TESOR
KVEJARR	5	0	1	6	0	0	0	0	0	5,09	AUDIOM
TE01615	3	0	6	8	0	0	0	0	0	5,66	TESOR
AG05170	3	0	1	6	0	0	0	0	0	5,07	AGENTES
AG03442	2	0	0	2	0	0	0	0	0	5,22	AGENTES
CH01928	0	0	0	9	0	0	0	0	0	4,71	BFEJECOM
GCARR01	8	0	0	3	0	0	0	0	0	5,15	ANALSRIE
SM01235	4	0	0	0	0	0	0	0	0	5,45	FUERVEN
N572748	5	0	11	14	1	1	1	0	0	5,62	TESOR
MIBANES	0	0	0	10	0	0	0	0	0	4,67	GRCCAS
EJ07010	3	0	0	0	0	0	0	0	0	5,48	EJECTAS
N470905	0	0	0	10	0	0	0	0	0	4,69	AGENTES
GAEDO	1	0	1	14	0	0	0	0	0	4,58	ATPUBS
MZAMORG	2	0	1	9	0	0	0	0	0	4,99	AUDIOM
EJ04249	1	1	0	17	0	0	0	0	0	4,40	EJECTAS
fraude JTAPI	1	0	0	5	2	1	0	1	1	5,21	AUDIOM
SM04011	4	0	0	14	0	0	0	0	0	4,48	FUERVEN
EJ01322	0	0	0	6	0	0	0	0	0	5,07	EJECTAS
TE01396	0	0	0	9	0	0	0	0	0	4,86	TESOR
JS03183	5	0	1	6	0	0	0	0	0	5,36	JEFEOPE
C929001	0	0	0	6	0	0	0	0	0	5,12	BFCAJERO
SM02243	6	0	0	3	0	0	0	0	0	5,42	FUERVEN
DJADELR	1	0	1	10	0	0	0	0	0	5,02	AGENTES
AG07170	5	0	1	10	0	0	0	0	0	5,01	AGENTES
SM02186	0	0	0	3	0	0	0	0	0	5,38	FUERVEN
MGOME11	11	1	5	21	2	1	1	0	0	4,28	PREGIO
EJ04487	10	7	0	8	0	0	0	0	0	6,15	EJECTAS
EJ11266	6	1	0	14	0	0	0	0	0	4,82	EJECTAS
SM02234	5	0	0	6	0	0	0	0	0	5,34	FUERVEN
EJ04261	2	0	1	7	0	0	0	0	0	5,37	EJECTAS
TE01283	8	0	8	19	0	0	0	0	0	5,40	TESOR
TE01614	0	0	0	10	0	0	3	0	0	4,97	TESOR
N451798	0	0	0	6	0	0	0	0	0	5,27	TARPAMPN
AV14929	0	0	0	8	0	0	0	0	0	5,14	BFCONSUL
EJ05339	5	3	0	15	0	0	0	0	0	4,99	EJECTAS
EJOU094	12	0	0	0	0	0	0	0	0	5,87	AUDIOM
MCANETS	0	0	0	6	0	0	0	0	0	5,30	BFCONSUL

ANEXO Q: LISTA COMPLETA DE SOSPECHOSOS ENCONTRADOS POR LOS TRES MODELOS

User_id	Datos básicos	Salidos	Ult. Movimientos	Loc. Personas	Talones	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiempo trans.	Perfil	N° Apariciones
GGALASS	1	0	1	5	0	1	0	0	0	3,79	GSENIOR	3
TE02248	3	0	1	7	2	2	1	0	0	2,22	TESOR	3
FLARA	0	0	0	6	0	1	0	0	0	1,61	SCLBTR	3
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS	3
HFATTAG	18	0	0	6	0	0	0	0	0	3,39	ANALSRIE	2
TE01904	2	0	1	12	0	0	0	0	0	2,99	BFTESSUC	2
SU01003	1	0	0	6	0	0	0	0	0	3,68	GRCCAS	2
TE01292	1	1	0	6	0	0	0	0	0	3,75	TESOR	2
TE01101	0	0	0	8	0	0	1	0	0	3,83	TESOR	2
GPOZOAR	2	2	0	3	0	0	0	0	0	2,42	GSENIOR	2
BSILV01	0	0	0	3	0	0	0	0	0	2,66	GSENIOR	2
EI03327	0	0	0	3	0	0	0	0	0	3,89	EJECTAS	2
AG01238	0	0	0	3	0	0	0	0	0	2,57	AGENTES	2
PDEMEDP	4	0	0	3	0	0	0	0	0	1,72	ANALSRIE	2
CRIVERP	3	0	1	4	0	0	0	0	0	2,00	EJECTAS	2
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR	2
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR	2
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR	2
EI23300	5	0	0	3	0	0	0	0	0	2,02	GRCEJE	2
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NEGINTE	2
EJOU052	3	0	0	0	0	0	0	0	0	1,83	AUDIOM	2
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVEN	2
DP61002	1	1	0	4	0	0	0	0	0	2,03	DADMCP0	2
C178001	3	0	0	1	0	0	0	0	0	2,04	SUPCATL	2
AR31300	6	0	0	0	0	0	0	0	0	1,71	ANALSRIE	2
CFUERTH	3	0	0	8	0	0	0	0	0	1,91	AGENTES	2
CMARIAN	7	0	0	0	0	0	0	0	0	1,12	JANALRIE	2
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVEN	2
CSUAZOD	5	0	1	8	0	0	0	0	0	2,08	OFCOMEX	2
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR	2
TE02043	6	1	3	9	0	0	0	0	0	1,35	TESOR	2
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR	2
JS01046	2	0	1	3	1	0	0	0	0	2,46	JEFE0PE	2
AG05297	1	0	1	2	0	0	0	0	0	2,00	AGENTES	2
TE01213	1	0	1	2	0	0	0	0	0	2,00	TESOR	2
CB16879	6	0	0	3	0	0	0	0	0	2,19	GRCCAS	2
AG02032	0	0	0	5	0	0	0	0	0	1,37	AGENTES	2
AG02157	0	0	0	3	0	0	0	0	0	1,40	AGENTES	2
N328981	3	0	0	4	0	0	0	0	0	2,35	EJECTAS	2
EI01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE	2
EI17170	0	0	0	3	0	0	1	0	0	1,38	GSENIOR	2
DLOPEZV	0	0	0	4	0	0	0	0	0	1,47	FMLOCAL	2
N596659	0	0	0	3	0	0	0	0	0	1,50	TESOR	2
NREYE01	0	0	0	5	0	0	0	0	0	1,46	SCLBTR	2
PALVARM	0	0	0	6	0	0	0	0	0	1,46	EJECTAS	2
JREYESI	0	0	0	3	0	0	2	0	0	1,87	ATPUBS	2
TE02198	0	0	0	6	0	0	0	0	0	1,53	TESOR	2
CB03860	0	0	0	4	0	0	0	0	0	1,74	GRCCAS	2
RPARRAP	0	0	0	3	0	0	0	0	0	1,82	EJECTAS	2
JS01300	0	0	0	5	0	0	3	0	0	2,00	JEFE0PE	2
EFIQUUC	0	0	0	5	0	0	0	0	0	1,81	EJECTAS	2
CF04913	0	0	0	7	0	0	0	0	0	1,77	GRCCAS	2
N451798	0	0	0	6	0	0	0	0	0	5,27	FARPAMPN	2
C929001	0	0	0	6	0	0	0	0	0	5,12	BFCAJERO	2
MCANETS	0	0	0	6	0	0	0	0	0	5,30	BFCONSUL	2
EI01322	0	0	0	6	0	0	0	0	0	5,07	EJECTAS	2
SM01179	0	0	0	4	0	0	0	0	0	4,91	FUERVEN	2
N373190	0	0	0	4	0	0	0	0	0	4,90	BFEJECOM	2
LBALMDE	0	0	0	4	0	0	0	0	0	4,86	AGENTES	2
SM02186	0	0	0	3	0	0	0	0	0	5,38	FUERVEN	2
N431695	2	0	0	5	0	0	0	0	0	4,76	AGENTES	2
AV10966	0	0	0	3	0	0	0	0	0	5,00	BFCONSUL	2
AV60999	0	0	0	3	0	0	0	0	0	4,94	BFCONSUL	2
TE01930	0	0	0	5	0	0	0	0	0	4,72	BFTESSUC	2
FBUSTAM	0	0	0	4	0	0	0	0	0	4,76	AGENTES	2
EIOU040	1	0	0	6	0	0	0	0	0	4,65	AUDIOM	2
CF11936	0	0	0	3	0	0	0	0	0	4,87	GRCCAS	2
N450601	0	0	0	4	0	0	0	0	0	4,67	JEFE0PE	2
AG03442	2	0	0	2	0	0	0	0	0	5,22	AGENTES	2
AV14929	0	0	0	8	0	0	0	0	0	5,14	BFCONSUL	2
C919003	0	0	0	4	0	0	0	0	0	1,95	BFCAJERO	2
EI01316	0	0	0	3	0	0	0	0	0	1,99	EJECTAS	2
CB1A345	0	0	0	4	0	0	0	0	0	2,00	GRCCAS	2
AI01332	0	0	0	5	0	0	0	0	0	1,96	CAPTADO	2
EI07012	18	1	6	80	0	0	2	0	0	17,44	EJECTAS	1
EI12411	20	5	0	186	0	0	1	0	0	14,10	EJECTAS	1
CCONT82	16	0	2	27	5	1	0	1	0	7,68	AUDIOM	1
MORDEP	74	1	76	102	12	12	0	0	0	9,98	GSENIOR	1
ADIA204	11	2	0	130	0	0	0	0	0	11,22	EJECTAS	1
EI29345	18	0	3	145	3	0	0	0	0	13,27	DADMCP0	1
KROJOMA	11	1	1	117	0	1	0	0	0	14,58	DADMCP0	1
AG02016	40	6	0	148	0	0	5	0	0	13,65	AGENTES	1
EI02237	16	1	11	221	2	3	0	0	0	10,93	EJECTAS	1
EI10024	18	3	6	196	0	0	29	0	0	10,47	EJECTAS	1
EI02325	30	5	0	125	0	0	1	0	0	14,91	EJECTAS	1
N470528	18	8	1	172	0	0	0	0	0	11,44	EJECTAS	1
EI06024	31	1	5	113	0	0	3	0	0	15,05	EJECTAS	1
N493886	28	2	7	141	0	0	2	0	0	15,87	EJECTAS	1
AS01119	21	0	4	142	0	1	1	0	0	14,10	ATPUBS	1
AROME01	9	0	0	6	11	2	0	0	0	30,54	ATPUBS	1

AG02345	3	0	0	117	0	0	0	0	0	7,49	GRCCOB	1
TE01924	3	0	16	6	0	0	0	0	0	22,45	BFTESSUC	1
TE01565	24	1	7	64	10	23	1	0	23	9,11	TESOR	1
N604834	7	0	0	135	8	0	0	0	0	13,65	ATPUBS	1
TE02331	1	0	1	3	0	0	0	0	0	7,00	TESOR	1
AS01148	3	0	0	122	4	0	1	0	0	16,55	ATPUBS	1
EJ06183	15	0	1	135	0	0	6	0	0	12,61	EJECTAS	1
KMOREIR	35	0	0	175	0	0	0	0	0	16,42	ANALISTA	1
TE01021	2	0	1	9	1	1	0	0	0	20,17	TESOR	1
N421760	9	0	0	25	4	0	0	0	0	16,20	ATPUBS	1
TE01293	4	0	3	18	0	0	1	0	1	6,89	TESOR	1
FCORTEZ	4	0	3	160	0	0	24	0	0	10,33	REGPAMPA	1
THERRER	13	0	0	76	0	0	0	0	0	13,68	EJECTAS	1
IRUBILZ	10	1	4	22	0	4	0	0	4	10,02	AUDIOM	1
ASANMAA	12	0	0	216	0	0	0	0	0	13,29	AGENTES	1
MGONZ13	21	1	8	243	1	2	0	0	0	13,24	DADMCPO	1
EJ02308	13	3	2	138	0	0	0	0	0	14,41	EJECTAS	1
SBUNSTE	14	0	0	0	0	0	0	0	0	19,78	AGENTES	1
EFLORES	16	0	0	204	0	4	0	0	0	15,68	DADMCPO	1
EPINOZA	18	0	0	2	0	0	0	0	0	30,70	CAPTADO	1
AMARQUM	6	0	2	29	4	1	0	1	0	20,20	ATPUBS	1
N570787	2	1	1	52	0	1	0	0	0	13,24	ATPUBS	1
GGUERRV	9	0	0	84	2	2	0	0	0	14,87	DADMCPO	1
JS01048	7	0	0	40	5	0	1	0	0	10,57	JEFOEPE	1
PCHACAM	7	0	6	174	0	0	0	0	0	15,01	CCENACT	1
AS07002	1	0	0	6	2	1	0	0	0	17,94	ATPUBS	1
MCRISOR	2	0	1	7	2	1	0	0	0	36,15	ATPUBS	1
SMELIPB	1	0	6	2	0	0	0	0	0	15,63	GSENIOR	1
SDURRUZ	2	1	0	4	0	1	0	0	0	10,57	GSENIOR	1
SJARA02	0	0	0	173	0	0	1	0	0	15,32	DEPCUE	1
OVASQUP	0	39	0	1	0	0	0	0	0	12,11	GSENIOR	1
SOCHOAP	0	0	0	27	0	0	3	0	0	17,95	FMLOCAL	1
AR26300	0	0	0	58	0	0	0	0	0	8,33	ANALSRIE	1
TE04300	0	46	0	0	0	0	0	0	0	18,96	TESOR	1
EJ02907	0	0	0	26	0	1	0	0	0	7,90	BFEJECOM	1
RMARTIR	0	0	0	165	0	0	0	0	0	17,26	BFEJECOM	1
N801015	0	0	0	66	0	1	0	0	0	14,77	BFEJECOM	1
AG01153	0	0	0	9	0	0	0	0	0	19,45	AGENTES	1
N604137	0	0	0	17	0	0	2	0	0	12,73	TESOR	1
SRAMIRA	0	0	0	12	0	12	0	0	0	9,65	SCLBTR	1
EJ06962	0	0	0	34	0	1	1	0	0	11,72	BFEJECOM	1
N364920	0	0	0	29	0	1	0	0	0	12,36	BFEJECOM	1
MZUNIGA	0	0	0	2	0	0	2	0	0	19,67	EJEVIR	1
C600001	20	0	0	25	0	0	0	0	0	9,46	SUPCATL	1
EJOU070	21	0	0	24	0	0	0	0	0	7,43	AUDIOM	1
N568224	30	0	0	26	0	0	0	0	0	10,08	AGENTES	1
AG03393	17	0	1	26	0	0	0	0	0	9,23	AGENTES	1
EJ01183	18	0	0	28	0	0	0	0	0	7,96	EJECTAS	1
C626001	21	0	0	15	0	0	0	0	0	7,76	SUPCATL	1
AR29300	18	0	0	30	0	0	0	0	0	9,73	ANALSRIE	1
AG03067	17	1	1	26	0	0	0	0	0	8,56	AGENTES	1
AG01002	25	3	1	20	0	0	0	0	0	9,81	AGENTES	1
AR38300	18	0	0	16	0	0	0	0	0	9,36	ANALSRIE	1
AG01247	17	0	0	18	0	0	0	0	0	8,51	AGENTES	1
N431443	27	0	0	14	0	0	0	0	0	10,45	ANALSRIE	1
N722241	22	0	0	36	0	0	0	0	0	9,98	AGENTES	1
AG02399	17	0	3	22	0	0	0	0	0	9,98	AGENTES	1
EJ02877	21	0	0	11	0	0	0	0	0	8,77	GRCEJE	1
AR11234	18	2	0	23	0	0	0	0	0	10,25	ANALSRIE	1
EJ05020	25	3	0	26	0	0	3	0	0	9,63	EJECTAS	1
C077003	29	0	0	33	0	0	0	0	0	10,71	CAJASTL	1
EJ32345	25	0	0	33	0	0	0	0	0	11,44	GRCEJE	1
AG02162	18	0	1	24	0	0	0	0	0	11,46	AGENTES	1
PGALVEM	16	0	0	27	0	0	0	0	0	10,64	ANALISTA	1
KPRADOD	27	0	4	36	0	0	0	0	0	10,19	AUDIOM	1
AG01861	28	0	0	10	0	0	0	0	0	9,88	GRCEJE	1
AG01467	18	0	0	22	0	0	0	0	0	11,47	AGENTES	1
AG01207	18	0	3	18	0	0	0	0	0	11,04	AGENTES	1
EJ05252	26	2	0	34	0	0	0	0	0	10,91	EJECTAS	1
EJ02873	21	0	0	11	0	0	0	0	0	10,72	GRCEJE	1
C055001	26	0	0	36	0	0	0	0	0	11,12	SUPCATL	1
EJOUT56	20	0	0	30	0	0	0	0	0	11,74	AUDIOM	1
AG01269	24	0	3	40	0	0	0	0	0	9,18	AGENTES	1
C004031	1	0	0	7	0	0	0	0	0	5,29	CAJASTL	1
RR0MERN	2	0	0	5	0	0	0	0	0	5,62	GSENIOR	1
TE03193	0	0	0	7	0	0	0	0	0	5,33	TESOR	1
C386001	1	0	0	4	0	0	0	0	0	5,71	SUPCATL	1
C043001	1	0	0	7	0	0	0	0	0	5,60	SUPCATL	1
LSEPU04	0	0	0	3	0	0	0	0	0	5,56	ANALSRIE	1
C207001	0	0	0	5	0	0	0	0	0	5,75	SUPCATL	1
AV01966	0	0	0	3	0	0	0	0	0	5,63	BFCONSUL	1
JS03220	0	0	0	4	0	0	1	0	0	5,51	JEFOEPE	1
N484117	0	0	0	6	0	0	0	0	0	5,80	BFJOFUC	1
EJ02970	0	0	0	8	0	0	0	0	0	5,41	BFEJECOM	1
fraude CCON643	4	1	1	4	1	1	0	0	0	1,00	AUDIOM	1
N484075	4	1	5	4	0	0	0	0	0	2,65	EJECTAS	1
JINZUNZ	6	0	0	0	0	0	0	0	0	2,09	GRCCOB	1
C040001	6	0	0	1	0	0	0	0	0	2,25	SUPCATL	1
AP02186	0	0	0	3	0	0	0	0	0	2,06	DADMCPO	1
N484327	1	0	1	2	0	0	0	0	0	2,33	TESOR	1

AS01231	1	0	2	2	0	0	0	0	0	2,50	ATPUBS	1
AC02234	1	0	1	4	0	0	0	0	0	2,22	EJECTAS	1
N449286	0	0	0	4	0	0	0	0	0	2,08	BFEJECOM	1
CGUTIEL	2	0	5	4	0	0	0	0	0	2,93	GSENIOR	1
FVTA167	2	0	0	3	0	0	0	0	0	2,23	FUERVERN	1
EJ02319	0	0	0	4	0	0	0	0	0	2,11	EJECTAS	1
TE02087	0	0	0	8	0	0	0	0	0	1,85	TESOR	1
CB71345	2	0	0	1	0	0	0	0	0	2,39	GRCCAS	1
SU02856	1	0	0	3	0	0	0	0	0	2,25	GRCCAS	1
JRIVE12	2	0	2	8	0	0	0	0	0	2,23	GSENIOR	1
MMIRA01	3	0	0	0	0	0	0	0	0	2,56	EJECTAS	1
PTAMAYO	0	0	0	4	0	0	0	0	0	2,22	DADMCP0	1
CF11913	0	0	0	5	0	0	0	0	0	2,15	GRCCAS	1
MBALT01	0	0	0	4	0	0	0	0	0	2,22	SCLBTR	1
EJ45345	9	0	0	2	0	0	0	0	0	2,45	GRCEJE	1
EJ04013	3	0	0	2	0	0	0	0	0	2,48	GRCEJE	1
RJ01288	0	0	0	5	0	0	0	0	0	2,21	JANALRIE	1
FOTONDO	0	0	0	4	0	0	0	0	0	2,31	JANALRIE	1
DLEPE	3	0	0	0	0	0	0	0	0	2,67	GSENIOR	1
CSAEZGO	1	0	1	2	0	0	0	0	0	2,67	PROCENT	1
AV13966	0	0	0	6	0	0	0	0	0	2,19	BFCONSUL	1
JSEIN	2	0	0	3	0	0	0	0	0	2,46	ANALISTA	1
TE01427	6	0	4	10	0	0	0	0	0	2,63	TESOR	1
TE01992	0	0	0	4	0	0	0	0	0	2,35	BFTESSUC	1
AG02248	1	0	0	6	0	0	0	0	0	2,20	AGENTES	1
CH02297	0	0	0	6	0	0	0	0	0	2,22	EJECTAS	1
TE01647	0	0	0	3	0	0	0	0	0	2,44	TESOR	1
MRICCB	2	0	0	2	0	0	0	0	0	2,57	ANALSRIE	1
EJ11874	1	0	0	2	0	0	0	0	0	2,56	GRCEJE	1
JS01903	0	0	0	3	0	0	0	0	0	2,50	BFJOPSUC	1
N364581	0	0	0	3	0	0	0	0	0	2,51	BFCAJERO	1
RECO142	0	0	0	3	0	0	0	0	0	2,53	AUDIOM	1
MESPINP	0	0	0	3	0	0	0	0	0	2,53	EJECTAS	1
MLLANOC	0	0	0	5	0	0	0	0	0	2,40	NEGINTE	1
EJ21288	0	0	0	3	0	0	0	0	0	2,54	EJECTAS	1
TE01032	0	0	0	3	0	0	0	0	0	2,57	TESOR	1
JS29300	4	0	1	8	0	0	1	0	0	2,45	JEFEPOE	1
AG01445	2	0	1	5	0	0	0	0	0	2,64	AGENTES	1
AC03297	0	0	0	4	0	0	0	0	0	2,52	EJECTAS	1
FVTA115	0	0	0	4	0	0	0	0	0	2,55	FUERVERN	1
AV04905	0	0	0	4	0	0	0	0	0	2,55	BFCONSUL	1
CH06300	0	0	0	3	0	0	0	0	0	2,64	EJECTAS	1
C045001	2	0	0	5	0	0	0	0	0	2,56	SUPCATL	1
AOYARCC	0	0	0	4	0	0	0	0	0	2,61	BFCONSUL	1
LROMERS	0	0	0	4	0	0	0	0	0	2,61	GSENIOR	1
JTORREA	1	0	0	3	0	0	0	0	0	2,72	GSENIOR	1
NGONZAD	5	0	0	0	0	0	0	0	0	3,03	GSENIOR	1
CF22843	8	0	0	0	0	0	0	0	0	3,07	GRCCAS	1
CB01870	2	0	0	3	0	0	0	0	0	2,81	GRCCAS	1
EJ11402	0	0	0	6	0	0	0	0	0	2,57	EJECTAS	1
CT66345	0	0	0	5	0	0	0	0	0	2,65	GRCCAS	1
AP01045	0	0	0	5	0	0	0	0	0	2,65	ADMCP0	1
C031006	0	0	0	7	0	0	0	0	0	2,52	CAJASTL	1
C182001	3	0	0	3	0	0	0	0	0	2,88	SUPCATL	1
MBILBAB	0	0	0	6	0	0	0	0	0	2,63	MERCAP	1
MMIQU01	0	0	0	3	0	0	0	0	0	2,86	BFCONSUL	1
TE01195	6	0	1	10	0	0	0	0	0	2,64	TESOR	1
GNEIR01	0	0	0	5	0	0	0	0	0	2,77	EJECTAS	1
VDQUUME	0	0	0	6	0	0	0	0	0	2,71	EJECTAS	1
AG01443	3	2	3	7	0	0	1	0	0	3,44	AGENTES	1
CF16823	0	0	0	3	0	0	0	0	0	2,95	GRCCAS	1
C174002	3	0	0	4	0	0	0	0	0	2,96	SUPCATL	1
C024001	3	0	0	6	0	0	0	0	0	2,86	SUPCATL	1
TE01192	1	0	1	2	0	0	0	0	0	3,33	TESOR	1
C284001	0	0	0	4	0	0	0	0	0	3,01	SUPCATL	1
RJ04012	0	0	0	6	0	0	0	0	0	2,86	JANALRIE	1
EJ03856	2	0	0	7	0	0	0	0	0	2,83	GRCEJE	1
EJ61345	3	0	0	0	0	0	0	0	0	3,39	GRCEJE	1
JRIVERZ	1	0	1	4	0	0	0	0	0	3,23	JEFEPOE	1
ARETE03	1	0	1	4	0	0	0	0	0	3,24	ATPUBS	1
C268006	10	0	0	1	0	0	0	0	0	3,44	CAJASTL	1
N346118	0	0	0	10	0	0	0	0	0	2,65	ADMGT5	1
AV01927	0	0	0	4	0	0	0	0	0	3,09	BFCONSUL	1
EJOU089	1	1	0	1	0	0	0	0	0	3,50	AUDIOM	1
EJ05297	0	0	0	6	0	0	0	0	0	2,96	EJECTAS	1
EJ20186	1	0	0	2	0	0	0	0	0	3,31	EJECTAS	1
JSTARKA	3	0	0	0	0	0	0	0	0	3,50	AGENTES	1
DDAROCG	0	0	0	12	0	0	0	0	0	2,52	AGENTES	1
EJOU112	0	0	0	4	0	0	0	0	0	3,18	AUDIOM	1
AG03874	5	0	0	0	0	0	0	0	0	3,57	GRCCOB	1
JNORAMV	1	0	0	4	0	0	0	0	0	3,21	GSENIOR	1
AR03288	0	0	0	3	0	0	0	0	0	3,27	ANALSRIE	1
JS01051	1	0	1	4	0	0	0	0	0	3,39	JEFEPOE	1
AC07300	3	0	0	2	0	0	0	0	0	3,41	GSENIOR	1
SM03017	3	0	0	1	0	0	0	0	0	3,55	FUERVERN	1
EV01911	0	0	0	10	0	0	0	0	0	2,85	BFEVASUC	1
MAGUJLA	0	0	0	4	0	0	0	0	0	3,30	BFANAADN	1
EJ05068	0	0	0	11	0	0	0	0	0	2,81	EJECTAS	1
CTRUL01	0	0	0	4	0	0	0	0	0	3,31	AUDIOM	1
TE01623	6	0	1	11	0	0	1	0	0	3,07	TESOR	1

BBRAVOH	0	0	0	6	0	0	0	0	0	3,19	GSENIOR	1
FVTA017	0	0	0	5	0	0	0	0	0	3,27	FUERVERN	1
JLEOND	0	0	0	3	0	0	0	0	0	3,42	GSENIOR	1
CCONTRC	0	0	0	5	0	0	0	0	0	3,30	NEGINTE	1
EJ04252	3	0	0	8	0	0	0	0	0	3,14	EJECTAS	1
TE01912	0	0	0	5	0	0	0	0	0	3,34	BFTESSUC	1
TE01273	0	0	0	10	0	0	0	0	0	3,00	TESOR	1
EJ05905	0	0	0	9	0	0	0	0	0	3,08	BFEJECOM	1
RBARRAP	2	0	0	7	0	0	0	0	0	3,27	ANALSRIE	1
EJ02050	3	1	0	12	0	0	0	0	0	3,08	EJECTAS	1
FVTA056	0	0	0	5	0	0	0	0	0	3,39	FUERVERN	1
MLEIVAF	0	0	0	7	0	0	0	0	0	3,28	MERCAP	1
TE01455	1	0	2	2	0	0	0	0	0	4,00	TESOR	1
EVALENZ	0	0	0	5	0	0	0	0	0	3,46	TALLSYG	1
CB05860	0	0	0	5	0	0	0	0	0	3,50	GRCCAS	1
TE01247	4	0	3	7	0	0	0	0	0	3,98	TESOR	1
JS01226	5	4	1	14	0	0	0	0	0	3,74	JEFEOPE	1
RSLAUGH	0	0	0	4	0	0	0	0	0	3,64	JANALRIE	1
SS12093	3	0	0	1	0	0	0	0	0	3,91	AGENTES	1
EJ02350	8	8	0	12	0	0	0	0	0	4,28	EJECTAS	1
TE01366	2	1	2	2	0	0	0	0	0	4,33	TESOR	1
EJ10288	0	0	0	10	0	0	0	0	0	3,26	EJECTAS	1
EJ05088	5	0	1	12	0	0	0	0	0	3,37	EJECTAS	1
JS01939	0	0	0	8	0	0	0	0	0	3,43	BFJOPSUC	1
AG01873	4	0	0	2	0	0	0	0	0	3,93	GRCEJE	1
MFUENTM	0	0	0	5	0	0	0	0	0	3,65	ADMGTS	1
JS02914	0	0	0	3	0	0	0	0	0	3,82	BFJOPSUC	1
AV06999	0	0	0	7	0	0	0	0	0	3,53	BFCONSUL	1
RVICARI	3	0	0	3	0	0	0	0	0	3,89	GSENIOR	1
SM02021	5	0	0	0	0	0	0	0	0	4,16	FUERVERN	1
AG01015	8	2	0	3	0	0	0	0	0	4,19	AGENTES	1
MCARVAD	3	0	0	0	0	0	0	0	0	4,14	UAIPERSO	1
MCALDEA	6	0	0	14	0	0	0	0	0	3,13	CCENACT	1
N573151	2	0	1	2	0	0	0	0	0	4,15	BFEJECOM	1
AR01300	8	0	0	0	0	0	0	0	0	4,20	ANALSRIE	1
DLEPPES	0	0	0	4	0	0	0	0	0	3,83	AGENTES	1
JSOTOU	6	0	0	1	0	0	0	0	0	4,11	EJECTAS	1
CO70001	0	0	0	3	0	0	0	0	0	3,92	SUPCATL	1
AP01003	8	0	2	12	0	0	0	0	0	3,67	PREGIO	1
ABURGOS	1	3	1	2	0	0	0	0	0	4,68	OFCOMEX	1
NFERRAN	5	0	0	0	0	0	0	0	0	4,25	GSENIOR	1
EJ01868	0	0	0	5	0	0	0	0	0	3,81	GRCEJE	1
AALBORN	5	0	1	10	0	0	0	0	0	3,70	SUPERVIS	1
GRIVERO	0	0	0	3	0	0	0	0	0	3,98	DELIVERY	1
EQUIROZ	0	0	0	7	0	0	0	0	0	3,70	DADMCPD	1
AR01011	9	0	0	0	0	0	0	0	0	4,30	ANALSRIE	1
MCARES	0	0	0	3	0	0	0	0	0	4,00	ADMAMP	1
FVTA065	0	0	0	3	0	0	0	0	0	4,00	FUERVERN	1
AG01214	0	0	0	10	0	0	0	0	0	3,47	AGENTES	1
GGONZAF	0	0	0	6	0	0	0	0	0	3,80	AUDIOM	1
EJOU109	6	0	0	10	0	0	0	0	0	3,55	AUDIOM	1
JS01908	0	0	0	7	0	0	0	0	0	3,75	BFJOPSUC	1
AG21300	0	0	0	5	0	0	0	0	0	3,88	AGENTES	1
BK01919	0	0	0	4	0	0	0	0	0	3,97	BFJOPSUC	1
CB04856	4	0	0	13	0	0	0	0	0	3,39	GRCCAS	1
EJ07297	1	0	1	3	0	0	0	0	0	4,24	EJECTAS	1
CF04916	11	0	0	0	0	0	0	0	0	4,45	GRCCAS	1
AI01392	0	0	0	8	0	0	0	0	0	3,73	CAPTADO	1
DGATI11	0	0	0	3	0	0	0	0	0	4,09	ADMGTS	1
EJ01297	0	0	0	14	0	0	0	0	0	3,30	EJECTAS	1
N730635	0	0	0	6	0	0	0	0	0	3,88	TESOR	1
PGALA01	0	0	0	12	0	0	0	0	0	3,44	DADMCPD	1
OX01288	3	0	1	6	0	0	0	0	0	4,09	NEGINTE	1
AV68999	0	0	0	3	0	0	0	0	0	4,12	BFCONSUL	1
SM01172	0	0	0	13	0	0	0	0	0	3,41	FUERVERN	1
EJ05822	0	0	0	5	0	0	0	0	0	3,99	BFEJECOM	1
EJ49300	1	0	1	7	0	0	0	0	0	4,04	EJECTAS	1
EJ01243	0	0	0	5	0	0	0	0	0	4,02	EJECTAS	1
ASELAME	3	1	2	8	0	0	0	0	0	4,34	ATPUBS	1
TPASSIM	4	3	0	3	0	0	0	0	0	4,69	AUDIOM	1
JS02966	0	0	0	5	0	0	0	0	0	4,04	BFJOPSUC	1
EJ06411	0	0	0	9	0	0	0	0	0	3,76	EJECTAS	1
DHERNAA	0	0	0	12	0	0	0	0	0	3,56	BFEJEDEM	1
PJERIAM	7	0	4	11	0	0	0	0	0	4,35	AGENTES	1
CFRED01	0	0	0	8	0	0	0	0	0	3,86	JANALRIE	1
FVTA061	0	0	0	5	0	0	0	0	0	4,10	FUERVERN	1
EJ01875	2	0	0	3	0	0	0	0	0	4,28	GRCEJE	1
AV05936	0	0	0	3	0	0	0	0	0	4,25	BFCONSUL	1
ATOLEDO	4	0	0	2	0	0	0	0	0	4,39	CLASIRIE	1
MVERGAM	1	0	0	11	0	0	0	0	0	3,59	JANALRIE	1
FVTA011	0	0	0	5	0	0	0	0	0	4,13	FUERVERN	1
SATRIA	0	0	0	6	0	0	0	0	0	4,06	GSENIOR	1
AC02322	2	0	5	10	0	0	0	0	0	4,59	EJECTAS	1
SCORTET	0	0	0	3	0	0	0	0	0	4,31	GRCCAS	1
EJ10241	7	0	0	0	0	0	0	0	0	4,62	GRCEJE	1
TE01291	1	0	1	7	0	0	0	0	0	4,23	TESOR	1
EJ80345	3	0	0	0	0	0	0	0	0	4,61	GRCEJE	1
EJ02252	6	2	0	3	0	0	0	0	0	4,76	EJECTAS	1
EJ08879	5	0	0	0	0	0	0	0	0	4,65	GRCEJE	1
SCID	5	3	2	8	0	0	0	0	0	4,89	ATPUBS	1

EJ59345	0	0	0	8	0	0	0	0	0	4,00	GRCEJE	1
EJ01237	1	0	0	17	0	0	0	0	0	3,35	GRCEJE	1
CF04930	0	0	0	6	0	0	0	0	0	4,17	GRCCAS	1
MFUCHSL	4	2	0	8	0	0	0	0	0	4,43	EJECTAS	1
C933001	0	0	0	3	0	0	0	0	0	4,39	BFCAJERO	1
TE01488	11	0	4	15	0	0	0	0	0	4,23	TESOR	1
TE01088	3	0	2	8	0	0	0	0	0	4,42	TESOR	1
EJ08288	0	0	0	4	0	0	0	0	0	4,34	EJECTAS	1
SM01222	0	0	0	8	0	0	0	0	0	4,06	FUERVEN	1
TE01604	0	0	0	19	0	0	4	0	0	3,23	TESOR	1
RLARREA	0	0	0	6	0	0	0	0	0	4,25	EJECTAS	1
JS04234	0	0	0	8	0	0	0	0	0	4,13	JEFEOPE	1
CGONZAE	2	0	0	2	0	0	0	0	0	4,62	CCENACT	1
AV16966	0	0	0	7	0	0	0	0	0	4,23	BFCONSUL	1
FVTA043	2	0	0	1	0	0	0	0	0	4,71	FUERVEN	1
TE01825	1	0	1	5	0	0	0	0	0	4,59	BFTESSUC	1
CCORONA	1	0	0	6	0	0	0	0	0	4,35	ANALSRIE	1
TE01180	1	0	1	16	0	0	0	0	0	3,79	TESOR	1
SGONZAT	0	0	0	5	0	0	0	0	0	4,44	DADMCPO	1
GSOTO	0	0	0	6	0	0	0	0	0	4,39	GSENIOR	1
PGALAZF	0	0	0	5	0	0	0	0	0	4,47	ADMCPPO	1
N562291	0	0	0	5	0	0	0	0	0	4,48	BFEJECOM	1
FVTA109	3	0	0	5	0	0	0	0	0	4,52	FUERVEN	1
C831001	0	0	0	6	0	0	0	0	0	4,42	BFCAJERO	1
SM01243	0	0	0	10	0	0	0	0	0	4,14	FUERVEN	1
PTORR12	3	0	2	6	0	0	0	0	0	4,82	PREGIO	1
JS01288	0	0	0	13	0	0	0	0	0	3,93	JEFEOPE	1
CF20843	0	0	0	3	0	0	0	0	0	4,66	GRCCAS	1
MCHAPRU	18	15	6	18	0	0	0	0	0	5,18	EJECTAS	1
FVTA008	8	0	0	2	0	0	0	0	0	4,81	FUERVEN	1
VCHAVEG	18	0	7	24	0	0	0	0	0	3,90	AUDIOM	1
CCON176	4	0	3	11	0	0	0	0	0	4,61	AUDIOM	1
AAGUILE	0	6	0	0	0	0	0	0	0	5,91	GSENIOR	1
PT01915	0	0	0	7	0	0	0	0	0	4,40	BFEJECOM	1
N359455	0	0	0	9	0	0	0	0	0	4,26	EJECTAS	1
CF01904	0	0	0	5	0	0	0	0	0	4,57	GRCCAS	1
TE01368	8	0	8	15	0	0	0	0	0	4,96	TESOR	1
CVARG02	5	0	1	13	0	0	0	0	0	4,23	NEGINTE	1
EJ02561	4	2	2	14	0	0	0	0	0	4,45	EJECTAS	1
TE01389	1	0	1	11	0	0	0	0	0	4,34	TESOR	1
N422347	2	0	0	7	0	0	0	0	0	4,46	AGENTES	1
CB01886	0	0	0	10	0	0	0	0	0	4,27	GRCCAS	1
JMUNO16	0	0	0	6	0	0	0	0	0	4,55	DIRERIES	1
SPIZARG	10	5	0	11	0	0	0	0	0	4,91	EJECTAS	1
AG07288	0	0	0	9	0	0	0	0	0	4,39	AGENTES	1
TE01910	0	0	0	10	0	0	0	0	0	4,33	BFTESSUC	1
MRATTOD	0	0	0	11	0	0	0	0	0	4,30	MERCAP	1
AI01297	0	0	0	11	0	0	0	0	0	4,30	CAPTADO	1
EJ03881	6	0	0	0	0	0	0	0	0	5,18	GRCEJE	1
TE02180	0	0	0	12	0	0	0	0	0	4,27	TESOR	1
C332001	11	0	0	0	0	0	0	0	0	5,21	SUPCATL	1
CF01936	0	0	0	7	0	0	0	0	0	4,67	GRCCAS	1
MSALAMA	2	1	1	2	0	0	0	0	0	5,40	NEGINTE	1
JS01365	0	0	0	8	0	0	0	0	0	4,61	JEFEOPE	1
AR02234	13	0	0	0	0	0	0	0	0	5,22	GRCEJE	1
SHENRIV	3	0	0	8	0	0	0	0	0	4,64	ANALSRIE	1
AALLEN5	0	0	0	10	0	0	0	0	0	4,48	DADMCPO	1
TE01363	7	0	1	8	0	0	0	0	0	4,93	TESOR	1
N571300	0	0	0	9	0	0	1	0	0	4,60	TESOR	1
KVEJARR	5	0	1	6	0	0	0	0	0	5,09	AUDIOM	1
TE01615	3	0	6	8	0	0	0	0	0	5,66	TESOR	1
AG05170	3	0	1	6	0	0	0	0	0	5,07	AGENTES	1
CH01928	0	0	0	9	0	0	0	0	0	4,71	BFEJECOM	1
GCARR01	8	0	0	3	0	0	0	0	0	5,15	ANALSRIE	1
SM01235	4	0	0	0	0	0	0	0	0	5,45	FUERVEN	1
N572748	5	0	11	14	1	1	1	0	0	5,62	TESOR	1
MIBANES	0	0	0	10	0	0	0	0	0	4,67	GRCCAS	1
EJ07010	3	0	0	0	0	0	0	0	0	5,48	EJECTAS	1
N470905	0	0	0	10	0	0	0	0	0	4,69	AGENTES	1
GAEDO	1	0	1	14	0	0	0	0	0	4,58	ATPUBS	1
MZAMORG	2	0	1	9	0	0	0	0	0	4,99	AUDIOM	1
EJ04249	1	1	0	17	0	0	0	0	0	4,40	EJECTAS	1
SM04011	4	0	0	14	0	0	0	0	0	4,48	FUERVEN	1
TE01396	0	0	0	9	0	0	0	0	0	4,86	TESOR	1
JS03183	5	0	1	6	0	0	0	0	0	5,36	JEFEOPE	1
SM02243	6	0	0	3	0	0	0	0	0	5,42	FUERVEN	1
DJADELR	1	0	1	10	0	0	0	0	0	5,02	AGENTES	1
AG07170	5	0	1	10	0	0	0	0	0	5,01	AGENTES	1
MGOME11	11	1	5	21	2	1	1	0	0	4,28	PREGIO	1
EJ04487	10	7	0	8	0	0	0	0	0	6,15	EJECTAS	1
EJ11266	6	1	0	14	0	0	0	0	0	4,82	EJECTAS	1
SM02234	5	0	0	6	0	0	0	0	0	5,34	FUERVEN	1
EJ04261	2	0	1	7	0	0	0	0	0	5,37	EJECTAS	1
TE01283	8	0	8	19	0	0	0	0	0	5,40	TESOR	1
TE01614	0	0	0	10	0	0	3	0	0	4,97	TESOR	1
EJ05339	5	3	0	15	0	0	0	0	0	4,99	EJECTAS	1
EJOU094	12	0	0	0	0	0	0	0	0	5,87	AUDIOM	1

ANEXO R: USUARIOS ETIQUETADOS COMO SOSPECHOSOS POR MÁS DE UN MODELO

User_id	Datos básicos	SalDOS	UIT. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
GGALASS	1	0	1	5	0	1	0	0	0	3,79	GSENIOR	3
TE02248	3	0	1	7	2	2	1	0	0	2,22	TESOR	3
FLARA	0	0	0	6	0	1	0	0	0	1,61	SCLBTR	3
AC01004	0	0	0	2	0	0	1	0	0	1,88	EJECTAS	3
HFATTAG	18	0	0	6	0	0	0	0	0	3,39	ANALSRIE	2
TE01904	2	0	1	12	0	0	0	0	0	2,99	BFTESSUC	2
SU01003	1	0	0	6	0	0	0	0	0	3,68	GRCCAS	2
TE01292	1	1	0	6	0	0	0	0	0	3,75	TESOR	2
TE01101	0	0	0	8	0	0	1	0	0	3,83	TESOR	2
GPOZOAR	2	2	0	3	0	0	0	0	0	2,42	GSENIOR	2
BSILV01	0	0	0	3	0	0	0	0	0	2,66	GSENIOR	2
EJ03327	0	0	0	3	0	0	0	0	0	3,89	EJECTAS	2
AG01238	0	0	0	3	0	0	0	0	0	2,57	AGENTES	2
PDEMEDP	4	0	0	3	0	0	0	0	0	1,72	ANALSRIE	2
CRIVERP	3	0	1	4	0	0	0	0	0	2,00	EJECTAS	2
HSEPULT	2	0	2	5	0	0	0	0	0	1,62	GSENIOR	2
DMARGAA	1	0	1	2	0	0	0	0	0	1,33	GSENIOR	2
EMUNOZG	3	0	0	0	0	0	0	0	0	1,56	GSENIOR	2
EJ23300	5	0	0	3	0	0	0	0	0	2,02	GRCEJE	2
FLYONAU	2	0	0	1	0	0	0	0	0	1,75	NEGINTE	2
EJOU052	3	0	0	0	0	0	0	0	0	1,83	AUDIOM	2
VT18093	1	0	0	4	0	0	0	0	0	1,67	FUERVERN	2
DP61002	1	1	0	4	0	0	0	0	0	2,0275	DADMCPO	2
C178001	3	0	0	1	0	0	0	0	0	2,04	SUPCATL	2
AR31300	6	0	0	0	0	0	0	0	0	1,71	ANALSRIE	2
CFUERTH	3	0	0	8	0	0	0	0	0	1,91	AGENTES	2
CMARIAN	7	0	0	0	0	0	0	0	0	1,12	JANALRIE	2
FVTA232	1	0	0	4	0	0	0	0	0	1,80	FUERVERN	2
CSUAZOD	5	0	1	8	0	0	0	0	0	2,08	OFCOMEX	2
CBAHAMO	2	0	0	1	0	0	0	0	0	1,94	GSENIOR	2
TE02043	6	1	3	9	0	0	0	0	0	1,35	TESOR	2
TE01299	0	0	0	4	0	0	0	0	0	1,13	TESOR	2
JS01046	2	0	1	3	1	0	0	0	0	2,46	JEFEPOE	2
AG05297	1	0	1	2	0	0	0	0	0	2,00	AGENTES	2
TE01213	1	0	1	2	0	0	0	0	0	2,00	TESOR	2
CB16879	6	0	0	3	0	0	0	0	0	2,19	GRCCAS	2
AG02032	0	0	0	5	0	0	0	0	0	1,37	AGENTES	2
AG02157	0	0	0	3	0	0	0	0	0	1,40	AGENTES	2
N328981	3	0	0	4	0	0	0	0	0	2,35	EJECTAS	2
EJ01866	0	0	0	3	0	0	0	0	0	1,42	GRCEJE	2
EJ17170	0	0	0	3	0	0	1	0	0	1,38	GSENIOR	2
DLOPEZV	0	0	0	4	0	0	0	0	0	1,47	FMLOCAL	2
N596659	0	0	0	3	0	0	0	0	0	1,50	TESOR	2
NREYE01	0	0	0	5	0	0	0	0	0	1,46	SCLBTR	2
PALVARM	0	0	0	6	0	0	0	0	0	1,46	EJECTAS	2
JREYESI	0	0	0	3	0	0	2	0	0	1,87	ATPUBS	2
TE02198	0	0	0	6	0	0	0	0	0	1,53	TESOR	2
CB03860	0	0	0	4	0	0	0	0	0	1,74	GRCCAS	2
RPARRAP	0	0	0	3	0	0	0	0	0	1,82	EJECTAS	2
JS01300	0	0	0	5	0	0	3	0	0	2,00	JEFEPOE	2
EFIGUEC	0	0	0	5	0	0	0	0	0	1,81	EJECTAS	2
CF04913	0	0	0	7	0	0	0	0	0	1,77	GRCCAS	2
N451798	0	0	0	6	0	0	0	0	0	5,27	TARPAMPN	2
C929001	0	0	0	6	0	0	0	0	0	5,12	BFCAJERO	2
MCANETS	0	0	0	6	0	0	0	0	0	5,30	BFCONSUL	2
EJ01322	0	0	0	6	0	0	0	0	0	5,07	EJECTAS	2
SM01179	0	0	0	4	0	0	0	0	0	4,91	FUERVERN	2
N373190	0	0	0	4	0	0	0	0	0	4,90	BFEJECOM	2
LBALMDE	0	0	0	4	0	0	0	0	0	4,86	AGENTES	2
SM02186	0	0	0	3	0	0	0	0	0	5,38	FUERVERN	2
N431695	2	0	0	5	0	0	0	0	0	4,76	AGENTES	2
AV10966	0	0	0	3	0	0	0	0	0	5,00	BFCONSUL	2
AV60999	0	0	0	3	0	0	0	0	0	4,94	BFCONSUL	2
TE01930	0	0	0	5	0	0	0	0	0	4,72	BFTESSUC	2
FBUSTAM	0	0	0	4	0	0	0	0	0	4,76	AGENTES	2
EJOU040	1	0	0	6	0	0	0	0	0	4,65	AUDIOM	2
CF11936	0	0	0	3	0	0	0	0	0	4,87	GRCCAS	2
N450601	0	0	0	4	0	0	0	0	0	4,67	JEFEPOE	2
AG03442	2	0	0	2	0	0	0	0	0	5,22	AGENTES	2
AV14929	0	0	0	8	0	0	0	0	0	5,14	BFCONSUL	2
C919003	0	0	0	4	0	0	0	0	0	1,95	BFCAJERO	2
EJ01316	0	0	0	3	0	0	0	0	0	1,99	EJECTAS	2
CB1A345	0	0	0	4	0	0	0	0	0	2	GRCCAS	2
AI01332	0	0	0	5	0	0	0	0	0	1,96	CAPTADO	2

ANEXO S: DISTANCIA DE SOSPECHOSOS DE LA LISTA FINAL A LA MEDIA DE LOS DATOS

User_id	Datos básicos	Saldos	Ult. Movimientos	Loc. Personas	Talonnarios	Cabecera Cuenta	Prod. y Sub. de OP.	Cheques	Ordenes no pago	Tiemp trans.	Perfil	N° Apariciones
GGALASS											GSENIOR	3
TE02248											TESOR	3
FLARA											SCLBTR	3
AC01004											EJECTAS	3
HFATTAG											ANALRIE	2
TE01904											BFTESSUC	2
SU01003											GRCCAS	2
TE01292											TESOR	2
TE01101											TESOR	2
GPOZOAR											GSENIOR	2
BSILV01											GSENIOR	2
EJ03327											EJECTAS	2
AG01238											AGENTES	2
PDEMEDP											ANALRIE	2
CRIVERP											EJECTAS	2
HSEPUlt											GSENIOR	2
DMARGAA											GSENIOR	2
EMUNOZG											GSENIOR	2
EJ23300											GRCEJE	2
FLYONAU											NEGINTE	2
EJOU052											AUDIOM	2
VT18093											FUERVERN	2
DP61002											DADMCP0	2
C178001											SUPCATL	2
AR31300											ANALRIE	2
CFUERTH											AGENTES	2
CMARIAN											JANALRIE	2
FVTA232											FUERVERN	2
CSUAZOD											OFCOMEX	2
CBAHAMO											GSENIOR	2
TE02043											TESOR	2
TE01299											TESOR	2
JS01046											JEFEOPE	2
AG05297											AGENTES	2
TE01213											TESOR	2
CB16879											GRCCAS	2
AG02032											AGENTES	2
AG02157											AGENTES	2
N328981											EJECTAS	2
EJ01866											GRCEJE	2
EJ17170											GSENIOR	2
DLOPEZV											FMLOCAL	2
N596659											TESOR	2
NREYED1											SCLBTR	2
PALVARM											EJECTAS	2
JREYESI											ATPUBS	2
TE02198											TESOR	2
CB03860											GRCCAS	2
RPARRAP											EJECTAS	2
JS01300											JEFEOPE	2
EFIGUEC											EJECTAS	2
CF04913											GRCCAS	2
N451798											TARPAMPN	2
C929001											BFCAJERO	2
MCANETS											BFCONSUL	2
EJ01322											EJECTAS	2
SM01179											FUERVERN	2
N373190											BFEJECOM	2
LBALMDE											AGENTES	2
SM02186											FUERVERN	2
N431695											AGENTES	2
AV10966											BFCONSUL	2
AV60999											BFCONSUL	2
TE01930											BFTESSUC	2
FBUSTAM											AGENTES	2
EJOU040											AUDIOM	2
CF11936											GRCCAS	2
N450601											JEFEOPE	2
AG03442											AGENTES	2
AV14929											BFCONSUL	2
C919003											BFCAJERO	2
EJ01316											EJECTAS	2
CB1A345											GRCCAS	2
A101332											CAPTADO	2

