

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Detección de Fraudes en Bodegas de Datos basado en los niveles de agregación

Karen Daniela Vergara Osorio

Universidad Nacional de Colombia
Departamento de Ciencias de la Computación y de la Decisión
Medellín, Colombia
2017

Detección de Fraudes en Bodegas de Datos basado en los niveles de agregación

Karen Daniela Vergara Osorio

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en Ingeniería de Sistemas

Director (a):
Ph.D. Francisco Javier Moreno A.

Línea de Investigación:
Bases de datos

Universidad Nacional de Colombia
Departamento de Ciencias de la Computación y de la Decisión
Medellín, Colombia
2017

(Dedicatoria o lema)

Este trabajo es dedicado a Dios y a mi madre, quien siempre ha estado conmigo y me ha brindado todo su apoyo.

Resumen

Los fraudes en las compañías a nivel mundial representan pérdidas de millones de dólares. La detección temprana de los fraudes evita que se continúen con estas acciones. La contabilidad forense incluye diversas técnicas para apoyar esta detección, en esta tesis se aplican las técnicas: ley de Benford, *SAME-SAME-SAME*, *SAME-SAME-DIFFERENTE* y *Relative Size Factor* a bodegas de datos usando los niveles de agregación de las dimensiones para analizar desde datos generales (e.g., a nivel de Año) hasta datos específicos (e.g., a nivel de día) la presencia de posibles fraudes. Con estas técnicas se definen operadores OLAP, los cuales se aplican en una bodega de datos y ofrecen información sobre la posible existencia de fraudes.

Palabras clave: bodega de datos, contabilidad forense, dimensión, fraude, nivel de agregación, operador OLAP.

Abstract

Frauds at companies globally represent losses of millions of dollars. Early detection of fraud prevents the continuation of these actions. Forensic accounting includes several techniques to support this detection, in this thesis are applied the techniques: Benford's law, SAME-SAME-SAME, SAME-SAME-DIFFERENT and Relative Size Factor to data warehouses using aggregation levels of the dimensions to analyze from general data (e.g., at year level) to specific data (e.g., at day level) the presence of possible frauds. These techniques define OLAP operators, which are applied in a data warehouse and offer information on the possible existence of fraud.

Keywords: aggregation level, data warehouse, dimension, forensic accounting, fraud, OLAP operator

Contenido

Resumen	VI
Lista de figuras	X
Lista de tablas	XI
Introducción	1
1. Técnicas de detección de fraudes	5
1.1. Inteligencia Artificial (IA).....	5
1.1.1 <i>Clustering</i>	5
1.2.1 Aprendizaje de máquinas.....	8
1.2. Finanzas	8
1.3. Contabilidad Forense	10
2. Aplicación de técnicas de detección de fraudes en una BD	19
2.1. Caso de estudio	20
2.2. Definición de operadores propuestos para la detección de fraudes en una BD	25
2.2.1 Ley de Benford.....	26
2.2.2 <i>SAME-SAME-SAME</i>	31
2.2.3 <i>SAME-SAME-DIFFERENT</i>	35
2.2.4 <i>Relative Size Factor</i>	39
2.3. Posibles escenarios	42
2.4. Resultados	45
2.4.1 Ley de Benford.....	45
2.4.2 <i>SAME-SAME-SAME</i>	50
2.4.3 <i>SAME-SAME-DIFFERENT</i>	56
2.4.4 <i>Relative Size Factor</i>	62
3. Conclusiones y recomendaciones	65
3.1. Conclusiones.....	65

3.2. Recomendaciones.....	66
A. Anexo: Resultados operadores <i>SAME-SAME-SAME</i> y <i>SAME-SAME-DIFFERENT</i> para cada mes de 2013	67
B. Anexo: Capturas de pantalla del prototipo.....	78
Bibliografía	81

Lista de figuras

Figura 0-1. Resultados Estudio Global de Fraude Fuente: (ACFE, 2016).....	3
Figura 1-1. <i>Clustering</i> Jerárquico Fuente: (Kouznetsov & Tsvetovat, 2011)	6
Figura 1-2. <i>Clustering</i> Particional. Fuente (Olaya, 2014)	6
Figura 2-1. Modelo conceptual – Ventas	21
Figura 2-2. Dimensión Tiempo y sus niveles	22
Figura 2-3. Cubo de ventas	23
Figura 2-4. Bosquejo de las operaciones OLAP: Roll-up y Drill-down.....	24
Figura 2-5. Ejemplo de la operación OLAP: <i>Slice & dice</i>	24
Figura 2-6. Distribución primer dígito ValorVenta para Diciembre de 2013.....	27
Figura 2-7. Distribución primer dígito ValorVenta para 12/12/2013.....	28
Figura 2-8. Distribución primer dígito ValorVenta para 2013.....	29
Figura 2-9. Extracción de tajada (sub-cubo)	31
Figura 2-10. Distribución primer dígito ValorVenta para 1993 y Octubre 1993	43
Figura 2-11. Distribución primer dígito ValorVenta para 1993 y Octubre 1993	44
Figura 2-12. Distribución primer dígito ValorVenta para 1993 y Octubre 1993	44
Figura 2-13. Distribución primer dígito ValorVenta para 1993 y Octubre 1993	45
Figura 2-14. Distribución primer dígito ValorVenta para 2013.....	45
Figura 2-15. Distribuciones primer dígito ValorVenta para Enero 2013 – Junio 2013	46
Figura 2-16. Distribuciones primer dígito ValorVenta para Julio 2013 – Diciembre 2013	47
Figura 2-17. Distribución primer dígito ValorVenta para Junio 2013 – Hombres.....	48
Figura 2-18. Distribución primer dígito ValorVenta para Junio 2013 – Mujeres.....	48
Figura 2-19. Distribución primer dígito ValorVenta para Junio 2013 – Hombres por Moneda.	49
Figura 2-20. Distribución primer dígito ValorVenta para Junio 2013 – Mujeres por Moneda	50
Figura B-1. Parámetros de entrada del operador.....	78
Figura B-2. Resultados iniciales del operador con parámetros de entrada	79
Figura B-3. Código SQL operador Benford Inicial.....	79
Figura B-4. Aplicación de <i>Slice & dice</i>	80
Figura B-5. Resultados de la aplicación de <i>slice & dice</i>	81

Lista de tablas

Tabla 0-1. Compañías afectadas por el fraude y vulnerables a él. Fuente: (Kroll, 2015) ..	2
Tabla 1-1. Probabilidades de la ley de Benford para el primer y segundo dígito. Fuente: (Benford, 1938)	11
Tabla 1-2. Técnicas para detección de fraudes	15
Tabla 2-1. Muestra de datos de la BD	25
Tabla 2-2. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de la ley de Benford	26
Tabla 2-3. Distribución primer dígito ValorVenta para Diciembre de 2013	27
Tabla 2-4. Distribución primer dígito ValorVenta para 12/12/2013	28
Tabla 2-5. Distribución primer dígito ValorVenta para 2013	29
Tabla 2-6. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de <i>SAME-SAME-SAME</i>	32
Tabla 2-7. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 2013	33
Tabla 2-8. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en Mayo 2013	34
Tabla 2-9. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 24/05/2013	34
Tabla 2-10. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de <i>SAME-SAME-DIFFERENT</i>	36
Tabla 2-11. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 2012	37
Tabla 2-12. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en Diciembre 2012	37
Tabla 2-13. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 30/12/2012	38
Tabla 2-14. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de <i>Relative Size Factor</i>	40
Tabla 2-15. <i>Relative Size Factor</i> para Agosto 2011	41
Tabla 2-16. <i>Relative Size Factor</i> para 2011	41
Tabla 2-17. <i>Relative Size Factor</i> para 08/08/2011	41
Tabla 2-18. Número de repeticiones para los valores de: Producto, Género, Nombre Moneda, ValorVenta, PrecioUnidad e Impuesto en 2013	51

Tabla 2-19. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Enero 2013.....	52
Tabla 2-20. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 - Hombres	52
Tabla 2-21. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres	53
Tabla 2-22. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar	53
Tabla 2-23. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar Australiano.....	54
Tabla 2-24. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar Canadiense.....	54
Tabla 2-25. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Libra Esterlina	54
Tabla 2-26. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar	55
Tabla 2-27. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar Australiano	55
Tabla 2-28. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar Canadiense	56
Tabla 2-29. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Libra Esterlina	56
Tabla 2-30. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en 2013.....	57
Tabla 2-31. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres	58
Tabla 2-32. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres	58
Tabla 2-33. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar.....	59
Tabla 2-34. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar Australiano	59
Tabla 2-35. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar Canadiense	59

Tabla 2-36. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Libra Esterlina 60

Tabla 2-37. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar..... 60

Tabla 2-38. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar..... 61

Tabla 2-39. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar..... 61

Tabla 2-40. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar..... 61

Tabla 2-41. *Relative Size Factor* para 2013..... 62

Tabla 2-42. *Relative Size Factor* para cada mes de 2013 62

Tabla 2-43. *Relative Size Factor* para Junio 2013 por género 63

Tabla 2-44. *Relative Size Factor* para Junio 2013 – Hombres por moneda. 63

Tabla 2-45. *Relative Size Factor* para Junio 2013 – Mujeres por moneda 63

Tabla A-1. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Enero 2013..... 67

Tabla A-2. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Febrero 2013..... 68

Tabla A-3. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Marzo 2013 68

Tabla A-4. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Abril 2013 69

Tabla A-5. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Mayo 2013 69

Tabla A-6. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 70

Tabla A-7. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Julio 2013..... 70

Tabla A-8. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Agosto 2013 71

Tabla A-9. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Septiembre 2013 71

Tabla A-10. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Octubre 2013 72

Tabla A-11. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Noviembre 2013..... 72

Tabla A-12. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Diciembre 2013 73

Tabla A-13. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Enero 2013 73

Tabla A-14. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Febrero 2013 74

Tabla A-15. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Marzo 2013	74
Tabla A-16. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Abril 2013.....	74
Tabla A-17. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Mayo 2013	75
Tabla A-18. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013	75
Tabla A-19. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Julio 2013.....	75
Tabla A-20. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Agosto 2013.....	76
Tabla A-21. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Septiembre 2013.....	76
Tabla A-22. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Octubre 2013	76
Tabla A-23. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Noviembre 2013.....	77
Tabla A-24. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Diciembre 2013	77

Introducción

Un fraude es “una acción contraria a la verdad y a la rectitud, que perjudica a la persona contra quien se comete”, según la RAE (RAE, 2014). Nótese que aunque en la definición el concepto de fraude está asociado con una persona, por extensión esta persona puede ser una compañía (persona jurídica).

En (Kroll, 2015) se presenta un informe a partir de encuestas hechas a altos ejecutivos de diversas compañías sobre el fraude. Se entrevistaron 768 ejecutivos y se concluyó que: “...el fraude sigue en aumento, el 75% de las compañías informan que han sido víctimas de un incidente de fraude en el último año (2014). Esto significa un incremento del 14% con respecto a la cifra que se informó hace solamente tres años...”. En la Tabla 0-1 se presenta el porcentaje de las compañías afectadas en 2014 según el tipo de fraudes.

Tabla 0-1. Compañías afectadas por el fraude y vulnerables a él. Fuente: (Kroll, 2015)

COMPAÑÍAS AFECTADAS POR EL FRAUDE Y VULNERABLES A ÉL		
TIPOS DE FRAUDE	PORCENTAJE DE COMPAÑÍAS AFECTADAS EN LOS ÚLTIMOS 12 MESES	PORCENTAJE DE COMPAÑÍAS QUE SE DESCRIBEN A SÍ MISMAS COMO EXTREMADA O MODERADAMENTE VULNERABLES A ESTO
Robo de activos físicos	22 %	62 %
Fraude de proveedores o proceso de compras	17 %	49 %
Robo de información	15 %	51 %
Conflicto de intereses de la gerencia	12 %	36 %
Violación por incumplimiento regulatorio o normativo	12 %	40 %
Corrupción y soborno	11 %	40 %
Fraude financiero interno	9 %	43 %
Malversación de activos de la compañía	7 %	40 %
Lavado de dinero	4 %	34 %
Robo de Propiedad Intelectual	4 %	37 %
Prácticas monopólicas	2 %	26 %

Otro estudio sobre fraudes, lo hace la ACFE (*Association of Certified Fraud Examiners*) (ACFE, 2016) donde se analizaron 2410 casos de fraude laboral entre enero de 2014 y octubre de 2015. Los principales hallazgos fueron 6.3 billones de dólares en pérdidas, 23% de los casos ocasionaron más de 1 millón de dólares en pérdidas para la compañía y la pérdida media *por caso* fue de 150 mil dólares.

Figura 0-1. Resultados Estudio Global de Fraude Fuente: (ACFE, 2016)



Estas dos investigaciones muestran cómo los fraudes afectan a las compañías y lo valiosas que podrían ser las acciones para disminuirlos o detectarlos a tiempo.

Por lo cual se desarrolla esta investigación, para la detección de posibles fraudes, usando operadores OLAP en una bodega de datos que permitan aplicar técnicas como la ley de Benford, la prueba de *Relative Size Factor*, la prueba *SAME-SAME-SAME* y la prueba *SAME-SAME-DIFFERENT*.

Para ilustrar acerca del desarrollo de este trabajo, se comienza considerando los conceptos de Dato e Información desde el área de la informática (de Pablos Heredero, López-Hermoso, Martín-Romo, & Medina, 2004):

- Dato: Elemento de conocimiento que carece de significado por sí mismo, o que está fuera de contexto.
- Información: conjunto de datos, elaborado y situado en un contexto, de forma que tiene un significado para alguien en un momento y lugar determinado.

Con estos conceptos definidos, se emprende la exploración y consolidación del estado del arte para la detección de fraudes en bodegas de datos, encontrando diversas áreas encargadas de implementar y fortalecer técnicas debido al aumento en las pérdidas de

dinero expuestas previamente; al indagar en la literatura se observa que la aplicación de las técnicas de detección de fraudes para bodegas de datos son pocas, lo cual presenta la oportunidad para ampliar esta área de conocimiento a través de esta investigación.

En el documento se encuentran las principales áreas involucradas en la detección de fraudes, exponiendo técnicas que tienen como principal característica la adaptabilidad para aplicarse en diversas compañías o escenarios. De las técnicas presentadas se eligen 4 para aplicarse a una bodega de datos: la ley de Benford, *Relative Size Factor*, *SAME-SAME-SAME* y *SAME-SAME-DIFFERENT*. Luego se continúa con la contextualización para realizar la aplicación de estas técnicas en una BD a través de un caso de estudio, donde se definen los operadores OLAP, se muestran los posibles escenarios cuando se apliquen los operadores y finalmente se presentan los resultados obtenidos para cada operador.

En los resultados de los operadores OLAP se presenta el detalle de las ejecuciones realizadas en la BD, donde se muestra que existen indicios de fraude en los datos, para complementar esta afirmación se deben aplicar técnicas de detección de fraudes adicionales.

1. Técnicas de detección de fraudes

En este capítulo se presentan las técnicas de detección de fraudes de las siguientes áreas de estudio: inteligencia artificial, finanzas y contabilidad forense. Se analizan estas áreas debido a que ofrecen un mayor número de técnicas concebidas o adaptadas *específicamente* para la detección de fraudes.

1.1. Inteligencia Artificial (IA)

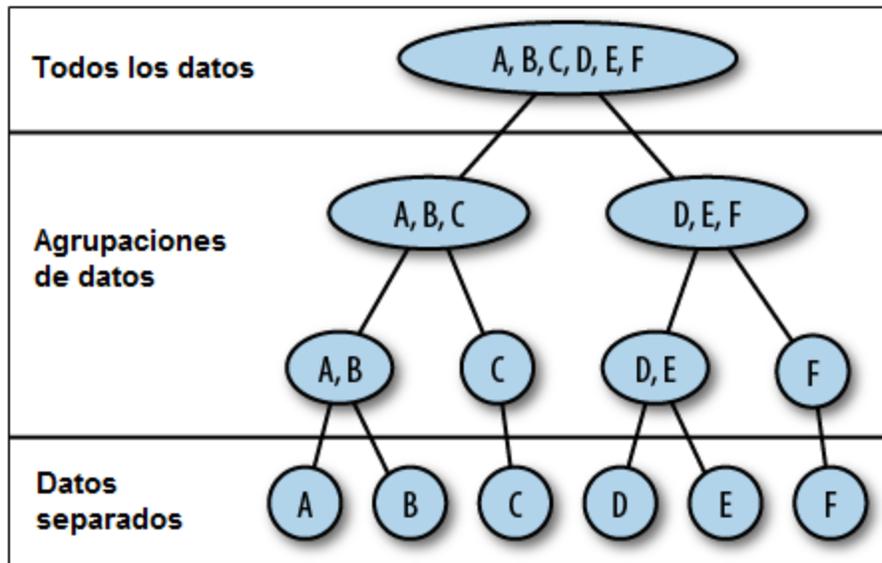
1.1.1 *Clustering*

El *clustering* (agrupamiento) se puede aplicar para la detección de datos atípicos (*outliers*), los cuales *podrían* ser fraudes. El *clustering* es el proceso de agrupar los datos por alguna característica en común, e.g., cercanía, color, forma; además es un método de clasificación *no supervisado*, i.e., no se tiene conocimiento de los datos ni de resultados de clasificaciones previas (Singh & Aggarwal, 2013).

Existen diversos métodos de *clustering* (Singh & Aggarwal, 2013):

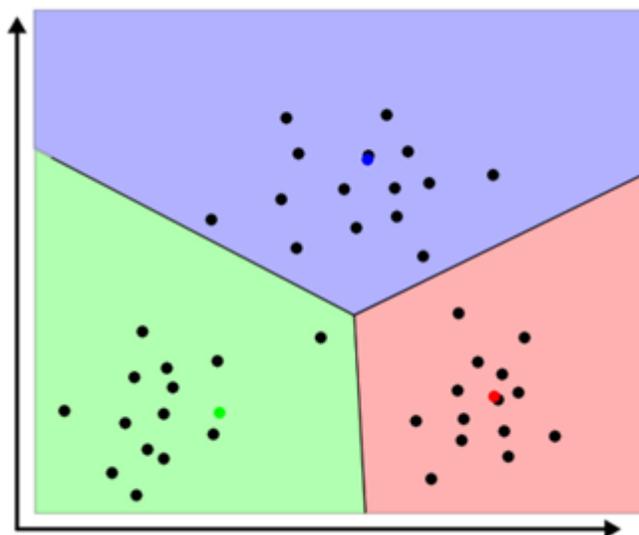
- Jerárquico: se agrupan los datos en una estructura jerárquica (árbol): se inicia desde los nodos hojas (i.e., con los datos no agrupados separados) hasta el máximo nivel de agrupación (raíz del árbol), ver Figura 1-1.

Figura 1-1. *Clustering* Jerárquico Fuente: (Kouznetsov & Tsvetovat, 2011)



- Particional: se agrupan los datos considerando una característica según algún criterio (usualmente la distancia). Un ejemplo se presenta en la Figura 1-2, donde se muestran tres *clusters* y los datos agrupados de acuerdo con su distancia. En esta figura, el punto azul, el verde y el rojo representan el *centroide* de cada *cluster*, el cual se puede considerar como el elemento representativo o promedio de los elementos de cada *cluster*.

Figura 1-2. *Clustering* Particional. Fuente (Olaya, 2014)



Para lograr la detección de valores atípicos los algoritmos de *clustering* suelen ser modificados. Por ejemplo, el algoritmo *K-means* (*K-means* es un método de agrupamiento particional, que divide un conjunto de n observaciones (datos) en k grupos, cada observación se asigna al grupo cuyo valor medio es más cercano) ha sido modificado por Ahmed & Mahmood (2013) para obtener los datos atípicos a partir de este método. Básicamente, los autores incorporan al algoritmo el cálculo de la suma de errores al cuadrado (SSE) y la suma total de cuadrados (SST) con el fin de reducir el error en los grupos asignados; cuando se identifican valores atípicos (los valores más lejanos de los valores medios), estos se separan y se recalculan los valores medios.

Una aplicación de *clustering* para detectar fraudes la hizo la empresa SAS Institute (SAS Institute & Federal Data Corporation, 1999), mediante su herramienta estadística SAS (*Statistical Analysis System*). Se analizaron las transacciones hechas con tarjetas de compra, se establecieron los *clusters* y valores atípicos al cotejarlos con expertos se descubrieron operaciones que se clasificaron como fraude. Una de las conclusiones de la investigación es que a través del *clustering* se puede crear una base de conocimiento de fraudes que puede ayudar a su detección.

Otro caso de aplicación de *clustering* para la detección de fraudes la hicieron Hilas, Mastorocostas, & Rekanos (2015), donde usan los datos de las transacciones de un usuario en las redes y en los servicios a los que accede (mediante llamadas telefónicas o páginas web). El trabajo buscaba identificar el comportamiento del usuario y sus preferencias; además identificar problemas potenciales de seguridad y descubrir el posible uso fraudulento de la red. Los autores identificaron en sus datos dos grupos importantes: i) usuarios de comportamiento mixto (el usuario ejecutaba acciones honestas y deshonestas (fraudulentas)) y ii) "ruido" en los datos (se presentaba un área "gris", debido a que los datos no se ajustaban a las caracterizaciones de usuario honesto ni deshonesto definidas por el experto inicialmente). A raíz de esto, los autores concluyen que el *clustering* no es suficiente para detectar si un usuario es honesto o deshonesto.

Sabau (2012) explica brevemente las técnicas de *clustering* para detectar fraudes; analiza más de 25 artículos, y concluye que las técnicas de *clustering* particional se usan en las tres cuartas partes de estos artículos. Además, indica que se prefieren las técnicas

de *clustering* sin modificarlas (adaptarlas) debido a que no se perciben beneficios significativos en la precisión de los resultados al modificarlos.

1.2.1 Aprendizaje de máquinas

Para detectar datos atípicos también se pueden aplicar algoritmos basados en el *aprendizaje de máquinas* (machine learning). Por ejemplo:

- Verleysen (2003) propone un método que se puede aplicar cuando se tienen varias dimensiones en un conjunto de datos (e.g., comportamiento de los precios del mercado de valores, rendimientos de las acciones, pronósticos). El método analiza la información que el algoritmo ha aprendido y a partir de ella identifica posibles anomalías o datos por fuera de los rangos normales que ha establecido. El algoritmo se aplicó a la predicción de series de tiempo de datos del mercado financiero y se encontró que el 75% de las predicciones se acercan a la realidad, los demás casos (25%) se deben analizar con más detalle para determinar si son fraudes o errores en las herramientas de predicción.
- El algoritmo ARVDH (Jacob, Sangeetha, & Karthik, 2014), el cual consta de tres fases: la preparación de los registros de datos (seleccionar y asignar puntajes a los datos de acuerdo con anomalías conocidas), eliminar datos no requeridos (i.e., los datos que se encuentran dentro de un rango de puntajes normales) y finalmente la detección de anomalías (se identifican los datos que están por fuera del rango de puntajes normales, i.e., los que podrían ser fraudes). Los autores concluyeron que la aplicación del algoritmo es eficiente para la detección de anomalías, las cuales con análisis adicionales se podrían identificar como fraudes.

1.2. Finanzas

Las finanzas es uno de los sectores donde los fraudes ocurren con más frecuencia. Esto ha llevado a la concepción de métodos especializados para su detección y prevención. Algunas aproximaciones tratan de *prevenir* los fraudes mediante distribuciones estadísticas y de esta forma evitar que se cometan transacciones ilegales (Chan & Stolfo,

1998; Stolfo, Fan, Prodromidis, Chan, & Lee, 1997). Otras se enfocan en su detección *a posteriori*, lo cual sirve a su vez para enriquecer los métodos estadísticos y para la identificación y penalización de criminales.

Una actividad que se destaca en este sector es el análisis de transacciones con tarjetas de crédito (Lu, Sirwongwattana, Huang, & Kou, 2004). Allí se propone el siguiente método: primero se clasifica el tipo de transacción (como usual o inusual, e.g., una transacción inusual puede ser una compra hecha en el exterior o en un horario en el que el usuario no suele hacer compras), luego se le asigna un costo a la transacción, el cual está relacionado con la capacidad (cupó) de la tarjeta y con el número de la transacción (primera, segunda, tercera, etc.) en una serie de transacciones consecutivas, e.g., llevadas a cabo durante la misma hora - el número de la transacción es importante porque en los fraudes se suele gastar todo el cupó de la tarjeta en 4 a 5 transacciones en promedio; por lo cual, descubrir si la primera transacción (de la serie de transacciones consecutivas) es un fraude es prioritario. Finalmente, mediante un árbol de decisión sensible al costo (Rokach & Maimon, 2014), (Sahin, Bulkan, & Duman, 2013) se decide si la transacción es un *posible* fraude. Estos métodos se implementan para monitorear las transacciones y disminuir los robos (Sahin et al., 2013). Sin embargo, un inconveniente de este tipo de estudios es el acceso a los datos de las transacciones debido a la protección bancaria (privacidad de los datos).

Por otro lado, en los últimos años las compras por internet han aumentado como lo muestra Fernández-Portillo *et al.* (2015). Por ejemplo, en España en 2001 las compras por Internet fueron de 127 millones de euros y en 2013 fueron de 12731 millones de euros. En los sitios de compras es usual que los usuarios lean los comentarios (reseñas) de los usuarios que ya compraron en dicho sitio. Sin embargo, estas reseñas no son necesariamente confiables ya que podrían haber sido manipuladas para mejorar la imagen del sitio. Esta manipulación es considerada por Hu, Liu, & Sambamurthy (2011) como un fraude.

También se han desarrollado modelos para detectar fraudes en informes de los estados financieros. Uno de ellos es el modelo de lenguaje integrado (Dong et al., 2014), el cual combina un modelo estadístico con un modelo de análisis semántico para detectar el uso estratégico de lenguaje engañoso en balances financieros (*financial statements*) tales

como ventas ficticias, reconocimiento de gastos indebidos, valoración incorrecta de activos, deudas camufladas y divulgación inadecuada de información financiera, entre otros. Los autores aplicaron su método en compañías chinas y detectaron fraudes en los balances financieros presentados: el modelo pudo capturar *patrones semánticos* que caracterizan a los balances financieros fraudulentos de los no fraudulentos.

1.3. Contabilidad Forense

En el área de contabilidad se ha desarrollado el concepto de *contabilidad forense*, que es “la ciencia de la recopilación y presentación de información financiera en una forma adecuada para una corte de jurisprudencia contra los autores de delitos económicos” (Wadhwa & Pal, 2012). De la integración de la contabilidad, la auditoría y las técnicas de investigación se obtiene la especialidad conocida como contabilidad forense que se centra en la detección o prevención de fraudes contables (Wadhwa & Pal, 2012).

Una técnica especializada para descubrir posibles errores o datos atípicos en conjuntos de datos es la *ley de Benford* (Tackett, 2007). En 1938, Frank Benford (Benford, 1938) observó que las primeras páginas de un libro de logaritmos mostraban más desgaste que las últimas páginas, lo que le sugirió que en muchas aplicaciones los números más usados comienzan con el dígito 1. Para probar su hipótesis, Benford analizó los primeros dígitos de datos de diferentes fuentes (e.g: alturas de montañas, direcciones urbanas, precios de acciones, poblaciones de ciudades, entre otros) y mostró que en conjuntos (muestras) de mínimo 10000 números hay una distribución logarítmica que rige la distribución de los primeros dígitos cuando los números se componen de cuatro o más dígitos.

En la Tabla 1-1 se muestran las probabilidades de los dígitos del 0 al 9 de aparecer en el primer y en el segundo dígito de un número de un conjunto de números, i.e., el dígito 1 tiene 30,1% de probabilidad de aparecer como primer dígito y 11,4% de probabilidad de aparecer como segundo dígito.

Tabla 1-1. Probabilidades de la ley de Benford para el primer y segundo dígito. Fuente: (Benford, 1938)

Digit	First Place	Second Place
0.....	0.000	0.120
1.....	0.301	0.114
2.....	0.176	0.108
3.....	0.125	0.104
4.....	0.097	0.100
5.....	0.079	0.097
6.....	0.067	0.093
7.....	0.058	0.090
8.....	0.051	0.088
9.....	0.046	0.085

Esta distribución se toma como referencia para analizar conjuntos de números y así detectar posibles anomalías. Diversas investigaciones han validado esta ley, e.g., en diversas ciencias se ha aplicado esta ley y se ha comprobado su cumplimiento usando datos numéricos de áreas tan diversas como campos geomagnéticos, terremotos, casos de enfermedades infecciosas a nivel global, flujos de fotones en objetos brillantes, entre otros (Sambridge, Tkalčić, & Jackson, 2010).

Sin embargo, también se ha demostrado que el éxito de esta ley es bajo cuando se aplica a datos que se encuentran relacionados entre ellos, debido a que la independencia en los datos es una característica fundamental para que se cumpla la ley de Benford (Ausloos, Herteliu, & Ileanu, 2014). Por ejemplo, en el estudio realizado por Ausloos et al (2014), se demuestra que para los datos sobre el número de nacimientos en Rumania entre Enero 01 de 1905 y Diciembre 31 de 2001, al analizar el primer dígito no se cumple la ley de Benford (sin embargo, al analizar el segundo dígito la ley *sí* se cumplió). Los autores concluyen que los datos analizados están relacionados de forma social, i.e., se ven afectados por las relaciones de las personas según su ambiente social (en el artículo se examina la población considerando sus creencias religiosas); por lo tanto, carecen de aleatoriedad que es una propiedad requerida por la ley de Benford.

Otro ejemplo donde *no* fue exitosa la ley de Benford, se presenta en (Breunig & Goerres, 2011) en Alemania, se tomaron los datos históricos registrados de elecciones de 15 años

con el fin de descubrir posibles fraudes en dichos sufragios. Los resultados, luego de aplicar la ley de Benford, *no sugirieron* la presencia de fraudes, debido a que la manipulación de grupos de personas pequeños no se refleja en los resultados globales.

Por otro lado, en el sector financiero y económico se encuentran *aplicaciones exitosas* de la ley de Benford para la detección de fraudes. En estudios de empresas en Johannesburgo (Saville, 2006), Europa (Henselmann, Scherr, & Ditter, 2013; Mir, Ausloos, & Cerqueti, 2014; Žgela & Dobša, 2011) y Medellín (Hernández G, 2009).

Por ejemplo, Saville (2006) aplicó la ley de Benford con datos de 17 empresas para probar la hipótesis de que la ley de Benford se puede aplicar para identificar informes falsos o fraudulentos de datos contables y los resultados que obtuvo apoyan este argumento. Los hallazgos fueron de particular relevancia para los auditores, accionistas, analistas financieros, gestores de inversión e inversores privados. En el artículo se analizaron 17 empresas sospechosas de haber cometido fraude, al aplicar la ley de Benford se encontró una violación de la ley en cada una de las 17 empresas, lo cual soportó las sospechas consideradas.

Otro estudio es el de la Universidad de Erlangen-Núremberg en Alemania (Henselmann et al., 2013) donde se analizaron informes financieros. Al aplicar la ley de Benford se descubrió que la distribución del primer dígito siguió la ley de Benford. Los autores concluyen que los resultados empíricos sugieren que la aplicación de la ley de Benford a los informes financieros podría ser una herramienta analítica útil para los inversores y también puede ser de interés para acciones de planificación.

En el estudio de Medellín (Hernández G, 2009), se aplicó la ley de Benford como método de auditoría para un sistema contable de una empresa y se detectaron fraudes en algunos meses del año. El estudio se hizo en varias sedes de una empresa, y en cada una se encontraron anormalidades, catalogadas como *posibles* fraudes y luego confirmados como tales. Hernández concluye que la ley de Benford fue una herramienta contable valiosa en este caso.

La mayoría de las investigaciones usan el primer dígito para aplicar la ley de Benford; sin embargo, existen aplicaciones que emplean el segundo, el tercero e incluso el cuarto

dígito y afirman que el análisis de estos dígitos ofrece mejores indicios de *posibles* errores o anomalías (Diekmann, 2004).

A continuación se presentan otros casos de estudio donde se ha aplicado la ley de Benford con el primer dígito. En el sector contable se considera como una herramienta útil para identificar posibles errores, fraudes potenciales o irregularidades (Durtschi, Hillison, & Pacini, 2004). Por ejemplo, al emplear la ley de Benford en un sistema de pagos de un banco de Croacia se encontró que esta ley es eficaz en la auditoría de los sistemas de información, específicamente para el análisis del sistema de pago extranjero porque se lograron identificar movimientos anormales que resultaron ser fraudes (Krakar & Žgela, 2009).

En el sector de la salud se ha aplicado a ondas cerebrales del ser humano. Como el seguimiento y el análisis automatizado en línea de la actividad eléctrica del cerebro se usa con frecuencia para la verificación de las enfermedades cerebrales y para la estimación de la profundidad de la anestesia en los sujetos que van a ser sometidos a cirugía, se quería probar si al aplicar la ley de Benford se podía detectar la modulación accidental o defectos en las señales neurofisiológicas. Se aplicó la ley a datos de ambiente artificial (*in vitro*) y a electroencefalogramas (EEG) y se encontró que en el sevoflurano (un anestésico), la distribución del primer dígito sigue la ley de Benford, mientras que para los datos de EEG la distribución es regular en todos los dígitos. También se analizaron los datos en presencia de ruido de alta frecuencia, y se muestra que la ley de Benford *no* se cumple allí (Kreuzer et al., 2014).

La ley de Benford también se ha aplicado para detectar entrevistadores engañosos. En (Bredl, Winker, & Kötschau, 2008) se hicieron cuestionarios en diferentes pueblos de Europa, se definieron variables como número de encuestas por entrevistador, el número de preguntas respondidas y no respondidas. Luego, al aplicar la ley para el primer dígito de estos datos se concluyó que esta es un “buen” punto de partida para detectar *posibles* fraudes debido a que se encuentran diferencias entre los resultados de un entrevistador honesto y uno que posiblemente está cometiendo fraude; sin embargo, se requieren otras técnicas o análisis adicionales si se desea establecer con certeza si un comportamiento es fraudulento o no.

Con los casos expuestos para la ley de Benford se puede afirmar que esta técnica ha sido usada en múltiples áreas del conocimiento, identificando dos resultados posibles: aplicable o no aplicable al escenario planteado, ambos resultados son válidos, por lo cual previamente a la aplicación de esta técnica se debe asegurar lo mencionado en algunos de los casos: los datos *no* deben estar relacionados entre sí, i.e. los datos deben ser aleatorios para que la técnica sea aplicable para la detección de fraudes.

Uno de los autores que más ha estudiado la ley de Benford es Mark Nigrini (Nigrini, 2011; Nigrini & Wells, 2012). Nigrini desarrolla además otras tres técnicas para la detección de *posibles* fraudes: *SAME-SAME-SAME*, *SAME-SAME-DIFFERENT* y *Relative Size Factor* (Nigrini, 1999, 2011). Las técnicas *SAME-SAME-SAME* y *SAME-SAME-DIFFERENT* están basadas en la detección de duplicados excesivos de los datos lo cual puede indicar posibles fraudes y/o errores.

Un ejemplo de la técnica *SAME-SAME-SAME* es el siguiente: supóngase que se analizan datos de transacciones de ventas que incluyen: valor de la compra, fecha, código de producto, código de vendedor y número de factura. Supóngase que varias ventas son idénticas, i.e., tienen los mismos datos en estos atributos (*SAME-SAME-SAME*), esto es un indicio de un fraude. Por su parte, la técnica de *SAME-SAME-DIFFERENT* funciona así: supóngase que en el ejemplo anterior varias ventas son idénticas en los datos de valor de compra, fecha, código de producto y código de vendedor, i.e., solo difieren en los números de factura. Esto podría sugerir un posible fraude, para determinarlo con certeza se deben hacer análisis y validaciones adicionales.

Finalmente, la técnica *Relative Size Factor* identifica la relación entre los dos valores más altos de un conjunto de datos. Por ejemplo, si se tienen las ventas de autos de un año y se toman los dos autos vendidos más costosos, el valor más alto se divide por el otro valor, si el resultado es mayor que 1 se sospecha de un posible fraude o de error en los datos.

A continuación, en la Tabla 1-2 se presenta un resumen comparativo de los trabajos analizados (ordenados cronológicamente).

Tabla 1-2. Técnicas para detección de fraudes

Referencia	Tipo de técnica usada para la detección de fraudes	Área de identificación de posibles fraudes	Principal conclusión sobre la detección de fraudes
Stolfo, Fan, Prodromidis, Chan, & Lee (1997)	Distribuciones estadísticas y aprendizaje de máquinas.	Fraudes en tarjetas de crédito.	Se redujo el número de falsos positivos y de posibles fraudes.
Chan & Stolfo (1998)	Distribuciones estadísticas	Fraudes en tarjetas de crédito.	Su método puede ayudar a reducir las pérdidas (de dinero) debido a transacciones ilegítimas.
Nigrini (Nigrini, 1999, 2011; Nigrini & Wells, 2012)	Ley de Benford	Contabilidad Forense	Este autor es el principal promotor de la ley de Benford, y ha detectado con ella fraudes principalmente en datos contables.
Verleysen (2003)	Algoritmo basado en aprendizaje de máquinas	Predicción de series de tiempo de datos del mercado financiero	Las predicciones son un 75% similar a la realidad, el otro 25% requiere de análisis adicionales para evaluar si son errores o fraudes.
Durtschi, Hillison, & Pacini (2004)	Ley de Benford	Contabilidad	Se valida esta ley como una herramienta útil para encontrar posibles errores, fraudes potenciales o irregularidades
Lu, Sirwongwattana, Huang, & Kou (2004)	Monitoreo y análisis de transacciones	Fraudes en tarjetas de crédito y detección de intrusiones en computadores.	Se resalta el monitoreo de los datos y el análisis de las transacciones como una herramienta eficaz para la detección de posibles fraudes.

Saville (2006)	Ley de Benford	Informes contables para la bolsa de valores	Se encontraron fraudes en los informes de algunas empresas.
Tackett (2007)	Ley de Benford	Cifras contables.	Se aplicó la ley de Benford con uno y dos dígitos y se detectaron posibles fraudes.
Bredl, Winker, & Kötschau (2008)	Ley de Benford	Entrevistadores	Los resultados no fueron concluyentes. Se necesitan otras técnicas para el análisis de posibles fraudes.
Hernández G (2009)	Ley de Benford	Auditoría en sistema contable	Se detectaron fraudes en datos de diversos meses del año.
Krakar & Žgela (2009)	Ley de Benford	Auditoría sistemas de pago	Se encontraron transacciones anormales, las cuales se confirmaron luego como fraudes.
Sambridge, Tkalčić, & Jackson (2010)	Ley de Benford	Ciencias Naturales	Valida la aplicación de esta ley en la detección de fraudes en diversos indicadores (profundidades de terremotos, emisiones de gases de efecto invernadero, número de casos de enfermedades infecciosas, entre otros) de las ciencias naturales.
Breunig & Goerres (2011)	Ley de Benford	Votaciones electorales	No se encontraron (posibles) fraudes aunque los autores afirman que los datos podrían estar correlacionados.
Hu, Liu, & Sambamurthy (2011)	Indicadores matemáticos	Comentarios (reseñas) en páginas web.	Se descubrieron manipulaciones en los comentarios de algunos productos en páginas como Amazon y Barnes & Noble y esto es considerado un fraude.

Nigrini (Nigrini, 2011)	<i>SAME-SAME-SAME</i>	Contabilidad Forense	Se identificaron fraudes en datos de facturas contables.
Nigrini (Nigrini, 2011)	<i>SAME-SAME-DIFFERENT</i>	Contabilidad Forense	Se identificaron fraudes en datos de facturas contables.
Nigrini (Nigrini, 2011)	<i>Relative Size Factor</i>	Contabilidad Forense	Se encontraron posibles anomalías que se deben analizar con más detalles para determinar si son fraudes.
Sabau (2012)	<i>Clustering</i>	Fraudes financieros en tarjetas de crédito, seguros médicos, entre otros	El <i>clustering</i> de tipo particional es el más usado (75%) para la detección de <i>posibles</i> fraudes.
Ahmed & Mahmood (2013)	<i>Clustering:</i> Modificación del algoritmo <i>k-means</i>	Datos médicos de infografías (radiografía de los vasos y ganglios linfáticos)	Se identifican <i>outliers</i> con mayor precisión (a comparación de la aplicación sin la modificación del algoritmo), los cuales podrían ser fraudes.
Henselmann, Scherr, & Ditter (2013)	Ley de Benford	Informes financieros	Se detectaron fraudes, los datos evidenciaron manipulación humana.
Sahin, Bulkan, & Duman (2013)	Árboles de decisión sensibles al costo	Fraudes en tarjetas de crédito.	Se mejoró la clasificación de las transacciones como fraudulentas y no fraudulentas.
Singh & Aggarwal (2013)	<i>Clustering</i>	Diversas áreas: sistemas de bases de datos, aplicaciones de software, entre otros	Se aplica el <i>clustering</i> para la detección de <i>outliers</i> , los cuales podrían ser fraudes.
Ausloos, Herteliu, & Ileanu (2014)	Ley de Benford	Número de nacimientos en Rumania	La ley de Benford no se cumple para estos datos.
Dong, Liao, Fang, Cheng, Chen, & Fan (2014)	Modelo para el análisis de patrones semánticos	Informes financieros	Se identificaron fraudes (lenguaje engañoso) en algunos informes.
Jacob,	Algoritmo basado en	Diversas	Se detectaron

Sangeetha, & Karthik (2014)	aprendizaje de máquinas	aplicaciones: fraudes en tarjetas de crédito, amenazas en <i>ciber</i> seguridad, entre otros	anomalías en los datos, las cuales podrían ser fraudes.
Kreuzer, Jordan, Antkowiak, Drexler, Kochs, & Schneider (2014)	Ley de Benford	Actividad eléctrica cerebral	Se encontraron comportamientos similares a la ley de Benford, pero cuando se analizaron los datos con ruidos de alta frecuencia no se cumplió.
Mir, Ausloos, & Cerqueti (2014)	Ley de Benford	Ingreso anual de impuestos en Italia	No se detectaron (posibles) fraudes con la aplicación de la ley.
Hilas, Mastorocostas & Rekanos (2015)	<i>Clustering</i>	Transacciones en redes (llamadas telefónicas y accesos a páginas web)	La técnica fue insuficiente para determinar <i>con certeza</i> si hubo fraudes.

Luego de la revisión de estas técnicas, en el siguiente capítulo se presenta la aplicación de algunas de ellas para la detección de fraudes en una bodega de datos.

2. Aplicación de técnicas de detección de fraudes en una BD

Hoy los volúmenes de datos generados por las aplicaciones (comentarios en redes sociales, reseñas de productos, mensajes de correo, registros de sensores, entre otros) son del orden de *exabytes* (10^{18} bytes). Estos datos se pueden almacenar en bases de datos o en archivos de texto (XML, JSON). Dentro de las bases de datos, una opción son las bodegas de datos (BD)¹, las cuales son concebidas para facilitar el análisis y la integración de datos provenientes de diversas fuentes (Verleysen, 2003).

En este trabajo se consideran como fuentes de datos para identificar posibles fraudes a las BD. Debido a su estructura, una BD ofrece la oportunidad para aplicar técnicas como las descritas para la detección de posibles fraudes. Se definen los siguientes conceptos:

- **Bodega de datos:** es un conjunto de datos orientado a uno o varios temas (e.g., ventas, crímenes, envíos; los cuales son definidos según los requisitos analíticos de la organización), es integrada (integra datos de diferentes fuentes, e.g., de diferentes sistemas de gestión de bases de datos), es no volátil (los datos se cargan de forma masiva, se acceden principalmente para lectura y rara vez son actualizados) y posee datos variables en el tiempo (los datos son usualmente históricos, i.e., suelen abarcar periodos de operación de uno, dos o más años) (Inmon, 2005)
- **OLAP (*Online Analytical Processing*):** es un conjunto de herramientas para explorar y analizar los datos. Estas herramientas permiten observar los datos desde diferentes perspectivas gracias al modelo de dimensiones (se explican luego) que soportan las BD (Information Resources Management Association, 2012). OLAP incluye

¹ Aunque la sigla BD usualmente se usa para referirse a base de datos, en esta tesis, esta sigla siempre significará bodega de datos.

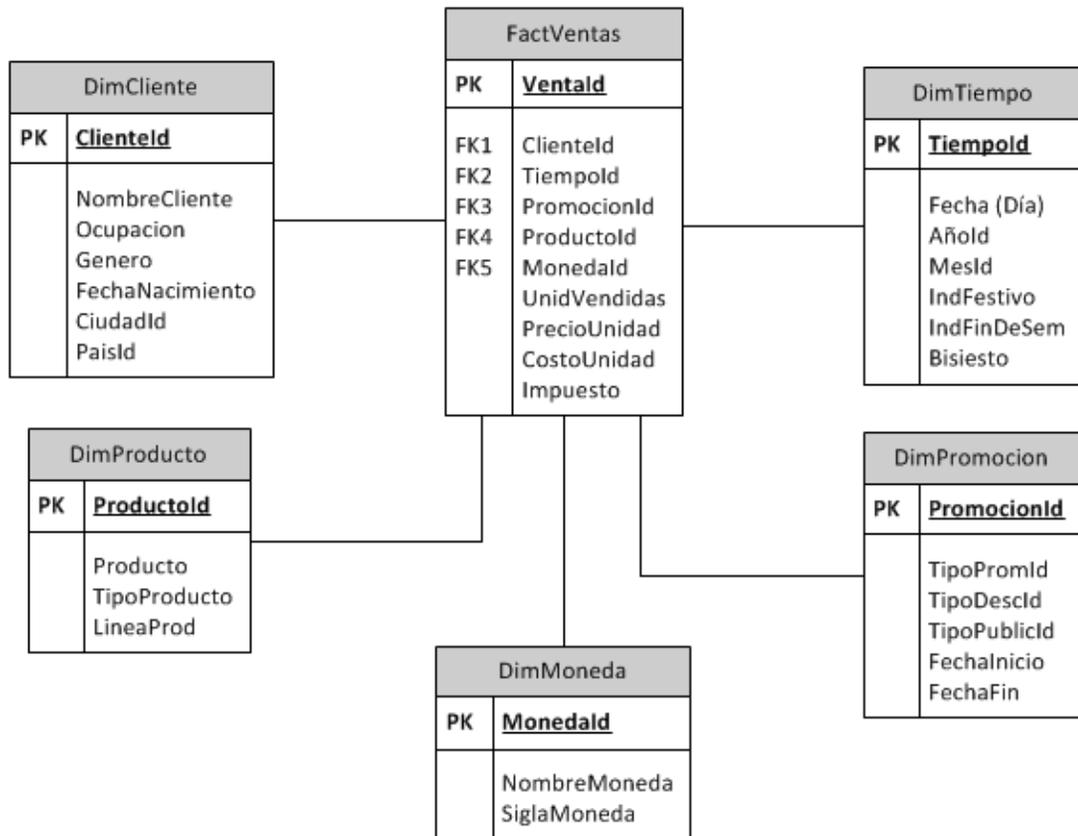
operadores como *roll-up*, *drill-down*, *pivot*, *slice and dice* (se explican luego), entre otros que facilitan la exploración y el análisis de los cubos de datos (se explican luego) de las BD.

En particular, hay pocas aplicaciones de la ley de Benford en el contexto de las BD y no se encontró ninguna que aplique técnicas como *SAME-SAME-SAME*, *SAME-SAME-DIFFERENT* y *Relative Size Factor*. Por ejemplo, Kraus y Valverde (2014) aplicaron la ley de Benford en una BD; sin embargo, en su trabajo simplemente extraen un conjunto de datos de la BD y aplican la ley, i.e., no aprovechan la estructura de la BD. Es decir, en dicho trabajo no se aprovechan características propias de una BD como la agregación de datos a través de los diferentes niveles de una dimensión (estos conceptos se definen más abajo) en una BD para la detección de posibles fraudes. Por ejemplo, se podría aplicar la ley de Benford a las ventas diarias, mensuales, anuales, etc. y analizar *si la ley se conserva en cada uno de estos niveles* (día, mes, año), e.g., si se detecta que la ley *no se cumple* en un mes específico se podría analizar su incidencia en un nivel superior (año) y en un nivel inferior (día). Nótese la conexión de esta idea con el trabajo de las elecciones en Alemania (Breunig & Goerres, 2011) descrito en la Sección 2.3.

2.1. Caso de estudio

Considérese el modelo conceptual de una BD de Ventas de la Figura 2-1.

Figura 2-1. Modelo conceptual – Ventas



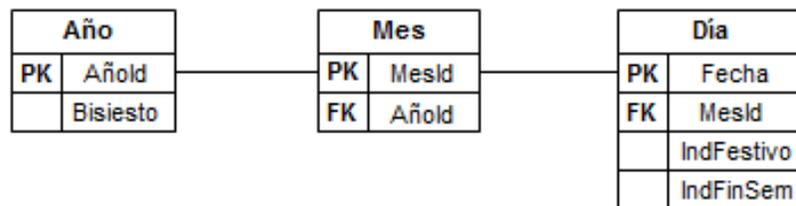
A continuación se presentan algunas definiciones esenciales en el contexto de las BD (Malinowski & Zimányi, 2008, Kimball & Ross, 2011):

- **Dimensión:** es una perspectiva para el análisis de los datos, e.g., los clientes, los productos, las promociones y el tiempo. Las dimensiones *caracterizan* a los *hechos* (ver siguiente definición), e.g., a las ventas. Además, cada dimensión está estructurada jerárquicamente en niveles (se definen más abajo).
- **Hecho:** es el sujeto, el tema de análisis, e.g., las ventas. Los hechos se almacenan en una tabla denominada *tabla de hechos*. Cada hecho está asociado con un valor de cada una de las dimensiones que lo caracteriza. Por ejemplo, considérese una venta con Ventald = 12, la cual se le hizo al cliente con Clienteld = 1, el cual compró el producto con Productold = 5, la venta ocurrió en la fecha con Tiempold = 213 (este número representa una fecha específica, e.g., el 19/Feb/2016) y fue asociada con la

promoción con PromociónId = 1. Además, los hechos poseen medidas (ver siguiente definición).

- **Medida:** es un indicador, usualmente numérico, de un hecho sobre los cuales se hacen las operaciones de agregación (totales, promedios, máximos) para el análisis de los hechos. En el ejemplo de la Figura 2-1, las medidas son UnidVendidas (unidades vendidas), PrecioUnidad (precio unitario) y CostoUnit (costo unitario). Así, es posible obtener, e.g., el total de unidades vendidas (UnidVendidas) de un producto en una fecha específica.
- **Nivel:** es un componente de una dimensión. Los niveles de una dimensión se estructuran en jerarquías (e.g., los días se agrupan en meses y los meses en años). Esta organización jerárquica permite analizar los hechos, desde un nivel general (e.g., año) hasta un nivel más detallado (e.g., día); así cada nivel representa un *nivel de granularidad*. Por ejemplo, en la dimensión Tiempo los hechos se pueden analizar a nivel de día, mes o año (e.g., total de unidades vendidas diaria, mensual y anualmente); en la dimensión Cliente los hechos se podrían analizar a nivel de ciudad o país. Cada nivel puede tener atributos, i.e., propiedades que lo caracterizan. En la Figura 2-2 se muestra la dimensión Tiempo con sus correspondientes niveles. Nótese que IndFestivo (indicador de día festivo) e IndFinDeSem (indicador de día de fin de semana) no son niveles, sino atributos del nivel Día.

Figura 2-2. Dimensión Tiempo y sus niveles



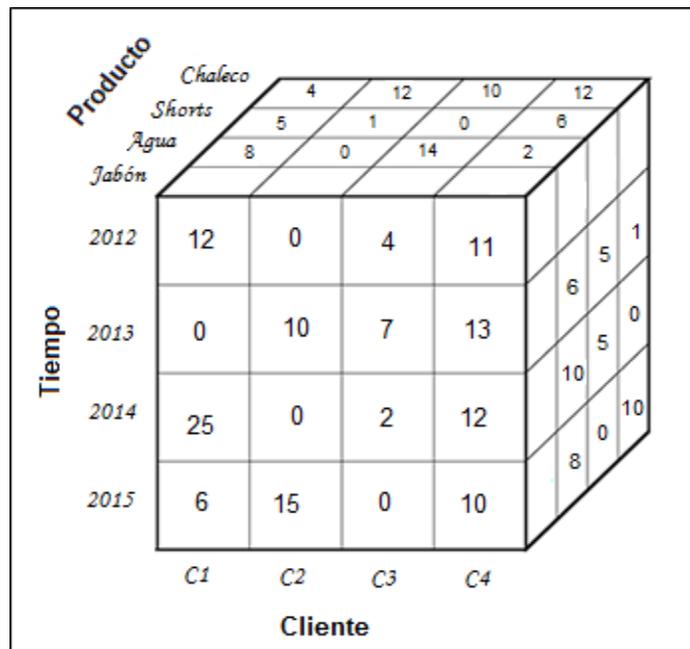
PK = Primary Key (clave primaria)

FK = Foreign Key (clave foránea)

- **Cubo de datos:** es una vista de los datos de la BD que facilita su análisis. En la Figura 2-3 se muestra un cubo de datos donde se presentan tres dimensiones Tiempo,

Cliente y Producto. Los valores numéricos en cada celda corresponden al número de unidades vendidas (UnidVendidas). Por ejemplo, del producto Jabón el cliente C1 en 2015 compró 6 unidades. Nótese que en la Figura 2-3 en la dimensión Tiempo, los datos se muestran agregados en el nivel de granularidad Año.

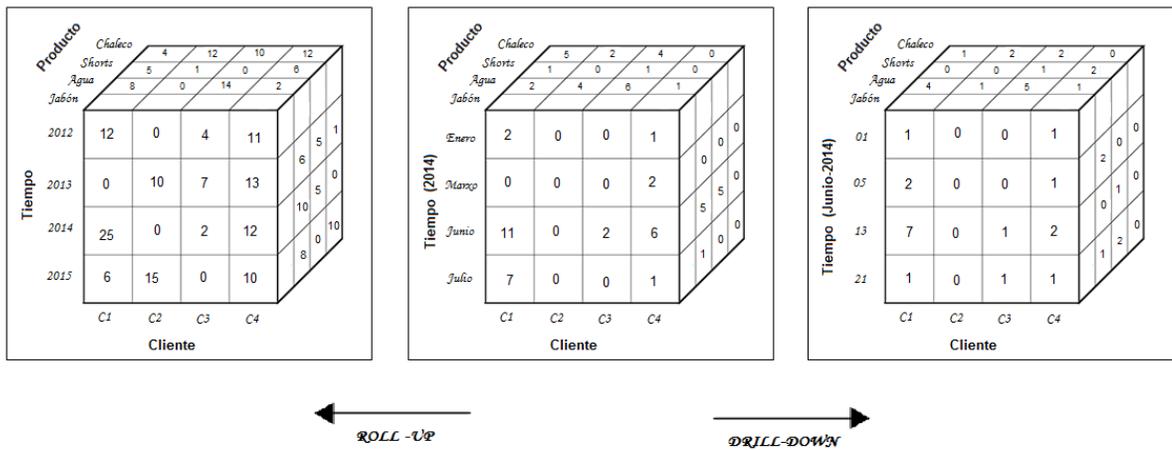
Figura 2-3. Cubo de ventas



- *Roll-up*: es una operación OLAP que transforma medidas agregadas de un nivel inferior (más específico) a un nivel superior (más general). Por ejemplo, transformar los totales *diarios* de unidades vendidas (UnidVendidas) a totales de unidades vendidas *anuales*.
- *Drill-down*: es la operación OLAP contraria a la operación *roll-up*.

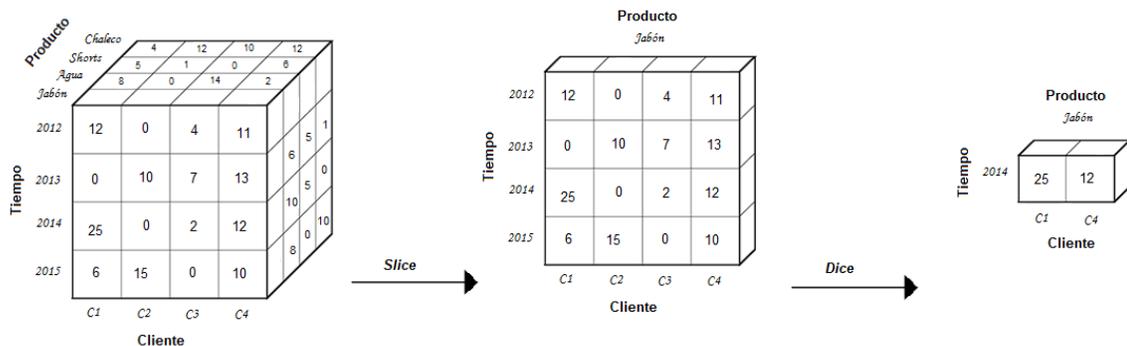
En la Figura 2-4 se bosquejan estas dos operaciones.

Figura 2-4. Bosquejo de las operaciones OLAP: Roll-up y Drill-down



- *Slice & dice*: es una operación OLAP que selecciona los datos de una o más dimensiones, generando un sub-cubo. Por ejemplo, seleccionar las unidades vendidas (UnidVendidas) para el producto Jabón, en el año 2014 para los clientes C1 y C4, el resultado se muestra en la Figura 2-5:

Figura 2-5. Ejemplo de la operación OLAP: *Slice & dice*



Las BD se presentan como un escenario propicio para aplicar técnicas de detección de (posibles) fraudes, en particular se considerará la aplicación de la ley de Benford, *SAME-SAME-SAME*, *SAME-SAME-DIFFERENT* y *Relative Size Factor*. El objetivo es analizar la detección de posibles fraudes de datos en los diferentes niveles de las dimensiones, e.g., si hay indicio de fraude a nivel de mes (e.g., las unidades vendidas en un mes dado *no* siguen la ley de Benford), analizar si este indicio de fraude se conserva en un nivel más general (e.g., año), i.e., aplicar un *slice & dice* a la BD, y analizar si es posible

determinar en un nivel más específico (e.g., día), i.e., aplicar nuevamente un *slice & dice*, cual es el día (o días) específico donde se presentó los posibles fraudes. Nótese que a pesar de que los datos que suelen conformar una BD han sido previamente sometidos a un proceso de depuración durante la etapa de ETL (extraer (*extract*), transformar (*transform*) y cargar (*load*)), esto no garantiza que estén exentos de fraudes. De esta forma, la aplicación de técnicas para la detección de fraudes en una BD se puede considerar como una depuración *a posteriori* de las tareas de ETL.

Los datos usados para el caso de estudio son datos libres de Microsoft (Microsoft Corp, 2013) y del Banco Mundial (Banco Mundial, 2014), los cuales alimentan la BD usada. Estos datos pertenecen a una compañía ficticia que vende bicicletas, accesorios y elementos necesarios para una salida de campo. Para el caso de estudio se usa esta BD debido a que cumple con las condiciones planteadas para la aplicación de las técnicas de detección de fraudes: aleatoriedad en los datos, cantidad de registros (mayor a 10.000 registros), estructura de la BD de acuerdo a las definiciones planteadas.

Tabla 2-1. Muestra de datos de la BD

Producto	Fecha	Mes	Año	NombreCliente	ApellidoCliente	ValorVenta	PrecioUnidad	Promoción
Fender Set - Mountain	2013-02-18 00:00:00.000	Febrero	2013	Jonathan	Robinson	21,98	21,98	No Discount
Sport-100 Helmet, Blue	2013-02-18 00:00:00.000	Febrero	2013	Jonathan	Robinson	34,99	34,99	No Discount
Long-Sleeve Logo Jer...	2013-02-18 00:00:00.000	Febrero	2013	Ruben	Lopez	49,99	49,99	No Discount
AWC Logo Cap	2013-02-18 00:00:00.000	Febrero	2013	Benjamin	Zhang	8,99	8,99	No Discount
Road-250 Black, 44	2013-02-18 00:00:00.000	Febrero	2013	Nicole	Butler	2443,35	2443,35	No Discount
Bike Wash - Dissolver	2013-02-18 00:00:00.000	Febrero	2013	Nicole	Butler	7,95	7,95	No Discount
Road-250 Black, 44	2013-02-18 00:00:00.000	Febrero	2013	Sebastian	Morris	2443,35	2443,35	Volume Di...
Mountain-200 Black, 38	2013-02-18 00:00:00.000	Febrero	2013	Victoria	Russell	2294,99	2294,99	No Discount
Mountain-200 Black, 38	2013-02-18 00:00:00.000	Febrero	2013	Chase	Cox	2294,99	2294,99	No Discount
Sport-100 Helmet, Blue	2013-02-18 00:00:00.000	Febrero	2013	Chase	Cox	34,99	34,99	No Discount
Touring-3000 Blue, 44	2013-02-18 00:00:00.000	Febrero	2013	Sebastian	James	742,35	742,35	No Discount

2.2. Definición de operadores propuestos para la detección de fraudes en una BD

Para los ejemplos de esta sección, se continuará con la BD de la Figura 2-1; para la definición de cada operador se comienza con un ejemplo específico. Para los experimentos se tienen las siguientes consideraciones: el nivel de Género presenta como

posibles valores Mujer y Hombre, y para el nivel del Cliente y el nivel de Producto se usarán los atributos NombreCliente y Producto (nombre del producto) para su identificación, respectivamente. Para el ejemplo se supone que no hay dos clientes ni dos productos con el mismo nombre.

2.2.1 Ley de Benford

Para aplicar este operador se debe elegir de la tabla de hechos (FactVentas) una *medida numérica* para analizar su primer dígito, aquí se usa ValorVenta, además se elige un nivel de una dimensión para aplicar inicialmente el operador. Aquí se usa la dimensión DimTiempo y se considera para la aplicación inicial del operador el nivel Mes de la venta (Diciembre de 2013). Luego, el operador se aplica a en un nivel inferior (Día 12/12/2013) y en un nivel superior (Año 2013). En la Tabla 2-2 se presenta una muestra de datos que se usará para el ejemplo.

Tabla 2-2. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de la ley de Benford

Fecha	ValorVenta
22/03/2013	34,99
03/04/2013	4,99
21/05/2013	3,99
24/05/2013	32,6
24/05/2013	3,99
08/09/2013	4,99
16/12/2013	32,6
01/01/2013	8,99
01/01/2013	7,95
01/01/2013	7,95
01/01/2013	21,98
01/01/2013	24,49

A continuación se presentan los pasos para aplicar el operador.

1. Indicar la medida numérica de la tabla de hechos (FactVentas): ValorVenta.
2. Indicar la dimensión que se usará: DimTiempo

3. Indicar el nivel para la aplicación inicial del operador: Mes y el valor de este nivel: Diciembre de 2013.

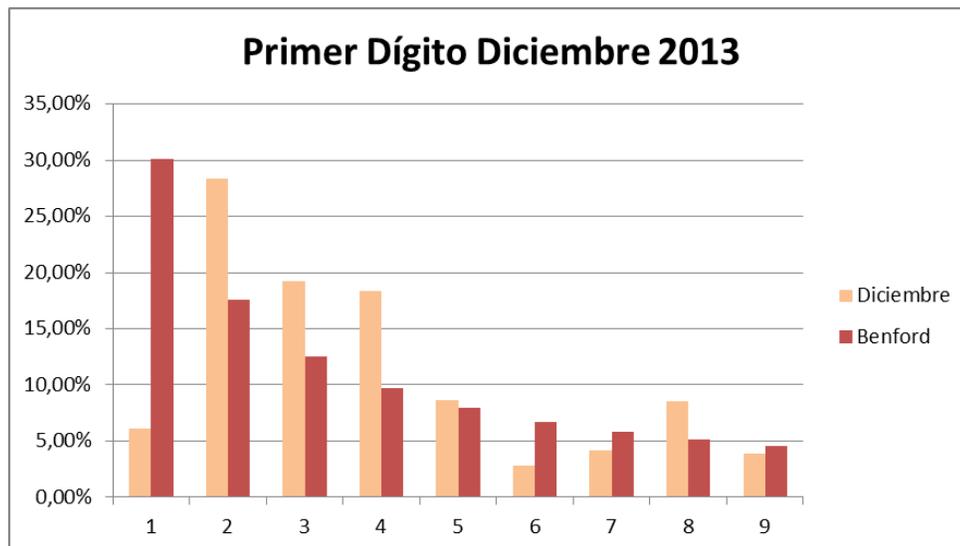
Los valores indicados en los pasos 1 al 3 son los parámetros de entrada del operador.

4. Se calcula la frecuencia del primer dígito en la medida numérica (ValorVenta) y se obtienen los resultados al aplicar la ley de Benford, ver Tabla 2-3 y Figura 2-6.

Tabla 2-3. Distribución primer dígito ValorVenta para Diciembre de 2013

Dígito	Benford	Diciembre
1	30,10%	6,09%
2	17,61%	28,37%
3	12,49%	19,24%
4	9,69%	18,32%
5	7,92%	8,68%
6	6,69%	2,77%
7	5,80%	4,15%
8	5,12%	8,51%
9	4,58%	3,88%

Figura 2-6. Distribución primer dígito ValorVenta para Diciembre de 2013



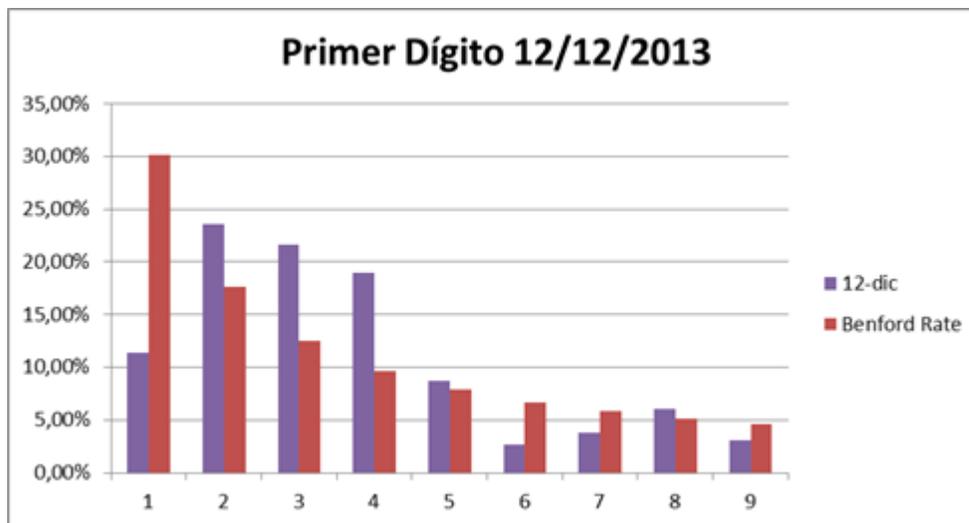
Los resultados muestran que *no se cumple* la ley de Benford. A continuación se aplica el operador en un nivel inferior (con respecto al nivel Mes) y en uno superior (también con respecto al nivel Mes) para analizar el comportamiento de la ley en estos niveles.

5. Indicar los niveles y sus valores correspondientes para aplicar el operador en un nivel inferior y en uno superior: Día (12/12/2013) y Año (2013), respectivamente.
6. Se calcula la frecuencia del primer dígito en la medida numérica (ValorVenta) y se obtienen los resultados al aplicar la ley de Benford, ver Tabla 2-4 y Figura 2-7.

Tabla 2-4. Distribución primer dígito ValorVenta para 12/12/2013

Dígito	Benford	12/12/2013
1	30,10%	11,41%
2	17,61%	23,57%
3	12,49%	21,67%
4	9,69%	19,01%
5	7,92%	8,75%
6	6,69%	2,66%
7	5,80%	3,80%
8	5,12%	6,08%
9	4,58%	3,04%

Figura 2-7. Distribución primer dígito ValorVenta para 12/12/2013

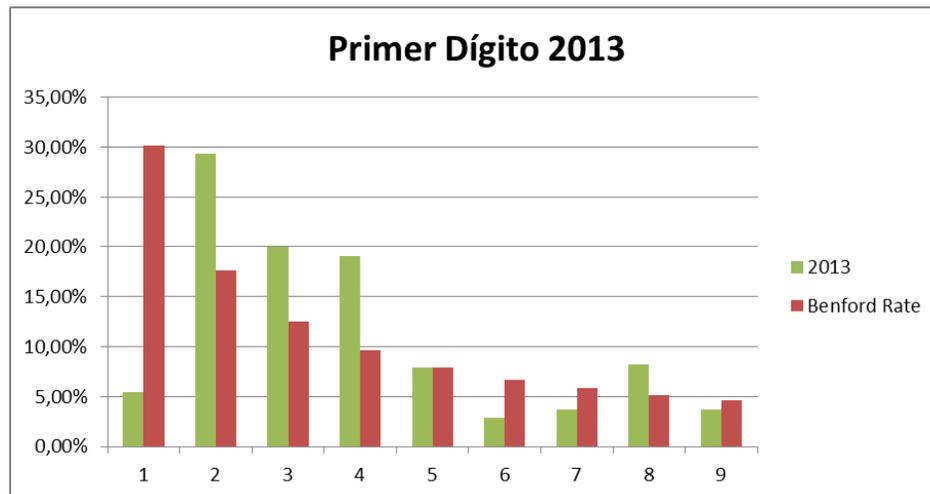


Los resultados a nivel de este día (12/12/2013) muestran un comportamiento similar al del mes de Diciembre 2013 (i.e., *no se cumple* la ley de Benford), en la Tabla 2-5 y en la Figura 2-8 se muestran los resultados a nivel del año 2013.

Tabla 2-5. Distribución primer dígito ValorVenta para 2013

Dígito	Benford	2013
1	30,10%	5,43%
2	17,61%	29,28%
3	12,49%	19,97%
4	9,69%	19,05%
5	7,92%	7,87%
6	6,69%	2,83%
7	5,80%	3,69%
8	5,12%	8,21%
9	4,58%	3,68%

Figura 2-8. Distribución primer dígito ValorVenta para 2013



7. Finalmente se tienen todos los resultados del operador, donde se observa para este ejemplo que no se cumple la ley de Benford, pero la distribución que se presenta para el primer dígito de ValorVenta es similar a nivel de Día (12/12/2013), Mes (Dic-2013) y Año (2013).

Con base en lo anterior se define el operador ley de Benford así:

Benford (N, m).

Donde:

- N es un conjunto de parejas $\{p_1, p_2, \dots, p_k\}$ y $p_i = (n_i, v_i)$ donde n_i es un nivel de una dimensión y v_i es un valor de dicho nivel.
Restricción: Sea n_i el nivel de una pareja p_i y sea n_j el nivel de una pareja p_j , $p_i \neq p_j$, entonces n_i y n_j son niveles de distintas dimensiones.
- m es una medida numérica.

Así, para el ejemplo anterior, si se desea aplicar el operador para el nivel Mes y para el valor Diciembre de 2013, con la medida ValorVenta entonces el operador se invoca así:

Benford ({(Mes: Dic/2013)}, ValorVenta).

De manera similar, si se desea aplicar el operador para el nivel año y para el valor 2013, con la medida ValorVenta:

Benford ({(Año: 2013)}, ValorVenta).

Ahora bien, si se desea aplicar, e.g., el operador para el nivel año con valor 2013, para el nivel línea de producto (LineaProd) con valor *Bike Accesorios*, con la medida ValorVenta:

Benford ({(Año: 2013), (LineaProd: Bike Accesories)}, ValorVenta).

De esta forma, el operador se aplica sobre un subconjunto de hechos (confróntese con el operador OLAP *Slice & Dice*): se extrae una tajada (o en general un sub-cubo) de la tabla de hechos donde a una medida numérica especificada se le aplica la ley de Benford.

Para ejemplificar esta extracción se muestra la acción del *Slice & Dice* en la Figura 2-9.

Figura 2-9. Extracción de tajada (sub-cubo)

Fecha	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	SignoMoneda	LineaProducto	NombreCliente
01/01/2011	3578,27	3578,27	286,2616	Mujer	Marco Alemán	Road-150 Red, 44	DEM	Road-150	Abigail Henderson
02/03/2011	3578,27	3578,27	286,2616	Mujer	Dólar	Road-150 Red, 56	USD	Road-150	Courtney Edwards
04/05/2011	3578,27	3578,27	286,2616	Hombre	Dólar Australiano	Road-150 Red, 44	AUD	Road-150	Jonathon Gutierrez
12/05/2011	3578,27	3578,27	286,2616	Mujer	Libra Esterlina	Road-150 Red, 44	GBP	Road-150	Larasha Alonso
02/01/2011	3578,27	3578,27	286,2616	Hombre	Dólar Australiano	Road-150 Red, 56	AUD	Road-150	Warren Jai
03/01/2011	3578,27	3578,27	286,2616	Mujer	Marco Alemán	Road-150 Red, 44	DEM	Road-150	Melanie Sanchez
24/10/2011	3578,27	3578,27	286,2616	Hombre	Libra Esterlina	Road-150 Red, 48	GBP	Road-150	Albert Blanco
05/12/2011	699,0982	699,0982	55,9279	Hombre	Dólar	Road-650 Black, 58	USD	Road-650	Kevin Huang
15/02/2011	3578,27	3578,27	286,2616	Mujer	Marco Alemán	Road-150 Red, 44	DEM	Road-150	Anne Dominguez
15/11/2011	3578,27	3578,27	286,2616	Mujer	Dólar Australiano	Road-150 Red, 44	AUD	Road-150	Laura Gonzalez
01/05/2012	1000,4375	1000,4375	80,035	Mujer	Dólar	Road-550-W Yellow	USD	Road-550-W	Kaitlyn Bennett
04/06/2012	2181,5625	2181,5625	174,525	Mujer	Dólar	Road-250 Red, 58	USD	Road-250	Jenna Mitchell
18/03/2012	2181,5625	2181,5625	174,525	Mujer	Dólar Australiano	Road-250 Black, 48	AUD	Road-250	Joselyn Adhe
08/01/2012	2443,35	2443,35	195,468	Mujer	Dólar Canadiense	Road-250 Red, 52	CAD	Road-250	Brianna Gray
01/01/2012	2443,35	2443,35	195,468	Mujer	Dólar	Road-250 Red, 52	USD	Road-250	Madeline Hall
25/01/2012	782,99	782,99	62,6392	Hombre	Dólar	Road-650 Black, 58	USD	Road-650	Erik Suarez
15/01/2012	2181,5625	2181,5625	174,525	Hombre	Libra Esterlina	Road-250 Red, 58	GBP	Road-250	Ross Martin
18/01/2012	2181,5625	2181,5625	174,525	Hombre	Dólar	Road-250 Red, 58	USD	Road-250	Jason Gonzales
18/01/2012	2181,5625	2181,5625	174,525	Mujer	Dólar	Road-250 Black, 44	USD	Road-250	Alicia Shen
18/01/2012	2049,0982	2049,0982	163,9279	Mujer	Dólar Canadiense	Mountain-200 Black	CAD	Mountain-200	Andrea Turner
05/12/2011	2443,35	2443,35	195,468	Mujer	Dólar Australiano	Road-250 Red, 52	AUD	Road-250	Angela Jai
19/02/2012	2181,5625	2181,5625	174,525	Mujer	Dólar	Road-250 Black, 58	USD	Road-250	Dana Dominguez
29/04/2012	2443,35	2443,35	195,468	Hombre	Dólar	Road-250 Red, 52	USD	Road-250	Jerome Torres
30/04/2012	2181,5625	2181,5625	174,525	Hombre	Dólar	Road-250 Red, 58	USD	Road-250	Dalton Parker
05/08/2012	2443,35	2443,35	195,468	Mujer	Dólar	Road-250 Red, 52	USD	Road-250	Olivia Garcia
19/11/2012	2443,35	2443,35	195,468	Hombre	Dólar	Road-250 Red, 52	USD	Road-250	Robert Hill
23/01/2013	9,99	9,99	0,7992	Mujer	Dólar Australiano	Water Bottle 300 cc	AUD	Water Bottle	Rob Verhoff
14/12/2013	9,99	9,99	0,7992	Mujer	Dólar Australiano	Mountain Bottle Ca	AUD	Mountain Bottle Ca	Rob Verhoff
13/07/2013	2,29	2,29	0,1832	Mujer	Dólar Australiano	Patch Kit/8 Patches	AUD	Patch kit	Rob Verhoff

Tabla de Hechos:
FactVentasTajada 1: DimTiempo
Nivel Año: 2012Tajada 2: DimTiempo
Nivel Mes: Ene/2012Tajada 3: DimTiempo
Nivel Día: 18/01/2012

2.2.2 SAME-SAME-SAME

Para aplicar este operador se debe elegir de la tabla de hechos (FactVentas) una medida, aquí ValorVenta, además se elige un nivel de una dimensión para aplicar inicialmente el operador. Aquí se usa la dimensión DimTiempo y se considera para la aplicación inicial el año de la compra (e.g., 2013), luego se aplicará con el mes de la compra (e.g. Mayo 2013) y el día de la compra (e.g., 24/05/2013). Para este operador se debe elegir un nivel de una o más dimensiones adicionales, aquí se usará: el Producto (DimProducto) y el NombreCliente (DimCliente), en la Tabla 2-6 se presenta una muestra de datos que se usará para este ejemplo.

Tabla 2-6. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de *SAME-SAME-SAME*

Fecha	Producto	NombreCliente	ValorVenta
01/01/2013	AWC Logo Cap	Maurice Shan	8,99
01/01/2013	Bike Wash - Dissolver	Amanda Sanders	7,95
01/01/2013	Bike Wash - Dissolver	Oscar Perry	7,95
01/01/2013	Fender Set - Mountain	Shannon Wang	21,98
01/01/2013	Half-Finger Gloves, S	Karl Andersen	24,49
01/01/2013	HL Road Tire	Amanda Sanders	32,6
01/01/2013	LL Mountain Tire	Kaylee Gray	24,99
01/01/2013	Mountain Bottle Cage	Cindy Patel	9,99
01/01/2013	Mountain Bottle Cage	Nicole Brown	9,99
01/01/2013	Mountain-200 Black, 42	Cindy Patel	2294,99
01/01/2013	Mountain-200 Black, 46	Shannon Wang	2294,99
01/01/2013	Mountain-200 Silver, 38	Nicole Brown	2319,99
01/01/2013	Mountain-500 Silver, 42	Kaylee Gray	564,99
01/01/2013	Patch Kit/8 Patches	Kaylee Gray	2,29

A continuación se presentan los pasos para aplicar el operador.

1. Indicar la medida numérica de la tabla de hechos (FactVentas): ValorVenta.
2. Indicar la dimensión que se usará: DimTiempo.
3. Indicar el nivel para la aplicación inicial del operador: Año y el valor de este nivel: 2013.
4. Indicar niveles adicionales de otras dimensiones y sus valores: DimCliente: NombreCliente y DimProducto: Producto, para ambos niveles se tomarán todos sus valores.

Los valores indicados en los pasos 1 al 4 son los parámetros de entrada del operador.

5. Se aplica la técnica *SAME-SAME-SAME*, y se obtiene el número de repeticiones para los valores iguales de: ValorVenta, Producto y NombreCliente en 2013, en la Tabla 2-7 se muestra el resultado obtenido.

Tabla 2-7. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 2013

Año	Producto	NombreCliente	ValorVenta	NumRepeticiones
2013	Touring Tire Tube	April Shan	4,99	15
2013	Patch Kit/8 Patches	Dalton Perez	2,29	10
2013	Patch Kit/8 Patches	Samantha Jenkins	2,29	10
2013	Road Tire Tube	Jennifer Simmons	3,99	9
2013	Patch Kit/8 Patches	Jennifer Simmons	2,29	8
2013	Patch Kit/8 Patches	Mason Roberts	2,29	8
2013	Road Tire Tube	Ryan Thompson	3,99	8
2013	Patch Kit/8 Patches	April Shan	2,29	7
2013	Patch Kit/8 Patches	Ashley Henderson	2,29	7
2013	Patch Kit/8 Patches	Henry Garcia	2,29	7
2013	Patch Kit/8 Patches	Luis Diaz	2,29	7
2013	Patch Kit/8 Patches	Nancy Chapman	2,29	7
2013	Road Tire Tube	Ashley Henderson	3,99	7
2013	Road Tire Tube	Charles Jackson	3,99	7
2013	Road Tire Tube	Dalton Perez	3,99	7
2013	Road Tire Tube	Fernando Barnes	3,99	7
2013	Patch Kit/8 Patches	Fernando Barnes	2,29	6
2013	Road Tire Tube	Daniel Davis	3,99	6

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD es 15, se aplicará la técnica a nivel de mes y día para analizar estos resultados.

6. Indicar otros niveles y valores correspondientes para aplicar el operador: Mes (Mayo 2013) y Día (24/05/2013).
7. Al aplicar la técnica *SAME-SAME-SAME* se obtiene el número de repeticiones para los valores iguales de: ValorVenta, Producto y NombreCliente en Mayo 2013, en la Tabla 2-8 se muestra el resultado obtenido.

Tabla 2-8. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en Mayo 2013

Mes	Producto	NombreCliente	ValorVenta	NumRepeticiones
201305	HL Road Tire	Ryan Thompson	32,6	2
201305	Long-Sleeve Logo Jersey, M	Gina Martin	49,99	2
201305	Patch Kit/8 Patches	Jordan Baker	2,29	2
201305	Patch Kit/8 Patches	Miranda Gonzales	2,29	2
201305	Road Bottle Cage	Anne Hernandez	8,99	2
201305	Road Tire Tube	Ana Perry	3,99	2
201305	Road Tire Tube	Fernando Barnes	3,99	2
201305	Road Tire Tube	Ryan Thompson	3,99	2
201305	Touring Tire Tube	April Shan	4,99	2
201305	Water Bottle - 30 oz.	Anne Hernandez	4,99	2

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD para Mayo 2013 es 2.

- Al aplicar la técnica *SAME-SAME-SAME* se obtiene el número de repeticiones para los valores iguales de: ValorVenta, Producto y NombreCliente el 24/05/2013, en la Tabla 2-9 se muestra el resultado obtenido.

Tabla 2-9. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 24/05/2013

Fecha	Producto	NombreCliente	ValorVenta	NumRepeticiones
24/05/2013	HL Road Tire	Ryan Thompson	32,6	2
24/05/2013	Road Tire Tube	Ryan Thompson	3,99	2

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD para 24/05/2013 es 2.

Con base en lo anterior se define el operador *SAME-SAME-SAME* así:

$$SSS(N, m)$$

Donde:

- N es un conjunto de parejas $\{p_1, p_2, \dots, p_k\}$ y $p_i = (n_i, v_i)$ donde n_i es el nivel de una dimensión y v_i es un valor de dicho nivel.

Restricción: Sea n_i el nivel de una pareja p_i y sea n_j el nivel de una pareja p_j , $p_i \not\sqsubseteq p_j$, entonces n_i y n_j son niveles de distintas dimensiones.

- m es un conjunto de medidas $\{m_1, m_2, \dots, m_k\}$.

Así, para el ejemplo anterior, si se desea aplicar el operador para el nivel Mes con el valor Mayo de 2013, y para los niveles adicionales Producto y NombreCliente con valor * (el asterisco indica que se deben seleccionar todos los datos del nivel) para ambos, con la medida ValorVenta entonces el operador se invoca así:

SSS ({(Mes: May/2013)}, {(Producto: *)}, {(NombreCliente: *)}, {ValorVenta})

De manera similar, si se desea aplicar el operador para el nivel año y para el valor 2013, los niveles Producto y NombreCliente, con la medida ValorVenta:

SSS ({(Año: 2013)}, {(Producto: *)}, {(NombreCliente: *)}, {ValorVenta})

Ahora bien, si se desea aplicar, e.g., el operador para el nivel año con valor 2013, para los niveles Producto con valor (*) y Promoción con valor *Volume Discount*, con las medidas ValorVenta y PrecioUnidad:

SSS ({(Año: 2013)}, {(Producto: *)}, {(Promoción: Volume Discount)}, {ValorVenta, PrecioUnidad}).

De esta forma, el operador se aplica sobre un subconjunto de hechos (confróntese con el operador OLAP *Slice & Dice*): se extrae una tajada (o en general un sub-cubo) de la tabla de hechos donde se aplica la técnica *SAME-SAME-SAME*.

2.2.3 SAME-SAME-DIFFERENT

Este operador es similar al anterior. Para aplicar este operador se debe elegir de la tabla de hechos (FactVentas) una medida, aquí ValorVenta, un nivel de una dimensión para aplicar inicialmente el operador: la dimensión DimTiempo y se considera para la aplicación inicial el año de la compra (e.g., 2012), luego se aplicará con el mes de la compra (e.g., Diciembre 2012) y el día de la compra (e.g., 30/12/2012). Para este operador se debe elegir un nivel de una o más dimensiones adicionales, aquí se usará: el Producto (DimProducto) y el NombreCliente (DimCliente), en la Tabla 2-10 se presenta una muestra de datos que se usará para este ejemplo.

Tabla 2-10. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de *SAME-SAME-DIFFERENT*

Fecha	Producto	NombreCliente	ValorVenta
01/01/2012	Mountain-200 Silver, 38	Connor Li	2071,4196
01/01/2012	Road-250 Black, 48	Erika Munoz	2181,5625
01/01/2012	Road-250 Black, 52	Alison Kumar	2181,5625
01/01/2012	Road-250 Red, 48	Mackenzie Ramirez	2443,35
01/01/2012	Road-250 Red, 52	Daniel Harris	2443,35
01/01/2012	Road-250 Red, 52	Keith Deng	2443,35
01/01/2012	Road-550-W Yellow, 38	Kaitlyn Bennett	1000,4375
01/01/2012	Road-650 Black, 62	Joe Jordan	782,99
01/01/2012	Road-650 Black, 62	Johnathan Kapoor	782,99
02/01/2012	Mountain-200 Black, 42	Casey Munoz	2049,0982
02/01/2012	Mountain-200 Silver, 46	Erick Sai	2071,4196
02/01/2012	Road-250 Black, 44	Manuel Garcia	2181,5625
02/01/2012	Road-250 Black, 48	Caleb Long	2181,5625
02/01/2012	Road-250 Black, 48	Stephanie Cooper	2181,5625
02/01/2012	Road-250 Red, 44	Michelle James	2443,35
02/01/2012	Road-250 Red, 44	Yolanda Sharma	2443,35
02/01/2012	Road-250 Red, 48	Gabriella Collins	2443,35
02/01/2012	Road-250 Red, 58	Sean Baker	2181,5625

A continuación se presentan los pasos para aplicar el operador.

1. Indicar la medida numérica de la tabla de hechos (FactVentas): ValorVenta.
2. Indicar la dimensión que se usará: DimTiempo.
3. Indicar el nivel para la aplicación inicial del operador: Año y el valor de este nivel: 2012.
4. Indicar niveles adicionales de otras dimensiones y sus valores: DimProducto: Producto y DimCliente: NombreCliente, para ambos niveles se tomarán todos sus valores.

Los valores indicados en los pasos 1 al 4 son los parámetros de entrada del operador.

5. Se aplica la técnica *SAME-SAME-DIFFERENT* y se obtiene el número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 2012, en la Tabla 2-11 se muestra el resultado obtenido.

Tabla 2-11. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 2012

Año	Producto	NombreCliente	ValorVenta	NumRepeticiones
2012	Mountain-200 Black, 46	Aaron Hughes	2049,0982	200
2012	Mountain-200 Black, 46	Aidan Powell	2049,0982	200
2012	Mountain-200 Black, 46	Alan Huang	2049,0982	200
2012	Mountain-200 Black, 46	Albert Moreno	2049,0982	200
2012	Mountain-200 Black, 46	Alejandro Kumar	2049,0982	200
2012	Mountain-200 Black, 46	Alejandro Lal	2049,0982	200
2012	Mountain-200 Black, 46	Alex Hernandez	2049,0982	200
2012	Mountain-200 Black, 46	Alexandra Collins	2049,0982	200
2012	Mountain-200 Black, 46	Alexia Henderson	2049,0982	200
2012	Mountain-200 Black, 46	Amanda Morris	2049,0982	200
2012	Mountain-200 Black, 46	Amber Roberts	2049,0982	200

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD es 200, se aplicará la técnica a nivel de mes y día para analizar estos resultados.

- Indicar otros niveles y valores correspondientes para aplicar el operador: Mes (Diciembre 2012) y Día (30/12/2012).
- Al aplicar la técnica *SAME-SAME-DIFFERENT* se obtiene el número de repeticiones para los valores iguales de: ValorVenta, Producto y NombreCliente en Diciembre 2012, en la Tabla 2-12 se muestra el resultado obtenido.

Tabla 2-12. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en Diciembre 2012

Mes	Producto	NombreCliente	ValorVenta	NumRepeticiones
201212	Mountain-200 Silver, 46	Kevin Hill	2071,4196	27
201212	Mountain-200 Silver, 46	Jorge Sun	2071,4196	27
201212	Mountain-200 Silver, 46	Darryl Wu	2071,4196	27
201212	Mountain-200 Silver, 46	Jessica Rivera	2071,4196	27
201212	Mountain-200 Silver, 46	Willie Chande	2071,4196	27
201212	Mountain-200 Silver, 46	Candace Chapman	2071,4196	27
201212	Mountain-200 Silver, 46	Samantha Martin	2071,4196	27
201212	Mountain-200 Silver, 46	Shannon Guo	2071,4196	27
201212	Mountain-200 Silver, 46	Jacquelyn Blanco	2071,4196	27
201212	Mountain-200 Silver, 46	Nicolas Raje	2071,4196	27
201212	Mountain-200 Silver, 46	Bobby Rodriguez	2071,4196	27
201212	Mountain-200 Silver, 46	Paula Moreno	2071,4196	27

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD para Diciembre 2012 es 27.

8. Al aplicar la técnica *SAME-SAME-DIFFERENT* se obtiene el número de repeticiones para los valores iguales de: ValorVenta, Producto y NombreCliente el 30/12/2012, en la Tabla 2-13 se muestra el resultado obtenido.

Tabla 2-13. Número de repeticiones para los valores de: ValorVenta, Producto y NombreCliente en 30/12/2012

Fecha	Producto	NombreCliente	ValorVenta	NumRepeticiones
30/12/2012	Water Bottle - 30 oz.	Latoya Nara	4,99	7
30/12/2012	Water Bottle - 30 oz.	Anne Hernandez	4,99	7
30/12/2012	Water Bottle - 30 oz.	Franklin Raji	4,99	7
30/12/2012	Water Bottle - 30 oz.	Frank Navarro	4,99	7
30/12/2012	Water Bottle - 30 oz.	Ruben Patel	4,99	7
30/12/2012	Water Bottle - 30 oz.	April Deng	4,99	7
30/12/2012	Water Bottle - 30 oz.	Kendra Romero	4,99	7
30/12/2012	Mountain Bottle Cage	Anne Hernandez	9,99	4
30/12/2012	Mountain Bottle Cage	Franklin Raji	9,99	4
30/12/2012	Mountain Bottle Cage	Frank Navarro	9,99	4
30/12/2012	Mountain Bottle Cage	April Deng	9,99	4
30/12/2012	Road Bottle Cage	Latoya Nara	8,99	4

En los resultados se observa que el mayor número de repeticiones para los valores registrados en la BD para 30/12/2012 es 7.

Con base en lo anterior se define el operador *SAME-SAME-DIFFERENT* así:

$$SSD(N, m, d)$$

Donde:

- N es un conjunto de parejas $\{p_1, p_2, \dots, p_k\}$ y $p_i = (n_i, v_i)$ donde n_i es el nivel de una dimensión y v_i es un valor de dicho nivel.
Restricción: Sea n_i el nivel de una pareja p_i y sea n_j el nivel de una pareja p_j , $p_i \not\sqsubseteq p_j$, entonces n_i y n_j son niveles de distintas dimensiones.
- m es un conjunto de medidas $\{m_1, m_2, \dots, m_k\}$.
- d es un nivel de una dimensión o una medida.

Restricciones:

- Sea d un nivel y sea n_i el nivel de una pareja p_i , entonces d y n_i son niveles de distintas dimensiones.
- Sea d una medida y sea m_i una medida, entonces d y m_i son medidas diferentes.

Así, para el ejemplo anterior, si se desea aplicar el operador para el nivel Mes y para el valor Diciembre de 2012, el nivel Producto con valor (*), con la medida ValorVenta y el nivel NombreCliente como d , entonces el operador se invoca así:

SSD ({(Mes: Dic/2012)}, (Producto: *)), {ValorVenta}, NombreCliente)

De manera similar, si se desea aplicar el operador para el nivel año y para el valor 2012, el nivel Producto con valor (*), con la medida ValorVenta y el nivel NombreCliente como d :

SSD ({(Año: 2012)}, (Producto: *)), {ValorVenta}, NombreCliente)

Ahora bien, si se desea aplicar, e.g., el operador para el nivel año con valor 2012, para los niveles Producto con valor *Road Tire Tube*, con la medida ValorVenta y la medida PrecioUnidad como d :

SSD ({(Año: 2012), (Producto: Road Tire Tube)}, {ValorVenta}, PrecioUnidad).

De esta forma, el operador se aplica sobre un subconjunto de hechos (confróntese con el operador OLAP *Slice & Dice*): se extrae una tajada (o en general un sub-cubo) de la tabla de hechos donde se aplica la técnica *SAME-SAME- DIFFERENT*.

2.2.4 Relative Size Factor

Para aplicar este operador se debe elegir de la tabla de hechos (FactVentas) una medida numérica, aquí se usa ValorVenta, además se elige un nivel de una dimensión para aplicar inicialmente el operador. Se usa la dimensión DimTiempo y se considera para la aplicación inicial del operador el nivel Mes de la venta (Agosto de 2011). Luego, el operador se aplica a en un nivel inferior (Día 08/08/2011) y en un nivel superior (Año 2011). En la Tabla 2-14 se presenta una muestra de datos que se usará para el ejemplo.

Tabla 2-14. Muestra de datos (fragmento de la tabla de hechos) para aplicar el operador de *Relative Size Factor*

Fecha	ValorVenta
01/01/2011	3578,27
02/01/2011	3578,27
02/01/2011	699,0982
03/01/2011	3578,27
04/01/2011	3578,27
04/01/2011	699,0982
05/01/2011	3578,27
05/01/2011	699,0982
06/01/2011	3578,27
06/01/2011	3399,99
06/01/2011	3374,99
07/01/2011	3578,27
07/01/2011	3399,99
08/01/2011	3578,27
09/01/2011	3578,27
09/01/2011	3399,99

A continuación se presentan los pasos para aplicar el operador.

1. Indicar la medida numérica de la tabla de hechos (FactVentas): ValorVenta.
2. Indicar la dimensión que se usará: DimTiempo
3. Indicar el nivel para la aplicación inicial del operador: Mes y el valor de este nivel: Agosto de 2011.

Los valores indicados en los pasos 1 al 3 son los parámetros de entrada del operador.

4. Se calcula el *Relative Size Factor* en la medida numérica (ValorVenta) y se obtienen los resultados, ver Tabla 2-15.

Tabla 2-15. *Relative Size Factor* para Agosto 2011

<i>Relative Size Factor</i>	
Agosto 2011	1,0524

El resultado es mayor que 1, por lo cual *podrían* existir fraudes en los datos. A continuación se aplica el operador en un nivel inferior (con respecto al nivel Mes) y en uno superior (también con respecto al nivel Mes) para analizar el resultado en estos niveles.

- Indicar los niveles y sus valores correspondientes para aplicar el operador en un nivel inferior y en uno superior: Día (08/08/2011) y Año (2011), respectivamente.
- Se calcula el *Relative Size Factor* en la medida numérica (ValorVenta) para el año 2011 y se obtienen los resultados, ver Tabla 2-16.

Tabla 2-16. *Relative Size Factor* para 2011

<i>Relative Size Factor</i>	
2011	1,0524

El resultado es mayor que 1, por lo cual podrían existir fraudes en los datos.

- Se calcula el *Relative Size Factor* en la medida numérica (ValorVenta) para el día 08/08/2011 y se obtienen los resultados, ver Tabla 2-17.

Tabla 2-17. *Relative Size Factor* para 08/08/2011

<i>Relative Size Factor</i>	
08/08/2011	1,06023

El resultado es mayor que 1, por lo cual podrían existir fraudes en los datos.

Con base en lo anterior se define el operador *Relative Size Factor* así:

RSF (N, m).

Donde:

- N es un conjunto de parejas $\{p_1, p_2, \dots, p_k\}$ y $p_i = (n_i, v_i)$ donde n_i es un nivel de una dimensión y v_i es un valor de dicho nivel.

Restricción: Sea n_i el nivel de una pareja p_i y sea n_j el nivel de una pareja p_j , $p_i \neq p_j$, entonces n_i y n_j son niveles de distintas dimensiones.

- m es una medida numérica.

Así, para el ejemplo anterior, si se desea aplicar el operador para el nivel Mes y para el valor Agosto 2011, con la medida ValorVenta entonces el operador se invoca así:

RSF ($\{(Mes: Ago/2011)\}$, ValorVenta).

De manera similar, si se desea aplicar el operador para el nivel año y para el valor 2011, con la medida ValorVenta:

RSF ($\{(Año: 2011)\}$, ValorVenta).

Ahora bien, si se desea aplicar, e.g., el operador para el nivel año con valor 2011, para el nivel de DimCliente (NombreCliente) con valor "Alexandra Thomas", con la medida ValorVenta:

RSF ($\{(Año: 2011), (NombreCliente: Alexandra Thomas)\}$, ValorVenta).

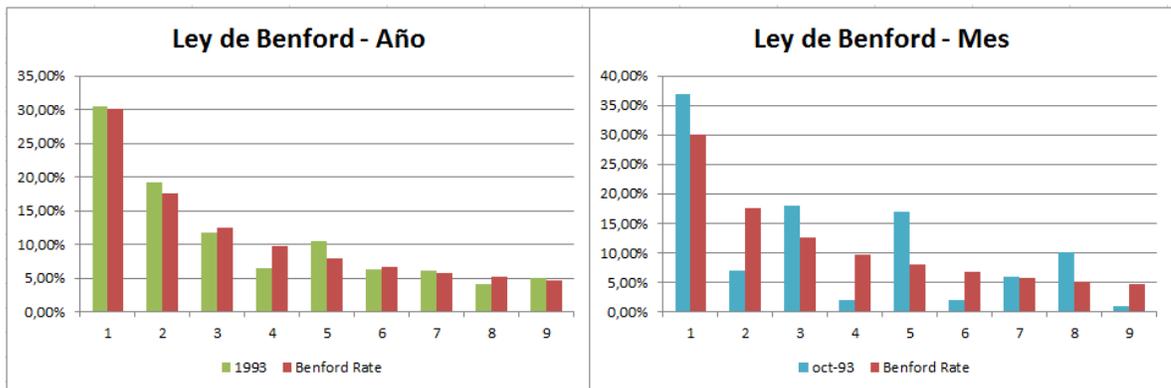
De esta forma, el operador se aplica sobre un subconjunto de hechos (confróntese con el operador OLAP *Slice & Dice*): se extrae una tajada (o en general un sub-cubo) de la tabla de hechos donde a una medida numérica especificada se le aplica *Relative Size Factor*.

2.3. Posibles escenarios

Se presentan algunos escenarios que podrían surgir al aplicar la ley de Benford en diferentes niveles de granularidad de una dimensión. Estos escenarios sirven como base para los análisis posteriores en el caso de estudio.

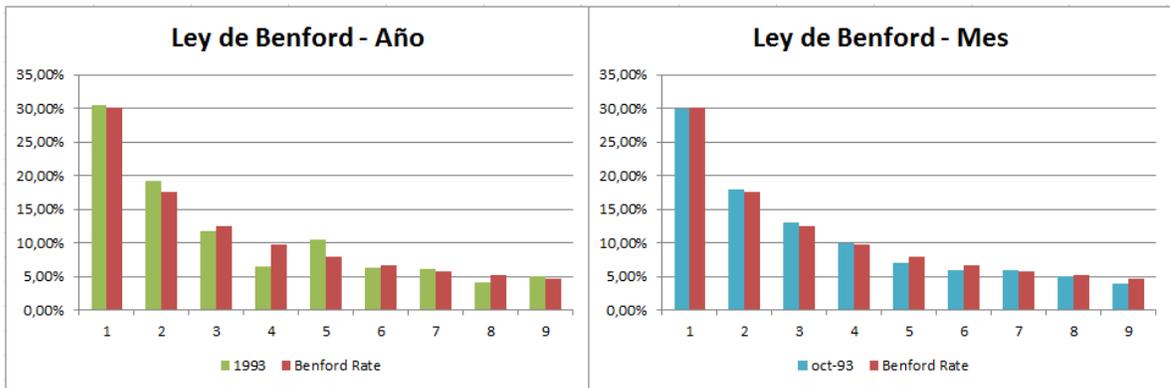
- Escenario 1: Este escenario se presenta cuando al aplicar el operador de Benford a los datos a nivel de año (e.g., 1993) esta se cumple, pero a nivel de mes (e.g., Octubre 1993) la ley no se cumple. Este escenario sugiere que en un nivel superior i.e. nivel de año, no se perciben fraudes pero al examinar con más detalle los datos sí *podría* haber fraudes. Es decir, los posibles fraudes se “camuflan” cuando se observan los datos con un nivel más alto, i.e., nivel de año, véase la Figura 2-10.

Figura 2-10. Distribución primer dígito ValorVenta para 1993 y Octubre 1993



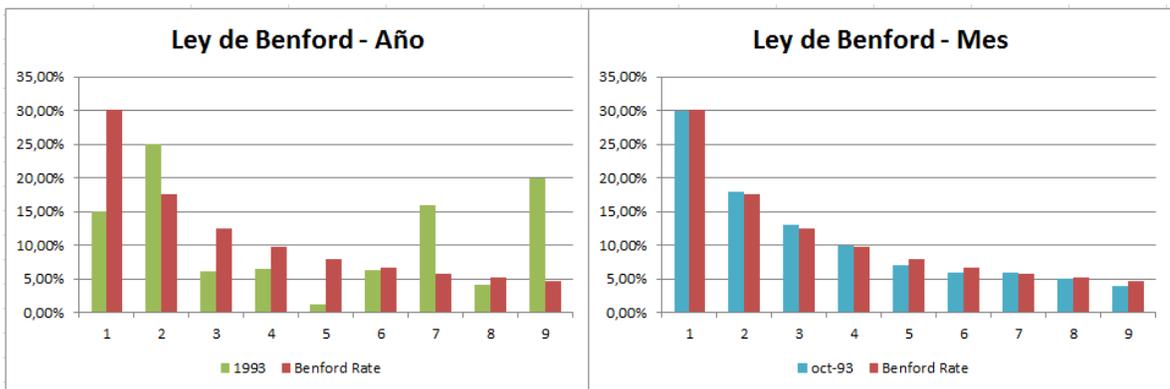
- Escenario 2: Este escenario se presenta cuando al aplicar el operador de Benford a los datos a nivel de año (e.g., 1993) esta se cumple y a nivel de mes (e.g., Octubre 1993) la ley *también* se cumple. Este escenario sugiere que ni en el nivel superior, i.e. nivel de año, ni en el inferior, i.e. nivel de mes, se perciben fraudes, véase la Figura 2-11.

Figura 2-11. Distribución primer dígito ValorVenta para 1993 y Octubre 1993



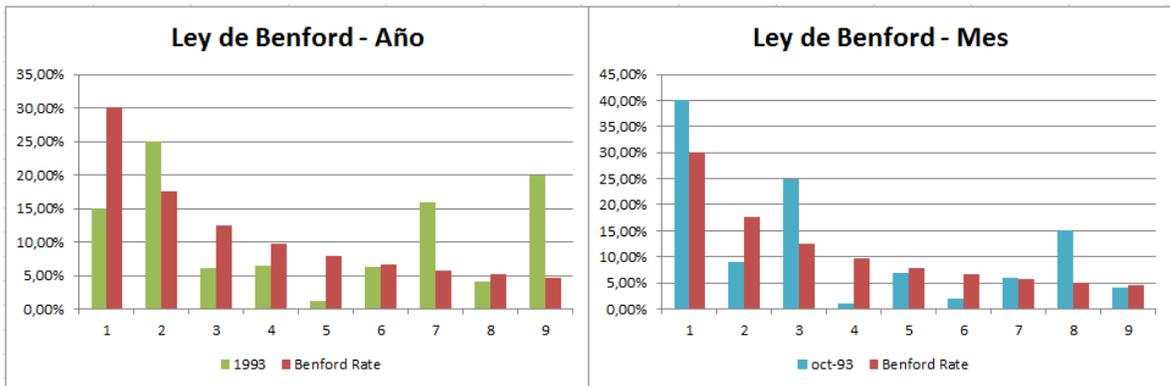
- Escenario 3: Este escenario se presenta cuando al aplicar el operador de Benford a los datos a nivel de año (e.g., 1993) esta no se cumple, pero a nivel de mes (e.g., Octubre 1993) la ley sí se cumple. Este escenario sugiere que en un nivel superior, i.e. nivel de año, se perciben fraudes pero al examinar con más detalle los datos no hay indicios de fraudes. Es decir, los posibles fraudes se perciben en un nivel superior, i.e. nivel de año, pero no en uno inferior, i.e. nivel de mes, véase la Figura 2-12.

Figura 2-12. Distribución primer dígito ValorVenta para 1993 y Octubre 1993



- Escenario 4: Este escenario se presenta cuando al aplicar el operador de Benford a los datos a nivel de año (e.g., 1993) esta no se cumple y a nivel de mes (e.g., Octubre 1993) la ley tampoco se cumple. Este escenario sugiere que en un nivel superior, i.e. nivel de año, se perciben fraudes y al examinar con más detalle los posibles fraudes se siguen percibiendo. Es decir, hay indicios de fraudes tanto en un nivel superior, i.e., nivel de año, como en uno inferior, i.e. nivel de mes, véase la Figura 2-13.

Figura 2-13. Distribución primer dígito ValorVenta para 1993 y Octubre 1993



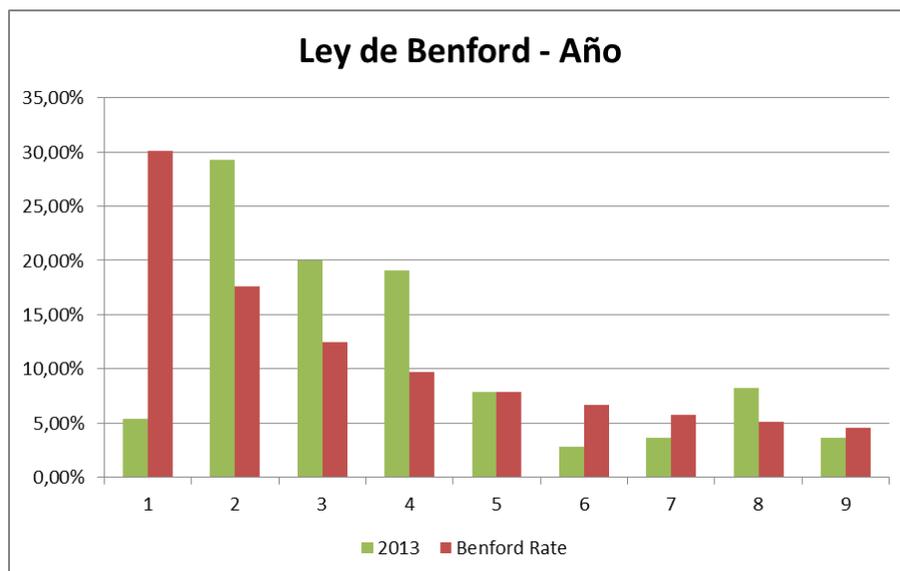
2.4. Resultados

Continuando con el análisis de la BD del caso de estudio planteado, se muestran los resultados obtenidos al aplicar los cuatro operadores definidos.

2.4.1 Ley de Benford

Para analizar la BD, se comienza aplicando el operador de la ley de Benford para el nivel Año y la medida ValorVenta: Benford ({{Año: 2013}}, ValorVenta). La distribución obtenida se muestra en la Figura 2-14:

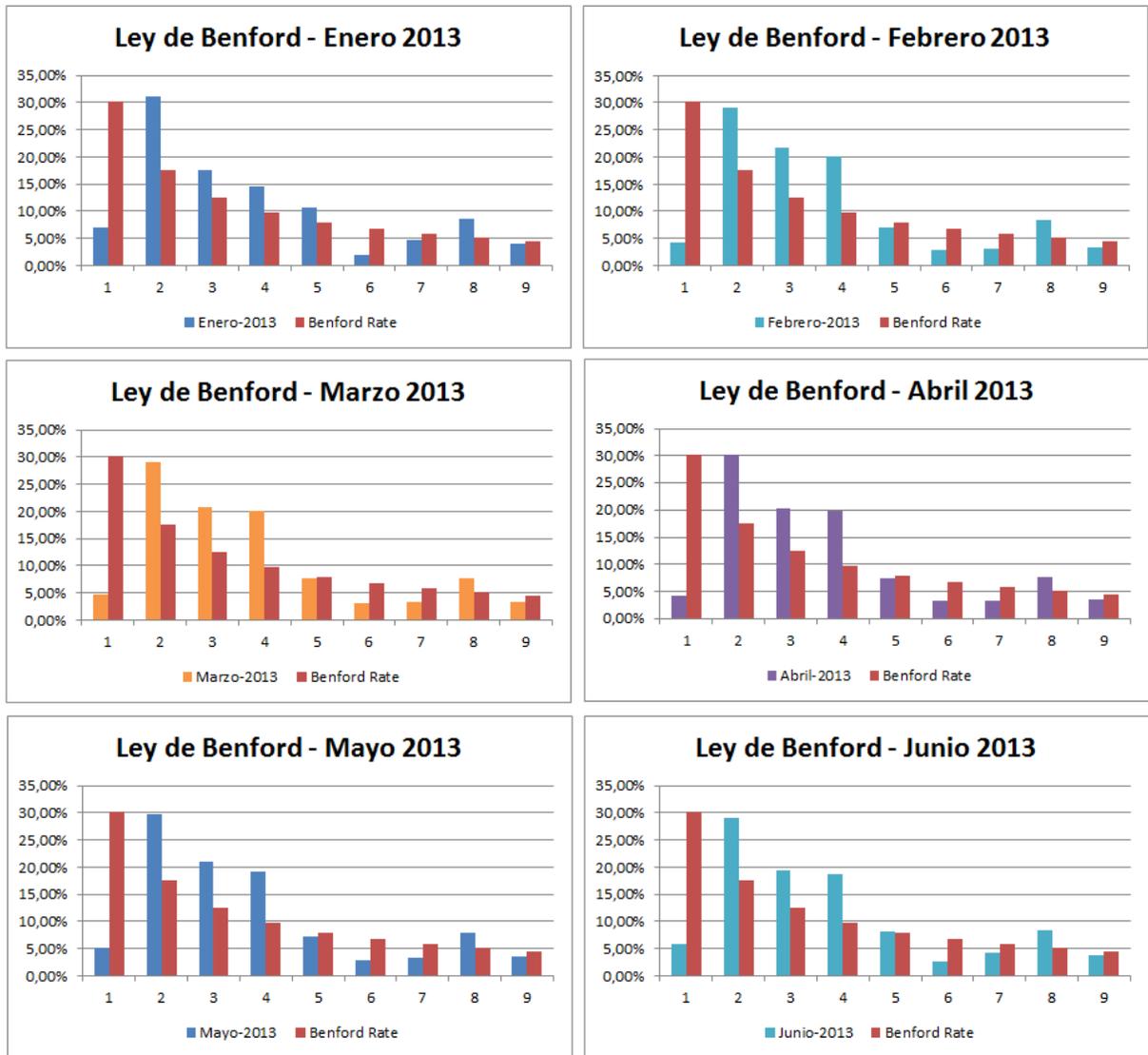
Figura 2-14. Distribución primer dígito ValorVenta para 2013



En la distribución obtenida *no se cumple* la ley de Benford, por lo cual se analizará a nivel de Mes para observar si este comportamiento se mantiene.

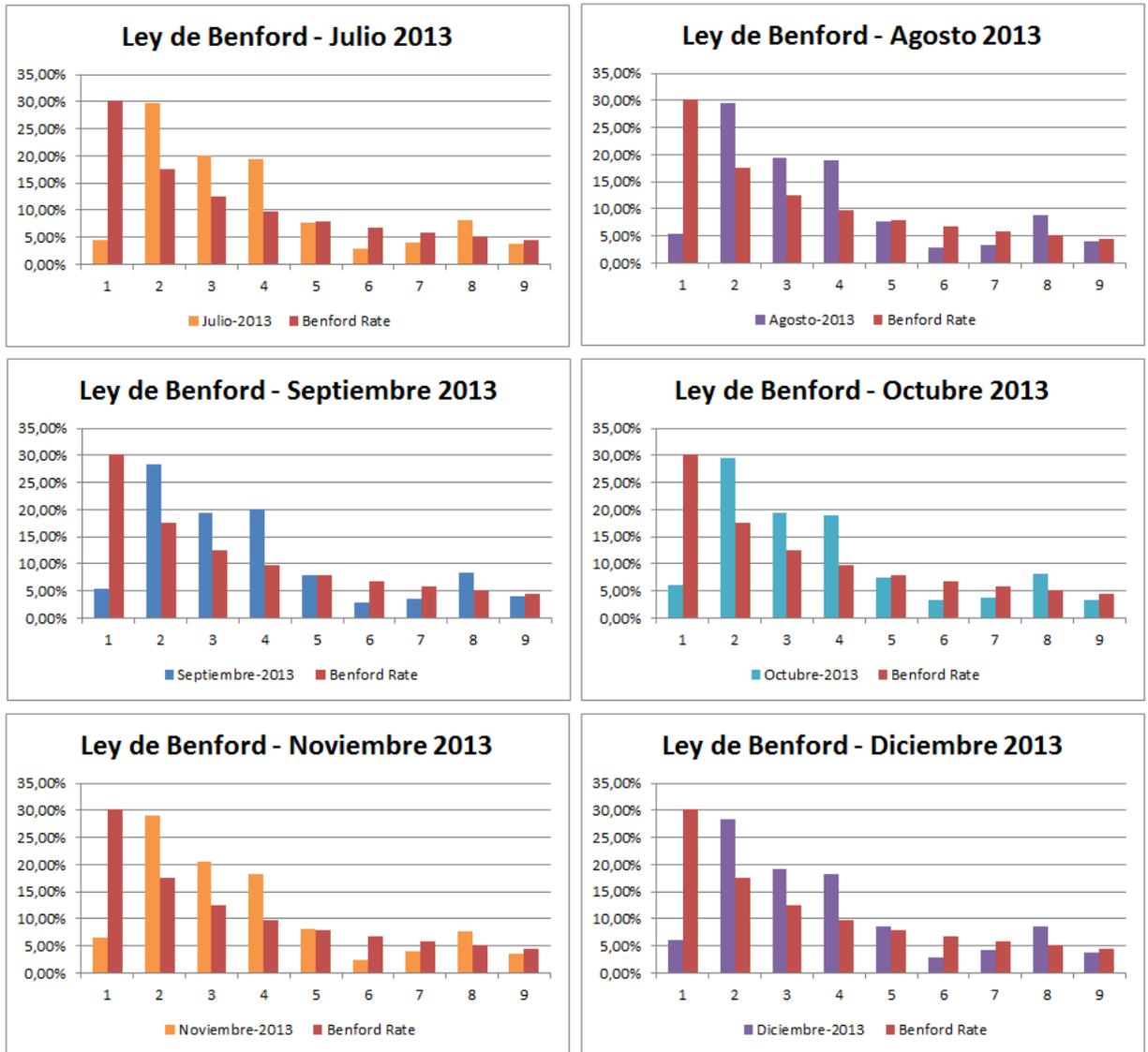
En las Figuras 2-15 y 2-16, se observan los resultados de las distribuciones Benford al aplicarlas a *cada mes* del año 2013, la aplicación del operador para Enero de 2013 es así: Benford (({Mes: Ene/2013}), ValorVenta) y es análogo para los demás meses.

Figura 2-15. Distribuciones primer dígito ValorVenta para Enero 2013 – Junio 2013



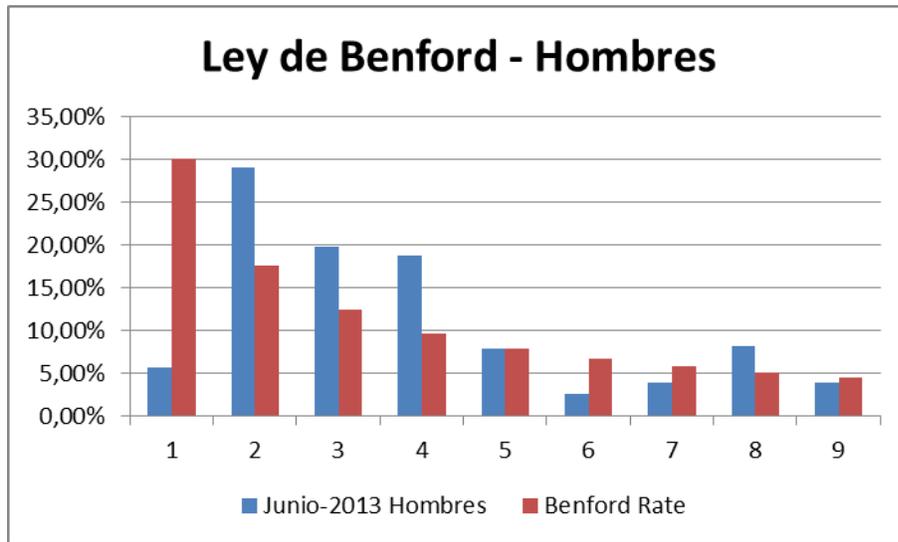
En las distribuciones para los primeros seis meses del 2013, se conserva el comportamiento de la distribución del 2013, i.e., no se cumple la ley de Benford.

Figura 2-16. Distribuciones primer dígito ValorVenta para Julio 2013 – Diciembre 2013



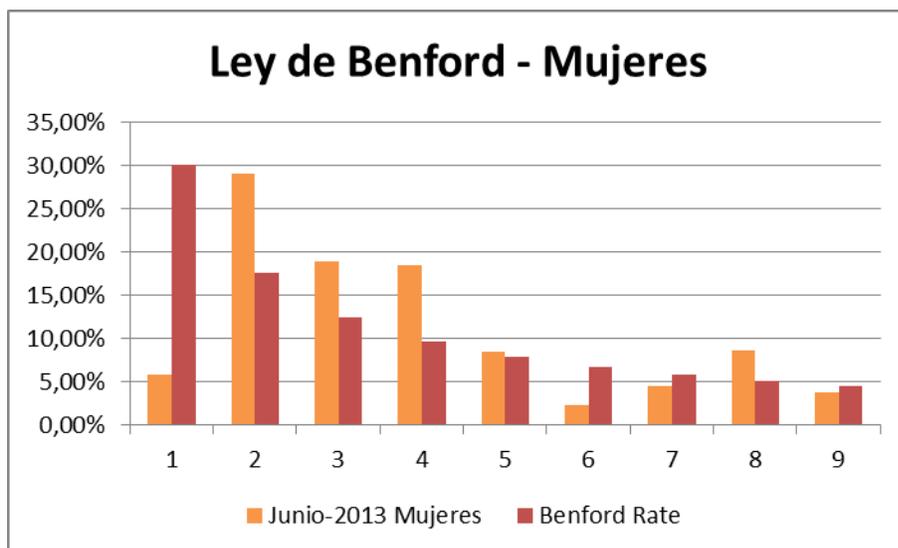
Al igual que en las distribuciones de los primeros seis meses del 2013, en los otros seis meses *tampoco* se cumple la ley de Benford. Se analizará con un nivel de detalle más específico, eligiendo el mes de Junio y separando por el Género del cliente que hizo la compra para aplicar el operador nuevamente: Benford ((Mes: Jun/2013), (Genero: Hombre)), ValorVenta), véase la Figura 2-17.

Figura 2-17. Distribución primer dígito ValorVenta para Junio 2013 – Hombres



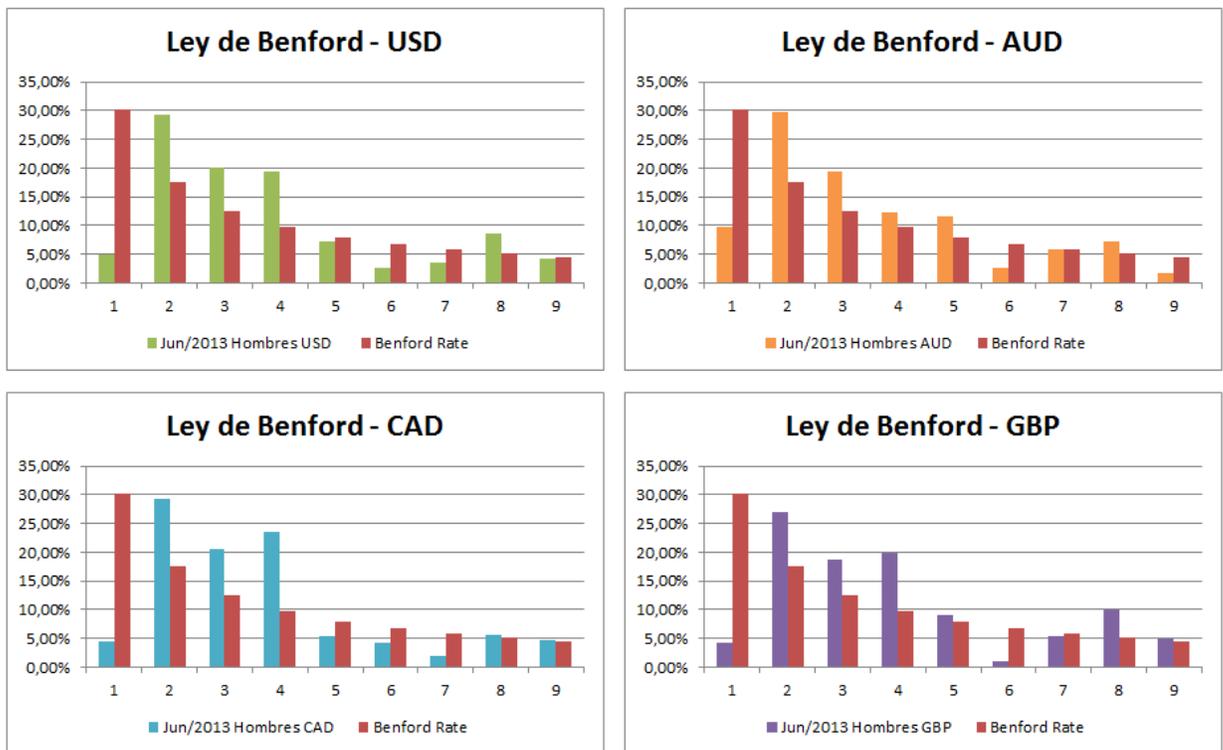
Se aplicó la ley de Benford para el mes de Junio 2013, solo para compras hechas por hombres y continúa sin cumplirse la ley de Benford como se observa en la Figura 2-17, se aplica nuevamente el operador pero ahora para el mes de Junio y las compras hechas por mujeres: Benford (({Mes: Jun/2013}), (Genero: Mujer)), ValorVenta), véase la Figura 2-18.

Figura 2-18. Distribución primer dígito ValorVenta para Junio 2013 – Mujeres



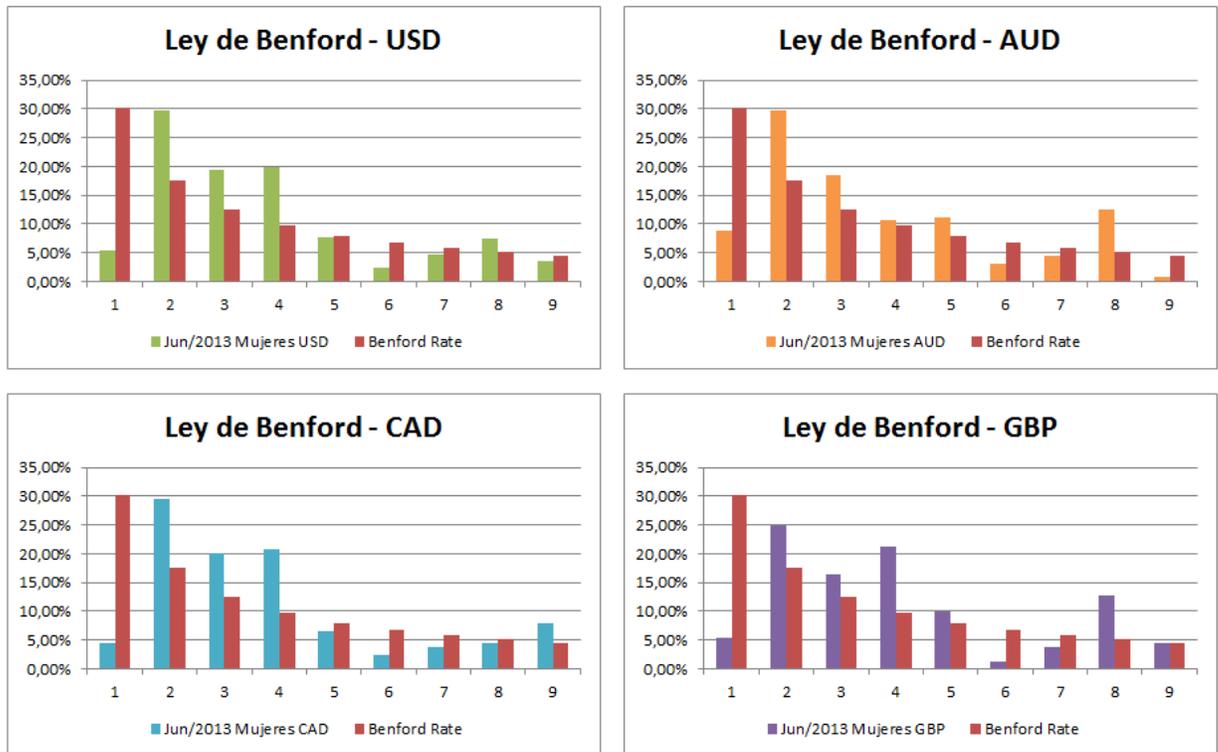
Como los resultados obtenidos en la distribución anterior, no han cambiado con respecto a las aplicaciones anteriores, i.e., sigue sin cumplirse la ley de Benford, se continuará aplicando el operador de Benford ahora para el Mes de Junio, compras realizadas por hombres y de acuerdo con la moneda con la cual pagó la compra (Dólar (USD), Dólar Australiano (AUD), Dólar Canadiense (CAD) o Libra Esterlina (GBP)), se aplica de la siguiente manera el operador Benford ((Mes: Jun/2013), (Genero: Hombre), (SiglaMoneda: USD)), ValorVenta), véase la Figura 2-19.

Figura 2-19. Distribución primer dígito ValorVenta para Junio 2013 – Hombres por Moneda.



Los resultados obtenidos nuevamente reflejan que *no se cumple* la ley de Benford, ahora se aplicará el operador para las Mujeres que hicieron compras el mes de Junio, analizándolo de acuerdo con la moneda con la cual se realizó la compra: Benford ((Mes: Jun/2013), (Genero: Mujer), (SiglaMoneda: USD)), ValorVenta), véase la Figura 2-20.

Figura 2-20. Distribución primer dígito ValorVenta para Junio 2013 – Mujeres por Moneda



Las distribuciones del mes de Junio para las mujeres con las correspondientes monedas de pago, continúa reflejando el no cumplimiento de la ley de Benford.

Finalmente, luego de las aplicaciones hechas, se concluye que los datos de la BD no cumplen la ley de Benford, lo cual es un indicio de fraudes, por lo cual se aplicarán los demás operadores definidos para continuar con el análisis.

2.4.2 SAME-SAME-SAME

Para este operador se deben definir los niveles y medidas a usar para el análisis, las cuales serán:

SSS (({Año: 2013}, {Producto: *}, {Genero: *}, {NombreMoneda, *}), {ValorVenta, PrecioUnidad, Impuesto})

En la Tabla 2-18 se presentan los resultados con el número de repeticiones para los niveles y medidas elegidos.

Tabla 2-18. Número de repeticiones para los valores de: Producto, Género, Nombre Moneda, ValorVenta, PrecioUnidad e Impuesto en 2013

Año	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
2013	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	1156
2013	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	1116
2013	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	816
2013	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	802
2013	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	802
2013	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	799
2013	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	607
2013	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	600
2013	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	598
2013	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	594
2013	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	585
2013	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	575
2013	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	574
2013	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	570
2013	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	570

De acuerdo con los niveles y medidas el número de repeticiones mayor es 1156, se analizará a nivel de Mes, conservando los demás niveles y medidas, para todos los meses del 2013:

SSS ({(Mes: Ene/2013), (Producto: *), (Genero: *), (NombreMoneda, *)}, {ValorVenta, PrecioUnidad, Impuesto}).

El mayor número de repeticiones presentadas en Enero es 37, en Febrero es 83, en Marzo es 100, en Abril es 92, en Mayo es 88, en Junio es 117, en Julio es 96, en Agosto es 117, en Septiembre es 111, en Octubre es 117, en Noviembre es 113 y en Diciembre es 123, en el Anexo A se muestran estos resultados de cada mes, en la Tabla 2-19 se muestra el resultado del mes de Enero.

Tabla 2-19. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Enero 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	37
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	35
201301	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	25
201301	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar	24
201301	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	22
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar Australiano	20
201301	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	20
201301	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	19
201301	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	18
201301	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	18
201301	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	17
201301	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Dólar	17
201301	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	17
201301	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	16
201301	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar Australiano	16
201301	32,6	32,6	2,608	HL Road Tire	Mujer	Dólar	15

Los meses que presentaron mayor número de repeticiones son Diciembre con 123 y Junio, Agosto y Octubre con 117.

Se analizan los datos de Junio 2013, diferenciándolos por el género del cliente que realizó la compra:

SSS ({(Mes: Jun/2013), (Producto: *), (Genero: Hombre), (NombreMoneda, *)}, {ValorVenta, PrecioUnidad, Impuesto}), se muestra el resultado en la Tabla 2-20.

Tabla 2-20. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 - Hombres

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	117
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	84
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	74
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	64
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	64
201306	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	64
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	62
201306	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	54
201306	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	50
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	48
201306	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar	46
201306	35	35	2,8	HL Mountain Tire	Hombre	Dólar	42
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar Canadiense	35
201306	4,99	4,99	0,3992	Touring Tire Tube	Hombre	Dólar	31

Se conserva el mayor de número de repeticiones de Junio, ahora se calcula el número de repeticiones para las mujeres:

SSS (({Mes: Jun/2013}), (Producto: *), (Genero: Mujer), (NombreMoneda, *)), {ValorVenta, PrecioUnidad, Impuesto}), se observa el resultado en la Tabla 2-21.

Tabla 2-21. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	110
201306	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	85
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	72
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	66
201306	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	57
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	52
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	51
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	50
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	50
201306	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	49
201306	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Dólar	43
201306	35	35	2,8	HL Mountain Tire	Mujer	Dólar	38
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Libra Esterlina	36

El mayor número de repeticiones es 110, se analiza tomando los datos por la moneda con la cual se hizo el pago: SSS (({Mes: Jun/2013}), (Producto: *), (Genero: Hombre), (NombreMoneda, Dólar)), {ValorVenta, PrecioUnidad, Impuesto}), los resultados para cada moneda se muestran en las Tablas: 2-22, 2-23, 2-24 y 2-25.

Tabla 2-22. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	117
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	84
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	74
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	64
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	64
201306	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	64
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	62
201306	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	54
201306	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	50
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	48
201306	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar	46
201306	35	35	2,8	HL Mountain Tire	Hombre	Dólar	42

Tabla 2-23. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar Australiano

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar Australiano	30
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar Australiano	23
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar Australiano	22
201306	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar Australiano	20
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar Australiano	20
201306	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar Australiano	15
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar Australiano	15
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar Australiano	15
201306	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar Australiano	14
201306	21,49	21,49	1,7192	LL Road Tire	Hombre	Dólar Australiano	13

Tabla 2-24. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Dólar Canadiense

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar Canadiense	35
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar Canadiense	30
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar Canadiense	29
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar Canadiense	19
201306	29,99	29,99	2,3992	ML Mountain Tire	Hombre	Dólar Canadiense	18
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar Canadiense	17
201306	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar Canadiense	15
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar Canadiense	14
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar Canadiense	13
201306	35	35	2,8	HL Mountain Tire	Hombre	Dólar Canadiense	11

Tabla 2-25. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Hombres – Libra Esterlina

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Libra Esterlina	27
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Libra Esterlina	16
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Libra Esterlina	16
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Libra Esterlina	15
201306	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Libra Esterlina	15
201306	3,99	3,99	0,3192	Road Tire Tube	Hombre	Libra Esterlina	14
201306	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Libra Esterlina	13
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Libra Esterlina	12
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Libra Esterlina	10
201306	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Libra Esterlina	10
201306	4,99	4,99	0,3992	Touring Tire Tube	Hombre	Libra Esterlina	10

Para los valores de Junio 2013 de los Hombres en las monedas de pago, en Dólar se conserva el mayor número de repeticiones, ahora se aplica este operador pero para las

Mujeres con las monedas de pago correspondientes, los resultados se muestran en las Tablas: 2-26, 2-27, 2-28 y 2-29.

Tabla 2-26. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	110
201306	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	85
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	72
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	66
201306	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	57
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	52
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	51
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	50
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	50
201306	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	49
201306	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Dólar	43

Tabla 2-27. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar Australiano

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Dólar Australiano	29
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar Australiano	28
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar Australiano	27
201306	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar Australiano	24
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar Australiano	18
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar Australiano	18
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar Australiano	16
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar Australiano	14
201306	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar Australiano	12
201306	24,99	24,99	1,9992	LL Mountain Tire	Mujer	Dólar Australiano	12
201306	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar Australiano	12

Tabla 2-28. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Dólar Canadiense

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar Canadiense	24
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar Canadiense	23
201306	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar Canadiense	22
201306	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar Canadiense	21
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar Canadiense	18
201306	35	35	2,8	HL Mountain Tire	Mujer	Dólar Canadiense	13
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar Canadiense	11
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar Canadiense	10
201306	29,99	29,99	2,3992	ML Mountain Tire	Mujer	Dólar Canadiense	9
201306	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar Canadiense	9
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar Canadiense	9

Tabla 2-29. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013 – Mujeres – Libra Esterlina

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Libra Esterlina	36
201306	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Libra Esterlina	21
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Libra Esterlina	19
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Libra Esterlina	18
201306	4,99	4,99	0,3992	Touring Tire Tube	Mujer	Libra Esterlina	16
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Libra Esterlina	15
201306	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Libra Esterlina	14
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Libra Esterlina	13
201306	3,99	3,99	0,3192	Road Tire Tube	Mujer	Libra Esterlina	12
201306	28,99	28,99	2,3192	Touring Tire	Mujer	Libra Esterlina	10
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Libra Esterlina	9

Los resultados obtenidos para las mujeres en Junio 2013 de acuerdo con sus monedas son números de repeticiones menores a los de los Hombres. Debido a que el número de repeticiones de Junio 2013 se conserva al hacer las tajadas (o sub-cubos) hasta Junio 2013 – Hombre – Dólar, lo cual es un indicio de fraude, se aplica el operador *SAME-SAME-DIFFERENT* para continuar con el análisis.

2.4.3 SAME-SAME-DIFFERENT

Para este operador se usarán los siguientes niveles y medidas:

SSD (({Año: 2013}), (Genero: *), (NombreMoneda: *)), {ValorVenta, PrecioUnidad, Impuesto}, Producto), véase Tabla 2-30.

Tabla 2-30. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en 2013

Año	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
2013	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	2349
2013	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	2349
2013	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	2349
2013	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	2318
2013	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	2318
2013	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	2318
2013	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	1729
2013	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	1729
2013	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	1729
2013	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	1663
2013	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	1663
2013	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	1663
2013	8,99	8,99	0,7192	Hombre	Dólar	AWC Logo Cap	1159
2013	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, L	1159
2013	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, M	1159
2013	8,99	8,99	0,7192	Hombre	Dólar	Road Bottle Cage	1159
2013	8,99	8,99	0,7192	Mujer	Dólar	AWC Logo Cap	1124

Los resultados al aplicar el operador para 2013, indican que el mayor número de repeticiones es 2349, los cuales pertenecen a tres productos. Se aplica el operador para cada mes del 2013 para continuar el análisis:

SSD ({{(Mes: Ene/2013)}, (Genero: *)}, (NombreMoneda: *)}, {ValorVenta, PrecioUnidad, Impuesto}, Producto)

El mayor número de repeticiones presentadas en Enero es 56, en Febrero es 167, en Marzo es 191, en Abril es 190, en Mayo es 197, en Junio es 232, en Julio es 208, en Agosto es 218, en Septiembre es 219, en Octubre es 241, en Noviembre es 231 y en Diciembre es 244; en el Anexo A se muestran los resultados de cada mes.

Los meses que presentaron mayor número de repeticiones son: Diciembre 244 y Octubre 241 y Junio 232.

Se analizan los datos del mes de Junio, ahora tomando los datos por género:

SSD ({{(Mes: Jun/2013)}, (Genero: Hombre), (NombreMoneda: *)}, {ValorVenta, PrecioUnidad, Impuesto}, Producto), véase Tabla 2-31.

Tabla 2-31. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	232
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	176
201306	8,99	8,99	0,7192	Hombre	Dólar	AWC Logo Cap	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, L	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, M	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Road Bottle Cage	125
201306	2,29	2,29	0,1832	Hombre	Dólar	Patch Kit/8 Patches	74

El mayor número de estos resultados es el mismo del mes Junio 2013, ahora se calcula para las mujeres:

SSD (({Mes: Jun/2013}), (Género: Mujer), (NombreMoneda: *)), {ValorVenta, PrecioUnidad, Impuesto}, Producto), véase Tabla 2-32.

Tabla 2-32. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	231
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	168
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	168
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	168
201306	8,99	8,99	0,7192	Mujer	Dólar	AWC Logo Cap	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Racing Socks, L	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Racing Socks, M	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Road Bottle Cage	103
201306	2,29	2,29	0,1832	Mujer	Dólar	Patch Kit/8 Patches	72
201306	21,98	21,98	1,7584	Mujer	Dólar	Fender Set - Mountain	57

El mayor número de estos resultados es 231

Para lograr un mayor nivel de detalle en el análisis se tomarán las compras realizadas por cada género (hombre, mujer) en el mes de Junio 2013 de acuerdo con la moneda con la cual se pagó:

SSD (({Mes: Jun/2013}), (Género: Hombre), (NombreMoneda: Dólar)), {ValorVenta, PrecioUnidad, Impuesto}, Producto), los resultados se muestran en las Tablas: 2-33, 2-34, 2-35 y 2-36.

Tabla 2-33. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	232
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	176
201306	8,99	8,99	0,7192	Hombre	Dólar	AWC Logo Cap	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, L	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, M	125
201306	8,99	8,99	0,7192	Hombre	Dólar	Road Bottle Cage	125
201306	2,29	2,29	0,1832	Hombre	Dólar	Patch Kit/8 Patches	74
201306	9,99	9,99	0,7992	Hombre	Dólar	Mountain Bottle Cage	62
201306	21,98	21,98	1,7584	Hombre	Dólar	Fender Set - Mountain	54

Tabla 2-34. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar Australiano

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	34,99	34,99	2,7992	Hombre	Dólar Australiano	Sport-100 Helmet, Black	57
201306	34,99	34,99	2,7992	Hombre	Dólar Australiano	Sport-100 Helmet, Blue	57
201306	34,99	34,99	2,7992	Hombre	Dólar Australiano	Sport-100 Helmet, Red	57
201306	4,99	4,99	0,3992	Hombre	Dólar Australiano	Mountain Tire Tube	49
201306	4,99	4,99	0,3992	Hombre	Dólar Australiano	Touring Tire Tube	49
201306	4,99	4,99	0,3992	Hombre	Dólar Australiano	Water Bottle - 30 oz.	49
201306	8,99	8,99	0,7192	Hombre	Dólar Australiano	AWC Logo Cap	33
201306	8,99	8,99	0,7192	Hombre	Dólar Australiano	Racing Socks, L	33
201306	8,99	8,99	0,7192	Hombre	Dólar Australiano	Racing Socks, M	33
201306	8,99	8,99	0,7192	Hombre	Dólar Australiano	Road Bottle Cage	33
201306	2,29	2,29	0,1832	Hombre	Dólar Australiano	Patch Kit/8 Patches	30
201306	53,99	53,99	4,3192	Hombre	Dólar Australiano	Short-Sleeve Classic Jersey, L	27

Tabla 2-35. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Dólar Canadiense

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Hombre	Dólar Canadiense	Mountain Tire Tube	72
201306	4,99	4,99	0,3992	Hombre	Dólar Canadiense	Touring Tire Tube	72
201306	4,99	4,99	0,3992	Hombre	Dólar Canadiense	Water Bottle - 30 oz.	72
201306	34,99	34,99	2,7992	Hombre	Dólar Canadiense	Sport-100 Helmet, Black	46
201306	34,99	34,99	2,7992	Hombre	Dólar Canadiense	Sport-100 Helmet, Blue	46
201306	34,99	34,99	2,7992	Hombre	Dólar Canadiense	Sport-100 Helmet, Red	46
201306	2,29	2,29	0,1832	Hombre	Dólar Canadiense	Patch Kit/8 Patches	29
201306	8,99	8,99	0,7192	Hombre	Dólar Canadiense	AWC Logo Cap	20
201306	8,99	8,99	0,7192	Hombre	Dólar Canadiense	Racing Socks, L	20
201306	8,99	8,99	0,7192	Hombre	Dólar Canadiense	Racing Socks, M	20
201306	8,99	8,99	0,7192	Hombre	Dólar Canadiense	Road Bottle Cage	20
201306	29,99	29,99	2,3992	Hombre	Dólar Canadiense	ML Mountain Tire	18

Tabla 2-36. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Hombres – Libra Esterlina

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Hombre	Libra Esterlina	Mountain Tire Tube	47
201306	4,99	4,99	0,3992	Hombre	Libra Esterlina	Touring Tire Tube	47
201306	4,99	4,99	0,3992	Hombre	Libra Esterlina	Water Bottle - 30 oz.	47
201306	34,99	34,99	2,7992	Hombre	Libra Esterlina	Sport-100 Helmet, Black	36
201306	34,99	34,99	2,7992	Hombre	Libra Esterlina	Sport-100 Helmet, Blue	36
201306	34,99	34,99	2,7992	Hombre	Libra Esterlina	Sport-100 Helmet, Red	36
201306	8,99	8,99	0,7192	Hombre	Libra Esterlina	AWC Logo Cap	30
201306	8,99	8,99	0,7192	Hombre	Libra Esterlina	Racing Socks, L	30
201306	8,99	8,99	0,7192	Hombre	Libra Esterlina	Road Bottle Cage	30
201306	2,29	2,29	0,1832	Hombre	Libra Esterlina	Patch Kit/8 Patches	16
201306	9,99	9,99	0,7992	Hombre	Libra Esterlina	Mountain Bottle Cage	15
201306	3,99	3,99	0,3192	Hombre	Libra Esterlina	Road Tire Tube	14
201306	49,99	49,99	3,9992	Hombre	Libra Esterlina	Long-Sleeve Logo Jersey, L	13

Los resultados muestran que en la Tabla 2-33, las compras hechas por hombres en Junio se conserva el mismo número de repeticiones, ahora se hace el mismo cálculo para la Mujer:

SSD (({Mes: Jun/2013}), (Genero: Mujer), (NombreMoneda: Dólar)), {ValorVenta, PrecioUnidad, Impuesto}, Producto), los resultados se muestran en las Tablas: 2-37, 2-38, 2-39 y 2-40.

Tabla 2-37. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	231
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	168
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	168
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	168
201306	8,99	8,99	0,7192	Mujer	Dólar	AWC Logo Cap	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Racing Socks, L	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Racing Socks, M	103
201306	8,99	8,99	0,7192	Mujer	Dólar	Road Bottle Cage	103
201306	2,29	2,29	0,1832	Mujer	Dólar	Patch Kit/8 Patches	72
201306	21,98	21,98	1,7584	Mujer	Dólar	Fender Set - Mountain	57

Tabla 2-38. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	8,99	8,99	0,7192	Mujer	Dólar Australiano	AWC Logo Cap	56
201306	8,99	8,99	0,7192	Mujer	Dólar Australiano	Racing Socks, L	56
201306	8,99	8,99	0,7192	Mujer	Dólar Australiano	Racing Socks, M	56
201306	8,99	8,99	0,7192	Mujer	Dólar Australiano	Road Bottle Cage	56
201306	34,99	34,99	2,7992	Mujer	Dólar Australiano	Sport-100 Helmet, Black	48
201306	34,99	34,99	2,7992	Mujer	Dólar Australiano	Sport-100 Helmet, Blue	48
201306	34,99	34,99	2,7992	Mujer	Dólar Australiano	Sport-100 Helmet, Red	48
201306	4,99	4,99	0,3992	Mujer	Dólar Australiano	Mountain Tire Tube	45
201306	4,99	4,99	0,3992	Mujer	Dólar Australiano	Touring Tire Tube	45
201306	4,99	4,99	0,3992	Mujer	Dólar Australiano	Water Bottle - 30 oz.	45
201306	2,29	2,29	0,1832	Mujer	Dólar Australiano	Patch Kit/8 Patches	27
201306	3,99	3,99	0,3192	Mujer	Dólar Australiano	Road Tire Tube	24

Tabla 2-39. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Mujer	Dólar Canadiense	Mountain Tire Tube	52
201306	4,99	4,99	0,3992	Mujer	Dólar Canadiense	Touring Tire Tube	52
201306	4,99	4,99	0,3992	Mujer	Dólar Canadiense	Water Bottle - 30 oz.	52
201306	34,99	34,99	2,7992	Mujer	Dólar Canadiense	Sport-100 Helmet, Black	29
201306	34,99	34,99	2,7992	Mujer	Dólar Canadiense	Sport-100 Helmet, Blue	29
201306	34,99	34,99	2,7992	Mujer	Dólar Canadiense	Sport-100 Helmet, Red	29
201306	9,99	9,99	0,7992	Mujer	Dólar Canadiense	Mountain Bottle Cage	23
201306	21,98	21,98	1,7584	Mujer	Dólar Canadiense	Fender Set - Mountain	22
201306	2,29	2,29	0,1832	Mujer	Dólar Canadiense	Patch Kit/8 Patches	18
201306	8,99	8,99	0,7192	Mujer	Dólar Canadiense	AWC Logo Cap	13
201306	35	35	2,8	Mujer	Dólar Canadiense	HL Mountain Tire	13
201306	8,99	8,99	0,7192	Mujer	Dólar Canadiense	Racing Socks, L	13

Tabla 2-40. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013 – Mujeres – Dólar

Mes	ValorVenta	PrecioUnidad	Impuesto	Género	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Mujer	Libra Esterlina	Mountain Tire Tube	57
201306	4,99	4,99	0,3992	Mujer	Libra Esterlina	Touring Tire Tube	57
201306	4,99	4,99	0,3992	Mujer	Libra Esterlina	Water Bottle - 30 oz.	57
201306	8,99	8,99	0,7192	Mujer	Libra Esterlina	AWC Logo Cap	43
201306	8,99	8,99	0,7192	Mujer	Libra Esterlina	Racing Socks, L	43
201306	8,99	8,99	0,7192	Mujer	Libra Esterlina	Road Bottle Cage	43
201306	34,99	34,99	2,7992	Mujer	Libra Esterlina	Sport-100 Helmet, Black	42
201306	34,99	34,99	2,7992	Mujer	Libra Esterlina	Sport-100 Helmet, Blue	42
201306	34,99	34,99	2,7992	Mujer	Libra Esterlina	Sport-100 Helmet, Red	42
201306	539,99	539,99	43,1992	Mujer	Libra Esterlina	Mountain-500 Black, 40	20
201306	539,99	539,99	43,1992	Mujer	Libra Esterlina	Mountain-500 Black, 42	20
201306	539,99	539,99	43,1992	Mujer	Libra Esterlina	Mountain-500 Black, 44	20

Los resultados obtenidos para las mujeres en Junio 2013 de acuerdo con sus monedas son números de repeticiones menores a las de los Hombres. Se refleja el mismo número de repeticiones en Junio 2013 y en la tajada de Junio 2013 – Hombre – Dólar, esto podría indicar un fraude, por lo cual se aplica el operador *Relative Size Factor*.

2.4.4 *Relative Size Factor*

Para aplicar este operador se usa el nivel de año y la medida ValorVenta:

RSF ({{Año: 2013}}, ValorVenta), ver Tabla 2-41.

Tabla 2-41. *Relative Size Factor* para 2013

<i>Relative Size Factor</i>	
2013	1,02486

El resultado obtenido es mayor que 1, por lo cual se aplica el operador para cada mes del año 2013: RSF ({{Mes: Ene/2013}}, ValorVenta), ver Tabla 2-42.

Tabla 2-42. *Relative Size Factor* para cada mes de 2013

<i>Relative Size Factor</i>	
Enero	1,02486
Febrero	1,02486
Marzo	1,02486
Abril	1,02486
Mayo	1,02486
Junio	1,02486
Julio	1,02486
Agosto	1,02486
Septiembre	1,02486
Octubre	1,02486
Noviembre	1,02486
Diciembre	1,02486

El valor para RSF en cada mes es el mismo: 1,02486, se aplica el operador tomando los valores por género de cliente: RSF ((Mes: Jun/2013), (Genero: Hombre)), ValorVenta), ver Tabla 2-43

Tabla 2-43. *Relative Size Factor* para Junio 2013 por género

<i>Relative Size Factor</i>	
Junio 2013 - Hombres	1,02486
Junio 2013 - Mujeres	1,02486

Para hombres y mujeres se conserva el valor de *RSF*, se aplicará el operador de acuerdo con la moneda con la que el cliente pagó:

RSF ((Mes: Jun/2013), (Genero: Hombre), (SiglaMoneda: USD)), ValorVenta), ver Tabla 2-44.

Tabla 2-44. *Relative Size Factor* para Junio 2013 – Hombres por moneda.

<i>Relative Size Factor</i>	
Junio 2013 – Hombres USD	1,02486
Junio 2013 – Hombres CAD	1,02486
Junio 2013 – Hombres AUD	1,02486
Junio 2013 – Hombres GBP	1,02486

RSF ((Mes: Jun/2013), (Genero: Mujer), (SiglaMoneda: CAD)), ValorVenta), ver Tabla 2-45.

Tabla 2-45. *Relative Size Factor* para Junio 2013 – Mujeres por moneda

<i>Relative Size Factor</i>	
Junio 2013 – Mujeres USD	1,02486
Junio 2013 – Mujeres CAD	1,02486
Junio 2013 – Mujeres AUD	1,02486
Junio 2013 – Mujeres GBP	1,02486

Se conserva el mismo valor para RSF en todas las aplicaciones realizadas.

Luego de la aplicación de los cuatro operadores la conclusión principal es que posiblemente existen anomalías en los datos (indicios de fraudes): no se cumple la ley de Benford, los operadores SSS y SSD conservan sus números de repetición luego de aplicarlos y finalmente el RSF da como resultado un número mayor a 1.

Los operadores OLAP definidos entregaron resultados que requieren de la aplicación de otras técnicas de detección de fraudes para confirmar los indicios de fraudes que se encontraron en la BD. Cabe resaltar que a través de la definición de los operadores OLAP, los escenarios planteados y los resultados del caso de estudio se logra estructurar un método escalable para la aplicación de las cuatro técnicas para detección de fraudes – ley de Benford, *Relative Size Factor*, *SAME-SAME-SAME* y *SAME-SAME-DIFFERENT* – para las BD.

El uso de los niveles de agregación en los operadores OLAP permite usar diversos enfoques de análisis de los datos almacenados, por ejemplo cuando se conocen a detalle los datos almacenados en la BD se reduce el tiempo empleado en las aplicaciones de los operadores.

Los escenarios planteados permiten identificar los posibles resultados al aplicar los operadores, los cuales complementan los resultados. Otro factor que aporta a este análisis, es el detalle que se presenta para cada aplicación de los operadores, razón por la cual se crea el anexo A, para mostrar el mayor número de ejecuciones realizadas durante la investigación que soportan los operadores OLAP propuestos.

Los resultados obtenidos para cada uno de los operadores OLAP inician el camino para continuar indagando y construyendo operadores que aprovechen las características de una BD y las diversas técnicas de detección de fraudes, además continuar la generación de conocimiento académico y aportar prototipos que se conviertan luego en herramientas para compañías de diferentes industrias.

3. Conclusiones y recomendaciones

Finalmente se obtienen de esta investigación las siguientes conclusiones y recomendaciones

3.1. Conclusiones

- Se identificaron diversas técnicas para la detección de fraudes; de cada una se presenta su utilidad en diferentes áreas: finanzas, inteligencia artificial, contabilidad forense; enumerando solamente algunas para ejemplificar la importancia de la detección de fraudes.
- Se definieron operadores OLAP para el uso de cuatro técnicas seleccionadas, explicando su significado y relación con una BD y sus elementos básicos para explorar los datos almacenados.
- Se programaron una serie de algoritmos los cuales hacen parte de un prototipo para la aplicación de los operadores definidos en una BD, las herramientas usadas son Excel y SQL Server Management, ver Anexo B.
- Los resultados obtenidos en los experimentos no son concluyentes para afirmar que existen fraudes en la BD analizada, los operadores aplicados dan indicios de posibles fraudes. Se deben complementar con otras técnicas para identificar si son verdaderamente fraudes (posiblemente se requiera de la intervención humana para ello).
- La ley de Benford, *Relative Size Factor*, *SAME-SAME-SAME* y *SAME-SAME-DIFFERENT* son técnicas útiles para iniciar el proceso de detección de *posibles* fraudes en una BD, pero deben complementarse con otras técnicas para identificar un fraude.

- Los resultados de la investigación impulsan la continuación de la misma, se requiere ampliar la base de conocimiento de las técnicas de detección de fraudes y de las características que deben continuar aprovechándose de las BD, además para complementar los resultados de este caso de estudio, realizar aplicaciones de los operadores OLAP propuestos en compañías reales.
- Esta investigación presenta la posibilidad de continuar alimentando los conocimientos construidos, aplicando otras técnicas para detección de fraudes en BD, además continuar la evolución del prototipo para enfocarlo para plataformas web y móviles.
- La metodología usada para el desarrollo de este trabajo permite identificar diversas líneas para trabajos futuros: implementaciones usando *frameworks* para aplicaciones web, construcción de operadores OLAP usando técnicas de finanzas como los modelos estadísticos o el modelo del lenguaje integrado, hasta la creación de nuevas técnicas para detección de fraudes.

3.2. Recomendaciones

- Las técnicas aplicadas al caso de estudio son herramientas útiles para un análisis inicial de posibles fraudes en los datos, estas técnicas se deben complementar con otro tipo de técnicas para verificar si hay fraudes o no.
- Se debe continuar explorando los operadores definidos y posiblemente definir otros que aprovechen la estructura de una BD con el fin de facilitar la identificación de posibles fraudes.
- La diversidad de técnicas y aplicaciones presentadas y encontradas durante la investigación muestra que la creación de técnicas transversales a las áreas del conocimiento y a las compañías es posible con los avances en la academia y en la tecnología.

A. Anexo: Resultados operadores *SAME-SAME-SAME* y *SAME-SAME-DIFFERENT* para cada mes de 2013

En este anexo se presentan los resultados para cada mes del 2013 al aplicar el operador: *SAME-SAME-SAME*:

Tabla A-1. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Enero 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	37
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	35
201301	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	25
201301	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar	24
201301	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	22
201301	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar Australiano	20
201301	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	20
201301	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	19
201301	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	18
201301	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	18
201301	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	17
201301	8,99	8,99	0,7192	Road Bottle Cage	Mujer	Dólar	17
201301	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	17
201301	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	16
201301	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar Australiano	16
201301	32,6	32,6	2,608	HL Road Tire	Mujer	Dólar	15

Tabla A-2. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Febrero 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201302	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	83
201302	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	66
201302	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	65
201302	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	63
201302	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	61
201302	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	56
201302	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	55
201302	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	45
201302	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	45
201302	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	42
201302	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	41
201302	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	41
201302	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	38
201302	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	36
201302	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	35

Tabla A-3. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Marzo 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201303	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	100
201303	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	83
201303	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	72
201303	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	63
201303	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	63
201303	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	53
201303	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	51
201303	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	50
201303	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	49
201303	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	46
201303	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	44
201303	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	44
201303	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	43
201303	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	42
201303	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	42
201303	8,99	8,99	0,7192	Road Bottle Cage	Hombre	Dólar	40

Tabla A-4. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Abril 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201304	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	92
201304	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	78
201304	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	70
201304	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	67
201304	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	67
201304	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	65
201304	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	56
201304	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	51
201304	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	50
201304	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	47
201304	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	47
201304	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	44
201304	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	42
201304	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	42
201304	4,99	4,99	0,3992	Touring Tire Tube	Mujer	Dólar	37
201304	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	36

Tabla A-5. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Mayo 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201305	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	88
201305	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	85
201305	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	82
201305	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	75
201305	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	72
201305	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	63
201305	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	62
201305	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	51
201305	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	50
201305	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	49
201305	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	49
201305	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	48
201305	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	47
201305	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	47

Tabla A-6. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Junio 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	117
201306	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	110
201306	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	85
201306	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	84
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	74
201306	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	72
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	66
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	64
201306	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	64
201306	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	64
201306	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	62
201306	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	57
201306	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	54
201306	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	52
201306	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	51

Tabla A-7. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Julio 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201307	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	96
201307	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	91
201307	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	80
201307	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	71
201307	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	71
201307	3,99	3,99	0,3192	Road Tire Tube	Mujer	Dólar	64
201307	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	63
201307	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	56
201307	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	55
201307	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	54
201307	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	54
201307	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	50
201307	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	48
201307	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	48
201307	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	48
201307	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	47

Tabla A-8. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Agosto 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201308	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	117
201308	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	101
201308	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	91
201308	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	78
201308	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	67
201308	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	64
201308	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	61
201308	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	60
201308	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	59
201308	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	57
201308	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	56
201308	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	51
201308	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	51
201308	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	49
201308	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	48

Tabla A-9. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Septiembre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201309	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	111
201309	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	109
201309	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	76
201309	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	69
201309	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	63
201309	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	61
201309	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	60
201309	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	56
201309	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	55
201309	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	55
201309	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	54
201309	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	54
201309	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	53
201309	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	53
201309	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	53

Tabla A-10. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Octubre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201310	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	117
201310	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	111
201310	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	92
201310	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	76
201310	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	75
201310	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	69
201310	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	64
201310	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	60
201310	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	60
201310	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	59
201310	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	59
201310	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	59
201310	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	57
201310	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	55

Tabla A-11. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Noviembre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201311	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	113
201311	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	110
201311	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	78
201311	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	74
201311	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	73
201311	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Mujer	Dólar	71
201311	21,98	21,98	1,7584	Fender Set - Mountain	Hombre	Dólar	71
201311	34,99	34,99	2,7992	Sport-100 Helmet, Red	Mujer	Dólar	69
201311	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	67
201311	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	67
201311	34,99	34,99	2,7992	Sport-100 Helmet, Black	Hombre	Dólar	62
201311	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	61
201311	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	60
201311	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	60
201311	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	58

Tabla A-12. Número de repeticiones para los valores de: Producto, Género, NombreMoneda, ValorVenta, PrecioUnidad e Impuesto en Diciembre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Producto	Genero	NombreMoneda	NumRepeticiones
201312	4,99	4,99	0,3992	Water Bottle - 30 oz.	Mujer	Dólar	123
201312	4,99	4,99	0,3992	Water Bottle - 30 oz.	Hombre	Dólar	116
201312	4,99	4,99	0,3992	Mountain Tire Tube	Mujer	Dólar	92
201312	2,29	2,29	0,1832	Patch Kit/8 Patches	Hombre	Dólar	84
201312	4,99	4,99	0,3992	Mountain Tire Tube	Hombre	Dólar	84
201312	2,29	2,29	0,1832	Patch Kit/8 Patches	Mujer	Dólar	72
201312	21,98	21,98	1,7584	Fender Set - Mountain	Mujer	Dólar	72
201312	8,99	8,99	0,7192	AWC Logo Cap	Mujer	Dólar	71
201312	9,99	9,99	0,7992	Mountain Bottle Cage	Mujer	Dólar	71
201312	9,99	9,99	0,7992	Mountain Bottle Cage	Hombre	Dólar	68
201312	34,99	34,99	2,7992	Sport-100 Helmet, Black	Mujer	Dólar	66
201312	8,99	8,99	0,7192	AWC Logo Cap	Hombre	Dólar	65
201312	34,99	34,99	2,7992	Sport-100 Helmet, Red	Hombre	Dólar	64
201312	34,99	34,99	2,7992	Sport-100 Helmet, Blue	Hombre	Dólar	64
201312	3,99	3,99	0,3192	Road Tire Tube	Hombre	Dólar	63

Además se presentan los resultados para cada mes del 2013 al aplicar el operador: *SAME-SAME-DIFFERENT*:

Tabla A-13. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Enero 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201301	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	56
201301	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	56
201301	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	56
201301	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	55
201301	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	55
201301	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	55
201301	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	51
201301	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	51
201301	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	51
201301	8,99	8,99	0,7192	Hombre	Dólar	AWC Logo Cap	46
201301	8,99	8,99	0,7192	Hombre	Dólar	Racing Socks, L	46

Tabla A-14. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Febrero 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201302	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	167
201302	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	167
201302	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	167
201302	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	154
201302	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	154
201302	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	154
201302	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	124
201302	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	124
201302	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	124
201302	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	109
201302	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	109

Tabla A-15. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Marzo 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201303	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	191
201303	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	191
201303	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	191
201303	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	179
201303	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	179
201303	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	179
201303	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	129
201303	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	129
201303	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	129
201303	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	127
201303	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	127

Tabla A-16. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Abril 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201304	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	190
201304	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	190
201304	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	190
201304	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	169
201304	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	169
201304	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	169
201304	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	111
201304	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	111
201304	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	111
201304	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	109
201304	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	109

Tabla A-17. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Mayo 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201305	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	197
201305	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	197
201305	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	197
201305	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	186
201305	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	186
201305	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	186
201305	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	150
201305	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	150
201305	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	150
201305	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	131
201305	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	131

Tabla A-18. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Junio 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201306	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	232
201306	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	232
201306	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	231
201306	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	231
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	176
201306	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	176
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	168
201306	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	168

Tabla A-19. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Julio 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201307	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	208
201307	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	208
201307	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	208
201307	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	202
201307	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	202
201307	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	202
201307	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	140
201307	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	140
201307	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	140
201307	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	132
201307	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	132

Tabla A-20. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Agosto 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201308	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	218
201308	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	218
201308	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	218
201308	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	209
201308	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	209
201308	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	209
201308	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	151
201308	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	151
201308	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	151
201308	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	149
201308	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	149

Tabla A-21. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Septiembre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201309	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	219
201309	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	219
201309	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	219
201309	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	201
201309	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	201
201309	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	201
201309	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	164
201309	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	164
201309	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	164
201309	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	147
201309	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	147

Tabla A-22. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Octubre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201310	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	241
201310	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	241
201310	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	241
201310	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	220
201310	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	220
201310	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	220
201310	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	174
201310	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	174
201310	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	174
201310	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	158
201310	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	158

Tabla A-23. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Noviembre 2013

Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201311	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	231
201311	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	231
201311	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	231
201311	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	228
201311	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	228
201311	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	228
201311	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	201
201311	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	201
201311	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Red	201
201311	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	189
201311	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	189

Tabla A-24. Número de repeticiones para los valores de: ValorVenta, PrecioUnidad, Impuesto, Género, NombreMoneda y Producto en Diciembre 2013

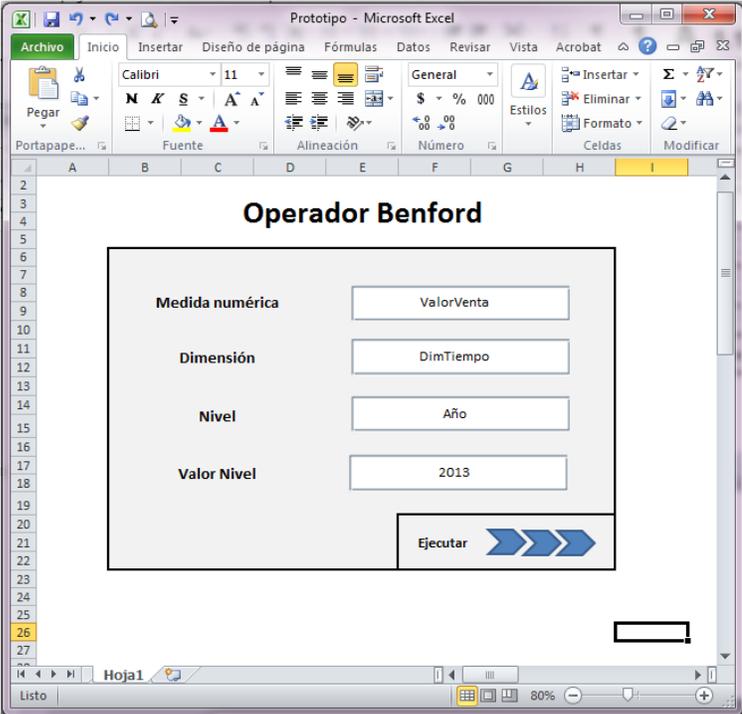
Mes	ValorVenta	PrecioUnidad	Impuesto	Genero	NombreMoneda	Producto	NumRepeticiones
201312	4,99	4,99	0,3992	Hombre	Dólar	Mountain Tire Tube	244
201312	4,99	4,99	0,3992	Hombre	Dólar	Touring Tire Tube	244
201312	4,99	4,99	0,3992	Hombre	Dólar	Water Bottle - 30 oz.	244
201312	4,99	4,99	0,3992	Mujer	Dólar	Mountain Tire Tube	243
201312	4,99	4,99	0,3992	Mujer	Dólar	Touring Tire Tube	243
201312	4,99	4,99	0,3992	Mujer	Dólar	Water Bottle - 30 oz.	243
201312	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Black	184
201312	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Blue	184
201312	34,99	34,99	2,7992	Hombre	Dólar	Sport-100 Helmet, Red	184
201312	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Black	180
201312	34,99	34,99	2,7992	Mujer	Dólar	Sport-100 Helmet, Blue	180

B. Anexo: Capturas de pantalla del prototipo

La implementación del prototipo se hizo con las herramientas Microsoft Excel y Microsoft SQL Server Management Studio. Excel se usó para la interfaz gráfica y en SQL se crearon los procedimientos almacenados con los operadores definidos en la Sección 2.2.

Para ejemplificar el uso del prototipo, se usará el operador de la ley de Benford, la Figura B-1 muestra la pantalla de ingreso de los parámetros de entrada iniciales del operador:

Figura B-1. Parámetros de entrada del operador



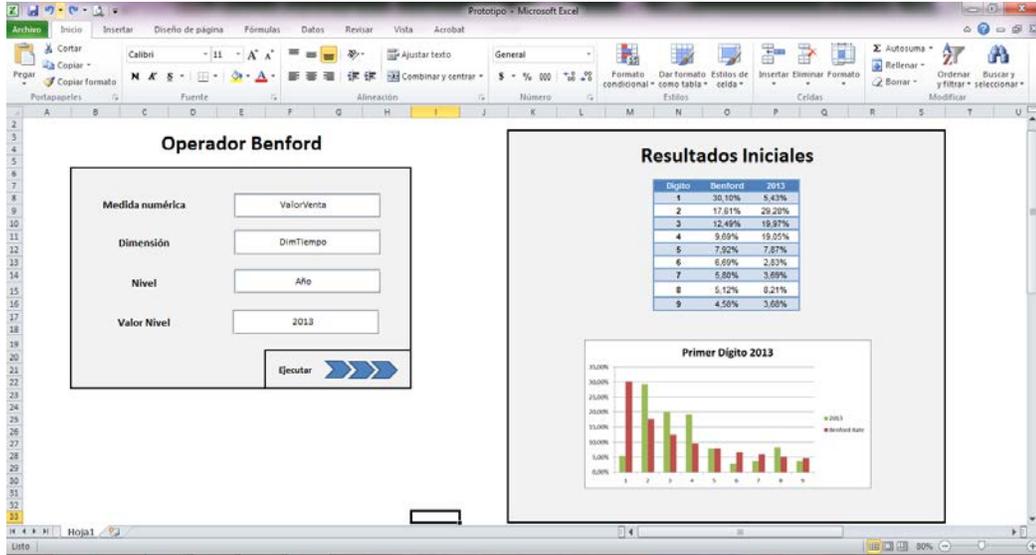
The screenshot shows a Microsoft Excel window titled "Prototipo - Microsoft Excel". The ribbon includes "Archivo", "Inicio", "Insertar", "Diseño de página", "Fórmulas", "Datos", "Revisar", "Vista", and "Acrobat". The main content area displays a form titled "Operador Benford" with the following fields and controls:

Medida numérica	<input type="text" value="ValorVenta"/>
Dimensión	<input type="text" value="DimTiempo"/>
Nivel	<input type="text" value="Año"/>
Valor Nivel	<input type="text" value="2013"/>
<input style="background-color: #4F81BD; color: white; border: none; padding: 5px 20px;" type="button" value="Ejecutar"/>	

The Excel status bar at the bottom shows "Hoja1" and "Listo".

Luego de ingresar Medida numérica, Dimensión, Nivel y Valor Nivel, se da clic en Ejecutar y el resultado se muestra en la Figura B-2:

Figura B-2. Resultados iniciales del operador con parámetros de entrada



Para presentar la lógica usada en el procedimiento almacenado para el operador Benford en sus resultados iniciales se muestra un fragmento del código en la Figura B-3:

Figura B-3. Código SQL operador Benford Inicial

```

-- Author: Daniela Vergara
-- Create date: 25/10/2012
ALTER PROCEDURE sp_OperadorBenfordInicial
-- Add the parameters for the stored procedure here
    @Medida NVARCHAR(50),
    @Dimension NVARCHAR(50),
    @Nivel NVARCHAR(50), @Valor1 NVARCHAR(50)
AS
BEGIN
    SET NOCOUNT ON;
    DECLARE @Consulta
    SET @Consulta = 'SELECT '+ @Medida+ ' AS Medida'
    'INTO Benford'
    FROM [AdventureworksDW2014].[dbo].[FactInternetSales] FIS
    LEFT JOIN [AdventureworksDW2014].[dbo].[DimDate] D ON FIS.OrderDateKey = D.DateKey
    WHERE ' + @Nivel + '=' + @Valor1

    exec sp_executesql @Consulta

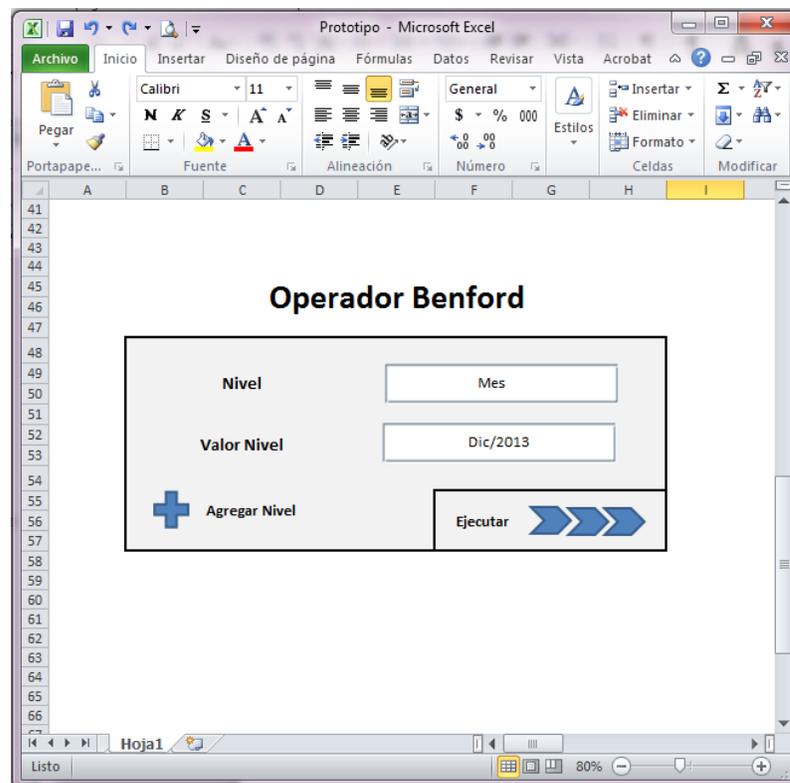
    SELECT LEFT(Medida,1) AS Dígito
    , COUNT(LEFT(Medida,1)) AS Frecuencia
    FROM Benford
    GROUP BY LEFT(Medida,1)
    ORDER BY LEFT(Medida,1)
    
```

Para los operadores implementados se crearon procedimientos almacenados en la herramienta Microsoft SQL Server Management Studio, usando el lenguaje SQL, estos procedimientos reciben los parámetros definidos en la sección 2.2, en los algoritmos se usaron comandos DDL (permiten crear y definir campos e índices) y comandos DML (permiten generar consultas para ordenar, filtrar y extraer datos).

Para cada operador se implementaron dos procedimientos: el primero recibe los parámetros de entrada iniciales y se ejecuta al dar clic en el botón Ejecutar de la Figura B-1; el segundo procedimiento permite aplicar *Slice & dice* al dar clic en el botón Ejecutar de la Figura B-4.

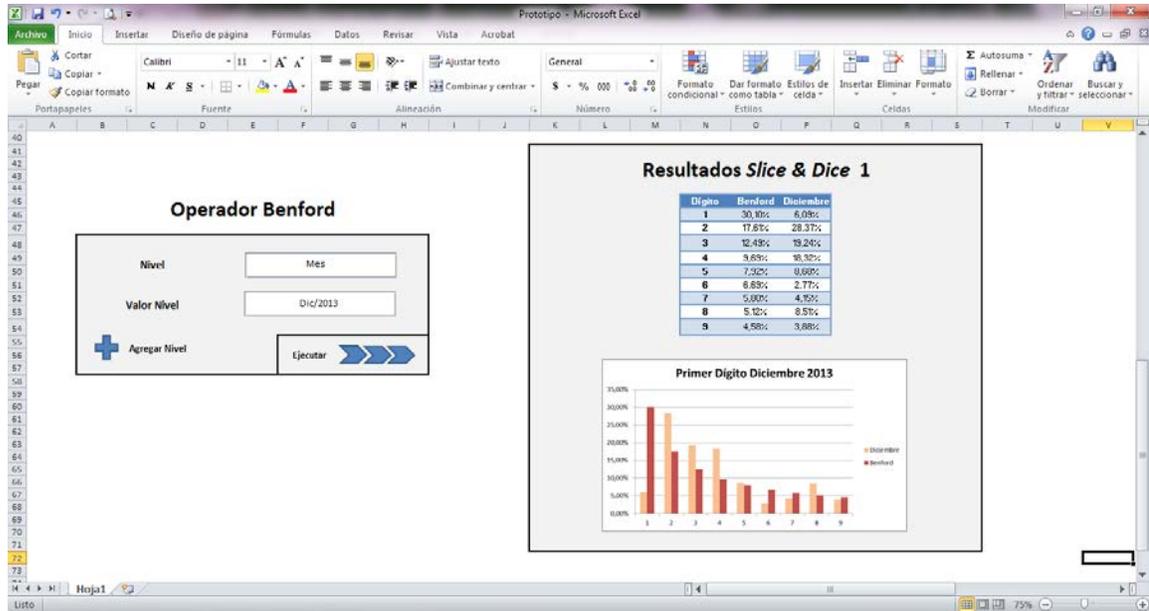
La aplicación de *slice & dice*, es decir, para analizar las tajadas (o sub-cubos) aparece la pantalla de la Figura B-4, donde se pueden agregar las parejas de Nivel y Valor del Nivel que se requieran para el análisis, el prototipo explicado en este anexo permite agregar hasta 5 niveles con sus respectivos valores.

Figura B-4. Aplicación de *Slice & dice*



Luego de ingresar los niveles y sus correspondientes valores, se da clic en Ejecutar y se obtiene el resultado del operador como se muestra en la Figura B-5.

Figura B-5. Resultados de la aplicación de *slice & dice*



Este prototipo es la parte inicial de una aplicación híbrida que permita la ejecución de los operadores OLAP definidos en esta investigación y de otros operadores que se construirán en un futuro para las BD, se plantea continuar con esta implementación para aplicarla a nivel empresarial, en primera instancia en una compañía de Contact Center.

Bibliografía

- ACFE. (2016). *Report to the Nations on Occupational Fraud and Abuse: 2016 Global Fraud Study*. Retrieved from <https://s3-us-west-2.amazonaws.com/acfe-public/2016-report-to-the-nations.pdf>
- Ahmed, M., & Mahmood, A. N. (2013). A Novel Approach for Outlier Detection and Clustering Improvement. *IEEE*, 577–582.
- Ausloos, M., Herteliu, C., & Ileanu, B. (2014). Breakdown of Benford's Law for Birth Data. Retrieved from <http://arxiv.org/abs/1410.1755>
- Banco Mundial. (2014). Sector financiero | Data. Retrieved March 25, 2016, from <http://datos.bancomundial.org/tema/sector-financiero>
- Benford, F. (1938). The Law of Anomalous Numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572. <http://doi.org/10.2307/984802>
- Bredl, S., Winker, P., & Kötschau, K. (2008). A statistical approach to detect cheating interviewers. *Justus Liebig University Giessen, Center for International Development and Environmental Research (ZEU)*, (39). Retrieved from <http://econpapers.repec.org/RePEc:zbw:zeudps:39>
- Breunig, C., & Goerres, A. (2011). Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany. *Electoral Studies*, 30(3), 534–545. <http://doi.org/10.1016/j.electstud.2011.03.005>
- Chan, P. K., & Stolfo, S. J. (1998). Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. *American Association for Artificial Intelligence*, 164–168.
- de Pablos Heredero, C., López-Hermoso, J. J., Martín-Romo, S., & Medina, S. (2004). *Informática y comunicaciones en la empresa*. ESIC Editorial.
- Diekmann, A. (2004). Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Swiss Federal Institute of Technology Zurich*, (October).
- Dong, W., Liao, S. S., Fang, B., Cheng, X., Chen, Z., & Fan, W. (2014). The Detection of Fraudulent Financial statement: An Integrated Language Model. *Association for*

- Information Systems*. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1004&context=pacis2014&sei-redirect=1&referer=http://scholar.google.com/scholar?start=10&q=fraud+detection+in+financiar+business&hl=es&as_sdt=0,5&as_ylo=2011&as_vis%253
- Durtschi, C., Hillison, W., & Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensics Accounting*, 99(99), 17–34.
- Fernandez-Portillo, A., Sánchez-Escobedo, M. C., Hernández-Mogollón, R., & Jiménez-Naranjo, H. V. (2015, September 5). La importancia de la innovación en el comercio electrónico. *Universia Business Review*. Retrieved from <https://ubr.universia.net/article/view/1528/importancia-la-innovacion-el-comercio-electronico>
- Henselmann, K., Scherr, E., & Ditter, D. (2013). Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SEC XBRL filings. *The Open Access Publication Server of the ZBW - Econstor*. Retrieved from <http://www.econstor.eu/handle/10419/88418>
- Hernández G, H. I. (2009). *Ley de Benford, Herramienta para "Detectar Fraude" en la auditoría de sistemas de información: Un caso práctico, en una empresa en la ciudad de Medellín*. EAFIT.
- Hilas, C. S., Mastorocostas, P. A., & Rekanos, I. T. (2015). Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A case Study. *Applied Mathematics & Information Sciences*, 9(4), 1709.
- Hu, N., Liu, L., & Sambamurthy, V. (2011). Fraud detection in online consumer reviews. *Decision Support Systems*, 50(3), 614–626. <http://doi.org/10.1016/j.dss.2010.08.012>
- Information Resources Management Association. (2012). *Data Mining: Concepts, Methodologies, Tools, and Applications*. IGI Global. <http://doi.org/10.4018/978-1-4666-2455-9>
- Inmon, W. H. (2005). *Building the data warehouse*. John Wiley & sons.
- Jacob, K. M., Sangeetha, K., & Karthik, S. (2014). Outlier Mining for Removing the Anomalies in High Dimensional Data Using ARVDH Algorithm. *International Journal of Computer Science and Mobile Computing*, 3(2), 552–557.
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.

- Kouznetsov, A., & Tsvetovat, M. (2011). *Social Network Analysis for Startups*. Retrieved from <https://www.safaribooksonline.com/library/view/social-network-analysis/9781449311377/ch04.html>
- Krakar, Z., & Žgela, M. (2009). Application of Benford's Law in Payment Systems Auditing. *Journal of Information and Organizational Sciences*, 33(1), 39–51.
- Kraus, C., & Valverde, R. (2014). A Data Warehouse design for the detection of fraud in the supply chain by using the Benford's Law. *American Journal of Applied Sciences*, 11(9), 1507–1518. <http://doi.org/10.3844/ajassp.2014.1507.1518>
- Kreuzer, M. Ms., Jordan, D. P., Antkowiak, B. P., Drexler, B. M., Kochs, E. F. M., & Schneider, G. M. (2014). Brain Electrical Activity Obeys Benford's Law. *Anesthesia & Analgesia*, 118(1), 183–191.
- Kroll. (2015). *Informe global sobre Fraude 2015/2016*.
- Lu, C.-T., Sirwongwattana, S., Huang, Y.-P., & Kou, Y. (2004). Survey of fraud detection techniques. In *IEEE International Conference on Networking, Sensing and Control, 2004* (Vol. 2, pp. 749–754). IEEE. <http://doi.org/10.1109/ICNSC.2004.1297040>
- Malinowski, E., & Zimányi, E. (2008). *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*. Springer eBooks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Retrieved from <http://ezproxy.unal.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat02704a&AN=unc.000787461&lang=es&site=eds-live>
- Microsoft Corp. (2013). Microsoft SQL Server Product Samples. Retrieved February 5, 2016, from <http://msftdbprodsamples.codeplex.com/>
- Mir, T. A., Ausloos, M., & Cerqueti, R. (2014). Benford's law predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. Retrieved from <http://arxiv.org/abs/1410.2890>
- Nigrini, M. (1999). I've Got Your Number. *Journal of Accountancy*. Retrieved from <http://www.journalofaccountancy.com/issues/1999/may/nigrini>
- Nigrini, M. (2011). *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. (John Wiley & Sons, Ed.) (1st ed.). Retrieved from <http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470890460.html>
- Nigrini, M., & Wells, J. T. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. (John Wiley & Sons, Ed.) (1st ed.). Retrieved from <http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118152859.html>
- Olaya, V. (2014). Sistemas de Información Geográfica. Retrieved April 30, 2015, from

- http://volaya.github.io/libro-sig/chapters/Estadistica_avanzada.html
- RAE. (2014). Diccionario Real Academia Española. Retrieved May 15, 2016, from <http://www.rae.es/>
- Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications* (Vol. 81). WORLD SCIENTIFIC. <http://doi.org/10.1142/9097>
- Sabau, A. S. (2012). Survey of clustering based financial fraud detection research. *Informatica Economica*, 16(1), 110.
- Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, 40(15), 5916–5923. <http://doi.org/10.1016/j.eswa.2013.05.021>
- Sambridge, M., Tkalčić, H., & Jackson, a. (2010). Benford's law in the natural sciences. *Geophysical Research Letters*, 37(22), n/a-n/a. <http://doi.org/10.1029/2010GL044830>
- SAS Institute, & Federal Data Corporation. (1999). Using Data Mining Techniques for Fraud Detection.
- Saville, A. (2006). Using Benford's Law to detect data error and fraud: An Examination of companies listed on the Johannesburg Stock Exchange. *South African Journal of Economic and Management Sciences*, 9(3). Retrieved from <http://hdl.handle.net/2263/3283>
- Singh, J., & Aggarwal, S. (2013). Survey on Outlier Detection in Data Mining. *International Journal of Computer Applications*, 67(19), 29–32.
- Stolfo, S. J., Fan, D. W., Prodromidis, A. L., Chan, P. K., & Lee, W. (1997). Credit Card Fraud Detection Using Meta-Learning : *AAAI Technical Report*, 83–90.
- Tackett, J. A. (2007). Digital analysis: A better way to detect fraud. *Journal of Corporate Accounting & Finance*, 18(4), 27–36. <http://doi.org/10.1002/jcaf.20305>
- Verleysen, M. (2003). Learning high-dimensional data. In *Limitations and Future Trends in Neural Computation* (pp. 141–162). Retrieved from <http://books.google.com/books?hl=es&lr=&id=WfFW8-jMuAcC&pgis=1>
- Wadhwa, L., & Pal, V. (2012). Forensic Accounting And Fraud Examination In India. *International Journal of Applied Engineering Research*, 7(11).
- Žgela, M., & Dobša, J. (2011). Analysis of Top 500 Central and East European Companies Net Income Using Benford's Law. *Journal of Information and Organizational Sciences*, 35(2), 215–228.