

Utilizar el análisis de datos para detectar el fraude con tarjetas de crédito

Las empresas emplean hoy en día técnicas analíticas para la detección precoz de los fraudes con tarjetas de crédito, un factor clave para mitigar los daños causados por el fraude. El tipo más común de fraude con tarjeta de crédito no implica el robo físico de la tarjeta, sino el de las credenciales de la tarjeta de crédito, que luego se utilizan para realizar compras en línea.

Imagine que le han contratado como analista de datos para trabajar en la división de tarjetas de crédito de un banco. Y su primera misión consiste en unirse a su equipo en el uso del análisis de datos para la detección precoz y la mitigación del fraude con tarjetas de crédito.

Para poder prescribir un camino a seguir, es decir, sugerir lo que debería hacerse para que el fraude se detecte a tiempo, necesita comprender qué aspecto tiene una transacción fraudulenta. Y para ello hay que empezar por examinar los datos históricos.

He aquí un conjunto de datos de muestra que recoge los detalles de las transacciones con tarjeta de crédito de unos cuantos usuarios.

IP Address	User ID	Account Number	Age	Shipping Address	Transaction Date	Transaction Time	Transaction Value	Product Category	Units Purchased
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	15-5-20	15:00:05	\$121.58	Clothing	1
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	10-6-20	10:23:10	\$79.23	Electronics	2
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	1-6-20	07:12:45		Home Décor	1
1.186.52.7	johnp	25671147	32	In-store	3-6-20	01:11:10	\$2,009.99	Electronics	10
	johnp	25671147	32	In-store	2020-06-03	01:15:12	\$4,131.00	Electronics	15
1.186.52.7	johnp	25671147	32	P.O. Box 1049	03-06-2020	01:22:24	\$3,010.50	Tools	20
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	15 May 2020	17:02:08	\$234.20	Furniture	1
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	18 May 2020	19:12:45	\$141.00	Kithcen Supplies	3
	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	01 June 2020	17:34:15	\$157.25	Car Spares	2
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	13 June 2020	18:02:10	\$59.99	Kithcen Supplies	1
172.165.10.1	ellend	11568528		P.O. Box 1322	07 June 2020	15:53:12	\$99.99	Clothing	1
172.165.10.1	ellend	11568528		P.O. Box 1322	08 June 2020	17:15:30	\$53.15	Beauty	1
1.167.255.10	ellend	11568528		P.O. Box 5401	02 July 2020	00:05:10	\$4,895.00	Laptop	1

Las técnicas descriptivas de análisis, es decir, las técnicas que le ayudan a comprender lo sucedido, incluyen la identificación de patrones y anomalías en los datos. Las anomalías significan una variación en un patrón que parece poco característico, o, fuera de lo normal. Las anomalías pueden producirse por razones perfectamente válidas y genuinas, pero justifican una evaluación porque pueden ser un signo de actividad fraudulenta.

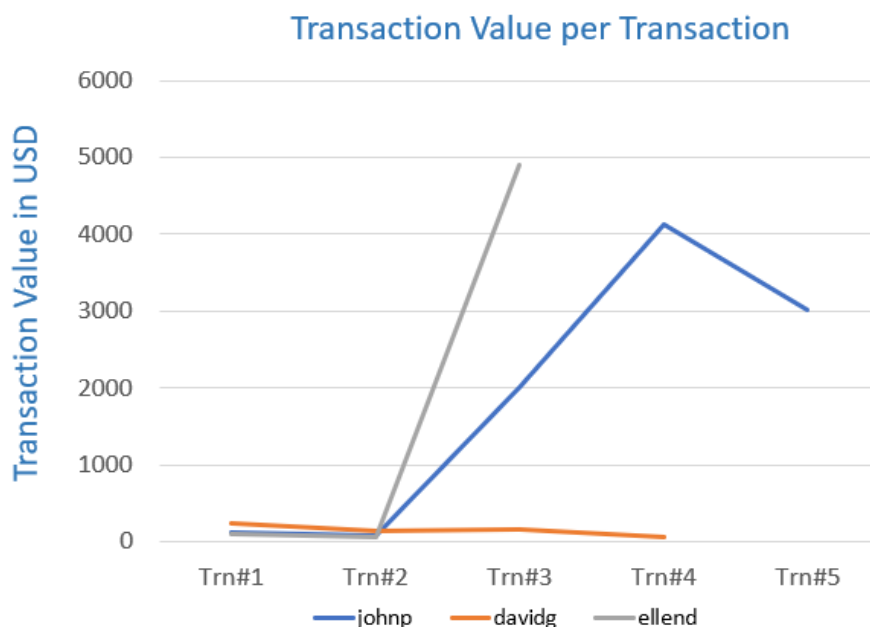
Estudios anteriores han sugerido que algunos de los sucesos comunes a los que debe estar atento incluyen:

- Un cambio en la frecuencia de los pedidos realizados; por ejemplo, un cliente que suele realizar un par de pedidos al mes de repente realiza numerosas transacciones en un breve espacio de tiempo, a veces en cuestión de minutos desde el pedido anterior.
- Pedidos significativamente superiores a la transacción media de un usuario.
- Pedidos masivos del mismo artículo con ligeras variaciones como el color o el tamaño, especialmente si esto es atípico en el historial de transacciones del usuario.
- Un cambio repentino en la preferencia de entrega, por ejemplo, un cambio de dirección de entrega en casa o en la oficina a entrega en tienda, almacén o apartado de correos.
- Una dirección IP no coincidente, o una dirección IP que no es de la localidad o zona general de la dirección de facturación.

Antes de poder analizar los datos en busca de patrones y anomalías, es necesario:

- **Identificar y reunir todos los puntos de datos que puedan ser relevantes para su caso de uso.** Por ejemplo, los detalles del titular de la tarjeta, los detalles de la transacción, los detalles de la entrega, la ubicación y la red son algunos de los puntos de datos que podrían explorarse.
- **Limpie los datos.** Tiene que identificar y solucionar los problemas en los datos que puedan dar lugar a conclusiones falsas o incompletas, como los valores de datos que faltan y los datos incorrectos. Puede que también necesite normalizar los formatos de los datos en algunos casos, por ejemplo, los campos de fecha.

Por último, cuando llegue a las conclusiones, creará visualizaciones apropiadas que comuniquen sus conclusiones a su audiencia. El gráfico siguiente muestra una de esas visualizaciones que usted utilizaría para captar una tendencia oculta en el conjunto de datos de muestra compartido anteriormente en el estudio de caso.



En la siguiente sección se le pedirá que responda a las siguientes 5 (cinco) preguntas basadas en este estudio de caso:

1. Enumere al menos 5 (cinco) puntos de datos necesarios para el análisis y la detección de un fraude con tarjeta de crédito. (3 puntos)
 - Desde mi punto de vista, los datos esenciales a considerar dentro de un movimiento de fraude es la **IP del cliente**, la **dirección de entrega**, **fecha y hora de la transacción**, **valor o monto total** de los productos adquiridos y por último con relación a los datos, se debe considerar el **ID o nombre del cliente**. Es lo mínimo para evaluar la razón o sospecha de fraude, que en base al historial de los movimientos del cliente se puede generar un buen análisis de fraude.
2. Identifique 3 (tres) errores/cuestiones que podrían afectar a la exactitud de sus conclusiones, basándose en una tabla de datos proporcionada. (3 puntos)
 - Los tres factores que puede afectar la exactitud de mis conclusiones es **no tener los registros completos**, **los formatos sean diferentes** en la columna donde se almacena el dato. por último, que **los datos no vengan de fuentes confiables**. Por todo lo anterior, considero que se debe brindar un proceso de limpieza de datos y comunicarse con los encargados de área para evaluar la confiabilidad de los registros.
3. Identifique 2 (dos) anomalías, o comportamientos inesperados, que le hagan pensar que la transacción puede ser sospechosa, basándose en una tabla de datos proporcionada. (2 puntos)
 - Las dos anomalías encontradas, es en particular el caso del usuario **johnp** que tiene dos compras de montos altos en lugares no registrados (IP Address) en compras anteriores con montos altos, en la misma fecha y diferente ubicación de entrega. Además, el caso de **ellend**, que mantiene registros de compras no mayores a \$100, y multiplico el valor de su compra por mucho más de lo esperado, y en diferente ubicación según los registros de la tabla.
4. Explique brevemente su conclusión clave a partir del cuadro de visualización de datos proporcionado. (1 punto)
 - En conclusión, se puede reconocer y evitar el fraude financiero, en base a registros anteriores. Por ello, se necesita buscar datos relevantes para un buen informe y control de actividades sospechosas, tal como lo muestra la gráfica que apoyan mi análisis y conclusión.

5. Identifique el tipo de análisis que está realizando cuando analiza datos históricos de tarjetas de crédito para comprender cómo es una transacción fraudulenta [Pista: Los cuatro tipos de análisis son: Descriptivo, Diagnóstico, Predictivo, Prescriptivo] (1 punto)
- El análisis diagnóstico es la técnica utilizada en este ejemplo, debido a que se desea comprender un evento, lo que nos lleva a investigar la causa o raíz del problema gracias a los datos históricos.