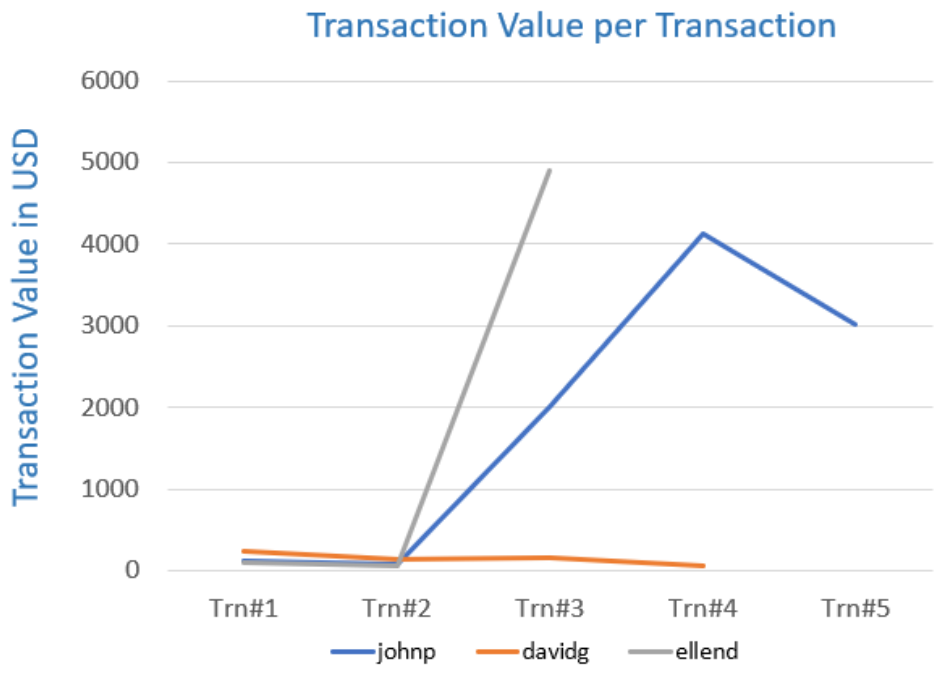
Using data analysis for the early detection and mitigation of credit card fraud

Sample data set that captures the credit card transaction details for a few users

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

The graph below samples one such visualization that you would use to capture a trend hidden in the sample data set shared earlier on in the case study



In the next section you will be asked to answer the following 5 (five) questions based on this case study:

List at least 5 (five) data points that are required for the analysis and detection of a credit card fraud. (3 marks)

- IP address
- User ID
- Account Number
- Shipping Address

- Transaction Date
- Transaction Time
- Transaction Value
- Units Purchased

Identify 3 (three) errors/issues that could impact the accuracy of your findings, based on a data table provided. (3 marks)

- Blank IP address
- Blank Transaction Value
- Transaction Date (Format)
- Blank Age

Identify 2 (two) anomalies, or unexpected behaviors, that would lead you to believe the transaction may be suspect, based on a data table provided. (2 marks)

- Johnp (User ID)
    - Fraud detection (Yes)
        - Different IP address (1.186.52.7) purchased 2k+ value and 10+ units
        - Different Shipping Address
        - All of the purchases were made on the same day and around the same time
- Davidg (User ID)
    - Fraud detection (No)
- Ellend (User ID)
    - Fraud detection (Yes)
        - 3rd purchase with different IP address, shipping address and 4k+ value

Briefly explain your key take-away from the provided data visualization chart. (1 mark)

- User Davidg looks like a normal banking account with everything being linear without any suspicious/fraud
- Users Johnp and Ellend has a sudden spike from transaction 2 which shows signs of suspicious/fraud

Identify the type of analysis that you are performing when you are analyzing historical credit card data to understand what a fraudulent transaction looks like. [Hint: The four types of Analytics include: Descriptive, Diagnostic, Predictive, Prescriptive] (1 mark)

- Descriptive
- Why ?
    - Techniques that help you gain an understanding of what happened, include the identification of patterns and anomalies in data
    - Anomalies signify a variation in a pattern that seems uncharacteristic, or, out of the ordinary
    - Anomalies may occur for perfectly valid and genuine reasons, but they do warrant an evaluation because they can be a sign of fraudulent activity