Data Analysis in Action

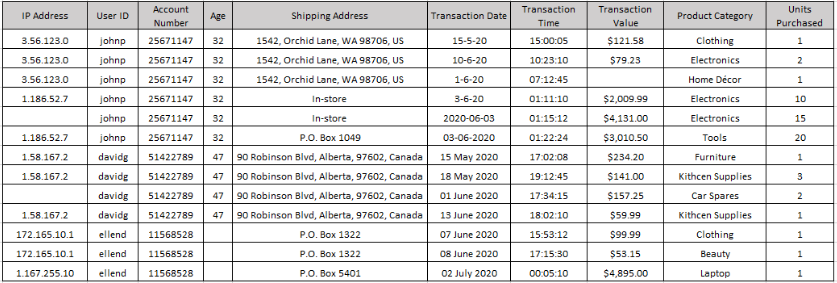
**Using Data Analysis for Detecting Credit Card Fraud**

Companies today are employing analytical techniques for the early detection of credit card frauds, a key factor in mitigating fraud damage. The most common type of credit card fraud does not involve the physical stealing of the card, but that of credit card credentials, which are then used for online purchases.

Imagine that you have been hired as a Data Analyst to work in the Credit Card Division of a bank. And your first assignment is to join your team in using data analysis for the early detection and mitigation of credit card fraud.

In order to prescribe a way forward, that is, suggest what should be done in order for fraud to get detected early on, you need to understand what a fraudulent transaction looks like. And for that you need to start by looking at historical data.

Here is a sample data set that captures the credit card transaction details for a few users.



Descriptive techniques of analysis, that is, 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.

Past studies have suggested that some of the common events that you may need to watch out for include:

* A change in frequency of orders placed, for example, a customer who typically places a couple of orders a month, suddenly makes numerous transactions within a short span of time, sometimes within minutes of the previous order.
* Orders that are significantly higher than a user’s average transaction.
* Bulk orders of the same item with slight variations such as color or size—especially if this is atypical of the user’s transaction history.
* A sudden change in delivery preference, for example, a change from home or office delivery address to in-store, warehouse, or PO Box delivery.
* A mismatched IP Address, or an IP Address that is not from the general location or area of the billing address.

Before you can analyze the data for patterns and anomalies, you need to:

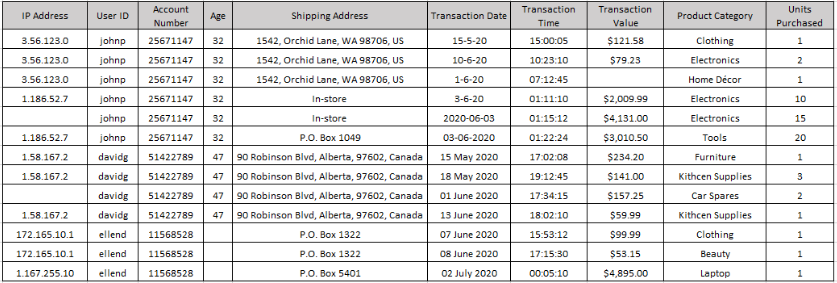
* **Identify and gather all data points that can be of relevance to your use case.** For example, the card holder’s details, transaction details, delivery details, location, and network are some of the data points that could be explored.
* **Clean the data.** You need to identify and fix issues in the data that can lead to false or incomplete findings, such as missing data values and incorrect data. You may also need to standardize data formats in some cases, for example, the date fields.

Finally, when you arrive at the findings, you will create appropriate visualizations that communicate your findings to your audience. 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.

**Graded Final Assignment**

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

* Card holder / Customer ID, Transaction date, Transaction time, Transaction value, Shipping address, IP address, Device model, Location

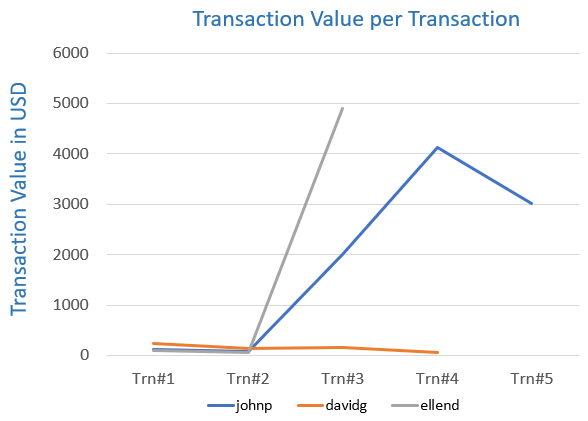
1. Identify 3 (three) errors/issues that could impact the accuracy of your findings, based on a data table provided. (3 marks)

* Missing transaction value
* Missing IP Address
* Date format inconsistency

1. 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)

* Significantly higher Transaction Value where Shipping Address has been changed from home/office address to P.O. Box.
* Higher Transaction Value and increased frequency of transactions.
* IP Address change and significantly higher Transaction Value.
* IP Address change and Shipping Address change

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



* The visualization depicts the transaction values per transaction for all three users. The key take-away from this visualization is the sharp rise in the transaction values for users johnp and ellend, which may be indicative of an anomaly.

1. 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 Analytics' as the type of analysis that they are performing when they are analyzing historical credit card data to understand what a fraudulent transaction looks like.