This is the final project I worked on to complete Coursera's "Introduction to data analytics" online course. The case study used is the problem of credit card fraud. Therefore, we have to answer some open questions based on our particular case scenario.

Case Study:

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.

Here is a 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

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.

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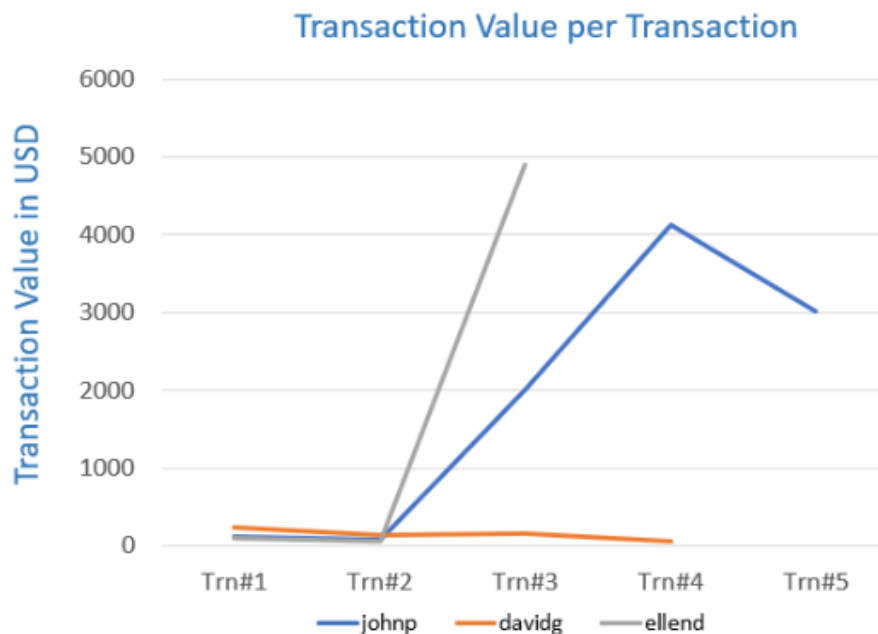
In the next section you will be asked to answer the following 5 (five) questions based on this case study:

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

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

3. 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)

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

5. 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)

Analysis Findings and Results:

1. List at least 5 (five) data points that are required for the analysis and detection of credit card fraud. (3 marks)
   - IP address
   - Username and account number
   - Shipping address
   - Transaction date and time
   - Transaction value and units

2. Identify 3 (three) errors/issues that could impact the accuracy of your findings, based on a data table provided. (3 marks)
   - Missing IP address in column 1, row number 5 and 9
   - Empty Transaction value in transaction value column, row number 3
   - Inconsistent date format in Transaction date column.

3. 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)
   - IP address changed for user Johnp and transaction value and unit number were significantly higher.
   - IP address and shipping address changer for user Johnp and transaction value and unit number were much higher than usual.
   - IP address and shipping address have been changed for user ellend and transaction value was much bigger than usual transaction.

4. Briefly explain your key take-away from the provided data visualization chart. (1 mark)
   - The visualization shows the transaction value per transaction for all three users. An important takeaway from this visualization is the sharp increase in transaction values for users johnp and ellend. This may indicate an anomaly.

6. 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)
   - I used descriptive analysis to explain what happened in this case study. There are two user accounts that show credit card fraud. This is evidenced by the very high transaction value of the average user's credit card usage.