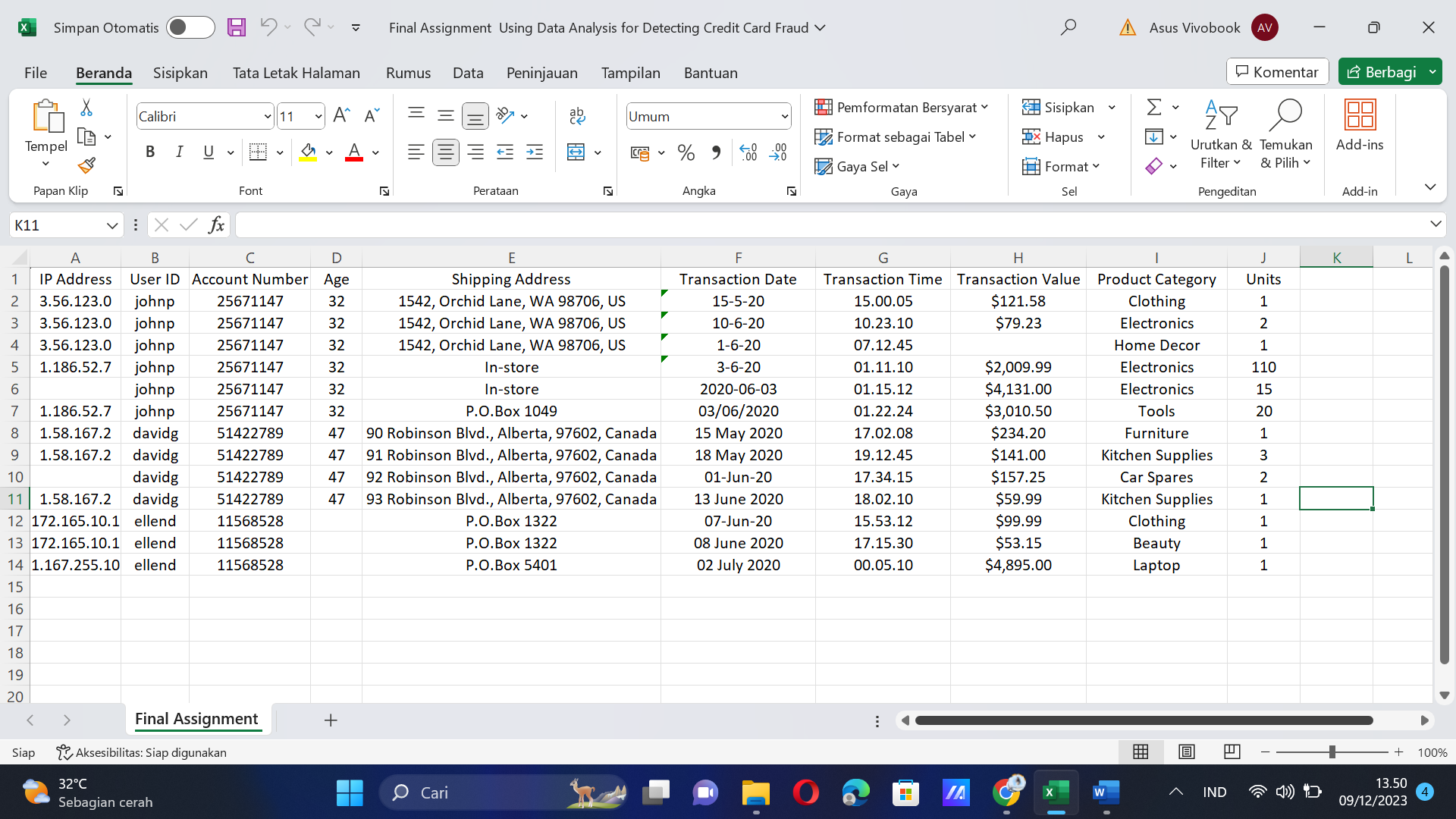
Final Assignment : Introduction to Data Analytics

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.

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

1. 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.
2. Orders that are significantly higher than a user’s average transaction.
3. 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.
4. A sudden change in delivery preference, for example, a change from home or office delivery address to in-store, warehouse, or PO Box delivery.
5. 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.**

**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 a credit card fraud.

1. Card Holder / Customer ID:

* This identifies the unique customer associated with the transaction, enabling comparison with past behavior and account details.
* It helps identify potential unauthorized access or use of the card.

2. Transaction Date & Time:

* Provides context for the transaction by pinpointing when it occurred.
* Helps detect suspicious activity outside of the cardholder's typical spending patterns.
* Enables identification of unusual timing patterns, such as late-night or weekend transactions.

3. Transaction Value:

* The amount spent in the transaction allows comparison with historical spending habits.
* Unusual spending spikes, particularly for large amounts, can be red flags for potential fraud.
* Helps identify transactions inconsistent with the cardholder's usual spending patterns.

4. Shipping Address:

* The address where the purchased items are delivered offers valuable information.
* Inconsistencies with the cardholder's registered address or unusual shipping destinations can be suspicious.
* Provides clues about the legitimacy of the transaction and the recipient of the goods.

5. IP Address:

* Identifies the internet connection used for the transaction, providing location information.
* Allows detection of transactions originating from unexpected locations, potentially indicating card theft or unauthorized use.
* Helps identify potential cases of remote access fraud.

6. Device model

* Information about the device used for the transaction can help identify suspicious activity. For example, a transaction from an unknown device type might be a sign of unauthorized access.

7. Location

* Knowing the approximate location where the transaction occurred can be valuable in detecting fraud. Inconsistencies with the cardholder's location or unusual activity in a particular region can be red flags.

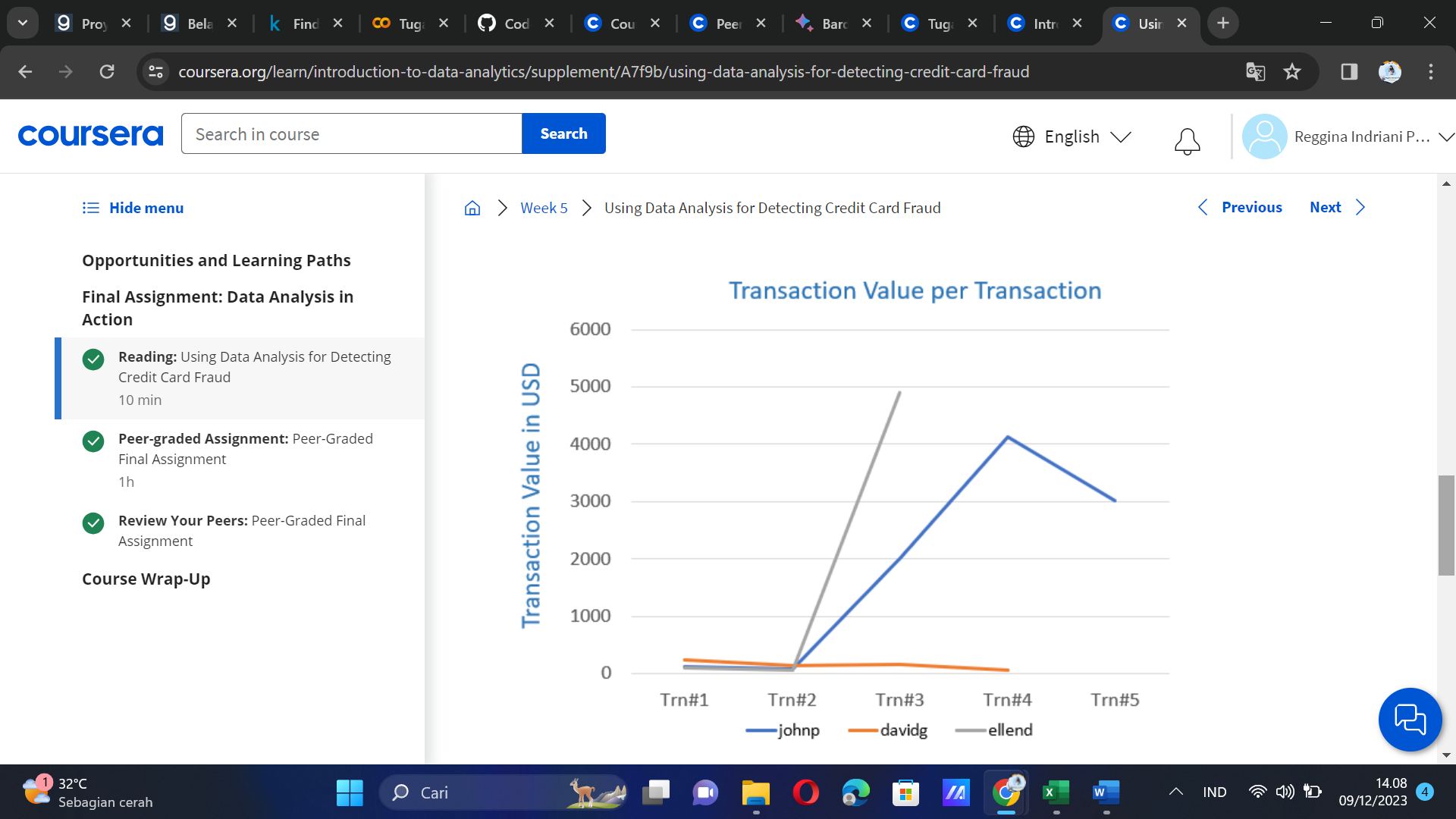
1. Identify 3 (three) errors/issues that could impact the accuracy of your findings, based on a data table provided.

* 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.

* 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.



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]

Descriptive Analytics because analyzing historical credit card data to understand what a fraudulent transaction looks like.