### **Worksheet: Quality Assurance for Fraud Detection Features**

#### **Objective:**

In this assignment, you will apply quality assurance (QA) techniques to the enriched credit card transaction data. Specifically, you will focus on the columns transactions\_last\_24h and transactions\_last\_2weeks. You will manually test these columns, write Python Behave test cases, and answer high-level QA questions.

#### **Features Explanation:**

* **transactions\_last\_24h**: This column represents the number of transactions made using a specific credit card in the last 24 hours from the current transaction's timestamp. It helps identify unusual spending patterns that may indicate fraudulent activity.
* **transactions\_last\_2weeks**: This column represents the number of transactions made using a specific credit card in the last two weeks from the current transaction's timestamp. It provides a broader view of the card's usage pattern, which is useful for detecting longer-term trends that might indicate fraud.

**Data Directory:** [CISC367 - Worksheet 2](https://drive.google.com/drive/u/0/folders/1-4nwvG4KXudA7_WnfgVdnAJ1Zs2txozD)

#### **Instructions:**

1. **High-Level QA Questions**:
   * Reflect on the importance of documentation and communication in the QA process. How would you document your QA findings and communicate them to stakeholders to ensure effective resolution of data quality issues? What are key aspects for you to include in your testing and why?

Answer Here

* + Consider you have 1000 features. How would you organize your testing efforts? Elaborate on feature similarities and differences, nomenclature, etc.   
      
    Answer Here
  + Name a few common biases that you may be exposed to while testing.

Answer Here

* + What impact may these biases have?

Answer Here

* + What is the difference between blackbox testing and whitebox testing?

Answer Here

1. **Manual Testing**:
   * Conduct a manual test of the transactions\_last\_24h and transactions\_last\_2weeks columns.
   * What are edge cases and anomalies you would look for in these features?

Answer Here

1. **Automated Testing**:
   * Conduct an automated test of the transactions\_last\_24h and transactions\_last\_2weeks columns using behave. You must use the same Given and When statement from class, but create a new then statement.  
       
     Paste your then statement code here
   * Explain: is this black box or white box testing and why?

Answer Here

1. **Feature Calculation and Fault Analysis**:
   * Analyze the provided example data, identifying any faulty values and explaining why it is incorrect. How are the calculations currently being calculated?  
     Answer Here
   * Discuss all aspects in which the current calculations are wrong and propose corrections.  
       
     Answer Here
2. **Potential Consequences of QA Failures**:

* What could go wrong if QA processes fail for the transactions\_last\_24h and transactions\_last\_2weeks features?  
  Answer Here
* How might inaccurate data/poor QA affect the performance of fraud detection algorithms?  
  Answer Here
* What are possible risks towards the validity of the outcome of the fraud detection model?  
  Answer Here
* Elaborate on some examples of how QA failures could occur.

Answer Here

* Propose strategies (process-oriented and communication-oriented) to mitigate the risks associated with QA failures.

Answer Here

### **Deliverables (Due by next Monday night):**

* Submit this document to canvas.
* Include a detailed report of your manual testing process, including any findings and screenshots, if applicable.
* Provide your Python Behave test cases in a .py file and .feature file (one file only please).

### **Evaluation Criteria:**

* Clarity and completeness of your answers to the high-level QA questions.
* Thoroughness and accuracy of your manual testing process.
* Correctness and coverage of your Python Behave test cases.
* Insightfulness and practicality of your responses to the additional questions.