# **Project Status Report**

## Name: Ashley Littles

## Date: 8-8-23

## Status Report

| **Status** | Progress |
| --- | --- |
| **Supporting Documentation** | * When revising the codebook, I only included the data that would be relevant to the research questions. I also renamed a couple of the data types to be better understood in the codebook. I removed the data: **step**-Maps a unit of time in the real world, **nameOrig**-The name of the customer who started the transaction, **nameDest**-The name of the customer who is the recipient of the transaction, **oldbalanceDest-**The initial recipient balance before the transaction without information for customers that start with M., and **newbalanceDest**-The recipient’s new balance after the transaction without information for customers that start with M. I chose to rename type to transactionType and amount to transactionAmount. |
| **Preliminary Data** | * The variables of interest that align with the research questions are: * **transactionType**: CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER. * **transactionAmount**: The amount the transaction is. * **oldbalanceOrg**: The starting balance of the transaction. * **newbalanceOrig**: The new balance after the transaction is made. * **isFraud**: Determines if the transaction is fraudulent (1) or non-fraudulent (0). * **isFlaggedFraud**: If a single transaction is trying to transfer more than $200,000 then it will be flagged as illegal. * For gathering the data, I began with duplicating the CSV record and revising it then save it as Revised\_FraudDatasetSubset. When revising the data, I remove any unnecessary information. I removed the data: step, nameOrig, nameDest, oldbalanceDest, and newbalanceDest. To make the data easier to understand I renamed type to transactionType and amount to transactionAmount. I then deleted row 1 which contained the variable names. The file was then saved and exported as a CSV file to be imported into mySQL workbench. I created a new schema called Frauddataset and then imported Revised\_FraudDatasetSubset into a new table. |
| **Data Analysis Requirements** | * The data that is available meets the data analysis requirements. There is plenty of data available to answer the research questions. Being able to match the transactionType to isFraud and isFlaggedFraud will be incredibly beneficial to solving the problem. Understanding the transactionAmount and the oldbalanceOrg and newbalanceOrig will allow us to figure out how obvious the fraud appears. |
| **Next Steps and Rationale** | * The available data and supporting documentation completely fulfill the data analysis requirements to address the organizational problem. The data provides direct answers to the research questions to ultimately solve the problem. * The data already presented is enough to support the organizational problem, but any extra additional data could help us dive deeper into our research. Knowing the location, the transaction was made could help give us more insight if that is connected to the fraudulent activity. |

## Supporting Documentation

The synthetic dataset was generated using the PaySim simulator. PaySim can create mobile money transactions from samples of real transactions from financial logs from a mobile money service being utilized in an African country. The original logs were given by a multinational company. This current company is a provider of mobile financial services that are actively running in more than 14 countries all over the world.

**Revised Codebook:**

1. transactionType: CASH-IN, CASH-OUT, DEBIT, PAYMENT, and TRANSFER.
2. transactionAmount: The amount the transaction is.
3. oldbalanceOrg: The starting balance of the transaction.
4. newbalanceOrig: The new balance after the transaction is made.
5. isFraud: Determines if the transaction is fraudulent (1) or non-fraudulent (0).
6. isFlaggedFraud: If a single transaction is trying to transfer more than $200,000 then it will be flagged as illegal.

## Preliminary Data