# **Project Status Report**

## Name: Alyssa Basile

## Date: December 6, 2021

## Status Report

| **Status** | Progress |
| --- | --- |
| **Supporting Documentation** | * I decided to focus on step, type, amount, and isFraud. I changed “step” to “timeHour”. I feel like “step” is more so for fitness (for example, using a Fitbit shows the amount of step), I don’t see it as time. I think the codes I have set up with the description is clear. Not only did I add the description with the column names, I added in the value range and the data type associated with the column. I also kept nameOrig, oldbalanceOrig, and newbalanceOrig. I kept nameOrig because the table needed a primary key. I kept oldbalanceOrig and newbalanceOrig because it would show why it’s considered fraud. I also removed isFlaggedFraud because I noticed all of them are 0, so they are not useful in this analysis. * The codebook is included below this table. |
| **Preliminary Data** | * step (“timeHour”) * type * amount * nameOrig * oldbalanceOrig * newbalanceOrig * isFraud * I looked over the CSV file to see what data is available to me. I figured out which data would be useful and which I should delete for the time being. After saving the revised CSV file, I was able to create a new table and import the file to SQL database and into the new table in SQL. |
| **Data Analysis Requirements** | * I would like to further understand what exactly “step” is. I am assuming it counts each hour every day for 30 days (744 total). It shows which one is a fraudulent transaction. None of them are flagged as fraud. It also does not tell me why or how it is happening, the only thing that seems common is the type and it is either “TRANSFER” or “CASH-OUT”. I also need to find out about the bank security system, whether it is put in place or not and if it is put in place, I need to see whether the system needs updated or has a weak spot. So, more analysis will be required. |
| **Next Steps and Rationale** | * The available data and supporting documentation help answer the one question I have, which is how often this happens. According to the data, 40 of the 4999 data is considered a fraudulent transaction. According to the time (if I am understanding this correctly), it appears it all happened in one day (from hour 1 to hour 5). The data provided does not help answer why it is happening and whether there are security defenses put in place. The data that has been given to me consists of the transactions and whether it’s a fraudulent transaction or not. Which is why further analysis will be required and not all answers I need are quantitative data. |

## Supporting Documentation

In addition to the status report, be sure to submit a distilled version of the codebook to help your stakeholders clearly understand what preliminary data is available.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Description** | **Value Range** | **Data Type** |
| timeHour | Unit of time in hours (1 = 1 hour) | 1, 2, 3…744 | INT |
| type | Type of transaction | CASH-IN, CASH-OUT, DEBIT, PAYMENT, TRANSFER | VARCHAR |
| amount | Amount of transaction in local currency | N/A | DECIMAL |
| nameOrig | Customer who started the transaction | N/A | VARCHAR |
| oldbalanceOrig | Initial balance before the transaction | N/A | DECIMAL |
| newbalanceOrig | Customer’s new balance after the transaction | N/A | DECIMAL |
| isFraud | Identifies a transaction as fraudulent | 1 = fraudulent  0 = non-fraudulent | INT |

## Preliminary Data

A picture containing text, screenshot, monitor, computer

Description automatically generated

Reference:

Southern New Hampshire University. Synthetic Financial Dataset Codebook for Fraud Detection. <https://learn.snhu.edu/content/enforced/900443-DAT-223-T2773-OL-TRAD-UG.21EW2/Course%20Documents/Synthetic%20Financial%20Dataset%20Codebook%20for%20Fraud%20Detection.pdf?_&d2lSessionVal=cQ5Mpz7R3rjQvoZzjoytPooux&ou=900443>.