# DAT 223 Project Two Client Proposal Part 2

**Organization Problem for:  
Analyst’s Name: Andrea Plunkett Jackson**

**Date:02/15/2024**

## Problem Statement

La Banca Central Bank has become aware of an increase in fraudulent transactions at viable financial institutions in regard to customer personal information and financial account compromise. The present objective is to identify activity parameters using various transactional data signatures that indicate probable fraud concentrating on monitoring and establishing identification of types of transactional data that are further liable to lead to duplicitous activity such as cash-outs, payments for goods and services, or fund transfers. With the vision to ultimately thwart risk as a whole.

## Research Questions

* How does La Banca Central Bank quality management affect the incidence and severity of bank fraud within the organization?
* How effective and ethical is pattern of fraud analysis in detecting and preventing point of sale or request for money movement fraud cases?
* How do data protection regulations affect the employee hiring process and outcome in the bank fraud sector?

## Data Types

The types of data required to be collected is both quantitative and qualitative in order to fully answer the above research questions addressing the organizational problem and vision for the future. Qualitatively identify customers through surveys using questioning on preferences, concerns, and baseline knowledge of current financial scams. Observation of current employee knowledge using in house resources of call monitoring and quiz base questioning establishing level of skill set for real time detection of fraud incidents presently occurring and textual data, to understand training and hiring procedure. In regard to quantitative data, continuous data is recommended as it allows for the measuring of fractional values that is incurred with financial analysis. This combination will allow us to address each question by providing objective and measurable evidence to locate the correlation between the variables and proceed with proper data driven resolution.

## Success Criteria

Success criteria for achieving positive outcomes based on La Banca Central bank vision to ultimately thwart risk from fraudulent transactions would be within the frequency parametric of 96%. The standards of success satisfying stakeholders through delivery of project outcome would be defined and measured through transparency and establishing benchmarks. Due to the sensitive and regulatory requirements of handling customer personal information and financial account data while cost will be minimal as the resources required are available both within the bank and public data sets, the scope and timeline must be considered in order to accommodate the large complex data.

## Variables of Interest or Units of Analysis

The variables of interest per codebook are as follows :

* Step: Timestamp variable to be used to look for outline of fraudulent behavior.
* Type: Characteristics of transactions to establish link between money movement type and fraud.
* Amount: The monetary value of the transaction to be used to evaluate possible link between fraud occurrence and transactional charges.
* NameDest: Receiver of transaction variable used to pinpoint possible commonality.
* IsFraud: Variable determining current fraud behavior in order to establish connections if applicable.
* IsFlaggedFraud: Review of account indicators on attempts on single transactions greater than $200,000 to establish controls in regard to Bank Secrecy Act.
* OldBalanceOrg: Based on the research question on internal possibility of fraud incidents review of balances prior to fraud activity to see if pattern can be determined.

For the Qualitative data, the unit of analysis that would be used would be a review of hiring criteria, training overview and call review to establish how fraud is currently handled within the bank.

## Data Gathering and Analysis Approach

Outline of general data analysis approach for gathering and analyzing the data provided in CSV file.

* Explore data quality by reviewing data types, missing values, outliers, duplicates, and the descriptive statistics through data cleaning to standardizing the numerical values and establishing table requirements.
* Import of CSV file into MySQL after defining objectives based on research questions and variables discussed above.
* Filter and group data in order to review patterns and correlations.
* Use visualization using matplotlib command to review the data through scatter plot to determine relationship between variables as well as box- and – whisker plot to see possible outliers that may be hidden within the information.

## Rationale

The research questions provide a full scope of bank fraud scenarios to encompass, and custom tailor data driven analysis required to meet or exceed the frequency parametric of 96% for success criteria. Pertaining to the variables provided within the dataset align with the required knowledge to cease fraudulent transactions. By completing an entire utilization through a correlation investigation, we will be able to address the business problem and meet the goal using statistical evidence. As customers are the life blood of banking, their hard earn dollars is what allows continuous servicing and the obligation to protect their assets with integrity. Through this process we will not only address the baseline business concerns but allow for future growth for the bank itself by establishing guidelines.

## Approval and Authority to Proceed

Managers to complete this section, please leave blank.

We approve the project as described above and authorize the team to proceed.

| Name | Title | Date |
| --- | --- | --- |
|  |  |  |