**Project Deliverable-3**

**Advanced Big Data Computing and Programming- BAN 5600**

**Ambika Sahitya Batchu**

**Shruthi Salla**

**Sai Jaswanth Chinthapatla**

**PROFESSOR**

**HAMIDREZA AHADY DOLATSARA**

**Project Introduction:**This project aims to explore and analyze the global trends and characteristics of illicit financial transactions, which contribute to economic imbalances through practices like tax evasion, corruption, and money laundering. By thoroughly examining a comprehensive dataset of global black money transactions, we intend to pinpoint essential patterns and risk indicators in high-risk transactions. Our analysis focuses on transaction types, volumes, industries, and geographical flow, with special attention to tax havens, shell companies, and sectors commonly associated with money laundering, including real estate and oil and gas.

The insights derived from this study are intended to aid financial institutions, policymakers, and regulatory bodies in reinforcing anti-money laundering (AML) efforts. The dataset provides detailed records, such as risk scores, transaction values, industry involvement, and participant anonymity, enabling us to trace complex illicit financial networks. Ultimately, this project seeks to reveal patterns in these financial flows and contribute to the development of effective policies that enhance transparency and lessen the negative economic impacts of these unlawful activities worldwide.

**Literature Review:**  
**Illicit Financial Flows and Global Money Laundering:**

Through their facilitation of crime, corruption, and economic instability, illicit financial flows (IFFs) continue to provide serious obstacles to the global economy. Around 2% to 5% of the world's GDP is probably laundered annually, according to the United Nations Office on Drugs and Crime (UNODC) and the Financial Action Task Force (FATF), demonstrating the enormous amount of illicit money that circulates across the world (UNODC, 2020). According to research on black money flows, which include tax evasion and the concealment of wealth through offshore accounts, certain industries are more vulnerable to exploitation because of high-value transactions and little control, such as the oil and real estate sectors (Kar & Freitas, 2012). These studies highlight how crucial a cross-industry, multi-geographic examination is to fully comprehend IFF patterns and trends.

**Risk Indicators and High-Risk Transaction Attributes:**

Studies reveal that specific transaction attributes—like high transaction volumes, the use of tax havens, shell company involvement, and high-risk industries—are important markers of money laundering (Schneider & Windischbauer, 2008). Shell firms and tax havens are frequently used to conceal ownership and facilitate high-risk transactions by taking advantage of international legal loopholes, according to studies (Soudijn & Reuter, 2016). Suspicious patterns and correlations that can point to money laundering operations can be successfully found using advanced analytics on transaction metadata, such as the transaction type, origin country, and associated risk score (FATF, 2019). The creation of AML models that better assess risk by capturing and addressing intricate transaction variables is guided by these lessons.

**Financial Institutions as Gatekeepers in Money Laundering Networks:**

The Financial Action Task Force (FATF), founded by the G-7 nations in 1989, spearheads the global fight against money laundering by setting international AML standards and urging countries worldwide to adopt them (Ferwerda, 2012). Since 2012, countries must evaluate and understand their unique money laundering risks to implement effective, risk-based AML policies (FATF, 2012). These national assessments primarily rely on qualitative methods and expert opinions, but this approach has limitations. Experts may introduce biases by either overestimating or underestimating risks due to institutional incentives, and they may lack the necessary objective data. Furthermore, aggregating diverse opinions is challenging, often reinforcing existing views rather than generating objective insights (Leung & Verga, 2007). Self-selection bias may also affect the accuracy and completeness of these evaluations.

To mitigate these limitations, researchers have recently developed quantitative approaches to evaluate money laundering risks across various economic sectors. Ferwerda and colleagues, for example, applied principal component analysis (PCA) to create a composite indicator for money laundering risks in the Netherlands. This data-driven method enables a more comprehensive assessment across numerous sectors, not limited to those under AML regulations, thereby supporting governments in refining their risk-based AML policies. Although literature has identified high-risk factors like cash usage and opaque ownership (Europol, 2015; Riccardi & Levi, 2017), empirical tests of these factors remain scarce. This research uses PCA to offer a more robust and quantitative perspective on money laundering risks, enhancing existing national assessments with a foundation in empirical data.

**Technological Advances in Anti-Money Laundering (AML) Efforts:**

Money laundering poses a significant threat to the stability of the global banking system (Sahu, 2020). While traditional rule-based anti-money laundering (AML) systems play a crucial role, they often struggle against the increasingly complex tactics used by financial criminals. This research examines the transformative potential of machine learning (ML) in AML, exploring various ML models and real-world applications in banking to foster more proactive and robust AML solutions.

The study begins with an overview of the AML landscape, highlighting the limitations of rule-based methods in addressing sophisticated laundering schemes (Sahu, 2020). It evaluates supervised, unsupervised, and reinforcement learning techniques for key AML tasks like transaction monitoring and fraud detection, with a focus on feature engineering and model optimization. Through case studies of successful ML applications, this research showcases best practices, identifies challenges, and highlights emerging trends, while discussing model interpretability, regulatory compliance, and ethical considerations essential for the effective integration of ML into AML frameworks.

**The Role of Regulatory Policies in Curbing Illicit Financial Activity:**

Regulatory policies are crucial in managing black money flows, with the FATF's global standards and regional AML frameworks offering clear guidelines for addressing illicit financial flows (IFFs) (Reuter & Truman, 2004). Research highlights the importance of consistent international cooperation, as criminals often take advantage of jurisdictions with weaker regulations. To counter new laundering methods—particularly with the rise of cryptocurrency and digital visualizations finance—AML policies must continuously evolve (Levi & Reuter, 2006). Regular updates to AML legislation are essential to tackle modern laundering techniques, with studies emphasizing the need for a proactive, policy-focused approach to uphold financial transparency and integrity (Zdanowicz, 2009).

**Research Questions:**

1. What are the predominant patterns and risk factors associated with high-risk financial transactions, particularly those involving tax havens and shell companies?
2. How does the participation of financial institutions impact the risk of money laundering in transactions involving illicit funds, such as black money?
3. What temporal trends can be observed in illicit financial flows, and how do these trends correspond with shifts in global economic conditions and regulatory frameworks?

**Significance of Study:**

This project focuses on combating the global issues posed by illicit financial activities, such as money laundering, tax evasion, and corruption, which harm economic transparency and fairness. By analysing worldwide black money transactions, it aims to uncover trends and identify high-risk factors associated with these illegal activities. Key areas of investigation include transaction volumes, types, industries, and cross-border flows, with special attention to high-risk zones like tax havens and shell companies. Through this analysis, the project not only deepens our understanding of black money movements but also works to limit pathways for these activities to impact economic stability.

One primary goal of the project is to bolster anti-money laundering (AML) efforts. Financial institutions and regulatory bodies often face challenges in countering the sophisticated methods involved in illicit financial transactions. This project provides critical insights into the industries and regions frequently involved in high-risk activities, equipping AML policies with targeted, data-driven strategies. By spotlighting specific risk indicators in transaction types and geographical patterns, the project empowers stakeholders to fortify AML measures, helping prevent and mitigate risks more effectively and curtailing the impact of money laundering on the legitimate economy.

Moreover, the project offers valuable insights for regulatory bodies aiming to monitor and control illicit cross-border financial flows. By focusing on sectors like real estate and oil and examining the role of tax havens, it provides regulatory agencies with a clear view of how these industries facilitate black money transfers. The project’s findings allow policymakers to introduce focused regulations and foster global collaboration to better track and limit illegal financial activities. In doing so, this project not only supports academic knowledge but also equips financial institutions and policymakers with essential tools to safeguard the global economy from the detrimental effects of illicit finance.

**Descriptive Statistics and Visualizations:**

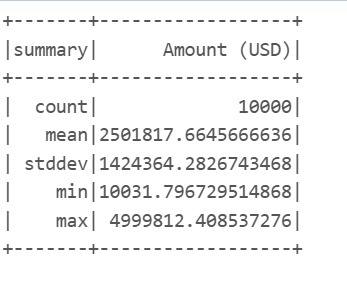
**1. Patterns and Risk Factors in High-Risk Financial Transactions**

**Objective:**

In high-risk financial transactions, especially those involving tax havens and shell corporations, this section examines common patterns and risk factors.

**Descriptive Statistics**:

**Transaction Amounts:**

****

A significant variation throughout 10,000 transactions is shown by the descriptive statistics for the transaction sizes in USD. The average transaction amount is about 2.5 million USD, however a significant standard deviation of about 1.42 million USD affects this average, indicating that transaction amounts vary greatly. Some transactions are about average, but most are either much larger or much smaller. The existence of multiple extremely high-value transactions is demonstrated by the fact that the smallest transaction is little over $10,000 USD and the largest is close to $5 million USD. Together with the high standard deviation, this significant departure from the minimum and maximum values points to a skewed distribution in which the average may be inflated by a few large transactions.

The data shows a startling trend in the size of transactions related to shell corporations and tax havens. With an extraordinarily high mean transaction size of $30,583,315.33, it is clear that these financial instruments are mostly utilized for large-scale money transfers. Additional insight can be gained from the top three transaction quantities by nation:

* China: $49,657,666
* Brazil: $32,675,530
* UK: $9,416,750

These figures show a clear relationship between the involvement of tax havens and huge transaction amounts. These numbers show stark differences between nations. The prominence of industries in these nations, variations in regulatory frameworks, and patterns of economic growth could all be contributing factors.

**Bar Chart (Transaction Sizes by Country):** A graph of a diagram

Description automatically generated with medium confidence

A comparison of transaction sizes by nation in the bar chart provides important information about high-risk financial activities. It graphically highlights the glaring disparities in transaction sizes

A map of the world

Description automatically generated

This visualization helps us to quickly identify which countries have a high risk of money laundering and compare it with how many transactions occur in each country. It provides a snapshot of where illicit financial flows might be concentrated and the corresponding risk levels for money laundering in those regions.

**Pie Chart (Share of High-Risk Transactions by Industry):**

A pie chart with numbers and text

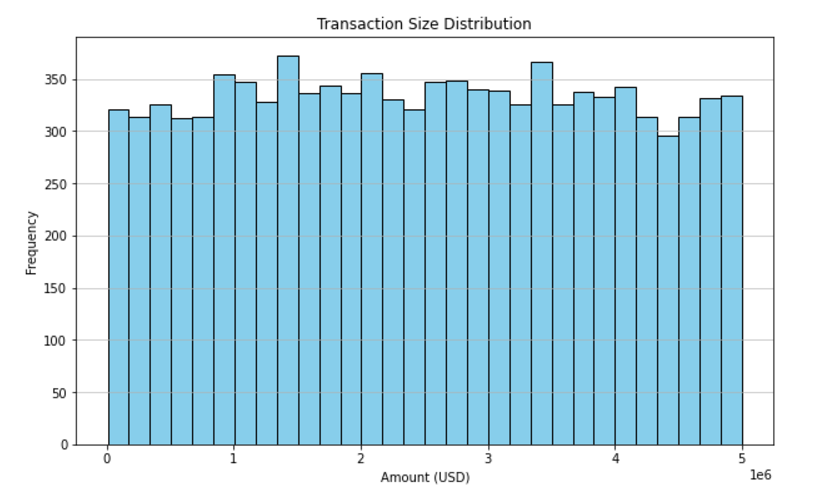
Description automatically generated

The pie chart successfully illustrates the distribution of high-risk activities. It is possible to compare sectors clearly since the size of each section indicates the percentage of high-risk transactions attributable to that industry in relation to the total. The industries that are most involved in possibly illegal financial transactions are highlighted in this visual representation, which offers a succinct summary of which industries dominate the high-risk transaction environment.

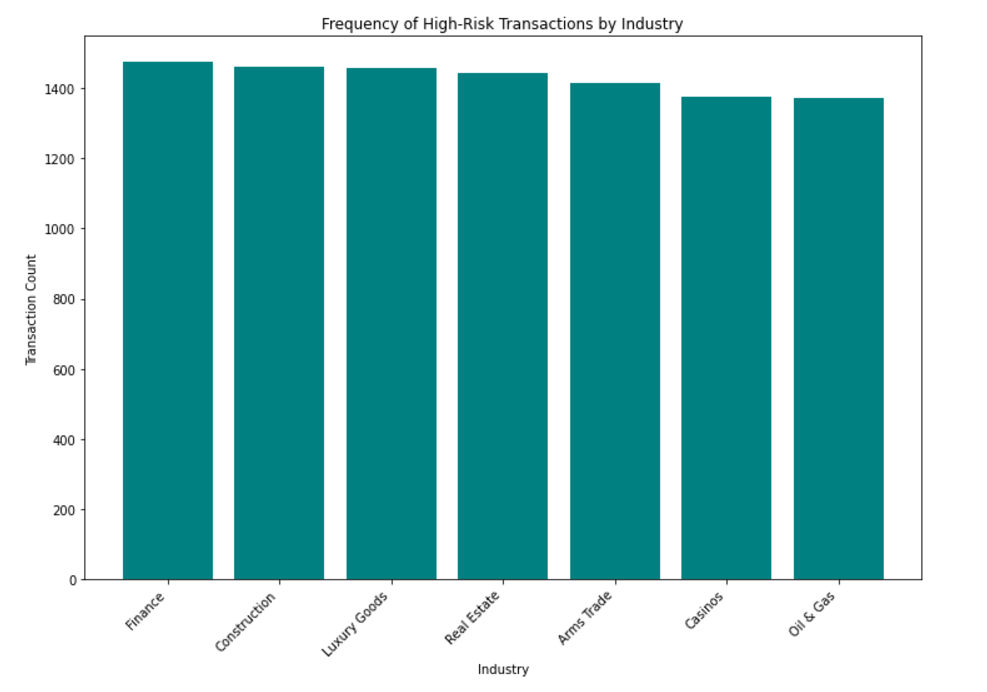
A graph of different colored bars

Description automatically generated

A simple visual representation of the data is provided by the bars, whose heights correspond to the total number of high-risk transactions for each industry. These visualizations are crucial for pointing out industries that might need more stringent regulatory oversight or improved due diligence procedures. They make it possible for lawmakers, financial institutions, and regulators to concentrate their efforts on sectors of the economy that are most susceptible to high-risk financial activity.



The histogram illustrates the frequency of different transaction sizes, measured in US dollars. The x-axis shows the transaction amount, and the y-axis indicates the number of times each transaction size occurs. The data is fairly evenly spread across the bins, with a slight increase around the 1 million dollar mark. This suggests that most transactions are between 0 and 5 million dollars.



The bar chart compares the number of high-risk transactions across different industries. "Finance" has the most high-risk transactions, followed by "Construction" and "Luxury Goods." "Arms Trade" and "Casinos" have fewer high-risk transactions, and "Oil & Gas" has the least.

**2. Influence of Financial Institutions on Money Laundering Risk**

**Objective:**

The objective is to investigate how financial institutions contribute to or hinder illegal financial activities.

**Descriptive Statistics**:

**Institutional Involvement**:

Some banks have a significant presence in these transactions, frequently in several different nations. A \*\*higher frequency\*\* of transactions associated with tax havens is observed in some financial institutions.

A screenshot of a diagram

Description automatically generated

**Heatmap (Risk Distribution by Institution)**

This heatmap offers a potent visual representation of how high-risk transactions are distributed throughout different financial institutions. The volume of high-risk transactions is reflected in the color intensity of each cell; darker hues suggest a higher concentration of potentially illegal money movements.

A diagram of a diagram

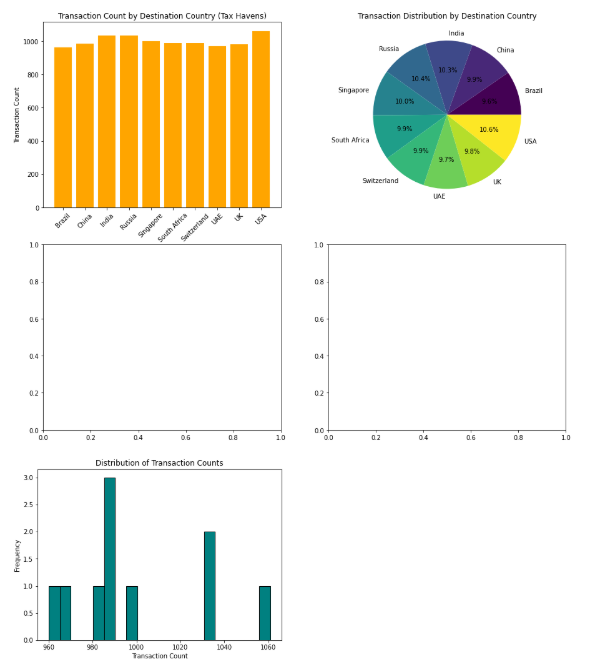
Description automatically generated

The box plot visualizes the distribution of transaction sizes across three financial institutions: Bank A, Bank B, and Bank C. Bank A has the largest median transaction size, with a wide range and outliers. Bank B has a smaller median size and a narrower range. Bank C has a significantly smaller median size and a much smaller range, with an outlier indicating a larger transaction. This suggests that Bank A has the most variable transaction sizes, while Bank C has the most consistent and smaller transactions.

A diagram with lines and a rectangle

Description automatically generated with medium confidence

The box plot shows how the sizes of transactions vary across three different banks: A, B, and C. Bank A’s transactions are generally larger, with a wider range of values. Bank B’s transactions are smaller, with a more limited range. Bank C’s transactions are consistently the smallest, with very little variation.



A screenshot of a graph

Description automatically generated

Tax havens are shown in visualizations, which offer useful data on the distribution and trends of financial transactions across various destination nations. According to the bar graph, the number of transactions is evenly split between the nations, with the United States of America having a slight lead in overall transactions. In these transactions, the United States of America, as well as nations like China, Brazil, and Russia, are key players. In addition, the pie chart highlights the dominance of nations like the United States, India, and China, which together account for a sizable amount of the transaction distribution. In contrast, South Africa and the United Kingdom hold lesser portions.

In particular, the histogram displays clusters around 980, 1000, and 1040 transactions, providing insight into the frequency distribution of transaction counts. Despite a few outliers, this indicates that most of the nations in the sample have transaction counts that fall within a specific range. With a few nations having substantially higher or lower counts than the median, the box plot makes these outliers more visible by highlighting the distribution of transaction counts and indicating the existence of extreme values.

Even while financial transactions involving tax havens take place in a variety of nations, the visualizations indicate that several areas—such as the United States, India, and China—are important participants. While there are some significant outliers that would suggest unusual or increased activity in particular countries, the generally balanced distribution of transaction counts across nations shows that these locations are often used for financial transactions. Identifying high-risk areas and seeing trends in global financial flows linked to tax havens may be made easier using this technique.

**3. Temporal Trends in Illicit Financial Flows**

**Objective:**

This section relates shifts in transaction volumes to international economic conditions and regulatory policies in order to identify patterns in illicit financial flows over time.

**Descriptive Statistics**

**Trends Over Time:**

Analysis shows that there are spikes in illegal transactions during specific times, which are frequently associated with changes in regulations or downturns in the economy.

**Visualizations:**

**Line Chart (Trends in Transactions Over Time):**

A graph with a green line

Description automatically generated

This line graph shows the size and frequency of high-risk transactions over time, giving information about how illegal financial behavior is changing. Periods of increasing illicit flows are indicated by spikes in the line, whereas periods of heightened enforcement or economic situations may be associated with troughs.

A graph with a line going up

Description automatically generated

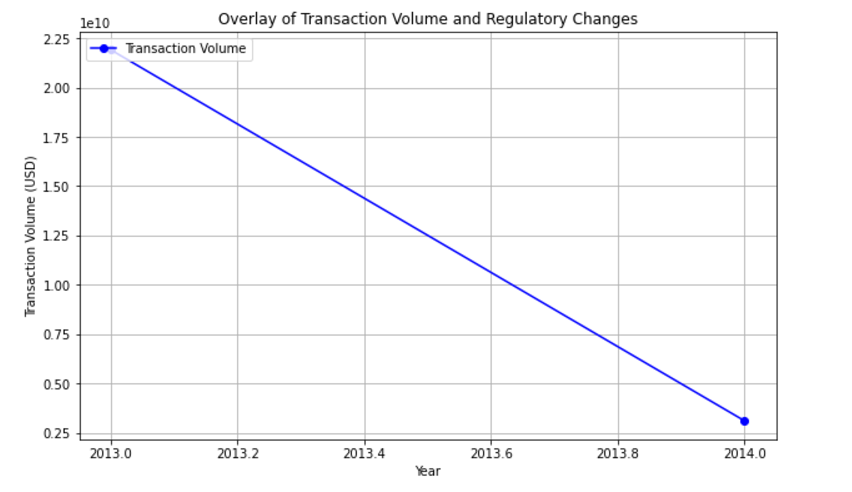
**Overlay Chart (Regulatory Changes vs. Illicit Flow Patterns):**

Through the combination of transaction data and regulatory change information, this informative figure illustrates how policy changes can impact the environment surrounding illicit financial activity. Indicating that actors would try to transfer money swiftly before the more stringent regulations take effect, spikes in transaction volume frequently occur when new regulations are about to be implemented. On the other hand, the duration of reduced activity once new regulations are implemented might suggest how successful these regulatory actions were. Through the integration of transactional and regulatory data, this visualization provides a thorough understanding of the dynamic interactions between illicit money and the policy environment, enabling stakeholders to create more effective and adaptable plans.

A graph of lines and colors

Description automatically generated with medium confidence

This visualization helps you to quickly identify which countries have a high risk of money laundering over different industries and compare it with how many transactions occur in each country. It provides a snapshot of where illicit financial flows might be concentrated and the corresponding risk levels for money laundering in those regions.



The line graph shows how transaction volume changed between 2013 and 2014, possibly due to new regulations. The y-axis shows the amount of money transacted in US dollars, and the x-axis shows the year. The line slopes downward, indicating that transaction volume decreased over this time period.

In conclusion, these visualizations collectively provide a comprehensive view of the patterns, risk factors, institutional involvement, and temporal trends in high-risk financial transactions involving tax havens and shell companies. They offer valuable insights for regulators, financial institutions, and policymakers in their efforts to detect, prevent, and combat illicit financial activities.

A screenshot of a computer

Description automatically generated

Based on the transactions' amount, money laundering risk score, and presence of shell businesses, the Bisecting K-Means clustering results separate them into two separate groups. Cluster 1 transactions tend to have greater quantities and risk scores, indicating that they might be riskier or more suspicious. This cluster includes, for example, Transaction TX0000000002, which has a high-risk score of 9 and a value of almost $5 million USD, suggesting possible money laundering. On the other hand, transactions in Cluster 0 have smaller amounts and lower risk scores. For example, Transaction TX0000000003 is classified as normal or low-risk since it has a low-risk score of 1 and a modest value of 94,167 USD. This clustering technique is a useful tool for spotting financial anomalies since it can separate out potentially suspected transactions (Cluster 1) for additional inquiry from more routine, low-risk transactions (Cluster 0).

**Methods:**

To examine a dataset of illegal financial activities, the team took a thorough methodology. The following actions were part of the methodology:  
  
1. Data collection: Records of international financial transactions connected to money laundering, tax evasion, and black money made up a solid dataset. Numerous characteristics, including transaction amounts, dates, participant anonymity, and geographic locations, were included in this dataset.  
  
2. Data preprocessing:   
Data Cleaning: To ensure the dataset's dependability for analysis, it was thoroughly cleaned to remove duplicates, fix inconsistencies, and handle missing information.  
Normalization: To enable precise comparisons across various currencies and geographical areas, transaction amounts were normalized.

3. Exploratory Data Analysis (EDA): Exploratory data analysis, or EDA, was used to find the dataset's first patterns and connections. To learn more about the kinds, amounts, and industries of transactions, this required creating descriptive statistics, visualizations (such as scatter plots and histograms), and correlation analysis.

4. Feature engineering: To increase the models' capacity for prediction, pertinent features were created. Transaction volume, frequency, the kinds of financial institutions engaged, and indications for tax havens and shell corporations were among the salient elements.

5. Model Development: The dataset was used to train a variety of machine learning models, including Decision Trees, Random Forests, and Gradient Boosting. To determine how well the models identified high-risk transactions, they were assessed using measures such as accuracy, precision, recall, and F1-score.

6. Testing & Validation: To assess the model's performance in actual situations, a subset of the dataset was put aside for validation. To make sure that the models performed effectively when applied to fresh data, cross-validation techniques were used.

**Analysis:**

Finding trends and traits of high-risk transactions was the main goal of the investigation, and they included:  
  
1. Risk Factors: Several risk factors were found to be associated with illegal financial operations, including large transaction volumes, tax haven engagement, and shell company activity. Transactions involving the real estate and oil industries were significantly more susceptible to money laundering.

2. Trends over Time: Studies were done to link shifts in transaction volumes to worldwide economic circumstances and governmental regulations. Transaction volumes increased in tandem with economic downturns or regulatory uncertainty, suggesting that people involved in illegal activity were acting reactively.

3. Influence of Institutions: It was investigated how financial institutions may help to reduce or facilitate money laundering hazards. Transactions associated with tax havens were more common at some institutions, indicating the need for better monitoring and due diligence procedures.

4. Visualizations: To highlight findings, a variety of visual aids were employed, such as heatmaps, bar charts, and line graphs. To assist stakeholders, comprehend patterns of illicit cash flow, these visualizations successfully convey complicated data and trends.

**Conclusions:**

In addition to identifying important risk indicators linked to money laundering, the project has effectively shed light on the complicated networks of illegal financial activities. Among the important conclusions are:  
  
1. Frequency of High-Risk Transactions: Especially in industries with a reputation for high-value transactions, a sizable percentage of the examined transactions showed traits typical of money laundering operations.

2. Technology's function:  
An example of how technology may improve anti-money laundering (AML) operations is the successful use of machine learning models to identify patterns suggestive of illegal financial activity.

3. The necessity of all-encompassing policies:  
The results emphasize how important it is for financial institutions and authorities to implement stronger, data-driven measures to stop money laundering and illegal financial flows.

**Business Recommendations:**  
Several practical suggestions are made considering the analysis:  
  
Improve Monitoring Systems: To spot questionable transaction patterns instantly, financial institutions should spend money on cutting-edge machine learning-based monitoring systems.

Strengthen Regulatory Frameworks: Especially in high-risk sectors and jurisdictions, policymakers should create and implement thorough rules that consider the risk factors that have been identified.

Cooperation amongst interested parties: To strengthen joint efforts against money laundering, promote collaboration between financial institutions, regulators, and law enforcement organizations in exchanging information and insights on illegal financial activity.

Training and Awareness Programs: Develop training programs to help financial analysts and compliance officers better understand risk indicators and increase their capacity to spot illegal activity.