

Project 4 Group 11

Black Money Data Preparation

Data Cleaning

During the data cleaning process, we began by renaming the columns to exclude any special characters and converted all letters to lowercase. We then checked for null values within our dataset, and fortunately, there were none. Lastly, we converted the transaction date from an object type to a proper date format. With the data cleaning portion complete, we were able to export the cleaned dataset for use in the machine learning phase of the experiment.

It is worth noting that during the data cleaning and exploratory data analysis (EDA) process, we discovered that our dataset was very well distributed across our categorical features, regardless of whether the transactions were legal or illegal. This made us realize that finding a strong correlation between features would be challenging. Refer to the appendix for images.

Machine Learning Experiment

To begin our machine learning experiment, we first dropped several columns that appeared unimportant to the model, including the transaction date, individuals involved, financial institution, year, and the transaction ID. These columns did not provide any meaningful impact on our machine learning model.

After dropping the unnecessary columns, we defined our pipeline by separating the features into three categories: categorical, binary, and numeric. First, we scaled the numeric features using a standard scaler. Since there were no null values, no imputation was necessary. Next, we transformed the binary features from objects to integers. Lastly, we applied one-hot encoding to the categorical features. Once the encoding was completed, we combined all the features into a final dataset, which was now ready for analysis.

After compiling the final dataset, we tested various models to determine which performed best. Given the nature of our dataset and its balanced distribution, we hypothesized that a model capable of identifying nonlinear relationships would be most suitable for this experiment. Linear regression, SVC, AdaBoost, GradientBoost, and K-nearest neighbors were all unable to identify any meaningful trends. This led us to test LGBM, XGBoost, and extra tree models. These models performed the best, with weighted accuracies ranging from the upper 50% to low 60%. Although LGBM and XGBoost were the top-performing models, due to the constraints of PythonAnywhere, we ultimately chose to use the extra tree model.

Grant Hagen
Marty Thompson
Sabina Kamalova
Jonathan YoannidesYoung

It is important to mention that we attempted to build a neural network to evaluate whether it would outperform the other models. In fact, it did, achieving an accuracy score in the high 60%. However, due to the limitations of PythonAnywhere, we were unable to move forward with the neural network. Additionally, our dataset was unbalanced, with a higher proportion of illegal transactions compared to legal ones. While we attempted to address this imbalance using SMOTE, time constraints prevented us from including this adjustment in our final model.

source_of_money			
		count	
		industry	source_of_money
transaction_type	source_of_money	Arms Trade	Illegal
			Legal
Cash Withdrawal	Illegal	Casinos	998
	Legal		416
Cryptocurrency	Illegal	Construction	994
	Legal		383
Offshore Transfer	Illegal	Finance	1003
	Legal		457
Property Purchase	Illegal	Luxury Goods	1026
	Legal		449
Stocks Transfer	Illegal	Oil & Gas	1030
	Legal		429
		Real Estate	Illegal
			Legal
source_of_money			
		count	
destination_country	source_of_money		source_of_money
			count
Brazil	Illegal	1	715
	Legal		311
China	Illegal	2	660
	Legal		287
India	Illegal	3	746
	Legal		289
Russia	Illegal	4	692
	Legal		310
Singapore	Illegal	5	663
	Legal		289
South Africa	Illegal	6	681
	Legal		303
Switzerland	Illegal	7	698
	Legal		280
UAE	Illegal	8	697
	Legal		286
UK	Illegal	9	748
	Legal		325
USA	Illegal	10	717
	Legal		303
source_of_money			
		count	
country	source_of_money		source_of_money
			count
Brazil	Illegal		748
	Legal		258
China	Illegal		711
	Legal		327
India	Illegal		695
	Legal		295
Russia	Illegal		701
	Legal		296
Singapore	Illegal		685
	Legal		310
South Africa	Illegal		728
	Legal		298
Switzerland	Illegal		687
	Legal		298
UAE	Illegal		685
	Legal		291
UK	Illegal		708
	Legal		306
USA	Illegal		669
	Legal		304

Data Visualization with Tableau

In this project, Tableau was utilized to transform complex datasets into interactive visualizations, making it easier to uncover meaningful insights. The data preparation phase involved organizing fields into dimensions, such as Country, Financial Institution, and Industry, and measures like Transaction Amount and Money Laundering Risk. This distinction allowed for a more structured approach to exploring the data. After preparing the data, various dashboards were built to visualize key trends, allowing us to interactively explore relationships across different regions, individuals, and industries.

One of the main goals of using Tableau was to identify potential money laundering hotspots. By mapping transactions geographically, we could track financial activities and observe patterns, particularly in regions where higher risks of illegal activities were suspected. The ability to drill down into specific locations and filter by different financial institutions or individuals allowed for a more in-depth analysis.

Additionally, the dashboards provided insights into the flow of money across borders, particularly to and from tax haven countries. By visualizing the financial streams, we gained a better understanding of how illicit funds could be moving between regions. The combination of heat maps, bar charts, and trend lines made it easier to identify anomalies in the financial data, helping to spot potential red flags that may indicate money laundering activity.

These visualizations not only simplified complex datasets but also provided a platform for generating actionable insights that could be used for further investigation or compliance monitoring.

Design Concept and Color Scheme

In designing the dashboards and reports, we carefully selected a color scheme to reflect the underlying theme of global financial crimes and money laundering. Each color was chosen with purpose to convey the serious, shadowy atmosphere often associated with hidden and illicit financial activities.

We chose steel blue to represent the cold, calculated nature of global financial transactions. This color highlights the detached, mechanical processes involved in transferring illicit funds across borders, emphasizing the precision and complexity of financial crimes. It sets an analytical tone, allowing us to dissect these covert operations visually.

Grant Hagen
Marty Thompson
Sabina Kamalova
Jonathan Yoannides Young

To contrast this, we incorporated burnt orange, signaling danger and urgency. This color represents the high-risk nature of illegal financial activity, creating tension alongside the steel blue. Together, these colors suggest that beneath the surface of seemingly routine transactions lies a constant threat of exposure and legal consequences. We used this combination to reinforce the idea of ordinary-looking, yet dangerous, financial dealings.

Dark maroon was another essential part of our palette, symbolizing hidden wealth and power. This color evokes secrecy, greed, and the accumulation of vast, untraceable assets, underscoring how illicit funds are often hidden behind complex layers of financial transactions. By using dark maroon, we aimed to visually represent the accumulation of wealth kept out of sight, beyond the reach of regulatory authorities.

Finally, we incorporated olive green to draw attention to the concept of "dirty" wealth—funds acquired through corrupt or illegal means. The green tones tie back to the traditional association with money, emphasizing how wealth and corruption are intertwined. This color was crucial in reinforcing the theme of money laundering and the corrosive impact it has on legitimate financial systems.

Together, the colors we selected create a cohesive visual language that mirrors the serious, calculated, and secretive world of global financial crimes. They enhance both the aesthetic and conceptual elements of our project by conveying the tension, urgency, and corruption inherent in illicit financial activities.

