BLACK MONEY

Marty Thompson, Grant Hagen , Jonathan YoannidesYoung, & Sabina Kamalova

Inspiration

We found the black money dataset particularly interesting because it uncovers hidden financial activities that have a significant impact on the global economy. Through the data, we can explore patterns of illicit wealth generation and the complex flow of black money across international borders.

Additionally, the dataset reveals the challenges involved in tracking and combating financial crime, highlighting gaps where policy improvements are needed. By studying this data, we gain valuable insights into the ethical, economic, and social implications of black money, making it a compelling subject for deeper investigation.

Design Concepts

- Steel blue represents the cold, calculated transactions in global financial crimes.
- Burnt orange signals danger and urgency, reflecting the risky nature of illicit funds.
- Dark maroon suggests hidden wealth and power, underscoring secrecy and greed.
- Olive green is linked to money and corruption, emphasizing “dirty” wealth.

Together, these colors convey a serious, shadowy atmosphere, perfect for the concept of hidden, illegal money.

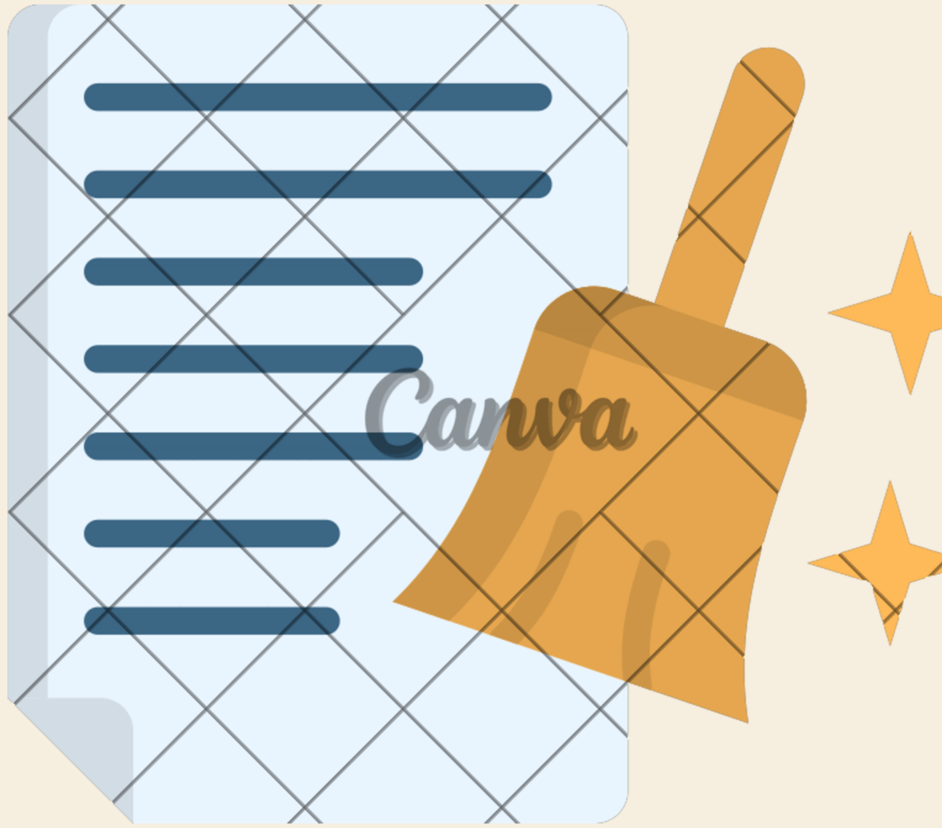
Research Questions?

- Is there any particular country who is prone to more illegal transactions?

- Is there a threshold that makes a transaction more sketchy?

- Are there transaction types that are more sketchy than others?

Data Cleaning



- Pretty clean dataset
- Changed data types for a few columns
- Dropped columns that provided no value
- No null values
- Tried to extract more value, no value added to model due to distribution of the dataset

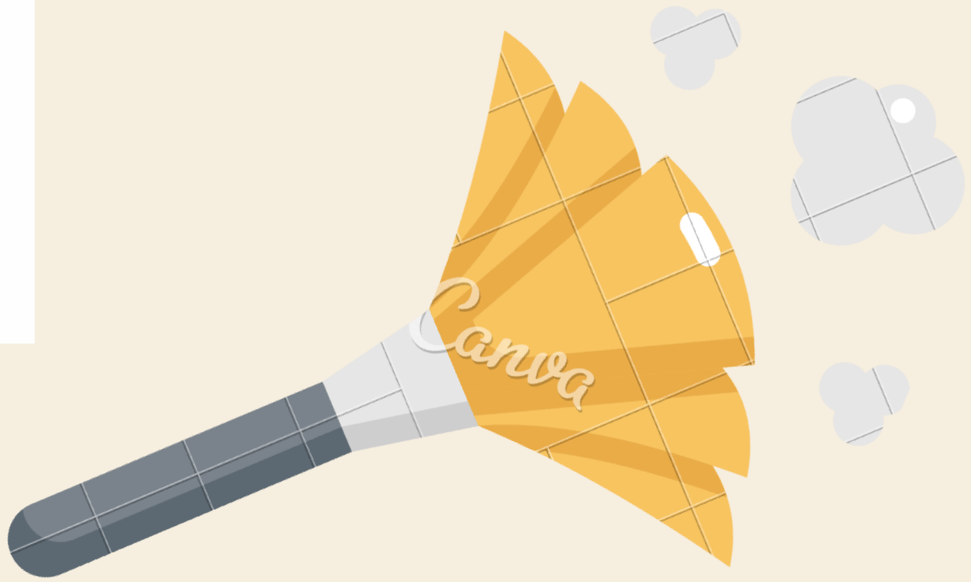
Data Cleaning Cont.

		source_of_money
		count
country	source_of_money	
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

		source_of_money
		count
transaction_type	source_of_money	
Cash Withdrawal	Illegal	1402
	Legal	576
Cryptocurrency	Illegal	1370
	Legal	603
Offshore Transfer	Illegal	1382
	Legal	598
Property Purchase	Illegal	1467
	Legal	619
Stocks Transfer	Illegal	1396
	Legal	587

		source_of_money
		count
industry	source_of_money	
Arms Trade	Illegal	998
	Legal	416
Casinos	Illegal	994
	Legal	383
Construction	Illegal	1003
	Legal	457
Finance	Illegal	1026
	Legal	449
Luxury Goods	Illegal	1030
	Legal	429
Oil & Gas	Illegal	970
	Legal	402
Real Estate	Illegal	996
	Legal	447

		source_of_money
		count
destination_country	source_of_money	
Brazil	Illegal	671
	Legal	289
China	Illegal	672
	Legal	314
India	Illegal	736
	Legal	296
Russia	Illegal	722
	Legal	313
Singapore	Illegal	700
	Legal	299
South Africa	Illegal	719
	Legal	270
Switzerland	Illegal	680
	Legal	307
UAE	Illegal	691
	Legal	279
UK	Illegal	654
	Legal	327
USA	Illegal	772
	Legal	289



Data Cleaning Cont.

reported_by_authority	source_of_money	count
False	Illegal	5616
	Legal	2379
True	Illegal	1401
	Legal	604

money_laundering_risk_score	source_of_money	count
1	Illegal	715
	Legal	311
2	Illegal	660
	Legal	287
3	Illegal	746
	Legal	289
4	Illegal	692
	Legal	310
5	Illegal	663
	Legal	289
6	Illegal	681
	Legal	303
7	Illegal	698
	Legal	280
8	Illegal	697
	Legal	286
9	Illegal	748
	Legal	325
10	Illegal	717
	Legal	303

shell_companies_involved	source_of_money	count
0	Illegal	735
	Legal	319
1	Illegal	686
	Legal	304
2	Illegal	693
	Legal	307
3	Illegal	673
	Legal	306
4	Illegal	715
	Legal	280
5	Illegal	738
	Legal	291
6	Illegal	712
	Legal	283
7	Illegal	675
	Legal	307
8	Illegal	699
	Legal	287
9	Illegal	691
	Legal	299

tax_haven_country	source_of_money	count
Bahamas	Illegal	1119
	Legal	509
Cayman Islands	Illegal	1178
	Legal	498
Luxembourg	Illegal	1190
	Legal	491
Panama	Illegal	1235
	Legal	508
Singapore	Illegal	1153
	Legal	491
Switzerland	Illegal	1142
	Legal	486

Machine Learning Experiment

- Defined our pipeline
- XGBoost and LGBM models performed the best
- Due to PythonAnywhere, we proceeded with ExtraTrees
- Moderate results due to our dataset

ML Experiment Cont.

- Future work could include the following
- SMOTE redistribution
- Neural network
- Grouping data in more ways

Live Demo

https://ghagen.pythonanywhere.com

Limitation #1

One limitation we faced was incomplete or missing data, which made it challenging to gain a comprehensive understanding of the full scope of black money activities. This gap in the dataset restricted our ability to draw accurate conclusions and fully map out the problem.

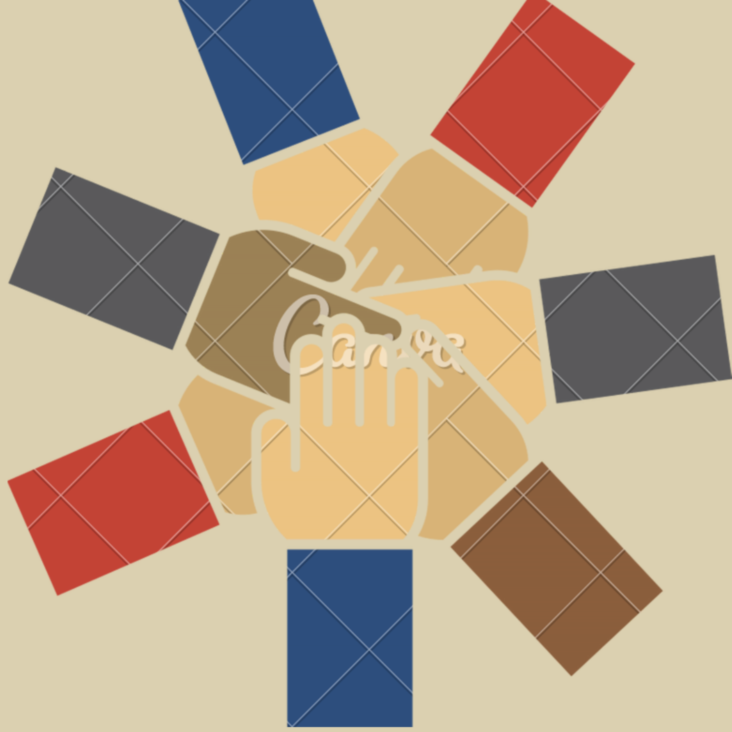
Limitation #2

Another issue was the lack of granular details in specific areas of the dataset. This limitation prevented us from conducting a deeper analysis of certain patterns or trends, especially when examining smaller-scale activities or more nuanced aspects of black money movement.

Limitation #3

A significant limitation was that some data entries were flagged solely based on the suspicions of officers, without concrete evidence. This introduced a level of subjectivity and potential bias into the dataset, which could affect the reliability of any conclusions drawn from these flagged entries.

Conclusion/Reflection



· The data set is imbalanced with low legal vs high illegal transactions. We were only able to achieve up to 60% accuracy.
· A big limitation is that we did not know whether the data was generated from actual illegal/illicit transactions or from Suspicious Activity Reports (SARs) that turned out to be legal.
· Even SARs that turn out to be legal can still point to illicit activities.
· Having SARs as legal vs actual illegal transactions would help to utilize resources in the proper transaction types (real estate, crypto-currency, cash, transfers and others) and the most corrupt countries.
· Our call to action is to use data with less limitations and bias to further analyze.
· We can conclude that Dark Money is a world-wide economic problem.

Thank You