Lumbago Edge Bank

Investigating Fraud Utilising Advanced Analytical Techniques

NBS BAC Hackathon

VSLAM

Ananya Balehithlu | Bai Shun Yao | Max Tan Zheyuan
Tan Kit Hon, Luke | Vinay Krishnaa Vinod

Table of Contents

01
Scope

Situational Analysis & Our Interpretation

Situation:

Whistleblower Fraud Accusation

Cause:

Unsatisfactory current salaries

Effect:

- Alleged '<u>Defrauding</u>'
- Involving accounts payable, employee expenses and the use of corporate credit cards.

Interpretation	Approach
Due to financial constraints and ease of defrauding, employees might engage in fraudulent activities in order to ease their <u>financial struggles</u>	Utilising advanced analytics techniques, comb through accounts payable, expenses and credit card transaction data to identify <u>suspicious activities</u> that render <u>financial advantages</u> to employees → **Develop Risk Profiles**

02
Risk Profiles

Risk Profiles

Disruptive Actors

Invalid entries

Ghost Actors

Fake employees/vendors

Malicious Actors

Insiders with malicious intent

Disruptive Actors

People who make invalid entries. Entries do not agree with data context

Examples of invalid entries include:

- Leaving transaction date empty
- Phone number contains alphabets
- Numeric names

Ghost Actors

Actors that have valid entries, but are not part of the organisation. This occurs mainly due to a lack of proper access control. Examples of such fraud are:

- Bank’s money is flowing out to outsiders
- Outsiders are initiating transactions without the proper authority
- Expenses are being incurred without valid tax codes

Malicious Actors

Actors that are part of the company but have abused their authority to embezzle funds from the bank. Examples of such fraud includes:

- Setting negative value transactions to cover up the previous transactions
- Making transactions during leave
- Making transactions after leaving the company

Malicious actors will always be switching up their methods to cover up for their illegal activities

03
Analysis Process

Original Data

VSLAM analytics

How data was originally organised:

Transaction Data	
PK	**employee_number**
	company
	vendor_location
	vendor_name
	custom_merchant_category
	creditor_merchant_category
	comment
	expense_date
	status
	tax_code
	expense_amount
	net_amount
	tax_amount
	authorised_by
	transaction_id
	country

Leave	
PK	**employee_number**
	position
	dept
	dept_description
	proj_division
	BU
	annual_leave_entitlement
	leave_code
	leave_type
	leave_type_description
	from
	to
	days
	total_no_of_leave_days
	remarks
	BU_entry_date
	employee_type

Employee	
PK	**Employee Master No. INT NO NULL**
	INCOME
	Job Description TEXT NO NULL
	Adm Plan(Junior/Senior/Worker) TEXT NO NULL
	BU Division TEXT NO NULL
	BU Description TEXT
	Date Join DATE NO NULL
	Bank Acct STR NO NULL
	Country TEXT NO NULL
	Gender (F/M) TEXT NO NULL
	Mobile Phone INT
	Home Phone INT
	Address TEXT

Income	
PK	**Employee Master No. INT NO NULL**
	Date DATE
	Division TEXT
	Grade TEXT
	Description TEXT
	Amount $INT
	Payment_Type(Deduction, Earning, Pension
	Payment_Sub_Type TEXT
	Relevant Income (Yes/No) TEXT
	Emplyoment Type(Daily/Monthly) NO NULL

payment	
	document_number varchar(16) NOT NULL
	payment_date DATE NOT NULL
	remarks varchar(256) NOT NULL
	total_amount DEC NOT NULL
	bank_number varchar(16) NOT NULL
	vendor_id varchar(8) NOT NULL
	invoice_id varchar(16) NOT NULL
	source varchar(16) NOT NULL

vendor	
PK	**ven_id varchar(16) NOT NULL**
	vendor_name varchar(32) NOT NULL
	vendor_source varchar(16)
	classification varchar(64) NOT NULL
	purchasing_department varchar(128)
	supplier_receiving_bank_account varchar(32)
	shipping_days int NOT NULL
	contact_details varchar(64)
	address varchar(64)
	country varchar(16)

invoice	
PK	**document_number varchar(16) NOT NULL**
	document_type varchar (64) NOT NULL
	payment_provider varchar(8) NOT NULL
	invoice_date DATE NOT NULL
	document_status varchar(16) NOT NULL
	payment_due_date DATE NOT NULL
	department_name varchar(128) NOT NULL
	currency varchar(4)
	account_provider varchar(16)
	source varchar(16)
	line_of_payable_list BIGINT NOT NULL
	amount_payable_taxed DEC NOT NULL
	amount_payable_untaxed DEC NOT NULL
	product_name varchar(128) NOT NULL
	document_number varchar(16) NOT NULL
	remarks varchar(256)

There is a need to re-organise data for efficient analysis

Normalising Data

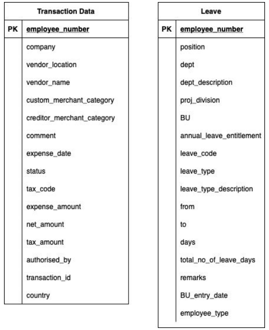
Decomposing Data (Normalisation)

Process

- Identify important attributes
- Remove redundant data

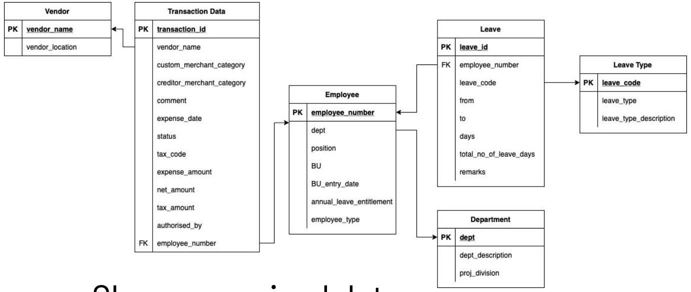
Allows us to:

- Conduct <u>multi-faceted analysis</u> on the same set of data
- Flag out <u>invalid entries</u> (Good starting point for analysis)
- Explicitly present <u>association</u> → decomposition of data does <u>not</u> mean we will lose association



Many columns, Messy

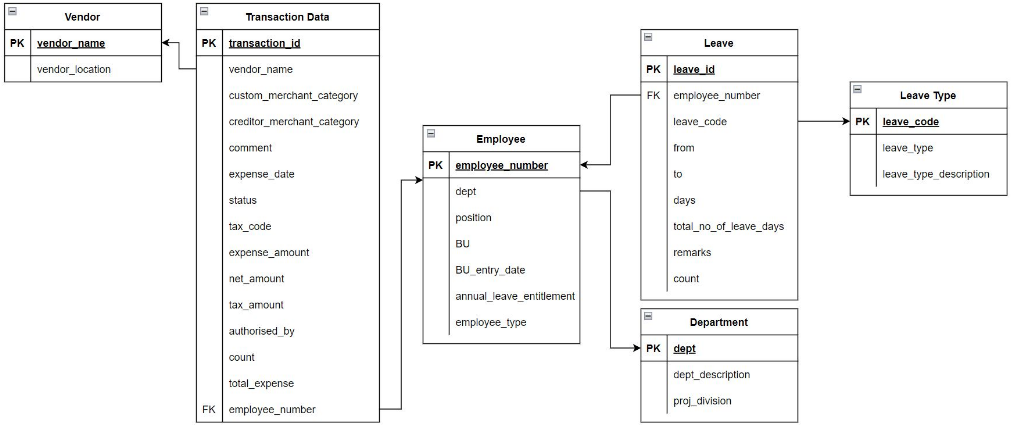
Normalisation

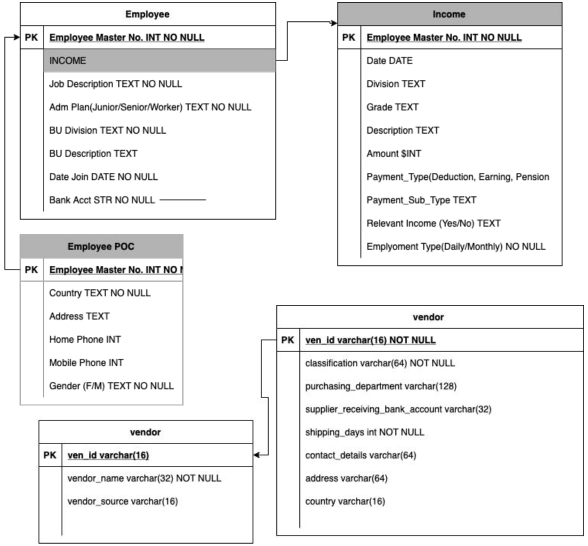


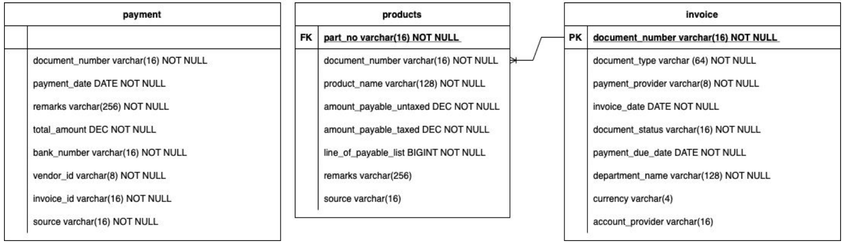
Clean, organised data

Normalising Data

Normalised Data (Re-organised)







Now analysis can be performed on the respective tables.

Scope | Risk Profiles | **Analysis Process** | Outcomes & Insights | Recommendations

Disruptive Actors

Proposed Checks & Rules

1. Invalid transaction IDs (Credit Card Data)
2. Invalid employee numbers (Credit Card Data)
3. Invalid tax codes (Credit Card Data)

Disruptive Actors

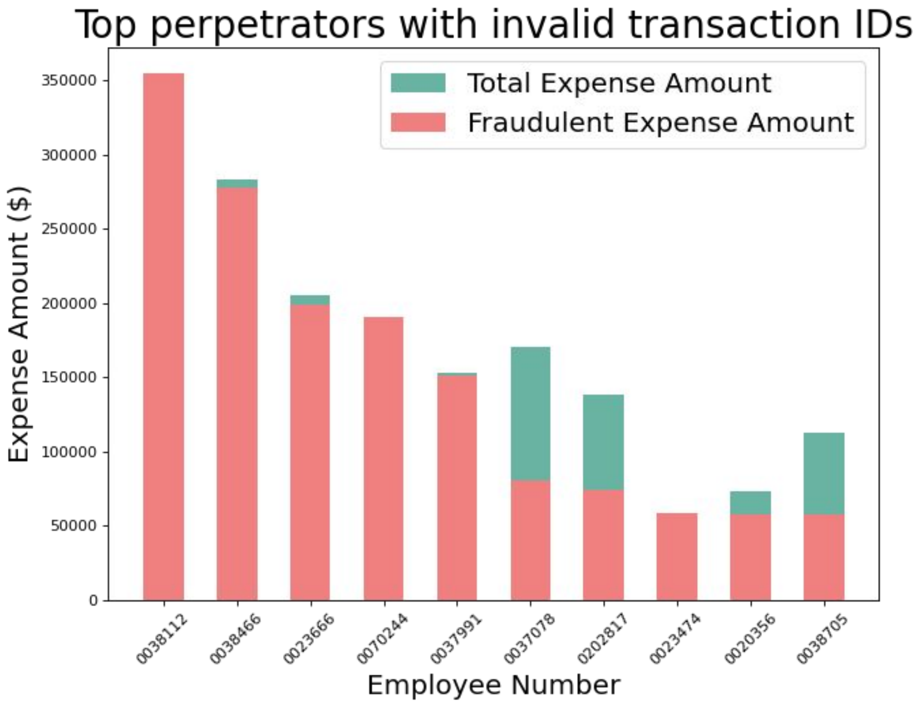
Rule 1: Invalid Transaction IDs

Invalid transaction IDs are:
- Non-numeric
- Zero/NA

→ <u>Flagged</u>
Example: "NA", "Q1", "inv2976"

Invalid Transaction IDs found: **11,342**

Total amount lost: **$ 4,715,186.84**



Scope | Risk Profiles | Analysis Process | Outcomes & Insights | Recommendations

Rule 2: Invalid Employee Numbers

Invalid employee numbers are:
- Non-numeric
- Zero/NA

→ Flagged
Example: "02186A"

3 instances of invalid Employee Numbers

Total amount lost: **$ 1978.28**

employee_number	vendor_name	custom_merchant_category
020186A	DATAWORLD PTY LTD	Telephones and Fax Office
020186A	FINSBURY GREEN PRNTING	Stationery
020186A	FINSBURY GREEN PRNTING	Stationery

Examples of invalid employee numbers from Credit Card Dataset

Rule 3: Invalid Tax Codes

Nature of valid tax codes

- P0, P1 or P2
- Tax code can be left blank if transaction status is "UNSUBMITTED"

No Invalid Tax Codes Found

Ghost Actors

Proposed Checks & Rules

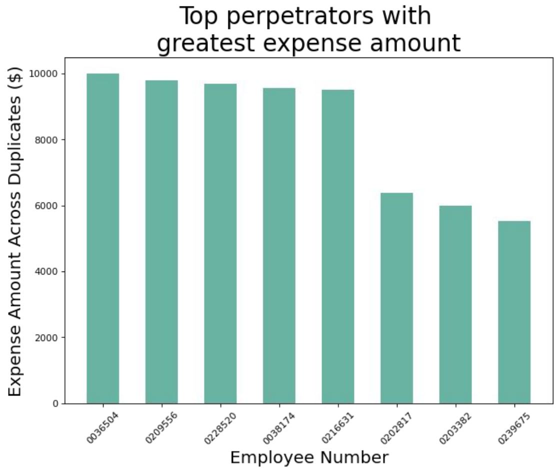
1. Duplicate rows (Credit Card Data)
2. Fake vendors (Accounts Payable Data)
3. Sharing of bank accounts between employees

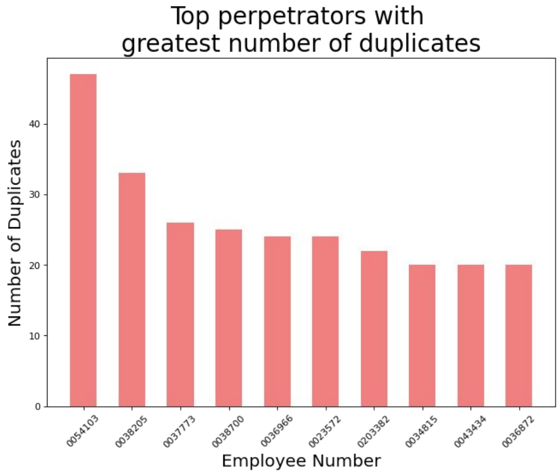
Ghost Actors

Rule 1: Duplicate rows

Rows that are 100% alike in all attributes are indicative of invalid/unauthorised transactions.

The same transactions are charged multiple times. This could indicate cash flow to dummy actors





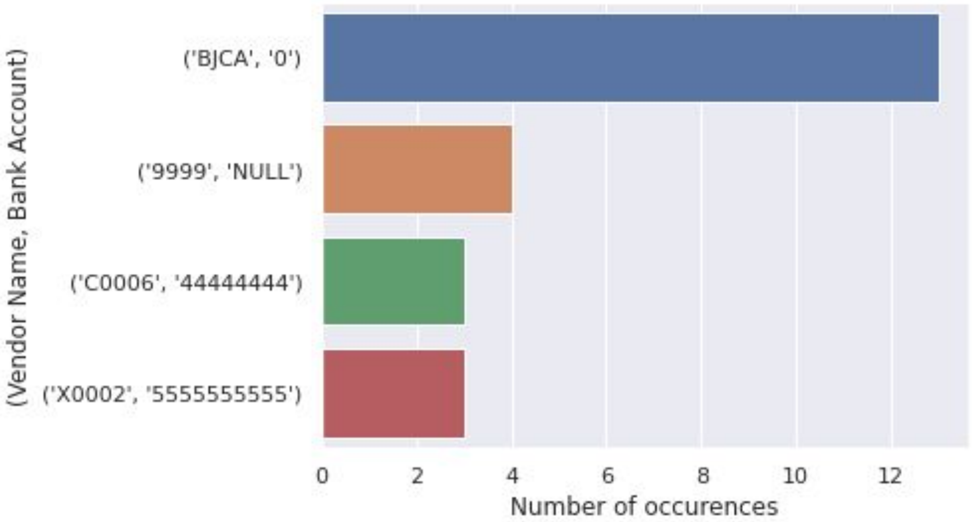
Total Amount Lost: **$ 205,786.48**

Scope | Risk Profiles | **Analysis Process** | Outcomes & Insights | Recommendations

Ghost Actors

Rule 2: Unlikely Vendor Bank Accounts

Some vendor have unlikely bank accounts listed,
suggesting that money could have been siphoned away as false payments .



Rule 3: Sharing of Bank Accounts

4 Employees have been found to share 2 bank accounts. The Employee IDs are:

20186/20186A

- Sharing Employee POC details
- Continuity in work terms between 2 IDs
- Country field from Singapore to Hong Kong
- $776.09 spending as *'186'* and $1978.28 spending as *'186A'* → spread out

Conclusion: Inter-company transfer - Ghost Employee ID

Total Amount Lost: **$ 2744.37**

33876/454690

- Sharing Employee POC details
- No continuity in work terms
- No credit card records as *'454690'*
- Bank account and mobile number are same
- Country field Singapore while address is US

Conclusion: Suspicious Employee with credit card transactions

Total Amount Lost: **$ 11,011.99**

Malicious Actors

Proposed Checks & Rules

1. Employees making transactions during leave (Credit Card Data)
2. Negative amounts (Credit Card Data)
3. Amounts should tally with each other (Credit Card Data)
4. Payments to discontinued/deactivated vendors (Accounts Payable)
5. Overpayments to vendors (Accounts Payable)
6. Embezzlement by employees who have already left Lumbargo

Malicious Actors

Rule 1: Employees making transactions during leave

Credit card transactions are being recorded while employees are on annual leave.
Represents a misappropriation of corporate credit card funds for personal expenses

Invalid transaction if:

- Employee on leave
- Employee makes transaction during leave period
- Duplicated records are aggregated to 1 record loss

Total Amount Lost: **$ 530,097.97**

	employee_number	expense_date	from	to	expense_amount
199403	0226757	2021-11-23	2021-11-19	2021-11-25	226.35
199494	0037076	2021-11-11	2021-11-10	2021-11-13	71.93
199593	0226757	2021-10-11	2021-10-10	2021-10-13	431
199681	0036627	2021-11-12	2021-11-08	2021-11-15	31.29
199737	0037140	2021-09-18	2021-09-17	2021-09-20	622

Examples of employees making transactions during leave

**Rules performed on datasets with disruptive and ghost actors removed*

Malicious Actors

Rule 2: Negative Transaction Amounts

Some transactions have been recorded with negative expense amounts.
These are dangerous as they can hide fraudulent transactions to be viewed as accounting errors.

Identified as negative if:

- Expense amount is negative
- Sum of net and tax amount is negative

Total Amount Lost: **$ 158,065.98**

expense_amount	net_amount	tax_amount	authorised_by
-886.049988	-805.500000	-80.550003	77778648
-537.000000	-488.179993	-48.820000	12374278
-3281.000000	-2982.729980	-298.269989	12362948
-3281.000000	-2982.729980	-298.269989	12359648

Examples of negative amounts

**Rules performed on datasets with disruptive and ghost actors removed*

Rule 3: Amounts should tally with each other

Checking the validity of transactions through accounting formula check.

Valid transactions should adhere to this formula if its status is anything but “UNSUBMITTED”:

Expense amount = net amount + tax amount

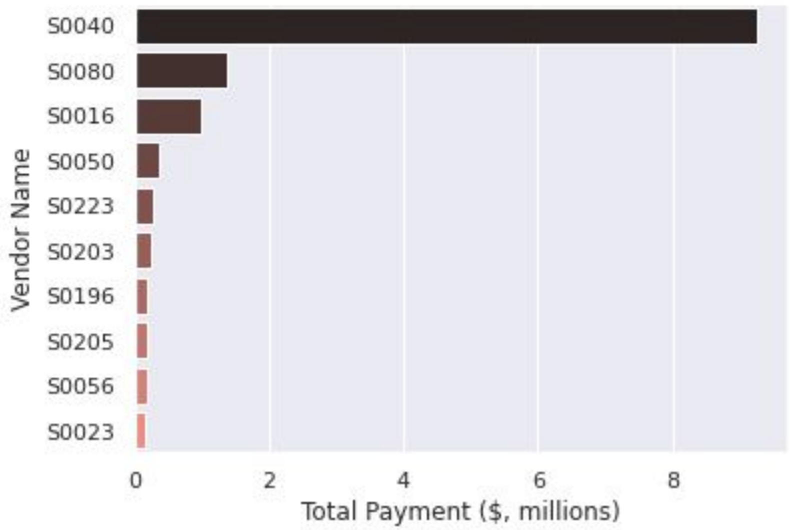
Transactions that do not adhere are flagged out

No findings*

Rules performed on datasets with disruptive and ghost actors removed

Rule 4: Payments to discontinued/deactivated vendors

Some vendors have received payments even though they are listed as discontinued or deactivated.



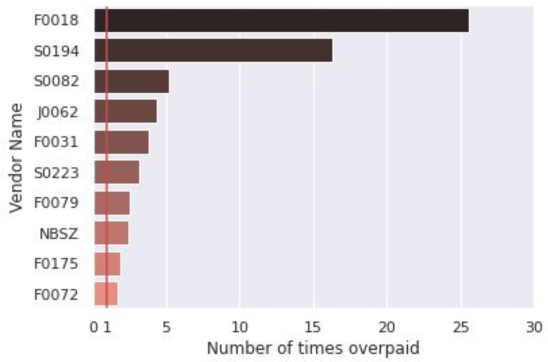
Total Amount Lost: **$ 14,445,195.10**

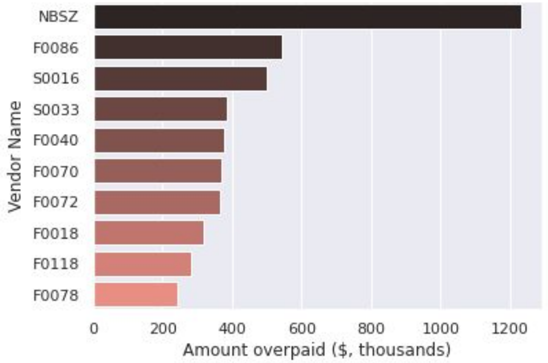
Malicious Actors

Rule 5: Overpayments to vendors

Some amounts that were under Accounts Payable Payments did not tally with the amounts that were owed to those vendors.

Fraud can occur if the bank is paying more than what is owed to the vendors.

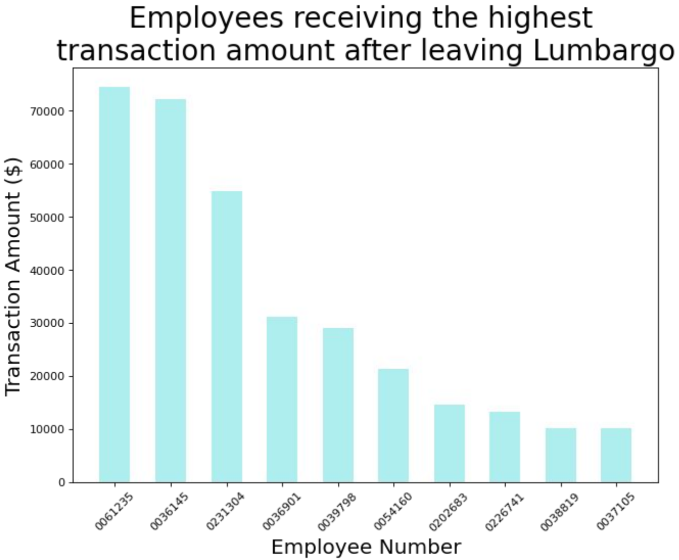


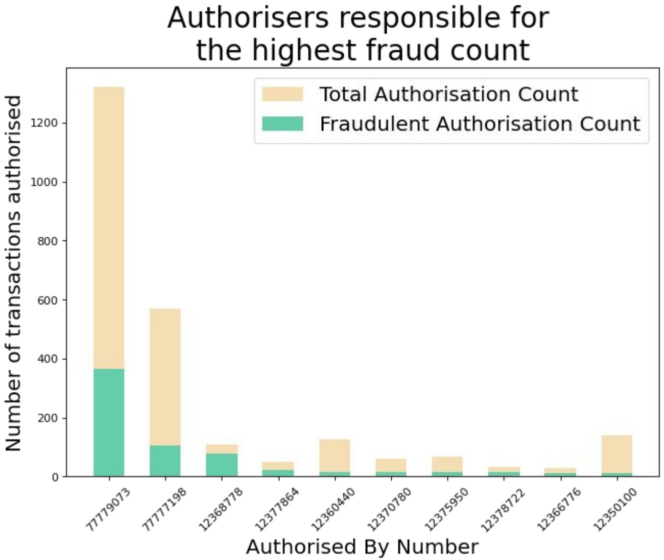


Total Amount Lost: **$ 2,066,166.93**

Rule 6: Embezzlement by employees who have left

Employees are still making transactions despite having already left the company as determined by either the date term specified in the dataset or the contractual end term, if they are hired on the contractual basis.





Total Amount Lost: **$358,596.62**

VSLAM analytics

DBSCAN: Density Based Spatial Clustering of Applications with Noise

Preliminary Stage

<u>Data Loading and Cleaning</u>

- DBSCAN limitations in text analysis - removal of string prevalent columns
- Changing string to int/float data type

<u>Data Processing</u>

- Standard Scaling & Gaussian Normalisation to make cross-variable comparisons

```
In [27]: full_data
Out[27]:
```

	Employee Number	Company	Vendor Location	Vendor Name	Custom Merchant Category	Creditor Merchant Category	Comment	Expense Date	Status	Tax Code	Expense Amount
37684	0036145	3000	WOLLONGONG	REGIONAL EXPRESS BSP	Travel Domestic Airfares	AIRLINES (EXCLUDING THOSE WITH	Durham/Sarah	2021-09-13	SENT_TO_GL	P1	263.60
41964	0056690	3000	WOLLONGONG	CALTEX STAR MART WOL	Travel Domestic Private MV Exp	SERVICE STATIONS	fuel for small plant	2021-09-12	SENT_TO_GL	P1	136.97
29669	0036910	3000	GOULBURN	AUTO ONE GOULBURN PTY	Repairs & Maintenance Motor Ve	MOTOR PARTS, ACCESSORIES STORE	20L PLASTIC 35070, DIESEL CAN	2021-11-15	SENT_TO_GL	P1	92.50
49047	0036679	3000	BEGA	STEEL SUPPLIES BEGA	Minor Equipment Under 5k	MISC & SPECIALTY RETAIL STORES	JN48951, Mild steel angle	2021-11-27	SENT_TO_GL	P1	88.35
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	FIRE DEPTS,	myE-Toll	2021-	[illegible]	[illegible]	[illegible]

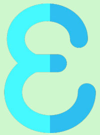
Snapshot of credit card dataset

Model Building - Bivariate Analysis

Model Building Stage

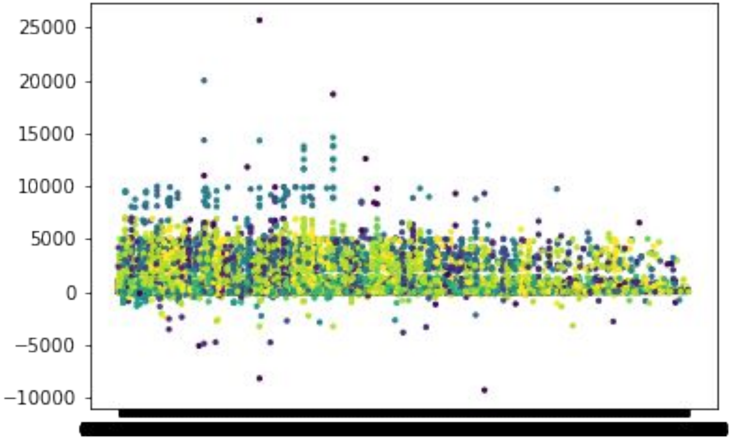
<u>Model Generation on normalised data</u>

- 2 Optimisation Hyperparameters introduced



Epsilon

Minimum Samples



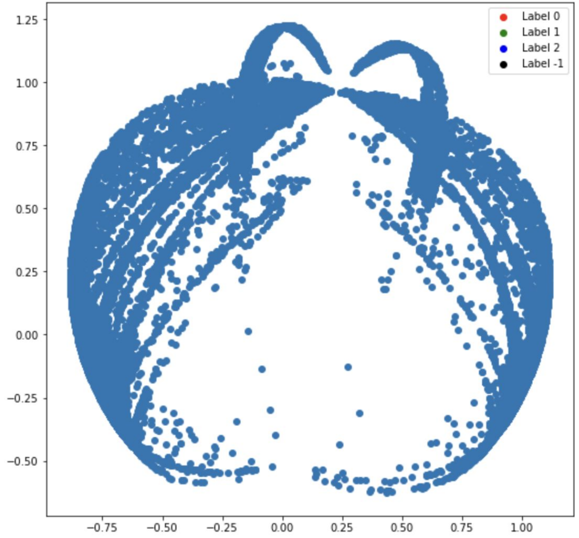
Unoptimized Model Generation - Expense Amount against EmployeeID

Model Building - Multivariate Analysis

Model Building Stage

<u>Model Generation on normalised data</u>

- 2 optimisation hyperparameters created
- Sklearn.decomposition is used to find the SVD of multiple variables
- Plotting of decomposed points in a 2D space for clustering



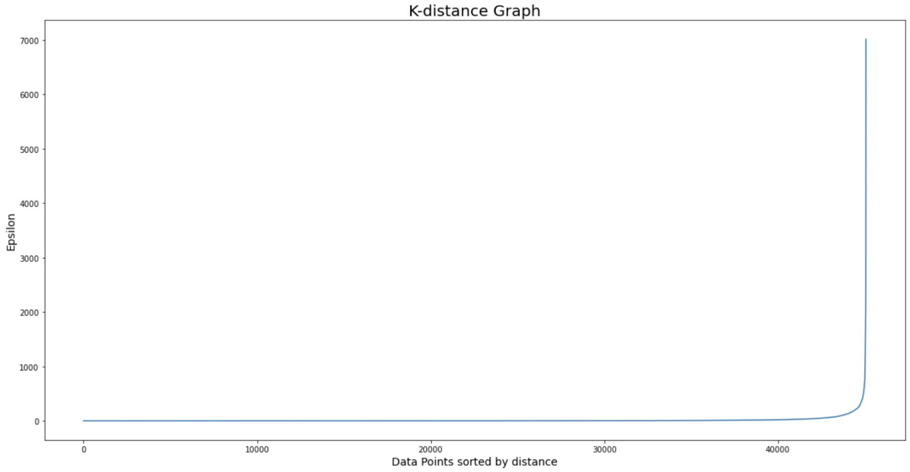
Unoptimized Model Generation - Analysis of all variables

Model Optimisation -Bivariate Analysis

Model Optimisation Stage

<u>Iterative Approach of reducing cost of Model</u>

- Optimisation of Epsilon Hyperparameter
- No optimisation of min. samples required for bivariate analysis
- Finding Epsilon at steepest gradient of K-Means Clustering



Finding Optimised Epsilon - Analysis of all variables

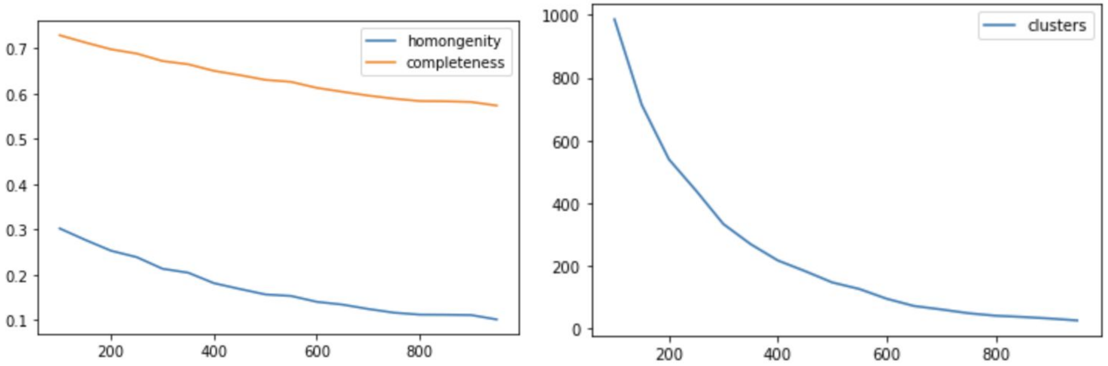
Model Optimisation -Bivariate Analysis

Model Optimisation Stage

<u>Iterative Approach of reducing cost of Model</u>

- Finding optimised Epsilon through 3 Parameters

- Model Completeness
- Model Homogeneity
- Number of Clusters



Comparison of all three Model Optimisation Parameters

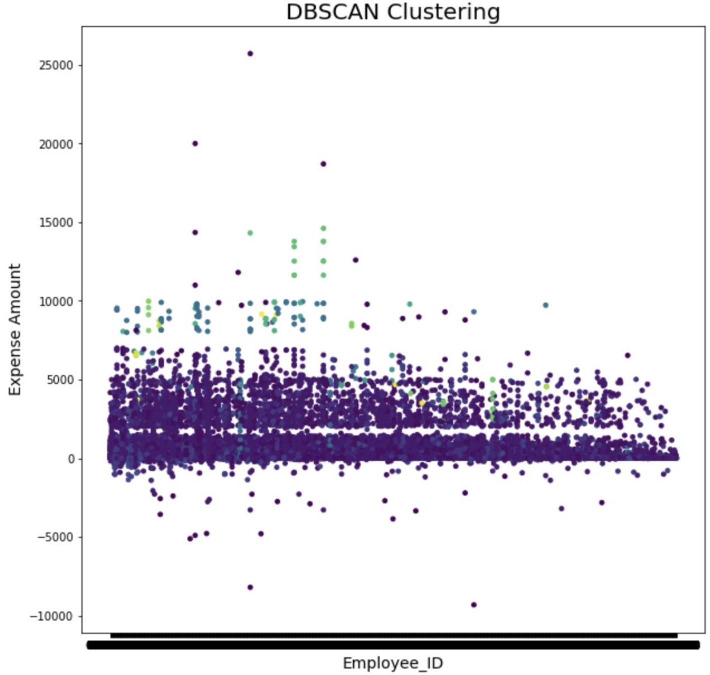
Optimised DBSCAN Model for Bivariate Analysis

Conclusion: Optimized Model Creation

- Optimized Epsilon: 300
- Optimized min. samples: 2

Steps Forward:

- Identification of anomalies through pycaret anomaly detection: prescriptive analysis of fraudulent behaviour



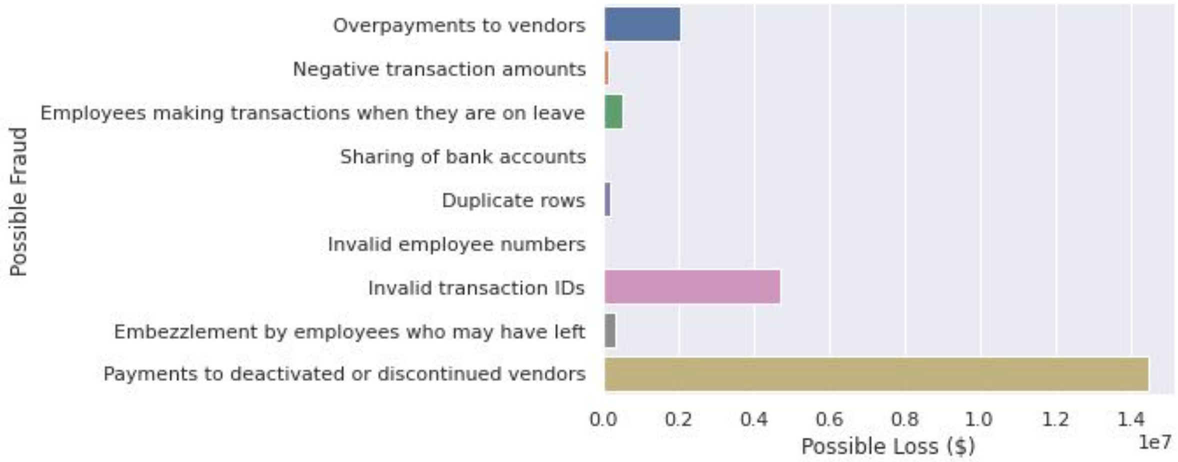
Creation of Model with best clustering parameters

04
Outcomes & Insights

$22,494,831

Suspected Fraud Valuation

Which rules are the most effective?



The rule which has the highest financial impact identified was the invalid transaction IDs under the Credit Card data set, with a total potential loss of 4,715,186.84.

Outcomes

Which risk profile is the most prolific?

The Malicious Actor risk profile is the most prolific, as such actors have the highest number of possible rules. We also observe the highest number of observations where employees are committing fraud by exploiting difficult to notice loopholes such as crediting the company for their transactions even after they have left the company, as well as having their fraudulent transactions being approved by a small group of "fraudulent authorisers."

These risk profiles address the danger of system administrators - such as identified authorisers - for approving large numbers of fraudulent transactions. By taking a identity agnostic approach using Machine Learning models such as DBSCAN, credit card transactions are flagged as fraud by their natural characteristics rather than the approval of other parties.

05
Recommendations

Recommendation

Handling Disruptive Actors

- Disruptive actors exist due to the lack of input control

- **Input control** can be implemented in the following ways:
    - User interface restricts type of input e.g. no alphabets in transaction ID input
    - User interface requires certain inputs to be keyed in for a transaction to be recorded
    - Normalised database implementation will restrict duplicate primary keys e.g. duplicate transaction IDs

Handling Ghost Actors

- **Strong internal controls**
  - Stringent access controls to authorise and verify employees
  - Multiple user access levels i.e. do not give access to employees that they do not need

  Aim: Prevents ghost actors from accessing system

- **Decentralised processes**
  - Actor making transaction should not be able to authorise his/her own transactions

  Aim: Prevents ghost actors from being able to make transactions from one point of access

- **White-listing**
  - List of pre-approved vendors should be updated on a regular basis

  Aim: Prevents money from flowing out to fake/outdated vendors

Handling Malicious Actors

- After sieving out disruptive actors and ghost actors, a more complex approach must be taken to detect malicious actors
- **Rules** developed in this case will be useful in flagging out future similar occurrences
- Increasing sophistication of malicious actors require an **adaptive & constantly evolving approach to fraud detection**
- **Regular data mining & analytics** is required to pick up on new trends and develop new rules
- A **strong whistleblower program** generates good starting points for data exploration and analytics

Thank You

Investigating Fraud Utilising Advanced Analytical Techniques

NBS BAC Hackathon

VSLAM
Ananya Balehithlu | Bai Shun Yao | Max Tan Zheyuan
Tan Kit Hon, Luke | Vinay Krishnaa Vinod