Members of Group 6:
Andra, Joan
Bernice, Lycia
Hui Ling, Shyan-Ee

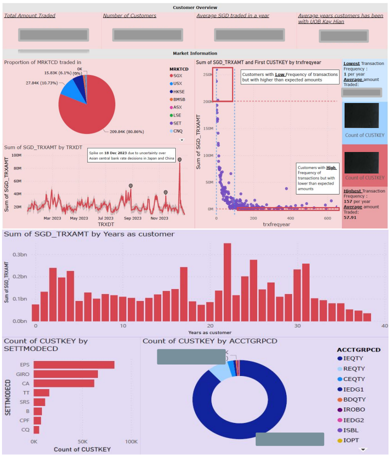
Background Information	UOBKH, a brokerage firm headquartered in Singapore, offering a wide range of financial services including stockbroking, derivatives trading, investment advisory, wealth management, and corporate finance. With a regional presence across Asia, it serves individual investors and institutional corporations.
Project Statement	Due to large voluminous transactions faced by UOBKH and the lack of efficient process in place to flag out irregularities, UOBKH needs a better process to detect anomalies in trading transactions, and identify trends in client's account activity.

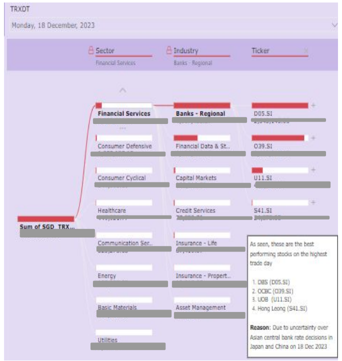
Our Approach

1 Cleaning

- Basic data cleaning & Data reformatting
- Filtering of Trade Receipts Information only to focus purely on the trading transactions
- Creation of Customer Subkey through the date opened variable to identify the number of accounts each customer have
- Addition of Sector and Industry Information extracted from Yahoo Finance for each trading transaction

2 Exploratory Data Analysis (EDA)

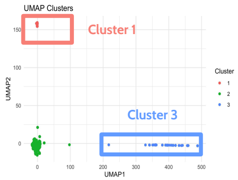




Dashboards were created to illustrate an overview of customers and transactions.
Main takeaway: Anomalies detected for high frequency and low amount transactions & low frequency and high amount transactions as seen from outliers

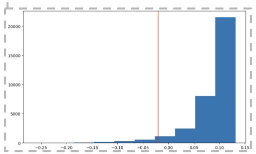
3 Customer Profiling

UMAP Clustering



Unsupervised clustering was performed on customer data to flag out any out-of-the-norm customer groups. Cluster 2 contained most of the customers while cluster 1 & 3 contained a small number of customers which we then flagged out as anomalous.

Isolation Forest

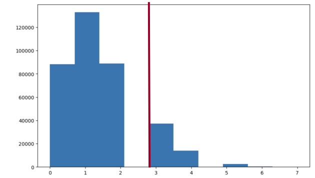


Dataset size: 34,291
Anomalies: 1,189
Percentage of anomalies: 3.47%

4 Classification Models/ Regression

Labelled data based on a rules-based scorecard
1. Check for Customer Suspension
2. Check for Transaction Frequency
3. Check for Transaction Amount
4. Check for Settlement Mode RIsk
5. Check for Sector Risk
6. Check for Security (company) Risk
7. Check for Customer Cluster

Based on histogram, we labelled all transactions with scores more than a set cutoff score as anomalous. 2 datasets were created with cutoff scores of 3 and 4.



After labelling, classification trees and logistic regression models were run on the labelled data.

Sensitivity	Random Forest	Logistic Regression	Lasso (lambda.1se)	XGBoost
Dataset 1	0.614	0.4484	0.406	0.662
Dataset 2	0.742	0.5449	0.5386	0.718
Dataset 3	0.115	0.0549	0.0344	0.133
Dataset 4	0.452	0.3331	0.2904	0.468

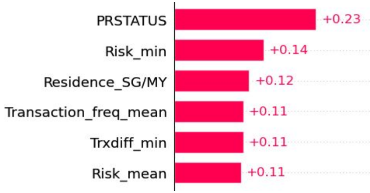
Based on classification trees and regression models, we extracted rules and significant variables and compared them with the rules set previously to see if there are any differences/similarities.

Insights

What can UOBKH gain?
By harnessing machine analysis of 2023 Transaction Receipts, UOBKH gains a novel perspective on anomalies detected by the machines. This enables potential modifications to enhance their own anomaly detection capabilities, thereby adding value to existing processes.

What have we learnt ?
Our approach extends beyond unsupervised learning to include supervised learning, which involves synthetically labeling the data. This expansion enhances the project by allowing for a more comprehensive analysis and deeper insights, particularly with the incorporation of customer-related data.

Variable Importance



Top features that indicates higher chance of anomalies for customer profiles.

It is interesting to note that transaction amount is not a leading variable in discerning anomalies.