

**Decode-X 2025**

**Team Name:** Debug Dynasty

**Team:** Kaushal Tuntune

Atif Ansari

Vaibhav Boudh

**Contents:**

|  |  |
| --- | --- |
| **Introduction** | **2** |
| **1 Business Objective**1.1 Background1.2 Business Case | **2** |
| **2 Project Management** 2.1 Work Distribution2.2 Project Plan2.3 Key Risks | **4** |
| **3 Data Preparation and Availability** 3.1 Data Preparation3.2 Data Availability3.3 Expansion and Explanation3.4 Methodology | **6** |
| **4. Proposed Approach for Detecting and Preventing Circular Trading**4.1 Data Analysis4.2 Daily Count of Anomalous Trades | **10** |
| **5. Conclusion and Findings** | **16** |
| **6. References** | **17** |

**Introduction:**

In the bustling global financial district, Global Stock Exchange, under CEO Jonathan Harris's leadership, commands attention as a pivotal hub for financial trading. However, recent murmurs of concern regarding front running—a disreputable exploitation of advanced trade knowledge—have clouded the exchange's reputation. Recognizing the urgency of the situation, Harris hastily convened a meeting with key stakeholders, including Sarah Martinez, Head of Compliance, and Emily Chang, Chief Data Scientist. Together, they faced the daunting task of addressing front-running threats to preserve the exchange's integrity and investor trust. Amidst the dialogue, a shared determination emerged to leverage innovative solutions and collaborative expertise to combat illicit activities. As the exchange braces against these challenges, it reaffirms its commitment to transparency, regulatory compliance, and market stability. In the face of uncertainty, Global Stock Exchange stands resolute, ready to navigate the complexities of the financial landscape with integrity and resolve.

1. **Business Objective**
   1. [**Background**](https://docs.google.com/document/d/1aNs6JZHElUxxOReII-lmS5PBJa9Kw0G5/edit#heading%3Dh.30j0zll)**:**

One of the primary objectives of the highly dynamic and quite complicated financial markets, of today, is the assurance of the purity of trade operations which is necessary for maintaining market efficiency and investor trust. In the past few years, the concerns have been becoming more and more about the practices of fraudulent trading that draw off the transparency of the market. If the abuse is not addressed correctly, it can grow into significant scandals and pose risks to the financial system stability. As such, preventing abusive traders from hobbling the price mechanism has long been a set the market operator(s) are interested in mastering.

Circular trading is a fraudulent trading strategy that involves parties working together to create the illusion of high trade volumes, and then sell at a higher price in a non-transparent market, thereby misleading the market. Traders who engage in circular trading use false trade transactions to generate trade volume without the need to involve any real capital, thereby contradicting the very essence of trade that is supposed to be the exchange of goods or services for money. This illusionary maneuver signifies the superficial and deceptive nature of this regulation in the capital market.

To a primary global stock exchange, abusive tactics can erode trust and result in significant financial and reputational damage. As a result, the development of methodologies that can trust the analytical processes for the detection and flagging of these cases can be obtained. Through the application of advanced data mining technologies, statistical anomaly detection, and network analysis, this project, therefore, is to establish a stronger fraud detection system. This system is devised to identify strange trading behaviours deliberately—for example, circular trading in order for the market to remain uncorrupted through facilitating intervention on time.

Indeed, the product is a solution that is oriented primarily to the prevention of stock exchange behaviour and secondly to the supply in time of the needed information to the compliance officers and regulators. The main purpose of the solution is to prevent illegal activities in the financial market. The shared solution has two parts, one is related to the detection of the problem and the other provides measures of resistance to change.

**1.2 Business Case**

The impetus for this project stems from the significant financial and reputational risks associated with fraudulent trading practices, particularly circular trading, in the stock market. When collusive behaviour manipulates trading volumes and prices, it distorts market signals, undermines transparency, and erodes investor confidence. This not only leads to decreased participation in the market but can also trigger regulatory interventions that further impact revenue and trust.

By developing a robust fraud detection system tailored to identify circular trading patterns, the stock exchange can proactively mitigate these risks. The benefits include safeguarding the integrity of market data, preventing artificial inflation of trade volumes, and ensuring that pricing remains reflective of true market demand. Moreover, a reliable detection framework enhances regulatory compliance, reduces the likelihood of costly legal disputes, and preserves the exchange’s reputation as a fair and transparent trading platform.

In essence, this project offers a strategic business advantage by enabling timely detection and intervention, thereby protecting investor interests and supporting sustained market growth. The proposed solution not only addresses immediate concerns but also lays the groundwork for long-term improvements in market surveillance and operational resilience.

1. **Project Management**

**2.1 Work Distribution**

|  |  |
| --- | --- |
| **Responsibilities** | **Team mate** |
| Research  Data Visualization - Power BI, Excel Report - Google Docs  Presentation - PPT | Kaushal |
| Research  Report - Google Docs  Data Interpretation & Visualization - Power BI | Atif |
| Data Cleaning Data Modelling  Algorithm Training & Analysis | Vaibhav |

**2.2 Project Plan**

As mentioned above, in section 2.1, responsibilities like Data Modelling, Algorithm Training, Data Visualization, Report making and Presentation making were distributed amongst the teammates. Here is the tabular breakdown of plan for creating the final product:

|  |  |
| --- | --- |
| **Phases** | **Activities** |
| Data Cleaning | Handling duplicate data Handling inconsistent data Handling data transformation Column Sanitization |
| Data Interpretation | Descriptive statistics Data visualisation Correlation analysis Inferential statistics |
| Algorithm Training | Algorithm Selection Training Data Preparation |
| Algorithm Testing | Test Data  Testing Performance  Evaluation  Fine-tuning |

**2.3 Key Risks**

This project, while addressing the issue of curative trading and cheating activities, which has a comprehensive approach to the detection of circular trading and fraudulent activities, on the other hand, faces some of the most crucial risks that must be managed if it is to succeed and be reliable in the end:

* Data Quality and Integrity: One of the major factors that influence the detection of fraud is the purity of the data and also the accuracy of its completeness, as every model is not 100% accurate and cannot be completely reliable if there is incomplete or lacking data. Inconsistent, missing, or wrong entries—this is particularly bothersome in fields such as timestamps, trade volumes, and order details—can bring additional complications, such as inducing the result in misleading fashion or increasing the rate of false signals. One activity that is necessary is the conduction of both data cleaning and validation procedures.
* Model Risk and Accuracy: Using a number of approaches, i.e. statistical outlier detection, machine learning (Isolation Forest), and networking analysis presents the model risk. Each model bears certain drawbacks, and the improper setup of them can cause either the missing of fraud (false negatives) or the flagging of legal exits as questionable (false positives). We must have congruent balances in the relation of sensitivity and specificity since this is the only way we can safeguard the potential trust of people in systems.
* Complexity of Fraud Schemes: Circular trading, a matter of complex nature, can facilitate the use of various sophisticated and flexible strategies that may change over time. First and foremost, the data analyses must be modified to capture brand new patterns. The danger still exists that subjects of static models who are changing it may render it less effective, and this will require ongoing updates and monitoring models then.
* Integration and Implementation Risks: The integrating of the detection system along with existing market surveillance and compliance workflows implies the facing of difficulties. The differences in data formats, system compatibility issues, and real-time data processing are some of the issues that could affect the solution’s operational performance.
* Regulatory and Compliance Risk: The handling of financial information and the possibility of fraud detection are strictly confidential issues and any system issues in the sector would result in compliance checks. The solution needs to be capable of reliably detecting, displaying, and documenting all the patterns that are qualified to follow as well as be able to produce a clear audit trail and documentation to establish the validity of any further investigations.
* Stakeholder Buy-In and Operational Impact: The project’s success comprises the willingness and competent implementation of the solution by the regulatory and compliance teams. Misrepresentation of the raised oddities and a lot of false alerts may cause a weak operation and this is the way that can erode confidence in the detection system.

Through proactive management of the above risks, design of the most flexible systems, and unambiguous communication with stakeholders either through a web survey, direct communications, social network, or other means the project can avoid the potential pitfalls and guarantee a robust, actionable fraud detection system which can be relied upon based on its output’s accuracy and confidence.

1. **Data Preparation and Availability**

**3.1 Data Preparation**

Our team received two essential datasets – the orders dataset and the trades dataset – which are critical for detecting fraudulent activities, particularly circular trading, in the stock market. Recognizing that the accuracy of our analysis hinges on the quality of the data, we implemented a rigorous cleaning and standardization process. This involved:

* **Standardizing Column Names:** Removing any extra spaces and converting all column headers to lowercase to ensure consistency.
* **Converting Date and Time Fields:** Transforming string-based dates (e.g., "13-Apr-07") into datetime objects and ensuring time fields (such as order\_time and trade\_time) are properly formatted as time objects.
* **Handling Missing Data:** Identifying and addressing gaps or anomalies in the data, thereby ensuring that our subsequent analysis is built on a reliable foundation.

This thorough data preparation allowed us to accurately merge and compare information across datasets, forming the basis for our advanced analytical models aimed at exposing irregular trading patterns.

**3.2 Data Availability**

**3.2.1 Order Dataset**

The order dataset, commonly referred to as the order book, provides a comprehensive record of all orders placed during the trading period. With 602,473 observations and 14 columns, this dataset includes critical details such as:

* **Unique Identifiers:** ORDER\_SEQUENCE, ORDER\_ID
* **Timestamps:** ORDER\_TIME, ORDER\_DATE
* **Security Identification:** SCRIP\_CODE
* **Participant Details:** MEMBER\_CODE, CLIENT\_ID
* **Transaction Type:** BUY\_OR\_SELL
* **Pricing and Quantities:** RATE, QUANTITY
* **Operational Details:** TRADER\_ID, TERMINAL\_ID, LOCATION\_ID

These attributes offer deep insights into the order placements, which are essential for understanding the initial intent behind each trade.

**3.2.2 Trade Dataset**

The trade dataset, or trade book, captures the execution details of orders, recording 414,957 observations across 20 columns. It contains vital information including:

* **Unique Identifiers:** TRADE\_SEQUENCE, TRADE\_NUMBER
* **Timestamps:** TRADE\_TIME, TRADE\_DATE
* **Security Identification:** SCRIP\_CODE
* **Participant Codes:** BUY\_MEMBER\_CODE, SELL\_MEMBER\_CODE
* **Client Information:** BUY\_CLIENT\_ID, SELL\_CLIENT\_ID
* **Order References:** BUY\_ORDER\_ID, SELL\_ORDER\_ID
* **Trading Personnel:** BUY\_TRADER\_ID, SELL\_TRADER\_ID
* **Execution Metrics:** TRADE\_QUANTITY, TRADE\_RATE, TRADE\_VALUE
* **Location Information:** BUY\_LOCATION\_ID, SELL\_LOCATION\_ID
* **Execution Timestamps:** BUY\_TIMESTAMP, SELL\_TIMESTAMP

This dataset is indispensable for analyzing how orders are executed, revealing the dynamics of trade volumes, pricing, and the interactions between market participants.

**3.3 Expansion and Explanation**

The order dataset provides a holistic view of the orders submitted, offering insights into who is placing orders, at what prices, and with what intentions. Conversely, the trade dataset details the actual market transactions, capturing the execution details that reveal the real trading behaviour. By merging these datasets, we can bridge the gap between order intent and trade execution.

Our analysis leverages this integrated data to detect irregularities indicative of circular trading. Techniques such as statistical outlier detection, network analysis, and machine learning are applied to identify:

* **Anomalous Trade Volumes:** Using Z-scores, refined IQR, and MAD-based methods.
* **Repeated Trading Patterns:** Identifying frequent reciprocal trades between the same buyer–seller pairs.
* **Network Structures and Cycles:** Constructing directed graphs to visualize relationships and detect cycles—a hallmark of collusion.
* **Temporal Spikes:** Aggregating trades into daily time series to pinpoint periods of unusually high trading activity.

Together, these approaches enable us to systematically narrow down a massive dataset into targeted segments of suspicious activity. This process provides compelling, multi-faceted evidence that supports further investigation into potential circular trading practices, ensuring that market surveillance is both comprehensive and precise.

This comprehensive explanation of our data preparation and availability establishes a strong foundation for our subsequent analyses. It demonstrates how the raw data is transformed and utilized to uncover critical patterns, making our solution both robust and uniquely tailored to the problem of detecting circular trading in the stock market.

**3.4 Methodology**

Our approach to detecting circular trading fraud integrates multiple analytical techniques that together form a robust and layered solution. This methodology is organized into several key phases, each designed to incrementally build evidence and refine our detection capabilities.

**3.4.1 Define the Problem and Approach for Detecting Circular Trading**

This deceptive practice encompasses a group of traders or accounts trading among themselves, hence termed circular trading. This collusive activity artificially inflates trading volume and may manipulate market price, creating a false appearance of market liquidity and activity. Such activities undermine market integrity, mislead investors, and can result in significant losses and regulatory penalties.

Tackling this issue is multi-pronged and systematic. We begin with data preparation, involving cleaning and standardizing the orders and trades data sets, where dates, times, and other critical fields need to be accurately formatted. This first step lays a vital foundation for our analysis since high-quality, consistent data is paramount in analysis.

**3.4.2 Data Preparation and Cleaning**

The foundation of our analysis is high-quality data. We begin by ingesting two primary datasets—orders and trades—and standardizing their structure. This involves:

* **Standardizing Column Names:** Removing extraneous spaces and converting all column names to lowercase to ensure consistency and avoid key errors.
* **Converting Dates and Times:** Transforming string representations of dates and times into standardized datetime and time objects. This step is critical as it allows for precise temporal analysis and accurate merging of datasets.
* **Handling Missing or Erroneous Data:** Applying validation and cleaning techniques to ensure the integrity of key fields such as order dates, trade times, and quantities.

**3.4.3 Exploratory Data Analysis (EDA) and Feature Engineering**

With clean data in hand, the next phase is to understand the underlying patterns:

* **Statistical Analysis:** We perform exploratory analysis on trade quantities and values using descriptive statistics and visualizations. This helps to identify initial anomalies or irregularities.
* **Time Series Analysis:** By constructing a unified datetime column for trades and resampling the data on a daily basis, we can detect abnormal spikes in trading activity—an early indicator of potential fraudulent behaviour.
* **Feature Engineering:** New features are created to capture aspects such as time differences between order placements and trade executions, frequency of trades between specific buyer–seller pairs, and rolling averages of trade volumes. These engineered features serve as critical inputs for subsequent anomaly detection.

**3.4.4 Multi-Method Anomaly and Pattern Detection**

Our detection framework integrates various statistical, machine learning, and network analysis methods for the identification of suspicious patterns that might indicate circular trading.

* **Statistical Outlier Detection:**
  + **Z-Score and IQR Methods:** The Z-scores of trade quantities were computed and flagged for those trades whose quantities were significantly above the average. Besides that, a refined interquartile range (IQR) was used with an increased multiplier to eliminate noise and better isolate true anomalies.
* **Machine Learning Approaches:**
  + **Isolation Forest:** An Isolation Forest was trained on certain key numerical features (such as trade quantity, trade rate, and trade value) to pinpoint subtle anomalies that might not show up by means of statistical methods alone.
* **Pattern Recognition and Network Analysis:**
  + **Repeated Trading Pair Analysis:** Trades are grouped by buyer, seller, and scrip code to find pairs repeatedly trading with each other, a classic sign of circular trading.
  + **Network Graph Construction and Cycle Detection:** A directed network graph was built from trading members as nodes and trading alliances weighed by volume as edges. More advanced techniques—like cycle detection and strongly connected component (SCC) analysis—were used to find closed loops in trading, which are very strong signals of collusion.

**4. Proposed Approach for Detecting and Preventing Circular Trading**

In response to the growing threat of circular trading, our solution harnesses a multi-layered analytical framework designed to identify and mitigate this fraudulent activity effectively. Our approach integrates advanced data analytics, feature engineering, machine learning, and network analysis to deliver a comprehensive detection system. Below is an outline of our proposed methodology:

**Advanced Data Analytics and Feature Engineering:**

Our solution begins with an in-depth exploration of the orders and trades datasets. We conduct thorough exploratory data analysis (EDA) to understand the inherent characteristics and variability in the data. This phase includes:

* **Data Exploration:** We examine distributions, summary statistics, and trends in trade quantities, prices, and timings to reveal initial anomalies.
* **Feature Engineering:** Recognizing that circular trading often manifests as repeated transactions between the same entities, we develop key features such as:
  + **Repetition Frequency:** Calculated by grouping trades by buyer, seller, and security code to flag pairs engaging in frequent reciprocal transactions.
  + **Temporal Patterns:** Metrics like the time gap between consecutive trades and order-to-trade delays help pinpoint rapid-fire transactions that deviate from normal market behaviour.
  + **Volume and Price Impact:** Derived variables measure abnormal trade sizes and discrepancies in transaction prices, serving as early warning signals of potential manipulation.

These engineered features enhance the discriminative power of our detection models, enabling a more precise isolation of suspicious activities.

**Machine Learning Anomaly Detection:**

To complement our statistical methods, we employ machine learning algorithms—most notably, the Isolation Forest—to detect outliers in a multi-dimensional feature space. By training the model on key indicators such as trade volume, rate, and value, we can capture subtle anomalies that may escape traditional statistical tests. The integration of machine learning not only increases our detection accuracy but also provides a probabilistic anomaly score, which is instrumental in ranking suspicious activities for further scrutiny.

**Network Analysis and Cycle Detection:**

Circular trading inherently involves a closed loop of transactions among colluding parties. To capture this phenomenon, we construct a directed network graph where each node represents a trading participant, and edges denote trading relationships weighted by trade volume. Within this network:

* **Repeated Trading Patterns:** We identify clusters of buyers and sellers who transact repeatedly.
* **Cycle Detection:** Utilizing algorithms like simple cycle detection and strongly connected component analysis, we uncover cycles within the network—an unequivocal indicator of circular trading.
* **Centrality Measures:** By calculating metrics such as degree centrality, we highlight key participants whose repeated interactions suggest orchestrated collusion.

This network-based perspective not only visualizes the structure of trading relationships but also strengthens the evidence of collusive behaviour through the identification of closed loops.

**Preventive Measures and Continuous Monitoring:**

The insights generated by our detection framework empower regulatory and compliance teams to:

* **Proactively Intervene:** Identify and investigate high-risk trading clusters before they escalate into systemic issues.
* **Enhance Surveillance:** Incorporate real-time monitoring tools that continuously feed into our analytical pipeline, ensuring that emerging patterns of circular trading are promptly flagged.
* **Support Regulatory Compliance:** Maintain an auditable trail of detections and interventions, reinforcing transparency and adherence to regulatory standards.

Through this comprehensive approach, our solution not only detects circular trading with high accuracy but also provides a scalable framework for continuous improvement—ensuring that the market remains transparent, fair, and resilient against fraudulent activities.

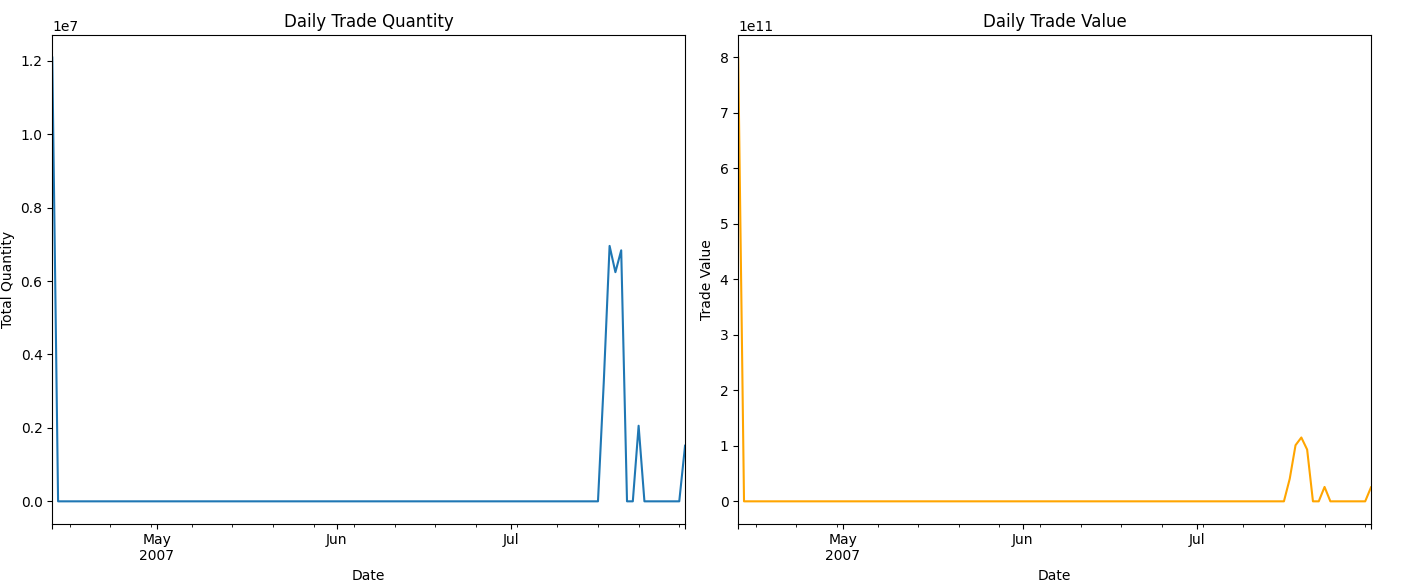
This proposed approach, by integrating advanced analytics, machine learning, and network analysis, forms a holistic framework designed to detect and prevent circular trading fraud effectively. Its multi-dimensional nature makes it unique and robust, meeting the stringent requirements of market surveillance and regulatory compliance.

**4.1 Data Analysis:**

**4.1.1 Data ingestion and cleaning :**

Through this data ingestion and cleaning step, all inconsistencies in naming conventions and date/time formats are resolved, and missing or erroneous fields are addressed. As a result, the dataset becomes uniformly structured and reliable, laying a strong foundation for the advanced analyses and anomaly detection processes that follow.

**4.1.2 Time-Series Analysis of Trading Activity:**



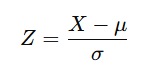
This figure presents two line charts side by side—one showing the daily aggregated trade quantity and the other showing the daily trade value. The charts illustrate trends over time, highlighting any significant spikes in trade volume or value that may indicate unusual trading activity, which could be a sign of circular trading manipulation.

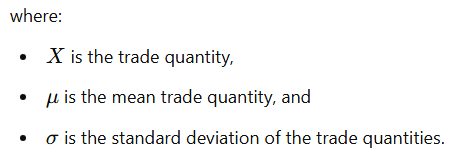
The line charts depict the aggregated trading data, allowing us to observe fluctuations in trade quantity and value throughout the period. Notable spikes, especially on certain dates such as mid-April or mid-July (as seen in this example), may signal potential manipulative activities because large, abnormal surges in trading can be linked to orchestrated efforts like circular trading. By identifying these spikes, we can narrow down specific days for further investigation into irregular trading patterns. This time-series analysis helps target timeframes where suspicious activities are more concentrated, making it an effective tool for uncovering fraudulent behaviours. for 5 seconds.

These time-series plots of daily aggregated trade quantity and value highlight overall trading patterns over the examined period. Noticeable spikes appear in mid-April and mid-July, suggesting potential anomalies or irregular trading behaviours—possibly indicative of circular trading. By pinpointing these unusual peaks, investigators can narrow their focus to specific dates for deeper scrutiny, thus enhancing the efficiency and precision of further fraud detection efforts.

**4.1.3** **Z-Score (Standard Deviation–Based Detection)**

* **Definition:** The Z-score measures how many standard deviations a data point is from the mean. Formally, it is defined as:

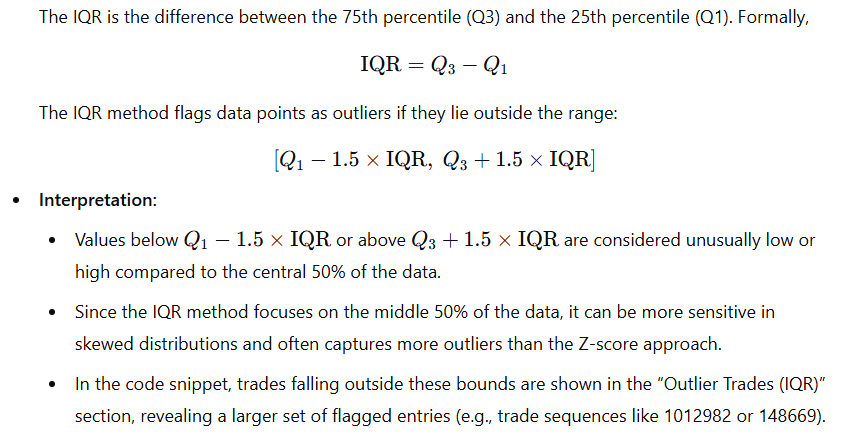




* **Interpretation:**
  + A high Z-score (e.g., > 3) indicates that the trade quantity is much larger than the average, suggesting it may be abnormally large and thus “suspicious.”
  + In the snippet, trades with volume\_zscore > 3 are flagged, yielding entries like trade\_sequence = 1126254 or 617649, which deviate significantly from normal market behaviour.

**4.1.4** **IQR (Interquartile Range) Method**

* **Definition:**



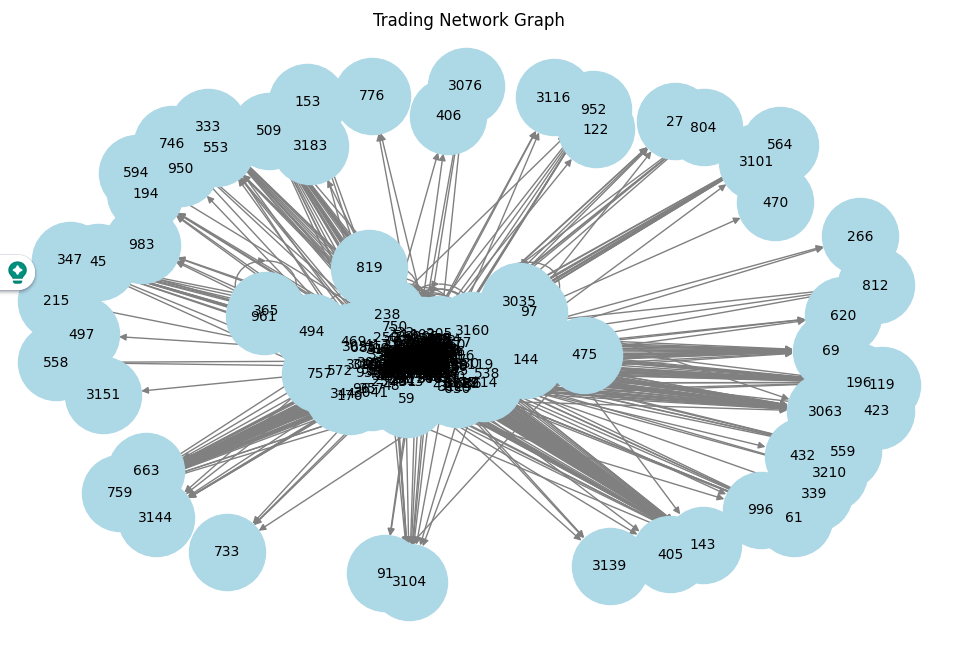
**Why These Methods Are Important:**

1. **Initial Screening for Suspicious Trades:** Both Z-score and IQR help highlight trades with abnormally high (or low) quantities. While a single outlier does not confirm fraud, it signals that the trade warrants closer inspection.
2. **Complementary Approaches:**
   * **Z-score** relies on the mean and standard deviation. It works well when data is roughly bell-shaped but may underestimate outliers in heavily skewed distributions.
   * **IQR** is more robust to skewed data and extreme values but can over-flag points if the distribution is very wide.
3. **Foundation for Further Analysis:** Once these outliers are identified, we can cross-reference them with other indicators (e.g., repeated buyer–seller pairs, network centrality, or ML-based anomaly scores) to build a stronger case for potential circular trading or other fraudulent activity.

**In Practice:**

* The snippet shows that **Z-score detection** flagged fewer trades (e.g., 858) because it specifically targets values that lie more than three standard deviations from the mean.
* The **IQR method** flagged a larger subset (e.g., 49,108 trades) because it can be more sensitive in distributions with a long tail or heavy skew.
* These flagged entries (sampled in the output) reveal trade sequences, member codes, and volumes that deviate significantly from normal market behaviour. Investigators or compliance teams then focus on these “outlier trades” for additional context—such as repeated patterns or collusive networks—before concluding if the trades are indeed fraudulent.

**4.1.4 Trading Network Graph diagram:**



This figure shows a network of trading participants (nodes) and the connections between them (edges). Each node corresponds to a buyer or seller, while each edge represents the flow of trades between two parties, weighted by the volume or frequency of their transactions. The layout visually clusters nodes that frequently trade with each other, enabling us to pinpoint potentially collusive groups.

**Central Clusters:** Notice the dense grouping in the middle. These highly interconnected nodes frequently exchange trades among themselves—often a red flag for collusion or circular trading.

**Peripheral Nodes:** Some participants (nodes) lie at the edges with fewer connections, suggesting they trade less frequently or only with a small subset of others. Such nodes may be less suspicious unless they exhibit other risk factors (e.g., large anomalous trades).

**Identifying Collusive Loops:** If you see circular patterns or small “loops” where a group of nodes trade only among themselves, it can be indicative of market manipulation (e.g., artificially inflating volumes). Investigators can focus on these loops for deeper scrutiny.

**Node Size or Color (If Applied):** In some versions, node size or color might reflect metrics like degree centrality or anomaly scores. Larger or more intensely colored nodes may indicate higher involvement in suspicious trades.

Overall, this network visualization helps compliance teams see potential “hotspots” where repeated trades might point to collusion, making it easier to prioritize further investigation.

**4.2 Daily Count of Anomalous Trades:**

This line chart plots the number of trades flagged as anomalous by our detection methods (e.g., Isolation Forest or statistical outlier detection) on each calendar day. The x-axis represents time, while the y-axis shows the daily total of detected anomalies.

* **Spikes and Trends:** The most striking feature is the sharp drop from mid-April to early June, followed by a smaller cluster of anomalies in mid-July. Such spikes often signal periods of heightened suspicious activity, where fraudulent or manipulative behaviours may be more prevalent.
* **Concentration of Anomalies:** By clustering on specific days, investigators can focus on those narrow time windows for deeper forensic analysis—scrutinizing which buyers and sellers contributed to the sudden rise in flagged trades.
* **Regulatory and Investigative Value:** Regulators and compliance teams can correlate these spikes with known market events (e.g., announcements or unusual price movements) to see if there is a link between public news and a surge in anomalous trades.
* **Long-Term Monitoring:** Tracking this anomaly trend over time allows the exchange to see whether suspicious activities are tapering off (possibly due to heightened oversight) or re-emerging under different patterns.

**5. Conclusion and Findings:**

Our integrated approach to detecting circular trading—encompassing statistical outlier detection, machine learning–based anomaly scoring, network analysis, and time-series monitoring—provides a practical, business-oriented framework for market surveillance. While it does not guarantee perfect prediction, the solution offers a clear starting point for focusing investigative efforts, thus delivering tangible benefits:

* **Targeted Investigations:**

By highlighting suspicious buyer–seller relationships and abnormal trading patterns, our system helps compliance teams and analysts narrow their search to specific segments of the dataset. This focus ensures that business stakeholders spend less time combing through massive volumes of records and more time evaluating truly high-risk areas.

* **Improved Efficiency and Time Savings:**

Rather than allocating resources to a blanket review of all trading activity, investigators can zero in on the most concerning trades flagged by outlier analysis and anomaly detection. This targeted workflow frees up valuable time and allows teams to concentrate on the most urgent threats.

* **Foundation for Future Enhancements:**

Although the current solution effectively guides where to look, it is not an all-encompassing predictor of fraud. Ongoing refinement—such as incorporating additional data sources, adjusting detection thresholds, or leveraging updated machine learning models—can further improve precision. As market behaviours evolve, the framework can be recalibrated to maintain its effectiveness.

* **Strengthened Market Integrity:**

Even with its inherent limitations, the system provides meaningful direction for pre-empting potential collusion and manipulative trading. By focusing investigative resources on likely hotspots, the exchange reinforces trust in its operations and bolsters investor confidence.

In summary, this solution offers a practical means to identify and prioritize suspicious activity, laying the groundwork for continuous improvement. Although it does not serve as a definitive proof of fraudulent behaviour, it significantly streamlines the investigative process, empowering teams to take swift, data-driven actions that safeguard the integrity of the trading environment.

**6. References:**

[1] <https://www.investopedia.com/terms/c/circulartrading.asp>

[2] <https://vajiramandravi.com/upsc-daily-current-affairs/prelims-pointers/what-is-circular-trading/>

[3] <https://www.geeksforgeeks.org/what-is-isolation-forest/>