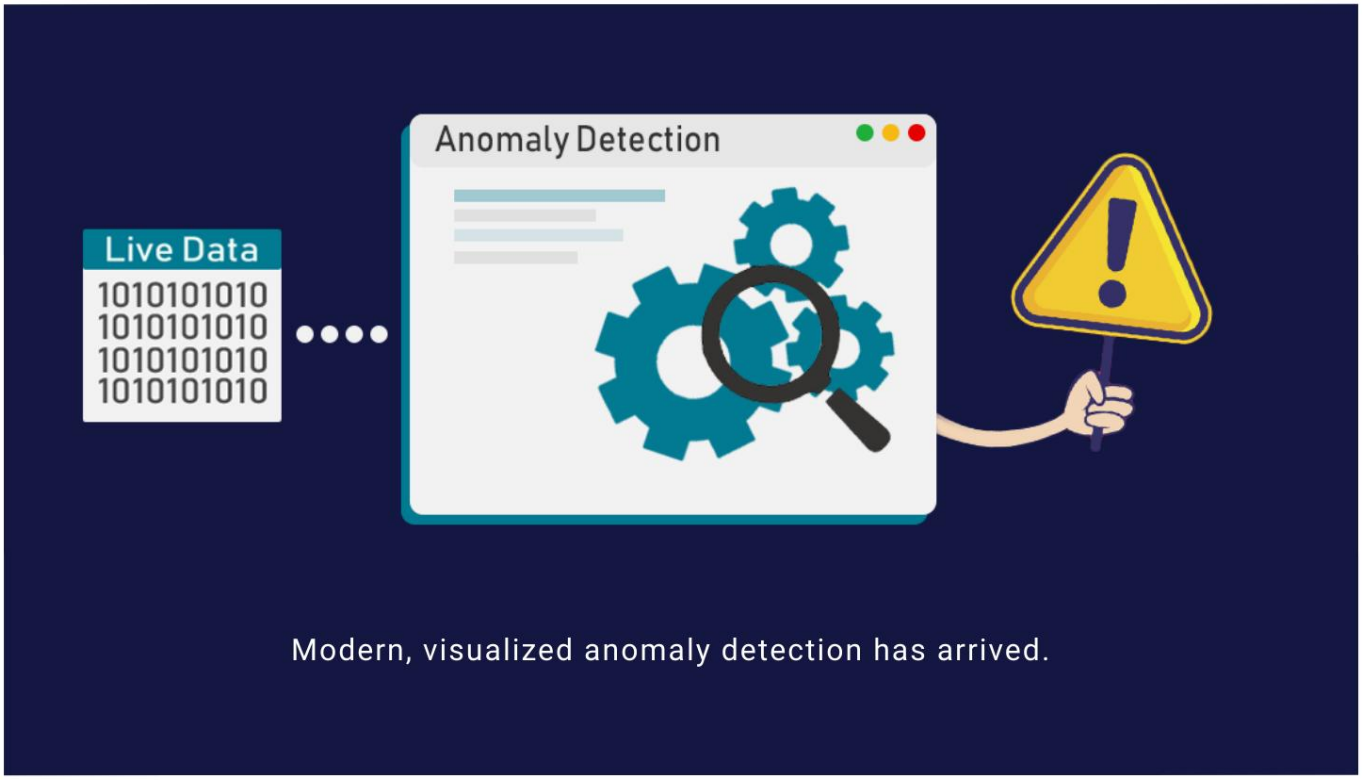


1

Detection of Anomalous Financial Behaviour

A comprehensive report on suspicious financial activity of several entities and inviduals as an outcome of an extensive analysis of available data.

Hypothesis

As part of our analysis of the given dataset, we've conjured the hypothesis that people who have transfered an amount of or over 200,000 Australian dollars in a short period of time have a very high chance of being anomalous or, in simpler terms, being engaged in fraudalent activities.

Scale of Discussion

A large number of people are involved in the concerned cluster, forming an entire chain

Employed Algorithm

As part of the analysis, some categorical features were chosen which were deemed to be key in categorizing in the activities of the involved individuals, e.g.

- Transaction type
- Reference
- Amount
- Weekday

After grouping the concerned people together, we employed the machine learning algorithm K-means clustering. We decided to use 16 clusters as we were dividing features into bins.

Viability of Algorithm

Should there be anomalies in a dataset such as the one provided, there will be some minority groups, which are referred to as collective outliers. The rest of the majority dataset would have a regular activity level without any anomalies in the same time period as that in which the anomalies were detected

K-means clustering, hence, was chosen as it allows us to detect collective outliers that allow us to find anomalies in the dataset.

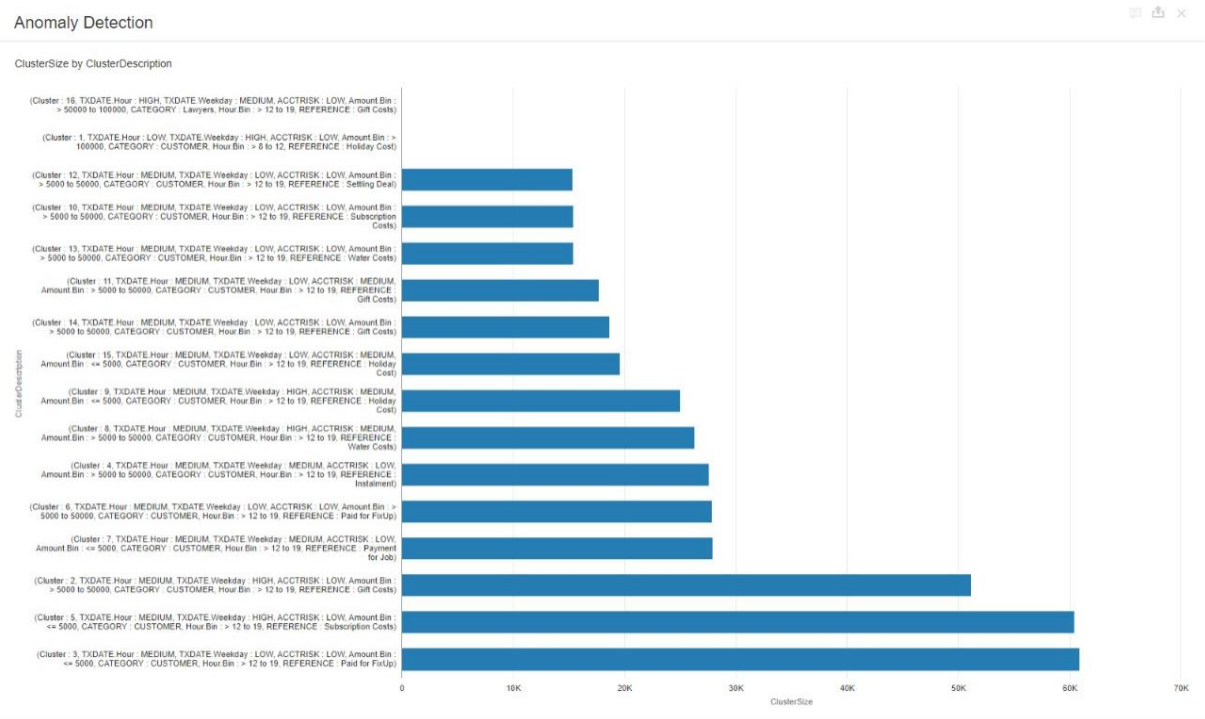
Process

We can separate all transactions into clusters based upon similarities with other transactions; we ended up with 16 clusters.

Two of those 16 were collective outliers of sizes 10 and 30 respectively. The latter was the one we chose to do a selective study of, which became the foundation of our final findings.

This cluster contained a very high spike in the Amount of money(200K AUD) being transferred amidst a very small number of transactions over a small period of time i.e. an hour.

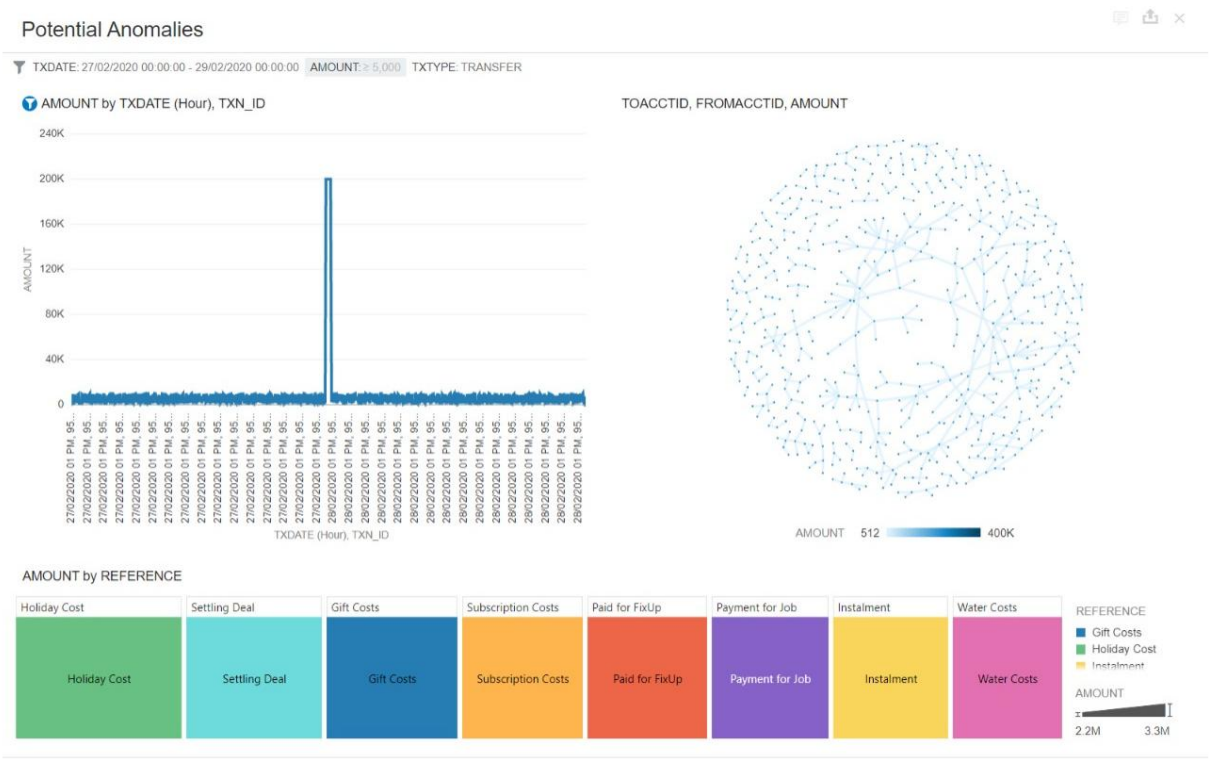
We hypothesized that those who have transactions of or greater than 200,000 AUD over the short period of time are anomalies; however, of course we had to confirm this.



Two of those 16 were collective outliers of sizes 10 and 30 respectively. The latter was the one we chose to do a selective study of, which became the foundation of our final findings.

This cluster contained a very high spike in the Amount of money(200K AUD) being transferred amidst a very small number of transactions over a small period of time i.e. an hour.

We hypothesized that those who have transactions of or greater than 200,000 AUD over the short period of time are anomalies; however, of course we had to confirm this.



As can be quite clearly seen in the graph above, usually transactions remain at a number far smaller than 40,000 AUD - according to specific data, the highest it usally goes to is 5,000 AUD.

However, on the 29th of February between 9AM and 10AM, there is a sudden spike that touches 200,000 AUD over an extremely small amount of time. This alone should be enough to raise concern.

For the cluster, we included the feature of Time and the bin of Amount. What added to our suspiciouns was the fact that, approximately, the

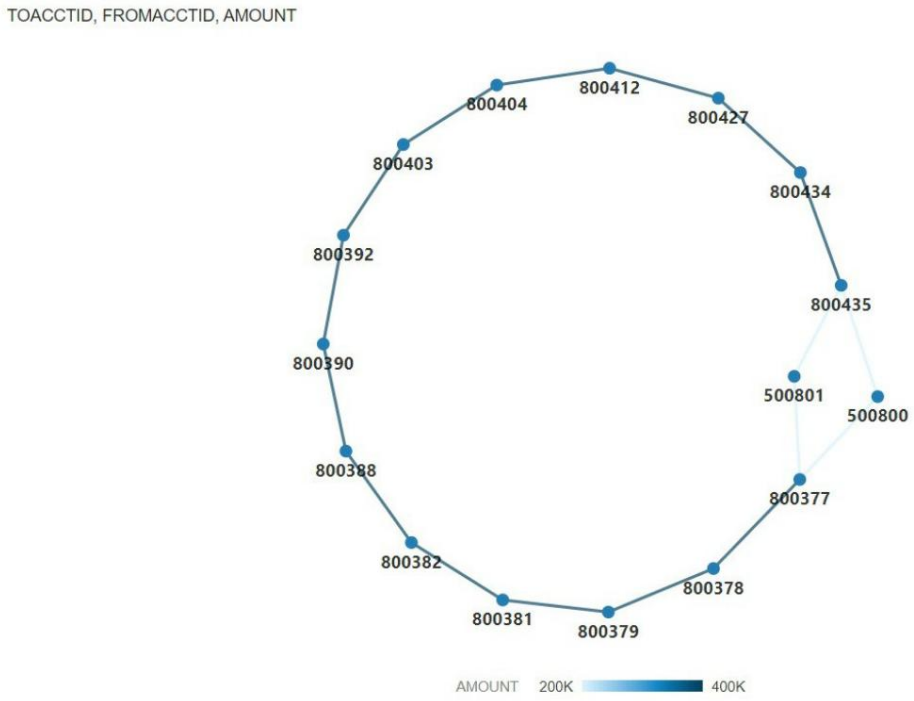
- median was 2x10^2
- mean was 2x10^3

whereas the maximum amount was 2x10^5. A very large difference, all caused by that one sudden, abnormal spike!

Furthermore, we saw that all those who had made transactions greater than 200,000 AUD during that period of time shared core features. Besides the common amount and time, they had been made via transfers, the majority reference provided was Holiday cost, which would further become distributed amongst other people.

Conclusion: Network of Suspicious Transactions

Upon selecting the spike in the above graph, the following chain emerged.



This quite clearly establishes that there is, in fact, a network of transactions involving several people transferring a large sum of money over a very short period of time which can only be categorized as highly suspect with a very high potential of fraudalent activities taking place, thus confirming our hypothesis.