ARIZONA Auditor General
Making a Positive Difference

# FRAUD PREVENTION ALERT
## Data Analytics: Part 4
December 2019

# Why are we issuing this alert?

As presented in previous Fraud Prevention Alerts, using risk-based assessments and data analytic techniques to identify trends, patterns, anomalies, and abnormal relationships in data can be an effective anti-fraud control.[1] Data analytic techniques can be separated into 3 categories for practical applications: basic, statistical, and advanced. This final alert in a 4-part series focuses on advanced analytical techniques and describes how using these techniques on an organization's data and investigating associated findings can help public officials identify fraudulent activity before it becomes significant.

## What are advanced data analytic techniques?

Advanced data analytic techniques involve mathematical and visual approaches to interpreting data. Deviations could indicate an error, irregularity, or fraud, and should be further investigated. This alert focuses on location analysis, trend analysis, data visualizations, and machine learning algorithms.

## Location analysis

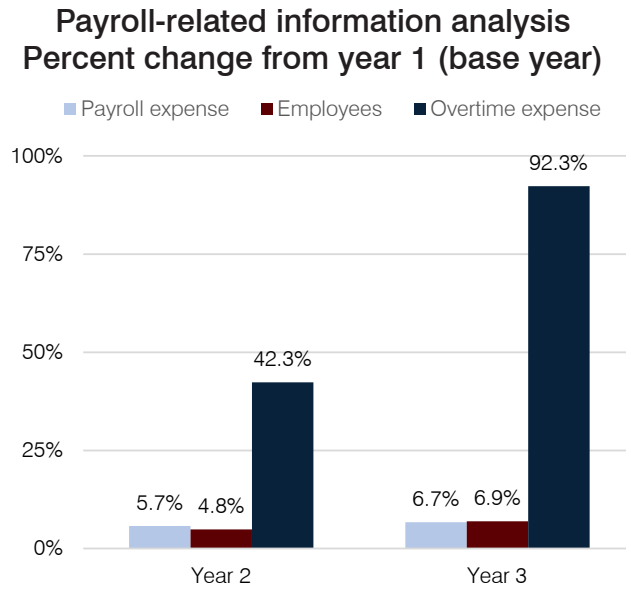
Location analysis reveals the relationship of a location to other data, such as events, transactions, or facilities. Specifically, a location's address can be converted into latitude and longitude coordinates (geocoding), allowing the address to be mapped and mathematically analyzed. For example, agencies can analyze the proximity of employee addresses to vendor addresses in order to reveal relationships that might otherwise go unnoticed. As illustrated in the graphic below, the mathematical calculation between 2 addresses shows that an accounts payable clerk lives within a quarter mile of a mailbox at a retail postal services store used by a vendor as its physical address. Further investigation is required to ascertain whether there is an employee-vendor relationship and its propriety.

**Employee and vendor address analysis**


[1] See Office of the Auditor General, *Fraud Prevention Alert—Data Analytics Part 1*, November 2017, Report 17-406, *Fraud Prevention Alert—Data Analytics Part 2*, October 2018, Report 18-406, and *Fraud Prevention Alert—Data Analytics Part 3*, July 2019, Report 19-404.

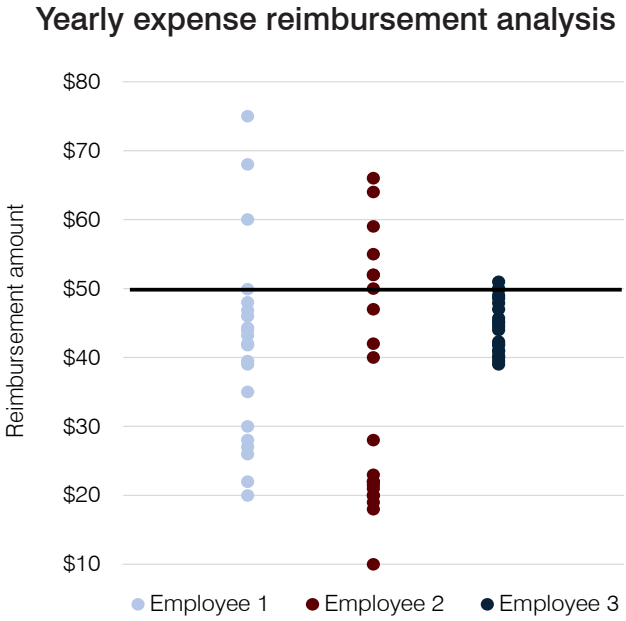
## Trend analysis

Trend analysis evaluates data over a sufficient period of time by focusing on line-item changes that may reveal patterns requiring action by management. For example, and as illustrated in the graph of payroll-related information, the percentage change of a given year is calculated by dividing the amount of change between the base year and the given year by the base year amount. This analysis of payroll expenses, employee counts, and overtime payments over a 3-year span reveals that overtime expenses rose by 42 percent in year 2 and 92 percent in year 3, although rises in payroll expense and employee counts were significantly less during that time frame. Further investigation is required to determine whether the overtime payments were authorized and appropriate.




## Data visualization

Data visualization is a graphical representation of information that helps users understand complex issues and identify patterns, trends, and correlations that would be difficult to notice using numerical representations alone. For example, if an agency has a reimbursement policy that requires supervisory approval for expense reimbursements over $50, visual representation of a year's worth of data by department or by employee may show patterns not observable in day-to-day accounting processes. For example, and as illustrated in the yearly expense reimbursement analysis, employees 1 and 2 show expected behavior, with multiple reimbursement amounts both below and above the threshold for supervisory review. Employee 3's data is unusual both for its concentration just under the reimbursement threshold and because most of the reimbursements are for amounts between $40 and $50. Although the pattern is unusual compared to other employees and should be further investigated, it may be explained by a circumstance unique to that employee, or it may indicate fraud.

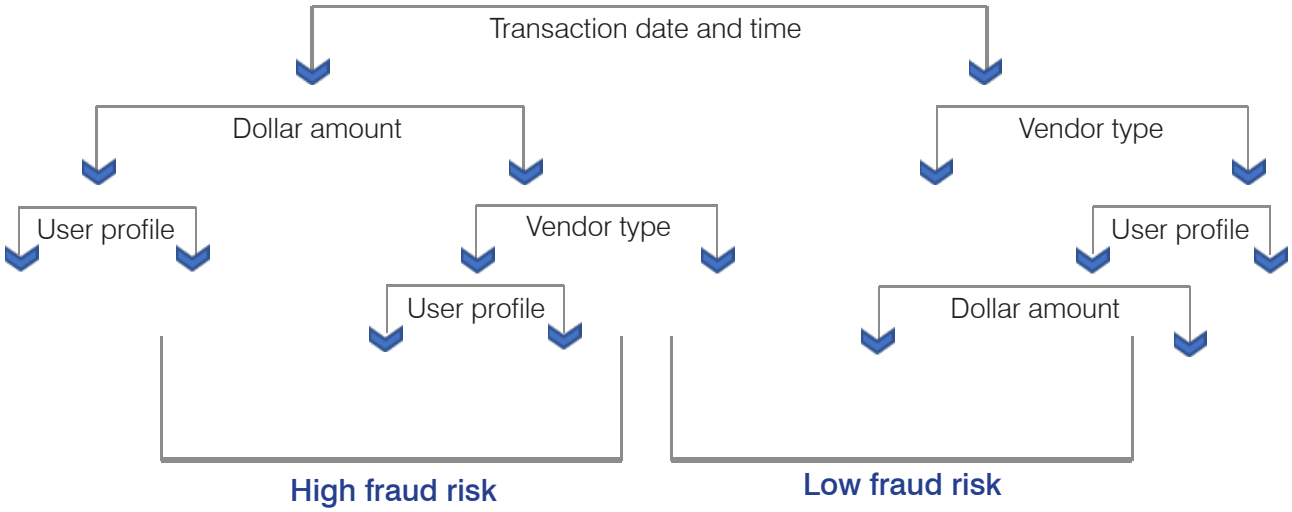



## Machine learning algorithms

Machine learning is the application of algorithms, or sequences of instructions, to predict an outcome. Although applying machine learning takes advanced technical expertise, it is becoming more important in fraud detection and deterrence and is widely used by banks and credit card companies to find suspicious transactions. In its simplest form, users provide a set of "training data," such as a set of transactions (e.g., transaction date and time, dollar amount, vendor type, and user profile) that include known fraudulent transactions. The algorithm develops a model to predict fraud based on the known fraud instances in the "training data." The algorithm model may then be applied to new datasets to identify potential fraudulent transactions that exhibit similar characteristics to the previously known fraudulent transactions. Applying machine learning algorithms to data over time facilitates a learning process that adjusts the algorithms to more accurately predict outcomes.

As visualized in the decision tree (see page 3), components of a transaction dataset may be classified into high or low fraud risk by applying machine learning algorithms. Each branch splits the data into 2 outflows based on some value of the individual transaction. Additional branches will add new information that can increase the prediction accuracy of the decision tree. For example, transactions can first be split on whether or not a transaction occurred outside normal business hours, and then the transaction dollar amounts can be split by an expected dollar range, even-dollar testing, or

Decision tree




utilizing Benford's Law.[2] Transactions are also split on whether or not a transaction is made at a type of vendor expected for that particular user. The transaction data continues through the decision tree until a final determination of high or low fraud risk is ascertained.

## Recommendations

To help protect public monies, public officials should establish proactive data monitoring and analysis programs designed to deter and detect fraud. Specifically, public officials should:

- Review all areas of operation and determine areas that may be susceptible to fraud, using a risk-based approach to prioritize the most vulnerable areas.
- Develop and regularly perform data analyses to detect fraud in the identified areas.
- Investigate findings in a timely manner using trained, knowledgeable employees to interpret results.
- Communicate the results to management, who should then take appropriate action.

[2] For more information on even-dollar testing and Benford's Law, see Office of the Auditor General, *Fraud Prevention Alert—Data Analytics Part 2*, October 2018, Report 18-406, and *Fraud Prevention Alert—Data Analytics Part 3*, July 2019, Report 19-404.