POLITECNICO DI TORINO

MASTER'S THESIS IN ENGINEERING & MANAGEMENT

HARNESSING THE POWER OF DATA ANALYTICS TO DRIVE BUSINESS VALUE: IMPLEMENTATION FOR FERRERO INTERNATIONAL S.A. OF A D&A AND CONTINUOUS MONITORING MODEL TO PREVENT FINANCIAL FRAUD.

SUPERVISOR

Prof.ssa Elisa Ughetto

CANDIDATE

Giorgia Gargani

Academic Year 2022 - 2023

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all the individuals who have supported and encouraged me throughout the journey of this master's thesis.

Firstly, I want to extend my thanks to my supervisor, Elisa Ughetto, for her guidance and unwavering support.

I am deeply thankful for my parents, Paola e Davide, and for their constant love and support. Their immovable presence has been the cornerstone of my academic journey, and I will be forever grateful for their effort in making me become a strong and independent woman. A special thanks also go to my brother, Alessandro, for being always a conscious and passionate guide standing by my side in every challenge I wanted to accomplish.

I would like to thank my grandparents, Gennaro and Marisa, for always believing in my dreams and for being a dedicated support at every stage of my university journey.

Lastly, I would like to express my gratitude to my friends who have been a strong and insightful presence during this incredible journey. Without their support, love and time reaching this goal would have been undoubtedly more arduous.

ABSTRACT

In contemporary organizations, Journal Entries serve as fundamental components of financial statements and hold valuable information. Nevertheless, these entries are frequently generated and managed by employees who have access to sensitive financial data. As a result, they become a significant source for carrying out fraudulent activities.

Implementing robust data analytics and continuous monitoring models has become crucial in preventing fraud and mitigating associated risks. These models aid organizations in identifying anomalies within their accounting systems. However, designing an effective data analytics model requires careful consideration of various factors, particularly the selection of Key Risk Indicators (KRIs) and the technical execution of the model. Considering these aspects is vital to ensure the success and efficacy of the data analytics framework.

The first part of this master thesis provides an extensive literature review, highlighting the evolution, concepts, tools, and techniques utilized in big data analytics nowadays. Additionally, it surveys the existing literature on the application of data analytics models in fraud detection and prevention to identify current trends and challenges in the field.

The second segment focuses on anomaly detection in accounting examining various types of financial fraud, their impact on organizations, and the traditional methods employed to mitigate such fraudulent activities. The third part emphasises proactive measures to identify potential fraud risks before they occur. It investigates the integration of data analytics into existing control systems, enabling organizations to detect early warning signs and develop effective preventive strategies. The thesis further explores the challenges and ethical considerations associated with utilizing data analysis for fraud prevention.

Lastly, a comprehensive case study is conducted to demonstrate the practical implementation of a data analytics and continuous monitoring model. The study presents a real-world scenario where big data analytics techniques are applied to detect and prevent financial fraud in an accounting setting. It evaluates the effectiveness of the model and highlights the benefits, limitations, and practical considerations involved in its implementation.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

1

BIG DATA ANALYTICS IN ACCOUNTING: TECHNIQUES AND APPLICATIONS

1.1 Defining Big Data Analytics: Literature Review

In recent years, the explosion of digital data has transformed the way organizations operate and make decisions. Accounting has been significantly impacted by the rise of big data and the availability of sophisticated analytics tools and techniques. Big data analytics in accounting refers to the process of extracting insights from vast and complex datasets to improve financial performance, reduce risk, and enhance decision-making. The use of big data analytics in accounting has grown rapidly, with many organizations recognizing its potential to provide valuable insights and competitive advantages.

The chapter discusses the various types of data analytics, including descriptive, diagnostic, predictive, and prescriptive analytics, and highlights the importance of data analytics in accounting. The chapter also examines the various applications of big data analytics in accounting, such as fraud detection and prevention, financial statement analysis, risk assessment and management, and audit automation and analysis. Furthermore, will be explored the different tools and techniques used in big data analytics for accounting, including data management and integration tools, data visualization and reporting tools, predictive modelling tools, and machine learning algorithms.

1.1.1 Four Vs of big data

Big data analytics refers to the process of extracting insights and meaningful information from large, complex datasets that would be impossible to process using traditional data processing tools (Oracle, 2022). Although the term "big data analytics" is commonly used, there is no standard definition of the term, and it can have different meanings.

One definition of big data analytics focuses on the four Vs of big data: volume, velocity, variety and veracity (Doug Laney, 2011). According to this definition, big data analytics involves the processing and analysis of large datasets that are too big, too fast, or too diverse for traditional data processing tools to handle. This definition emphasizes the need for specialized tools and techniques to manage and analyse large datasets effectively. Understanding these four Vs is critical to understanding the challenges and opportunities of big data analytics. Below are the 4Vs explained in detail:

- **Volume:** Volume refers to the sheer amount of data that is generated and collected every day. Traditional data sources typically involve structured data that is stored in databases, spreadsheets, and other structured formats. In contrast, big data involves large volumes of both structured and unstructured data from a wide range of sources, such as social media, sensors, mobile devices, and more. Managing and processing such large volumes of data requires specialized tools and techniques, such as distributed storage systems, cloud computing, and data warehouses.

- **Velocity**: Velocity refers to the speed at which data is generated and processed. In the age of big data, data is being generated and collected in real-time or near-real-time, which means that it must be processed and analysed quickly to extract insights and value. For example, online retailers need to analyse customer behaviour in real-time to deliver personalized recommendations and promotions. To manage and process such high-velocity data, organizations need to deploy real-time analytics tools, such as streaming analytics, in-memory databases, and event processing.

- **Variety**: Variety refers to the diversity of data sources and formats that are part of big data. In addition to traditional structured data, big data includes unstructured data, such as text, images, and video, as well as semi-structured data, such as XML and JSON. This variety of data sources and formats poses challenges for traditional data processing tools, which are designed to work with structured data. To analyse and interpret such diverse data, organizations need to deploy specialized tools, such as natural language processing, sentiment analysis, and image recognition.

- **Veracity**: The fourth V of big data is veracity, which refers to the reliability and accuracy of the data. With the increasing volume, velocity, and variety of data, ensuring data quality and accuracy is becoming increasingly challenging. Inaccurate or unreliable data can lead to incorrect or misleading insights, which can have serious consequences for organizations. Veracity is important in big data analytics because the data comes from a variety of sources and may not always be accurate or consistent. For example, social media data may contain false information, incomplete or inconsistent data. Similarly, data from sensors may be affected by environmental factors, measurement errors, or calibration issues. To ensure the veracity of big data, organizations need to implement data governance and data quality frameworks. Data governance involves defining policies and procedures for managing fraud throughout its lifecycle, including data acquisition, storage, processing, and analysis. Data quality frameworks involve assessing the quality of the data and ensuring that it meets certain standards for accuracy, completeness, consistency, and timeliness.

Overall, the four Vs of big data are important because they highlight the unique challenges and opportunities of analysing large, complex, and diverse datasets. By understanding these characteristics, organizations can develop effective strategies for managing and analysing big data to gain insights and drive business value.

1.1.2 Advanced Analytics Techniques in Big Data Analytics and NLP

Another definition of big data analytics emphasizes the role of advanced analytics techniques in uncovering insights from large datasets. This definition focuses on the use of statistical models, machine learning algorithms, and other advanced techniques to interpret large volumes of data extracting valuable information. Also, this definition recognizes the need for more and more specialized skills and expertise in data science and analytics the use of statistical models and machine learning algorithms is particularly important because these techniques can help organizations identify patterns and relationships that are not immediately apparent in the data. These techniques involve both the use of mathematical models and

complex algorithms that can learn from large databases and make predictions or decisions based on that learning.

In addition to statistical models and machine learning algorithms, many other advanced techniques are used in big data analytics. For example, natural language processing (NLP) is a technique that is used to analyse and interpret unstructured data, such as text data from social media or customer reviews which can be difficult to analyse using traditional data analysis techniques. One of the main applications of NLP in big data analytics is *sentiment analysis*, which involves analysing text data to determine the sentiment or emotional tone of the content. This can be useful for companies that want to understand how customers feel about their products or services. For example, a company can use sentiment analysis to monitor social media conversations about their products and services and gain insights into customer satisfaction levels.

Another application of NLP in big data analytics is *topic modelling*, which involves identifying the topics or themes present in a large corpus of text data. This can be useful for companies that want to understand the topics that are most important to their customers or to identify emerging trends in their industry. For example, a company that sells sports equipment can use topic modelling to identify the most popular sports and activities among their customers and adjust their product offerings accordingly.

1.1.4 Leveraging Insights and Opportunities for Business Value and Competitive Advantage: the crucial role of the Analysts

A third strand of big data analytics literature focuses on the value that can be generated by analysing large datasets regarding the identification of opportunities that can drive business value and competitive advantage. This definition emphasizes the need for a *strategic approach* to big data analytics, focused on addressing business challenges and opportunities. Starting from that, the true value of big data is not in the dataset itself but in the business opportunities that can be derived from it (Oracle, 2022).

This concept highlights the importance of taking a strategic approach to big data analytics companies must first focus on extracting key business challenges and opportunities, and then use big data analytics to drive solutions that deliver measurable business value. This requires a deep understanding of the business, its goals, and its customers, as well as the ability to translate those insights into concrete actions that can be taken to drive business value.

To harness the full potential of big data analytics for business value, companies must ensure they have the necessary technological infrastructure and analytical capabilities. This involves having the capacity to gather and store significant amounts of data, as well as the ability to process and analyze that data using advanced analytical techniques. It also requires having the right people with the right skills in place h human element in data analytics is essential for several reasons. Human analysts bring a contextual understanding to the data analysis process, leveraging their domain knowledge and expertise to interpret data in the context of the business, industry, or specific problem at hand. They can identify relevant variables, consider external factors, and apply critical thinking to ensure accurate and meaningful analysis. Moreover, human analysts can exercise judgment and intuition when making decisions based on data, taking into account factors that may not be captured solely through automated algorithms. They can also ask insightful questions, explore different perspectives, and uncover nuances that automated systems may overlook. Furthermore, analysts provide the necessary *ethical oversight*, ensuring that data privacy, fairness, and transparency are upheld throughout the analytics process. Their involvement ensures a balanced approach that combines the power of machines with human insights to more accurate, reliable, and actionable results in data analytics.

Overall, by identifying key business challenges and opportunities and leveraging advanced analytical techniques to generate insights and solutions, companies can unlock the full potential of big data and drive real value for their organization.

1.2 Big Data Analytics in Accounting: categories and most insightful application

Big data analytics has become increasingly important in the field of accounting due to the enormous volume of data generated by financial transactions and the need for real-time insights into financial performance. Accounting data is particularly well-suited to big data

analytics, as it often involves structured data that can be easily analysed using analytical tools and techniques. One of the key benefits of big data analytics in accounting is its ability to help organizations improve their financial performance. By analysing large volumes of financial data, companies can identify trends and patterns that can be used to optimize financial performance and make more informed business decisions. This can include identifying areas where costs can be reduced, where revenue can be increased, or where operational efficiencies can be improved (KPMG, 2022).

Big data analytics can also be used to improve financial reporting and compliance. By analysing financial data in real time, companies can identify potential compliance issues and take action to address them before they become larger problems. This can help to ensure that financial reporting is accurate and timely, reducing the risk of regulatory penalties and other legal issues. Another key benefit of big data analytics in accounting is its ability to enhance risk management. By analysing financial data, companies can identify potential risks and take action to mitigate them before they become significant issues. This can include identifying potential fraud, waste, or abuse, as well as identifying areas where financial controls may be weak and implementing measures to strengthen them. On balance, the importance of big data analytics in accounting cannot be overstated. By leveraging the power of big data analytics, organizations can improve financial performance, enhance compliance, and risk management, and make more informed business decisions. This requires a combination of advanced analytical tools and techniques, as well as the right people with the right skills to analyse the data and derive meaningful insights from it. As the volume of financial data continues to grow, the role of big data analytics in accounting will only become more critical in the years ahead.

At a high level, there are four main types of data analytics: descriptive, diagnostic, predictive, and prescriptive (Harvard Business School, 2021):

- *Descriptive analytics* is the most basic type of analytics and involves simply summarizing historical data to gain insights into past performance. This can include generating reports and dashboards to visualize data and identify trends, as well as performing basic statistical analyses to quantify key performance indicators.

- *Diagnostic analytics* is a step up from descriptive analytics and involves digging deeper into historical data to identify the root causes of specific outcomes. This can involve using techniques like regression analysis, correlation analysis, and causal inference to identify the factors that are driving unexpected outcomes. Diagnostic analytics is particularly useful for identifying opportunities for improvement and optimizing performance.
- *Predictive analytics* is focused on using historical data to make predictions about future outcomes. This involves building statistical models and machine learning algorithms that can be used to forecast trends, identify patterns, and anticipate changes in the business environment. Predictive analytics is particularly useful for identifying emerging opportunities and risks, and for optimizing decision-making in real time.
- *Prescriptive analytics* is the most advanced type of analytics and involves using predictive models and other advanced techniques to recommend specific actions to take. This can involve using optimization algorithms to identify the best course of action given a set of constraints or using simulation techniques to evaluate the potential impact of different decisions. Prescriptive analytics is particularly useful for identifying optimal strategies and making decisions in complex and uncertain environments.

While all four types of data analytics - descriptive, diagnostic, predictive, and prescriptive - have the potential to prevent fraud, predictive and prescriptive analytics are especially valuable in identifying and stopping fraudulent activities before they occur. Predictive analytics utilizes statistical models and machine learning algorithms to scrutinize historical data and uncover patterns that could indicate future fraudulent activity. This approach can identify potential instances of fraud in transactional data, alerting organizations to take preventive measures. On the other hand, prescriptive analytics takes predictive analytics a step further by recommending actions to prevent or mitigate fraudulent activity. Using artificial intelligence and machine learning algorithms, prescriptive analytics evaluates potential outcomes and suggests the most effective course of action. This proactive approach to preventing fraud can be highly effective for organizations.

1.3. Implementing Data Analytics for Risk Assessment

One of the most prominent applications of big data analytics in accounting is fraud detection and prevention as it will be described more in detail in the next chapter. In fact, by analysing large volumes of transactional data, big data analytics can identify patterns and anomalies that could indicate fraudulent activity, enabling organizations to take proactive measures to prevent or mitigate the risk of fraud. For example, big data analytics can be used to flag unusual patterns of spending or transfers of funds, and alert auditors or managers to take a closer look.

Another important application of big data analytics in accounting is financial statement analysis. By analysing large volumes of financial data, big data analytics can provide insights into financial performance, trends, and areas of opportunity or risk. This can help organizations make more informed decisions about resource allocation, investment strategies, and financial planning.

Traditionally, financial statement analysis has been a time-consuming and labour-intensive process that relies on manual data entry and manipulation. However, with the advent of big data analytics, this process can be automated and streamlined, allowing analysts to identify trends and patterns quickly and accurately in financial data. Using big data analytics tools, financial analysts can analyse large volumes of financial data from multiple sources, such as financial statements, annual reports, and news articles, to gain a more comprehensive understanding of a company's financial health. This analysis can include identifying key performance indicators (KPIs) and trends over time, such as revenue growth, profitability, and debt levels.

Data analytics models can also help to identify potential red flags in financial statements, such as irregularities or inconsistencies that may indicate fraudulent activity or errors in financial reporting. By identifying these red flags early, financial analysts can investigate further and take appropriate actions to address any issues. In addition, big data analytics can help financial analysts to conduct more accurate and comprehensive financial statement analyses by providing a broader range of data sources and more sophisticated analysis tools.

Risk assessment and management are where big data analytics can be particularly useful. The use of data analytics in corruption and fraud risk assessments can be linked to other objectives, such as effectiveness and efficiency, and can be applied across different levels and competencies within an organization. While procurement officials may collect data to support tendering and contract performance assessment, project managers, risk managers, and auditors may utilize the same data to identify potential vulnerabilities and assess risks in the control environment. The purpose of the data value chain is to convert data into information and knowledge that can address a range of questions related to fraud and corruption risks, as summarized in *Table 1*. Therefore, data analytics has the potential to contribute to achieving broader strategic objectives within an organization.

	Hindsight	**Insight**	**Foresight**
Information	What happened? (Reporting)	What is happening now? (Alerts)	What will happen? (Extrapolation)
Knowledge	How and why did it happen? (Modelling, experimental design)	What is the next best action? (Recommendation)	What's the best/worst that can happen? (Prediction, optimisation, simulation)

Table 1: Key questions that data analytics can address (Davenport, Harris & Morison, 2010)

The table presents a comprehensive framework consisting of three key dimensions: Hindsight, Insight, and Foresight, each associated with different levels of information and knowledge. In the *Hindsight* category, the focus is on understanding the past. It involves reporting and analysing information to discern what exactly happened in a given situation. This aspect is crucial for retrospective assessments and historical analysis. Moving on to *Insight*, the emphasis shifts to the present and involves gaining real-time awareness and insights into ongoing events. This enables the ability to identify and respond to current circumstances, often through alerts and notifications. It helps decision-makers stay updated and make informed choices promptly. Finally, *Foresight* entails a forward-looking perspective. It involves extrapolating from availed information and knowledge to anticipate future outcomes. This dimension explores the possibilities of what might happen and

facilitates prediction, optimization, and simulation. It also allows for the identification of the next best course of action and offers recommendations based on existing knowledge and predictive models.

The conceptual framing provided by *Table 1* for the use of data analytics in assessing corruption and fraud risks is more specific and risk-based when applied in practice. Specific questions related to risk assessment help extract information from different areas of operations to make informed decisions and take actions to reduce an organization's vulnerability to fraud and corruption. Linking the questions that data analytics will answer to risk assessments ensures that the analytics process serves broader strategic objectives rather than just being driven by data or technological tools (Cotton, Sandra, & Leslie, 2016).

1.3.1 Steps for using data analytics to effectively assess risks

The decision to adopt a data-driven approach for risk assessments and the use of data analytics involves investing in data governance, data integrity, and other quality controls that come with such efforts. Ideally, data should be interoperable, accessible, discoverable, and open, to enable continuous production, collection, sharing, and re-use. While the steps outlined in *Figure 1* do not require this ideal scenario, they do assume that an institution has the basic data architecture, infrastructure, and skills to implement them. Some institutions may need to make systemic improvements to their data collection process or engage external experts for successful implementation. Nevertheless, the specific application of data analytics in corruption risk assessments offers a targeted and focused approach to getting data in order, which can also serve broader strategic objectives.

Figure 1: Steps for carrying out a data-driven risk assessment (OECD, Baesens, Van Vlasselaer &Verbecke, 2015)

As depicted in the figure, pre-processing consists of four stages, which can be more time-consuming than the actual analytics process itself. The time required for each step may vary depending on the circumstances and the level of maturity of an organization's data governance. An organization with a centralized data warehouse may require less time to obtain the necessary data for analysis. Conversely, an organization that needs access to data held by an external entity, such as another government agency, may need to devote significant time and resources to establishing the necessary processes and procedures to obtain the data.

The following presents a summary of the sequential steps and supplementary factors involved. It's important to note that these steps are not always strictly linear and may involve

repeated activities, such as assessing the reliability and validity of data. However, they are presented in sequence to emphasize important considerations and aspects:

1. **Defining risk-based objectives** is crucial for effective data analysis and risk assessment. By shaping objectives based on identified risks, the team can target areas where fraud and corruption risks are most likely to occur. This alignment ensures that the data analytics plan is an informative tool for robust risk assessment. Perception-based risk assessments, input from experts, workshops, focus groups, audit reports, media coverage, and previous data analyses can inform the development of specific questions and indicators. It is essential to define objectives that are descriptive, diagnostic, predictive, or prescriptive. Well-defined objectives and risk identification are crucial for the following steps, including identifying the relevant data custodians and datasets, acquiring, and collecting the data, cleaning the data, and conducting appropriate analytical tests. Before implementing analytics tests, it is necessary to understand programme rules and processes, including what is considered normal behaviour. By doing so, the analytics team can develop refined analytics tests that may produce fewer false positives, thus saving time in reviewing results.

2. **Identify data needs and sources.** The subsequent step in the corruption risk assessment process is to determine the requisite data to identify the fraud or corruption risks outlined in the previous step, as well as identify the sources of that data. This can entail gathering data that is present within the organisation or data from external organisations. The precise data required for analysis will depend on the analytics objectives and the specific indicators that the analytics tests will be employed to identify.

3. **Select and obtain the data**. The subsequent phase involves the acquisition of data that is required for conducting the analysis. In cases where data is procured from external entities, it is essential to have a basic comprehension of the data before creating a formal request for it. A data request typically contains details such as the data format, the intended purpose of the data, a set of sample fields, control totals for

validating data completeness, and any constraints associated with the data. It is worth noting that the employment of data analytics for corruption risk assessment often runs concurrently with, and can initiate, extensive enhancements to data governance, particularly when multiple institutions are engaged in data collection, creation, or sharing. After gathering the data, it is crucial to obtain or construct a data dictionary. A *data dictionary* provides explanations of every field within the data and is a key source of information for data analysts. It ensures that individuals use uniform definitions and comprehend when different terms are used for the same concept, or when the same term has different meanings across government entities or programs. This stage also entails addressing other data governance issues such as data semantics, interoperability, and metadata.

4. **Understand data and assess reliability.** The subsequent stage involves evaluating the dependability and validity of the data and implementing measures to clean and format the data if required, to guarantee its suitability for analysis. Data analysts may conduct one or several data-validation tests to authenticate the reliability and comprehensiveness of the data, including cross-checking the data types against the record layout and data dictionary, matching the record count with the control totals received, checking the hash totals of numeric fields with the control totals received, detecting missing data or duplicate data, reconciling the data to accounting records, conducting reasonability tests, and periodic testing to determine if the data covers the requested period. It is imperative to resolve any discrepancies or anomalies identified during the data validation and cleaning process before performing the analysis, as these anomalies may potentially signify fraudulent or corrupt activities:

    - Verify the data types against the record layout and data dictionary (for example, text fields contain text).
    - Confirm the record count with the control totals received.
    - Confirm the hash totals of numeric fields with the control totals received.
    - Identify missing data (for example, blank fields or gaps in sequences).

- Check for duplicate data and confirm whether any duplicates identified are false positives.
- Reconcile the data to accounting records.
- Perform period testing to determine if the data cover the requested period (INTOSAI Guidelines IT Audit, 2016).

Any discrepancies identified should be addressed before performing the analysis, which may include re-requesting the data (INTOSAI Guidelines IT Audit, 2016). Some suggest that data cleaning should only be done when performing the analysis as cleaning might delete "interesting outliers" (Kimball, 2014). Care should be taken at this step to understand and assess any discrepancies or outliers identified as a result of data validation tests or data cleaning procedures as outliers or anomalies may be indicative of fraud or corruption.

5. **Develop an analysis plan, including specific analytics tests.** This stage entails crafting an analysis plan that delineates the data to be scrutinized, the particular analytics methodology that the team will execute, and the frequency of such an approach. A data analytics work plan can encompass a few weeks or can be part of a more comprehensive risk assessment. The latter can last for several months as it might encompass multiple data sources and concurrent risk management activities. When utilizing data analytics to maintain integrity, particularly in fraud detection and evaluating the efficacy of internal controls, government entities may consider the following actions:

— *Analyse all relevant data*: Comprehensive analysis of relevant data: Government entities can apply analytics tests to the entire dataset. While random sampling helps to help inconsistently across data populations, it may not be sufficient to detect fraud or corruption, as these activities do not typically happen randomly. (ACL, 2013).
— *Design data analytics tests based on the identified fraud indicators*: The analytics team can translate specific indicators of fraud or corruption, identified during the initial analysis, into precise analytical procedures.

— *Determine whether the analysis will be conducted on an ad hoc, repetitive, or continuous basis*: Government entities can choose to conduct data analytics tests on an ad hoc, repetitive, or continuous basis. The frequency of these tests depends on the purpose for which analytics are being employed. For instance, ad hoc data analytics tests can be performed to identify potential issues that may indicate the presence of fraud opportunities. (ACL, 2013). This approach may be sufficient for a project manager using data analytics to identify risks and analyse the effectiveness of control activities relative to specific operational areas or contract types. However, programme managers using data analytics to maintain programme integrity could automate data analytic tests to monitor fraud indicators on a continuous, real-time basis, if possible (U.S. Government Accountability Office, 2015). If it is not possible to automate data analytics tests to be conducted continuously, such as in cases where data is available only periodically, performing these tests on a regular and periodic basis can still provide valuable insights. For instance, incorporating data analytics tests into monthly transaction cycles can help ensure that risks are addressed throughout the year rather than solely on an annual basis. This approach allows for ongoing monitoring and mitigation of risks on a more frequent basis.

6. **Perform the analysis** – During this phase, the analytics team executes the analysis plan to carry out the desired analysis. Several software programs are available to aid in conducting these analyses. Microsoft Access and Excel are commonly used tools among auditors and accountants and can be suitable for less extensive or intricate analyses, depending on the specific context. However, when dealing with large or complex datasets, such as analysing "big data" to identify fraudulent healthcare providers by matching healthcare records, the volume of data may consist of hundreds of millions of rows. It is important to note that Excel has a maximum row limit of approximately one million rows, which may necessitate the use of alternative software programs capable of handling larger datasets. As an alternative, there are other software programmes (e.g., ACL, IDEA, SAS and open-source tools like R and Python, etc.) that can handle larger datasets and complex procedures.

7. **Interpret, communicate and act** – Interpreting the results of analytics in the context of risk assessments entails an iterative process of assessing the output relative to the initial objectives. To what extent are the data answering the stated research questions? Can the tests be refined further to increase the clarity of the results and determine corrective actions, if any? Is there a logical explanation for the results or signs of potentially fraudulent activity? Data analytics tests do not confirm fraud in the procurement cycle; however, they signal specific cases that look suspicious and could require addition additional investigation. Thus, involving individuals with sound judgement, experience, expertise and scepticism are all critical for the evaluation of results. They also provide an effective communication tool to communicate results to relevant parties, including the project managers, internal audit function and investigative bodies who may follow up on instances of potential fraud and corruption.

The effective implementation of data analytics to detect integrity issues is an ongoing and iterative process. As circumstances change and new information becomes available, it is important to update data analytics processes accordingly. Incorporating a feedback loop into the analytics plan can help enhance its effectiveness, where the results of analytics are utilized to inform the design of future analytics tests. Depending on the sophistication of the data analytics system, the feedback loop can be integrated manually or automatically. In a manual feedback loop, anti-fraud or anti-corruption experts analyse the results that correspond to predetermined fraudulent patterns. Through their experience, analysts refine these patterns to improve the identification of new cases in the future. In more advanced analytics systems, analysts validate whether a specific identified case is indeed fraudulent, and this information is used to automatically refine the model employed by the system. This type of machine learning enables continuous improvement of analytics, increased efficiency, and reduced false positives. (KPMG, 2016).

1.4 Overview of Analytical Approaches

The choice of utilizing one or several analytic techniques is contingent upon various factors such as the goals of the analysis, resources, expertise, data quality, and data accessibility. Objectives that intend to make inferences about findings to a larger population or predict patterns of fraudulent or corrupt activities necessitate the application of more sophisticated analytical techniques and statistical knowledge, which can lead to an increase in detection rates and enable the analysis of structured data. *Figure 3* presents an outline of the factors to consider when selecting a data analytics technique based on three factors, namely, detection rates, complexity, and value.

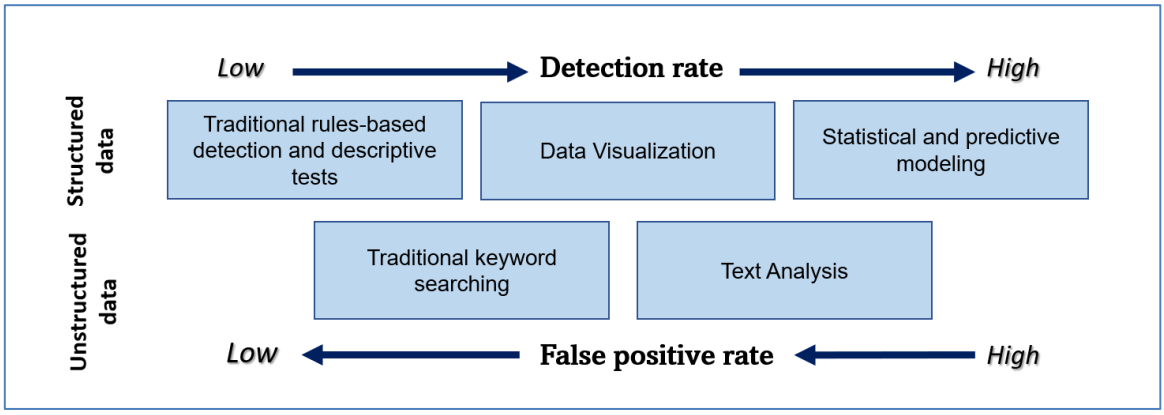


Figure 2: Selecting data analytics techniques based on detection rate, complexity and value (EY, 2016)

In general, rules-based testing using methods like data matching and data mining can be used to achieve descriptive objectives like estimating the likelihood of fraud and corruption in a population and emphasizing red flags for additional inquiry. This technique entails creating pre-set rules to filter or mine data to spot anomalous behaviour. Additionally, information on previous fraud cases and program regulations can be used to spot warning signs that guide the queries run on data sets. A rules-based analytics method, for instance, could help an organization examine procurement data to find bidders who had been awarded sole-source contracts for military contracts during advertising periods of the competition. Similarly, queries can be created to examine transaction data for recognizing numerous transactions

made by the same cardholder from the same vendor on the same day if procurement rules limit people from making purchases above a specific threshold amount. Such rules-based testing, often referred to as breakpoint clustering, can be used to assess the likelihood that suppliers or workers will separate transactions to exceed spending caps when evaluating invoice payments or purchases. While easier than other methods, these analyses could produce more false positives and lower detection rates (U.S. Government Accountability Office, 2015).

Anomaly or outlier detection seeks to find odd or unexpected behaviours, which can help identify potential corruption or fraud when precise patterns are unknown. This is in contrast to rules-based detection, which concentrates on detecting particular known fraud or corruption schemes. Finding figures that are higher or lower than anticipated could be a sign of fraud. Different strategies can be used to implement anomaly or outlier identification. By grouping data based on a comparable attribute, such as location, cluster analysis, for instance, can be used to spot anomalies or outliers in comparison to what would be predicted based on that group.

As previously mentioned, risk assessments can provide managers valuable insights into the control environment and risk mitigation measures by any predictive questions. Predictive analytics involves the creation of models that detect attributes or patterns that exhibit a strong correlation with known instances of fraud. These models are then applied to incoming transactions to determine whether they bear any resemblance to previously known cases of fraud (Henderson & Hammersburg, 2020). modelling, which is particularly adept at identifying complex patterns in data, could be employed to detect potentially fraudulent claims or transactions before they are paid out (U.S. Government Accountability Office, 2015). Furthermore, predictive models can aid managers in scoring transactions according to the likelihood that they represent corrupt or fraudulent behaviour prioritizing further reviews of contracts or transactions.

In contrast to the techniques that utilize structured data, there exist other methods that can aid in assessing risks in unstructured data. Text mining is one such approach that can be employed to recognize patterns in unstructured data, including emails, reports, and social media, and

derive valuable insights from them (Henderson & Hammersburg, 2020). In the context of evaluating corruption risks in infrastructure, text mining can also be used by a line ministry to evaluate internal fraud or corruption risks by scanning emails or social media to detect red flags such as procurement officials spending beyond their means or using specific keywords. To increase the effectiveness of text analytics, entities may refer to the fraud triangle to compile a list of keywords specific to the industry, relevant fraud risks, and the dataset at hand (Association of Certified Fraud Examiners, 2022).

2

FINANCIAL FRAUD AND ANOMALY DETECTION IN ACCOUNTING

This thesis aims to support the idea that data analytics can help find unusual financial transactions and prevent the risk of financial fraud. Financial fraud refers to any deliberate and deceptive action taken by an individual or group of people to obtain money or assets in an illegal or unethical manner. This chapter focuses on financial fraud and anomaly detection in accounting and explores various forms of fraudulent activities, including misappropriation of assets, financial statement fraud, and corruption. By understanding these types of fraud, organizations can better assess the risk of fraud within their operations. The chapter also introduces a Fraud Risk Assessment Framework, which provides a structured approach to identify, analyse, and mitigate fraud risks. By implementing this framework, organizations can proactively assess their vulnerability to fraud and take appropriate measures to prevent and detect fraudulent activities. In this chapter will be emphasized the importance of effectively assessing the risk of fraud and highlights the significance of anomaly detection techniques in identifying suspicious patterns or behaviours that may indicate fraudulent activities. Through a comprehensive understanding of financial fraud and utilizing the Fraud Risk Assessment Framework, organizations can enhance their ability to safeguard their assets, maintain the integrity of financial statements, and mitigate the risks associated with fraudulent activities.

2.1. Occupational Frauds

There are various types and categories of fraud, with new schemes emerging each day. Some common types of fraud include credit card fraud, investment fraud, insurance fraud, and cybercrime. Investment fraud involves misleading or deceptive investment practices, while insurance fraud involves making false claims to an insurance company. Cybercrime, on the other hand, consists of various schemes that include hacking, phishing, and malware, among

others. In this chapter, the focus will be on Occupational Fraud, which term refers to frauds that are committed by individuals against the organizations that employ them, according to the annual report "Occupational Fraud 2022: A Report to the Nations" of ACFE (Association of Certified Fraud Examiners).

The ACFE in the 2012 Report to the Nation outlines the three categories of occupational fraud and their subcategories in the following figure:

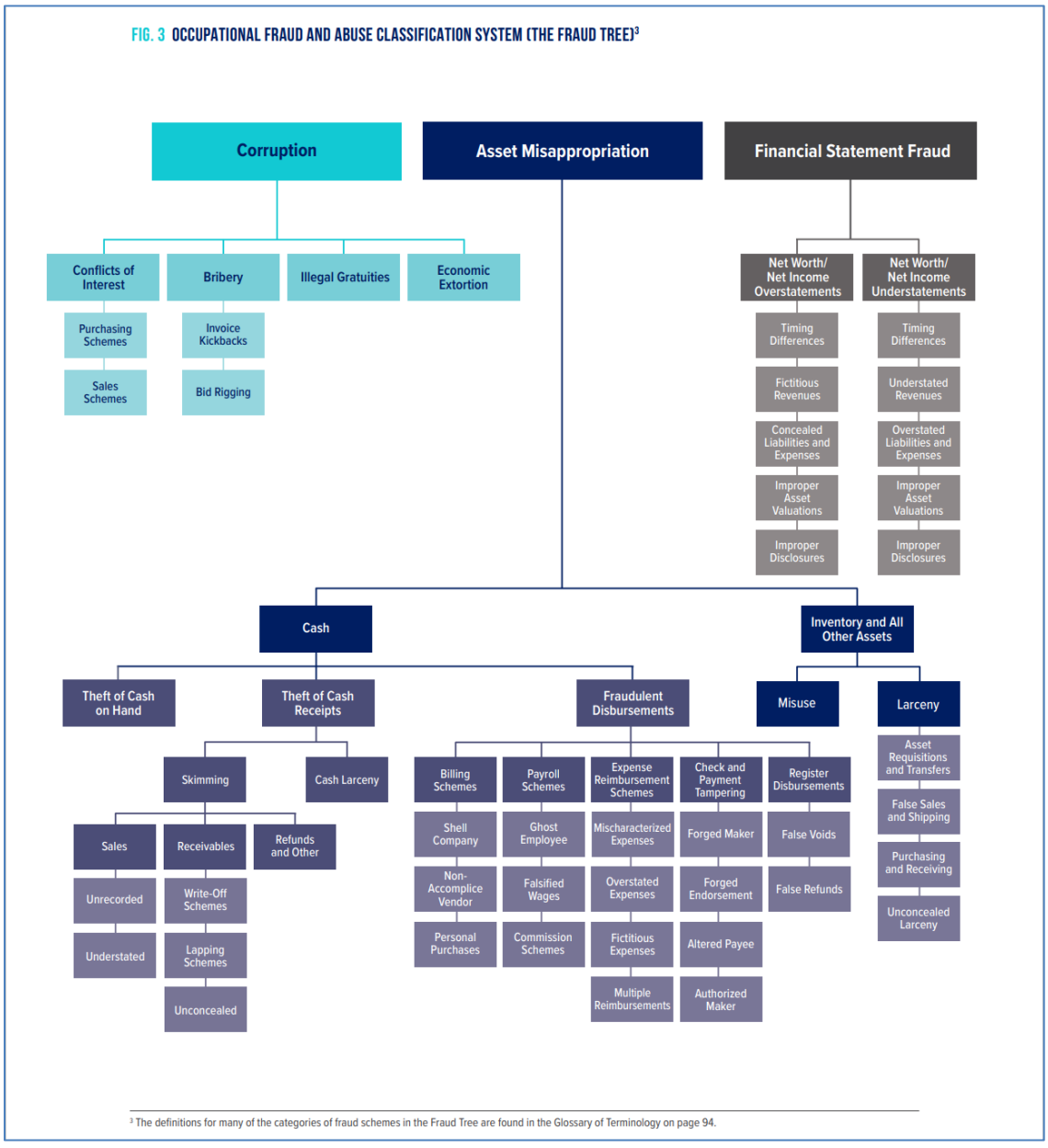


Figure 3: Occupational Fraud and abuse classification system (ACFE, 2022)

2.1.1. Misappropriation of assets

As the figure below shows, obtained from the annual "Report to the Nation" issued by ACFE in 2022, 40% involved more than one of the three primary categories of occupational fraud, 32% of fraudsters committed both asset misappropriation and corruption schemes as part of their crime while 2% misappropriated assets and committed financial statement fraud.

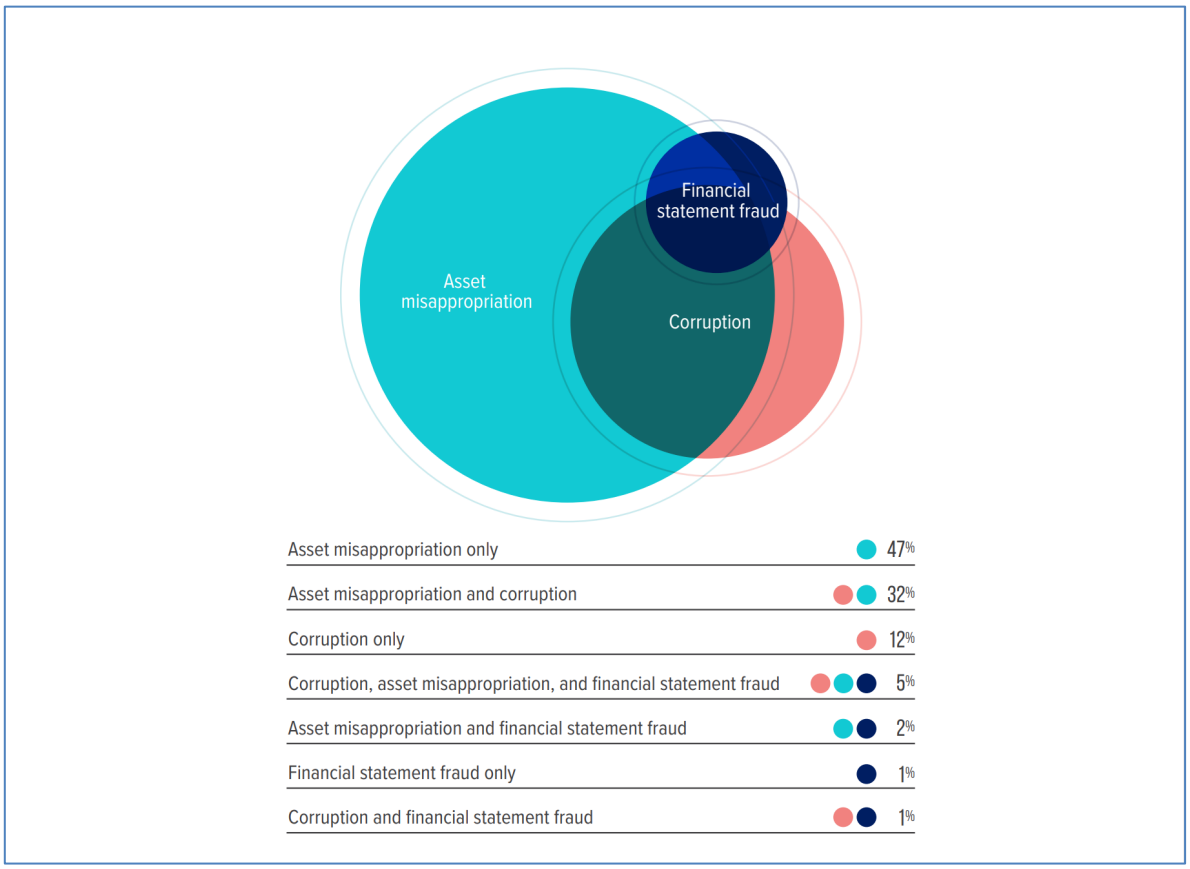


Figure 4: Frequency of fraudsters committing more than one type of occupational fraud (ACFE, 2022)

Asset misappropriation, which is a form of fraud that involves the theft or misuse of an organization's assets by employees or other insiders, is the most common with 86% of cases falling under this category.

It is a pervasive category of fraud that has significant financial and reputational impacts on businesses across different industries, however, tends to contend with the lowest median loss at USD 100,000 per case (“Occupational Fraud 2022: A Reports to the nations of ACFE).

Asset misappropriation can occur in various ways, including embezzlement, skimming, larceny, and false invoicing. Embezzlement is the fraudulent misappropriation of funds by an employee entrusted with their management, while skimming involves the theft of cash or inventory at the point of sale. Larceny, on the other hand, is the theft of an organization's assets without the employee being entrusted with their management. False invoicing refers to the creation of fake invoices to obtain payment for goods or services that were never supplied.

The ACFE also identified that approximately 85% of all asset misappropriation cases involved 11 the misuse or theft of cash. Indicators of fraud are commonly known as "red flags". Research conducted indicated the red flags for misappropriation of assets:

1. Accounting anomalies, such as faulty journal entries, inaccuracies in ledgers, or fictitious documents.
2. Internal control overrides and breakdowns.
3. Analytical fraud symptoms, including procedures or relations that are unusual or too unrealistic to be plausible. For example, transactions or events that happen at odd times or places; that are performed by or involve people who would not normally participate; or that include odd procedures, policies or practices. They may also include transaction amounts that are too large or too small.
4. Lifestyle indicators are observed in individuals who engage in fraudulent activities, as they tend to fulfil their immediate needs initially and subsequently begin to enhance their lifestyles gradually. Unusual behaviours of people (people who are involved in fraud often feel stress and, as a result, change their behaviours to cope with this stress).
5. Tips and complaints that something is suspicious.

2.1.2. Financial statement fraud

Financial statement fraud is the least frequent form of occupational fraud but by far the costliest. Financial statement fraud schemes, in which the perpetrator intentionally causes a material misstatement or omission in the organization's financial statements, are the least common (9% of schemes) but costliest (USD 593,000) category (ACFE, 2022). This type of fraud usually has a greater probability of giving rise to a material misstatement and of being committed by upper management.

The research concluded and enumerated fraudulent behaviour of the following types:

1. Inappropriate accounting reallocations, including transfers from flows to stocks. For example, significant transfers were made from what was effectively a suspense expenditure account, "Prepaid Capacity Costs" to a "Construction in Progress" account, which was treated as capital expenditure.
2. Accounting treatments designed to influence disclosure rather than recognition. For example, line costs were transferred to accounts that rolled up into "Selling, General and Administrative Expenses (SG&A)."
3. Adjustments that may have not changed the reported profits but did change the allocation between gross and net profit disclosures.
4. Inappropriate journal entries are often accompanied by failures in documentation and breaches in normal internal controls.
5. Adjustments are almost universally being carried out at the corporate level. In many cases, however, these "top side" adjustments made at the corporate level required adjustments in operating divisions and international operations.

2.1.3 Corruption

The Organisation for Economic Co-operation and Development (OECD) has suggested the The following definition of 'corruption': *the active or passive misuse of the powers of public officials (appointed or elected) for private financial or other benefits.* The third category, corruption, encompasses various offences such as collusion with a vendor to make false

payments for goods or services that were never delivered, collusion with a health care provider to create false health insurance claims, kickbacks, in which an employee receives payments from a third party in exchange for business advantages or bribery, in which an employee uses company funds to provide benefit to another business or individual in exchange for business advantage or personal gain.

In terms of frequency and losses, it occupies an intermediate position. Corruption schemes are found in approximately 50% of cases and typically result in a median loss of USD 150,000. Corruption can be particularly damaging as it undermines trust among stakeholders, erodes the reputation of the organization, stifles competition and can lead to significant financial losses. Companies can mitigate the risk of corruption by implementing a strong code of ethics, promoting transparency and accountability, training employees on anti-corruption measures, and conducting regular audits and investigations to identify and prevent instances of corruption.

Allocating compliance resources efficiently is crucial for companies due to their limited nature. Conducting a comprehensive risk assessment of corruption not only enhances efficiency but also adds credibility to the company's anti-corruption compliance endeavours. By undertaking such an assessment, a company positions itself to demonstrate diligent care in evaluating its risks, even when unforeseen issues arise.

The initial stage of the corruption risk assessment should primarily concentrate on identifying the actual risks associated with the company's operations. This involves evaluating factors such as the nature of its operations, the revenue generation process, the extent of business conducted with governmental entities, the utilization of agents and intermediaries, the countries of operation, the regulatory landscape, and other relevant considerations. The subsequent stage involves assessing the company's existing policies and controls aimed at mitigating corruption risks and analysing their effectiveness or any potential gaps. This evaluation allows for a comprehensive understanding of the current risk mitigation measures in place. The third stage entails developing a plan to establish an anti-corruption compliance program that is both effective and efficient. This plan should be based on the identified risks,

the existing controls, and the additional resources available to ensure reasonable assurance of compliance.

A thorough risk assessment instils confidence by ensuring that corruption risks are adequately identified and analysed. Additionally, more robust risk assessments may involve transaction testing at both the corporate level and in high-risk locations. The output of such assessments should comprise detailed recommendations for modifying the design and implementation of the anti-corruption compliance program to address identified risks effectively.

In organizations with anti-fraud programs, instances of fraud can still occur. To gain a deeper understanding of the factors contributing to occupational fraud, participants in the survey were asked to identify the primary internal control weakness that allowed the fraud to take place. The predominant factor identified in the study was the absence of sufficient internal controls, with 29% of victim organizations lacking adequate measures to prevent fraud. Additionally, 20% of cases involved the override of existing internal controls, indicating that the victim organizations had implemented protective mechanisms, but the perpetrators found ways to bypass them. This data highlights the potential for nearly half of the fraud cases in the study to have been prevented with a more robust anti-fraud control system in place.

The levels of authority within an organization influence individuals' access and influence, consequently affecting their ability to commit fraud. The analysis examined how the internal control weaknesses varied based on the position of the perpetrators. Unsurprisingly, schemes perpetrated by owners and executives were most commonly associated with a deficient "tone at the top." Among staff-level employees and mid-level managers, the most prevalent control weakness was the lack of internal controls, accounting for 34% and 29% of cases, respectively.

2.2 Assess the Risk of Fraud

Even within organizations that have implemented anti-fraud programs, instances of fraud can still occur. The Report for the Nation 2022 from ACFE exposes that the predominant factor of internal control weakness was the absence of a sufficient number of internal controls, as 29% of victim organizations lacked adequate measures to prevent fraud. Furthermore, in 20% of the cases, internal controls were overridden, signifying that the victim organizations had established protective mechanisms, but the perpetrators managed to bypass them. These findings indicate that a stronger system of anti-fraud controls could have potentially prevented nearly half of the frauds observed in the study.

The levels of authority held by individuals within an organization play a significant role in determining their access and influence, subsequently influencing their capacity to commit fraud. An analysis was conducted to examine how internal control weaknesses varied based on the position of the perpetrator. Schemes perpetrated by owners and executives were most associated with a deficient "tone at the top." Among staff-level employees and mid-level managers, the lack of internal controls emerged as the most prevalent control weakness, accounting for 34% and 29% of cases, respectively.

Arguably, the complete elimination of fraud risk in any given area is an unfeasible feat, except through total avoidance of said area. Companies may opt not to engage with specific vendors or purchasers, abstain from acquiring assets requiring high levels of protection, or refrain from expanding or conducting business in politically unstable nations. Conversely, if the risk assessment deems the cost to outweigh the benefits, an exit strategy may be selected. Some risks may be deemed acceptable and assumed without additional control features as the cost of implementation would surpass the expected loss. For instance, banks issuing credit cards can potentially reduce fraudulent charges through the incorporation of new high-tech security measures; however, the financial or customer inconvenience costs may outweigh the expenses incurred through fraudulent transactions. Fraud is a cost of conducting business that requires a cost-to-benefit or return-on-investment analysis. A risk assessment aids in

determining the necessary level of controls to implement while striking a balance between acceptable risk tolerance and the cost of reducing the risk.

$$Risk = Impact \times Probability\ (threats\ and\ vulnerabilities)$$

Equation 1: Risk formula

In general, companies strive to mitigate risks by implementing controls that may consist of preventative, monitoring, or detection measures. Additionally, risk reduction can be achieved by purchasing insurance or mandating bonds for certain employees. In some instances, it may be determined that the costs of preventing fraud in a specific area outweigh the benefits. However, investments in measures to detect fraud may be an acceptable risk due to lower costs and the reduced likelihood of significant losses. In any risk assessment, detective measures must also be factored in. It is primarily a management decision as to the level of response to adopt in response to the risk of fraud. This decision will be primarily based on the rationale behind conducting the fraud risk assessment in the first place. Was it due to regulatory or audit requirements, the need to assess the internal control system, or to reduce costs associated with fraud? A risk assessment serves to identify potential areas of fraud, whether they are internal or external, direct or indirect, and the degree of vulnerability or likelihood of occurrence.

The probability component is determined by various factors, such as:

- The industry or nature of the business
- The values and ethics of senior management and employees
- Internal controls—preventive and detective
- Business environment—local versus multinational, small versus large, brick-and-mortar versus Internet, geographic location, economic conditions
- Industry trends
- History
- Resources
- Internal control

  Complexity

- Volume
- Standards

2.2.1 Fraud Risk Assessment Framework

In addition to identifying potential areas of fraud, risk assessments must also consider the possible way of mitigating risks and improving the actual risk controls implemented.
A valuable resource for organizations is the *fraud-risk assessment framework* offered by The American Institute of Certified Public Accountants. This framework possesses the capability to function as an advantageous model, incorporating the existing controls that have been implemented by the organization, as well as providing potential future strategies to address and mitigate risks. With the aid of this framework, employees and the corresponding top management can acquire a comprehensive understanding of the primary sources of fraudulent activities within their company's operational processes, along with the specific business areas that are most vulnerable to such risks. The distinctive feature of this framework lies in the requirement to input predefined responses in certain columns, such as Likelihood (remote, reasonably possible, or probable), and Significance (immaterial, significant, or material), while other columns necessitate a qualitative description.

Below it is possible to observe an overview of the framework including the two examples of identifying the Fraud Risks "*Holding books open*" and "*Roundtrip transactions*":

Identified Fraud risks and Schemes (1)	Likelihood (2)	Significance (3)	People and/or Department (4)	Existing anti-fraud Controls (5)	Controls effectiveness assessment (6)	Residual risks (7)	Fraud risk response (8)
Financial Reporting • Holding books open	Reasonably possible	Material	Accounting	The standard monthly close process Reconciliation of invoice register to the general ledger Established procedures for shipping, invoicing, and revenue recognition. Established process for consolidation	Tested by management	Risk of management override	Testing of late journal entries Cut-off testing by IA
• Roundtrip transactions	Remote	Insignificant	Accounting	N/A	Tested by internal auditors	Risk of management override	N/A

Table 2: Fraud Risk Assessment Framework (AICPA, 2020)

The framework is based on 8 fields described in detail below:

1. **Identified Fraud Risks and Schemes**: This column aims to present a comprehensive inventory of the various fraud risks and schemes that an organization may encounter. The list of potential risks and schemes will vary depending on the organization's characteristics and circumstances, and should be developed based on several sources, including (a) research on the industry and its trends, (b) feedback from employees and other stakeholders through interviews or surveys, (c) brainstorming sessions involving experts in relevant fields, and (d) information received through the organization's whistle-blower hotline. By compiling and

analysing this information, the organization can better understand its exposure to fraud and take steps to mitigate and prevent it.

2. Likelihood of Occurrence: Creating an effective fraud risk management program involves evaluating the probability of the potential fraud risks identified, which allows the organization to establish appropriate anti-fraud measures for the risks that are considered most probable. To conduct this assessment, it is sufficient to categorize the likelihood of each risk as remote, reasonably possible, or probable. This risk evaluation approach enables the organization to prioritize its resources and efforts and allocate them where they are most needed to prevent, detect, and respond to fraud incidents.

3. Significance to the Organization: Assessing the significance of fraud risks to an organization requires analysing both quantitative and qualitative factors. While some fraud risks may not have a significant direct financial impact on the organization, they could still have a profound effect on its reputation, making them more substantial risks. Therefore, it is essential to evaluate the significance of fraud risks based on both their financial and non-financial impacts. To perform this assessment, risks can be classified as immaterial, significant, or material, depending on their overall impact on the organization's operations and objectives. This approach ensures that the organization can prioritize its resources and focus on mitigating the risks that pose the greatest threat to its sustainability and success.

4. People and/or Department Subject to the Risk: It is critical to identify the people inside and outside the company who may be vulnerable to fraud risks during the identification and assessment of fraud risks. Such information will help to tailor the organization's reaction to fraud threats, including putting in place the correct job segregation, setting up the right review and approval procedures, and putting in place proactive fraud audit procedures.

5. Existing Anti-fraud Internal Controls: The next stage is to match the existing controls to the pertinent fraud risks after identifying and assessing the likelihood and magnitude of the fraud risks. It's crucial to remember that this mapping procedure shouldn't happen before the initial evaluation of fraud risks without taking internal controls into account.

6. Assessment of Internal Controls Effectiveness: Organizations need to establish a procedure to assess the effectiveness of their identified controls in effectively operating and addressing fraud risks as intended. Companies that fall under the scope of Section 404 of the U.S. Sarbanes-Oxley Act of 2002 are required to have such a process in place. However, organizations that are not subject to Sarbanes-Oxley should still consider implementing suitable review and monitoring procedures to ensure confidence in the proper functioning of their internal control framework.

7. Residual Risks: Upon careful examination of the internal control structure, it may be concluded that certain fraud risks are not sufficiently mitigated for various reasons. These reasons may include the absence of properly designed controls to address specific fraud risks or the ineffective operation of identified controls. To address these remaining risks, the organization should thoroughly evaluate them during the development of the fraud risk response.

8. Fraud Risk Response: The organization should assess residual risks and develop fraud risk responses to effectively address any remaining risks. These responses may involve implementing extra controls, implementing proactive fraud auditing techniques, or mitigating the risk by discontinuing the activity. The organization should carefully consider which combination of these options is most appropriate for addressing the identified risks.

2.3 The Fraud Triangle and reporting methods

The motivation behind an employee's engagement in fraudulent activities is a complex matter. A well-established theory, commonly referred to as the "fraud triangle," offers insights into the factors contributing to fraud. This theory, initially documented in 1950 by UCLA professors who conducted interviews with inmates at the Illinois state prison in Joliet, highlights three key elements: pressure, opportunity, and rationalization. These factors

collectively contribute to an individual's likelihood of committing fraud. Below is a schematic representation of the triangle theory:

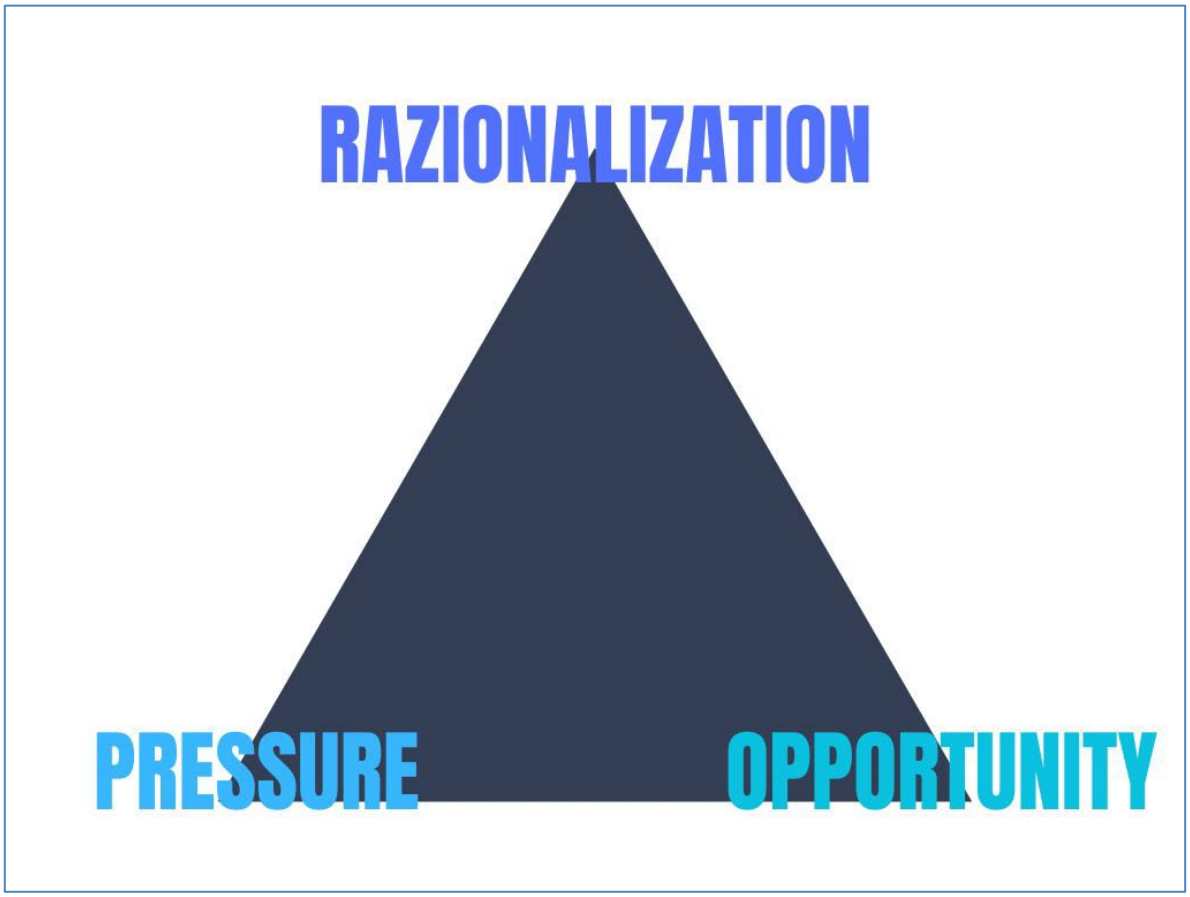


Figure 5: The Fraud Triangle

Pressure assumes a consequential role across diverse domains of life, and its influence can act as a driving force behind employees resorting to fraudulent practices. The desire for financial stability or the presence of personal difficulties, such as escalating medical bills or an extravagant standard of living, can serve as compelling factors that instigate individuals to engage in fraudulent behavioural. Furthermore, organizations themselves can contribute to this pressure by imposing impractical performance targets or fostering an environment that encourages the suppression of unfavourable outcomes, all to preserve a positive public image.

The *opportunity* represents another significant factor in understanding the occurrence of fraud and is often more evident than pressure. For the smooth functioning of any organization, a certain level of trust must be extended to its employees. However, the element of opportunity arises when an entrusted employee breaches that trust by exploiting their position to engage in fraudulent activities.

Rationalization constitutes the third factor and generally serves as an individual's justification for committing fraud. In numerous instances, perpetrators convince themselves that their actions are mere temporary borrowings from the organization. Alternatively, they may rationalize their behaviour by adopting thoughts such as "they won't notice the missing funds" or "the organization deserves the consequences it is facing." Such rationalizations enable individuals to morally reconcile their fraudulent actions and mitigate any guilt or remorse.

Starting from the Fraud Triangle theory, David T. Wolfe and Dana R. Hermanson in the book "*The Fraud Diamond: Considering the Four Elements of Fraud*" extended the fraud triangle framework by introducing a fourth component, namely the *individual's capability*, to enhance the effectiveness of fraud prevention and detection. While opportunity provides the gateway for fraudulent behavioural, motivation and rationalization can lure individuals towards it. However, the crucial aspect lies in the individual's capability to recognize the open gateway as an opportunity and exploit it repeatedly, rather than just a one-time occurrence.
The elements comprising the fourth dimension, so individual capability, are as follows:

— **Position/Function**: Individuals in influential roles such as CEOs or divisional presidents possess the authority to determine the timing of contracts or deals, leveraging their positional power.
— **Intellectual Acumen**: Fraudsters demonstrate intelligence in understanding and exploiting weaknesses in internal controls. They effectively utilize their position, function, or authorized access to their maximum advantage.
— **Confidence/Ego**: Successful fraudsters exhibit a strong ego and unwavering confidence in their ability to avoid detection. They believe they can talk their way out of trouble if confronted.

— **Coercion Skills**: A proficient fraudster possesses the ability to coerce others into participating in or concealing fraudulent activities. They can convincingly persuade others to comply or turn a blind eye.
— **Effective Deception**: A skilled fraudster excels at lying convincingly and consistently. To evade detection, they must confidently deceive auditors, investors, and other stakeholders, while maintaining consistency in their fabricated story.

It is important to explicitly assess the capabilities of top executives and key personnel. Emphasizing capability necessitates that organizations and their auditors develop a deeper understanding of individual attributes and proficiencies among employees. When evaluating fraud risk or endeavouring to prevent and detect fraud, members of the audit committee, corporate accountants, or auditors should concentrate on the personality traits and competencies of key executives and individuals responsible for high-risk areas. Conducting routine background checks on new hires can help uncover any prior criminal convictions. In the assessment of individuals' attributes and abilities, several approaches to gathering information can prove beneficial. There is no substitute for investing time with the individual, as frequent interactions encompassing various business and social contexts can yield a meaningful assessment of their capabilities.

Also, when an individual's capabilities pose a significant risk factor, it is imperative to respond with stronger controls or enhanced audit testing. For instance, if the sales vice president exhibits an excessively aggressive and competitive nature, along with an obsession with meeting monthly sales quotas, it may be necessary to implement stricter controls regarding revenue recognition or conduct expanded testing of sales during the annual audit. Furthermore, introducing a periodic rotation of routine yet crucial tasks among staff members can reduce opportunities for fraud stemming from long-term familiarity with specific functions and their associated controls.

During this phase of response, it is crucial to direct special attention to situations that not only involve incentive and rationalization but also encompass the combination of opportunity and capability. In essence, the question to be asked is, "*Are there any vulnerabilities to fraud*

that can be exploited by individuals possessing the necessary skills?" If such vulnerabilities exist, these areas present an exceptionally high risk since all the elements required for a fraud opportunity to materialize are present.

Another important aspect to consider is that reassessing the capabilities of top executives and key personnel could prevent the risk of committing corruption fraud. The assessment of capability and the subsequent response should not be regarded as one-time endeavours. It is necessary to continuously update the assessment of capability and the corresponding response for two primary reasons. Firstly, individuals have the potential to develop new capabilities over time, particularly as they progress in their careers and gain professional growth. Merely because someone lacked the power or knowledge to engage in fraudulent activities in the past does not guarantee that they will not acquire such power or knowledge in the future. Their capacity to commit fraud may increase, necessitating additional controls or scrutiny.

Secondly, organizational processes, controls, and circumstances change over time. Consequently, certain individuals may become better suited to commit fraud in the new environment, even if they were previously incapable under different conditions. For instance, consider a company that has recently implemented a sophisticated new IT system. This system may render employees with limited digital proficiency incapable of exploiting its controls. Conversely, individuals with strong IT skills may see their capability to commit fraud enhanced by this change. This newfound capability must be taken into account, and appropriate responses should be implemented accordingly.

2.3.1 Motivating Employees to speak up: Strategies for promoting organizational voice

ACFE offered an interesting analysis regarding the method of tip submission. The subsequent figure illustrates a considerable decline in the utilization of telephone hotlines in the past while the usage of email and web-based/online reporting has surpassed it. These findings highlight the diverse and evolving preferences of whistleblowers when it comes to reporting

fraud, particularly on online and electronic platforms. Consequently, organizations are advised to maintain multiple channels to accommodate various forms of fraud reporting.

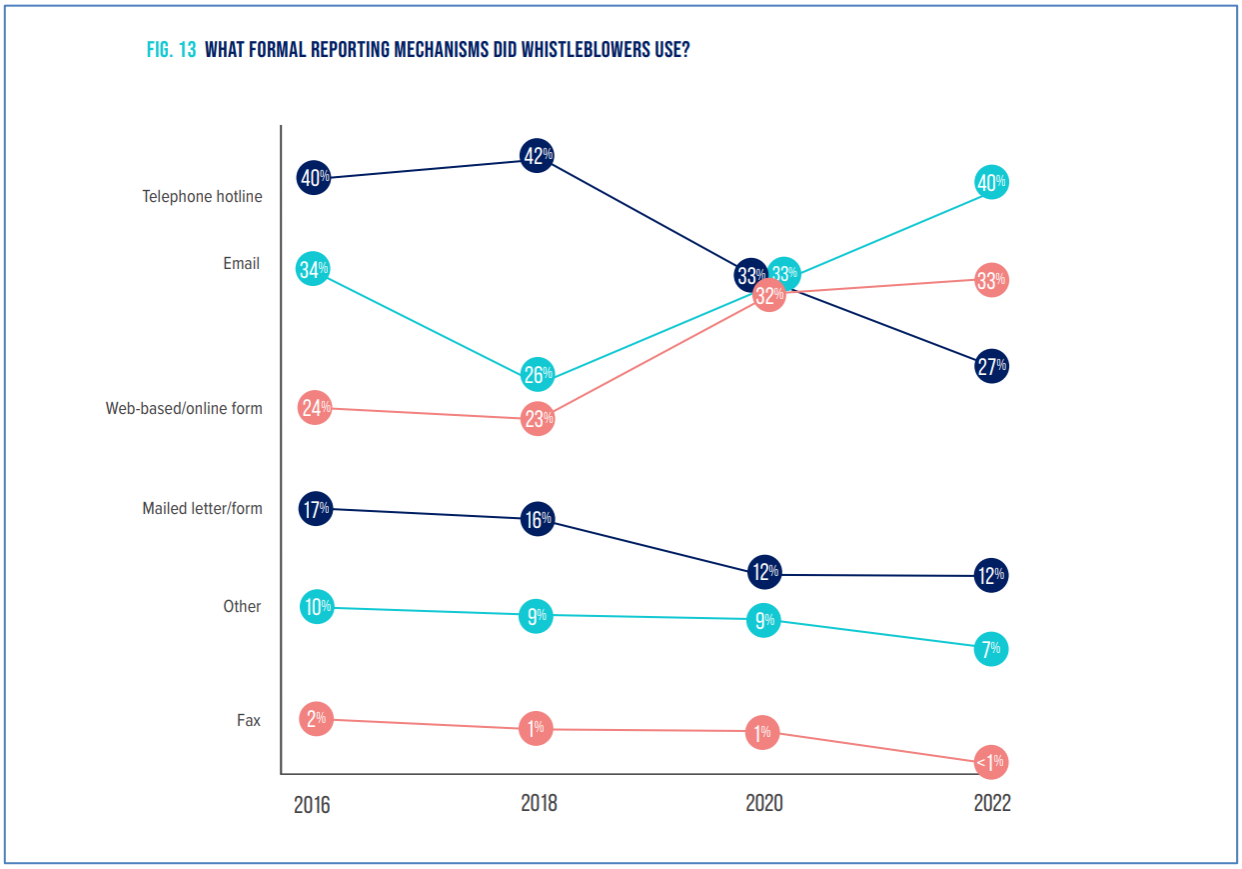


Figure 6: Formal reporting Mechanisms used by whistleblowers, ACFE.

Also, not all reports regarding suspected fraud are channeled through formal reporting mechanisms. Some reports are informally conveyed to individuals within the organization. Whistleblowers who opt not to use hotline mechanisms commonly share their concerns with their immediate supervisors (30%). However, it has been shown that whistleblowers may approach various other parties, including executives, internal audit teams, fraud investigation units, or their colleagues. Given that reports can potentially be received by anyone within an organization, it is crucial to provide comprehensive guidance to all staff members regarding the organization's protocols for handling fraud allegations and what steps to take if they receive a report concerning suspected fraud.

To prevent white-collar crime, both employees and managers mustn't remain confronted with financial fraud.

First, organizations tend to excessively rely on a limited range of compliance and control tools to mitigate misconduct and promote disclosure. Despite the potential benefits of these tools, the outcomes were largely unsatisfactory. A limited number of individuals dared to voice their concerns. The underlying reasons for this phenomenon are intricately tied to the impact of sanctioning systems, which tend to distort our cognitive processes and deter us from pursuing morally right actions. The apprehension of facing consequences such as the jeopardization of career advancements, incentives, or salary increments often triggers a self-preservation mindset. Consequently, individuals tend to adopt a business-oriented perspective rather than a moral one when making decisions about their course of action.

Also, when organizations formulate compliance policies and codes of conduct, they intend to evoke a sense of obligation and ethical responsibility within individuals, encouraging them to report any observed instances of misconduct. However, these measures fail to effectively motivate a significant number of individuals to come forward and voice their concerns. There is the erroneous assumption that specific demographic groups or personality traits, such as extroversion, optimism, or leadership qualities, inherently predispose individuals to speak up against wrongdoing. Research in behavioural science indicates that neither gender nor extroversion is a reliable predictor of whistleblowing behavioural, regardless of the industry or occupation involved. The notion that there exists a specific gender, disposition, age, or personality type that possesses a unique ability to speak up is unfounded.
The main interesting strategies addressed by experts seem to be the following:

1) **Segregation of duties**. One of the primary preventative measures of utmost importance is the segregation of accounting responsibilities, particularly those associated with executing outgoing payments. It is imperative to assign distinct employees to perform the tasks of approving, recording, and reporting transactions. Furthermore, it is crucial to ensure that the individual responsible for generating payment checks or authorizing invoices is not entrusted with the task of signing checks or initiating online payments.

Likewise, the staff member responsible for making bank deposits should not be tasked with reconciling the organization's bank statements. If the scale of operations does not permit complete segregation of duties, alternative measures can be considered. These include periodically rotating staff members across various duties or involving a board member to oversee the process. Additionally, implementing a mandatory vacation policy can pose challenges for fraudulent employees attempting to conceal their illicit schemes.

2) **Redesigning reporting tools**. Redesigning reporting tools is imperative for organizations as it allows them to separate the motivation behind speaking up from the management of the issues raised. These two aspects possess distinct objectives and require different modes of thinking. It is essential to thoroughly assess and enhance the effectiveness of their existing reporting mechanisms. For instance, they should consider investing in training programs aimed at empowering employees to identify and appropriately respond to instances of misconduct that they witness as bystanders. Often, individuals are unsure how to react in such situations, whether the wrongdoing is trivial or fraudulent, accidental or intentional. Notably, similar training initiatives in educational institutions have proven successful in enabling people to recognize situations that may lead to sexual assaults, consequently reducing their occurrence. Managers ought to review zero-tolerance policies that inadvertently discourage employees from admitting their mistakes or shortcomings. These policies may create an environment where individuals are reluctant to disclose instances where they have overcharged customers or conducted trials resulting in unsatisfactory outcomes. Simplicity should be prioritized when establishing reporting procedures. Hotlines must guarantee cyber-security, external accessibility, anonymity, and ease of use. Moreover, these channels for disclosure should facilitate collective reporting, as my research indicates that most individuals prefer reporting instances of misconduct alongside allies rather than doing so alone. When individuals bring forth issues directly to you, it is crucial to refrain from blaming the messenger. Human nature tends to instinctively reject negative news, often leading to an "auto-immune response" where such messengers are disregarded, as claimed by experts in the field of whistleblowing. Instead, take a moment to reflect on how promptly, sensitively, and effectively you address the raised concerns. It is important to note that investing in

reporting tools alone will yield limited results without cultivating an organizational culture based on integrity.

3) **Intensive training for employees**. Organizations that implement antifraud training programs tend to experience reduced financial losses and shorter durations of fraudulent activities compared to those without such programs (ACFE, 2022). It is therefore crucial for businesses and other organizations to provide targeted training on fraud awareness, not only for managers but also for employees. Also, the ACFE recommends that organizations educate their staff on what actions constitute fraudulent behavioural, how fraud negatively impacts the entire organization and the proper procedures for reporting suspicious activities. Managers and employees should also receive training on recognizing behavioural red flags exhibited by potential perpetrators and be encouraged to remain vigilant. Examples of red flags include employees who seem to be living beyond their means or those who consistently refuse to take time off. Moreover, certain insurance providers may offer discounts if a significant majority of staff members undergo specific anti-fraud training programs.

3

DATA ANALYSIS FOR PREVENTING FRAUD

In today's business environment, fraudulent activities pose a significant risk to companies, with increasingly sophisticated fraud schemes causing substantial financial and reputational damage. Fraudulent activities can be challenging to detect and prevent, making it necessary for companies to use advanced techniques to identify and mitigate the risks associated with fraudulent activities. One such technique is the use of data analytics. By leveraging data analytics tools and techniques, companies can identify patterns and anomalies in large volumes of data to detect potential fraud. These analytics tools can provide insights into employee behavioural, customer transactions, and supplier relationships, enabling companies to identify fraud risks and develop targeted risk management strategies. Successful analytics-based fraud detection involves a combination of skilled personnel, advanced technology, and effective processes. The use of analytics can help companies stay ahead of fraudsters and minimize the risk of financial loss, regulatory penalties, and reputational damage. In summary, the successful use of analytics can be a valuable tool for companies looking to detect and prevent fraudulent activities, enabling them to safeguard their business operations, assets, and reputation.

This chapter focuses on the importance of data analysis in preventing fraud. With the increasing amount of data being generated in today's digital age, fraud has become more sophisticated and challenging to detect. Data analysis techniques, such as data mining and analytics, can help organizations identify potentially fraudulent activities by analyzing patterns, trends, and anomalies in data. However, to effectively prevent fraud, it's essential to establish trust in the accuracy and reliability of the data being analyzed. This chapter will discuss the four anchors of trust, including data quality, transparency, expertise, and ethics, and how they can help build trust in the analytics process. Additionally, the chapter will address the low usage of data analytics in fraud prevention and its reasons and implications.

Overall, this chapter aims to highlight the power of data analysis in preventing fraud and the importance of establishing trust in the analytics process.

3.1 The power of Trust in Analytics

Currently, a multitude of important decisions that affect individuals, businesses, and societies rely on complex analytics. The emergence of artificial intelligence (AI) promises faster and better decision-making capabilities, leading to substantial investments across business sectors, including consumer targeting, risk assessment, and self-driving cars. However, as more and more decisions depend on data and analytics (D&A), questions are arising about the trust placed in the data, analytics, and controls that underlie this new approach. Unfortunately, only a few organizations are considering the concept of trust in their D&A, and those that do tend to focus narrowly on accuracy, reliability, and security. With complex analytics, it is unclear how to determine whether a result is correct or whether automated decisions are making the right choices. Furthermore, what constitutes a "right" decision, and who is qualified to make that judgment? As analytics becomes more widespread across all sectors, including regulators, policymakers, and those who protect consumer rights, trust becomes an increasingly pressing issue. A heightened focus on trust is emerging, and as algorithms make more decisions on behalf of people, it will become a defining characteristic of D&A. The growing recognition of the high stakes involved underscores the need to address trust issues.

The increasing financial power of analytics and their potential political and moral implications are causing concern. Those who can ensure trusted analytics will have an advantage in decision-making and strong customer relationships. While well-designed algorithms often make more accurate decisions than humans, complex and opaque data and analytics pose a risk for poor design, errors, underperformance, misuse, and unintended, exploitative consequences. Individuals naturally desire to place their trust in accurate data and analytics, employed in a manner that they comprehend, by individuals whom they trust, for a purpose that they endorse and find valuable. However, providing satisfactory answers

to these queries can be challenging, as there is no universally accepted framework for assuring in this context. (KPMG, 2016).

The value of trusted analytics varies depending on the context, with medical and financial industries being particularly important. Previously, concerns about trusted analytics were mainly limited to high-tech companies and high-risk industries. However, with the increasing complexity of the analytics ecosystem and democratization of analytics power, regulators are paying attention and becoming more concerned about the potential for misuse of consumer data and abuse of analytics. Trust is moving up the agenda, and distributed innovation is raising concerns about who controls or has access to our data and analytics. The new environment weakens transparency rather than strengthening it.

Measuring and quantifying trust is a challenging task, making it difficult to ascertain the exact value of trusted analytics. However, organizations that earn a reputation for trustworthiness may reap the benefits of improved results, stronger relationships, and better decision-making. The impact of trust is particularly evident when consumers are involved, as the loss of trust can significantly impact a brand's reputation and sales performance. Despite this, many consumers remain unaware of how organizations use their data, and most businesses have little incentive to be transparent beyond regulatory compliance.

As consumers become more aware of the power their data commands, the value of the trust will increase and become more tangible. For instance, some automotive insurance carriers now offer customers the chance to save on premiums by allowing diagnostic devices to monitor their driving behavioural, demonstrating how consumers may be willing to trade privacy for benefits when they trust the organization.

3.1.1 Closing the trust gap: the four anchors of trust

In light of the increasingly complex ecosystem surrounding analytics, several questions are asked by literature on how organizations ensure trust throughout the analytics lifecycle. KPMG's 2016 article "*The Power of Trust in Analytics*" identifies four key anchors for trusted analytics.

The first is quality, which refers to the accuracy, provenance, freshness, consistency, and completeness of the data used in analytics and data management processes. Within numerous

organizations, concerns may arise regarding the "lineage" of data, which pertains to understanding the data's origin and the path it followed to reach the analytics system. Data quality is also influenced by factors such as consistency and completeness, making them crucial aspects to consider.

The second is accepted use, which requires organizations and analytics experts to ensure that the intended analytical approach is appropriate for the context in which it is being used and that the way the data is manipulated is appropriate and defensible. Crime statistics, for example, could be used as a proxy for economic vibrancy in a specific geographic area, but only if the right statistics are being leveraged and in the right context. Knowing when and how to appropriately apply data and analytics to various scenarios is a key competence for companies nowadays (KPMG,2016).

Once assurance is established in the first two anchors, executives and analytics experts will need to ensure that the analytics work and achieve their intended purpose — that the predictions and insights are accurate and reflect reality — the third anchor of trusted analytics. If decisions are made or consumers are targeted using unreliable predictions, it will rapidly erode consumer trust and undermine the confidence of executives who depend on these predictions to make well-informed decisions.

The fourth and final anchor of trust in analytics is ethics and integrity. In an era where data plays an increasingly central role in decision-making processes, organizations must prioritize the ethical use of data and predictions. This entails ensuring that the collection, analysis, and application of data are conducted in a manner that is transparent, unbiased, and non-discriminatory. Maintaining ethics and integrity is crucial to build and sustain trust with stakeholders, including customers, employees, and regulatory bodies. Organizations must be mindful of the potential reputational risks associated with unethical practices in data analytics, as the consequences can be severe and far-reaching.

The ethical considerations surrounding data analytics encompass a wide range of issues. One significant concern is the protection of individual privacy and data rights. As organizations collect and analyze vast amounts of personal data, it is essential to implement robust security measures and comply with relevant data protection regulations. Additionally, organizations must ensure that data analytics processes are conducted with utmost transparency, providing

clear explanations of the methodologies used and the basis for any predictions or decisions derived from the analysis. Furthermore, there is a growing recognition of the potential for bias and discrimination within data analytics. Biased algorithms or skewed datasets can perpetuate existing inequalities and lead to unfair treatment or decision-making outcomes. Organizations must actively address and mitigate these biases by regularly evaluating and auditing their analytics models and data sources. Additionally, fostering diversity and inclusivity within data analytics teams can help minimize the risk of unconscious biases and enhance the overall integrity of the analytical processes.

Given the rapidly evolving nature of the field, legislation, and regulation about ethics in data analytics are likely to emerge. Organizations should proactively stay informed about these developments and ensure compliance with emerging standards. By adhering to ethical principles, being transparent in their practices, and actively addressing biases, organizations can foster trust, mitigate reputational risks, and pave the way for responsible and trustworthy data analytics practices.

3.1.2 Implications for the analytical enterprise

Ensuring trust in analytics is not a singular event or a compliance requirement, but rather an ongoing and comprehensive effort that must encompass the entire organization. Building trust in analytics requires a lifecycle approach that spans from data sourcing and preparation to the generation of insights and value. This endeavour also affects multiple areas of corporate management, such as talent management, compliance, sourcing, and strategy development. As trust becomes essential, organizations and executives will need to continually assess their approaches and controls. The trust paradigm will undoubtedly evolve, and it is not yet clear how new and innovative uses of data and analytics will impact it. Therefore, maintaining vigilance over the anchors of trust is critical. To fully realize the benefits of analytics, trust in analytics must be addressed immediately at the societal, governmental, and consumer levels. This is a major business issue with significant implications. The eventual approach to engendering trust will vary across industries and contexts, and there is no one-size-fits-all solution. The debate about trust in analytics is

already underway, and businesses and consumers will need to participate in operationalizing and embedding trust concepts and controls into practice.

The strategies and methods employed to establish and foster trust will differ based on the industry and specific context. Industries handling personal health information or financial data, for instance, will require more expediency, rigorousness, and transparency compared to those involved in resource extraction. No universally applicable solution or approach fits all situations. Discussions surrounding trust in analytics are already taking place, primarily within academic and policy circles. However, businesses and consumers also have a role to play, especially in terms of effectively implementing and integrating trust concepts and controls into practical applications.

3.2 Low usage of such a powerful tool: reasons and implications

The low usage of analytics is a matter of concern because analytics can be an indispensable tool in the highly complex world of fraud detection. This is especially important considering the huge cost of fraud. For example, a 2016 global survey of over 40,000 certified fraud examiners revealed that fraud accounted for US$6.3 billion in losses, with the typical organization losing 5% of its revenues annually to fraud (2016 Report to the Nations on Occupational Fraud and Abuse, ACFE).

A considerable number of companies are not effectively utilizing analytics to identify fraudsters, and the reasons for this can vary. Some corporate leaders may not fully comprehend the benefits of analytics, while others may be hesitant to invest in its implementation due to cost concerns. Additionally, some decision-makers may believe that implementing advanced analytics is only worthwhile after a significant fraud incident occurs. This lack of adoption can be attributed to a "trust deficit," wherein there is a lack of confidence in the accuracy and effectiveness of the underlying data, analysis, and business interpretation in distinguishing legitimate transactions from fraudulent activity in a cost-efficient manner (KPMG, 2020).

This indicates a general distrust in the ability of fraud detection processes to identify untrustworthy employees and business partners. However, if these trust issues are properly

addressed, analytics can prove to be a powerful addition to any company's anti-fraud program, mitigating potential financial and reputational losses from fraudulent activity and communicating to potential fraudsters that the risks of getting caught are too high. This is why trusted analytics is an essential tool for reducing security and reputation risks.

3.2.1 Successful analytics requires high-quality components

The initial trust anchor pertains to incorporating processes that ensure the **quality** of the components involved in the analytics program, where employees play a significant role. The data that should be analyzed should be specifically relevant to identifying potentially fraudulent activities, suspicious transactions, or abnormalities in regular processes. This includes any information that could provide indications of fraud or raise concerns. Therefore, the sources of data for analysis should include the processes in which an employee could influence a transaction, such as employee expense reports, accounts payable and any transaction that includes the handling of cash. The data has to be accurate and up-to-date, and the sources of the data need to be known and understood.

The data analysis program needs to exhibit consistency and comprehensiveness. Its design should align with the specific task and be based on relevant processes, such as transaction types and involved functions. These principles apply to all forms of analytics, including the detection of fraud, particularly instances involving intentionally manipulated data. Due to the enormous quantity of data produced in contemporary times, it is reasonable to assume that analytics could be utilized to identify fraudulent activities. Most anomaly detection techniques, including those based on machine learning, rely on identifying unusual patterns in an otherwise homogenous population. Nevertheless, the effectiveness of these analytical methods, especially when fraud is infrequent, hinges on the capacity to determine what constitutes normal behavioural. A successful program for detecting fraud using analytics must encompass identifying anomalies while having a firm grasp of what is considered normal.

The failure of analytics-based fraud detection programs usually stems from a lack of understanding within the implementation platform regarding what constitutes normal

behavioural, rather than a lack of analytical rigour. It is easier to identify honest individuals who exhibit transparent behavioural than it is to detect those who engage in fraudulent activities. This concept can be compared to lowering the water level of a muddy river to better see the rocks on the riverbed, a principle successfully applied in lean manufacturing systems.

The second trust anchor pertains to the effectiveness of the process for analyzing transactions and ensuring that the output is accurate and useful. An effective anti-fraud analytics process aims to achieve a balance between generating an excessive number of red flags and too few, necessitating meticulous calibration and iterative refinement of the algorithm through trial and error. Achieving an optimal rate of fraud alerts could take several months in a large and complex organization, and data analysts must manage expectations to avoid frustration among decision-makers. If the process generates too many false positives, corporate leaders may lose confidence in it, and stakeholders could lose faith in their employer if each potential case is aggressively investigated. On the other hand, too few red flags could result in cases of fraud escaping detection, which could be equally or more harmful if executives begin to doubt the process's effectiveness and seek alternative methods. In general, detecting too many false positives is preferable because it demonstrates vigilance, even if the anomaly investigated does not lead to any significant outcome, which can build trust rather than erode it.

More companies are deploying data analytics for fraud detection. Yet, as observed earlier, the global survey of fraudsters found that only 3% of successful detections used analytics (KPMG, 2016). One reason for the gap is that the long-term operational control (trust anchor n.3) of the analytics processed may not have been established, let alone optimized with the result that the detection rate is less than expected. While it requires a high level of expertise and technology to integrate advanced analytics into business processes, such resources are indispensable. Without possessing that skill, the organization runs the risk of losing confidence in the program's ability to operate as intended, potentially leading to a decline in commitment to the program. To ensure the effectiveness of an analytics program, it is not enough to design an algorithm and let it run indefinitely without any updates. Regular updates are essential as circumstances change. The program should be vigilant in identifying routines

that generate a significant number of false positives, which consume valuable time and resources for investigation. Incorporating cognitive and machine learning systems allows companies to continuously enhance their analytics and increase efficiency in achieving their objectives. These techniques require considerable time and effort by the company (KPMG, 2020).

The fourth trust anchor of trusted anti-fraud analytics any anti-fraud program will be more effective concerns the ethical integrity of the process. Is its use considered acceptable by such stakeholders of the company's stakeholders, most notably as employees, suppliers, customers, business partners and regulators? This anchor holds significant importance among the four anchors as it addresses critical aspects of the relationship between a company and its stakeholders, where trust plays a vital role. It extends beyond mere legal compliance. Even if a company adheres to the law, adopting an excessively strict approach to fraud detection could potentially undermine trust among employees and other relevant parties involved in the detection process.

These concerns are particularly relevant in the evolving field of behavioural analytics. While fraud detection through analytics traditionally focused on transactions, there is a growing recognition of the need to analyze employee behavioural. This introduces an additional layer of anti-fraud measures by monitoring employees for potential behavioural anomalies that may indicate fraudulent activities. An example of this is seen in the realm of threat analysis, where certain US federal agencies and contractors are implementing programs to monitor employees' computers used to manage insider threats. To ensure the effectiveness of any anti-fraud program, it is crucial to operate with the trust and consent of the company's stakeholders, including its employees and third-party business partners. However, this can be a delicate issue and must be handled with care, taking into consideration the prevailing culture of the organization. Many executives and employees are wary of analytics being used to monitor their activities, which can damage the trust between the company and its employees. To address this, the company's leadership must cite the purpose of the anti-fraud analytics program, emphasizing that its initial intent is to protect the reputation of the company, rather than to victimize individuals or particular groups. This can be easier to

explain after a significant case of fraud has been uncovered when people are more receptive to the idea of preventing a recurrence. Additionally, companies may choose to anonymize portions of the data under analysis and only disclose information about the individual responsible for a pattern of suspicious behavioural if it raises a red flag. Transparency is crucial in maintaining the trust of stakeholders, as long as the program is operated with clear and honest intentions.

In the present scenario, it is crucial to maintain a balance between surveillance and transparency. The level of surveillance and transparency can vary depending on the relationship between the organization and its stakeholders. For instance, an organization dealing with sensitive information or handling large sums of money is likely to have a stronger surveillance program. However, if an organization is transparent about its analytics program and operates it ethically, it can build trust with its stakeholders. Today, there is a growing trend towards greater transparency, with stakeholders expecting more openness from organizations. Nevertheless, transparency must be managed with caution, as disclosing too much information could aid fraudsters in avoiding detection. Striking the right balance between transparency and surveillance is a challenging task for organizations.

3.3 Fraud Detection Techniques

Detecting occupational fraud can be challenging, as employees who commit fraud are often familiar with the company's policies and procedures and know how to work around them. These employees are typically trusted members of the organization who have access to various systems and understand how they work, which can allow them to identify weaknesses in the system. They may have even previously worked around normal procedures to solve a problem on behalf of the company, which exposes these weaknesses. While policies and procedures are important for identifying errors and mistakes, they are not enough to prevent intentional circumvention of systems by employees committing fraud. In addition, these employees may use various methods to conceal their actions, such as falsifying documents or misrepresenting transaction records. It is important to strike a balance between preventing fraud and allowing for efficient business operations, as too many restrictions or controls can

impede the productivity of honest employees. Despite efforts to prevent fraud, errors may still occur due to flaws in the system or unintentional mistakes by employees.
Below are listed possible techniques that can be implemented to use data analytics to detect financial fraud.

- **Detection algorithms** - One of the most common techniques is to use anomaly detection algorithms to identify unusual patterns of behaviour or transactions. This technique relies on the assumption that fraudulent transactions are likely to be outliers compared to normal transactions. For example, if an employee suddenly starts making large transactions or transferring money to unusual accounts, an anomaly detection algorithm may flag these transactions for further investigation, and this is the case of the project explained in this thesis.
- **Predictive analytics** - Another technique is to use predictive analytics to identify potentially fraudulent behavioural before it occurs. This technique involves building models that can predict the likelihood of fraudulent behaviour based on historical data. For example, a model might look at the spending patterns of a company's employees and identify those who are more likely to engage in fraudulent behaviour.
- **Network analysis** is another technique that can be used to detect financial fraud. This technique involves analyzing the connections between individuals or entities involved in a transaction. For example, if two employees are found to have a connection to a vendor that is suspected of fraudulent behaviour, this may indicate collusion between the employees and the vendor.
- **Text mining and natural language processing** can also be used to detect financial fraud. This technique involves analyzing unstructured data such as emails or chat logs for indicators of fraudulent behavioural. For example, if an employee is found to be communicating with a vendor suspiciously, this may indicate collusion or fraudulent behavioural.
- **Machine learning** can be used to detect financial fraud. This technique involves training models on large datasets to identify patterns of behaviour associated with fraudulent activity. For example, a machine learning model might identify patterns of behaviour associated with embezzlement or money laundering.

In the realm of detecting occupational fraud, only indicators, symptoms, or red flags can be observed. Upon identification, it is imperative to investigate these indicators to determine whether fraud has indeed taken place. However, given the substantial volume of false positives, many of these warning signs may be neglected despite their importance. It is worth noting that even after addressing an identified symptom in a particular area, other red flags within that same domain may be overlooked. Such red flags can manifest as irregularities in internal control, accounting anomalies, analytical irregularities, tips, and behavioural changes. Business systems are established to ensure the efficient functioning of a business, including the recording of transactions. These processes involve various measures to guarantee the smooth operation of the enterprise, the protection of assets, and the accuracy of reporting and recording. One of the primary objectives of internal controls is to prevent, deter, and detect fraud. However, instances of internal control overrides or weaknesses are typically associated with the most common types of fraud and can undermine the goal of fraud prevention and deterrence. In certain cases, however, there may be valid reasons for circumventing internal control, such as when an unforeseen scenario arises that was not considered in the original control design. Employees may then actively search for ways to carry out their responsibilities and maintain the business process, although such actions may or may not be officially sanctioned. Effective internal control encompasses several key components, including but not limited to:

- Segregation of duties, which minimizes the risk of collusion and the circumvention of controls.
- Physical safeguards to protect assets, including confidential information stored within computer systems.
- Independent checks, such as regular monitoring and audits, to ensure ongoing compliance with internal controls.
- Proper documentation and supporting materials that validate transactions and provide a clear audit trail.

- Appropriate authorization for all transactions, records, and other relevant activities, to ensure that approvals and control measures adhere to established authorization limits.

The focus of detection techniques should be on identifying weaknesses in internal controls. Any irregularities should be thoroughly examined, and appropriate actions should be taken and documented. The documentation can help implement corrective measures to internal controls, if necessary.

Accounting anomalies are defined as unusual items associated with the accounting system, including entries and backup documents. Journal entries, by their nature, are used to adjust unusual items outside the normal day-to-day flow of the accounting system. Since journal entries present a high-risk area for concealing fraudulent activities, manual journal entries should be carefully reviewed, and automated journal entries should be tested. Many accounting anomalies also fall under analytical anomalies such as outliers, as the case study presented in this thesis.

Analytical anomalies are common occurrences in business systems that lack integration. Unlike enterprise resource planning (ERP) systems that automatically populate related modules, many organizations have disjointed business systems that do not communicate with one another. Consequently, extra care must be taken when transferring data from one system to another, especially when it is done manually. To reduce the number of false positives, one must distinguish between high-risk and low-risk anomalies, and only investigate those that have fraud potential. This requires an understanding of the business systems, the business itself, and the industry it operates in. Internal auditors are expected to have in-depth knowledge of the workings of the business, while external auditors, forensic accountants, consultants, and investigators must familiarize themselves with the business entity and its industry.

3.3.1 Benford's Law

Benford's law was first exposed by Frank Benford, a physicist who presented the paper "*The Law of Anomalous Numbers*" (Benford, 1938) to the American Philosophical Society serves as a valuable tool for assessing data quality and identifying anomalous data. While researching the common logarithm table, the scholar made an intriguing observation regarding its physical condition. He noticed that the initial pages of the table exhibited greater wear and tear, appearing more damaged compared to the later pages. From this observation, he deduced that the early pages, which presented logarithms of lower digits such as 1, 2, and 3, were consulted more frequently than the later pages. This conclusion aligned with his further observation that the occurrence of lower initial digits (e.g., 1, 2, and 3) surpassed that of higher digits (e.g., 7, 8, and 9).

Motivated by this insight, the scholar embarked on a comprehensive study encompassing twenty distinct datasets, covering various domains such as population figures, river areas, atomic weights, mathematical tables, cost data, street addresses, and even random numbers extracted from newspapers and issues of Reader's Digest. To ensure a robust analysis, he meticulously curated these datasets to ensure diversity and multiple sources.

Through his investigation, he established that the expected frequency of the digits 1 and 2 appearing as the first digit in a number averaged 30.6% and 18.5%, respectively, cumulatively accounting for 49.1%. In other words, the probability of the first digit being 1 or 2 in any given number was determined to be 30.6% and 18.5%, respectively.

While there are no explicit guidelines available, certain characteristics are typically found in datasets that adhere to Benford's Law. Experimental evidence suggests that datasets containing reasonably large numbers with four or more digits and encompassing more than 1,000 records tend to exhibit good conformity.

Here are some general guidelines for a dataset to follow Benford's Law:

- The records should reflect the scale of events and facts being measured. Examples of such records include population data, trading data from a stock exchange, daily revenue of a company, and similar instances.

- Each record should represent an entire population rather than being a sample from that population.
- The dataset should not primarily consist of identification numbers, numerical labels, numbers influenced by human thought, or specially generated numbers. Examples of such numbers include National Identification Numbers, vehicle registration numbers, credit card numbers, bank account numbers, invoice numbers, and purchase orders.
- The dataset should not be skewed towards lower or higher records, such as salary data that predominantly includes similar amounts for a large group of lower-level staff. Research has indicated that datasets with an average (mean) greater than the median tend to exhibit better conformity to Benford's Law.
- The records should not be subject to any restrictions or limitations, such as a tax record that only includes salary or property income above a certain threshold for taxation. Another example would be a record consisting of fixed commission income earned regardless of the sales amount.

Adhering to these guidelines can increase the likelihood of a dataset conforming to Benford's Law. However, it is important to note that these guidelines serve as general indicators, and the specific characteristics of datasets may vary in practice.

In the realms of auditing, forensic accounting, and fraud examination, Benford's Law is extensively employed to analyze records and assess potential deviations between the observed frequencies of digits in a dataset and the expected frequencies dictated by Benford's Law. Any such deviations may serve as indications of anomalies, thereby raising concerns regarding a heightened risk of fraud, manipulation, or error. This analytical approach holds significant value in identifying irregularities and guiding further investigations in these fields.

First-Position-Digit Test

To assess the adherence of a dataset to Benford's Law, several types of digit tests can be conducted, with a specific focus on the first position, second position, and first two positions of the digits within the dataset. The first position-digit test involves comparing the actual frequencies of digits occurring in the first position (1, 2, 3, ..., 9) against the expected frequencies outlined by Benford's Law (Figure 1) for those positions. This test provides a

broad overview of the dataset, but it is important to note that any meaningful insights gleaned from deviations in this test alone may be limited. Further digit tests or statistical analyses are typically necessary to draw conclusive inferences.

Second Position-Digit Test

The second position-digit test involves analyzing the digit positioned second from the left in a number. For instance, in the numbers 25000 and 124678, the second digits are 5 and 2, respectively. In this test, the actual proportion of second digits within a dataset is examined and compared to the expected proportion based on Benford's Law (Figure 2). This test provides a high-level assessment and offers a general overview of the data's reasonability. Any significant deviations between the observed occurrence of second digits and the expected proportion warrant further examination to identify specific reasons behind such anomalies. It is worth noting that in payment or price-related data, an increased occurrence of 0s and 5s can be a normal phenomenon, given the prevalence of rounding numbers such as 500, 1000, 1,500, 2000, and 2,500. However, significant variations of other digits in the second position from the expected proportion in the same price-related data may indicate abnormal duplication or other anomalies that necessitate further investigation.

First-two-positions-digit test

The first two-positions-digit test involves comparing the actual frequencies of digits appearing in the first two positions of numbers with the expected proportion based on Benford's law. For example, in the numbers 25000 and 124678, the first two digits are 25 and 12, respectively. This test focuses on identifying duplications and specific trends within the dataset. Deviations between the actual and expected frequencies may provide insights into various possibilities:

- Excessive amounts just below specific cutoff thresholds (e.g., 300, 400, 500, ..., 1000) can explain the higher frequencies observed at digits like 29, 39, 49, 59, ..., 99. Such variations are often observed in price-related data for retail brands.

- In cases of employee expense reimbursement, where more stringent approval procedures are required for amounts above a certain minimum threshold, employees may tend to keep their claim amounts just below that threshold to avoid stricter scrutiny. This behaviour can lead to a high frequency of certain first-two-digit combinations that correspond to the minimum threshold. For example, if the reimbursement policy does not require supporting documents for claims below 5,000, employees may aim to keep their claims just below this threshold, such as at 4,900, resulting in a spike in the first two-digit combination of 49.
- Financial institutions may exhibit high frequencies at specific first two digits for loan figures just below a threshold that triggers more rigorous assessment and approval procedures.
- State-owned procuring agencies may tend to keep their procurements just below certain minimum thresholds, such as Rs. 500,000 or Rs. 3,000,000, to avoid mandatory publication on the PPRA website or in print media, as per PPRA rules. In such cases, the first two digits may show spikes at 49 and 29, respectively.

When conducting a first-two-positions-digit test on declared income in tax returns for the years 2021-2022, spikes may be observed at digits like 39 or 40, considering the maximum threshold of Rs. 400,000 above which income becomes taxable for business individuals and associations of persons. By analyzing the first two digits of the dataset and considering these possibilities, valuable insights can be gained regarding the underlying patterns and specific factors influencing the observed frequencies.

3.3.2 Data Mining vs Data Analytics

As it is already known, the abundance of data has made it necessary to develop techniques to extract valuable insights and knowledge from it. Two popular techniques used to extract knowledge from data are data mining and data analysis, described in the following paragraphs. Despite the overlap in their purpose, these techniques differ in their approach, tools, and methods.

Data mining refers to the process of discovering hidden patterns and relationships in large datasets using machine learning, statistical analysis, and data visualization techniques. Data mining is a branch of data analysis, which falls under the broader umbrella of data analytics. Data analytics typically involves the formation of a hypothesis, which is then either confirmed or proven false based on findings derived from data. Data analytics can be further categorized into exploratory data analysis (EDA), confirmatory data analysis (CDA), and qualitative data analysis (QDA). During the EDA stage, data is explored, and hypotheses are formed, while CDA involves testing the hypotheses and determining whether they are accurate. In addition, in the CDA category are frequently used OLAP (Online Analytical Processing) tools which are software applications that allow users to analyse multidimensional data from different perspectives. They enable users to view, analyse, and manipulate data from different angles or dimensions, such as time, geography, product, or customer. OLAP tools use a multidimensional database to store data, which allows for fast querying and analysis of large datasets. They typically provide features such as drill-down, slice-and-dice, and pivot tables, which allow users to explore data hierarchically and generate reports and visualizations that facilitate decision-making. QDA, on the other hand, is used to conclude nonquantitative or non-numerical data, such as images or text. Data analytics provides valuable insight into datasets, discovers underlying data relationships and structures, tests assumptions and hypotheses, identifies variables of causal relationships, and detects anomalies.

Anyways, both data mining and data analytics can be effective in preventing financial fraud. Data mining is particularly effective in detecting patterns and anomalies in large datasets, while data analytics is more focused on using statistical methods to analyse data and make informed decisions.

4

FERRERO S.A.: A CONTINUOUS MONITORING MODEL IMPLEMENTATION

This chapter focuses on the data analytics (D&A) lifecycle of the continuous monitoring model (CMM), which includes various stages such as analytics design, data collection and management, analytics execution, and follow-up on detected outliers. The chapter highlights the importance of each stage in the D&A process and explores how these stages are interconnected. In addition, it will be discussed the use of the Power BI dashboard as a tool for communication and reporting data, showcasing some examples of how the dashboard can be used to visualize and present insights to stakeholders. Overall, this chapter provides a comprehensive overview of the D&A lifecycle, emphasizing the critical role played by each stage in the success of the case study.

To evaluate the optimal continuous monitoring model to be implemented for the Ferrero Group, a comprehensive analysis of the company's organizational structure and accounting system, including the process of recording journal entries, is imperative. This assessment aims to gain a thorough understanding of how the company operates and maintains its financial records.

The Ferrero Group is a renowned multinational confectionery company that has made a significant impact on the global market. Founded in 1946 by Pietro Ferrero in Alba, Italy, the company has grown into one of the largest and most successful players in the confectionery industry. One of the key factors contributing to Ferrero's success is its unwavering commitment to quality. The company's stringent quality control measures ensure that only the finest ingredients are used in its products, resulting in superior taste and consumer satisfaction. Ferrero has also invested heavily in research and development to continuously improve its products and introduce new flavours and varieties to cater to evolving consumer preferences. Innovation has been a driving force behind Ferrero's growth and market dominance. The company has consistently introduced groundbreaking products that have captured the imagination of consumers worldwide.

Ferrero Group's latest approved financial statement reveals an impressive financial performance, showcasing its strength in the confectionery industry. The company reported a substantial increase in revenue, reaching $12.5 billion, representing a remarkable growth of 8% compared to the previous year. This revenue growth can be attributed to the strong demand for Ferrero's iconic brands, such as Nutella, Ferrero Rocher, Kinder, and Tic Tac, which experienced double-digit sales growth in key markets. The company also demonstrated effective cost management, leading to an operating profit of $1.8 billion, reflecting an improvement in profitability margins. Ferrero's net profit margin stood at a healthy 14%, indicating efficient control over costs and strong sales performance. Furthermore, the company generated significant cash flows, with an operating cash flow of $2.3 billion, enabling investments in research and development, strategic initiatives, and geographical expansions. These impressive financial results affirm Ferrero's robust financial position and its ability to thrive in the competitive confectionery market.

4.1 D&A Lifecycle

The business case focuses on the implementation of a Data Analytics and Continuous Monitoring Model to prevent the risk of unusual manual high journal entries posted in the accounting system of an Italian company. In particular, the company object for this study is a food-sector multinational with more than 55 branches all over the world. It is the second largest group in the world in the chocolate confectionery market.

The D&A Lifecycle, as shown in *Figure 8* below, is essentially based on the following macro processes that will be described in detail in this chapter:

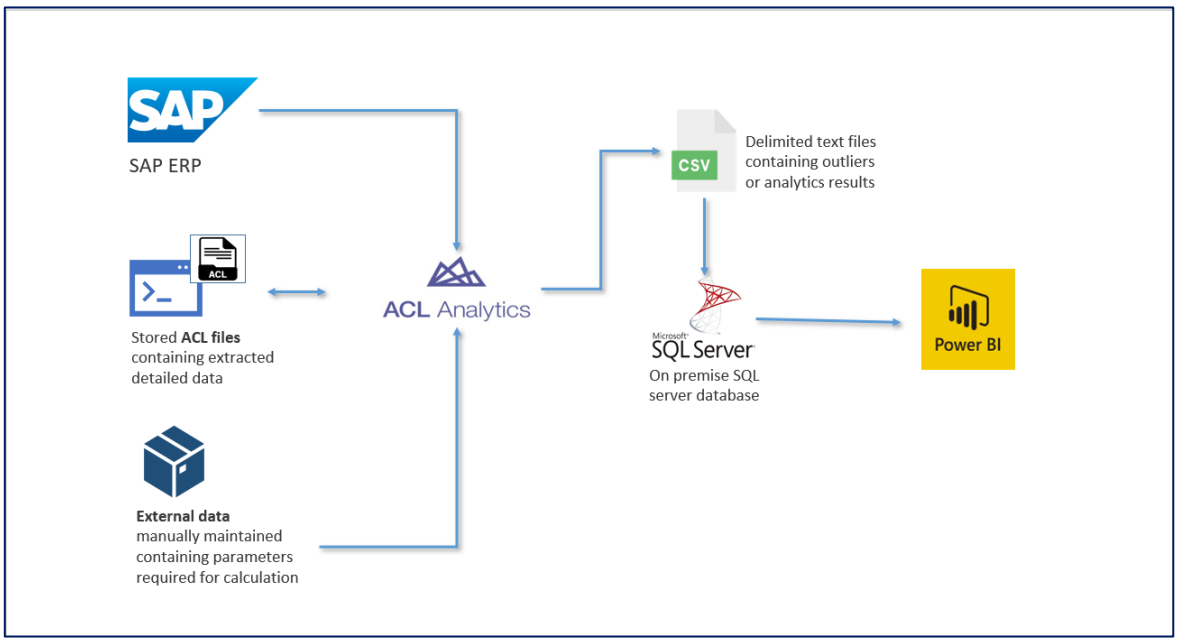


Figure 7: D&A Lifecycle

1. **Analytics design and implementation.** The first phase consists of the design of an analytics model based on control objectives and business performance indicators and the implementation of designed analytics.

2. **Collection of data and data Management**. The subsequent stage of the process involves gathering data, which serves as the primary input for the entire process. The primary sources of data collection comprise the enterprise resource planning (ERP) system, the company's Data Lake, and external data. Furthermore, to enhance the accuracy of the model, it is necessary to update the gathered data regularly, thus enabling the execution of analytics.

3. **Analytics Execution** and **follow-up on detected outliers** so the technical execution of implemented analytics and the consequent analysis of the obtained results. In this phase, a Root cause analysis is done to understand the possible sources of risks and it

is performed validation of detected outliers to confirm red flags underlying the risk scenario; the last phase is the action plans definition and follow-up on detected outliers.

4. **Communication and reporting.** The last process of the business case focuses on the creation of an ad-hoc dashboard to display the results. In this way, data and gained insights are more easily understood and used by stakeholders and Top Management. For this project, the PowerBi cloud reporting dashboard has been implemented.

4.1.1 Analytics Design and Implementation

The analysis is conducted on the Journal entries, that are in financial accounting any entries made directly within general ledger systems that are used to record transactions, allocations, adjustments, and corrections (KPMG, 2019). They include:

- Standard journal entries used to record recurring transactions and adjustments; and
- Non-standard journal entries used to record non-recurring, unusual transactions, or adjustments.

The business case test of journal entries and other adjustments is done to respond to the significant risk from management override of controls through the recording of inappropriate or unauthorized journal entries and other adjustments. By their nature, journal entries represent a unique opportunity for management to override controls and perpetrate fraud by manipulating accounting records and preparing fraudulent financial statements. Because overrides can happen unpredictably, they pose a risk of fraud and are therefore considered a significant risk. It is possible to categorize journal entries in two ways:

- **Automated journal entries,** which are standard journal entries that are typically automatically initiated, authorized, recorded, and processed in the general ledger. The

use of automated journal entries reduces the risk of management bypassing controls because there is less opportunity for manual intervention in the process and procedures.

- **Manual journal entries** are journal entries that are initiated by an individual and manually entered by an IT system.

Example characteristics of fraudulent journal entries and other adjustments may include:

- Journal entries and other adjustments made to unusual or seldom-used accounts.
- Journal entries and other adjustments created or posted by users and accounts who typically do not create or post journal entries.
- Journal entries and other adjustments recorded at the end of the period or as post-closing entries that have little or no explanation or description.
- Journal entries and other adjustments made either before or during the preparation of financial statements that do not have account numbers.
- Journal entries and other adjustments that consistently use round numbers or have consistent ending numbers.
- Journal entries and other adjustments that contain unusual combinations of debits and credits.

There are several reasons why manual journal entries can be risky for a company:

- Incorrect entries: Manual journal entries are prone to errors, such as transposition errors, calculation errors, and recording the wrong amounts. These errors can lead to incorrect financial statements, which can have serious consequences for the company, such as incorrect tax payments and misleading information for decision-making.
- Lack of documentation: Manual journal entries often lack proper documentation, such as receipts or invoices. This can make it difficult to verify the accuracy of the entries and can lead to fraud or mismanagement.

- Time-consuming: The process of manually entering journal entries can be time-consuming and labour-intensive, which can lead to delays in the financial reporting process and a higher risk of errors.
- Limited audit trail: Manual journal entries often lack a comprehensive audit trail, making it difficult to track changes and identify any potential issues.
- Increased risk of fraud: The manual process of journal entries can make it easier for fraudulent activity to go undetected, as there are fewer safeguards in place to detect and prevent it.

Overall, manual journal entries can be risky for a company because they increase the risk of errors, lack proper documentation, are time-consuming, have a limited audit trail, and can increase the risk of fraud. It is generally recommended that companies use automated accounting systems to help mitigate these risks.

4.1.2 KRI: approaches and applications

The first step to design the Data Analytics and Continuous monitoring model is to identify a Key Risk Indicator. A KRY is a measure used by management to indicate how much risk is associated with an activity and it serves as an indicator that highlights the potential occurrence of an unfavourable event in the future. It ideally offers early warning signals or subsequent confirmation when risks (both strategic and operational) shift in a manner that could hinder the attainment of Key Performance Indicators (KPIs).

There are two main approaches to selecting KRIs:

- Top-Down: The senior management takes the initiative to choose key indicators for monitoring purposes throughout the organization. This top-down approach is particularly effective when it comes to strategic Key Risk Indicators (KRIs). It enables the consolidation of risks and enhances management's comprehension of shared risks that have an impact on the overall strategy and business objectives.

- Bottom-Up: The lines of business are responsible for selecting and monitoring indicators that are specifically relevant to their operational processes. This bottom-up approach ensures that key risks are identified and monitored at a detailed level, enabling the lines of business to effectively manage risks that are most tangible and directly applicable to their operations.

In our scenario, the choice fell on the top-down approach the priority for the company was to aggregate two types of risks, the one related to high imports of manual journal entries and the one related to the small ones.

The KRI to be tested for the case study described in the following paragraphs is *Unusual manual high entries*. From the functionalities point of view, the scope of the KRI is to highlight anomalous high journal entries, made on a single account. It consists of two parts to identify separately very small amounts and very high amounts.

KRI Outlier definition

The objective of the analytics model is to identify outliers and develop a strategy to detect and prevent the presence of high manual journal entries in the future. The KRI is calculated using the following formula:

$$\boldsymbol{K = \frac{Entry\ Amount\ - Average\ (Amount)}{StdDev\ (Amount)}}$$

Equation 2: KRI Outlier calculation

- **Amount** is calculated over the same account on the same company for the full period (e.g., 1.09.2022 - 31.05.2023).

- **Entry Amount** is the absolute value of a single journal entry made over the same account.

For the company under analysis, the period considered is calculated as follows:

Time period = Current fiscal year YTD + N° previous complete FYs (N = "N° of Fiscal Year" execution parameter, default 2).

Equation 3: Time period formula

In our scenario, manual JE are considered as outliers if K > 2, as shown in the following figure:

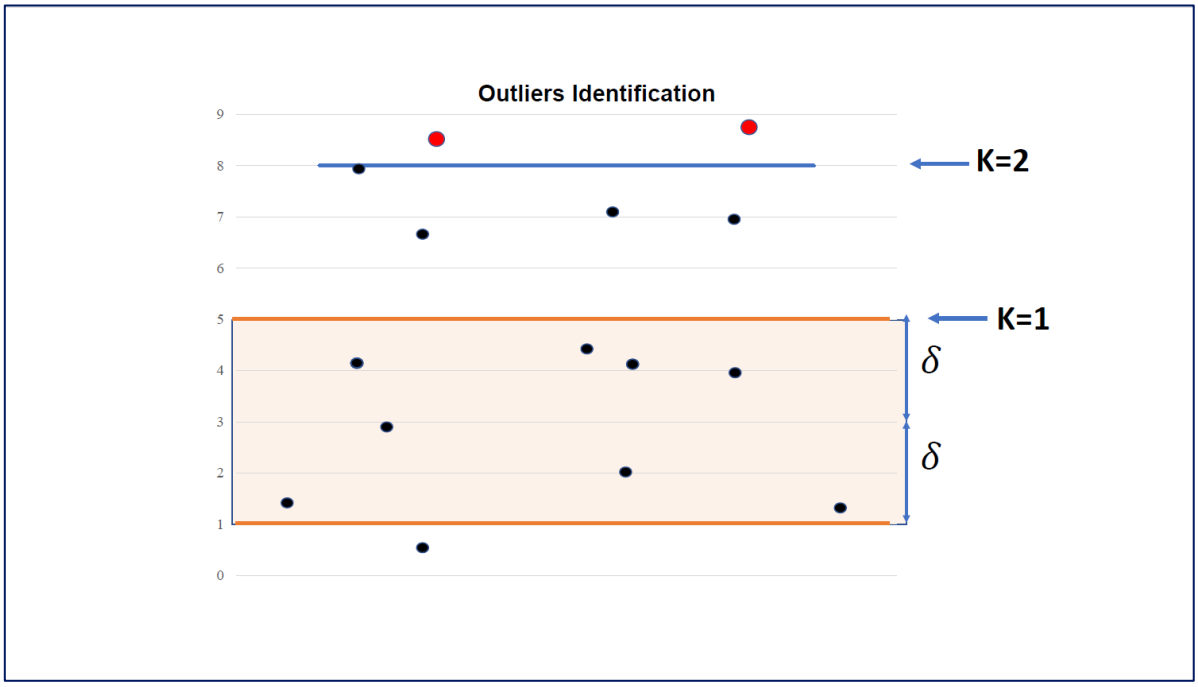


Figure 8: Outliers Identification (KPMG, 2022)

As shown in the previous figure, the data analytics model implemented should monitor manual journal entries recorded during the month to identify any unusual/unauthorized posting that should be further investigated so the ones having a K greater than 2.

4.1.3 Collection of data and data management

The procedure of analyzing data, alternatively referred to as the steps of data analysis, encompasses the collection of all relevant information, its processing, exploration, and utilization for identifying patterns and other valuable insights. The process of data analysis encompasses several stages, namely:

• **Data Requirement Gathering**: This stage involves determining the purpose of the analysis, the type of data to be used, and the data to be analyzed.

• **Data Collection**: After identifying the requirements, it is necessary to collect data from various sources such as case studies, surveys, interviews, questionnaires, direct observation, and focus groups. The collected data must be organized for further analysis.

• **Data Cleaning**: Since not all collected data is useful, this stage entails removing irrelevant data such as white spaces, duplicate records, and basic errors. Data cleaning is mandatory before sending the information for analysis.

• **Data Analysis**: This stage involves the use of data analysis tools such as Excel, Python, R, Looker, Metabase and Microsoft Power BI to interpret and understand the data and arrive at conclusions.

• **Data Interpretation**: Once the results are obtained, they must be interpreted to determine the best course of action based on the findings.

• **Data Visualization**: Data visualization entails presenting data in a graphical format such as charts, graphs, maps, bullet points, or other methods that enable easier interpretation of the data. Visualization helps to derive valuable insights by facilitating the comparison of datasets and observation of relationships.

For the project described in this chapter, the step related to the data collection is strictly correlated with collecting data from the company's data lake. A data lake is a central repository of all data in an organization, and it can be used to store structured, semi-structured, and unstructured data. To collect data from a data lake, several steps need to be followed.

The first step is to identify the type of data required for the analysis. This has been achieved by reviewing the analysis objectives and determining the specific data sets that will be used, in the project's case are manual journal entries, both high amount and very small amounts. Once the data sets have been identified, the next step is to locate them within the data lake. This can be done by reviewing the data lake architecture and data schema, which will provide information on where specific data sets are located within the data lake.

In the context of Data collection, the data sets came from JE produced in three different ways that are ERP systems, Data Lake, and External Data as is possible to observe in Figure 8. The first source comprises SAP Transactions, which are automated and cannot be modified by users. JE are automatically generated in SAP to represent accounting for logistics/managerial events that occur outside the Finance (FI) SAP Module, such as material movements or invoice registration. Enterprise Resource Planning (ERP) systems, such as SAP, are the main source of data collection for many organizations because they are designed to capture and store all critical business data in one central location. Collecting data directly from the ERP system ensures access to accurate and timely information about business operations. This data can be used for various purposes, such as financial reporting, forecasting, budgeting, inventory management, and customer analysis.
In many cases, data collected from other sources, such as external databases, can also be integrated into the ERP system. This was done for the model under analysis, providing a more comprehensive view of the business and enabling informed decisions based on a broader range of data. Overall, the SAP system is a valuable data collection source as it provides a single, centralized location for all critical business data, along with built-in analysis tools that can help organizations make better decisions based on their data.

The second main source of data collection is the company's Data Lake organized in ACL files. A data lake is a central repository where raw data from various sources is stored in its native format until it is needed. The purpose of a data lake is to provide a platform for the organization and analysis of large amounts of data from different sources. A data lake enables the organization to store all types of data, both structured and unstructured, and to use various analytics tools to extract insights and knowledge from this data. ACL files, on the other hand,

are a type of data file that is used in data analytics. ACL stands for Audit Command Language and is a software tool that provides a comprehensive range of data analysis functions. ACL files contain data extracted from various sources, such as ERP systems, accounting systems, or spreadsheets, and are used to perform various types of data analysis, including data mining, fraud detection, and compliance testing. ACL files are particularly useful in data analytics because they allow analysts to easily manipulate and analyse large amounts of data from different sources.

A third source of data is the External data manually maintained containing parameters required for calculation. To use them in the model designed for this case study, there are a few steps that can be followed:

1. *Determine the required parameters*: First, identify the specific parameters that are needed for the data analysis. This could include variables such as dates, customer IDs, or product SKUs.
2. *Obtain the external data*: Once the required parameters are identified, obtain the external data containing the parameters. This could come from various sources, such as spreadsheets, text files, or databases.
3. *Clean and format the data*: Next, clean and format the external data to ensure it is in a usable format for data analytics. This could involve removing unnecessary columns, checking for inconsistencies, and ensuring the data is properly formatted.
4. *Merge the external data with the main data set*: After the external data is cleaned and formatted, it can be merged with the main data set. This can be done using various tools, such as Excel, SQL, or Python.
5. *Use the parameters in calculations*: Finally, the parameters from the external data can be used in calculations for the data analysis. This could involve using formulas in Excel or writing custom code in a programming language such as Python.

In the context of data analytics, the accuracy and reliability of external data must be ensured. Inaccuracies or inconsistencies present within such data sets could result in erroneous conclusions and potentially deleterious business decisions. Hence, it becomes necessary to conduct a comprehensive evaluation of external data sources before integrating them into

analytical processes. By subjecting external data sources to meticulous scrutiny, any potential inaccuracies, inconsistencies, or redundancies can be identified and resolved before proceeding with data analysis. This proactive approach helps to minimize the possibility of data-related errors or oversights that could adversely affect the analytical outcomes. Therefore, it is essential to exercise due diligence in evaluating external data sources and ensuring that the collected data aligns with the desired data quality standards for data analytics.

4.1.4 Analytics Execution in ACL

After locating the data sets, the next step is to extract them from the data lake or the ERP system. In the scenario under analysis, where an organization wants to use analytics to identify potentially fraudulent transactions in their financial data, the company has already gathered their data into a structured format and has selected a data analytics tool to help them with the analysis. In this case, ACL Analytics has been used. **Audit Command Language** (ACL) analytics is a data extraction and analysis software used for fraud detection and prevention, and risk management.

The first step in the technical execution of this analytics implementation is to import the data into ACL. This involves connecting to a database or loading data from a spreadsheet or text file. Once the data is loaded, the team would begin to explore and clean the data using ACL's data preparation tools. This might involve identifying and removing duplicates, filling in missing values, and standardizing data formats. An advantage of ACL is that is possible to use regression analysis capabilities to identify correlations between different variables in the data.

As the analysis proceeds, the team would begin to identify potential fraudulent transactions using ACL's exception detection tools. These tools allow the team to define specific criteria for what constitutes an exception or anomaly in the data. The scenario described in this thesis has been defining an exception as a transaction in registering manual journal entries that is significantly larger or significantly smaller than usual.

Once potential exceptions have been identified, is up to the team to determine whether or not they represent actual instances of fraud. Different techniques could be implemented to validate the detected outliers to confirm red flags underlying risk scenarios. This might involve looking at additional data sources, consulting with experts in the field or exploiting statistics tools. Statistical analysis techniques can be used to identify data points that are significantly different from the norm, so as in the case presented in this thesis exploiting the standard deviation formula to detect outliers. If the transactions have a value that is more than 2, this could be an indication of fraudulent activity.

In the end, with ACL Analytics, organizations can gain deep insights into their financial data and identify potential areas of risk or fraud. By following a structured approach to analytics implementation, organizations can make data-driven decisions that improve their operations and reduce risk. Finally, the extracted data must be stored in a suitable location for analysis. This can be a local database or a cloud-based analytics platform such as Google BigQuery or Microsoft Azure. The data must be stored in a format that is compatible with the analysis tools that will be used for analysis.

4.2 Reporting results

The last step of the project consists of the team reporting their findings to stakeholders using ACL's reporting tools and creating a dashboard that collects all the information gained during the analytics. It might involve creating tables, charts, and other visualizations to help communicate the results of the analysis.

Creating a dashboard with all the data collected during the analytics is crucial for organizations to effectively communicate insights and make data-driven decisions. A dashboard presents a visual representation of key performance indicators (KPIs) and metrics that matter to a business, allowing executives and decision-makers to identify trends and patterns quickly and easily in the data. By centralizing all relevant information on a single

platform, a dashboard provides a comprehensive overview of an organization's operations, making it easier to identify areas of concern or improvement.

A well-designed dashboard allows stakeholders to drill down into specific data points, providing greater context and insights into a business's performance. Dashboards can also provide real-time updates and alerts, enabling organizations to make informed decisions quickly in response to changing market conditions or emerging opportunities. Additionally, dashboards provide a common language for teams to discuss and understand data, fostering collaboration and communication across an organization. By providing a comprehensive, visual representation of key metrics and KPIs, a dashboard allows stakeholders to make informed decisions quickly and proactively address areas of concern or opportunities for growth.

An interesting theory by A. Cairo exposed in the book "*The Functional Art: An introduction to information graphics and Visualization*" is that visualization should be seen as a technology. Although this may appear unconventional, the concept of technology typically evokes thoughts of machines, such as MP3 players, cars, refrigerators, electric toothbrushes, and computers. However, these devices share a common essence, not in their physical attributes, but in their underlying nature:

1. *They serve as extensions of ourselves*. This notion was first put forth by Canadian media thinker Marshall McLuhan over half a century ago. A lawn mower assists us in maintaining a tidy garden without the need for manual labor. An MP3 player not only functions as a playback device but also aids in preserving the songs that hold significance during the best and worst moments of our lives. The data visualization process often requires the use of computational techniques, algorithms, and programming languages to transform raw data into visually understandable forms. Similar to other technological processes, data visualization involves the application of specific methodologies and tools to achieve desired outcomes.

2. *They serve as means to accomplish goals.* While some technologies have singular purposes others encompass multiple objectives. Consider a computer, whose functionality relies on the integration of other technologies, including installed software. In this way, technologies can incorporate and support other technologies. It is common knowledge that data visualization plays a crucial role in aiding decision-making processes. By presenting data in a visually intuitive manner, it enhances the comprehension and interpretation of information, enabling users to make informed decisions more effectively and efficiently. This aspect aligns with the fundamental purpose of technology, which is to streamline various aspects of human activities and help reach goals.

Data visualization exhibits characteristics and functionalities that align with the concept of technology. Its reliance on advanced tools and techniques, data processing capabilities, impact on decision-making, and facilitation of communication all contribute to its classification as a technology in the context of data analysis and information visualization.

In the scenario described above, Power BI is a popular tool that could be used to report and represent the results obtained from the analytics. Power BI is a Microsoft service designed for business analytics, offering features that enable users to create their reports and dashboards with interactive visualizations. It provides capabilities for business intelligence and allows end-users to easily navigate the interface and utilize its functionalities. Power BI allows the creation of interactive and visually appealing reports and dashboards that can be customized to fit the needs of the user and their stakeholders. In the last chapter, will be presented some examples of the results obtained through the dashboard.

5

RESULTS ACHIEVED AND FOLLOW-UP STRATEGIES

5.1 KRI Analysis and visual representation of the Results

This chapter will describe various instances of dashboards that have been generated using the PowerBi tool. Moreover, a numeric table will be presented, providing a concise overview of the principal indicators categorized by year and company. The KPIs under analysis and explained in detail in the next paragraph are the following:

- *N° of Outliers by Company Code (SAPcode & Name).*
- *N° of Outliers by Profit & Loss and Balance Sheet Accounts.*
- *N° of JE reversed in the same Fiscal Year*
- *Outliers amount (absolute value) by Fiscal Year & SAP period*
- *N° of outliers by Business Transaction & Company Code*
- *N° Outliers amount by account*
- *% Of Manual Journal entries compared to the entire population*
- *Number of documents by Financial Statement*
- *% Of Manual Journal entries by Company Code*
- *Number of documents and total amount in M € by Fiscal Year and Fiscal Period*
- *Number of documents BY Document type*

The displayed dashboard (*Figure* 10 below) presents various analytics concerning the Key Risk Indicator (KRI) "*F01.02 Analytics Full Population*". These analytics provide insights into the risks associated with analyzing the underlying population of the analysis, specifically the total number of documents issued by the company that contain the manual journal entries under examination. Several analyses have been conducted, taking into account variables such as year, source of financial documents, source of company documents, and the proportion of manual journal entries in relation to automated entries.

In addition, the dashboards described are documented in *Annex A* and *Annex B* at the end of the thesis.

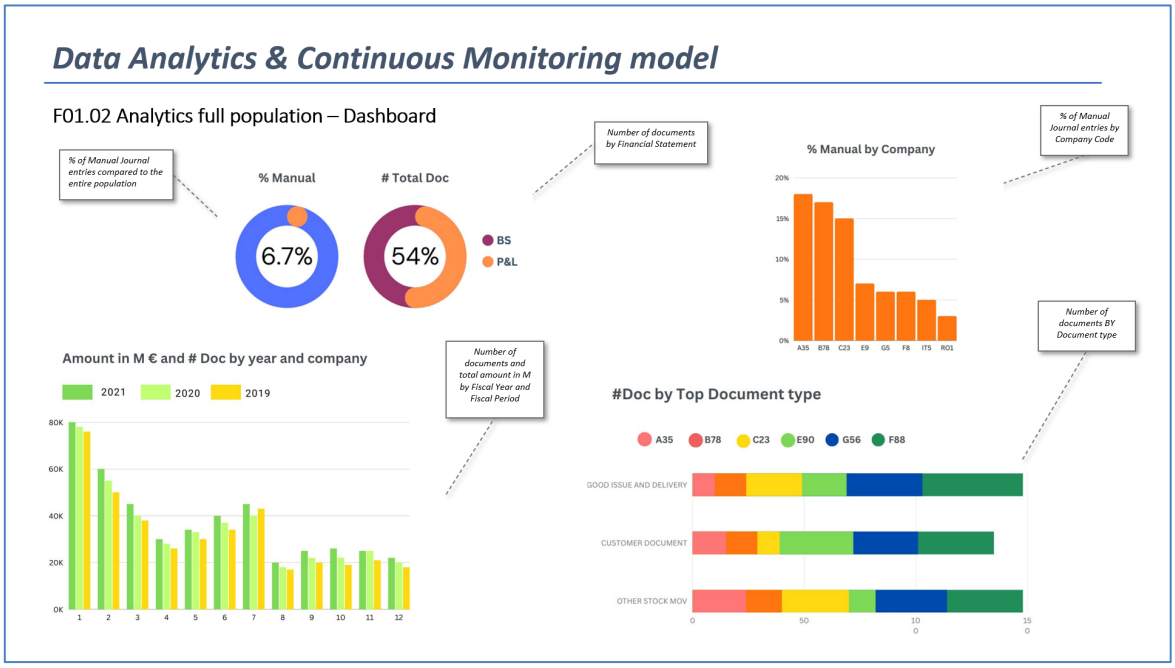


Figure 9: KRI F01.02 Analytics full population

As visible from the figure above, the KRI under analysis is composed of 5 principal KPIs that are the following:

- *% Of Manual Journal entries compared to the entire population*

The presented pie charts are related to the KPI *Number of documents by Financial Statement* and indicate that manual journal entries account for 6.7% of the entire document population. Among these manual entries, 54% are derived from the Profit and Loss statement, while the remaining percentage originates from the balance sheet. Based on this data, several conclusions can be drawn. Firstly, among the proportion of Manual Journal Entries, it could be said that manual journal entries constitute only 6.7% of the overall document population for 2021.

Also, regarding the second pie chart, the significant share of manual entries originating from the profit and loss statement (54%) suggests that the P&L accounts may be subject to more

complex or nuanced transactions that necessitate manual adjustments or entries. Conversely, the relatively smaller percentage of manual entries originating from the balance sheet implies that this area is more inclined to automation. The higher level of automation in the Profit and Loss statement suggests that the company's revenue and expense recognition processes are relatively standardized.

In the end, the data from the pie chart can also serve as a starting point for evaluating process efficiency and identifying opportunities for improvement. By reducing the reliance on manual journal entries, the company can potentially streamline its financial reporting processes and minimize the risk of errors.

- *Number of documents and total amount in M by Fiscal Year and Fiscal Period*

The bar chart reveals a conspicuous surge in the number of documents during the initial fiscal period (1 and 2), with a recorded count of 80,000 documents by 2021. Subsequently, the data suggest a relatively stable trend throughout the remaining fiscal years, except for a notable positive spike observed in the seventh period. Upon considering the three-year timeframe encompassing 2019, 2020, and 2021, the graph exhibits a discernible alignment in the number of documents across these years.

- *Number of documents by Document type*

The bar chart depicted herein showcases the quantified distribution of documents categorized by their respective types and further segregated based on company codes. Among the various document types, it is noteworthy that the most prevalent ones include "*good issue and delivery,"* "*customer document*", and "*other stock movement*." Good issue and delivery primarily include documents related to the issuance and delivery of goods or products. Examples could include shipping orders, delivery receipts, or packing slips that provide detailed information about the goods being sent to customers or other recipients.

Customer document comprises documents specifically associated with customers. Examples could include sales invoices, purchase orders, contracts, or agreements between the company

and its customers. These documents often contain details about the products or services purchased, pricing information, payment terms, and other relevant customer-related data. Other stock movement category instead encompasses various types of documents that capture movements and changes within the company's stock or inventory. Examples could include stock transfer forms, stock adjustment records, stock reconciliation reports, or inventory count sheets. These documents track the movement of goods within different locations, adjustments made to stock quantities, or the results of regular inventory audits.

Manual intervention arises within the document categories due to their inherent complexity and potential impact on the financial and operational aspects of a company. In the case of "good issue and delivery," auditors often need to verify the accuracy of inventory movements, ensuring that goods are properly issued, accounted for, and delivered to the intended recipients. This requires meticulous examination of supporting documentation, cross-referencing with sales orders and delivery records to detect any discrepancies or irregularities. Similarly, "customer documents" necessitate manual intervention as auditors meticulously review sales invoices, purchase orders, and contracts to ensure their accuracy, completeness, and adherence to applicable accounting standards. This involves verifying pricing details, and payment terms, and confirming the proper recognition of revenue. Auditors must assess the consistency and reliability of customer-related information to maintain financial transparency and mitigate potential risks. The category of "other stock movement" encompasses diverse documents that record changes and adjustments within the company's inventory. Auditors must manually scrutinize stock transfer forms, adjustment records, and inventory count sheets to verify the accuracy of stock movements, identify any anomalies, and ensure the integrity of inventory valuation. These documents require careful examination to maintain proper inventory control and prevent potential errors or fraudulent activities.

Overall, manual intervention for auditors within these document categories is essential to maintain data integrity, detect errors, prevent fraud, and comply with regulatory requirements. The intricate nature of these documents and their significant impact on

financial reporting necessitates thorough scrutiny, making manual intervention a crucial aspect of the auditing process.

- *% Of Manual Journal Entries by Company Code*

The bar chart highlights that the Company with code A35 has the highest percentage of manual journal entries, accounting for 18% of its total entries followed by company B78 with 17% and C23 with 15%. Companies with more complex operations may encounter a higher volume of transactions that require manual adjustments. For instance, businesses involved in multiple lines of business, international operations, or complex intercompany transactions may have a greater need for manual journal entries to accurately record and account for these activities.

Regarding the maturity level of Systems and Processes, companies with outdated or less automated financial systems may rely more on manual interventions. Legacy systems or insufficient technology infrastructure may limit automation capabilities, leading to a higher proportion of manual journal entries. Also, when there are decentralized or fragmented accounting functions across multiple subsidiaries or business units' financial documents may experience more manual entries. Lack of standardized processes or coordination among various entities can contribute to an increased reliance on manual adjustments.

Company A35 e B78 have also experienced a merger process with some of their subsidiary and it contributes a higher percentage of manual journal entries. The integration of disparate systems, harmonization of accounting policies, and consolidation of financial data can often require manual interventions during the transition phase.

The second KRI under analysis is the "*F01.01 Anomalous journal entries, made on a single account*":

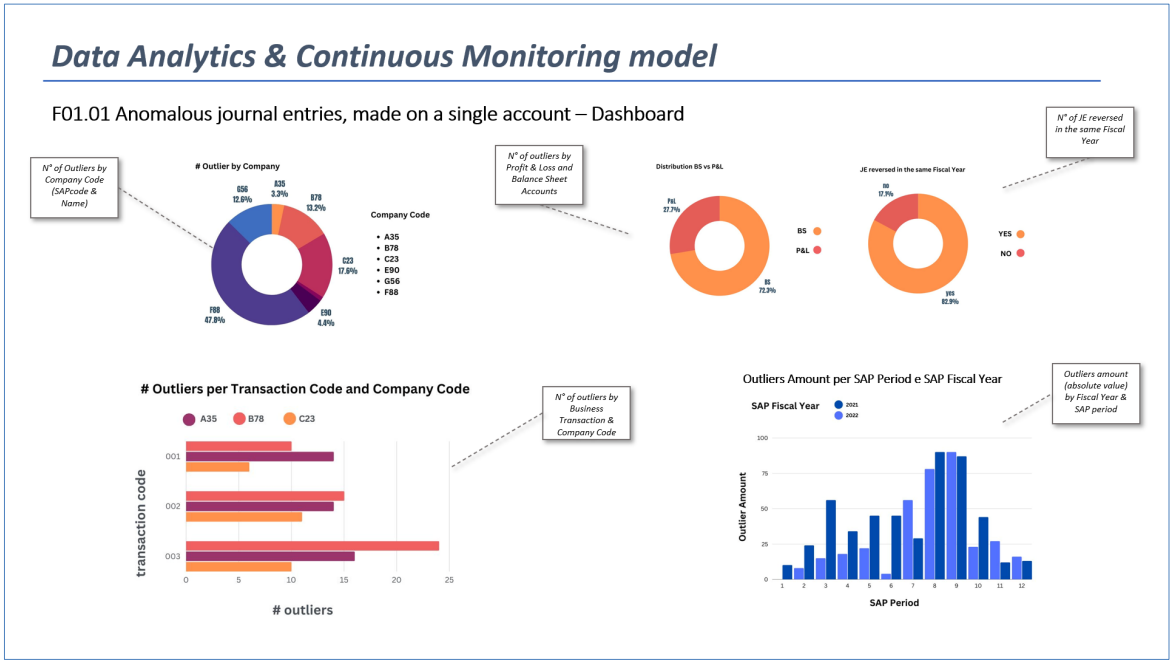


Figure 10: KRI F1.01 Anomalous journal entries made on a single account

As visible from the figure above, the KRI under analysis is composed of 5 principal KPIs that are the following:

- *N° of Outliers by Company Code (SAPcode & Name)*

As evident from the pie chart presented on the left side of the above figure, the calculation of the number of outliers by company code KPI took into account the percentage by which each company contributes to the total number of identified outliers. There is an evident major source of outliers, and it is within the F88 company. Primarily, it is imperative to acknowledge that the size and complexity of a company play a significant role, as they directly impact the likelihood of encountering outliers. Specifically, larger companies with expansive operations and intricate business processes are more prone to encountering outliers. This propensity can be attributed to the sheer volume of data entering the enterprise resource planning (ERP) system. The scale and diversity of their activities contribute to a

greater number of exceptional data points, thereby increasing the probability of encountering outliers.

- *N° of outliers by Profit & Loss and Balance Sheet Accounts*

The primary cause of outliers in the detection of journal entries resides in the balance sheet as opposed to the profit and loss (P&L) document. The balance sheet offers a concise representation of an enterprise's financial state at a precise moment, while the P&L statement concentrates on the financial performance across a given period. In terms of identifying anomalies within journal entries, the balance sheet typically holds greater significance, as it encompasses all journal entries about the registration of SAP transactions, such as the recording of invoices for goods entry and invoice payments for warehouse procurement. Instances of balance sheet outliers may manifest as erroneous values associated with SAP transactions or inconsistencies among various balance sheet accounts.

- *N° of JE reversed in the same Fiscal Year*

The Key Performance Indicator (KPI) "Number of Journal Entries (JE) Reversed in the Same Fiscal Year" serves as a metric to assess the frequency and extent of journal entries that are reversed within the same fiscal year. This KPI holds significant importance in evaluating the accuracy and reliability of financial reporting processes. Reversing journal entries indicate the *correction of errors or adjustments made to previously recorded transactions*. A higher number of JE reversals within the same fiscal year can imply a higher incidence of errors or inaccuracies in the initial journal entries. Consequently, it is crucial for organizations to closely monitor and analyze this KPI to identify areas for improvement in financial controls and data integrity. An excessive number of reversals may signify weaknesses in internal controls, inadequate training of accounting personnel, or deficiencies in the recording and validation processes. Additionally, a lower number of JE reversals demonstrates the effectiveness of controls and the reliability of financial information, instilling confidence among stakeholders and investors. Based on the obtained results, it is apparent that the company should prioritize the implementation and enhancement of internal auditing controls, as well as initiate training programs for employees.

In the concluding paragraph of this chapter, various mitigation strategies will be proposed. These strategies will serve as recommendations to alleviate the identified weaknesses and bolster the internal control framework. The proposed measures will encompass a comprehensive approach, taking into consideration areas such as process enhancements, system integrations, automation, continuous monitoring, and regular performance evaluations.

- *N° of outliers by Business Transaction & Company Code*

In the context of manual journal entries, various business transactions can potentially serve as the source of outliers. However, certain transactions are more prone to generating outliers than others. For instance, non-routine transactions often involve a higher risk of errors or discrepancies, such as transactions related to mergers and acquisitions, restructuring activities, significant asset valuations or impairments, changes in accounting policies, and unusual or exceptional events impacting financial statements. Additionally, transactions involving estimates, judgments, or subjective assessments, such as provisions for bad debts, valuation of inventory, or recognition of revenue, can also introduce a higher likelihood of outliers in manual journal entries. These transactions are inherently more susceptible to errors or misinterpretation, potentially resulting in outliers in the recorded financial data.
It is important to note that rigorous internal controls, proper documentation, and adequate review processes can help minimize the occurrence of outliers in manual journal entries and enhance the overall accuracy and reliability of financial reporting.

- *Outliers amount (absolute value) by SAP period*

The period of September (09/23) has consistently exhibited a higher incidence of outliers, both in 2021 and 2022. This pattern can be attributed to several factors, primarily stemming from the culmination of the fiscal year at the end of September and the subsequent requirements for year-end adjustments. During this period, management is tasked with executing significant modifications to their financial records, encompassing accruals, provisions, and revaluations. These adjustments entail intricate calculations and estimations,

consequently augmenting the probability of errors or discrepancies that may become apparent through outlier detection. Moreover, the audit and review processes conducted by external auditors or internal review teams gain heightened prominence in September as the fiscal year draws to a close. These meticulous scrutiny procedures encompass a comprehensive examination of financial records.

5.2 Overview of the Total Outliers for 2021 and 2022

Below there is a sum-up table that shows the number of outliers for the period in scope (1.09.2021 - 31.08.2022) calculated on the same KRI "*F01.01 Anomalous journal entries, made on a single account*". The table was built considering the main goals that the company wanted to achieve regarding the control monitoring system.

Company / Year	N° of Total ITEMS		N° of Manual ITEMS		N° of ITEMS OUTLIERS (K>1)		N° of ITEMS OUTLIERS (K>2)	
	2021	2022	2021	2022	2021	2022	2021	2022
FR10	2.842.017	2.842.017	56.021	39.215	2.665,0	2.500,0	1,97%	1,4%
FR40	131.145	131.348	12.373	8.661	544,0	467,8	9,43%	6,6%
GB30	2.489.956	2.489.777	86.026	60.218	3.360,0	3.300,0	3,45%	2,4%
GB40	913.095	913.095	27.150	19.005	1.231,0	1.200,0	2,97%	2,1%
IT60	3.456.890	2.344.678	197.596	138.317	2.130,0	2.100,0	5,72%	4,0%
IT80	234.998	235.003	100.329	70.230	4.649,0	4.520,0	42,69%	29,9%
IT90	653.354	760.556	23.889	16.722	1.304,0	1.300,0	3,66%	2,6%
Total	**10.721.455**	**9.716.474**	**27.004**	**18.903**	**15883,00**	**15387,84**	**0,25%**	**0,20%**

Table 3: Outliers identified on a single account (K>2)

The different variables are explained in detail:

— **Total number of items**: The table provides information on the total number of items for each company and year. This data is used to compare the scale of operations among the companies or track any significant changes in the volume of transactions over time.

— **Manual items**: The number of manual items indicates the extent to which companies rely on manual processes for data entry or other accounting activities.

— **Outlier items**: The table presents the number of outlier items based on two different thresholds (K>1 and K>2). These outlier items indicate potential anomalies or irregularities in the data. Companies with a higher percentage of outlier items may need to further investigate and address the underlying causes.

— **Yearly comparisons**: By comparing the data between the years 2021 and 2022 for each company, trends and patterns can be observed. Decreases in the number of manual items or outlier items may suggest improvements in data quality, process efficiency, or fraud detection measures.

The table shown above presents a set of data for different companies for the years 2021 and 2022, showing the total number of items, the number of manual items and the number of outliers (with values above a certain threshold) for each company. Some conclusions can be drawn.

FR10 and GB30, have a relatively low number of outlier items compared to the total number of items, indicating greater conformity and accuracy in their financial operations. On the other hand, companies like IT80 show a significantly higher percentage of outlier items, suggesting a higher likelihood of anomalies or fraud in their accounting process.

One of the companies, in particular, identified as FR10, had a high number of total items, with a decrease from 2021 to 2022. However, the number of manual and outlier items decreased in both years. This is because company FR10, in the light of the implementation of the continuous monitoring model, has specialized in greater automation reducing the number of activities in which is necessary to register journal entries manually. Following the data analysis, companies like FR10 exhibited responsive and proactive behavioural by implementing robust control mechanisms and automated processes for managing journal entries. This proactive approach ensures that potential anomalies or fraudulent activities in the journal entries are promptly identified and addressed. By leveraging advanced technologies and data analytics tools, FR10 has been able to detect and prevent irregularities and maintain the integrity of financial data. This strategic focus on control and automation reflects the company's commitment to maintaining transparency, accuracy, and compliance in its financial operations, ultimately enhancing overall organizational efficiency, and safeguarding against potential financial risks.

Similarly, other companies show variations in their data between 2021 and 2022. Some companies show a decrease in both the number of manual and outlier items, which is a positive sign of improved internal processes. Conversely, other companies have an increase in both the number of manual and outliers' items from 2021 to 2022 resulting in red flags for potentially fraudulent activities:

- IT80: The number of manual items increased from 100,329 in 2021 to 70,230 in 2022, and the number of outliers' items increased from 4,649 to 4,520 during the same period.
- IT90: The number of manual items increased from 23,889 in 2021 to 16,722 in 2022, and the number of outliers' items increased from 1,304 to 1,300 from 2021 to 2022.

In addition, the percentage values represent the proportion of items that are considered outliers based on the specific threshold K>1 or K>2. Higher percentage values indicate a larger proportion of items deviating from expected values. When it comes to fraud detection, higher percentage values of anomalies suggest a higher likelihood of irregular or fraudulent activities within the accounting systems. Therefore, the percentage values of anomalies serve

as crucial indicators in the context of detecting outliers and preventing fraudulent activities, prompting organizations to focus their attention on areas that exhibit a higher concentration of anomalies.

5.3 Milestones of the CM Model and possible mitigation strategies

As evident from the data presented in *Table 3*, there has been a significant reduction in the occurrence of outliers in the year 2022, accounting for a **0.20%** of the total. This noteworthy observation gives rise to several implications. Notably, it serves as a positive indication for both the company itself and the auditors, suggesting that the decisions made regarding the implementation of the continuous monitoring model have been validated and deemed appropriate. However, it is important to acknowledge that the overall number of items in 2022 is smaller, while the number of manual items remains relatively consistent.

The three primary areas of improvement have been derived from the identification of the three main weaknesses, as visible in *Table 4* below.
It is important to acknowledge that the overarching objective of the project primarily pertains to the prompt rectification of errors or discrepancies, rather than providing strategies and suitable solutions for the Group as a whole.

Deficiency	Possible Solution
High frequency of manual interventions	**Automation** **Standardization and Templates** **Integration of Systems**
ERP Systems are not error-oriented	**Data Validation and Error Check**
Inexpert internal auditors	**Training and Knowledge Sharing**

Table 4: Summarization of main areas of improvement

The initial discrepancy observed pertains to the frequency of manual interventions required during the registration of journal entries in financial statement documentation. To mitigate this risk and reduce the overall number of manual journal entries (thereby decreasing the likelihood of potential risks), several strategies could be implemented:

1. Need for deeper Automation: Implementing automated systems and software can significantly reduce the reliance on manual entry. Robotic Process Automation (RPA) or specialized accounting software can be used to automate repetitive and rule-based journal entry tasks.

2. Standardization and Templates: Establishing standardized templates for journal entries can enhance efficiency and reduce the likelihood of errors. These templates can include pre-defined fields and guidelines for data input, ensuring consistency and reducing the need for manual intervention.

3. Systems integration: Integrating different systems, such as accounting software, enterprise resource planning (ERP) systems, and data sources, can facilitate seamless data flow and reduce the need for manual entry.

With regards to the configuration of the ERP systems, it would be advantageous to incorporate resilient validation and error-checking mechanisms capable of early detection and rectification of errors, also in the first phase of the process so the data collection. This may encompass automated checks for data accuracy, completeness, and adherence to predefined rules or thresholds. By proactively identifying and resolving errors, the number of outliers identified by the Continuous Monitoring Model (CCM) can be significantly reduced.

One notable challenge that organizations may encounter is the lack of expertise among internal auditors when it comes to registering manual journal entries. This knowledge gap can have significant implications, potentially leading to the occurrence of outliers and false positives in fraud detection. Now it is clear that manual journal entries require a deep understanding of accounting principles, policies, and controls to ensure accurate recording and classification of financial transactions. When internal auditors lack the necessary expertise in this area, they may inadvertently introduce errors or misinterpretations, resulting in outliers that deviate from expected patterns. Furthermore, false positives in fraud detection may arise due to the auditors' limited understanding of the complexities and subtleties involved in identifying fraudulent activities within journal entries. Therefore, organizations must provide comprehensive training and ongoing professional development opportunities to internal auditors, equipping them with the requisite knowledge and skills to effectively

navigate the intricacies of a manual journal entry registrations and enhance the accuracy and reliability of financial reporting and fraud detection processes.

In the end, the Data Analytics & Continuous Monitoring Model seems to be an effective choice in preventing financial fraud for the company analyzed. The studies and further analysis conducted in this thesis have shed light on the need for internal auditors to be conscious of the risk of financial fraud and consequently implement proactive strategies to mitigate that risk. To address this issue, organizations can employ several strategies such as designing and specializing in accounting software that can minimize reliance on manual entry and improve efficiency. Additionally, standardization and the use of templates for journal entries can enhance consistency, streamline processes, and reduce the likelihood of errors.

However, organizations may face challenges related to the expertise of internal auditors in registering manual journal entries. The lack of understanding of accounting principles, policies, and controls among auditors can lead to errors and misinterpretations, resulting in outliers and false positives in fraud detection. Therefore, organizations must provide comprehensive training and ongoing professional development opportunities to internal auditors. By equipping auditors with the necessary knowledge and skills, organizations can enhance the accuracy and reliability of financial reporting and fraud detection processes.
In the end, the implementation of automation, standardization, and systems integration, coupled with the development of internal auditors' expertise, can significantly reduce the reliance on manual interventions during journal entry registration, improve efficiency, and strengthen financial reporting and fraud detection mechanisms. All these strategies contribute to the overall effectiveness and reliability of an organization's financial statement documentation processes, reducing the likelihood of errors and enhancing stakeholders' trust in the reported financial information.

BIBLIOGRAPHY

[1] Oracle, 2022, "The Evolution of Big Data and the Future of the Data Lakehouse" - How organizations use lakehouses to get more value from data, http://*analytics (oracle.com)*

[2] Ripon Patgiri, and Arif Ahmed, 2016, "Big Data: The V's of the Game Changer Paradigm", *IEEE 18th International Conference on High-Performance Computing and Communications.*

[3] Laney Doug, 2001, "3D data management: Controlling data volume, velocity and variety".

[4] Jenna Dutcher, 2022, "How important is the human element behind business analytics?", article on *Fortune - Education. https://fortune.com/education/articles/how-important-is-the-human-element-behind-business-analytics/*

[5] Catherine Cote, Harvard Business School, 2021, article "4 types of data analytics to improve decision-making", *Business Insights*.

[6] Davenport, Harris & Morison, 2010, "Analytics at Work: Smarter Decisions, Better Results", *Harvard Business Press.*

[7] KPMG LLP, 2006, "Fraud Risk Management: Developing a Strategy for Prevention, Detection, and Response"

[8] The Institute of Internal Auditors, The American Institute of Certified Public Accountants, The Association of Certified Fraud Examiners, 2008, "Managing the Business Risk of Fraud: A Practical Guide".

[9] Cotton L.D, Sandra J, Leslye G, 2016, "Fraud Risk Management Guide", *Committee of Sponsoring Organizations of the Treadway Commission (COSO).*

[10] OECD, Baesens, Van Vlasselaer &Verbecke, 2015, "Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection.

[11] INTOSAI, 2016," Guidelines on IT Audit", *ISSAI 5300, Professional Standards Committee*

[12] R. Kimball, 2014, "The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics".

[13] ACL, 2013, Institute of Internal Auditors, "Global Technology Audit Guide 13:

Fraud Prevention and Detection in an Automated World".

[14] U.S. Government Accountability Office, 2015, "Fraud Reduction and Data Analytics Act of 2015", *PUBL186.PS (congress.gov).*

[15] KPMG, Global Data & Analytics, 2016, "Steps for using data analytics to effectively assess risks", *issue 4.*

[16] KPMG, Global Data & Analytics, 2016, "The power of trust in analytics", *issue 1 Forensic and Third Party Risk Global Community (kpmg.com).*

[17] Global Forensic Data Analytics Survey EY, 2016, "Shifting into high gear: mitigating risks and demonstrating returns"

[18] Henderson & Hammersburg, 2013, "An Enterprise Approach to Fraud Detection and Prevention in Government Programs", https://vdocuments.mx/an-enterprise-approach-to-fraud-detection-and-prevention-an-enterprise-approach.html?page=1

[19] ACFE, 2016, " Report to the nation on occupational fraud", *Global Fraud study.*

[20] ACFE, 2022, "Occupational Fraud 2022: A Report to the Nations"

[21] KPMG, 2020, "Using analytics successfully to detect-fraud", *Forensic Data Analytics (kpmg.com)*

[22] KPMG, 2019, "Journal Entries principle of accounting prevision ISA Italy 240", *Forensic Data Analytics (kpmg.com)*

[23] ICAI, 2020, "Data Analytics and Continuous Controls Monitoring".

[24] Gee S., 2015, "Fraud and fraud detection: a data analytics approach", *Wiley Corporate F&A*

[25] Compliance Core article, 2020, "5 Components of effective key risk indicator", *https://compliancecore.com/blog-key-risk-indicators/*

[26] Marcello R., 2020, "The use of Big data analytics and artificial intelligence tools to preven fraud in the audit field: A conceptual frame", *Rivista italiana di ragioneria e di economia aziendale.*

[28] Arvaniti, V., 2016, "Data mining journal entries discovering unusual financial transaction", *Eindhoven University of Technology*.

[29] OECD, 2019, "Analytics for Integrity: Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments", *Analytics for Integrity OECD.*

[30] Islam S., "Three Essays on Information Security Breaches and Big Data Analytics:

Accounting and Auditing Perspective", *Louisiana Tech Digital Commons.*

[31] Walsh, 2021, "How to Encourage Employees to Speak Up When They See Wrongdoing", *Harvard Business Review, https://hbr.org/2021/02/how-to-encourage-employees-to-speak-up-when-they-see-wrongdoing*

[32] J. Heer, M. Bostock,V.Ogievetsky, 2010 "The Value of Visualization in Data Science", *A survey of powerful visualization techniques, from the obvious to the obscure,* https://dl.acm.org/doi/10.1145/1794514.1805128

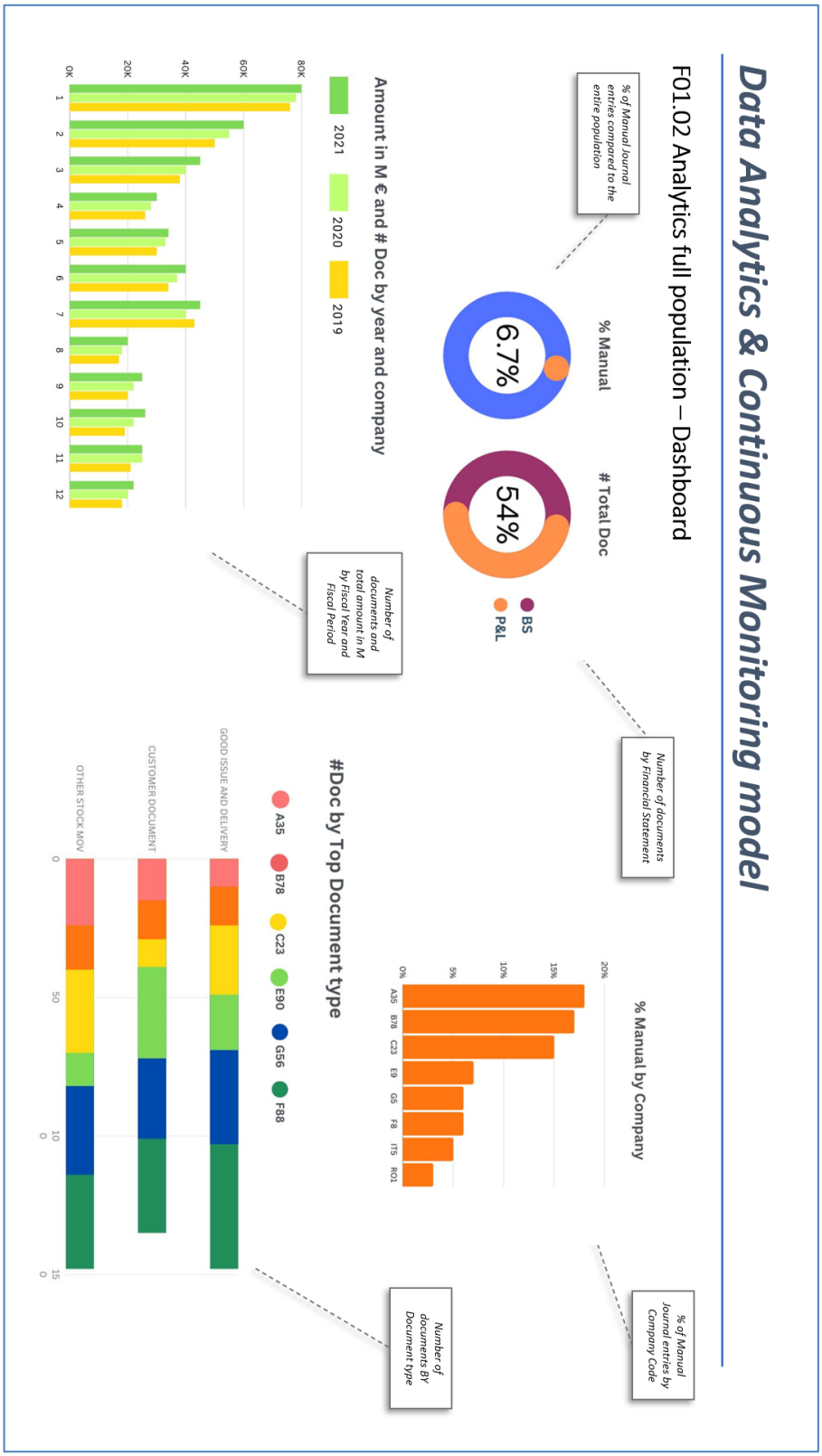
[34] A. Cairo, 2013, "The Functional Art - An Introduction to Information Graphics and Visualization".

[35] A. Jamain, 2001, "Benford's Law", *Imperial College of London Department of Mathematics.*

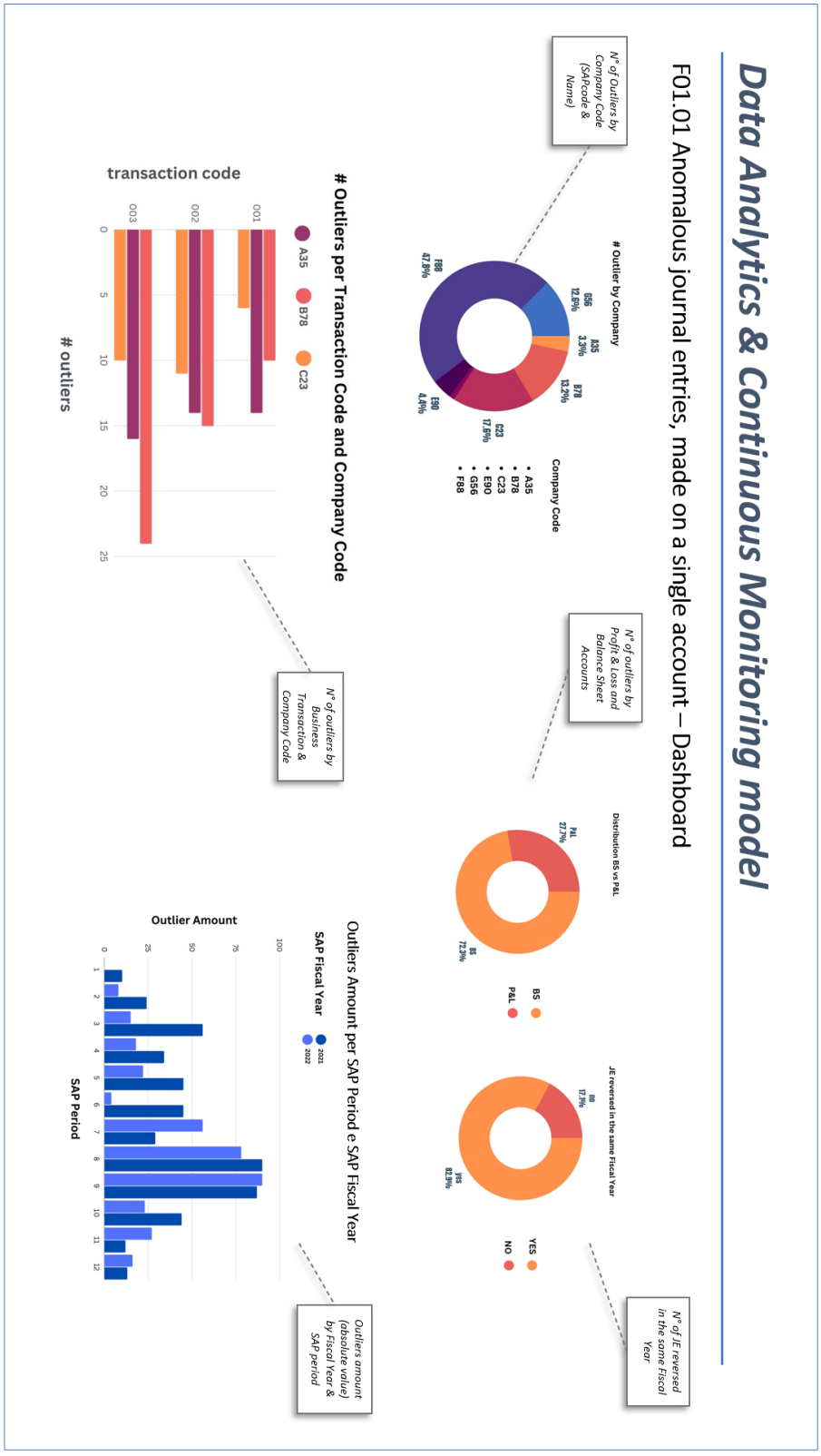
[36] N. Mansor, 2015, "Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent For Future Research", *International Journal of Academic Research in Accounting Finance and Management Sciences.*

[37] D.T.Wolfe, D.R.Hermanson,2004, "The Fraud Diamond: Considering the four elements of fraud", *Kennesaw State University.*

APPENDIX A


Data Analytics & Continuous Monitoring model
F01.02 Analytics full population – Dashboard
% of Manual Journal entries compared to the entire population
% Manual
6.7%
Total Doc
54%
BS
P&L
Number of documents by Financial Statement
% Manual by Company
% of Manual Journal entries by Company Code
Amount in M € and # Doc by year and company
2021
2020
2019
Number of documents and total amount in M by Fiscal Year and Fiscal Period
#Doc by Top Document type
A35
B78
C23
E90
G56
F88
GOOD ISSUE AND DELIVERY
CUSTOMER DOCUMENT
OTHER STOCK MOV
Number of documents BY Document type

APPENDIX B


Data Analytics & Continuous Monitoring model
F01.01 Anomalous journal entries, made on a single account – Dashboard
N° of Outliers by Company Code (SAPcode & Name)
Outlier by Company
G56 12.8%
A35 3.3%
B78 13.2%
C23 17.8%
E90 4.4%
F88 47.8%
Company Code
A35
B78
C23
E90
G56
F88
N° of outliers by Profit & Loss and Balance Sheet Accounts
Distribution BS vs P&L
P&L 27.7%
BS 72.3%
P&L
BS
N° of JE reversed in the same Fiscal Year
JE reversed in the same Fiscal Year
NO 17.7%
YES 82.3%
NO
YES
N° of outliers by Business Transaction & Company Code
Outliers per Transaction Code and Company Code
transaction code
001
002
003
A35
B78
C23
outliers
0
5
10
15
20
25
Outliers amount (absolute value) by Fiscal Year & SAP period
Outliers Amount per SAP Period e SAP Fiscal Year
SAP Fiscal Year
2021
2022
Outlier Amount
0
25
50
75
100
SAP Period
1
2
3
4
5
6
7
8
9
10
11
12