Issues in Information Systems
Volume 24, Issue 4, pp. 161-175, 2023

DOI: https://doi.org/10.48009/4_iis_2023_113

Anomaly identification through data visualization: regression analysis revisited

Assion Lawson-Body, *University of North Dakota, assion.lawsonbody@und.edu*
Patrick Glenn, *University of North Dakota, patrick.glenn@und.edu*
Laurence Lawson-Body, *University of North Dakota, laurence.lawsonbody@und.edu*
Abdou Illia, *Eastern Illinois University, aillia@eiu.edu*
Emmanuel Ufiteyezu, *University of Rwanda, ufitemmy54@gmail.com*

Abstract

In the literature, all aspects of data visualization do not guarantee anomaly detection and practical forecasting recommendations. To fill this gap in the literature, this research aims at explaining data visualization characteristics (visibility, mass notification, information sharing, emergency management, and accessibility) combined with regression methods that allow organizations to detect anomalies and make forecasting. This research revisits and provides new insights into regression analysis because of the advent of the characteristics of data visualization. A database containing 9994 records was used to test the regression model to detect anomalies using data visualization. The regression model for data visualization focused on the relationships between sales and profits. An anomaly was detected because the furniture and office supplies sold in the central region had negative profits. To further visualize, the forecast model showed that sales and profit values seemed to go much higher just before the end of each year. Discussion and conclusion were offered. This research contributes to the data visualization literature.

Keywords: data visualization, anomaly detection, regression analysis, forecasting

Introduction

Data science and analytics are essential in converting big data into strategically valuable knowledge using data visualization (Kourtit and Nijkamp, 2018; Lawson-Body et al., 2022). These present data in a format that is easier to explore, analyze, and use to support decision-making processes (Lea *et al.*, 2018). Data visualization should preferably present this strategic information in a client-friendly way via graphical design or a representation of interactive visual mappings through analytics and metrics in the form of graphs, charts, histograms, etc. (Kourtit and Nijkamp, 2018; Lawson-Body et al., 2022). In addition, data visualization should offer graphical diagnostic capabilities, colorful graphical indicators, and easy-to-read gauges. Moreover, data visualization allows users to navigate, select, and display data via an interface often used as a component of data analytics (Janvrin *et al.*, 2014; Lawson-Body et al., 2022). Furthermore, data visualization is constructed of numerous tools, such as regression methods and resources, to be utilized appropriately for the specific analytical task. This can help the organizations monitor their progress, detect anomalies, make forecasting, and identify when they must change direction to improve the productivity of their employees (Lea *et al.*, 2018).

Unfortunately, there is no one-size-fits-all approach to which data visualization tools to implement or not. Like big data and the inability to define what big data means to an individual or company, anomalies, and forecasting in data themselves are not contained in a designated box. In attempting to develop a prescription for future success, data visualization must be understood, analyzed, and potentially modified to identify anomalies and forecast accurately. In the literature, all parts or aspects of data visualization do not guarantee anomaly detection and forecasting recommendations. This is because many misunderstandings of data visualization deviate from its fundamental and original goals, anomaly identification, and forecasting. Not long ago, the ability to create innovative data visualizations was a nice-to-have skill. Primarily, it benefited data-minded managers who deliberately decided to invest in acquiring it. That has changed. Now data visualization is a must-have skill for all managers because it is often the only way to make sense of their work (Berinator, 2016).

The ability to detect valid anomalies through data visualization is much more important than evaluating data tables. Visually viewing an anomaly is much more 'user-friendly' than historical, descriptive statistics such as variance, standard deviation, skewness, and kurtosis, to highlight a few. Now, descriptive statistics are valuable tools to be used in conjunction with data visualization. Still, one does not need advanced arithmetic to identify an anomaly or question historical data when a visual aid is prepared. In addition, data visualization characteristics alone are insufficient to detect anomalies and make forecasting. Furthermore, the traditional multiple regression analyses use unit-weighted items of factor measures, ignoring the effect of measurement errors (Basco et al., 2022).

To fill these literature gaps, this research aims to explain data visualization characteristics (visibility, mass notification, information sharing, emergency management, and accessibility) combined with regression methods that allow organizations to detect anomalies and make forecasting. This research revisits and provides new insights into regression analysis because of the advent of the characteristics of data visualization.

Data visualization has helped exponentially by providing dynamic, multi-visual plots to illustrate various descriptive statistics including regression analyses. Simple time-series plots to more in-depth clustering algorithms have helped multiple companies increase revenue and reduce operating expenses. As such, anomaly detection is not for the faint of heart. Good experience in analytics and understanding what is being analyzed lead to substantial value propositions. Decisions made from forecasting techniques after data modification from inferred anomalies can lead to incorrect conclusions and costly choices. This research aims to explain anomaly detection and forecasting, the weight both tools hold on analytics, and the risk of inexperienced analysis.

This article is divided into five parts: the first part deals with related work and the theoretical background of data visualization using the technology acceptance model (TAM). The second part bears on the purpose of the study. The third part illustrates the research methodology. The fourth part presents the analysis and results. Finally, the last part deals with the discussion and conclusion.

Theoretical Background

The TAM is a widely accepted theory in information systems (IS) (Abdullah *et al.*, 2016). TAM explains how individuals come to accept and utilize innovation. Employees are inclined to accept a new technology that fits their cognitive capacity more effectively**.** Many authors used the TAM as a theoretical framework to explain user behaviors in the information technology (IT) adoption context (Lin *et al.*, 2011). For instance, Lin *et al.* (2011) concluded that websites could be valuable resources if they can help organizations gather information or complete administrative procedures quickly, efficiently, and effectively. Following Lin et al. (2011), TAM is used as a theoretical underpinning to explain the acceptance of data visualization to detect anomalies in organizations.

Related Work

Characteristics of Data Visualization

Data visualization is "the use of visual representations to explore, make sense of, and communicate data" (Sharda et al., 2018, p 101). Data visualization is utilizing data and illustrating it in various: graphs, tables, maps, and charts. The act of visualization appears to be second nature, but the appearance of the data and story the data is attempting to show weighs heavily on how it is displayed. Selecting the proper visualization and identifying anomalies decides if the data is communicated effectively. An anomaly is any observation that deviates so much from the other observations in the data set to arouse suspicion (Garg, 2020). Put to “arouse suspicion” leads to the intent that visual aid can help you identify issues and anomalies. Anomaly detection is proper forecasting through understanding the historical data, reconciling verified anomalies, and visualization of this data and future forecasts to present a story. Data visualization, anomaly detection, and forecasting are necessary in today’s business world and arguably the backbone of all financial decisions.

Data visualization was around before the 17th century AD (Sharda et al., 2018). In the 17th century, visualizations were from point A to point B, maps, roads, and resources. It was until 1644 that the notably recognized first visual representation of mathematical statistics was created by Michael Florent Van Langren, a renowned Flemish astronomer (Insightsoftware, 2022, p 1). At that time, the use of charts was assigned to the names of the individual astronomers associated with their estimates. In the 18th century, thematic mapping and abstract graphics appeared (Insightsoftware, 2022, p 1; Sharda et al., 2018). Thematic mapping included geologic, economic, and medical data, whereas the abstract renditions included errors, many functions, and empirical data. Mr. William Playfair, arguably the inventor of such well-known plots: line, bar, pie, histogram, time series, contour, and scatter, to highlight some of the first time-series multi-range graphs in 1821 (Insightsoftware, 2022, p 1; Sharda et al., 2018). In 1854, John Snow created the visual mapping of the outbreak of Cholera in London during that epidemic, and later, in 1869, Charles Minard developed the ever-famous chart of Napoleon’s army: size and location, as well as temperature and fatalities (Insightsoftware, 2022, p 1; Sharda et al., 2018).

The early 20th century led to heavy scrutiny of the lack of specifics evolving from statisticians, but later in the 20th century, predominantly due to the vivid growth of computer technology, arose the “rebirth of data visualization” (Insightsoftware, 2019, p 5). Using computer servers, hardware, software, and dashboards has made analytic tools quite popular. The days of physically plotting a static statistic are long gone, and the days of dynamic interaction have just begun.

Data visualization characteristics relevant to anomaly detection and forecasting are visibility, information sharing, mass notification, emergency management, and accessibility. Each of these is discussed below.

Visibility

Visibility is the degree to which the innovation is visible to others (Lee and Kozar, 2008). Also, visibility represents the degree to which data visualization is important to others (Lee and Kozar, 2008). Moreover, visibility is the extent to which data visualization is perceived to possess the appropriate aesthetic elements such as images, colors, labels, fonts, sizes, and others (Yigitbasioglu and Velcu, 2012; Bashirzadeh *et al.*, 2021; Lawson-Body et al., 2022).

Furthermore, visibility enriches a message with visual elements uniquely (Bashirzadeh *et al.,* 2021). Processing different visibility elements in data visualization requires distinct cognitive resources (Bashirzadeh *et al.*, 2021). The underlying philosophy of data visualization is "a picture is worth a thousand words" (Lea *et al.,* 2018). This philosophy explains that information load is an essential issue in data visualization. The visibility elements used to fix the information load issue are animations and pictographs in digital communication (Bashirzadeh *et al.,* 2021; Lawson-Body et al., 2022). An animation consists of images that create a perception of movement (Bashirzadeh *et al.,* 2021). Pictographs are small images interspersed throughout the text and serve as nonverbal cues to express emotions, ideas, or gestures (Bashirzadeh *et al.*, 2021). Employing animations or pictographs alone can be beneficial. However, using them simultaneously or too much in data visualization may be counterproductive.

Information Sharing

Information sharing is the extent to which crucial or proprietary information is available to other members through data visualization (Pham et al., 2019). Data visualization is considered a channel of information sharing. Dimensions of information sharing includes the types of shared information and channels of information sharing (Pham et al., 2019).

The TAM theory explains that graphically shared information is associated with spatial information, while tabular deliveries are for symbolic information (Yigitbasioglu and Velcu, 2012). Notably, data visualization can print graphical and tabular deliveries of information (Dyczkowski *et al.,* 2014; Karami *et al.,* 2017; Lawson-Body et al., 2022). Instead of printing those deliveries, some employees might export information to spreadsheets or Word software (Dyczkowski *et al.,* 2014; Karami *et al.,* 2017; Lawson-Body et al., 2022). In addition, most data visualization techniques are equipped with filtering and sorting features, allowing the selection of relevant reports that managers will use to make informed decisions (Dyczkowski *et al.*, 2014; Karami *et al.,* 2017; Lawson-Body et al., 2022).

From another perspective, sharing information through data visualization allows managers to stay in touch and in contact with the stakeholders (Reinking *et al.,* 2020). Those managers have mobile devices like smartphones that data visualization systems can use to share information. In addition, many data visualizations are designed to fit the size of a smartphone and are compatible with smartphone operating systems and browsers, allowing them to be accessible anywhere (Reinking *et al.,* 2020).

Mass Notification

The mass notification system is a mechanism to turn the focus to the exceptions, outliers, and data highlights (Yigitbasioglu and Velcu, 2012; Karami *et al.,* 2017; McBride *et al.*, 2020; Roud, 2021; Lawson-Body et al., 2022). Whether embedded in the data visualization or presented separately, mass notifications can be used as extra layers of abstraction to make data visualization more useful (Karami *et al.,* 2017). Through the lens of TAM theory, we posit that data visualization equipped with mass notification systems would require lower mental effort or ability than data visualization with no alerting systems. Additionally, data visualization contains mass notification features preventing sudden performance decrease in any organization. Moreover, data visualization is desirable to notify users when performance targets are missed and possible actions are advised (Yigitbasioglu and Velcu, 2012). That means employees do not need to panic in an exceptional situation. Finally, mass notification systems protect them against any surprising conditions.

Emergency management

Data visualization is necessary for emergency management in organizations. This is because emergencies are characterized by ambiguity and high stress (Roud, 2021). Also, responding to emergencies requires the capability to explain problems using text and images (Lawson-Body et al., 2022). In consequence, emergency management planning is appropriate. This planning approach must encompass collective improvisation (Roud, 2021), the capacity to determine the timing of emergencies, and the anticipation of the next step (Lawson-Body et al., 2022). Moreover, scientists and emergency managers recognized that proper data visualization might positively affect emergency management planning in organizations. This is because the messaging places the emergencies in an adequate context (Lawson-Body et al., 2022). Furthermore, the messaging provides timely information to emergency managers and coordinated responses to various user groups (McBride *et al.,* 2020; Roud, 2021; Lawson-Body et al., 2022).

Using easy-to-understand visualizations and summarized tables, authors developed data science projects for the geographic distribution of incidents and the detection of temporal patterns in the data in the public sector (Pérez-González *et al.,* 2019). They concluded that the geographic and temporal incident distribution might interest emergency services managers in designing public policies concerning security and health matters.

Accessibility

Accessibility is the effort required to access the data visualization system (Reinking et al., 2020). Also, various studies based on the diffusion of innovation (DOI) and TAM theories have empirically verified that diffusion is faster when innovations are high accessibility (Yoon *et al.,* 2020). Because of the rapid proliferation of data visualization, it is believed that almost every office worker has data visualization, whether it is accessible difficultly or easily (Yigitbasioglu and Velcu, 2012; Lawson-Body et al., 2022). In a different vein, there needs to be a balance between the complexity of accessing data visualization and its perceived ease of use. Excessive features might negatively affect the decision to access valuable data visualization (Yigitbasioglu and Velcu, 2012). According to Vilarinho *et al.* (2018), data visualizations are part of a business intelligence system that includes a broad set of complex tools and technologies. The level of accessibility of data visualization can require exceptional cognitive effort from workers. Therefore, the accessibility of innovation is usually negatively associated with its adoption and acceptance (Yoon *et al.,* 2020). Accessibility has also been an essential determinant in adopting and accepting data visualization innovations (Yoon *et al.,* 2020**).**

Data Visualization and Regression Analysis

Regression coefficients represent the mean change in the dependent variable for one single unit of change in the independent variable while controlling the constant in the model (Basco et al, 2022). The only difference between linear simple and multiple regression is that the latter incorporates more than one independent variable. Multiple regression analysis studies the relationship between a dependent variable (y) with two or more predictor or independent variables. The regression equation defines this relationship with an error term (Basco et al, 2022, Halim et al., 2023). This is an unresolved issue because neglecting measurement error can over or underestimate the impacts of relationships among the independent and dependent variables. Authors have proposed residual assumptions to develop a good regression model in response to this weakness in the multiple regression analysis (Basco et al, 2022). In this model, the errors are assumed to have independent and identical normal random distribution with zero mean and a common variance (Basco et al, 2022, Halim et al., 2023). However, by using this method, variables that have high leverage and large residuals are neglected. The technique used to solve the problem of the residual

assumptions is the robust regression for identifying outliers and influential observations (Basco et al, 2022, Halim et al., 2023). Furthermore, linear regression models based on weighted least squares regression are used to minimize the variance of errors. Finally, the residual standard error can determine whether a regression model fits most observed data (Halim et al., 2023).
Linear and multiple regression are descriptive statistics used to identify anomalies in data visualization. Both are considered as anomalies detection methods for data visualization. Of course, data visualization fills the gap between the measurement errors and weaknesses revealed using regression analysis. Regression analysis becomes complementary to data visualization. In other words, data visualization brings new insights into the descriptive statistics domain.

Furthermore, the new characteristics of data visualization have created a complementary package for organizations to take advantage of regression analysis. Undoubtedly, there are strong associations between data visualization and regression analysis. Those associations allow better anomaly detection in organizations. First, most data visualization tools can produce regression equations and analysis. Second, the additional features of data visualization make the measurements errors of regression analysis insignificant and irrelevant.

Purpose of the study

This research discusses data visualization characteristics that allow us to detect anomalies and make forecasting in organizations. These characteristics are visibility, mass notification, information sharing, emergency management, and accessibility. More importantly, descriptive analytics such as simple and multiple regression analyses are combined with these characteristics to identify organizational issues, anomalies, and flaws. As a result, data visualization brings new insights into the descriptive statistics domain. Specifically:

- Data visualization makes the measurement errors of regression analysis insignificant and irrelevant.
- Multiple case study examples illustrate the appropriate use of data visualization for anomaly detection and forecasting.
- Point, collective, and contextual anomalies are provided to identify organizational anomalies.

Research Methodology

This research has utilized two complementary methodologies to illustrate anomaly detection through data visualization: the regression method and the type or nature of the anomaly method.

In analytics, regression analysis is a visualization process for estimating the relationships between two variables (linear regression) or among more than two variables (multiple regression). The linear regression equation is a straight-line equation with an intercept across the y-axis and a slope ($\Delta y/\Delta x$) (Basco et al, 2022, Halim et al., 2023). This research focuses on the volume of sales and profits dataset as dependent and independent variables respectively to generate equation using linear regression. This research works illustrates the regression analysis using the Tableau Sample Superstore dataset.

Regarding the type of anomaly method, point, collective and contextual anomalies are demonstrated via data visualization. To understand the value of anomaly detection and put into practice what can happen, two examples were found to be presented. The first example of machine learning in action delves into an undisclosed company in Scotland that provides facility management services for more than three dozen locations. A second example case was realized from personal experience, with data from a local, anonymous oil and gas production company.

Point, Collective, and Contextual Anomaly

An "anomaly is any observation which deviates so much from the other observations in the data set to arouse suspicion." (Garg, 2020). Put to "arouse suspicion" leads to the intent that visual aid can help you identify issues without understanding why they may be an issue. Data visualization and anomaly detection are relatively simple topics to grasp, both with their usefulness in business and their application in daily life.

Thus denoted, anomalies and detection come in various shapes and sizes. This research presents three distinct anomaly types: point anomaly, collective anomaly, and contextual anomaly (Garg, 2020). As read, a point anomaly is a "single anomalous instance in a larger dataset" (Garg, 2020). A single or point anomaly can occur from a unique actual occurrence or result from an error. This anomaly is the most recognizable and easily understood anomaly as this single instance stands alone from otherwise explainable data. The following is a collective anomaly. This anomaly is when multiple occurrences are in line with one another but offset from the dataset. This anomaly is identified due to the economic circumstances and noted as such. This anomaly may be used in various sensitivities but may not be considered in historical data as part of any near-term forecasting measure. The most challenging anomaly to identify is contextual anomalies. A contextual anomaly may be an anomaly that using descriptive statistics falls within the normal distribution but proves to be an anomaly based on when it occurs. As with streaming and perpetual analytics, specific windows of time may or may not be identified to note these variances.

Data Collection

Tableau Corporation is an American company that developed and commercialized in-memory database analytics software (Hoelscher and Moltimer, 2018). Tableau has an academic version of its software that students and teachers can download for free and use for educational purposes. Tableau Desktop and Tableau prep builder are the most downloaded products. This two software are accompanied by a large dataset called sample superstore, which can be found at the following web address: https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls. The sample superstore is an Excel file that contains about 9994 records and has been adopted to test the regression equation using data visualization in this research.

Analysis and Results

Regression Analysis using Tableau Sample Superstore Dataset

Profitability Analysis

This example was focused on profit margin analysis. We edited the profit to show negative and positive profits to develop this analysis further. To do so, we calculated the profit margin considering profit divided by sales. This analysis allowed us to identify anomalies in the profitability analysis. The finding from data visualization considering all items revealed that Furniture and Office Supplies in the Central Region are unprofitable. The company has $163,797 in Total Sales for Furniture in the Central Region, resulting in a Profit margin of -$89.4. The company has 167,026 in Total Sales for Office Supplies in the Central Region, resulting in a Profit margin of -$225.9. Most products had very high sales compared to and positive profit margin.

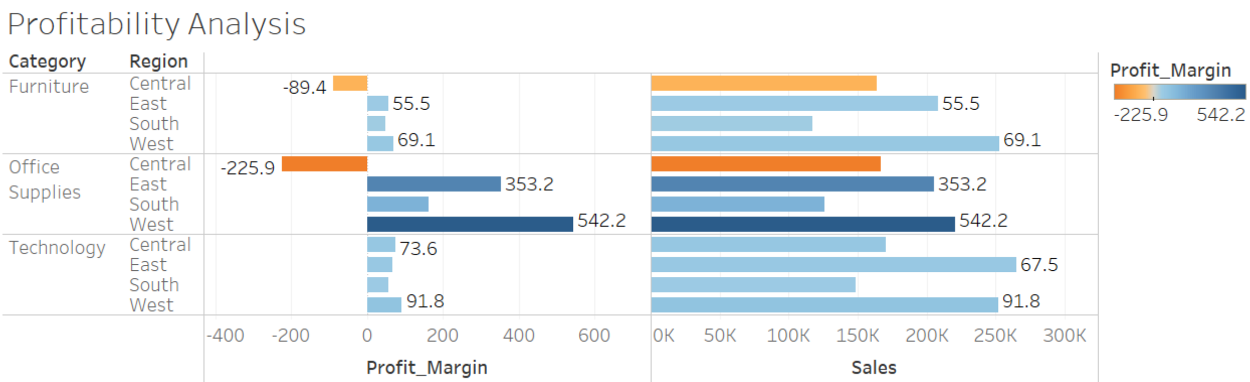

Figure 1: Profitability Analysis

Forecast of Future Values for Sales and Profit

To further visualize, the forecast model shows that sales and profit values seem to go much higher just before the end of each year. The trends of the sum of sales (actual and forecast) for Order Date Month are broken down by segment. The color shows details about the category of products and the Forecast indicator. The data is filtered on Region and Order Date. The Region filter keeps Central, East, South, and West. The Order Date filter ranges from January 3, 2018, to December 30, 2021.

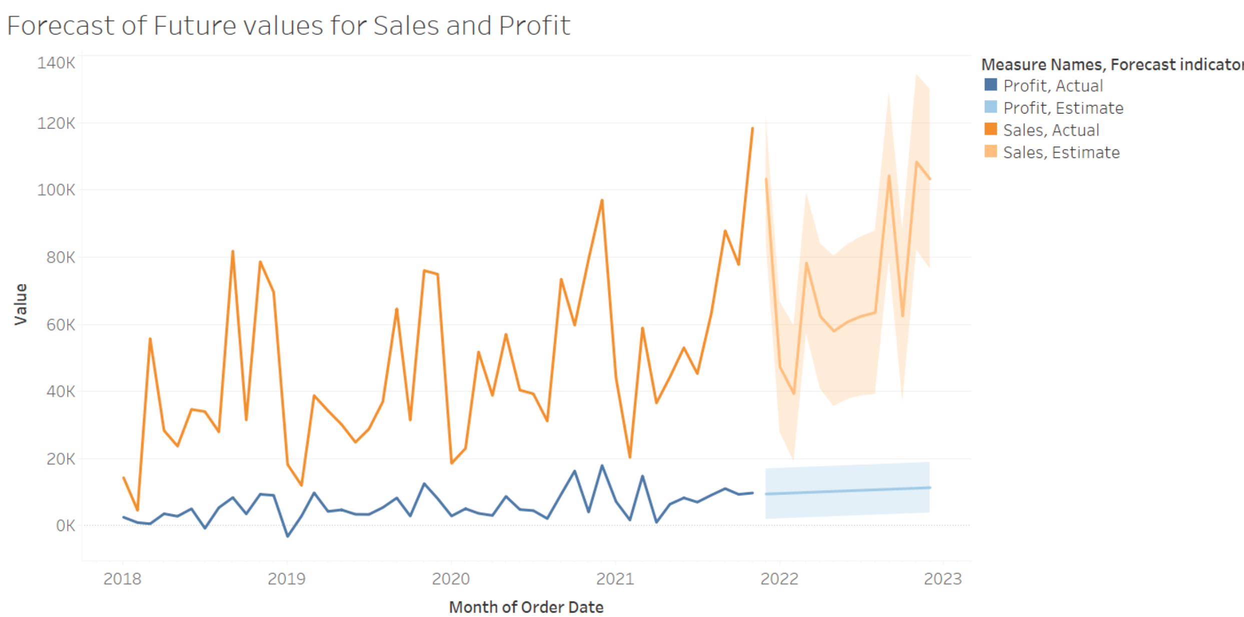

Figure 2: Forecast of Future Values

Comparison between sales and profit values

Figure 3 shows how scatter plots visualize the relationships between sales and profit variables. These two variables are numerical ones. This visualization allowed us to compare sales to profit, resulting in a chart analogous to a Cartesian graph with x and y coordinates. The three categories of products are furniture, office supplies, and technology. The number of distinct regions is multiplied by the number of departments. The linear regression provides a statistical definition of the relationship between sales and profit. Table 1 shows sales (y) and profit (x) as the dependent and independent variables. The value of the slope of the data ($\Delta y/\Delta x$) is 4.44659. The value of the Y-axis intercept or constant (b) is 47343.9. That is the value of sales when the profit equals zero. In order words, it is the break-even or the minimum level of sales the company must have to survive. The total number of samples in the dataset (n) is 9994.

Overall, the coefficient signs are positive (+). That indicates the sales increase as the profit increases. The value of the sales increases with the rate of 4.44659, while the value of profit increases too. The coefficient of determination (R-Squared) is the proportion of variance for sales, or the variability of sales explained by profit. The P-value is 0.0495233, which indicates a P-value less than 0.05 ($p < 0.05$). This result suggests a significant P-value. We conclude that sales impact the profit because changes in the sales' value are related to changes in the profit's value.

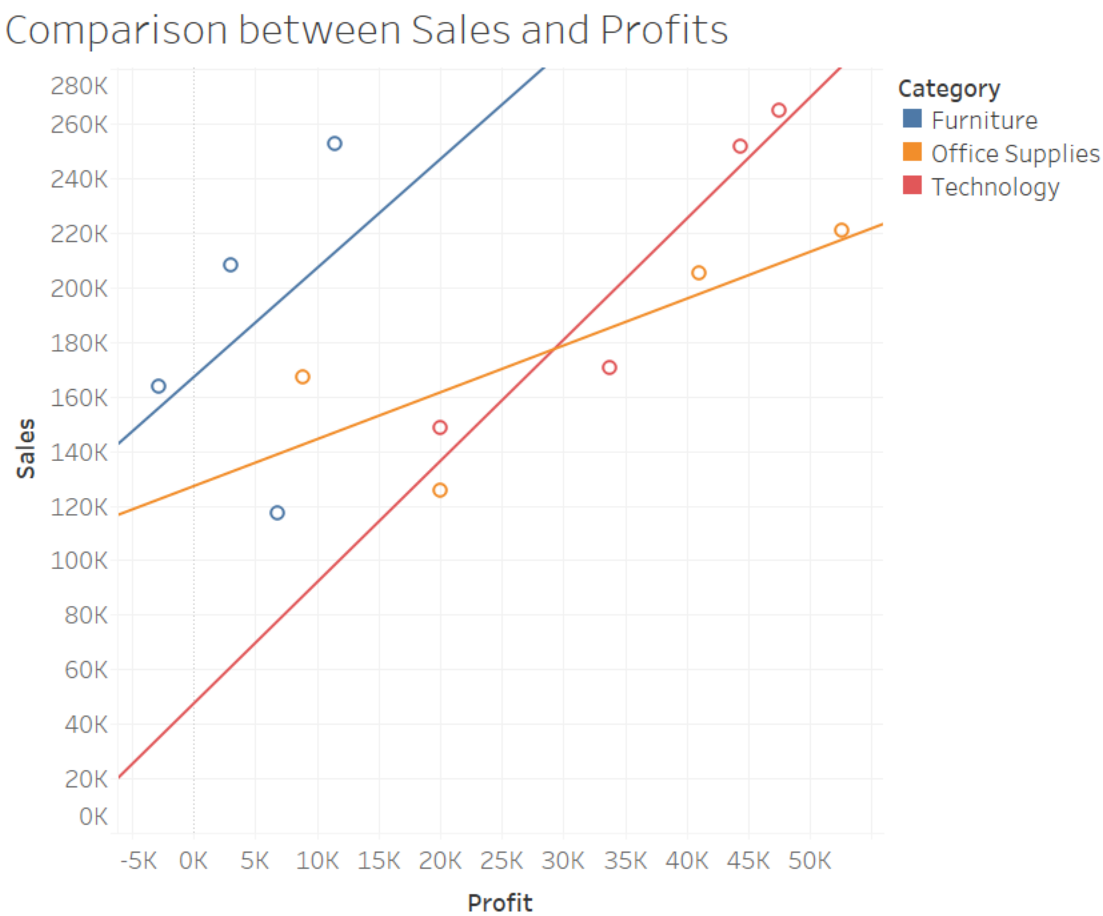

Figure 3: Comparison between Sales and Profit

Table 1: Regression Analysis Results

Sales = 4.44659*Profit + 47343.9
R-Squared: 0.903406
P-value: 0.0495233

Illustration of Anomaly Detection using Case Examples of Electrical and Oil/Gas Production Companies

Electrical consumptions case study

To set the context, the essence of this case study is looking into the truth behind 'low energy' buildings. The promises of low-energy buildings in the United Kingdom often consume much more resources than promised; in some instances, anywhere from 2-11% excess power is consumed (Cui and Wang, 2017). A task presented to the management service was constantly monitoring electrical consumption to identify any noticeable anomalous events. Before implementing the machine learning algorithm, employees realized a multitude of hours to review near real time data via tabular format. By employing multiple historical fits to the weekly power consumption, and breaking the models based on best-fit analysis, two time-series forecasts were matched, one for the weekdays and one for the weekends. By understanding traffic on location, identifying two models to fit business hours was appropriate here. An error interval was allowed based on variance and applied. With the forecasting model in place, two groups of anomalies were detected, point and collective anomalies (Cui and Wang, 2017). It was then understood that point anomalies were results of brief meter errors, whereas the collective anomalies assisted in identifying issues with scheduled usage (See Figure 4). Cost savings were realized quickly by identifying a timing issue associated with scheduled heating and promptly addressed. Further, this implementation, the discussion on 'low energy' buildings leads to further discussion that could be human error rather than construction shortcomings.

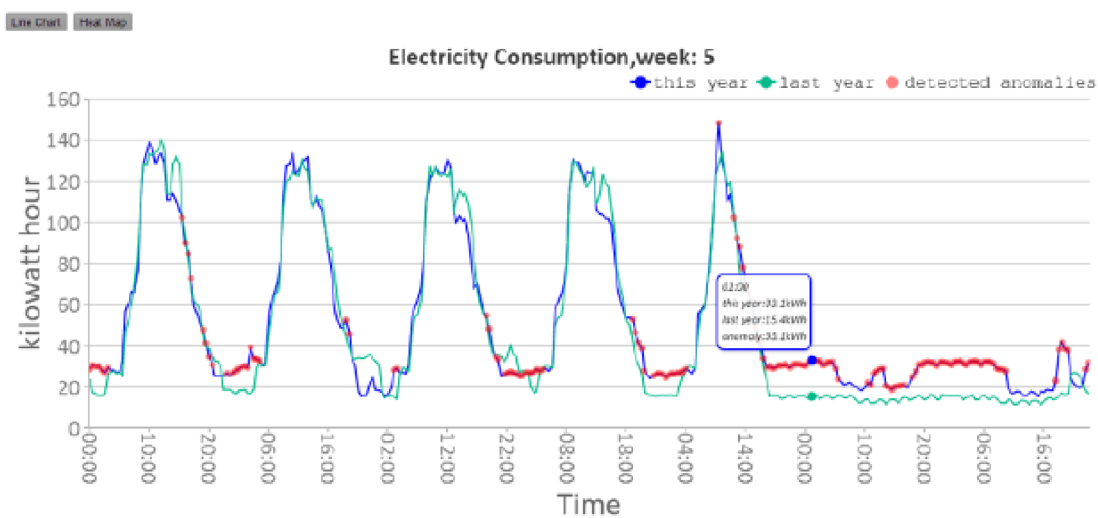

Figure 4. Electrical Consumption with Anomalies Identified

Anonymous oil and gas production company case study

For context, the streaming analytic period illustrated is twelve months, whereas a longer time-series plotting was required. In reviewing the visual line plot of monthly operating expense per category in Figure 5, two easily identified point anomalies are evident. The first point is in April 2021, with a substantial credit of $247,101; the second is in July 2021, with a total value of $266,426. The first anomaly is in the property tax subcategory, and the second is direct field wages. Maintaining these data points through additional information was essential, though this seems unlikely without understanding the data. The reasons to incorporate this data and not cull it is due to the annual short-term incentive plan bonus' distributed in July each year and renegotiation on tax value each year from previous years' payment. These are historically realized annually recurring payments (credits), essential to incorporate. As such, forecasting methods that consider a sort of seasonality appear to be the first place to start for pertinent data. By including this information, the data quality is withheld, the context is realized, and minimal bias is translated into the forecasting (Stikeleather, 2017).

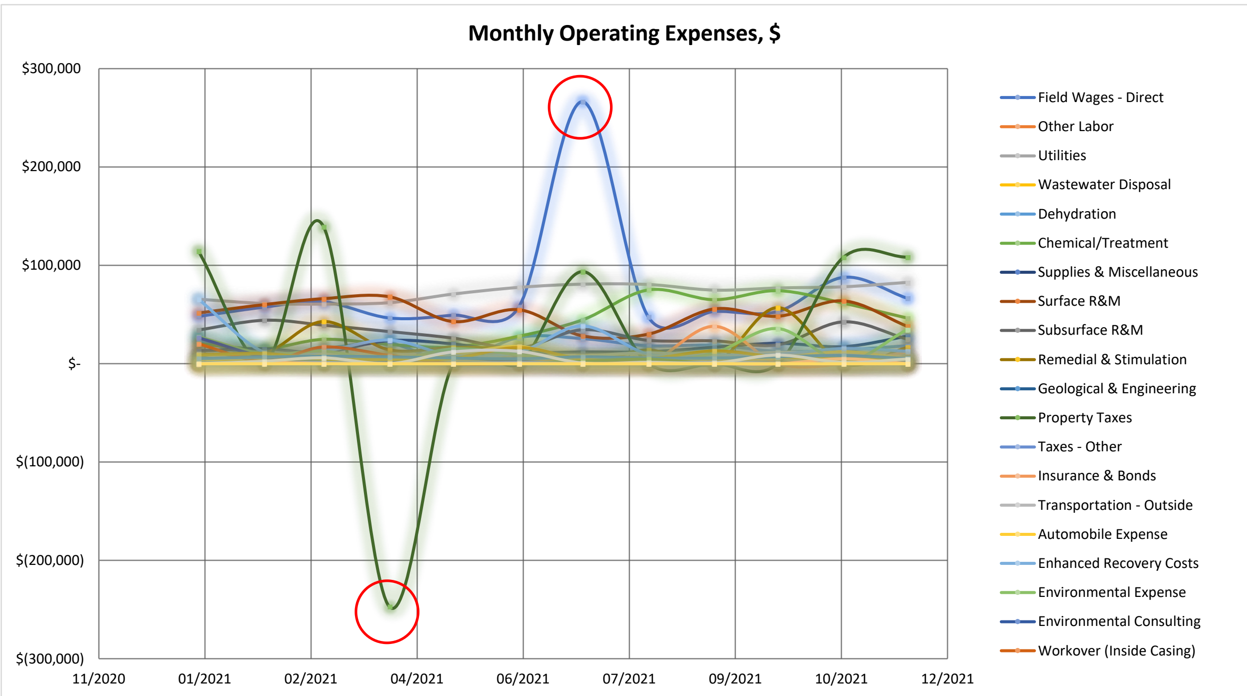

Figure 5. Monthly Operating Expenses with Anomalies Identified

Discussion

Thus far, we have discussed three common anomaly detections and regression analysis methods for data visualization. But we have not focused on the importance of understanding the data itself. Without proper knowledge of what is being analyzed, an adequate timeframe in which data is accessed, visible, shared, and diagnosed, and without validation of assumptions utilized, heavy misses can occur, both in description and prediction. So often, as visualization is a potent tool, creators can get caught up in appearance and lose sight of the message, in agreement with Stikeleather (2017), who indicated that visualization shows beautiful exercises in special effects that show off statistical skills but do not serve an informing purpose. Without a clear purpose, or ability to convey a message, visualization is simply art rather than a potential business solution. This research's guidelines for effective data visualization include visibility, mass notification, information sharing, emergency management, and accessibility. To further confirm the above, visibility in data visualization shows elements like a video that generates more interest, attracts user cognitive ability, and promotes digital smiles like emojis to improve relationship strengths (Bashirzadeh *et al.,* 2021).

Information sharing should be present in data visualization because it allows relevant information to co-workers, managers, and stakeholders (Yigitbasioglu and Velcu, 2012; Lea *et al.,* 2018; Schotter *et al.*, 2018). This corroborates the assumption of Kim *et al. (*2020), who stipulate that data visualization allows transferring knowledge to employees to enhance their productivity. Many scholars have investigated data visualization modalities, such as audio, video, and text, on the performance and efficiency of the information-sharing process. Concerning mass notification, this research stipulates those exceptions can be raised when researching anomaly detection and forecasting through visualization—mass notification involves presenting the correct story to raise questions and further ingrain understanding of the business model. The ability to identify issues over various periods and the potential for algorithms to provide warning of excess deviation from expected are necessary components of data visualization.

Based on the analysis in this study, easy emergency management makes data visualization more useful in organizations. This aligns with the thoughts of McBride *et al. (*2020) and Roud (2021), who indicated that data visualization facilitates coordinated responses in emergencies. Also, one of the main tasks of emergency management in data visualization is to boost the predictive and forecasting endeavor of organizations.

Finally, the accessibility of data visualization allows managers to stay in touch and contact with the information provided through the system (Reinking et al., 2020). Yoon *et al. (*2020) added that appropriate accessibility allows data visualization to be perceived as relatively easy to understand and use. Accessibility becomes a priority because it helps to get the message across about what is happening and what to do about it. Without proper understanding and training on how to apply various techniques, anomalies may not be anomalies, incorrect forecasting without external influence, and worthless visual representation of data may be unhelpful and worse, lead to catastrophic prescriptions.

Conclusion

Because all aspects of data visualization do not guarantee anomaly detection and practical forecasting recommendations, this research aims at explaining data visualization characteristics (visibility, mass notification, information sharing, emergency management, and accessibility) combined with regression methods that allow us to detect anomalies and make forecasting. A database containing 9994 records was used to test the regression model to detect anomalies and make forecasting using data visualization. The regression model for data visualization focused on the relationships between sales and profits. Anomalies were found because the furniture and office supplies sold in the central region have negative profits. However, most products had very high sales and positive profit margins. To further visualize, the forecast model showed that sales and profit values seemed to go much higher just before the end of each year.

Further multiple case study examples are provided to illustrate the appropriate use of data visualization for anomaly detection and forecasting—the first case study of machine learning algorithms for electrical consumption timing anomalies. The second case study highlights how relevant data consideration is, and the potential errors associated with misunderstanding. The third case study outlines an anomaly related to a region with negative profits despite having the highest quantity of products sold. The last and fourth example is forecasting.

The basis of this paper is proper forecasting through understanding the historical data, reconciling verified anomalies, and visualization of this data and future forecasts to present a story. Understanding the characteristics of data visualization can be used by many, inexpensively and with incredible value propositions. To reiterate, data visualization has been around for some time. However, with the advent of big data, exponential advancement of technology, and affordability of the solutions providers, caution must be an essential parameter to keep in mind while utilizing said tools.

References

Abdullah, F., Ward, R., & Ahmed, E. (2016). Investigating the influence of TAM's most commonly used external variables on students' Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) of e-portfolios, *Computers in Human Behavior, 63*, 75-90, http://dx.doi.org/10.1016/j.chb.2016.05.014

Basco, R., Hair, J. F., Ringle, C. M., & Sarstedt, M. (2022). Advancing family business research through modeling nonlinear relationships: comparing PLS-SEM and multiple regression, *Journal of Family Business Strategy, 13,* https://doi.org/10.1016/j.jfbs.2021.100457

Bashirzadeh, Y., Mai, R., & Faure, C. (2021). How rich is too rich? Visual design elements in digital marketing communications, *International Journal of Research in Marketing*, In press, https://doi.org/10.1016/j.ijresmar.2021.06.008

Berinato, S. (2016). Analytics and Data Science: Visualizations that really work. *Harvard Business Review,* Retrieved October 15, 2022, from https://hbr.org/2016/06/visualizations-that-really-work

Cui, L., & Wang, H. (2017). Anomaly Detection and Visualization of School Electricity Consumption Data, *In Proceedings of 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Institute of Electrical and Electronics Engineers (IEEE), Held 10-12 March 2017, Beijing, China.

Dyczkowski, M., Korczak, J., Dudycz, H. (2014). Multi-criteria Evaluation of the Intelligent Dashboard for SME Managers based on Scorecard Framework, *Federated Conference on Computer Science and Information Systems*, 1147–1155

Garg, S. (2020). Algorithm selection for anomaly detection. *Analytics Vidhya,* Retrieved October 18, 2022, from https://medium.com/analytics-vidhya/algorithm-selection-for-anomaly-detection-ef193fd0d6d1

Halim, G. A., Agustin, P., Adiwijayanto, E., & Ohyver, M. (2023). Estimation of cost of living in particular city using multiple regression analysis and correction of residual assumptions through appropriate methods, *Procedia Computer Science, 216,* 613-619, http://dx.doi.org/10.1016/j.procs.2022.12.176

Hoelscher, J., & Moltimer, A. (2018). Using Tableau to Visualize data and Drive Decision-Making, Journal of Accounting Education, 44, 49-59. https://doi.org/10.1016/j.jaccedu.2018.05.002

Janvrin, D. J., Raschke, R, N., & Dilla, W. N. (2014). Making sense of complex data using interactive data visualization, *Journal of Accounting Education*, *32,* 31-48, https://doi.org/10.1016/j.jaccedu.2014.09.003

Karami, M., Langarizadeh, M., & Fatehi, M.. (2017). Evaluation of Effective Dashboards: Key Concepts and Criteria, *The Open Medical Informatics Journal*, *11*, 52-57

Kim, J., Shin, S., Bae, K., Oh, S., Park, E., & del Pobil, A. P. (2020). Can AI be a content generator? Effects of content generators and information delivery methods on the psychology of content consumers. *Telematics and Informatics, 55,* Article e101452. https://doi.org/10.1016/j.tele.2020.101452

Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities – An experiment on Stockholm. *Land Use Policy, 71,* 24-35. https://doi.org/10.1016/j.landusepol.2017.10.019

Insightsoftware. (2022, November 4). *A brief history of data visualization.* insightsoftware. Retrieved June 4, 2023, from https://insightsoftware.com/blog/a-brief-history-of-data-visualization/

Lawson-Body, A., Lawson-Body, L., & Illia, A. (2022). Data Visualization: Developing and Validating Dashboard Measurement Instruments, *Journal of Computer Information Systems*, https://doi.org/10.1080/08874417.2022.2073295

Lea, B-R., Yu, W-B., & Min, H. (2018). Data visualization for assessing the biofuel commercialization potential within the business intelligence framework. *Journal of Cleaner Production, 188,* 921-941. https://doi.org/10.1016/j.jclepro.2018.02.288

Lee, Y., & Kozar, A. K. (2008). An empirical investigation of anti-spyware software adoption: a multitheoretical perspective. *Information and Management*. *45(2),* 109-119. https://doi.org/10.1016/j.im.2008.01.002

Lin, F., Fofanah, S. S., & Liang, D. (2011). Assessing citizen adoption of eGovernment initiatives in Gambia: a validation of technology acceptance model in information systems success. *Government Information Quarterly*, *28(2),* 271-279. https://doi.org/10.1016/j.giq.2010.09.004

McBride, S. K., Bostrom, A., Sutton, J., de Groot, R.M., Baltay,A. S., Terbush, B., Bodin, P., Dixon, M., Holland, E., Arba, R., Laustsen, P., Lui, S., & Vinci, M. (2020). Developing post-alert messaging for ShakeAlert, the earthquake early warning system for the West Coast of the United States of America, *International Journal of Disaster Risk Reduction, 50,* Article e101713, https://doi.org/10.1016/j.ijdrr.2020.101713

Pérez-González, P. J., Colebrook, M., Roda-García, J., & Rosa-Remedios, C. B. (2019). Developing a data analytics platform to support decision making in emergency and security management, *Expert Systems With Applications, 120*, 167-184, https://doi.org/10.1016/j.eswa.2018.11.023

Pham, C. H., Nguyen, T-T, McDonalds, S., & Tran-Kieu, N. Q. (2019). Information sharing in logistics firms: An exploratory study of the Vietnamese logistics sector, *The Asian Journal of Shipping and Logistics*, *35(2),* 87-95. http://dx.doi.10.1016/j.ajsl.2019.06.001

Reinking, J., Arnold, V., & Sutton, S. G. (2020). Synthesizing enterprise data through digital dashboards to strategically align performance: Why do operational managers use dashboards? *International*

Journal of Accounting Information Systems, *37*, Article e100452. http://dx.doi.10.1016/j.ajsl.2019.06.001

Roud , E. (2021). Collective improvisation in emergency response, *Safety Science, 135*, Article e105104, https://doi.org/10.1016/j.ssci.2020.105104

Schotter, A. P.J., Buchel, O., & Vashchilko, T. (2018). Interactive visualization for research contextualization in international business. *Journal of World Business*, *53*, 356-372. http://dx.doi.org/10.1016/j.jwb.2017.01.006

Sharda, R., Delen, D., Turban, E., Aronson, J. E., Liang, T.-P., & King, D. (2018). *Business Intelligence, analytics, and Data Science: A Managerial Perspective* (4th ed.). Pearson.

Stikeleather, J. (2017). Analytics and Data Science: when data visualization works - and when it doesn't, *Harvard Business Review,* Retrieved October 15, 2022, from https://hbr.org/2013/03/when-data-visualization-works-and

Vilarinho, S., Lopes, I., & Sousa, S. (2018). Developing dashboards for SMEs to improve performance of productive equipment and processes, *Journal of Industrial Information Integration, 12,* 13-22. https://doi.org/10.1016/j.jii.2018.02.003

Yigitbasioglu, O. M. & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research, *International Journal of Accounting Information Systems, 13*, 41-59. https://doi:10.1016/j.accinf.2011.08.002

Yoon, C., Lim, D. & Park, C. (2020). Factors affecting adoption of smart farms: The case of Korea, *Computers in Human Behavior, 108*, Article e106309. https://doi.org/10.1016/j.chb.2020.106309