**Data Breaches Affecting Washington State Residents**

**Guillermo J. Figueroa**

**Ph.D., Business Analytics, University of South Alabama**

**BUS 754 Applied Business Analytics II**

**Dr. Ermanno Affuso**

**11/17/2024**

**Abstract**

Data breaches have become a challenge for the current business environment. A data breach is the unauthorize acquisition of data that compromises the security, confidentiality, integrity of personal information maintained by a person, or entity. During the year 2023 Washington State had ($4 million) the third highest total number of data breach notices. (Data.wa.gov) September 2024. The Attorney General’s office (AGO) received 133 data breaches notification. This study analyzes a secondary dataset in Washington State. The State law requires entities impacted by a data breach to notify individual whose personal information has been hacked or stolen, and to comply with the Attorney General if more than 500 residents are impacted by the breach. The number of Washingtonians that experience data breaches were analyzed and including the financial performance for the entities that disclose their record publicly. We analyzed the days to identify the breach and the days elapsed before the notification. There is no correlation, within the independent and dependent variables. Washington, businesses revenue was not impacted by the data breaches. Factors such as the nature of the breach, customer loyalty, and other initiatives may drive the potential economic impact on company performance.

***Keywords***data breaches, cybersecurity, consumer privacy, data privacy, internet

**Introduction**

The rise in data breaches threatens Washingtonians privacy and undermines public trust in businesses. As we continue to witnessed these incidents, there is limited research on the broader scale of data breaches on consumer behavior, businesses and performance. Over the past 25 years the internet brought significant changes to the traditional business model by creating new ways of doing business. Digital Marketing has been a main force, increasing the rate of consumer data collection. These initiatives have generated high volume of online activities and transactions between consumers and businesses. Consumers are active on social media, health care portals to communicate and schedule online doctors’ appointments. The financial industry has been promoting customer mobile and digital services to meet their financial needs. The current economic environment and the usage of internet has raised concerns as the shared data becomes accessible to other devices and companies. This represents a risk for potential data breaches. Private information stored in company data bases is vulnerable to hackers. (Markos et al., 2023).

The current high-tech evolution and information environment has led to the increase volume of data to be targeted. Cybersecurity has become a very important discipline and advancements continue to take place in technology and open AI. Cyber security protects the security, integrity, confidentiality of communication, life, integration, tangible or intangible assets, and protect the data. In summary, cyber security ensures the security of virtual life on cyber networks. (Aslan et al., 2023). To overcome threats it is important to understand the situation and be prepared to implement tools to help prevent potential data breaches.

One of the most impactful data breaches in the financial industry occurred at Equifax in 2017, a credit reporting agency. The breach affected 143 million customers, exposing credit card accounts, social security numbers and drivers’ licenses. (Landmann, 2023).

The purpose of the research explores the following:

The number of residents in Washington State impacted by the Data Breach. Days to identify the breach and Days elapsed before notification. Our objective is to evaluate how data breaches affects consumer trust in local businesses and revenue.

**Literature Review**

The development and expansion of network technologies has created critical security problems. Various attempts have been made to provide a cyber security environment to protect assets, of the different institutions, organizations and individuals. (El Amin et al., 2024). In today’s digital transformation world, privacy has become the most protected human right and privilege (Aiello, 2024). As the technology enables the exchange of information, the need to protect personal and private data. We have seen through the work of previous researchers violations lead to decreased trust in a company or website and trustworthiness and perceived information safety determine a consumer’s willingness to share personal information with a business. (Markos et al., 2023).

It is required for consumers to make online purchases, interact on social media platforms, or connect with health care providers, and exchange personal information through these platforms. The growth that e-commerce and the smartphone has generated concerns about online data security as breaches continue to rise. E-Commerce expansion was accelerated due to social distancing measures changing consumer to spend more time online. (IBISWorld, August 2024). Consumers feel unprotected when a social contract is disrupted by a data breach, resulting in the loss in personal information. Digital transformation, and connectivity, and intelligent systems, digital breaches have become an area of concerned. Such attacks may cause financial damage to individuals, institutions, big companies, and even state governments (Aslan et al., 2023). The lack of effective cybersecurity measures has proven to cause increasing losses and can range from financial damages, highlighting the important need for robust cyber risk management. The goal of this research is to assess the impact of Public and Private data breaches on the number of residents affected in Washington State. The Law requires entities impacted by a data breach to notify the Attorney General’s Office (AGO) when more than 500 Washingtonians personal information was compromised as a result of the breach. (data.wa.gov). The remainder of the paper is organized as follows. The following section presents the subsequent hypothesis development. The description of the methodology employed to test the hypotheses and report the findings. Finally, we discuss the findings in the context of prior studies and examine the implications for theory and practice.

**Hypotheses**

Hypothesis 1 (H₁): The number of Washington Residents impacted by the data breach will affect the revenue of companies in Washington State.

Null Hypothesis (Ho): There is no relationship between the number of Washington residents impacted by the data breach and the revenue of companies in Washington State.

Alternative Hypothesis (H₁): There is a relationship between the number of Washington residents ᶟᶟimpacted by the data breach and the revenue of companies in Washington State.

We fail to reject the Ho would indicate that there is a significant relationship

Hypothesis 2 (H²): The time that it takes to identify the data breach will affect consumer satisfaction and the willingness to continue to do business with local companies.

Null Hypothesis (Ho): The time to identify the data breach has no effect on consumer satisfaction or the willingness to continue doing business with local companies.

Alternative Hypothesis (H₁): The time to identify the data breach does affect consumer satisfaction or the willingness to continue business with local companies.

We fail to reject Ho would imply that the time to identify the breach does not significantly affect consumer satisfaction or willingness to continue doing business.

Hypothesis 3(H3). The days elapsed before notification of the data breach will affect company reputation and revenues.

Null Hypothesis (Ho): The days elapsed before the notification of the data breach has no effect on company reputation or revenues.

Alternative Hypothesis (H₁): The days elapsed before the notification of the data breach does affect company reputation or revenues.

We fail to reject the Ho would indicate that the days elapsed until notification do not significantly affect company reputation or revenues.

Before examining the Research questions, we outline a brief descriptive analysis of data breach trends. Firs, it provides the evidence of the depth of data breaches as part of this research. Second, this descriptive analysis allows us to compare and identify which type of entity are most effected. Then we can refer to these trends as we develop our hypothesis around how different types of information impact the willingness of consumers to do business with the company, and their behavior following the data breach.

**Methodology**

Research Questions

RQ1: To what level the data breaches affecting Washingtonian residents influence the revenues of the companies in the State

RQ2: Is there a statistically significant relationship between the frequency of data breaches and corporate revenue for entities located in Washington State?

RQ3: What are the average days that it takes to identify the breach?

**Dependent Variable (DV)**

The dependent variable is the revenue of the company or entities affected by the data breaches. This variable will measure how the revenue of these companies changes following the data breach event. Through the research we want to measure and assess the impact of data breach. The measurement will be based on an ordinal scale based on the numbers of Washingtonians affected by the data breach.

**Independent Variable (IVs)**

Through the research I want to measure and assess the Days to Identify the breach. The measurement will be based on an ordinal scale based on the numbers of Washingtonians affected by the data breach, days to identify data breach and days elapsed to notify the residents about the data breach incidents.

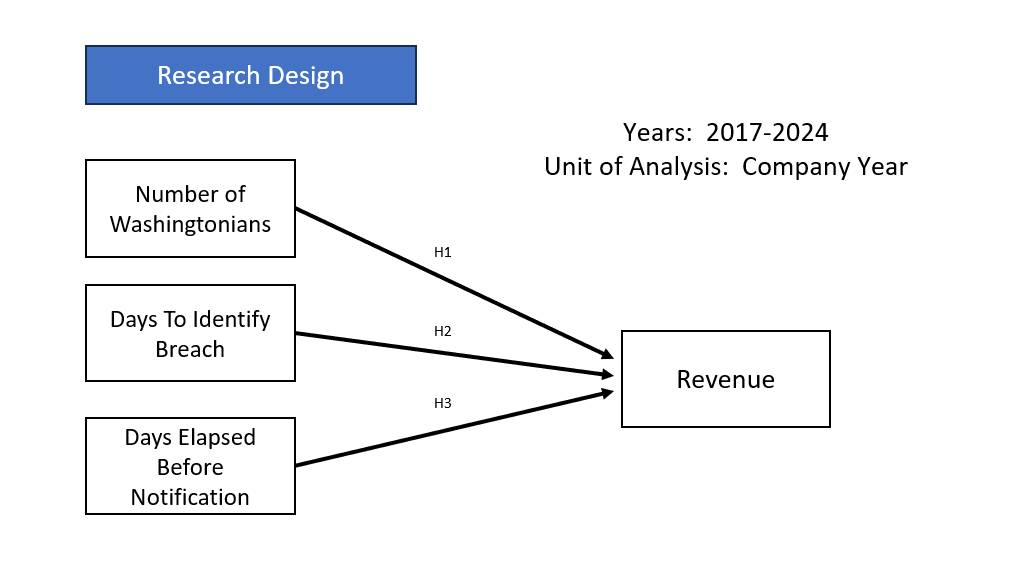
**Unit of Analysis**

The unit of analysis is the individual residents.

Samples of Population:

The estimated or known number of Washington residents whose information was affected by the data breach. (Figure 1. Model Variable illustrates the description).

Figure 1 illustrates the theoretical model and Research Design

****

**Fig. 1**. **Model Variable**

**Data Description**

The data used is provided by Washington State Attorney General’s Office Consumer Protection Division. It’s composed by 1,209 rows and 25 columns. The period of time is from August 2015 to present. For this research we considered the data through 2023. The variables used for the study are Washingtonians Affected which is the estimated or known number of Washington residents whose information was affected by the data breach. Days to Identify Breach is the total number of days it takes on an entity to discover that a breach of consumer data has occurred after the breach has begun. The Days Elapsed Before Notification is the number of days that elapsed before the notifying entity submitted notice to the AGO. Revenue- The revenue for the entities involved in data breach was not provided in the data, however additional research was done to gather the information published utilizing the financial and economic resources to the public.

Preprocessing

After loading the dataset, it required data cleaning to handle missing values, remove duplicates and fix the inconsistencies with some of the entries. The data was split into the training and testing sets to evaluate the mode’s performance.

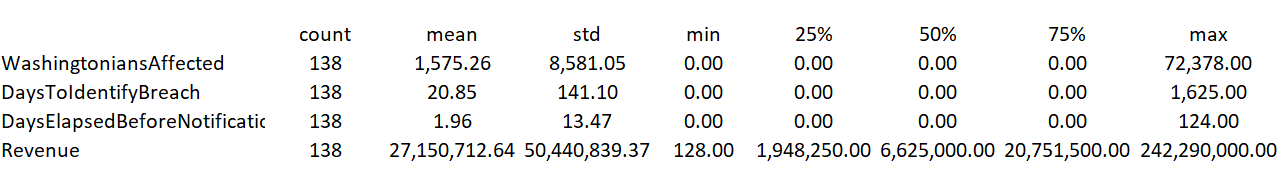
Analytical Methods- For this research we used Machine Learning, Python, and analytical techniques were used to determine, predict and measure current situation to help improved and protect Washingtonians from cybersecurity threats. The objective is to use these techniques to identify trends and help improve the safety for individuals and the organizations.

Validation

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MSE | RMSE | Overfitting |
| Kfold Train | 5.20163E+15 | 72122357.24 | No |
| Kfold Test | 2.56143E+15 | 50610533.12 |  |
| OLS Train | 5.69963E+15 | 23867022.82 | No |
| OLS Test | 3.06281E+15 | 55342650.61 |  |

The KFold model was used to cross validation process. It provides the capability by splitting the dataset into multiple subsets to test each one. The KFold Test error MSE of 2.56143E+15 is smaller than the KFold Train error MSE 5.20163E+15. This indicates that the model is performing better than on the training data. The KFold model is underfitting. For the OLS model, the training error MSE of 5.66963E+15 is higher than the test error MSE of 3.06281E+15. Typically the training error is lower than the test error. The model is underfitting the training data.

**Results and Discussion**

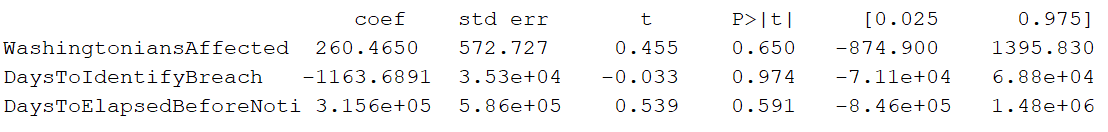


On average there is about 1,575 Washington Residents affected per incident. The standard deviation is much larger than the mean. Some incidents have a much higher of individuals that were affected. On average it takes 21 days to identify a breach. The standard deviation of 141 indicates that the number of days to identify breaches varies with some being identified much later than others. The average notification is being done on almost immediately. While the average is low, some notifications are taking longer period of time. The longest time before making a notification was 124 days. The average revenue across all the incidents was $27 million. The revenue varies among the entities that were affected by the breaches.

**Correlation Matrix**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| The -0.521 correlation indicates that as the number of Washingtonians affected by a breach increase, there is a slight tendency for the revenue to decrease. There is no linear relationship between the two variables. The entities with higher revenues tend to take less time to identify breaches. The correlation is -0.0691. This means that the correlation is low indicating that there is not a strong meaningful relationship between Revenue and Days to Identify the Breach. The correlation of 0.0498 shows that there is a weak positive correlation between Washingtonians Affected and Days to Identify Breach. As more people are affected by the breach, it may take longer time to identify the breach. Since the correlation is weak, there is a minor relationship. The correlation of -0.0010 shows that there is no correlation between Washingtonian Affected and Days Elapsed Before the Notification. There is no linear relationship. The correlation of 0.0976 is a positive weak correlation between Days to Identify Breach and Days Elapsed Before Notification. The breaches that take longer to identify tend to have a slightly longer time before a notification is sent. |  |  |  |  |

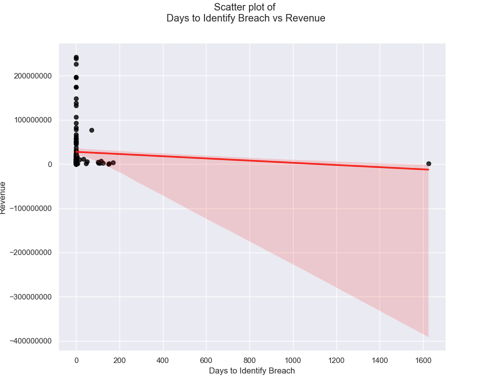
***Regression Analysis***

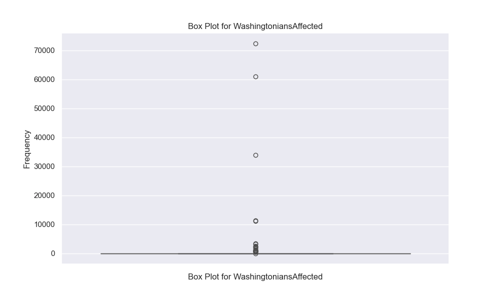
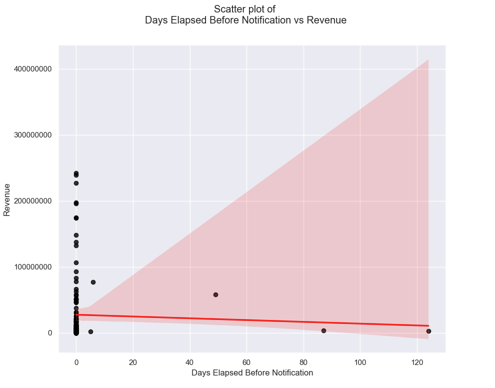


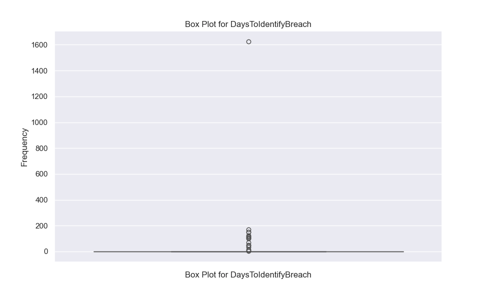
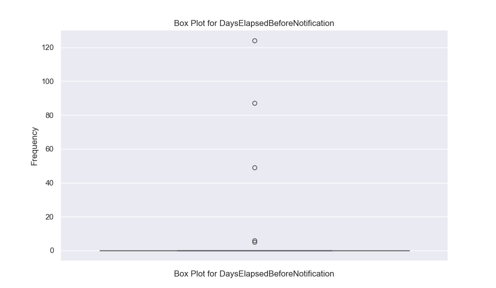
The predictor variables do not show a statistically significant relationship with the dependent variable, as the p-values are greater than 0.05. The lack of statistical significance presents that these variables cannot provide an outcome variable in this model.

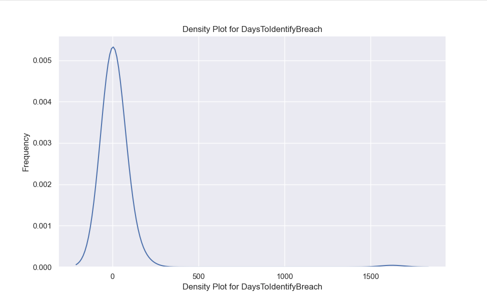
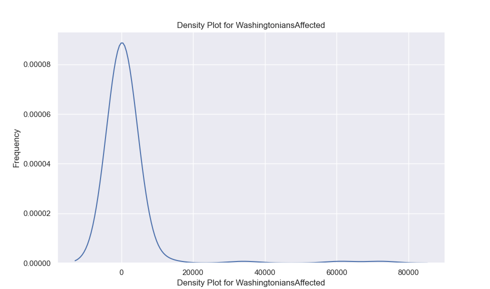
Scatter Plots and Density Analysis using Python Coding

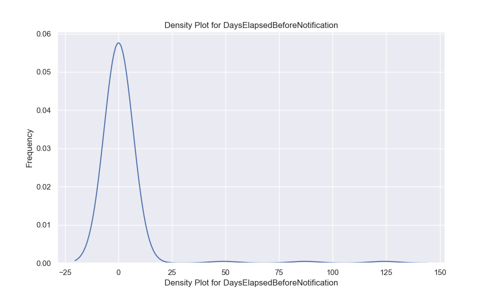
**Distributions and Analytica Methods**











The X-axis (Days Elapsed Before Notification): This represents the number of days taken to notify residents after the breach.

Y-axis (Frequency): This indicates how often different values of days elapsed occur in the dataset. A higher density means that more breaches had a similar number of days before notification.

The X-axis (Days to identify breach): This represents the number of days taken to notify residents after the breach.

Y-axis (Frequency): This indicates how often different values of days elapsed occur in the dataset. A higher density means that more breaches had a similar number of days to identify breach.

**Limitations and Future Research**

Like other research studies, our findings have limitations. This dataset even though it includes key data composed of 1,209 rows and 25 columns it does not include other variables that can help build a predictable model. We will approach the Washington State department and the office of the Attorney General to request permission to work with them and obtain additional variables to complete a Research Paper. The plan is to contribute to the value of knowledge in the field of cybersecurity to protect consumers, local businesses from future data breaches.

**Conclusion**

In conclusion there is no correlation, within the variables analyzed in Washington, despite public concern regarding the reputational and financial consequences of data breaches, no statistically significant correlation was identified between data breaches and changes in the revenue of the companies involved. Factors such as the nature of the breach, customer loyalty, and other initiatives may drive the potential economic impact on company performance. Future research will be focused on exploring additional variables that can provide significant relationship and predictable model.

**APPENDIX A**

**Table 1.**

*Descriptive Statistics.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Mean | Std | Min | Q1 | Median | Q3 | Max |
| Washingtonians Affected | 12513.968919 | 2643.085816 | 9561.000000 |  |  |  |  |
| Days to Identify Breach | 116.784810 | 218.625351 | -503.000000 |  |  |  |  |
| Days Elapsed Before Noti. | 723.000000 | 102.109736 | 0.000000 |  |  |  |  |

**Table 2.**

*Correlations.*

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | (1) | (2) | (3) |
| Washingtonians Affected | 1.000000 | -0.025976 | -0.60966 |
| Days to Identify Breach | -0.025976 | 1.000000 | -0.079481 |
| Days Elapsed Before Noti. | -0.060966 | -0.079481 | 1.000000 |
|  |  |  |  |
|  |  |  |  |

**Table 3.**

*Distributions of the variables.*

|  |  |
| --- | --- |
| Days to Identify Breach | Long trails shift toward the right- Indicating that it takes more time to identify the breach. |
| Days to Elapsed Before Notification | Trails are shifting towards the right- Indicating significant delays in the notification process. |
|  |  |

**APPENDIX B**

**Attorney General Law Provision to enhanced data security**

Attorney General calls for enhanced data security law to protect consumers as part of 2015 legislative agenda

**FOR IMMEDIATE RELEASE:**

**OLYMPIA —** Washington State Attorney General Bob Ferguson has introduced bipartisan legislation that will strengthen Washington’s data breach notification law to help Washingtonians protect their personal information.

The Attorney General request legislation [Senate Bill 5047](http://app.leg.wa.gov/billinfo/summary.aspx?bill=5047&year=2015)is sponsored by Sen. John Braun, R—Centralia. The companion bill, [House Bill 1078](http://app.leg.wa.gov/billinfo/summary.aspx?bill=1078&year=2015), is sponsored by Rep. Zack Hudgins, D—Tukwila.

**Reference**

Annotated Bibliography:

Aiello, S. (2024). Privacy Principles and Harms: Balancing Protection and Innovation. *Journal of Cybersecurity Education, Research and Practice*, *2024*(1). <https://doi.org/10.62915/2472-2707.1167>

This paper delves into the fundamental principles of privacy and explores the potential harm that can arise from the mishandling of personal data. It emphasizes the delicate balance between safeguarding individuals' privacy rights and fostering innovation in a data-driven society. By analyzing key privacy principles and their implications, this paper explores the foundational privacy principles that define the concept of privacy while delving into the potential harms that can arise when these principles are violated.

Alkhazi, B., Alshaikh, M., Alkhezi, S., & Labbaci, H. (2022). Assessment of the Impact of Information Security Awareness Training Methods on Knowledge, Attitude, and Behavior. *IEEE Access*, *10*, 132132–132143. <https://doi.org/10.1109/ACCESS.2022.3230286>

This study suggests that all methods raise knowledge equivalently. However, having more than one delivery method to convey the same message has a greater impact on users’ attitudes. When it comes to behavioral change, however, text-based and game-based training formats performed better than their counterparts.

Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. *Electronics*, *12*(6), 1333. <https://doi.org/10.3390/electronics12061333>

The paper first extensively explains the main reasons for cyber-attacks. Then, it reviews the most recent attacks, attack patterns, and detection techniques. Thirdly, the article discusses contemporary technical and nontechnical solutions for recognizing attacks in advance.

Ben Salamah, F., Palomino, M. A., Craven, M. J., Papadaki, M., & Furnell, S. (2023). An Adaptive Cybersecurity Training Framework for the Education of Social Media Users at Work. *Applied Sciences*, *13*(17), 9595. <https://doi.org/10.3390/app13179595>

Bergadano, F., & Giacinto, G. (Eds.). (2023). *AI for Cybersecurity: Robust models for Authentication, Threat and Anomaly Detection*. MDPI - Multidisciplinary Digital Publishing Institute.

*Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security Danda B. Rawat , Senior Member, IEEE, Ronald Doku, and Moses Garuba Abstract— ‘‘Knowledge is power” is an old adage that has been found to be true in today’s information age. Knowledge is derived*. (n.d.).

A technology being applied in various field. AI is being used in text, image, video, audio and code.

Dominik Molitor, Wullianallur Raghupathi, Aditya Saharia and Viju Rahgupathi. (n.d.). *Exploring Key Issues in Cybersecurity Data Breaches: Analyzing Data Breach Litigation with ML-Based Text Analytics.*

This study serves as a foundational piece for comprehensively analyzing large textual datasets. The findings hold significant implications for both researchers and practitioners in cybersecurity, especially, those grapping with the challenges of data breaches.

El Amin, H., Samhat, A. E., Chamoun, M., Oueidat, L., & Feghali, A. (2024). An Integrated Approach to Cyber Risk Management with Cyber Threat Intelligence Framework to Secure Critical Infrastructure. *Journal of Cybersecurity and Privacy*, *4*(2), 357–381. <https://doi.org/10.3390/jcp4020018>

Such information is the product of a cyber threat intelligence process that proactively delivers knowledge about cyber threats to inform decision-making and strengthen defenses. In this paper, we overview risk management and threat intelligence frameworks.

Gatzlaff, K. M., & McCullough, K. A. (2010). The Effect of Data Breaches on Shareholder Wealth. *Risk Management and Insurance Review*, *13*(1), 61–83. <https://doi.org/10.1111/j.1540-6296.2010.01178.x>

This article examines the stock market's assessment of the cost of data breaches at publicly traded companies in which personal information such as customer and/or employee data are exposed.

Gwebu, K. L., Wang, J., & Wang, L. (2018). The Role of Corporate Reputation and Crisis Response Strategies in Data Breach Management. *Journal of Management Information Systems*, *35*(2), 683–714. <https://doi.org/10.1080/07421222.2018.1451962>

Drawing on cognitive dissonance theory and the research on cue diagnostic and crisis management, this study examines the relative efficacy of firm reputation and a range of post-breach response strategies.

Horner, C. (Ed.). (2020). *Cybersecurity in Business Valuation: Addressing the Impact of Data Breaches on Value (a BVR Briefing)*. Business Valuation Resources, LLC.

Li, T., Deng, W., & Wu, J. (Eds.). (2023). *Advanced Machine Learning Applications in Big Data Analytics*. MDPI. <https://doi.org/10.3390/books978-3-0365-8487-4>

Cybersecurity in Business Valuation: Addressing the Impact of Data Breaches on Value (a BVR Briefing)

Markos, E., Peña, P., Labrecque, L. I., & Swani, K. (2023). Are data breaches the new norm? Exploring data breach trends, consumer sentiment, and responses to security invasions. *Journal of Consumer Affairs*, *57*(3), 1089–1119. <https://doi.org/10.1111/joca.12554>

This study show that data breaches are frequent, vary across industries, and consumer attitudes and actions vary by data type compromised.

Ng, K. C., Zhang, X., Thong, J. Y. L., & Tam, K. Y. (2021). Protecting Against Threats to Information Security: An Attitudinal Ambivalence Perspective. *Journal of Management Information Systems*, *38*(3), 732–764. <https://doi.org/10.1080/07421222.2021.1962601>

From a theoretical standpoint, our work contributes to the information security literature by incorporating attitudinal ambivalence, which arises from the intrapersonal and interpersonal appraisal processes, into PMT.

Pricop, E., Stamatescu, G., & Fattahi, J. (Eds.). (2021). *Advanced Topics in Systems Safety and Security*. MDPI - Multidisciplinary Digital Publishing Institute.

Advanced in tech Systems Safety and Security to leverage data protection of individuals.

Saeed, S., Suayyid, S. A., Al-Ghamdi, M. S., Al-Muhaisen, H., & Almuhaideb, A. M. (2023). A Systematic Literature Review on Cyber Threat Intelligence for Organizational Cybersecurity Resilience. *Sensors*, *23*(16), 7273. <https://doi.org/10.3390/s23167273>

knowledge with us and providing insights that can be drawn upon by a wide audience.

Scheibmeir, J. A., & Malaiya, Y. K. (2021). Social media analytics of the Internet of Things. *Discover Internet of Things*, *1*(1), 16. <https://doi.org/10.1007/s43926-021-00016-5>

Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A Systematic Review on Deep-Learning-Based Phishing Email Detection. *Electronics*, *12*(21), 4545. <https://doi.org/10.3390/electronics12214545>

The review identifies gaps in the literature and informs the development of future research questions and areas of focus. With the increasing sophistication of phishing attacks, applying deep learning in this area is a critical and rapidly evolving field. This systematic literature review aims to provide insights into the current state of research and identify areas for future research to advance the field of phishing detection using deep learning.

Trim, P. R. J., & Lee, Y.-I. (2024). Advances in Cybersecurity: Challenges and Solutions. *Applied Sciences*, *14*(10), 4300. <https://doi.org/10.3390/app14104300>

Tuckfield, B. (2023). *Dive into data science: Use Python to tackle your toughest business challenges*. No Starch Press.

The research findings encourage the cooperative spirit of the researchers.

Wang, Q., Ngai, E. W. T., Pienta, D., & Thatcher, J. B. (2023). Information Technology Innovativeness and Data-Breach Risk: A Longitudinal Study. *Journal of Management Information Systems*, *40*(4), 1139–1170. <https://doi.org/10.1080/07421222.2023.2267319>

This study shows the importance of understanding organizational learning in risk assessment and change management, as well as the critical role of contextual factors in moderating the unintended security related consequence liked to IT innovations.