Final Project: First Deliverable

Saint Louis University

Professor Tim Hough

AA-5200 Information Visualization & Presentation

Group-07

Harika Aakula

Keerthi Madhuri Vaddepalli

Tejaswini Bollikonda

Rajesh Bandari

September 07, 2025

Introduction:

In today's digital society, data breaches have become a dangerous threat that impacts individuals, organisations, governments, and businesses. Breaches have been growing worldwide, driven by ransomware, stolen credentials, and third-party vulnerabilities. As of 2025, ransomware is present in 44% of breaches affecting small and medium-sized businesses (Hylender et al., 2025).

In this project, our team has chosen to explore the "**Data Breach Notifications Affecting Washington Residents (Personal Information Breakdown)"** dataset. This paper aims to provide a clear vision of why we chose this dataset, what goal we will address, possible research questions, describe the dataset, identify the prospective audience with use cases, and suggest visuals, followed by a final reflection.

Goal of the Project:

Practical objectives provide quantitative support for management decisions (Doran, 1981). This project focuses on the evolving issue of data breaches and their impact on industries in Washington State. Its primary goal is to analyse breach causes, affected industries, the number of residents impacted, and breach lifecycles to equip the Washington Attorney General's Office with actionable insights for consumer protection and cyber risk management. Following Doran's emphasis on the SMART model, goals must be clear and time-specific to ensure research leads to actionable outcomes (Doran,1981), the goal is specific to Washington breaches, measurable through key

variables like frequency and scale, achievable with both quantitative and qualitative data, relevant due to the social and economic impact of breaches (Fotis, 2024), and time-bound, covering 2014–2026, to identify long-term trends.

Research Questions:

Below are the research questions that add value if analysed effectively.

1. How did the number of incidents and impacted Washingtonians affect the frequency and scale of data breaches from 2014 to 2026?
2. What is the relationship between the industry impacted by a breach and its cause of breach? Which industries are the most susceptible to a particular kind of attack?
3. Which categories of personal data are most compromised, and how can they be related to the industry-specific business type and the cause of the breach?
4. Is there a relationship between the breach lifecycle, the number of affected individuals, or the type of data compromised? Are specific industries or breach causes associated with faster or slower response times?
5. Which specific cyberattack is the most common cause of breach, and has the prevalence of these specific types changed over the years?

Dataset Description & Variables:

The dataset titled “Data Breach Notifications Affecting Washington Residents (Personal Information Breakdown)” was created and provided by the Washington State Attorney General’s office Consumer Protection Division to notify the Attorney

General's office when more than 500 Washingtonians' personal information was compromised due to the breach. The dataset was created on June 1, 2021, with the latest updated date being September 6, 2025, and it covers data breach incidents reported from September 2014 to the present. The dataset comprises 6386 rows and 16 columns (Data Breach Notifications Affecting Washington Residents (Personal Information Breakdown) | Data. WA | State of Washington, 2025).

Table 1

Displaying dataset variables with detailed descriptions

Variable Name	**Type of the Variable**	**Range of Values**	**Explain why you are considering including in your analysis**
DataBreach Cause	Categorical	Theft or Mistake, Cyberattack, Unauthorised access	This variable is important to understand the primary reason of the data breach
Washingtonians Affected	Numerical	15 to 328889	This variable is crucial in determining the scale and severity of data breach
Industry Type	Categorical	Business, Finance, Government, Health, Education, Non-Profit/Charity	This Variable helps categorise the type of organisations experienced data breach

Information Type	Categorical	Biometric Data, Driver's License or Washington ID Card Number, Social security number etc.	It is essential for analysing types of personal information compromised
Year	Numerical (Ordinal variable)	2016 to 2026	It is important variable as it helps to perform time series analysis of data breach over specific period
BreachLifecycleRange	Categorical-Range type	1-99 days, 100-199 days, 365+ days etc	It is useful in understanding the duration of data breach from the data breach has been identified.
Name	Categorical	Delta airlines, H&R block, Microsoft etc.	Useful in identifying the organisation affected by breach.
Cyber attack type	Categorical	Malware, Phishing, Ransom, Unclear/unknown, Others	Provide a deeper granular view of cyberattacks to pinpoint specific threat which allow us to identify the recommendations.

Business Type	Categorical	Software, Retail, transport, Home, hospitality etc	More specific category of industry type can be useful to identify trends within broad sector.

Prospective audience, Use Cases & Suggested visuals:

The insights gained from this dataset can benefit various stakeholders, including state policy makers and regulators on cybersecurity incidents, such as members of the Washington Attorney General's office. This project can also be valuable to business executives and risk managers, mainly in industries vulnerable to data breaches, such as healthcare, finance and retail. Cybersecurity professionals, including CISOs, analysts, incident response teams, SOC teams, and consumer advocates, can also consider this dataset to derive insights into privacy rights and consumer protection.

The analysis has various use cases, as the CISO can use the report to identify the industry's most commonly occurring cyberattack types. The Verizon report highlights the detailed analysis of incident classification patterns, such as System intrusion and social engineering (Hylender et al., 2025). Similarly, a CISO can find the trends that allow them to justify the budget for specific defences such as ransomware countermeasures. Another use case is that policymakers can use it to identify specific vulnerabilities that require legislative action and introduce new policies within Washington State. For instance, if data highlights that breaches are caused mainly by phishing attacks, legislators can introduce new policies or advocate for public awareness

campaigns. Supporting this, the Verizon report highlights social engineering attacks such as phishing, which have increased as a percentage of overall breaches (Hylender et al., 2025).

Various visuals can be used to communicate findings effectively. Line graphs to display breach frequency and scale over time, while a stacked bar chart illustrates causes of breach across industries, business types and years. Heat maps highlight the relationship between industries, cyberattack type, breach causes, and affected residents. A clustered column/bar can display information types and companies with their frequency, and a boxplot depicts variability in breach lifecycles. These strategies help make complex data accessible for decision makers (Hylender et al., 2025).

Reflection:

In conclusion, this project provides valuable insights into data breaches affecting Washington residents, including trends in breach frequency, causes, affected industries and breach lifecycle. The insights gained can guide policymakers in creating targeted regulations, help organisations allocate resources effectively for cybersecurity, and help CISOs improve strategies. Further, with a complex data interpretation through visuals, the project enhances decision-making while strengthening consumer protection and providing long-term resilience against evolving cyber threats.

References:

Hylender, C. D., Langlois, P., Pinto, A., Widup, S., Verizon DBIR team, Verizon Threat Research Advisory Centre (VTRAC) team, & U.S. Secret Service. (2025). Verizon 2025 Data Breach Investigations Report. https://www.verizon.com/business/resources/T21f/reports/2025-dbir-data-breach-investigations-report.pdf

Doran, G. (1981). There is a S.M.A.R.T. way to write Management's goals and objectives. Management Review, 70(11), p. 35. Permalink: https://ezp.slu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=6043491&site=eds-live

Fotis, F. (2024). Economic Impact of Cyber Attacks and Effective Cyber Risk Management Strategies: A light literature review and case study analysis. Procedia Computer Science, 251, 471–478. https://doi.org/10.1016/j.procs.2024.11.135

Data breach Notifications affecting Washington residents (Personal Information Breakdown) | Data.WA | State of Washington. (2025, September 7). https://data.wa.gov/Consumer-Protection/Data-Breach-Notifications-Affecting-Washington-Res/padd-mby7/about_data