Data Breach Affecting Washington State Residents

Research Design

Introduction

Over the past 25 years the internet created new ways of doing business which has brought increase in data across various applications. Digital Marketing has been a main driver to increase the rate of consumer data collection. These initiatives have generated significant online activities and transactions between consumers and businesses. Consumers are actively participating on social media, health care portals to schedule online doctors' appointments, requesting direct and indirect loans to meet their financial needs. The current economic environment and the usage of IoT internet of things has raised concerns as the shared data becomes accessible to other devices and companies that represent a risk for potential data breaches (Aguirre et al., 2015; Markos et., 2018). Private information stored in company data bases is vulnerable to hackers. (Chang, 2018; Privacy Rights Clearinghouse, 2019). We are now part of the big data era as the numbers continue to increase exponentially creating an alarming rate. The growth of data is now moving from terabytes to petabytes. (Laney 2001) implemented the concept of the V's which is directly associated with big data. The terms are known for Volume, Velocity and Variety. There is a fourth V which is veracity. The purpose of our research is threefold and explores the following:

1- The number of residents in Washington State impacted by the Data Breach. 2- Days to identify the breach 3- Days elapsed before notification. To achieve this, we utilize the datasets regarding data breach from 2015 to present published in the Washington state portal on September 2024 (data.wa.gov) This data was provided by the Washington State Attorney General's Office Protection Division.

Literature Review

Nearly half breaches involved customer personal identifiable information (PII), which can include tax identification (ID) numbers, emails, phone numbers and home addresses. According to previous research the cost-benefit analysis is not always accurate, because consumers cannot determine how their privacy may be impacted when shared with a third party (Jozani et al., 2020). Research does consistently find that online privacy violations lead to decreased trust in a company or website (Bougoure et al., 2016; Martin, 2018), and trustworthiness and perceived information safety determine a consumer’s willingness to share personal information with a business (Rainie & Maeve, 2016; Whitler & Farris, 2017). It is required for Consumers to make online purchases, interact on social media platforms, or connect with health care providers, information like name, date of birth, and health background information to complete the process. The growth that e-commerce and the smartphone has generated concerns about online data security as breaches continue to rise. E-Commerce expansion surged due to social distancing measures changing consumer to spend more time online. (IBISWorld, August 2024). Consumers feel vulnerable when a social contract is disrupted by a data breach, resulting in the loss in personal information. (Martin et al., 2017). My research is based on the Data Breach notifications affecting Washington Residents. The goal is to assess the impact of breaches on the number of residents affected in Washington State. Washington 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.

Hypotheses

H1. The data lost in a breach will affect consumers' satisfaction.

H2. The days to identify breach will affect revenue

H3. The days elapsed before notification will affect revenue

H4. The number of Washingtonians will be financially impacted

Research Questions:

1- What is the average number of Washingtonians affected by data breaches during the past 7 years?

2- What are the average days that it takes to identify the breach?

3- What are the average days that it takes to notify Washingtonians about the data breach?

4- How do the data breaches impact Washingtonians Financially?

Dependent Variable (DV):

The main focus of my research is on data breaches affecting Washington residents as the dependable variable. Through the research I 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 impact of data 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 2. Model Variable illustrates the description).

Figure 1 illustrates the theoretical model and Research Design

Methodology

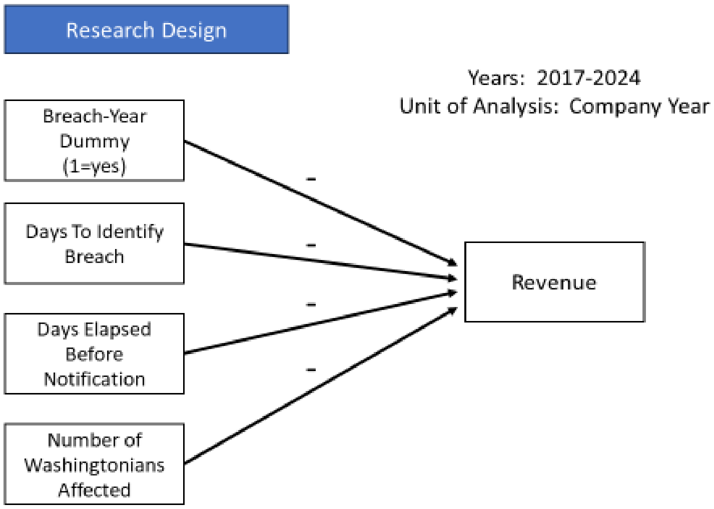

Research Design
Years: 2017-2024
Unit of Analysis: Company Year
Breach-Year Dummy (1=yes)
Days To Identify Breach
Days Elapsed Before Notification
Number of Washingtonians Affected
-
-
-
-
Revenue

Fig. 1. **Model Variable**

Figure 2 illustrates the theoretical model and Research Design

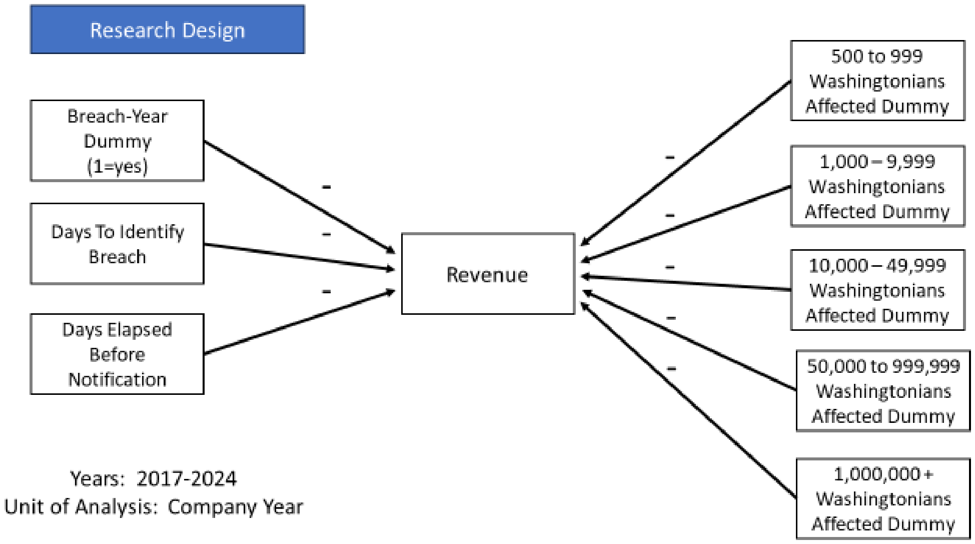

Research Design
Breach-Year Dummy (1=yes)
Days To Identify Breach
Days Elapsed Before Notification
-
-
-
Revenue
-
-
-
-
-
500 to 999 Washingtonians Affected Dummy
1,000 – 9,999 Washingtonians Affected Dummy
10,000 – 49,999 Washingtonians Affected Dummy
50,000 to 999,999 Washingtonians Affected Dummy
1,000,000 + Washingtonians Affected Dummy
Years: 2017-2024
Unit of Analysis: Company Year

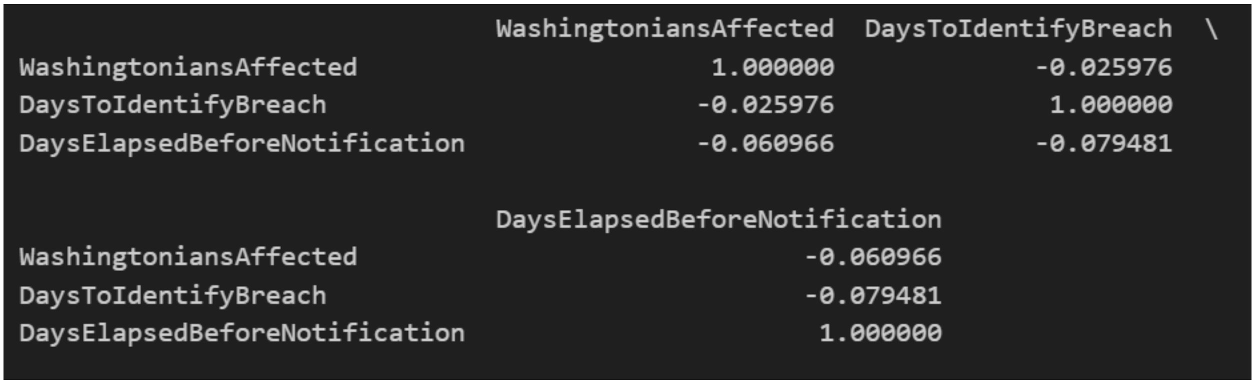
Fig. 2. Model Variable

Statistical Tables:

Descriptive statistics

We turn to the Washington Residents affected by the data breach.

Correlations.

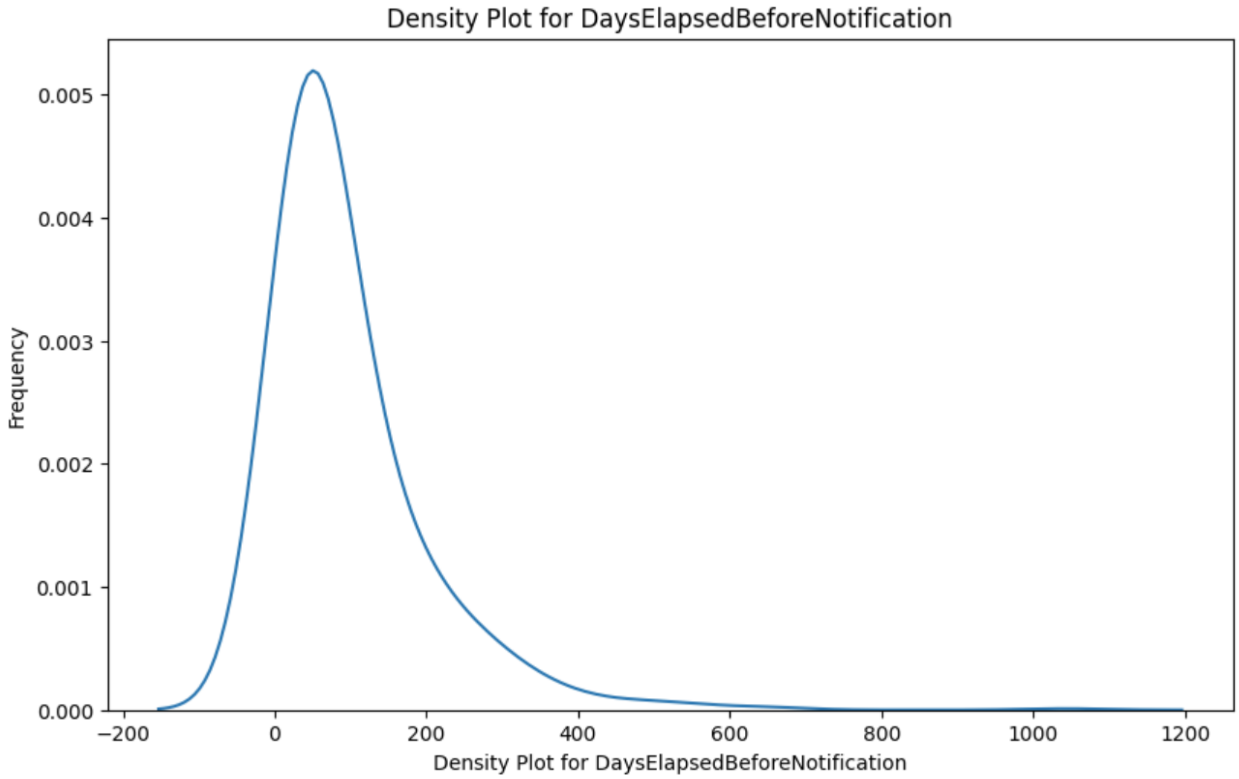


```
                              WashingtoniansAffected  DaysToIdentifyBreach  \
WashingtoniansAffected                      1.000000             -0.025976
DaysToIdentifyBreach                       -0.025976              1.000000
DaysElapsedBeforeNotification              -0.060966             -0.079481

                              DaysElapsedBeforeNotification
WashingtoniansAffected                            -0.060966
DaysToIdentifyBreach                              -0.079481
DaysElapsedBeforeNotification                      1.000000
```

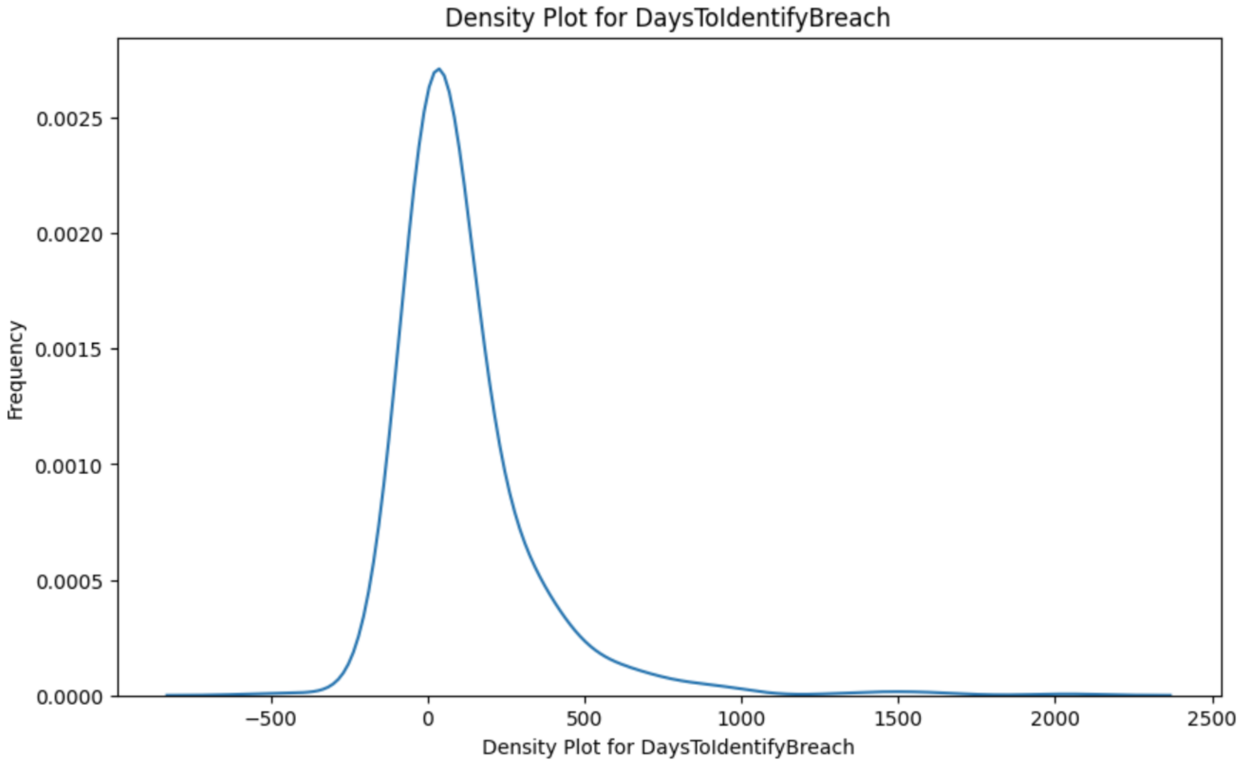
There is a strong positive correlation coefficient on more individuals affected which corresponds to more days to identify. There is a negative correlation (more affected individuals to days to identify and days Elapsed before notification.

Distributions.



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

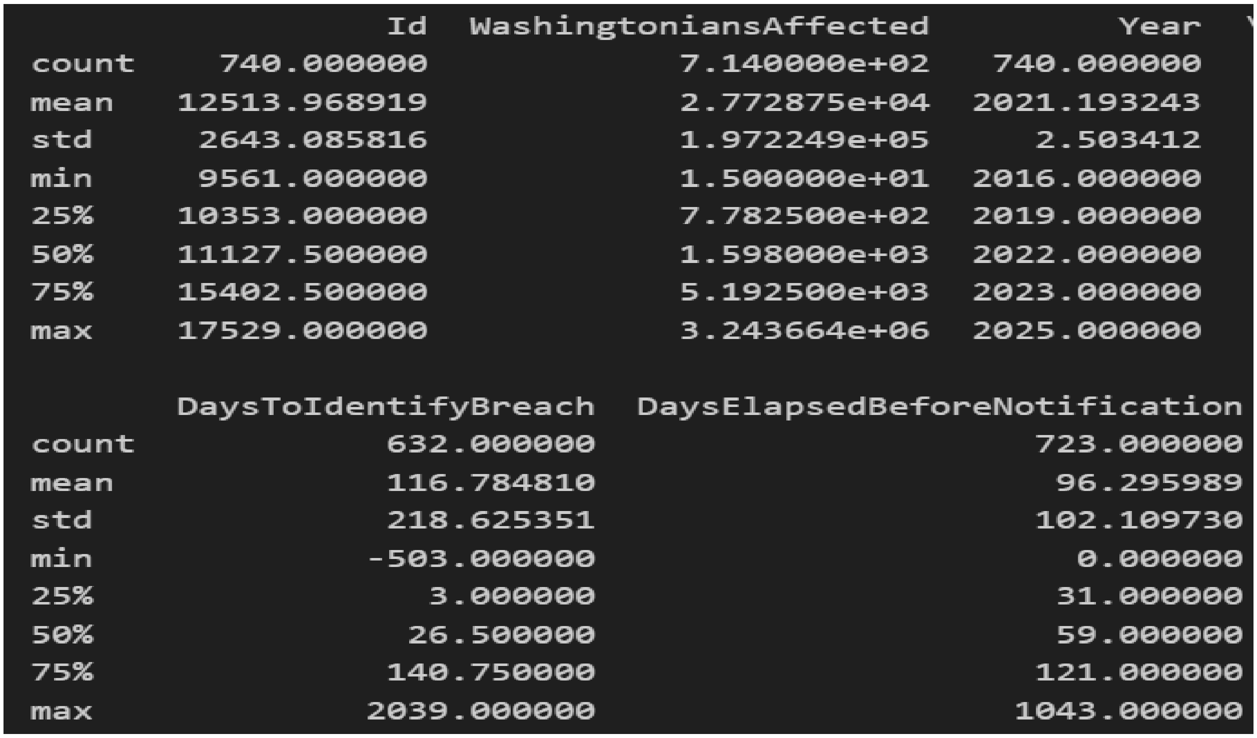
Y-axis (Frequency): This indicates how often different values of days elapsed occur in the dataset. A higher density means that more breaches had a similar number of days before notification.



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

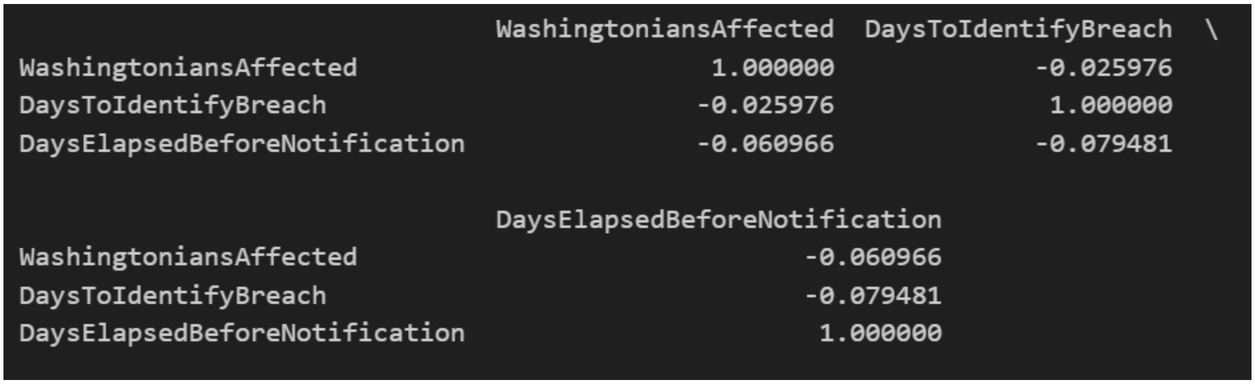
Y-axis (Frequency): This indicates how often different values of days elapsed occur in the dataset. A higher density means that more breaches had a similar number of days to identify breach.

Analytical Methods



```
                 Id  WashingtoniansAffected         Year
count    740.000000            7.140000e+02   740.000000
mean   12513.968919            2.772875e+04  2021.193243
std     2643.085816            1.972249e+05     2.503412
min     9561.000000            1.500000e+01  2016.000000
25%    10353.000000            7.782500e+02  2019.000000
50%    11127.500000            1.598000e+03  2022.000000
75%    15402.500000            5.192500e+03  2023.000000
max    17529.000000            3.243664e+06  2025.000000

       DaysToIdentifyBreach  DaysElapsedBeforeNotification
count            632.000000                     723.000000
mean             116.784810                      96.295989
std              218.625351                     102.109730
min             -503.000000                       0.000000
25%                3.000000                      31.000000
50%               26.500000                      59.000000
75%              140.750000                     121.000000
max             2039.000000                    1043.000000
```



```
                               WashingtoniansAffected  DaysToIdentifyBreach  \
WashingtoniansAffected                       1.000000             -0.025976
DaysToIdentifyBreach                        -0.025976              1.000000
DaysElapsedBeforeNotification               -0.060966             -0.079481

                               DaysElapsedBeforeNotification
WashingtoniansAffected                             -0.060966
DaysToIdentifyBreach                               -0.079481
DaysElapsedBeforeNotification                       1.000000
```

APPENDIX

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

Figure 1.

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.