GENBA 894

Project Proposal

Andrew Egelhof

KSU Summer 2018

I will be doing my project on project 2B, “Investigating the History of Data Breaches in U.S. using Data Science Techniques”. I will be cleaning, transforming, analyzing and interpreting the provided data set.

**Business Problem:** Businesses and organizations face daily threats of loss of data. Whether from malware, hacking, unintended disclosure or insider beaches. Organizations need to understand what the threats are, what the common methods of loss are and how they can identify potential losses and what they can to prevent them.

**Clients**: This analysis should server to cover a broad range of customers. Any commercial company, government or non profit should be able to gain some information from this report.

Some of the business questions I will address:

* Information about what types of breaches are most popular.
* What are the main causes of the breaches.
* What types of organizations are most likely to have a data breach.
* Information about data breaches by state.
* What organizations are more vulnerable to what type of breaches
* How breaches change over time
* What organizations experience repeated incidents.

**Expected Benefits**: Once an organization understands the threats they face, they can better prepare for and protect themselves from them. The information provided in this analysis should help prepare to defend themselves against any data breach or loss.

**Data Set**: The main dataset is provided by PrivacyRights.org at <https://www.privacyrights.org/data-breaches> . A CSV file. I will investigate gathering additional information from other sources.