## 

## 

# Data Visualization on Data Breaches in India

Aman Patel

Department of Applied Data Science, San Jose State University

DATA 230: Data Visualization

Prof. Andrew H. Bond

December 11, 2022

## 

**Abstract**

Data breaches, which happen daily in too many places, result in the loss of sensitive, private, and crucial personal, financial, and health information. When weighing other, more significant concerns like customer confidence, social trust, and personal safety in addition to the potential financial penalty, the cost of the data breach is considered. In this essay, we examine the greatest data breaches in history, those that occurred between 2004 and 2017 and caused the loss of more than 30,000 records. To reveal the truth and the hidden aspects of the data breach event, the data visualization technique is used from a number of angles. Based on the case study, we once again demonstrate how the true risks of a data breach are out of control. Based on the analysis and visualization, we learn a number of fascinating facts that very few academics have previously focused on, and it seems that the true repercussions of data breach are much more severe than people know.

Table of Contents

[**Data Visualization on Dat**](#_heading=h.1v6155pxhu4h)**a Breaches in India**

[1. INTRODUCTION](#_heading=h.gjdgxs)

[i. Project Summary:](#_heading=h.30j0zll)

[ii. Purpose:](#_heading=h.1fob9te)

[iii. Intended Audience](#_heading=h.2et92p0)

[2. Project Background](#_heading=h.2s8eyo1)

[i. Project Planning and Scheduling:](#_heading=h.3rdcrjn)

[ii. Project Development Approach](#_heading=h.35nkun2)

[a) Tableau](#_heading=h.1ksv4uv)

[b) Tableau Features](#_heading=h.fqzzkudr8873)

[**3. DATASET**](#_heading=h.4ho28s6yl65)

[i. Dat](#_heading=h.4i7ojhp)a Breaches Dataset

[4. DATA VISUALIZATION](#_heading=h.3as4poj)

[i. Process](#_heading=h.1pxezwc)

[a) Process to import dataset to tableau :](#_heading=h.49x2ik5)

[b) Creating of different interactive worksheets :](#_heading=h.43ogz1w0rhmx)

[c) Worksheet 1:](#_heading=h.4gg8s9l7djcv)

[d) Worksheet 2:](#_heading=h.1tzbd2kbsc7i)

[e) Worksheet 3:](#_heading=h.vf75cobuhtqo)

[f) Worksheet 4:](#_heading=h.et6ig3hjnyld)

[g) Worksheet 5:](#_heading=h.sh0rnuayuet7)

[h) Worksheet 6:](#_heading=h.9ujiij953j2n)

[ii. Dashboards](#_heading=h.1hmsyys)

[a. Dashboard 1:](#_heading=h.bb3ifybp6ggk)

[5. Conclusion](#_heading=h.43ky6rz)

[REFERENCE](#_heading=h.2iq8gzs)

# INTRODUCTION

## Project Summary:

* In this project, there are several visualizations made with the help of Data breach dataset, about the cyber crimes taking place in India between year 2004 to 2017.
* This project focuses on different visualizations like which were the methods used to breach data, which sector was the main target of cyber crimes and, in which year the greatest number of data breaches took place.
* Moreover, the visualizations also focused on was the method practiced the entity, which can be also defined as the main source of the data leak and most importantly what to leak records.

## Purpose:

* The main purpose to choose the topic as Data breach, was to spread awareness about the vastly spreading cybercrimes, and let people know how to prevent it by securing your data.
* To use any of the third-party sites, you must know its user policy, because nowadays web is the key sector, where the greatest number of data leaks is taking place. Data plays significant role in a person’s life or an institution or else a multi-national company.
* The major goal of this project is to examine this through various visualizations in order to highlight these visualizations for the benefit of citizens and the cybercrime branch in order to get the most out of preventing data breaches.

## Intended Audience

* Professor
* Project Supervisor
* Class Students
* Cyber Crime Investigation Team
* Targeted citizens

# Project Background

## Project Planning and Scheduling:

* In this project, Tableau was used to show data pertaining to data breaches that occurred over a number of years. This project focuses on various visualizations, such as the methods used to leak data, the industry that was the main target of cybercrime, the year that saw the most data breaches, the organization that can also be considered the primary source of the data leak, and most importantly, what records were leaked. The main objective of this project is to analyze this through various visualizations so that people may benefit from emphasizing these images to cybercrime team to investigate and prevent the data leak.

## Project Development Approach

### Tableau

* Unquestionably, data visualizations play a significant role in how we convey information about our data. It might be challenging to choose which tool to employ within your organization since the majority of teams now use data visualization tools to gain insight into their raw data. After eight years of working in the field of data visualization, I found myself more and more inclined to suggest Tableau to my clients who were looking for more effective methods to use their data. Tableau is also known for its speed, and the powerful integration technique as well as scalability.
* Tableau is a prominent data visualization application for corporate intelligence and data analysis. Tableau is an excellent tool for reporting and analyzing massive volumes of data in the business intelligence and data visualization fields. Tableau was established in the United States in 2003, and Salesforce acquired it in June 2019. It helps users create various graphs, maps, dashboards, and stories to visualize and analyze data to support business decision-making. Due to the abundance of intriguing and unique features it provides, Tableau is one of the most popular business intelligence applications (BI).

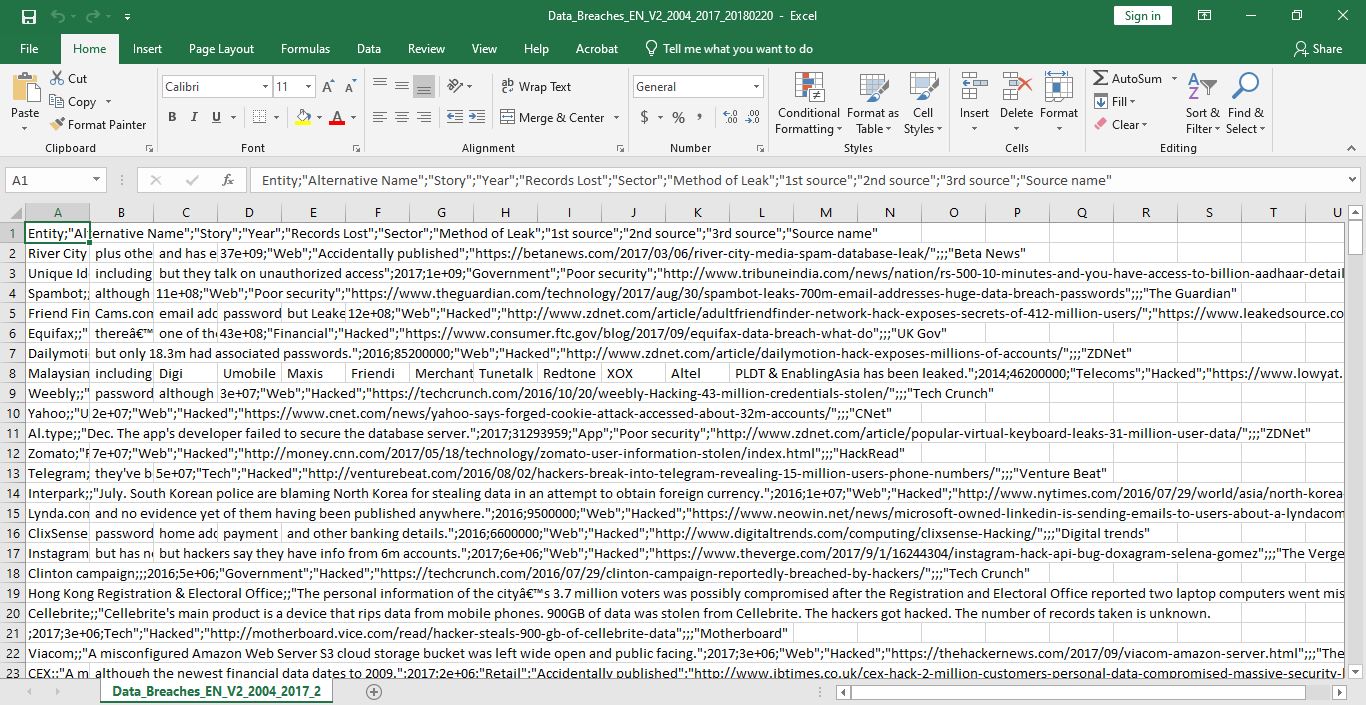
### Tableau Features

* Data and user security are given great consideration by Tableau. It has a complete security system that relies on authentication and permission systems for user access and data connections.
* Tableau guarantees connectivity to both real-time data sources and external data sources for in-memory data extraction. As a result, the user has complete freedom to use data from various sources without any limitations.
* Tableau Dashboards use text, graphic objects, visualizations, and many other tools to give you a complete picture of your data.
* Tableau offers simple ways for users to work together and rapidly share data in the form of visualizations, sheets, dashboards, etc. You can use it to safely transfer data from many different data sources, including hybrid, on-premise, and cloud.
* Tableau's Ask data function has increased its popularity among consumers worldwide. With this functionality, manipulating data is as easy as performing standard Google searches.

# DATASET

## Data Breaches dataset

* Data Breaches 2004-2017 dataset
* Source: Kaggle
* Dataset size: 97.94 MB
* Dataset Columns: 11
* Dataset Rows: 270



# DATA VISUALIZATION

## Process

### Process to import dataset to tableau:

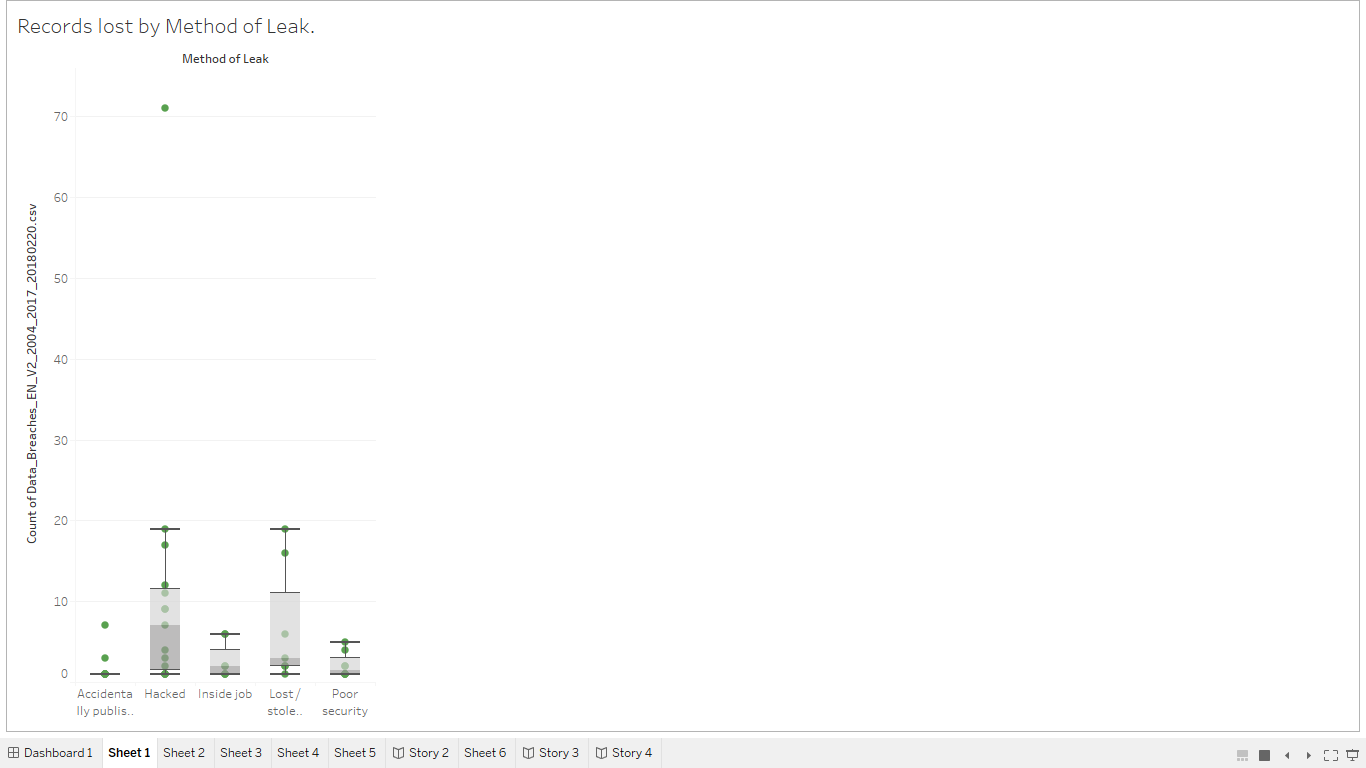
* As the first step we will locate Tableau Text File Connector.
* Then we will choose the Tableau CSV Connector.
* Moving to third step we will import the desired CSV File from our System.
* Lastly, will configure the Tableau CSV Text File Properties.

### Creating of different interactive worksheets:

* I have performed, total 6 of different visualizations which will be shown below in the form of worksheets, where the visualizations totally revolve around data breach activities taken place between year 2004 to 2017.
* There have been key visualizations, which were made regarding records, sectors affected, entity, methods of leak, and many more.

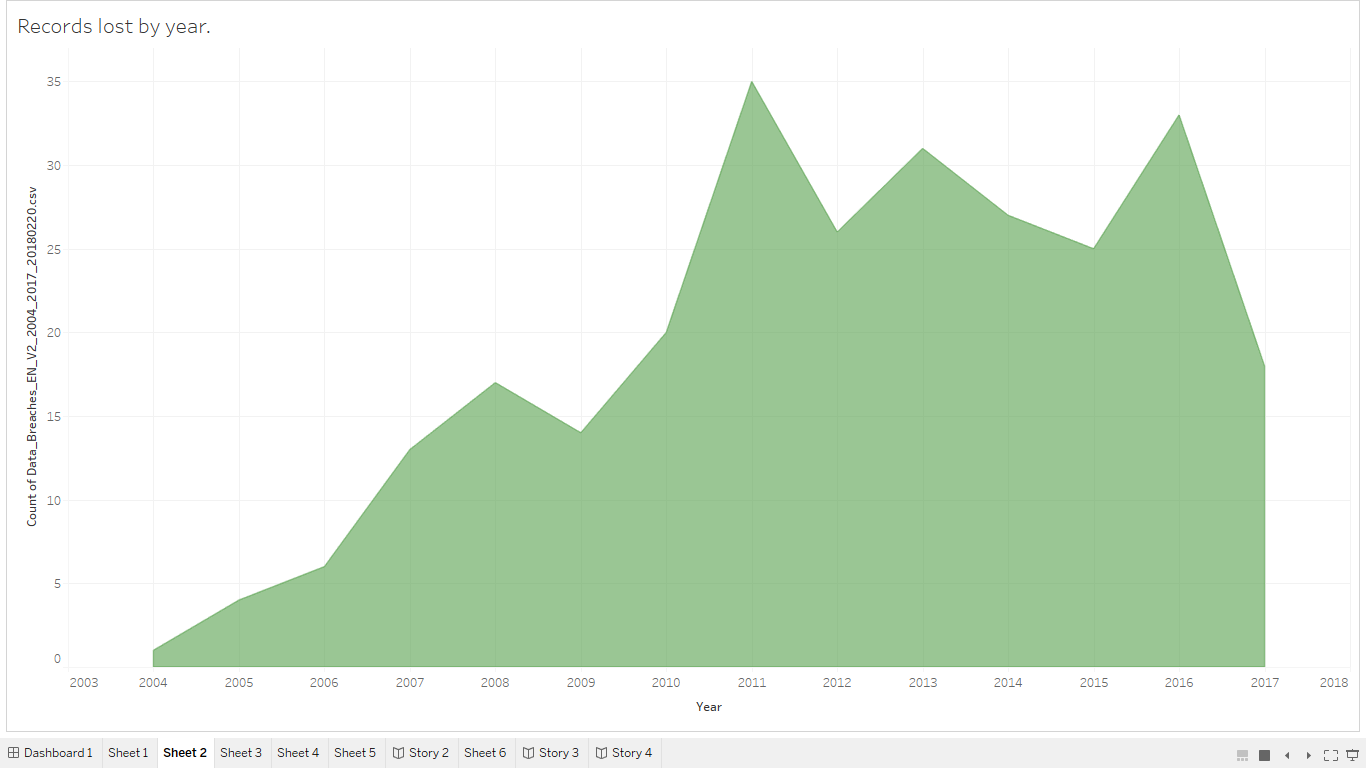
### Worksheet 1:

### The below shown visualization depicts the major methods which were practiced to breach data between year 2004 to 2017, where hacking dominated with most number of count, followed by the lost or stolen devices.



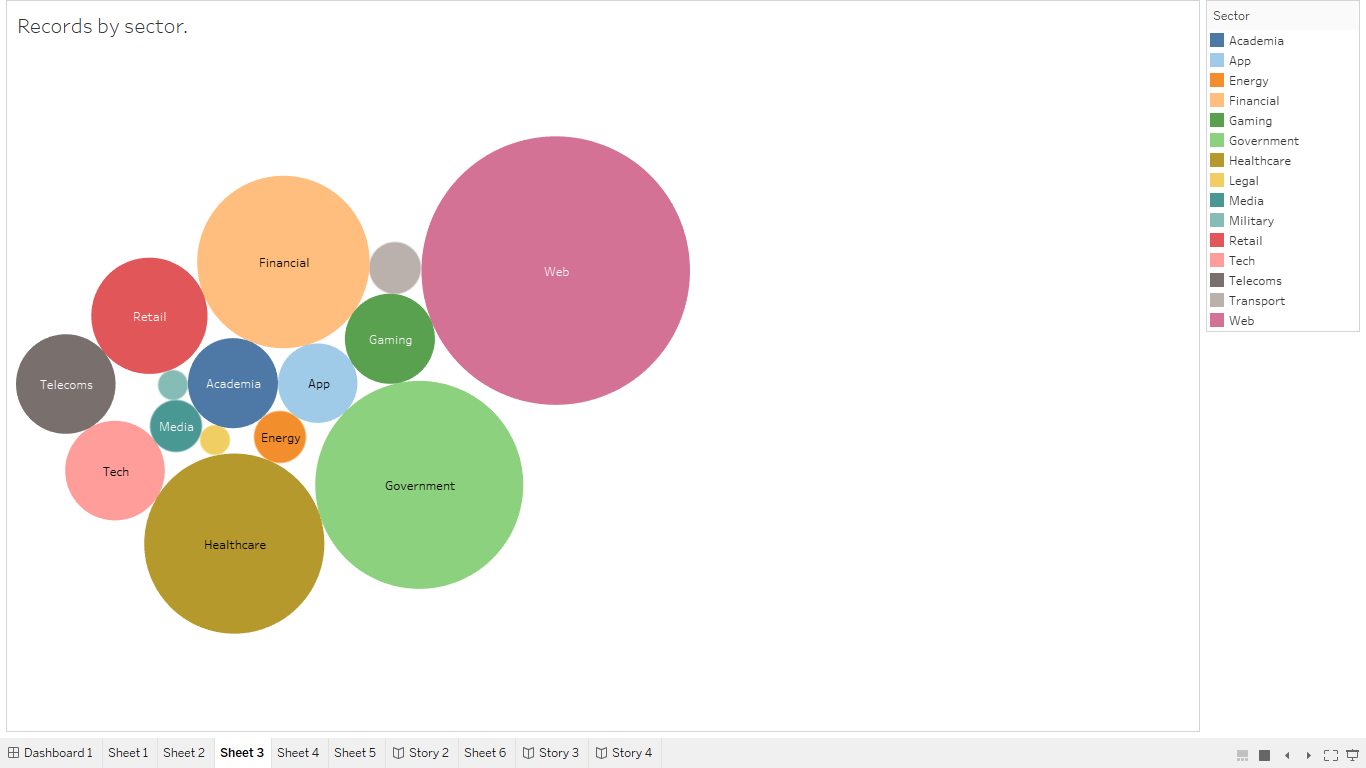
### Worksheet 2:

* The visualization shows, which of the following year faced the greatest number of cybercrimes from 2004-2017. The count touched to 36 between year 2010 and 2011, which then was followed by 33 in year 2017. Thus, this shows the steady increase in data breaches after year 2010, which are not a good sign, as we need to take precautions to keep our data safe.



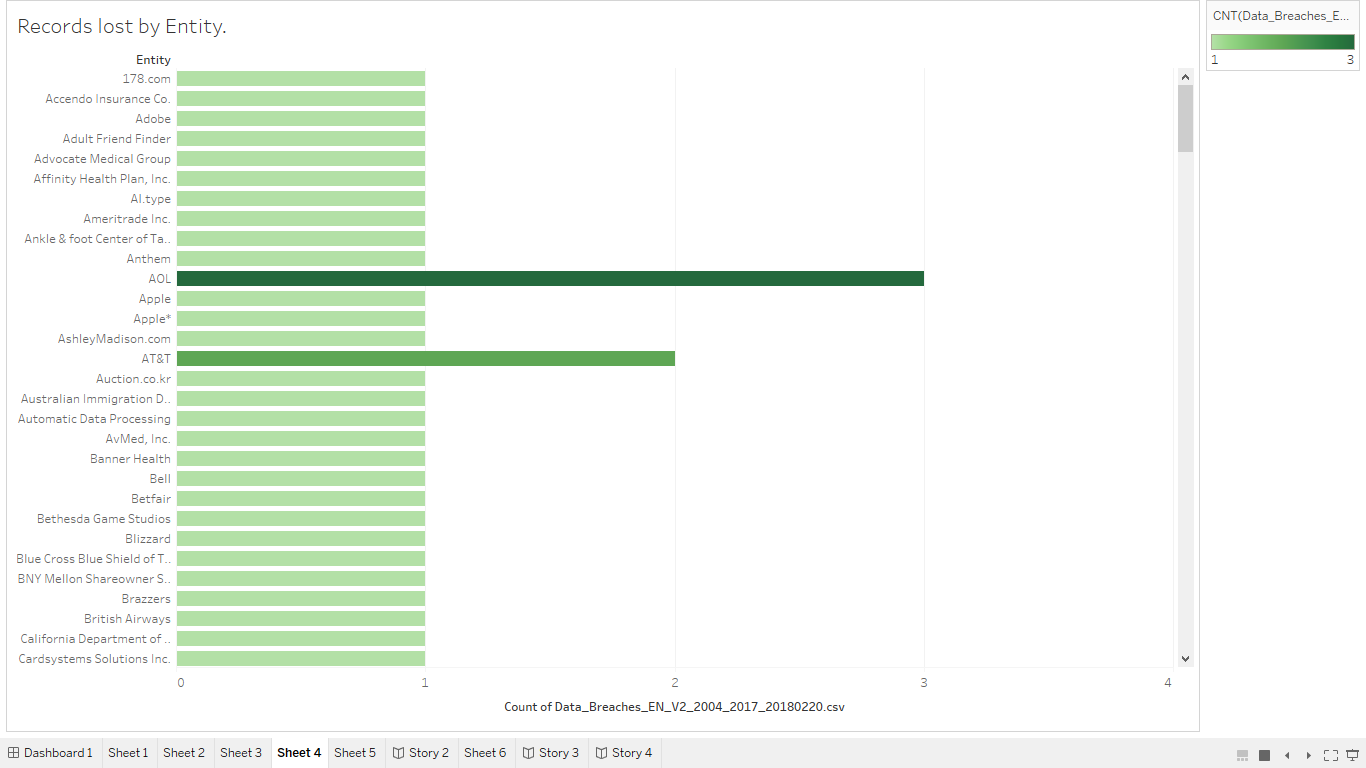
### Worksheet 3:

* This visualization performed below, shows the sectors which lead to a data leaks, where web was the leading sector, followed by government and finance sector.



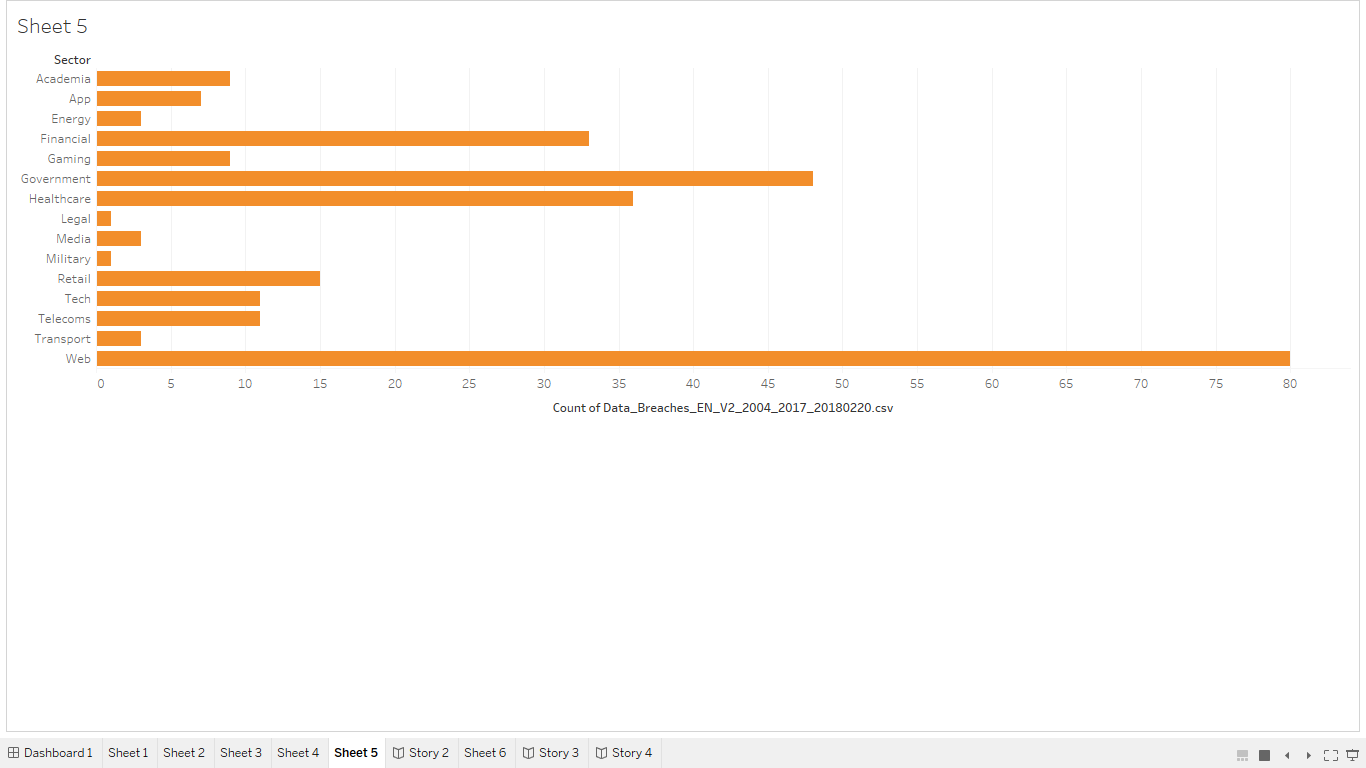
### Worksheet 4:

* This sheet shows the entity or the source through which data leak took place, where AOL and Yahoo, were the main source for data crime. Thus, before visiting any of the third-party app or sites, one must se its user policy, and check if they are accessing your data, if yes you must prevent clicking on the fake or scam links.



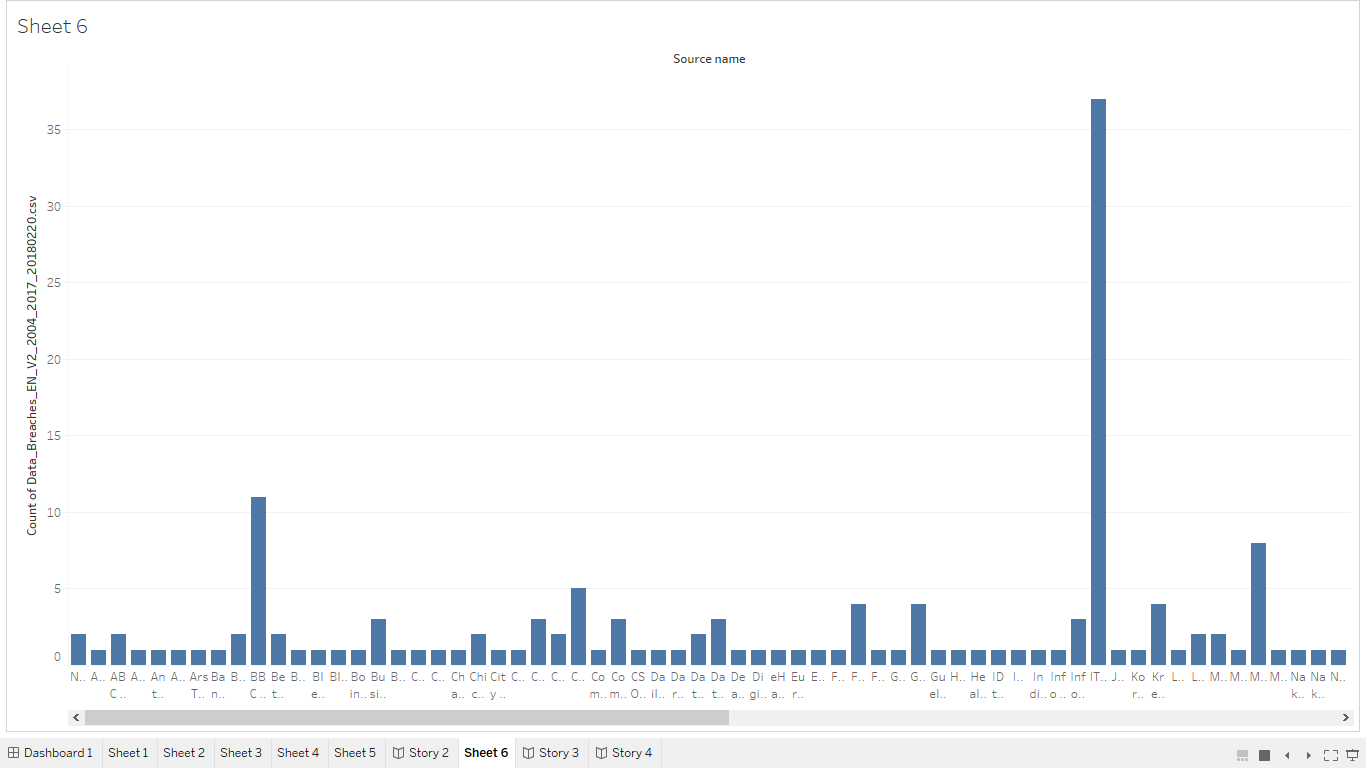
### Worksheet 5:

* This sheet shows the sectors affected most by the data leaks which took place between year 2004-2017.



### Worksheet 6:

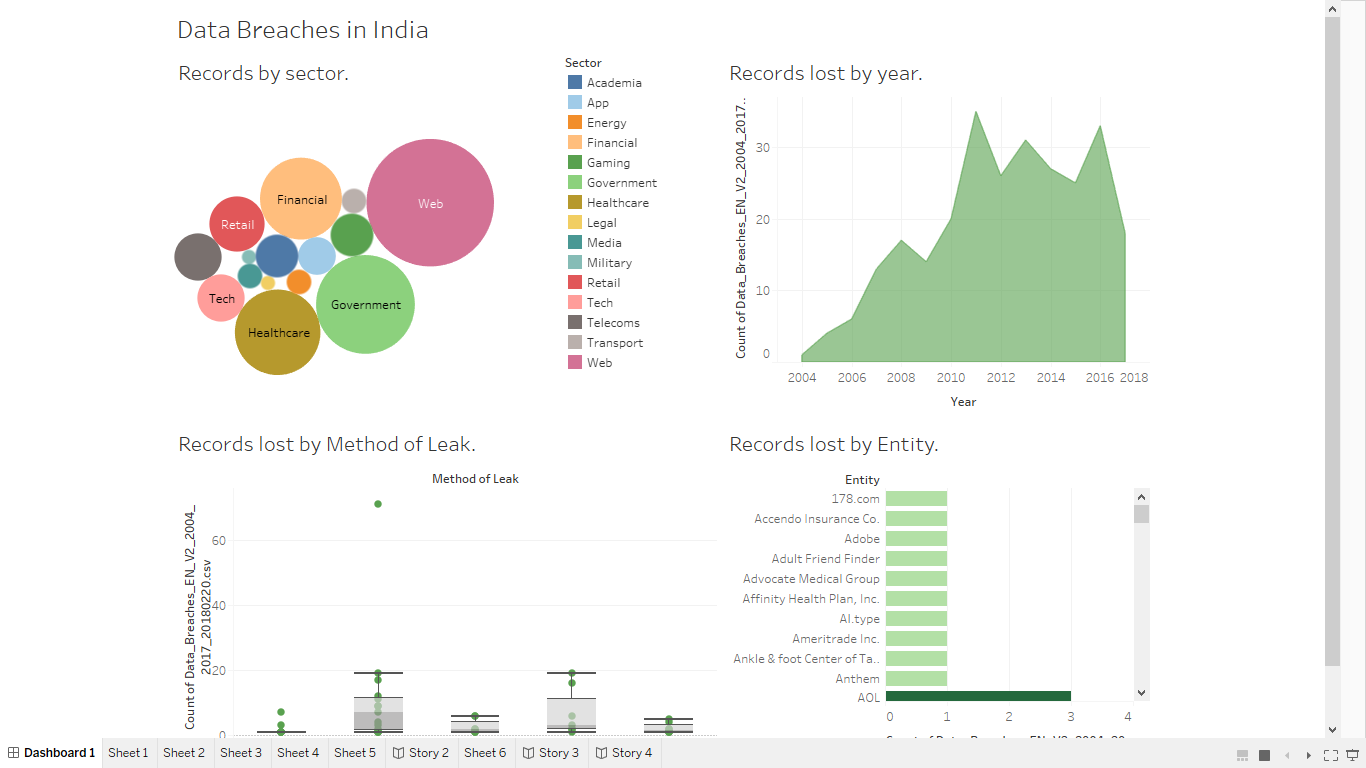
* The below worksheet shows the visualization performed to identify which was the main source of data leak.



## Dashboards

### Dashboard 1:

* This Dashboard shows the different visualizations performed to analyze data breaches between year 2004-2017.



# Use Cases

* Records by Sector
* Records lost by method of leak
* Records lost by year
* Records by entity

# Conclusion

The data breach visualizations performed on the data which was collected from year 2004 to 2017, showed a great impact, where these visualizations and interactive dashboard can lead to a successful conclusion of this project, where year 2010,2011 and 2017 were the years with the greatest number of cyber-attacks. Moreover, visualizations performed to determine the main factor of method leak, where hacking and lost/stolen devices dominated. Lastly, many third-party sites or apps, were the key reason which lead to data breaches. Thus, it taught us to be aware of the fake messages as well as scam sites, so don’t directly click on it, and take precautions to keep your data safe.

# Future Goals

* Will try to work on the live data in future, which would make a great impact on the public awareness on the data crimes.
* Lastly, will try to make many more interactive dashboards, which would make the data more readable, or understandable.

## REFERENCE

1. https://www.kaggle.com/datasets
2. https://public.tableau.com/app/discover

GitHub: https://github.com/amanp109/DV\_project.git