Data Analysis for Managers - CIA-1

Hemanath Kumar J - 21121015

**List of Top Data Breaches (2004 - 2021)**

<https://www.kaggle.com/hishaamarmghan/list-of-top-data-breaches-2004-2021>

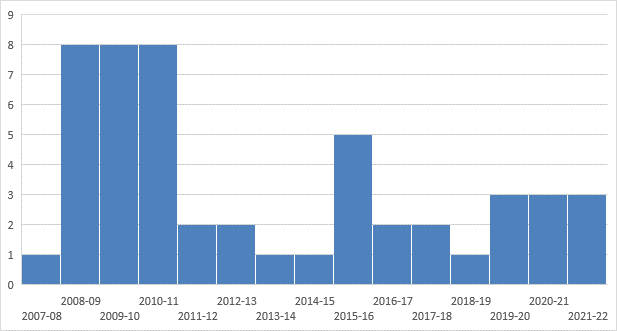


Data breaching is still a major issue in today’s world. There are data breaches even in some of the big companies in the world. Technology can be either a boon or a bane. The data selected studies this issue in great depth and analyzes the trends so far. The data which I took was in ratio scale. Among all the sectors, I chose healthcare sector (converted on the basis of nominal scale).

**Bar graph:**

**Inference:** We can see from the bar graph that in 2021 there was maximum breaches from Anthem INC. (aka Blue Cross Blue Shield). This healthcare is present in many states inside the US. While some of the healthcare mentioned has only one or few branches. There is also a sudden increase in number of breaches in 2021, which maybe because of poor maintenance.

**Histogram:** No. of hacked in healthcare vs Years



**Inference:** From this histogram, we can see that the hacking was the major issue, compared to other minor issues like poor security & stolen media. Hacking is reduced when compared with a decade before, but still need more focus as it’s being constant over the past years.

**Pie chart:** Year and Percentage of hacks at healthcare sector.

**Inference:** This pie chart shows that the hacking count was more at 2009-2010 at healthcare industries.

**Scatter plot:** No. of breaches happening through hacking vs remaining types(in all sectors)

**Inference:** This data doesn’t seem to resemble any type of pattern. So for this there is no visible relation between the two variables.

**Contingency table:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **Count of healthcare** | **Count of hacked** | **Sum of 2200000** |  | **Range** |  |  |
| **2016** | **9** | **9** | **67018816** |  | **Low** |  |  |
| **2017** | **4** | **4** | **89747877** |  | **Low** |  |  |
| **2018** | **8** | **8** | **338730465** |  | **Medium** |  |  |
| **2019** | **9** | **9** | **507839740** |  | **Medium** |  |  |
| **2020** | **16** | **16** | **727621779** |  | **High** |  |  |
| **2021** | **4** | **4** | **125247000** |  | **High** |  |  |
| **Grand Total** | **50** | **50** | **1856205677** |  |  |  |  |

**Inference:** The attack began with phishing e-mails sent to a handful of its employees. This was the main reason of increase in data breaches from last 2 years.

**Conclusion:** There are some reports saying that there are many records which has not been exposed yet. Most breaches were the result of hacking, but in second place was insider theft or employee error/negligence. Data breach costs rose from USD 3.86 million to USD 4.24 million, the highest average total cost in the 17-year history, which is proved through this dataset.

**Suggestions:** We can avoid this by increasing IT infrastructure, frequent risk analysis and limited access to patient’s information.