



Shikhar Saxena
September 2021

CONTENT

What is Data Breach ?

Exploratory approach

Insights from EDA

Conclusions and Recommendations

What is Data Breach ?

A data breach is a security violation in which sensitive, protected, or confidential data is copied, transmitted, viewed, stolen, or used by an individual unauthorized to do so.

Data breaches may involve financial information such as credit card & debit card details, bank details, or vulnerable unstructured data like personal health information, files, documents, sensitive information, etc.

Consequences of Data Breach:-

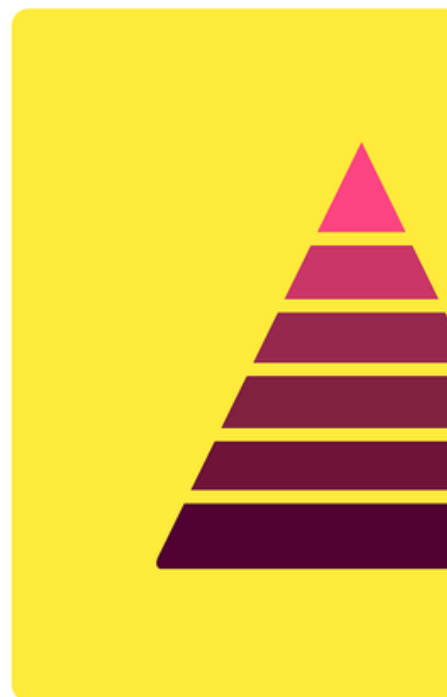
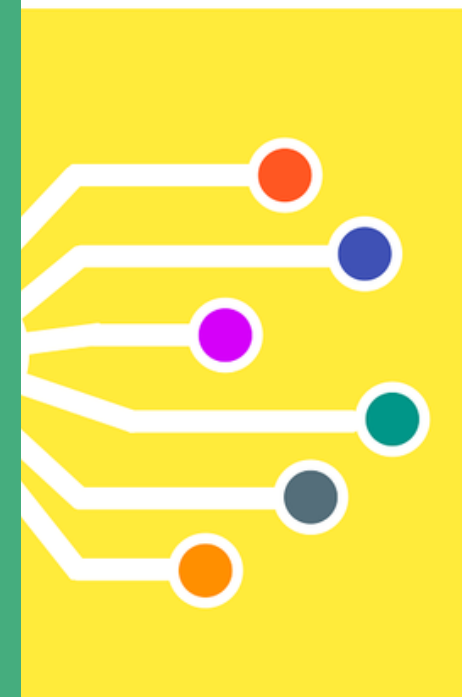
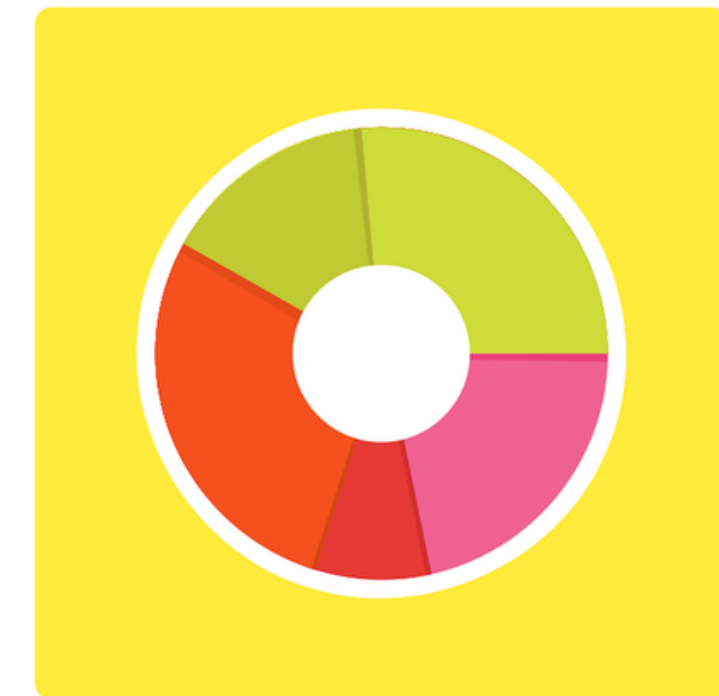
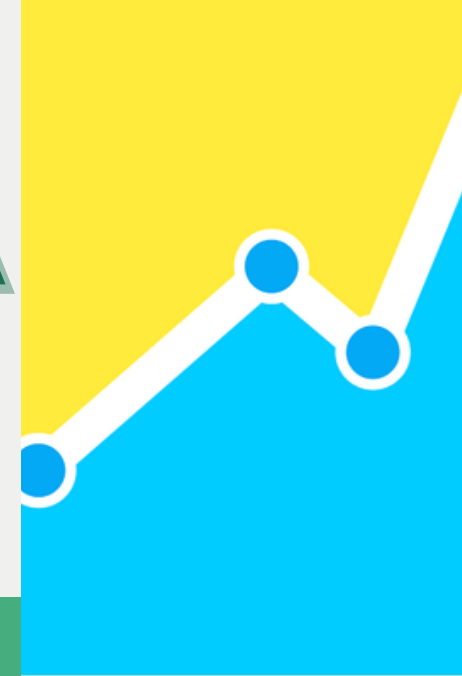
- Financial Loss
- Brand Value and Reputation Damage
- Operational Downtime
- Legal Action and Investigation Procedures
- Loss of Sensitive Data
- Lost Revenue and Impact on Finances

2.7
billion

Record- In 2019, a collection of 2.7 billion identity records, consisting of 774 million unique email addresses and 21 million unique passwords, was posted on the web for sale.

EXPLORATORY DATA ANALYSIS STEPS

1. Importing the required libraries
2. Load the Dataset
3. Data Pre-processing
4. Understanding the Data
5. Data Analysis and Visualization
6. Final Conclusions

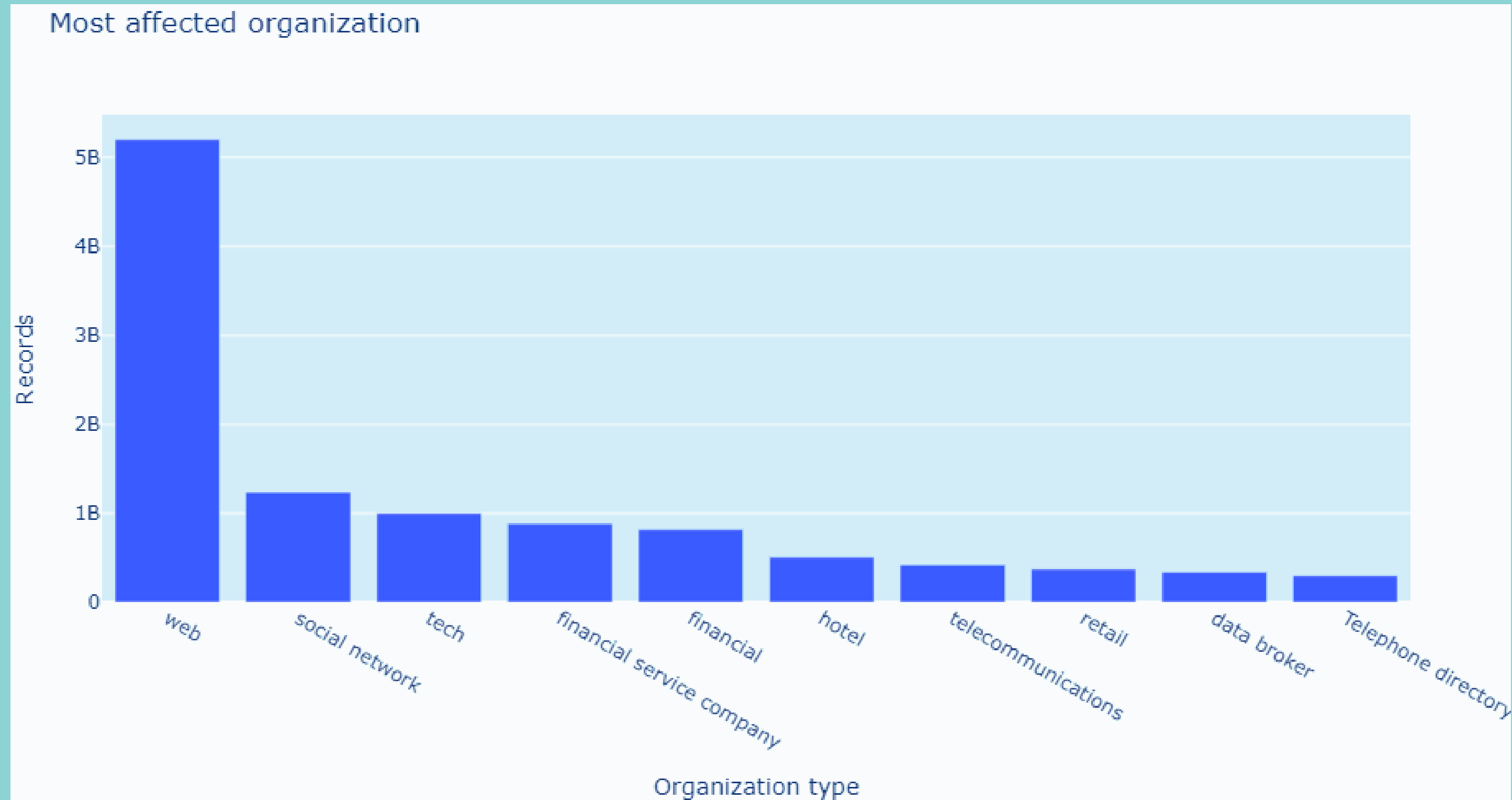


Dataset

	Entity	Year	Records	Organization type	Method
0	21st Century Oncology	2016	2200000	healthcare	hacked
1	500px	2020	14870304	social networking	hacked
2	Accendo Insurance Co.	2020	175350	healthcare	poor security
3	Adobe Systems Incorporated	2013	152000000	tech	hacked
4	Adobe Inc.	2019	7500000	tech	poor security

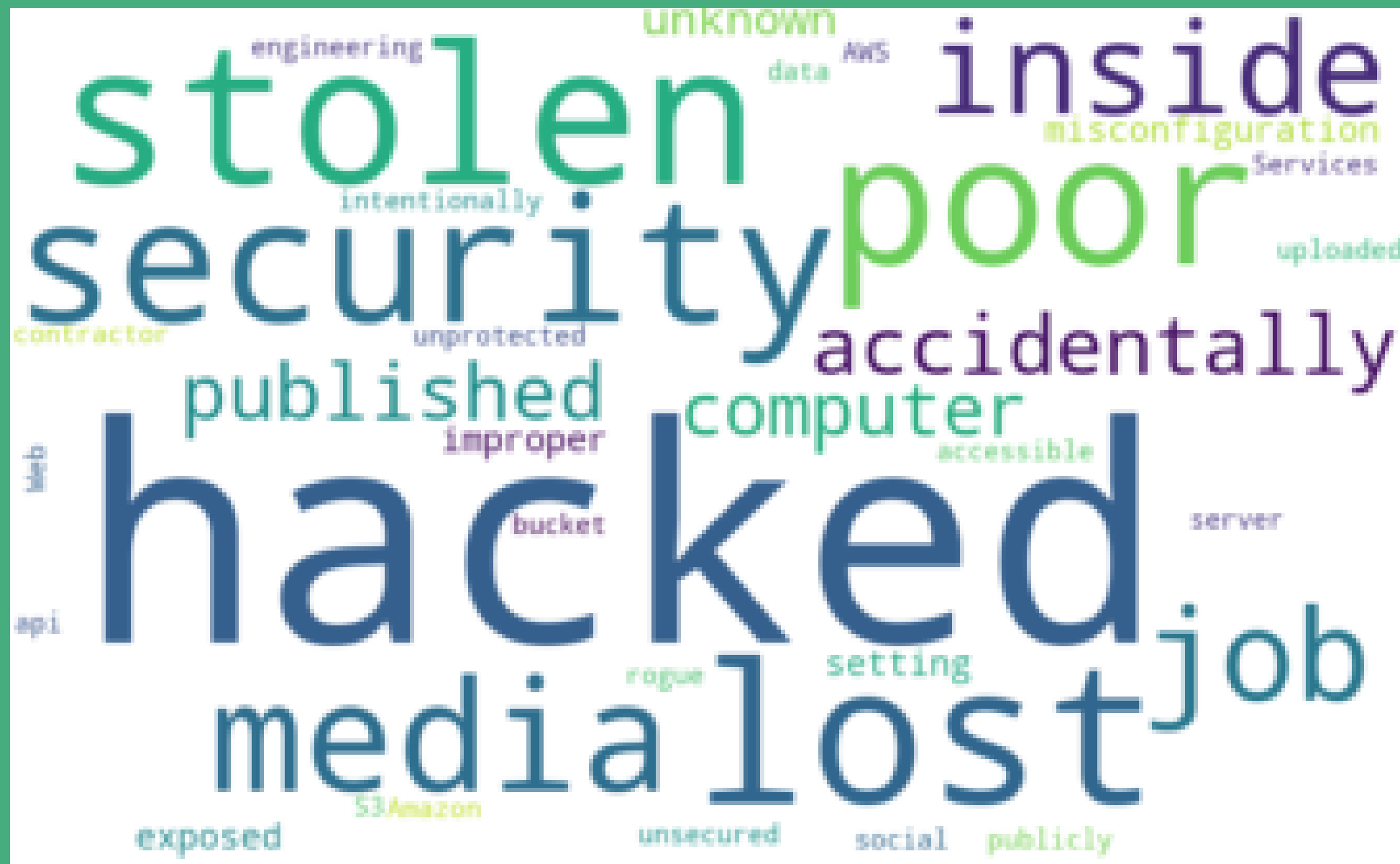
- Entity:- Names of all companies
- Year:- In what year did the data breach took place (2004-2021)
- Records:- Total No. of Data Breaches happened with the company.
- Organization type:- all the domains/ sectors that are affected by Data Breach.
- Method:- all the possible ways or methods used for Data Breach.

Visualizing top 10 most affected Organizations



Insights:- Web organizations top the chart with more than 5B data breaches, followed by social networks, tech.

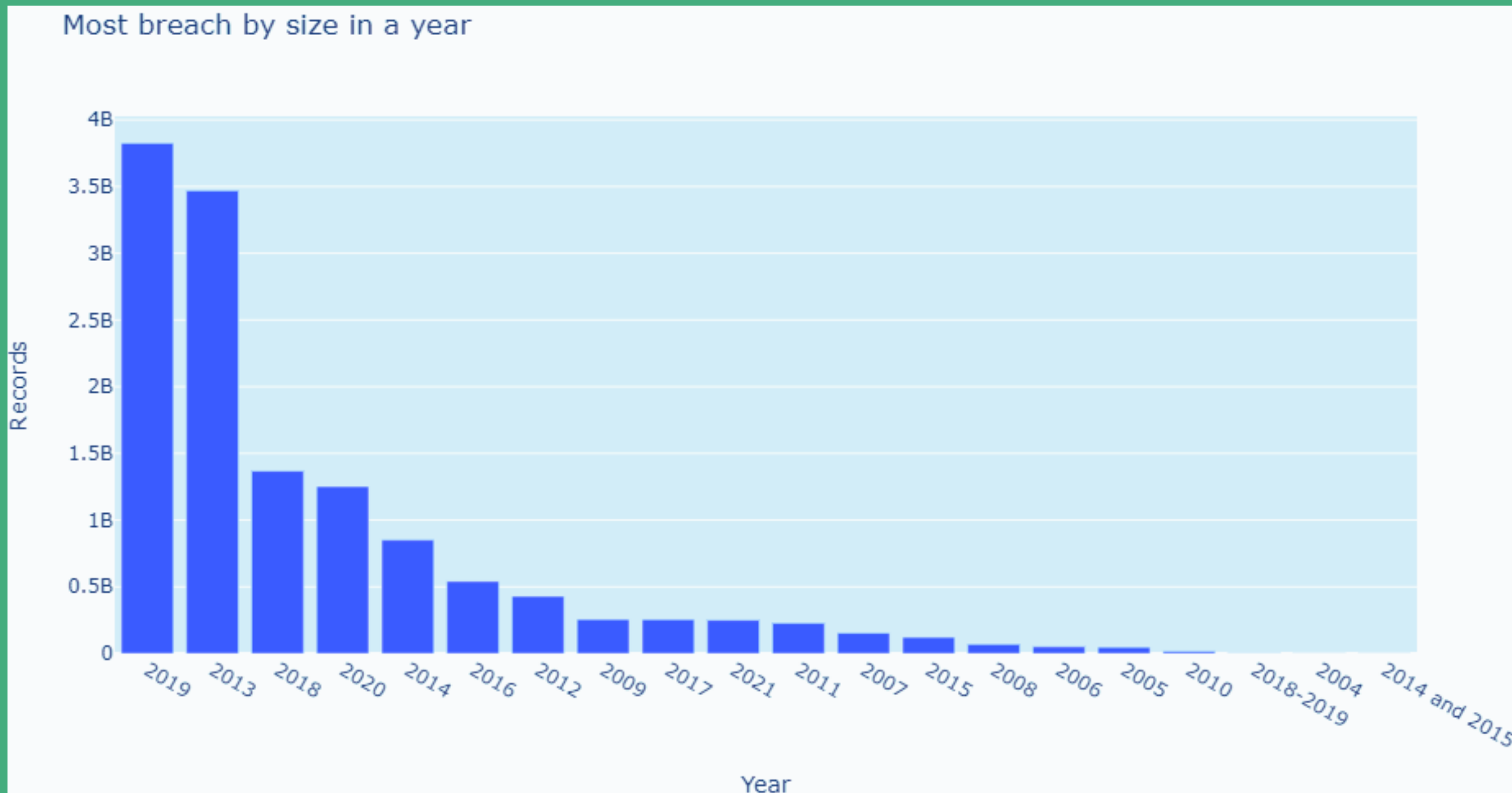
Visualizing the different Methods used for Data Breach



Insights:-

- Hacking method is the top priority for data breachers.
- Poor Security is also a frequent method for Data Breach.
- In some companies, an Inside job is also responsible for data breaches.

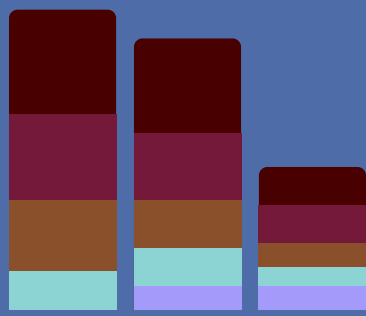
Visualizing Most Data Breaches by size in a year



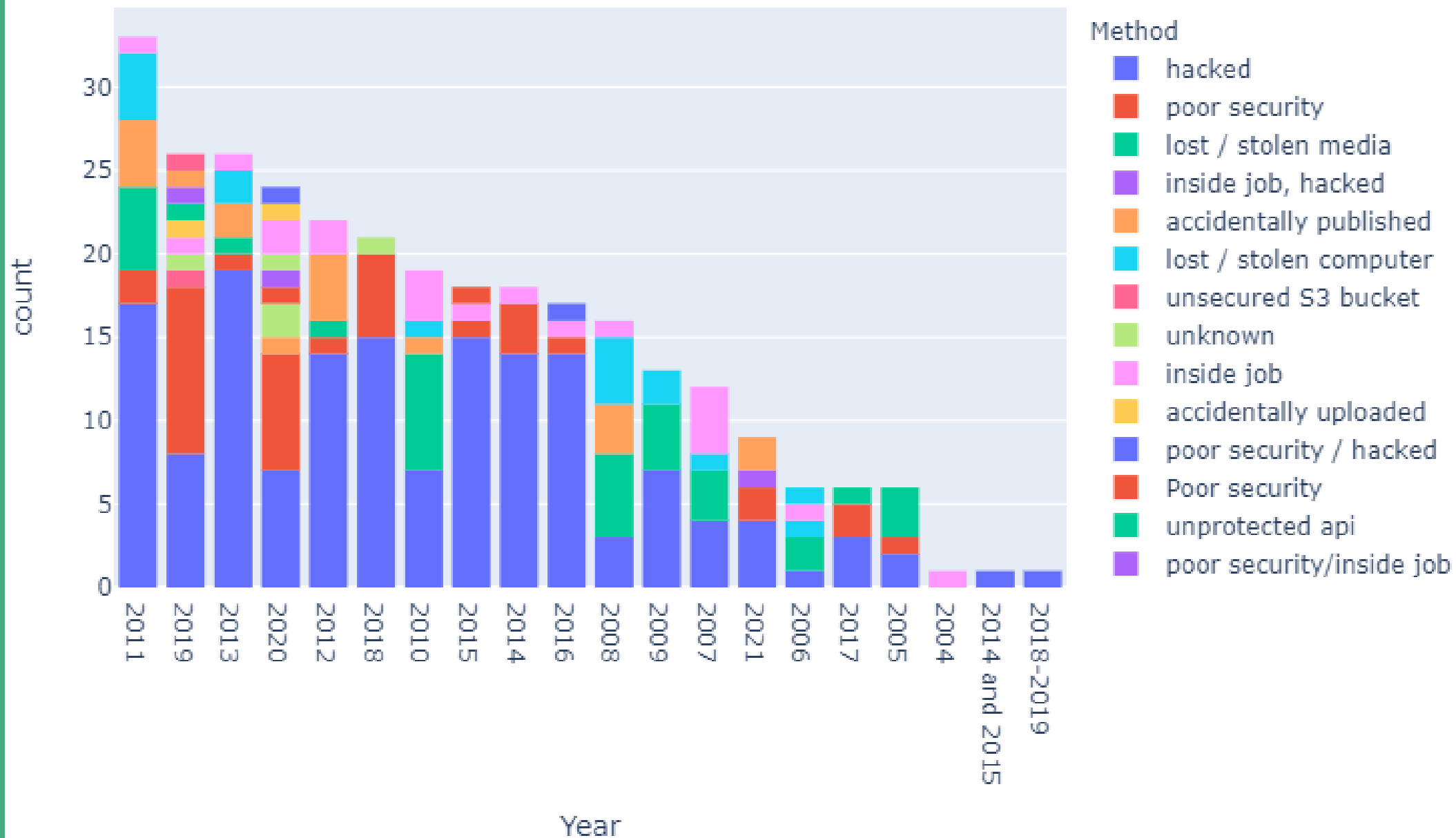
Insights:-

- Year 2019 has the highest number of data breaches. More than 3.5B user's data has been leaked in the year 2019.
- Year 2004 has the least data breaches.

Analyze different methods used for data breaches in a particular year



Methods used for Data Breach in a year

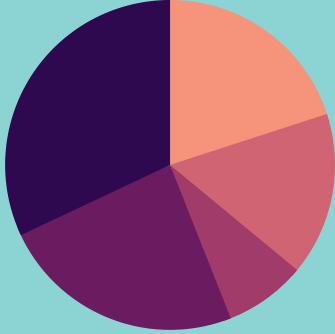


Insights:-

- In 2011, there are 17 cases of data breaches due to hacking, 2 due to poor security, 5 due to lost/stolen media, 4 due to accidentally published, 4 due to lost/stolen computer, and 1 due to inside job.
- In 2019, 8 cases of data breaches due to hacking, 10 due to poor security, 1 due to unsecured S3 bucket, 1 due to inside job, 1 due to unprotected api and 1 due to data exposed by misconfiguration.

- The Hacking Method is most used for Data Breach in the last 7 to 8 years.

Most Affected Companies in each Sector



Most Affected Companies in each Sector



Insights:-

- Yahoo is the most affected company by Data Breaches, followed by Facebook.
- In the Web Sector, Yahoo is the most affected company, followed by Friend Finder Networks, Wattpad, and eBay.
- In Social Network Sector, Facebook is the most affected company, followed by Instagram.
- In Tech Sector, MongoDB is the most affected company, followed by Microsoft.
- In Financial Sector, First American Corporation is the most affected company.
- In Hotel Sector, Marriott International is the most affected company.

Final Conclusions and Recommendations

- Web Organizations top the chart with more than 5B data breaches, followed by social networks, tech. If you are a frequent web user, then you should be more careful.
- Year 2019 has the highest number of data breaches. More than 3.5B user's data has been leaked in the year 2019
- The hacking method is the top priority for data breaches, followed by Poor Security, so when you are online be careful from hackers.

Final Conclusions and Recommendations

- In Web Sector Yahoo is the most affected company, in Social Network Sector Facebook is the most affected, in Tech Sector MongoDB is the most affected, in the financial Sector, First American corporation is most affected, in the hotel sector Marriott International is most affected, in the telecommunications sector, Airtel is the most affected company.
- Be very careful whenever you do any online bank transactions, be careful from hackers while using the Internet, do not share your bank and personal details with anyone, keep your data protected.

REFERENCES

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

GitHub link for code: https://github.com/Shikharsaxena195/DataBreach_EDA

Websites referred:

https://en.wikipedia.org/wiki/Data_breach

<https://plotly.com/python/bar-charts/>