Can cyber security stop the hack before it happens?

Most of our transactions are done online these days, making life easier for us. However, it also poses safety risks, as our confidential information is exposed to fraud and hackers. In fact, hackers detected several ways to steal important information, which makes our lives more difficult, they can access secure information by only entering the phone number connected to the bank account.

Due to rising risks associated with electronic payment systems and the increase in piracy operations around the world, the need to tighten and develop regulatory systems is increasing.

Questions:

Q 1: Which month has the most penetration?

Q2: How many times has the anonymous recording?

Q3: What number of hacks have there?

Q4: How long has there been no penetration of server?

Dataset :

I'll work with the Kaggle dataset of Malicious Server Hacks, I propose that by using knowledge of cybersecurity, we can predict whether a server will be hacked before it happens.

```
In [91]: import pandas as pd

In [96]: Malicious_Server = pd.read_csv("Train.csv")

In [103]: Malicious_Server.head(10000)
```

Out[103]:

	INCIDENT_ID	DATE	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_10	X_11	X_12	X_13	X_14	X_15	MALICIOUS_OFFENSE
0	CR_102659	04-JUL-04	0	36	34	2	1	5	6	1	6	1	174	1.0	92	29	36	0
1	CR_189752	18-JUL-17	1	37	37	0	0	11	17	1	6	1	236	1.0	103	142	34	1
2	CR_184637	15-MAR-17	0	3	2	3	5	1	0	2	3	1	174	1.0	110	93	34	1
3	CR_139071	13-FEB-09	0	33	32	2	1	7	1	1	6	1	249	1.0	72	29	34	1
4	CR_109335	13-APR-05	0	33	32	2	1	8	3	0	5	1	174	0.0	112	29	43	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9995	CR_18276	20-JUL-93	0	49	49	6	5	4	2	2	6	1	102	0.0	24	29	43	1
9996	CR_155592	19-APR-12	0	5	5	3	5	5	6	1	6	1	316	1.0	103	80	34	1
9997	CR_168953	29-NOV-14	0	21	23	4	1	5	6	1	6	1	316	1.0	103	76	34	1
9998	CR_129357	18-AUG-08	0	14	17	9	0	5	6	0	5	1	174	0.0	103	29	23	1
9999	CR_118701	23-MAR-06	7	45	45	10	3	9	8	0	5	1	249	1.0	112	93	34	1

Figure 1

Here is what real data looks like!

This is a subset of the full data loaded as pandas which contains 18 features and 23856 observations

The Features:

1 - INCIDENT_ID represents Unique identifier for an incident log

2 – DATE represents the Data of incident occurrence

3 - X*1* - *X*15 represents Anonymized logging parameters

4 - MALICIOUS_OFFENSE represents [Target] Indicates if the incident was a hack [1: Yes; 0:No]

Tools

This study will be utilizing tools such as pandas, matplotlib, numpy for data analysis, and a Jupyter notebook for its execution.

TO DO

I'll follow this workflow on the dataset step by step:

1 - Data cleaning

2 - Exploratory data analysis

3 - Feature engineering

4 - Machine learning

5 - Model evaluation and data visualization