Internship Report

Honeypot attacks data analysis

Wang Lei
01/26/2021

Dataset

First 5 rows of the dataset

	date	time	access_ip	host_ip	request_line	status_code	match_result	
0	[2020-03-01	00:01:37+0900]	134.209.184.77	closedbeta.net:80	GET /blog/wp-login.php HTTP/1.1	200	1011	R0VUIC9ibG9nL3dwLWxvZ2luLnBocCBIVFRQLz
1	[2020-03-01	00:01:38+0900]	134.209.184.77	closedbeta.net:80	POST /blog/closedbeta.net/blog/wp-login.php HT...	200	1011	UE9TVCAvYmxvZy9jbG9zZWRiZXRhLm5ldC9ibG
2	[2020-03-01	00:11:53+0900]	193.106.30.99	closedbeta.net:80	GET / HTTP/1.1	200	False	R0VUIC8gSFRUUC8xLjEKSG9zdDogY2xvc2VkYm
3	[2020-03-01	00:19:47+0900]	165.22.222.119	closedbeta.net:80	POST /code.conflicts.php HTTP/1.1	200	1037	UE9TVCAvY29kZS5jb25mbGljdHMucGhwIEhUVF
4	[2020-03-01	00:19:55+0900]	189.240.124.61	closedbeta.net:80	POST /work.clases.php HTTP/1.1	200	1037	UE9TVCAvd29yay5jbGFzZXMucGhwIEhUVFAvM

Dataset

8 columns of the dataset

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 563883 entries, 0 to 199532
Data columns (total 8 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   date           563883 non-null  object
 1   time           563883 non-null  object
 2   access_ip      563883 non-null  object
 3   host_ip        563883 non-null  object
 4   request_line   563883 non-null  object
 5   status_code    563883 non-null  int64
 6   match_result   563883 non-null  object
 7   request_all    563883 non-null  object
dtypes: int64(1), object(7)
memory usage: 38.7+ MB
```

Feature engineering

14 new variables created

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 563383 entries, 0 to 199532
Data columns (total 14 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   month          563383 non-null  object
 1   day            563383 non-null  int64
 2   hour_JP        563383 non-null  int64
 3   country        563383 non-null  object
 4   country_code   563383 non-null  object
 5   ip_country     563383 non-null  object
 6   region         563383 non-null  object
 7   city           563383 non-null  object
 8   latitude       563383 non-null  float64
 9   longitude      563383 non-null  float64
 10  timezone       563383 non-null  int64
 11  hour_Local     563383 non-null  int64
 12  method         563383 non-null  object
 13  http_version   563383 non-null  object
dtypes: float64(2), int64(4), object(8)
memory usage: 64.5+ MB
```

Categories of variables

The variables were divided into 3 categories

- Time related variables
- Location related variables
- Technique related variables

Categories of variables

- **Time related category**:
  - date
  - month
  - day
  - hour at Japan time
  - hour at local time

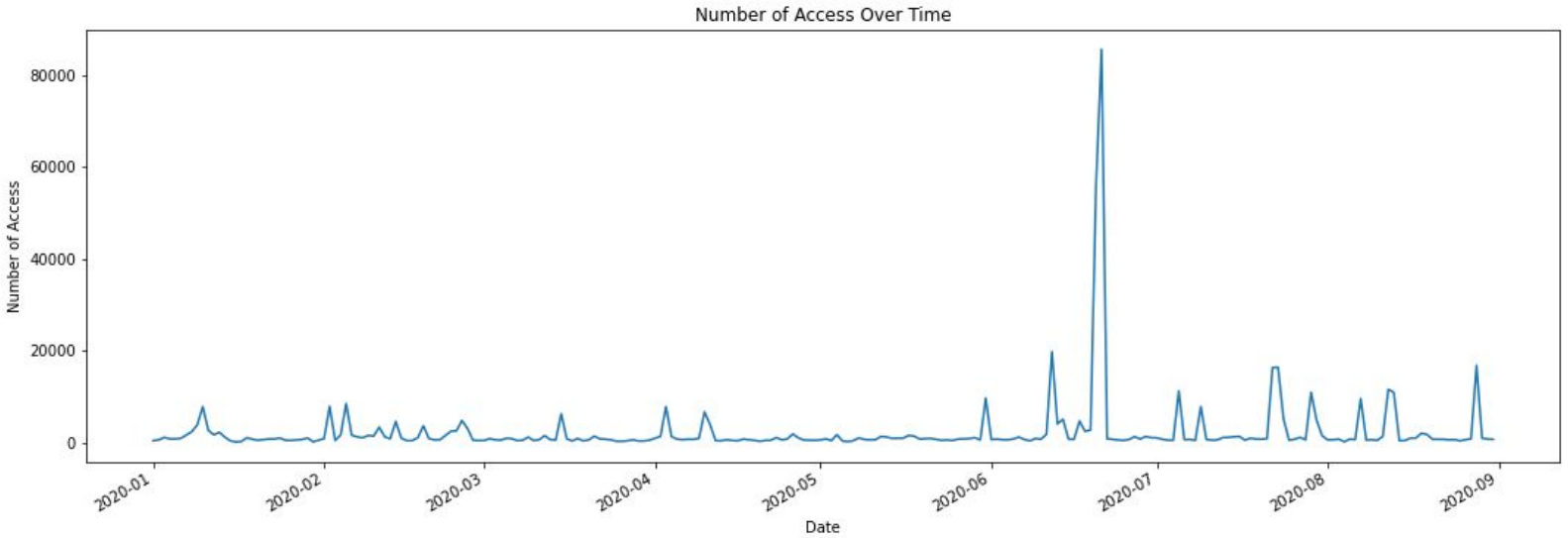
Categories of variables

- **Location related variables:**
  - **IP**
  - **country**
  - **city**

Categories of variables

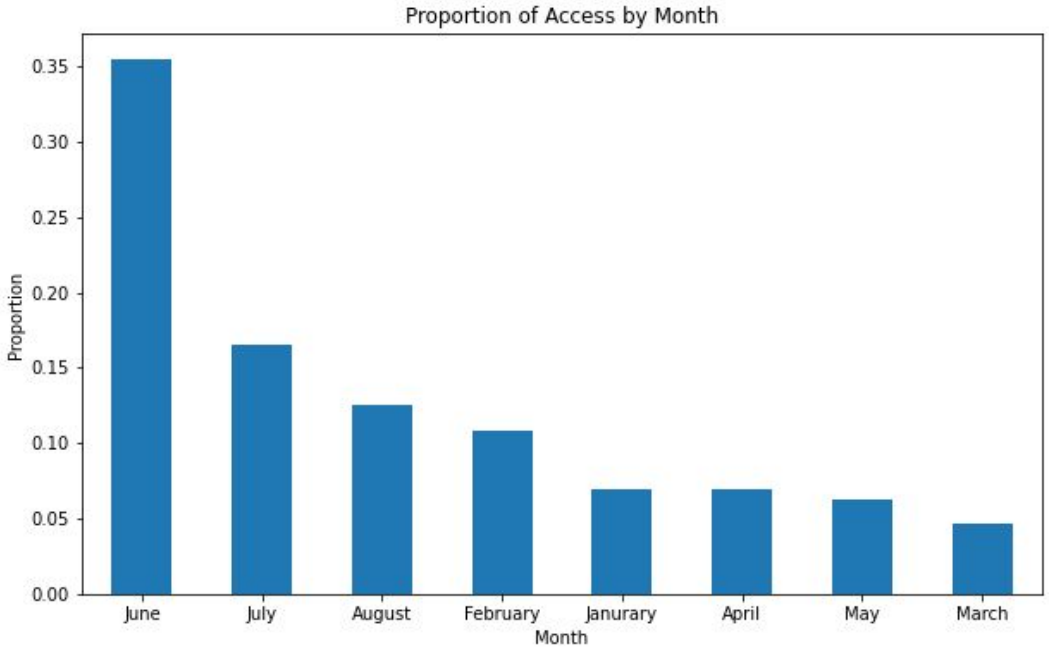
- テクニックのカテゴリー:
  - method
  - Http version

Time related variables — date



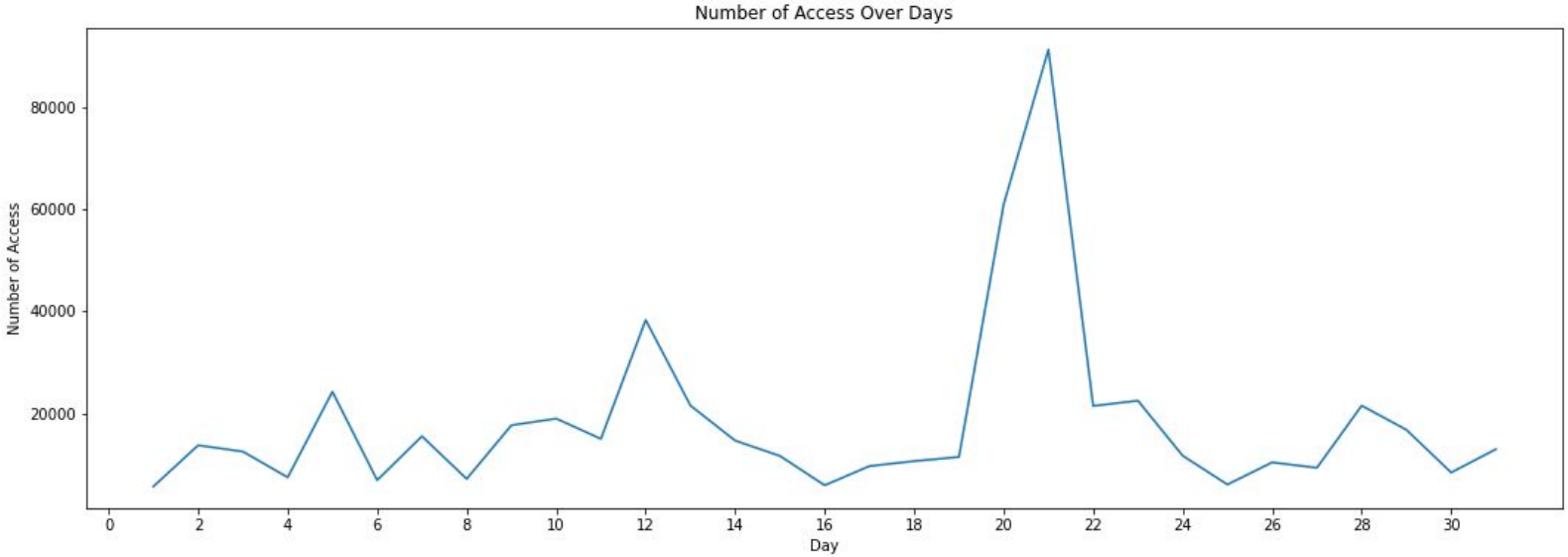
There are fluctuations. The peak occured in June.

Time related variables— month



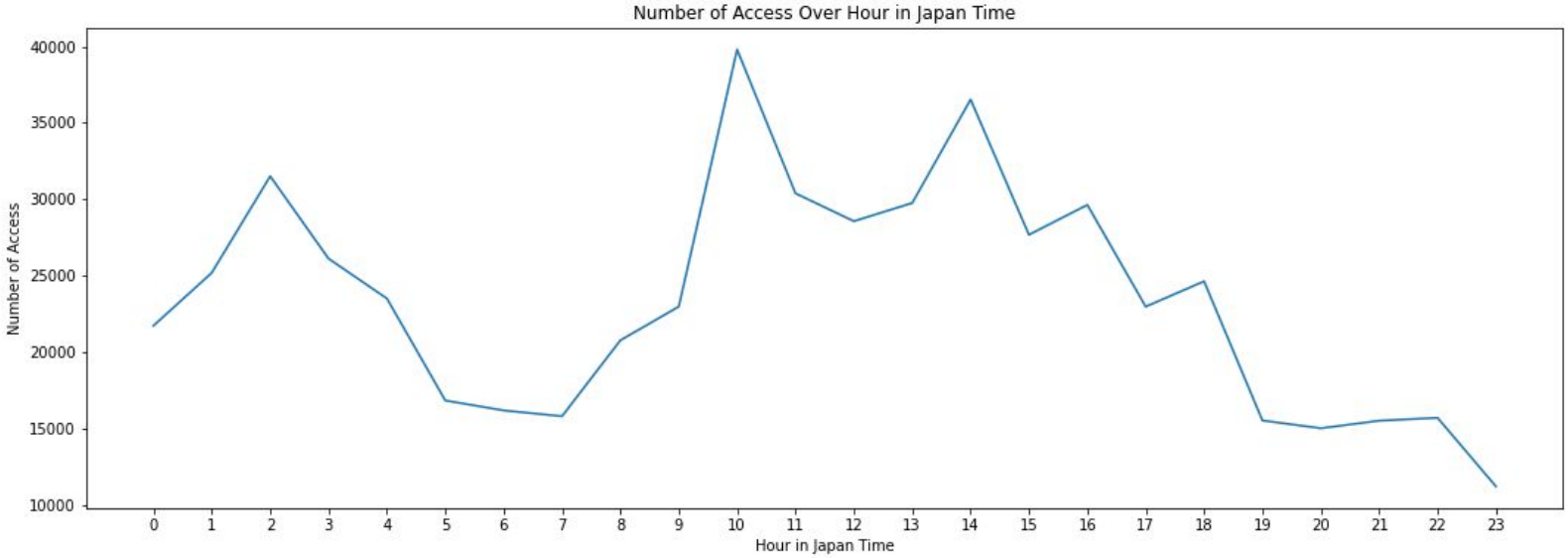
June has the most attacks. July and August also have relatively more attacks than the other months.

Time related variables — day



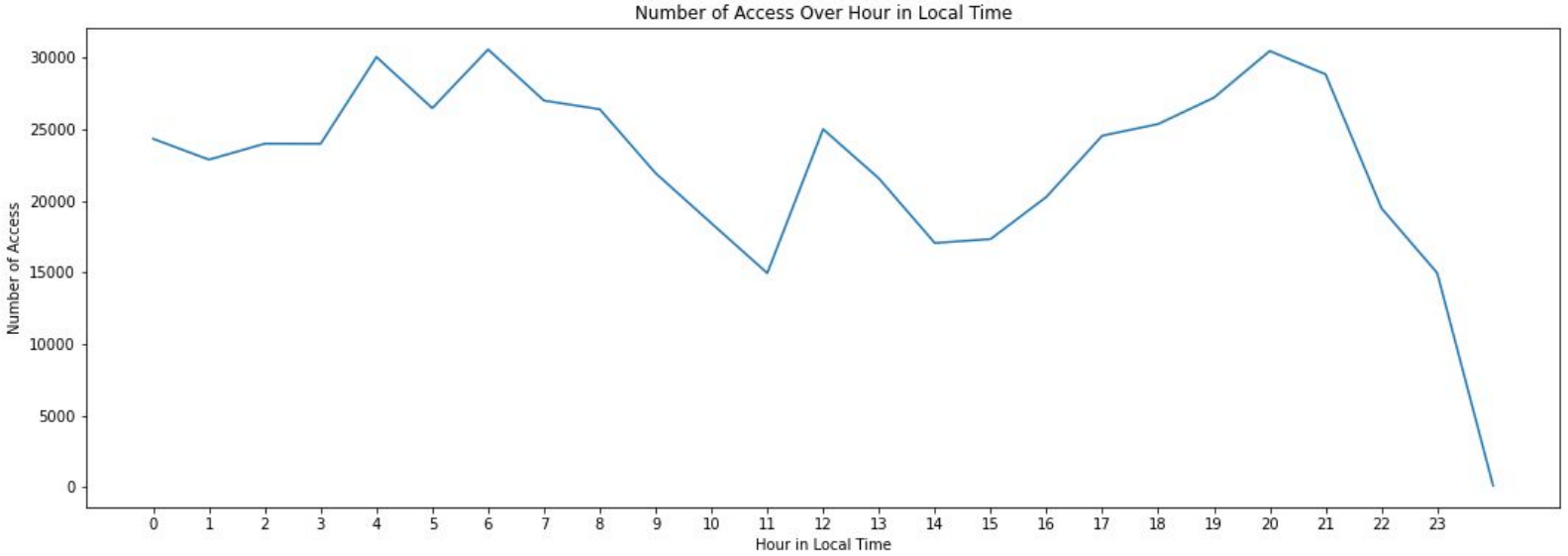
Regarding the day of month, 21 and 12 each month had relatively more attacks.

Time related variables — hour (Japan time)



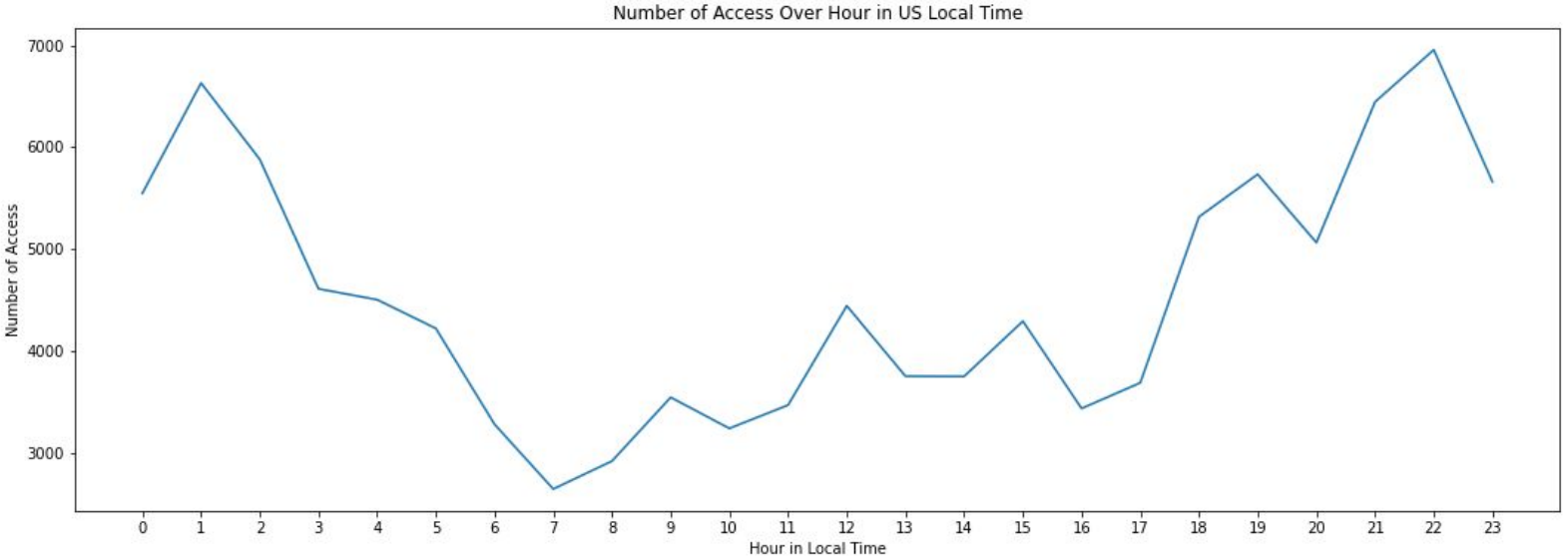
More attacks occurred at 10Am and 2PM in Japan local time

Time related variables — hour (local time all)



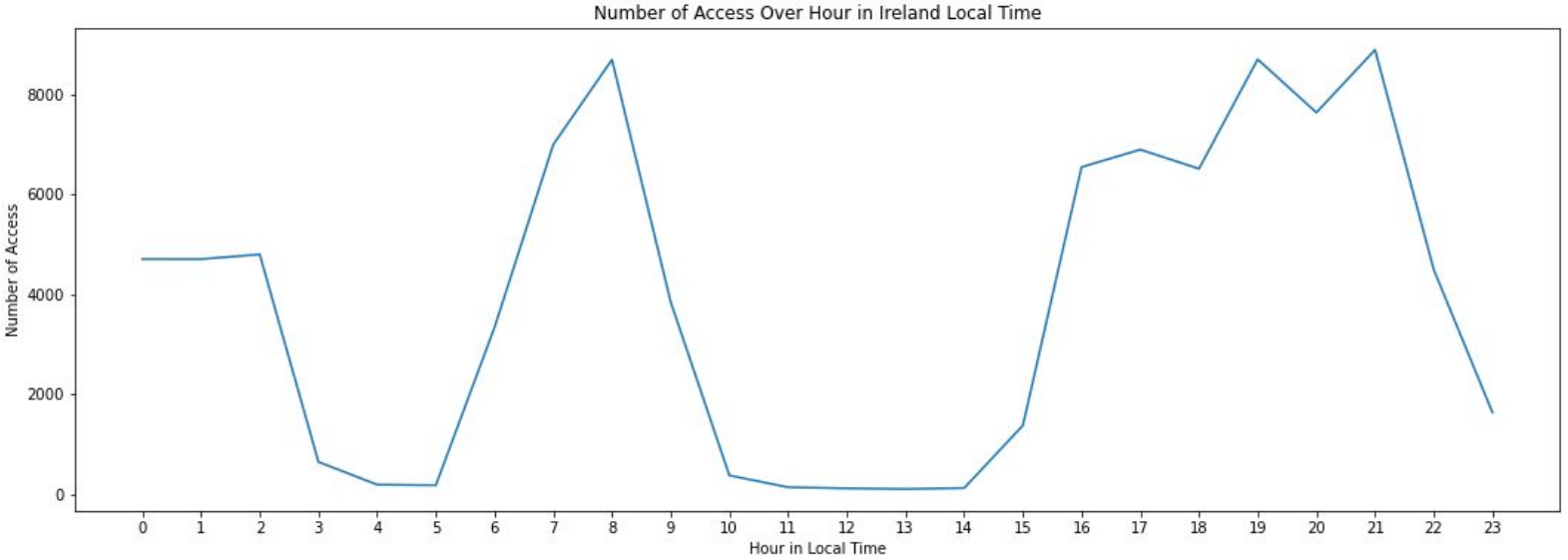
Attackers prefer attacking early in the morning or at night rather than in daytime.

Time related variables— hour (US local time)



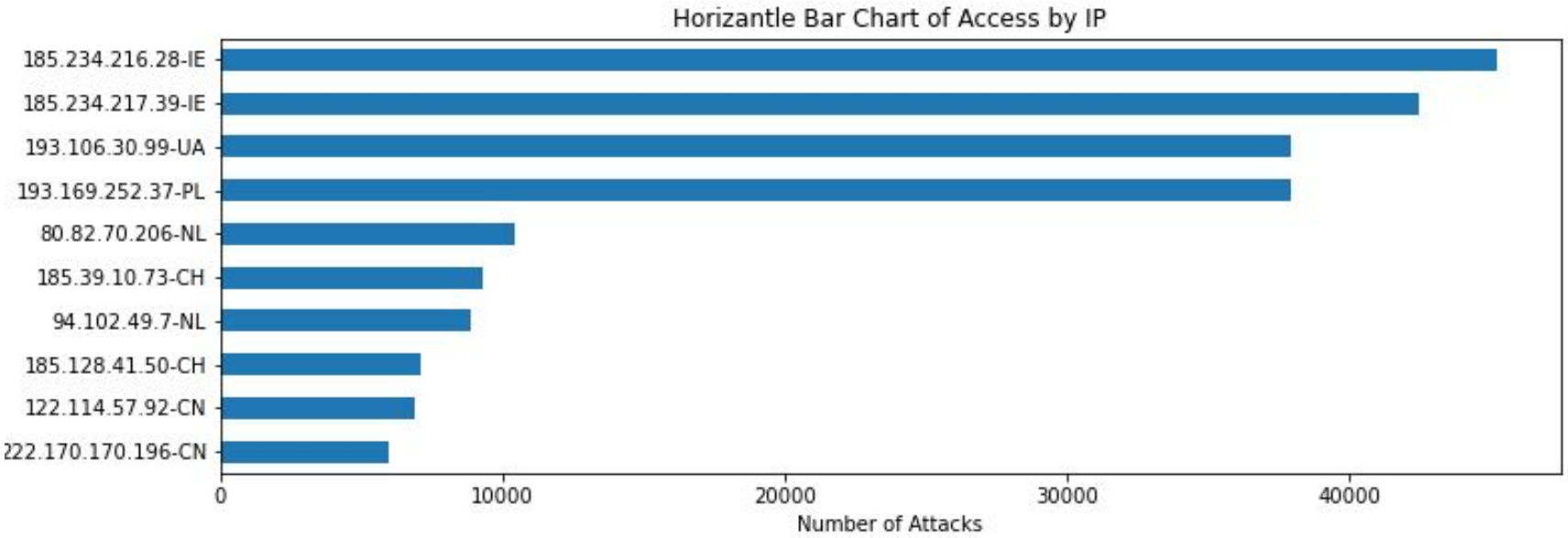
Most of the attacks from US occurred at 1AM and 10PM in US local time

Time related variables — hour (Ireland local time)



Most of the attacks from Ireland occurred at 8AM, 7PM, 9PM.

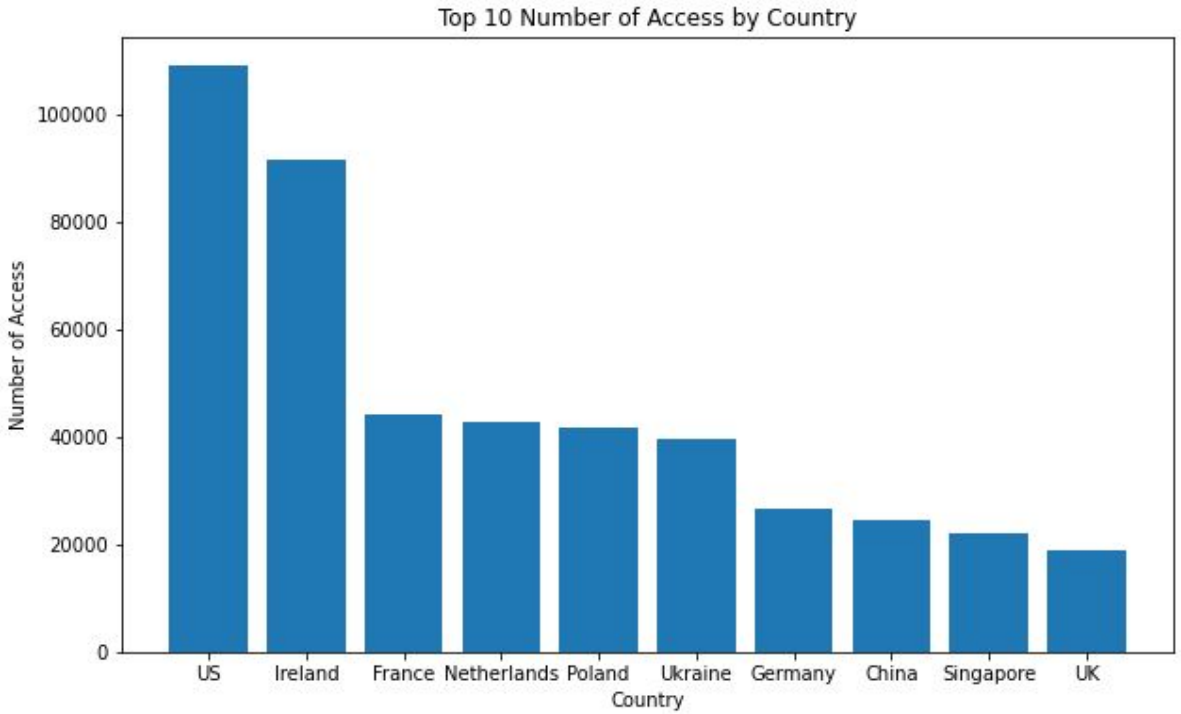
Location related variables— IP



Location related variables — IP

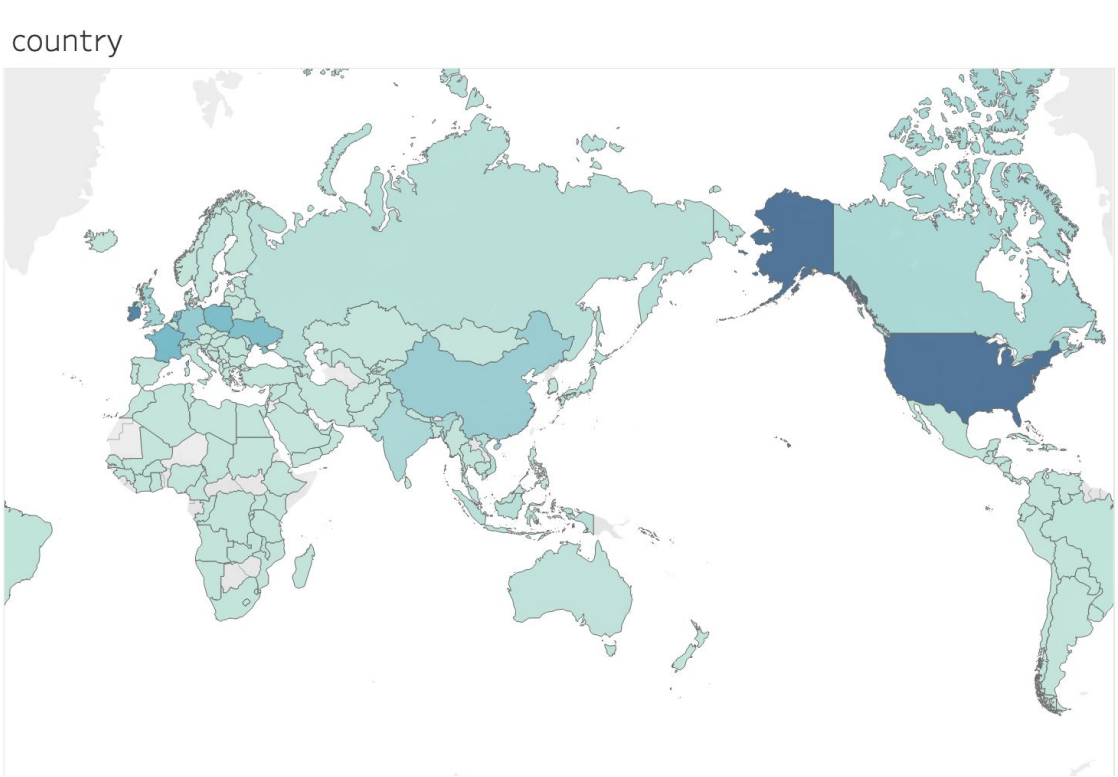
ip_country	country	
185.234.216.28-IE	Ireland	45191
185.234.217.39-IE	Ireland	42428
193.106.30.99-UA	Ukraine	37926
193.169.252.37-PL	Poland	37913
80.82.70.206-NL	Netherlands	10444
185.39.10.73-CH	Switzerland	9280
94.102.49.7-NL	Netherlands	8840
185.128.41.50-CH	Switzerland	7116
122.114.57.92-CN	China	6840
222.170.170.196-CN	China	5952

Location related variables — country



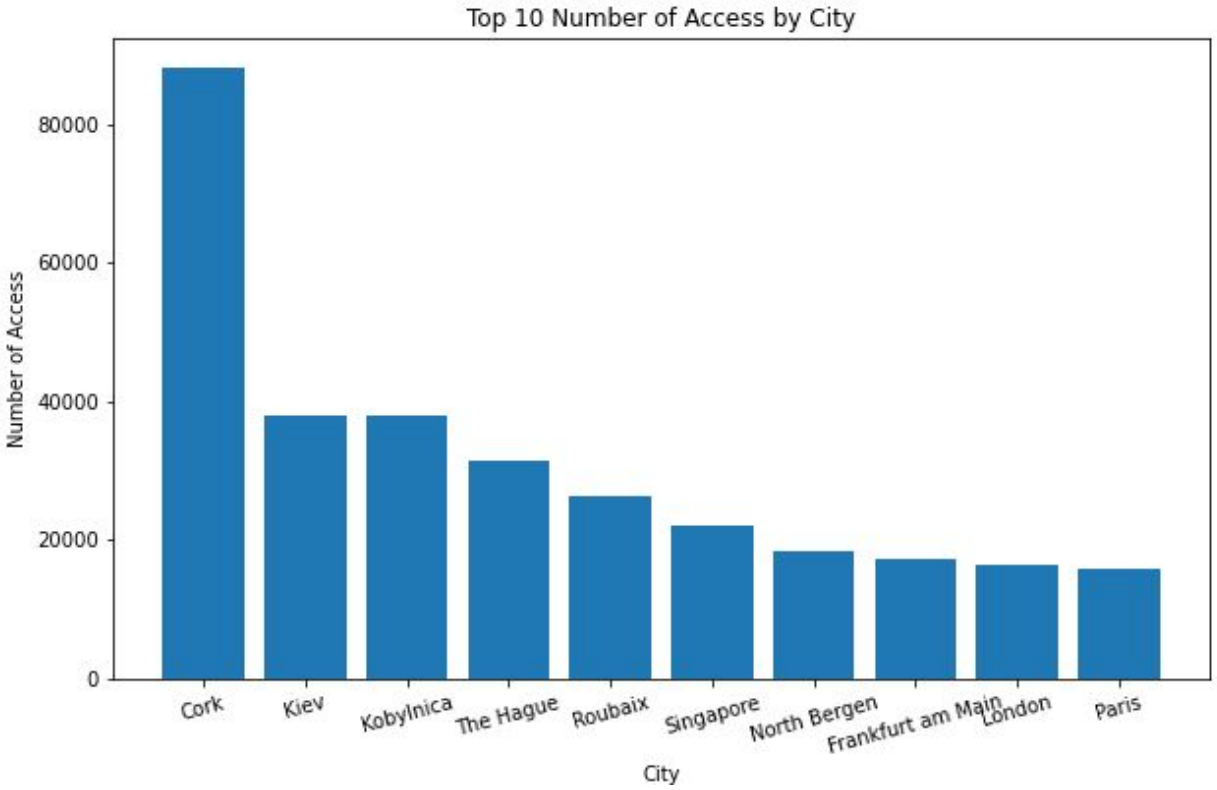
	country	count
0	US	108978
1	Ireland	91565
2	France	44303
3	Netherlands	42770
4	Poland	41847
5	Ukraine	39682
6	Germany	26578
7	China	24729
8	Singapore	22071
9	UK	18940

Location related variables — country



	country	count
0	US	108978
1	Ireland	91565
2	France	44303
3	Netherlands	42770
4	Poland	41847
5	Ukraine	39682
6	Germany	26578
7	China	24729
8	Singapore	22071
9	UK	18940

Location related variables — city

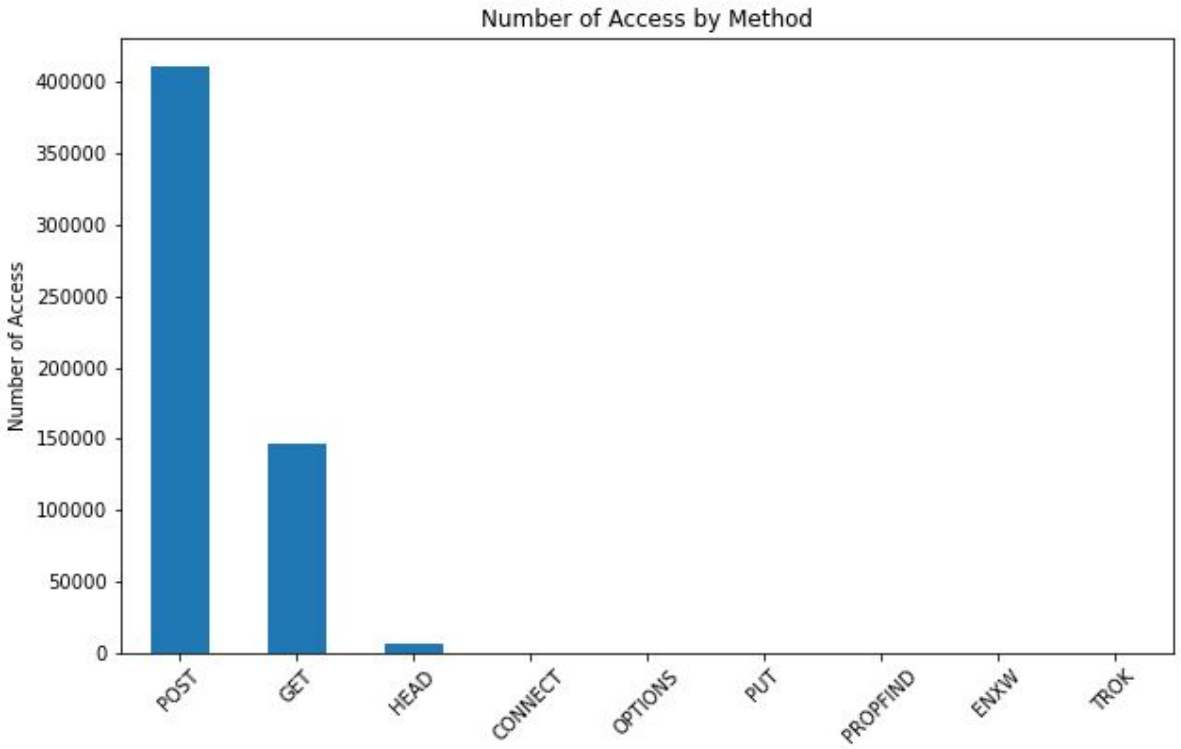


	city	count
0	Cork	88017
1	Kiev	37985
2	Kobylnica	37977
3	The Hague	31504
4	Roubaix	26215
5	Singapore	22071
6	North Bergen	18417
7	Frankfurt am Main	17266
8	London	16537
9	Paris	15948

Location related variables — city

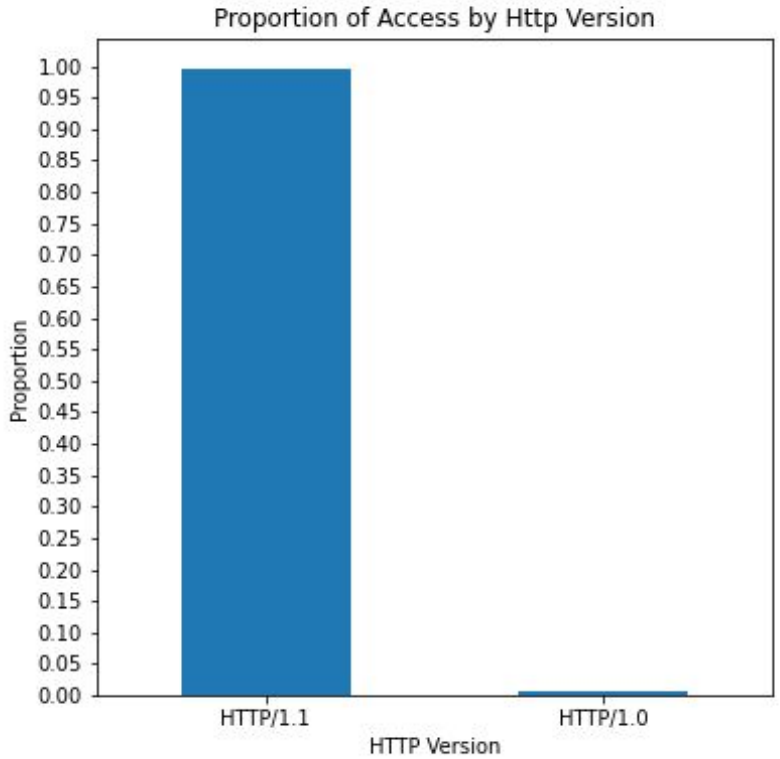
	city	count
0	Cork	88017
1	Kiev	37985
2	Kobylnica	37977
3	The Hague	31504
4	Roubaix	26215
5	Singapore	22071
6	North Bergen	18417
7	Frankfurt am Main	17266
8	London	16537
9	Paris	15948

Technique related variables— method



Most of the method are POST and GET.

Technique related variables — http version



99% of http versions are HTTP/1.1

THANK YOU

Q & A