



# Ultimate Technologies

Part I

Exploratory Data Analysis :

Exploration of login time series data



# Exploratory Data Analysis - Errors

93137	1970-04-13 18:50:19
93138	1970-04-13 18:43:56
93139	1970-04-13 18:54:02
93140	1970-04-13 18:57:38
93141	1970-04-13 18:54:23

- This is our login data with the following issues
- No column for aggregation (number of logins)
- Time stamp not indexed
- Note, all the way on the left the numerical index is in order
- The time portion of the time stamp all the way on the right is not in sequence

# Exploratory Data Analysis - Errors

Each row is a login, so we can count it as one and add a column for aggregation

	login_time	logins
0	1970-01-01 20:13:18	1
1	1970-01-01 20:16:10	1
2	1970-01-01 20:16:37	1
3	1970-01-01 20:16:36	1
4	1970-01-01 20:26:21	1
...	...	...
93137	1970-04-13 18:50:19	1
93138	1970-04-13 18:43:56	1
93139	1970-04-13 18:54:02	1
93140	1970-04-13 18:57:38	1
93141	1970-04-13 18:54:23	1
93142 rows × 2 columns		

# Exploratory Data Analysis – Data Wrangling

We sort the values so that they are in sequential order

	login_time	logins
93125	1970-04-13 18:16:48	1
93126	1970-04-13 18:18:26	1
93129	1970-04-13 18:35:43	1
93132	1970-04-13 18:36:53	1
93130	1970-04-13 18:36:55	1
93133	1970-04-13 18:40:31	1
93135	1970-04-13 18:40:40	1
93131	1970-04-13 18:43:19	1
93138	1970-04-13 18:43:56	1
93134	1970-04-13 18:46:06	1
93136	1970-04-13 18:48:52	1
93137	1970-04-13 18:50:19	1
93139	1970-04-13 18:54:02	1
93141	1970-04-13 18:54:23	1
93140	1970-04-13 18:57:38	1

# Exploratory Data Analysis – Data Wrangling

- Set time stamp as Index

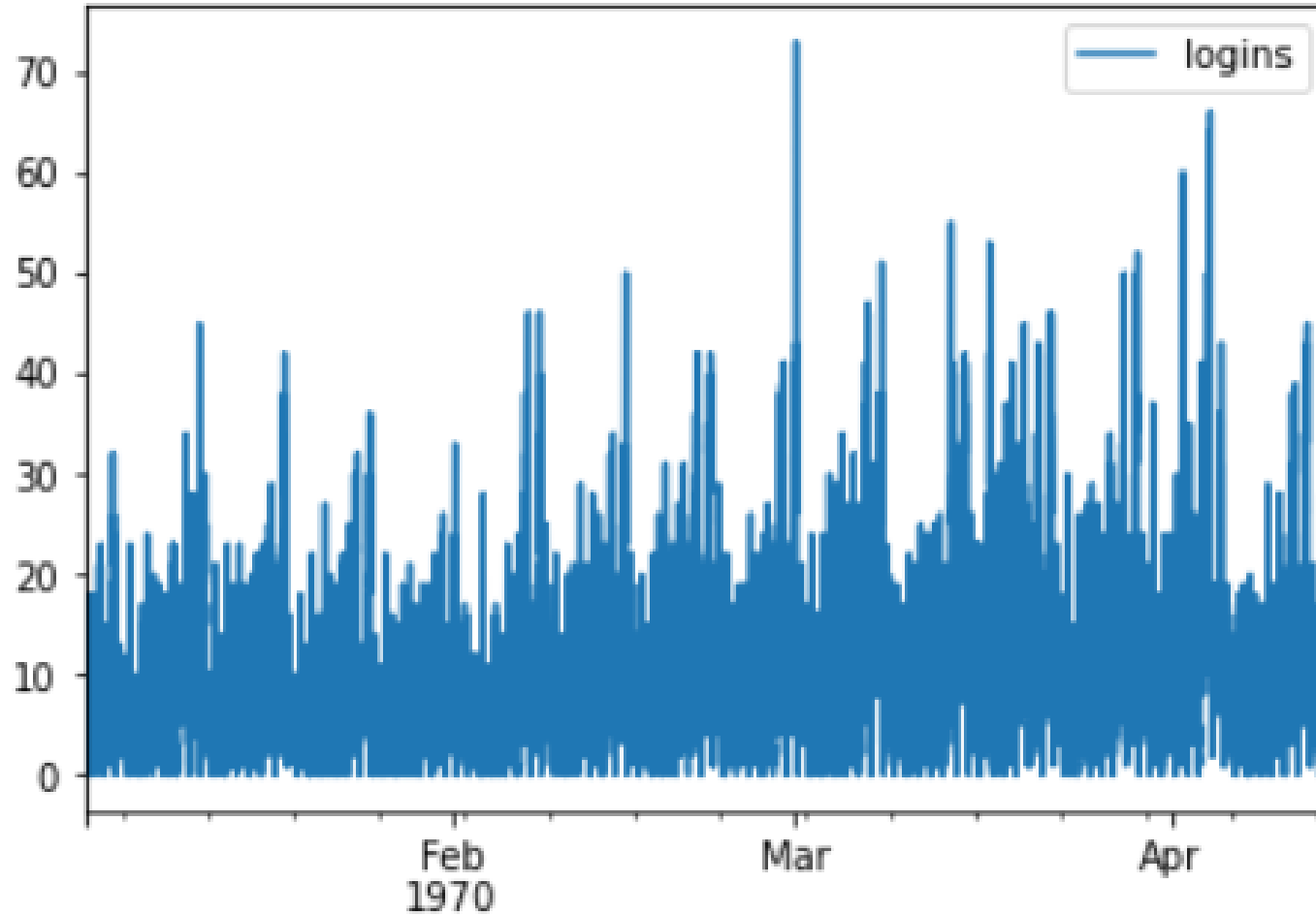
logins	
login_time	
1970-01-01 20:12:16	1
1970-01-01 20:13:18	1
1970-01-01 20:16:10	1
1970-01-01 20:16:36	1
1970-01-01 20:16:37	1
1970-01-01 20:21:41	1
1970-01-01 20:26:05	1
1970-01-01 20:26:21	1
1970-01-01 20:31:03	1
1970-01-01 20:34:46	1

- Now we can aggregate the logins in 15-minute intervals

logins	
login_time	
1970-04-13 15:15:00	4
1970-04-13 15:30:00	3
1970-04-13 15:45:00	2
1970-04-13 16:00:00	7
1970-04-13 16:15:00	8
1970-04-13 16:30:00	4
1970-04-13 16:45:00	3
1970-04-13 17:00:00	5
1970-04-13 17:15:00	3
1970-04-13 17:30:00	9

# Exploratory Data Analysis

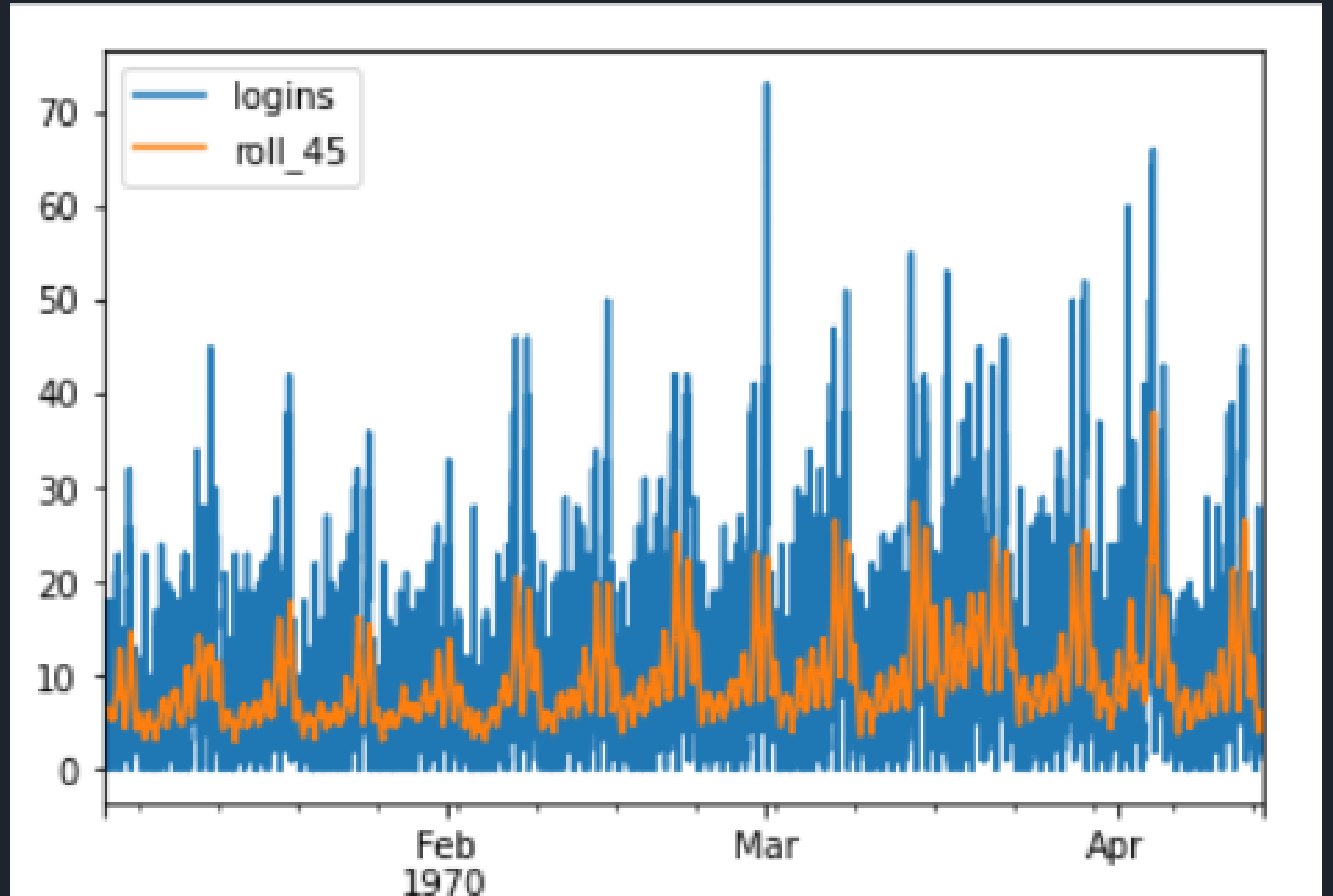
- Not very easy to see a pattern
- This is around 2 months and a half



# Exploratory Data Analysis

Apply a 45-day rolling mean on top of the graph. Now we can see a pattern with valleys perhaps at the beginning of the week and peaks on the weekend. Increases from month to month with the highest peak in April.

Let's zoom into each month.

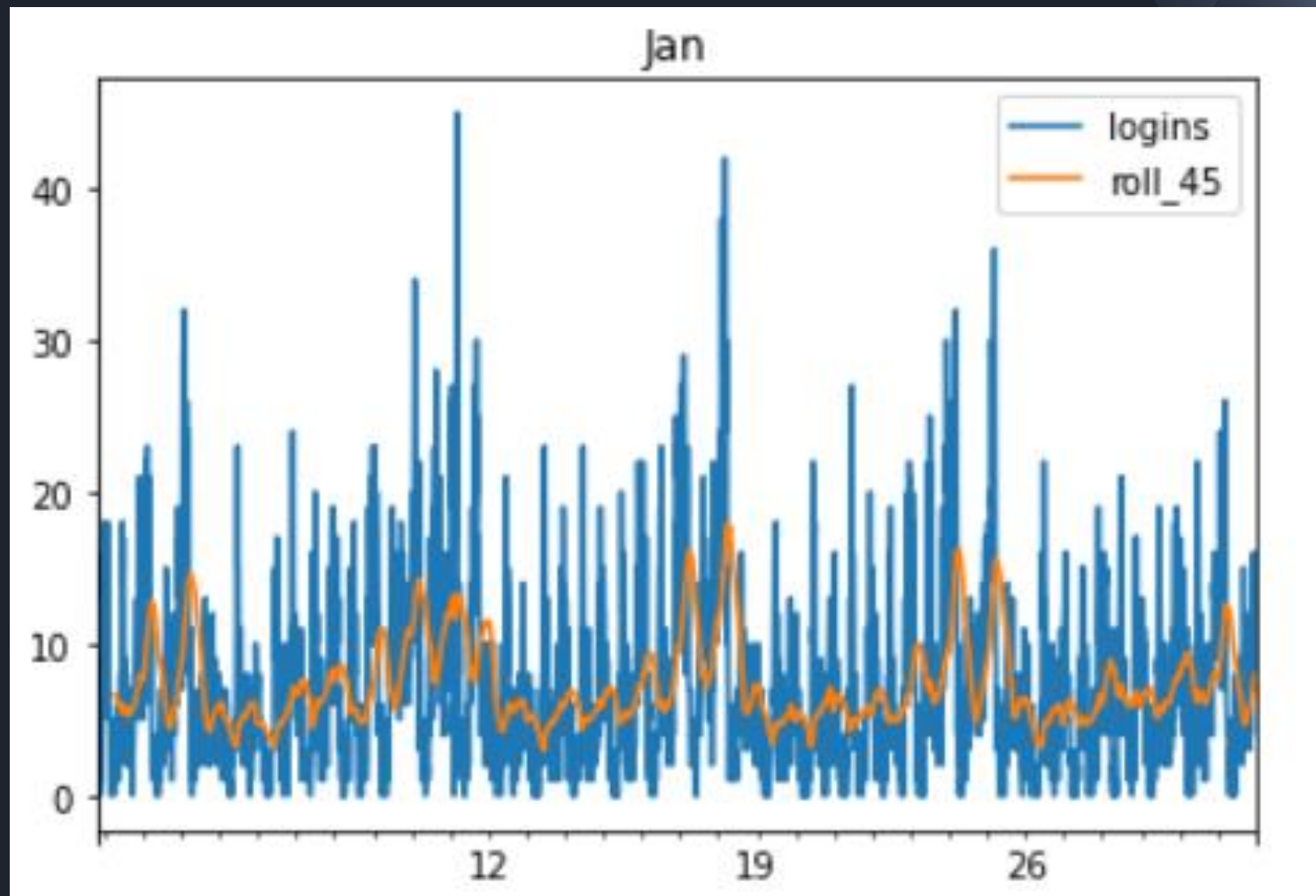


# Exploratory Data Analysis

Zoom into January 1970:

We clearly see a pattern there.

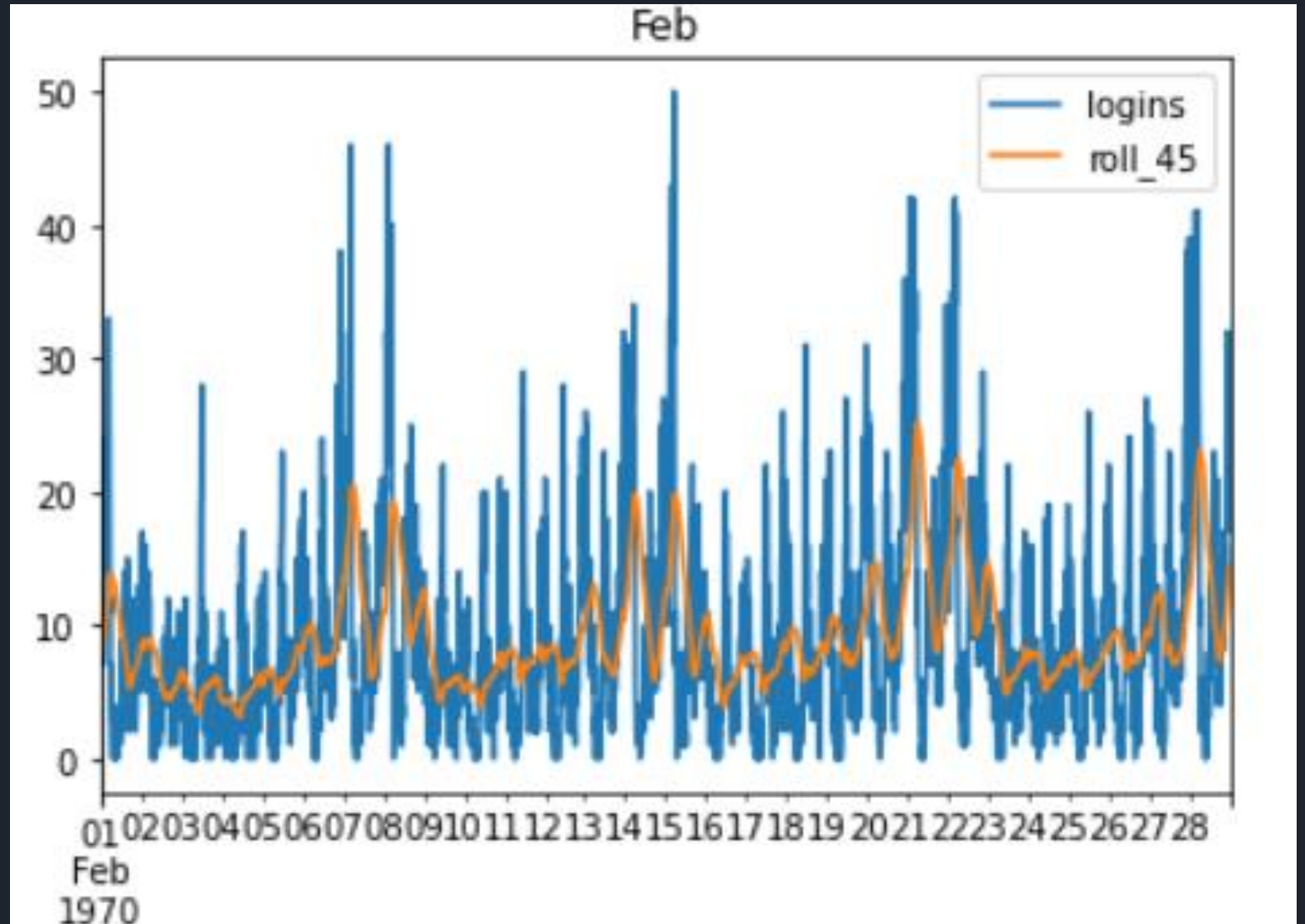
We confirm a 7-day pattern, valleys on the weekdays and peaks close to the weekend, note the peaks are around 15 logins, we'll inspect even closer but let's compare months for similarities.





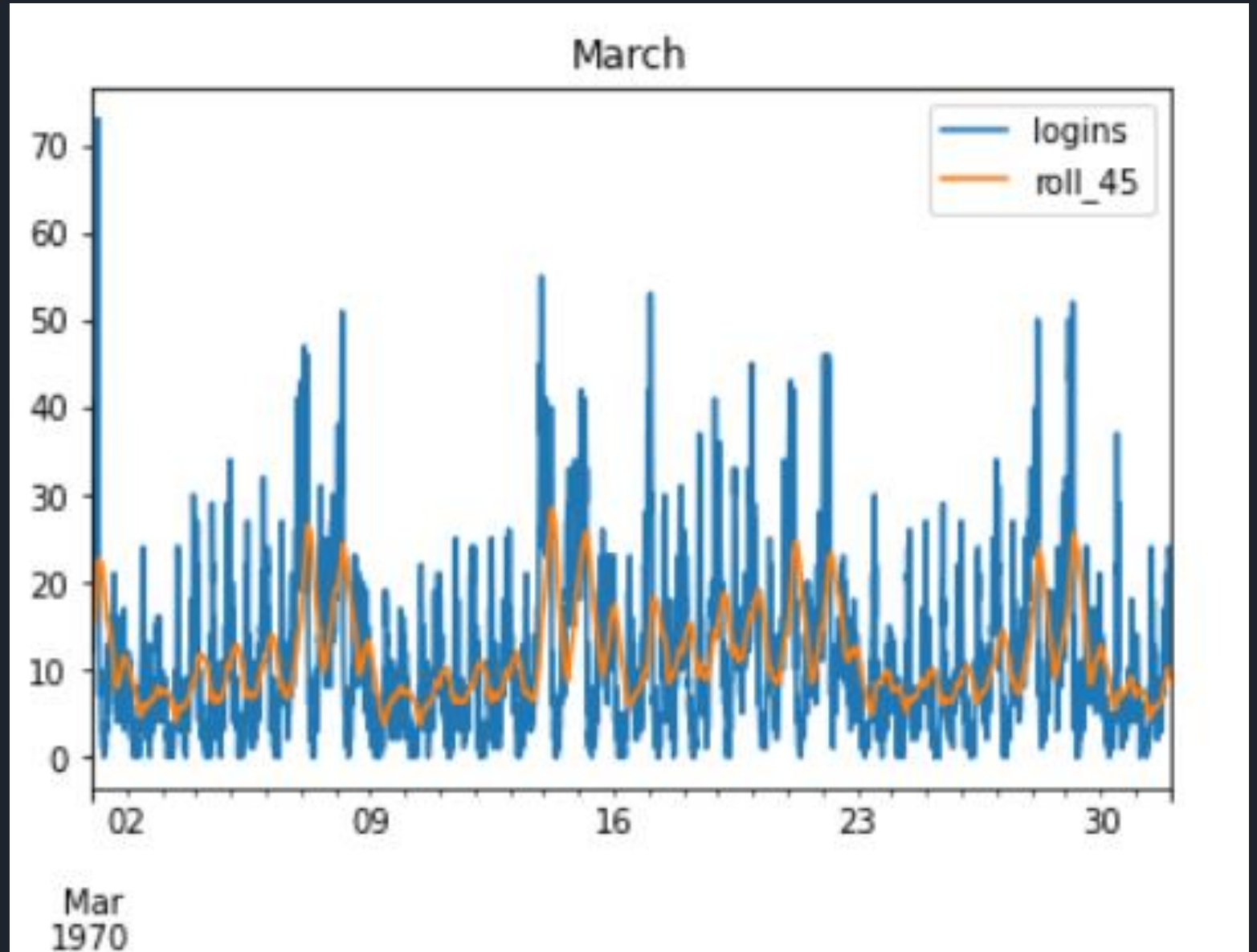
# Exploratory Data Analysis

February: Similar pattern but the rolling mean is overall higher, note the peaks are around 20 logins (more than 20 on the last 2 weeks).



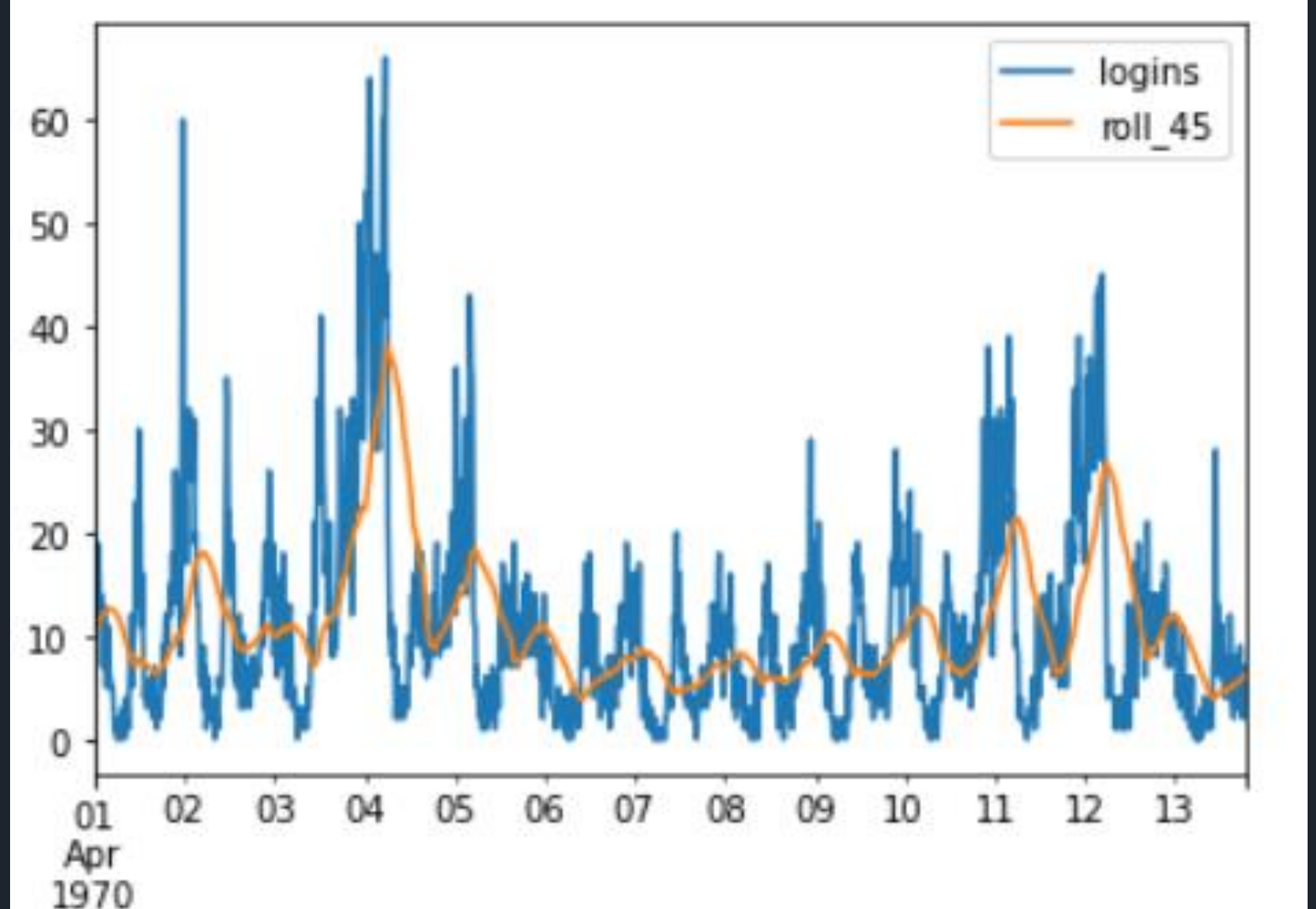
# Exploratory Data Analysis

March: Peaks all above 20. same pattern.



# Exploratory Data Analysis

April: we have a significant surge on logins on that first week peak, the rolling mean seems to be around 38. The data only goes up to the 13<sup>th</sup> of the month.



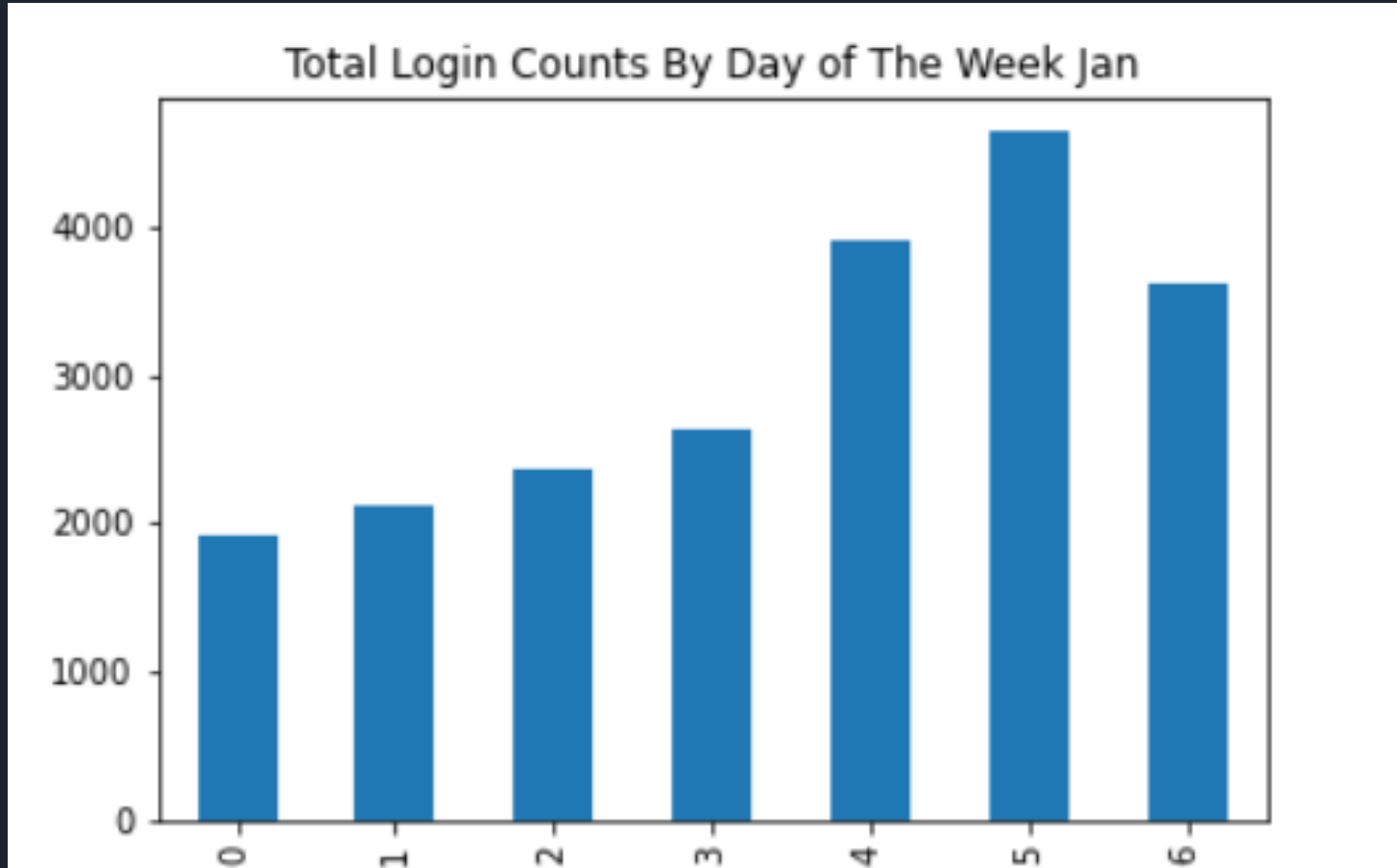
# Exploratory Data Analysis

January 1970: what are the total logins by day of the week?

Looks like Friday, Saturday and Sunday  
Accumulated the most logins over the month

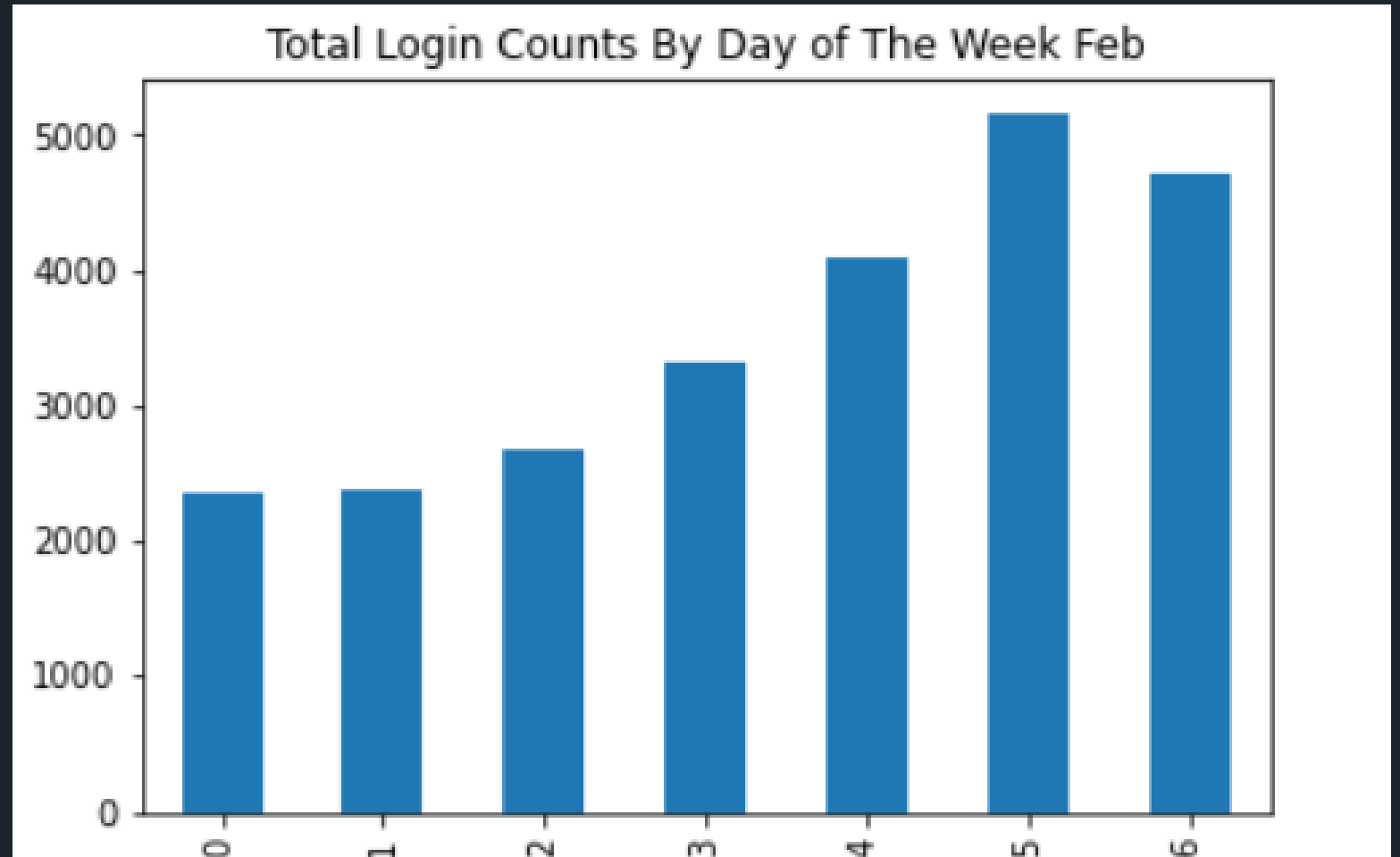
The logins counts gradually increase from  
the beginning of the week and peak at the  
end

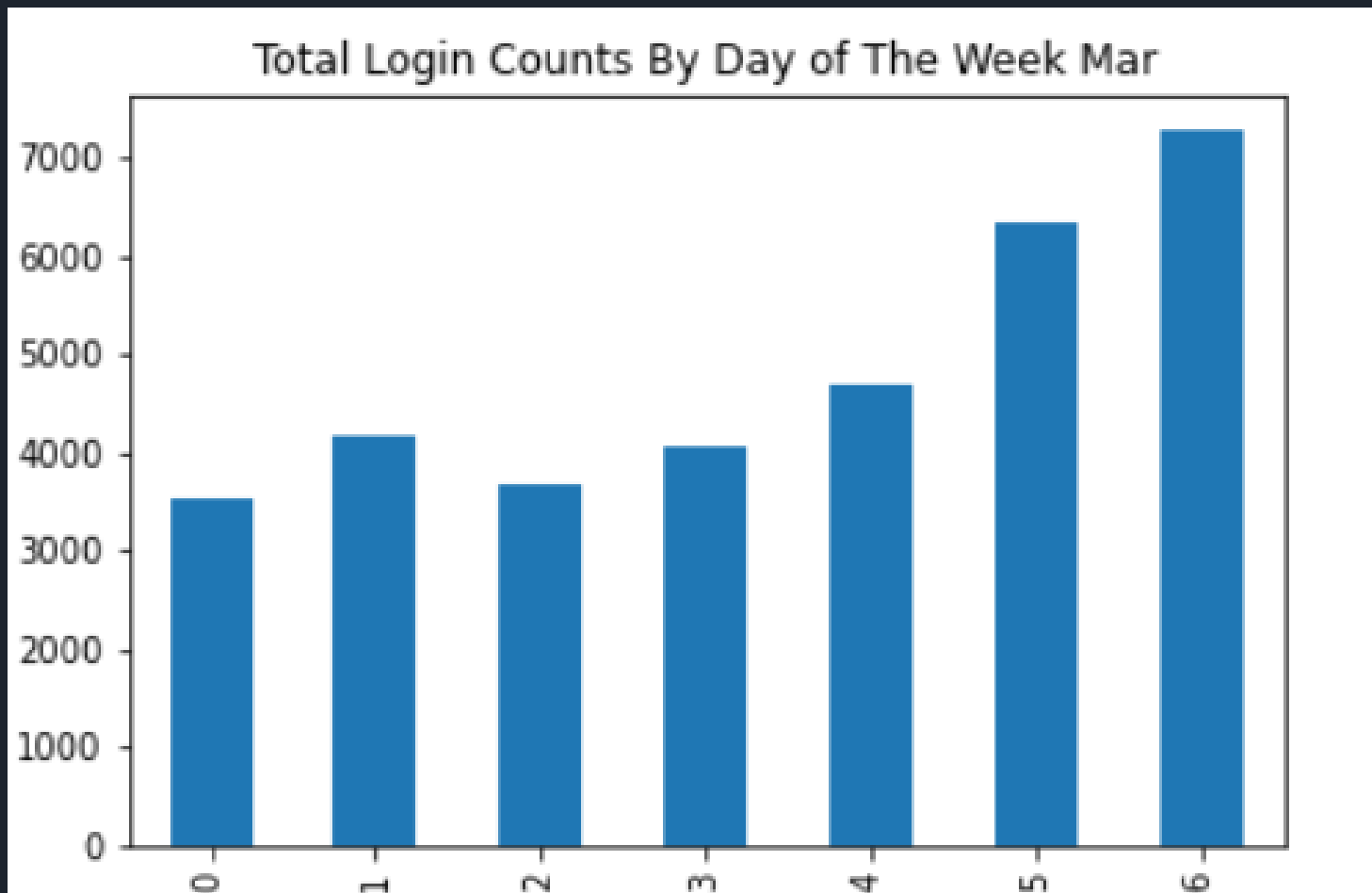
Let's have a look at the other months



# Exploratory Data Analysis

Feb: similar pattern





# Exploratory Data Analysis

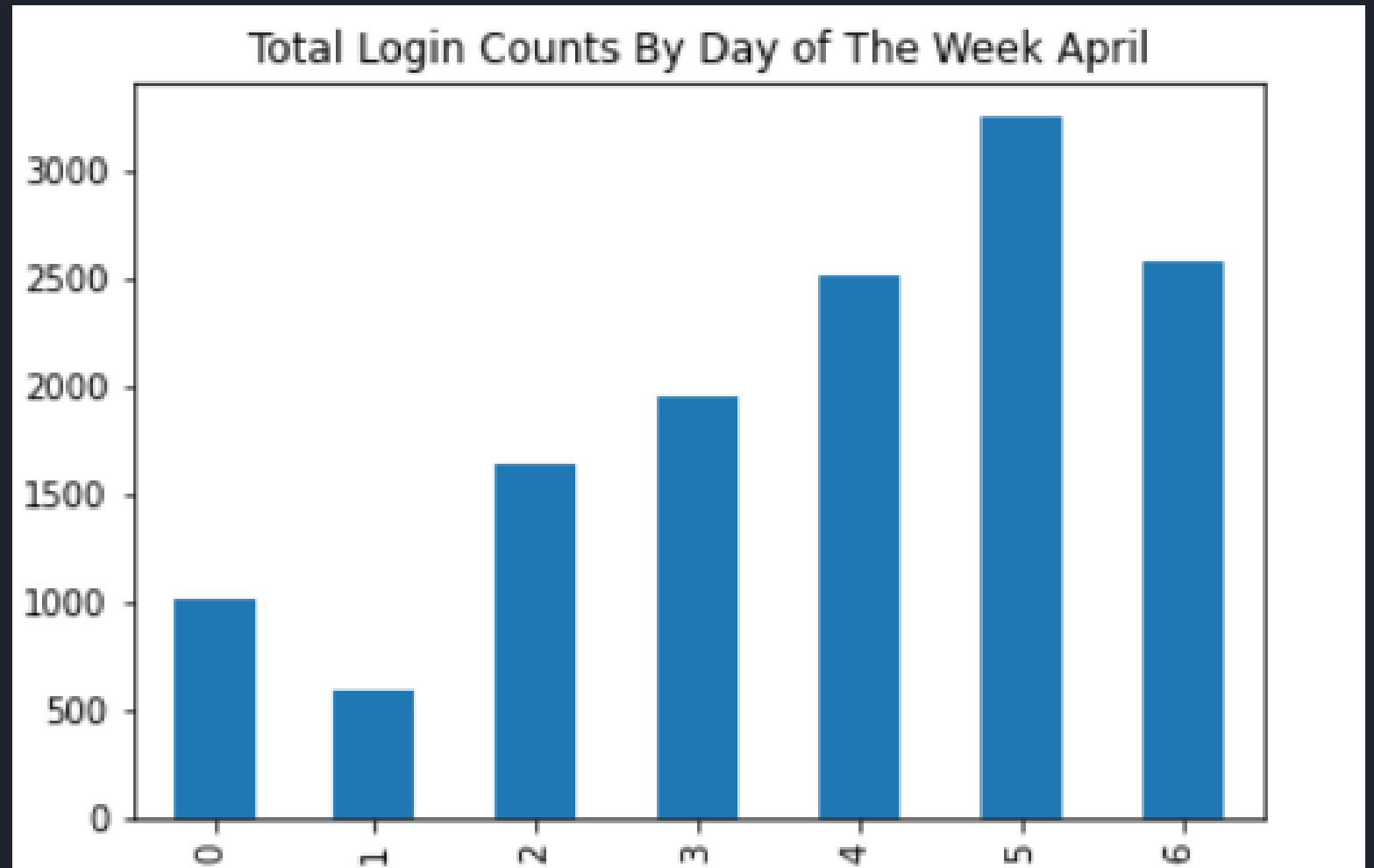
March: although Friday, Saturday and Sunday are still the highest, Tuesday is getting close to Friday here, so there's a bit of a bump up on Tuesday logins and Sunday logins

# Exploratory Data Analysis

April: but remember we only had up to the 13<sup>th</sup>

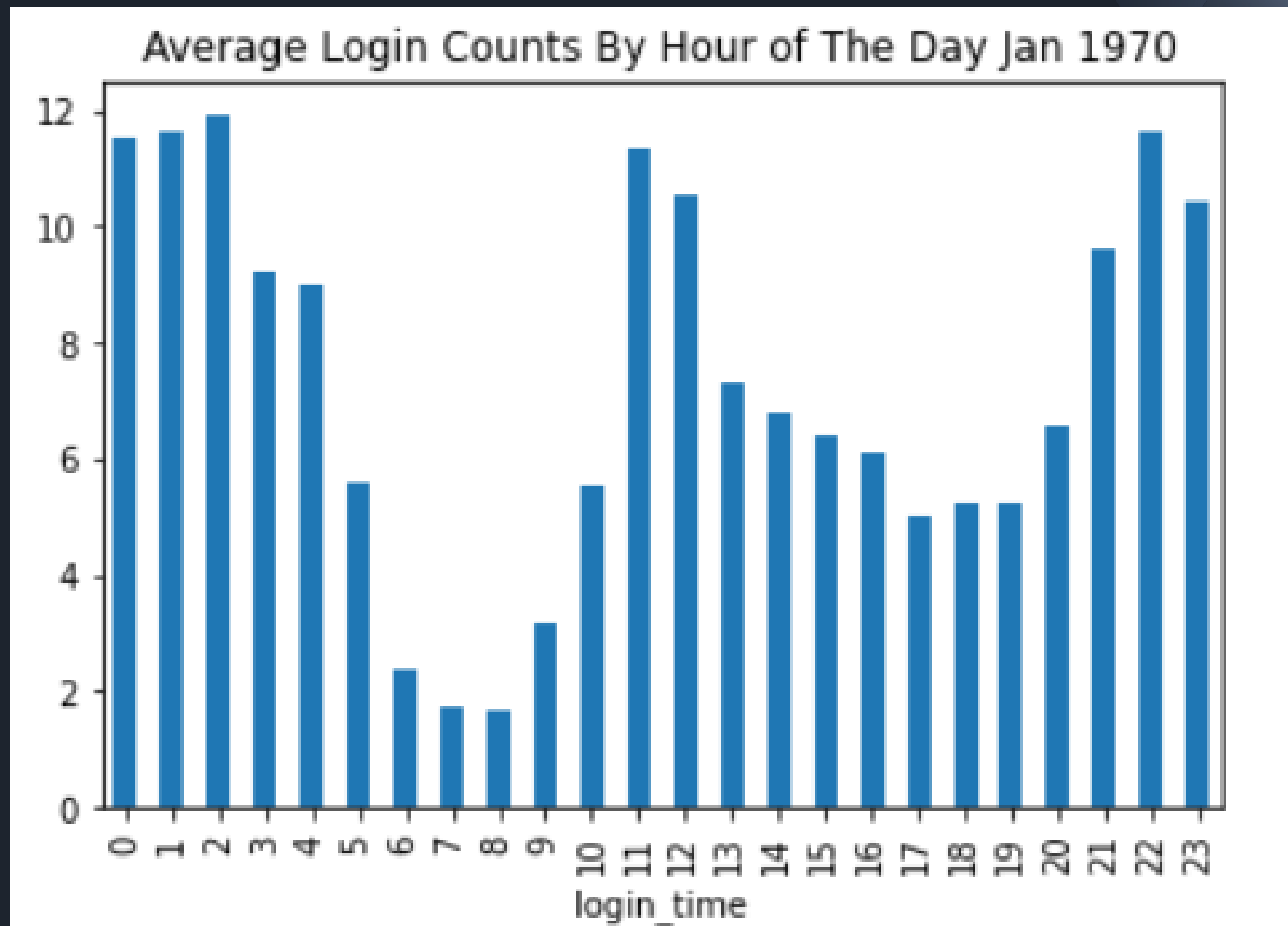
We see a dip on Tuesday and a bump up on Saturday

It looks like the days that accumulated the most counts overall on all months are Friday Saturday and Sunday but at what times of the day?



# Exploratory Data Analysis

- We established the highest total counts are seen on Friday, Saturday and Sunday
- Let's look at Average logins by hour of the day for Jan. We'll focus on peak times.
- Peak avg times seem to have three clusters
- 12 am to 2am
- 11am to 12 pm
- 9pm to 11pm

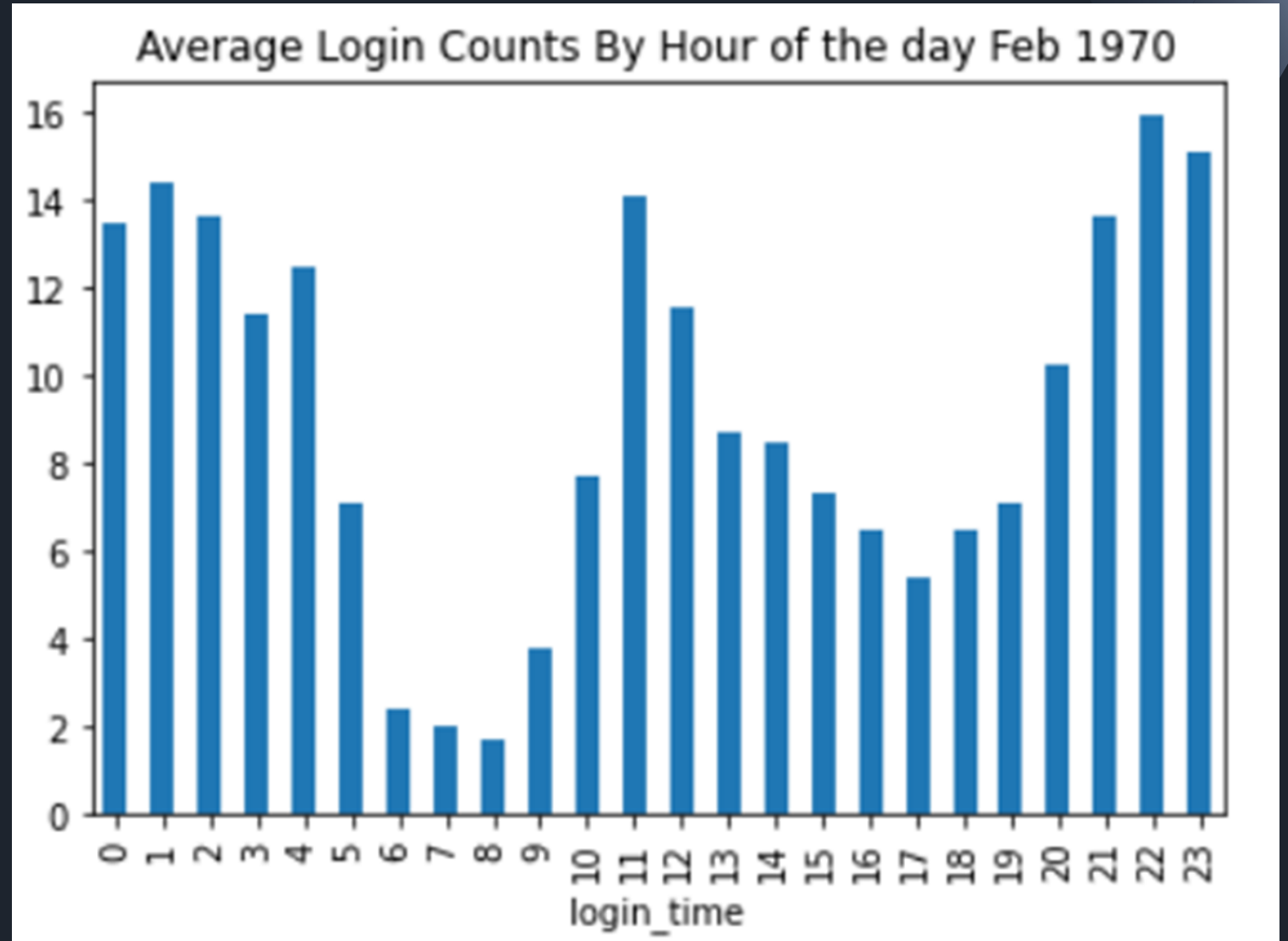


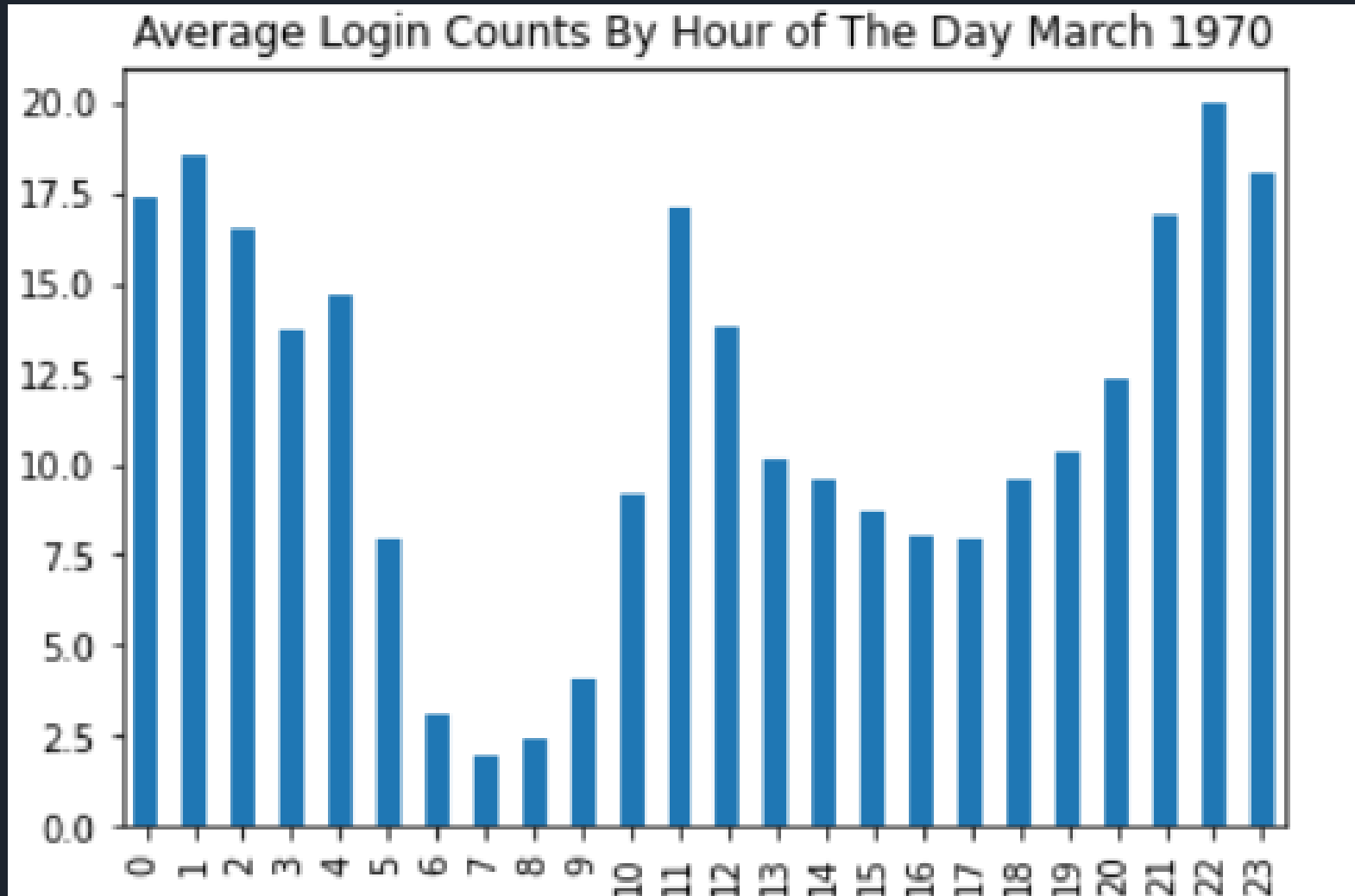


# Exploratory Data Analysis

Feb peak Avg times

- 12am to 2 am
- 11 am to 12pm
- 9 p to 11 pm





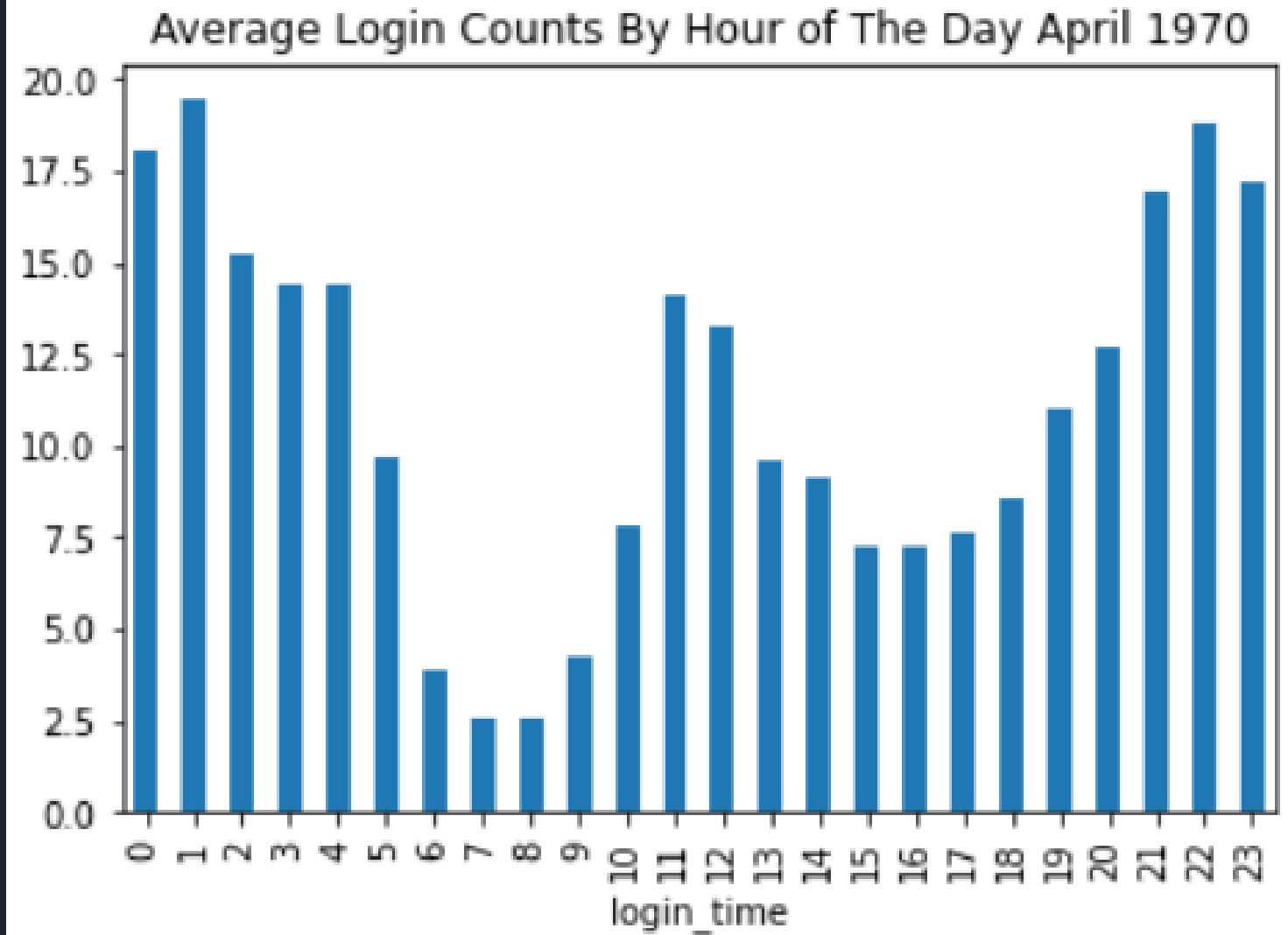
# Exploratory Data Analysis

March has similar peak avg logins



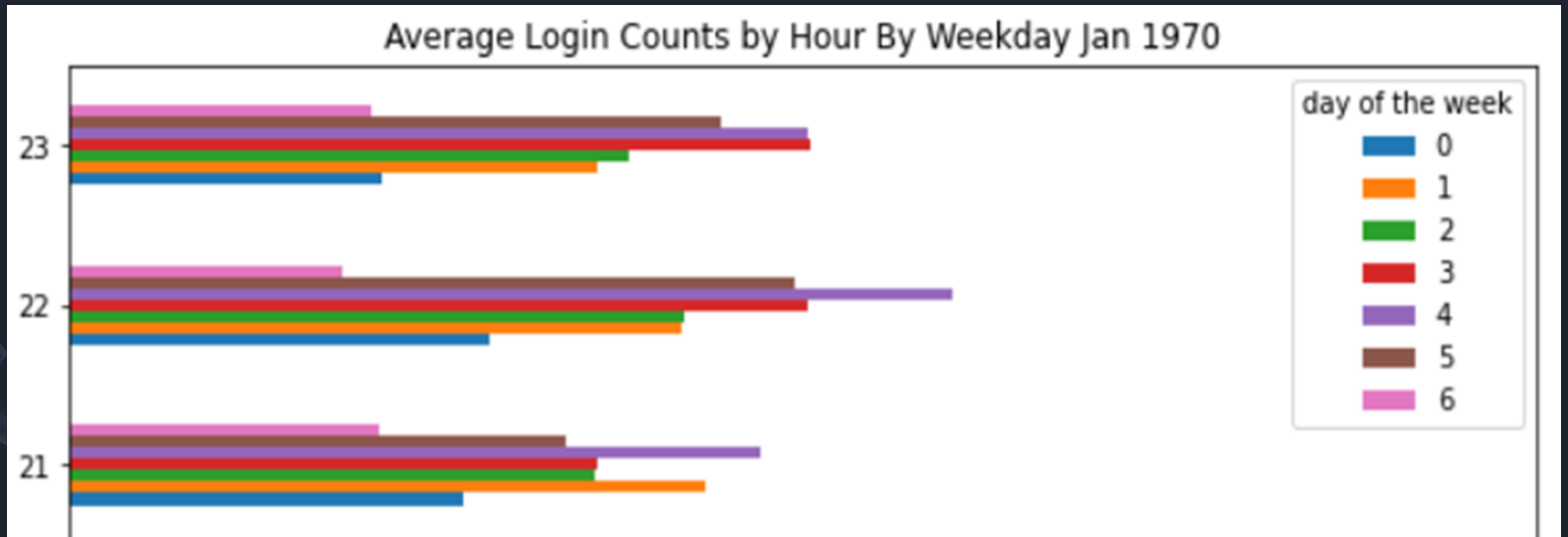
# Exploratory Data Analysis

April also has similar pattern.



# Exploratory Data Analysis

Finally, let's explore deeper and look at average logins by day of the week, by hour and for each month and focus on peak average times. Military time is displayed on the left. Color is the day of the week. The X axis is average logins. A chart has been prepared in the following slide as a nice data summary since these charts are too long would be easier to read in the Jupyter Notebook that accompanies these slides since it can scroll down length wise. This is mainly to show how data was broken down.



# Exploratory Data Analysis

Here is the chart, showing the highest average logins as the day or simply “not a peak time” for valleys (lower values).

Note the difference on which days at which hours we have the highest average logins.

This is different depending on the month.

Particularly note the time brackets from 12am to 5am and 11am to 12pm and compare by month.

Highest Average Logins by Hour by Weekday

Times (Military)	Jan	Feb	March	April
0	Friday, Sunday	Saturday	Saturday	Saturday
1	Sunday	Saturday	Saturday	Saturday
2	Sunday	Saturday	Saturday	Saturday
3	Sunday	Sunday	Saturday	Saturday
4	Sunday	Sunday	Sunday	Sunday
5	Sunday	Sunday	Sunday	Saturday
6	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
7	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
8	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
9	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
10	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
11	Tuesday	Thursday	Wednesday	Wednesday
12	Friday	Friday	Thursday	Friday
13	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
14	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
15	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
16	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
17	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
18	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
19	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
20	Not A Peak Time	Not A Peak Time	Not A Peak Time	Not A Peak Time
21	Friday	Friday	Friday	Friday
22	Friday	Friday	Friday	Friday
23	Tuesday, Friday	Friday	Friday	Friday

# Summary

- Most logins occur on Friday, Saturday and Sunday
- On average most logins occur at certain times
  - 12 am to 5 am
  - 11 am to 12 pm
  - 9 to 11 pm
- The times when most logins occur on certain days and certain hours are not always the same depending on the month
- There were additional peak times on Saturday and Sunday that were picked up by aggregating by hour and weekday that were not picked up by aggregating by hour.
- One finding not noticeable from the previous chart, but visible here on the right from around 12-5am, is where Saturday has the highest average logins by a large margin in April

