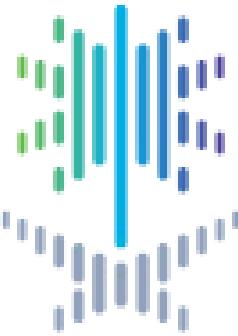


الهيئة العامة للإحصاء  
المملكة العربية السعودية



# SOCIAL MOOD ON ECONOMY INDEX

---

Abdullah Alqahtani | 439051203 •

# Outline

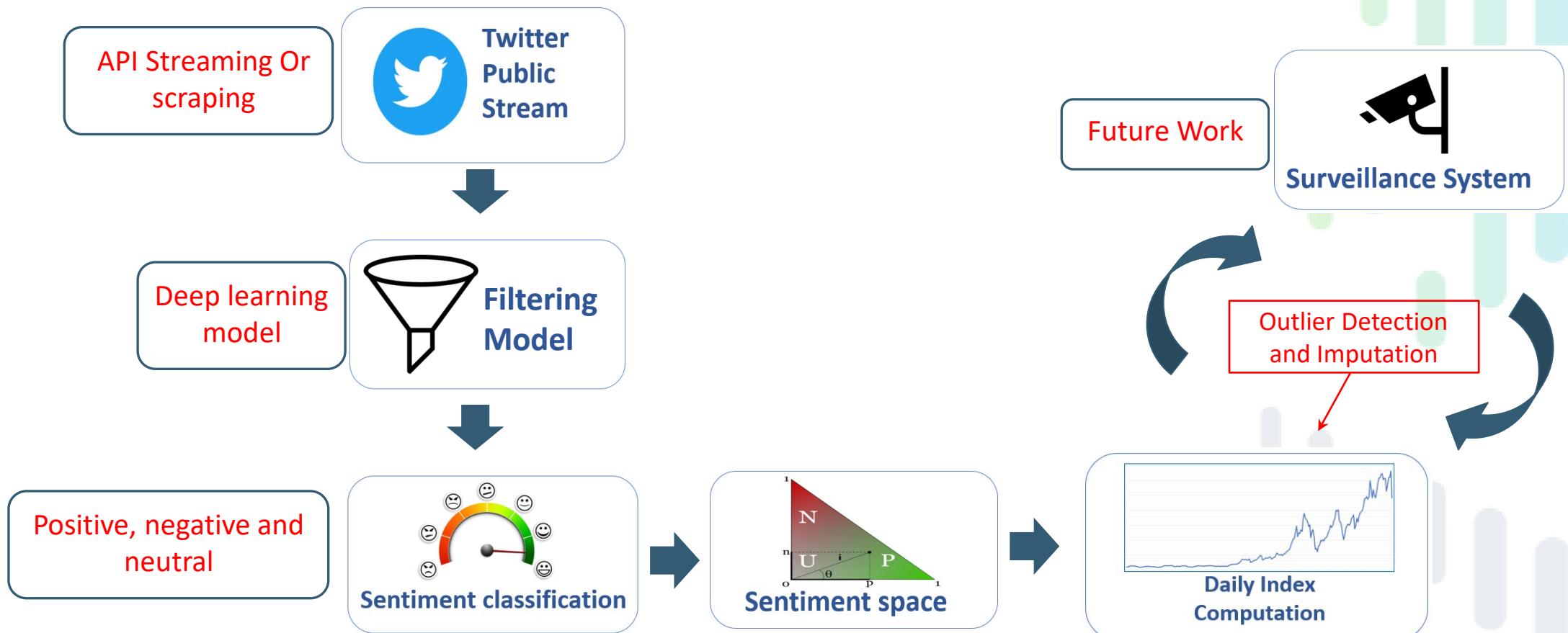
- **Introduction**
- **Pipeline processing**
- **Data collection : stream & scraping**
- **AI filtering model**
- **Sentiment classification model**
- **Sentiment space**
- **Visualization**
- **Future work**



## Objective

- Collecting data containing Saudi tweets with locations.
- Building a deep learning model to extract economy tweets only
- Building a deep leaning sentiment analysis model to classify data
- Measuring Saudi citizens and residents satisfaction about the economy
- Plotting social moods on the economy index

# Pipeline constructing





Twitter  
Public  
Stream

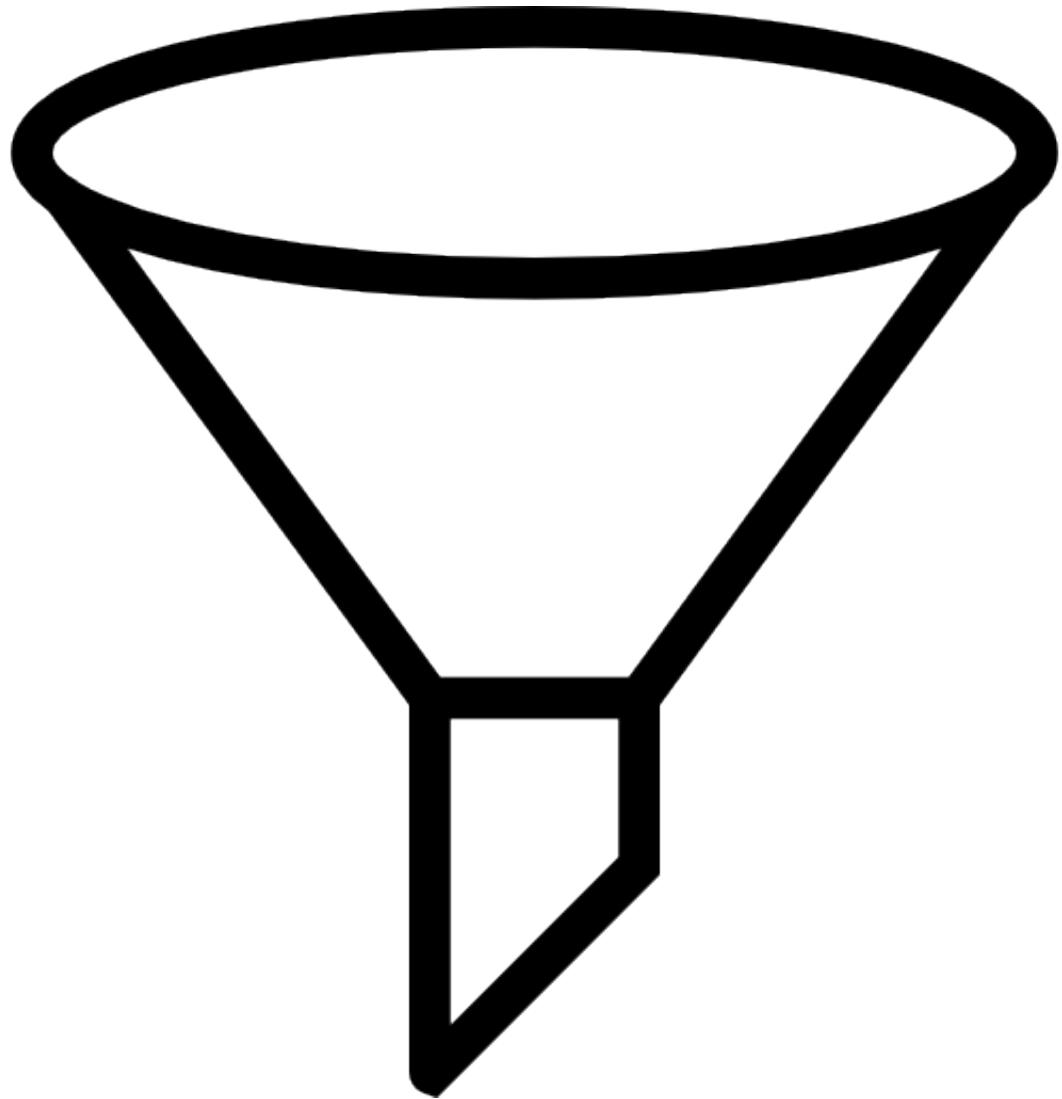


## Data Collection: Streaming/scraping

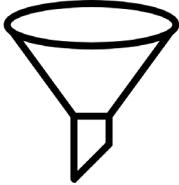
- Connecting to Twitter through Twitter API and filtering by Saudi Arabia cities coordinates

like follow sample....

Coordinates	Place	Tweet_time	Tweet
[[45.52122, 25]]	روضة سدير, المملكة العربية السعودية	2022-07-25 17:	الحمدلله على نعمة الاتحاد
[[42.375135, 1]]	بيشة, المملكة العربية السعودية	2022-08-03 01:	لبي ذا الديره ومن سكن فيها . ❤️❤️❤️❤️❤️❤️❤️❤️
[[44.605401, 2]]	الزلفي, المملكة العربية السعودية	2022-07-29 20:	لا حول ولا قوة الا بالله النصر مشاكله و هو شاته حتى بالوديات يبون المها او م
[[38.929659, 2]]	جدة, المملكة العربية السعودية	2022-08-03 21:	اغنية اللحظة " وش تبي " راشد الماجد □
[[49.416654, 2]]	الاحساء, المملكة العربية السعودية	2022-07-22 10:	متى أصير غني عشان اقول كلمة ترانزيت ، مليت من كلمة محطة بائزين

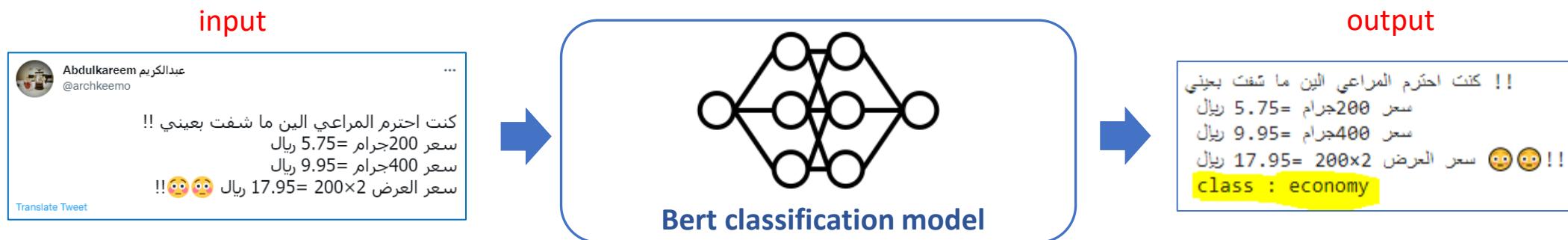


# Filtering Model



## Filtering model

- The most accurate deep learning model until now is The Arab-BERT model so we use it to classify data into 5 category [sport , religion , politic , economy]
- Than we going to take economy tweet only



- Note : the accuracy of this model is good but we have to increase it by fine tuning the model with our data



# Sentiment classification

# Sentiment classification

- In sentiment classification we use The Bert-Base-Arabic model as one of the best models in classification for 3 labels [negative, neutral , positive]



# Some sample of Sentiment classification

input

{ فهد زعلان من سعر كوب القهوة في ستار بوكس }

{ عروض الجمعة الصفراء في متجر نون لا تفوتك }



output

{ 'label': 'negative'  
'score': 0.6068520545959473 }

{ 'label': 'neutral'  
'score': 0.8241519927978516 }



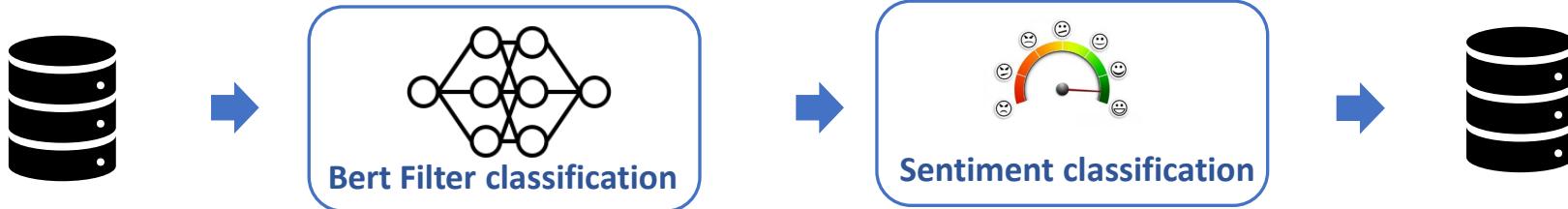
{ الاقتصاد السعودي اقتصاد قوي ومتين و الدليل جائحة  
كورونا فارنا بالدول الاخرى و تعرف وش اقصد }

{ 'label': 'positive'  
'score': 0.8995946049690247 }

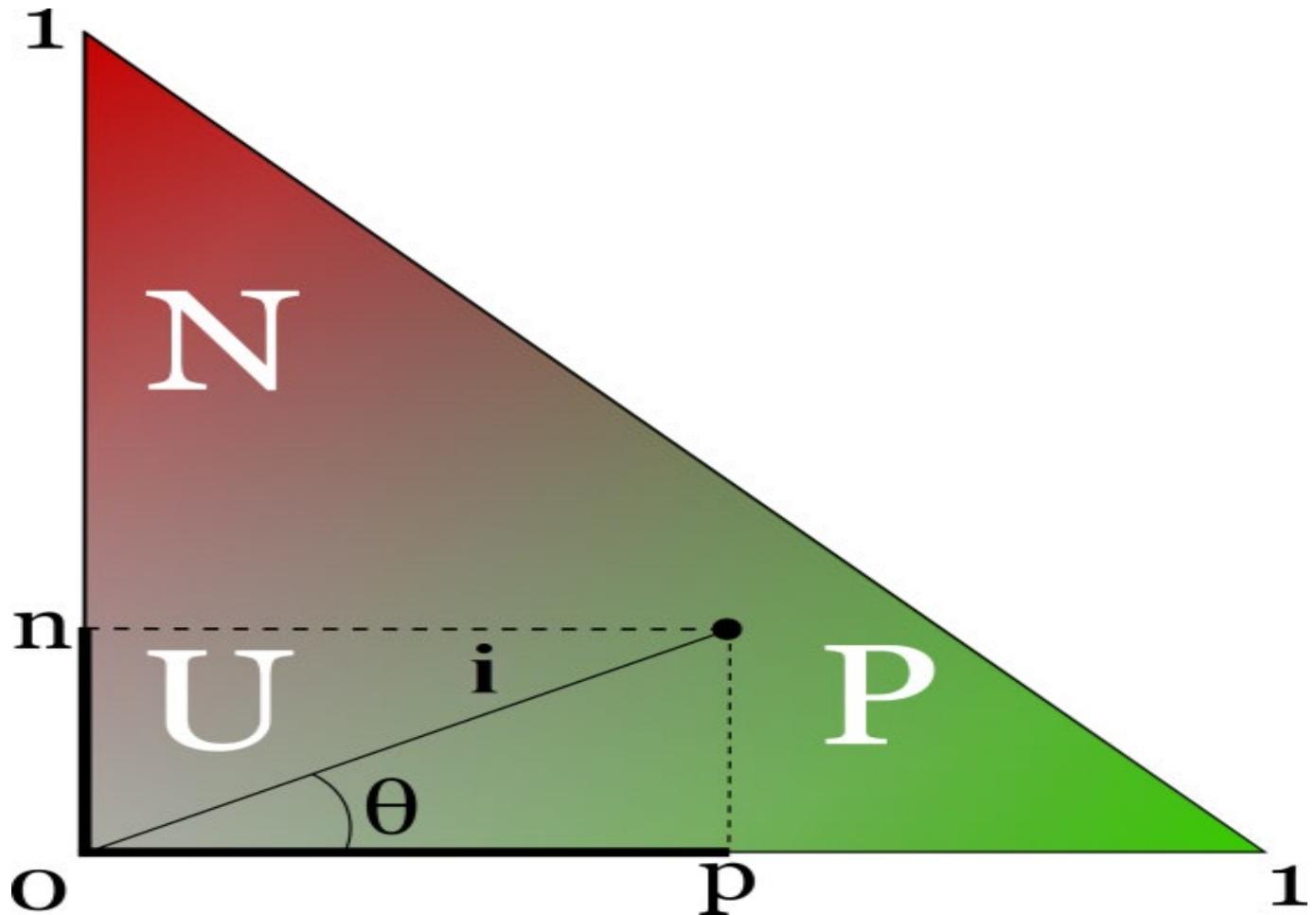
{ انخفاض ارباح سهم شركة اسمنت القصيم بنسبة 20  
بالمئة الله يعين حطيت فيها كل راس مالي }

{ 'label': 'negative'  
'score': 0.9897794127464294 }

# Data after double prediction



score	label	class	Coordinate	Place	Tweet time	Tweet
0.655076	positive	economy	[46.3043]	Riyadh, Kingdom of Saudi Arabia	7/20/2022	ماركة عطور سعودية خيل 🐾 و على ضمالي ❤️
0.974189	negative	economy	[46.3043]	Riyadh, Kingdom of Saudi Arabia	7/20/2022	عشان لا أحد يقول الموضوع شخصته ارتقى 100% مغلق
0.958232	negative	economy	[43.2591]	البخاري، المملكة العربية السعودية	7/20/2022	أنا مش شاهم أيه الإكتفاء الذاتي اللي أنا فيه دا أنا بمقتنش ط
0.894256	positive	economy	[46.3043]	الرياض, المملكة العربية السعودية	7/20/2022	برlad سعودي جنيد جوده عليه @Aldowayesh وياسعار مناسبة



# Sentiment space

# Sentiment space

- To calculate the daily index we use the ISTAT formula called “Sentiment space”

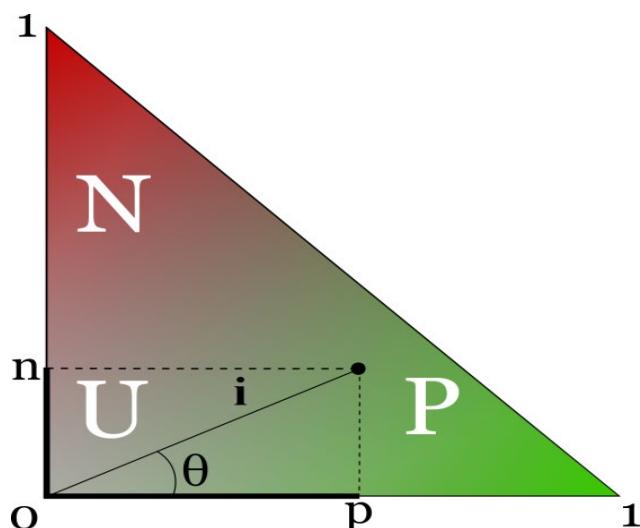
- In Sentix, positive (**p**) and negative (**n**) sentiment

- scores of lemmas are constrained as follows:

$$\begin{cases} p \in [0, 1] \\ n \in [0, 1] \\ p + n \leq 1 \end{cases}$$

- Therefore Sentix maps lemmas to points

- belonging to the sentiment triangle:



- From  $(p, n)$  coordinates we can pass to polar coordinates  $(i, \theta)$  and derive two *additional* sentiment scores:

✓ **Polarity**       $\omega = 1 - 4\theta/\pi$      $\omega \in [-1, 1]$

✓ **Intensity**       $i = \sqrt{p^2 + n^2}$      $i \in [0, 1]$

- This way Sentix lemmas are mapped to a **4D sentiment space**

- How do we use sentiment space values to calculate the index?

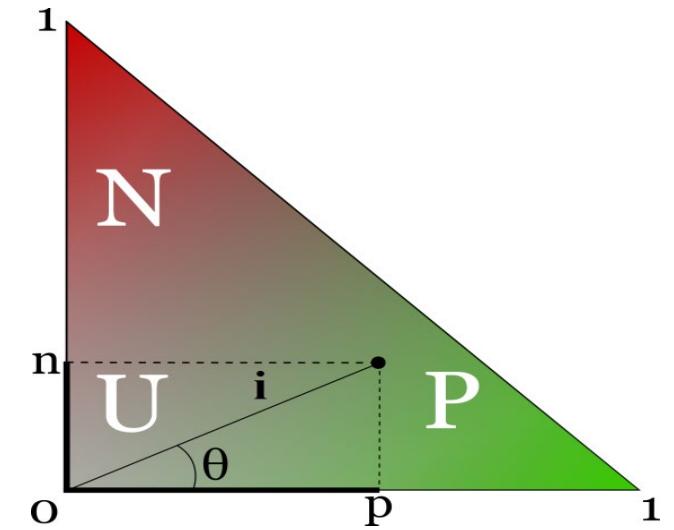


# Clustering and Calculation of the Index

- How do we use sentiment space valubles to calculate the index?



- Once sentiment scores ( $p, n, \omega, i$ ) are available for all the tweets of a daily block...
- Lastly we compute the **daily index value (S)**, which depends on the distribution of tweets within the Positive, Neutral and Negative classes



$$S = \bar{\omega}_i = \frac{\sum_t i_t \omega_t}{\sum_t i_t} = \frac{\sum_{t \in P} i_t \omega_t + \sum_{t \in N} i_t \omega_t}{\sum_t i_t} \quad \text{where } \omega_t \stackrel{\text{def}}{=} 0 \quad \forall t \in \text{Neutral}$$

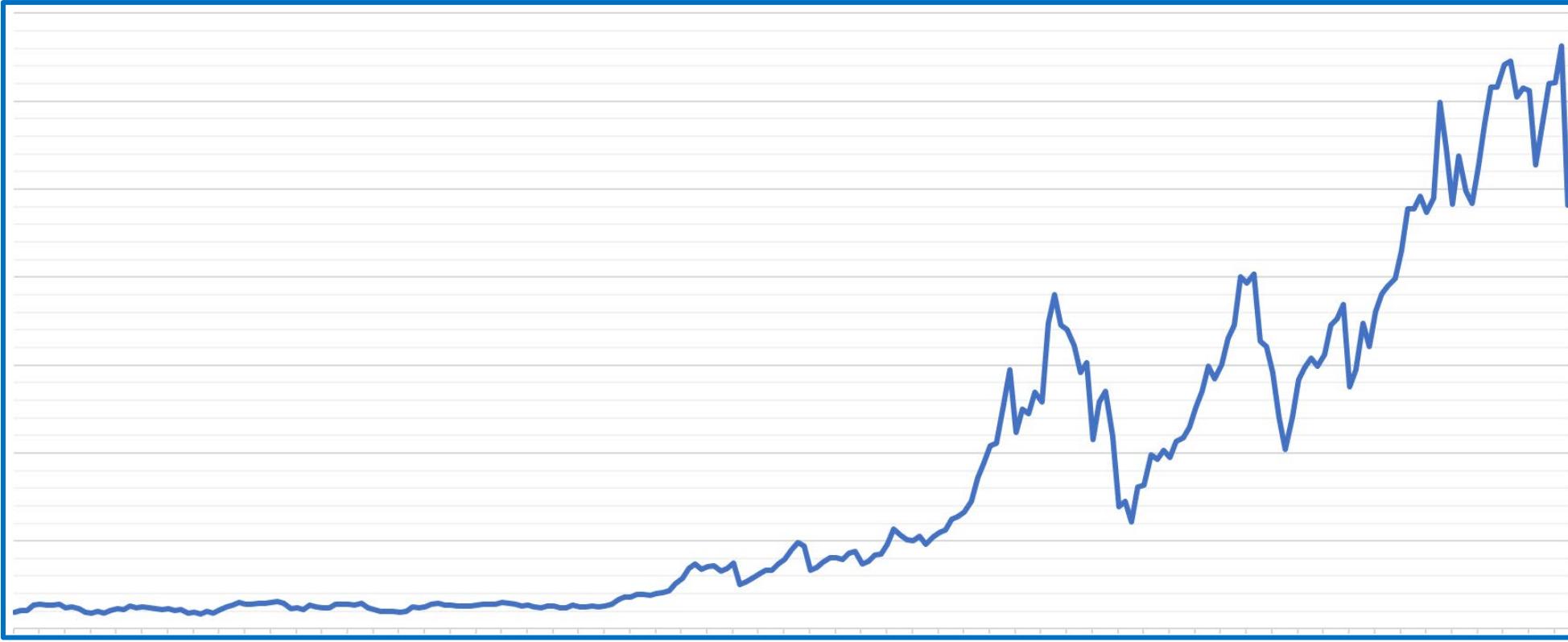
# Data after apply formula

$$S = \bar{\omega}_i = \frac{\sum_i i_t \omega_t}{\sum_i i_t} = \frac{\sum_{t \in P} i_t \omega_t + \sum_{t \in N} i_t \omega_t}{\sum_i i_t}$$

$$S = \bar{\omega}_i = \frac{\sum_t i_t \omega_t}{\sum_t i_t} = \frac{\sum_{t \in P} i_t \omega_t + \sum_{t \in N} i_t \omega_t}{\sum_t i_t}$$

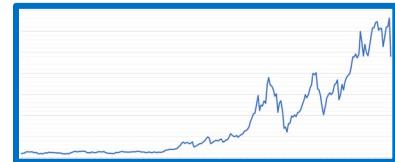
where  $\omega_t \stackrel{\text{def}}{=} 0 \quad \forall t \in \text{Neutral}$

daily_index	Polarity	theta	i	neg_weighted_average	Pos_weighted_average	Tweet_time
1.011527011	0.003138047	0.782934	0.272234	0.218497544	0.162389393	7/20/2022
0.793536377	-0.059303629	0.831975	0.287235	0.224779246	0.178824989	7/21/2022
1.081992616	0.023515797	0.766929	0.286804	0.253675911	0.133809477	7/22/2022
1.082782814	0.022962171	0.767364	0.277378	0.236118153	0.145557688	7/23/2022
0.922656147	-0.022174398	0.802814	0.286699	0.238567262	0.159002896	7/24/2022
1.030187751	0.00822903	0.778935	0.272595	0.221265259	0.159215924	7/25/2022
0.961025938	-0.010560784	0.793693	0.27097	0.209982893	0.171264957	7/26/2022
0.650576564	-0.104502827	0.867474	0.299072	0.230987237	0.189971214	7/27/2022
0.909550179	-0.025290291	0.805261	0.279606	0.222635025	0.169153865	7/28/2022
1.045223499	0.011744842	0.776174	0.259707	0.194337305	0.172280403	7/29/2022
0.816229324	-0.053219253	0.827196	0.289596	0.23249747	0.172657958	7/30/2022
1.081329481	0.022540597	0.767695	0.277152	0.235541863	0.146058369	7/31/2022
1.505255721	0.118114881	0.692631	0.233772	0.182463433	0.146139201	8/1/2022
1.264522926	0.068787202	0.731373	0.260042	0.218114651	0.141591308	8/2/2022
2.183020724	0.233907682	0.601687	0.197721	0.150543294	0.128180306	8/3/2022

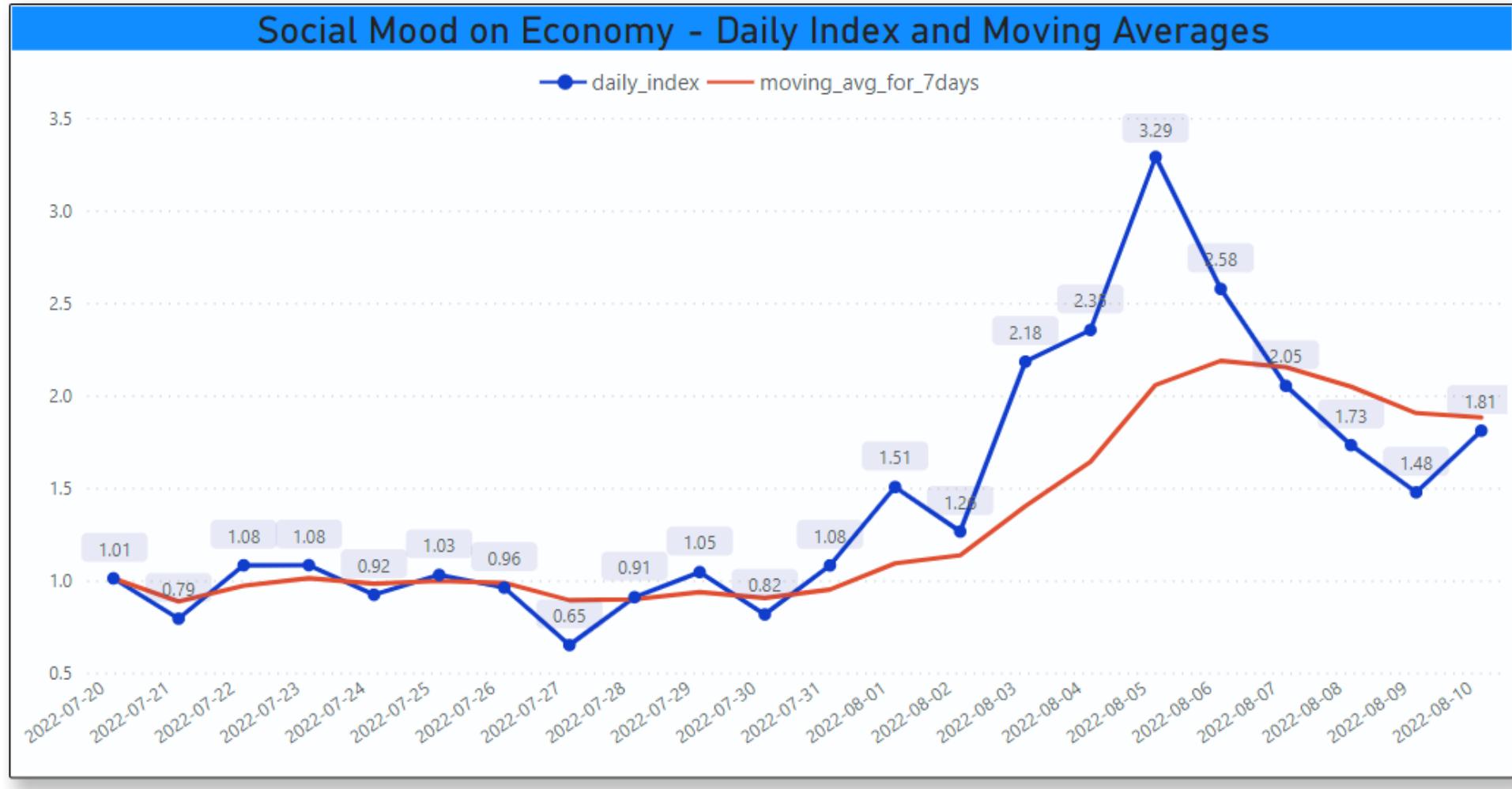


# Daily Index Computation

# Visualisation

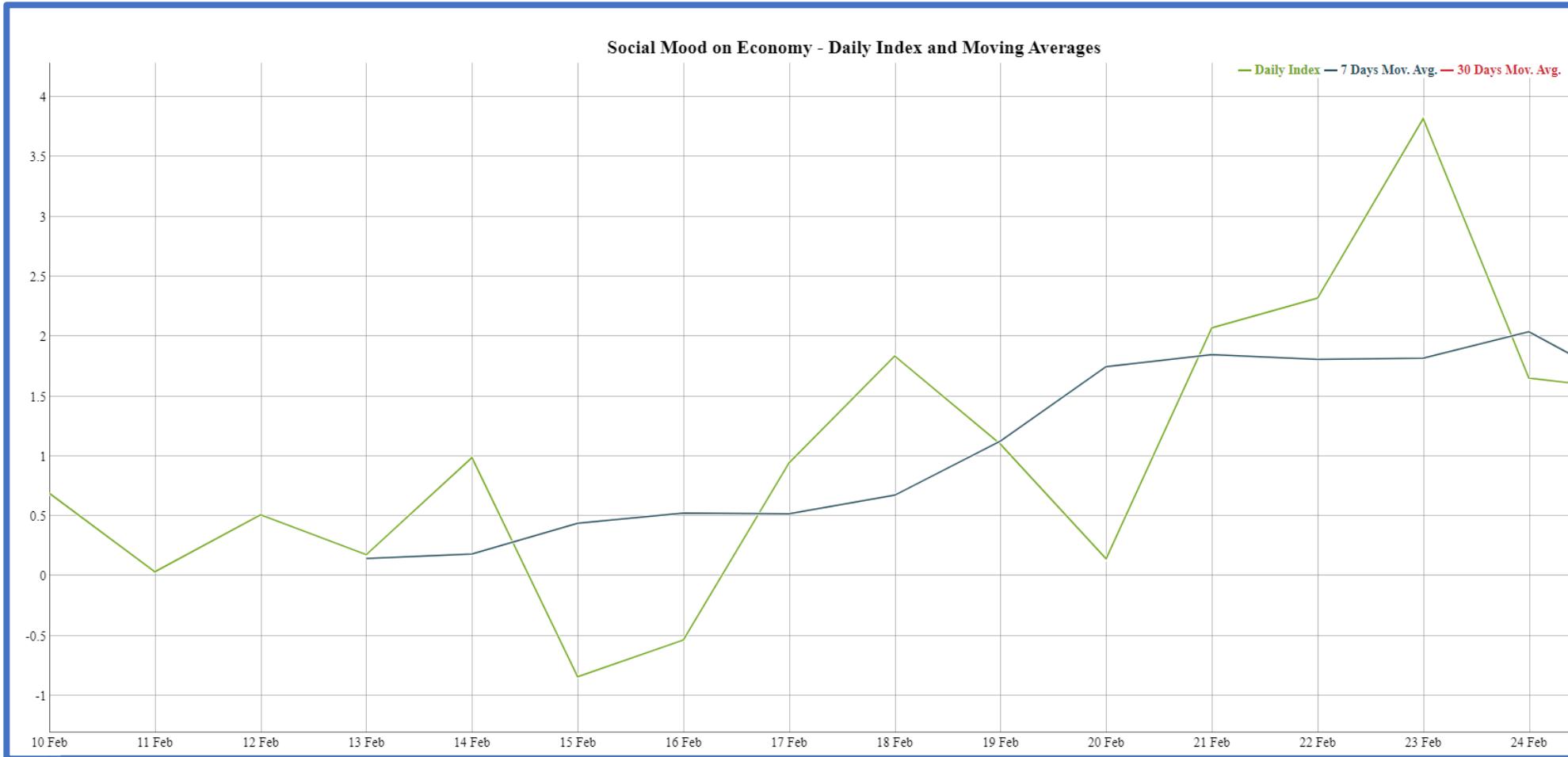


- There is a plot of 22 days of data collection

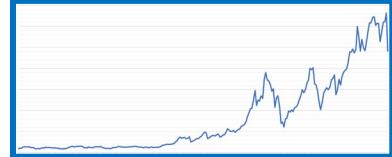


# Visualisation of ISTAT

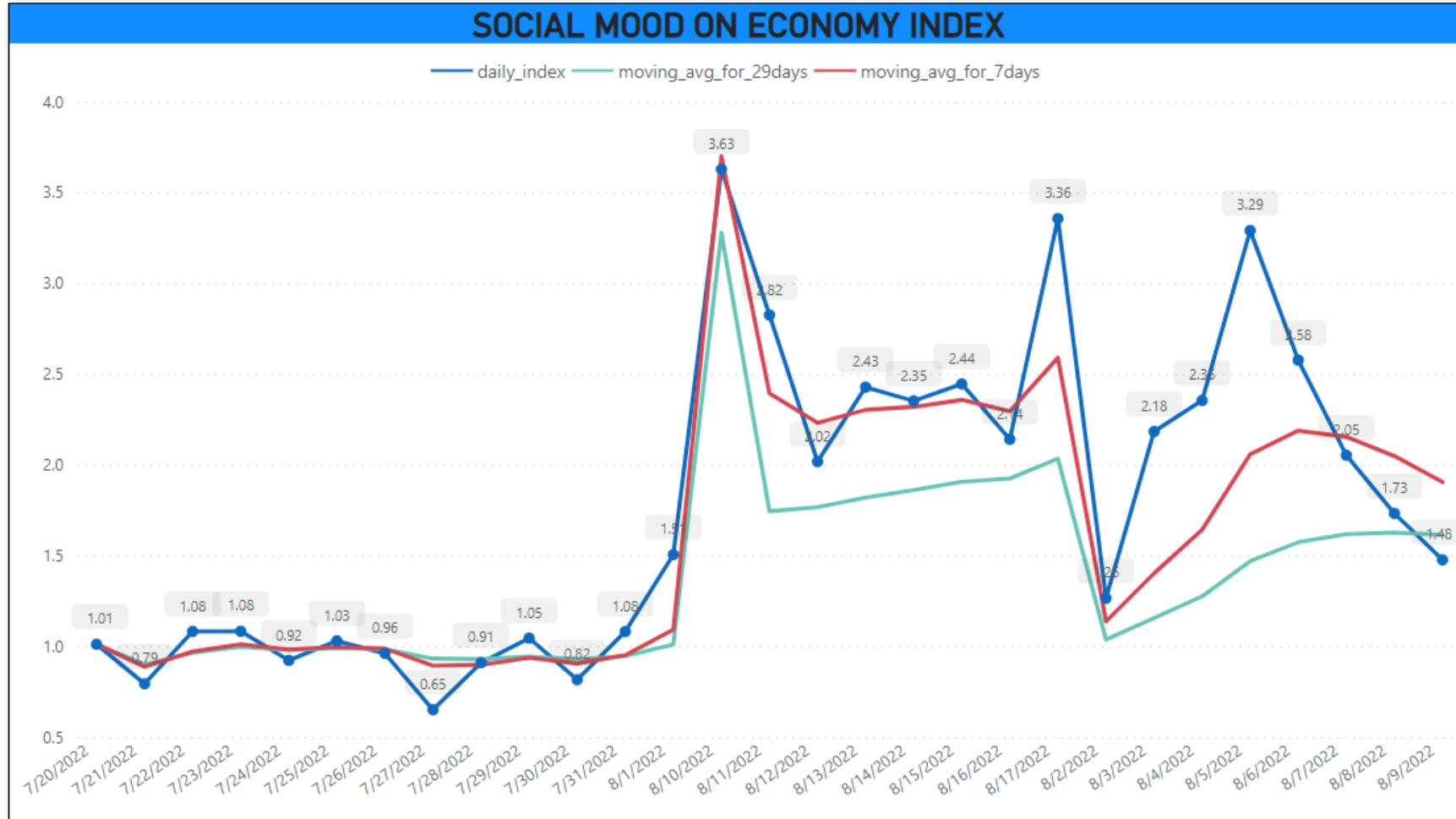
- There is a plot of 22 days of data collection

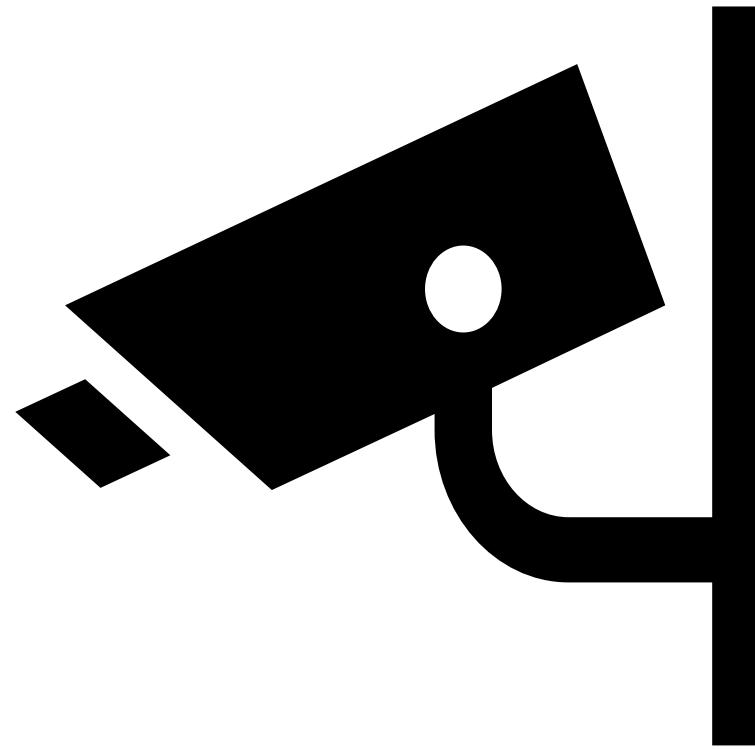


# Visualisation



- There is a plot of 29 days of data collection





# Surveillance System



## Surveillance System

- The system will be monitored from time to time in the event of a sudden and strong rise in the indicator or a sudden and strong decrease in the indicator
- And that will be considered as Future Work