# Twitter Challenge



Sophia Perticarà

#### Obiettivo

Vince chi scarica il maggior numero di tweets relativi al percorso dell'ITALIA agli Europei in corso, proponendo l'analisi più interessante e distintiva.





# Ricerca per hashtag

Data: 20/03/2020 - 1 anno fa

Si chiamerà Euro 2020 o Euro

2021? La decisione della Uefa



Euro 2020 è diventato ufficialmente Euro 2021. Solo nelle tempistiche però, perché il nome resterà sempre lo stesso: "Sebbene avrà luogo dall'11 giugno all'11 luglio 2021, Euro 2020 sarà ancora noto come Uefa Euro 2020". Lo ha comunicato la Uefa con un comunicato ufficiale, dopo aver preso la decisione di posticipare il torneo di un anno a causa dell'emergenza Coronavirus.

#euro2021
#euro2020
#europei2021
#europei2020
#uefa2020
#uefa2021



### API Twitter (Tweepy)



### Tweepy.Cursor

```
for ind, tweet in enumerate(tweepy.Cursor (api.search, q=query, count=100, lang="it", since= "2021-06-23", until = "2021-06-26", tweet_mode="extended").items() ):
```

- Tweets fino ad un massimo di 7 giorni antecedenti la data corrente
- Since-until: range via via di 3 giorni per evitare il blocco da parte di Twitter

Tweets totali scaricati: 121.984

Finestra temporale : 23 Giugno - 1 Luglio

Campi: Created\_at, Text





### **API Twitter**



Per i tweet più vecchi di 7 giorni:

 GetOldTweet3 Package (non mi funziona)



• Selenium scraping (da fare...)





# Pulizia e selezione (Pandas e Regex)



• selezione dei post che parlano dell'Italia:

```
keyword_ita = ["italia", "azzur"]
df_italia = df[df[cols_list].stack().str.contains('|'.join(keyword_ita),
case=False,na=False).any(level=0)]
```

- identificazione nuove var (re package):
  - Retweeted(?<=RT\s)(@[A-Za-z]+[A-Za-z0-9-\_]+)</li>
  - Mentioned (?<!RT\s)(@[A-Za-z]+[A-Za-z0-9-\_]+)</p>
  - Hashtags(#[A-Za-z]+[A-Za-z0-9-\_]+)
- eliminazione RT

Tweets totali: 47.681

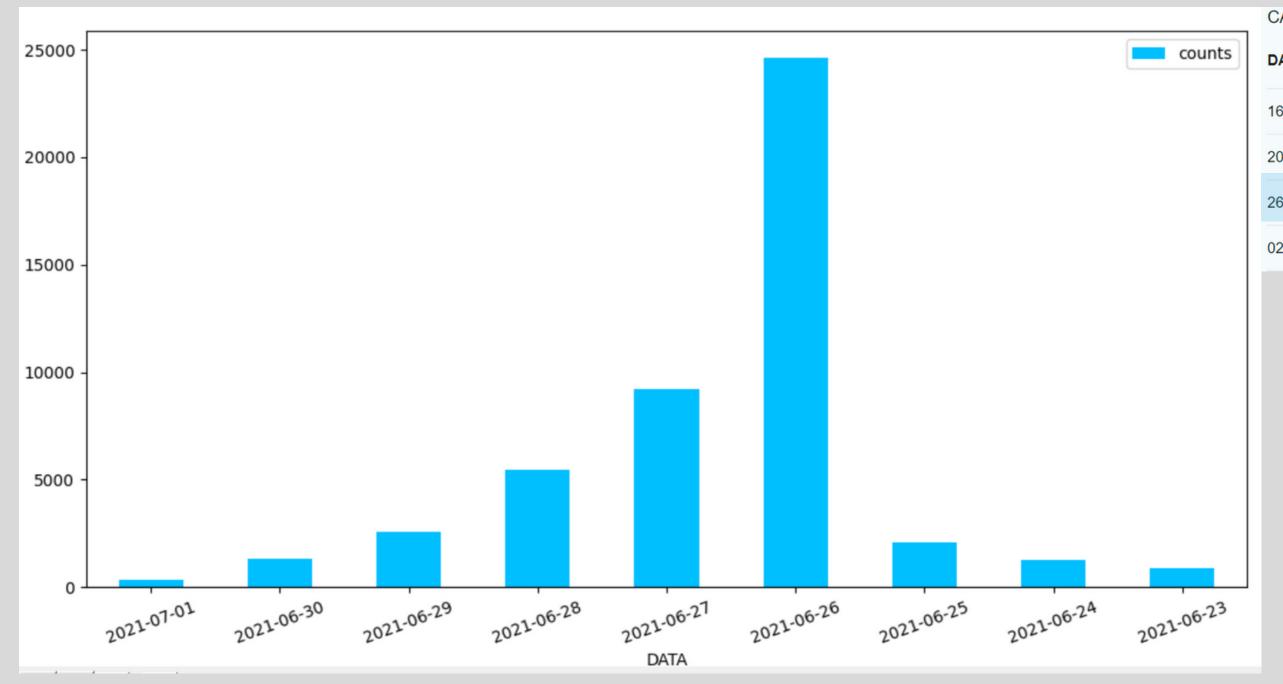


#### Euro\_challenge.csv

```
2021-07-01 10:36:31;#Euro2020, Belgio-Italia sold out. I biglietti
   2021-07-01 10:35:36;RT @Vivo_Azzurro: #Nazionale πEURO 2016 🔙 🛍 La
    ;2021-07-01 10:31:42;L'80% degli italiani ha seguito gli ottavi di
    ;2021-07-01 10:30:17;RT @Vivo Azzurro: #Nazionale π 🤂 Ieri sera l'
   ;2021-07-01 10:30:00;"#EURO2020 è più nitido che mai grazie alla t
   ;2021-07-01 10:29:44;RT @Vivo_Azzurro: #Nazionale πEURO 2016 🔙 🛍 L
    ;2021-07-01 10:28:27;RT @Vivo_Azzurro: #Nazionale π 🔂 Ieri sera l'
    ;2021-07-01 10:27:31;RT @SpotifyItaly: Siete tu, i tuoi amici e #I
_____;2021-07-01 10:26:46;RT @Vivo_Azzurro: #Nazionale πEURO 2016 🔙 🛍 L
🗖 🔲 ;2021-07-01 10:26:00;RT @Vivo_Azzurro: #Nazionale πΕURO 2016 🔙 🛍 L
ೡ;2021-07-01 10:25:13;RT @Vivo_Azzurro: #Nazionale π 🔂 Ieri sera l'
    ;2021-07-01 10:24:05;@BGaggiottini @Vivo Azzurro @SkySport @BGaggi
    ;2021-07-01 10:22:30;RT @Vivo_Azzurro: #Nazionale πEURO 2016 🔙 🛍 L
    ;2021-07-01 10:19:58;RT @Vivo_Azzurro: 🔕 Ciao, Giovanni #DiLorenzo
    ;2021-07-01 10:19:22; "RT @siamonoitv2000: ""Io sacerdote, lui calc
    ;2021-07-01 10:18:57; A ME INTERESSA SOLO CHE DOMANI SIATE 11 CANI
    ;2021-07-01 10:17:43;EURO 2020, La squadra arbitrale di Belgio-Ita
   ;2021-07-01 10:16:29;RT @Vivo Azzurro: #Nazionale # 🤂 Ieri sera l'
_____;2021-07-01 10:16:29;RT @Vivo_Azzurro: #Nazionale πEURO 2016 🔙 🛍 L
    ;2021-07-01 10:15:04;Italia-Belgio, spettacolo all'aperto per spin
    ;2021-07-01 10:14:53;RT @DiMarzio: #Euro2020, #ITA | Verso #Belgio
   ;2021-07-01 10:14:52;RT @Vivo_Azzurro: #Nazionale π 🖫 La vittoria
   ;2021-07-01 10:14:31;RT @Vivo_Azzurro: #Nazionale #EUR💋 2016 🔙 🛍 L
    ;2021-07-01 10:13:06;RT @4everAnnina: Non quarderò la partita #Bel
```

# Distribuzione tweet per date





CALENDARIO ITALIA EURO 2020				
DATA	ORA	PARTITA		
16/06	21.00	Italia-Svizzera 3-0		
20/06	18.00	Italia-Galles 1-0		
26/06	21.00	Italia-Austria 2-1 dts		
02/07	21.00	Belgio- <b>Italia</b>		



# I top 10 tweet più retweettati?



▼ TEXT	counts		
RT @robymancio: L'abbiamo voluta, l'abbiamo vinta e adesso abbracciamoci tutti insieme! #Euro2020 #ITA #ItaliaAustria https://t.co/3MMvZakH	1901	Roberto Mancini	
RT @Vivo_Azzurro: #EURO2020	1331	Official Twitter of Roberto Mancini	
SÌ! SIAMO AI QUARTI!			
ттат #ItaliaAustria 2□-1□		s	
#ITA #ITAAUT #Nazionale #Azzurri #VivoAzzurro https://t.co/		Nazionale Italiana ★★★★ ♥  @Vivo_Azzurro	
RT @trash_italiano: Il rigore che ha fatto uscire la Francia dall'#Euro2020 but con il sottofondo che ha fatto perdere la Francia all'Eurov	1104	Twitter ufficiale delle Nazionali Italiane  di Calcio  Hashtag: #VivoAzzurro English: @Azzurri Arabic: @Azzurri_Ar   Forza #Azzurri	
RT @juventusfc: BRAVO FEDEEEEEEEEEEEEEEEEEE! 💜 🧷	1104		
@federicochiesa in gol e @Vivo_Azzurro ai quarti di #EURO2020 т https://t.co/5xwjL			
RT @Vivo_Azzurro: Tat #ItaliaAustria 2-1 (0-0 d.t.r.)	946	Federico Chiesa (**) @federicochiesa	
#Chiesa 95', #Pessina 105', #Kalajdzic 114'		Giocatore della Juventus e della Nazionale Italiana di Calcio.	
Gli #Azzurr  RT @IlContiAndrea: Sicuramente l'immagine più bella di stasera #ItaliaAustria #Euro2020 https://t.co/AUJRwle2BH	609		
RT @chiellini: SIAMO AI QUARTI!!! Ho sofferto, ho gioito, ho incitato i miei compagni e alla fine ho esultato come ogni tifoso italiano! rr	335	Andrea Conti 🕗	
		@IlContiAndrea	
RT @Vivo_Azzurro: #Nazionale ɪт	317	Con il Web e coltivo i miei sogni. Giornalista 🥷 Scrivo di Spettacolo su @FQMagazinelt del @fattoquotidiano e collaboro con @Fanpage Sport: 🟋 🧎	
#Belgio Vs #Italia		gramagazment der grantequetadane e conducte con grantpage sporta in	
Quarti di finale  Venerdì 2 luglio, ore 21 Football Arena Munich Monaco di B			
RT @GiorgiaMeloni: L'Italia ha vinto! In piedi, contro ogni polemica e imposizione. Si vola ai quarti di finale, forza azzurri ɪт 💪	292	TM TM	
		EURO2020	

# Gli hashtag più popolari?

pythor

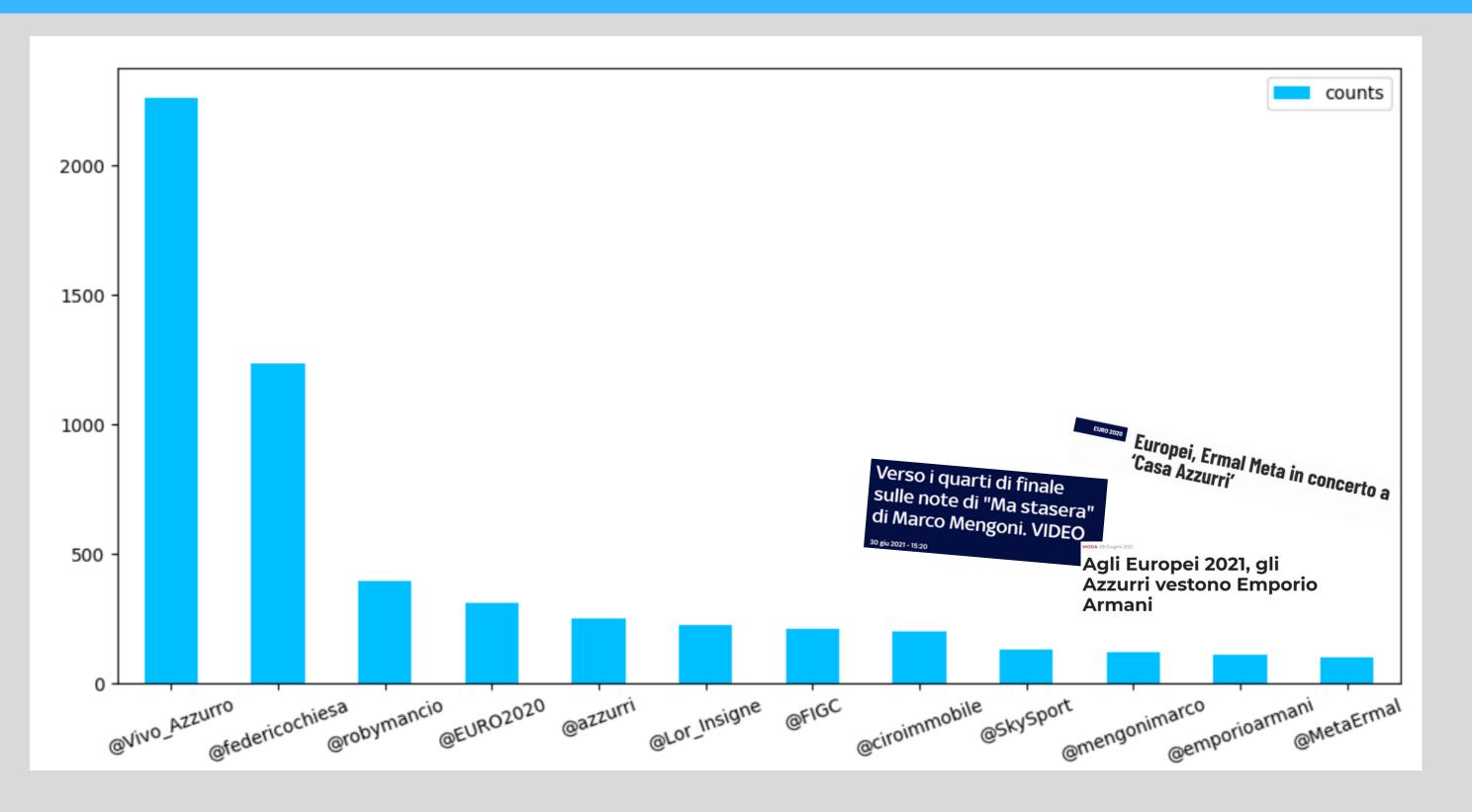
- WordCloud Package
- Totali Hashtag distinti: 4.026
- Distribuzione di frequenza (con anche i RT)



Column	▼ hashtag	counts
0	#EURO2020	31685
1	#ItaliaAustria	24200
2	#ITA	13321
3	#ITAAUT	12802
4	#Euro2020	11257
5	#VivoAzzurro	9081
6	#Azzurri	8828
7	#Nazionale	7658
8	#Italia	4681
9	#Chiesa	2290
10	#FranciaSvizzera	2030
11	#Austria	1919
12	#SkySport	1820
13	#Pessina	1736
14	#euro2020	1731
15	#Wembley	1728
16	#FRASUI	1674
17	#Euro2021	1557
18	#europei2021	1545
19	#ItaliaBelgio	1449
20	#Londra	1260

# Chi viene menzionato di più?





<b>▼</b> mentioned	counts
@Vivo_Azzurro	2264
@federicochiesa	1235
@robymancio	395
@EURO2020	309
@azzurri	249
@Lor_Insigne	228
@FIGC	213
@ciroimmobile	202
@SkySport	133
@mengonimarco	120
@emporioarmani	111
@MetaErmal	103
@bonucci_leo19	91
@figc	83
@OptaPaolo	78
@UEFA	77
@repubblica	75
@robersperanza	74
@RaiSport	73
@RaiUno	72
@EUCouncil	71
	@Vivo_Azzurro @federicochiesa @robymancio @EURO2020 @azzurri @Lor_Insigne @FIGC @ciroimmobile @SkySport @mengonimarco @emporioarmani @MetaErmal @bonucci_leo19 @figc @OptaPaolo @UEFA @repubblica @robersperanza @RaiSport @RaiUno

## Tweet che sfuggono alla selezione...

Si può dire che il Portogallo è stato anche sfortunato, il Belgio alla fine è battibile. Sicuramente una signora squadra che merita rispetto, ma ce la dobbiamo giocare senza paura... si può vincere! #BelgioPortogallo #EURO2020

Dopo i francesi, facciamo fuori i tedeschi. #ENGGER #EURO2020"

Mo' facciamo entrare Pessina e Chiesa, che stiamo aspettando? #FRASUI #EURO2020"

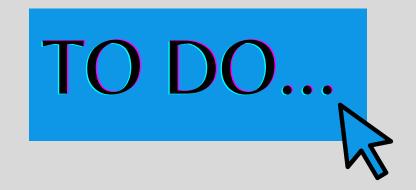
certo che però facciamo schifo, c'è per praticamente tutto il tempo nessuno ha fatto gol #europei2021"

se non facciamo 3-0 per la terza volta non siamo contentx #EURO2020

ampliare set di ricerca keyword = ["italia", "azzurri"] nomi dei giocatori della formazione?

# Le parole più usate nei tweet?

- Package Nltk
- Pulizia testo del tweet (nltk.corpus.stopwords.words('italian'))
  - stopwords: lista delle parole di uso più comune del linguaggio naturale che non apportano nessuna informazione utile
- applicazione di modelli per capire di cosa si parla e qual è il sentiment



```
ad
al
allo
ai
agli
all
agl
alla
alle
con
col
coi
da
dal
dallo
dai
dagli
dall
dagl
dalla
dalle
```

## Github (codici)

https://github.com/sophiaperticara/Master\_Bicocca\_BigData

- Scarico\_tweet.py
- 1. Analisi.py

