

PROGETTO DI DATA MANAGEMENT

LACASA DE PAPEL

EL FENOMENO

Stefano Colombo 838941
Marco Savino 793516
Antonella Zaccaria 848647



EL FENOMENO

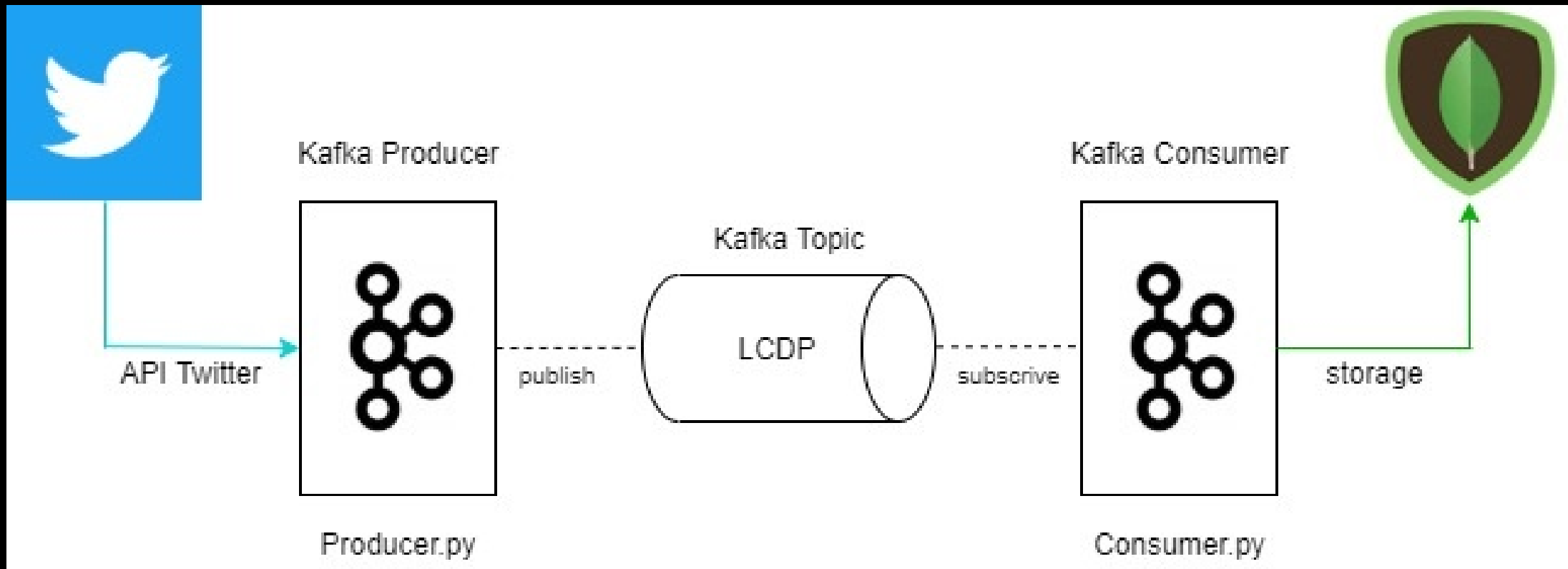


OBIETTIVI

- Implementare un'architettura che consentisse l'utilizzo di due (Velocità, Varietà) delle 3V caratteristiche dei *Big Data*
- Comprendere quali fossero le polarità suscitate negli utenti rispetto ai diversi personaggi citati, in modo da confrontare i personaggi più amati e quelli più odiati in riferimento ai *tweet* in lingua italiana, inglese e spagnola ed analizzare le caratteristiche degli attori che interpretano tali personaggi.



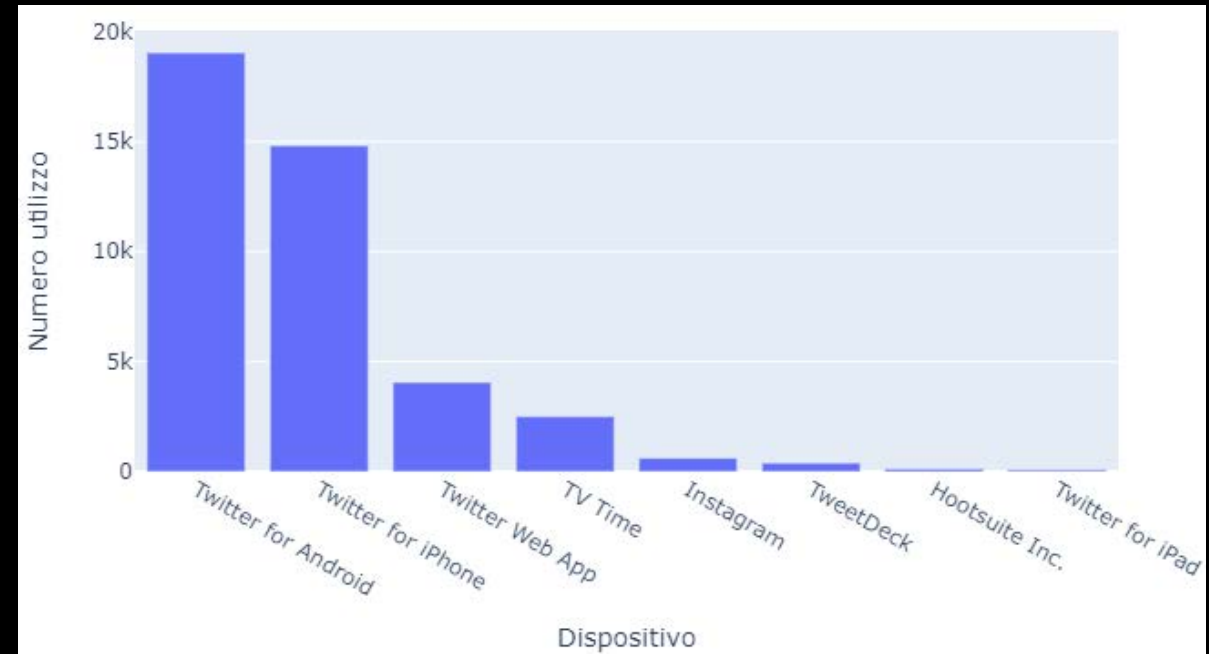
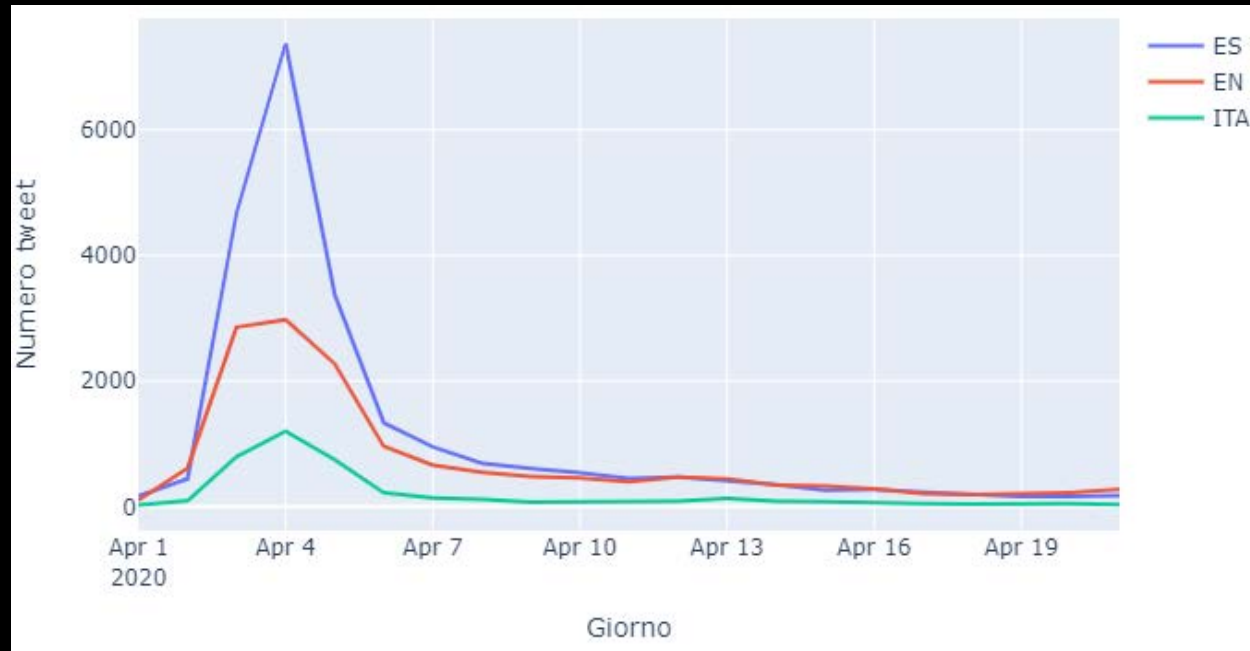
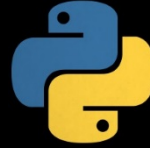
VELOCITA'



- #LaCasaDePapel
- Dal 01-04-2020 al 21-04-2020

- 4.226 *tweet* in lingua italiana
- 15.298 *tweet* in lingua inglese
- 23.297 *tweet* in lingua spagnola

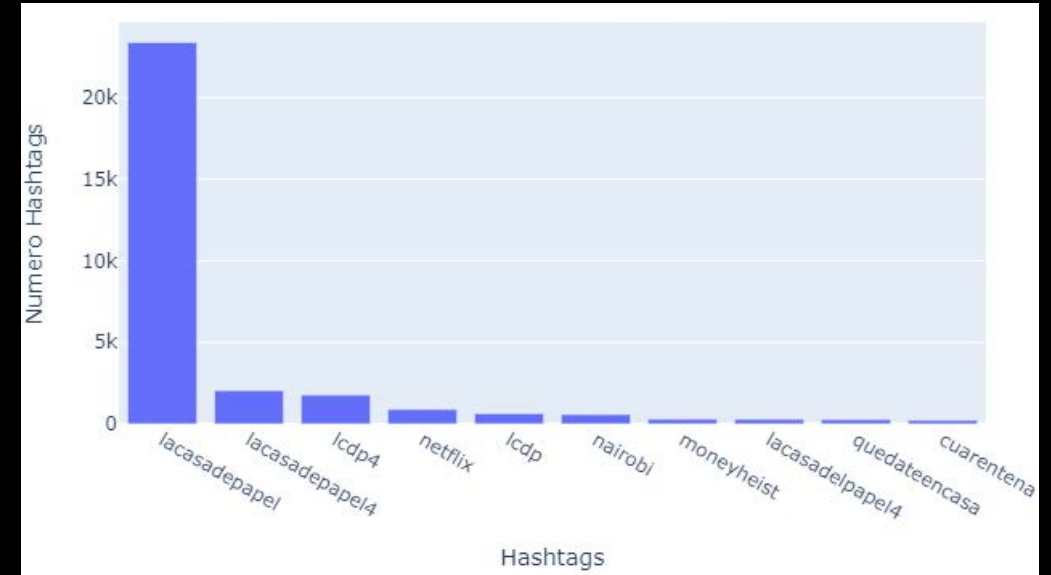
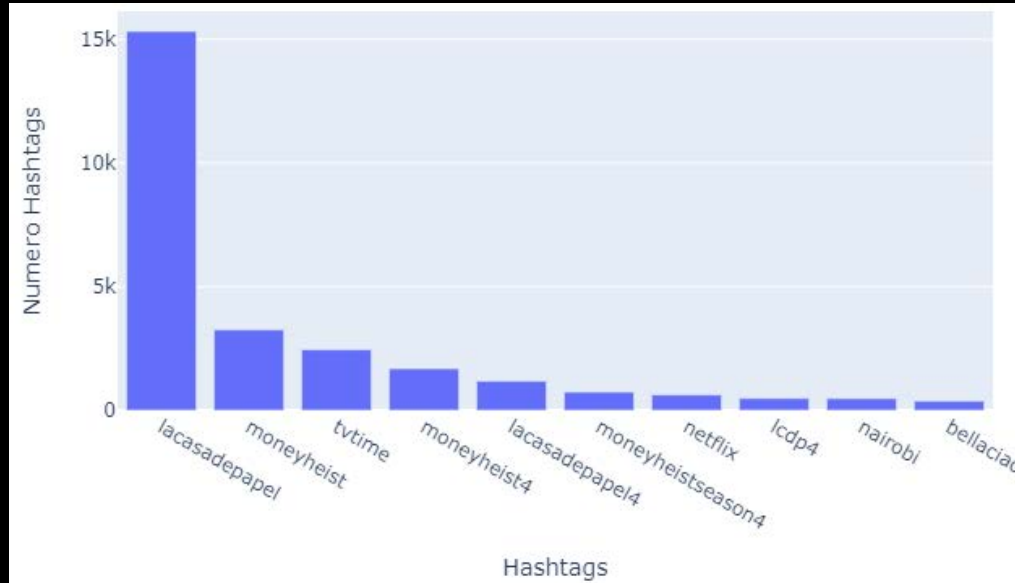
ANALISI PRELIMINARI



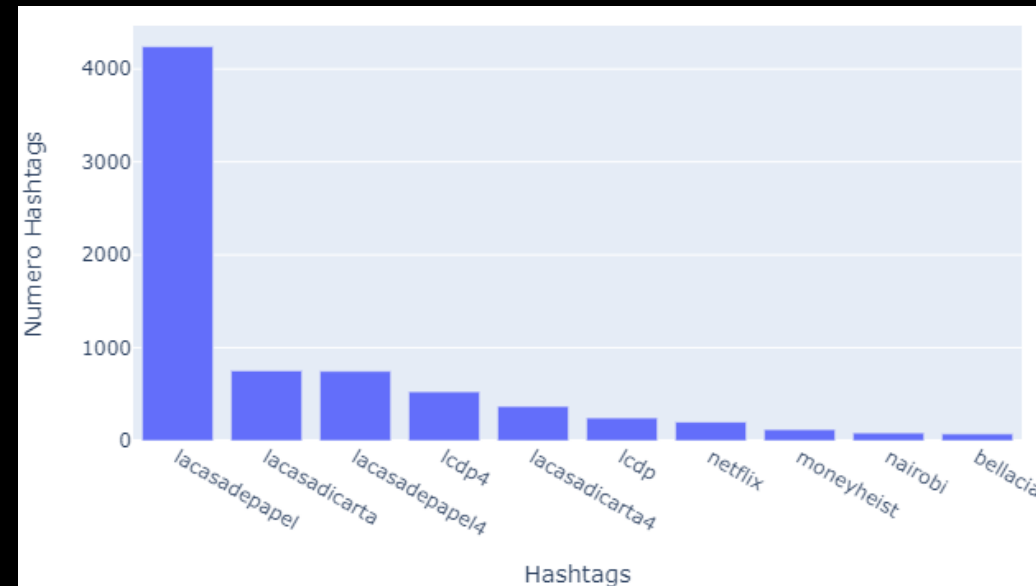
#EN



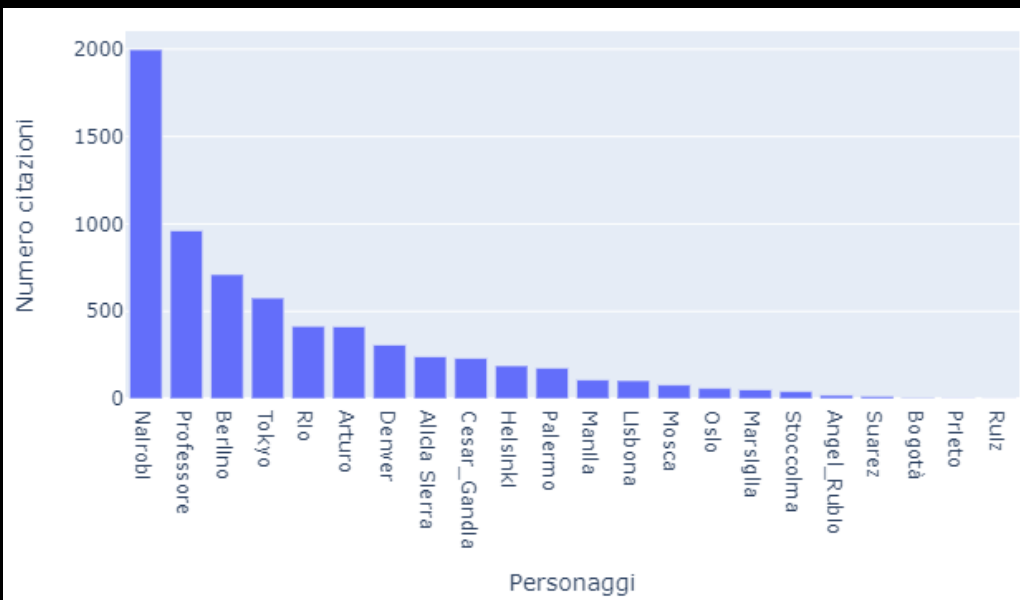
#ES



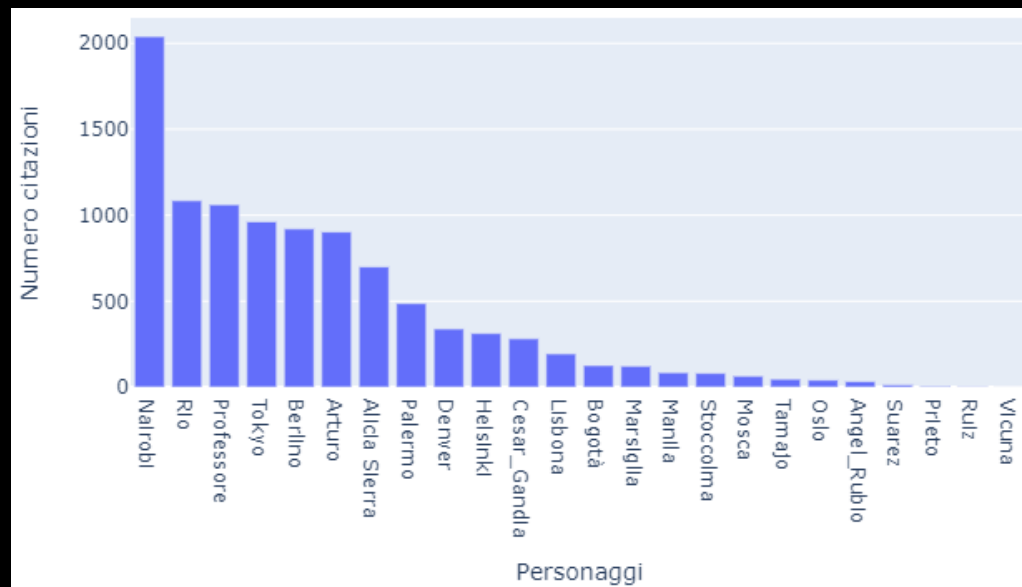
#IT



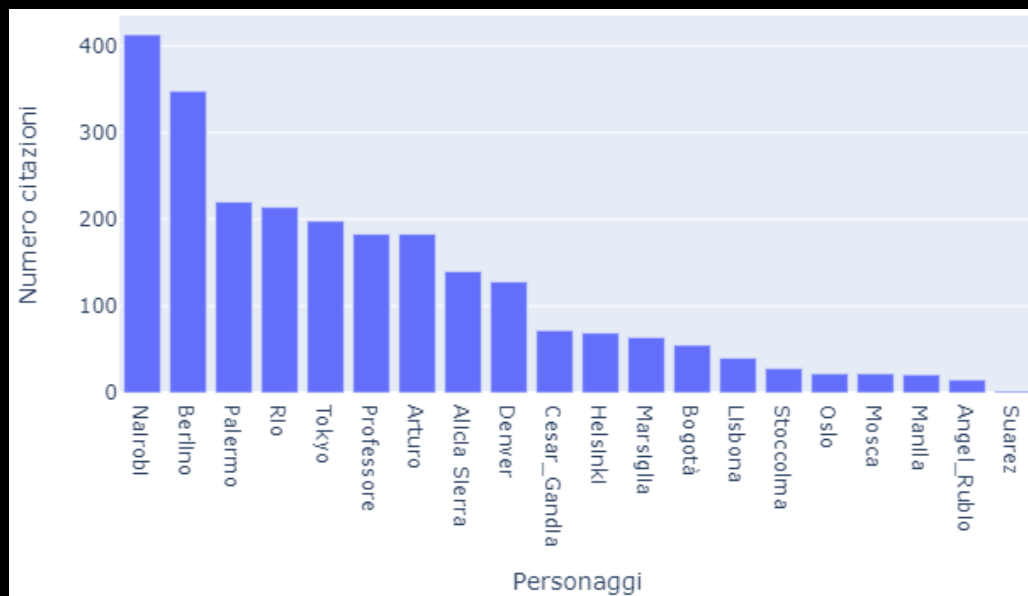
EN



ES



IT



WORDCLOUD



SENTIMENT ANALYSIS TWEET INGLESI

- Lessico VADER [-1; +1]
- $-1 \leq \text{sentiment} \leq -0.5$: F_NEG
- $-0.5 < \text{sentiment} < 0$: NEG
- Sentiment = 0: NEUTRO
- $0 < \text{sentiment} < 0.5$: POS
- $0.5 \leq \text{sentiment} \leq 1$: F_POS
- Normalizzazione per corretto confronto

Watching Nairobi take a bullet through the head broke me. Fuck you to the writers of money heist. [-0.8689]

We will always miss u Nairobi. [-0.1531]

For Nairobi #moneyheist #lacasadepapel [0.0]

Nairobi was my only favourite character [0.25]

Nairobi is the best character on lacasadepapel [0.6369]





SENTIMENT ANALYSIS TWEET ITALIANI E SPAGNOLI

- Googletrans (Python)
- Lessico VADER [-1; +1]

*Sappiate che avete tutto il mio odio profondo
#lacasadepapel [-0.5719]*

*Vorrei solo sottolineare il fatto che anche questa
volta è colpa di Tokyo, se non avesse lasciato
Rio sull'isola tutto questo non sarebbe successo
[-0.0516]*

*Ho visto l'episodio S04E06 di Money Heist!
#lacasadepapel [0.0]*

*Comunque io aspetto solo che Denver e
Stoccolma tornino insieme nella quinta stagione
e adottino Rio [0.1779]*

*Non so voi ma la amo #aliciasierra
#lacasadepapel [0.7783]*



*Tokio q! aweona quiere cagar todo el plan
#lacasadepapel. [-0.5574]*

*Ya no voy a aceptar ninguna propuesta de
matrimonio sin una escena como esta.
He dicho. #lacasadepapel [-0.1803]*

*Acabo de ver el episodio s04e02 de money heist
#lacasadepapel [0.0]*

*Igual esta buena la mina en jefe de la polic'ia
#lacasadepapel [0.4404]*

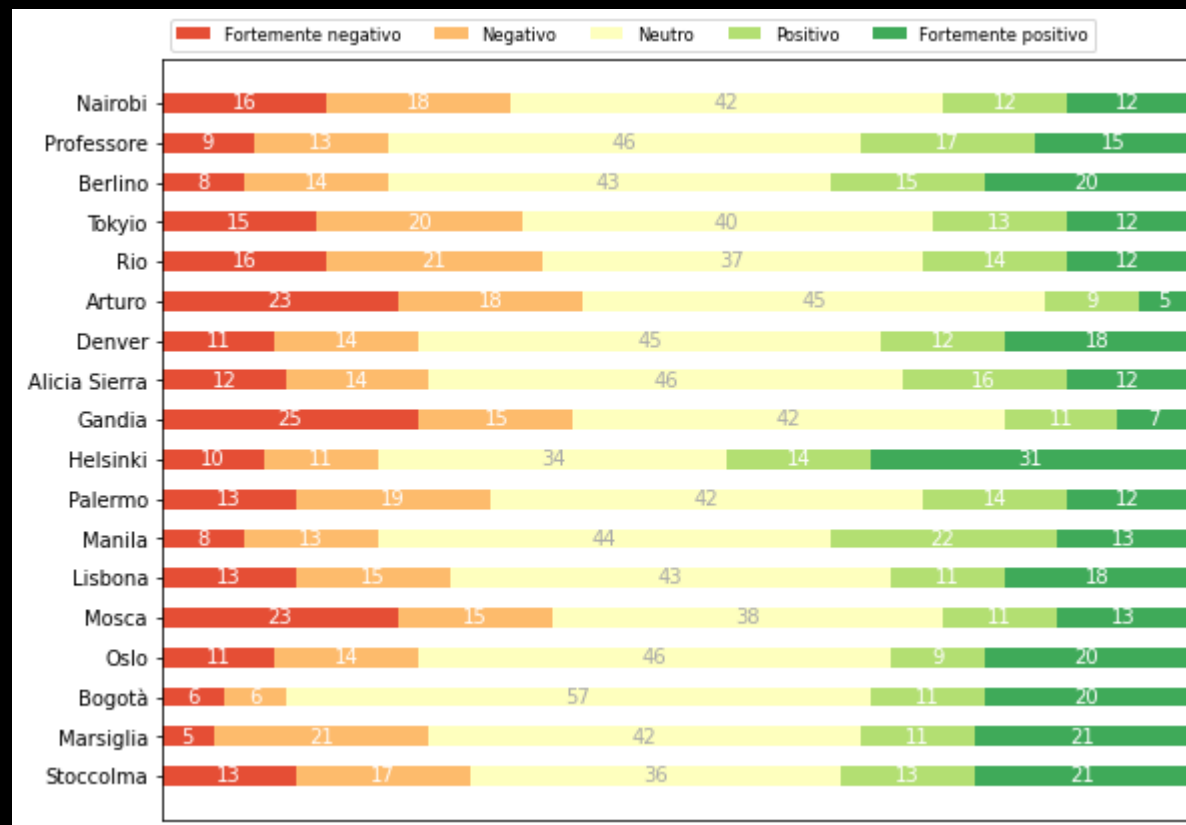
*Termine de ver la segunda temporada de
#lacasadepapel
que genial final wn la raja [0.6249]*



IT



ES

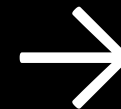


VARIETA'

SCRAPING:

- BeautifulSoup (Python) → Biografia degli attori
- Web Scraper (Google Chrome) → Premi + Film + Serie TV attori

Biografia
Premi
Film
Serie TV
Tweet



Nodi: Attore, Cinema,
Riconoscimenti, TV, Personaggio,
Tweet

Relazioni: *ha_recitato_in* (Attore-
Cinema),
ha_ricevuto_il_premio (Attore-
Riconoscimenti),
ha_recitato_nel_programma_tv
(Attore-TV),
ha_interpretato_il_ruolo_di_ne_LC
DP (Attore-Personaggio) e
citato_nel_tweet (Tweet-
Personaggio).

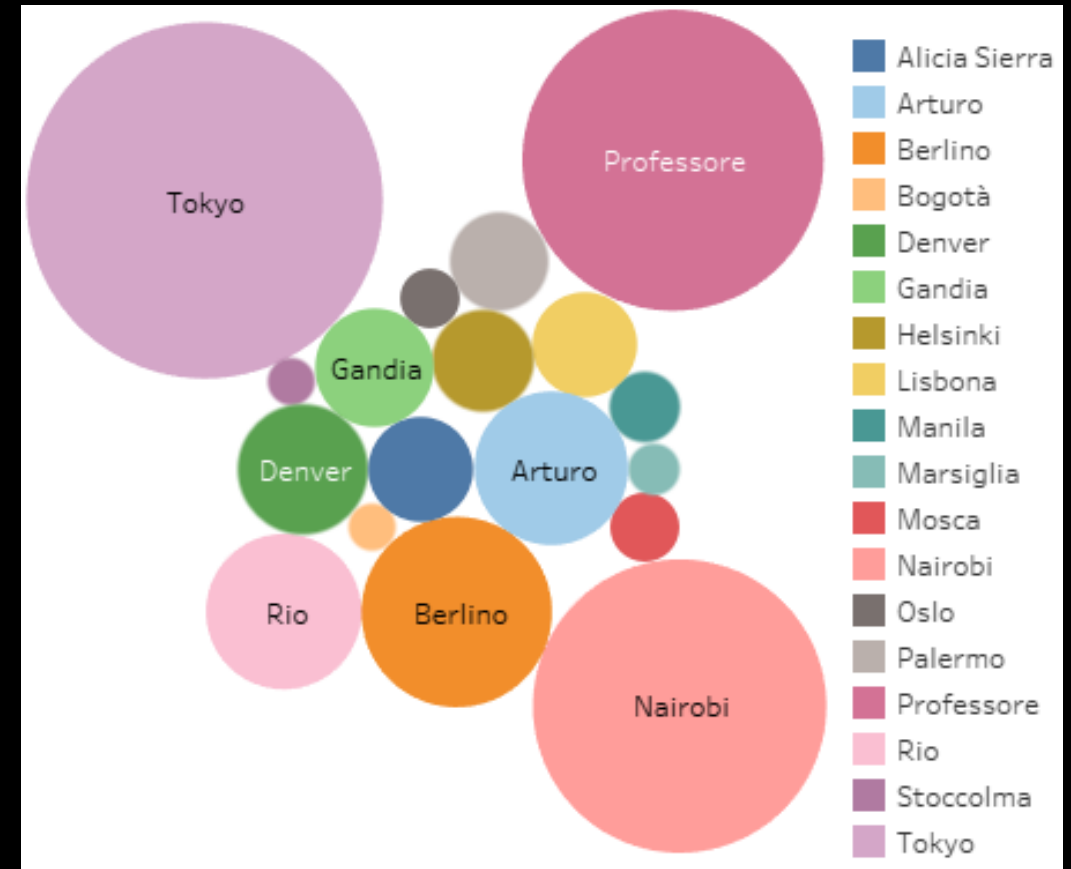
SCORE DI IMPORTANZA DEI PERSONAGGI

$$\text{Score} = \text{NTPF} * 1 + \text{NTP} * 0.5 + \text{NTN} * 0.5 + \text{NTFN} * 1$$

- NTFP = Numero di Tweet Fortemente Positivi
- NTP = Numero di Tweet Positivi
- NTN = Numero di Tweet Negativi
- NTFN = Numero di Tweet Fortemente Negativi

Se il personaggio ha ricevuto premi:
Score * (2*numero premi ricevuti)

Altrimenti:
Score invariato



CONCLUSIONI

- La scelta di svolgere un progetto sulla Casa De Papel è legata alla curiosità di analizzare una serie tv così famosa da diventare un vero e proprio fenomeno mondiale, soprattutto durante il periodo di *lockdown* in cui le persone disponevano di più tempo da dedicare alla visione di serie tv.
- La *Sentiment Analysis* ha evidenziato i personaggi preferiti e quelli meno apprezzati dagli utenti di *Twitter*, che rispecchiano le interpretazioni ed i comportamenti dei personaggi durante la quarta stagione della *Casa De Papel*.
- La varietà dei dati raccolti ha inoltre permesso il calcolo di uno *score* di importanza dei vari personaggi, ottenuto tenendo conto di un numero maggiore di fattori.

GRAZIE PER L'ATTENZIONE

