

#BBC100Women



Antonicelli Giorgia, 872784

Lorgna Lorenzo, 829776

Maccario Alessandro, 865682

BBC 100 Women

Ogni anno, la **BBC** pubblica la **lista delle 100 donne** più ispiratrici ed influenti, provenienti da ogni angolo del mondo.



Obiettivi



- Quanta risonanza hanno su Twitter la pubblicazione della lista e la stagione delle donne della BBC? È effettivamente vero che si discute attivamente del tema delle donne?
- Che tipologia di donne vengono premiate? Da dove provengono? In cosa si sono distinte? Quanti anni hanno? La loro popolarità aumenta dopo esser state inserite nella lista?
- Nel corso degli anni, l'evento è diventato più popolare? È cambiato il tipo di donne inserite nella lista?



Data management

Nella realizzazione del progetto l'intento è stato quello di affrontare almeno due delle 3V:

VELOCITÀ



VARIETÀ



Per quanto riguarda la velocità, è stata simulata la raccolta dati, in questo caso tweets, provenienti da una fonte streaming. Per quanto riguarda invece la varietà i dati raccolti dalla fonte streaming sono stati integrati con ulteriori dati provenienti da diverse fonti.





Scaricamento tweets



Per ogni anno, dal 2015 al 2020, sono stati scaricati tutti i tweets che soddisfassero la seguente query:

```
( @BBC100Women OR #BBC100Women OR #BBC100women OR #bbc100women  
OR #Bbc100Women OR #bbc100WOMEN OR #bBc100women OR #BBC100WOMEN )
```

Successivamente, considerando i soli anni 2015 e 2019, **per ogni donna** inserita nella lista si sono scaricati i tweets che rispettassero la seguente query:

```
( USERNAME_TWITTER_DONNA OR ( NOME_DONNA AND ( HASHTAGS_CONSIDERATI ) ) )
```



BBC News Africa @BBCAfrica · 29 nov 2020
"Death is when you are afraid to speak the truth."

In the midst of the #endsars protests in October, we spoke to @AishaYesufu who says she is dedicating her life to fighting for justice. The Nigerian activist is one of this year's #BBC100Women.



Vanessa Nakate @vanessa_vash · 24 nov 2020

I am hugely honored to be included in this amazing list of inspiring and influential women. This is precisely why we need to help educate and empower all girls. When we do, they grow up to become the leaders we so desperately need. #BBC100Women



Nimco Ali (OBE) @NimkoAli · 19 nov 2018

I am honoured to be named on this years #BBC100Women alongside so many incredible women. And being in the top 5 has just left me speechless. Thank You to the bbc for this honour.





Simulazione velocità



KAFKA Producer

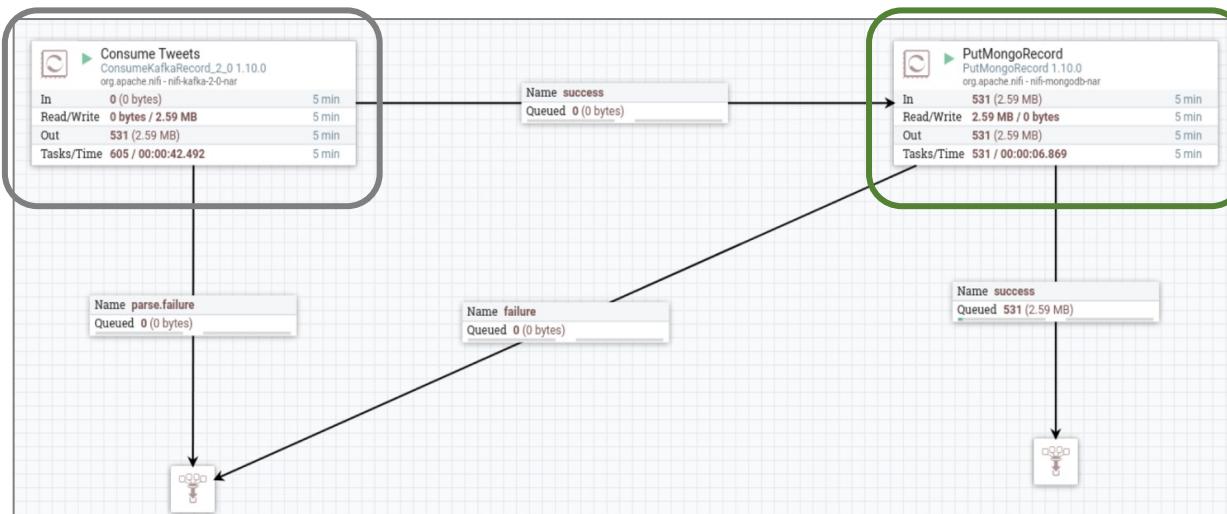
```
from kafka import KafkaProducer
import json
import time

# producer creation
tweets_2020_producer = KafkaProducer(
    bootstrap_servers=[ "kafka:9092" ],
    value_serializer=lambda v: json.dumps(v).encode("UTF-8"))

# lettura file json tweets scaricati
tweets_final = json.load(open("tweets_2020.json"))

# tweets -> topic -> mongodb
for tweet in tweets_final:
    tweets_2020_producer.send(topic='tweets', value=tweet)
    # time.sleep(0.005)
```

NIFI Consumer



Mongo DB database





Varietà



Una volta ottenuti i tweets si è proceduto con una **fase di arricchimento** dei dati. Le operazioni di linking tra diverse fonti dati sono state eseguite tramite **script** appositi in **Python**, ottenendo come output finale dei files **csv** che sono stati utilizzati per creare delle visualizzazioni in modo da agevolare la risposta alle domande di ricerca.



Identikit Donne

Abc	Id
Abc	Nome Donna
#	Numero Tweets Pre Uscita
#	Numero Tweets Post Uscita
#	Età
Abc	Categoria
🌐	Stato di provenienza
Abc	Regione
Abc	Username Twitter
Abc	Hashtag
📅	Anno
Abc	Account verificato

Indicatori Stati

🌐	Stato
📅	Anno
#	Pil
#	Pil Pro Capita
#	Gender Gap
#	Hdi
Abc	Id donna

Numero Tweets

Data
Numero tweets

BBC 100 Women:
<https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-12-08/women.csv>

GDP / GDP_per_capita:
<https://www.imf.org/external/datasets/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD>

Gender Global Gap:
https://reports.weforum.org/Human_Development_Index/

<http://hdr.undp.org/en/indicators/137506>



Data visualization





100 Women Un evento organizzato annualmente dalla BBC che a partire dal 2013 ha come obiettivo quello di valorizzare il ruolo della donna nella società moderna. L'idea nasce in seguito allo stupro di gruppo nei confronti di una ragazza 23enne di Delhi avvenuto nel 2012. A seguito di numerose manifestazioni di protesta, alcuni personaggi di rilievo della BBC decisero di creare una serie incentrata sulle questioni e le conquiste delle donne nella società di oggi.

Nei mesi finali dell'anno viene solitamente definita la **lista delle 100 donne** che meglio si sono distinte e che sono riuscite a fare la differenza nel corso dell'anno in modo tale che possano essere di ispirazione per il mondo intero.

Dopo la pubblicazione della lista prende il via la **"stagione delle donne della BBC"**, della durata di 3 settimane: trasmissioni, documentari, dibattiti online incentrati sul tema delle donne.

Le donne di tutto il mondo sono incoraggiate a partecipare tramite **Twitter**, commentando la lista, nonché seguendo le interviste e i dibattiti che si svolgono dopo la pubblicazione di essa.

Di seguito viene mostrata la risonanza mediatica dell'evento e una panoramica descrittiva delle donne nominate in due anni distinti considerandone caratteristiche e provenienza. Infine, viene analizzata l'influenza dell'evento sulla popolarità delle donne su Twitter.

Quanto se ne parla su Twitter?

Sono stati presi in considerazione i tweets con l'hashtag **#BBC100women** e la menzione **@BBC100Women**

Selezione l'anno da esplorare

Anno

- 2015
- 2016
- 2017
- 2018
- 2019
- 2020

Numero tweet



Per evidenziare maggiormente la distribuzione del numero di tweets, le celle sono state colorate considerando il logaritmo del numero di tweets

mese	January	February	March	April	May	June	July	August†	September	October	November	December
giorno												
1	0	0	0	0	4	0	0	0	29	3	3	3,418
2	0	0	1	0	0	3	0	1	23	2	6	334
3	0	1	6	0	1	0	1	1	8	1	6	144
4	1	1	1	2	0	5	0	1	5	0	4	103
5	0	0	3	1	2	3	0	18	2	3	0	56
6	0	0	0	0	2	2	0	9	1	2	4	31
7	0	5	4	1	1	0	0	8	3	0	1	67
8	0	0	3	1	0	0	0	0	2	1	1	40
9	0	0	0	1	0	3	0	7	3	0	5	37
10	1	1	2	0	1	3	1	3	5	29	0	29
11	0	1	0	0	3	2	0	2	4	0	6	21
12	0	0	1	0	0	0	2	5	5	0	1	32
13	0	1	0	0	2	0	0	5	2	3	2	20
14	0	0	0	0	3	0	135	0	22	1	0	24
15	2	0	1	3	1	0	31	0	10	3	1	66
16	1	1	1	1	0	0	9	3	10	1	2	23
17	1	3	0	0	0	0	2	2	3	0	24	29
18	0	0	0	0	1	1	7	2	3	0	232	66
19	1	1	0	0	1	1	1	0	3	0	766	37
20	0	1	0	0	0	0	1	0	1	1	185	12
21	1	0	0	0	2	0	8	1	1	3	69	7
22	1	0	0	0	0	0	7	3	4	7	80	6
23	0	0	2	3	0	1	2	0	7	2	586	2
24	1	1	0	1	0	6	7	10	7	0	136	0
25	0	0	4	0	2	1	0	12	0	0	194	1
26	1	0	0	0	2	1	1	3	2	0	129	1
27	0	1	0	0	3	0	3	1	0	19	176	1
28	2	0	0	0	3	0	0	1	11	64	122	6
29	0	0	0	2	0	0	1	0	19	3	148	3
30	0	0	0	3	1	2	2	1	7	2	285	4
31	1	2	0	0	1	1	25	0	3	0	5	

Fonti dati:

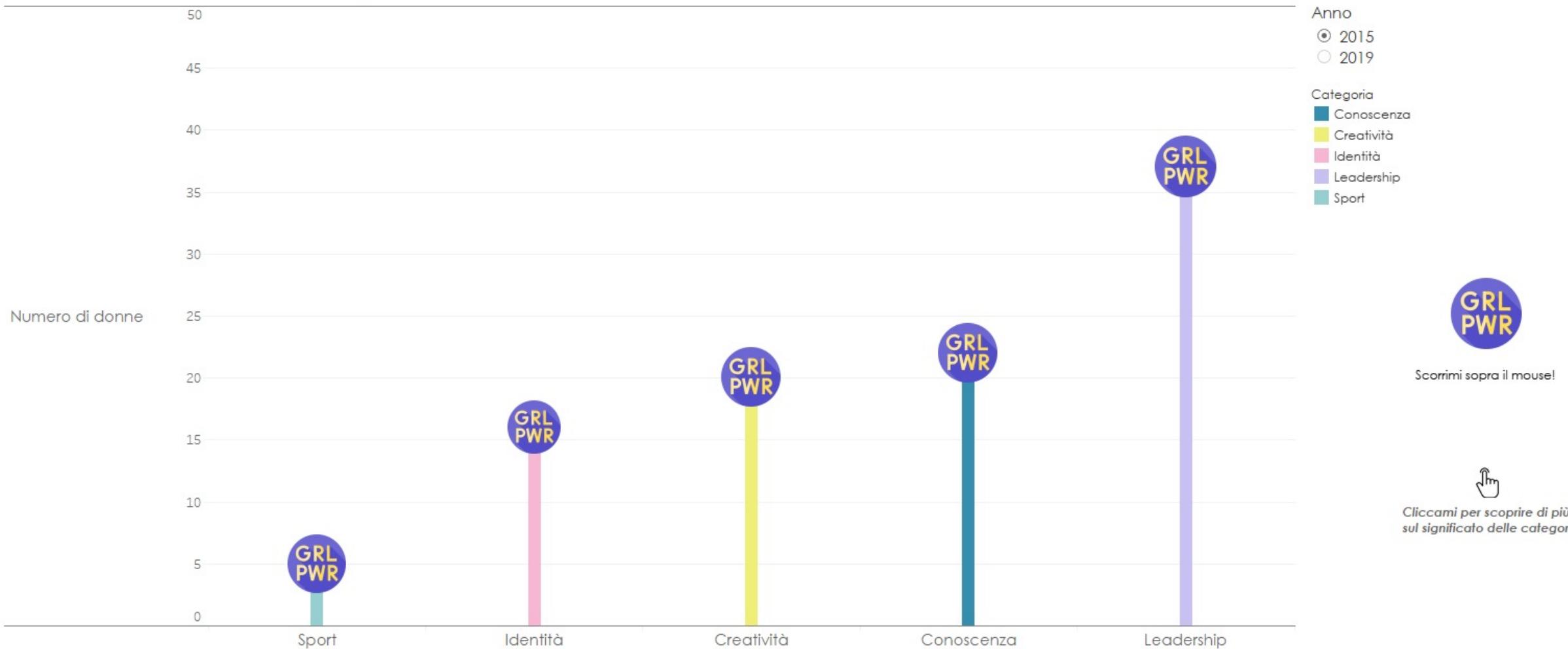
[#BBC100women](https://twitter.com)

[https://en.wikipedia.org/wiki/100_Women_\(BBC\)](https://en.wikipedia.org/wiki/100_Women_(BBC))

Categorie rappresentative ed età

Qual è la classe più rappresentativa delle donne in lista?

Quali sono le età mediane* per ciascun gruppo?



Fonti dati:

<https://www.flaticon.com/>

<https://www.bbc.com/news/world-50042279>

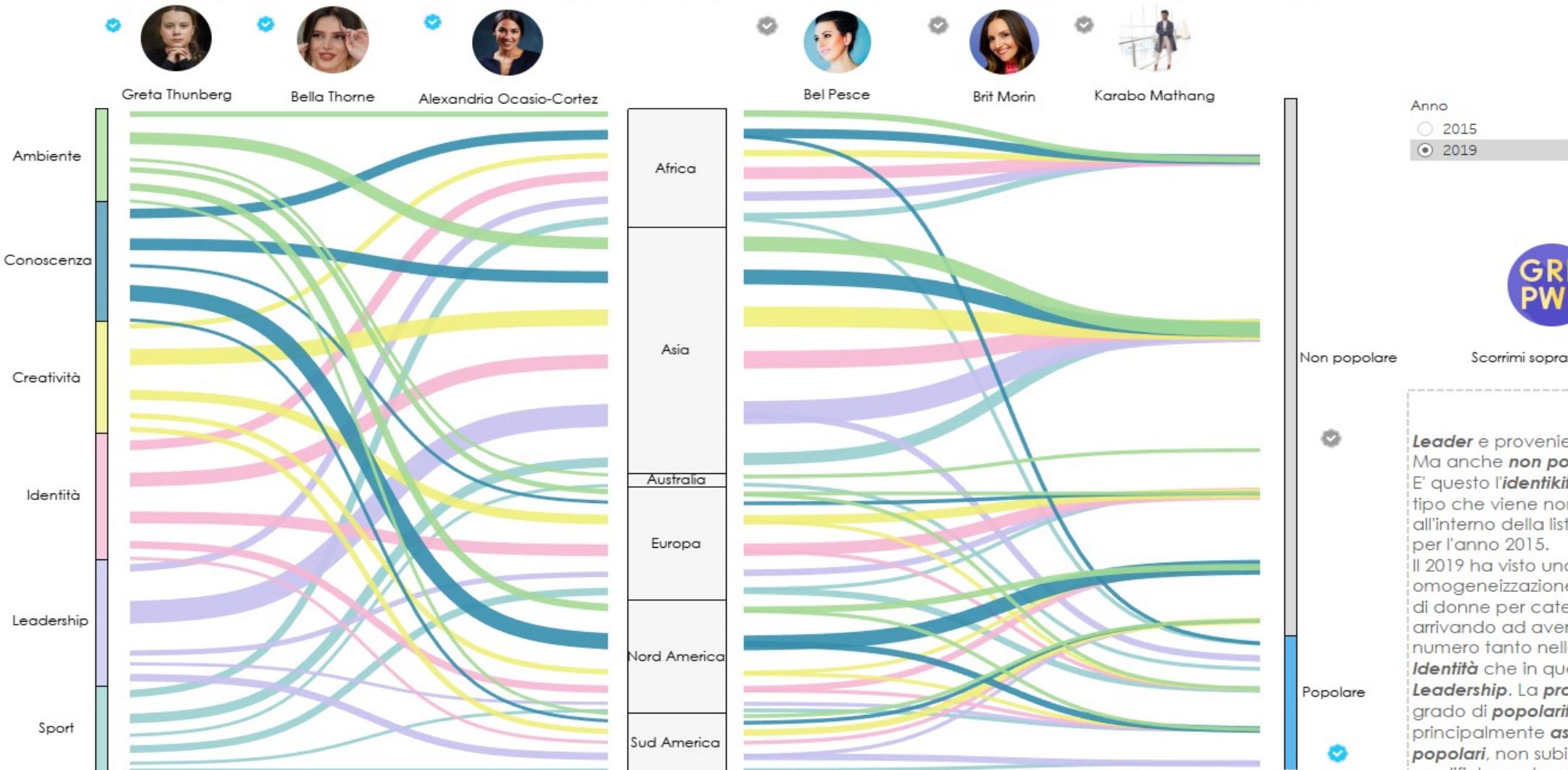
*NOTA: cos'è la misura MEDIANA?

Sinteticamente, la mediana è il valore che occupa il posto centrale in una serie di dati disposti in ordine crescente o decrescente.

Appartenenza e provenienza delle donne in lista

Come cambia, nei due anni, il rapporto tra categorie e provenienza geografica?

Si hanno sostanziali differenze? Qual è la categoria col più alto numero di donne? E la regione di provenienza? E la loro popolarità?



Scorrimi sopra il mouse!

Leader e provenienti dall'**Asia**. Ma anche **non popolari**. E' questo l'**identikit** della donna tipo che viene nominata all'interno della lista, almeno per l'anno 2015. Il 2019 ha visto una omogeneizzazione del numero di donne per categorie, arrivando ad averne lo stesso numero tanto nella categoria **Identità** che in quella **Leadership**. La **provenienza** e il grado di **popolarità**, principalmente **asiatiche** e **non popolari**, non subiscono modifiche sostanziali.

Fonti dati:

GrlPwr (icona): <https://www.freepik.com>

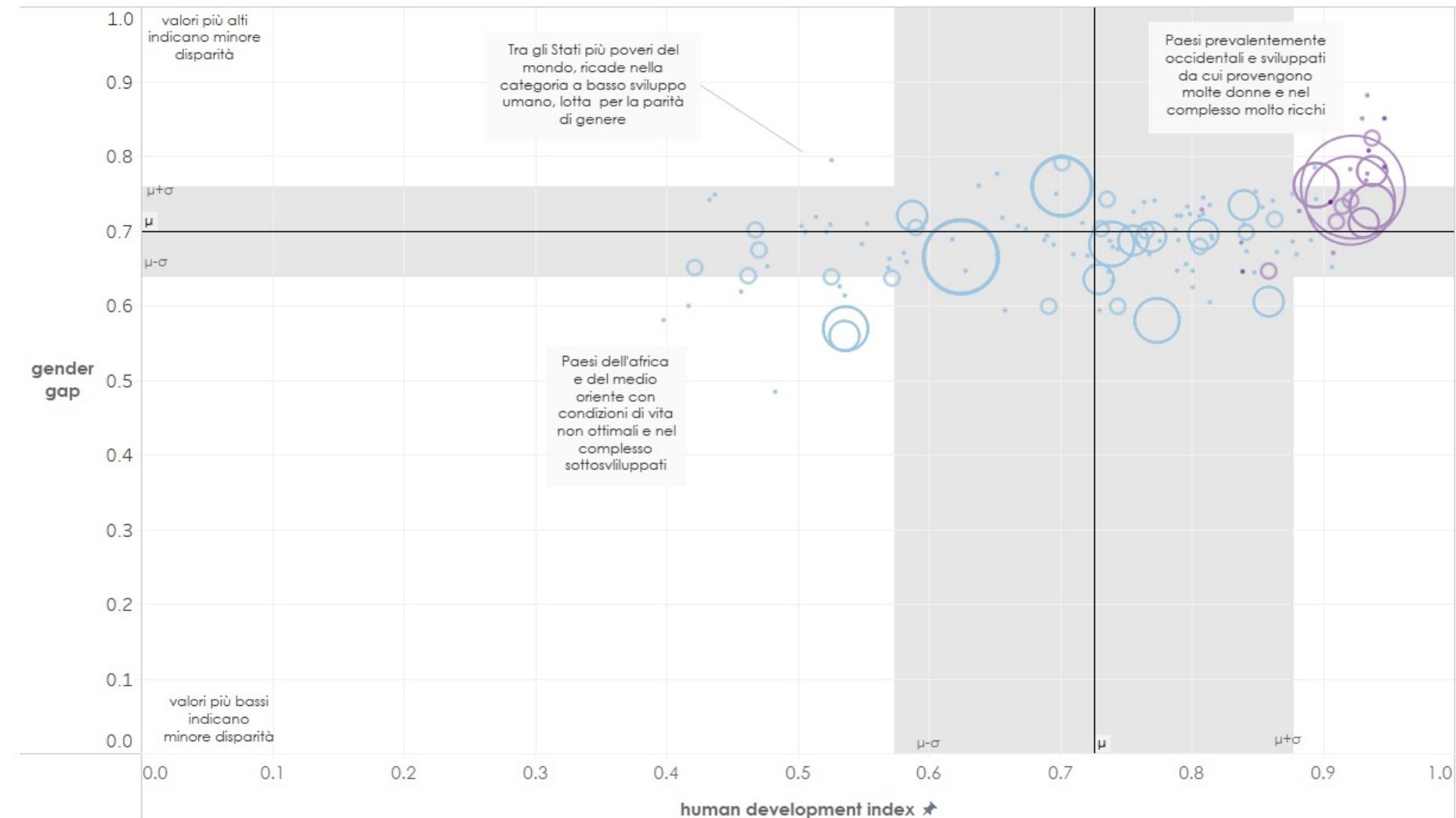
<https://www.bbc.com/news/world-50042279>

[https://en.wikipedia.org/wiki/100_Women_\(BBC\)](https://en.wikipedia.org/wiki/100_Women_(BBC))

<https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>

Da quali Stati provengono le donne nominate nella lista della BBC? Che caratteristiche hanno questi Stati?

Rapporto tra indice di disparità di genere, indice di sviluppo umano e pil pro capite



Passa con il mouse sopra gli Stati per leggerne le caratteristiche.

Puoi anche selezionare l'anno dall'elenco qui sotto e cercare differenze tra i due anni

Anno
 2015
 2019

pil pro capite
(in dollari statunitensi)
270 [color bar] 119,408

numero donne nominate per Stato

0
1
2
3
5
7
8
11

Fonti dati:

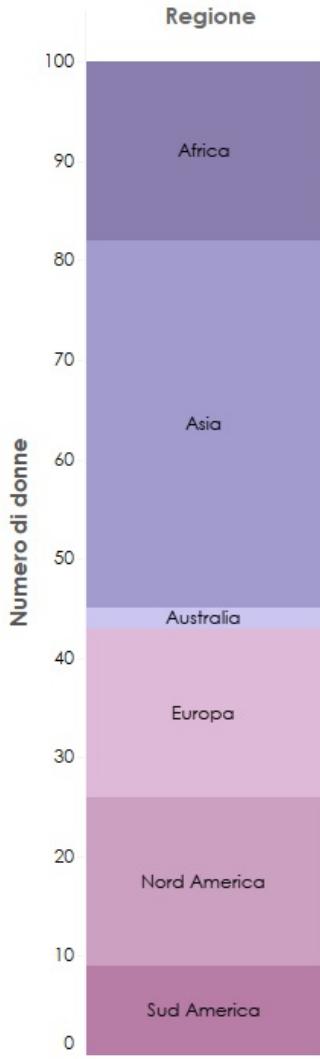
pil per capite: <https://www.imf.org/external/datamapper/NGDPDPC@WEO/OEMDC/ADVEC/WEOWORLD>

global gender gap: <https://reports.weforum.org>

human development index: <http://hdr.undp.org/en/indicators/137506>

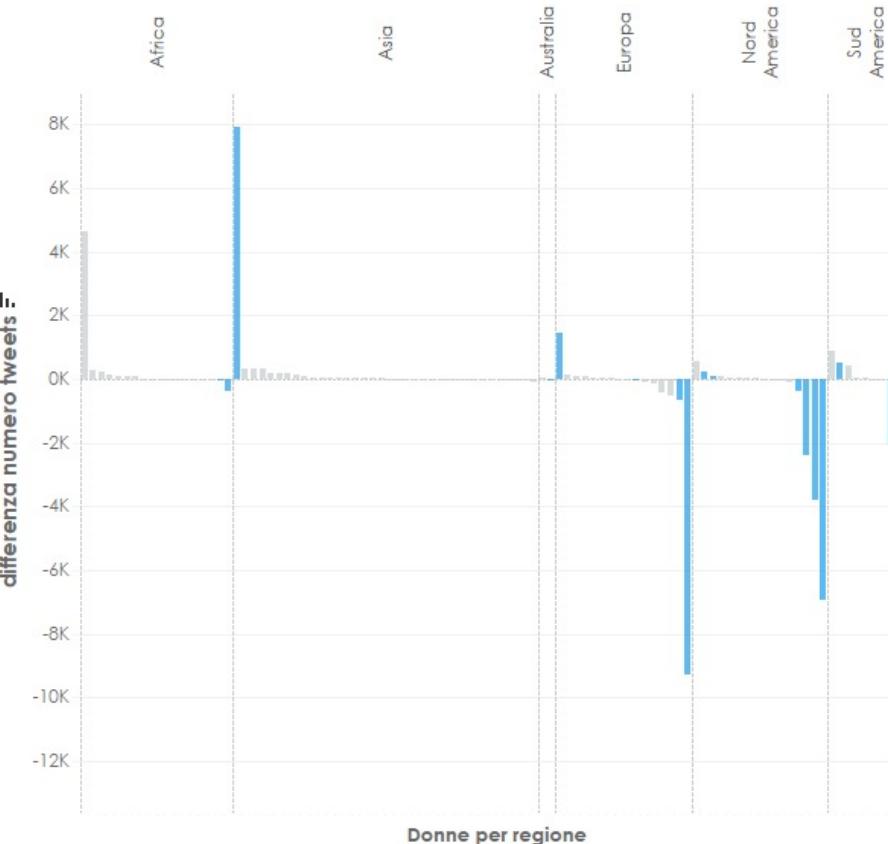
Essendo nominate dalla BBC, la popolarità delle donne è stata influenzata ?

Distribuzione della differenza del numero di tweets in cui vengono menzionate le donne, divise per regione di provenienza



Scorri qui sopra
con il mouse!

Anno
● 2015
● 2019



I tweets per ciascuna donna sono stati scaricati utilizzando come filtro di ricerca: hashtag #BBC100Women, username (menzione @), hashtag associati alla donna, nome completo

Dal grafico a barre sono state escluse tre donne, che vengono riportate nel grafico qui sotto (outliers 2019), in quanto hanno una diminuzione del numero di tweets molto elevata rispetto alle altre donne.

Outliers 2019

differenza numero tweets

Nel complesso le donne con account verificato hanno presentato una differenza nel numero di tweets in termini assoluti più alta, mentre la maggior parte delle donne con account non verificato, hanno avuto un incremento del numero di tweets che le riguardano nel periodo successivo alla pubblicazione della lista.

Nel 2019 risultano esserci più donne popolari con una differenza negativa rispetto al 2015.

L'evento ha un impatto positivo su donne meno conosciute che ottengono un boost di popolarità grazie alla nomina, mentre le donne con una certa notorietà non ne giovanano altrettanto.

Grazie per l'attenzione