

# PROGRAMMI TELEVISI IN PRIMA SERATA: CORRELAZIONE FRA AUDIENCE E TWEET

Marco Braga, Marco Maugeri, Anna Nava



# Indice

Obiettivo

Raccolta Dati

Arricchimento Dati

Infografiche

# Indice

---

## Obiettivo



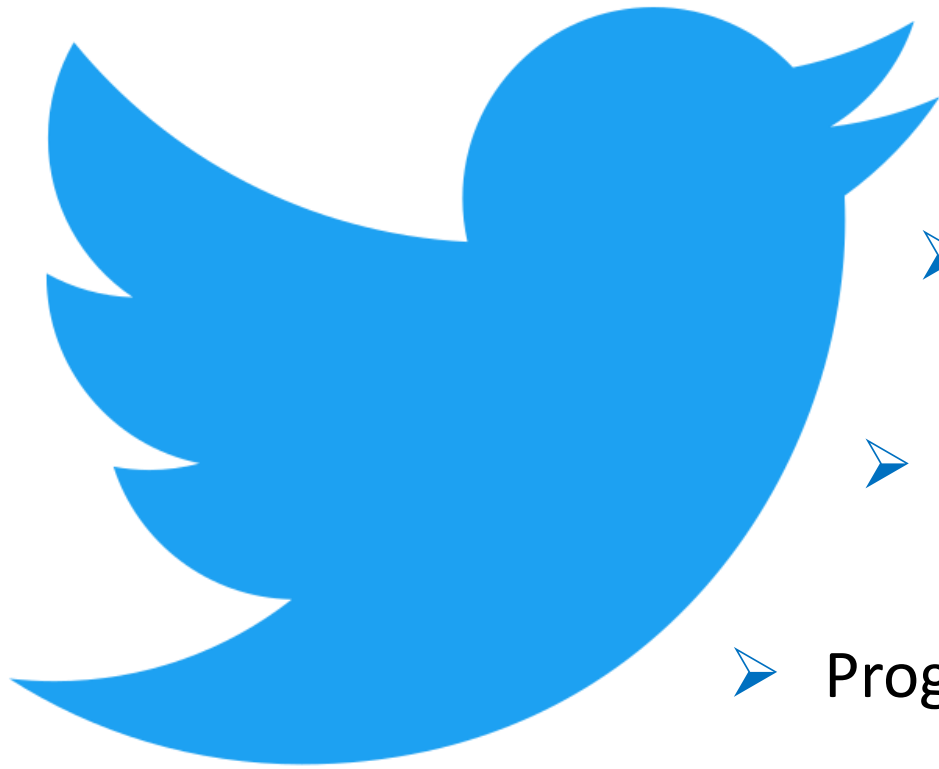
Raccolta Dati

Arricchimento Dati

Infografiche

Conclusioni

# SET THE RESEARCH GOAL

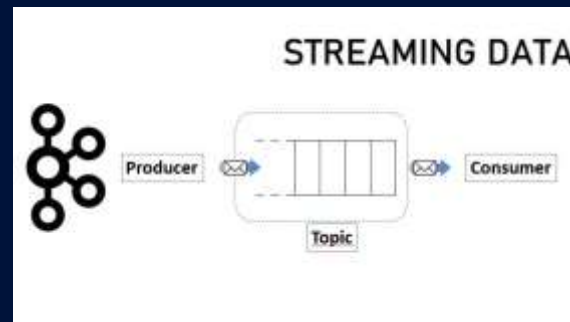


- Esistenza correlazione fra audience e tweet
- Correlazione spuria o causalità
- Programmi più social

Indice

Obiettivo

## Raccolta Dati



Arricchimento Dati

Infografiche

Conclusioni

# SCRAPING DATA



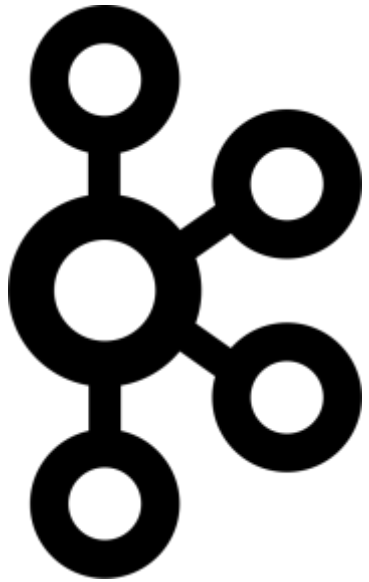
BeautifulSoup



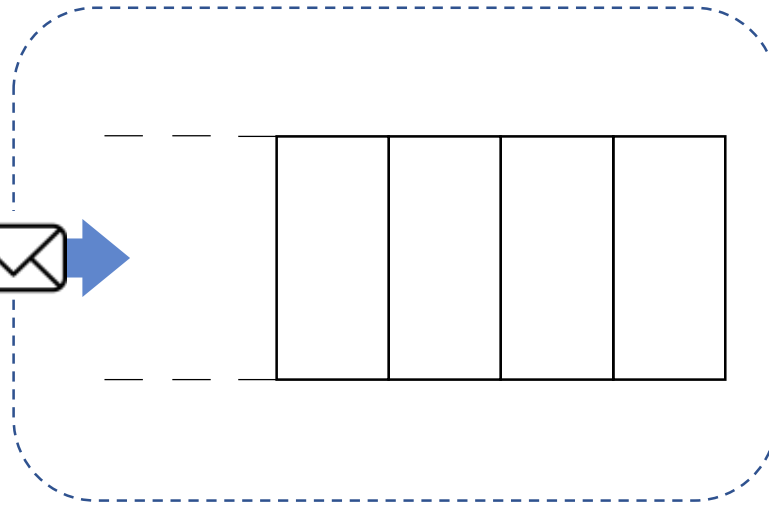
*Json*

```
[
  {
    nome_programma: "",
    data: "",
    orario: "",
    canale: "",
    parole_chiave: [
      nome_programma + "sera",
      nome_programma + "stasera",
      nome_programma + "oggi",
      nome_programma + canale
    ]
  }
  ...
]
```

# STREAMING DATA



**Producer**



**Topic**



**Consumer**

# SCRAPING DATA



**DAVIDEMAGGIO**  
*La TV dietro le quinte*



BeautifulSoup



**Nome programma**  
Spettatori  
Share

N.B. Solo i programmi più visti

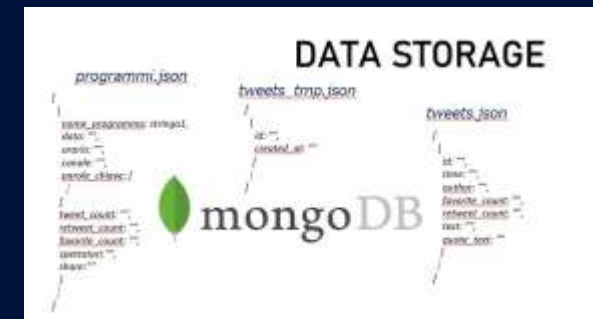
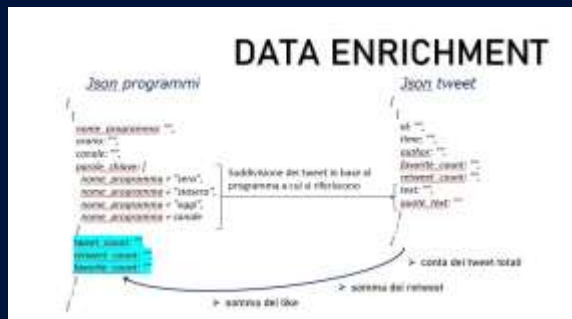


Indice

Obiettivo

Raccolta Dati

## Arricchimento Dati



Infografiche

Conclusioni

# DATA ENRICHMENT

## *Json programmi*

```
[  
  {  
    nome_programma: "",  
    data: "",  
    orario: "",  
    canale: "",  
    parole_chiave: [  
      nome_programma + "sera",  
      nome_programma + "stasera",  
      nome_programma + "oggi",  
      nome_programma + canale  
    ],  
    tweet_count: "",  
    retweet_count: "",  
    favorite_count: ""  
  }  
  :  
]
```

Suddivisione dei tweet in base al  
programma a cui si riferiscono

## *Json tweet*

```
[  
  {  
    id: "",  
    time: "",  
    author: "",  
    favorite_count: "",  
    retweet_count: "",  
    text: "",  
    quote_text: ""  
  }  
  :  
]
```

➤ conta dei tweet totali

➤ somma dei retweet

➤ somma dei like

# DATA ENRICHMENT

*Json programmi*

```
[
{
  nome_programma: stringa1,
  data: "",
  orario: "",
  canale: "",
  parole_chiave: [
    :
  ]
  tweet_count: "",
  retweet_count: "",
  favorite_count: "",
  spettatori: "",
  share: ""
}
:
]
```

Confronto con

**Fuzzywuzzy**

*Ascolti su Davide Maggio*

**nome programma: stringa1\_bis**



Spettatori

Share

Match se:

- fuzz.partial\_ratio(stringa1, stringa1\_bis) > 65
- stringa con somiglianza massima

# DATA STORAGE

*programmi.json*

```
[  
  n  
  {  
    nome_programma: stringa1,  
    data: "",  
    orario: "",  
    canale: "",  
    parole_chiave: [  
      :  
    ]  
    tweet_count: "",  
    retweet_count: "",  
    favorite_count: "",  
    spettatori:"",  
    share:""  
  }  
  :  
]
```

*tweets\_tmp.json*

```
[  
  {  
    id: "",  
    created_at: ""  
  }  
  :  
]
```

*tweets.json*

```
[  
  {  
    id: "",  
    time: "",  
    author: "",  
    favorite_count: "",  
    retweet_count: "",  
    text: "",  
    quote_text: ""  
  }  
  :  
]
```



mongoDB

Indice

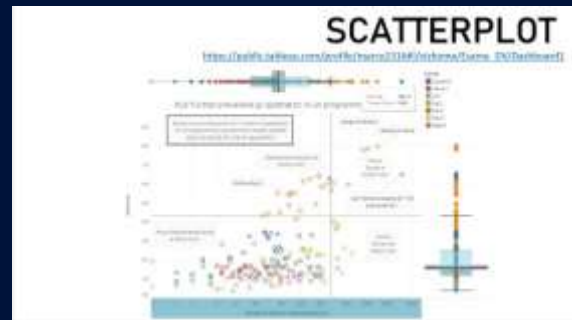
Obiettivo

Raccolta Dati

Arricchimento Dati

---

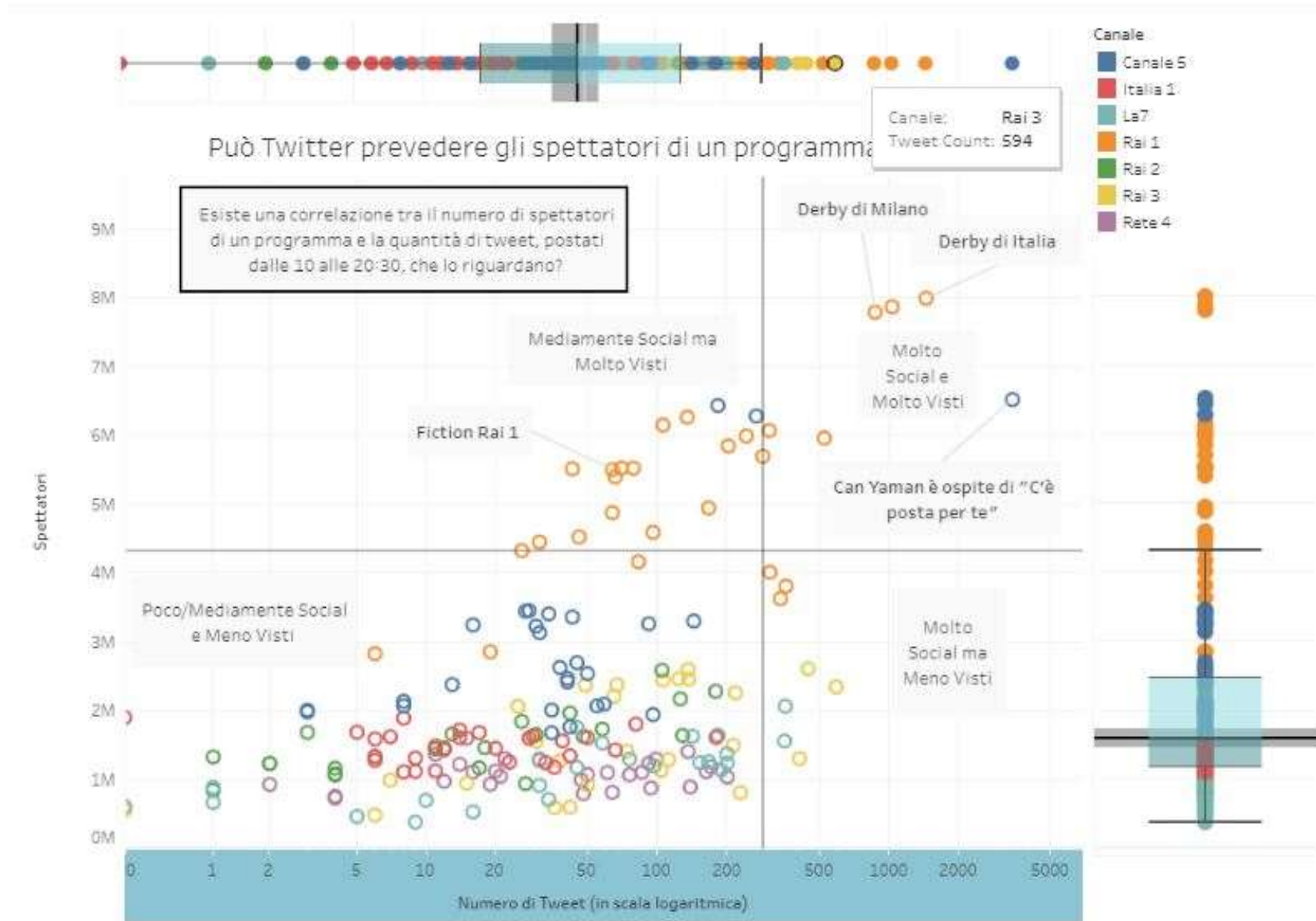
## Infografiche



Conclusioni

# SCATTERPLOT

[https://public.tableau.com/profile/marco2316#!/vizhome/Esame\\_DV/Dashboard1](https://public.tableau.com/profile/marco2316#!/vizhome/Esame_DV/Dashboard1)

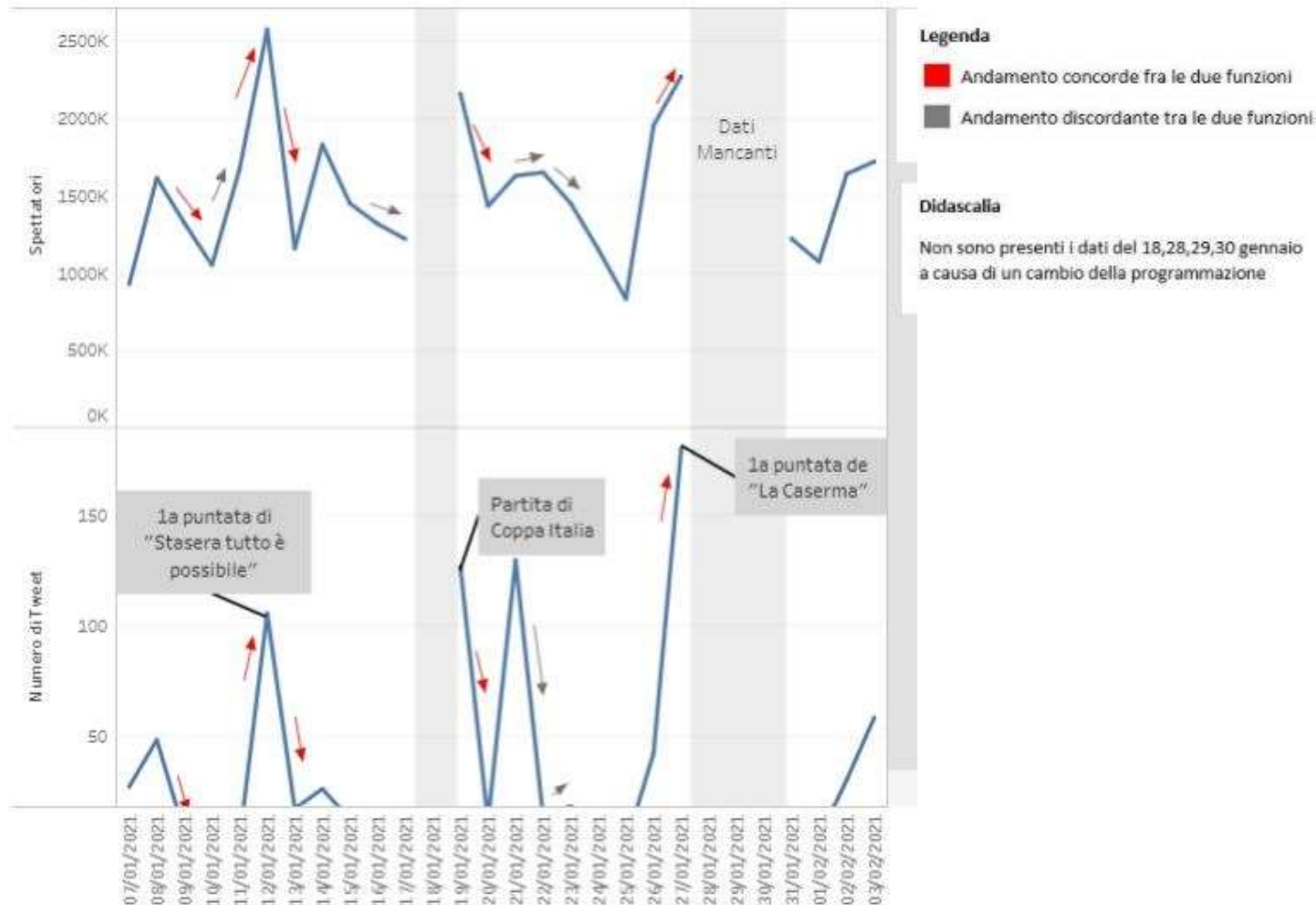


# SERIE TEMPORALI

[https://public.tableau.com/profile/anna5773#!/vizhome/Esame\\_DV\\_Doppie\\_Serie\\_temporale\\_finali\\_16128875947680/Dashboard2](https://public.tableau.com/profile/anna5773#!/vizhome/Esame_DV_Doppie_Serie_temporale_finali_16128875947680/Dashboard2)

E' davvero possibile prevedere l'audience dei programmi in prima serata su Rai 2?

Abbiamo cercato una correlazione tra il numero di spettatori di un programma e il numero di tweet, pubblicati tra le 10 e le 20:30, che lo riguardano. Questo è il risultato!



Indice

Obiettivo

Raccolta Dati

Arricchimento Dati

Infografiche

Valutazioni infografiche

Conclusioni





# CRITICITÀ

## RACCOLTA DATI

### ➤ Dati mancanti

Cambi di programma

Errori: 30/01

Mancanza dati audience

### ➤ Tweet spazzatura

Nomi programmi generici

*Stasera Italia, DiMartedì ...*

## PREVISIONE

### ➤ Differenza utenza Twitter - spettatori televisivi

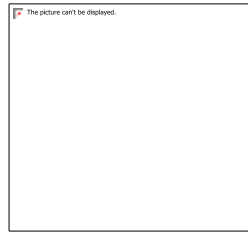
### ➤ Categorie programmi TV più social

### ➤ Previsioni con poco anticipo

# RISULTATI

➤ **Scarsa correlazione**

➤ **Programmi più social**



➤ **Effetto social degli ospiti speciali**



➤ **Serie temporali poco significative**