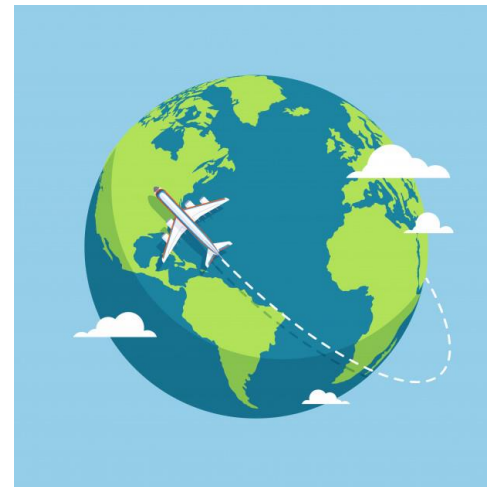


XLI CONFERENZA ITALIANA DI SCIENZE REGIONALI

Web Conference, 2-4 settembre 2020

# COMPETITIVENESS CITY INDEX FOR TRAVELLERS



Antonella Bianchino - Istat , Daniela Fusco - Istat , Daniele Pisciotano

– Univ. Federico II, Dario Volpicelli – Univ. Parthenope

- Obiettivo
- Le misurazioni del turismo
- Aspetti metodologici
- Competitiveness City Index for Travellers: i primi risultati
- Conclusioni

## Dati ufficiali



## Opinioni



Scopo del lavoro è la costruzione di un indice sintetico che consenta di valutare la competitività delle città turistiche italiane anche dal punto di vista del viaggiatore.

- Secondo il WTO, nel 2018, il numero di arrivi di turisti internazionali in tutto il mondo ha raggiunto 1,4 miliardi. L'Italia è nella top ten dei paesi con maggiore competitività turistica grazie soprattutto alle sue risorse naturali e culturali.
- Secondo l'Istat, il 2018 ha segnato un nuovo record per il turismo italiano: 428,8 milioni di clienti negli esercizi ricettivi, in crescita di due punti percentuali rispetto all'anno precedente.
- Roma è la città più gettonata con 29 milioni di presenze, seguita da Venezia e Milano (entrambe con 12,1 milioni). Per trovare una città del sud bisogna arrivare all'undicesimo posto della graduatoria con i 3,7 milioni di presenze nella città di Napoli.

La rilevazione "Movimento dei clienti negli esercizi ricettivi" viene svolta dall'Istat con periodicità mensile dal 1956 e rappresenta la principale fonte di informazione sul turismo interno disponibile in Italia. La rilevazione quantifica, per ciascun mese e per ciascun comune, gli arrivi e le presenze dei clienti (residenti e non). L'Istat provvede, inoltre, al calcolo degli indici di utilizzazione dei posti letto e delle camere delle strutture ricettive di tipo alberghiero.

Unità di rilevazione sono gli esercizi ricettivi presenti sul territorio nazionale:

- esercizi alberghieri: alberghi classificati in cinque categorie distinte per numero di stelle e residenze turistico-alberghiere;
- esercizi extra-alberghieri: campeggi, villaggi turistici, forme miste di campeggi e villaggi turistici, alloggi in affitto gestiti in forma imprenditoriale, agriturismi, ostelli per la gioventù, case per ferie, rifugi di montagna, bed and breakfast e altri esercizi ricettivi n.a.c.



Considera una tipologia di strutture ricettive non incluse tra le unità di osservazione dell'indagine Istat. Attraverso la piattaforma "Inside AirBnB" è possibile ottenere moltissime informazioni in merito agli alloggi presenti nelle più importanti mete turistiche italiane.

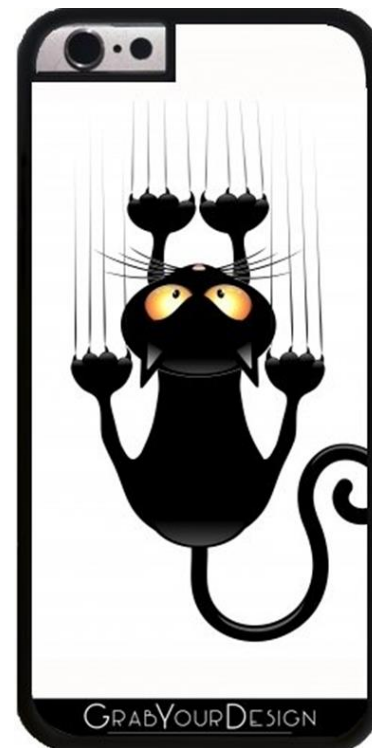
Twitter è un servizio di micro-blogging con due caratteristiche principali:

- i suoi utenti inviano messaggi (tweet) di massimo 140 caratteri composti solitamente da parole chiave (sotto forma di hashtag), linguaggio naturale e abbreviazioni comuni.
- ogni utente può seguire (follow) altri utenti affinché la propria timeline sia popolata dai loro tweet



Variabili «scrapate» su Twitter:

- La lunghezza del tweet
- Il numero di commenti
- Il numero di retweet
- Il numero di likes
- Gli hashtags
- Se è presente o meno un elemento multimediale (foto/video/gif)
- La lingua in cui è scritto
- Se attivata dall'utente, la geolocalizzazione del tweet
- Data ed ora di pubblicazione
- Numero di follower e di following del profilo che ha twittato
- Numero di tweet del profilo che ha twittato
- La geolocalizzazione del tweet
- Gli Hashtag



La Sentiment Analysis, conosciuta anche come Opinion Mining, è un campo all'interno del Natural Language Processing (NLP), il cui scopo è l'analisi di un testo con il fine di identificare e classificare l'informazione presente nello stesso. Essa ha consentito di valutare:

- Polarità: opinione positiva o negativa
- Oggetto: ciò di cui si parla
- Opinion holder: la persona o entità che esprime il parere.

In altri termini l'analisi del sentimento serve per conoscere la brand perception (dove per brand si intende qualsiasi oggetto di cui si voglia esprimere una opinione) attraverso gli scambi di interazione degli utenti nei social network o più in generale nel web.



Sono state selezionate le seguenti informazioni:

- Il testo del tweet
- Il numero di follower e di following dell'account autore del tweet
- Il numero di post pubblicati dall'account autore del tweet
- Il numero di retweet
- Il numero di like
- Il sentiment che scaturisce dal tweet

I testi sono stati sottoposti ad un importante pre-processamento, necessario per l'elaborazione degli stessi a fini statistici. Oltre agli spazi vuoti, ai collegamenti e ai simboli, sono state rimosse le così dette stop words, ossia l'insieme di parole comunemente usate in qualsiasi lingua, come congiunzioni e avverbi, che creano “rumore” nell'analisi.

## Indicatori oggettivi:

- Densità ricettiva: numero di posti letto/km<sup>2</sup> (la possibilità di una città turistica di accogliere più o meno turisti)
- Risultati economici: numero di presenze/numero di posti letto (anno) (quanto fruttano i posti letto disponibili nella città nel corso dell'anno)
- Sostenibilità sociale: numero di presenze/popolazione
- Accessibilità media ai BnB: giorni di apertura medi su 365/prezzo medio degli Air BnB

### Indicatori soggettivi:

- Popolarità della fonte: follower/account seguiti (indica quanto l'account autore del tweet sia popolare e possa di conseguenza influenzare un numero elevato di persone)
- Regolarità del profilo: numero di post pubblicati (un profilo più attivo è un profilo più autorevole, i suoi tweet sono più visibili al pubblico e risultano più credibili)
- Polarità: tweet positivi/tweet totali (permette di capire quanti turisti siano rimasti soddisfatti dalla città da loro visitata)
- Diffusione: (retweet + like)/numero di tweet totali (un tweet con un alto numero di like e condivisioni appare in diversi profili e può raccogliere consensi ed influenzare il pensiero altrui)
- Engagement del profilo: (likes + retweet)/numero di follower (la capacità di un profilo di generare interesse nei propri follower attraverso un tweet)

L'indice sintetico scelto è basato sul metodo delle penalità per coefficiente di variazione (Mazziotta Pareto index) può essere scritto, in forma generalizzata, nel seguente modo:

$$MPI_{i+/-} = M_{zi} +/ - S_{zi} cv_i$$

Dove M è la media della matrice delle osservazioni

S ne è la varianza

cv è il coefficiente di variazione.

La sintesi degli indicatori mediante il metodo proposto consente di realizzare, in modo semplice e immediato, analisi descrittive finalizzate a confronti temporali, oltre che spaziali, dello stato di fenomeni complessi.

# Le città scelte



**ROMA**



**VENEZIA**



**MILANO**



**FIRENZE**

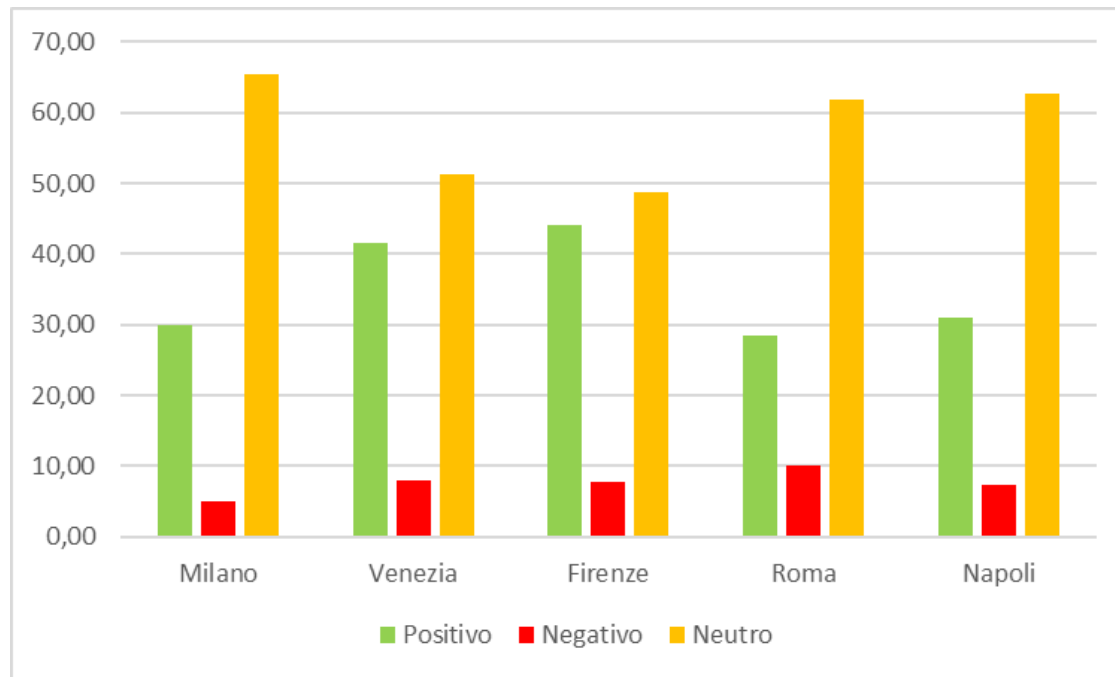


**NAPOLI**

## Indicatori oggettivi:

<b>Città</b>	<b>Densità rice ttiva</b>	<b>Risultati economici</b>	<b>Sostenibilità sociale</b>	<b>Accessibilità media ai BnB</b>
<i><b>Milano</b></i>	39,6	251,9	4,1	1,5
<i><b>Venezia</b></i>	20,4	725,1	18,6	1,9
<i><b>Firenze</b></i>	12,9	337,3	9,6	2,2
<i><b>Roma</b></i>	34,3	175,5	5,3	2,4
<i><b>Napoli</b></i>	11,8	1020,8	4	3,9

## Polarità dei tweet



## Indicatori soggettivi:

<b>Città</b>	<b>Popolarità fonte</b>	<b>Regolarità profilo</b>	<b>Polarità</b>	<b>Diffusione</b>	<b>Engagement</b>
<i>Milano</i>	7,27	22.164,34	0,29	1,53	0,02
<i>Venezia</i>	8,71	11.141,61	0,42	9,96	0,01
<i>Firenze</i>	4,39	25.316,52	0,44	1,37	0,01
<i>Roma</i>	4,6	15.930,01	0,29	5,32	0,02
<i>Napoli</i>	11,14	26.766,46	0,31	5,18	0,01



## Graduatoria



- Gli indicatori soggettivi sono complementari agli indicatori oggettivi, in quanto consentono di valutare le eventuali divergenze tra ciò che le persone riferiscono e ciò che viene catturato dagli indicatori oggettivi. La considerazione di tali indicatori permette di avere una visione più articolata e completa dei fenomeni
- Twitter rappresenta la fonte ideale per analisi di questo tipo, per la normativa sulla privacy del sito che lascia come impostazione predefinita la natura pubblica delle informazioni, ed in secondo luogo per la grossa potenzialità degli hashtag che etichettano ogni tweet in base all'argomento a cui si riferisce, dando la possibilità di riunire discussioni relative allo stesso tema, anche se avviate da utenti che non hanno alcun legame tra loro
- Necessario delimitare lo scraping a coloro che effettivamente sono turisti e non residenti che commentano la città in cui vivono.
- Utilizzo di informazioni non limitate ad una settimana, ma relative ad un intero anno, rendendo l'analisi delle opinioni degli utenti più veritiera.

# Grazie per l'attenzione!

