

XLI CONFERENZA ITALIANA DI SCIENZE REGIONALI

COMPETITIVENESS CITY INDEX FOR TRAVELLERS (CCIT)

Antonella Bianchino¹, Daniela Fusco², Daniele Pisciotano³, Dario Volpicelli⁴

SOMMARIO

Nel 2018, secondo l'Organizzazione mondiale del turismo (WTO, 2018), il numero di arrivi di turisti internazionali in tutto il mondo ha raggiunto 1,4 miliardi, il che rappresenta un enorme potenziale per il settore e le economie a livello globale. Secondo il WTO l'Italia è nella top ten dei paesi con maggiore competitività turistica grazie soprattutto alle sue risorse naturali e culturali.

Tuttavia, nel mondo globalizzato, ciò che conta per il turista è l'opinione dei viaggiatori, il numero di volte in cui viene menzionata una località e come viene considerata dall'influencer marketing.

Lo scopo del lavoro è quello di creare un indice sintetico che consenta di valutare la competitività turistica delle città italiane valutando sia i dati sulla ricettività che le opinioni dei viaggiatori. Per farlo sono stati presi da un lato i dati ufficiali dell'Istat dall'altro i Big Data, in particolare le informazioni provenienti dalla principale piattaforma di case vacanza e le opinioni dei viaggiatori espresse su Twitter. I risultati consentono di costruire una graduatoria di gradimento delle città turistiche italiane.

¹ Istat, Dipartimento per lo sviluppo di metodi e tecnologie per la produzione statistica, Potenza, e-mail: bianchin@istat.it

² Istat, Dipartimento per lo sviluppo di metodi e tecnologie per la produzione statistica, Napoli, e-mail: dafusco@istat.it (corresponding author).

³ Università degli Studi di Napoli Federico II, Napoli, e-mail: daniele.pisciottano@studenti.unina.it.

⁴ Università degli Studi di Napoli PARTHENOS, Napoli, e-mail: dario.volpicelli@studenti.uniparthenos.it.

1. Introduzione

Fino a qualche tempo fa, la scelta della destinazione di un viaggio era dettata dal passaparola, dalle offerte proposte dalle agenzie, dalle foto dei cataloghi. Con l'avvento di Internet e dei social network chi viaggia trova sempre più ispirazione dagli influencer, che con la condivisione delle proprie esperienze, attraverso un approccio di promozione partecipativa dal basso, suggeriscono le mete ai turisti.

Per questa ragione, conoscere il parere delle persone su una meta turistica fornisce le informazioni di carattere qualitativo che spingono le scelte dei viaggiatori. Da questo punto di vista i social network hanno aperto una nuova frontiera, poiché dall'analisi dei post è possibile capire, al di là dei numeri, quali siano le mete che piacciono di più e per quali ragioni.

Si consideri Twitter: il numero di commenti, il numero di retweet, il numero di likes, consentono di costruire indicatori utili a valutare quali emozioni abbia potuto stimolare una meta vacanziera, quanti utenti ne abbiano parlato, da dove provenivano e che influenza hanno all'interno del social network.

Scopo del lavoro è la costruzione di un indice sintetico che dall'analisi di queste variabili, insieme ai dati ufficiali, consenta di valutare la competitività delle città turistiche italiane anche dal punto di vista del viaggiatore. Dopo l'estrazione delle informazioni da Twitter è stata effettuata la Sentiment Analysis dei tweet per valutare le opinioni degli utenti sulle città italiane. Le altre informazioni sono state utilizzate per costruire indicatori in grado di determinare il peso tweet, la sua visibilità e, laddove possibile, la provenienza geografica.

I risultati consentono di costruire una graduatoria di gradimento delle città turistiche italiane. Per testare l'indice si è scelto di costruirlo solo sulle principali città turistiche italiane: Roma, Venezia, Milano e Firenze, oltre che la città di Napoli, prima città del mezzogiorno per presenze. Sono stati analizzati i tweet del mese di luglio, mese in cui viene registrato il picco di presenze turistiche in Italia.

2. Le misurazioni del turismo

Il 2018 ha segnato un nuovo record per il turismo italiano: 428,8 milioni di clienti negli esercizi ricettivi (Istat, 2019), 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. Per comprendere i reali meccanismi che regolano il turismo, è però necessario analizzare il fenomeno sia da un punto di vista qualitativo che quantitativo.

Da un punto di vista quantitativo l'Istat, attraverso la Rilevazione Viaggi e vacanze e la Rilevazione Movimento dei clienti negli esercizi ricettivi (Regolamento per le Statistiche del Turismo 692/2011), fornisce numerose informazioni con un dettaglio di tipo comunale.

La rilevazione "Movimento dei clienti negli esercizi ricettivi" viene svolta con periodicità mensile. Le statistiche mensili vengono elaborate regolarmente dall'Istat a partire dal 1956 e rappresentano 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) secondo la categoria di esercizio e il tipo di struttura e secondo il paese estero o la regione italiana di residenza. L'aspetto di maggior interesse dei dati risiede nella possibilità di articolare il movimento turistico dei clienti secondo tutte le possibili combinazioni delle variabili considerate, in modo da consentire un'analisi approfondita delle relazioni che intercorrono tra queste. 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, ripartiti tra strutture alberghiere e strutture extra-alberghiere:

- 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.

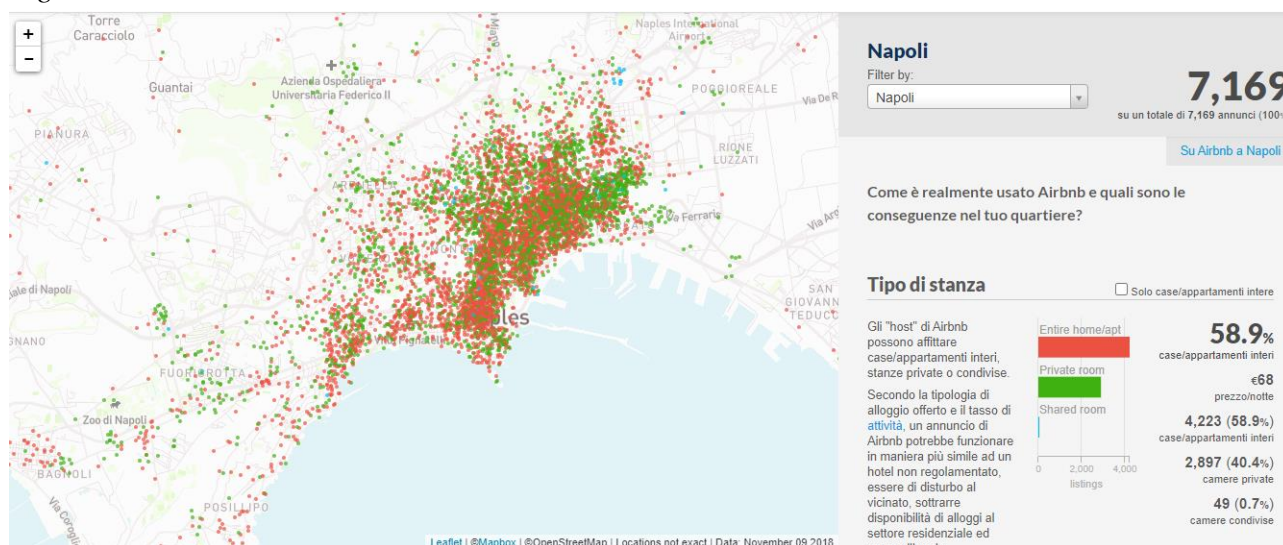
La raccolta dei dati è affidata agli organi intermedi. La rilevazione viene condotta secondo le regole contenute nelle circolari annuali dell'Istat.

Se i dati Istat permettono di avere una visione generale dell'andamento del turismo in Italia, quelli provenienti dagli AirBnB (soluzione molto adottata tra gli under 50) consentono di ottenere ulteriori informazioni sulle scelte dei turisti, considerando 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, tra cui proprio Milano, Roma, Venezia, Firenze e Napoli, oggetto dell'indagine.

Inoltre è facile rappresentare tali BnB in maniera grafica su una cartina geografica della città di riferimento (Figura 1), qualora ci sia bisogno di tener conto della localizzazione, magari per metterla in relazione con il prezzo (è lecito pensare che alloggi ubicati in centro città siano meno economici rispetto a quelli presenti in periferia).

Figura 1 – Localizzazione dinamica delle strutture AirBnB



Fonte: Inside AirBnB

3. Le carenze informative: potenzialità dei Big Data

Attraverso i soli dati provenienti dalle fonti ufficiali è possibile stimare la portata di un fenomeno esclusivamente da un punto di vista quantitativo, ma ai fini di una più accurata indagine è necessario tener conto anche di informazioni di carattere qualitativo.

È in questo contesto che ritorna utile l'utilizzo dei Big Data. Questi dati forniscono infatti informazioni complementari potenzialmente rilevanti rispetto agli standard, essendo basati su set di informazioni piuttosto diversi, sono inoltre disponibili in tempo reale (o quasi) e, generalmente, non sono soggetti a successive revisioni.

Per Big Data si intende una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore (Curry, 2016).

Il passaggio da data a Big data è possibile grazie alla crescita di potenza dei sistemi odierni e della loro capacità computazionale. Connessioni più diffuse, facili e veloci sono fondamentali per questo passaggio.

Le 3 caratteristiche principali di questa fonte di dati già accennate sono tradizionalmente chiamate le “3 V”:

- Volume, una quantità enorme di dati.
- Velocità, con cui i dati si rendono disponibili, è così elevata da rendere necessario l'utilizzo di nuovi strumenti.
- Varietà, diversità delle fonti e dei formati dovuta alla loro mancata strutturazione.

Tener conto delle “recensioni” delle persone permette di capire cosa possa attrarre di una città e quali problemi potrebbero venire risolti, dando la possibilità di avere un feedback immediato ad un costo decisamente contenuto, soprattutto tenendo conto di quanto costerebbe un'indagine del genere svolta con i metodi tradizionali.

Nonostante possano sembrare così utili, i Big Data non sono affatto strumenti senza difetti, portano infatti con sé nuove sfide a partire dall'accesso dei dati. Poiché la qualità dell'analisi dipende dalla qualità delle informazioni analizzate, una delle sfide principali nell'uso dei Big Data è proprio far fronte al potenziamento della “cura” di tali dati al fine di semplificarne l'usabilità (Freitas, 2016).

L'analisi della fonte da cui provengono i dati è pertanto fondamentale. Sfortunatamente la proprietà di molti Big Data è di fornitori privati o sono collegati ad aspetti di carattere privato, quindi una volta identificata una fonte potenzialmente utile di Big Data, bisogna valutare il costo dell'informazione, la presenza di metadati che possano fornire informazioni aggiuntive ed evitare violazioni di privacy.

Dati tradizionali e Big Data risultano in definitiva due facce della stessa medaglia, non si può pensare di condurre un'indagine basandosi sulla mutua esclusività dell'una o dell'altra, è importante trovare un compromesso che possa portare a stime sempre più precise sopperendo alle mancanze di una fonte attraverso l'altra.

4. Aspetti metodologici

Il lavoro si pone l'obiettivo di sintetizzare, attraverso un set di indicatori di carattere sia oggettivo che soggettivo, le predilezioni verso le principali mete turistiche italiane.

In questa fase sperimentale si scelto di porre l'attenzione solo su quelle che, in base ai risultati dell'indagine sul Movimento turistico dell'Istat, sono le principali mete turistiche italiane, ossia Roma, Venezia, Milano, Firenze e Napoli.

Per la graduazione delle città si è scelto di utilizzare, da un punto di vista metodologico, un indice sintetico, il Competitive City Index for Travellers (CCIT) in grado di sintetizzare gli aspetti indicati nel paragrafo 2.

L'uso degli indici sintetici per analizzare il turismo è consolidato in letteratura. Si fa riferimento in particolare al World Trade Organization che pubblica biennialmente il Travel & Tourism (T&T) Competitiveness Index. Esso confronta la competitività T&T di 140 economie e misura l'insieme di fattori e politiche che consentono lo sviluppo sostenibile del settore T&T, che a sua volta contribuisce allo sviluppo e alla competitività di un paese. Esso è costituito da 4 sotto indici (Capacità di Sviluppo, Policy, Infrastrutture, Risorse naturali e culturali), 14 domini e 90 indicatori distribuiti nei differenti domini (WTO, 2019).

Di seguito si specificano e si analizzano la fonte Big Data utilizzata, la metodologia di sintesi scelta per la costruzione del CCIT e la scelta degli indicatori.

4.1. Twitter come fonte informativa

Per ottenere dei dati rappresentativi dell'opinione popolare è utile ricorrere alle procedure di scraping dei social network.

Con scraping intendiamo l'estrazione di dati da un sito internet attraverso l'utilizzo di software o algoritmi pensati per questo utilizzo. Diversi social network sono tuttavia molto restrittivi in merito al rilascio di dati, soprattutto dopo il 2016 a seguito dello scandalo di Cambridge Analytica che coinvolse Facebook. Da allora è pressappoco impossibile ricavare dati dalle piattaforme appartenenti a Mark Zuckerberg come Facebook o Instagram, tuttavia c'è ancora un social network piuttosto disponibile in merito allo scraping: Twitter.

Twitter infatti si presta perfettamente a questa procedura visto il limite dei caratteri molto circoscritto e i suoi post prettamente testuali. Inoltre, è possibile richiedere accesso alle API tramite un form compilabile direttamente sul sito, senza dover ricorrere a terze parti.

Il form richiede una compilazione molto specifica e numerosi dati da inserire, ma in tempi ragionevoli, e con le giuste motivazioni, è possibile ottenere l'accesso all'account developer, che permette di sviluppare la propria app di scraping.

È tuttavia necessario specificare che l'accesso alle API gratuite soffre di diverse limitazioni, tra cui, particolarmente di ostacolo all'indagine, quella della possibilità di accesso ai soli tweet risalenti al massimo ai 7 giorni precedenti alla ricerca.

In questo caso tutte le procedure informatiche come lo scraping, l'analisi del sentiment, ecc. sono state effettuate attraverso degli algoritmi scritti in Python (Mitchell, 2018). Python è un linguaggio di programmazione orientato ad oggetti nato con lo scopo di essere di più facile comprensione ed utilizzo rispetto ai propri concorrenti come Java.

La sua forza risiede proprio nella sua versatilità e facilità di apprendimento, è inoltre dotato di una community di utenti piuttosto vasta che propone librerie open source capaci di assolvere a numerosissime funzioni.

Ci sono diverse variabili osservabili in un tweet, non tutte sono fondamentali nella fase di raccolta dati, ogni variabile può essere più o meno utile a seconda dello scopo perseguito.

È comunque utile analizzarle:

- La lunghezza del tweet (attualmente il numero massimo di caratteri è 280)
- 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

Attraverso un algoritmo scritto in Python è stato possibile raccogliere tutte queste informazioni da un numero cospicuo di tweet, selezionandoli in base alla città grazie a due fattori:

- La geolocalizzazione del tweet
- Gli Hashtag

È stato utile per la ricerca tenere in considerazione anche l'utilizzo degli hashtag in merito alla città di riferimento in quanto diversi tweet non sempre risultavano provenire dalla città geolocalizzata (ad esempio alcuni tweet geolocalizzati a Napoli riguardavano in realtà Sorrento).

Tramite due librerie Python, TextBlob e Pandas, è stato possibile effettuare l'analisi del sentiment proveniente da ogni tweet.

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. Di solito, oltre a identificare l'opinione, questi sistemi estraggono gli attributi dell'espressione come:

- 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.

È stato necessario selezionare le informazioni più utili tra quelle estratte, in modo da costruire gli indicatori sintetici soggettivi che verranno successivamente analizzati, tali informazioni riguardano:

- 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.

4.2 Gli indici sintetici

La definizione comune di indice è quella di un valore che esprime la sintesi di diversi indicatori, riassumendo informazioni più complesse ed analizzando le relazioni intercorrenti fra diversi fenomeni.

È però interessante approfondire la nozione di indice sintetico. Quest'ultimo si occupa di misurare il livello di un fenomeno complesso, non direttamente misurabile applicando una combinazione di indicatori elementari.

Uno degli indici sintetici più importanti sviluppati in Istat negli ultimi anni è il BES, Bisogno Equo Sostenibile, costruito con l'obiettivo di valutare il progresso della società non soltanto dal punto di vista economico, ma anche sociale ed ambientale. L'analisi dei 130 indicatori che compongono il BES, mira a rendere il Paese maggiormente consapevole dei propri punti di forza e delle difficoltà da superare per migliorare la qualità della vita dei cittadini, ponendo tale concetto alla base delle politiche pubbliche e delle scelte individuali (Istat, 2019).

È interessante citare in questo contesto un recente lavoro, il Semantic Brand Score (SBS) (Fronzetti, 2018). Tale indice sintetico misura l'importanza di un brand qualora sia possibile analizzare fonti di dati testuali (particolarmente improntato verso i big data).

Sfruttando la teoria dei grafi, il text mining e la social network analysis, tale misura combina 3 indicatori fondamentali (effettuando una somma standardizzata tra di loro):

- prevalenza (misura quanto un brand è menzionato in un discorso)
- diversità (eterogeneità delle associazioni testuali del brand)
- connettività (il potere di connessione del brand, quanto è nel cuore del discorso)

Tutti gli indicatori considerati hanno la stessa importanza e solo combinati sono utili a misurare l'importanza di un brand. Ogni parola può essere così trasformata in un brand ed analizzata in modo da assegnarle un valore, tramite codici già esistenti e importabili in Python.

Com'è noto, la metodologia applicabile per la costruzione di un qualsiasi indice sintetico prevede le seguenti fasi (OECD, 2008):

- a) definizione del fenomeno oggetto di studio
- b) selezione degli indicatori elementari;
- c) standardizzazione degli indicatori elementari;
- d) ponderazione e aggregazione degli indicatori standardizzati in uno o più indici sintetici

La costruzione di un indice sintetico avviene secondo diverse metodologie, una delle quali è il metodo Mazziotta-Pareto Index (MPI) che sintetizza un insieme di indicatori non completamente sostituibili.

Esso si basa su una funzione non lineare che, partendo dalla media aritmetica degli indicatori normalizzati, introduce una penalità per le unità con valori sbilanciati degli indicatori (De Muro, 2011).

Esistono due versioni dell'indice, MPI e MPI aggiustato (AMPI), la prima versione è più indicata per un'analisi "statica" (ad esempio annuale), mentre la seconda è più utile in caso di analisi "dinamiche" (ad esempio pluriennale). L'indice è espresso dalla seguente formula:

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

Dove M è la media della matrice delle osservazioni

S ne è la varianza

cv è il coefficiente di variazione

Se l'indicatore è di tipo crescente o positivo, ossia se a variazioni crescenti dell'indicatore corrispondono variazioni positive del fenomeno (per esempio, lo sviluppo di un'area geografica), si utilizza la versione con penalità negativa. Viceversa, se l'indicatore è di tipo decrescente o negativo, ossia se a variazioni crescenti dell'indicatore corrispondono variazioni negative del fenomeno (per esempio, la povertà di un'area geografica), si ricorre alla formula con penalità positiva. Nel primo caso, il coefficiente di penalità corregge la media degli indicatori standardizzati "spingendola" verso il basso, mentre nel secondo la corregge "spingendola" verso l'alto.

L'indicatore composito dovrebbe idealmente misurare multi-concetti dimensionali che non possono essere catturati da un singolo indicatore.

Numerosi sono i metodi disponibili per la sintesi di indicatori elementari. Gli applicativi che consentono il calcolo di tali indici di sintesi sono solitamente legati alla loro implementazione. Per tale motivo, il confronto tra i risultati ottenuti con un metodo piuttosto che con un altro generalmente non è agevole.

In questo lavoro è stato utilizzato il software i.Ranker, un sistema on line sviluppato in Istat. Esso consente una computazionale user friendly ed un'adeguata valutazione, comparando i risultati prodotti attraverso più metodi di sintesi statistica degli indicatori elementari riconosciuti in letteratura.

Il software prende in considerazione e consente di implementare otto diversi metodi: il metodo Mazziotta-Pareto Index (MPI) nelle due varianti (positivo e negativo); il metodo tassonomico di Wroclaw (Wroclaw); la media della media dei valori standardizzati (M1Z); il metodo delle graduatorie (Grad.RNK); 2 il metodo degli indici relativi (IR); il metodo della media aritmetica dei numeri indici base media (ANIM); il metodo della media geometrica dei numeri indici base media (GNIM); il metodo della media quadratica dei numeri indici base media (QNIM).

4.3 La scelta degli indicatori

Una volta raccolti i dati provenienti dalle fonti istituzionali ed i Big Data provenienti da Twitter, sono stati costruiti gli indicatori semplici divisi in due categorie: soggettivi ed oggettivi.

Gli indicatori soggettivi sono stati costruiti attraverso i dati raccolti da Twitter, in modo da ottenere un feedback diretto dei turisti in merito alle città da loro visitate, di seguito la loro descrizione:

- 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)

Gli indicatori oggettivi sono frutto dei dati raccolti attraverso le indagini Istat e le informazioni presenti sul sito "Inside AirBnB":

- Densità ricettiva: numero di posti letto/km² (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

Come già descritto nei paragrafi precedenti, tali indicatori hanno contribuito alla costruzione del CCIT.

5. Competitive City Index for Travellers: i primi risultati

I risultati degli indicatori oggettivi mostrano che le città con maggiore densità ricettiva sono quelle con maggiore flusso turistico, ossia Milano e Roma, mentre la città con meno posti letto per km² è Napoli (Tabella 1).

In merito all'indicatore di sostenibilità sociale è la città di Venezia ad avere il maggiore flusso turistico rispetto alla popolazione residente, situazione che ingolfa evidentemente le attività ordinarie dei residenti. Interessante è l'analisi dei risultati economici raggiunti: la città di Napoli, pur avendo un minore quantitativo di turisti, ha contato, nel 2018, ben 1.021 presenze per ciascun posto letto, risultato che conferma l'andamento di crescita turistica che la città sta vivendo negli ultimi tempi.

Tabella 1 – Indicatori di carattere oggettivo per città, anno 2018

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

Fonte: Nostra elaborazione su dati Istat e AirBnb

L'accessibilità media ai BnB mostra che è ancora la città di Napoli ad essere la più competitiva in relazione alla media dei giorni di strutture occupate rispetto al prezzo della singola struttura (3,8).

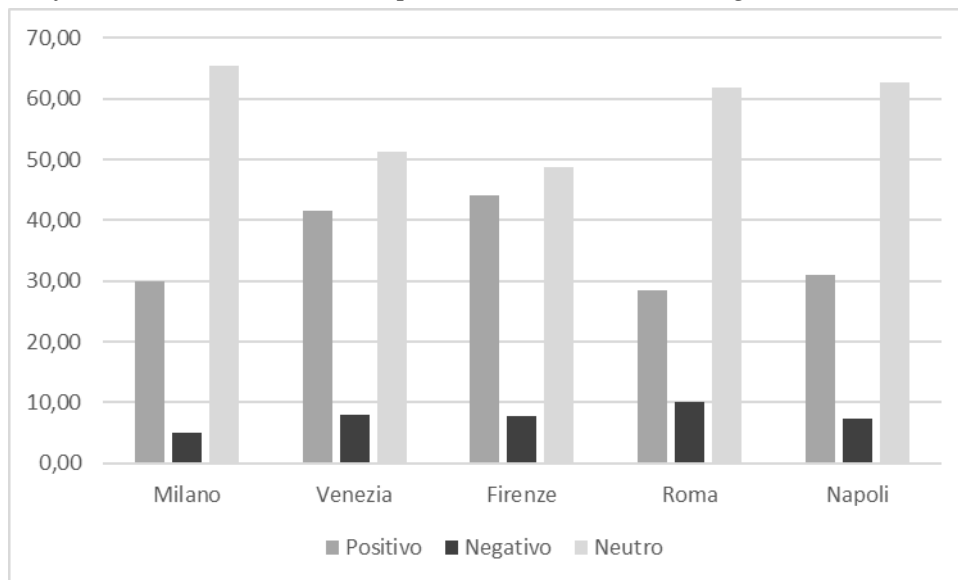
Per l'analisi del sentiment sono stati presi in considerazione i tweet pubblicati nella prima settimana di luglio 2020. Come già specificato nei paragrafi precedenti, la scelta del mese è dipesa dal fatto che luglio è il mese di maggiore affluenza per il turismo italiano. Ovviamente, nell'anno considerato, l'avvento della pandemia ha portato una riduzione del flusso turistico sia locale, ma soprattutto straniero. Per tale ragione si è deciso di focalizzare l'attenzione sui tweet in lingua italiana.

In totale sono stati analizzati circa 1.000 tweet, relativi prevalentemente alla città di Milano (36%), seguita da Roma (24%), Firenze (17%) e Napoli e Venezia, entrambe con l'11%.

Considerando la polarità dei tweet, ossia come sono state considerate le città visitate nel periodo considerato, essa tende prevalentemente alla neutralità. Infatti, se si considera il totale dei tweet, nel 60% dei casi gli utenti si sono espressi con parole che non esprimono né entusiasmo, né di disapprovazione.

Le città per le quali gli utenti hanno mostrato maggiori sentimenti positivi sono Firenze (44%) e Venezia (42%), mentre quella per cui prevale un sentimento negativo rispetto al totale dei tweet è Roma (10%) (Grafico 1).

Grafico 1 – Polarità dei tweet per città, 1 settimana di luglio 2020



Fonte: Nostra elaborazione su dati Twitter

La tabella 2 mostra i risultati degli indicatori soggettivi considerati. La polarità è stata sintetizzata in un unico indicatore che esprime il numero dei tweet positivi sul totale.

Partendo dal primo indicatore, popolarità della fonte, è chiaro che gli utenti più popolari, nel periodo considerato abbiano menzionato la città di Napoli. Si tratta anche degli utenti più autorevoli, poiché dall'indicatore Regolarità profilo emerge che gli utenti considerati hanno anche una maggiore presenza sul social network in termini di tweet pubblicati. I tweet maggiormente retweettati, e che quindi hanno raggiunto un maggior numero di utenti, sono quelli relativi alla città di Venezia, seguita da Roma e Napoli. In ogni caso i tweet considerati non hanno suscitato un grande interesse (indicatore Engagement); leggermente superiori valori per le città di Milano e Roma.

Tabella 2 – Indicatori soggettivi per città, 1 settimana di luglio 2020

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

Fonte: Nostra elaborazione su dati Twitter

La sintesi degli indicatori ottenuta con il MPI index con penalità negativa, descritto nel paragrafo 4.2, consente di ottenere la graduatoria finale delle città italiane in base al CCIT. A tutti gli indicatori è stato dato segno positivo, poiché contribuiscono positivamente all'indice sintetico. L'unico a cui è stato dato segno negativo è la sostenibilità sociale poiché valori elevati dell'indicatore mostrano la presenza di un flusso turistico che potrebbe non essere sopportabile per la popolazione residente.

Nel periodo analizzato è la città di Napoli la meta preferita dai viaggiatori, seguita nell'ordine da Milano, Roma, Venezia e Firenze. È interessante sottolineare che tutti i metodi applicati (Metodo delle graduatorie, Media dei valori standardizzati, Media degli indici relativi, Metodo MPI-, Metodo tassonomico di Wroclaw) portano la città di Napoli sempre prima in graduatoria e le graduatorie ottenute sono del tutto simili.

6. Conclusioni

Il lavoro mostra come aspetti di carattere soggettivo, grazie all'alto valore informativo e analitico, possano contribuire ad analizzare fenomeni quali il turismo.

Gli indicatori soggettivi sono complementari agli indicatori strettamente 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.

Nel caso specifico, in base ai dati ufficiali, è Roma la principale meta turistica italiana, seguita da Milano e Venezia. Considerando anche le opinioni dei viaggiatori la prima città è Napoli, undicesima se si considera solo il flusso turistico.

Nell'analisi descritta il numero degli indicatori scelti è esiguo, questo a causa delle limitazioni poste dall'utilizzo delle API gratuite nello scraping dei dati.

Una versione a pagamento consentirebbe di utilizzare un numero maggiore di variabili e quindi di costruire ulteriori indicatori soggettivi. Inoltre consentirebbe di scaricare informazioni non limitate ad una settimana, ma relative ad un intero anno, rendendo l'analisi delle opinioni degli utenti più veritiera. In questo modo sarebbe possibile utilizzare gli indicatori soggettivi riferiti allo stesso periodo di riferimento di quelli oggettivi, che inevitabilmente risentono di ritardi di diffusione rispetto al periodo di rilevazione.

In ultima analisi, per lo sviluppo futuro del CCIT sarà necessario delimitare lo scraping a coloro che effettivamente sono turisti e non residenti che commentano la città in cui vivono. L'informazione sulla città di residenza non è tuttavia sempre espressa dagli utenti e ciò potrebbe rappresentare un limite per la ricerca.

In ogni caso, i risultati della ricerca hanno dimostrato che Twitter rappresenta la fonte ideale per analisi di questo tipo, innanzitutto 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.

7. Bibliografia

- Istat (2019). Movimento turistico in Italia, anno 2018. Statistica report – 7 novembre 2019. Roma.
- Istat (2019). *BES 2019 – Il benessere equo sostenibile in Italia*. Roma, 2019.
- WTO (2019). *The Travel & Tourism Competitiveness Report 2019*. Published by the World Economic Forum's Platform for Shaping the Future of Mobility. Ginevra.
- Mitchell R. (2018). *Web Scraping with Python. Collection more data from the modern web*. O'Reilly Media, Inc., Sebastopoli USA, 2018.
- Fronzetti Colladon A. (2018). *The Semantic Band Score*. Journal of Business Research Volume 88, July 2018, Pages 150-160.
- Curry E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe*. Jose' Mari'a Cavanillas Edward Curry Wolfgang Wahlster Editors. Ginevra: Springer Open.
- Freitas A. Curry E. (2016). Big Data Curation. *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe*. Jose' Mari'a Cavanillas Edward Curry Wolfgang Wahlster Editors. Ginevra: Springer Open.
- De Muro, P., Mazziotta, M., Pareto, A. (2011) *Composite Indices of Development and Poverty: An Application to MDGs*. Soc Indic Res 104, 1–18 (2011).
- OECD (2008). *Handbook on Constructing Composite Indicators METHODOLOGY AND USER GUIDE*. OECD publications.

ABSTRACT

In 2018, according to the World Tourism Organization (WTO), the number of arrivals of international tourists worldwide reached 1.4 billion, which represents enormous potential for the sector and global economies. According to the WTO, Italy is in the top ten of the countries with the greatest tourist competitiveness thanks above all to its natural and cultural resources.

Today, in the globalized world, tourists are pressed by the opinion of travelers, the number of times that a location is mentioned and in which way influencers marketing consider it.

The aim of the work is to create a synthetic index that allows to evaluate the tourist competitiveness of Italian cities by evaluating both the data on the receptivity and the opinions of travelers. To do this, the official data of Istat have been taken together with Big Data, in particular information from the main holiday home platform and the opinions of travelers expressed on Twitter. Subjective and objective indicators have been produced. The results allow you to build a rating list of Italian tourist cities