

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

SCUOLA DI SCIENZE

Corso di Laurea in Informatica per il Management

**Social Media Analysis:**  
**Progettazione ed implementazione di  
piattaforme per la brand comparison ed  
applicazione in ambito turistico**

**Relatore:**  
**Chiar.mo Prof.**  
**MARCO DI FELICE**

**Presentata da:**  
**FEDERICA MERENDI**

**Correlatore:**  
**Chiar.mo Prof.**  
**MARCELLO MARIANI**

**Sessione I**  
**Anno Accademico 2017-2018**



*Alle mie nonne*



# Introduzione

Al giorno d'oggi il web e i social media svolgono un ruolo cruciale nella vita di tutti i giorni. Secondo i dati Eurostat relativi all'utilizzo di Internet nell'Unione Europea nel 2016 [14] l'87% dei cittadini europei tra i 16 e i 74 anni, la percentuale aumenta al 96% se ci riferiamo alla sola Italia, ha utilizzato Internet almeno una volta al giorno. Il 78% degli italiani in particolare ha utilizzato Internet per l'invio e la ricezione di e-mail, il 70% per visualizzare contenuti video, il 60% ha utilizzato social network e letto notizie online e, infine, il 40% ha utilizzato servizi online per viaggi e alloggi.

I dati, quindi, non solo evidenziano come Internet sia effettivamente utilizzato su base quotidiana, ma sottolineano anche come il digitale stia progressivamente prendendo il posto dell'analogico. Uno sguardo più attento ai dati relativi al solo utilizzo dei social network mette in evidenza come siano i più giovani tra i 16 e i 24 anni, in particolare l'85% di essi, a frequentare maggiormente le piattaforme social.

La nascita dei social media ha posto l'utente al centro dell'attenzione permettendogli di far sentire la propria opinione attraverso commenti, post e tweet e attraverso i 'mi piace' o i 'non mi piace' al contenuto visualizzato. È da sottolineare, inoltre, come gli abbia dato la possibilità di crearne a sua volta.

In questo quadro è evidente come le imprese abbiano la possibilità di utilizzare il web e i social media per allargare il proprio raggio di influenza, ottenere riscontri diretti dai clienti e aumentarne la fidelizzazione. Solo il 44% delle imprese italiane [15], però, fa utilizzo di questi strumenti e l'utilizzo non

è sempre ottimale. Le imprese devono, quindi, analizzare il proprio utilizzo dei social media e il pubblico a cui si riferiscono per capire quali siano i modi migliori per instaurare un rapporto duraturo con chi li segue e definire una strategia adeguata. Ad esempio, se un utilizzo dei contenuti multimediali ha generalmente un impatto positivo sul cliente rispetto a un contenuto solo testuale, l'impresa dovrebbe iniziare ad utilizzare più contenuti del primo tipo per coinvolgere maggiormente l'utente e aumentare la sua fidelizzazione al brand.

Uno degli obiettivi di questo elaborato è realizzare una piattaforma software che fornisca alle imprese un metodo semplice e immediato per analizzare la propria performance sui social media, in particolare in questa sede si utilizzeranno Facebook e Twitter. La piattaforma permetterà agli utenti di inserire un account da analizzare e dopo l'elaborazione mostrerà sia l'analisi dell'account, che quella dei post o tweet estratti. Infatti, metterà in evidenza tramite grafici quali siano i mesi o i giorni migliori per la pubblicazione e quale lunghezza e tipologia di contenuti sia la più apprezzata dai propri contatti. La piattaforma permetterà infine il confronto diretto tra due diversi account, per poter analizzare il proprio comportamento online in relazione a quello dei concorrenti.

In secondo luogo, si fornirà un esempio applicativo analizzando gli account ufficiali dedicati alla promozione turistica delle regioni italiane per poter estendere gli studi effettuati al riguardo anche al social media Twitter. L'argomento, infatti, è già stato affrontato in *"Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations"* da Mariani et al. [37] in cui però la ricerca è circoscritta al social network Facebook.

Il primo capitolo dell'elaborato è dedicato alla presentazione dello stato dell'arte, verranno inizialmente introdotti i *big data*, in cui rientrano gli innumerevoli dati generati dai social media, poi si tratteranno i social media e il Web 2.0 ponendo l'accento sui due in analisi: Facebook e Twitter. Verranno poi illustrati i rapporti tra imprese e social media, con particola-

re riferimento al turismo e infine si completerà lo studio analizzando alcune delle piattaforme per l'analisi dei social media esistenti.

Il secondo capitolo tratterà la progettazione della piattaforma software, con particolare attenzione alle specifiche progettuali e l'architettura del progetto, presentando la piattaforma realizzata.

Il terzo capitolo è focalizzato sull'implementazione del progetto, verranno inizialmente illustrate le tecnologie utilizzate per lo sviluppo e successivamente si discuteranno i dettagli implementativi relativi all'estrazione dei dati e alla visualizzazione degli stessi.

Infine, il quarto e ultimo capitolo presenterà l'analisi degli account ufficiali per la promozione turistica delle regioni italiane, in particolare, verranno presentate le modalità di raccolta dei dati e le metriche utilizzate. Successivamente, si analizzeranno i dati estratti da un punto di vista quantitativo e si analizzeranno i risultati della regressione sui dati. Infine, si confronteranno i risultati ottenuti sui due social media in esame per poter stabilire quali siano le buone prassi nell'utilizzo degli stessi.



# Indice

<b>Introduzione</b>	<b>i</b>
<b>Elenco delle figure</b>	<b>xi</b>
<b>Elenco delle tabelle</b>	<b>xi</b>
<b>1 Stato dell'arte</b>	<b>1</b>
1.1 Big data . . . . .	1
1.2 Social media . . . . .	3
1.2.1 Facebook . . . . .	4
1.2.2 Twitter . . . . .	5
1.3 Imprese e Social Media . . . . .	6
1.3.1 Social Media e Turismo . . . . .	8
1.4 Tool per <i>brand comparison</i> sui social media . . . . .	11
<b>2 Progettazione</b>	<b>15</b>
2.1 Specifiche del progetto . . . . .	15
2.2 Architettura del progetto . . . . .	16
2.2.1 Visualizzazione dati di una pagina Facebook . . . . .	17
2.2.2 Confronto tra due pagine Facebook . . . . .	18
2.2.3 Visualizzazione dati di un profilo Twitter . . . . .	19
2.2.4 Confronto tra due profili Twitter . . . . .	20
<b>3 Implementazione</b>	<b>23</b>
3.1 Tecnologie utilizzate . . . . .	23

---

3.1.1	Python . . . . .	23
3.1.2	Javascript . . . . .	25
3.1.3	HTML, CSS e Bootstrap . . . . .	26
3.1.4	MySQL . . . . .	26
3.1.5	Facebook API . . . . .	27
3.1.6	Twitter API . . . . .	29
3.2	Dettagli implementativi . . . . .	31
3.2.1	Estrazione dei dati . . . . .	31
3.2.2	Visualizzazione dei dati . . . . .	36
<b>4</b>	<b>Analisi dei dati</b>	<b>43</b>
4.1	Modalità di raccolta dei dati . . . . .	43
4.2	Metriche utilizzate nell'analisi . . . . .	46
4.3	Analisi quantitativa dei dati . . . . .	48
4.3.1	Analisi quantitativa dati di Facebook . . . . .	48
4.3.2	Analisi quantitativa dati di Twitter . . . . .	59
4.4	Analisi della regressione sui dati . . . . .	69
4.4.1	Regressione sui dati di Facebook . . . . .	71
4.4.2	Regressione sui dati di Twitter . . . . .	77
4.5	Confronto analisi tra Facebook e Twitter . . . . .	81
	<b>Conclusioni</b>	<b>83</b>
	<b>Bibliografia e Sitografia</b>	<b>92</b>

# Elenco delle figure

1.1	Le 4V dei big data. Fonte: IBM (www.ibmbigdatahub.com)	2
1.2	Utilizzo dei social media da parte delle imprese in base alla tipologia di social. Dati Eurostat.	6
1.3	Scopo utilizzo social media da parte delle imprese. Dati Eurostat.	7
2.1	Pagina principale.	16
2.2	Visualizzazione dati di una singola pagina Facebook.	17
2.3	Dettaglio visualizzazione interazioni dal 10 agosto 2017 al 20 agosto 2017 per la pagina dell'Alto Adige.	18
2.4	Visualizzazione dati per il confronto tra pagine Facebook.	19
2.5	Visualizzazione dati di un singolo profilo Twitter.	19
2.6	Dettaglio visualizzazione interazioni dal 15 marzo 2018 al 4 aprile 2018 per il profilo della Sicilia.	20
2.7	Visualizzazione confronto tra due profili Twitter.	21
3.1	API Graph Explorer.	29
3.2	Richiesta dati API Graph di Facebook.	31
3.3	Parsing dati relativi alla pagina Facebook.	32
3.4	Inserimento dati pagina Facebook nel database.	32
3.5	Parsing dati relativi ai post di una pagina Facebook.	33
3.6	Inserimento dati post Facebook nel database.	33
3.7	<i>Stored procedure</i> per il calcolo dell'engagement dei post di una pagina Facebook.	34

---

3.8	Richiesta dati Twitter API. . . . .	35
3.9	Algoritmo per effettuare più chiamate alla Twitter API. . . . .	35
3.10	Parsing dati relativi al profilo Twitter. . . . .	36
3.11	Parsing dati relativi ai tweet di un profilo Twitter. . . . .	36
3.12	Inserimento dati profilo Twitter nel database. . . . .	36
3.13	Inserimento dati tweet nel database. . . . .	36
3.14	Definizione URL in Flask. . . . .	37
3.15	Indirizzamento alla pagina di visualizzazione dei dati. . . . .	38
3.16	Estrazione dei dati relativi a una singola pagina Facebook per la visualizzazione. . . . .	39
3.17	Interrogazione al database per l'estrazione dei dati relativi a una pagina Facebook. . . . .	39
3.18	Esempio metodo per generare grafico relativo all'engagement mensile di una pagina Facebook. . . . .	40
3.19	Metodo per salvare il grafico come immagine. . . . .	41
3.20	Metodi preparare i dati e per salvarli in un file csv. . . . .	41
4.1	Engagement delle pagine Facebook per la promozione turistica regionale. . . . .	51
4.2	Engagement normalizzato delle pagine Facebook per la pro- mozione turistica regionale. . . . .	52
4.3	Distribuzione post su base mensile e relativo engagement me- dio per la regione Sicilia. . . . .	52
4.4	Distribuzione post durante l'anno in base alla zona geografica di provenienza della regione. . . . .	53
4.5	Numero di post per stagione e relativo engagement medio. . . . .	53
4.6	Distribuzione post in base al giorno della settimana e relativo engagement. . . . .	54
4.7	Distribuzione post in base al momento della giornata. . . . .	54
4.8	Numeri di post per tipologia della regione Sardegna e relativo engagement. . . . .	55
4.9	Numero di post in base alla tipologia a livello regionale. . . . .	56

---

4.10 Numeri di post per tipologia e relativo engagement. . . . .	56
4.11 Numero di caratteri per post a livello regionale. . . . .	57
4.12 Numero di caratteri per post della regione Sicilia e relativo engagement. . . . .	57
4.13 Numero di post in base alla lunghezza e relativo engagement. .	58
4.14 Numero totale di reazioni, commenti e condivisioni su base regionale. . . . .	58
4.15 Migliori post del 2017 per reazioni, commenti e condivisioni. .	60
4.16 Engagement dei profili Twitter per la promozione turistica regionale. . . . .	63
4.17 Engagement normalizzato dei profili per la promozione turi- stica regionale. . . . .	63
4.18 Distribuzione tweet in base al giorno della settimana. . . . .	64
4.19 Distribuzione tweet in base al momento della giornata. . . . .	64
4.20 Numero di caratteri per tweet per ogni regione. . . . .	65
4.21 Numero di caratteri per tweet della regione Sicilia e relativo engagement. . . . .	65
4.22 Lunghezza tweet e relativo engagement. . . . .	66
4.23 Numero di risposte sui 500 tweet estratti per ogni regione. . .	67
4.24 Distribuzione giornaliera dei soli tweet della regione Sardegna. .	67
4.25 Distribuzione giornaliera di tweet e risposte della regione Sar- degna. . . . .	68
4.26 Numero di likes e retweet totali ai tweet su base regionale. . .	68
4.27 Migliori tweet del 2017 per numero di <i>likes</i> e retweet . . . . .	70
4.28 Codice per la regressione sui dati di Facebook. . . . .	74
4.29 Analisi della regressione sui dati relativi a Facebook. . . . .	75
4.30 Stima dei parametri relativi a Facebook. . . . .	76
4.31 Codice per la regressione sui dati di Twitter. . . . .	78
4.32 Analisi della regressione sui dati relativi a Twitter. . . . .	79
4.33 Stima dei parametri relativi a Twitter. . . . .	80



# Elenco delle tabelle

1.1	Tool disponibili online per analisi social media . . . . .	12
4.1	Account ufficiali per la promozione turistica delle regioni italiane su Facebook e Twitter. . . . .	44
4.2	Numero di reazioni per post. . . . .	49
4.3	Numero di commenti per post. . . . .	49
4.4	Numero di condivisioni per post. . . . .	50
4.5	Panoramica sulle pagine Facebook in analisi: numero fan, post pubblicati, engagement e engagement normalizzato percentuale	50
4.6	Numero di reazioni, commenti e condivisioni per ogni regione.	59
4.7	Numero di 'mi piace' per tweet. . . . .	61
4.8	Numero di retweet per tweet. . . . .	61
4.9	Panoramica sui profili Twitter in analisi: numero <i>followers</i> , <i>following</i> , tweet totali pubblicati, engagement ed engagement normalizzato percentuale. . . . .	62
4.10	Numero di 'mi piace' e retweet totali per ogni regione. . . . .	69
4.11	Hashtag più utilizzato da ogni regione. . . . .	71
4.12	Hashtag con engagement più alto per ogni regione. . . . .	72



# Capitolo 1

## Stato dell'arte

In questo primo capitolo si illustreranno le diverse ricerche nel campo delle imprese e i social media. Il paragrafo 1.1 introdurrà i *big data*, il paragrafo 1.2 analizzerà i social media, con particolare attenzione a Facebook e Twitter, il paragrafo 1.3 presenterà l'utilizzo dei social media da parte delle imprese con attenzione all'ambito turistico. Nel paragrafo 1.4, infine, verranno analizzati alcuni *tool* esistenti per la *brand comparison* sui social media.

### 1.1 Big data

I Big data sono grandi volumi di dati che non possono essere gestiti in modo ottimale utilizzando metodi tradizionali. De Mauro et al. [11] (2016) ne danno la seguente definizione:

*“I Big data sono informazioni caratterizzate da un così grande volume, velocità e varietà da richiedere specifiche tecnologie e metodi analitici per essere trasformato in valore.”*

Doug Laney, nel 2001, definisce le tre caratteristiche principali dei Big data, chiamate ‘tre V’ [32]:

- Volume: la grande quantità di dati generati ogni secondo. I dati possono essere generati da utenti o automaticamente dalle macchine, tramite sensori o strumenti specifici ed è necessario salvarli velocemente.

- **Varietà:** la diversità dei dati, poiché vengono considerati dati in formati diversi che provengono da diverse fonti e che avevano diversi scopi. I dati possono essere sia strutturati, se ben definiti e rappresentabili con numeri o categorie, oppure non strutturati, se è complicato definirli.
- **Velocità:** la velocità con cui si generano nuovi dati, basta pensare al numero di post pubblicati ogni secondo su Facebook, al numero di transizioni di e-commerce o a un sensore che rileva la velocità a cui si sta spostando una macchina.

Successivamente, alle tre caratteristiche di Laney, ne è stata aggiunta un'altra:

- **Veridicità:** quanto i dati sono affidabili e la loro qualità in base alle fonti. Alcuni dati sono generati internamente all'organizzazione che ne fa uso, però in altri casi sono generati esternamente ed è quindi necessario valutarne la provenienza. I dati, infatti, possono provenire da sistemi operazionali interni alle aziende, da sensori e strumenti scientifici, venendo quindi generati automaticamente da macchine, oppure possono provenire dal Web, e quindi da siti, social media, ecc. Questi dati, possono essere incorretti nel caso si ricorra in errori di *sensing* o, nel caso del web, anche degli utenti. Eliminare le anomalie, chiamate *outlier*, però, non è sempre consigliabile poiché si potrebbero cancellare elementi potenzialmente utili [36, 51].



Figura 1.1: Le 4V dei big data. Fonte: IBM ([www.ibmbigdatahub.com](http://www.ibmbigdatahub.com))

In campo aziendale l'utilizzo dei big data è fondamentale per analizzare la propria clientela, infatti raccogliendo le grandi moli di dati provenienti dai

social media, dal web e dall'azienda stessa è possibile condurre analisi per profilare il proprio pubblico nei migliori dei modi. Lo studio dei propri clienti non è utile solo a fine pubblicitari, ma al contrario è utile anche nell'invenzione e progettazione di nuovi prodotti in modo da poter produrre ciò che i clienti desiderano.

Per estrarre queste informazioni si fa utilizzo del *data mining*, “*un processo aziendale per esplorare grandi somme di dati con lo scopo di scoprire pattern e regole significative*[35].”

## 1.2 Social media

Per poter parlare di social media è necessario introdurre il concetto di Web 2.0. Il Web 2.0 è un nuovo metodo di utilizzo del web che dal 2004 non ha più visto un individuo pubblicare contenuti isolati, ma al contrario si è venuta a creare una collaborazione continua tra gli utenti. Al posto dei semplici siti web solo informativi, ad esempio, ora si utilizzano i blog in cui è possibile per gli utenti commentare e interagire con l'autore. Questa partecipazione degli utenti nella pubblicazione e generazione di contenuti viene spesso indicata come *User Generated Content* (UGC)<sup>1</sup>.

Kaplan e Haenlein [29] hanno definito i social media come:

*“gruppo di applicazioni basate su Internet, costruite sulle ideologie e i fondamenti tecnologici del Web 2.0, e che permettono la creazione e lo scambio di User Generated Content.”*

I social media sono quindi piattaforme web in cui le persone possono entrare in contatto tra loro e con enti esterni, come le imprese, creando delle comunità online di individui con gli stessi interessi e valori.

Il primo social media, *Geocities*, fu creato nel 1994 e consentiva agli utenti di creare una propria pagina personale.

Le applicazioni dei social media ricadono ora in cinque diverse categorie[29]:

---

<sup>1</sup>User Generated Content(UGC): tradotto come contenuto generato dagli utenti.

- Collaborative, quando gli utenti collaborano per la creazione di contenuti, ad esempio Wikipedia.
- Blog e Microblog, i blog sono generalmente gestiti da una sola persona, ma gli utenti possono interagire con essa tramite i commenti. Un esempio di microblog è Twitter.
- Multimedia, comunità che condividono contenuti multimediali, alcuni esempi sono Youtube, nel caso dei video, Flickr, nel caso delle foto, e SlideShare, quando si tratta di presentazioni PowerPoint.
- Social network, che sono uno spazio utile a creare legami con altre persone, l'esempio più famoso è sicuramente Facebook. Il concetto di social network è basato sulla teoria dei sei gradi di separazione, secondo cui una persona può essere collegata a qualunque altra attraverso una catena di conoscenze e relazioni di massimo 6 persone[12].
- Mondi virtuali, sia nel caso dei videogame che nel caso di mondi sociali virtuali, in cui gli utenti possono assumere sembianze di *avatar*. Due esempi sono World of Warcraft e Second Life.

### 1.2.1 Facebook

Facebook<sup>2</sup> è stato fondato all'inizio del 2004 da Mark Zuckerberg con i suoi collaboratori Dustin Moskovitz, Chris Hughes e Eduardo Saverin. La piattaforma è nata con l'intento di mettere in contatto gli studenti di Harvard e si è poi diffusa in altre università fino a permettere, il 26 settembre 2006, il libero accesso a chiunque avesse più di 13 anni. Facebook è una piattaforma per il social networking che permette alle persone di essere in contatto, scoprire cosa accade nel mondo e condividere ciò che ritengono importante. L'utente può modificare la propria privacy per permettere al social media di accedere solo ai dati a cui si dà il consenso. Il social media inoltre permette di pubblicare foto, video, sondaggi, liste, creare eventi ed entrare a far parte

---

<sup>2</sup><https://www.facebook.com/>

di gruppi con altri utenti che appartengono alla stessa community. Da ottobre 2016 permette di mettere in vendita o comprare articoli per mezzo del Marketplace e da marzo 2017 permette di creare raccolte fondi [16].

Facebook, quindi, non permette solo di rimanere in contatto con i propri amici, ma consente di entrare a far parte di comunità virtuali e amplificare in diversi modi le proprie conoscenze tramite un'interfaccia estremamente intuitiva. Secondo i dati Eurostat relativi all'anno 2017 in Italia si hanno più di 30 milioni di utenti attivi su Facebook mensilmente, affermando il social media come piattaforma più usata del territorio [8].

Il modello di business di Facebook è basato sulla pubblicità, la piattaforma mette infatti a disposizione delle aziende la possibilità di sponsorizzare i propri post e mostrarli al target di riferimento. La sola sponsorizzazione però non sempre è sufficiente a creare una buona strategia di marketing, tramite il social le aziende possono infatti interagire direttamente con i propri clienti instaurando un rapporto più diretto con essi. Se il pubblico apprezza i post dell'azienda lo mostrerà condividendo, commentando o lasciando una reazione come 'Mi piace', 'Love', 'Wow', 'Ahah', 'Sigh' e 'Grrr' e aumentando la propria fidelizzazione verso l'impresa.

### 1.2.2 Twitter

Twitter<sup>3</sup> è stato fondato a marzo 2006 da Jack Dorsey, Noah Glass, Biz Stone e Evan Williams e lanciato a luglio dello stesso anno.

Il social media permette agli utenti di postare brevi messaggi chiamati tweet, da novembre 2017 si hanno 280 caratteri massimi; seguire altri utenti, indicati come *following*, per vedere i loro tweet ed essere a propria volta seguiti da altri, i *followers*. Gli utenti possono inoltre pubblicare foto, video e sondaggi, rispondere ai tweet e *retweettare*, cioè condividere il tweet, o mettere un 'mi piace' ai tweet. La caratteristica principale di Twitter, oltre alla brevità dei messaggi, è l'utilizzo degli *hashtag*, delle parole chiave anticipate dal carattere “#” che facilitano le ricerche telematiche nel social media.

---

<sup>3</sup><https://twitter.com/>

Le imprese tramite il social media, possono non solo sponsorizzare i tweet, ma anche gli *hashtag* facendo in modo che questo diventino primi nelle ‘Tendenze’, che rappresentano la classifica degli *hashtag* più utilizzati in quel momento.

### 1.3 Imprese e Social Media

Secondo i dati Eurostat [15] in Europa il 47% delle imprese ha utilizzato almeno una tipologia di social media nel 2017, in Italia questa percentuale scende a 44%, in particolare il 27% delle imprese ha utilizzato una sola tipologia social media, mentre il 17% ne ha utilizzate due o più. I dati mostrano inoltre quali sono le categorie di social media più utilizzate - figura 1.2 - si può notare come i social network, come Facebook e LinkedIn, siano al primo posto, seguiti dalle piattaforme multimediali, come Youtube, e dai blog o microblog, come Twitter, che in Italia sono utilizzati molto meno rispetto la media europea.

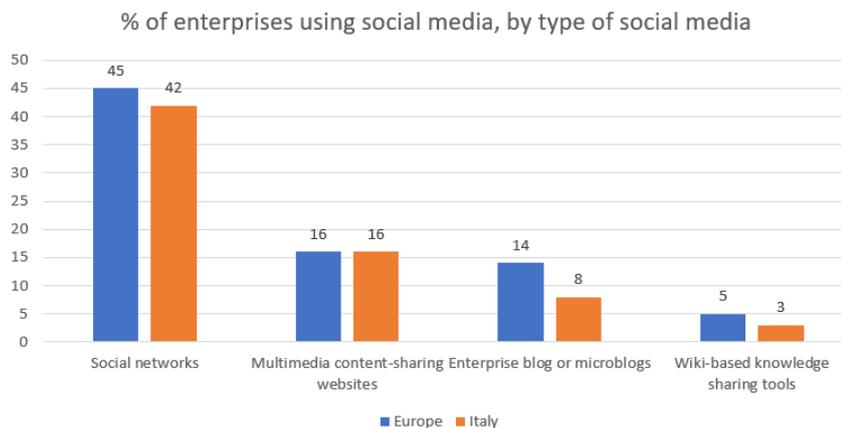


Figura 1.2: Utilizzo dei social media da parte delle imprese in base alla tipologia di social. Dati Eurostat.

Le imprese quindi utilizzano diversi social media per entrare in contatto con i propri clienti e lo studio presenta anche i motivi principali per cui vengo-

no utilizzati, la figura 1.3 infatti evidenzia che in Italia il 37% delle imprese utilizza i social media per migliorare la propria immagine e sponsorizzare prodotti, il 23% per interagire con i clienti, il 13%, un punto percentuale maggiore rispetto la media europea, per coinvolgere i clienti nello sviluppo e l'innovazione dei propri prodotti e servizi e il 7% per collaborare con altre imprese.

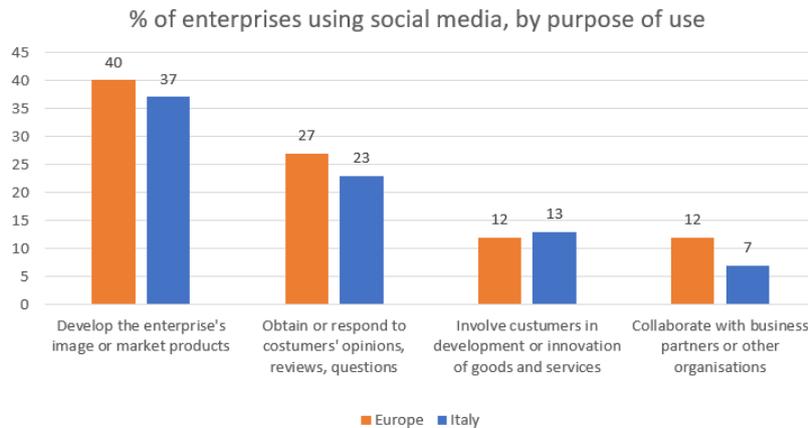


Figura 1.3: Scopo utilizzo social media da parte delle imprese. Dati Eurostat.

Gli obiettivi delle imprese sui social media, infatti, riguardano principalmente [2, 58] :

- Costruire e mantenere una relazione con i clienti, poiché è possibile comunicare e dialogare direttamente con essi facendoli sentire più coinvolti.
- Interagire con i clienti per ascoltarne domande, richieste, problemi e idee.
- Aumentare la *brand awareness* e il passaparola, cioè aumentare il bacino di persone che conosce l'impresa, un buon esempio è l'utilizzo di concorsi o messaggi virali che si diffondono rapidamente. Un altro esempio si ha con un utente Facebook che parla di un prodotto che ha

apprezzato e lo raccomanda ai propri contatti; questi potrebbero, se interessati, cercare l'azienda aumentando la conoscenza del brand.

- Amplificare la propria audience e migliorare il target, poiché raggiungendo più persone si viene a contatto con potenziali clienti che abitano in aree geografiche differenti. Si possono inoltre analizzare, se messi a disposizione dai social media, le demografiche dei clienti per migliorare le proprie strategie.
- Promuovere prodotti, servizi e l'impresa stessa, utilizzando messaggi creativi e persuasivi pensati appositamente per il proprio target, per fare in modo che i clienti non li trovino fastidiosi, ma al contrario li condividano con amici e parenti.
- Aumentare l'engagement, cioè il grado di coinvolgimento degli utenti con ciò che viene pubblicato sui social media; più un utente è coinvolto più sarà fedele all'impresa.

Si può quindi notare come ogni social media offra all'impresa diversi modi per interagire con i propri clienti. Facebook potrebbe essere utilizzato per pubblicare foto, sponsorizzare eventi, creare sondaggi e raccogliere opinioni. Twitter, invece, spesso viene utilizzato non solo per la promozione, ma anche per il supporto tecnico, ad esempio alcune compagnie telefoniche, come Vodafone, Tim, Wind e 3<sup>4</sup>, danno la possibilità ai propri clienti di contattare direttamente l'assistenza con un semplice tweet.

### 1.3.1 Social Media e Turismo

Questo elaborato propone un'analisi relativa all'utilizzo dei social media Facebook e Twitter da parte delle DMO delle regioni italiane. Per questo motivo nei paragrafi successivi verranno illustrati gli studi relativi all'utilizzo dei social media in ambiente turistico.

---

<sup>4</sup>I profili Twitter italiani delle compagnie: <https://twitter.com/VodafoneIT>, [twitter.com/TIM\\_Official](https://twitter.com/TIM_Official), [twitter.com/WindItalia](https://twitter.com/WindItalia) e [twitter.com/Tre\\_It](https://twitter.com/Tre_It)

### Utilizzo dei social media da parte dei turisti

Secondo i dati Istat il 43% dei residenti in Italia nel 2017 ha prenotato le proprie vacanze via Internet, il dato è la media tra il 41,9% dei residenti che ha prenotato vacanze con al massimo tre notti e il 44% che le ha prenotate da quattro o più notti. È utile evidenziare come anche il 55,2% dei viaggi di lavoro sia stato prenotato tramite Internet [25].

Questi dati, evidenziano come l'utilizzo del web da parte dei turisti sia in crescita, poiché la quota dei viaggi prenotati tramite Internet è aumentata di circa 5 punti percentuali rispetto al 2016. Quindi, prima di poter parlare dell'utilizzo dei social media da parte degli enti turistici è necessario capire come li utilizzino i turisti.

Lo studio "How 'social' are destinations? Examining European DMO social media usage" [62] evidenzia come i turisti facciano un uso dei social media, prima, durante e dopo il viaggio. Prima del viaggio il turista ricerca le possibili mete, i trasporti, le strutture dove alloggiare e le attività che offre il luogo, valutando le alternative possibili. Durante il viaggio, non vengono generalmente effettuate ricerche ma al contrario si creano contenuti, ad esempio scattando foto e pubblicandole sui social. Infine, avviene l'utilizzo principale: i turisti, una volta a casa, usano i social media per condividere le loro esperienze, raccomandando o meno i luoghi visitati ai propri contatti e pubblicando recensioni, foto o video [9, 45].

### DMO

L'Organizzazione Mondiale del Turismo (OTM), definisce il Destination Management come [47]:

*“gestione coordinata di tutti gli elementi che compongono una destinazione (attrazioni, accesso, marketing, risorse umane, immagine e prezzi). Il Destination Management adotta un approccio strategico per collegare le loro entità molto diverse per una migliore gestione della destinazione.”*

Le Destination Management Organization (DMO) sono quindi delle organizzazioni che definiscono una strategia per la promozione turistica, in Italia coincidono con gli enti del turismo, e ne esistono di diverse tipologie:

- National Tourism Authorities (NTAs) o Organisations (NTOs), responsabili per la gestione e il marketing del turismo a livello nazionale.
- Regional, provincial o state DMOs (RTOs), responsabili per la gestione e il marketing del turismo in una regione geografica limitata, come le regioni o le provincie.
- Local DMOs, responsabili della gestione e il marketing del turismo a livello cittadino.

I vantaggi di una buona gestione delle DMO sono [47]:

- Vantaggio competitivo rispetto ai rivali, che è definito dall'offerta di esperienze uniche e una buona permanenza da parte del visitatore.
- Assicurare la sostenibilità del turismo, in modo che la destinazione mantenga la sua integrità ambientale e protegga le attrazioni principali.
- Estendere i benefici del turismo, per supportare lo sviluppo della comunità territoriale.
- Migliorare il rendimento del turismo, attraverso pubblicità mirate al fine di aumentare il numero di visitatori.
- Costruire una forte identità aziendale, in modo da fidelizzare i turisti e farli tornare a visitare la destinazione.

Al giorno d'oggi, poiché i visitatori fanno un ampio uso delle piattaforme online è necessario che le DMO si adattino al marketing online, definendo una strategia integrata sia online che offline. Le strategie di marketing offline sono necessarie ad aumentare la notorietà del brand, mentre quelle online ad aumentare il numero di persone raggiunte e diffondere la destinazione. L'utilizzo di più canali di comunicazione facilita l'esperienza del turista che può

rapidamente passare, ad esempio, da un post su Facebook al sito per la prenotazione o l'acquisto. Nella promozione e comunicazione, inoltre, si devono mettere in evidenza le attrazioni e i punti di forza della propria destinazione in modi creativi, l'obiettivo è infatti quello di attirare l'attenzione del turista poiché è il cliente stesso che in prima persona ricerca le informazioni [13] .

In un'ottica italiana, Mariani et al.(2016)[37] hanno studiato l'uso di Facebook nelle DMO regionali italiane. Lo studio è basato su un'analisi dei dati provenienti dalle pagine Facebook delle regioni e arricchito da interviste ai manager delle DMO in questione. Dalla ricerca è emerso come l'utilizzo di concorsi, quiz e sondaggi aiuti la pagina a crescere, così come i contenuti multimediali. In secondo luogo, è emerso che la differenza tra nord e sud nell'utilizzo degli strumenti digitali è decrescente e che le regioni italiane utilizzano una strategia top-down al posto di aumentare l'*user-generated content* (UGC) che invece aiuterebbe la DMO a fidelizzare gli utenti. Infine, nello studio è stata condotta una regressione lineare per identificare quali fossero i fattori che influivano positivamente o negativamente sull'engagement, l'alta frequenza di post e la pubblicazione di questi alla mattina ha un impatto negativo su di esso, mentre i contenuti multimediali e la lunghezza maggiore hanno un impatto positivo.

## 1.4 Tool per *brand comparison* sui social media

Per avere un quadro completo sullo stato attuale dell'analisi dei social media si sono studiati alcuni dei siti che offrono l'analisi degli account di Facebook e/o Twitter in modo totalmente o almeno in parte gratuito.

La tabella 1.1 mostra una raccolta dei tool disponibili online, i collegamenti ad essi e i social media che ogni piattaforma mette a disposizione per l'analisi.

Purtroppo in questa sede, per motivi di copyright, non è possibile analizzarli tutti in dettaglio. Verranno quindi considerati i soli tre tool: Follo-

Piattaforma	Collegamento	Social Media
Agorapulse	<a href="https://www.agorapulse.com/">https://www.agorapulse.com/</a>	Facebook, Twitter, altri
Audiense	<a href="https://audiense.com">https://audiense.com</a>	Twitter, altri
Buffer	<a href="https://buffer.com/">https://buffer.com/</a>	Facebook, Twitter, altri
Cyfe	<a href="https://www.cyfe.com/">https://www.cyfe.com/</a>	Facebook, Twitter, altri
Followerwonk	<a href="https://moz.com/followerwonk">https://moz.com/followerwonk</a>	Twitter
Keyhole	<a href="http://keyhole.co/">http://keyhole.co/</a>	Facebook, Twitter, altri
Klear	<a href="https://klear.com">https://klear.com</a>	Facebook, Twitter, altri
Likealyzer	<a href="https://likealyzer.com/">https://likealyzer.com/</a>	Facebook
Quintly	<a href="https://www.quintly.com/">https://www.quintly.com/</a>	Facebook, Twitter, altri
Rival IQ	<a href="https://www.rivaliq.com/">https://www.rivaliq.com/</a>	Facebook, Twitter, altri
SocialBakers	<a href="https://www.socialbakers.com/">https://www.socialbakers.com/</a>	Facebook, Twitter, altri
SocialRank	<a href="https://socialrank.com/">https://socialrank.com/</a>	Twitter, altri
SumAll	<a href="https://sumall.com/">https://sumall.com/</a>	Facebook, Twitter, altri
Sprout Social	<a href="https://sproutsocial.com/">https://sproutsocial.com/</a>	Facebook, Twitter, altri
TweetReach	<a href="https://tweetreach.com/">https://tweetreach.com/</a>	Twitter
Twitonomy	<a href="https://www.twitonomy.com/">https://www.twitonomy.com/</a>	Twitter

Tabella 1.1: Tool disponibili online per analisi social media

werwonk, Keyhole e Likealyzer, che mostrano almeno una parte dell'analisi gratuitamente e senza richiedere l'accesso al sistema.

### **Followerwonk**

Followerwonk [43] permette l'analisi dei soli profili Twitter ed è focalizzata sull'analisi dei followers e following. La piattaforma permette, infatti di:

- ricercare tramite una parola chiave dei profili Twitter;
- confrontare i followers e following tra due o tre profili in base al numero di giorni dall'iscrizione, al numero medio di nuovi follower giornalieri e di tweet totale e settimanale;

- analizzare i followers o following di un profilo in base a genere, locazione e attività su Twitter.

La piattaforma mette a disposizione anche altre funzionalità, ma richiedono l'accesso al sistema e quindi non sono state considerate nell'analisi.

### **Keyhole**

Keyhole [30] permette l'analisi dei profili Twitter, Facebook, Instagram e Youtube. Per accedere ad alcune delle metriche è richiesto l'accesso al sito, per questo motivo si sono analizzate le sole funzionalità pubbliche e relative ai social media Facebook e Twitter.

La piattaforma permette di effettuare una ricerca su Twitter relativa agli hashtag, a una parola chiave, alle menzioni di un account e all'utilizzo di un URL.

Keyhole offre l'analisi di un profilo Twitter ed infatti mostra i dati relativi al profilo considerato, i dati relativi all'engagement giornaliero degli ultimi 99 tweet, i migliori tweet e i migliori hashtag in base all'engagement.

Infine, permette l'analisi di una pagina Facebook mostrando inizialmente i dati della pagina e mostrando l'engagement mensile dell'ultimo anno, con un massimo di 99 post, i migliori post in base all'engagement e la lunghezza ottimale dei post.

### **LikeAlyzer**

LikeAlyzer [34] fornisce l'analisi delle sole pagine Facebook e permette di ampliare le funzionalità tramite l'accesso al sito.

La piattaforma propone un'analisi generica della pagina Facebook analizzando le informazioni fornite all'utente e come sia possibile migliorarle, analizza poi l'attività della pagina, le interazioni con gli utenti e l'engagement della pagina.



# Capitolo 2

## Progettazione

In questo capitolo si tratterà la progettazione della piattaforma realizzata: nel paragrafo 2.1 verranno esplicitate le specifiche del progetto, mentre nel paragrafo 2.2 la sua architettura.

### 2.1 Specifiche del progetto

Il progetto consiste nella progettazione e l'implementazione di una piattaforma software per l'estrazione, l'analisi e la visualizzazione di dati pubblici provenienti da pagine Facebook e profili Twitter.

L'obiettivo della piattaforma è quello di fornire all'utilizzatore uno strumento che permetta di analizzare i propri profili social e quelli dei concorrenti, consentendo anche confronti diretti tra i dati.

L'estrazione dei dati avviene per mezzo di librerie messe a disposizione dagli stessi social media e riguardano sia gli account, che i post o i tweet.

La piattaforma software permette di scegliere il social media, tra Facebook o Twitter, e l'account di cui analizzare i dati e mostra le informazioni relative al profilo scelto, consentendo all'utente di:

- scegliere quale delle analisi proposte visualizzare graficamente;
- scegliere, in alcuni casi, l'arco temporale dei dati da visualizzare;

- salvare i grafici come immagini;
- salvare i dati in un file CSV <sup>1</sup>;
- confrontare due profili social differenti.

Le analisi proposte riguardano l'engagement, il numero di interazioni, la tipologia dei post, l'utilizzo di hashtag e la lunghezza dei messaggi pubblicati.

## 2.2 Architettura del progetto

Il progetto può essere diviso in due parti principali, un *back end*, che gestisce l'estrazione dei dati, la gestione del database e il reindirizzamento delle schermate, e un *front end*, che si occupa della visualizzazione dei dati.

La piattaforma permette - nella pagina principale mostrata in figura 2.1 - di scegliere il social media, Facebook o Twitter, e l'account di cui analizzare i dati. Una volta inseriti i dati *back end* viene gestita l'estrazione degli stessi, che sia tramite nuova richiesta all'API del social media o tramite il database.

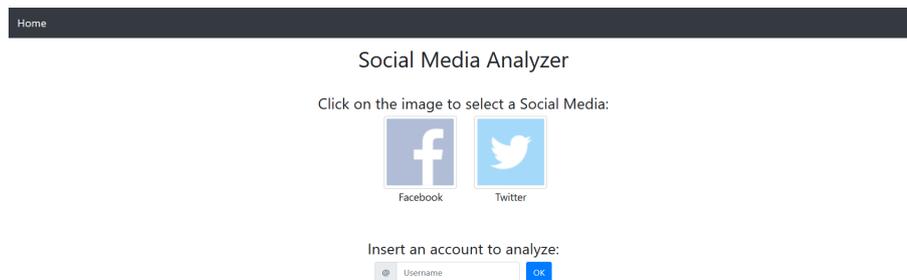


Figura 2.1: Pagina principale.

<sup>1</sup>Comma-separated values (CSV): è un formato di file di testo utile a salvare dati in record memorizzati su righe separate, i cui attributi sono separati da virgole [53].

Per quanto riguarda la visualizzazione dei dati le schermate presentano una struttura uniforme. In alto a sinistra è presente il pulsante ‘home’ utile per tornare alla homepage e effettuare una nuova ricerca. Sono poi presentati i dati relativi all’account in analisi: nome, nome utente, descrizione, dati relativi ai fan, l’engagement medio e l’engagement medio normalizzato.

Un menù a tendina permette di selezionare la tipologia del grafico che si desidera visualizzare. Vi è poi indicato il tipo di grafico visualizzato, una breve descrizione dello stesso e della sua utilità e viene mostrato l’arco temporale dei dati in analisi. In alcuni casi è possibile selezionare un intervallo di date di cui visualizzare il dettaglio.

Infine, è presente il grafico con i dati e, a destra, un pannello con gli strumenti utili a operare sul grafico e a salvare i dati come immagine o come file CSV. Le informazioni relative a una pagina o profilo vengono estratti una sola volta e non è possibile aggiornarli.

### 2.2.1 Visualizzazione dati di una pagina Facebook

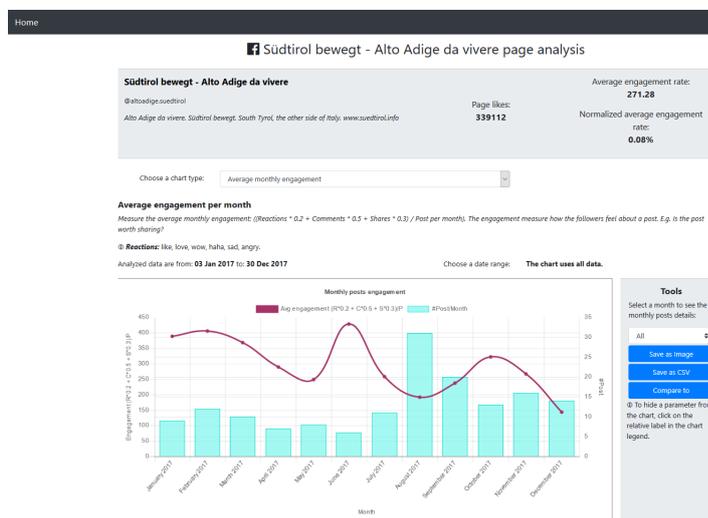


Figura 2.2: Visualizzazione dati di una singola pagina Facebook.

La figura 2.2 rappresenta l’analisi di una sola pagina Facebook. Tra i dati della pagina, oltre ai campi precedentemente definiti, è presente il numero

totale di fan della pagina.

Per quello che concerne la visualizzazione dei dati è possibile analizzare i grafici relativi: all'engagement medio mensile; al totale di interazioni divise in reazioni, commenti e condivisioni; l'engagement medio in base alla tipologia del post e l'engagement medio in base alla lunghezza del testo. Per le prime due tipologie di grafici è possibile selezionare un mese per vederne il dettaglio giornaliero e poter scendere ancora più nel dettaglio scegliendo un intervallo di date interne al mese in esame. Il pulsante 'compare to' mostra l'opzione per inserire un secondo account con cui confrontare quello in analisi. Un esempio è visibile in figura 2.3.

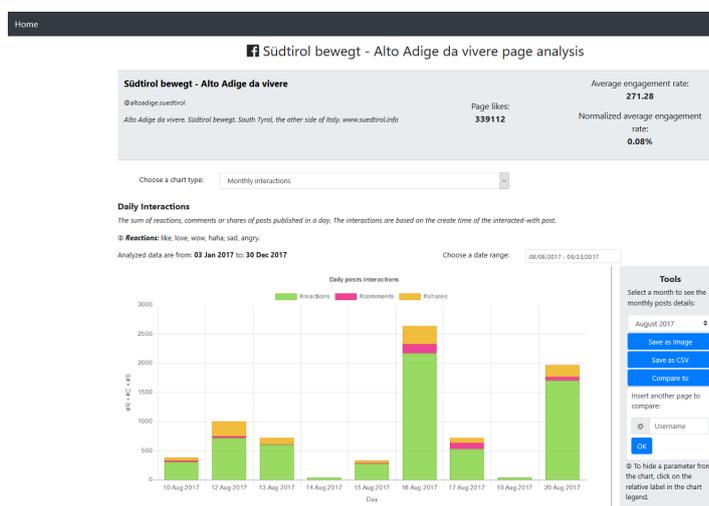


Figura 2.3: Dettaglio visualizzazione interazioni dal 10 agosto 2017 al 20 agosto 2017 per la pagina dell'Alto Adige.

## 2.2.2 Confronto tra due pagine Facebook

La figura 2.4 rappresenta la schermata di confronto tra due pagine Facebook. Come per la visualizzazione di una sola pagina, per ogni account è presente il numero di fan.

Il confronto avviene su base mensile e permette di visualizzare numero di post al mese e il relativo engagement medio e il numero di interazioni

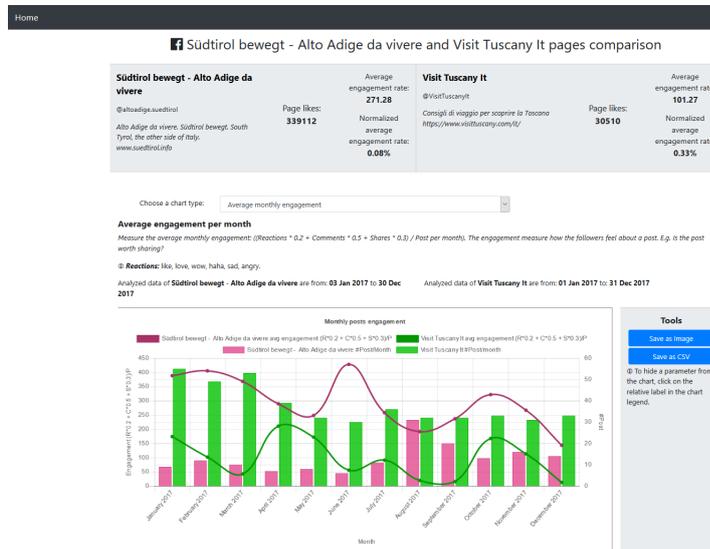


Figura 2.4: Visualizzazione dati per il confronto tra pagine Facebook.

mensili. Confrontando i dati non è possibile modificare l'arco temporale dei dati visualizzati.

### 2.2.3 Visualizzazione dati di un profilo Twitter

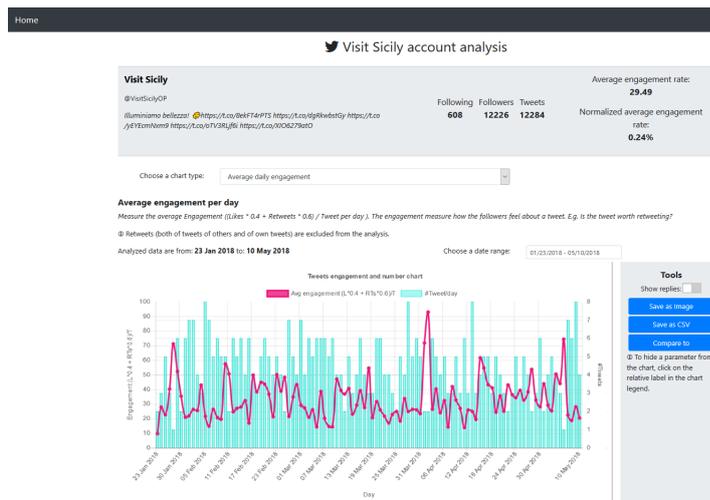


Figura 2.5: Visualizzazione dati di un singolo profilo Twitter.

La figura 2.5 rappresenta la schermata di analisi di un profilo Twitter. Nella sezione iniziale dedicata ai dati relativi al profilo sono presenti il numero di *following*, *followers* e tweet totali pubblicati dalla pagina.

Scegliendo dal menù a tendina è possibile visualizzare: l'engagement medio giornaliero; il numero di reazioni giornaliere, divise in 'mi piace' e retweet; gli hashtag che hanno prodotto un engagement più alto e l'engagement in base alla lunghezza del tweet.

Come per l'analisi di Facebook, è possibile scegliere un intervallo di tempo per le prime due tipologie di grafico. Dal pannello con gli strumenti è inoltre possibile scegliere se includere nel grafico anche le risposte ad altri tweet. Tramite il pulsante 'compare to', come in precedenza, è possibile inserire il nome utente del profilo Twitter con cui si vuole effettuare il confronto. Un esempio si ha in figura 2.6.

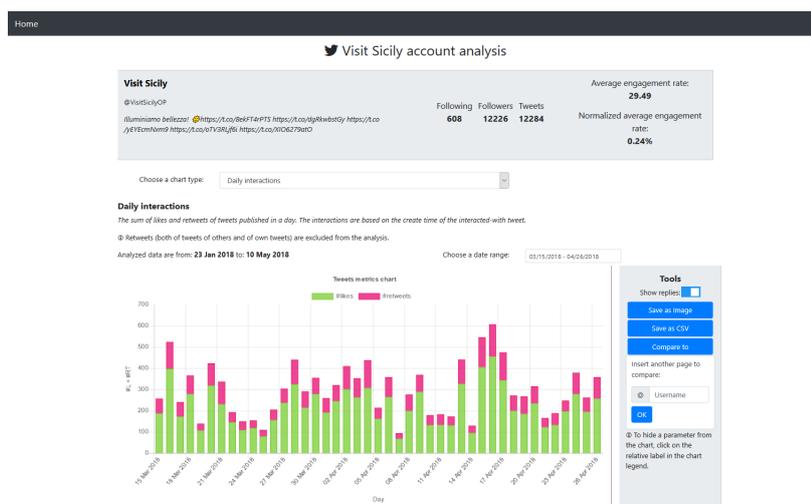


Figura 2.6: Dettaglio visualizzazione interazioni dal 15 marzo 2018 al 4 aprile 2018 per il profilo della Sicilia.

## 2.2.4 Confronto tra due profili Twitter

La figura 2.7 rappresenta la schermata di confronto tra due profili Twitter. Per ogni pagina viene mostrato il numero di *following*, *followers* e tweet totali

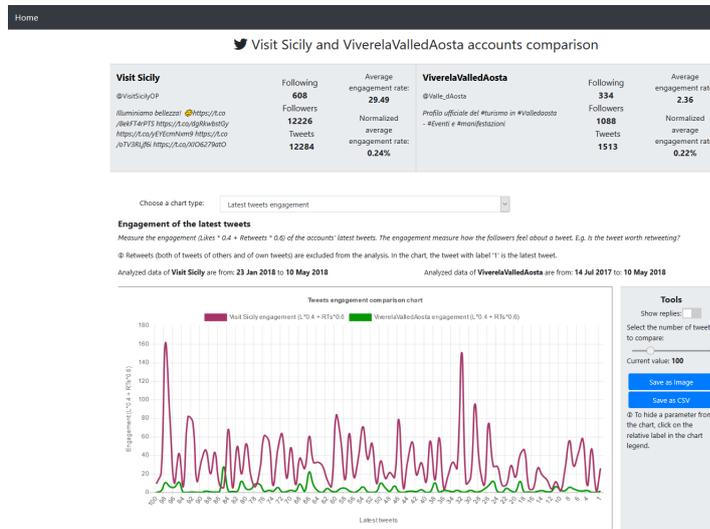


Figura 2.7: Visualizzazione confronto tra due profili Twitter.

pubblicati.

Il confronto tra le pagine avviene in base agli ultimi cinquecento tweet e risposte pubblicati, poiché non è possibile definire un intervallo di tempo definito di cui estrarre i dati.

È possibile confrontare i tweet in base all'engagement medio o in base al numero di interazioni e i tweet vengono mostrati dal più datato al più recente, cioè il tweet identificato con l'etichetta '1' è l'ultimo pubblicato prima dell'estrazione dei dati. Inizialmente viene mostrato il confronto tra cento tweet, ma viene lasciata la possibilità di variarne il numero dal pannello con gli strumenti.

Come nel caso dell'analisi singola, inoltre, è possibile includere le risposte ai tweet nell'analisi.



# Capitolo 3

## Implementazione

In questo capitolo viene discussa l'implementazione della piattaforma realizzata. Nel paragrafo 3.1 vengono introdotte le tecnologie utilizzate nello sviluppo del sistema, mentre nel paragrafo 3.2 vengono discussi i dettagli implementativi.

### 3.1 Tecnologie utilizzate

#### 3.1.1 Python

Python [49] è stato creato da Guido van Rossum, che ne ha iniziato l'implementazione nel 1989 e l'ha rilasciato per la prima volta nel 1991, al *Stichting Mathematisch Centrum* in Olanda. Il nome del linguaggio è un omaggio al gruppo comico Monty Python e in particolare alla commedia “Monty Python's Flying Circus”, di cui Van Rossum stava leggendo il copione in quel periodo. Attualmente lo sviluppo di Python è gestito dall'organizzazione no-profit Python Software Foundation e la versione più recente è la 3.4. Nello sviluppo della piattaforma si è fatto uso della versione 2.7.

Python è un linguaggio:

- interpretato, poiché un interprete analizza il codice sorgente e se è sintatticamente corretto lo esegue senza creare eseguibili;

- interattivo, poiché permette di eseguire frammenti di codice direttamente da terminale;
- orientato agli oggetti.

Python è caratterizzato dall'utilizzo di moduli, che sono l'equivalente delle librerie in altri linguaggi di programmazione, eccezioni, tipizzazione forte e dinamica, strutture dati di alto livello e classi. Il linguaggio è estendibile in *C* o *C++* ed è utilizzabile come linguaggio di estensione per programmi che necessitano di un'interfaccia di programmazione. Python è molto elegante e altamente leggibile poiché per delimitare i blocchi di programma utilizza l'indentazione<sup>1</sup>. Infine, il linguaggio è portabile e gira su molte varianti di Unix, su Mac e su Windows.

Per questo progetto sono stati utilizzati i moduli:

- Requests, che permette l'invio di richieste HTTP, senza aggiungere manualmente stringhe all'URL. Si è utilizzato il modulo per effettuare richieste alle API di Facebook e Twitter;
- MySQL, che permette la connessione ai database MySQL;
- Time, utile alla gestione dei dati temporali;
- Base64, per poter gestire la codifica e la decodifica degli URL. Si è utilizzato il modulo per poter ottenere le credenziali da utilizzare con l'API di Twitter;
- Urllib, modulo utile alla gestione degli URL, utilizzato anch'esso per ottenere le credenziali per l'API di Twitter.

Inoltre, si è utilizzato il microframework Flask [18] creato da Armin Ronacher nel 2010 e basato sul *template engine* Jinja2 e sul *toolkit* WSGI<sup>2</sup>

---

<sup>1</sup>L'indentazione è un sistema che per mezzo di spazi e tab raggruppa blocchi di codice sotto l'istruzione logica dalla quale dipendono, senza l'utilizzo di parentesi o parole chiave.

<sup>2</sup>Web Server Gateway Interface (WSGI): Specifica che descrive come un web server comunica con le applicazioni web e come queste possano essere incatenate tra loro per processare una richiesta [63].

Werkzeug. Flask mantiene un nucleo semplice ma estensibile, per mezzo di librerie esterne, e non impone scelte implementative come, ad esempio, il database da utilizzare. Il microframework è stato utilizzato per implementare la parte server-side del programma.

Python e i moduli relativi sono stati impiegati nell'estrazione dei dati dalle API dei social media, l'inserimento dei dati relativi nel database e nell'implementazione *back end* del progetto software.

### 3.1.2 Javascript

Javascript [26], proposto da Brendan Eich nel 1995 e diventato standard ECMA nel 1997, è nato come linguaggio di scripting<sup>3</sup> client-side, ma attualmente è utilizzato per tutti gli aspetti della programmazione anche quelli server-side. Javascript è conforme alle specifiche di ECMAScript e allo standard ECMA-262. ECMAScript è un linguaggio di programmazione orientato agli oggetti utile a eseguire computazioni e manipolare oggetti computazionali in un ambiente.

Javascript è estendibile tramite librerie e nella realizzazione del progetto sono state utilizzate le seguenti:

- jQuery, utile a manipolare documenti HTML;
- FileSaver.js, utile per il salvataggio di file client-side, utilizzato sia per salvare file CSV che le immagini;
- Canvas-toBlob.js, per convertire i grafici visualizzati nel canvas, un elemento HTML, in immagini;
- Chart.js, utile a creare i grafici.

L'utilizzo di Javascript è stato circoscritto alla parte client-side e di rappresentazione del programma.

---

<sup>3</sup>Linguaggio di scripting: *“Un linguaggio di scripting è un linguaggio di programmazione che è utilizzato per manipolare, personalizzare e automatizzare le funzioni di un sistema esistente[24].”*

### 3.1.3 HTML, CSS e Bootstrap

HTML<sup>4</sup> è un linguaggio di markup, che definisce come disporre gli elementi all'interno di una pagina web utilizzando dei 'tag'<sup>5</sup> e stabilire dei collegamenti tra pagine HTML. La sintassi di HTML è definita dal W3C (World Wide Web Consortium), un'organizzazione non governativa il cui scopo è sviluppare tutte le potenzialità del web.

CSS<sup>6</sup> è un linguaggio utile a descrivere la presentazione di pagine web per quanto concerne colori, struttura e *font*. Permette di rendere il contenuto della pagina adattabile ai diversi schermi ed è anch'esso gestito dal W3C. CSS separa il contenuto della pagina HTML dalla sua formattazione. Il linguaggio utilizza dei selettori per identificare gli elementi della pagina web e blocchi di dichiarazioni per modificarne lo stile.

Bootstrap[4], invece, è un *toolkit*, cioè un insieme di strumenti, sviluppato inizialmente per Twitter e poi reso open source, che mette a disposizione modelli grafici utili a modificare l'aspetto di una pagina web in modo semplice e veloce.

Nello sviluppo del progetto HTML, CSS e Bootstrap sono stati utilizzati per realizzare l'interfaccia grafica della piattaforma software.

### 3.1.4 MySQL

MySQL [44] fornisce un veloce, multithread, multiutente, e robusto database server SQL (Structured Query Language). MySQL è un software di Oracle ed ha una doppia licenza, una open source e una commerciale. MySQL è un RDBMS<sup>7</sup>, cioè un sistema software in grado di gestire collezioni di dati grandi, condivise e persistenti, il sistema è inoltre in grado di assicurare

---

<sup>4</sup>HyperText Markup Language (HTML): tradotto come linguaggio di marcatura per gli ipertesti.

<sup>5</sup>Il tag è un elemento racchiuso tra due <>, un esempio è <p> testo </p> in cui il tag <p> è utilizzato per definire un paragrafo.

<sup>6</sup>Cascading Style Sheets(CSS): tradotto come fogli di stile a cascata.

<sup>7</sup>Relational Database Management System (RDBMS): tradotto come sistema di gestione di basi di dati relazionali.

affidabilità e privacy in modo efficiente e efficace per mezzo del modello relazionale. Il modello relazionale permette di organizzare i dati in insiemi di record a struttura fissa. Gli insiemi sono chiamati relazioni e rappresentati come tabelle; le righe di queste tabelle rappresentano specifici record e le colonne rappresentano i campi, cioè gli attributi, del record.

MySQL è uno delle specializzazioni dello standard SQL, elaborato da Donald Chamberlin nel 1974 presso l'IBM, e permette l'esecuzione di tutte le operazioni necessarie alla gestione e all'utilizzo di un database.

Nella realizzazione del progetto, MySQL è stato utilizzato per rendere persistenti i dati raccolti, per permetterne l'analisi e rendere più efficiente l'utilizzo della piattaforma.

### 3.1.5 Facebook API

Facebook mette a disposizione degli sviluppatori diverse API<sup>8</sup> basate sui servizi che offre, ad esempio, ne esiste una per la gestione delle pubblicità: l'API Marketing, una per la trasmissione di dirette video: l'API Live ed altre ancora in base ai possibili bisogni degli sviluppatori.

In questo elaborato si è utilizzata l' API Graph che è un API a basso livello basata su HTTP<sup>9</sup> per poter effettuare query sui dati, pubblicare notizie e gestire account. L'API Graph fornisce la rappresentazione delle informazioni presenti su Facebook sotto forma di *social graph*<sup>10</sup>, composta da:

- nodi, cioè oggetti singoli come utenti e pagine, utile per ottenere dati su un oggetto specifico;

---

<sup>8</sup>Application Programming Interface (API): tradotto come interfaccia di programmazione di un'applicazione.

<sup>9</sup>HyperText Transfer Protocol (HTTP): tradotto come protocollo di trasferimento di un ipertesto.

<sup>10</sup>Social Graph: "rete di connessioni esistenti attraverso cui le persone(nodi) comunicano e si scambiano informazioni." Dave Morin, 2007 Facebook [19].

- segmenti, cioè connessioni tra una raccolta di oggetti e un singolo oggetto come i commenti a una foto, utili per ottenere raccolte di oggetti su un singolo oggetto;
- campi, cioè dati relativi a un oggetto come il nome di una pagina o la sua descrizione, utile per ottenere dati su un singolo oggetto o su ogni oggetto di una raccolta.

Per utilizzare l'API Graph è necessario possedere un token d'accesso, una stringa opaca che identifica un utente, un'applicazione o una pagina:

- Token d'accesso dell'utente, necessario quando un'applicazione effettua una chiamata API per leggere, modificare e scrivere dati Facebook di un utente per suo conto, per ottenerli è necessario che l'utente dia il consenso. Questo token può avere breve durata, circa due ore, o lunga durata, circa 60 giorni.
- Token d'accesso dell'applicazione, necessario per modificare e leggere le impostazioni dell'applicazione, generata tramite una chiave segreta tra l'applicazione e Facebook. Questo token permette di visualizzare solo dati pubblici.
- Token d'accesso della Pagina, necessario quando un'applicazione effettua una chiamata API per leggere, modificare e pubblicare dati relativi a una Pagina Facebook. Per ottenerne uno è necessario ottenere un token d'accesso dell'utente e poi richiedere l'autorizzazione come gestore di una pagina.

Nella realizzazione del progetto si è creata un'applicazione Facebook e si è utilizzato un token d'accesso dell'applicazione, poiché, oltre a utilizzare unicamente dati pubblici sulla piattaforma, non rende necessaria la registrazione dell'utente alla sistema per ottenerne il token.

Facebook inoltre mette a disposizione un tool di esplorazione, il Graph API Explorer, per poter esplorare il social graph di Facebook senza dover implementare esternamente una chiamata HTTP. L'applicazione permette di

scegliere il tipo di token d'accesso e ne genera uno per l'utente utilizzatore e permette di selezionare il tipo di azione da svolgere (GET, POST o DELETE) e la versione dell'API che si vuole utilizzare.

Ad esempio, nel caso in cui si voglia effettuare una GET, per estrarre dei dati, è necessario specificare quali dati si vuole richiedere. La richiesta può essere perfezionata tramite un pannello laterale con tutti i campi che è possibile estrarre. Una volta inviata la richiesta si possono visualizzare i dati estratti nel pannello apposito in formato JSON. Un esempio di utilizzo del Graph API Explorer è visibile in figura 3.1.

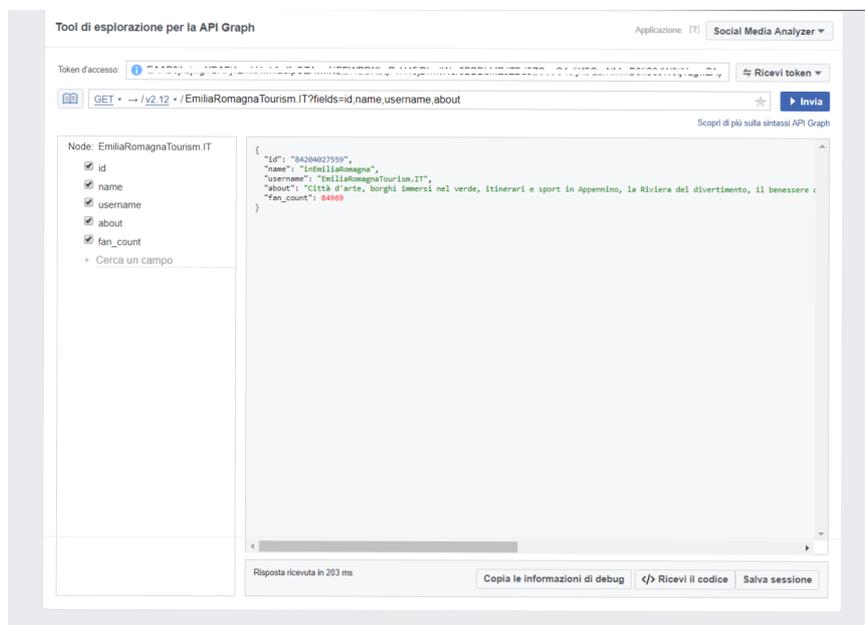


Figura 3.1: API Graph Explorer.

Per estrarre i dati si è utilizzato Python e il modulo *requests* per comporre la richiesta all'API Graph.

### 3.1.6 Twitter API

Twitter, come Facebook, mette a disposizione diverse API per gli sviluppatori, non tutte però sono gratuite poiché l'azienda ha definito un sistema di accesso ai dati a tre livelli:

- Standard, è il livello gratuito, può utilizzare solo API standard ed incorre in limitazioni per quello che concerne il numero di tweet che è possibile estrarre e le finestre di tempo utili all'estrazione.
- Premium, ha il costo medio e fornisce l'accesso a un numero superiore di dati.
- Enterprise, è il livello più costoso, fornisce un accesso completo ai dati, pacchetti dedicati, un gestore personale dell'account e supporto tecnico.

Twitter richiede inoltre un'autorizzazione OAuth<sup>11</sup>, che può essere:

- Autenticazione dell'utente, in cui si richiede l'accesso dell'utente al sistema e si permette all'applicazione di agire per suo conto.
- Autenticazione solo dell'applicazione, l'applicazione fa richieste all'API senza l'intervento dell'utente. Utile se è necessario accedere alle sole informazioni pubbliche.

Nello sviluppo della piattaforma si è utilizzato un accesso ai dati standard e un'autorizzazione della sola applicazione. Il metodo che si è utilizzato della Twitter API è *GET statuses/user\_timeline*, che restituisce una collezione dei tweet più recenti postati da un utente in base al nome o l'identificativo dello stesso. Il metodo restituisce al massimo 3200 dei tweet più recenti di un utente e nel totale vengono inclusi i retweet, di tweet propri o altrui, e le risposte ad altri tweet.

---

<sup>11</sup>OAuth: Protocollo che consente autorizzazioni sicure in un modo semplice e standardizzato per le applicazioni web, mobile e desktop [46].

## 3.2 Dettagli implementativi

### 3.2.1 Estrazione dei dati

#### Estrazione dati dall'API Graph di Facebook

Come anticipato nel paragrafo 3.1.5, per estrarre le informazioni dalla piattaforma Facebook si è utilizzata l'API Graph con una chiave dell'applicazione. Per poter ottenere una chiave dell'applicazione è necessario registrarsi come sviluppatori sulla piattaforma *Facebook for developers*<sup>12</sup> e creare un'applicazione.

```
# Il metodo definisce la richiesta all'API Graph.
def req_facebook(account):
    try:
        query = account + "?fields=id%2Cname%2Cfan_count%2Cabout%2Cusername%2Cposts.since(01%2F01%2F2017).until(01%2F01%2F2018)" \
            ".limit(999){created_time%2Ctype%2Cid%2Cshares%2Ccomments.limit(0)." \
            "summary(true)%2Creactions.limit(0).summary(true)%2Cmessage}"
        request = requests.get("https://graph.facebook.com/v2.12/" + query, {'access_token': app_token})
        return request
    except Exception as e:
        print e.message
```

Figura 3.2: Richiesta dati API Graph di Facebook.

La richiesta all'API, visibile in figura 3.2 è composta dal nome utente della pagina<sup>13</sup> che si vuole analizzare e dall'interrogazione sui dati che si vogliono estrarre, in questo caso: l'identificativo della pagina, il nome della pagina, il totale dei fan della pagina, una breve descrizione della pagina, il nome utente della pagina e tutti i post pubblicati tra il 1 gennaio 2017 e il 1 gennaio 2018 escluso. Di ogni post si è estratta la data di pubblicazione, il codice identificativo, il numero di condivisioni, il numero di commenti e il numero di reazioni (intesi come: *Mi piace*, *Love*, *Wow*, *Ahah*, *Sigh*, *Grrrr*) e il testo del post. L'API Graph permette di estrarre circa 600 post riferiti a un solo anno per ogni pagina Facebook.

Il risultato della richiesta è convertito in formato JSON<sup>14</sup> per poter filtrare

<sup>12</sup><https://developers.facebook.com/>

<sup>13</sup>Nome utente pagina facebook: <https://www.facebook.com/nomeutente>

<sup>14</sup>JSON è un formato per lo scambio di dati. Basato su coppie nome/valore ed elenchi ordinati di valori [27].

più agevolmente i dati. La figura 3.3 mostra il *parsing*<sup>15</sup> dei dati relativi alla pagina Facebook, in particolare viene verificata la presenza del campo *about*, poiché non tutte le pagine forniscono una breve descrizione. Si verifica poi che siano presenti dei post. Per ogni post - come mostrato in figura 3.5 - controlla che sia presente un testo e, in caso di successo, ne calcola la lunghezza; se il testo non è presente la lunghezza viene posta uguale a 0. Anche le condivisioni necessitano di un controllo aggiuntivo, poiché se non sono presenti il campo *shares* risulta mancante.

```
#controllo che l'API non torni un errore
if 'error' in results.keys():
    return 'NULL PAGE NAME'
else:
    page_id = results['id']
    username = results['username']
    screen_name = results['name']

    # controllo che esista un campo 'about'
    if 'about' in results.keys():
        about = results['about']
    else:
        about = " "

    fan_count = results['fan_count']

# chiamata per inserire dati pagina nel database
insert_facebook_page(page_id, username, screen_name, about, fan_count)
```

Figura 3.3: Parsing dati relativi alla pagina Facebook.

```
insert_query = "INSERT INTO FACEBOOK_PAGE (page_id, username, screen_name, about, tot_page_likes) " \
               " VALUES (%s, %s, %s, %s, %s)"
cursor.execute(insert_query, (page_id, username, screen_name, about, fan_count))
```

Figura 3.4: Inserimento dati pagina Facebook nel database.

Le figure 3.4 e 3.6 mostrano le funzioni utilizzate per inserire i dati estratti nel database.

Infine, una volta estratti i dati, viene chiamata una *stored procedure*<sup>16</sup> sul database - visibile in figura 3.7 - per calcolare l'engagement di tutti i post dell'account in analisi.

<sup>15</sup>Parsing: analisi sintattica.

<sup>16</sup>La *stored procedure* è un programma scritto in SQL mantenuto dal database.

```

# controlla che siano presenti post
if 'posts' in results.keys():
    i = 0
    while i < len(results['posts']['data']):
        try:
            user_id = results['id']
            id_post = results['posts']['data'][i]['id']
            created_time = results['posts']['data'][i]['created_time']
            post_type = results['posts']['data'][i]['type']

            # controlla che il post abbia un testo e ne calcola la lunghezza
            if 'message' in results['posts']['data'][i].keys():
                post_length = len(results['posts']['data'][i]['message'])
            else:
                post_length = 0

            # controlla che il post sia stato condiviso
            if 'shares' in results['posts']['data'][i].keys():
                shares_count = results['posts']['data'][i]['shares']['count']
            else:
                shares_count = 0

            likes_count = results['posts']['data'][i]['reactions']['summary']['total_count']
            comments_count = results['posts']['data'][i]['comments']['summary']['total_count']

            insert_facebook_post(id_post, user_id, created_time, post_length, likes_count, comments_count,
                                shares_count, post_type)

            i += 1
        except Exception as e:
            if i >= len(results['posts']['data']):
                print "end"
            else:
                print "ERROR : ", str(e)
            break
        except ValueError as e:
            print 'VALUE ERROR : ', str(e)
    else:
        return "VOID ERROR"
# inserisco l'engagement dei post
set_facebook_post_engagement(account)

```

Figura 3.5: Parsing dati relativi ai post di una pagina Facebook.

```

insert_query = "INSERT INTO FACEBOOK_POST (post_id, post_author, created_time, post_length, tot_likes, " \
              "tot_comments, tot_shares, type) VALUES (%s, %s, %s, %s, %s, %s, %s, %s)"
cursor.execute(insert_query, (id_post, user_id, created_time, post_length, likes_count, comments_count, shares_count,
                              post_type))

```

Figura 3.6: Inserimento dati post Facebook nel database.

## Estrazione dati dalla Twitter API

Per poter estrarre i dati da Twitter è necessario registrarsi come sviluppatore e creare un'applicazione su *Twitter Apps*<sup>17</sup>. La creazione del token dell'applicazione, per poter accedere i dati, richiede più fasi:

1. Codificare la *consumer key* e quella *secret* nella codifica URL, concatenare le chiavi codificate tramite un carattere ':' e codificare la stringa ottenuta in Base64, un sistema di codifica che consente di tradurre i

<sup>17</sup><https://apps.twitter.com/>

```

DELIMITER $
CREATE PROCEDURE set_facebook_post_engagement (IN _username VARCHAR(50))
BEGIN
  DECLARE engagement_sum DOUBLE;
  DECLARE finish INT DEFAULT 0;
  DECLARE likes INT DEFAULT 0;
  DECLARE comments INT DEFAULT 0;
  DECLARE shares INT DEFAULT 0;
  DECLARE post VARCHAR(50) DEFAULT "";

  DECLARE cur CURSOR FOR SELECT post_id, tot_likes, tot_comments, tot_shares
    FROM FACEBOOK_POST, FACEBOOK_PAGE
    WHERE post_author = page_id AND FACEBOOK_PAGE.username = _username;

  DECLARE CONTINUE HANDLER FOR NOT FOUND SET finish = 1;

  OPEN cur;
  getPost: LOOP
  FETCH cur INTO post, likes, comments, shares;

    IF finish = 1 THEN
      LEAVE getPost;
    END IF;

    SET engagement_sum = (likes * 0.2 + comments * 0.5 + shares * 0.3);
    INSERT INTO FACEBOOK_METRICS VALUES (post, engagement_sum);

  END LOOP getPost;
  CLOSE cur;
END;
$

```

Figura 3.7: *Stored procedure* per il calcolo dell'engagement dei post di una pagina Facebook.

dati binari in stringhe di testo ASCII rappresentandoli sulla base di 64 ASCII diversi.

2. Ottenere un Bearer Token<sup>18</sup>, inviando una richiesta HTTP di tipo POST e includendo l'autorizzazione del codice in Base64 e un content-type.
3. Utilizzare il token ottenuto per effettuare la richiesta all'API.

La richiesta all'API - mostrata in figura 3.8 - è composta dal numero di tweet da estrarre, il nome utente<sup>19</sup> del profilo di cui estrarre i dati, l'esclusione dei retweet, l'id massimo, cioè quello più datato, da cui partire a estrarre i dati.

L'API restituisce al massimo duecento tweet a richiesta e nel totale sono inclusi retweet ai tweet di altri e ai propri, perciò vengono effettuate più

<sup>18</sup>Bearer Token: Una chiave di sicurezza con la proprietà di garantire l'uso della chiave a chiunque ne sia in possesso senza richiedere l'autenticazione.

<sup>19</sup>Nome utente profilo Twitter: <https://twitter.com/nomeutente>

```
# metodo per effettuare la richiesta all'API di Twitter
def req_twitter(account, count, latest_tweet):
    headers = {'Authorization': 'Bearer {}'.format(access_token)}
    params = {'count': count, 'screen_name': account, 'include_rts': 'false', 'max_id': latest_tweet}
    response = requests.get(SEARCH_TWEETS_URL, params=params, headers=headers)
    if response.status_code != requests.codes.ok:
        raise Exception('Invalid request')
    return response
```

Figura 3.8: Richiesta dati Twitter API.

chiamate all'API - figura 3.9 - fino a quando non viene raggiunto il totale di 500 rilevazioni, compresi sia i tweet che le risposte del profilo in questione ad altri tweet. La richiesta non comprende le risposte al tweet pubblicato poiché l'API non ne consente l'estrazione diretta.

```
total_tweets = len(statuses)
while total_tweets != 500:
    if latest_tweet is None:
        break
    remaining_tweet = 500 - total_tweets
    statuses = req_twitter(account, remaining_tweet, latest_tweet - 1).json()
    latest_tweet = parse_tweets(statuses)
    total_tweets = total_tweets + len(statuses)
```

Figura 3.9: Algoritmo per effettuare più chiamate alla Twitter API.

Il risultato della richiesta all'API è stato convertito anch'esso in formato JSON. Dai dati restituiti sono estratti - figura 3.10 - l'identificativo dell'utente, il numero totale di followers, il numero totale di following, il numero totali di tweet pubblicati dal profilo considerato, il nome utente del profilo, il nome del profilo e una breve descrizione. Per ogni tweet - come mostrato in figura 3.11 - viene estratto: l'identificativo del tweet, l'autore del tweet, la data di creazione, la lunghezza del tweet, il numero 'mi piace', il numero di retweet, se il tweet è una risposta o meno e gli hashtag utilizzati.

Le figure 3.12 e 3.13 mostrano le funzioni utilizzate per inserire i dati estratti nel database di riferimento.

Gli hashtag utilizzati vengono inseriti in una tabella separata del database, in modo da evitare duplicati nella tabella relativa ai tweet.

Anche in questo caso, come per Facebook, una volta popolato il database viene richiamata la *stored procedure* per calcolare ed inserire l'engagement.

```
# controlla che l'account abbia pubblicato tweet
if len(statuses) == 0:
    return 'VOID ERROR'

user_id = statuses[0]['user']['id']
followers_count = statuses[0]['user']['followers_count']
following_count = statuses[0]['user']['friends_count']
statuses_count = statuses[0]['user']['statuses_count']
about = statuses[0]['user']['description']
username = statuses[0]['user']['screen_name']
screen_name = statuses[0]['user']['name']

# inserisce dati account twitter
insert_twitter_account(user_id, username, screen_name, about, statuses_count, followers_count, following_count)
```

Figura 3.10: Parsing dati relativi al profilo Twitter.

```
def parse_tweets(statuses):
    latest_tweet = None
    for status in statuses:
        try:
            tweet_id = status['id']
            post_author = status['user']['id']
            created_time = time.strftime('%Y-%m-%d %H:%M:%S',
                                         time.strptime(status['created_at'], '%a %b %d %H:%M:%S +0000 %Y'))
            tweet_length = len(status['text'])
            favourite_count = status['favorite_count']
            retweet_count = status['retweet_count']

            # controlla se il tweet è una risposta
            if status['in_reply_to_status_id'] is None:
                is_reply = 0
            else:
                is_reply = 1
            latest_tweet = int(tweet_id)

            insert_twitter_tweet(tweet_id, post_author, created_time, tweet_length, favourite_count, retweet_count,
                                is_reply)

            # salva gli hashtag del tweet
            for tweet_hashtag in status['entities']['hashtags']:
                hashtag = tweet_hashtag['text']
                insert_tweet_hashtag(tweet_id, hashtag)

        except Exception as e:
            print 'ERROR: ', e
    return latest_tweet
```

Figura 3.11: Parsing dati relativi ai tweet di un profilo Twitter.

```
insert_query = "INSERT INTO TWITTER_ACCOUNT (user_id, username, screen_name, about, statuses_count, " \
              "followers_count, following_count) " \
              "VALUES (%s, %s, %s, %s, %s, %s, %s)"
cursor.execute(insert_query, (user_id, username, screen_name, about, statuses_count, followers_count,
                              following_count))
```

Figura 3.12: Inserimento dati profilo Twitter nel database.

```
insert_query = "INSERT INTO TWEET (tweet_id, post_author, created_time, tweet_length, favourite_count, " \
              "retweet_count, is_reply) " \
              "VALUES (%s, %s, %s, %s, %s, %s, %s)"
cursor.execute(insert_query, (tweet_id, post_author, created_time, tweet_length, favourite_count, retweet_count,
                              is_reply))
```

Figura 3.13: Inserimento dati tweet nel database.

### 3.2.2 Visualizzazione dei dati

Per permettere la visualizzazione dei dati si è utilizzato Flask[18]. Il microframework permette di specificare diversi URL per attraversare la piat-

taforma dinamicamente, un esempio si ha in figura 3.14.

```
@app.route("/")
def index():
    session.pop('social', None)
    session.pop('account', None)
    return render_template('index.html')
```

Figura 3.14: Definizione URL in Flask.

Le schermate sono già state presentate in precedenza nel paragrafo 2.2 relativo all'architettura del progetto, in questo paragrafo ne verrà trattata solo l'implementazione.

### Pagina principale

La pagina principale è la prima schermata che viene mostrata. Quando l'utente clicca sul tasto 'ok' per confermare i dati inseriti, questi vengono controllati e in caso di successo si viene reindirizzati alla pagina di visualizzazione dei dati. La figura 3.15 mostra i controlli effettuati sulla selezione del social media e l'inserimento del nome utente. Innanzitutto, entrambi i campi devono essere compilati. Successivamente, si verifica se l'account è già stato analizzato o meno: se l'account è già presente nel database i dati non vengono estratti nuovamente e si viene direttamente reindirizzati alla pagina relativa, in caso contrario, viene effettuata una richiesta all'API - già affrontata nel paragrafo precedente - e solo successivamente reindirizzati.

### Visualizzazione dati raccolti

La figura 3.16 mostra il codice di uno dei metodi utilizzati per generare la pagina di visualizzazione dei dati. Vengono estratti dal database tramite interrogazioni - un esempio viene mostrato in figura 3.17 - tutti i dati relativi all'account che si vuole analizzare e vengono passati al relativo *template*<sup>20</sup> per la visualizzazione.

<sup>20</sup>Il *template* è un modello predefinito di un documento.

```

@app.route("/get_account", methods=['GET'])
def get_account():
    try:
        error = None
        if request.method == 'GET':
            account = request.args.get('account', None)
            social = request.args.get('social', None)

            if not account and not social:
                error = "You have to select a social and insert an account."
                return render_template('index.html', error=error)
            else:
                if not account:
                    error = "You have to insert an account."
                    return render_template('index.html', error=error)

                if not social:
                    error = "You have to select a social."
                    return render_template('index.html', error=error)
                elif social == 'facebook':
                    check_account = check_facebook_page(account)
                    if check_account[0] > 0:
                        account = check_account[1]
                    else:
                        account = get_facebook_data(account)
                        if account == 'NULL PAGE NAME':
                            error = "There is no page with this account, please insert another or check the spelling."
                            return render_template('index.html', error=error)
                        elif account == 'VOID ERROR':
                            error = "The page has no posts, please insert another one or check the spelling."
                            return render_template('index.html', error=error)
                        session['social'] = 'facebook'
                        session['account'] = account
                        return redirect('facebook/' + str(account))
                elif social == 'twitter':
                    check_account = check_twitter_account(account)
                    if check_account[0] > 0:
                        account = check_account[1]
                    else:
                        account = get_twitter_data(account)
                        if account == 'NULL PAGE NAME':
                            error = "There is no account with this name, please insert another or check the spelling."
                            return render_template('index.html', error=error)
                        elif account == 'VOID ERROR':
                            error = "The account has no tweets, please insert another one or check the spelling."
                            return render_template('index.html', error=error)
                        session['social'] = 'twitter'
                        session['account'] = account
                        return redirect('twitter/' + str(account))
                return render_template('index.html', error=error)
    except Exception as e:
        print "ERROR: ", str(e.message)

```

Figura 3.15: Indirizzamento alla pagina di visualizzazione dei dati.

Una volta estratti i dati e generato il *template* relativo si ha la gestione client-side della creazione dei grafici e del salvataggio come immagine o come file CSV.

La figura 3.18 rappresenta un metodo per la generazione di un grafico, quello relativo all'engagement mensile di una pagina Facebook, per mezzo della libreria Chart.js [6].

La figura 3.19 rappresenta il codice per salvare il grafico come immagine.

Infine, la figura 3.20 rappresenta il metodo per preparare i dati da inserire nel file CSV e il codice per il salvataggio del file.

```

@app.route("/facebook/<string:account>")
def single_facebook(account):
    try:
        if account:
            # page data
            page_dict = get_facebook_page_data(account)
            engagements = get_facebook_page_engagement(account)
            page_dict['engagement'] = engagements[0]
            page_dict['engagement_norm'] = engagements[1]

            # post daily data
            posts = get_posts_day(account)
            posts_day_engagement = {
                'labels': posts[0],
                'post_value': posts[1],
                'engagement_value': posts[2]
            }

            # post daily interactions
            posts_interactions = get_daily_post_interactions(account)

            # post monthly data
            posts_month = get_posts_month(account)
            post_month_engagement = {
                'labels': posts_month[0],
                'post_value': posts_month[1],
                'engagement_value': posts_month[2]
            }

            # post daily interactions
            post_monthly_interactions = get_monthly_post_interactions(account)

            # post type engagement
            posts_type_engagement = get_post_engagement_type(account)

            # post length engagement
            posts_length_engagement = get_post_engagement_length(account)

            return render_template('single_page.html', social='facebook', account=page_dict,
                                   post_day_engagement=posts_day_engagement, post_interactions=posts_interactions,
                                   post_month_engagement=post_month_engagement,
                                   post_monthly_interactions=post_monthly_interactions,
                                   posts_type_engagement=posts_type_engagement,
                                   posts_length_engagement=posts_length_engagement)
        else:
            error = "Something went wrong, please try again."
            return render_template('index.html', error=error)
    except Exception as e:
        print "ERROR : ", str(e.message)

```

Figura 3.16: Estrazione dei dati relativi a una singola pagina Facebook per la visualizzazione.

```

# metodo che estrae i dati relativi alla pagina Facebook dal database.
def get_facebook_page_data(page):
    try:
        conn = mysql.connect()
        cursor = conn.cursor()
        query = "SELECT * " \
                "FROM FACEBOOK_PAGE " \
                "WHERE username = %s;"
        cursor.execute(query, page)
        page_data = cursor.fetchall()
        page_dict = {
            'username': page_data[0][1],
            'screen_name': page_data[0][2],
            'about': page_data[0][3],
            'fan_count': page_data[0][4]
        }
        cursor.close()
        conn.close()
        return page_dict
    except Exception as e:
        print 'ERROR: ', str(e)

```

Figura 3.17: Interrogazione al database per l'estrazione dei dati relativi a una pagina Facebook.

```

// funzione che crea il grafico di engagement mensile
function post_engagement_monthly_chart(labels, posts, engagement) {
  typeChart = "avg_monthly_eng";
  let mychart = document.getElementById("myChart");
  postChart = new Chart(mychart, {
    type: 'bar',
    data: {
      labels: labels,
      datasets: [{
        label: 'Avg engagement (R*0.2 + C*0.5 + S*0.3)/P',
        type: "line",
        yAxisID: 'left',
        backgroundColor: "rgb(165, 51, 103)",
        borderColor: "rgb(165, 51, 103)",
        data: engagement,
        fill: false
      }, {
        label: '#Post/Month',
        type: 'bar',
        yAxisID: 'right',
        backgroundColor: "rgba(140, 247, 239, 0.8)",
        borderColor: "rgb(0, 211, 196)",
        borderWidth: 1,
        data: posts
      }
    ]
  }
},

```

(a)

```

options: {
  title: {
    display: true,
    text: 'Monthly posts engagement'
  },
  scales: {
    xAxes: [{
      scaleLabel: {
        display: true,
        labelString: 'Month'
      }
    }],
    yAxes: [{
      scaleLabel: {
        display: true,
        labelString: 'Engagement (R*0.2 + C*0.5 + S*0.3)/P'
      },
      id: 'left',
      position: 'left',
      ticks: {
        beginAtZero: true
      }
    }, {
      scaleLabel: {
        display: true,
        labelString: '#Post'
      },
      id: 'right',
      position: 'right',
      ticks: {
        beginAtZero: true
      }
    }
  ]
}
});
}

```

(b)

Figura 3.18: Esempio metodo per generare grafico relativo all'engagement mensile di una pagina Facebook.

```
// gestione salvataggio grafico
$("#save-image-btn").click(function(){
    $("#myChart").get(0).toBlob(function(blob) {
        var fileName = page + "_" + typeChart + "_fb_single.png";
        saveAs(blob, fileName);
    });
});
```

Figura 3.19: Metodo per salvare il grafico come immagine.

```
//funzione che ordina i dati per creare il file csv
function create_csv_interactions_array(labels, first_value, second_value, third_value, typeChart){
    csv_array = [];
    if(typeChart == "post_daily_interactions"){
        var header_array = ["Day", "#reactions_day", "#comments_day", "#shares_day"];
    } else if (typeChart == "post_monthly_interactions"){
        var header_array = ["Month", "#reactions_month", "#comments_month", "#shares_month"];
    }
    csv_array.push(header_array);

    for(var i = 0; i < labels.length; i++){
        var inner_array = [labels[i], first_value[i], second_value[i], third_value[i]];
        csv_array.push(inner_array);
    }
}

//salvataggio file csv
$("#save-csv-btn").click(function(){
    console.log('csv');
    var csvContent = "";
    console.log(csv_array);
    csv_array.forEach(function(row, index){
        csvContent += row.join(",") + "\n";
    });

    var blob = new Blob([csvContent], { type: 'text/csv;charset=utf-8;' });
    var fileName = page + "_" + typeChart + "_fb_single.csv";
    saveAs(blob, fileName);
});
```

Figura 3.20: Metodi preparare i dati e per salvarli in un file csv.



# Capitolo 4

## Analisi dei dati

In questo ultimo capitolo viene trattata l'analisi dei dati e si è considerato come esempio di *brand comparison* l'analisi degli account ufficiali per la promozione turistica delle regioni italiane. In particolare, il paragrafo 4.1 tratta delle modalità di raccolta dei dati, il paragrafo 4.2 delle metriche utilizzate nell'analisi, il paragrafo 4.3 dell'analisi quantitativa dei dati, il paragrafo 4.4 della regressione effettuata sui dati e infine il paragrafo 4.5 confronta i risultati delle analisi tra le due piattaforme Facebook e Twitter.

### 4.1 Modalità di raccolta dei dati

Per poter raccogliere i dati relativi ai soli account turistici ufficiali delle regioni italiane si è effettuata una ricerca riguardante i siti web degli enti regionali in modo da ottenere i collegamenti diretti agli account ufficiali ed escludere le pagine gestite da altre entità. Nel caso in cui il sito web non contenesse i collegamenti si è effettuata una ulteriore ricerca direttamente nel social media di riferimento e si è verificata l'ufficialità tramite la descrizione dell'account. Le regioni Campania e Molise non hanno né pagine Facebook, né profili Twitter ufficiali attualmente attivi, per questo motivo le due regioni sono state escluse dall'analisi. La regione Calabria, invece, ha solo la pagina Facebook ufficiale per cui non è presente nell'analisi relativa a Twitter. Infine,

la regione Trentino - Alto Adige presenta due diverse pagine di riferimento per la promozione turistica ufficiale, una per il Trentino e una per l'Alto Adige, nell'analisi le due vengono considerate entrambe e separatamente. Si è effettuata l'analisi considerando gli account italiani e la tabella 4.1 riporta per ogni regione considerata il collegamento alla pagina Facebook e al profilo Twitter da cui si sono estratti i dati utilizzando la piattaforma sviluppata.

<b>Regione</b>	<b>Account Facebook</b> <i>http://www.facebook.com/</i>	<b>Account Twitter</b> <i>https://twitter.com/</i>
Abruzzo	visit.abruzzo	YourAbruzzo
Alto Adige	altoadige.suedtirol	altoadige_info
Basilicata	BasilicataTuristica	Basilicata_Tur
Calabria	TurismoRegioneCalabria	—
Emilia - Romagna	EmiliaRomagnaTourism.IT	turismoER
Friuli - Venezia Giulia	friuliveneziagiulia.turismo	FVGlive
Lazio	visitlazio	visit_lazio
Liguria	turismoinliguria	TurismoLiguria
Lombardia	inLOMBARDIA	inLOMBARDIA
Marche	marche.tourism	MarcheTourism
Piemonte	VisitPiemonte	Visit_Piemonte
Puglia	viaggiareinpuglia.it	viaggiarepuglia
Sardegna	VisitSardinia	SardegnaTurismo
Sicilia	visitsicily.info	VisitSicilyOP
Toscana	VisitTuscanyIt	VisitTuscanyIT
Trentino	visittrentino	VisitTrentino
Umbria	Umbriatourism	UmbriaTourism
Valle D'Aosta	visitdaosta	Valle_dAosta
Veneto	TurismoInVeneto	TurismoVeneto

Tabella 4.1: Account ufficiali per la promozione turistica delle regioni italiane su Facebook e Twitter.

Per ogni pagina Facebook si sono estratti:

- *id*: identificativo univoco per la pagina;
- *name*: nome della pagina;
- *username*: nome utente della pagina, utile a trovare la stessa più facilmente sul social;
- *about*: breve descrizione della pagina;
- *tot\_page\_likes*: numero di persone che hanno messo ‘mi piace’ alla pagina.

Per ogni post della pagina si è estratto:

- *id*: identificativo univoco del post;
- *post\_author*: id della pagina che ha pubblicato il post;
- *created\_time*: giorno e ora a cui è stato pubblicato il post;
- *post\_length*: lunghezza del post;
- *tot\_reactions*: numero totale di reazioni (‘Mi piace’, ‘Love’, ‘Wow’, ‘Ahah’, ‘Sigh’ e ‘Grrr’) ricevute dal post;
- *tot\_comments*: numero totale di commenti al post;
- *tot\_shares*: numero totale di condivisioni del post;
- *type*: tipologia di post pubblicato.

Per i profili Twitter invece si sono estratti:

- *id*: identificativo univoco per il profilo;
- *name*: nome del profilo;
- *username*: nome utente del profilo;
- *about*: breve descrizione del profilo;

- *tot\_tweets*: numero di tweet totali pubblicati dal profilo;
- *followers*: numero di account che seguono il profilo, cioè ne sono fan;
- *following*: numero di account che il profilo segue, cioè i profili di cui l'utente è fan.

Per ogni tweet del profilo si è estratto:

- *id*: identificativo univoco del tweet;
- *tweet\_author*: id del profilo che ha pubblicato il tweet;
- *created\_time*: giorno e ora a cui è stato pubblicato il tweet;
- *tweet\_length*: lunghezza del tweet;
- *tot\_favourites*: numero totale di 'mi piace' ricevuti dal tweet;
- *tot\_retweet*: numero totale di condivisioni del tweet;
- *is\_reply*: se il tweet è una risposta o meno.

Con Twitter inoltre si è memorizzato ogni hashtag presente in un determinato tweet salvando l'id del tweet e l'hashtag utilizzato.

Oltre ai dati mostrati dalla piattaforma, engagement, tipologia e lunghezza dei post, si è approfondita l'analisi interrogando direttamente il database per ottenere ulteriori dati e poter analizzare singolarmente e confrontare l'utilizzo dei due social media in esame.

## 4.2 Metriche utilizzate nell'analisi

Per l'analisi si è utilizzata la metrica<sup>1</sup> dell'engagement, una delle metriche più diffuse che misura il grado di coinvolgimento dell'utente in relazione a un

---

<sup>1</sup>Una metrica è un sistema di misurazione che quantifica caratteristiche statiche o dinamiche e che quindi descrive uno stato, un processo, un trend o un'evoluzione. Le metriche vengono utilizzate per definire degli obiettivi da parte dell'impresa, ne viene misurato il grado e successivamente l'impresa si impegna a migliorarle[48].

determinato account o contenuto. Il calcolo dell'engagement varia per ogni social media, poiché ogni piattaforma mette a disposizione degli utenti diversi strumenti per interagire con i contenuti pubblicati. La formula generica per l'engagement è rappresentabile dalla formula:

$$\text{Engagement} = \frac{(\#likes + \#comments + \#shares)}{\#posts} \quad (4.1)$$

In questo elaborato si è utilizzata la formula (4.2) proposta da Mariani et al.(2016) [37] in cui l'engagement viene calcolato utilizzando diversi pesi per i 'Mi piace' (20%), i commenti (50%) e le condivisioni (30%). I diversi pesi sono stati definiti in base al coinvolgimento dell'utente nello svolgerli, cioè il tempo che richiede l'operazione e quanto questa guadagni visibilità. Mettere un 'mi piace' a un post e condividerlo sono azioni che richiedono un semplice click, la condivisione porta però più visibilità al post e quindi ha un peso maggiore. I commenti, al contrario, richiedono più coinvolgimento dell'utente poiché esso deve impiegare tempo nello scriverli per poter condividere un suo pensiero.

$$\text{Engagement Facebook} = \frac{(\#reactions * 0.2 + \#comments * 0.5 + \#shares * 0.3)}{\#post} \quad (4.2)$$

Anche l'engagement normalizzato (4.3) è stato calcolato seguendo questa formula, ma normalizzando il risultato ottenuto al numero di *likes* alla pagina.

$$\text{Engagement Facebook normalizzato} = \frac{(\#reactions * 0.2 + \#comments * 0.5 + \#shares * 0.3)}{\#post * \#page\_likes} \quad (4.3)$$

Per mantenere una linea coerente nel calcolo delle metriche si è applicato il calcolo ponderato anche a Twitter. Tuttavia, poiché in Twitter non si sono

potute estrarre le risposte - cioè i commenti al tweet - si sono modificati i pesi dell'equazione assegnando ai 'mi piace' un peso del 40% e ai retweet un peso del 60%. Così facendo si è mantenuta la logica secondo cui i *likes* portino meno diffusione del tweet rispetto alle condivisioni. Le formule risultanti sono quindi la formula (4.4) e formula (4.5). Nel caso dell'engagement normalizzato si è considerato il numero di *followers*, che corrispondono ai fan del profilo.

$$\begin{aligned} & \textbf{Engagement Twitter} \\ & \frac{(\#likes * 0.4 + \#retweets * 0.6)}{\#tweet} \end{aligned} \tag{4.4}$$

$$\begin{aligned} & \textbf{Engagement Twitter normalizzato} \\ & \frac{(\#likes * 0.4 + \#retweets * 0.6)}{\#tweet * \#followers} \end{aligned} \tag{4.5}$$

Si sono utilizzati esclusivamente i dati pubblici, è quindi probabile che un post o un tweet abbia ricevuto più interazioni di quelle rilevate dalla piattaforma a causa delle impostazioni relative alla privacy definite dagli utenti.

## 4.3 Analisi quantitativa dei dati

### 4.3.1 Analisi quantitativa dati di Facebook

Come anticipato, le regioni Campania e Molise non possiedono pagine turistiche ufficiali e sono perciò state escluse dall'analisi, mentre la regione Trentino - Alto Adige è presente con due pagine separate che vengono quindi considerate come entità a sé stanti.

I dati estratti sono relativi all'anno solare 2017 e comprendono post pubblicati dal 1 gennaio 2017 al 31 dicembre 2017, compreso. In totale si sono estratti 9.526 post, che hanno generato più di 5 milioni di interazioni totali suddivise in: 4.007.759 reazioni, 93.371 commenti e 921.373 condivisioni.

In particolare, è emerso che 54 post (0,57% del totale) non hanno ricevuto nessuna interazione, 73 post non hanno nessun ‘mi piace’ (0,77% del totale), 724 post (7,60% del totale) non sono stati condivisi e 2.195 post non sono mai stati commentati (23,04% del totale). Le tabelle 4.2, 4.3 e 4.4, rappresentano rispettivamente la distribuzione del numero di reazioni, di commenti, e di condivisioni sui post.

Nu mero reazioni	Numero post	% su totale post
0 – 10	653	6,85
11 – 100	2778	29,16
101 – 500	3927	41,22
501 – 1000	1328	13,94
1001 – 5000	788	8,27
5001 – 10000	42	0,44
> 10000	10	0,10

Tabella 4.2: Numero di reazioni per post.

Numero commenti	Numero post	% su totale post
0 - 10	7353	77,19
11 – 100	2049	21,51
101 – 500	120	1,26
501 – 1000	1	0,01
> 1000	3	0,03

Tabella 4.3: Numero di commenti per post.

Analizzando in dettaglio i dati relativi alle singole regioni - tabella 4.5 - è emerso che Alto Adige, Sicilia, Trentino, Marche, Lombardia e Veneto sono le regioni ad avere più fan, contando ognuna più di 200.000 *likes* alla propria pagina Facebook.

Numero condivisioni	Numero post	% su totale post
0 – 10	2824	29,65
11 – 100	4596	48,25
101 – 500	1842	19,34
501 – 1000	147	1,54
1001 – 5000	114	1,20
5001 – 10000	1	0,01
> 10000	2	0,02

Tabella 4.4: Numero di condivisioni per post.

Regione	Fan della pagina	#Post	Engagement	Engagement normalizzato (%)
Abruzzo	132247	598	75,68	0,06
Basilicata	82599	584	79,02	0,10
Calabria	6390	593	2,76	0,04
Emilia – Romagna	81475	593	38,15	0,05
Friuli - Venezia Giulia	93456	356	227,03	0,24
Lazio	159982	598	70,07	0,04
Liguria	161593	557	196,08	0,12
Lombardia	223779	602	178,01	0,08
Marche	242668	590	236,15	0,10
Piemonte	25987	152	7,06	0,03
Puglia	149433	613	88,01	0,06
Sardegna	49353	308	27,32	0,06
Sicilia	323664	615	179,91	0,06
Toscana	30510	455	101,27	0,33
Trentino	257652	571	112,92	0,04
Alto Adige	339112	157	271,28	0,08
Umbria	44631	458	23,26	0,05
Valle D'Aosta	65979	574	120,48	0,18
Veneto	203247	552	223,79	0,11

Tabella 4.5: Panoramica sulle pagine Facebook in analisi: numero fan, post pubblicati, engagement e engagement normalizzato percentuale

Se si considera l'engagement, le regioni che prevalgono sono Alto Adige, Lombardia, Friuli - Venezia Giulia e Veneto - figura 4.1.



Figura 4.1: Engagement delle pagine Facebook per la promozione turistica regionale.

Mentre, normalizzando l'engagement al numero dei fan della pagina la Toscana si classifica prima, seguita da Friuli - Venezia Giulia e Valle d'Aosta - figura 4.2. È importante notare come tutte le regioni con un engagement alto siano regioni del nord d'Italia.

Sicilia, Puglia e Lombardia sono state, invece, le più attive sul social pubblicando ciascuna più di 600 post.

### Distribuzione dei post nel tempo

La divisione su base mensile della regione Sicilia è visibile in figura 4.3 - l'immagine è stata generata utilizzando la piattaforma sviluppata - dove è possibile osservare anche l'engagement medio di ogni mese.

Come si può notare già dalla sola analisi della regione Sicilia, il numero di post diminuisce nella seconda metà dell'anno nel periodo successivo al mese

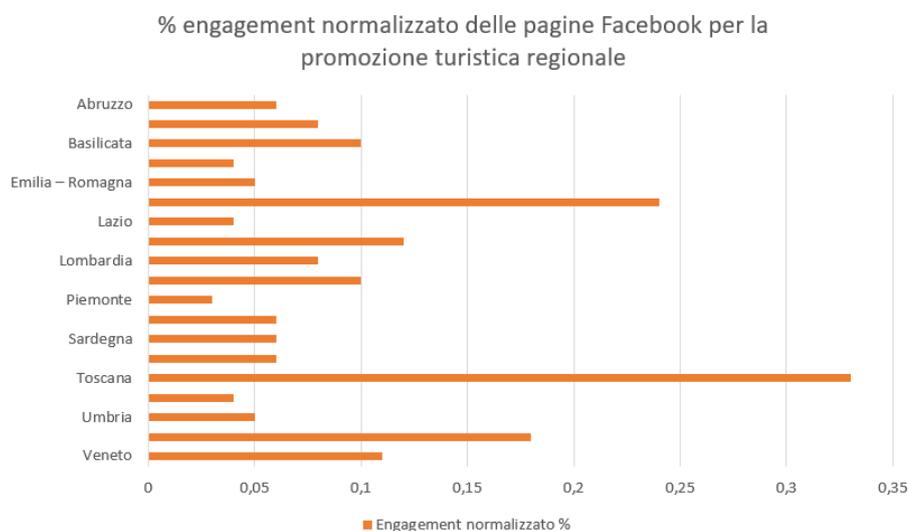


Figura 4.2: Engagement normalizzato delle pagine Facebook per la promozione turistica regionale.

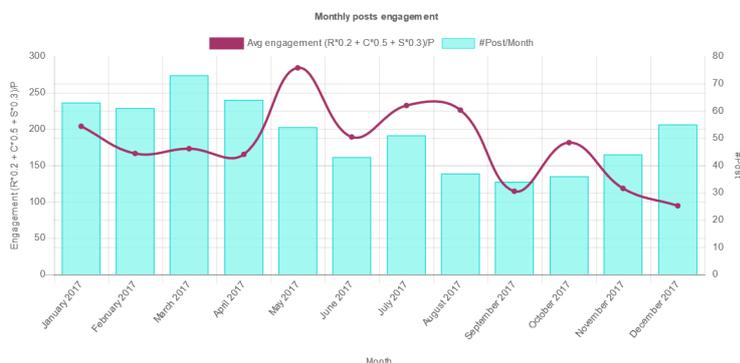


Figura 4.3: Distribuzione post su base mensile e relativo engagement medio per la regione Sicilia.

di agosto, una probabile causa di questa diminuzione di post è la fine della stagione estiva.

La figura 4.4 esamina ulteriormente questa relazione, infatti si può notare come le regioni del nord siano le più attive nei primi mesi dell'anno, in concomitanza con il turismo invernale, e quelle del sud siano quelle a postare di più in agosto, quando si ha più turismo balneare. Il calo di pubblicazioni

nei mesi autunnali è invece comune a tutte le zone geografiche.

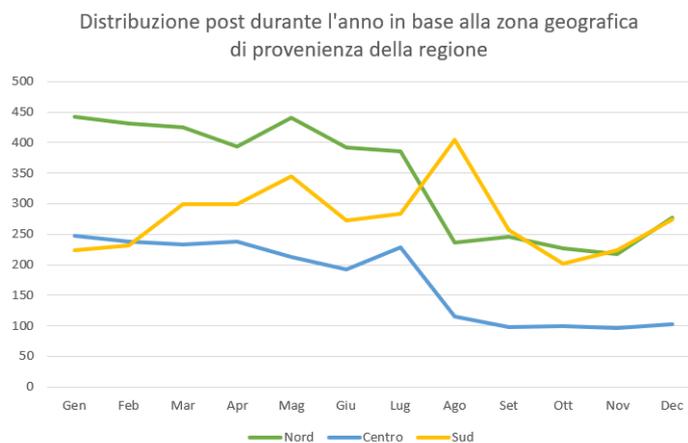


Figura 4.4: Distribuzione post durante l'anno in base alla zona geografica di provenienza della regione.

Anche l'engagement segue questo andamento, si può infatti notare dalla figura 4.5 come i post pubblicati in autunno siano quelli che hanno coinvolto meno gli utenti.

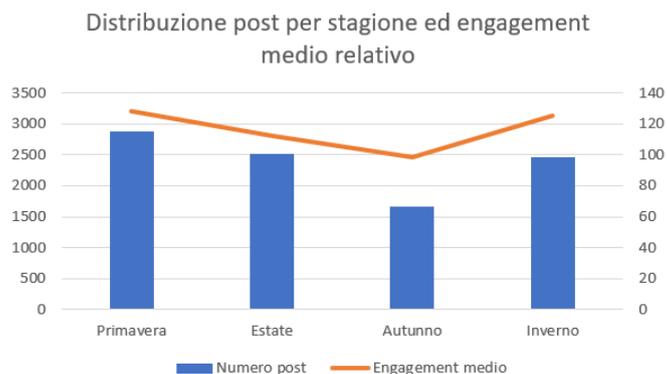


Figura 4.5: Numero di post per stagione e relativo engagement medio.

Le figure 4.6 e 4.7 rappresentano rispettivamente la distribuzione dei post in base ai giorni della settimana e quella in base al momento della giornata. Si può notare come siano stati pubblicati più post nei giorni infrasettimanali

e che si ha engagement maggiore il lunedì, il mercoledì e la domenica. Per quanto riguarda il momento della giornata sono stati pubblicati più post di mattina e pomeriggio, ma l'engagement maggiore si ha con i post pubblicati di mattina e sera.

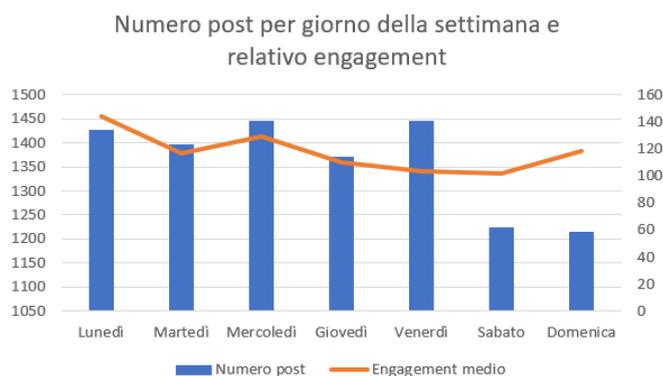


Figura 4.6: Distribuzione post in base al giorno della settimana e relativo engagement.

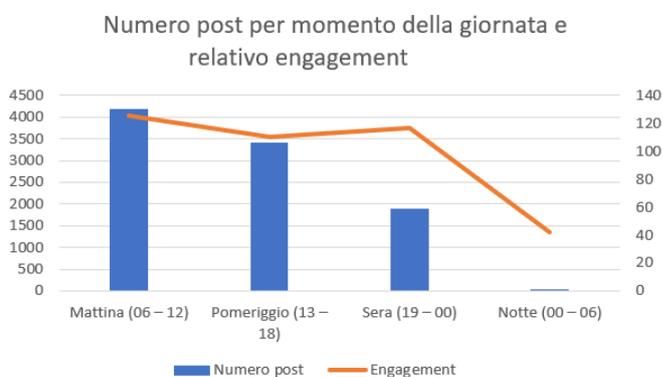


Figura 4.7: Distribuzione post in base al momento della giornata.

### Distribuzione dei post in base alla tipologia

Relativamente alla tipologia di post possiamo innanzitutto prendere in esame la regione Sardegna - nella figura 4.8 estratta dalla piattaforma - che

è l'unica, insieme alla regione Basilicata, ad aver utilizzato tutte le tipologie di post rilevate: foto, video, stato, link e nota. Possiamo notare come le foto siano di gran lunga la tipologia di post più utilizzata e come questa tendenza sia confermata anche a livello generale, come si può notare dalla figura 4.9. Video e link si classificano rispettivamente al secondo e terzo posto. È importante notare come l'Alto Adige pubblici, in proporzione, molti più video rispetto le altre regioni e come Toscana, Calabria e Friuli - Venezia Giulia siano quelle a utilizzare in maggior parte i link.

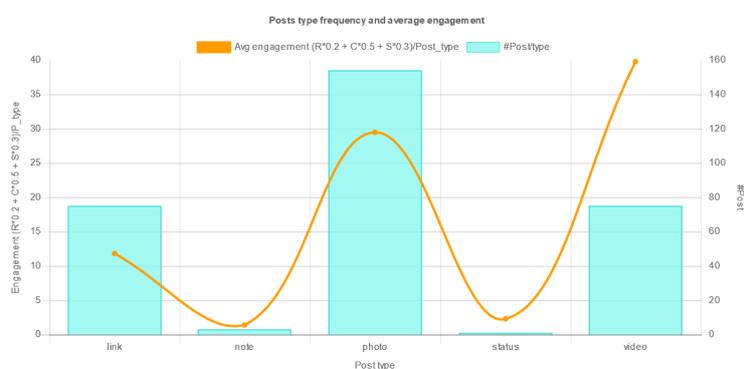


Figura 4.8: Numeri di post per tipologia della regione Sardegna e relativo engagement.

Lazio, Abruzzo, Veneto e Toscana sono le regioni che pubblicano più foto e le prime due, insieme a Sicilia e Lombardia, sono quelle che utilizzano il maggior numero di contenuti multimediali.

I contenuti multimediali sono anche la tipologia di post preferita dagli utenti - come si nota dalla figura 4.10 - le foto infatti si classificano al primo posto e i video al secondo se si considera l'engagement.

### Distribuzione dei post in base alla lunghezza

Esaminando il numero di caratteri dei post, cioè la loro lunghezza, si può notare come la tendenza principale sia la pubblicazione di post di bassa-media lunghezza, il cui messaggio è scritto in 100-400 caratteri. La figura 4.11 mostra la distribuzione del numero di caratteri per post a livello regionale,

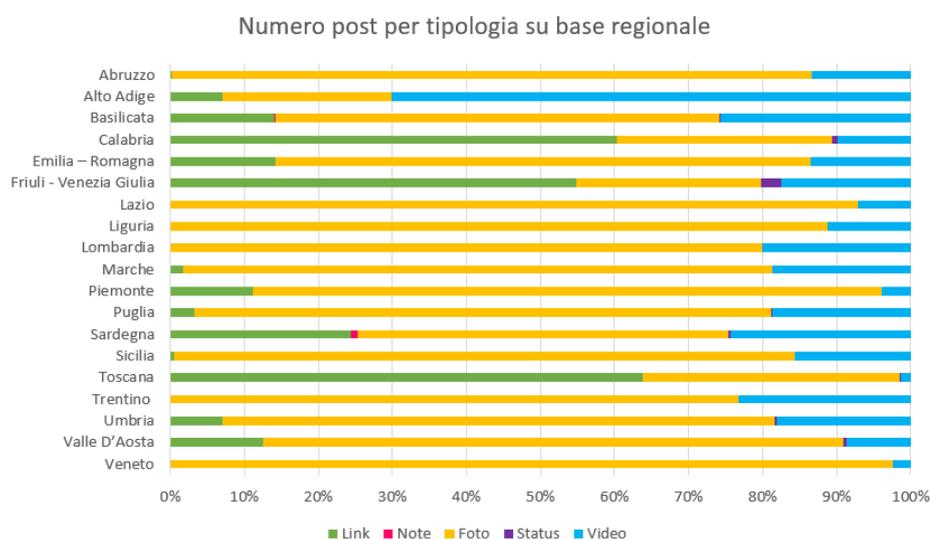


Figura 4.9: Numero di post in base alla tipologia a livello regionale.

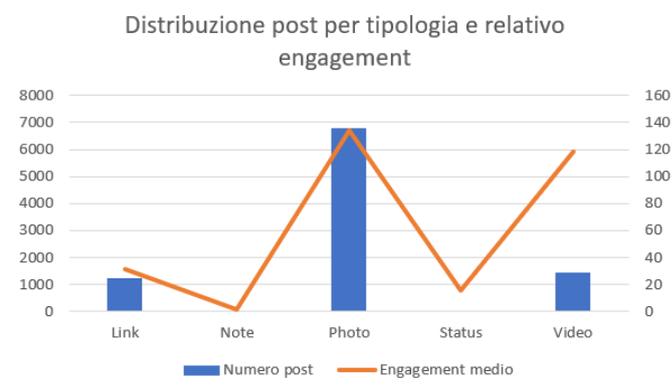


Figura 4.10: Numeri di post per tipologia e relativo engagement.

mentre la figura 4.12 mostra i dati relativi alla sola regione Sicilia tramite il grafico generato dalla piattaforma.

Osservando la figura 4.13 si può notare come l'engagement medio maggiore sia generato dai post molto brevi o di media lunghezza.

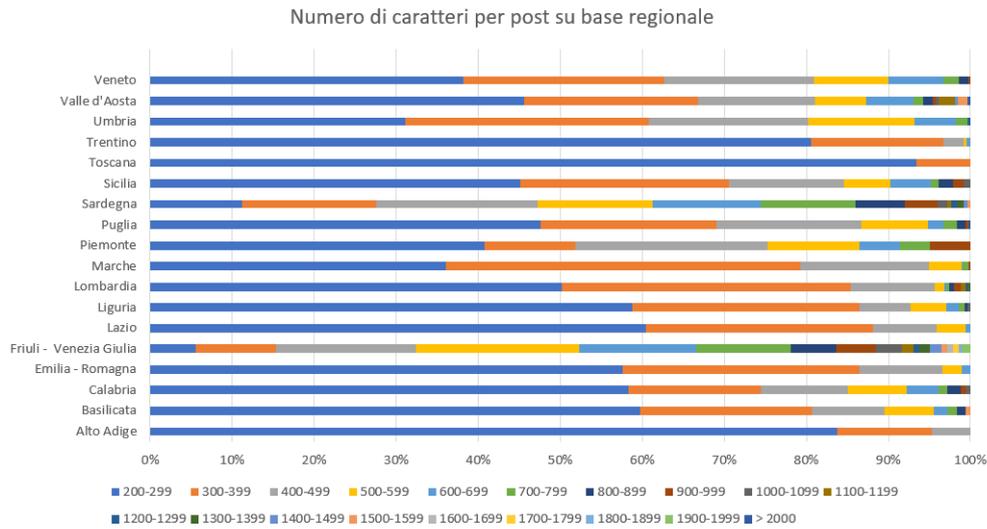


Figura 4.11: Numero di caratteri per post a livello regionale.

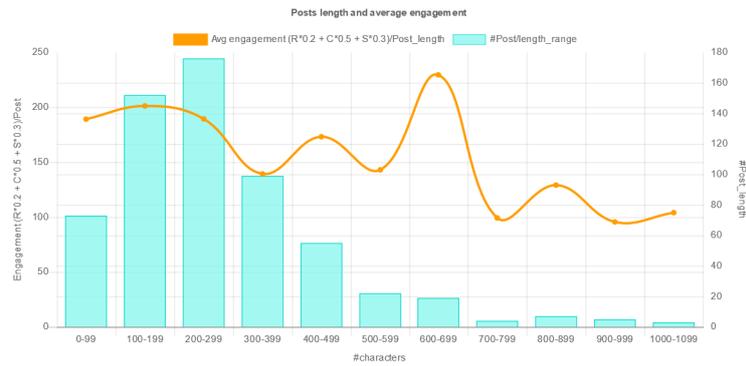


Figura 4.12: Numero di caratteri per post della regione Sicilia e relativo engagement.

### Analisi delle interazioni

In conclusione, è necessario analizzare il numero di reazioni, commenti e condivisioni ricevuti dai post. Innanzitutto, possiamo notare in figura 4.14 e tramite la relativa tabella 4.6 la distribuzione delle interazioni.

La regione Marche, seguita da Veneto e Lombardia, è la regione che ha ricevuto il più alto numero di interazioni. La regione infatti si classifica prima

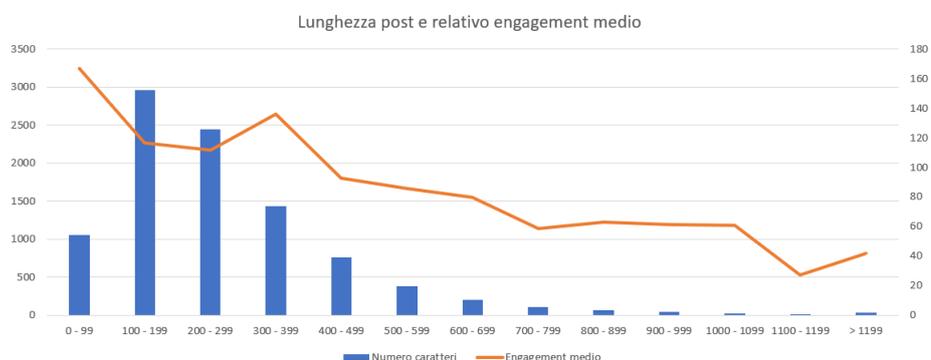


Figura 4.13: Numero di post in base alla lunghezza e relativo engagement.

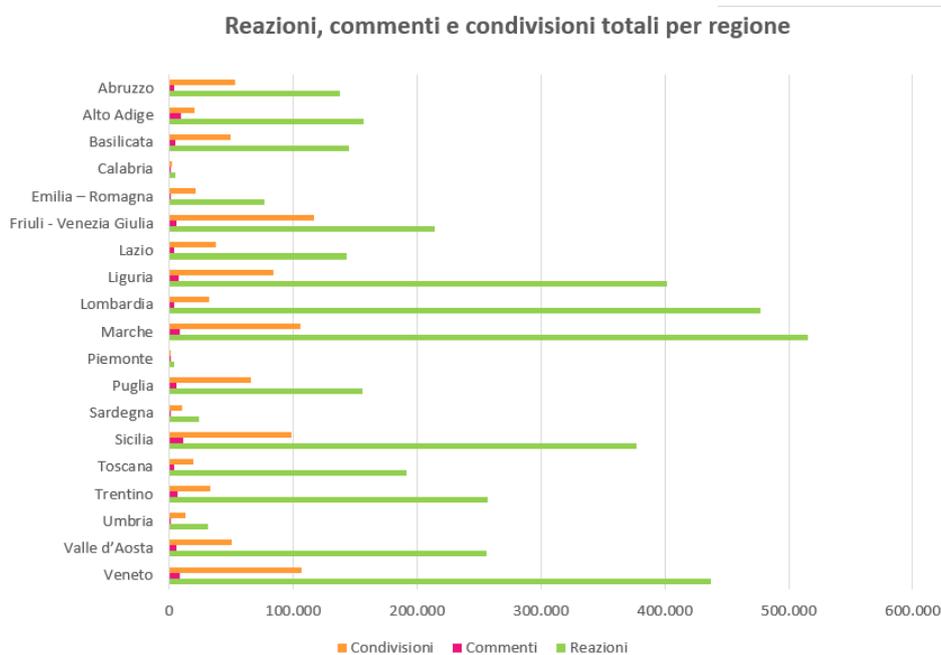


Figura 4.14: Numero totale di reazioni, commenti e condivisioni su base regionale.

per quanto concerne il numero di reazioni ricevute, e terza sia per il numero di commenti, che per il numero di condivisioni.

La Lombardia, seconda classificata per il numero di reazioni ricevute, è anche la regione ad aver pubblicato il post che ha ricevuto più reazioni nel

Regione	Numero reazioni	Numero commenti	Numero condivisioni
Abruzzo	137921	3584	52925
Alto Adige	157138	9871	20759
Basilicata	145248	4535	49439
Calabria	4841	157	1968
Emilia - Romagna	76946	1648	21364
Friuli - Venezia Giulia	214350	5953	116584
Lazio	143417	4079	37269
Liguria	401425	7592	83777
Lombardia	477079	4167	32201
Marche	515674	8659	106204
Piemonte	4202	85	634
Puglia	155649	5832	66346
Sardegna	24449	840	10346
Sicilia	376684	11661	98255
Toscana	191525	3623	19866
Trentino	256543	6514	33040
Umbria	31374	783	13284
Valle D'Aosta	256247	5662	50242
Veneto	437050	8126	106870

Tabella 4.6: Numero di reazioni, commenti e condivisioni per ogni regione.

2017 - figura 4.15 (a) .

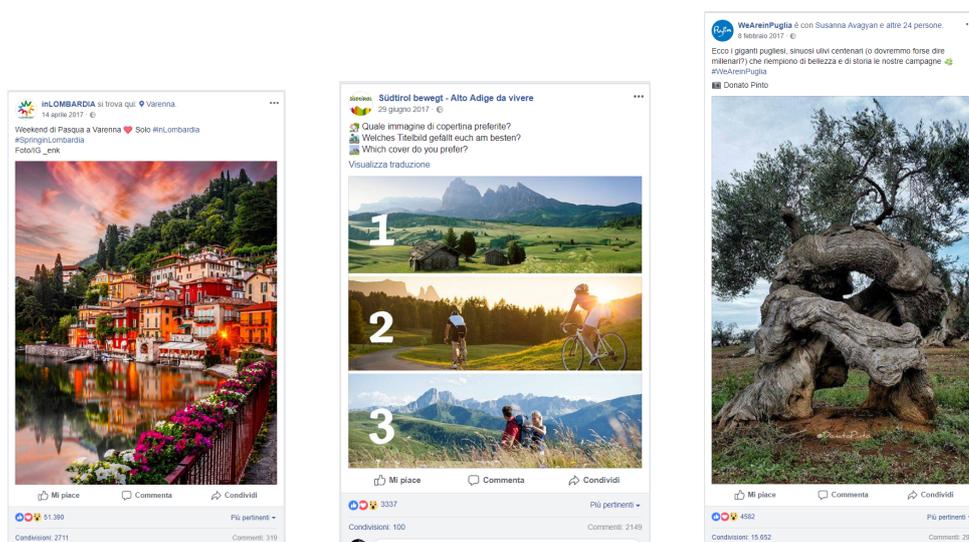
Per quanto riguarda i commenti, la Sicilia è la regione ad averne ricevuti di più seguita da Alto Adige, che possiede anche il primato per il post più commentato - figura 4.15 (b), e Marche.

I post più condivisi sono quelli del Friuli-Venezia Giulia, al secondo posto si classifica il Veneto e al terzo le Marche. Il post più condiviso, però è stato pubblicato dalla regione Puglia ed è visibile in figura 4.15 (c).

### 4.3.2 Analisi quantitativa dati di Twitter

Anche analizzando Twitter, come anticipato, le regioni Campania e Molise non sono presenti nell'analisi e Trentino e Alto Adige sono considerate come entità divise. In questo caso, inoltre, non è presente neanche la regione Calabria poiché il profilo di riferimento non è attivo da luglio 2017.

Per ogni regione sono stati estratti gli ultimi 500 tweet e risposte pubblicate dal profilo, ignorando i retweet ai tweet di altri utenti e ai propri tweet.



(a) Post che ha ricevuto il maggior numero di reazioni nel 2017.

(b) Post che ha ricevuto il maggior numero di commenti nel 2017.

(c) Post che è stato condiviso più volte nel 2017.

Figura 4.15: Migliori post del 2017 per reazioni, commenti e condivisioni.

Date le limitazioni poste da Twitter, infatti, non è stato possibile limitare i tweet estratti a un periodo di tempo definito; l'analisi comprende comunque i soli anni 2017 e 2018, il tweet più datato estratto è stato pubblicato dalla regione Piemonte il 30 gennaio 2017, mentre quelli più recenti sono del 10 maggio 2018, il giorno dell'estrazione.

Si sono estratti 9000 tweet totali e circa 231 mila interazioni divise in 165.936 'mi piace' ai tweet e 64.987 retweet. In particolare, è emerso che 363 tweet (4,03% del totale) non hanno ricevuto nessuna interazione, 477 tweet (5,3% del totale) non hanno nessun 'mi piace' e 1229 tweet (13,66% del totale) non sono mai stati condivisi. Le tabelle 4.7 e 4.8, rappresentano rispettivamente la distribuzione del numero di 'mi piace' e di retweet sui tweet.

La tabella 4.9 presenta una panoramica degli account regionali su Twitter. Possiamo innanzitutto notare come Marche, Emilia-Romagna, Trentino e

Numero 'mi piace'	Numero tweet	% su totale tweet
0 – 10	4346	48,29
11 – 50	3988	44,31
51 – 100	562	6,24
101 – 200	99	1,1
> 200	5	0,06

Tabella 4.7: Numero di 'mi piace' per tweet.

Numero retweet	Numero tweet	% su totale tweet
0 – 10	6942	77,13
11 – 50	2022	22,47
> 50	36	0.4

Tabella 4.8: Numero di retweet per tweet.

Puglia siano le regioni con il più alto numero di fan, contando ognuna più di 30.000 *followers*. Puglia, Marche e Umbria sono invece le regioni che "seguono" il maggior numero di profili.

Le regioni più attive sulla piattaforma sono Emilia-Romagna, Marche e Lombardia, il numero di tweet totali però non è totalmente rappresentativo poiché non tutte le regioni hanno creato il profilo sul social allo stesso momento.

Sicilia, Veneto, Abruzzo, Puglia e Liguria sono le regioni con l'engagement più alto - figura 4.16 - mentre se si considera l'engagement normalizzato - figura 4.17 - il Piemonte è il primo classificato, seguito da Sicilia, Valle d'Aosta, Abruzzo e Veneto. È utile sottolineare come le regioni che generano un coinvolgimento maggiore sono distribuite tra il nord e il sud.

### Distribuzione dei tweet nel tempo

In questo caso, al contrario di Facebook, non si sono analizzate le distribuzioni su base mensile poiché, non avendo un periodo di tempo delimitato,

Regione	Followers	Following	#Tweet	Engagement	Engagement normalizzato (%)
Abruzzo	15552	4880	15874	18,82	0,12
Alto Adige	13532	1678	12514	10,33	0,08
Basilicata	11283	1295	18427	7,26	0,06
Emilia - Romagna	68773	3663	66428	6,31	0,01
Friuli - Venezia Giulia	16979	1821	21148	11,02	0,06
Lazio	20883	1594	20943	15,26	0,07
Liguria	16869	2226	18930	17,29	0,10
Lombardia	22804	2775	25743	5,14	0,02
Marche	108643	9146	38279	9,97	0,01
Piemonte	725	478	892	3,50	0,48
Puglia	30292	10060	18536	17,45	0,06
Sardegna	16867	2310	11452	5,89	0,03
Sicilia	12226	608	12284	29,49	0,24
Toscana	13812	3217	4813	3,49	0,03
Trentino	31370	261	19071	9,57	0,03
Umbria	16759	5249	12388	11,61	0,07
Valle D'Aosta	1088	334	1513	2,36	0,22
Veneto	23659	825	18319	25,96	0,11

Tabella 4.9: Panoramica sui profili Twitter in analisi: numero *followers*, *following*, tweet totali pubblicati, engagement ed engagement normalizzato percentuale.

non sarebbero state attendibili. Si è comunque analizzata la distribuzione dei post in base al giorno della settimana - figura 4.18 - e in base al momento della giornata - figura 4.19.

Dal quadro emerge come siano pubblicati più tweet nei giorni infrasettimanali e nelle ore diurne. L'engagement però, come si può notare, è più alto nel fine settimana. Per quanto riguarda il momento della giornata, invece, i tweet pubblicati alla sera e alla mattina hanno un engagement più elevato.

### Distribuzione dei tweet in base alla lunghezza

I tweet, da novembre 2017 dispongono di 280 caratteri di lunghezza massima, prima il limite era posto a 140 caratteri. Nonostante i tweet siano

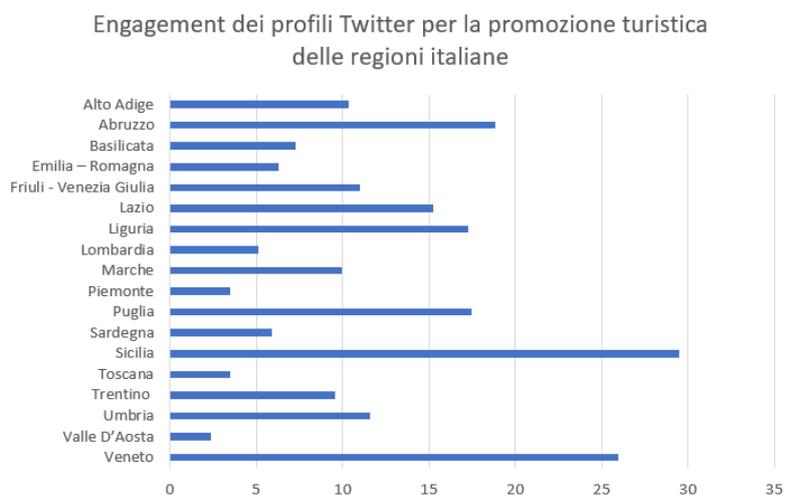


Figura 4.16: Engagement dei profili Twitter per la promozione turistica regionale.

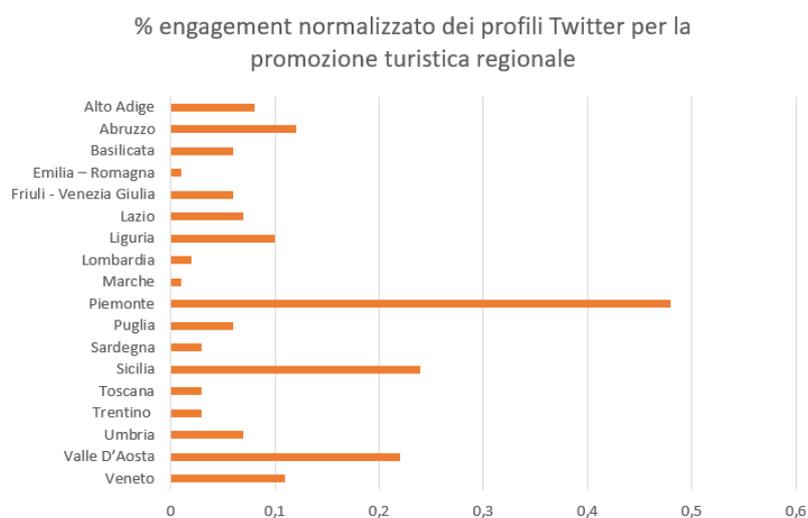


Figura 4.17: Engagement normalizzato dei profili per la promozione turistica regionale.

composti da pochi caratteri rispetto ai post di Facebook, si è analizzata comunque la loro lunghezza per capire quale fosse la lunghezza più frequente. La figura 4.20 mostra per ogni regione il numero di tweet in base al numero

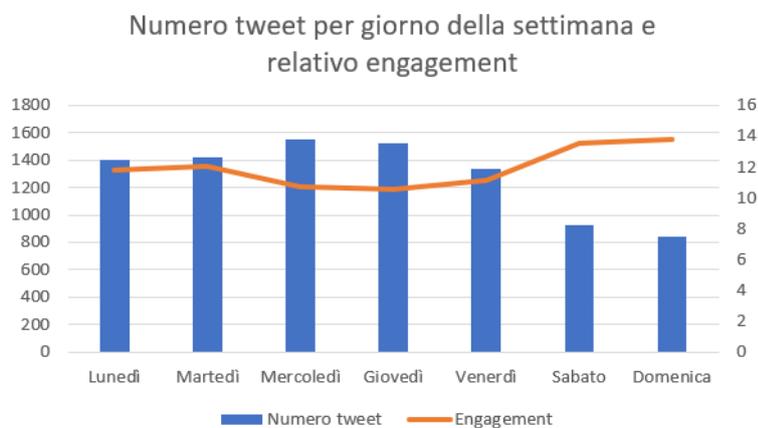


Figura 4.18: Distribuzione tweet in base al giorno della settimana.

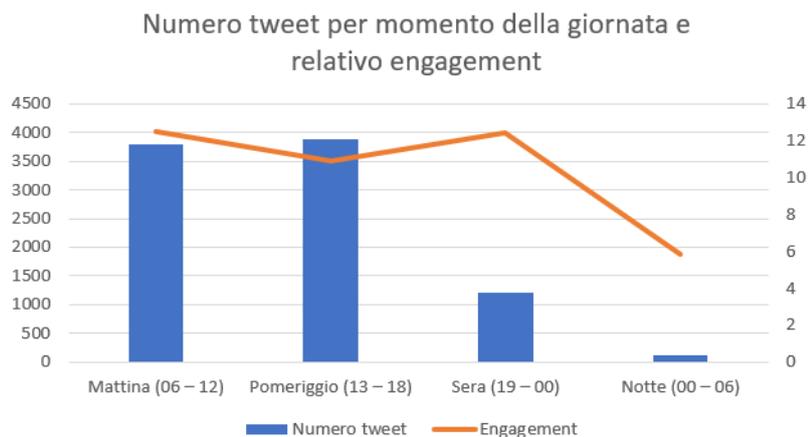


Figura 4.19: Distribuzione tweet in base al momento della giornata.

di caratteri e analizzando il grafico si evince come la maggior parte dei tweet abbia tra i 100 e i 150 caratteri. La figura 4.21 rappresenta la lunghezza dei tweet per la regione Sicilia, che con Piemonte e Alto Adige, è la sola ad avere utilizzato più di 150 caratteri.

L'engagement relativo alla lunghezza dei post - figura 4.22 - mostra come i tweet più apprezzati siano quelli più lunghi.

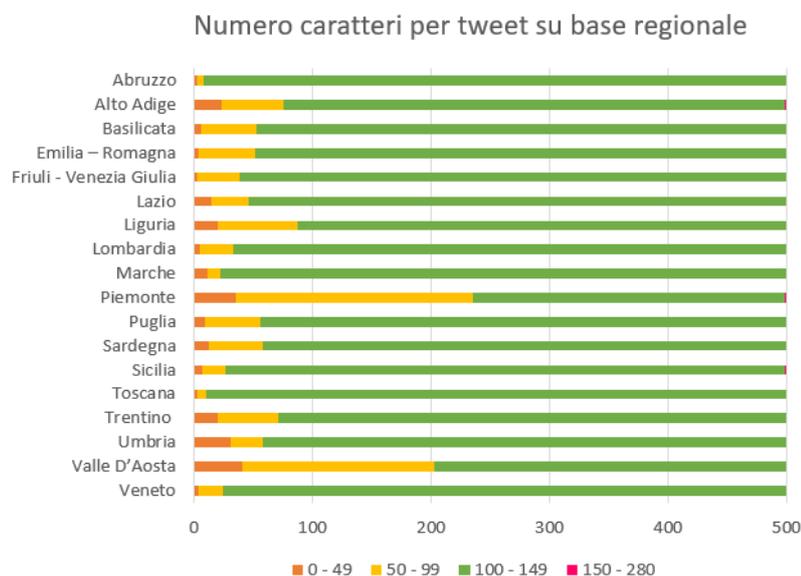


Figura 4.20: Numero di caratteri per tweet per ogni regione.

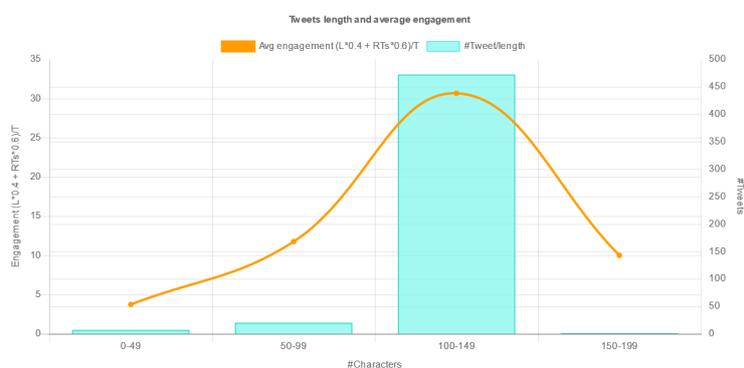


Figura 4.21: Numero di caratteri per tweet della regione Sicilia e relativo engagement.

### Analisi delle risposte

Si è effettuata, inoltre, un'analisi approfondita sulle sole risposte ad altri tweet pubblicate dai profili regionali. In totale si sono estratte 487 risposte (5,41% dei tweet estratti), delle quali 133 (27,31% delle risposte) non hanno ricevuto interazioni, 138 (28,34% delle risposte) non hanno ricevuto nessun

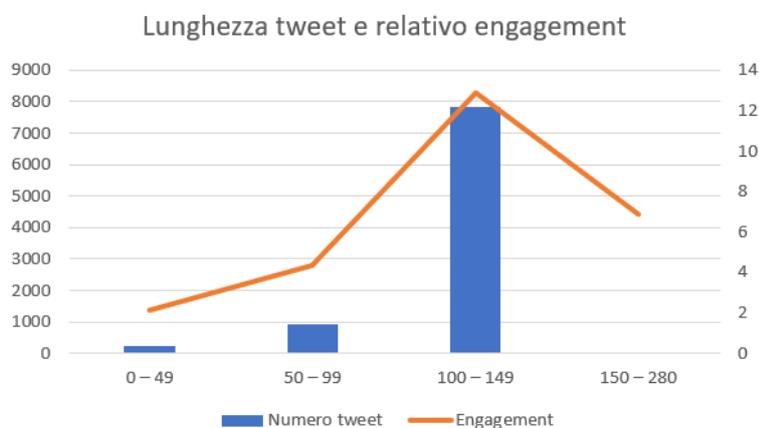


Figura 4.22: Lunghezza tweet e relativo engagement.

‘mi piace’ e 374 (76,80% delle risposte) non sono state condivise tramite retweet. Si può notare dalla figura 4.23 che la Sardegna è la regione più attiva nel rispondere, con 108 risposte sui 500 tweet estratti. Le figure 4.24 e 4.25 - estratte dalla piattaforma sviluppata - rappresentano rispettivamente la distribuzione giornaliera dei tweet estratti senza e considerando le risposte per la regione Sardegna.

### Analisi delle interazioni

Per quanto riguarda le interazioni con i tweet - figura 4.26 e tabella 4.10 - si può notare come la Sicilia sia la regione i cui tweet hanno ricevuto più interazioni totali. La Sicilia è infatti la regione ad aver ricevuto più ‘mi piace’ e anche quella ad essere stata più condivisa, al secondo e terzo posto, in entrambe le classifiche si piazzano rispettivamente Veneto e Abruzzo.

Il tweet ad avere ricevuto il maggior numero di ‘mi piace’ è stato pubblicato dalla Sicilia ed è visibile in figura 4.27 (a) con un totale di 274 likes. Ad essere stati condivisi il maggior numero di volte sono invece due tweet, uno è stato pubblicato ancora una volta all’account della regione Sicilia ed è in figura 4.27 (b) , l’altro invece è stato pubblicato dalla regione Sardegna, figura 4.27 (c).

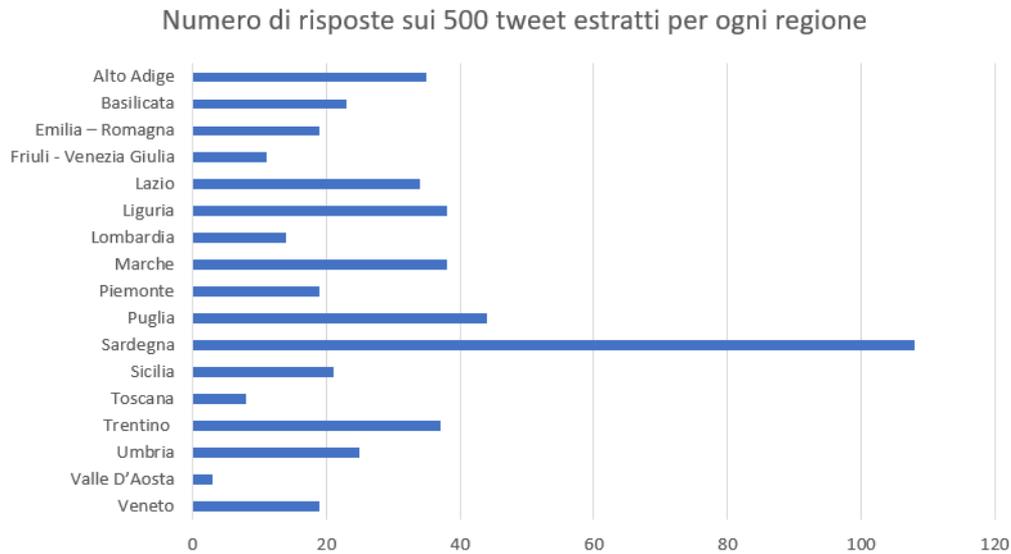


Figura 4.23: Numero di risposte sui 500 tweet estratti per ogni regione.

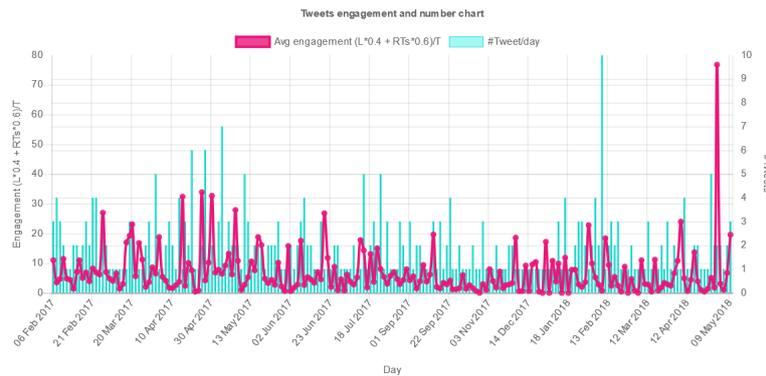


Figura 4.24: Distribuzione giornaliera dei soli tweet della regione Sardegnna.

### Analisi hashtag

In conclusione, è necessario analizzare gli hashtag. Nei tweet estratti sono stati utilizzati 3828 hashtag diversi per un totale di 6707 hashtag utilizzati. Per ogni regione - tabella 4.11 - si sono considerati gli hashtag più utilizzati. È importante notare come quasi tutte le regioni utilizzino con più frequenza hashtag in cui è presente il nome della stessa o l'account di riferimento sui social.

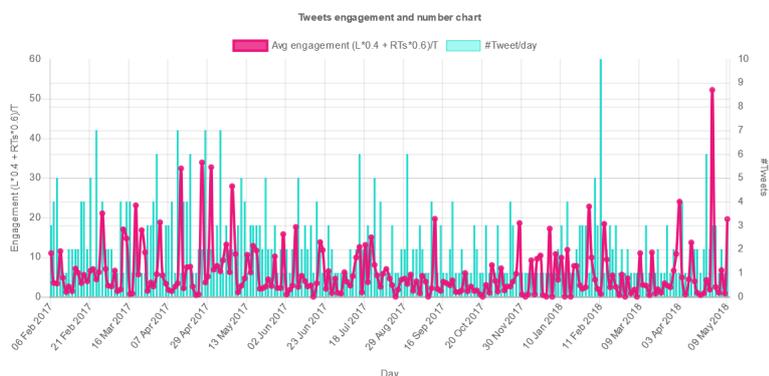


Figura 4.25: Distribuzione giornaliera di tweet e risposte della regione Sardegna.

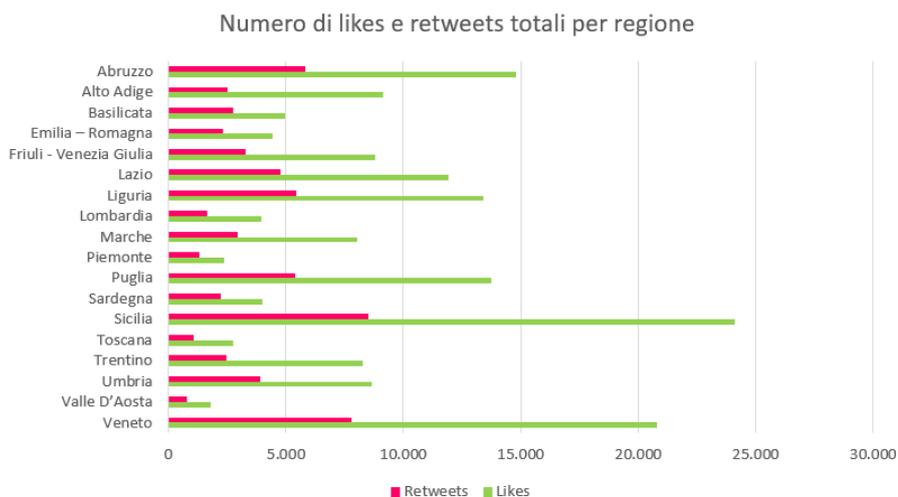


Figura 4.26: Numero di likes e retweet totali ai tweet su base regionale.

Un'altra analisi condotta sugli hashtag è relativa agli hashtag per ogni regione che hanno realizzato un engagement più alto, visibili in tabella 4.12.

In questo caso è importante notare come siano i nomi di località turistiche e ricorrenze particolari a generare più engagement.

Regione	Numero 'mi piace'	Numero retweet
Abruzzo	14812	5808
Alto Adige	9117	2533
Basilicata	4949	2754
Emilia – Romagna	4430	2302
Friuli - Venezia Giulia	8821	3300
Lazio	11940	4758
Liguria	13416	5466
Lombardia	3973	1638
Marche	8039	2946
Piemonte	2375	1334
Puglia	13720	5396
Sardegna	4003	2242
Sicilia	24089	8515
Toscana	2748	1074
Trentino	8262	2468
Umbria	8671	3898
Valle D'Aosta	1773	785
Veneto	20798	7770

Tabella 4.10: Numero di 'mi piace' e retweet totali per ogni regione.

## 4.4 Analisi della regressione sui dati

Per approfondire l'analisi sui dati estratti si è deciso di svolgere una regressione lineare multipla sui dati. La regressione serve ad analizzare la relazione tra una variabile  $Y$ , chiamata dipendente, e una o più variabili  $X$ , chiamate indipendenti. La regressione può essere utilizzata per effettuare previsioni sulla variabile dipendente o per analizzare l'influenza delle variabili indipendenti su quella dipendente.

Il modello di regressione lineare multipla assume la formula (4.6), dove



(a) Tweet che ha ricevuto il maggior numero di likes nel 2017.

(b) Tweet della Sicilia che ha ricevuto il maggior numero di retweet nel 2017.

(c) Tweet della Sardegna che ha ricevuto il maggior numero di retweet nel 2017.

Figura 4.27: Migliori tweet del 2017 per numero di *likes* e retweet

$\beta_0$  è l'intercetta, cioè l'effetto medio nel caso in cui tutte le variabili fossero uguali a 0,  $X_1, X_2, X_n$  sono le variabili indipendenti,  $\beta_1, \beta_2, \dots, \beta_n$  sono i coefficienti di regressione delle variabili indipendenti,  $\epsilon$  è la componente di errore del modello e  $Y$  è la variabile dipendente.

### Regressione lineare multipla

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (4.6)$$

In questa tesi si è voluta analizzare l'influenza di alcuni fattori, come l'ora di pubblicazione o la posizione geografica, sull'engagement di un post o di un tweet pubblicati dai profili turistici ufficiali delle regioni italiane. Si sono quindi innanzitutto analizzati i dati per decidere quali fattori considerare, si è poi preparato un file CSV contenente tutti i dati necessari all'analisi e infine si è utilizzato il software SAS [52] per effettuare la regressione.

Per preparare il file CSV si sono utilizzate delle variabili *dummy*, variabili binarie che indicano la presenza (1) o l'assenza (0) dell'attributo in ciascuna delle categorie considerate.

SAS è un insieme di prodotti software integrati utili all'analisi dei dati. Per l'analisi si è utilizzato SAS University Edition, gratuito per la ricerca

Regione	Hashtag	Numero utilizzi
Abruzzo	Abruzzo	100
Alto Adige	Winter	57
Basilicata	Basilicata	324
Emilia – Romagna	inEmiliaRomagna	160
Friuli - Venezia Giulia	Fvglive	122
Lazio	NelCuoredItalia	31
Liguria	orgogliopesto	71
Lombardia	inLombardia	351
Marche	Destinazionemarche	141
Piemonte	Visitpiemonte	302
Puglia	WeAreinPuglia	156
Sardegna	Sardegna	105
Sicilia	Visitsicilyinfo	129
Toscana	ToscanaOvunqueBella	79
Trentino	Trentino	60
Umbria	BuongiornoUmbria	45
Valle D'Aosta	Valledaostadavivere	357
Veneto	VisitVeneto	71

Tabella 4.11: Hashtag più utilizzato da ogni regione.

accademica, che permette di accedere ai dati, manipolarli, analizzarli per mezzo di tecniche statistiche e presentarne i risultati tramite report.

#### 4.4.1 Regressione sui dati di Facebook

Per effettuare la regressione sui dati relativi alle pagine Facebook dei profili turistici ufficiali delle regioni italiane si sono considerate sei variabili indipendenti: la tipologia di post, la lunghezza del post, il momento della giornata, il giorno della settimana e la stagione in cui il post è stato pubblicato e la zona geografica di provenienza della regione. Alcuni dati sono

Regione	Hashtag	Engagement
Abruzzo	abruzzonelcuore	62,40
Alto Adige	TreCime	74,00
Basilicata	GiornataMondialedelTurismo	49,60
Emilia – Romagna	CapitaleItalianaDellaCultura	37,79
Friuli - Venezia Giulia	Freccetricolori   live	59,60
Lazio	NataleDiRoma	74,10
Liguria	Nervi	71,80
Lombardia	MilanoDesignWeek   PalazzoLombardia	26,20
Marche	Gradara	31,33
Piemonte	Musicainquota	31,40
Puglia	GiornataMondialedellAmbiente	60,80
Sardegna	equinozio	59,60
Sicilia	CapitaleItalianadellaCultura2018PA	120,40
Toscana	BandiereBlu	30,32
Trentino	NewYear   happynewyear	51,00
Umbria	Buonanno	35,60
Valle D'Aosta	chiccopellegrino	27,40
Veneto	CappelladegliScrovegni	91,40

Tabella 4.12: Hashtag con engagement più alto per ogni regione.

stati divisi in categorie per poter considerare un numero limitato di opzioni alternative, ad esempio la lunghezza di un post avrebbe utilizzato più di 10 opzioni diverse, tramite la classificazione se ne utilizzano solo tre. La lunghezza del post, infatti, è stata suddivisa in:

- *length\_short*: se il post ha tra gli 0 e i 250 caratteri;
- *length\_medium*: se il post ha tra i 251 e i 500 caratteri;
- *length\_long*: se il post ha più di 500 caratteri.

Il giorno della settimana è stato suddiviso tra:

- *weekday*: se il post è stato pubblicato in un giorno infrasettimanale, cioè lunedì, martedì, mercoledì, giovedì o venerdì;

- *weekend*: se il post è stato pubblicato nel fine settimana, cioè sabato o domenica.

La tipologia di post è rimasta invariata; vengono considerati foto, video, note, link e stati in modo indipendente. Così come le stagioni: *season\_spring*, *season\_summer*, *season\_fall* e *season\_winter* e i momenti della giornata: *time\_morning*, *time\_afternoon*, *time\_evening* e *time\_night*.

La zona geografica di provenienza della regione è stata suddivisa in:

- *geography\_north*: se il post è stato pubblicato da Valle d'Aosta, Piemonte, Lombardia, Trentino, Alto Adige, Veneto, Friuli - Venezia Giulia, Liguria o Emilia - Romagna;
- *geography\_center*: se il post è stato pubblicato da Toscana, Marche, Lazio o Umbria;
- *geography\_south*: se il post è stato pubblicato da Abruzzo, Puglia, Calabria, Basilicata, Sicilia o Sardegna.

Osservando l'analisi quantitativa - già affrontata nel paragrafo 4.3.1 - si sono formulate le seguenti ipotesi:

- H1.** La tipologia di post influisce sull'engagement, in particolare foto e video hanno un impatto positiva su di esso.
- H2.** La lunghezza del post influisce sull'engagement, in particolare post più corti hanno un effetto positivo su di esso.
- H3.** Il momento della giornata in cui il post è stato pubblicato influisce sull'engagement, in particolare pubblicare post alla mattina ha un impatto positivo su di esso.
- H4.** Il giorno di pubblicazione influisce sull'engagement, in particolare i post infrasettimanali hanno un effetto positivo su di esso.
- H5.** La stagione in cui il post è stato pubblicato influisce sull'engagement, in particolare pubblicare post in primavera ha un effetto positivo su di esso.

**H6.** La zona di provenienza (nord, centro o sud) della regione influisce sull'engagement, in particolare le regioni del nord hanno un impatto positivo su di esso.

Per verificare le ipotesi si è effettuata una regressione lineare sui dati utilizzando il codice in figura 4.28.

```
proc reg data=work.facebook_regression_data PLOTS(MAXPOINTS=NONE);

model engagement = time_night time_morning time_afternoon time_evening
                    weekday weekend
                    season_winter season_spring season_summer season_fall
                    geography_north geography_center geography_south
                    length_short length_medium length_long
                    type_note type_photo type_status type_video type_link;

run;
```

Figura 4.28: Codice per la regressione sui dati di Facebook.

Il modello proposto, però, risulta non determinante poiché il coefficiente di determinazione  $R^2$  è del 2,95% - figura 4.29 - e quindi i regressori non predicono bene il valore della variabile dipendente.

Per considerare le ipotesi proposte, è necessario specificare che una variabile è accettabile se  $p\text{-value} < 0,05$ , dove il  $p\text{-value}$  è il livello di significatività assegnato e più piccolo è, tanto più significativa è l'ipotesi.

La figura 4.30 indica le stime dei parametri per ogni variabile mostrando il coefficiente beta e il livello di significatività. Si può notare come nella tabella siano presenti valori uguali a 0, le variabili sono state lasciate nel modello appositamente per mostrare tutte le alternative possibili e sono calcolate tramite una combinazione lineare delle altre categorie della stessa tipologia.

Considerando le ipotesi iniziali si può affermare che:

**H1.** L'ipotesi H1 non può essere rifiutata poiché *type\_photo* e *type\_video* hanno entrambe un  $p\text{-value}$  estremamente significativo (pari al 0,001).

**H2.** L'ipotesi H2 non può essere rifiutata poiché, anche in questo caso il  $p\text{-value}$  di *length\_short* è estremamente significativo.

Num. di osservazioni lette	9526
Num. di osservazioni usate	9526

Analisi della varianza					
Origine	DF	Somma dei quadrati	Media quadratica	Valore F	Pr > F
Modello	15	22139684	1475979	19.30	<.0001
Errore	9510	727229522	76470		
Totale corretto	9525	749369206			

Radice MSE	276.53206	R-quadro	0.0295
Media dip.	118.06101	R-quadro corr	0.0280
Coeff var	234.22810		

Figura 4.29: Analisi della regressione sui dati relativi a Facebook.

- H3.** L'ipotesi H3 non può essere rifiutata poiché il p-value di *time\_morning* è estremamente significativo.
- H4.** L'ipotesi H4 non può essere rifiutata, il *p-value* di *weekday* è significativo, ma valendo 0,04 lo è meno rispetto i valori precedenti.
- H5.** L'ipotesi H5 è rifiutata, poiché il *p-value* di tutte le variabili è molto maggiore di 0,05.
- H6.** L'ipotesi H6 non può essere rifiutata poiché, il p-value di *geography\_north* è estremamente significativo.

Concludendo, quindi dall'analisi emerge che gli utenti preferiscono contenuti visivi come foto e video con un testo breve, il dato è sicuramente giustificato dal fatto che i contenuti multimediali permettono alle regioni di interagire direttamente con i propri fan, un esempio concreto è quello del post che ha ricevuto più commenti nel 2017 - in figura 4.15 (b) - la regione

Stime dei parametri					
Variabile	DF	Stima dei parametri	Errore standard	Valore t	Pr >  t
Intercept	B	-71.27783	15.90288	-4.48	<.0001
time_night	B	-12.63640	46.16556	-0.27	0.7843
time_morning	B	31.69182	7.83364	4.05	<.0001
time_afternoon	B	14.99717	8.09714	1.85	0.0640
time_evening	0	0	.	.	.
weekday	B	13.59130	6.52384	2.08	0.0372
weekend	0	0	.	.	.
season_winter	B	12.28569	8.89175	1.38	0.1671
season_spring	B	11.82488	8.63608	1.37	0.1710
season_summer	B	5.04304	8.78385	0.57	0.5659
season_fall	0	0	.	.	.
geography_north	B	62.98808	6.57596	9.58	<.0001
geography_center	B	30.34686	7.81321	3.88	0.0001
geography_south	0	0	.	.	.
length_short	B	45.46282	10.24675	4.44	<.0001
length_medium	B	29.60375	10.75445	2.75	0.0059
length_long	0	0	.	.	.
type_note	B	-10.69661	138.61655	-0.08	0.9385
type_photo	B	96.52886	8.78314	10.99	<.0001
type_status	B	-38.83871	60.91590	-0.64	0.5238
type_video	B	86.65776	10.81609	8.01	<.0001
type_link	0	0	.	.	.

Figura 4.30: Stima dei parametri relativi a Facebook.

Alto Adige, infatti ha fatto scegliere ai propri fan l'immagine di copertina e questi hanno risposto in gran numero. Gli utenti sono inoltre più attivi la mattina e durante la settimana lavorativa, per questo dato si può pensare che, dato anche il diffuso utilizzo degli smartphone, questi si colleghino durante la pausa pranzo. Infine, l'engagement aumenta se la regione che ha pubblicato il post è del nord e la stagione in cui il post è pubblicato non influisce sull'engagement.

#### 4.4.2 Regressione sui dati di Twitter

La regressione sui dati relativa ai profili Twitter per la promozione turistica delle regioni italiane è stata effettuata considerando cinque variabili indipendenti: la lunghezza del tweet, il momento della giornata e il giorno della settimana in cui è stato pubblicato il tweet, la zona geografica di provenienza della regione che ha pubblicato il tweet e se il tweet è una risposta. Per quanto riguarda la divisione in categorie è stata rispettata la divisione già utilizzata per Facebook, bisogna sottolineare però come in questo caso non si consideri nè la stagione, nè la tipologia di tweet e che per quanto riguarda le regioni del sud la Calabria non sia presente. È utile anche definire come sia stata suddivisa la lunghezza del tweet:

- *length\_short*: se il post ha tra gli 0 e i 50 caratteri;
- *length\_medium*: se il post ha tra i 51 e i 140 caratteri;
- *length\_long*: se il post ha più di 140 caratteri.

Twitter considera invece la variabile relativa alla risposta: *reply\_yes* rappresenta i tweet che sono risposte, mentre *reply\_no* rappresenta quelli che non lo sono.

Come nel caso precedente si è analizzata l'analisi quantitativa dei dati e si sono formulate le seguenti ipotesi:

- H1.** La lunghezza del tweet influisce sull'engagement, in particolare tweet più lunghi hanno un effetto positivo su di esso.

- H2.** Il momento della giornata in cui il tweet è stato pubblicato influisce sull'engagement, in particolare pubblicare tweet alla mattina e alla sera ha un impatto positivo su di esso.
- H3.** Il giorno di pubblicazione influisce sull'engagement, in particolare pubblicare tweet nel fine settimana ha un effetto positivo su di esso.
- H4.** La zona di provenienza (nord, centro o sud) della regione influisce sull'engagement, in particolare le regioni del sud hanno un impatto positivo su di esso.
- H5.** Le risposte influiscono sull'engagement e in particolare hanno un impatto negativo sull'engagement.

Per verificare le ipotesi si è effettuata una regressione lineare multipla utilizzando il codice in figura 4.31.

```
proc reg data=work.twitter_regression_data PLOTS(MAXPOINTS=NONE);  
  
model engagement = time_night time_morning time_evening time_afternoon  
                    weekend weekday  
                    geography_north geography_center geography_south  
                    length_short length_medium length_long  
                    reply_yes reply_no;  
  
run;
```

Figura 4.31: Codice per la regressione sui dati di Twitter.

Anche in questo caso, nonostante sia comunque migliore del precedente, il modello proposto non è determinante poiché il coefficiente di determinazione  $R^2$  è del 9,26% - figura 4.32 - e i regressori non predicano in modo ottimale il valore della variabile dipendente.

Come nel caso di Facebook, si sono valutate le ipotesi in base al valore del *p-value*. La figura 4.33 indica le stime dei parametri per ogni variabile, anche in questo caso sono presenti i valori uguali a 0, per mostrare tutte le alternative possibili.

Considerando le ipotesi iniziali quindi, si può affermare che:

Num. di osservazioni lette		9000			
Num. di osservazioni usate		9000			
Analisi della varianza					
Origine	DF	Somma dei quadrati	Media quadratica	Valore F	Pr > F
Modello	9	154351	17150	101.90	<.0001
Errore	8990	1513068	168.30566		
Totale corretto	8999	1667419			

Radice MSE	12.97327	R-quadro	0.0926
Media dip.	11.70740	R-quadro corr	0.0917
Coeff var	110.81254		

Figura 4.32: Analisi della regressione sui dati relativi a Twitter.

- H1.** L'ipotesi H1 non può essere rifiutata poiché entrambi i  $p$ -value di *length\_short* e *length\_medium* sono estremamente significativi. Il loro coefficiente negativo implica che pubblicare tweet lunghi sia ottimale.
- H2.** L'ipotesi H2 non può essere rifiutata, il  $p$ -value di *time\_morning* è estremamente significativo, mentre quello di *time\_evening* è significativo.
- H3.** L'ipotesi H3 non può essere rifiutata, il  $p$ -value di *weekend* è estremamente significativo.
- H4.** L'ipotesi H4 non può essere rifiutata, il  $p$ -value di *geography\_north* e *geography\_center* è estremamente significativo. Come nel caso precedente, il loro coefficiente negativo implica che le regioni del sud aumentino l'engagement.
- H5.** L'ipotesi H5 non può essere rifiutata e il  $p$ -value di *reply\_yes* è estremamente significativo.

Stime dei parametri					
Variabile	DF	Stima dei parametri	Errore standard	Valore t	Pr >  t
Intercept	B	18.17835	0.53392	34.05	<.0001
time_night	B	-4.38871	1.26578	-3.47	0.0005
time_morning	B	1.82868	0.29658	6.17	<.0001
time_evening	B	1.37437	0.43120	3.19	0.0014
time_afternoon	0	0	.	.	.
weekend	B	2.37075	0.34662	6.84	<.0001
weekday	0	0	.	.	.
geography_north	B	-5.86528	0.32613	-17.98	<.0001
geography_center	B	-5.74123	0.39568	-14.51	<.0001
geography_south	0	0	.	.	.
length_short	B	-7.35083	0.98455	-7.47	<.0001
length_medium	B	-3.18781	0.50111	-6.36	<.0001
length_long	0	0	.	.	.
reply_yes	B	-10.90599	0.64343	-16.95	<.0001
reply_no	0	0	.	.	.

Figura 4.33: Stima dei parametri relativi a Twitter.

Dall'analisi quindi emerge che gli utenti preferiscono i tweet lunghi, in questo contesto è utile considerare che un tweet lungo sia comunque di soli 280 caratteri massimi; sono più attivi alla mattina e alla sera e nei fine settimana ed inoltre non sono interessati alle risposte. L'engagement aumenta se la regione che pubblica il tweet è del sud d'Italia.

## 4.5 Confronto analisi tra Facebook e Twitter

L'ultimo obiettivo di questo elaborato è effettuare un confronto tra l'utilizzo dei due social media in esame, Facebook e Twitter. In questo paragrafo, quindi, si confronteranno i risultati ottenuti dall'analisi dei profili ufficiali per la promozione turistica delle regioni italiane, mettendo in evidenza elementi comuni e differenze nell'utilizzo dei due social media da parte delle regioni e nella risposta dei fan sui rispettivi social. Ovviamente, gli elementi non comuni tra i due social media: gli hashtag, la tipologia di post, le risposte e le stagioni in cui sono pubblicati i post, non sono considerati in questo confronto.

Innanzitutto, si possono confrontare il numero di fan per gli account di promozione turistica regionale sui due social - visibili alla tabella 4.5 per Facebook e alla tabella 4.9 per Twitter - si può notare come il Trentino sia - in entrambi i social - una delle regioni con più fan, su Facebook è anticipato da Alto Adige e Sicilia, mentre su Twitter da Marche ed Emilia-Romagna. La Lombardia, invece, risulta essere la terza le regioni con più messaggi pubblicati in entrambi i social, su Facebook è anticipata da Sicilia e Puglia, mentre su Twitter da Emilia - Romagna e Marche. La Valle d'Aosta, infine, è la terza regione migliore per engagement normalizzato in entrambi i social media.

Per quanto concerne i giorni della settimana - figure 4.6 e 4.18 - il mercoledì è il giorno in cui gli account sono più attivi per entrambi i social media, però si può notare che mentre per Facebook non c'è molta differenza di engagement rispetto al giorno di pubblicazione, per Twitter il fine settimana ha un impatto positivo rispetto ai giorni infrasettimanali.

In relazione al momento della giornata, mattina e pomeriggio sono per entrambi i social media i momenti di maggiore pubblicazione - figure 4.7 e 4.19 - e in entrambi i casi la mattina e sera sono il migliore momento della giornata per pubblicare.

Se si considera la lunghezza dei post - figure 4.13 e 4.22 - si può, invece, notare come la *length\_short* preferita dagli utenti su Facebook, coincida con

la lunghezza dei tweet, con particolare riferimento al numero di caratteri più utilizzato: tra i 100 e i 200.

Infine, per quanto riguarda le regioni che hanno ricevuto più interazioni - con riferimento alla tabella 4.6 per Facebook e 4.10 per Twitter - possiamo notare come Veneto e Sicilia siano particolarmente amate in entrambi i social media. In particolare il Veneto è la terza regione per numero di reazioni e la seconda per numero di condivisioni su Facebook e la seconda sia per numero di 'mi piace' che per numero di retweet su Twitter, invece la Sicilia è prima pagina per numero di commenti su Facebook e il primo profilo per numero di 'mi piace' e retweet su Twitter.

# Conclusioni

Questa tesi ha trattato l'implementazione di una piattaforma per l'analisi della performance delle imprese sui social media Facebook e Twitter, fornendo anche un esempio applicativo in ambito turistico, tramite il confronto degli account ufficiali per la promozione turistica delle regioni italiane.

Si è infatti implementata una piattaforma software che permette alle imprese e ai singoli individui di analizzare la performance delle pagine Facebook e dei profili Twitter. Il tool è stato realizzato con un design elementare in modo che chiunque possa utilizzarlo con semplicità; una volta selezionato il social e inserito l'username dell'account da analizzare, infatti, la piattaforma si occupa di scaricare i dati, elaborarli e mostrarli nel modo più chiaro possibile, fornendo brevi descrizioni di ciò che si sta visualizzando. Poiché la piattaforma permette di salvare i dati estratti è possibile anche ampliare le proprie analisi esternamente al sistema. È importante sottolineare che, com'è emerso dall'analisi di altre piattaforme esistenti, non si hanno online dei sistemi che forniscano questi servizi gratuitamente. La piattaforma, infine, permette il confronto tra pagine e profili differenti permettendo un confronto diretto con i propri concorrenti.

Oltre all'implementazione della piattaforma, si è svolta un'analisi - per la quale si è utilizzato il tool creato - dei profili turistici promozionali delle regioni italiane. Per quello che concerne Facebook, dall'analisi è emerso come la pubblicazione di post da parte delle DMO non sia costante per tutto l'anno; nei mesi autunnali, infatti, cala in numero di post pubblicati, il momento è compreso tra la fine delle vacanze estive e l'inizio di quelle invernali, quindi

nel momento in cui generalmente non ci sono persone in viaggio. Un ulteriore dato emerso è relativo al momento della giornata più utilizzato da parte delle DMO per pubblicare i propri post: la mattina, e come questo sia anche il momento della giornata in cui i fan sono più attivi; il dato è importante poiché in contrasto con lo studio precedentemente effettuato nel 2013, in cui la pubblicazione mattutina aveva riscontri negativi sull'engagement, la spiegazione per questo fenomeno potrebbe essere attribuita agli smartphone che, essendosi diffusi sempre di più, hanno permesso agli individui di collegarsi ai social media in qualsiasi ora e luogo, senza dover aspettare di utilizzare un computer. Questa analisi contrasta con lo studio del 2013 anche nel caso della lunghezza dei post, poiché in quella ricerca post più lunghi avevano un impatto migliore sull'engagement rispetto quelli corti, mentre da questa analisi è emerso come quelli più corti, oltre ad essere più frequenti, hanno anche un effetto migliore sull'engagement rispetto i lunghi. L'ultimo dato con particolare rilevanza rispetto al social network Facebook è la tipologia di post utilizzata; dall'analisi è infatti emerso come i post con elementi visivi come foto e video, siano sia quelli più utilizzati, sia quelli che creano un maggior coinvolgimento con gli utenti, e come l'utilizzo di sondaggi e quiz sia effettivamente positivo per la pagina.

Per quanto riguarda Twitter, invece, è emerso che la regione Sicilia è sicuramente quella con l'account più coinvolgente per gli utenti, infatti risulta prima in quasi tutte le classifiche riguardanti le interazioni con i tweet e l'engagement. Un secondo dato rilevante è che nonostante le DMO pubblichino più tweet nei giorni lavorativi, gli utenti sono più attivi e coinvolti durante il fine settimana. Infine, è importante sottolineare due dati riguardanti gli hashtag, innanzitutto gli hashtag più utilizzati da ogni regione contengono in 15 casi su 18 il nome della regione o l'username del profilo di riferimento e il secondo dato è che gli hashtag che generano più engagement da parte degli utenti sono quelli relativi a località turistiche e ricorrenze speciali.

I primi limiti di questa tesi sono nella piattaforma software sviluppata, in particolare i dati estratti da Twitter, dato che l' API del social media

non consente di estrarre dati su una base temporale predefinita non è stato possibile, come nel caso di Facebook, analizzare l'andamento annuale delle pubblicazioni; un secondo limite riguarda l'estrazione delle risposte, infatti non è possibile estrarle insieme al tweet, ma sarebbe necessario effettuare una ricerca *ad hoc* per individuare inizialmente tutte le risposte degli utenti e poi associarle al tweet a cui si riferiscono.

Per i dati estratti da Facebook, invece, il limite maggiore è quello dell'estrazione circa di 600 tweet riferiti a un solo anno di una determinata pagina Facebook, fortunatamente, però, la maggior parte delle DMO ha pubblicato meno di 600 post nell'anno 2017.

Per quanto riguarda la piattaforma in sé, uno dei limiti maggiori è il non poter più aggiornare i dati una volta estratti, poiché una tale opzione permetterebbe all'utente di aumentare i dati utili all'analisi.

L'esempio di analisi in ambito turistico, invece, trova il suo limite più grande nella circoscrizione dell'analisi alle sole DMO regionali italiane e al solo anno 2017 e ai primi cinque mesi del 2018 per Twitter. Ampliando il bacino delle DMO regionali, considerando ad esempio anche quelle degli altri stati europei, si potrebbero confrontare tra di loro non solo il turismo tra i diversi stati, ma anche l'utilizzo dei social media da parte dei cittadini di stati diversi.

Gli sviluppi futuri, sia per la piattaforma che per l'analisi, sono innumerevoli. Per quanto riguarda l'analisi si potrebbero aumentare le tipologie di dati analizzati, introducendo ad esempio analisi semantiche, o aumentare gli anni considerati per lo studio. Per la piattaforma, invece, si potrebbero innanzitutto aumentare i social media analizzati, aggiungendo ad esempio Instagram e Youtube, e aumentare il numero di grafici proposti per l'analisi aggiungendo, ad esempio, i dati relativi al momento migliore della giornata per pubblicare i post. Infine, un ulteriore modo per estendere l'analisi è quello di chiedere all'utente l'accesso al social media per accedere a più dati e permettere un'analisi più accurata.



# Bibliografia e Sitografia

- [1] Agorapulse. Agorapulse: Simple & affordable social media management. <https://www.agorapulse.com/>.
- [2] Syed Zamberi Ahmad, Norita Ahmad, and Abdul Rahim Abu Bakar. Reflections of entrepreneurs of small and medium-sized enterprises concerning the adoption of social media and its impact on performance outcomes: Evidence from the uae. *Telematics and Informatics*, 2017.
- [3] Audiense. Twitter analytics | audiense. <https://audiense.com/>.
- [4] Bootstrap. Bootstrap · the most popular html, css, and js library in the world. <https://getbootstrap.com/>.
- [5] Buffer. Buffer: Social media management platform. <https://buffer.com/>.
- [6] Chart.js. Chart.js | open source html5 charts for your website. <https://www.chartjs.org/>.
- [7] Rival IQ Corporate. Rival iq: Competitive social media analytics for digital marketers. <https://www.rivaliq.com/>.
- [8] Vincenzo Cosenza. Social media in italia: analisi dei flussi di utilizzo del 2017. <http://vincos.it/2018/04/20/social-media-in-italia-analisi-dei-flussi-di-utilizzo-del-2017/>.

- [9] Carmen Cox, Stephen Burgess, Carmine Sellitto, and Jeremy Buultjens. The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, 18(8):743–764, 2009.
- [10] Cyfe. Cyfe: All-in-one business dashboard. <https://www.cyfe.com/>.
- [11] Andrea De Mauro, Marco Greco, and Michele Grimaldi. A formal definition of big data based on its essential features. *Library Review*, 65(3):122–135, 2016.
- [12] Peter Sheridan Dodds, Roby Muhamad, and Duncan J Watts. An experimental study of search in global social networks. *science*, 301(5634):827–829, 2003.
- [13] Josep Ejarque. *Social Media Marketing per il turismo: Come costruire il marketing 2.0 e gestire la reputazione della destinazione*. Hoepli Editore, 2015.
- [14] Eurostat. Digital economy & society in the EU. <http://ec.europa.eu/eurostat/cache/infographs/ict/index.html>.
- [15] Eurostat. Social media - statistics on the use by enterprises. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Social\\_media\\_-\\_statistics\\_on\\_the\\_use\\_by\\_enterprises](http://ec.europa.eu/eurostat/statistics-explained/index.php/Social_media_-_statistics_on_the_use_by_enterprises).
- [16] Facebook. Company info - facebook newsroom. <https://newsroom.fb.com/company-info/>.
- [17] Facebook. Facebook for developers. <https://developers.facebook.com/>.
- [18] Flask. Welcome | flask (a python microframework). <http://flask.pocoo.org/>.
- [19] John Furrier. Official definition of facebook's social graph – dave morin from facebook – graphing social patterns

- day 2. <https://podtech.wordpress.com/2007/10/09/official-definition-of-facebooks-social-graph-dave-morin-from-facebook-graphing-social-patterns-day-2/>.
- [20] Mircea Georgescu and Daniela Popescu. Social media—the new paradigm of collaboration and communication for business environment. *Procedia Economics and Finance*, 20:277–282, 2015.
- [21] Eun Go and Kyung Han You. But not all social media are the same: Analyzing organizations’ social media usage patterns. *Telematics and Informatics*, 33(1):176–186, 2016.
- [22] Tamara Grizane and Inguna Jurgelane. Social media impact on business evaluation. *Procedia Computer Science*, 104:190–196, 2017.
- [23] Clyde W Holsapple, Shih-Hui Hsiao, and Ram Pakath. Business social media analytics: Characterization and conceptual framework. *Decision Support Systems*, 110:32–45, 2018.
- [24] ECMA International. *Standard ECMA-262 ECMAScript 2018 Language Specification*. ECMA International, 2018.
- [25] Istat. Viaggi e vacanze in italia e all’estero. <https://www.istat.it/it/archivio/208998>.
- [26] JavaScript. Javascript.com. <https://www.javascript.com/>.
- [27] json. Json. <https://www.json.org/>.
- [28] Rohayah Kahar, Faizal Yamimi, Ghazali Bunari, and Hadina Habil. Trusting the social media in small business. *Procedia-Social and Behavioral Sciences*, 66:564–570, 2012.
- [29] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

- [30] Keyhole. Hashtag tracking for twitter, instagram and facebook - keyhole. <https://keyhole.co/>.
- [31] Klear. Klear: Influencer marketing software. <https://klear.com>.
- [32] Doug Laney. Application delivery strategies. *META Group, Stamford*, 2001.
- [33] In Lee. Social media analytics for enterprises: Typology, methods, and processes. *Business Horizons*, 2017.
- [34] Likealyzer. Meltwater - likealyzer. <https://likealyzer.com/>.
- [35] G.S. Linoff and M.J.A. Berry. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Books 24x7 IT PRO. Wiley, 2011.
- [36] A. Maheshwari. *Big Data*. 2017.
- [37] Marcello M Mariani, Marco Di Felice, and Matteo Mura. Facebook as a destination marketing tool: Evidence from italian regional destination management organizations. *Tourism Management*, 54:321–343, 2016.
- [38] Marcello M Mariani, Matteo Mura, and Marco Di Felice. The determinants of facebook social engagement for national tourism organizations' facebook pages: A quantitative approach. *Journal of Destination Marketing & Management*, 2017.
- [39] Roberto Marmo. *Social media mining: estrarre e analizzare informazioni dai social media*. HOEPLI EDITORE, 2016.
- [40] Margaret McCann and Alexis Barlow. Use and measurement of social media for smes. *Journal of Small Business and Enterprise Development*, 22(2):273–287, 2015.
- [41] Nikolaos Misirlis and Maro Vlachopoulou. Social media metrics and analytics in marketing—s3m: A mapping literature review. *International Journal of Information Management*, 38(1):270–276, 2018.

- 
- [42] Sebastian Molinillo, Francisco Liébana-Cabanillas, Rafael Anaya-Sánchez, and Dimitrios Buhalis. Dmo online platforms: Image and intention to visit. *Tourism Management*, 65:116–130, 2018.
- [43] Moz. Followerwonk: Tools for twitter analytics, bio search and more. <https://moz.com/followerwonk>.
- [44] MySQL. Mysql. <https://www.mysql.com>.
- [45] Hossein Nezakati, Asra Amidi, Yusmadi Yah Jusoh, Shayesteh Moghadas, Yuhanis Abdul Aziz, and Roghayeh Sohrabinezhadtalemi. Review of social media potential on knowledge sharing and collaboration in tourism industry. *Procedia-social and behavioral sciences*, 172:120–125, 2015.
- [46] OAuth. Oauth community site. <https://oauth.net/>.
- [47] World Tourism Organization. *A practical guide to tourism destination management*. World Tourism Organization, 2007.
- [48] Kay Peters, Yubo Chen, Andreas M Kaplan, Björn Ognibeni, and Koen Pauwels. Social media metrics—a framework and guidelines for managing social media. *Journal of interactive marketing*, 27(4):281–298, 2013.
- [49] Python. Welcome to python.org. <https://www.python.org/>.
- [50] Quintly. Social media analytics & competitor benchmarking | quintly. <https://www.quintly.com/>.
- [51] A. Rezzani. *Big Data: Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*. PerCorsi di studio. Apogeo Education, 2013.
- [52] SAS. Sas: Analytics, business intelligence and data management. <https://www.sas.com/>.

- [53] Yakov Shafranovich. Common format and mime type for comma-separated values (csv) files. 2005.
- [54] Sprout Social. Sprout social: Social media management software. <https://sproutsocial.com/>.
- [55] Socialbakers. Socialbakers: Ai-powered social media & digital marketing solution. <https://www.socialbakers.com/>.
- [56] SocialRank. Socialrank. <https://socialrank.com/>.
- [57] SumAll. Sumall: Automated social media. <https://sumall.com/>.
- [58] Georgios Tsimonis and Sergios Dimitriadis. Brand strategies in social media. *Marketing Intelligence & Planning*, 32(3):328–344, 2014.
- [59] TweetReach. Tweetreach: How far did your tweets travel? <https://tweetreach.com/>.
- [60] Twitonomy. Twitonomy: Twitter #analytics and much more... <https://www.twitonomy.com/>.
- [61] Twitter. Twitter developers. <https://developers.twitter.com/>.
- [62] Ahmet Uşaklı, Burcu Koç, and Sevil Sönmez. How'social'are destinations? examining european dmo social media usage. *Journal of destination marketing & management*, 6(2):136–149, 2017.
- [63] WSGI. Wsgi - wsgi.org. <http://wsgi.readthedocs.io/en/latest/>.
- [64] Benxiang Zeng and Rolf Gerritsen. What do we know about social media in tourism? a review. *Tourism Management Perspectives*, 10:27–36, 2014.
- [65] Mingli Zhang, Lingyun Guo, Mu Hu, and Wenhua Liu. Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation. *International Journal of Information Management*, 37(3):229–240, 2017.