

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Informatica per il management

**Facebook per promozione turistica:
Sviluppo di una piattaforma software
per l'analisi delle recensioni**

Tesi di Laurea in Basi di Dati

Relatore:
Chiar.mo Prof.
MARCO DI FELICE

Presentata da:
VALENTINA DUCA

Correlatore:
Chiar.mo Prof.
MARCELLO MARIANI

Sessione I
Anno Accademico 2016/2017

Indice

Elenco delle figure	4
Introduzione	7
1 Data Mining, Big-Data, DBMS e NoSQL	11
1.1 Data Mining	11
1.1.1 Cos'è il Data Mining	12
1.1.2 Cosa può fare il Data Mining	13
1.2 Big-Data	15
1.2.1 Cos'è il Big Data	15
1.2.2 Peculiarità principali	16
1.2.3 Utilizzo e opportunità dei Big-Data	18
1.2.4 Criticità e rischi derivanti dai Big-Data	19
1.3 DBMS	20
1.3.1 Peculiarità principali	22
1.4 Database NoSQL	26
1.4.1 Definizione di Database NoSQL	26
1.4.2 Peculiarità principali dei database NoSQL	26
1.4.3 Modelli principali di NoSQL	30
2 Social Media e Turismo	32
2.1 I social media	32
2.1.1 Facebook	33
2.1.2 L'effetto "Mi Piace"	34

2.2	Il turista 2.0	36
2.2.1	Il potere delle raccomandazioni	38
2.2.2	Nuovi fattori di scelta	39
2.3	L'engagement	40
2.3.1	Strategie: l'utilizzo di testimonial	41
2.4	Le DMO	42
2.4.1	Un nuovo scenario competitivo	42
2.4.2	Definizione di DMO	43
2.4.3	I ruoli delle DMO	44
2.5	Studi correlati	45
3	Progettazione	55
3.1	Specifiche del progetto	55
3.2	Architettura del progetto	56
3.3	Facebook APIs Documentation	58
3.3.1	API Graph	58
3.3.2	Facebook for developer tool	60
4	Implementazione	61
4.1	Tecnologie utilizzate	61
4.1.1	Python	61
4.1.2	HTML	63
4.1.3	JavaScript	63
4.1.4	PHP	65
4.1.5	SQL	66
4.2	Estrazione dati: dettagli implementativi	69
4.2.1	Chiamate API Graph	69
4.2.2	Il campo <i>next</i>	69
4.2.3	Filtraggio post	71
4.2.4	Interfaccia grafica	74
4.3	Visualizzazione dati: dettagli implementativi	75
4.3.1	La Home	76
4.3.2	Le Regioni	76

5	Analisi dei Dati	78
5.1	Reperire i dati	78
5.2	Analisi dati annuale	80
5.3	Utenti attivi e metriche di engagement	87
5.3.1	Engagement medio	91
5.3.2	Best Practice del 2016	93
5.4	Grafici per tipo di post	95
	Conclusioni	96
	Bibliografia	99
	Ringraziamenti	102

Elenco delle figure

1.1	Modello delle 3V big data	17
1.2	Architettura a livelli DBMS	21
1.3	Teorema del CAP	30
1.4	Modelli NoSQL	31
2.1	Età utenti italiani	34
2.2	Comunicazione digitale	36
2.3	L'evoluzione dei modelli di turismo	38
2.4	Fattori di scelta	39
2.5	Campagna pubblicitaria mediante testimonial	42
2.6	Ruoli fondamentali DMO	44
2.7	Utilizzo di strumenti di comunicazione di marketing su internet	47
2.8	Numero complessivo dei post delle DMOs italiane nel 2013	49
2.9	Fonti di pianificazioni di viaggio	50
2.10	Conoscenza dei brand	51
2.11	Motivi per cambiare programmi fedeltà con altro	51
2.12	Utilizzo dello smarthphone nei vari processi	52
2.13	Utilizzo dello smarthphone nei vari processi	53
2.14	Utilizzo dello smarthphone nei vari processi	53
3.1	Interfaccia grafica visualizzazione dati	57
3.2	Interfaccia grafica estrazione dati	58
3.3	Esempio di chiamata API Graph	59
3.4	Esempio di token d'accesso e richiesta API Graph	60

3.5	Esempio di richiesta API Graph	60
4.1	Procedura PHP	66
4.2	Esempio di chiamata Ajax	67
4.3	Utilizzo del linguaggio SQL	68
4.4	Esempio di query	68
4.5	Esempio di query 2	68
4.6	JSON restituito dalle richieste API Graph	70
4.7	Controllo del campo next	70
4.8	Algoritmo estrazione parte 1	72
4.9	Algoritmo estrazione parte 2	73
4.10	Algoritmo estrazione parte 3	74
4.11	Realizzazione grafico	75
4.12	Codice salvataggio grafico formato .png	77
4.13	Scelta range data	77
5.1	Tabella 1 - Link fan page e numero fan	80
5.2	Tabella 2 - Post totali anno 2016	81
5.3	Tabella 3 - Post e numero di "mi piace"	81
5.4	Tabella 4 - Post e numero di commenti	82
5.5	Tabella 5 - Post e numero di condivisioni	82
5.6	Tabella 6 - Post totale vs. Post delle DMOs	83
5.7	Tabella 7 - Parametri totale regionali annuali	84
5.8	Post totali per ogni mese dell'anno 2016	85
5.9	Post totali per ogni mese	85
5.10	Totale parametri annuali regionali	86
5.11	Totale parametri mensili	86
5.12	Formula engagement	87
5.13	Formula engagement normalizzato	87
5.14	Algoritmo calcolo metriche	88
5.15	Algoritmo calcolo metriche	88
5.16	Engagement annuale Veneto	89
5.17	Engagement normalizzato annuale Veneto	89
5.18	Confronto engagement	90
5.19	Confronto engagement normalizzato	90

5.20 Engagement medio mensile	91
5.21 Engagement medio annuale di ogni regione	91
5.22 Engagement normalizzato medio mensile	92
5.23 Engagement normalizzato medio annuale di ogni regione . .	92
5.24 Post con più commenti	93
5.25 Best practice 2016	94
5.26 Post con più commenti	95

Introduzione

Il numero elevato di utenti iscritti alle piattaforme social, come ad esempio Facebook¹, Twitter² o Instagram, ha intensificato il fenomeno dell'interazione sociale. Si noti, infatti, come, oltre che le relazioni tra turista e offerta, la tecnologia abbia trasformato quelle tra turista e turista: il Web 2.0³ e i social media hanno reso possibile nuovi livelli di interazione grazie alla possibilità di interagire in qualsiasi momento e luogo, anche durante il viaggio stesso. Il fatto che i turisti sfruttino le tecnologie nelle varie fasi della pratica turistica deriva dal fatto che il turista oggi è un turista esperto, in quanto ha accumulato esperienza nel corso degli anni e dei viaggi, e informato, in quanto la quantità di informazione, sia attraverso il passaparola, sia sui mezzi di comunicazione generalisti e specializzati, è ampia e in continua crescita. Il fatto che la ricerca delle informazioni sia stata profondamente trasformata dai social media può, in un certo senso, non sorprendere. Prima di prendere una decisione di acquisto un individuo assume informazioni attraverso la rete di amici, conoscenti e parenti,

¹www.facebook.com

²www.twitter.com

³Web 2.0 Il termine, apparso nel 2005, indica genericamente la seconda fase di sviluppo e diffusione di Internet, caratterizzata da un forte incremento dell'interazione tra sito e utente: maggiore partecipazione dei fruitori, che spesso diventano anche autori (blog, chat, forum, wiki); più efficiente condivisione delle informazioni, che possono essere più facilmente recuperate e scambiate con strumenti peer to peer o con sistemi di diffusione di contenuti multimediali come Youtube; affermazione dei social network. Nuovi linguaggi di programmazione consentono un rapido e costante aggiornamento dei siti web anche per chi non possiede una preparazione tecnica specifica. Il fenomeno è ancora in fortissima evoluzione.

compera riviste specializzate e cataloghi. Internet è dunque un altro strumento che si aggiunge a quelli tradizionalmente utilizzati e amplia le possibilità di informazione degli individui.

Nasce quindi l'esigenza di definire una nuova tipologia di dati dovuta dall'enorme volume e molteplicità dell'informazione diffusa in rete attraverso i social media. Da qui la nuova definizione di dati, i cosiddetti *big data*. Essi indicano i dati prodotti in grande quantità, con notevole rapidità nei formati più diversi, la cui elaborazione richiede tecnologie e risorse che vanno ben al di là dei sistemi convenzionali di gestione e immagazzinamento dei dati.

Tali tecnologie non hanno influenzato solo il comportamento dei turisti, ma anche quello dell'offerta. Il progresso della tecnologia permette di individuare e soddisfare i bisogni del turista. Facebook, Twitter e altre piattaforme, possono essere utilizzate dalle strutture turistiche (hotel, musei, linee aeree) per monitorare ed intercettare eventuali problemi di qualità nel servizio da loro offerto e risolvere tempestivamente con le modifiche necessarie. La tecnologia inserita in vari modi nell'ambiente della destinazione turistica, può arricchire l'esperienza turistica e migliorare così la competitività della destinazione.

A differenza del passato, si è rivelata molto importante la fase del post-arrivo e quella del ritorno a casa. In queste fasi si costruisce la reputazione della destinazione e la fidelizzazione dei turisti. Il processo di decisione del viaggio non termina più dopo aver stabilito data e durata, ma persiste anche durante e dopo il viaggio. Per queste motivazioni è importante capire che per una destinazione turistica essere social non vuol dire possedere una pagina Facebook, bensì saper utilizzare a proprio favore questi strumenti, non solo a livello informatico ma anche per dialogare e co-creare con il turista.

Come cita Ejarque⁴:

«Gli obiettivi di una destinazione sono di suscitare interesse, convincere e trasformare l'interessi in decisione, fidelizzare, far parlare di sè, rendere il turista il proprio promotore [...]»

Uno degli obiettivi della mia tesi verte sull'analisi dati delle pagine fan ufficiali per la promozione turistica regionale a livello italiano. Analizza ed esamina le varie reazioni ai post, essi possono essere foto, video, link, o altro, da parte degli utenti che seguono tali pagine che hanno come scopo quello di promuovere la propria regione come meta turistica. La piattaforma a cui si fa riferimento in questo caso specifico è Facebook. In particolare si vogliono analizzare i periodo durante l'arco di un anno in cui c'è più risposta da parte dell'utente, e quali tipi di post suscitano più reazioni. Inoltre, c'è la possibilità di confrontare il coinvolgimento annuale tra due regioni differenti.

Altro obiettivo della tesi è quello di sviluppare un tool che permetta di estrarre i dati necessari per delle analisi. Quindi inizialmente la tesi fornirà le informazioni utili al conseguimento di tale scopo. Una volta descritto il tool per l'estrazione dei dati, si analizzeranno le metodologie adottate per l'analisi degli stessi, e si presenteranno i risultati.

Prima di presentare l'analisi è stato stilato un capitolo che parla dei big data, riportando la definizione, l'utilizzo, i vantaggi e gli svantaggi. Sono poi stati descritti i DBMS, partendo da una descrizione generale per andare poi più sullo specifico, trattando anche le varie limitazioni. Il capitolo si conclude con la descrizione di uno dei più importanti DMBS, MySQL.

Il secondo capitolo descrive lo stato dell'arte, dove viene affrontato il discorso social media in relazione al turismo. Vengono infatti discussi vari articoli che riportano degli studi effettuati sul medesimo argomento, presi

⁴Ejarque, 2012a:4

da riviste come *Tourism Management*, una delle riviste leader a livello internazionale. Sono stati riportati i risultati ottenuti da questi studi e poi confrontati con i risultati di questo elaborato.

La progettazione viene discussa nel terzo capitolo. Vengono descritte con precisione le specifiche del progetto, la struttura del software da implementare e infine si parlerà anche delle varie APIs Facebook utilizzate utili al conseguimento dello scopo. Il quarto capitolo riguarda l'implementazione vera e propria della piattaforma software. L'introduzione comprende un accenno alle varie tecnologie utilizzate sia per la parte relativa all'estrazione dei dati, sia per la parte relativa alla visualizzazione grafica dei dati. Verrà descritto il database realizzato durante la fase di estrazione dei dati e infine una descrizione dettagliata della parte di visualizzazione.

Il quinto e ultimo capitolo sarà dedicato esclusivamente all'analisi dei dati ottenuti. Verranno riportati graficamente e in forma tabellare, tutti i risultati ottenuti dall'analisi dei dati precedentemente estratti. Successivamente verranno tratte specifiche conclusioni in merito ai risultati ottenuti.

Capitolo 1

Data Mining, Big-Data, DBMS e NoSQL

1.1 Data Mining

Nonostante si siano percepiti progressi in ogni aspetto della vita, i settori dell'archiviazione e della memorizzazione delle informazioni sono stati quelli in cui i cambiamenti si sono percepiti maggiormente e ad un ritmo più sostenuto. La crescita vertiginosa della mole di dati ha creato nuove opportunità, ma allo stesso tempo ulteriori problematiche, in ogni campo, dall'industria manifatturiera alla medicina e al marketing.

Quali sono le dimensioni di un database oggi? Le dimensioni raggiungono l'ordine dei Terabyte, 17 Terabyte è anche la dimensione del database del settore consegne UPS per la gestione delle spedizioni.[1]. Sembra incredibile, ma questa è la conseguenza naturale di operazioni sempre più automatizzate.

Generalmente, questi dati non vengono raccolti per essere analizzati o per ricavarne un modello di previsione, ma per migliorare l'efficienza delle operazioni a cui sono sottoposti. Una volta raccolti, però, costituiscono una grande quantità di informazioni preziose per i processi decisionali in qual-

siasi settore, come ad esempio la gestione dei rapporti con i clienti.

Con l'intensificarsi degli scambi sul Web, vengono messe a disposizione una gran quantità di dati raccolti e la necessità di analizzarli. Ciò è possibile grazie ai siti Web che riescono a tener traccia delle informazioni. L'analisi dati diventa, quindi, un'esigenza sempre più pressante.

I *data miner* sono coloro che applicano sofisticati algoritmi, conoscenza del campo applicativo e intuizione, per trarre da questi dati delle informazioni che possono essere fondamentali per un'azienda.

1.1.1 Cos'è il Data Mining

La disciplina del Data Mining è spesso accostata a quella della Business Intelligence¹ ed indica il procedimento di esplorazione ed analisi di un insieme di dati per individuare eventuali regolarità, estrarre nuove conoscenze e ricavare regole ricorrenti e significative. Obiettivo del Data Mining è di "estrarre informazioni" utili da un database e trasformarle in una struttura dati (pattern) per ulteriori utilizzi di indagine.

Secondo il libro *Data Mining Techniques*[4]

« Il data mining è il processo di esplorazione e analisi, automatico o semiautomatico, di un'ampia mole di dati al fine di scoprire modelli e regole significativi. »

Definirlo un processo automatico o semiautomatico però ha portato molte persone a pensare erroneamente che il data mining sia un prodotto in vendita, più che una tecnica o una disciplina da apprendere. La parte su cui focalizzare l'attenzione è sul processo di data mining. I dati sono disponibili in molte forme, in molti formati e da più sistemi.

¹Business Intelligence (BI) comprende l'insieme di strategie, processi, applicazioni, dati, tecnologie e architetture tecniche che vengono utilizzati dalle imprese [citazione necessaria] per sostenere la raccolta, l'analisi dei dati, la presentazione e la diffusione delle informazioni aziendali. Fonte: Wikipedia

Le tecniche di data mining sono state utilizzate per scoprire modelli nascosti e prevedere tendenze e comportamenti futuri nei mercati finanziari e non solo. I vantaggi concorrenziali ottenuti con l'estrazione dei dati includono un aumento delle entrate, un costo ridotto e una maggiore sensibilità e sensibilità al mercato.

1.1.2 Cosa può fare il Data Mining

Spesso si parla a sproposito di data mining, il termine viene usato per una serie ben precisa di attività che riguardano tutte le estrazioni di informazioni significative dei dati grezzi.

Le sei attività, o se preferiamo compiti, sono:

1. *Classificazione* - La classificazione consiste nell'esaminare le caratteristiche di un oggetto presentato di recente e assegnarlo ad un'insieme di classi predefinito. Gli oggetti da classificare sono generalmente rappresentati da record in una tabella all'interno di un database o in un file e l'atto di classificazione consiste nell'aggiungere una nuova colonna con un codice di classe di qualche tipo.
2. *Stima* - La stima si occupa di risultati costantemente valutati e viene spesso utilizzata per eseguire un'attività di classificazione.
3. *Previsione* - La previsione è la stessa della classificazione o della stima, tranne che i record sono classificati in base a un comportamento futuro previsto o al valore futuro stimato. In un compito di previsione, l'unico modo per verificare l'esattezza della classificazione è quello di attendere e vedere. La ragione principale per il trattamento della predizione come compito separato dalla classificazione e dalla stima è che nella modellazione predittiva ci sono ulteriori problemi relativi al rapporto temporale delle variabili di input o alla variabile di destinazione.
4. *Raggruppamento per affinità o regole di associazione* - Compito del raggruppamento per affinità è stabilire quali oggetti o voci possono

abbinarsi. L'esempio classico è stabilire quali prodotti si trovano insieme in un carrello del supermercato. La grande distribuzione utilizza tale tecnica per pianificare la distribuzione dei prodotti sugli scaffali o nei cataloghi in modo che gli articoli, che di solito vengono acquistati insieme, si trovino il più possibile vicini. Si occorre al raggruppamento per affinità anche per identificare le possibilità di vendite incrociate e per comporre pacchetti di prodotti o servizi.

5. *Clustering* - Il termine si riferisce alla segmentazione di un gruppo eterogeneo in sottogruppi (detti *cluster*) più omogenei. Ciò che distingue il clustering dalla classificazione è che non si fa ricorso a classi predefinite. Il clustering introduce molto spesso ad altre forme di data mining o modellazione: per esempio, potrebbe essere il primo passo da affrontare in un eventuale tentativo di segmentazione del mercato. Alla domanda "Verso quale tipo di promozione i clienti si dimostrano più ricettivi?", invece di individuare una regola valida per tutti, si suddivide la base della clientela in cluster o fasce con abitudini di consumo simili e quindi ci si ripropone di stabilire quale promozione interessa di più ciascun gruppo.

Dopo aver analizzato i dati attraverso le tecniche e gli algoritmi di Data Mining, esiste il problema di come mostrare i risultati all'utente finale. In alcuni casi lo scopo del data mining è offrire una semplice descrizione di un database complesso, in modo da analizzare e capire meglio individui, prodotti o processi in base ai quali i dati sono stati generati inizialmente. La descrizione efficace di uno specifico comportamento spesso ne offre anche una spiegazione, o per lo meno indica da dove partire per cercare una spiegazione. La visualizzazione dei dati è una forma molto efficace di data mining descrittivo. Non è sempre facile ottenere una visualizzazione significativa, ma un quadro chiaro della situazione vale molto più di innumerevoli regole di associazione, visto che è molto più immediato ricavare informazioni da dati visivi. [2]

1.2 Big-Data

1.2.1 Cos'è il Big Data

Big data è il termine usato per descrivere una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore.[5]

Già la parola stessa introduce una definizione di **big data**, ovvero **grandi dati**. Si parla davvero di una mole enorme di dati, di qualsiasi natura e generati da qualsiasi mezzo: telefono, computer, carte di credito, [...]. I big data non riguardano solamente il campo dell'Information Technology, ma possono essere necessari anche in qualsiasi mercato business, come le automobili o la medicina. L'unica rivoluzione non sta nella gran quantità di dati, ma nella capacità di usare tutte queste informazioni per elaborare, analizzare e sfruttare tali dati per scopi di vario genere, come in questo caso lo studio dei dati per le promozioni turistiche.

L'elaborazione di tali dati richiede tecnologie diverse e risorse che superano gli strumenti convenzionali di gestione e conservazione dei dati. Ciò permette di ottenere informazioni utilizzabili nei processi decisionali di un'azienda, è necessario utilizzare strumenti alternativi ed elevate capacità di calcolo.[6] Grazie al mondo open source sono state fornite soluzioni in grado di rendere utilizzabili i big data ai fini di estrapolare le informazioni. Queste soluzioni open source, inoltre, permettono di utilizzare hardware *commodity*², che rende l'investimento meno costoso.

Una definizione comune di *big data* è quella offerta da Doug Laney, che si basa sul "**Paradigma delle 3 V**", cioè *volume, velocità e varietà*. Il modello afferma che con il passare del tempo aumentano: volume e varietà dei dati, e la loro velocità di generazione. Modello tutt'ora usato per definire le principali caratteristiche dei big data. [7]

²Hardware facile da reperire e dal costo limitato

1.2.2 Peculiarità principali

Come sopracitati i fattori che caratterizzano i big data sono *volume*, *velocità* e *varietà*. Nello specifico:

- Volume.
Come già introdotto, un aspetto caratterizzante dei big data è la sua vasta quantità di dati. Gli strumenti mediante i quali scaricare grandi mole di dati possono essere generati automaticamente da macchina, possono essere sensori o strumenti scientifici, oppure generati direttamente dall'utente attraverso il Web 2.0.
- Velocità.
Una volta estratti i dati devono essere analizzati tempestivamente, affinché non divengano inutili, e quindi non considerabili per prendere una "decisione". La velocità di acquisizione e accesso dei dati richiesti è dunque fondamentale. Basta pensare che è molto frequente la necessità di reperire dati in tempo reale, per elaborarli con cadenza giornaliera, se non intra-giornaliera.
- Varietà.
I dati hanno natura fortemente eterogenea, ed ogni formato necessita di un trattamento dedicato. Tale caratteristica dei Big Data può richiedere operazioni di *scaling* o classificazioni convenzionali.³

Tali peculiari caratteristiche e specifiche richiedono che, rispetto alla memorizzazione, i database costituenti i Big Data, siano di tipo strutturato e non strutturato, e siano espressi su differenti scale di misura o siano anche di tipo qualitativo. Fortunatamente, i progressi compiuti dalla ricerca scientifica e tecnologica in ambito hardware e software hanno garantito soddisfacenti prestazioni in termini di efficienza, di velocità di accesso ai Big Data e di potenza ed efficacia di elaborazione.

Dal punto di vista delle architetture e della ingegnerizzazione dei dataset e delle strutture dei dati, i modelli più recenti dei Big Data sono basati su

³Manyika et al., 2011

metodologie ad elevata scalabilità⁴, e su soluzioni di tipo *NoSQL*. Si intende per NoSQL un insieme di tecnologie formanti un nuovo sistema di gestione dei dati diverso dal tradizionale RDBMS, in quanto non viene utilizzato il modello relazionale, non ha uno schema esplicito e il sistema è progettato per funzionare subito e bene in cluster.

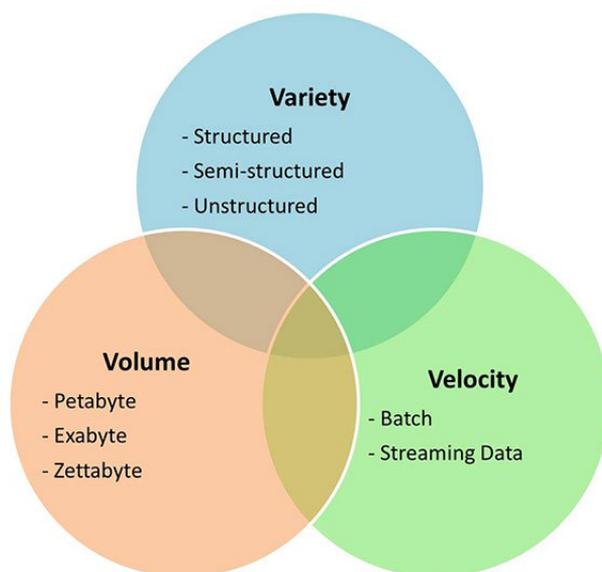


Figura 1.1: Modello delle 3V big data

⁴Il termine scalabilità, nelle telecomunicazioni, nell'ingegneria del software, in informatica e in altre discipline, si riferisce, in termini generali, alla capacità di un sistema di "crescere" o diminuire di scala in funzione delle necessità e delle disponibilità. Un sistema che gode di questa proprietà viene detto scalabile. Fonte: Wikipedia

1.2.3 Utilizzo e opportunità dei Big-Data

Le opportunità che hanno portato l'utilizzo dei big data possono essere viste sotto vari aspetti:

- > *Il business*: i big data danno la possibilità di perseguire nuovi modelli di business o di ottenere notevoli vantaggi competitivi sul business tradizionale dell'azienda.
- > *La tecnologia*: la gran mole e la complessità dei dati richiedono tecnologie apposite ed adeguate.
- > *L'aspetto finanziario*: l'utilizzo dei big data ha indubbiamente portato dei vantaggi economici alle aziende adottando questa soluzione. C'è da tener conto però, come citato nel punto precedente, che per gestire questa gran quantità servono delle tecnologie su misura per esse. Quindi come prima cosa occorre valutare a priori i costi da sostenere per implementare queste soluzioni.

I Big Data sono utilizzabili anche per l'e-learning, e la diffusione degli ambienti virtuali per l'apprendimento avranno un'impatto sullo sviluppo e la diffusione di questi sistemi, che potranno offrire nuove tecniche di apprendimento on line. Al giorno d'oggi assistiamo alla confluenza dei Big Data, il Data Mining, la Statistica, la Matematica, i Data Warehouse⁵, l'intelligenza artificiale, e le reti neurali in un nuovo paradigma, che prende il nome di Data Science[6].

Pertanto, il ruolo dei Big Data non sta solo nella capacità di poter gestire velocemente grandi volumi differenti di dati di vario tipo, ma è dato dall'opportunità che queste tecnologie offrono per nuove e notevoli applicazioni, anche in ambito educativo.

⁵In informatica un data warehouse (acronimo DW o DWH, traducibile come "magazzino di dati") è un archivio informatico contenente i dati di un'organizzazione, progettati per consentire di produrre facilmente analisi e relazioni utili a fini decisionali-aziendali. Fonte: Wikipedia

1.2.4 Criticità e rischi derivanti dai Big-Data

Un aspetto critico relativo ai Big Data è che si teme l'utilizzo predittivo delle correlazioni tra dataset aventi come oggetto le persone. Più precisamente, va sottolineato come sia divenuta una consuetudine la verifica (come, ad esempio, nel caso di una candidatura di lavoro) del profilo Facebook dei candidati da parte della commissione esaminatrice. I social network sono la vera miniera dei dati umani per Big Data.[3]

Altri aspetti critici riguardano l'affidabilità e la qualità dei dati e le questioni legate alla privacy.

Affidabilità e qualità dei dati.

Gli attributi che determinano la qualità del dato sono:

- *Completezza*: tutti i dati necessari per descrivere un'entità, una transazione o un evento, devono essere presenti.
- *Consistenza*: non devono essere presenti contraddizioni nei dati.
- *Accuratezza*: la correttezza dei valori. Il punto di partenza dell'accuratezza risiede nella loro *integrità*.
- *Assenza di duplicazione*: campi, record o tabelle non devono essere duplicati, questo, oltre a necessitare una doppia manutenzione, incide negativamente sulla qualità, poiché è possibile una mancata sincronizzazione tra le copie dei dati.
- *Integrità*: per garantire l'integrità dei dati vengono imposti alcuni vincoli. Quando si parla di questo aspetto ci si riferisce solitamente ai database relazionali.

Un aspetto da tenere fortemente in considerazione riguarda l'ambiguità di alcune parole, non sempre si possono distinguere significati diversi alle parole, ad esempio la parola "pesca" può riferirsi sia al frutto che allo sport. La sfida che i big data pongono è dunque legata alla rilevanza e all'attinenza che i dati hanno rispetto all'analisi. Come bisogna anche stare attenti

alla veridicità dell'informazione. Non sempre le notizie o i documenti contengono affermazioni e/o dati veritieri.

La qualità dei dati deve essere controllata e verificata, e il processo di *data quality* deve indicare come output quali sono i dati con livelli di integrità, completezza, consistenza e accuratezza ritenuti accettabili e quali sono invece da migliorare. Il concetto che sta alla base di data quality è semplice: se basiamo le nostre analisi e, di conseguenza, i nostri processi decisionali, su dati di scarsa qualità, il risultato che otterremo sarà anch'esso di scarsa qualità: si rischierebbe di prendere decisioni inefficienti, o addirittura del tutto errate.

Privacy e proprietà dei dati.

La diffusione e la reperibilità di gran quantità di dati, non indica che sia etico utilizzarli. Come anticipato precedentemente Facebook, e i social media in generale, danno la possibilità di estrarre dati sensibili, che potrebbero essere usati anche in modo inappropriato. Non solo, molto spesso vengono lasciate le cosiddette tracce elettroniche dei propri spostamenti, c'è quindi la possibilità che i nostri movimenti vengano monitorati.[6]

1.3 DBMS

Un DBMS (Data Base Management System) è un sistema software che è in grado di gestire collezioni di dati *grandi, condivise e persistenti*, in maniera *efficiente e sicura*. Il DBMS, pertanto, è un insieme di programmi, che sono rivolti alla gestione di dati memorizzati in archivi. Ovviamente, tra Database e DBMS esiste una forte iterazione per cui spesso si tende a confonderli, ma sono comunque due cose ben diverse e distinte.

E' sufficiente pensare al significato dell'acronimo DBMS (Sistema gestione della basi di dati), per capire che la principale funzione di un BDMS è quella di governare ogni operazione di accesso al Database, sia per quanto riguarda aggiornamenti o ricerche effettuate sugli archivi, sia per le richieste da parte di più applicazioni contemporaneamente.

La caratteristica principale del DBMS consiste nel mantenimento di un modello di dati, ovvero di un'astrazione logica tramite la quale l'utente può vedere i dati non in termini di record, file, campi, ma bensì in termini di *unità informative*, direttamente riferibili agli oggetti della realtà descritti dall'informazione stessa. Tramite i DBMS, è possibile implementare un paradigma di separazione di dati ed applicazioni. Le applicazioni non necessitano di conoscere la struttura fisica dei dati, ovvero come e dove sono memorizzati i dati, ma solamente cosa rappresentano, cioè la struttura logica.

In pratica, un DBMS può essere visto come un'architettura software a 3 livelli.

1. **Schema Esterno:** Descrive come si presenta il DB.
2. **Schema Logico:** Descrive cosa rappresenta il DB.
3. **Schema Fisico:** Descrive come/dove i dati vengono memorizzati.

Architettura a 3 livelli del DBMS [1]

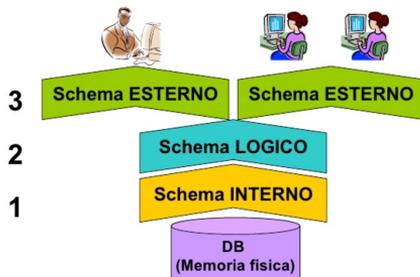


Figura 1.2: Architettura a livelli DBMS

Inoltre deve possedere le seguenti caratteristiche fondamentali:

- Facilità di accesso: il ritrovamento dei dati è facilitato e veloce.

- Indipendenza dalla struttura logica: posso modificare la struttura dati senza modificare il software.
- Indipendenza dalla struttura fisica: è possibile modificare i supporti con cui vengono registrati i dati e la modalità di accesso alla memoria di massa, senza modifiche alle applicazioni.
- Eliminazione della ridondanza: impedire che i dati si ripetano all'interno della struttura dati, che causa l'*incongruenza*, nel caso in cui lo stesso dato sia aggiornato in un archivio e non in un altro .
- Eliminazione della inconsistenza: (causata dall'incongruenza) evitare che i dati a disposizione non siano più affidabili, perché non si sa in modo certo quale dei diversi valori sia quello corretto.

1.3.1 Peculiarità principali

Un DMBS deve essere in grado di:

- Permettere la creazione di nuove base di dati tramite il linguaggio **DDL** (*Data Definition Language*), linguaggio che permette di definire la struttura dei dati.
- Facilitare gli utenti nella cancellazione, variazione e inserimento di dati tramite il linguaggio di manipolazione dei dati **DML** (*Data Manipulation Language*).
- Data un interrogazione al Data Base, deve rendere possibile l'estrazione di informazioni, tramite il **QL** (*Query Language*).

I componenti di un DBMS sono:

Efficienza: i DBMS forniscono adeguate strutture dati per organizzare i dati all'interno dei file, e per supportare le operazioni di ricerca/aggiornamento. In genere, parliamo di strutture dati ad albero o tabelle hash. L'*indice* è la struttura che contiene informazioni sulla posizione di memorizzazione delle tuple sulla base del valore del campo chiave.

Concorrenza: In molti sistemi è fondamentale gestire le operazioni concorrenti di accesso ai dati. La maggior parte dei DBMS forniscono un livello di *granularità di locking* più fine di quello convenzionale. Allo stesso tempo, un DBMS deve garantire il fatto che accessi da parte di applicazioni diverse non interferiscano tra loro, lasciando il sistema in uno stato inconsistente. Per prevenire tali situazioni, i DBMS implementano algoritmi di controllo della concorrenza. Il *Lock Manager* è quel componente del DBMS responsabile di gestire i lock alle risorse del DB, e di rispondere alle richieste delle transazioni, è quindi quell'elemento che consegna i permessi di lettura e di scrittura alle transazioni per le risorse condivise.

Affidabilità: Alcune operazioni sui dati sono particolarmente delicate, e devono essere gestite in maniera opportuna, secondo la regola del tutto o niente, ovvero o l'operazione viene eseguita fino alla fine o non deve essere tenuta in considerazione. Ad esempio nei sistemi bancari nel momento in cui si sta effettuando una transazione, o viene eseguita fino in fondo se si verifica un problema durante l'operazione quest'ultima viene annullata. Per questo, i DBMS devono fornire appositi strumenti per annullare operazioni non completate e fare **roll-back** dello stato del sistema. In molti casi i DBMS dispongono di strumento ed algoritmi ad hoc per garantire la persistenza dei dati anche in presenza di malfunzionamenti hardware/software. Un esempio possono essere i *log*, nei quali vengono indicate tutte le operazioni svolte dal DBMS, tramite i log è quindi possibile fare *do/undo* delle operazioni.

Sicurezza: La maggior parte dei DBMS implementa politiche di controllo degli accessi ai dati mediante sistemi di permessi. I permessi consistono nello stabilire quali sono le operazioni a cui l'utente è abilitato; le operazioni possono essere: lettura o scrittura.

In sostanza, con il DBMS si realizza una gestione centralizzata e controllata della base di dati, i vantaggi principali sono sottoelencati:

- Riduzione della ridondanza dei dati, dove con ridondanza s'intende la possibilità che i dati si presentino ripetutamente nella base di dati.
- Eliminazione dell'incongruenza.
- Condivisione dei dati da parte di tutte le applicazioni che ne facciano richiesta.
- Sicurezza e riservatezza delle informazioni, per ridurre il rischio di danneggiare i dati e di accessi non autorizzati.
- Ottimizzazione della struttura del data base, che ne facilita l'accesso, la manipolazione e la manutenzione.
- Indipendenza dei dati dalle applicazioni

Modelli logici DBMS

I DBMS possono differire sulla base del modello logico dei dati che supportano:

Modello Relazionale.

Il modello relazionale fu proposto per la prima volta nel 1970. Fin da allora ha avuto un successo crescente, dovuto fondamentalmente alla sua semplicità e alla sua flessibilità. La struttura di questo modello è molto semplice e potente. Al tempo stesso però impone un certo grado di rigidità. In ogni relazione possiamo rappresentare solo tuple corrispondenti allo schema della relazione stessa. In una base di dati, è opportuno evitare il più possibile che vengano inseriti dati privi di senso o sbagliati. Per delimitare che ciò accada, durante la creazione del database vengono stabiliti dei *vincoli di integrità*. Viene quindi introdotto il concetto di **chiave**. La *chiave primaria* è l'attributo che permette di individuare univocamente la tupla della relazione, mentre la *chiave esterna* è l'attributo di una tabella figlia che fa riferimento alla chiave primaria di una tabella detta madre.

Modello Gerarchico.

Un data base gerarchico è un insieme di *archivi*. Gli archivi sono composti da record chiamati *segmenti*. I segmenti sono in rapporto gerarchico tra loro attraverso legami di tipo padre-figlio. La struttura ad albero che caratterizza tale modello si basa sulla possibilità di individuare un segmento principale, denominato padre o radice, dal quale dipendono gli n segmenti figli, che a loro volta si trasformano in padri per altri figli, ecc.. . Non è possibile dal figlio risalire al padre. Questa struttura non è adatta ad una gestione moderna e dinamica delle basi di dati.

Modello Reticolare.

Il modello gerarchico rappresenta una prima soluzione al problema della gestione di grosse moli di dati, ma la sua intrinseca rigidità ne limita le potenzialità; per questo nasce il modello reticolare che, dotato di maggiore flessibilità, può adattarsi a situazioni più complesse. Il modello reticolare può essere visto come un'estensione del modello visto in precedenza, al quale sono state apportate miglioramenti rilevanti. In quest'ultimo modello ogni record può avere un numero qualsiasi di record particolari, chiamati record di collegamento (member), che formano delle catene tra le varie parti del sistema. Le strutture utilizzate nel modello reticolare sono due, *il record* e *il set*, che permette di correlare i record, per mezzo di catene di puntatori.

Modello ad Oggetti.

Una base di dati Object Oriented è un sistema con funzionalità e caratteristiche di un linguaggio di programmazione ad oggetti, che integra la tecnologia delle basi dati con paradigma ad oggetti, offrendo la possibilità di rendere persistenti i dati che gestisce, fornendo funzionalità di accesso e ricerca degli stessi.

1.4 Database NoSQL

1.4.1 Definizione di Database NoSQL

NoSQL è l'acronimo di "*Not Only SQL*" ed è usato per identificare tutti quei database che non fanno uso di un modello di dati relazionale, difatti concetti come tabella, indici, chiavi primarie o esterne non è detto che siano presenti.

Il movimento NoSQL nasce nel 2009 ed è in rapida crescita, sono infatti già molte le aziende, dalle piccole realtà alle grandi multinazionali che decidono di passare a questo tipo di database. Il motivo principale di questi cambiamenti è che permette alla base di dati e di conseguenza alle applicazioni corrispondenti, di scalare orizzontalmente superando in maniera egregia le limitazioni dei classici DB relazionali.

Tra i precursori troviamo Google con il suo BigTable e Amazon che ha dato vita a DynamoDB, entrambi progetti proprietari hanno dimostrato la necessità di uscire dai classici schemi relazionali per permettere un'efficiente scalabilità dei loro servizi.[8]

E' proprio da questi esperimenti che sono nati i primi database NoSQL, quasi tutti Open Source, ognuno caratterizzato da strutture dati diverse e altamente ottimizzate, ma che essenzialmente hanno in comune queste caratteristiche principali:

- schema-free
- la gestione di un'enorme quantità di dati garantendone comunque un rapido accesso
- la facilità di replicazione
- l'utilizzo di api per l'interfacciamento con i più noti linguaggi di sviluppo e abbandono di SQL

1.4.2 Peculiarità principali dei database NoSQL

I database NoSQL adottano DBMS non basati su modelli relazionali. Nell'elenco sottostante vengono riportate tutte le principali caratteristiche dei

database NoSQL:

Sono database *distribuiti*, ovvero ho la base di dati che non è implementata su una singola macchina.

Sono strumenti generalmente *open source*, questo aspetto ha contribuito alla sua diffusione al suo successo.

Non dispongono di schema (*schemeless*), non devono avere tabelle necessariamente con dati omogenei.

Non supportano le operazioni di *join*, ovvero un'operazione che permette di selezionare dati da più tabelle, evidentemente correlate tra loro.

Sono *scalabili orizzontalmente*.

Sono in grado di *gestire una gran mole di dati* e supportano le *repliche* di dati.

Non implementano le transazioni e le proprietà ACID dei database relazionali, ma delle proprietà dette BASE; l'acronimo sta ad indicare:

- *Basic Available*: il servizio è sempre disponibile, anche nel caso in cui il sistema sia soggetto a guasti.
- *Soft State*: la consistenza dei dati non sempre è garantita, due copie in due server possono differire in certi istanti di tempo.
- *Eventually Consistent*: prima o poi il sistema diventa consistente, ma non so quando nel tempo.

Queste proprietà sono dovute principalmente al fatto che questa tipologia di database è nata per essere veloce, flessibile e distribuita.

Elementi connessi alla diffusione dei database NoSQL

Vengono elencati ora i fattori che hanno contribuito alla diffusione dei database NoSQL.

Gestione dei Big-Data.

Come già precedentemente spiegato i big data comprendono una gran mole di dati, eterogenea, destrutturata, difficile da gestire attraverso tecnologia tradizionali (RDBMS). I sistemi NoSQL riescono a provvedere al problema dei big data su tutti i loro aspetti, quali volume, velocità e varietà.

Limitazioni del modello relazionale: questo modello è stato in grado di superare quelle limitazioni che rendono rigido il modello relazionale

1a Limitazione: il modello relazionale presuppone una rappresentazione tabellare, se i nostri dati non si presentano in tale forma, potrebbero causare problemi. Il modello NoSQL è *schemeless* privo di struttura e accetta una forma dei dati molto più libera, si passa dal parlare di tabelle con righe e colonne a parlare di collezioni e insiemi di documenti.

2a Limitazione: Alcune operazioni non possono essere implementate in SQL. Ad esempio in SQL non è possibile memorizzare un grafo e calcolare il percorso minimo tra due punti, mentre nel modello NoSQL si.

3a Limitazione: C'è stato un aumento esponenziale degli utenti che iniziavano a consumare enormi quantità di informazioni collegandosi tramite internet a basi di dati sempre più grandi. Scalare verticalmente stava diventando troppo costoso e insufficiente a mantenere la disponibilità dei servizi accettabile; l'obbligo di gestione dei vincoli, l'impossibilità di replicare i dati, una difficile gestione delle transazioni e la necessità di soddisfare le proprietà ACID, sono problemi che vengono superati con l'utilizzo della *scalabilità orizzontale*.

Teorema del CAP.

Tale teorema sostiene l'impossibilità per un sistema informatico distribuito

di fornire simultaneamente tutte e tre le seguenti garanzie:

1. Consistency (Consistenza): dopo ogni operazione il sistema si trova in uno stato consistente: tutti i nodi vedono gli stessi dati nello stesso momento.
2. Availability (Disponibilità): il servizio è sempre disponibile; l'eventuale fallimento di un nodo non blocca il sistema.
3. Partition Tolerance (Tolleranza di partizione): il sistema continua a operare anche in seguito ad un partizionamento della rete.

A seconda se il sistema decida di sacrificare Availability o Consistency o Partition Tolerance è possibile classificare i database distribuiti in 3 categorie: CA, CP, AP.

CASO 1: Proprietà garantite Consistency e Availability.

La probabilità che il server vada in crash dipende dal numero di nodi che compongono il cluster. Risulta evidente che i tradizionali RDBMS possono essere classificati come sistemi CA, cioè garantiscono consistenza e disponibilità ma non tolleranza alla perdita dei messaggi.

CASO 2: Proprietà garantite Availability e Partition Tolerance.

Non ho la consistenza, i dati replicati non sono aggiornati.

CASO 3: Proprietà garantite Consistency e Partition Tolerance.

Significa che il sistema non è sempre disponibile e la query potrebbe non andare a buon fine.

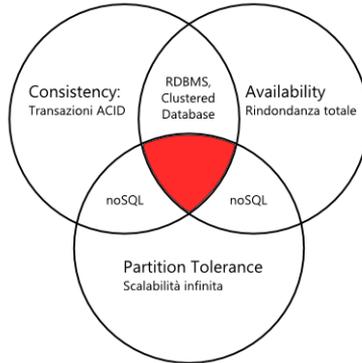


Figura 1.3: Teorema del CAP

1.4.3 Modelli principali di NoSQL

Tra i vari database non relazionali si possono distinguere principalmente 4 grandi gruppi:

1. **Key/Value:** modello chiave-valore, sono definiti da un semplice dizionario/mappa che permette all'utente di recuperare e aggiornare il valore memorizzato conoscendo la sua chiave.
2. **Document:** memorizza le informazioni come collezioni di documenti. Un documento può contenere informazioni annidate ed ha un formato riconosciuto (JSON, XML, ecc) che permette poi al server di eseguire delle query sui dati.
3. **Graph-oriented:** Rappresentano perfettamente una realtà composta da una fitta rete di connessioni e la modellano sotto forma di nodi e rami di un grafo. Ai nodi come ai singoli rami vengono associate le informazioni attraverso Key-Value store. Se togliamo le relazioni (ovvero i rami) assomigliano a tutti gli effetti ad un database documentale.
4. **Column-oriented:** Sistemi che utilizzano ancora le tabelle ma che non fanno alcun tipo di join, le informazioni non sono memorizzate

per riga bensì per colonna. Ogni colonna può avere un certo numero di righe e non necessariamente dello stesso numero delle altre colonne. I vantaggi di strutturare un database come insieme di colonne rispetto alle righe sono:

- lo schema è flessibile;
- se ho tantissime colonne ma lavoro solo su un sotto-insieme, avendole memorizzate su più file, aumento l'efficienza.

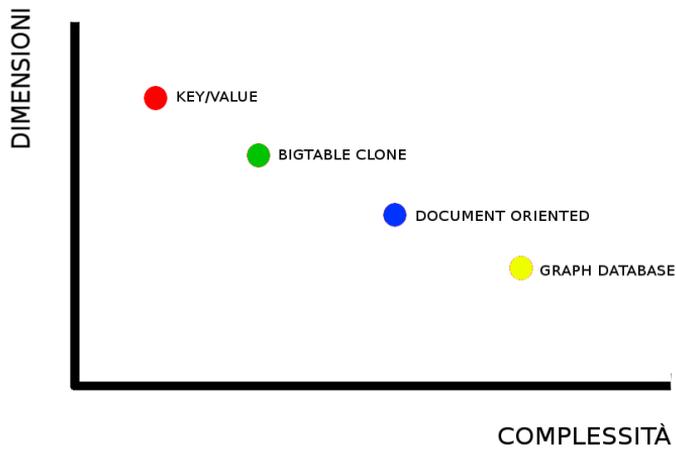


Figura 1.4: Modelli NoSQL

Capitolo 2

Social Media e Turismo

L'impatto dei social media nel settore turistico è diventato decisamente rilevante e non è possibile non tenerne conto, dal momento che hanno una notevole influenza nel processo decisionale del viaggiatore. La notevole rilevanza dei social media come strumento di marketing per le destinazioni fa sì che i DMOs cercano di creare un vantaggio competitivo per le proprie regioni sostenibile con la competitività presente dell'arena globale del turismo. Questa condizione ha portato alcune DMOs a esplorare i social media con strumenti relativamente a basso costo ma con una ricchezza globale. [13].

2.1 I social media

Più nello specifico, Rheingold definisce una comunità virtuale come [11]:

« un gruppo di persone che possono incontrarsi o meno faccia a faccia e che scambiano parole e idee attraverso la mediazione di bacheche e reti di computer. »

Alla base dei social media vi è la teoria dei 6 gradi di separazione secondo cui ogni individuo è legato a tutti gli altri attraverso un massimo di 6

persone [17]. Un social media rappresenta il nodo in cui questi interessi possono essere connessi tra loro attraverso differenti familiarità che vanno da acquisizioni occasionali ad affinità familiari. Il punto forte dei social media è il senso di appartenenza, le persone sentono di appartenere ad un gruppo. In base ai propri gusti, alle proprie caratteristiche gli utenti seguono e si iscrivono a gruppi di persone che rispecchiano le nostre caratteristiche, donandoci così la sensazione di far parte di un gruppo. Gli elementi sociali per costruire un social media efficace sono:

- Bisogni funzionali.
- Flusso di attività giornaliero
- Pulsioni aggreganti

2.1.1 Facebook

Facebook¹ nasce originariamente come piattaforma utile per tracciare le relazioni personali sorte tra i banchi dell'università americana. E' da quasi una decina di anni il social network più frequentato e visitato, solo in Italia conta più di 25 milioni di utenti, sono oltre un miliardo nel mondo. Mese dopo mese, grazie alla varietà di servizi offerti e alla facilità di accesso, oltre il 50% degli utenti attivi si collega ogni giorno. Da tenere in considerazione che il range di età degli utenti è davvero ampio. [15]

Una ragione del successo è sicuramente l'interfaccia intuitiva e semplice da usare, anche per i meno esperti. Inoltre, non offre semplicemente un profilo personale, mediante il quale sviluppare e gestire relazione con gli amici, ma un vero e proprio ambiente sociale dove condividere immagini, foto, video, pensieri, poter conversare in tempo reale, aderire a gruppi. Far parte di un gruppo, anche se in una comunità virtuale, dà senso di appartenenza e identità collettiva. Tutto ciò aumenta autostima e fiducia. Ciascun soggetto, però, partecipa in maniera differente alla vita della comunità. Facebook non è una piattaforma di nicchia, ma di massa. Per questo motivo chiunque

¹www.facebook.com

Confronto età utenti 2015 - 2017

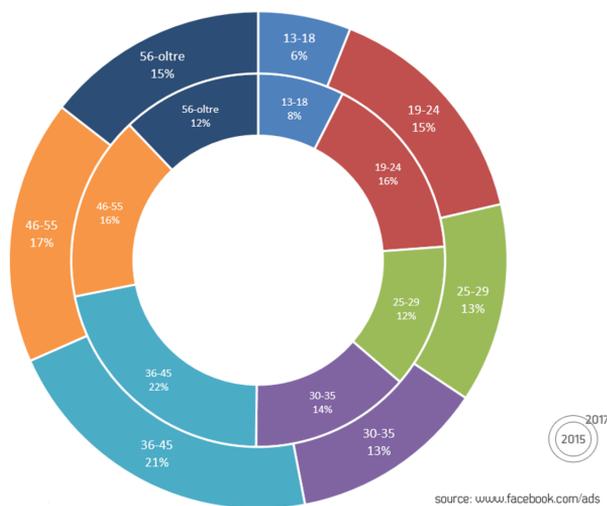


Figura 2.1: Età utenti italiani

voglia pianificare al meglio i propri investimenti in comunicazione e marketing su internet, non può non investire tempo, energie e parte del denaro per farsi pubblicità nei social network.

2.1.2 L'effetto "Mi Piace"

I social network diventano uno strumento potente di massimizzazione dei risultati positivi e la minimizzazione dei costi. Le aziende hanno quindi la possibilità di scoprire non soltanto il profilo di un potenziale consumatore, ma anche il suo network personale e relazionale. Facebook è uno strumento di marketing diverso da qualunque altro con il quale il mondo abbia avuto a che fare fino a questo momento, perché *la positività è parte integrante del suo DNA* [15].

I mi piace di Facebook:

- *Diminuiscono i costi di marketing e fanno crescere le vendite.*
Si ottiene molta più visibilità aumentando così l'engagement della

pagina, di conseguenza il costo della campagna Facebook diminuisce e il valore della pagina fan aumenta.

- *Permettono di controllare le conversazioni dei clienti.*

E' possibile vedere verso dove sono più orientati i gusti dei clienti, in base al contenuto che viene pubblicato e alla risposta che si ha ai vari post. Si acquisiscono così delle preziose informazioni di mercato, che contribuiscono al giusto orientamento per le varie offerte da proporre.

- *Provano che le persone stanno prestando attenzione.*

Ovviamente ogni qual volta viene pubblicato qualcosa, se le risposte sono tante significa che le persone stanno prestando attenzione ai contenuti pubblicati nella pagina. Cosa molto difficile in questi tempi, dal momento che le persone sono esposte a enormi quantità di informazioni disponibili in pochissimo tempo. Perciò l'attenzione dell'utente viene frammentata dalle varie attrazioni che la piattaforma offre. Catturare la loro attenzione, anche solo per 30 secondi nell'arco della giornata, mediante dei video magari, costituisce un'importante vittoria. Numerosi studi dimostrano che il venire a conoscenza di un prodotto o di una azienda, avviene grazie a Facebook. Questo denota anche l'importanza, per le imprese, di possedere una pagina fan su Facebook. Farsi conoscere e ottenere l'attenzione dei clienti è una di quelle battaglie che si devono vincere.

- *Contribuiscono alla fidelizzazione del cliente.*

In un ambiente sempre più competitivo, come quello dei social media, si agisce tramite messaggi, loghi, foto e video. Pubblicare dei post che possano coinvolgere gli utenti significa creare delle micro connessioni emozionali [15]. I clienti fidelizzati sono più propensi all'acquisto di prodotti.

- *Creano un passaparola positivo.*

Il *word-of-mouth*² è uno strumento potentissimo, che influisce in maniera determinante sulla fase decisionale dei clienti. Il passaparola è il

²tradotto: passaparola

risultato della socializzazione degli individui e può essere definito come *la trasmissione di informazioni positive o negative su beni e servizi offerti* [17]. Il passaparola nasce dall'esperienza personale di un individuo che può influenzare altre persone trasmettendo la sua delusione o la sua soddisfazione. Non tutti i passaparola negativi creano effetti dannosi sull'azienda. Ad ampliare tale fenomeno è il coinvolgimento del consumatore, ossia il significato personale, sociale ed economico che egli attribuisce agli acquisti. Più elevato il coinvolgimento più il consumatore ricercherà informazioni, prenderà in esame le caratteristiche del "prodotto", si farà una sua opinione e, soprattutto, sarà partecipe.

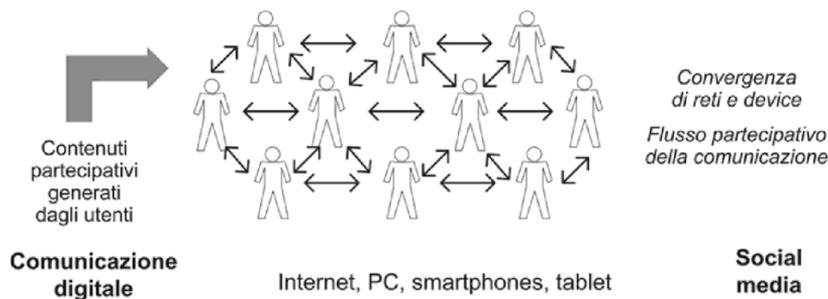


Figura 2.2: Comunicazione digitale

2.2 Il turista 2.0

Il turista di oggi è anche social. E' guidato da un nuovo modo di scegliere la destinazione, di sfruttarla e di mantenere con essa una relazione, durante e dopo il soggiorno. Infatti il viaggio si struttura in 5 fasi distinte che avvengono prima, dopo e durante il viaggio.

1. Dreaming: è la fase prima del viaggio, dove internet svolge un ruolo fondamentale perché diviene sempre più fonte di ispirazione. Mentre una volta ci si basava su esperienze di familiari e amici, ora la prima

cosa che si fa quando si vuole organizzare un viaggio è cercare online delle mete valide dove andare, leggendo recensioni, guardando foto messe in rete dai viaggiatori.

2. Planning: la fase in cui viene effettivamente prenotato il viaggio, come per la fase di dreaming anche qui le opinioni di altri utenti sono decisive. Si può dire che ora le scelte del turista sono influenzate dalla sua *community*. Secondo uno studio realizzato da Google, " *Travelers Road to decision*" del 2012, il 96% dei turisti cerca informazioni sulle destinazioni in rete e di questi il 63% lo fa attraverso social e motori di ricerca.
3. Booking: sono cambiate anche le modalità di prenotazione. Se si tratta di soggiorni inferiori ai cinque giorni solitamente vengono effettuate tutte online. In questa fase il turista si fida più delle recensioni e delle opinioni di altri utenti, pur non conoscendoli.
4. Living: la fase durante la permanenza del viaggiatore. Questa fase ha subito un cambiamento epocale. Oggi il viaggiatore condivide in tempo reale la propria vacanza, grazie al possesso di smartphon aventi Wi-Fi, 3G o 4G. Secondo *Text100DigitalIndex* il 72% realizza foto e video che pubblica sui social media e bel il 70% aggiorna il proprio stato su Facebook mentre è in vacanza.
5. Sharing: è la fase post vacanza, dove il turista tende a lasciare commenti, recensioni, foto e video per condividere la propria esperienza.

[10]

Turismo moderno	Turismo tradizionale	Turismo moderno	Turismo Post-moderno
Funzione sociale Vacanze	Rilassarsi Ricreazione	Sviluppo della propria identità Edonismo	Incontro con soggetti simili
Atteggiamenti esterni di fronte al turismo	Turismo come indicatore sociale dello status	Turismo indicatore di esperienza e di appartenenza	Turismo come strumento di soddisfazione personale ed emozionale
Modelli di target	Segmentazione sociodemografica	Segmentazione a seconda degli stili di vita	Segmentazione comportamenti emozionali e per esperienza

Figura 2.3: L'evoluzione dei modelli di turismo

2.2.1 Il potere delle raccomandazioni

Il 53% dei viaggiatori non consulta i canali ufficiali della destinazione (TNS Digital Life 2011) perché si fidano più dei turisti che sono stati nella destinazione, senza tener conto se siano amici o sconosciuti. La maggior parte delle persone che trovano strutture prive di recensioni tendono a non prenotarle. La condivisione di informazioni, la diffusione di piattaforme dedicate ai viaggi e delle *reti social* è una diretta conseguenza dell'importanza che il turismo riveste per la società post-moderna e, in particolare, dell'importanza che la vacanza ha per la singola persona. Questo contesto ha quindi facilitato la nascita di *community* di viaggiatori che si riuniscono intorno a destinazioni, condividendo viaggi, esperienze e passioni. Il turista oggi cerca la comodità e la facilità, fa parte ormai della sua quotidianità l'utilizzo dello smartphone, in quanto consente di comunicare con parenti e amici, oltre alla condivisione di foto e video sui social. [10]

2.2.2 Nuovi fattori di scelta

I criteri che spingono un turista a scegliere una destinazione sono decisamente cambiati da quelli tradizionali che ispirano la pubblicità. Il turista 2.0 dà all'immagine un peso maggiore rispetto alle parole, in quanto può vedere personalmente le caratteristiche del posto che a parole non sono sempre facilmente esprimibili o comprensibili. Quindi con che criterio viene scelta una meta? Il 50% si concentra su una destinazione alla quale è già interessato, il 41% la sceglie tra poche alternative e il resto decide in base alle offerte che trova in rete. Quello che caratterizzi soprattutto il turista moderno è l'*autonomia*, le persone vogliono poter organizzare le vacanze in maniera autonoma, senza fretta e/o pressioni. Questo spiega perché si preferisce ascoltare opinioni di altri turisti in quanto sicuramente daranno un'opinione sincera e più attendibile non vincolata da secondi fini. [10]

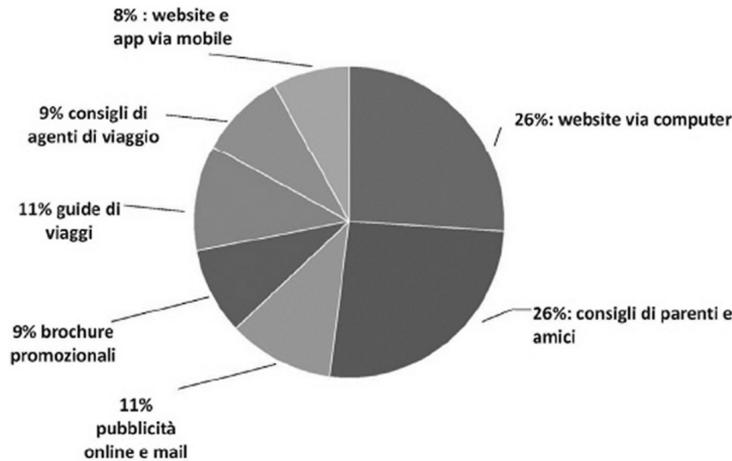


Figura 2.4: Fattori di scelta

2.3 L'engagement

Quando un brand pubblica sui suoi profili social contenuti ritenuti di qualità per gli utenti che costituiscono il suo target di riferimento, è facile che essi decidano di fare *following*³ nei suoi confronti. Nel momento in cui viene a costituirsi una community di follower attivi, cioè che partecipano in prima persona alle discussioni online è più facile per il brand (azienda, impresa, ecc..) acquisire una buona reputazione.

Un'azienda che gode di una buona reputazione riesce più facilmente a veicolare iniziative di marketing nei confronti della sua community; questa tenderà ad accettarle e, in molti casi, a diventare essa stessa veicolo di promozione.

Gli strumenti di social media marketing utili a fare engagement sugli utenti di una community variano da piattaforma a piattaforma.

Per esempio, hanno avuto particolare successo:

- *le applicazioni Facebook*: quiz e giochi che fanno leva sul social graph di Facebook per spingere i fan della pagina a particolari promozioni ed eventi;
- *i contest lanciati su instagram e twitter*: spingono le persone a scattare foto tematica e aggregarle sotto con uno specifico hashtag⁴, magari lanciato proprio dal brand stesso.

Il coinvolgimento di una community creata intorno a un brand è, dal punto di vista del marketing, l'azione che consente di trasformare i membri di una community in *leads*, cioè in potenziali clienti del prodotto/servizio promosso dall'impresa attraverso i social media. A differenza delle tradizionali attività di marketing, quelle svolte con il supporto dei social network consentono di acquisire in maniera più approfondita i dati di profilazione di ogni singolo utente e, inoltre, rendendo facilmente misurabile e valutabile

³tradotto: seguire

⁴Un hashtag è un tipo di etichetta (tag) utilizzato su alcuni servizi web e social network come aggregatore tematico, la sua funzione è di rendere più facile per gli utenti trovare messaggi su un tema o contenuto specifico. *Fonte: Wikipedia, www.wikipedia.com*

ogni promozione commerciale.

In questo modo è possibile costruire, nel corso del tempo, un'attività di *customer relationship management*⁵ che sia orientata alle effettive esigenze dei clienti e che favorisca la crescita e il rafforzamento della *loyalty*, ossia della fedeltà espressa dai clienti nei confronti del brand.

2.3.1 Strategie: l'utilizzo di testimonial

. Una buona strategia per aumentare il coinvolgimento degli utenti può essere quella di pubblicizzare il brand mediante testimonial. Come nel caso della pagina fan della regione Marche, Marche Tourism⁶, che ha coinvolto personaggi noti per far pubblicità alla regione. La campagna consiste nel pubblicare una foto con il volto del personaggio e una sua citazione riguardante la regione in questione. Ciò crea un alto tasso di coinvolgimento da parte degli utenti, sia da chi stima il testimonial e sia da chi è contrario. Fondamentale la tempestiva risposta da parte degli amministratori della pagina a qualsiasi feedback lasciato dagli utenti, positivo o negativo che sia, ma soprattutto negativo. E' fondamentale chiarire qualsiasi dubbio dell'utente onde evitare equivoci e soprattutto per confermare l'attendibilità dell'informazione [16].

⁵tradotto: Gestione dei rapporti con i clienti.

⁶<https://www.facebook.com/pg/marche.tourism/>



(a) Esempio di testimonial



(b) Botta e risposta utente-amministratore

Figura 2.5: Campagna pubblicitaria mediante testimonial

2.4 Le DMO

2.4.1 Un nuovo scenario competitivo

Per i turisti scoprire destinazioni nuove e sempre più lontane rappresenta un'esigenza di fondo insita nel concetto stesso di vacanza. Le **nuove destinazioni** più lontane hanno spesso costi inferiori e quindi prezzi più competitivi. Il settore turistico è costretto sempre ad adeguarsi al costante allargamento dei mercati. A livelli quantitativo l'*offerta* è cresciuta enormemente. L'allargarsi della *domanda* modifica il comportamento di viaggio. La *globalizzazione* dei mercati turistici ha indotto l'intensificarsi della concorrenza:

- Si è cercato di fare fronte alle sfide del mercato mondiale stimolando il turismo di qualità.
- Con l'evoluzione di tali tendenze è nata l'esigenza di nuove figure:

Destination Management e Destination Marketing.

Il prodotto turistico è da *gestire* (D. Management) e da *commercializzare* (D. Marketing) nella sua globalità.

Destination Management.

Il *Destination Management* nasce dall'esigenza di integrare all'interno di un processo strategico, le azioni necessarie per gestire i fattori di attrattiva e i servizi turistici per interessare la domanda di mercato e di posizionare la destinazione in ambiti competitivi adeguati rispetto alle caratteristiche del territorio. Un efficace processo di Destination Management deve quindi, da un lato, analizzare, definire e gestire i fattori di attrattiva e le differenti componenti imprenditoriali del sistema locale, dall'altro organizzare tutti questi elementi in proposte di offerta in grado di intercettare la domanda di mercato in maniera competitiva ed adeguata alla capacità del territorio.

2.4.2 Definizione di DMO

Trattasi di DMO (Destination Management Organization) le organizzazioni che "*svolgono un ruolo cruciale, attraverso efficaci scelte di posizionamento, aiutando le imprese del sistema locale, ovvero tutto il territorio, a creare, rafforzare e rendere sostenibile il proprio vantaggio competitivo*" (Sainaghi, 2005). Il World Tourism Organization le definisce come "l'organizzazione responsabile per il management e il marketing della destinazione" e in generale rientrano in una delle seguenti categorie:

- Autorità/organizzazioni nazionali per il turismo, responsabili per la gestione e la promozione del turismo a livello nazionale;
- DMO di livello regionale o provinciale;
- DMO locale, responsabile per la gestione e il marketing del turismo con competenze su una circoscritta area geografica o città.

Per quanto riguarda il mio elaborato le DMO di cui si parlerà faranno riferimento a quelle di livello regionale. Tradizionalmente le DMO sono state definite come destination marketing organization: tuttavia, negli ultimi

anni, pur riconoscendo che le azioni del marketing sono fondamentali per questo tipo di organizzazione, un'analisi più approfondita evidenzia l'esistenza di un insieme di attività che vanno da un orientamento al marketing o promozione ad uno che suggerisce un'azione più generale che include le funzioni tipiche del marketing ed aggiunge altre più attinenti con lo sviluppo della destinazione, tanto che spesso la DMO si configura come *destination developer*, ovvero come un soggetto che agisce per facilitare i processi di sviluppo e di difesa della competitività di un determinato territorio. [9]

2.4.3 I ruoli delle DMO

Le DMO hanno il compito di creare le condizioni culturali, strategiche ed organizzative favorevoli allo sviluppo turistico della destinazione. Di seguito viene proposto un possibile framework allo scopo di schematizzare l'azione organizzativa di tale ente.

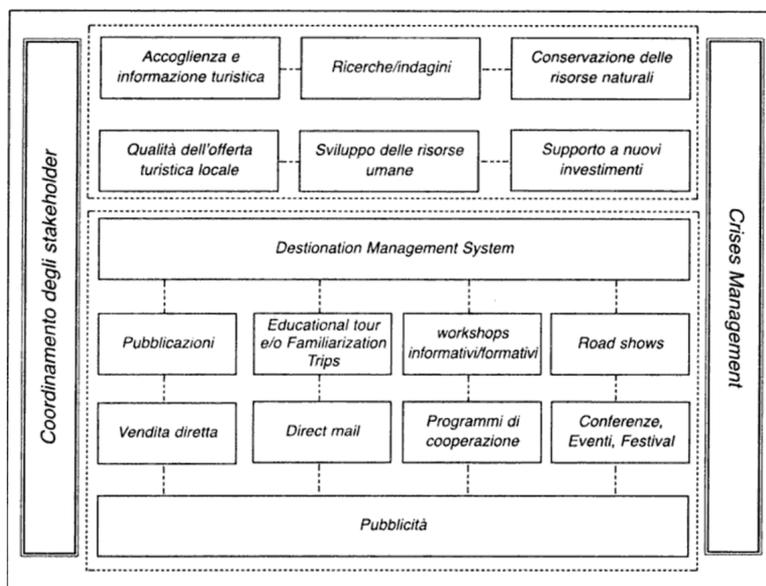


Figura 2.6: Ruoli fondamentali DMO

Diverse sono le attività che rientrano nel macro-gruppo delle funzioni di gestione della destinazione. La DMO può gestire punti di accoglienza ed informazione turistica come può focalizzarsi sulla Resource Stewardship, ossia lo sviluppo sostenibile della meta turistica che tiene a cuore le caratteristiche bio-fisiche dell'ambiente naturale. La DMO svolge attività di audit, di controllo su tutte le iniziative portate avanti dai singoli operatori, affinché lo standard qualitativo dei servizi offerti sia sempre ottimale. Un'indagine sulla soddisfazione del turista potrebbe essere il giusto riscontro della bontà della vacanza. La DMO incentiva l'imprenditorialità locale e indirettamente gestisce le risorse umane del sistema turistico evidenziando eventuali lacune di skills nell'offerta. [19]

2.5 Studi correlati

Prendendo in considerazione sempre siti web ufficiali e pagine fan delle DMO regionali, sia italiane che non, è stato possibile stabilire quali fattori determinino l'engagement e in che modo vengono usate e interpretate le strategie social media su Facebook.

L'articolo "*Internet marketing communication of tourist destination: Review of DMO website in Europe*"[12] ha concentrato l'elaborato sull'analisi di strumenti e tecniche di comunicazione di internet marketing usate dalle DMO nei loro siti web ufficiali.

I risultati ottenuti da tale studio stabiliscono che le tecniche più utilizzate dalle DMO per aumentare l'engagement degli user si dividono in 5 categorie principali: *la pubblicità, le pubbliche relazioni, promozione delle vendite, vendita personale e il marketing diretto*. Nello specifico queste categorie comprendono tecniche e strumenti di comunicazione marketing come ad esempio:

- Collegamenti ai siti web dei social media e servizi del Web 2.0. Il social media marketing utilizza le piattaforme web 2.0 come proprie per creare comunità virtuali. Ci sono diversi servizi del web che permettono di caricare dei contenuti, anche multimediali, e alcuni di essi

fanno sì che l'utente carichi il suo nuovo contenuto, come ad esempio un articolo, in modo tale che altri membri possano votarlo.

- Contest e lotterie. I contest sono basati su delle specifiche abilità dell'utente e il premio viene ricevuto sulla base di una valutazione di una giuria, mentre le lotterie contengono un elemento di fortuna. Solitamente i premi sono viaggi che includono soggiorni, pacchetti, biglietti, ecc.
- Tour virtuali e foto panoramiche. I contenuti interattivi e multimediali sono estremamente piacevoli per la presentazione delle offerte turistiche. Una foto panoramica consiste in un numero di foto che fanno sperimentare all'utente lo spazio intorno di 360 gradi. Un tour virtuale permette all'utente di muoversi virtualmente nello spazio e vedere le foto panoramiche di un luogo specifico. I tour virtuali avanzati sono basati sulla tecnologia della geo-localizzazione.⁷
- Disponibilità della lingua. Si è notato come la lingua con cui è possibile consultare il sito sia un fattore determinante. Ovviamente più sono le lingue disponibili più c'è la possibilità che i follower, anche stranieri, aumentino.

Da questa ricerca si evince che le tecniche e le strumentazioni più utilizzate per la comunicazione di marketing sono: le pubblicità e il marketing diretto (70%), seguite poi dalle public relation (32.31%), vendita personale (20%), mentre le tecniche meno utilizzate sono quelle riguardanti la promozione delle vendite (19%). Lo studio si conclude affermando che le DMOs non usano pienamente in maniera efficiente le tecniche e gli strumenti di comunicazione di marketing nella loro comunicazione con il mercato e il pubblico nei loro siti ufficiali. [12]

L'articolo sopra citato risale al 2011, mentre *Facebook as destination marketing tool: Evidence from Italian regional Destination Management Organization* è un articolo del 2013 che concentra l'attenzione su come vengono

⁷La geolocalizzazione è l'identificazione della posizione geografica nel mondo reale di un dato oggetto, come ad esempio un telefono cellulare o un computer connesso o meno ad Internet, secondo varie possibili tecniche. Fonte: Wikipedia.

Marketing communication instrument	Frequency	Maximum frequency	Relative frequency (in %)	Value on scale
Advertising	17	17	100%	Very often
PR	82	221	37,10%	Sometimes
Sales promotion	16	85	18,82%	Rarely
Personal selling	5	17	20,41%	Rarely
Direct marketing	133	187	71,12%	Often
TOTAL	253	527	48,01%	Sometimes

Source: Authors calculation

Figura 2.7: Utilizzo di strumenti di comunicazione di marketing su internet

usate le strategie social media su Facebook e che impatto hanno sull'engagement degli user.

Osservando più attentamente i dati di engagement emerge che la distinzione è tra le DMOs che usano e interpretano i social media in maniera attiva e quelle che lo utilizzano e interpretano in maniera passiva. Le DMOs attive (Trentino, Toscana, Marche e Sicilia) non utilizzano solamente Facebook ma numerose piattaforme social media come YouTube, Flickr e Pinterest. Inoltre, le regioni passive, come ad esempio il Molise, sono presenti su Facebook con un profilo ordinario, ciò comporta innumerevoli limitazioni come ad esempio avere solo 5000 amici, e usano i social media semplicemente per condividere informazioni e novità che sono già state pubblicate sul sito ufficiale della regione senza coinvolgere minimamente l'utente. A differenza di 3 anni ora tutte le regione hanno una propria pagina fan dedicata alla promozione turistica, ma sono comunque rimaste delle regioni più attive e altre meno.

In secondo luogo, sembra che le regioni come Trentino, Toscana e Sicilia abbiano sviluppato alcune buone pratiche gestionali che potrebbero esse-

re sfruttate con efficacia da altre regioni meno esperte, che attualmente si limitano a replicare sui contenuti sociali i contenuti pubblicati sul loro funzionario siti web. Come riportato anche nell'altro articolo, tali strategie includono la creazione di sondaggi di opinione, giochi, quiz e concorsi che comportano contenuti visivi (ad esempio, per selezionare l'immagine di copertina della pagina di Facebook). I risultati di tale studio indicano chiaramente che beneficerebbe anche altri DMO implementare strategicamente contenuti visivi per ottenere l'impegno degli utenti. Dato che è risaputa la preferenza da parte degli utenti di condividere contenuti visivi (Munar and Jacobsen, 2014), questo approccio potrebbe aumentare l'attività sociale nelle loro pagine. Inoltre, l'incorporazione di tali contenuti visivi in sondaggi o quiz, sembra produrre una "chiamata d'azione" ancora più forte per gli utenti, che aumenta ulteriormente l'attività sociale. Tuttavia, come dimostra questa analisi, solo pochi gestori di social media sono attualmente consapevoli del fatto che il valore edonistico derivante dai contenuti generati direttamente dall'utente potrebbe essere utilizzato per influenzare le decisioni di pianificazione del viaggio degli utenti (Ayeh et al., 2013) o per rafforzare la reputazione della loro destinazione creando un passaparola online positivo.

Terzo aspetto, le regioni italiane in genere utilizzano ancora i social media con un metodo tradizionale top-down anziché esplorare modi per rafforzare un approccio basato sui contenuti generati dall'utente quindi dal basso verso l'alto per interagire meglio e coinvolgere di più gli utenti, considerati anche come consumatori.

Quarto punto, non è ancora chiaro come i gestori dei social media DMO misurino il successo e l'efficacia delle loro strategie e attività su Facebook, anche se c'è da dire che i più grandi DMO regionali hanno fatto importanti investimenti con lo scopo di migliorare la cultura riguardante l'aspetto digitale dell'azienda.

Region	Total posts	Total posts DMO	% Posts DMO on Total posts
Trentino	8578	2690	31.4%
Emilia-Romagna	3289	2295	69.8%
Marche	2782	1494	53.7%
Liguria	2755	1305	47.4%
Tuscany	2188	790	36.1%
Sicily	1778	510	28.7%
Abruzzo	1726	988	57.2%
Puglia	1538	1084	70.5%
Basilicata	1534	993	64.7%
Piedmont	1474	1474	100.0%
Sardinia	1406	1065	75.7%
Friuli	1146	808	70.5%
Veneto	887	467	52.6%
Calabria	857	793	92.5%
Umbria	816	624	76.5%
Lombardy	613	510	83.2%
Valle D'Aosta	191	151	79.1%
Campania	35	1	2.9%
Molise	4	4	100.0%
Overall Total	33,597	18,046	53.7%

Figura 2.8: Numero complessivo dei post delle DMOs italiane nel 2013

The 2014 Traveler's Road to Decision[14] è un ulteriore studio svolto dalla Ipsos MediaCT, una compagnia indipendente di ricerca di marketing, commissionato da Google nel 2014. Consiste nell'intervistare 5000 che hanno viaggiato almeno una volta per motivi personali (3500 persone) o almeno tre volte per scopi commerciali (1500 persone) negli ultimi sei mesi. Sono state intervistate anche 1500 persone che hanno un reddito familiare superiore a 250 000 dollari e che hanno fatto un viaggio per motivi personali negli ultimi sei mesi. Gli intervistato hanno un'età compresa tra i 21 e i 64 anni, vivono negli USA, non coprono un ruolo importante all'interno delle aziende e sono coinvolti nei processi decisionali del viaggio. L'intervista è focalizzata sulle abitudini e gli atteggiamenti di viaggio.

Nella fase dell'*ispirazione* il 65% delle persone che viaggiano per piacere e il 69% delle perone che viaggiano per lavoro, cercano informazioni online prima di decidere la meta o come voler viaggiare. Le fonti di ispirazione sono quindi per il 65% le ricerche online, mentre per il 62% sono la famiglia, i collegi, gli amici.

Per quanto riguarda la ricerca, fase post-ispirazione, la parola chiave è "digitale". Internet è la fonte principale sia per la pianificazione di vacanze che per i viaggi di lavoro. Il 57% degli individui che viaggiano per svago e il 64% dei viaggiatori d'affari, iniziano le loro prenotazioni di viaggio e i processi di acquisto con delle ricerche. I motori di ricerca sono tra le fonti di pianificazione online più popolari per i viaggiatori, in particolare tra i viaggiatori che viaggiano per motivi di piacere.

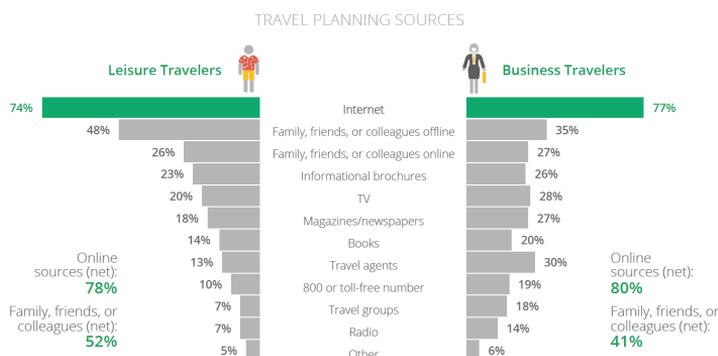


Figura 2.9: Fonti di pianificazioni di viaggio

La maggior parte dei viaggiatori ha indecisioni sul marchio e sulle differenze che si sono tra le marche. Alcuni viaggiatori che viaggiano per svago prenotano il loro viaggio senza prediligere alcun brand o compagnia. Que-

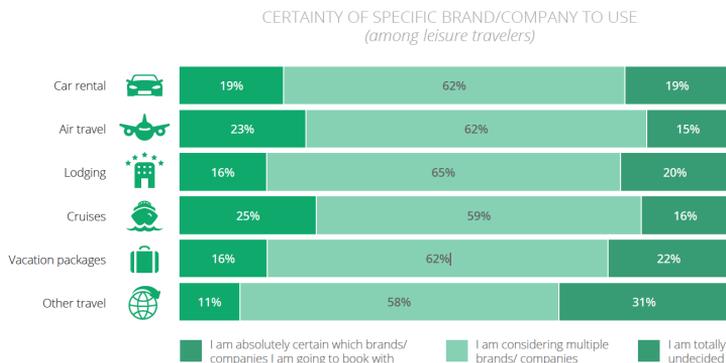


Figura 2.10: Conoscenza dei brand

sto perché essi non hanno ancora ben chiaro le differenze che ci sono tra le marche. I viaggiatori di lavoro sono più propensi a vedere delle differenze, ma non tanto di più rispetto agli altri viaggiatori.

L'influenza dei programmi fedeltà per i viaggi d'affari è notevolmente diminuita. In generale i viaggiatori passerebbero a programmi con vantaggi migliori e percorsi più facili che prevedono premi.

Personal travelers	Business travelers	Affluent travelers	Reasons for switching to a different loyalty/rewards program
88%	93%	92%	Would switch for any reason (net)
54%	55%	65%	Better perks
52%	54%	65%	Faster/easier to earn free flights, hotels rooms, rental cars, etc.
31%	40%	33%	Exclusive benefits
27%	34%	36%	Larger network/multiple routes or locations
26%	34%	34%	Ability to leverage points with other specific companies/partner hotels
13%	21%	10%	Personalized recommendations either pre-trip or during trip
4%	4%	3%	Other

Figura 2.11: Motivi per cambiare programmi fedeltà con altro

Gli smartphone svolgono un ruolo fondamentale nelle varie fasi del processo di viaggio. Il grafico sottostante riporta quali sono i mezzi con cui vengono affrontate le varie fasi, dall'ispirazione alla fase post-viaggio, facendo una distinzione tra viaggiatori d'affari e non.

	Leisure Travelers		Business Travelers	
	Computer/ Tablet	Smartphone	Computer/ Tablet	Smartphone
Used during any phase (Net)	94%	67%	97%	78%
Inspiration The time when you identified you wanted or needed to book travel	73%	31%	74%	39%
Research The time when you actively looked and researched your trip or travel plans	88%	27%	86%	36%
Purchase/booking The time when you booked your trip	81%	14%	84%	28%
Experiencing/traveling Any behavior you may have participated in during your trip	57%	50%	67%	54%
Post traveling Any behavior you may have participated in after you took your trip	63%	37%	67%	42%

Figura 2.12: Utilizzo dello smartphone nei vari processi

Inoltre gli smartphone danno la possibilità di prenotare i viaggi mediante le apposite applicazioni. Solitamente le app vengono sfruttate più da chi viaggia per lavoro, chi organizza viaggi per svago tende a prenotarli tramite smartphone, ma dai siti web. Tendenzialmente anche dopo aver scaricato le app, i viaggiatori utilizzano comunque i siti. L'83% dei viaggiatori non di lavoro si sono imbattuti in siti di viaggio privi di versioni compatibili con apparecchi mobili. Questa è una delle motivazioni che costringe il viaggiatore a cercare altrove e ha un impatto negativo sul marchio.

METHOD OF BOOKING ON SMARTPHONE <i>(among those who booked on the device)</i>	Leisure Travelers	Business Travelers
Online (Net)	71%	86%
Via the website using a browser	45%	55%
Used mobile apps (applications)	40%	63%
Phone call	44%	44%

Figura 2.13: Utilizzo dello smartphone nei vari processi



Figura 2.14: Utilizzo dello smartphone nei vari processi

Anche i video online hanno assunto un ruolo fondamentale nella fase decisionale del viaggio. Il 97% dei viaggiatori guardano video online che sono stati pubblicati nell'ultimo mese, per avere una maggiore consapevolezza della meta di destinazione. Precisamente, il 66% dei viaggiatori guarda video online quando stanno pensando di fare un viaggio e il 65% li guarda quando sta scegliendo una destinazione.

Per i viaggi di famiglia la metà dei viaggiatori decide di viaggiare senza avere ancora stabilito alcuna destinazione, che verrà scelta tenendo in forte

considerazione le preferenze dei bambini.

Nonostante gli studi siano relativamente recenti, in campo informatico, a distanza di cinque o sei anni possono verificarsi svariati cambiamenti. Soprattutto, con il progresso nel campo tecnologico, si tende sempre a migliorare le tecniche e le strategie di marketing mediante l'ausilio dei social media.

Capitolo 3

Progettazione

Questo capitolo tratterà più nel dettaglio le specifiche del progetto e la sua architettura, ovvero come è stato strutturato. Infine si parlerà della API di Facebook che è stata utilizzata utile al prelievo dei dati.

3.1 Specifiche del progetto

Il progetto consiste nello sviluppo di una piattaforma software per l'estrazione ed analisi dei dati proveniente dalle pagine Facebook delle DMO regionali italiane. Potenziali utilizzatori del tool potrebbero essere le DMO regionali stesse che desiderano capire le migliori strategie da adottare per una migliore pubblicazione dei contenuti, e valutare la propria presenza nel social media. I dati in questione sono stati prelevati facendo delle richieste mediante una specifica libreria fornita dalla piattaforma stessa.

La prima fase del progetto farà riferimento all'estrazione dei dati. Per dati si intendono: id dei post, data creazione, numero di mi piace, commenti, condivisioni e tipo di post. Tali informazioni sono relative all'anno solare 2016, quindi da Gennaio 2016 a Gennaio 2017 non compreso. Questo permette di avere già un'idea dei periodi, durante l'anno, in cui gli utenti vengono maggiormente coinvolti dai post pubblicati dalle pagine. L'obiettivo sarà quello di sviluppare un tool che permetta la raccolta di tali dati.

La seconda fase, invece, riguarda la visualizzazione grafica e nel contempo l'analisi dei dati raccolti. Un altro obiettivo del progetto è quello di fornire delle previsioni future in ambito turistico, analizzando le reazioni che gli utenti hanno di fronte ai contenuti pubblicati dalle singole pagine. L'analisi non consiste solo nel calcolo del coinvolgimento mensile degli utenti, ma anche in relazione al tipo di post che viene pubblicato (foto, video, link, ecc.).

3.2 Architettura del progetto

Come già precedentemente detto, il progetto è strutturato in due parti. La prima parte è dedicata ad uno degli obiettivi cardine dell'elaborato: l'estrazione dei dati dalle Pagine Facebook. Quindi è stato sviluppato uno script che si occuperà esclusivamente del raccoglimento e del salvataggio dei dati su un database creato ad hoc. Successivamente è stata sviluppata un'apposita interfaccia grafica che permette di estrarre i dati scegliendo regione e periodo temporale.

Le funzionalità previste dal tool sono le seguenti:

- Possibilità di scegliere la regione italiana di cui si vogliono conoscere le informazioni
- Possibilità di scegliere l'arco temporale
- Possibilità di visualizzare graficamente i dati analizzati

La seconda parte riguarda la visualizzazione dei dati raccolti. L'interfaccia grafica permette la visualizzazione dei grafici. Tali grafici possono essere di vario tipo: visualizzazione dei dati annuali, divisi per regione, visualizzazione dei dati secondo il tipo di post o visualizzazione dei dati di una singola regione relativi agli ultimi dodici mesi. Ovviamente è possibile scegliere la regione di cui visualizzare i dati e inoltre si può scegliere un periodo temporale a piacere. Inoltre, c'è la possibilità di confrontare a livello grafico gli engagement di due regioni a scelta. Questo script permette solo di

visualizzare le informazioni contenute nel database precedentemente riempito dallo script dedicato all'estrazione dei dati, non è possibile richiedere ulteriori dati che non siano già presenti sul database.

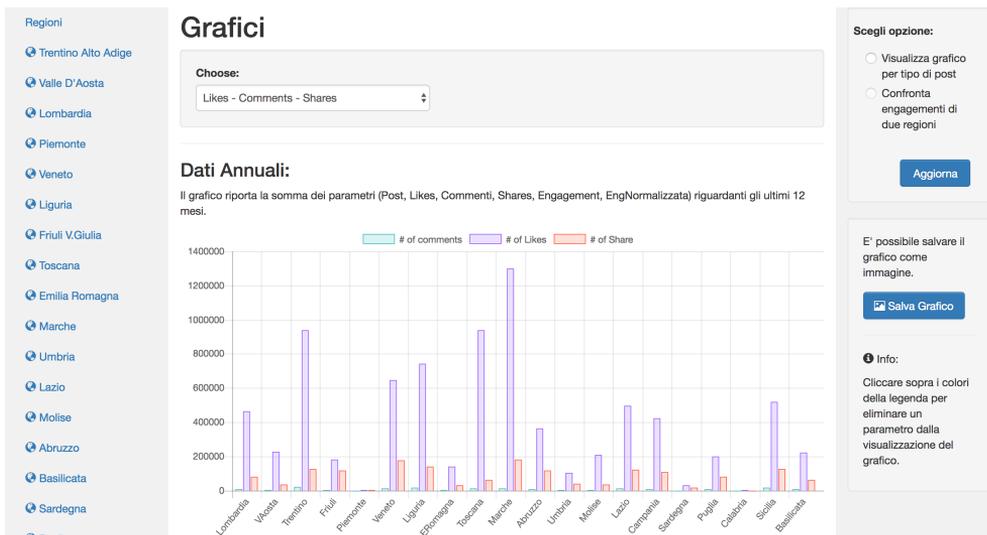


Figura 3.1: Interfaccia grafica visualizzazione dati

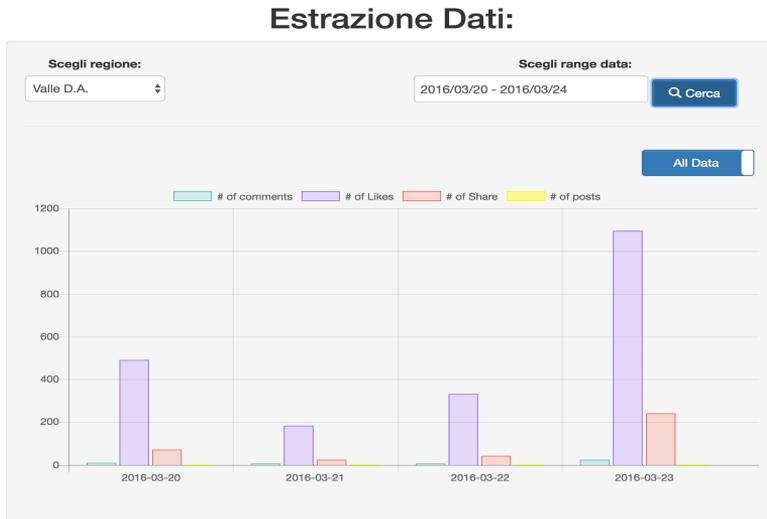


Figura 3.2: Interfaccia grafica estrazione dati

3.3 Facebook APIs Documentation

Le richieste eseguite per il prelievo dei dati sono state effettuate mediante una libreria specifica di Facebook l'**API Graph**. Per l'utilizzo di tale libreria è stato fondamentale consultare la documentazione presente nel sito <https://developers.facebook.com/docs>.

3.3.1 API Graph

"L'API Graph è il metodo principale per estrarre e inserire dati nella Piattaforma Facebook. Si tratta di una API di basso livello basata su HTTP che puoi usare per richiedere dati, pubblicare notizie, gestire inserzioni, caricare foto ed effettuare altre operazioni implementabili in un'app in modo programmatico."¹

¹fonte: <https://developers.facebook.com/docs/graph-api>

Il nome API Graph deriva dal concetto di "social graph", ovvero una rappresentazione delle informazioni presenti su Facebook composta da:

- **Nodi:** elementi come un utente, una foto, una pagina o un commento.
- **Segmenti:** le connessioni tra gli elementi sopra descritti, come le foto di una pagina o i commenti a una foto.
- **Campi:** informazioni sugli elementi, come il compleanno dell'utente o il nome di una pagina.

L'API Graph è basata su HTTP, pertanto funziona con qualsiasi linguaggio dotato di una libreria HTTP, come cURL e urllib. Per tale progetto il linguaggio utilizzato è **Python** e la libreria utile per le richieste API Graph è "**requests**". Nella maggior parte delle richieste, è necessario usare i *token d'accesso*, che possono essere generati automaticamente implementando Facebook Login oppure viene generato automaticamente attraverso il tool di Facebook for developer.[20]

```
response = requests.get("https://graph.facebook.com/v2.8/" + page[region[i]] + "?\n\nfields=feed.since(2016-01-01).until(2017-01-01).limit(25) \n\n{likes.limit(250),comments.limit(70),shares,created_time,from,id,type}&access_token="+token)\nresult = json.loads(response.content)
```

Figura 3.3: Esempio di chiamata API Graph

3.3.2 Facebook for developer tool

Tool di esplorazione per la API Graph è una pagina² che permette di prelevare un token d'accesso, inoltre dà la possibilità di testare le varie richieste API Graph.



Figura 3.4: Esempio di token d'accesso e richiesta API Graph

Il tool fornisce l'elenco di tutti i nodi di cui poter reperire le informazioni, e una volta formulata la richiesta il risultato viene restituito in formato json.



Figura 3.5: Esempio di richiesta API Graph

²<https://developers.facebook.com/tools/explorer/>

Capitolo 4

Implementazione

4.1 Tecnologie utilizzate

4.1.1 Python

Guido Van Rossum fornisce tale definizione del linguaggio:

« Python è un linguaggio di programmazione interpretato, interattivo e orientato agli oggetti. Incorpora al proprio interno moduli, eccezioni, tipizzazione dinamica, tipi di dati di altissimo livello e classi. Python combina un'eccezionale potenza con una sintassi estremamente chiara. Ha interfacce verso molte chiamate di sistema, oltre che verso diversi ambienti grafici, ed è estensibile in C e in C++. Inoltre è usabile come linguaggio di configurazione e di estensione per le applicazioni che richiedono un'interfaccia programmabile. Da ultimo Python è portabile: può girare su molte varianti di Unix, su Mac, su PC con MS-DOS, Windows, Windows NT e OS/2. »

Guido Van Rossum fu colui che nel 1989 decise di scrivere un linguaggio di programmazione che, a suo parere, risolvesse tutti i problemi presenti negli altri linguaggi di programmazione. Il suo linguaggio ebbe subito un notevole successo, inizialmente tra i colleghi del centro di ricerca dove lavorava

e poi successivamente alla pubblicazione su Usenet nel febbraio del 1991 il successo fu a livello mondiale.

Il termine Python deriva dalla passione dell'ideatore per il gruppo comico britannico degli anni sessanta, i Monty Python, che a loro volta scelsero questo nome per il loro gruppo perché "suonava divertente".

Python è un linguaggio intuitivo, interpretativo ed estremamente leggibile. Per leggibile si intende ordinato e ben indentato, dal momento che per delimitare i blocchi di codice non sono necessarie le parentesi, ma è sufficiente l'indentazione. Questo linguaggio permette di eseguire direttamente il codice sorgente che è stato scritto (da qui il termine *interpretativo*) o, addirittura, di scrivere il linguaggio direttamente sul prompt dei comandi, senza bisogno di creare o modificare il file sorgente (per questo motivo viene definito anche *interattivo*).

Le librerie si chiamano *moduli*. In questo progetto i moduli utilizzati sono:

- json: il modulo *json* è stato utile per parsare il risultato della chiamata API Graph, mediante il metodo "**json.loads(result)**", in modo da poter accedere ai dati in maniera diretta.
- requests: il modulo *requests* è stato utile per effettuare le richieste API Graph, mediante il metodo "**requests.get(richiesta API Graph)**".
- pymysql: tale modulo è utile per la creazione e gestione dei database su cui ho inserito tutte le informazioni ottenute dalle varie richieste.
- decimal: il modulo *decimal* è servito al fine di calcolare in maniera corretta tutte le metriche: *engagement* ed *engagement normalizzata*.
- flask: Flask è un *microframework* scritto in Python e basato su toolkit Werkzeug e Jinja2 motore di template. Flask è definito un microframework perchè non necessita di particolari strumenti o librerie. Non ha vincoli sul tipo di database a cui fare riferimento.

Python è stato utilizzato per prelevare i dati dalle pagine Facebook e, successivamente, per creare una piccola interfaccia grafica che desse la possi-

bilità all'utente di decidere da che regione e soprattutto di che periodo voler visualizzare i dati.

4.1.2 HTML

HTML è l'acronimo di *HyperText Markup Language*¹. L'HTML è un linguaggio di formato utilizzato per dichiarare la forma (struttura) che un documento ipermediale² deve assumere.

Il processo di standardizzazione del linguaggio è controllato da una organizzazione denominata *World Wide Web Consortium* (W3C).

Un documento HTML è un normalissimo file di testo e di conseguenza è visibile e modificabile con qualunque programma di word processing. I file HTML si distinguono dai normali file di testo tramite l'estensione *.html*. Un documento HTML è, per tanto, un unico file codificato in ASCII che include sia il contenuto che vogliamo far visualizzare, sia le istruzioni usate per elaborare quel contenuto.

Relativamente a questo progetto l'HTML è servito per creare la struttura dell'interfaccia grafica, con l'ausilio della libreria "**bootstrap**".

4.1.3 JavaScript

Javascript è un linguaggio di scripting client-side³.

Le caratteristiche principali di JavaScript sono riassunte nel seguente elenco:

- è un linguaggio case-sensitive; cioè fa differenza tra lettere maiuscole e minuscole.
- è un linguaggio interpretato
- il nucleo del linguaggio è stato incorporato all'interno del browser

¹tradotto: linguaggio per la marcatura di ipertesti

²documento contenente insieme di informazioni di tipo testuale, visivo e sonoro (testi, immagini, musiche, filmati) connesse fra loro e strutturate in modo interattivo

³tradotto: lato cliente

- esiste almeno in due varianti: lato client e lato server.

La variante lato client è quella più comunemente utilizzata e permette l'inserimento di contenuti eseguibili all'interno di pagine web. In questo modo le pagine Web non sono più *statiche*, ma diventano *pagine attive* o *dinamiche*, cioè possono comprendere programmi che interagiscono con l'utente, controllando il browser e creando dinamicamente nuovi contenuti HTML. Alcuni linguaggi di scripting (come il PHP) vengono eseguiti dal web server e sono detti, appunto, linguaggi sever-side. Javascript invece viene eseguito sul nostro personal computer dal browser, ecco perché viene definito client-side.

E' un linguaggio di scripting e non di programmazione. Pur avendo molti aspetti in comune, questi due tipi di linguaggi presentano notevoli differenze.

Linguaggio di scripting.

Viene utilizzato per realizzare gli **script**. Uno script è un piccolo programma interpretato dal browser. Un linguaggio di scripting può controllare il comportamento del browser e il suo contenuto, ma non può effettuare elaborazioni grafiche o gestire la rete.

Linguaggio di programmazione.

Un esempio di tale linguaggio può essere Java. Il linguaggio di programmazione non ha alcuna possibilità di controllare il browser nella sua interezza, ma può gestire la grafica, la rete e possiede funzionalità di *multithreading*, che permette di creare e gestire processi paralleli.

La sintassi JavaScript può essere inserita direttamente all'interno di una pagina HTML. E' sufficiente inserire i tag "`<script type="text/javascript"> .. </script>`" che delimitano il linguaggio JavaScript. Inoltre, non è compilato quindi è possibile visualizzare in qualsiasi momento il codice di una pagina HTML e leggere anche le righe di codice JavaScript. La sua sintassi non essendo compilata viene interpretata dal browser che, tramite un appo-

sito motore di scripting⁴, legge le parti di codice Javascript.[21]

Javascript è stato utilizzato esclusivamente nella seconda fase, ovvero per la realizzazione dell'interfaccia grafica. Tale linguaggio è rimasto utile alla gestione delle varie azioni presenti nell'interfaccia e delle richieste da parte del client. Grazie all'importazione delle libreria *Chart.js* è stato possibile realizzare dei grafici che hanno permesso la visualizzazione dei dati analizzati.

Le richieste al database sono state effettuate mediante chiamate Ajax che richiamano uno script PHP apposito per l'interrogazione del database.

4.1.4 PHP

L'acronimo sta ad indicare *Hypertext Processor* ed è un potentissimo linguaggio di scripting server-side che permette di creare complesse applicazioni.

E' un linguaggio multiplatforma, questo significa che funziona correttamente su server web con differenti sistemi operativi. Questo linguaggio è indicato soprattutto per lo sviluppo web e può essere tranquillamente integrato nel codice HTML. La peculiarità del linguaggio PHP è che il codice viene eseguito nel server, generando HTML che sarà poi inviato al client. Nonostante il linguaggio sia davvero semplice, facilmente utilizzabile anche dai principianti, offre prestazioni avanzate a programmatori esperti.[22] Per testare in locale il mio script PHP ho utilizzato il software MAMP.

Nell'elaborato, PHP è stato utilizzato come mediatore tra le richieste del client e il database. A parte, un file apposito nominato "*connessione.php*", permetteva la connessione dello script PHP con il database contenente i dati prelevati precedentemente. Tramite le chiamate Ajax al file "*procedure.php*", sviluppate in linguaggio JavaScript contenute nella pagina HTML, il PHP interroga il database e restituisce i risultati, sotto forma di stringa,

⁴cioè di visualizzazione

al client. In caso di esito positivo, i dati restituiti vanno a creare i vari dataset, utili per la realizzazione dei grafici.

```
/*
 * Metodo che esegue la query
 */
function request($query)
{
    include 'connessione.php';

    try
    {
        $stmt_latlon = $conn->prepare($query);
        $stmt_latlon->execute();
        $result = $stmt_latlon->fetchAll();
        $righe = $stmt_latlon->rowCount();
        if($righe!=0)
        {
            echo json_encode($result);
        }
        unset($stmt_latlon);
        unset($result);
    }
    catch(PDOException $e)
    {
        echo "Connection failed: " . $e->getMessage();
    }
} //Fine request
```

Figura 4.1: Procedura PHP

4.1.5 SQL

Le origini di SQL risalgono all'inizio degli anni 70 in California, con lo sviluppo del System R da parte della società IBM, un applicativo per la gestione dei dati, il cui linguaggio veniva chiamato Sequel. Venne poi sviluppato alla fine degli anni '70 un prodotto chiamato DB2 sempre con lo scopo di gestire le basi di dati, che utilizzava una primordiale versione di SQL. Da allora si sono succeduti un gran numero di prodotti che implementano questo linguaggio e ogni produttore, aggiungendo delle variazioni e estensioni proprie, ha contribuito alla creazione di quello che oggi viene chiamato SQL.

```

{
  sql = "SELECT engagement, engNorm FROM metriche WHERE region = '"+r+"'";
  $.ajax({
    method:"GET",
    url:"procedure.php",
    data: {query: sql},
    success: function(data)
    {
      var dati = JSON.parse(data);
      for(var j=0; j<dati.length; j++)
      {
        datasetEngB[j] = dati[j]["engagement"];
        datasetEngNormB[j] = dati[j]["engNorm"];
      }
    },
    error: function(a, b, c)
    {
      alert("Problemi ai click");
    }
  }); //Fine ajax
}

```

Figura 4.2: Esempio di chiamata Ajax

E' un linguaggio che serve per eseguire varie operazioni sia sui dati che sulle strutture che li contengono. La sigla, acronimo di *Structure Query Language*, è ormai diventata sinonimo di linguaggio standard per la gestione dei database relazionali.

SQL è dunque un linguaggio per la gestione di database relazionali, quindi assolve alle funzioni di Data Description Language (linguaggio di descrizione dei dati e delle strutture che li conterranno), di Data Manager Language (linguaggio per la manipolazione dei dati) e di linguaggio di interrogazione, Query Language.

Questo linguaggio viene definito **dichiarativo**, perché permette di svolgere le operazioni dichiarando *cosa* si deve ottenere e non *come* si deve ottenere. Forse la causa del suo grande successo sta nella sua semplicità di utilizzo. Non bisogna però farsi ingannare, perché se da un lato SQL è intuitivo e semplice, da un altro, per essere capito a fondo richiede di essere studiato con attenzione per coglierne tutte le sue sfumature e sfruttare a pieno le notevoli potenzialità che possiede.

All'interno dell'elaborato il linguaggio SQL è stato utilizzato per creare le tabelle e manipolarle. Le tabelle presenti nel database relativo al progetto fanno riferimento alle regioni di cui sono stati analizzati i dati, quindi sono state riempite con i dati estratti dalla piattaforma Facebook. Successivamente, una volta riempiti i database, sono state calcolate le metriche annuali di ogni regione e inserite poi in un unico database generale, contenente, quindi, tutte le informazioni di ogni regione. Per distinguere le informazioni da una regione all'altra è stato inserito un campo apposito che riporta il nome della regione.

```
# Create table as per requirement
sql = """CREATE TABLE BasilicataData (
  id_post CHAR(50) NOT NULL,
  data_creation DATE,
  type CHAR(50),
  shares_count INT,
  likes_count INT,
  comments_count INT )"""
#Eseguo la query
cursor.execute(sql)
```

(a) Creazione tabella

```
#Preparo la query per inserire il nuovo elemento nella tabella MarcheData
sql = "INSERT INTO BasilicataData(id_post, \
  data_creation, type, shares_count, likes_count, comments_count) \
VALUES ('%s', '%s', '%d', '%d', '%d')" % \
(feed['data'][x]['id'], feed['data'][x]['created_time'], \
  feed['data'][x]['type'], shares_count, likes_count, comments_count)
```

(b) Inserimento dati nella tabella

Figura 4.3: Utilizzo del linguaggio SQL

Una volta conclusa la fase di riempimento dei database, sono state effettuate delle richieste al database esclusivamente per far tornare dei dati. Ovviamente le query potevano essere di diversa natura, in base all'esigenza dell'utente.

```
sql = "SELECT region, SUM(tot_likes) as likes,SUM(tot_comments) AS comm, SUM(tot_shares) AS shares, SUM(tot_posts) AS posts, SUM(engagement) AS tot_eng, SUM(engNorm) AS tot_engN FROM Metriche WHERE region='"+region[i]+"'" GROUP BY region;"
```

Figura 4.4: Esempio di query

```
var sql = "SELECT * FROM Metriche WHERE region='"+region+"'" ;
```

Figura 4.5: Esempio di query 2

4.2 Estrazione dati: dettagli implementativi

4.2.1 Chiamate API Graph

Come precedentemente detto tutte le informazioni raccolte, sono state prelevate mediante chiamate API Graph. La chiamata è strutturata dal nome a dominio della Pagina Facebook, dai campi che si vogliono prelevare con i relativi limiti, il periodo temporale ed infine la chiave di accesso. Una grossa limitazione di questa libreria è che non si possono prelevare più di 250 informazioni alla volta, ciò causa un messaggio di errore per via dei troppi dati richiesti. E' stato quindi difficile trovare il giusto equilibrio tra i limiti, alla fine la soluzione è stata quella di porre un limite di 25 al numero di post da poter prelevare, 250 è il limite dei *likes* e 70 quello dei *commenti*.

Questa combinazione di limiti ha evitato un altro grande problema che si è riscontrato durante lo sviluppo del progetto, ovvero quello relativo alla scadenza della sessione. La chiave di accesso dopo un periodo di tempo cambia, quindi se lo script non finisce di estrarre i dati entro tale periodo, la sessione scade e la chiave di accesso con cui si effettuano le richieste non è più valida.

4.2.2 Il campo *next*

Le richieste che vengono fatte sono tutte dei *link* e il risultato ottenuto è sempre in formato JSON. A differenza degli *share* che vengono già calcolati e restituiti sotto forma di numeri, i *commenti* e i *likes* vengono restituiti come ulteriori array JSON. Questo perché vengono fornite alcune informazioni sugli utenti che hanno messo una reazione, quindi per ogni like vengono riportati l'id e il nome dell'utente, in più per i commenti vengono riportati anche la data di creazione del commento, il messaggio e l'id del messaggio.

Ciò comporta che il calcolo del numero dei "mi piace" e dei commenti per ogni post deve essere effettuato manualmente. E' prevedibile che un post possa avere più di mille "mi piace" e, avendo messo un limite di 250 "mi piace" alla volta, può capitare che non tutti risiedano sulla stessa pagina.

```

"likes": {
  "data": [
    {
      "id": "923397987684623",
      "name": "Dimitri Kozlovskij"
    },
    {
      "id": "285813711584819",
      "name": "Anna ..."
    },
    {
      "id": "147171449073117",
      "name": "Anna ..."
    },
    {
      "id": "10152065501916958",
      "name": "Anna ..."
    }
  ],

```

(a) Likes

```

"comments": {
  "data": [
    {
      "created_time": "2017-06-08T21:08:09+0000",
      "from": {
        "name": "Dimitri Kozlovskij",
        "id": "1054800687865228"
      },
      "message": "Love it.❤️",
      "id": "1647556125285158_1676425612398209"
    },
    {
      "created_time": "2017-06-08T20:57:07+0000",
      "from": {
        "name": "Anna ...",
        "id": "896652753682550"
      },
      "message": "Yes painted there",
      "id": "1647556125285158_1676409092399861"
    }
  ],

```

(b) Commenti

Figura 4.6: JSON restituito dalle richieste API Graph

Infatti, alla fine dell'array relativo ai likes, ma lo stesso vale anche sia per i commenti che per i post, c'è un campo chiamato "next" che contiene il link alla pagina successiva. E' stato necessario, quindi, effettuare un controllo se il campo "next" fosse presente o meno all'interno dell'array JSON che viene restituito dalla chiamata API Graph, per far sì che tutte le pagine vengano esaminate e tutti i dati calcolati correttamente.

```

#-> print(str(x)+"\n link: "+str(feed['data'][x]['link']))
#se un post ha dei like conto quanti sono
if 'likes' in feed['data'][x]:

    likes_count = len(feed['data'][x]['likes']['data'])
    paging = feed['data'][x]['likes']['paging']

    #i like potrebbero essere segnati su più pagine
    #quando sono tanti, quindi scorro tutte le
    #pagine dove sono presenti i like
    while 'next' in paging:
        url = paging['next']
        response = requests.get(url)
        json_s = json.loads(response.content)
        likes_count = likes_count + len(json_s['data'])
        paging = json_s['paging']
    print("likes_count: "+str(likes_count))

```

Figura 4.7: Controllo del campo next

Nel momento in cui non è più presente il campo "next", significa che la pagina che stiamo analizzando è l'ultima a contenere delle informazioni.

4.2.3 Filtraggio post

Non è sempre detto che ogni giorno gli amministratori pubblichino dei post, può capitare che in un giorno, o addirittura per un certo periodo temporale, la pagina non pubblichi nulla. Quindi uno dei controlli che è essenziale fare per il buon funzionamento dello script è quello di vedere se, nel JSON di ritorno dalla richiesta, è presente il campo "**feed**". Ciò sta ad indicare che c'è del contenuto da poter analizzare.

Può anche capitare che un post non abbia ricevuto alcun tipo di attenzione e quindi sia privo sia di "mi piace" che di commenti o condivisioni. Perciò un altro controllo che deve essere assolutamente fatto è quello di verificare la presenza dei campi "likes", "comments" e "shares" all'interno nell'array chiamato "data" contenuto nel campo "feed". Nel caso in cui lo script effettui delle operazioni su delle variabili nulle l'intero processo di esecuzione verrà introdotto causando un'eccezione. Ecco perché è molto importante prima effettuare tali controlli.

Altra cosa da tenere fortemente in considerazione è il fatto che all'interno delle pagine Facebook, nella sezione post, non sempre sono pubblicati contenuti dagli amministratori della pagina, ma da qualsiasi utente che segue la pagina. In questo progetto però le specifiche richiedono l'analisi delle reazioni degli utenti ai post pubblicati dalle pagine stesse, ecco perché tra i vari controlli è previsto anche quello che verifica che l'autore del post sia la pagina stessa e non terzi. Ciò avviene controllando se il campo "name" coincide con il nome della pagina, in caso contrario il post non viene preso in considerazione e si passa a quello successivo.

```

if 'feed' in result:
    feed = result['feed']
#Eseguo l'analisi finchè flag==true oppure finchè non viene superato
#il numero massimo di paging
while flag:

    conta = conta+1
    print("Paging: "+str(conta))

    #scorro i post presenti nella pagina
    #per prelevare solo le informazioni
    #utili per lo studio
    for x in range(len(feed['data'])):
        likes_count = 0
        comments_count = 0
        shares_count = 0
        str_comment = []

        #Questa prima scrematura permette di prelevare
        #solo post pubblicati dalla pagina e non da terzi
        if feed['data'][x]['from']['name'] == "Visit Abruzzo":

            print(str(x)+" : "+str(feed['data'][x]['created_time']))

            #se un post ha dei like conto quanti sono
            if 'likes' in feed['data'][x]:

                likes_count = len(feed['data'][x]['likes']['data'])
                paging = feed['data'][x]['likes']['paging']

                #i like potrebbero essere segnati su più pagine
                #quando sono tanti, quindi scorro tutte le
                #pagine dove sono presenti i like
                while 'next' in paging:
                    url = paging['next']
                    response = requests.get(url)
                    json_s = json.loads(response.content)
                    likes_count = likes_count + len(json_s['data'])
                    paging = json_s['paging']
                print("likes_count: "+str(likes_count))

```

Figura 4.8: Algoritmo estrazione parte 1

```

#se un post è stato condiviso prelevo il numero di condivisioni
if 'shares' in feed['data'][x]:
    shares_count = feed['data'][x]['shares']['count']
    print("shares_count: "+str(shares_count))

#se un post ha dei commenti
if 'comments' in feed['data'][x]:

    comments_count = len(feed['data'][x]['comments']['data'])
    paging = feed['data'][x]['comments']['paging']

    #come per i like, anche i commenti potrebbero trovarsi su
    #più pagine, quindi scorro tutte le pagine
    while 'next' in paging:
        j=0
        url = paging['next']
        response = requests.get(url)
        json_s = json.loads(response.content)
        comments_count = comments_count + len(json_s['data'])
        paging = json_s['paging']
    print("comments_count: "+str(comments_count))

#Preparo la query per inserire il nuovo elemento nella tabella MarcheData
sql = "INSERT INTO AbruzzoData(id_post, \
data_creation, type, shares_count, likes_count, comments_count) \
VALUES ('%s', '%s', '%s', '%d', '%d', '%d')" % \
(feed['data'][x]['id'], feed['data'][x]['created_time'], feed['data'][x]['type'], \
shares_count, likes_count, comments_count)
#->print(sql)
try:
    # Execute the SQL command
    cursor.execute(sql)
    # Commit your changes in the database
    conn.commit()
except:
    # Rollback in case there is any error
    conn.rollback()

```

Figura 4.9: Algoritmo estrazione parte 2

```

#Controllo se la pagina ha ancora post da analizzare
if 'paging' in feed:
    url = feed['paging']['next']
    print(url)
    try:
        response = requests.get(url)
        feed = json.loads(response.content)
    except:
        print(response)
        flag = False
else:
    flag = False

```

Figura 4.10: Algoritmo estrazione parte 3

4.2.4 Interfaccia grafica

Lo script che estrae i dati inizialmente era esclusivamente *back-end*⁵, successivamente è stata sviluppata un'interfaccia grafica per renderla *front-end*.

Tramite quest'interfaccia, l'utente ha la possibilità di scegliere la regione di cui si vogliono scaricare i dati e il periodo temporale. Una volta cliccato sul tasto della ricerca viene prima verificato che tali dati non siano già presenti all'interno della tabella contenente i dati di ogni regione, in caso contrario viene richiesto all'utente se vuole avviare il processo di estrazione dati. L'operazione richiede un po' di tempo, ovviamente maggior il range temporale richiesto, maggiore sarà il tempo di esecuzione.

L'interfaccia non prevede solo di estrapolare informazioni in tempo reale, ma fornisce anche una visualizzazione mediante grafici del numero totale dei like, commenti e share per ogni post e il numero dei post stessi giornalieri.

⁵I termini front-end e back-end denotano, rispettivamente, la parte visibile all'utente e con cui egli può interagire (interfaccia utente) e la parte che permette l'effettivo funzionamento di queste interazioni. Il front-end, nella sua accezione più generale, è responsabile dell'acquisizione dei dati di ingresso e della loro elaborazione con modalità conformi a specifiche predefinite e invarianti, tali da renderli utilizzabili dal back end. Il collegamento del front end al back end è un caso particolare di interfaccia. Fonte: it.wikipedia.org

Inoltre, effettua il calcolo dell'engagement ed engagement normalizzato e lo visualizza graficamente.

4.3 Visualizzazione dati: dettagli implementativi

Come già precedentemente detto, l'interfaccia grafica per la visualizzazione dei dati raccolti, e poi analizzati, interroga il database e non permette l'estrazione in tempo reale delle informazioni. Quest'interfaccia è strutturata da una *Home* dove è possibile visualizzare il grafico dei dati annuali di tutte e venti le regioni prese in considerazione. Poi c'è la possibilità di selezionare una regione a piacimento per vedere più nello specifico le informazioni relativi ad essa. E infine è possibile fare un confronto sull'engagement fra due regioni scelte.

L'interfaccia è stata realizzata in HTML con l'ausilio delle librerie **bootstrap** per rendere più piacevole alla vista la grafica. Gli eventi sono stati gestiti in JavaScript, così come anche le richieste al server. I grafici sono stati realizzati grazie alla libreria "*Charts.js*".

```
/*
 * Funzione che crea il grafico con tutti i dati (Likes, comments, shares)
 */
function chart(asseX, datasetAllComments, datasetAllLike, datasetAllShares)
{
    // Labels (lunghezza dataset: "asseX.length")
    greenBackground = [];
    greenBorder = [];
    violetBackground = [];
    violetBorder = [];
    orangeBackground = [];
    orangeBorder = [];

    for (var i=0; i<asseX.length; i++)
    {
        greenBackground[i] = "rgba(75, 192, 192, 0.2)";
        greenBorder[i] = "rgba(75, 192, 192, 1)";
        violetBackground[i] = "rgba(153, 182, 255, 0.2)";
        violetBorder[i] = "rgba(153, 182, 255, 1)";
        orangeBackground[i] = "rgba(255, 99, 71, 0.2)";
        orangeBorder[i] = "rgba(255, 99, 71, 1)";
    }

    // Fine del dettaglio colori
    var ctx = document.getElementById("myChart");
    var myChart = new Chart(ctx, {
        type: 'bar',
        data: {
            labels: asseX,
            datasets: [{
                label: '# of comments',
                data: datasetAllComments,
                backgroundColor: greenBackground,
                borderColor: greenBorder,
                borderWidth: 1
            },
            {
                label: '# of Likes',
                data: datasetAllLike,
                backgroundColor: violetBackground,
                borderColor: violetBorder,
                borderWidth: 1
            },
            {
                label: '# of Share',
                data: datasetAllShares,
                backgroundColor: orangeBackground,
                borderColor: orangeBorder,
                borderWidth: 1
            }
        ]
    },
    options: {
        scales: {
            yAxes: [{
                ticks: {
                    beginAtZero: true
                }
            }
        ]
    }
});
}
//Fine function chart LCS
```

(a) Parte 1

```
},
{
    label: '# of Likes',
    data: datasetAllLike,
    backgroundColor: violetBackground,
    borderColor: violetBorder,
    borderWidth: 1
},
{
    label: '# of Share',
    data: datasetAllShares,
    backgroundColor: orangeBackground,
    borderColor: orangeBorder,
    borderWidth: 1
}
]
},
options: {
    scales: {
        yAxes: [{
            ticks: {
                beginAtZero: true
            }
        }
    ]
}
});
}
//Fine function chart LCS
```

(b) Parte 2

Figura 4.11: Realizzazione grafico

4.3.1 La Home

Più nello specifico, la Home è composta dall'*header* contenente il titolo dell'elaborato, un menù laterale fisso che riporta l'elenco di tutte le regioni, i nomi delle regioni a loro volta sono dei link che se cliccati portano alle informazioni specifiche di tale regione. Sempre nella Home, il menù laterale destro, invece, presenta due contenitori, in uno c'è una *radio button* che ci permette di aggiornare il grafico visualizzando i dati annuali secondo il tipo di post pubblicato (foto, link, video ecc.). L'altra opzione della *radio button*, è selezionabile solo se siamo all'interno della pagina di una singola regione e prevede di scegliere un'ulteriore regione con cui fare il confronto dell'engagement. Nel secondo contenitore, oltre ad esserci le informazioni relative all'uso del grafico, c'è un bottone che permette il salvataggio in formato ".png" del grafico.

La parte centrale della pagina è dedicata al grafico, in alto un menù a tendina per decidere che tipo di dati nell'asse delle x visualizzare, la scelta è fra "mi piace-commenti-shares" o "post-engagement-engagement normalizzato". Sotto il menù c'è un contenitore uguale per tutti i grafici della pagina definito mediante l'elemento HTML "*canvas*" a cui sono stati assegnati altezza e larghezza standard. Per finire nel fondo della pagina è dedicato al *footer* contenenti le informazioni dello sviluppatore.

4.3.2 Le Regioni

Lo scheletro della pagina rimane invariato, quindi "header", "menù laterali" e "footer" saranno sempre gli stessi. L'unica cosa che varia è il contenuto della parte centrale, ovvero se viene selezionata una singola regione, a fianco al menù a tendina comparirà un "*range data picker*" che permette di scegliere un'arco temporale delle informazioni da visualizzare. Alla prima apertura il grafico visualizzerà i dati degli ultimi dodici mesi della regione specificata, nell'asse delle X troveremo i mesi, ma rimane sempre la possibilità di scegliere che tipi di dati visualizzare grazie al menù a tendina in alto.

L'opzione aggiuntiva possibile solo quando si apre una singola regione, pre-

vede di scegliere una regione tra la lista, eccetto se stessa ovviamente, per confrontare gli engagement e gli engagement normalizzati. Ogni grafico presente nell'interfaccia può essere salvato sotto forma di immagine formato .png.

```

/*
 * Funzione che parte quando clicchiamo sul bottone "Salva grafico"
 * e salva il grafico come immagine in formato .png
 */
$("#saveChart").click(function(){
  var ctx = document.getElementById("myChart");
  ctx.toBlob(function(blob){
    saveAs(blob, "chart_1.png");
  });
}); //Fine click saveChart

```

Figura 4.12: Codice salvataggio grafico formato .png

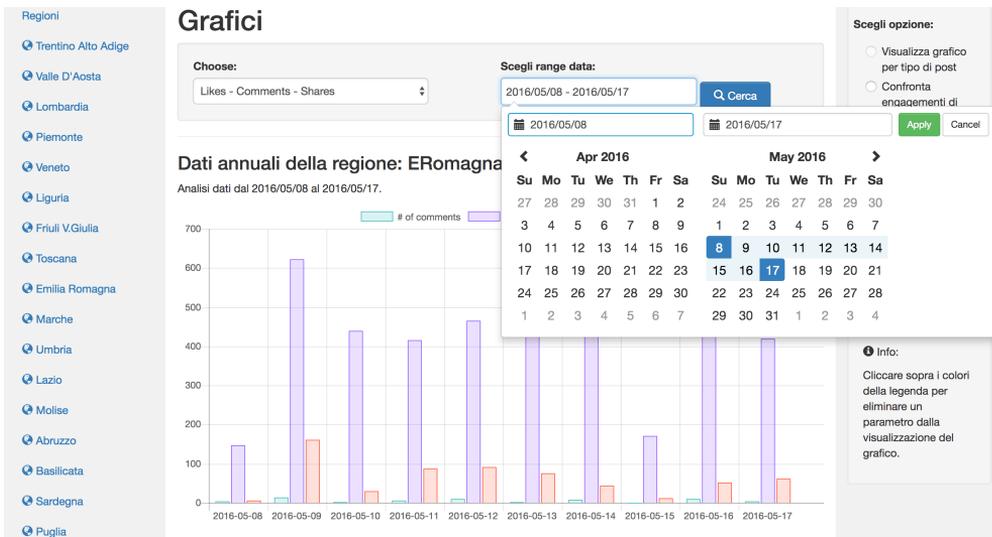


Figura 4.13: Scelta range data

Capitolo 5

Analisi dei Dati

Nel quinto e ultimo capitolo verranno analizzato i dati estratti. Nello specifico verranno rappresentati in forma tabellare e graficamente i dati; successivamente i risultati verranno analizzati e commentati al fine di determinare conclusioni sulla piattaforma Facebook, in merito al suo utilizzo e al coinvolgimento degli utenti.

5.1 Reperire i dati

Per effettuare tutti i calcoli e ottenere tutti i risultati di cui si aveva bisogno è stato sufficiente interrogare il database, che una volta popolato nella fase di estrazione dei dati, sono state create due tabelle contenenti tutte le informazioni necessarie. Una prima tabella nominata "AllData2016", la struttura di tale tabella è la seguente:

- ***id***: corrisponde all'identificativo della riga nella tabella, è un campo univoco.
- ***day***: è il campo che contiene data in cui è stato pubblicato il post.
- ***tot likes***: contiene il numero totale di likes per quello specifico post.
- ***tot comments***: contiene il numero totale di commenti per quello specifico post.

- **tot shares**: contiene il numero totale di condivisioni per quello specifico post.
- **type**: indica di che tipo è il post, se è una foto, un video, un link o altro.
- **region**: contiene la regione corrispondente del post.

Questi dati sono sufficienti per effettuare diversi tipi di analisi.

La seconda tabella che è stata creata utile all'analisi è quella relativa alle metriche chiamata, appunto, "Metriche2016", la struttura di tale tabella è la seguente:

- **id**: corrisponde all'identificativo della riga nella tabella, è un campo univoco.
- **month**: è il campo che contiene il mese in cui il post è stato pubblicato.
- **tot likes**: contiene il numero totale di likes per quello specifico mese.
- **tot comments**: contiene il numero totale di commenti per quello specifico mese.
- **tot shares**: contiene il numero totale di condivisioni per quello specifico mese.
- **tot post**: contiene il numero totale di post pubblicati in quello specifico mese.
- **engagement**: indica l'engagement calcolato per quello specifico mese.
- **engNorm**: indica l'engagement normalizzato calcolato per quello specifico mese.
- **region**: contiene la regione corrispondente del post.

Questa tabella riporta i valori di engagement mensile, da cui è possibile effettuare altri tipi di analisi.

5.2 Analisi dati annuale

Negli articoli precedenti si è parlato delle tecniche con cui le DMOs pubblicizzano e promuovono la loro regione attraverso i social media e i siti web ufficiali, in questo caso particolare si farà riferimento a Facebook. Gli studi avevano come obiettivo quello di verificare se le DMOs utilizzavano efficacemente le tecniche di promozione turistica e di conseguenza se coinvolgevano effettivamente gli user seguaci delle varie pagine fan di ogni regione.

Nella tabella sottostante sono state specificate le pagine ufficiali delle DMOs per la promozione turistica della loro regione e i relativi seguaci.

REGIONI	LINK	NUMERO FAN
Trentino Alto Adige	www.facebook.com/visittrentino/	217.673
Valle D'Aosta	www.facebook.com/pg/visitdaosta/	62.489
Piemonte	www.facebook.com/121408547869162/	4398
Lombardia	www.facebook.com/inLOMBARDIA/	178.588
Veneto	www.facebook.com/TurismoInVeneto	180.229
Liguria	www.facebook.com/turismoinliguria	146.021
Friuli Venezia Giulia	www.facebook.com/friuliveneziagiulia.turismo/	85.583
Emilia Romagna	www.facebook.com/EmiliaRomagnaTourism.IT/	62.380
Toscana	www.facebook.com/pg/VisitTuscany/	572.414
Marche	www.facebook.com/marche.tourism/	216.658
Umbria	www.facebook.com/pg/Umbriatourism/	30.656
Abruzzo	www.facebook.com/visit.abruzzo/	121.003
Lazio	www.facebook.com/visitlazio/	153.101
Molise	www.facebook.com/molipuntose/	41.318
Campania	https://www.facebook.com/campaniadavivere/	65.552
Basilicata	www.facebook.com/BasilicataTuristica/	65.823
Puglia	www.facebook.com/pg/viaggiareinpuglia.it	137.634
Calabria	www.facebook.com/TurismoRegioneCalabria/	4367
Sardegna	www.facebook.com/VisitSardinia/	43.985
Sicilia	www.facebook.com/visitsicily.info/	310.768

Figura 5.1: Tabella 1 - Link fan page e numero fan

Per effettuare questo tipo di analisi sono state estrapolate tutte le informazioni necessarie da ogni post, ovvero il numero dei *likes*, il numero dei *commenti* e degli *shares* e si è studiato il loro andamento nei mesi dell'anno solare del 2016.

Per la precisione l'analisi si è basata su un campione 21444 post pubblicate dalle DMOs nelle loro rispettive pagine fan su Facebook, di cui solo 33, (0,15%), non hanno riscontrato alcuna reazione da parte degli utenti, mentre 38 post non hanno ricevuto alcun "mi piace" (0,18%), 6508 nessun commento (30,3%) e 2560 post non sono stati condivisi (11,9%). I restanti 21411 post hanno almeno una reazione di qualsiasi tipo, per un totale di 8.941.289 reazioni, di cui 7.265.627 "mi piace", 151.643 commenti e 1.524.019 condivisioni.

TOTALI	#
# Totale post 2016	21444
# Totale post senza reazioni	33
# Totale post senza like	38
# Totale post senza commenti	6508
# Totale post senza shares	2560

Figura 5.2: Tabella 2 - Post totali anno 2016

# Likes	# Posts	% Sul totale posts
> 10 000	25	0,12
> 5000	82	0,39
< 1000	19976	93,16
< 100	8844	41,25
< 10	1907	8,9

Figura 5.3: Tabella 3 - Post e numero di "mi piace"

# comments	# Posts	% Sul totale posts
> 1000	0	0
> 100	100	0,47
< 100	21342	99,53
< 50	21064	98,23
< 10	17163	80,04

Figura 5.4: Tabella 4 - Post e numero di commenti

# shares	# Posts	% Sul totale posts
> 10 000	2	0,01
> 5000	6	0,03
< 1000	21315	99,4
< 100	17750	82,78
< 10	7463	34,81

Figura 5.5: Tabella 5 - Post e numero di condivisioni

Troviamo in forma tabellare la percentuale dei post pubblicati dalle DMOs rispetto ai post totali di ogni pagina dell'anno 2016.

REGIONI	POST TOTALI 2016	POST DMO	%POST DMO SUL TOTALE
Trentino Alto Adige	2196	1128	51,37
Valle D'Aosta	1345	739	54,95
Piemonte	1287	1176	91,38
Lombardia	2011	1519	75,54
Veneto	1559	616	39,52
Liguria	2727	1762	64,62
Friuli Venezia Giulia	1082	566	52,32
Emilia Romagna	2025	1441	71,17
Toscana	324	324	100
Marche	2111	1320	62,53
Umbria	2059	1773	86,11
Abruzzo	2742	1130	41,22
Lazio	1596	1596	100
Molise	1174	1027	87,48
Campania	2558	2360	92,26
Basilicata	1805	1036	57,4
Puglia	1450	721	49,73
Calabria	225	195	86,67
Sardegna	627	390	62,21
Sicilia	1243	639	51,41
TOTALI	32146	21458	66,76

Figura 5.6: Tabella 6 - Post totale vs. Post delle DMOs

La tabella sottostante riporta il numero di "mi piace", commenti e condivisione annuali di ogni regione.

REGIONI	#Likes (L)	#Comments (C)	#Shares (S)
Trentino Alto	992563	22109	130527
Valle D'Aosta	224167	4276	36060
Piemonte	4205	65	1149
Lombardia	279812	6697	68865
Veneto	692395	12774	196808
Liguria	623682	15029	88082
Friuli Venezia	152684	5084	102149
Emilia Romagna	152380	2620	30101
Toscana	577948	7938	43717
Marche	1063945	12356	163222
Umbria	103443	2096	40000
Abruzzo	410739	11631	133473
Lazio	514297	12039	133227
Molise	292968	6485	40029
Campania	224762	4105	64478
Basilicata	249284	6046	60721
Puglia	177256	5983	59670
Calabria	1741	30	581
Sardegna	39066	1098	15791
Sicilia	488290	13182	115369
TOTALE	7265627	151643	1524019

Figura 5.7: Tabella 7 - Parametri totale regionali annuali

Il grafico sotto stante riporta il numero totale di post pubblicati dalle DMOs per ogni mese dell'anno 2016.

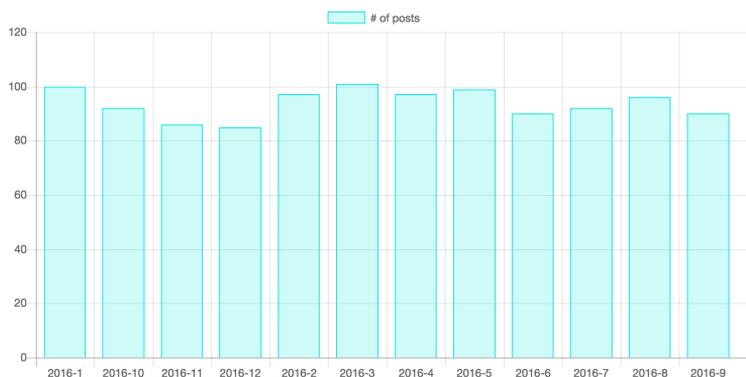


Figura 5.8: Post totali per ogni mese dell'anno 2016

	gennaio	febbraio	marzo	aprile	maggio	giugno	luglio	agosto	settembre	ottobre	novembre	dicembre
Nord	692	756	821	738	794	837	812	713	775	692	673	643
Centro	516	451	503	447	461	412	478	379	428	433	396	492
Sud	521	398	410	428	435	401	434	379	444	517	458	494

Figura 5.9: Post totali per ogni mese

Nei mesi invernali le regioni del Nord Italia pubblicano il minor numero di post rispetto ad altri mesi dell'anno, presumibilmente questi dati possono essere giustificati dal fatto che Dicembre e Gennaio siano i mesi di punta per i paesi che ospitano strutture sciistiche, perciò anche con poca promozione turistica si riscontrano numerose reazioni.

Oltre ad una analisi annuale dove nell'asse delle X troviamo i 12 mesi dell'anno, l'analisi annuale dei dati può fare anche riferimento alle 20 regioni italiane, ovvero quante reazioni sono state calcolate annualmente per ogni regione.

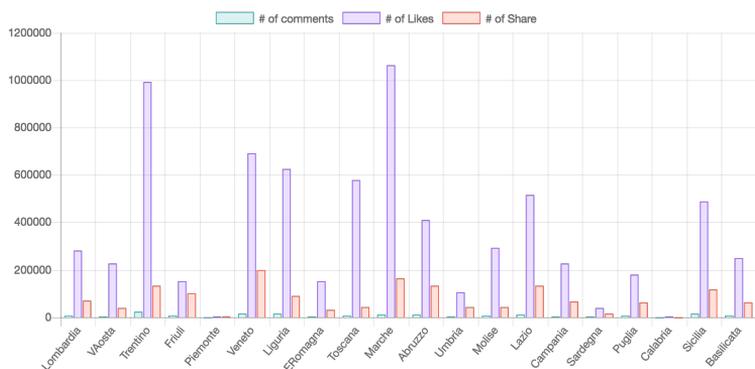


Figura 5.10: Totale parametri annuali regionali

Quest'ulteriore grafico rappresenta le reazioni ottenute dai post in base al mese dell'anno 2016.

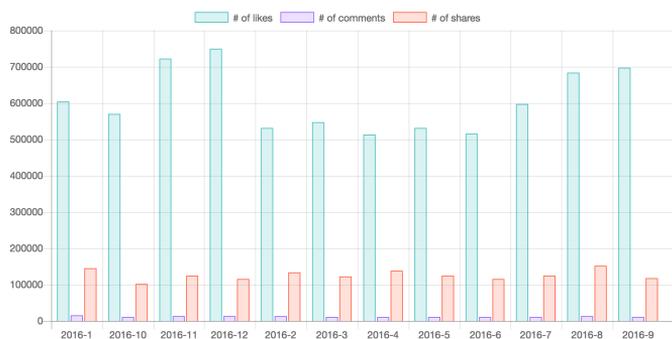


Figura 5.11: Totale parametri mensili

5.3 Utenti attivi e metriche di engagement

I KPI (Key Performance Indicator) sono una serie di metriche utili per misurare le performance e l'efficienza di un'organizzazione. In questo elaborato sono stati calcolati due indicatori chiave di performance: l'engagement e l'engagement normalizzato.

I tre parametri "likes", "commenti", "condivisioni" vengono moltiplicati per tre pesi differenti, in base al tipo di impegno che quell'azione implica, ovviamente il commento richiede il maggior "impegno" e quindi viene moltiplicato per il peso massimo di 0,5, mentre il "mi piace" viene messo con il minimo sforzo e quindi viene moltiplicato con il peso minore, cioè 0,2. Le due formule sono simili, con l'unica differenza che nell'engagement normalizzato non si divide solamente per il numero dei post, ma ad essi viene moltiplicato il numero dei fan di tale pagina.

Engagement

$$\frac{\#likes * (0.2) + \#comments * (0.3) + \#shares * (0.5)}{\#Posts}$$

Figura 5.12: Formula engagement

Engagement Normalizzato

$$\frac{\#likes * (0.2) + \#comments * (0.3) + \#shares * (0.5)}{\#Posts * \text{Fan count}}$$

Figura 5.13: Formula engagement normalizzato

Per calcolare gli engagement mensili di ogni regione, è stato sviluppato un algoritmo che permette di prelevare i dati necessari allo scopo, mese per mese, dalle tabelle di ogni regione e una volta calcolati gli engagement e gli engagement normalizzati sono stati inseriti nella tabella "Metriche2016".

```
#Creo un ciclo for per ciclare l'array dei mesi
for i in range(12):
    #Restituisce l'oggetto dato dal grafo come dizionario
    profile = graph.get_object(page[region[x]])

    #Mi ritornano il numero totale dei fan della pagina
    fan_count = profile['likes']
    print(fan_count)

    month = str(i+1)
    key = key+1
    sql = "SELECT SUM(likes_count) FROM "+region[x]+"Data WHERE data_creation BETWEEN '2016-"+month+"-01' AND '2016-"+month+"-31'"
    sql1 = "SELECT SUM(comments_count) FROM "+region[x]+"Data WHERE data_creation BETWEEN '2016-"+month+"-01' AND '2016-"+month+"-31'"
    sql2 = "SELECT SUM(shares_count) FROM "+region[x]+"Data WHERE data_creation BETWEEN '2016-"+month+"-01' AND '2016-"+month+"-31'"
    sql3 = "SELECT COUNT(id_post) FROM "+region[x]+"Data WHERE data_creation BETWEEN '2016-"+month+"-01' AND '2016-"+month+"-31'"
    m = "2016-"+month
```

Figura 5.14: Algoritmo calcolo metriche

```
#Preparo la query per inserire un nuovo elemento
query = "INSERT INTO Metriche2016(id, month, \
    tot_likes, tot_comments, tot_shares, tot_posts, engagement, engNorm, region) \
    VALUES (%d, '%s', %d, %d, %d, %d, '%s', '%s', '%s')" % \
    (key, m, tot_likes, tot_comments, tot_shares, tot_posts, round(engagement,3), round(engNorm,9), region[x])
```

Figura 5.15: Algoritmo calcolo metriche

Sono stati, quindi, calcolati gli engagement per ogni regione nei 12 mesi dell'anno 2016. Nel grafico viene riportato l'engagement annuale della regione Veneto, la regione con più alto engagement medio.

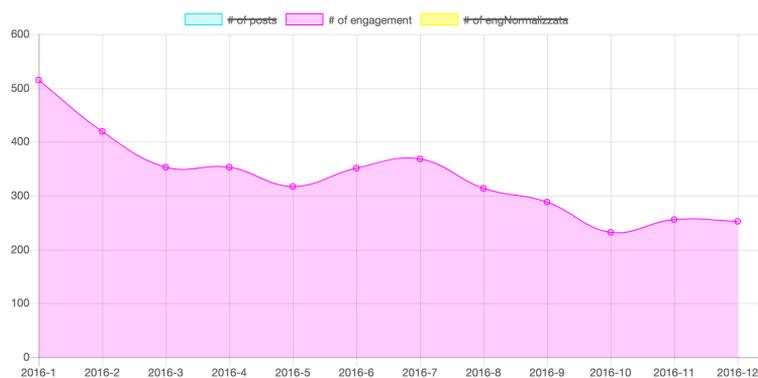


Figura 5.16: Engagement annuale Veneto

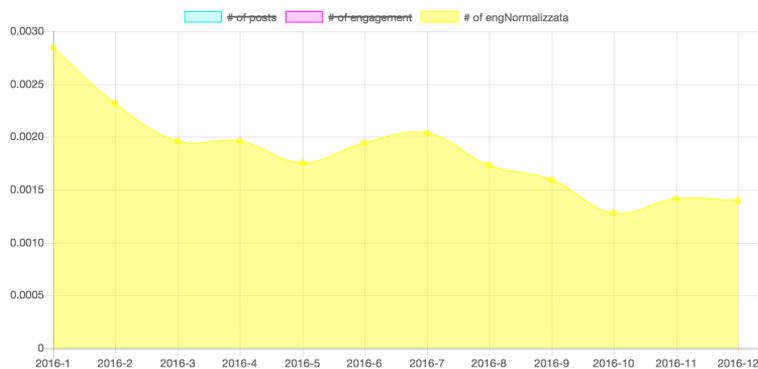


Figura 5.17: Engagement normalizzato annuale Veneto

Inoltre, è stata implementata, a livello grafico, la possibilità di confrontare l'engagement e l'engagement normalizzato di due regioni a scelta. Il grafico sottostante rappresenta il confronto fra la regione Toscana e la regione Veneto, non sono state scelte a caso, ma sono le regioni con l'engagement annuale maggiore rispetto a tutte le altre. Si può notare come l'asse delle Y cambi drasticamente.

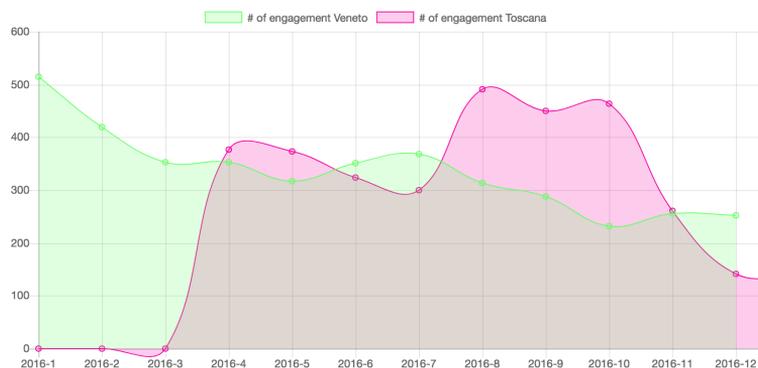


Figura 5.18: Confronto engagement

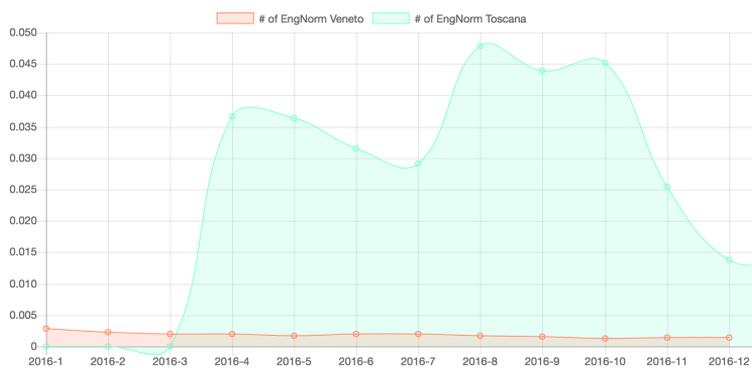


Figura 5.19: Confronto engagement normalizzato

5.3.1 Engagement medio

Per una migliore analisi sono stati calcolati gli engagement medi secondo due parametri: i mesi e le regioni. Gli engagement medi secondo i mesi prevedono la media di tutti gli engagement di ogni regione per ogni mese, mentre se prendiamo come riferimento le regione, l'engagement medio è la media degli engagement mensili di ogni regione.

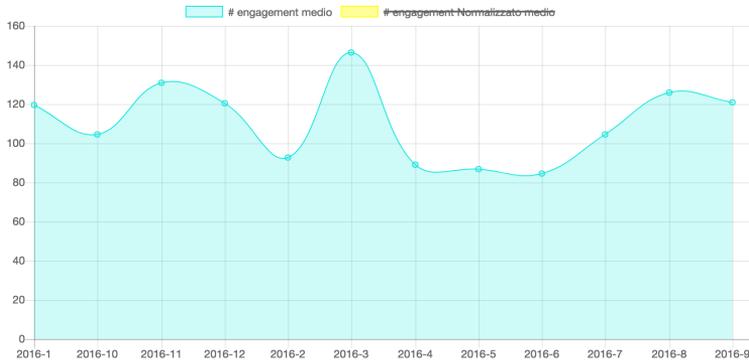


Figura 5.20: Engagement medio mensile

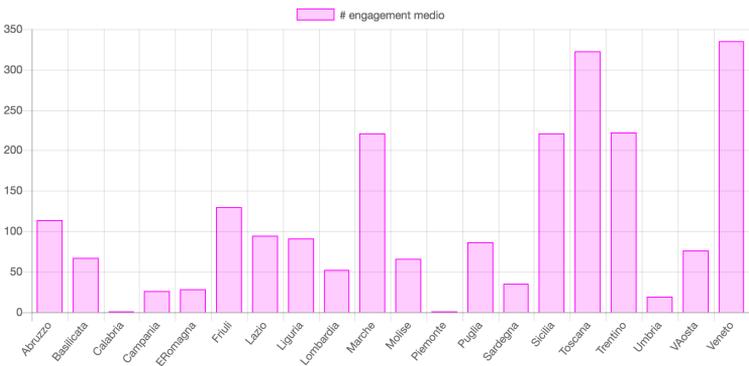


Figura 5.21: Engagement medio annuale di ogni regione

I grafici sottostanti, invece, rappresentano gli engagement normalizzati medi, sempre secondo i due parametri.

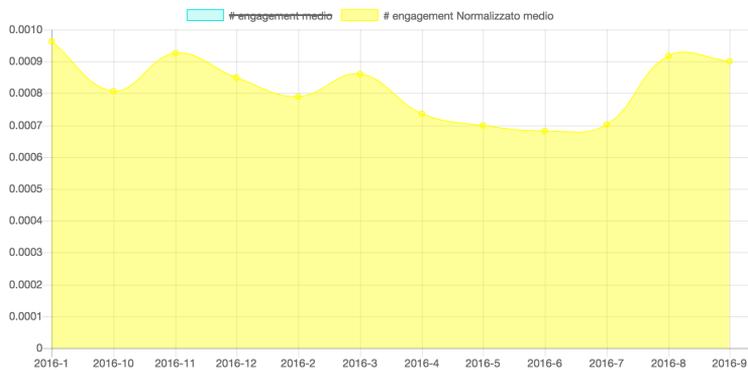


Figura 5.22: Engagement normalizzato medio mensile

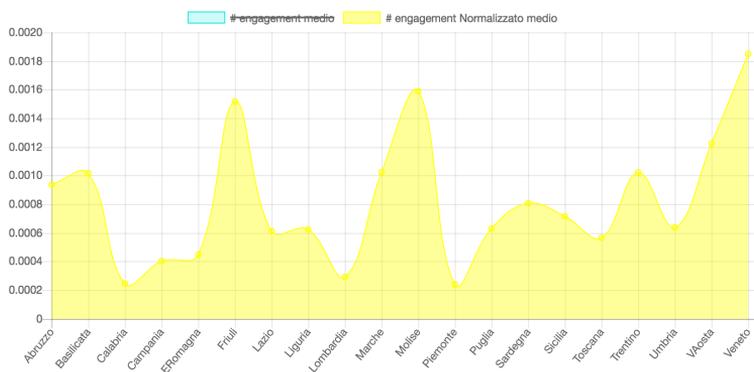


Figura 5.23: Engagement normalizzato medio annuale di ogni regione

5.3.2 Best Practice del 2016

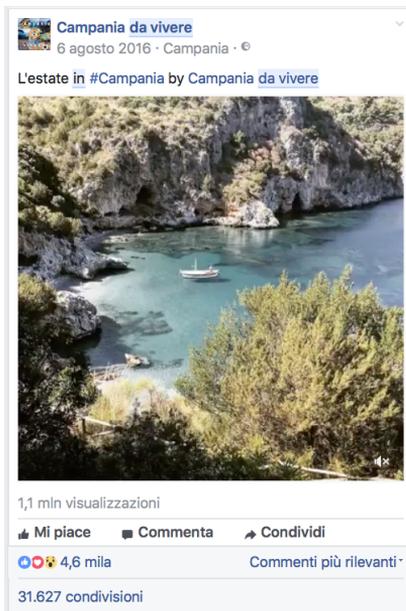
Si è voluto riportare i 3 post che hanno ricevuto rispettivamente: più mi piace, più commenti e più condivisioni, fra tutti i post pubblicati dalle varie regioni.

La figura sottostante riporta il caso di post con maggior numero di commenti.



Figura 5.24: Post con più commenti

Seguono le immagini con il maggior numero di "mi piace" e condivisioni.



(a) Post con più condivisioni



(b) Post con più "mi piace"

Figura 5.25: Best practice 2016

Si può notare come tutti e tre i post siano dei contenuti multimediali, per la precisione un video, una foto, e nel caso del post con più commenti, un video in diretta; consiste nel visualizzare un video nel momento preciso in cui viene registrato, questo permette di lasciare commenti in tempo reale. Una novità introdotta da meno di due anni che ha comportato un aumento

notevole di iterazioni in real time tra gli utenti.

5.4 Grafici per tipo di post

Un'altro fattore da non sottovalutare per adottare le giuste strategie, è capire a quale tipo di post l'utente presta più attenzione. E' già stato fatto presente che gli user sono più attratti dai contenuti multimediale, ma nello specifico a quale tra i tanti? Si è voluto analizzare i tipi di post più pubblicati durante l'anno 2016 da parte delle DMOs regionali italiane.

Dati annuali

Dati annuali secondo il tipo di post.

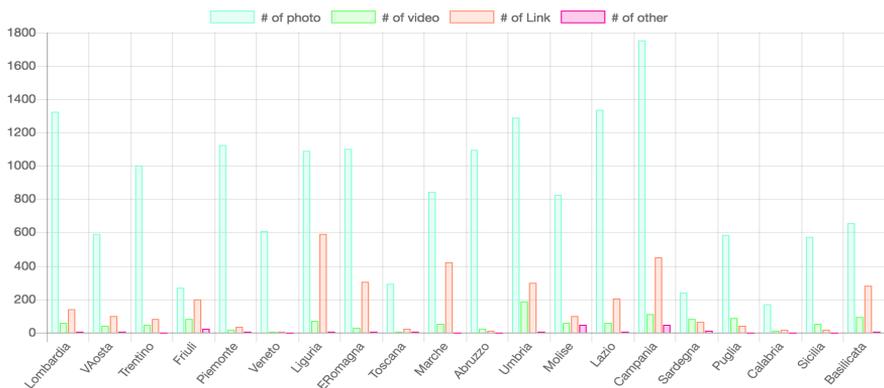


Figura 5.26: Post con più commenti

Conclusioni

Il contributo che vuole dare la mia tesi è scoprire come le DMOs utilizzano le piattaforme social media per trarne dei vantaggi nelle promozioni turistiche future. In particolare si è voluto fornire un tool che permettesse di estrarre, da tali piattaforme, i dati utili per effettuare poi le relative analisi. Il tool è stato pensato per far sì che possa essere utilizzato da qualsiasi utente, infatti fornisce una semplice interfaccia grafica. L'analisi dei dati fa riferimento a 20 pagine fan di Facebook relative alle regioni italiane. Lo studio si basa principalmente sulle reazioni ai post ottenute nell'anno 2016, per la precisione sono stati reperiti 21458 post.

Il primo punto da verificare riguardava l'aumento dell'utilizzo della piattaforma Facebook da parte delle DMOs, come è stato riportato nello studio del 2013 i post totali pubblicati dalle DMOs sono 18 046, mentre nell'anno 2016 ne sono stati pubblicati 21 458. Già da questi primi dati possiamo dedurre che c'è stato un aumento dell'utilizzo della piattaforma da parte delle DMOs. Un aspetto emerso dall'analisi è che ci sono alcuni periodi dell'anno in cui si pubblicano più post e altri meno, ciò varia da regione a regione. Se prendiamo per esempio le regioni del Sud nel mese di Agosto pubblicano molti meno post rispetto alle regioni del Nord, viceversa nel mese di Gennaio. Da quest'analisi si può dedurre che le località di mare nei mesi estivi necessitano di meno post per coinvolgere gli utenti, mentre i paesi ospitanti strutture sciistiche necessitano di meno pubblicità, per coinvolgere il turista, nei mesi in cui è possibile andare a sciare.

Già da questi dati si introduce un'altro argomento di analisi, cioè il divario tra Nord e Sud. Mentre lo studio del 2013 affermava che le piccole regioni del Sud non avevano molta attività nelle loro pagine Facebook, solitamente non ne possedevano nemmeno una, in questo elaborato possiamo vedere che le regioni come Molise e Campania abbiano davvero un gran movimento all'interno delle loro pagine fan. Solo la Calabria e la Sardegna detengono un numero nettamente inferiore di post pubblicati rispetto alle altre regioni, mentre la regione con maggior numero di post pubblicati

nell'anno 2016 è l'Abruzzo. Questi risultati indicano come le DMOs italiane hanno preso in seria considerazione l'utilizzo delle piattaforme social media, Facebook in questo caso specifico, ma anche altre come Twitter, Flickr e YouTube, come strumenti di promozione turistica efficiente.

Dietro la gestione delle pagine fan ci sono dei veri e propri team, come ad esempio in Trentino Alto Adige si occupa il *Trentino Marketing S.r.l.* della pagina "Visit Trentino", per le Marche c'è il *Social Media Team Marche*, questo fa capire quanto le regioni siano state portate nel corso del tempo ad investire denaro anche nella promozione turistica mediante social media. Anni fa non veniva tenuta in considerazione come ora, si prediligevano i siti web ufficiali e si faceva a meno della pagina Facebook.

Passiamo ora ad analizzare le reazioni degli utenti, quindi quanto il lavoro effettuato dai gestori delle pagine fan sia davvero efficiente nel coinvolgere gli utenti. Lo studio del 2011 mostrava le tecniche promozionali che attiravano di più la clientela, ovviamente la pubblicità è al primo posto, ma ci sono diversi modi di fare pubblicità come ad esempio attraverso contest e lotterie, attraverso foto panoramiche e tour virtuali ecc.. . Da questo elaborato è emerso che i post con più engagement sono i contenuti visivi e multimediali. Nello specifico video, foto e una nuova tipologia di contenuto multimediale: il video in diretta. Quest'ultimo permette all'utente di commentare il video nel momento stesso in cui viene registrato, quindi chi pubblica il video ha la possibilità di leggere in tempo reale le opinioni di chi lo sta guardando. Questo dà all'utente un senso di vicinanza alla pagina, perché commenta con la convinzione che qualcuno dall'altra parte legga di sicuro e che quindi potrà anche rispondergli nel giro di poco tempo o, addirittura, direttamente dal video che sta mandando in diretta.

L'approccio top-down si sta un po' smussando e si tenta anche un approccio bottom-up, ovvero si tengono molto in considerazione le opinioni e i contenuti generati dagli utenti. Lo dimostrano le tempestive risposte degli amministratori delle pagine ai commenti degli utenti, non solo negativi, ma anche ai commenti positivi. E' fondamentale rispondere ai

commenti negativi, perché permette alla pagina di spiegarsi e di dare una buona impressione, non ignorando tali commenti. Se l'utente riceve una risposta esauriente alla sua "lamentela" o al suo dubbio, può sempre generare un passaparola positivo a riguardo. Allo stesso tempo è importante anche rispondere ai commenti positivi, fanno sì che la pagina dimostri davvero dell'interessamento verso i suoi seguaci sotto ogni aspetto.

Il limite di questa analisi, sta nell'aver analizzato solamente le pagine Facebook delle DMOs italiane e prendendo un solo anno di riferimento: il 2016. Questo perché è stata svolta un'analisi di confronto per evidenziare le differenze rispetto agli studi precedenti. Un'analisi più approfondita si potrebbe fare prendendo come campione di dati non solo quello relativo all'Italia ma di tutta l'Europa, o addirittura di tutto il mondo. Oppure si potrebbe replicare lo studio per un'altra nazione e mettere a confronto i risultati ottenuti. In ogni caso il tool è generico, da quindi la possibilità di estrarre dati e ripetere le analisi su qualsiasi finestra temporale si voglia.

Un altro grande limite è dovuto dalle API Graph, che non permette di prelevare più di un certo numero di commenti, post, like alla volta, quindi la raccolta di un gran numero di dati richiede un notevole lasso di tempo. Inoltre, andrebbe implementato il metodo di richiesta automatica della chiave di sessione per l'estrapolazioni dei dati che richiedono diverso tempo, altrimenti la chiave di sessione scade e il processo di raccolta di dati s'interrompe.

Questa tesi potrebbe essere sviluppata per molteplici scopi, come ad esempio si può effettuare un'analisi che stabilisce in quale periodo della giornata gli utenti prestano più attenzione ai contenuti pubblicati. Oppure ampliare la ricerca su più anni di riferimento, per esempio gli ultimi 5 anni, per vedere i cambiamenti avvenuti e gli eventuali progressi nel corso degli anni. Infine, si potrebbe estendere la ricerca non solo nell'ambito turistico, ma in qualsiasi ambito, ossia generalizzare la fase di estrazione dati, dando all'utente la possibilità di indicare profili Facebook non necessariamente legati al contesto della promozione turistica regionale.

Bibliografia

- [1] *Data Mining*
Di Michael J. A. Berry, Gordon S. Linoff
APOGEO Editore, 2001
- [2] *Application of Data Mining In Marketing*
Radhakrishnan B, Shineraj G, Anver Muhammed K.M
International Journal of Computer Science and Network, Volume 2,
Issue 5, October 2013
- [3] *The big data e la conoscenza nella societa' del web 2.0.*
Andrea Parafioriti
Roma Tre, 2014
- [4] *Data Mining Techniques*
Gordon S. Linoff, Michael J. A. Berry
Wiley Publishing Inc. 2011
- [5] *A Formal definition of Big Data based on its essential Features.*
Andrea De Mauro, Marco Greco e Michele Grimaldi
Emerald Group Publishing Limited, 2016
- [6] *BIG DATA: Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati.*
Alessandro Rezzani, APOGEO, 2013

- [7] *3D Data Management: Controlling Data Volume, Velocity and Variety.*
Doug Laney
Meta Group, 2001
- [8] *Database NoSql analisi prestazionale di Redis*
Gabriele Casagrande, 2014
- [9] *Destination Management Organization. Ruolo, organizzazione ed indicatori di performance.*
Angelo Presenza
Franco Angeli, 2007
- [10] *Social media marketing per il turismo. Come costruire il marketing 2.0 e gestire la reputazione della destinazione.*
Josep Ejarque
HOEPLI, 2015
- [11] *The virtual community: Home standing on the eletronic fronter.*
Rheingold H.,
Reading, MA: Addison-Wesley., 1993
- [12] *Internet marketing communication of tourist destinations: review of DMO website in Europe.*
Iva Silvar, Danijela Krizman Pavlovic
Economic Research, Spec. Edition (2012)
- [13] *Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations.*
Marcello M. Mariani, Marco Di Felice, Matteo Mura
- [14] *The 2014 Traveler?s Road to Decision* Ipsos MediaCT, Google, 2014
- [15] *Facebook marketing. Comunicare e vendere con il social network n.1.*
Luca Conti, Cristiano Carriero
HOELPI, 2014

- [16] *Social media e comunicazione di marketing. Presidiare la Rete, costruire relazioni e acquisire clienti con strumenti del web 2.0.*
Alessandro Pruntesti
Franco Angeli, 2016
- [17] *Comportamento del consumatore.*
Luca Petruzzellis, Jean-Charles Chebat
Pearson ,2010
- [18] *La divulgazione dei contenuti e la promozione online in ambito turistico.*
Matteo Venturini
Primiceri Editore, 2013
- [19] *Destination Management: Un nuovo approccio al turismo.*
Enrico Marchesin, 2014
- [20] *Documentazioni ufficiali Facebook API Graph*
<https://developers.facebook.com/docs/graph-api>
- [21] <http://www.html.it/>.
- [22] *php.net. Che cos'è il PHP?. URL <http://php.net/manual/it/intro-what-is.php>.*

Ringraziamenti

Ringrazio infinitamente il professor Di Felice per la sua pazienza e la sua disponibilità, ma soprattutto per tutto il tempo che ha dedicato al mio lavoro. Grazie.

Ringrazio la mia famiglia, grazie a loro ho realizzato uno dei più importanti obiettivi della mia vita. Grazie di cuore