



Corso di laurea magistrale in Marketing Analysis and Metrics

Cattedra Customer Intelligence and Big Bata

***GEOSTYLE: ARTIFICIAL INTELLIGENCE FOR DISCOVERING
FASHION TRENDS IN SOCIAL MEDIA DATA FOR FASHION
LUXURY INDUSTRY***

Prof. Marina Paolanti

RELATORE

Prof. Villarroel Ordenes Francisco Javier

CORRELATORE

Giordana Andriani – Matr. 750521

CANDIDATO

ANNO ACCADEMICO 2023/2024

INDICE

Introduzione.....	3
Capitolo 1 – La moda incontra l’intelligenza artificiale.....	5
1.1 La digitalizzazione: tra sfide e innovazioni tecnologiche	5
1.2 L’intelligenza artificiale e il marketing della moda.....	15
Capitolo 2 – Geografia, Moda e Lusso.....	25
2.1 L'influenza geografica sulle tendenze globali	25
2.2 Il potere della geolocalizzazione	31
2.3 La Geografia della Moda	36
2.4 Il settore del Fashion Luxury	40
2.5 Obiettivo di ricerca	41
Capitolo 3 – Materiali e Metodi	42
3.1 Strumenti	42
3.2 Il Dataset.....	45
3.3 Metodologia.....	47
3.3.1 Data Pre-Processing.....	47
3.3.2 Data Exploration.....	50
3.3.3 Data Mining	52
Capitolo 4 – Risultati e Discussioni	57
4.1 Performance e risultati.....	57
4.1.1 Risultati di Clustering su Scala Mondiale	57
4.1.2 Profili di Consumo: Caratteristiche distintive delle comunità online di moda.....	61
4.2 Discussione e implicazioni di Marketing	69
4.3 Limitazioni e sviluppi futuri	74
Conclusione	76
Bibliografia.....	79
Appendice.....	84

GEOSTYLE: ARTIFICIAL INTELLIGENCE FOR DISCOVERING FASHION TRENDS IN SOCIAL MEDIA DATA FOR FASHION LUXURY INDUSTRY

INTRODUZIONE

Nell'epoca dell'Industria 4.0, caratterizzata dalla crescente digitalizzazione e dall'ampia adozione di tecnologie avanzate, il settore della moda ha subito profonde trasformazioni. L'introduzione dell'Intelligenza Artificiale (IA) e la capacità di acquisire e analizzare grandi quantità di dati (Big Data) hanno aperto nuove prospettive per le aziende, che stanno rivoluzionando i loro approcci operativi e la loro interazione con i clienti. Questo fenomeno è visto come un'opportunità unica per comprendere più a fondo i comportamenti dei consumatori e offrire esperienze di shopping altamente personalizzate.

Questa trasformazione, accelerata da fattori come la globalizzazione, la crescente digitalizzazione e gli impatti della pandemia, ha posto l'attenzione sulle nuove dinamiche di consumo e sulle strategie adottate dalle imprese per rimanere competitive in un mercato in continua evoluzione.

Questa tesi si propone di esplorare in dettaglio l'impatto dell'IA e della geolocalizzazione nel settore della moda e delle borse, con un focus particolare sull'analisi dei comportamenti dei consumatori in diverse regioni del mondo. L'elaborato è strutturato in quattro capitoli con l'obiettivo di analizzare al meglio il contesto, gli studi condotti fino a oggi e utilizzare i dati disponibili per comprendere come il marketing possa adattare le proprie decisioni strategiche e operative in maniera diversificata ed efficiente.

In particolare, il primo capitolo si concentra sull'evoluzione del settore della moda in un contesto di digitalizzazione e crescente utilizzo di tecnologie avanzate. Esplora come l'introduzione dell'intelligenza artificiale e della geolocalizzazione stiano influenzando la produzione e il marketing dei prodotti di moda.

Il secondo capitolo consiste in una revisione della letteratura riguardo al tema della geolocalizzazione, esaminando attentamente gli studi precedentemente condotti in questo campo e considerando come i fattori culturali e geografici influenzano i comportamenti dei consumatori. L'obiettivo principale è di creare una base solida su cui costruire lo studio in questione, e allo stesso tempo, individuare le lacune o aree di ricerca che non sono state ancora affrontate in modo esaustivo. Questo consentirà di proporre nuove direzioni per l'analisi e di formulare la domanda di ricerca.

Nel terzo capitolo, vengono descritti i dettagli dell'approccio metodologico adottato per condurre questa ricerca, inclusi il dataset utilizzato, le varie fasi affrontate e gli algoritmi di Intelligenza Artificiale selezionati per comprendere il comportamento dei consumatori su scala globale.

Infine, nell'ultimo capitolo di questo elaborato, è analizzata la performance dell'analisi condotta, seguita da una valutazione delle implicazioni manageriali e di marketing che ne conseguono. Questo consentirà di trarre conclusioni significative sulla relazione esistente tra fattori culturali, geografici e comportamentali dei consumatori nel settore della moda e fornirà una visione chiara delle diverse caratteristiche dei consumatori in tutto il mondo.

In conclusione, questo studio fornisce un'importante analisi delle potenziali opportunità e sfide che le aziende del settore possono affrontare nel mercato globale.

Capitolo 1 – La moda incontra l'intelligenza artificiale

1.1 La digitalizzazione: tra sfide e innovazioni tecnologiche

Le radici della digitalizzazione nell'economia risalgono intorno agli anni '90, quando si sono aperte le porte alla terza fase della rivoluzione industriale caratterizzata dall'emergere di Internet, dall'aumento dell'automazione attraverso l'uso di computer ed elettronica, dalla scoperta dell'energia nucleare e, più in generale, dall' "automazione dei processi industriali". In quest'epoca arrivarono anche i progressi nelle telecomunicazioni e questo aprì la strada alla globalizzazione (Heriel, 2021). La nascita della digitalizzazione ha, fin da subito, trasformato la struttura economica e industriale di tutti i paesi (Schwab, 2017) e la sua rapida evoluzione ha portato, oggi giorno, a parlare di una quarta fase della rivoluzione industriale (o Industria 4.0), segnata da trasformazioni digitali, dispositivi personali (anche Wi-Fi personale), analisi dei dati e tecnologie AI (Yuan et al, 2021). In un contesto del genere, risulta necessario distinguere due accezioni differenti del termine digitalizzazione, in inglese "digitization" and "digitalization". Per *digitization*, si fa riferimento alla pura conversione dall' analogico al digitale di dati e documenti esistenti; i dati stessi non vengono modificati, sono semplicemente codificati in un formato digitale. Invece, come la definisce Gartner, per *digitalization* si intende << l'uso delle tecnologie digitali per modificare un modello di business e fornire nuove opportunità di guadagno e di produzione di valore >> (Gartner). In poche parole, è il processo di passaggio a un'attività digitale. Quindi, la "digitalization" va oltre la "digitization", sfruttando la tecnologia dell'informazione digitale per trasformare completamente i processi di un'azienda, valutando, riprogettando e re immaginando il modo in cui si fanno gli affari (McKinsey & Company, 2021). La digitalizzazione ha avuto un impatto significativo su numerosi settori, tra cui business, sanità, istruzione, intrattenimento e altro ancora (Angevine C. et al, 2021). Con l'avanzare delle tecnologie digitali, sono emerse nuove innovazioni come smartphone, cloud computing¹, analisi dei big data, intelligenza artificiale e Internet of Things (IoT)² che hanno contribuito alla crescita esponenziale di questo processo consentendo una maggiore connettività, capacità di elaborazione dati e automazione. In breve tempo la digitalizzazione è

¹ Cloud Computing: funzionalità digitali complete fornite tramite Internet per consentire alle organizzazioni di operare, innovare e servire i clienti (Mckinsey & Company, 2022).

² Internet of Things: oggetti fisici incorporati con sensori e attuatori che comunicano con i sistemi informatici tramite reti cablate o wireless, consentendo al mondo fisico di essere monitorato digitalmente o addirittura controllato (Mckinsey & Company, 2022).

diventata parte integrante della vita quotidiana dell'uomo, trasformando il modo in cui lavoriamo, comunichiamo e interagiamo con il mondo che ci circonda (Heriel, 2021). Inoltre, ha rivoluzionato le industrie, creato nuovi modelli di business e aperto nuove possibilità di innovazione e di crescita (Angevine C. et al, 2021).

A contribuire alla rapida evoluzione di questo processo e di conseguenza all'introduzione in questa nuova fase della rivoluzione industriale vi è sicuramente la pandemia Covid-19, che ha messo a dura prova le aziende in vari settori di attività (McKinsey & Company, 2022). Molte di queste organizzazioni sono state costrette ad adottare nuove pratiche di lavoro interne e nuove modalità nell'offrire i propri prodotti/servizi ai clienti (Almaida et al, 2020). Allo stesso tempo è risultato necessario ridisegnare i modelli di gestione e collaborazione per far sì che nessuno all'interno delle organizzazioni fosse lasciato indietro o si sentisse escluso da questo processo. Questa accelerazione può essere osservata in varie dimensioni di business, che, spinte dalla necessità urgente di adattarsi alle sfide poste dalla pandemia hanno dovuto apportare diversi cambiamenti strutturali e metodologici (McKinsey & Company, 2022). In primo luogo, l'improvviso passaggio al lavoro a distanza ha richiesto l'adozione di strumenti e di tecnologie digitali per consentire ai dipendenti di lavorare efficacemente da casa. Le aziende hanno implementato in maniera rapida piattaforme di videoconferenza, software di project management e sistemi di condivisione di documenti basati su cloud per facilitare la collaborazione virtuale e mantenere la produttività. Questa transizione verso il lavoro a distanza non solo ha richiesto l'infrastruttura tecnologica, ma ha anche spinto la creazione di canali di comunicazione digitali, promuovendo una cultura della collaborazione online e del lavoro di squadra a distanza (Valaskova et al, 2021).

Inoltre, la pandemia ha evidenziato l'importanza di digitalizzare i processi aziendali tradizionali; con gli uffici fisici chiusi e con l'esigenza di adottare misure di precauzione e sicurezza, le aziende sono state obbligate a digitalizzare anche le attività amministrative, ad implementare soluzioni di cloud storage³ e ad automatizzare i flussi di lavoro. Questo cambiamento ha aumentato l'efficienza operativa e ridotto le necessità di interazioni fisiche e di documenti cartacei (Valaskova et al, 2021).

Un'altra tendenza importante che ha caratterizzato questo periodo è l'utilizzo degli e-commerce da parte dei clienti e l'ingente investimento delle aziende in questo canale di vendita. Infatti, con i negozi fisici che affrontano restrizioni e i consumatori che cercano alternative più sicure, le aziende hanno rapidamente ampliato la loro presenza online e le loro capacità di e-commerce

³ Cloud Storage: Sistema di archiviazione di dati non strutturati (TechTarget).

creando piattaforme online, ottimizzando le supply chain per la vendita e la consegna online, infine, migliorando la Customer Experience attraverso interazioni online personalizzate. Questo passaggio ai canali di vendita digitali ha aperto nuovi mercati, ampliato la portata dei clienti e creato opportunità di crescita, trasformando il modo in cui le aziende conducono il business (McKinsey & Company, 2020). Anche in termini di coinvolgimento dei clienti si è assistito a una rivoluzione digitale; la necessità di ridurre e contenere le interazioni di persona, ha portato le aziende a rafforzare i loro canali digitali per interagire con i clienti in modo efficace. Inoltre, le aziende si trovavano di fronte ad un consumatore diverso, con nuovi bisogni e nuove necessità: un consumatore vulnerabile, preoccupato, bisognoso di sicurezza e di vicinanza. Le aziende hanno dovuto modificare, perciò, le proprie strategie di coinvolgimento per adattare a questo nuovo contesto e hanno dovuto ricercare metodi sempre più innovativi di engagement per non perdere i propri clienti e per acquisirne di nuovi, sostituendo così l'interazione umana (Harvard Business School, 2022).

Affrontare queste sfide poste dal Covid 19 è stato reso possibile, in gran parte, grazie alla nascita e allo sviluppo dell'intelligenza artificiale. Infatti, il coinvolgimento dei clienti e l'intelligenza artificiale, ad esempio, sono strettamente intrecciati nell'era digitale, in quanto le tecnologie basate sull'IA offrono modi innovativi per le aziende di interagire con i propri clienti (McKinsey & Company, 2020). Con il termine intelligenza artificiale si intende la capacità di una macchina di eseguire le funzioni cognitive tipiche delle menti umane, come percepire, ragionare, apprendere, interagire con un ambiente, risolvere problemi e persino esercitare la creatività (McKinsey & Company, 2023). L'origine dell'intelligenza artificiale (IA) risale alla metà del XX secolo, quando scienziati e ricercatori iniziarono a esplorare il concetto di creare macchine in grado di imitare l'intelligenza umana e replicare, in questo modo, i comportamenti intelligenti tipici dell'uomo. Il termine "intelligenza artificiale" fu coniato durante un workshop tenutosi al Dartmouth College nel 1956, dove scienziati di spicco come Marvin Minsky, John McCarthy e Claude Shannon si riunirono per discutere le possibilità di sviluppare questo tipo di macchine e si arrivò, così, alla nascita di molti dei sottocampi che costituiscono le fondamenta dei moderni sistemi di intelligenza artificiale e, in particolare, delle macchine di apprendimento (Chalmers et al, 2020). Nei decenni successivi, il campo dell'IA ha vissuto periodi di espansione e di contrazione che hanno portato a sviluppi sporadici (Haenlein & Kaplan, 2019). Si definisce "Inverno dell'IA", il periodo in cui i progressi erano lenti, i finanziamenti diminuiti e le aspettative sulle capacità dell'IA erano irrealistiche. Tuttavia, negli anni '90 si è assistito a una rinascita della ricerca sull'IA con l'affermarsi dell'apprendimento

automatico, un sottocampo che si concentra nello sviluppo di algoritmi in grado di apprendere modelli dai dati. In questa fase, algoritmi come le reti neurali e gli alberi decisionali, hanno guadagnato importanza. L'interesse e gli studi in questo ambito sono cresciuti ancor di più negli anni successivi grazie alla disponibilità dei big data e l'aumento della potenza di calcolo che hanno spinto i progressi dell'IA (Brock & von Wangenheim, 2019). In questi stessi anni è nato ad esempio, l'apprendimento profondo, o Deep Learning (DL), un ramo dell'apprendimento automatico che utilizza reti neurali a più livelli e che ha rivoluzionato l'impiego dell'IA in settori come il riconoscimento delle immagini, l'elaborazione del linguaggio naturale e la guida autonoma (Chalmers et al, 2020). Negli ultimi anni, l'IA è diventata sempre più integrata nella nostra vita quotidiana e i rapidi progressi nelle tecniche di machine learning statistico hanno ampliato la portata delle applicazioni di intelligenza artificiale e gli usi commerciali che ora abbracciano diverse aree come il marketing (Davenport et al., 2019), la scoperta di molecole (Gawehn et al., 2016), la produzione automobilistica (Luckow et al., 2018) oltre che diversi campi tra cui sanità, finanza, trasporti, robotica e assistenti virtuali.

Il mondo degli affari è certamente uno dei campi in cui l'intelligenza artificiale è ampiamente e maggiormente utilizzata (Chalmers et al, 2020). Le aziende hanno imparato ad impiegare queste tecniche in vari modi per aumentare la produttività, risparmiare tempo e ridurre i costi. In particolare, le aziende multinazionali di successo utilizzano funzionalità di IA come l'automazione, l'analisi dei big data e l'elaborazione del linguaggio naturale per ottenere informazioni sulla propria attività e renderla più efficiente e pertinente per la propria base di clienti (Yuan et al, 2021), mentre aziende emergenti stanno registrando una rapida crescita e un ampio successo con strategie e piani adeguati grazie all'introduzione di tecniche di questo tipo (Haenlein & Kaplan).

Alcuni dei vantaggi e degli ambiti in cui l'IA rappresenta un ottimo strumento nelle aziende riguarda sicuramente il servizio clienti e assistenza: grazie all'inserimento di chatbot⁴ e di assistenti virtuali le aziende sono in grado di fornire un'assistenza ai clienti istantanea e personalizzata, rispondendo alle loro domande, fornendo loro informazioni di qualsiasi tipo e risolvendo problemi (Forbes, 2022). Ciò, oltre a ottimizzare i servizi offerti dalle aziende, garantisce una migliore Customer Experience (Yuan et al, 2021). Inoltre, l'intelligenza artificiale migliora la gestione della supply chain e della logistica, ottimizzando i livelli di inventario, prevedendo le fluttuazioni della domanda e ottimizzando i percorsi logistici (Forbes,

⁴ Chatbot: software che simula ed elabora le conversazioni umane (scritte o parlate), consentendo agli utenti di interagire con i dispositivi digitali come se stessero comunicando con una persona reale (OCI).

2022). Si parla anche di analisi predittiva in quanto l'intelligenza artificiale consente alle aziende di fare previsioni e analisi accurate sulla base di dati storici e in tempo reale: gli algoritmi di apprendimento automatico possono prevedere il comportamento dei clienti, i modelli di domanda e le tendenze del mercato, consentendo alle aziende di ottimizzare la gestione delle scorte, la pianificazione della produzione e la previsione delle vendite. In questo modo l'IA gioca un ruolo centrale, ad esempio, nel mercato azionario, uno dei mercati più popolari e aleatori a causa della sua natura dinamica, ma allo stesso tempo di grande interesse per gli investitori in quanto redditizio. L'IA, e in particolare le tecnologie di apprendimento come l'SVM⁵ e l'ANN⁶, permettono di prevedere questi mercati e offrire risultati positivi, riducendo quella imprevedibilità e quella percentuale di rischio tipico di questa realtà (Dutta, 2021).

Di notevole aiuto anche nel campo di rilevamento delle frodi, gli algoritmi di intelligenza artificiale analizzano grandi volumi di dati per rilevare anomalie, modelli e potenziali frodi in tempo reale. Si tratta di un aspetto cruciale nei servizi finanziari e nelle assicurazioni, dove i sistemi basati sull'intelligenza artificiale possono identificare le transazioni fraudolente, valutare i profili di rischio e prevenire le perdite finanziarie (Dutta, 2021).

Ma, l'intelligenza artificiale è in grado di fornire, anche, supporto decisionale a dirigenti e manager analizzando i dati, fornendo approfondimenti, formulando raccomandazioni e di conseguenza dando assistenza e supporto nella pianificazione strategica, nell'allocazione delle risorse, nell'analisi degli scenari e nel prendere decisioni informate e consapevoli (Forbes, 2022).

In sostanza, l'intelligenza artificiale non solo è diventato strumento chiave nella vita di tutti noi, ma è anche diventato un mezzo fondamentale per le aziende nel guidare e svolgere il proprio business (McKinsey & Company, 2023). Non a caso, un sondaggio condotto da Forbes mostra che la maggior parte degli imprenditori ritiene che l'intelligenza artificiale andrà a vantaggio delle loro attività; il 64% degli intervistati prevede che l'IA migliorerà le relazioni con i clienti e aumenterà la produttività, mentre il 60% si aspetta che l'IA guidi la crescita delle vendite. Per quanto riguarda le molteplici modalità di utilizzo dell'IA, il 44% degli intervistati la percepisce

⁵ Support Vector Machines (SVM): Algoritmo di deep learning che esegue l'apprendimento supervisionato per la classificazione o la regressione di gruppi di dati (Scikit Learn).

⁶ Artificial Neural Network (ANN): Algoritmo di deep learning che cerca di riconoscere le relazioni sottostanti in un insieme di dati attraverso un processo che imita il modo in cui opera il cervello umano (Scikit Learn).

come una risorsa per migliorare il processo decisionale, il 53% come risorsa per ridurre i tempi di risposta ed il 48% come risorsa per evitare errori. Le aziende si aspettano inoltre che l'intelligenza artificiale le aiuti a risparmiare sui costi (59%) e a semplificare i processi di lavoro (42%) (Forbes, 2023).

Come visto, l'intelligenza artificiale rappresenta un mezzo straordinario, che può portare a grandi risultati e vantaggi, non solo alle aziende per lo svolgimento del loro business, ma più in generale nella vita di tutti i giorni di una persona; tuttavia, non si potrebbe godere dei benefici che questa genera se non vi fosse la possibilità, oggi, di entrare in contatto con una enorme quantità di dati. Infatti, i big data svolgono un ruolo cruciale nell'efficacia e nel successo delle applicazioni di intelligenza artificiale (Yuan et al, 2021).

Nell'era digitale odierna, i grandi sviluppi di internet, i social media, i dispositivi mobili, i sensori e moltissimi altri strumenti tecnologici che si utilizzano giornalmente permettono di acquisire un'ampia quantità di informazioni. Si è sviluppato, così, il concetto di big data, un termine che si riferisce a set di dati così grandi e complessi che i metodi tradizionali di elaborazione dei dati risultano inadeguati (Mikalef et al, 2018). Le ingenti masse di dati si definiscono "big data" se presentano almeno una delle tre peculiarità: volume, velocità e varietà (Piva, 2019). Il volume si riferisce alla grande mole di informazioni frutto delle numerose attività digitali che si svolgono nella vita quotidiana e che rendono impossibile la loro raccolta tramite tecnologie classiche. I dati sono in perenne crescita e il volume di questi registra incrementi continui; Secondo quanto dichiarato dall'Idc⁷ nel 2019 l'ecosfera dei dati a disposizione, a livello globale, raggiungeva i 40 Zettabyte⁸ e nel 2020 il valore era quantificato in 44 di più rispetto all'ecosfera dei dati dell'anno precedente. Inoltre, è stato stimato che entro il 2025 questo valore è destinato a quintuplicarsi (The World Economic Forum, 2019).

La velocità, invece, si riferisce alla rapidità con cui vengono originati e acquisiti questi dati. Si tratta di una velocità senza precedenti, che è il risultato della profilazione di dispositivi dotati di sensori. Le aziende devono dunque cimentarsi in nuove sfide: non solo devono raccogliere i dati, che registrano un'acquisizione sempre più veloce, ma anche analizzare i big data in real-time. L'ultimo parametro con cui è possibile identificare i Big Data è la varietà, ovvero la

⁷ International Data Corporation: Principale fornitore globale di market intelligence, servizi di consulenza ed eventi per i mercati della tecnologia dell'informazione, delle telecomunicazioni e della tecnologia di consumo (Idc).

⁸ Zettabyte: Unità di misura della quantità di informazione pari a 270 byte (Repubblica).

differente tipologia di dati disponibili, che provengono cioè da fonti eterogenee, sempre più numerose. Tali fonti in crescita possono essere sensori IoT, social media, open data, oltre ai gestionali aziendali e ai sistemi transazionali (Piva, 2019). La varietà delle fonti eterogenee era già stata stigmatizzata nel 2008, agli albori della scalata dei social network, da Chris Anderson sul magazine Wired, con l'espressione "*More isn't just more. More is different*". Significava che più non è solo più, in termini quantitativi, ma "più" vuol dire aumentare la varietà delle fonti eterogenee (Wired, 2008).

Nel tempo, con l'aumento degli studi, dell'analisi e dell'utilizzo dei dati come fonte di informazione e di arricchimento, si tende a parlare di ulteriore due elementi per definire e descrivere i cosiddetti big data, ovvero il valore e la veridicità (Piva, 2019). A volte la raccolta dei dati non è sufficiente, perché essa non assicura la possibilità di estrarre conoscenza, ma solo informazioni. I dati rappresentano un'entità, una transazione, un fenomeno o un avvenimento in maniera codificata, però l'informazione non è una rappresentazione codificata, bensì un processo di data analytics. Spesso un'informazione ha significato solo per chi trae valore dalla generazione di quei dati e a quel punto si parla di conoscenza. Quindi la conoscenza emerge quando le informazioni aiutano il processo decisionale, dunque ad assumere provvedimenti, prendendo decisioni da concretizzare in azioni. La conoscenza è frutto del processo che porta le informazioni alla messa in pratica e all'utilizzo concreto dei dati raccolti nella fase iniziale. Seguendo tale criterio è importante che quando si parla di Big Data ci si riferisce a tutti quei dati da cui si può estrarre non solo un'informazione bensì una conoscenza (Mikalef et al, 2018). Per veridicità invece, si fa riferimento all'integrità e all'equità; con l'avvento dei big data, la gestione dei dati è diventata una sfida più complessa e difficoltosa, tuttavia, fondamentale, in quanto è necessario che le informazioni estrapolate siano affidabili e rappresentative della realtà al fine di effettuare analisi attendibili e proficue (Piva, 2019).

I Big Data, in questo senso, rappresentano il carburante necessario per addestrare, perfezionare e potenziare i modelli di intelligenza artificiale. Infatti, gli algoritmi AI imparano dai dati, analizzano modelli, correlazioni e relazioni all'interno di questi per identificare approfondimenti e fare previsioni accurate (Rialti et al, 2020). Quindi, più grande è il set di dati, migliore è la capacità dell'AI di identificare modelli significativi e fare previsioni affidabili. I big data forniscono un set di informazioni ricco e diversificato che consente ai modelli di intelligenza artificiale di catturare la complessità degli scenari del mondo reale e adattarsi a un'ampia gamma di contesti. Inoltre, i modelli di intelligenza artificiale addestrati sui big data tendono ad avere una maggiore accuratezza e migliori prestazioni rispetto ai modelli addestrati su set di dati più piccoli, e aiutano a ridurre i pregiudizi, a migliorare la significatività statistica

e a comprendere meglio le complessità dei vari fattori che influenzano i risultati. Di conseguenza, le previsioni e gli approfondimenti generati dai modelli di intelligenza artificiale sono più affidabili e robusti (Mikalef et al, 2018).

Questo approccio basato sui dati aiuta le aziende a prendere decisioni informate e a massimizzare il ritorno sugli investimenti. Non a caso, le iniziative Big Data sono state valutate come "estremamente importanti" per il 93% delle aziende (The Data Analytics Report, 2022).

Nel frenetico ambiente aziendale odierno, le informazioni in tempo reale sono cruciali. Con i big data, l'IA può elaborare e analizzare i dati in tempo reale, consentendo alle aziende di rispondere rapidamente alle mutevoli condizioni del mercato e alle esigenze dei clienti. L'analisi in tempo reale basata sull'intelligenza artificiale consente di prendere decisioni tempestive, individuare interazioni personalizzate con i clienti e a realizzare campagne di marketing agili (Piva, 2019).

In conclusione, i big data risultano fondamentali per sbloccare il pieno potenziale dell'intelligenza artificiale; grazie alla combinazione di questi due concetti le aziende possono stare al passo con la concorrenza, offrire ai clienti esperienze eccezionali e cogliere le opportunità non appena si presentano (Mikalef et al, 2018).

L'AI e i Big Data rappresentano il futuro di qualsiasi settore, ma questo è ancor più vero nel campo del marketing dove si presentano come un vero e proprio fattore di svolta (Forbes, 2021). Secondo quanto riportato da Statista, nel 2021 il mercato dell'intelligenza artificiale (AI) nel marketing è stato stimato di 15,84 miliardi di dollari (Statista, 2023) e si prevede che il valore aumenterà a oltre 107,5 miliardi di dollari entro il 2028 rivoluzionando così il modo in cui le aziende si relazionano con i clienti, raccolgono informazioni e ottimizzano le loro campagne (Statista, 2023).

Gli approcci di big data possono essere utilizzati per raccogliere, elaborare e analizzare enormi volumi di dati comportamentali come registri di ricerca, clic e-mail e cronologia del carrello di spedizione. Questo tracciamento, combinato con l'uso deliberato da parte dei clienti delle stesse piattaforme su molti dispositivi (mobile, desktop, ecc.), fornisce agli esperti di marketing una visione "omnicanales" del loro customer journey senza che questi si registrino sul sito web del rivenditore (Agrawani, 2022).

In seguito, gli strumenti basati sull'intelligenza artificiale analizzano i dati dei clienti per identificare modelli, preferenze e comportamenti di acquisto, consentendo campagne di marketing mirate, raccomandazioni personalizzate, contenuti specifici che portano a tassi di coinvolgimento e conversione più elevati; inoltre permettono di ottimizzare le strategie di

prezzo, prevedere il turn-over dei clienti e identificare le opportunità di cross-selling⁹ e up-selling¹⁰ (Haleem A. et al, 2022)

Oggi, l'apprendimento automatico e la capacità di calcolo basata su GPU¹¹, consentono di creare modelli sempre più complicati e più accurati per un'ampia gamma di attività nel campo del marketing.

L'area in cui l'intelligenza artificiale sta avendo un impatto significativo è sicuramente la segmentazione e il targeting dei clienti; le strategie di marketing tradizionali si basano spesso su dati demografici generalizzati, dando vita a campagne di ampio respiro che raggiungevano un pubblico vasto. L'intelligenza artificiale, invece, è in grado di analizzare grandi quantità di dati dei clienti, come il comportamento di navigazione, la cronologia degli acquisti e le interazioni sui social media, per creare segmenti di clienti più precisi. Ciò consente alle aziende, poi, di indirizzare i propri messaggi ad un gruppo di consumatori preciso e rendere più efficienti e proficui i propri sforzi di marketing.

Un altro compito fondamentale dell'IA riguarda, come descritto precedentemente, l'analisi predittiva; gli algoritmi possono analizzare grandi insiemi di dati per identificare modelli e tendenze, consentendo alle aziende di anticipare il comportamento dei clienti e fare previsioni accurate sui risultati futuri. Sfruttando l'analisi predittiva, le aziende possono ottimizzare le campagne di marketing, determinare i canali e i tempi migliori per raggiungere il pubblico di riferimento e allocare le risorse in modo efficace.

L'intelligenza artificiale svolge un ruolo significativo anche nella generazione e ottimizzazione di contenuti; gli algoritmi di elaborazione del linguaggio naturale possono analizzare grandi quantità di testo e generare contenuti coerenti e coinvolgenti, inoltre aiutano a generare consigli personalizzati sui prodotti, a scrivere testi pubblicitari convincenti e persino a creare campagne e-mail automatizzate. Questo non solo fa risparmiare tempo e risorse rendendo il lavoro dei creatori di contenuti notevolmente più semplice ed efficiente, ma consente alle aziende di fornire contenuti più pertinenti e mirati ai clienti, migliorando l'esperienza complessiva del marchio.

⁹ Cross-Selling: Offrire al cliente un prodotto o un servizio supplementare rispetto a quello che sta per comprare o ha già acquistato (SalesForce).

¹⁰ Up-Selling: Offrire al cliente un prodotto più complesso rispetto a quello inizialmente scelto (SalesForce).

¹¹ Graphic Processing Unit (GPU): circuito elettronico progettato per accelerare la creazione di immagini in un frame buffer, destinato all'output su un dispositivo di visualizzazione (TechTarget).

Inoltre, l'intelligenza artificiale facilita la determinazione dei prezzi dinamici e le offerte in tempo reale nello spazio pubblicitario; gli algoritmi possono analizzare le condizioni di mercato, la domanda dei consumatori e i prezzi della concorrenza per determinare il prezzo ottimale di prodotti e servizi. Ciò consente alle aziende di offrire prezzi personalizzati in base ai profili dei singoli clienti e alle dinamiche di mercato. Nel campo della pubblicità digitale, le piattaforme di offerta basate sull'intelligenza artificiale consentono aste in tempo reale, in cui gli inserzionisti possono fare offerte per il posizionamento degli annunci in base alla probabilità di conversione, massimizzando l'efficienza della spesa pubblicitaria.

In conclusione, l'intelligenza artificiale ha rivoluzionato il panorama del marketing, consentendo alle aziende di comprendere meglio i propri clienti, di personalizzare la messaggistica, di ottimizzare le campagne e di offrire esperienze migliori ai clienti.

Come visto, le nuove tecnologie possono essere utilizzate in vari modi nel marketing e ciascuno di questi ha il proprio insieme di vantaggi che possono essere quantitativi (come l'aumento delle vendite) o qualitativi (come la riduzione del rischio, l'aumento della velocità, l'aumento della soddisfazione del cliente). Tuttavia, vi sono una serie di vantaggi comuni a tutte le applicazioni AI:

- Aumento del ROI della campagna: gli strumenti di intelligenza artificiale possono formulare giudizi rapidi su come allocare al meglio le spese tra i canali multimediali o valutare i posizionamenti degli annunci così che i clienti siano continuamente coinvolti e che le campagne ne possano trarre il massimo valore (vedi Figura 1)
- Migliori relazioni con i clienti: i clienti vogliono che le organizzazioni comprendano le loro esigenze e aspettative, in questo modo l'IAI grazie alla personalizzazione, permette di raggiungere il cliente nel momento giusto della sua vita, garantendo loro massima soddisfazione, un'esperienza ottimale e aumentando, così il tasso di fidelizzazione. Inoltre, grazie a queste tecnologie è possibile individuare i clienti a rischio così da attuare iniziative apposite e ridurre il tasso di abbandono.
- Misurazione avanzata del marketing: Molte aziende faticano a tenere il passo con il volume di dati generato dagli sforzi digitali, rendendo difficile attribuire il successo a campagne specifiche, così le dashboard basate sull'intelligenza artificiale forniscono, una prospettiva più completa di ciò che funziona, consentendo un'allocazione appropriata delle risorse.
- Maggior velocità nel processo decisionale: L'intelligenza artificiale può eseguire analisi tattiche dei dati più velocemente degli umani e utilizzare l'apprendimento automatico per arrivare a decisioni rapide basate sulla campagna e sul contesto del cliente. Ciò consente ai

membri del team di concentrarsi su progetti strategici, che possono essere successivamente utilizzati per informare il marketing basato sull'intelligenza artificiale (Agrawani S., 2022).

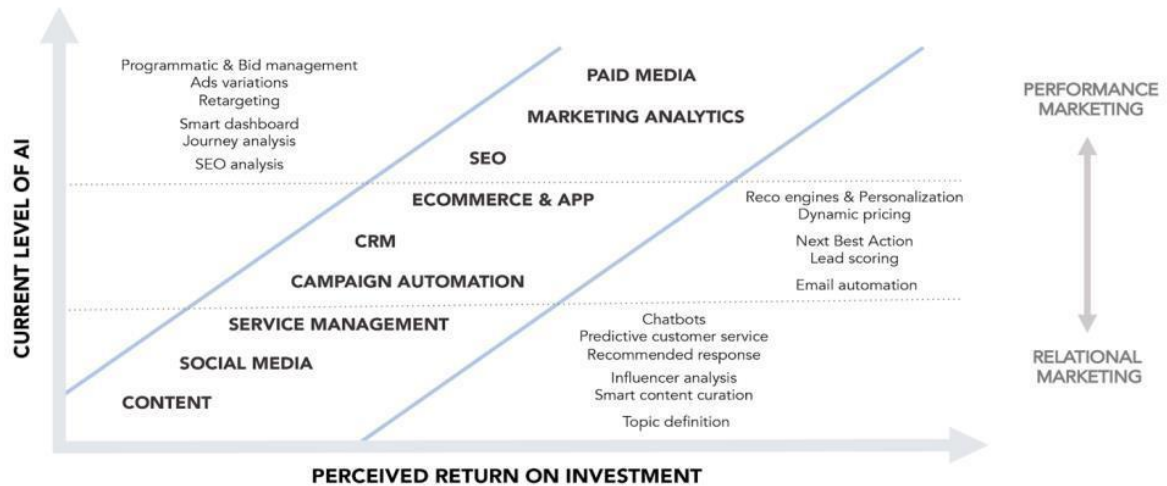


Figura 1. Sofisticazione dell'intelligenza artificiale e del ROI in tutte le aree funzionali.

1.2 L'intelligenza artificiale e il marketing della moda

Il processo di digitalizzazione ha cambiato radicalmente tutti i settori industriali, tra cui il settore della moda.

L'industria della moda e dell'abbigliamento (F&A) è una delle maggiori economie del mondo. Secondo quanto riportato da Statista, questo settore contribuisce per il 38% all'Asia Pacifico, per il 26% all'Europa e per il 22% al Nord America (Statista, 2019), mentre si prevede che i ricavi di questo mercato raggiungeranno i 701,40 miliardi di dollari a fine del 2024 (Statista, 2023) con il lusso che guiderà la crescita di questo settore. Infatti, si prevede che i consumi nella moda del lusso cresceranno tra il 5% e il 10% nel 2023, trainato dal forte slancio in Cina e negli Stati Uniti (McKinsey & Company, 2023).

La moda, oltre ad essere un settore altamente redditizio, rappresenta uno dei settori più dinamici al mondo, in continuo mutamento per soddisfare le esigenze dei consumatori e per rispondere alle rapide variazioni di tendenze di stile, guidate dalle nuove preferenze dei consumatori, dalle influenze sociali e dai cambiamenti culturali (Sukhmani Agrawani, 2022).

Ad esempio, con l'avvento del Covid 19 le aziende di moda hanno dovuto affrontare numerosi cambiamenti per far fronte e adattarsi alle nuove esigenze dei propri clienti. Prima della pandemia il consumatore era, infatti, orientato maggiormente verso l'aspetto estetico, le tendenze di moda e l'espressione individuale dello stile; cercava prodotti di marca, capi di abbigliamento alla moda e accessori che riflettessero la propria personalità e rappresentassero

il proprio status sociale; inoltre il consumatore pre - pandemico era guidato da un forte interesse verso l'abbigliamento formale e le esperienze di shopping in store, che venivano considerate un momento di svago e di socializzazione (Harvard Business School, 2021). Tuttavia, con l'avvento del Covid, le aziende si sono trovate ad operare in un mercato composto da consumatori che non ricercano più esteticità e tendenza bensì comfort e praticità, così da adattarsi alle nuove esigenze di lavoro da casa e ad uno stile di vita più rilassato. Infatti, si stima che tra le conseguenze “permanenti” della pandemia vi sia l'adozione, da parte delle aziende, di una modalità di lavoro ibrida, che prevede cioè una parte della settimana lavorativa svolta da remoto e una parte in ufficio. Di conseguenza l'abbigliamento formale e l'immagine non sono più prioritari per il consumatore che ormai predilige un abbigliamento casual e versatile. Man mano che si è usciti dalla condizione di restrizione sociale che ha caratterizzato gli anni pandemici, il gusto per l'estetico e per il glamour è ritornato ad esistere e a crescere, ma si parla comunque di un “Casual Glamour” (Sukhmani Agrawani, 2022). Inoltre, se prima l'attenzione era sulla moda come status symbol, ora l'attenzione è spostata sulla sostenibilità e sull'etica e anche l'interesse verso lo shopping fisico è venuto meno; infatti, le restrizioni di spostamento e i lockdown hanno portato ad un aumento delle esperienze di shopping virtuale, con i consumatori che cercano facilità di acquisto online, opzioni di consegna sicura e una buona esperienza complessiva (McKinsey & Company, 2022).

Questa rappresenta una sfida ancor più grande per le aziende di moda di lusso che devono rispondere reattivamente alle preferenze dei consumatori in evoluzione e abbracciare l'innovazione, mantenendo, al col tempo, la loro identità di marca distintiva e i propri valori di esclusività, artigianalità e qualità. (Kettydo, 2022).

Soffermandoci, quindi, sul consumatore della moda di lusso del 2023, le tendenze che stanno guidando il settore sono:

- *Sostenibilità*: L'adozione di pratiche etiche e sostenibili sono due driver importanti nelle scelte di acquisto del consumatore. Quest'ultimo predilige marchi che dimostrano un impegno verso pratiche rispettose dell'ambiente, come l'utilizzo di materiali organici o riciclati, l'implementazione di processi di produzione sostenibili e il sostegno al commercio equo e alle pratiche di lavoro etiche. Inoltre, è diventato un elemento che giustifica l'high price di un prodotto. Secondo un sondaggio condotto dalla Repubblica la metà dei consumatori è disposto a spendere dal 5% al 20% in più per capi di abbigliamento con un significativo minor impatto ambientale e sociale (Repubblica, 2021). Ma il consumatore non ricerca la sostenibilità solo in questo senso, bensì ricerca anche il benessere dei

dipendenti, la trasparenza della filiera e una riduzione del consumo di energia e acqua nella produzione (McKinsey & Company, 2020).

- *Personalizzazione*: I clienti, ora come non mai, apprezzano l'individualità e l'espressione di sé nelle loro scelte di moda; sono interessati a stili unici e personalizzati che riflettono la loro personalità e valori e che permettono loro di creare una dichiarazione di moda più personale, distinta ed esclusiva. Inoltre, il consumatore desidera sempre più essere posto al centro delle strategie di business dei brand, per cui, attraverso l'offerta di opzioni personalizzate, i clienti si sentono più coinvolti nel processo di progettazione e sviluppano una connessione emotiva e un senso di appartenenza maggiore verso il brand (McKinsey & Company, 2020).
- *Situazioni digitali*: I clienti preferiscono sempre più la comodità e l'accessibilità dello shopping online, che permette loro di fare acquisti comodamente da casa e scegliere tra un più ampio assortimento di quello che troverebbero in store, nonché di individuare informazioni dettagliate sul prodotto, recensioni dei clienti e confrontare prezzi e stili con i competitors. In tal senso le aziende si sono mosse potenziando e ottimizzando i canali e-commerce e i market place sui vari social network (McKinsey & Company, 2022). Bisogna poi tenere in considerazione che i clienti della moda di oggi si affidano fortemente alle piattaforme dei social media per trovare ispirazione, tendenze di stile, scoprire i marchi e prendere decisioni di acquisto. (Sukhmani Agrawani, 2022). Inoltre, nel frenetico mondo digitale di oggi, i clienti tendono a preferire le aziende aperte alle innovazioni digitali.

In questo contesto, l'accelerazione dell'innovazione tecnologica risulta strategica e indispensabile per i brand di moda, che ottengono nuove e migliori opportunità per servire i clienti creando al contempo un business più efficiente, responsabile e reattivo ai cambiamenti del mercato (McKinsey & Company, 2022).

L'intelligenza artificiale è emersa come un'importante frontiera dell'innovazione tecnologica (Hager, Bryant, Horvitz, Mataric e Honavar, 2017) e, così come man mano che se ne sono scoperti i benefici e i vantaggi ha cominciato a prendere parte in innumerevoli realtà, ha trovato applicazione anche nel settore della moda del lusso e non (Amazon Fashion, 2017; Wong & Liu, 2018). Infatti, questa può offrire miglioramenti significativi in termini di velocità, costi e flessibilità in tutta la filiera, dal processo di progettazione e di ottimizzazione della supply chain, a raccomandazioni personalizzate, marketing influencer e iniziative etiche.

L'IA sta trasformando il modo in cui la moda viene creata, commercializzata e consumata ed è ormai fondamentale per reinventare il design, il merchandising e il marketing (McKinsey &

Company, 2018). Secondo quanto emerso dal Global Fashion Survey, della testata Business of Fashion in collaborazione con McKinsey, il 20% degli executive che vi hanno preso parte crede che, l'uso dell'IA in queste attività, sarà una tendenza molto importante nel futuro prossimo del settore (McKinsey & Company, 2022). Insomma, l'intelligenza artificiale aiuta e consente ai marchi di moda di navigare nel loro panorama dinamico, soddisfare le aspettative dei clienti e guidare la crescita in un settore che prospera sull'innovazione e sul rimanere costantemente all'avanguardia.

Analizzando più nel dettaglio gli ambiti di applicazione delle nuove tecnologie in questo affascinante settore, l'IA può acquisire una duplice funzione all'interno delle aziende: driver nella definizione delle scelte strategiche e tool essenziale nell'attuazione di scelte tattico-operative (McKinsey & Company, 2018).

Per quanto riguarda le scelte strategiche, gli algoritmi di intelligenza artificiale permettono di prevedere insights utili dei consumatori, utilizzabili poi dalle aziende per prendere decisioni giuste e consapevoli nell'obiettivo di accrescere il proprio business. Ad esempio, le odierne tecniche di machine learning permettono di prevedere la domanda e le tendenze di moda: la previsione dei micro (6-12 mesi) e macro (> 12 mesi) trend è da sempre un processo essenziale nelle aziende di moda, e per molto tempo è stato svolto con metodi tradizionali di previsione, limitati però dai pregiudizi umani e dalla soggettività. Gli algoritmi di intelligenza artificiale, invece, analizzano grandi quantità di dati storici sia relativi alle vendite, che alle tendenze di mercato, ad esempio gli hashtag utilizzati o le ricerche effettuate dagli utenti sui social media, per fare delle previsioni accurate e consentire a designer e commercianti di ideare e pianificare le collezioni in modo appropriato e consapevole. Questo aiuta i marchi di moda a ottimizzare la gestione delle scorte, la pianificazione della produzione, le strategie di prezzo oltre ad assumere un ruolo fondamentale nel supportare i designer e la creatività (NFT Paris, 2022).

Gli algoritmi di IA inoltre ottimizzano e aiutano le aziende nella fase di segmentazione e di targeting, processo che si trova alla base di qualsiasi scelta strategica e operativa che si vuole intraprendere. Gli algoritmi di apprendimento automatico basati sui dati permettono di dividere i clienti in gruppi omogenei in base alle loro caratteristiche, comportamenti, esigenze e preferenze, costruendo così segmenti accurati, pertinenti e dinamici. In particolare, le tecniche di clustering possono essere utilizzate per raggruppare i clienti in base alle loro somiglianze e differenze e assegnarli a segmenti in base al loro valore, fedeltà o potenziale, mentre le tecniche di classificazione possono essere utilizzate per prevedere gli attributi dei clienti, come età, sesso, reddito o personalità in base alle loro interazioni con il brand, garantendo così un targeting ottimizzato e personalizzato. L'utilizzo dell'IA per questo scopo comporta diversi

vantaggi per l'azienda tra cui migliorare la comprensione e la soddisfazione dei clienti, aumentare l'efficienza e l'efficacia del marketing, migliorare il vantaggio competitivo e l'innovazione. Può anche ridurre i tempi e i costi della segmentazione e del targeting e migliorare le prestazioni di marketing e il ritorno sull'investimento (Deloitte, 2022).

Si possono utilizzare gli algoritmi di intelligenza artificiale anche per task di geolocalizzazione, termine che si riferisce all'uso della tecnologia e dei dati per determinare la posizione geografica degli individui, i cui dati sono analizzati, in modo da migliorare vari aspetti del business della moda. Immaginando un'azienda di moda di lusso il cui business detiene una copertura globale, risulta necessario per questa, comprendere a pieno ognuno dei suoi mercati e gestire la propria attività considerando le differenti abitudini e esigenze che li contraddistinguono. In questo modo, la geolocalizzazione permette alle aziende di individuare e comprendere meglio le caratteristiche dei diversi mercati geografici e di identificare i segmenti di mercato specifici in ogni regione (Targeting geografico). Di conseguenza, questa tecnica offre alle aziende la possibilità di creare esperienze personalizzate che tengono conto delle specificità culturali dei diversi mercati nonché di offrire contenuti, offerte e servizi che si adattano alle preferenze e alle tradizioni culturali locali. Questo livello di personalizzazione crea una connessione più profonda con i clienti e contribuisce a creare un'esperienza positiva. (Summa Linguae, 2022).

Non bisogna inoltre dimenticare che il settore della moda rappresenta uno dei maggiori produttori di rifiuti a livello globale (Giri et al, 2019) a causa di problemi come sovrapproduzione e resi di prodotti. Secondo quanto riportato da McKinsey il settore è responsabile delle emissioni globali di CO2 per una quota tra il 4% e il 15% e consuma più del 20% dell'acqua per usi industriali, inquinando, in seguito, soprattutto per l'utilizzo di prodotti chimici e per i processi di tintura e concia. E ancora, l'abbigliamento rappresenta la maggior fonte (35%) di microplastica nei mari (McKinsey & Company, 2020). In questo senso, i sistemi basati sull'IA possono analizzare l'impatto ambientale di diversi materiali, processi di produzione e pratiche della supply chain, consentendo ai brand della moda di prendere decisioni più consapevoli in materia di approvvigionamento sostenibile, produzione e riduzione degli sprechi (Bishman, 2023).

Questi sono alcuni dei metodi in cui l'IA può essere utilizzata per identificare scelte strategiche consapevoli e vantaggiose; Infine, una volta che son stati raccolti grandi quantità di dati e che queste son state definite, l'intelligenza artificiale trova massima applicazione nell'implementazione di attività e tattiche di marketing, che permettono alle aziende sia di creare esperienze sempre più coinvolgenti sia di massimizzare la Customer Experience. Ad esempio, l'istallazione di chatbots sui canali online permette di fornire consigli personalizzati,

rispondere a domande e aiutare i clienti a trovare i prodotti che stanno cercando (Forbes, 20213). Questo è ciò che ha fatto, Levi's, il noto brand di abbigliamento specializzato in pantaloni Jeans, che ha cercato di ottimizzare e personalizzare l'esperienza di vendita dei propri clienti con la realizzazione di un vero e proprio virtual stylist, una chatbot abilitata che assiste il cliente nella ricerca del prodotto (vedi figura 2).

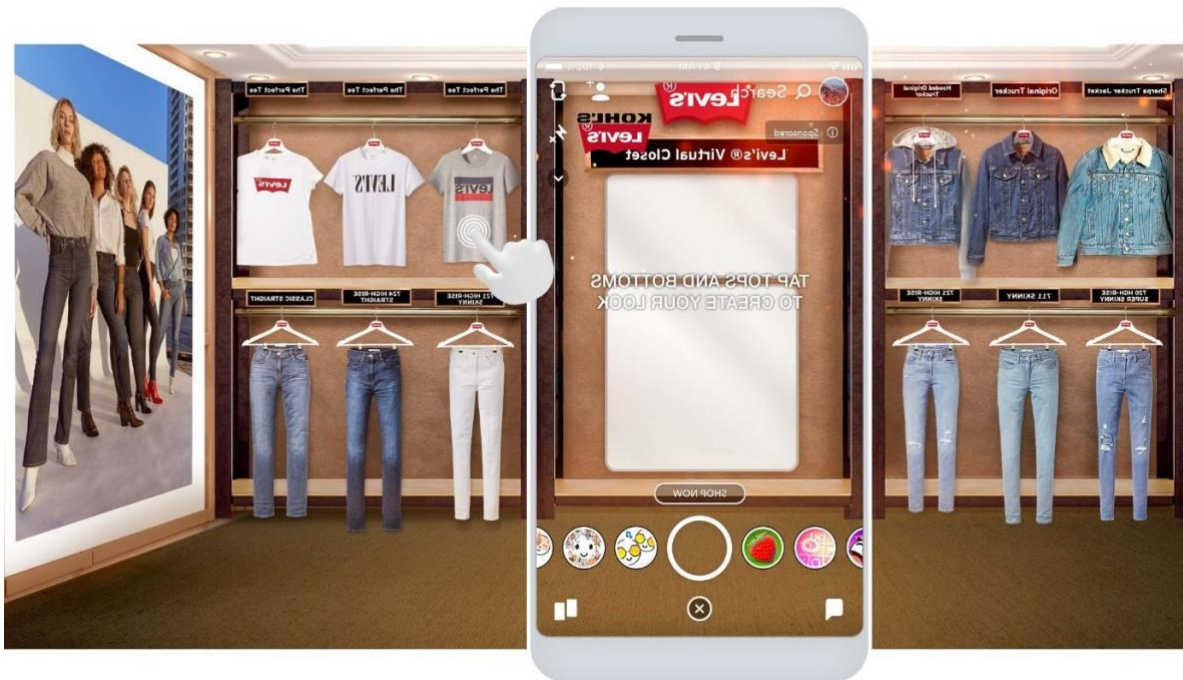


Figura 2 – Levi's Virtual Stylist.

Il virtual stylist utilizza l'elaborazione del linguaggio naturale per determinare ciò che ogni cliente sta cercando in un paio di jeans; così, il bot apprende lo stile di vita e le preferenze di adattamento di ogni cliente e fornisce consigli affidabili. Questi dati, poi, sono integrati con una vera formazione che ricevono gli stilisti umani e, infine, i clienti possono condividere la raccomandazione della chatbot per un paio di jeans con i loro amici per ottenere il loro Feedback (Sukhmani Agrawani, 2022).

Anche le recenti soluzioni di prova virtuale (o Virtual Try-On), che utilizzano le tecniche di realtà virtuale, hanno rivoluzionato l'esperienza dello shopping online; i clienti possono virtualmente provare capi di abbigliamento, accessori, e anche sperimentare stili e colori diversi. Questa ultima tecnologia non solo migliora la soddisfazione del cliente e la fiducia nelle decisioni di acquisto, ma migliora anche i tassi di rendimento dei marchi di moda (Giri et al, 2019). In questo senso, uno dei primi brand ad aver integrato l'intelligenza artificiale con successo nella propria attività è sicuramente Nike (Sukhmani Agrawani, 2022), leader mondiale

nell'abbigliamento sportivo con una quota del 14.4% (Pwc, 2021). In particolare, Nike, ha utilizzato l'intelligenza artificiale in due modi diversi per migliorare il proprio business:

- Nike Fit
- Personalizzazione della Customer Experience

Secondo Nike, il 60% delle persone indossa scarpe della misura sbagliata e ogni anno 500.000 persone affermano di aver acquistato scarpe della misura sbagliata (McKinsey & Company, 2020). Così il noto brand di abbigliamento sportivo ha combinato le tecnologie più all'avanguardia per realizzare Nike Fit, una app che aiuta i clienti ad individuare rapidamente la vestibilità corretta, risparmiando tempo e denaro sull'elaborazione dei resi e sui costi di rifornimento (vedi Figura 3).

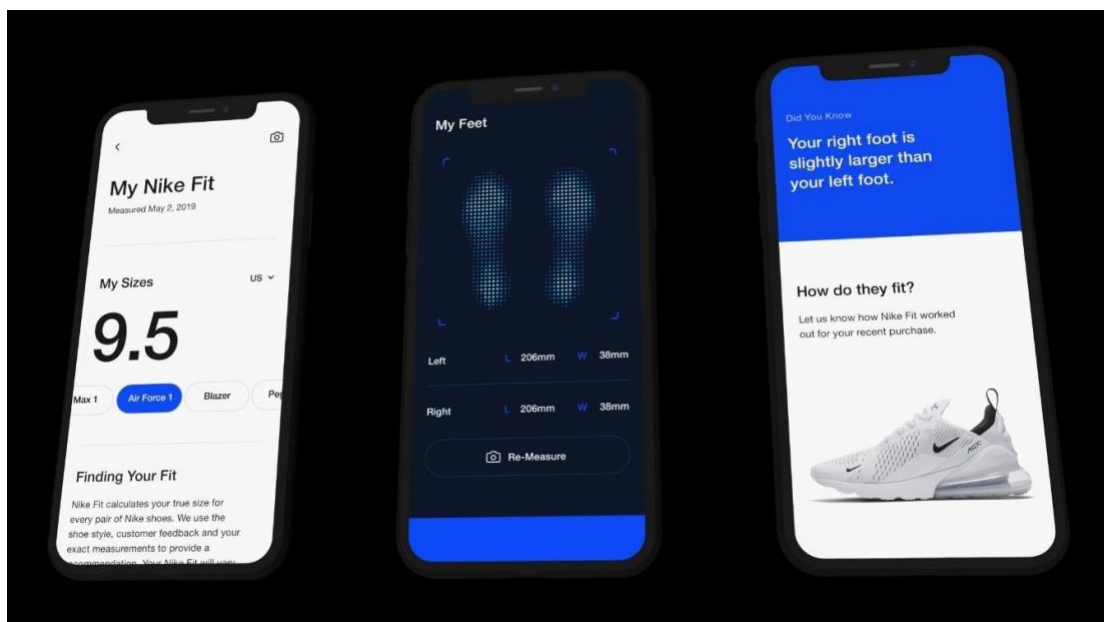


Figura 3. My Nike Fit App.

L'utente scatta una foto con il proprio smartphone e Nike Fit utilizza la visione artificiale per scansionare i piedi del cliente e raccogliere 13 punti di dati visivi, ottenendo così un modello 3D accurato in grado di estrarre i dati di lunghezza, larghezza e altezza.

Nike, inoltre, si contraddistingue per aver compreso e sfruttato in maniera rapida le potenzialità dell'IA anche in termini di personalizzazione, pillar fondamentale nella strategia go to market attuale del brand (Sukhmani Agrawani, 2022).

Infatti, in quest'era in cui la capacità di attenzione del cliente è solo di pochi secondi, una forma di comunicazione altamente personalizzata, consegnata al momento giusto e nel posto giusto permette di ottenere massimi risultati. Soprattutto nel settore della moda e dell'abbigliamento, che si muove e cambia alla velocità ritmo della luce, c'è bisogno di un impegno costante e di

un'interazione personalizzata per distinguersi dall'affollato mercato (Kettydo, 2022). Il gigante delle calzature ha voluto, perciò, migliorare la sua capacità di prevedere il comportamento dei clienti e ottimizzare le scorte nelle operazioni di back-office, andando oltre le collaborazioni al dettaglio e all'ingrosso e decidendo di ottenere il talento e gli strumenti di cui ha bisogno acquisendo, dal 2018, quattro società di data science e analisi: Invertex¹², Collect¹³, Zodiac¹⁴ (Sukhmani Agrawani, 2022).

Dopo aver apportato tali novità nel business Il COO dell'azienda ha dichiarato che la sezione Internet di Nike è esplosa in popolarità nel corso degli anni. Tra il 2017-18 e il 2018-19, i ricavi delle vendite dell'azienda sono aumentati del 12%, mentre nel 2020, le vendite su Internet dell'azienda sono aumentate di un sorprendente 82% (Forbes, 2019).

Inoltre, le tecnologie e gli sviluppi dell'IA vengono applicati alla moda come fonte di collegamento tra la vendita tradizionale offline e quella online. I clienti hanno imparato a conoscere l'interfaccia user-friendly e l'alta usabilità dello shopping in linea, e quindi cercano di vedere le seguenti tecnologie e innovazioni nei centri commerciali offline (D' Vaz - Sterling A, 2021). In tal senso, Alibaba, il più grande mercato di e-commerce al mondo, in un'epoca in cui le vendite offline sono in calo mentre online sta registrando una notevole crescita, si è presentato con un'innovazione del tutto centrata: l'*AI Fashion Assistant*. Si tratta di assistenti virtuali, installati in 13 centri commerciali durante il single day in Cina, che hanno permesso di unire il mondo online e offline portando l'azienda a registrare con successo un totale di 25,4 miliardi di dollari di vendite tra i due canali (Helen James, 2022). In particolare, i vestiti nei negozi, essendo incorporati con chip, permettono allo schermo digitale di identificarli. Quando una persona entra in un camerino dotato di schermo, il sistema identifica immediatamente i capi di abbigliamento e inizia a suggerire capi simili o complementari. Idealmente, l'assistente Fashion AI aiuta l'acquirente a trovare più opzioni sia online che offline. In seguito, effettuata una scelta, un semplice tocco sullo schermo permette di chiamare un assistente per portare l'articolo desiderato (Vedi Figura 4) (Sukhmani Agrawani, 2022).

¹² Invertex: Start-up Israeliana di Computer Vision.

¹³ Collect: Società di analisi predittiva e rilevamento della domanda.

¹⁴ Zodiac: Società di analisi dei dati dei consumatori.

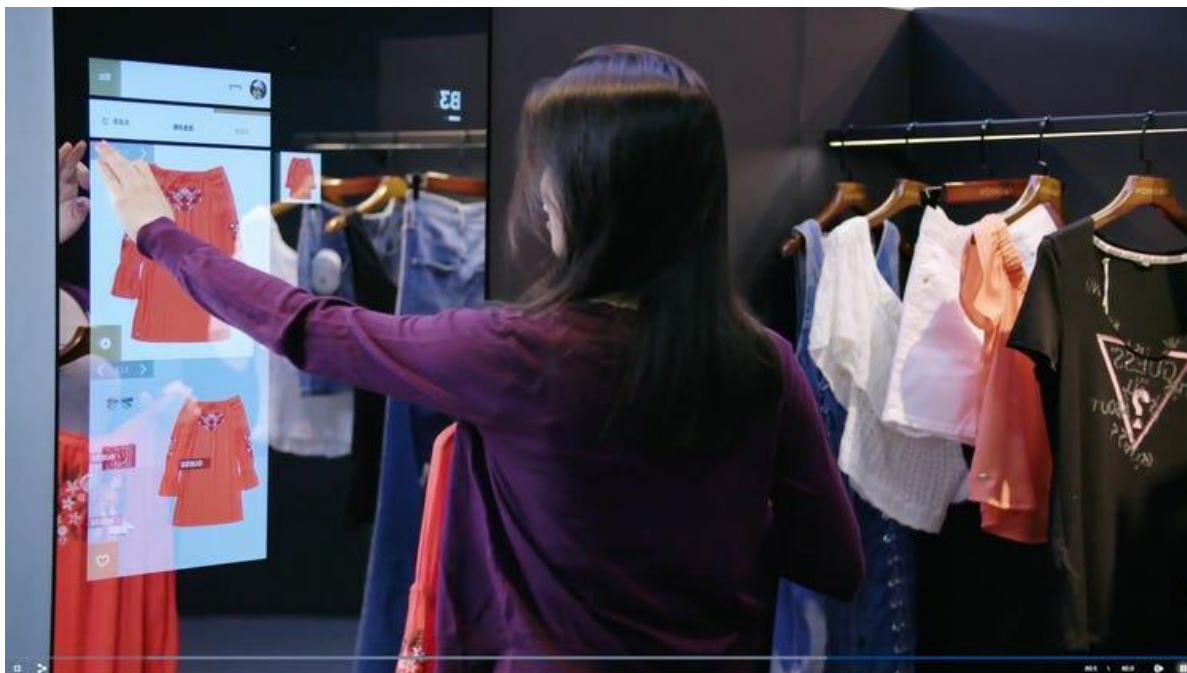


Figura 4 – Scegliere il tuo stile con un tocco grazie all'assistente Fashion AI di Alibaba.

Insomma, l'utilizzo e le capacità dell'intelligenza artificiale rappresentano una vera e propria svolta innovativa e dirompente per le aziende di moda, che ne stanno facendo sempre più un vantaggio competitivo e un sinonimo di differenziazione (Forbes, 2023). Inoltre, queste tecnologie hanno una prospettiva di crescita enorme e oggi assistiamo a continue e sorprendenti novità di applicazione delle stesse, basti pensare all'introduzione del concetto di Metaverso da parte di molti brand nel proprio business (Rathore B, 2023).

Questo è ancor più vero, come sottolineato precedentemente, per i noti brand del lusso che, con un modello di business basato sulla fornitura di un servizio eccellente e un'esperienza memorabile in negozio, utilizzano queste tecnologie per offrire lo stesso livello di esperienza, di soddisfazione e di esclusività, ma con un approccio nuovo, innovativo, contemporaneo e necessario (Toppan, 2021).

Tuttavia, se gli algoritmi di intelligenza artificiale vengono ormai quotidianamente utilizzati per automatizzare i processi e per implementare decisioni di marketing tattiche/operative, per quanto riguarda le scelte strategiche, ancora poche sono le aziende di fashion luxury che utilizzando la geolocalizzazione per personalizzare e rendere più mirate ed efficaci le proprie strategie di marketing.

In un mondo globalizzato, le aziende di moda si trovano sempre più a dover comunicare con persone provenienti da diverse località geografiche. Quindi, comprendere l'importanza di rivolgersi in maniera differente a persone di diversa cultura e località è fondamentale per

costruire relazioni solide, ottenere successo nei mercati internazionali e favorire la reciproca comprensione. Questo è ancora più vero in questo periodo storico in cui i consumatori sono molto esigenti e in particolare nelle aziende di moda di lusso, dove comunicare in maniera differente a persone di diverse località geografiche è di fondamentale importanza per il successo e la rilevanza globale del marchio. Le aziende di moda di lusso si trovano ad affrontare un pubblico internazionale impegnativo e sofisticato, proveniente da culture diverse e con aspettative uniche, ecco perché la comunicazione differenziata riveste un ruolo centrale (McKinsey & Company, 2022).

In conclusione, lo studio proposto si concentrerà su questa tematica, esplorando le possibilità e i vantaggi dell'integrazione dell'intelligenza artificiale nella in aziende di moda di lusso e analizzando come questa tecnica possa aprire loro, la strada a nuove opportunità di innovazione e di crescita.

Capitolo 2 – Geografia, Moda e Lusso

2.1 L'influenza geografica sulle tendenze globali

Negli ultimi decenni, l'interconnessione globale ha portato ad un aumento della diffusione delle tendenze in tutto il mondo. Tuttavia, è fondamentale riconoscere che gli aspetti culturali e geografici giocano un ruolo significativo nell'influenzare la diffusione di queste ultime. In particolare, questi fattori modellano i comportamenti, gli stili di vita e le preferenze delle persone, influenzando così l'adozione e la diffusione delle varie tendenze (Leiserowitz et al., 2006).

Ci sono diversi punti in cui gli aspetti culturali e geografici influenzano la diffusione di stili e idee:

- **Valori culturali:** Come concordato dai noti studiosi Nisbett e Cohen ogni società, a causa delle diverse condizioni geografiche, ecologiche e storiche, ha sviluppato il proprio modello di orientamenti di valore culturale, che a sua volta riflette il suo unico modo di pensare e di comportarsi (Cohen & Nisbett, 1997). Queste differenze influenzano ciò che è considerato desiderabile o accettabile. Quando si parla di valori culturali si fa riferimento a convinzioni, norme e aspettative condivise da una società, che sono ormai così profondamente radicate nel modo in cui le persone pensano, si comportano e si relazionano tra loro, da modellare le loro preferenze e la loro propensione ad adottare nuove tendenze (Leiserowitz et al., 2006).

Edward Said, nel suo libro "Orientalismo" del 1978, parla di "imperialismo culturale" sostenendo che le potenze occidentali hanno creato un'immagine distorta e stereotipata delle culture orientali, in particolare dell'Islam, per giustificare il loro dominio politico ed economico sulla regione (Edward Said, 1978). Ma in senso più generale, l'"imperialismo culturale" è una teoria secondo cui una cultura dominante influenza e modifica le tradizioni, le pratiche e i valori di culture subordinate. Questo comporta l'omogeneizzazione di valori, norme, pratiche, credenze e prodotti culturali, che possono portare a perdere o emarginare le tradizioni e le identità locali. L'imperialismo culturale è spesso associato all'occidentalizzazione o all'americanizzazione, poiché le influenze culturali degli Stati Uniti hanno avuto un impatto significativo su molti Paesi del mondo. Ciò può essere osservato attraverso il dominio dei media americani e delle industrie dell'intrattenimento, come i film, la musica e i programmi televisivi di Hollywood. Tuttavia, lo studio di Elteren del 2003, mostra quanto questo fenomeno possa verificarsi più in alcuni paesi piuttosto che

in altri. Ad esempio, le società più progressiste e aperte al cambiamento sono spesso più inclini ad abbracciare nuovi trend. Non a caso nelle società occidentali, l'individualismo e l'espressione personale sono spesso enfatizzati, portando l'adozione di diversi stili di moda. Al contrario, le culture collettiviste possono dare la priorità al conformismo e alla tradizione, determinando tendenze più uniformi. Le società conservative e tradizionali tendono a mantenere uno status quo e sono riluttanti all'accettazione di cambiamenti significativi nella loro cultura (Elteren, 2003).

A tal proposito, lo studio di R. McCann et al., confronta i membri di tre culture nazionali lungo le dimensioni chiave dell'individualismo-collettivismo. Il campione utilizzato è un gruppo di studenti provenienti da paesi con condizioni geografiche, ecologiche e storiche radicalmente diverse: Thailandia, Giappone e Stati Uniti. Secondo questo studio gli studenti americani sono caratterizzati da un "sé" importante e autonomo, dalla competizione e dalla disuguaglianza (individualismo verticale). I Giapponesi invece prediligono un modello culturale in cui l'individuo vede il sé come parte di un gruppo e in cui tutti sono simili tra loro (collettivismo orizzontale). Infine, per i Tailandesi, il sé è interdipendente con gli altri membri del gruppo ma la disuguaglianza tra i membri del gruppo è accettata (collettivismo verticale) (Mc Cann et al., 2010).

Tra i valori culturali che possono influenzare la diffusione delle tendenze vi sono certamente anche le leggi e la politica; nei paesi in cui vi è alta distanza di potere, ovvero dove l'autorità e la gerarchia sono rispettate, è più probabile che vengano adottate tendenze approvate da figure o istituzioni influenti. Al contrario, nelle culture a bassa distanza di potere, gli individui possono essere più aperti ad adottare nuove tendenze in modo indipendente. Le culture, invece, con un alto livello di evitamento dell'incertezza tendono a valorizzare la stabilità, la prevedibilità e le pratiche tradizionali riluttando e rendendo difficile l'adozione delle nuove tendenze, poiché potenzialmente viste come poco familiari o rischiose. D'altra parte, le culture con un basso grado di evitamento dell'incertezza possono essere più ricettive alle nuove tendenze e idee.

Il livello di attenzione al passato, al presente o al futuro, anche, gioca un altro ruolo importante nell'attitudine ad adottare i più comuni trends: le culture con una forte enfasi sul passato possono aggrapparsi alle pratiche tradizionali, rendendo più difficile l'accettazione delle nuove tendenze, mentre le culture con un orientamento temporale futuro possono essere più lungimiranti, aperte all'adozione di nuove tendenze e a soluzioni all'avanguardia. Infine, nelle società con rigide norme di genere, alcune tendenze possono essere ritenute appropriate per un genere ma non per l'altro. Ciò può limitare l'attrattiva e la diffusione di

tendenze specifiche, poiché potrebbero incontrare resistenza da parte di coloro che aderiscono strettamente ai ruoli di genere tradizionali (Hofstede & Bond, 1984).

Diversi accademici hanno condotto approfondite ricerche e studi sul ruolo e l'importanza dei valori culturali nel plasmare le tendenze e le abitudini di una nazione nel corso del tempo. Ad esempio, Caprar & Neville indagano perché l'adozione di pratiche sostenibili varia così ampiamente tra le organizzazioni appartenenti a culture differenti. In un mondo che lotta per raggiungere gli Obiettivi di Sviluppo del Millennio o per concordare un quadro appropriato per la mitigazione dei cambiamenti climatici, è infatti probabile che le aspettative sulle imprese di ridurre al minimo i danni e massimizzare il bene sociale e ambientale aumentino. Lo studio mostra come, a spiegare questa relazione, giochino un ruolo centrale i valori culturali del paese in cui si trova l'organizzazione, i quali provocano un effetto sia di "norma" che di "conformità". Nell'effetto "normativo" la cultura gioca un ruolo antecedente nella generazione e nell'assimilazione delle istituzioni per la sostenibilità, mentre nell'effetto "conformante" la cultura modera l'effetto di queste istituzioni sulla società, vale a dire, la conversione delle pressioni istituzionali in effettiva adozione della sostenibilità. Di conseguenza, nelle culture libere può essere più efficiente concentrarsi su meccanismi alternativi per promuovere la sostenibilità (come l'enfaticizzazione dei benefici economici, ad esempio), poiché il meccanismo istituzionale può avere un impatto limitato (Caprar & Neville, 2012). Allo stesso modo, Nguyen et al. sostengono che i valori culturali del collettivismo e la prospettiva a lungo termine dei consumatori hanno un'influenza significativa sull'intenzione di acquisto dei consumatori per prodotti ecocompatibili (Nguyen et al., 2017).

- **Influenza sociale:** Le persone tendono ad adottare le mode e le idee prevalenti all'interno dei loro circoli sociali, in particolare quando si trovano nella stessa area geografica. Questo fenomeno è amplificato da quello che viene definito "*Bandwagon Effect*", un bias cognitivo che si riferisce alla tendenza delle persone ad adottare un certo comportamento, stile o atteggiamento semplicemente perché lo fanno tutti gli altri. Più persone adottano una particolare tendenza, più è probabile che anche altre persone salgano sul "carrozzone" (Hayoung Sally Lim et al., 2020). Ad esempio, Eunjoo Choo et al. nel loro studio del 2021 trovano nell'effetto carrozzone, ancor più che nel senso di unicità ed esclusività, la spiegazione della sempre più crescente tendenza della Gen Z ad acquistare prodotti di lusso (Eunjoo Choo et al., 2021).

Nel contesto dell'influenza sociale, l'effetto carrozzone spiega come le persone tendano ad adottare le mode e le idee prevalenti: se una determinata idea o moda diventa popolare in un certo contesto culturale o geografico, le persone all'interno di quel contesto potrebbero sentirsi spinte ad adottarle per conformarsi al gruppo o per evitare di sentirsi estranee. Questo fenomeno è particolarmente evidente nelle aree urbane, che fungono da centri culturali in cui le tendenze emergono e si diffondono più rapidamente a causa della fitta popolazione e delle diverse interazioni.

Altri ricercatori, come Mayners et al. (2023), attraverso 5 studi differenti analizzano l'effetto della vicinanza geografica sull'influenza sociale e dimostrano che non solo la vicinanza sociale, ma anche l'omofilia¹⁵ percepita può innescare l'influenza sociale. Scoprono che questo effetto vale per rappresentazioni alternative della distanza geografica ed è confermato per una gamma di servizi diversi e persino per beni fisici. Inoltre, gli autori mostrano che la vicinanza geografica ha un effetto relativo perché l'influenza sociale di un mittente più vicino è più forte di quella di un mittente più distante, indipendentemente dalle distanze assolute (Mayners et al., 2023).

Uno studio interessante condotto da Christakis e Fowler nel 2007 ha indagato l'influenza sociale su scala globale, analizzando i dati di una grande rete di amicizie tra studenti di un college statunitense; hanno scoperto che le persone erano più propense ad adottare comportamenti come il fumo, l'obesità o la felicità in base alle decisioni dei loro amici. Inoltre, gli autori hanno constatato che quest'effetto di influenza era più forte tra gli amici più vicini geograficamente (Christakis & Fowler, 2007). Un altro studio condotto da Becker et al. nel 2012 ha evidenziato come l'influenza sociale possa variare a seconda del contesto culturale. In questo studio, i ricercatori hanno analizzato l'effetto di influenzare le scelte e i comportamenti degli altri, come ad esempio le donazioni in beneficenza, in diversi paesi (Stati Uniti, Gran Bretagna, la Germania, la Svizzera, la Corea del Sud, la Giappone) e hanno scoperto che l'influenza sociale era più forte nelle culture collettiviste dell'Asia orientale rispetto alle culture individualiste dell'Occidente (Becker M. et al., 2012).

- **Clima e ambiente:** fattori geografici, come il clima e le condizioni ambientali, giocano un ruolo cruciale nell'influenzare le tendenze e le scelte delle persone e questo è vero per molti e diversi settori. Uno studio pubblicato nel 2021 da Ionnaou et al., ha esaminato l'influenza del clima sulla scelta di abbigliamento da parte dei consumatori; la ricerca ha dimostrato

¹⁵ Omofilia: tendenza degli individui di associarsi e di creare legami con altri considerati simili per alcune peculiari caratteristiche (Mayners et al., 2023).

come le temperature più elevate aumentino la preferenza per indumenti leggeri e colorati, mentre le basse temperature spingano verso l'uso di abiti più pesanti e scuri (Ioannou et al., 2021).

Nel 2012, dei ricercatori hanno analizzato come il clima influenzi le tendenze alimentari nelle diverse regioni geografiche. Lo studio ha evidenziato che le peculiarità climatiche di un luogo possano favorire o limitare la disponibilità di certi alimenti, influenzando così le abitudini alimentari delle persone in quelle aree (Vermeulen J. S. et al. 2012). Un'indagine del 2021 condotta da Mertens et al. ha esaminato, invece, l'effetto delle condizioni ambientali sulla scelta del mezzo di trasporto delle persone; i risultati hanno mostrato come un clima favorevole possa incoraggiare l'utilizzo di mezzi di trasporto sostenibili, come biciclette e mezzi pubblici, rispetto all'utilizzo di veicoli privati (Mertens et al., 2015). Infine, lo studio del 2023 di Chen et al. ha approfondito l'effetto delle condizioni climatiche sulla preferenza per le attività all'aperto; la ricerca ha dimostrato come una temperatura moderata e un clima piacevole possano stimolare maggiormente l'adozione di attività sportive all'aperto rispetto a condizioni climatiche estreme (Chen et al., 2023).

- **Media e tecnologia:** i contesti culturali e geografici modellano le abitudini di consumo dei media e l'accesso alla tecnologia. Regioni diverse possono avere diverse forme di media (ad es. TV, radio, piattaforme di social media) con contenuti e portata distinti.

Prima di tutto, i contesti culturali sono fondamentali per comprendere come le persone utilizzano i media e le tecnologie. Ad esempio, in alcuni paesi orientali, come il Giappone, la cultura del rispetto e della discrezione può riflettersi nell'uso moderato dei social media e nella preferenza per la comunicazione faccia a faccia rispetto alle interazioni online. D'altra parte, in paesi occidentali come gli Stati Uniti, l'individualismo e la cultura del "self-branding" spingono le persone ad utilizzare i social media come strumenti per promuovere sé stessi e le proprie idee (Talen, 2004).

Alcune culture, poi, possono dare la priorità alle forme tradizionali di media, come i giornali e la radio, mentre altre possono essere più inclini ai media digitali. Inoltre, in alcune culture potrebbero esserci restrizioni o preferenze per generi specifici di contenuto oppure potrebbero utilizzare gli attuali mezzi e strumenti tecnologici per scopi differenti. Ad esempio, Nai e Kirkup hanno esplorato come i contesti culturali influenzino l'adozione del computer in Cina e Inghilterra, evidenziando che gli inglesi utilizzano il computer principalmente per la comunicazione e l'intrattenimento, e inoltre sono più esperti e sicuri di sé riguardo le loro capacità tecnologiche avanzate rispetto ai cinesi, i quali utilizzano il

computer prevalentemente come strumento di produttività ad esempio per lo studio. Ciò è attribuibile alle differenze culturali e alla diversa percezione della tecnologia in Cina e in Inghilterra (Nail & Kirkup, 2015).

I contesti geografici, invece, possono influenzare la disponibilità e l'accessibilità economica dei media: in alcune regioni, potrebbe esserci un accesso limitato ai media o una mancanza di diversità nelle opzioni dei media a causa di fattori come la censura, il controllo del governo o la limitata produzione di media locali.

L'indagine condotta da Khan N. et al. nel 2022 ha analizzato come i contesti geografici influenzino l'accesso a Internet nelle zone rurali del Pakistan. La ricerca ha rivelato che l'accesso ridotto a internet in queste aree è dovuto a fattori geografici come la mancanza di infrastrutture e a fattori socioculturali, come l'istruzione limitata e l'insufficiente alfabetizzazione digitale (Khan et al., 2022). La ricerca di Alsaleh A. et al. ha analizzato, invece, le differenze nella predisposizione dell'utilizzo dei social media online, in particolare Instagram, tra U.S.A. e Kuwait, scoprendo che gli abitanti americani, notoriamente ricchi di connessioni internet ad alta velocità, avevano maggiore interesse ad utilizzare questo tipo di piattaforme anche per svago piuttosto che gli abitanti del Kuwait che spesso potevano affidarsi solo a connessioni più lente e costose (Alsaleh et al., 2019).

La globalizzazione dei media e delle tecnologie ha sicuramente aperto le porte alla comunicazione in tempo reale tra persone provenienti da diverse parti del mondo. Tuttavia, nonostante questa interconnessione globale, i contesti culturali, geografici, economici e politici di ogni paese influenzano la modalità in cui vengono utilizzati questi potenti mezzi di comunicazione e risulta essenziale riconoscere che ci sono delle differenze, considerato anche il fatto che i media e le piattaforme tecnologiche influenzano la diffusione delle tendenze mettendo in mostra e promuovendo determinate idee, stili o prodotti. Ne è a dimostrazione lo studio di Guan e Yan che confronta le abitudini di consumo dei media e il comportamento ambientale in Cina e negli Stati Uniti rilevando che il contesto culturale ha influenzato il rapporto tra l'uso dei media e la preoccupazione per l'ambiente. I partecipanti cinesi che erano grandi consumatori di media avevano maggiore probabilità di esibire un comportamento ecologico, mentre negli Stati Uniti i consumatori di media pesanti hanno mostrato un comportamento meno ecologico (Guan & Yan, 2019).

Nel complesso, questi aspetti culturali e geografici modellano le percezioni, le interazioni e la disponibilità delle risorse delle persone, influenzando in modo significativo la diffusione delle

tendenze in varie parti del mondo. Comprendere queste influenze aiuta a spiegare perché alcune tendenze prosperano in regioni specifiche mentre non ottengono lo stesso successo altrove.

In questo contesto, risultano quindi, estremamente importanti e strategici gli studi sull'evoluzione dei trend in quanto, questi ultimi, rappresentano le direzioni in cui si muovono le preferenze e le mode di una determinata popolazione o settore, condizionando profondamente sia le scelte dei consumatori sia le strategie di marketing delle aziende.

2.2 Il potere della geolocalizzazione

In un'era globalizzata in cui l'analisi dei trend riveste un ruolo fondamentale per comprendere come idee, comportamenti e prodotti si diffondano tra le popolazioni di diversi paesi, ciò che rende affascinante e complesso questo studio è il riconoscimento che i trend non si diffondono uniformemente in tutto il mondo ma, come visto precedentemente, bisogna prendere in considerazione in modo significativo i fattori geografici e culturali specifici di ciascun paese.

In particolare, si parla di geolocalizzazione o localizzazione, un processo che consente di identificare e registrare la posizione geografica di un individuo, un oggetto o un evento, utilizzando dispositivi come smartphone, applicazioni o diversi metodi di tracciamento. Successivamente una volta che il dispositivo ha ottenuto le informazioni sulla sua posizione utilizzando uno o più di queste metodologie, le informazioni vengono inviate al dispositivo o al servizio che ha richiesto la posizione. Questo processo aiuta le aziende di vario tipo a distinguere i propri utenti in base alla posizione in cui si trovano e, di conseguenza, li supporta nel processo di adattamento del proprio prodotto volto a soddisfare i requisiti culturali, linguistici e di altro tipo di un mercato target (Ding et al., 2018) (Vedi Figura 5).



Figura 5. Social Network basati sulla posizione.

Con il rapido sviluppo dell'ingegneria dell'informazione e dei social network, le posizioni delle persone possono essere comodamente rilevate dalla tecnologia spaziale (Quian et al., 2017) e oggi giorno esistono vari modi per ottenere e conoscere la posizione di una persona.

Tra i più comuni vi sono:

- **GPS:** Forma più diffusa di tecnologia di geolocalizzazione utilizzata sui dispositivi mobili e che si basa su satelliti di comunicazione che orbitano attorno alla terra e che trasmettono continuamente il loro stato, la posizione esatta e l'ora precisa;
- **Wi-Fi:** Quando una persona si connette a una rete Wi-Fi, il dispositivo può rilevare le reti Wi-Fi circostanti e utilizzare i dati per triangolare la posizione approssimativa;
- **Bluetooth:** Si usano i Beacon Bluetooth¹⁶ a basso consumo energetico, per ricevere segnali dai dispositivi posizionati all'interno di un'area specifica per fornire una maggiore precisione sulla posizione di una persona;
- **Cellular Networks:** Le reti cellulari sono abilitate da una rete di torri cellulari che trasmettono le onde radio utilizzate per la comunicazione mobile (Incognia, 2023).

	GPS	Cellular networks	Wi-Fi positioning	Bluetooth low energy
Environment	Outdoor	Outdoor	Indoor & outdoor	Indoor mainly
Accuracy	10 - 100 m	0.75 miles	5 - 15 m	1 - 2 m
Battery consumption	High	Low	Medium	Low

Figura 6. Differenza dei diversi metodi di localizzazione in base all'ambiente in cui operano, l'accuratezza dei risultati ottenuti e il consumo di risorse impiegato.

In questo contesto, le piattaforme dei social media stanno assumendo sempre più un ruolo centrale, in quanto permettono di raccogliere le informazioni sulla localizzazione delle persone e della diffusione dei trend (Williams et al., 2017).

Con miliardi di utenti attivi quotidianamente, le piattaforme come Facebook, Twitter, LinkedIn, Instagram e TikTok offrono un enorme bacino di dati geografici e comportamentali e permettono di conoscere non solo cosa pensa un utente, ma anche quando e dove si trova.

¹⁶ Beacon Bluetooth: piccoli dispositivi wireless che funziona grazie alla tecnologia Bluetooth Low Energy e monitorare la presenza di dispositivi mobili per trasmettere dati entro un raggio di azione che varia dai 10 centimetri a circa 70 metri (Pugno M., 2020).

Gli utenti possono fornire i propri dati di localizzazione tramite molteplici modalità:

- Geo Tagging: Gli utenti di piattaforme come Instagram e Twitter possono geotaggare i loro post o le loro foto, fornendo così informazioni precise sulla loro posizione;
- Check-in: molte piattaforme di social media offrono una funzione di check-in che consente agli utenti di indicare la propria presenza in un luogo specifico, come un ristorante o un particolare, e di condividerla con amici, familiari o followers;
- Geo-stickers: Social Network come Instagram permettono di inserire degli adesivi o dei tag all'interno di post e storie, in cui è possibile indicare la posizione in cui ci si trova il momento.
- Real-Time Position: Applicazioni come Whatsapp e Snapchat permettono di inviare la posizione in tempo reale o attuale ad altri utenti;
- Hashtag: Per molti utenti dei social network è solito includere nei propri post, hashtag tramite cui vengono fornite informazioni, seppur approssimative, del momento in cui l'immagine è stata scattata o il post condiviso;

Le aziende, gli esperti di marketing e gli studiosi possono, in questo modo, mappare geograficamente la loro rete di utenti (vedi figura 7), ottenere preziose informazioni sulla posizione geografica degli stessi e comprendere come i trend si diffondano in diverse regioni del mondo (Dev.to)

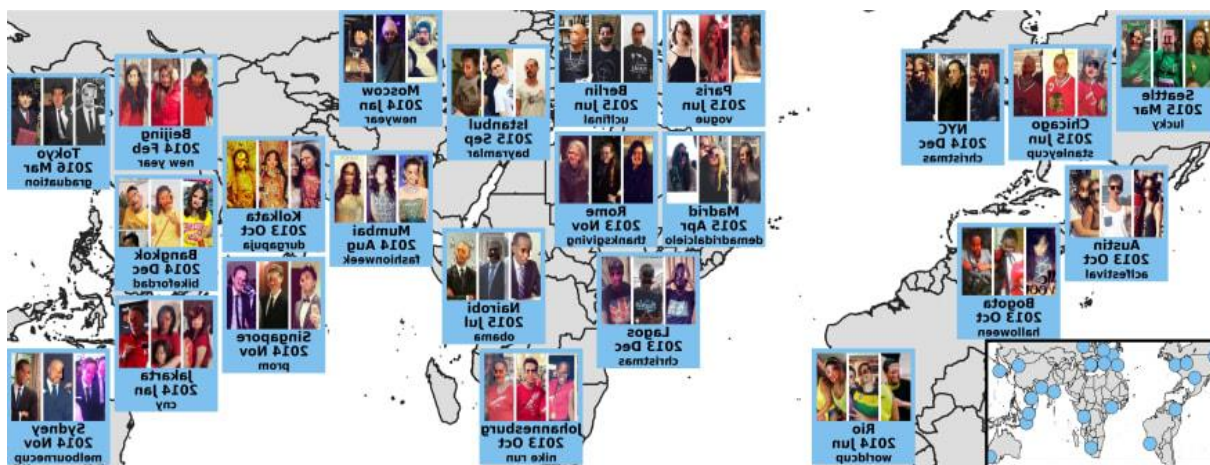


Figura 7. Informazioni specifiche relative all'utente social e alla sua posizione geografica.

A rivoluzionare il panorama della localizzazione, vi è sicuramente l'intelligenza artificiale che attraverso la traduzione automatica, l'elaborazione del linguaggio naturale, la personalizzazione dei contenuti, gli approfondimenti basati sui dati e altro ancora, consente alle aziende di

connettersi in modo efficace con il pubblico globale e permette di ricavare la posizione in tempo reale da un'immagine, da un post, o da una semplice scritta, in maniera accurata e sempre più rapida (Kesprov, 2023).

Nel corso del tempo, molti studiosi si sono impegnati ad analizzare il fenomeno della geolocalizzazione utilizzando diverse fonti e diversi strumenti. Per quanto riguarda le fonti si fa riferimento sia al tipo di social media preso in considerazione per eseguire il task in questione, sia all'interno di esso, quali sono gli elementi considerati per reperire i dati (es. Immagini, testo, interazione sociale), ad esempio, Martinez-Letorre et al., hanno effettuato uno studio di geolocalizzazione su dati forniti da social media incentrati su immagini come Flickr per individuare le risorse turistiche nella città a cui sono più interessati i turisti della città di Saragozza (Spagna) (Martinez-Letorre et al., 2014). Chi et al., invece, hanno sfruttato le immagini e i contenuti testuali presenti nel social network di Twitter per prevedere la posizione degli utenti (Chi et al., 2016). Francalanci et al., nel loro studio finalizzato a supportare i servizi di emergenza in caso di disastri naturali, hanno introdotto algoritmi che sfruttano non solo le immagini, che non sempre sono georeferenziate, ma vari indizi contestuali inclusi nei post dei social media per contribuire ad aumentare la proporzione di post che possono essere geolocalizzati (Francalanci et al., 2018). Lo studio di Hassan et al., ha analizzato l'influenza sociale nella scelta della posizione dell'attività e nei modelli di stile di vita, utilizzando i dati di geolocalizzazione del social media Foursquare ricavati dai servizi di check-in (Hassan et al., 2016).

Infine, lo studio innovativo di Zheng et al., ha superato i tradizionali approcci basati sui social media dove gli algoritmi imparano singolarmente dai contesti testuali e dalle reti sociali proponendo invece una nuova struttura end-to-end, Hybrid-attentive User Geolocation (HUG), che modella congiuntamente i testi dei post e le interazioni degli utenti nei social media su tre set di dati presi dal social media Twitter (Zheng, 2020).

Per quanto riguarda gli strumenti invece Sarker et al., ha utilizzato le tecniche di machine learning, in particolare decision tree, k-nearest neighbors, random forest, support Vector machine, per identificare, tramite post di 3 anni su Twitter, i posti in cui si ha un maggior abuso di Oppioidi. Gli autori hanno utilizzato l'elaborazione del linguaggio naturale per individuare tutti i post e la comunicazione presente sul social media in questione relativo all'argomento. I risultati hanno mostrato la Pennsylvania come paese in cui si ha un maggior abuso di oppioidi (Sarker et al., 2019). Ancora, lo studio di Chi et al., per prevedere la posizione degli utenti e dei tweet di Twitter. ha utilizzato il classificatore Naive Bayes (Chi et al., 2016).

Lourentzou et. al., nel 2017, motivati dal successo delle tecniche di Deep Learning per attività quali la visione artificiale, il riconoscimento vocale e l'elaborazione del linguaggio naturale, hanno deciso di valutarne e confermarne il successo anche in ambito di geolocalizzazione utilizzando un set di dati da Twitter (Lourentzou et al., 2017). Infine, Rahimi et al., offrono un modello di geolocalizzazione multi-vista basato su Graph Convolutional Networks, che utilizza sia il testo che il contesto di rete (Rahimi et al., 2018).

Nonostante i numerosi studi presenti in questo ambito, emerge il ripetitivo utilizzo da parte dei ricercatori della stessa fonte per raccogliere i dati e formulare le proprie analisi, ovvero il social media Twitter. Questo può essere spiegato dalla sua enorme popolarità e dalla facilità di accedere a dati geolocalizzati dagli utenti. Tuttavia, bisognerebbe esplorare anche altri social network in quanto ognuno di questi ha le sue peculiarità e il suo preciso bacino di utenti, i quali possono presentare comportamenti diversi.

Inoltre, sarebbe interessante, in questo contesto, sondare maggiormente il campo dell'apprendimento non supervisionato per ottenere un quadro più completo del fenomeno, aprire nuove prospettive di studio e arricchire ulteriormente la letteratura esistente.

La geolocalizzazione è ampiamente utilizzata per una varietà di scopi, tra cui la navigazione e il tracciamento dell'utente, la personalizzazione dei contenuti in base alla posizione e la fornitura di servizi basati sulla localizzazione, come la ricerca di ristoranti o negozi nelle vicinanze.

In generale questa imponente tecnica offre numerosi vantaggi sia per gli utenti stessi che per le aziende: per gli individui essere in grado di geolocalizzare sé stessi o gli altri significa poter facilmente trovare indicazioni stradali, scoprire punti di interesse nelle vicinanze, condividere la propria posizione con amici o familiari per motivi di sicurezza, e ricevere notifiche personalizzate basate sulla propria posizione; mentre consente alle aziende di migliorare il proprio targeting di marketing, coinvolgere i clienti, fornire un servizio di assistenza clienti migliore, ottimizzare le risorse e migliorare la logistica, nonché raccogliere e analizzare dati per prendere decisioni informate (Incognia, 2023).

2.3 La Geografia della Moda

L'industria della moda è particolarmente nota per la sua natura dinamica e transitoria che continua a cambiare nel tempo e nello spazio (TutorialsPoint, 2023).

Come disse, infatti, una delle più note e prestigiose stiliste di questo settore:

“La moda non è qualcosa che esiste solo nei vestiti. La moda è nel cielo, nella strada; la moda ha a che fare con le idee, con il modo in cui viviamo, con ciò che sta accadendo...”

Coco Chanel, 1883-1971

Di conseguenza, l'attività delle aziende che operano in questo settore si basa sulla capacità di prevedere le tendenze del futuro e cavalcare quelle del presente, così da garantire il successo dei propri prodotti e rimanere sempre al passo con i desideri e le aspettative dei consumatori.

Rispetto ad altri settori, la previsione delle tendenze nella moda presenta però una sfida unica e difficile da affrontare: le preferenze e le richieste di stile possono variare rapidamente nel tempo, spesso senza un'apparente o chiaro elemento scatenante. Pertanto, per le aziende della moda, è di vitale importanza conoscere i fattori che influenzano le tendenze per poter adattare la produzione, le vendite e il marketing di conseguenza.

Comfort, gusto, occasione ma più in generale norme sociali, mass media, critica sociale, ambiente, costumi, religione, ricchezza, clima, umore della città in cui si vive e più recentemente fashion influencers, KOL¹⁷, celebrità e icone di stile rappresentano solo alcuni dei fattori chiave che possono influire sull'evoluzione delle tendenze (vedi figura 8) (Halal & German, 2020).

¹⁷ KOL: Figure politiche, scrittori, celebrità visti come esperti di un certo settore o argomento e percepiti come persone reali e di conseguenza i loro consigli come genuini, sinceri e disinteressati e in grado di influenzare un ampio bacino di utenti (Lucesoli, 2021).



Figura 8. Tutti i fattori che influenzano la moda.

Inoltre, i piccoli marchi di moda locali si trovano a vivere in un mercato sempre più dominato da brand internazionali che personalizzano e diversificano la propria offerta permettendo loro di raggiungere qualsiasi realtà, e che, di conseguenza, incentivano qualsiasi brand di fashion ad adottare una strategia di livello globale. In questa ottica risulta essenziale, quindi, analizzare i macro-fattori che determinano e alterano lo sviluppo delle tendenze di moda nelle diverse parti del mondo.

Con l'espressione "*geografia della moda*" si fa riferimento allo studio interdisciplinare che esamina la connessione esistente tra geografia, società e industria della moda. Questo campo di ricerca analizza come la moda sia influenzata dal contesto geografico, esplorando i fenomeni come la produzione, la distribuzione e il consumo di abbigliamento in diverse parti del mondo. Il concetto di "*geografia della moda*" nasce nel 1900, quando la moda diventò un fenomeno globale e l'industria dell'abbigliamento si sviluppò a livello internazionale assumendo piano piano sempre più rilevanza (Patchett & Williams, 2021). Pioniere di questo argomento vi è la geografa *Louise Crewe* con il suo libro "*The Geographies Of Fashion*" in cui analizza il rapporto tra moda e geografia. Crewe ha sfidato il rifiuto della moda da parte della disciplina come una preoccupazione femminile frivola e minore, mostrando, dai suoi primi lavori nei quartieri della moda (Crewe, 1996) e spazi commerciali e lo shopping (Crewe, 2000; Crewe, 2001; Brooks et al., 2002) al suo sviluppo e all'applicazione delle teorie della scala e della relazionalità per mappare e connettere i diversi siti e architetture globali delle industrie della

moda (Crewe, 2008; Crewe, 2010; Crewe, 2016; Crewe, 2017), perché la moda è geograficamente rilevante e perché i geografi dovrebbero interessarsene. Come, infatti, sostiene nel suo libro *“l'industria della moda ha utilizzato attivamente strategie geografiche di associazione e dissociazione per creare e riprodurre gerarchie di valore: mettendo in rilievo determinati spazi, luoghi e persone mascherando le disuguaglianze globali, gli abusi degli standard di lavoro e le catastrofi ambientali che sono alla base dell'industria”* (Crewe, 2017). Sulla base di questa affermazione Crewe indaga come la moda è influenzata e come contribuisce allo sviluppo dei vari paesi. L'autrice affronta diverse tematiche come la produzione e il consumo di moda in diverse regioni, le sue dimensioni culturali, economiche e sociali e il ruolo delle città e degli spazi urbani in questo settore (Patchett & Williams, 2021). Da sempre Parigi ricopre egregiamente il ruolo di capitale mondiale della moda; infatti, è stata per lungo tempo il centro nevralgico in cui erano allocate le più prestigiose case di moda, dove si ospitavano le sfilate più importanti e da cui, si riteneva si divulgassero tutte le maggiori tendenze, ad esempio il concetto di "haute couture" e l'introduzione del concetto di "chic" (Halah & German, 2020).

Successivamente la geografia della moda si è evoluta e si è ampliata, includendo molte altre città e regioni del mondo.

Ben presto, grandi città come New York, Milano, Londra e Tokyo sono diventate importanti hub creativi e commerciali, dando vita a un vero e proprio sistema di moda internazionale e rientrando così nella denominazione di capitali della moda. Allo stesso tempo, città meno note come Berlino, Stoccolma, Copenaghen e Mumbai hanno acquisito una crescente rilevanza nella mappa della moda globale (Halah & German, 2020).

Il libro curato da Breward e Gilbert *“Fashion's world cities”* ha mostrato, poi, come lo status di “Capitale della moda” venisse spesso mitizzato e sfruttato dagli urbanisti per promuovere e valorizzare l'“economia culturale” delle grandi città, mentre il sistema di produzione dell'abbigliamento dell'industria veniva ristrutturato e globalizzato grazie al più ampio passaggio culturale (Breward & Gilbert, 2006). Ancora, Weller nel 2007 ha evidenziato che le capitali della moda erano “territori di influenza estetica” e “conoscenza viziosa” che hanno esercitato un'enorme influenza sulla ristrutturazione globale dell'industria, ma che dagli anni '70 ha visto il trasferimento della produzione lontano dai distretti dell'abbigliamento nelle città industrializzate centrali e semi-periferiche (Weller, 2007).

Quindi, la geografia della moda si è anche integrata con le dinamiche dei flussi migratori e delle produzioni tessili a livello globale. Ad esempio, molte aziende di moda del Nord America e dell'Europa occidentale hanno spostato la produzione in paesi a basso costo come Cina,

Bangladesh e Vietnam (Smith, 1996). Questo ha portato a una nuova mappa della produzione tessile e ad una ridefinizione dei ruoli che le diverse regioni del mondo ricoprono nella catena di approvvigionamento.

Inoltre, con l'avvento di Internet e dei social media, la diffusione delle tendenze e delle influenze nel mondo della moda si è accelerata. Oggi, le informazioni sulle ultime tendenze e i prodotti di moda raggiungono immediatamente un pubblico globale. Ciò ha aperto nuove opportunità per gli stilisti emergenti in tutto il mondo e ha spostato l'attenzione dalla geografia tradizionale della moda verso nuovi centri creativi (TutorialsPoint, 2023).

In questo contesto, un interessante studio introduce un approccio volto a rilevare quali città influenzano altre città in termini di propagazione dei loro stili, basandosi su immagini quotidiane di persone che indossano abiti e utilizzando come mezzo di strumento un'analisi di geolocalizzazione. Lo studio analizza quali città esercitano e ricevono maggiore influenza sulle altre, quali influenzano maggiormente le tendenze globali, quali contribuiscono alla preminenza di un determinato stile e come il grado di influenza di una città è cambiato nel tempo (Vedi Figura 9), con l'obiettivo di sfruttare queste informazioni per prevedere con precisione le tendenze future per località (Halah & German, 2020).

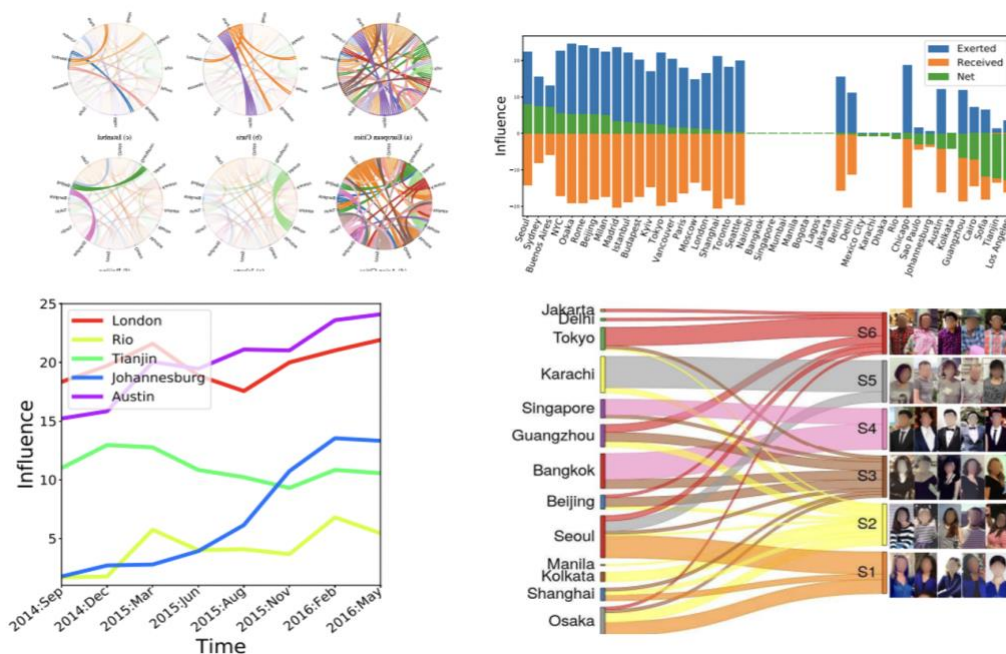


Figura 9. *Top Left*: Lo stile influenza le relazioni scoperte dal modello tra le città europee e asiatiche. *Top Right*: Classifica mondiale delle principali città in base alla loro influenza di moda sui loro coetanei. *Bottom Left*: Analisi dinamica che rivela i cambiamenti temporali delle città nella forza di influenza. *Bottom Right*: influenza delle città asiatiche sulle tendenze globali di 6 stili di moda dove l'ampiezza della connessione è relativa al peso dell'influenza di quella città rispetto ad altre città influencer dello stesso stile.

2.4 Il settore del Fashion Luxury

Le organizzazioni nelle industrie creative, come nella moda, affrontano un'enorme pressione a causa del cambiamento dei gusti (Hirsch, 1972). Mitigano l'incertezza selezionando e combinando accuratamente elementi culturali per progettare i loro prodotti (Verganti, 2009) e il successo o il fallimento di queste imprese è quindi determinato dalla loro comprensione delle dinamiche degli elementi culturali.

La geolocalizzazione, quindi la capacità di determinare la posizione geografica di qualunque persona, offre l'opportunità unica ai brand di capire le differenze esistenti tra i vari consumatori permettendogli di diversificare l'offerta e di personalizzare la comunicazione rivolta loro in maniera rapida ed efficiente. Non a caso, si tratta di uno strumento che i brand di qualsiasi settore, compresa l'industria della moda del lusso, hanno cominciato ad integrare nelle proprie strategie di mercato. Infatti, se, come abbiamo visto, le aziende di moda incontrano notevoli difficoltà nell'adattarsi alle più recenti tendenze, a causa della loro intrinseca natura mutevole e dinamica, questo è ancor più vero per il mondo del lusso. Infatti, riflettere le maggiori tendenze del momento e allo stesso tempo rimanere fedeli ad un posizionamento di mercato di fascia alta che raggiunge un segmento di nicchia, rappresenta la più grande sfida di un brand di lusso (TrendsFormative, 2022).

Il lusso si basa su un paradosso: per avere successo i brand devono avere un'elevata consapevolezza ma, allo stesso tempo, risultare rari ed esclusivi. Il concetto di "lusso" si riferisce a qualcosa che è intrinsecamente irraggiungibile, ed è spesso associato ad un prezzo premium e alla capacità di evocare il sogno tra gli individui. L'offerta dei brand di fascia alta si concentra sulla creazione di prodotti o servizi di alta qualità, unici e distintivi, che soddisfano le esigenze di una clientela esigente e disposta a pagare di più per avere qualcosa di esclusivo. Ed è proprio al fine di evocare e mantenere questo status di esclusività che i brand di questo settore si sono concentrati sempre su una offerta poco differenziata seppur a livello globale, proprio perché focalizzano la propria strategia sulla qualità, sullo status sociale, sul senso di rarità e su una clientela ben selezionata e specifica (Amaldoss e Jain, 2005; Ko et al., 2019).

Tuttavia, oggi giorno, questo settore sta attraverso un processo di democratizzazione guidato dalla crescita sostanziosa della propria clientela (TrendsFormative, 2022).

Dopo il crollo del 23% registrato nel 2020, anno ancora strettamente coinvolto nella pandemia, il mercato mondiale del lusso è cresciuto di quasi il 15% nel 2021, con un fatturato stimato di 1.400 miliardi di euro e con una previsione di crescita del settore tra il 6% e l'8% annuo entro il 2025 (Bain & Company, 2021). Uno dei principali fattori a guidare questa crescita è

sicuramente l'aumento della clientela in Cina e in America, che oggi giorno rappresentano i due più grandi mercati di consumo di brand di lusso. Due paesi, tuttavia, non molto simili tra loro. Inoltre, se finora i clienti dei marchi di lusso si sono concentrati nei grandi agglomerati urbani, molti acquirenti si stanno stabilizzando in città più piccole, acquisendo nuove abitudini, nuove consuetudini e diverse necessità (Bain & Company, 2021). Di conseguenza, i rinomati marchi di lusso sono obbligati ad adattare il loro consolidato e storico principio di esclusività, distintività e unicità, abbracciando la necessità di comunicare con un pubblico diverso, più esigente, il quale richiede una maggiore personalizzazione e ricerca un brand in grado di soddisfare i propri valori culturali e il proprio stile di vita.

2.5 Obiettivo di ricerca

Alla luce di quanto esplorato nel presente capitolo, dall'importanza degli elementi geografici e culturali nella diffusione delle tendenze a livello globale, al conseguente potere dello strumento della glocalizzazione, al concetto di geografia della moda, all'analisi del delicato e articolato settore del lusso e ad una generale revisione della letteratura in questo ambito, verrà effettuato uno studio volto ad indagare maggiormente questo fenomeno, fornendo, di conseguenza, importanti contributi di marketing alle aziende nel settore della moda di lusso.

La ricerca presente nel seguente elaborato si impegna ad arricchire la letteratura esistente in campo di geolocalizzazione, espandendo il pool di social media utilizzati per estrarre i dati e sondando il campo dell'apprendimento non supervisionato.

In particolare, attraverso un approccio data – driven, verranno identificate potenziali differenze tra paesi, riguardo alle scelte dei consumatori nell'acquisto di accessori di moda.

Si vuole cioè analizzare, se esiste una correlazione tra i contenuti dei post dei social media su diversi tipi di borse quando gli utenti provengono dalla stessa area geografica.

Verranno quindi indagate le seguenti ipotesi:

H1: Esiste una correlazione tra i contenuti dei post relativi a diversi tipi di borse pubblicati in posizioni geografiche vicine.

H2: Non vi è correlazione tra i contenuti dei post relativi a diversi tipi di borse pubblicati in differenti e distanti aree geografiche.

Capitolo 3 – Materiali e Metodi

3.1 Strumenti

L'avvento della digitalizzazione e il sempre maggior avanzamento tecnologico sta portando un enorme incremento nella disponibilità di dati in tempo reale per persone e aziende.

Questo aumento consistente di informazioni ha dato origine alla scienza dei dati, un campo interdisciplinare che combina elementi di statistica, matematica, informatica e scienze sociali per estrarre conoscenza e informazioni significative da grandi quantità di dati.

Come ogni disciplina scientifica, l'analisi dei dati segue un rigoroso processo graduale, definito "analisi dei dati", volto ad individuare le informazioni utili e valide presenti nell'insieme dei dati a disposizione. Questo processo costa di 5 fasi, ognuna delle quali richiede capacità e know-how differenti (vedi figura 10):



Figura 10. Schema delle attività che costituiscono il processo di analisi.

La prima fase consiste nell'individuazione del problema manageriale che si vuole risolvere e di conseguenza l'obiettivo di ricerca che si vuole raggiungere tramite l'analisi. In gergo tecnico si parla di "*problem statement*" e, durante questa fase, viene solitamente definita l'ipotesi di ricerca che si vuole confermare o confutare e il metodo da utilizzare per farlo.

Una volta stabilito l'obiettivo, si raccolgono e aggregano i dati appropriati. In questa fase risulta fondamentale determinare quali dati sono necessari; questi dati possono essere quantitativi (numerici) o qualitativi (descrittivi), ovvero acquisiti tramite ricerche di tipo qualitativo, ad esempio tramite sondaggio. Inoltre, i dati possono essere raccolti per via diretta dall'azienda in questione (First-party data), possono essere dati primari di altre organizzazioni (Second-party data), ad esempio i dati reperiti tramite social media, oppure possono essere dati acquistati da terze parti ovvero da organizzazioni la cui attività consiste nell'aggregare e raccogliere un'ampia quantità di dati, che solitamente sono Big Data (Third Party Data).

Una volta che i dati necessari sono disponibili, il passaggio successivo consiste nell'individuare e risolvere i problemi di qualità dei dati che potrebbero influire sull'accuratezza delle applicazioni di analisi. Si tratta di una fase molto importante a cui i data scientist dedicano circa

l'80% del tempo dell'analisi (Dodds, 2020). Ciò include l'esecuzione di attività di profilazione e pulizia dei dati per garantire che le informazioni in un set di dati siano coerenti e che gli errori e le voci duplicate vengano eliminate (TechTarget).

Uno degli ultimi passaggi del processo consiste nell'analisi e nella manipolazione dei dati; all'inizio di questa fase viene fatta un'analisi esplorativa: vengono identificati pattern, tendenze o relazioni significative attraverso tecniche di visualizzazione dei dati, come grafici, diagrammi o mappe di colori. In seguito, vengono utilizzati algoritmi o modelli statistici per individuare relazioni significative nei dati e per rispondere alla domanda di ricerca iniziale. In base all'obiettivo identificato inizialmente si può eseguire una analisi diagnostica, ovvero volta alla comprensione del motivo per cui è accaduto un determinato fenomeno; un'analisi predittiva, che consente di identificare le tendenze future sulla base dei dati storici; un'analisi descrittiva, in cui viene studiato un fenomeno già accaduto o un'analisi prescrittiva che consente di formulare raccomandazioni per il futuro.

Si arriva poi all'ultima fase che consiste nell'interpretazione dei dati analizzati: si tratta di un momento cruciale per le aziende in quanto permette loro di trarre valore effettivo dai passaggi precedenti. In questa fase si analizzano i risultati per comprendere le tendenze, i modelli e le relazioni tra i dati raccolti e questo consente all'azienda di trarre conclusioni e fare previsioni o suggerimenti per il futuro.

È importante, che ognuna delle fasi del processo di data Analysis sia svolta in maniera approfondita, dedicando risorse e competenze adeguate, poiché solo attraverso un'analisi accurata si potranno ottenere informazioni utili per prendere decisioni strategiche e migliorare le performance aziendali (Willer, 2023; Pickell, 2021).

Le iniziative di analisi dei dati supportano un'ampia varietà di usi aziendali, dall'analisi dei dati dei clienti per una miglior comprensione degli stessi, all'analisi dei dati finanziari per valutare le performance, analisi dei dati di marketing per rendere più efficaci le campagne e massimizzare il ritorno sugli investimenti e all'analisi dei dati di produzione per ridurre i costi di produzione e migliorare la qualità di produzione. Queste possibilità offrono molteplici benefici per le imprese, consentendo loro di ottimizzare le operazioni, identificare nuove opportunità, prevedere tendenze e rischi e prendere scelte più consapevoli, diventando così una componente essenziale per il successo aziendale.

Nel presente elaborato, al fine di comprendere l'esistenza di caratteristiche simili tra le abitudini d'acquisto di borse in diverse regioni geografiche, verrà analizzato un set di dati utilizzando le

metodologie di data mining e avvalendosi del potenziale offerto dal linguaggio di programmazione Python.

Il data mining è uno dei modi per manipolare i dati al momento della fase di “analisi dei dati” precedentemente descritta. Si tratta di un approccio che utilizza tecniche di intelligenza artificiale, statistica e machine learning per scoprire modelli nascosti o relazioni precedentemente sconosciute nei dati. Questa tecnica di estrazione delle conoscenze può essere utilizzata per scopi diversi, come la previsione delle tendenze di vendita, l'identificazione di anomalie o la creazione di segmenti di clientela.

Python invece, è uno dei linguaggi di programmazione di alto livello più usati al mondo, creato da Guido Van Rossum nel 1989 e rilasciato nel 1991. Per linguaggio di programmazione si intende il linguaggio (C, C++, Java, Python) in cui sono state scritte le istruzioni da parte dello sviluppatore affinché un computer esegua un determinato compito. Le statistiche mostrano che nel 2020 circa il 66% dei data scientist utilizza Python quotidianamente e l'84% lo usa come linguaggio principale. Da subito ha cominciato ad avere un grande successo tanto che gli autori del “2022 Octoverse report”, hanno evidenziato come Python sia cresciuto di oltre il 22% su base annua con oltre quattro milioni di sviluppatori su GitHub che lo hanno utilizzato a un certo punto nel 2022. Il successo di questo linguaggio di programmazione può essere spiegato dalla sua natura semplice e dalla sintassi leggibile che lo rendono un linguaggio eccellente per innumerevoli applicazioni, per l'automazione di attività, per la creazione di siti Web o software e per l'analisi dei dati; inoltre dispone di una libreria standard ampia, è un linguaggio open source¹⁸, multiplatforma e portatile¹⁹ (Scarlett, 2023).

¹⁸ Open Source: il codice sorgente del linguaggio di programmazione Python è disponibile liberamente e può essere consultato, modificato e utilizzato da chiunque.

¹⁹ Portatile: lo stesso codice può essere eseguito su qualsiasi sistema operativo con un interprete Python (differentemente, alcuni linguaggi richiedono la modifica del codice per l'esecuzione su piattaforme diverse).

3.2 Il Dataset

Per condurre lo studio di geolocalizzazione proposto in questo elaborato, è stato utilizzato un dataset di post Instagram relativi alle borse dal 2021 al 2023, selezionati in base agli hashtag associati.

Il seguente dataset è costituito da 16.788 osservazioni e 71 features²⁰ relative a informazioni specifiche del post (ora di creazione, hashtag, commenti, likes etc.), informazioni sull'immagine in questione (peso, larghezza), informazioni specifiche del dispositivo dal quale è stato pubblicato il post, informazioni dettagliate sull'utente (nome, followers, stato, url etc.) e sulla sua posizione (latitudine, longitudine, id etc.).

La scelta di adottare Instagram come social network da cui reperire i dati e, di conseguenza, come fonte su cui basare la ricerca è stata dettata dai seguenti motivi:

- Instagram è il terzo social network con più utenti al mondo (2 trilioni) al mondo, successivo solo a Facebook e YouTube, mentre è secondo tra i social media preferiti al mondo dopo WhatsApp (vedi Figura 11). Inoltre, nel 2023 Instagram si rivela leader in termini di crescita: l'app conta 522 milioni di utenti in più (+35,3%) rispetto l'anno precedente (Gonzalez, 2023).

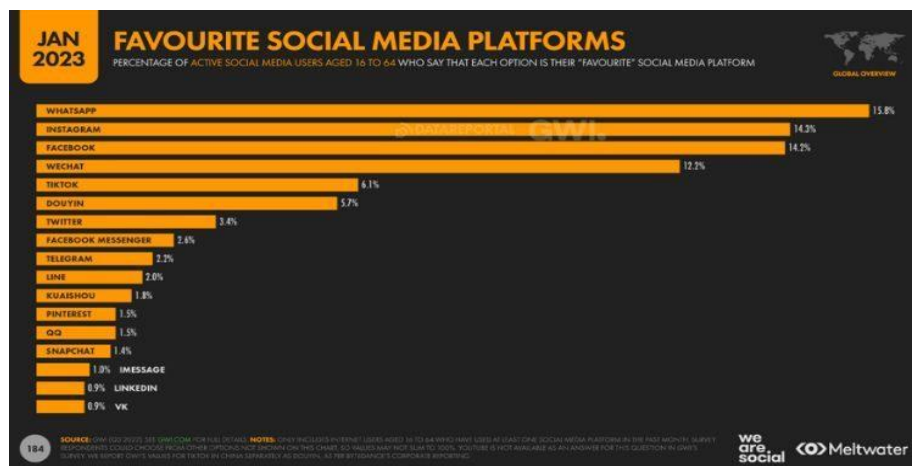


Figura 11 – Classifica dei social media preferiti tra gli utenti di età compresa tra i 16 e i 64 anni, nel 2023.

- Instagram è uno dei social network più diffusi a livello globale, con una vasta base di utenti in tutto il mondo. Gli Stati Uniti registrano il più grande numero di utenti, con circa 116

²⁰ Feature: una parte di informazione che è rilevante per risolvere un obiettivo computazionale in relazione a una certa applicazione.

milioni di persone che utilizzano la piattaforma, seguiti dall'India con 73 milioni e infine dal Brasile con 72 milioni di utenti. Tuttavia, Instagram è popolare e ampiamente utilizzato in tutti i paesi, senza differenze significative nell'uso tra le varie nazioni e questa uniformità nell'utilizzo, consente all'analisi di geolocalizzazione e ai risultati che ne derivano di assumere una portata e una valenza globale.

- La ricerca ha come obiettivo di offrire insight di geolocalizzazione utili nel business delle borse di lusso. Di conseguenza, con il mercato del lusso in fase di democratizzazione e la necessità di connettersi con una clientela più giovane, Instagram offre un'opportunità senza pari per le aziende del settore. Infatti, Instagram è particolarmente popolare tra i giovani e gli adulti giovani. Come affermato da Barnhart:
 - il 75% delle persone di età compresa fra 18 e 24 anni usa Instagram
 - il 57% delle persone di età compresa fra 25 e 30 anni usa Instagram

Il target group di Instagram è, quindi, principalmente giovane, anche se è ancora molto popolare tra gli utenti di età compresa fra 35 e 54 anni (Barnhart, 2021).

- Instagram è una piattaforma fortemente basata sulle immagini, il che rende più attraenti e coinvolgenti i contenuti pubblicati. Questo permette anche di riflettere in modo autentico gli interessi e le tendenze locali, facendo in modo che l'analisi delle immagini su Instagram offra una prospettiva preziosa per comprendere le preferenze e i comportamenti del pubblico in diverse regioni geografiche.
- Gli utenti di Instagram tendono ad essere molto attivi e coinvolgenti, con un alto tasso di interazione con contenuti e post geo riferiti. Questo può fornire dati geolocalizzati più affidabili e completi rispetto ad altri social network meno popolari o meno coinvolgenti.
- Instagram incorpora nella sua attività diversi metodi di geolocalizzazione, dal geo tagging, al check-in, a geo stickers, ad geo hashtag; questo può facilitare notevolmente la raccolta di dati geo riferiti per analisi e studi.

3.3 Metodologia

Nel seguente elaborato è proposta un'analisi di geolocalizzazione per esplorare l'esistenza di differenti tendenze nei comportamenti di acquisto di borse di lusso a livello globale, utilizzando il metodo A.I. proposto di seguito (vedi Figura 12).



Figura 12. Overview of AI Fashion Luxury geolocation analysis method adopted.

3.3.1 Data Pre-Processing

Per avviare il processo di data preprocessing del dataset, è stata effettuata una fase iniziale di *Data Cleaning*, ovvero è stata eseguita una pulizia preliminare all'interno del dataset eliminando le features poco rilevanti per lo studio in questione; sono perciò state eliminate le feature *iphone_struct_is_favorite*, *iphone_struct_is_private*, *iphone_struct_is_unpublished*, *iphone_struct_profile_pic_id*, *Iphone_struct_profile_pic_url*, *iphone_struct_username*, *iphone_struct_has_primary_country_in_profile*, *iphone_struct_can_be_reported_as_fraud*, *friendship_status_following*, *friendship_status_is_bestie*, *friendship_status_is_feed_favorite*, *friendship_status_is_restricted*, *friendship_status_outgoing_request*, *owner_is_unpublished*, *an_club_info_fan_club_id*, *fan_club_info_fan_club_name*, *media_to_caption_text*, *width*, *fan_club_info_is_fan_club_referral_eligible*, *height*, *caption*, *imgs*, *comments*, *location_has_public_page*, *fan_club_info_fan_consideration_page_revamp_eligibility*, *fan_club_info_is_fan_club_gifting_eligible*, *iphone_struct_has_anonymous_profile_picture*, *iphone_struct_has_highlight_reels*, *iphone_struct_is_favorite*, *iphone_struct_is_private*, *iphone_struct_is_unpublished*, *iphone_struct_profile_pic_id*, *iphone_struct_profile_pic_url*,

iphone_struct_username, display_resource_1, display_resource_2, display_resource_3, media_to_comment_has_next_page, owner_blocked_by_viewers, owner_followed_by_viewers, owner_has_blocked_viewer, owner_requested_by_viewer, owner_is_private, owner_is_verified.

Sono state poi eliminate dall'analisi anche le variabili di carattere strettamente personale (*id, id_row, owner_id, owner_profile_pic_url*) e ridondanti (*location_name; location_slug; location_id; location*).

In seguito, sono stati eliminati i valori duplicati (3834) e i valori nulli (8143) al fine di preservare l'integrità dell'analisi. Inoltre, sono state eliminate le variabili il cui contenuto prevalentemente mancante avrebbe potuto influenzare negativamente la validità dello studio: *titolo_post* (12713), *owner_profile_pic_url_hd* (12469), *edge_media_to_sponsor_user* (12469).


Infine, utilizzando il metodo *sns.boxplot*, è stata condotta un'analisi per identificare eventuali valori anomali (outlier) da eliminare.

Si è ottenuto così un dataset di nuove dimensioni: 5095 osservazioni x 7 features (*create_at; hasht_cor; hashtag_ricerca; edge_liked_by_count; media_to_comment_count; location_lat; location_lng*), sulla base del quale è stata avviata la fase di *Data Trasformation* per consolidare i dati di qualità in forme alternative modificandone il valore, la struttura o il formato.

In questa fase, considerando la natura dello studio da condurre, sono stati determinati e individuati i paesi soggetti all'analisi. Per fare ciò, è stata utilizzata la funzione *geolocator*²¹ per creare una nuova caratteristica denominata "*country*", che indica il paese in cui ogni post è stato pubblicato in base alle rispettive coordinate geografiche. Così, sono state eliminate le osservazioni aventi informazioni nulle nel campo in questione e sono state eliminate le caratteristiche relative alle coordinate (*location_lat; location_lng*), considerando solo la variabile "*country*" nel proseguimento dello studio.

²¹ Geo locator: libreria che fornisce funzionalità per la geolocalizzazione inversa (ottenere le coordinate dati un indirizzo) e per la geolocalizzazione diretta (ottenere l'indirizzo dati le coordinate), utilizzando i dati forniti dai servizi di geocoding come Google Maps, Bing Maps, Nominatim, etc.

Successivamente, tramite la libreria *Pycountry*²², sono stati unificati i valori di paesi simili come "Niger" e "Nigeria", considerandoli come un'unica entità; ed è stata utilizzata la libreria *Googletrans*²³ per tradurre i nomi dei paesi nella lingua inglese.

Inoltre, il valore "THE MOST KNOWN " è stato individuato come una scrittura mongola non riconosciuta; perciò, questo valore è stato sostituito con "China" in modo da eliminare l'ambiguità legata alla scrittura non riconosciuta ed in modo di individuare correttamente il paese coinvolto nello studio di geolocalizzazione.

Questi passaggi hanno permesso di creare un insieme coerente e omogeneo di paesi, riducendo la possibilità di duplicazione, semplificando la comprensione e la comparazione dei dati attraverso diverse regioni geografiche.

In seguito, sempre nella fase di trasformazione dei dati, si è svolta la codifica e la standardizzazione dei dati.

La codifica dei dati, svolta tramite la funzione *LabelEncoder()*, permette di convertire le variabili categoriche (in questo caso *create_at*, *hashtag_ricerca*, *hasht_corr*, *country*) in variabili numeriche. Questa operazione è importante perché fornisce un formato e una struttura adeguati ai dati per l'elaborazione da parte dei modelli, garantendo una migliore compatibilità, efficienza e precisione.

La standardizzazione invece, tramite la funzione *StandardScaler()*, riduce la scala dei dati in modo tale che ogni caratteristica in questione abbia media 0 e varianza 1. Questa operazione garantisce la comparabilità dei dati, eliminando le distorsioni indesiderate, migliorando le prestazioni degli algoritmi di analisi e facilitando l'interpretazione dei risultati.

In generale queste due analisi, risultano essenziali in quanto permettono di rappresentare e trattare dati omogenei e coerenti tra loro.

A conclusione del processo di preparazione dei dati è stata, infine, eseguita la *Dimensionality Reduction*, ovvero una tecnica statistica per ridurre la quantità di variabili casuali in un problema ottenendo un insieme di variabili principali che preservano quante più informazioni significative possibili. La riduzione della dimensionalità è utile per affrontare problemi legati

²² *Pycountry*: libreria Python utilizzata per fornire un'ampia gamma di informazioni sui paesi, come i loro nomi ufficiali, codici ISO, lingue parlate, fusi orari e altro ancora

²³ *Googletrans*: libreria Python che fornisce un'interfaccia per l'utilizzo dell'API (Application Programming Interface) di traduzione di Google Translate, consentendo di tradurre testo da una lingua all'altra.

all'alta dimensionalità dei dati, come il sovraccarico computazionale, l'overfitting o la difficoltà di visualizzazione.

All'interno dell'ambito della Dimensionality Reduction, nel presente studio, è stata eseguita l'analisi dei componenti principali (PCA), una tecnica che calcola una combinazione lineare delle variabili originali, ottenendo nuove variabili, chiamate componenti principali, ordinate in modo da rappresentare la massima varianza dei dati. Quindi, la prima componente principale rappresenta la direzione della massima varianza, la seconda componente principale rappresenta la direzione della seconda massima varianza e così via.

In questo modo è stata semplificata la complessità del dataset a disposizione, creando due componenti principali, rispettivamente Col1 e Col2, con le quali è stato eseguito lo studio.

3.3.2 Data Exploration

In seguito alla fase di preelaborazione dei dati si è ottenuto il dataset definito su cui basare l'analisi di dimensioni 1487 osservazioni x 6 features.

Concentrandoci sulle variabili selezionate per l'analisi di geolocalizzazione, queste includono:

- *Create_at*: l'ora e il giorno in cui è stato creato il post su Instagram
- *Hasht_cor*: un termine o una frase che viene utilizzata nel contesto di un post Instagram per associare il contenuto ad un argomento specifico o a un tema più ampio. Ad esempio, se si sta pubblicando una foto del mare, si può utilizzare l'hashtag correlato #summer o #hot per indicare che il post è relativo al tema dell'estate o di una vacanza.
- *Hashtag_ricerca*: un termine o una frase che viene utilizzato per scoprire o cercare specifici contenuti all'interno dell'app. Gli utenti di Instagram possono cercare hashtag come #food o #fashion per trovare post correlati o per esplorare contenuti relativi a questi temi.
- *Edge_liked_by_count*: numero di "mi piace" ottenuti nel post Instagram specifico.
- *Media_to_comment_count*: numero di "commenti" presenti nel post Instagram specifico.
- *Country*: paese in cui è stato generato il post Instagram.

Di conseguenza, una volta che il dataset è stato pulito, elaborato e manipolato si è passati ad una fase di esplorazione dei dati, al fine di acquisire una visione completa e dettagliata dei dati e al fine di visualizzare / indagare eventuali relazioni significative all'interno del dataset a disposizione.

Quindi, sono state analizzate le statistiche descrittive per ciascuna variabile del dataset: *create_at* (count: 1487.000000; mean: 410.5; std: 212.9; min: 0; 25%: 225.5; 50%: 462; 75%:582.0); *hashtag_ricerca* (count: 1487.000000; mean: 4.1; std: 2.5; min: 0; 25%: 3.0; 50%: 3.0; 75%:7.0); *hasht_cor* (count:1487.000000; mean: 328.1; std: 181.7; min: 0; 25%: 194; 50%: 312; 75%: 469.5); *edge_liked_by_count* (count:1487.000000; mean: 28.2; std: 29.2; min: 0; 25%: 3.0; 50%: 14.0; 75%: 56.0); *media_to_comment_count* (count: 1487.000000; mean: 6.3; std: 8.2; min: 0; 25%: 0; 50%: 2.0; 75%: 12.0); *country* (count:1487.000000; mean: 25.8; std: 15.8; min: 0; 25%: 15.0; 50%: 20.0; 75%: 37.0).

Successivamente, è stata condotta un'analisi della matrice di correlazione, espressa anche tramite una mappa a colori termici, volta ad esaminare le relazioni esistenti tra le varie variabili. Come mostrato di seguito (vedi Figura 13), le variabili presentano generalmente una bassa correlazione tra loro, ad eccezione di *edge_liked_by_count* e *media_to_comment_count* che presentano una correlazione di 0.79. Questa relazione può essere spiegata dal fatto che entrambe le variabili rappresentano indici di interazione al post, di conseguenza è plausibile che un maggior numero del numero dei like in un post, comporti un aumento del numero di commenti, riflettendo così l'interazione più intensa da parte degli utenti su quel post.

Inoltre, è emersa un'associazione significativa tra "Media_to_Comment_Count" e "country" con un coefficiente di correlazione di 0.52, risultato che apre uno spazio interessante per ulteriori analisi approfondite nel corso dello studio.

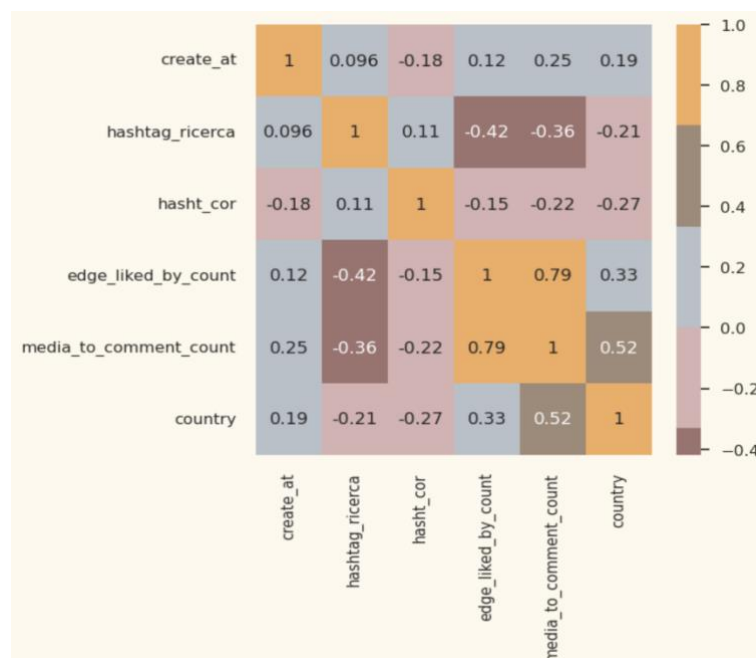


Figura 13. Matrice di correlazione

In aggiunta, utilizzando la funzione *countplot*, che consente di rappresentare graficamente la distribuzione dei valori di una variabile categorica all'interno di un dataset, è stata indagata maggiormente la variabile *country*, individuando così i specifici paesi coinvolti nell'analisi (vedi figura 14).

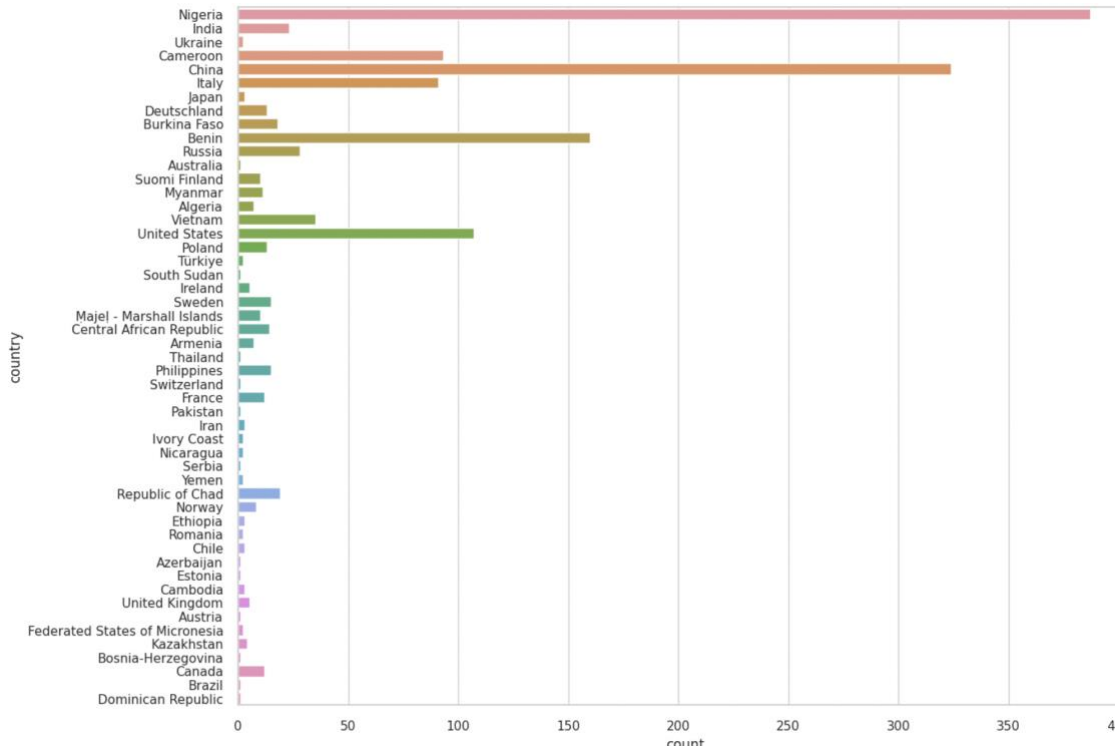


Figura 14. Countplot dei paesi presenti nella variabile “country”

3.3.3 Data Mining

Quando si parla di Machine Learning si fa riferimento ad un sottoinsieme dell'intelligenza artificiale che si occupa di trovare pattern simili per andare a fare predizioni, permettendo la risoluzione di problematiche non risolvibili con la mera statistica di base. Più nello specifico, il termine ML esprime la capacità, da parte della macchina, di imparare ed eseguire determinati compiti, sulla base di algoritmi che apprendono dai dati in modo iterativo. Come detto precedentemente, affinché una macchina sia in grado generare al proprio interno le regole per elaborare i dati input, sono necessari algoritmi²⁴ appositi (in questo caso Python), moltissimi dati di esempio, capacità computazionale e memoria.

²⁴ Algoritmo: serie di operazioni semplici, diverse fra loro, che, eseguite in un ordine prestabilito, permettono di giungere ad un risultato atteso.

Così come gli esseri viventi apprendono e imparano diversamente, allo stesso modo per gli algoritmi di machine learning sono previste diverse modalità di apprendimento: apprendimento supervisionato, apprendimento non supervisionato e apprendimento rinforzato.

Nel caso dell'apprendimento supervisionato (Supervised Learning), il programma viene addestrato su coppie di dataset input/output predefinite dai trainer in cui le risposte giuste sono già state correttamente etichettate. Al contrario, nel caso dell'apprendimento non supervisionato (Unsupervised Learning), i dati di input non presentano etichetta, di conseguenza, l'algoritmo deve trovare da sé le relazioni non lineari all'interno dei dati non classificati e capire quali consentono di raggiungere i risultati migliori rispetto agli obiettivi del modello. Infine, l'apprendimento con rinforzo (Reinforcement Learning), è utilizzato soprattutto in applicazioni che devono interagire con contesti altamente dinamici e incerti. In questo caso al modello sono posti degli obiettivi da raggiungere (come nell'apprendimento supervisionato) e un sistema di ricompense o penalità, ma non è fornito alcun suggerimento sugli oggetti che può incontrare o sulle regole del gioco adottate nell'ambiente; quindi, il modello apprende quali sono le decisioni giuste da prendere, in maniera iterativa, mediante tentativi ed errori, ricevendo un feedback ad ogni performance (Cervelli, 2022).

Considerata la numerosità di ricerche di geolocalizzazione condotte utilizzando approcci di apprendimento supervisionato, nel seguente studio è stato sondato il dominio dell'apprendimento non supervisionato.

In particolare, in questo studio viene utilizzata la tecnica del clustering, ovvero una tecnica di data mining che raggruppa dati non etichettati in base alle loro caratteristiche comuni, quindi in base alle loro somiglianze o differenze. L'output ottenuto da una procedura di Clustering consiste nella suddivisione del set di dati iniziale in gruppi di dimensioni minori e omogenei, noti come cluster, che massimizzino la similarità intra-cluster e minimizzino la similarità inter-cluster.

Quindi, nel seguente elaborato saranno impiegati algoritmi di clustering e metodi statistici descrittivi per rilevare possibili somiglianze nei comportamenti dei consumatori di borse di lusso a livello globale. Ciò avverrà seguendo il modello concettuale illustrato di seguito (vedi Figura 15).

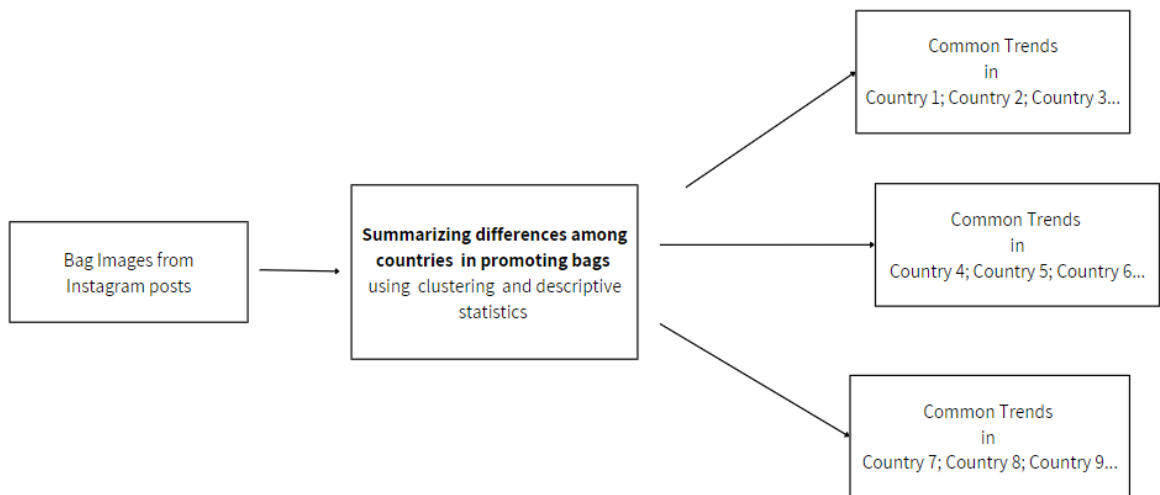


Figura 15. The conceptual framework for analyzing trends in luxury handbag consumption

Iniziando la fase di *Data Mining* è stato individuato il numero di cluster ottimale per procedere all'analisi. Infatti, un passaggio fondamentale per qualsiasi algoritmo non supervisionato è determinare il numero ottimale di cluster in cui i dati possono essere raggruppati, inoltre questa operazione risulta fondamentale per l'applicazione di alcuni algoritmi di clustering, perché influisce direttamente sulla qualità dei risultati e sull'interpretabilità dei modelli scoperti. Infatti, se viene scelto un numero insufficiente di cluster, i modelli potrebbero essere semplificati eccessivamente e perdere distinzioni più fini all'interno dei dati; d'altra parte, se si scelgono troppi cluster, i risultati potrebbero diventare eccessivamente complessi e difficili da interpretare.

Di conseguenza, nel presente studio il numero ottimali di cluster è stato calcolato attraverso l'Elbow Method, o metodo del gomito, che consiste nel tracciare una misura della qualità del clustering (come variazione spiegata) rispetto al numero di cluster e cercare un punto che assomigli ad un "gomito", indicatore di un buon compromesso tra il numero di cluster e la qualità del clustering.

È stato, così, individuato il punto di gomito utilizzando due misure differenti: WCSS (Within-Cluster Sum of Squares) e il Distortion Score. La differenza principale è nel modo in cui quantificano la qualità del clustering: WCSS utilizza la somma delle distanze al quadrato, mentre il punteggio di distorsione utilizza la distanza al quadrato media.

Come visibile nella Figura 16, l'analisi basata esclusivamente sulla misura di distorsione suggerisce che il numero ottimale di cluster sia 5. Allo stesso tempo, utilizzando la somma dei quadrati all'interno dei cluster come misura, emergono 3 cluster ottimali.

Di conseguenza, entrambi questi risultati saranno tenuti in considerazione durante il proseguimento del processo di data mining.

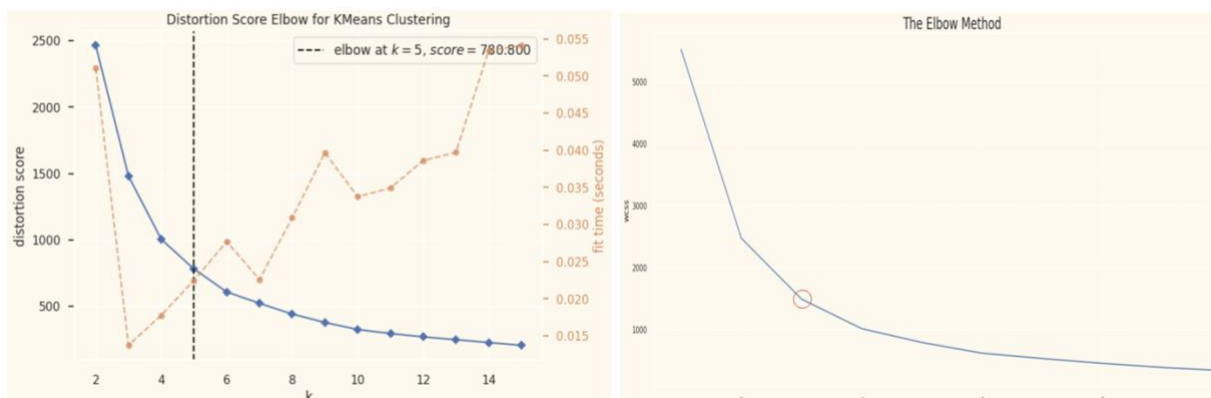


Figura 16. Metodo del punto di gomito con punteggio di distorsione (left) e la somma dei quadrati all'interno dei cluster (right).

Tuttavia, considerato l'obiettivo dello studio, si ritiene cruciale riuscire a catturare in maniera precisa e accurata la complessità dei dati provenienti da 51 paesi. Pertanto, una suddivisione del dataset in 5 cluster assicura una maggiore granularità nella divisione, consentendo di identificare tendenze più specifiche e sottolineare le sfumature nei comportamenti dei clienti nelle diverse regioni geografiche. Questa approfondita analisi consentirà, inoltre, una migliore comprensione delle relazioni tra i contenuti dei post sulle borse e le posizioni geografiche vicine, migliorando così la precisione delle conclusioni e la rilevanza delle strategie di marketing.

Di conseguenza, per il proseguimento del processo di data mining, verranno presi in considerazione 5 cluster.

Per quanto riguarda gli algoritmi, sono stati applicati 4 algoritmi di clustering:

- K - Means: è l'algoritmo più utilizzato e più semplice; si tratta di un algoritmo basato sul centroide che ha come obiettivo quello di ridurre al minimo la varianza dei punti dati all'interno di un cluster. Inizia selezionando casualmente k centroidi, dove k è il numero desiderato di cluster. Successivamente, assegna ogni punto dati al cluster il cui centroide è più vicino e ricalcola i centroidi come la media di tutti i punti nel cluster. Questi due passaggi vengono ripetuti fino a quando i centroidi smettono di cambiare o per un numero massimo di iterazioni (McGregor, 2020).
- Agglomerative Clustering: è il tipo più comune di algoritmo di clustering gerarchico; si tratta di una forma di clustering dal basso verso l'alto che permette di costruire una

gerarchia di cluster in cui ogni punto inizia come un cluster separato e viene successivamente unito a cluster vicini fino a creare un singolo cluster o un numero desiderato di cluster (McGregor, 2020).

- Spectral Clustering: è una tecnica EDA (Exploratory Data Analysis) non convenzionale che riduce set di dati multidimensionali complessi in cluster di dati simili in dimensioni più rare. Si tratta di un algoritmo di clustering che utilizza la teoria dei grafi per raggruppare insieme campioni simili e che utilizza l'approccio di connettività al clustering, in cui le comunità di nodi (cioè i punti dati) che sono collegati o immediatamente uno accanto all'altro sono identificati in un grafico. I nodi vengono quindi mappati in uno spazio a bassa dimensione che può essere facilmente segregato per formare cluster. Lo Spectral Clustering utilizza le informazioni dagli autovalori (spettro) di matrici speciali (es. matrice di affinità, matrice dei gradi e matrice laplaciana) derivate dal grafico o dal set di dati (Chatterjee, 2022).
- DBSCAN: è un algoritmo di clustering basato sulla densità che funziona sul presupposto che i cluster siano regioni dense nello spazio separate da regioni di densità inferiore. Raggruppa, quindi, punti dati "densamente raggruppati" in un singolo cluster. La caratteristica più interessante del clustering DBSCAN è che è robusto per i valori anomali e che non richiede che venga comunicato in anticipo il numero di cluster. DBSCAN utilizza due parametri per determinare come vengono definiti i cluster: *MinPts* (il numero minimo di punti dati che devono essere raggruppati insieme affinché un'area sia considerata ad alta densità) ed *Eps* (la distanza utilizzata per determinare se un punto dati si trova nella stessa area degli altri punti dati) (Sharma, 2023).

Capitolo 4 – Risultati e Discussioni

4.1 Performance e risultati

4.1.1 Risultati di Clustering su Scala Mondiale

Al fine di esaminare la differenza nelle tendenze di moda relative alle borse tra i vari paesi del mondo e di identificare le ragioni di eventuali similitudini tra questi paesi, è stata condotta un'analisi su un dataset di post pubblicati su Instagram, selezionati in base agli hashtag associati. Le caratteristiche considerate per questa analisi includono il momento e la posizione della pubblicazione dei vari post, gli hashtag indicati e associati ad essi, e informazioni sulla popolarità del post, come il numero di like e i commenti ricevuti.

Una volta completate le fasi di preparazione del dataset, comprese l'analisi, la pulizia dei dati, la trasformazione delle variabili categoriche in numeriche, la standardizzazione dei dati, la riduzione della dimensionalità delle features a disposizione; nonché terminata la fase di esplorazione dei dati tramite analisi statistiche che ci hanno permesso di conoscere meglio le relazioni esistenti tra le caratteristiche scelte, si è proseguito con l'attuazione del processo di Clustering.

Per un'analisi più esaustiva, durante questa fase finale della sperimentazione sono state confrontate le performance di 4 differenti algoritmi di clustering: K-means clustering, Agglomerative Hierarchical clustering, Spectral clustering e DBSCAN clustering.

Per valutare, invece, la qualità dei cluster formati da questi algoritmi, sono state utilizzate tre metriche quantitative che valutano la qualità dei cluster formati: il Silhouette Score, il Davies Bouldin Index e il Calinski Harabaz Index (Vedi Figura 17).

Il Silhouette score è un indice che varia da -1 e 1 e calcola la qualità dei cluster utilizzando la distanza intra cluster; quindi, verificando quanto sono distanti e separati tra loro, i cluster ottenuti. Si tratta della metrica generalmente e più comunemente utilizzata. Il *Davies Bouldin Index*, invece, indica la misura media di somiglianza di ciascun cluster con il suo cluster più simile, dove la somiglianza è il rapporto tra le distanze all'interno del cluster e le distanze tra essi. Il valore minimo dell'indice è 0, mentre un valore più piccolo (più vicino a 0) rappresenta un modello migliore. Infine, il Calinski Harabax Index, anche definito CH score, valuta la separazione tra i cluster trovati e tiene conto della dispersione all'interno dei cluster utilizzando la varianza inter e intra cluster. Valori più alti di questo indice mostrano cluster più compatti e separati.

QUANTITATIVE METRICS	KMEANS	AGGLOMERATIVE	SPECTRAL	DBSCAN
<i>Silhouette score</i>	0.48	0.43	0.51	0.36
<i>Devies-Bouldin Index</i>	0.71	0.76	0.55	1.39
<i>Calinski Harabaz Index</i>	2140.00	1665.52	1175.20	125.58

Figura 17. Confronto dei risultati degli algoritmi K-means, Agglomerative, Spectral e DBSCAN.

Sulla base dei risultati ottenuti, l'algoritmo Spectral Clustering ha dimostrato di avere una performance superiore in termini di silhouette score (0.51) e Devies-Bouldin Index (0.55), mentre l'algoritmo K-Means Clustering ha raggiunto risultati migliori in termini di Calinski Harabaz Index (2140.00).

Quindi, lo Spectral Clustering ha mostrato di avere una maggiore coesione all'interno dei cluster e una migliore separazione tra di essi, come evidenziato dal valore, considerevole sufficientemente buono, del silhouette score. D'altra parte, K-Means Clustering ha mostrato una maggiore efficacia nell'assegnare i dati ai cluster e nell'individuare cluster densi con una buona separazione tra di essi.

Per una comprensione più dettagliata dei risultati e per offrire una visualizzazione intuitiva di come i dati sono stati suddivisi nei cluster, abbiamo adottato un approccio visuale. In particolare, abbiamo esplorato la distribuzione dei cluster generati da ciascun algoritmo e come i dati sono distribuiti al loro interno attraverso grafici di dispersione bidimensionali (Scatterplot) come mostrato nella Figura 18.

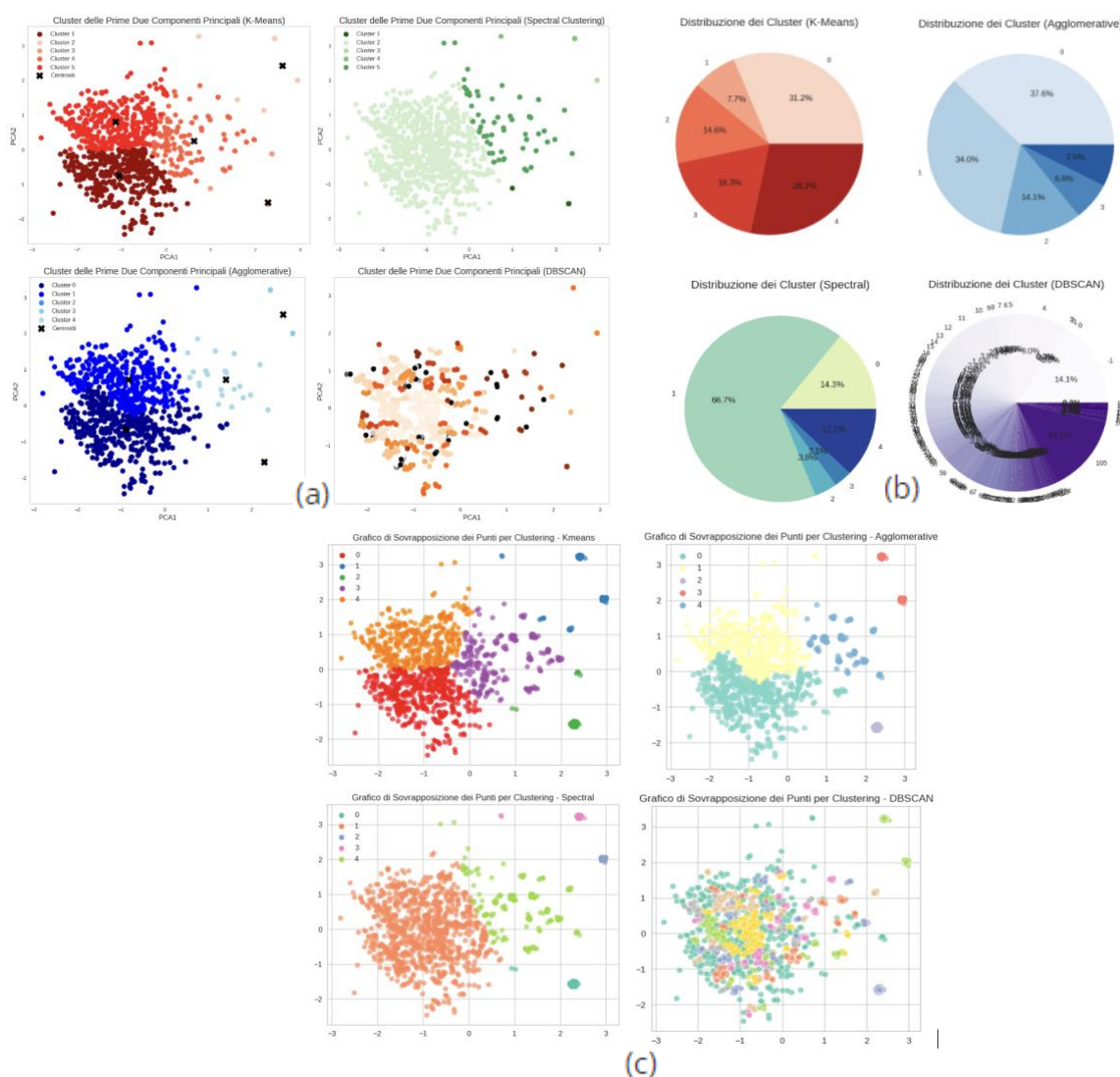


Figura 18. *Top Left:* Scatterplot dei cluster risultanti dagli algoritmi. *Top Right:* Diagramma a torta dei cluster risultanti dagli algoritmi. *Bottom:* Scatterplot con trasparenza per individuare sovrapposizioni tra i punti dei cluster.

Prendendo in considerazione gli output generati dagli algoritmi K-Means e Spectral, in quanto hanno performato meglio in termini quantitativi, i grafici mostrano una distribuzione dei dati più uniforme in K-Means (Figura 17a) e, considerato la natura e l'obiettivo dello studio, i punti appaiono comportarsi in maniera più coerente all'interno dei cluster individuati dal K Means Clustering anche in termini di distribuzione (Figura 17b). Invece, in generale, non sembra esservi alcun problema di sovrapposizione di punti (Figura 17c).

Di conseguenza, considerando il punteggio più elevato della metrica Calinski Harabax, l'uniformità maggiore dei cluster e una dispersione più adatta, possiamo concludere che l'algoritmo K-means ha dimostrato di avere una prestazione superiore in questo contesto di studio.

Individuato l'algoritmo di clustering che ha ottenuto migliori prestazioni, abbiamo creato una heatmap globale (vedi Figura 19) che mostra la distribuzione dei cluster su scala mondiale: ogni cluster è rappresentato da un colore univoco, consentendo di visualizzare facilmente le regioni del mondo in cui sono concentrati i dati appartenenti a ciascun gruppo individuato.



Figura 19. Mappatura dei cluster con codifica cromatica su scala globale.

Questa rappresentazione grafica ci fornisce un quadro geografico dei cluster, aiutandoci a identificare tendenze geografiche e concentrazioni di dati in diverse parti del mondo.

Possiamo notare che vi sono aree di alta concentrazione (o punti di hotspot) nei pressi del continente europeo e dell'Africa centrale, mentre si osservano aree di bassa concentrazione (o aree fredde) in Sud Africa, Nord America e Russia.

È importante sottolineare che i dati a nostra disposizione sembrano coprire praticamente tutte le aree del mondo, suggerendo che questa analisi abbia una portata globale. In particolare, i risultati del nostro studio hanno rivelato connessioni intriganti tra paesi geograficamente vicini, ad esempio, è possibile notare che paesi europei come Francia, Italia, Germania e Inghilterra condividono uno stesso cluster, indicando una certa uniformità nei comportamenti dei consumatori in questa regione. In modo simile, abbiamo rilevato che paesi dell'America del Sud come il Cile, il Brasile e la Repubblica Dominicana e il Nicaragua sembrano condividere un cluster simile, indicando una somiglianza nei comportamenti e nei gusti relativi alle borse.

In sostanza, questa rappresentazione geografica ci fornisce preziose informazioni sulle tendenze e le concentrazioni dei dati in tutto il mondo, evidenziando sia similitudini che differenze tra le diverse regioni.

4.1.2 Profili di Consumo: Caratteristiche distintive delle comunità online di moda

Dopo una prima visione generale della distribuzione dei cluster su scala globale e dei paesi coinvolti in ciascuno di essi, si è proseguiti ad un'analisi cluster-specifica così da associare loro un profilo e conoscerne maggiormente le caratteristiche intrinseche. Quindi, in questa fase, le features, sono state utilizzate come indicatori significativi per l'identificazione di pattern di tendenze di moda in ciascun gruppo di paesi.

Innanzitutto, sono state eseguite analisi più approfondite sui paesi caratterizzanti ciascun cluster.

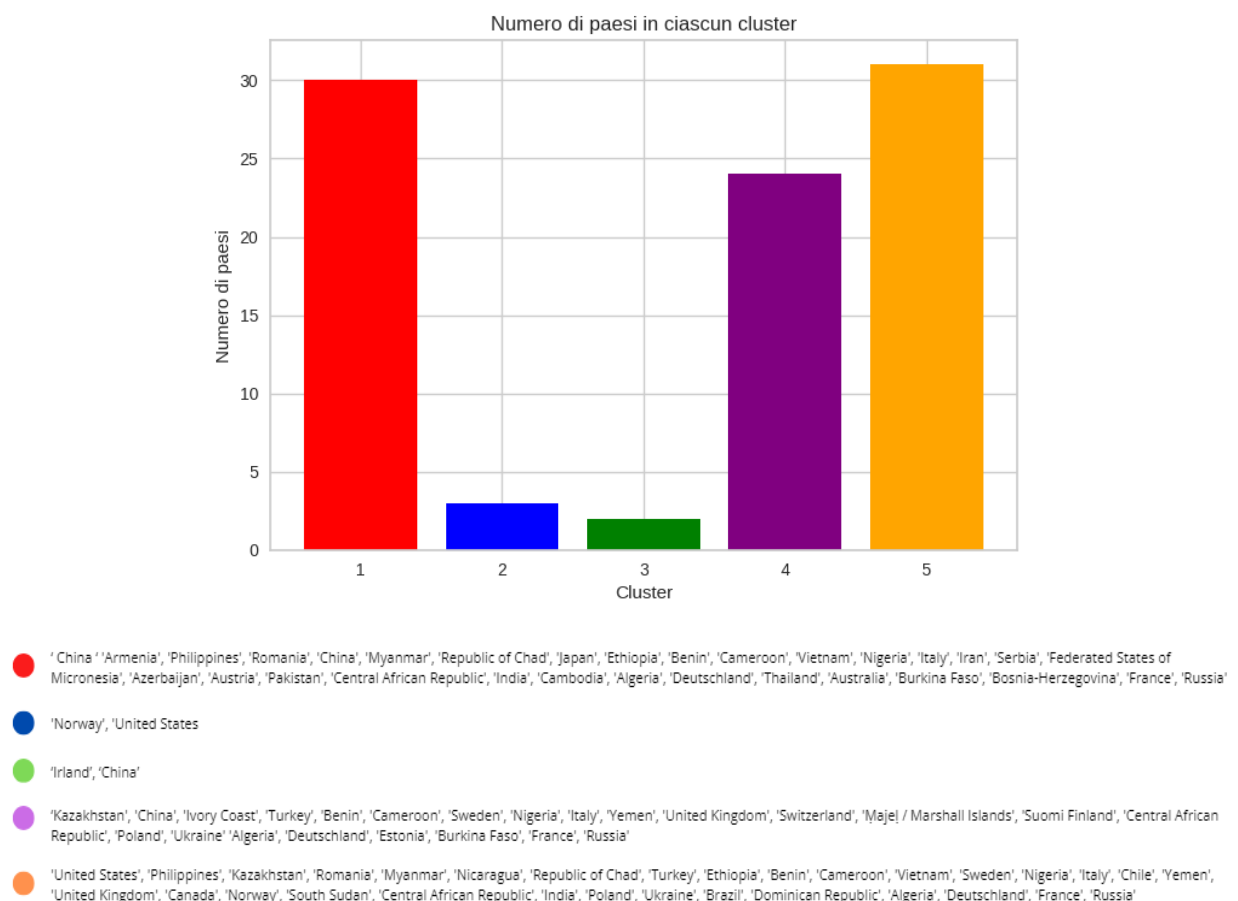


Figura 20. Distribuzione dei Paesi per Cluster: Grafico a Barre con Dettagli.

Guardando alla Figura 20, il cluster 5 sembra essere il più numeroso, con poco più di 30 paesi inclusi, a cui segue il cluster 1 con una leggera differenza. Il cluster 4 conta circa tra i 20 e i 24 paesi e infine i cluster 2 e 3 hanno un numero significativamente inferiore di paesi, oscillando

tra 0 e 5. Per riassumere, il cluster 5 è il più grande, seguito dal cluster 1, mentre i cluster 4, 2 e 3 hanno una presenza inferiore.

Inoltre, dall'elenco dei paesi stilato è possibile notare la presenza di paesi come "United States; Italy; United Kingdom; Russia, China" in più cluster.

Sono, poi, state esaminate le preferenze delle persone riguardo alla tipologia di borsa in ogni paese, al fine di comprendere i loro gusti (vedi Figura 21).

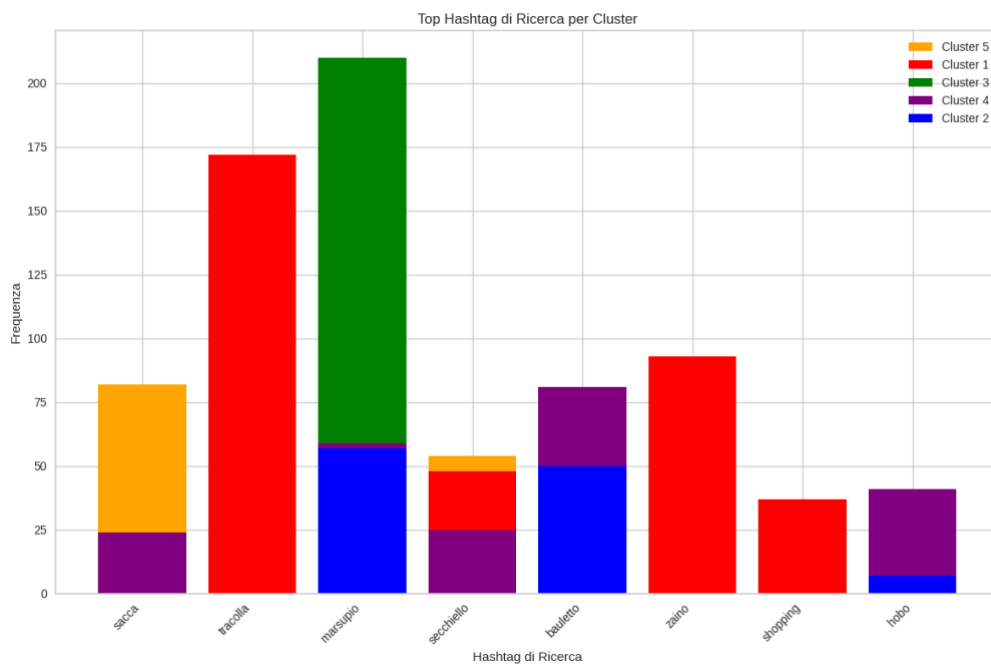


Figura 21. Grafico a barre degli hashtag di ricerca presenti in ciascun cluster.

Nel *cluster 1* emerge una serie di hashtag associati a diverse tipologie di borse, tra cui borse a tracolla, secchielli, zaini e shopping bag. Questo suggerisce un gruppo di consumatori orientato verso la praticità e la funzionalità, poiché queste borse sono note per la loro capacità di trasportare comodamente una varietà di oggetti senza compromettere la comodità. Inoltre, all'interno di questo cluster, si osserva una preferenza per uno stile più informale e versatile sia nell'abbigliamento che negli accessori: borse come quelle a tracolla e i secchielli si adattano facilmente a diverse occasioni, riflettendo un approccio alla moda flessibile. In sintesi, questi consumatori cercano soluzioni "hands-free" adattabili a uno stile di vita dinamico e pratico.

Il *cluster 2*, caratterizzato dalla ricerca di hashtag come "marsupio", "hobo", "bauletto" rappresenta, invece, un gruppo di consumatori che predilige borse di qualità, in quanto utilizzano termini e tipologie di borse solitamente tipiche dell'alta moda. Questo cluster potrebbe rappresentare individui che attribuiscono grande importanza all'aspetto distintivo e che sono orientate verso soluzioni eleganti e variegate.

Il *cluster 3*, focalizzato sul termine "marsupio" come hashtag rilevante, suggerisce un gruppo di consumatori con un interesse particolare per questo tipo di borsa.

Il *cluster 4*, con la predominanza di hashtag come "secchiello," "sacca," "bauletto," e "hobo," indica un gruppo di consumatori con una tendenza più eterogenea e non focalizzata su uno stile specifico. Infatti, questi acquirenti sembrano essere attratti da borse classiche e di design, come il "secchiello" e il "bauletto," che sono spesso considerate eleganti e adatte a varie occasioni. La menzione di "sacca", invece, suggerisce una predisposizione per borse spaziose e funzionali, mentre l'“hobo” potrebbe rappresentare la propensione verso uno stile boho-chic²⁵

Infine, i consumatori del *cluster 5*, con un alto livello di menzione per la borsa “sacca” sembrano orientati verso borse spaziose, funzionali e dalla forma versatile, ideali per le loro esigenze quotidiane.

In un secondo momento, è stata eseguita un'analisi prendendo in considerazione la variabile “hasht_cor”. L'analisi degli hashtag correlati rappresenta un componente cruciale nella ricerca, focalizzata sull'esplorazione delle dinamiche delle comunità online. In particolare, ho identificato gli hashtag più comuni all'interno di ciascun cluster tematico (Vedi Figura 22). Questo processo ha permesso di rivelare i trend e interessi predominanti all'interno di ciascun gruppo di utenti, contribuendo a una comprensione più profonda delle caratteristiche distintive delle diverse cluster communities.

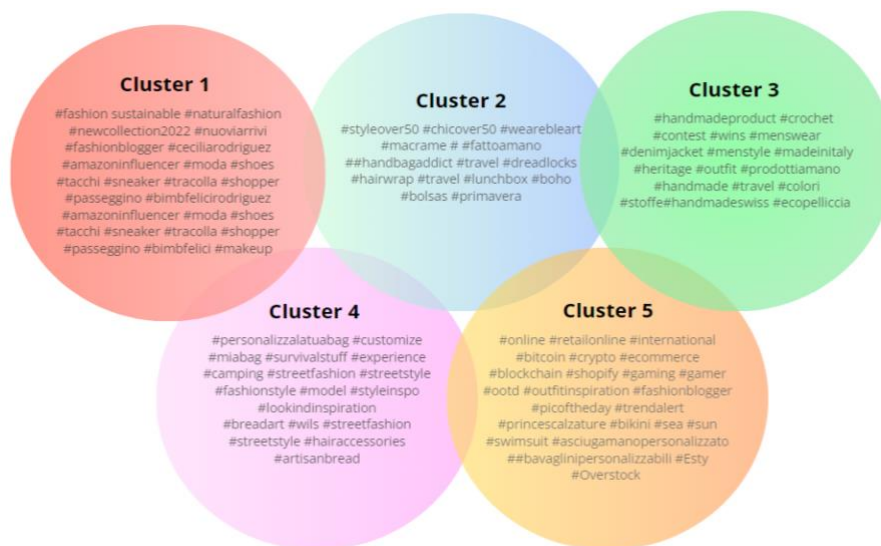


Figura 22. Elenco degli hashtag correlati, più frequenti, per ciascun cluster.

²⁵ Boho-chic: combinazione di elementi dello stile bohémien e dello stile chic, nato a Parigi e ispirato alla cultura Bohemien e alla cultura hippie degli anni '60 e '70. È caratterizzato da tessuti naturali, sovrapposizioni, influenze etniche, accessori vintage e oversize, creando un look bohémien sofisticato.

- All'interno del cluster 1, emerge un forte interesse per la sostenibilità, evidenziato dalla frequente comparsa degli hashtag #eco #fashionsustainable e #fashionnatural. Inoltre, il cluster sembra concentrarsi principalmente sulla moda e su accessori che riguardano sia borse che scarpe (#mocassino; #sneakers), con una particolare attenzione alle collezioni stagionali e ai nuovi arrivi (#newcollections #news). La presenza di hashtag legati all'infanzia (#passegino #infanziafelice) potrebbe suggerire che gli utenti dei post siano genitori (25-40 anni). La presenza di hashtag promozionali, come #promo2022 #promozioni, poi suggerisce una propensione ad acquistare durante le offerte speciali. Infine, la ripetizione degli hashtag legati agli influencer o alle celebrità indica un interesse per le tendenze di moda promosse da personalità di rilievo.
- Il cluster 2, invece, risulta interessato alla moda e agli accessori, in particolare borse con hashtag come #bolsacintura" e #handbagaddict. Inoltre, sembra esservi un interesse per stili personali e moda per persone sopra i 50 anni. Viaggi e avventure sono evidenziati con menzioni come #vanlife" e #travel." L'arte e la creatività emergono con l'hashtag #macrame #art. Infine, gli hashtag legati agli stili di capelli, come "dreadlocks" e "hairwrap" suggeriscono un interesse per stili di capelli unici.
- Il cluster 3 si contraddistingue per un forte interesse verso l'artigianato con hashtag come "handmade" e "crochet" molto frequenti. Vi è un interesse per contest e concorsi legati a prodotti fatti a mano. La fotografia di viaggio e l'esplorazione sono anch'essi rilevanti con hashtag come "travel" e "offroad." Infine, l'ampia menzione di "#madeinitaly" suggerisce un legame con l'alta moda italiana, confermata anche dall'utilizzo dell'hashtag "Heritage" che è un asset tipico dei prodotti di lusso.
- I consumatori appartenenti al cluster 4 sono caratterizzati da una varietà di interessi e passioni. Sono individualisti che amano esprimere la propria personalità attraverso la moda personalizzata, come visibile dall'uso frequente di hashtag #personalizzalabag #customize. Queste persone apprezzano la creazione di prodotti fatti a mano, come borse o accessori per capelli, e mostrano un interesse per la cucina artigianale (#breadart #countrybread). Allo stesso tempo, sono avventurosi, amanti della natura e del campeggio, evidenziando un desiderio di esperienze all'aperto e selvagge con hashtag #wild #survivalstuff #camping. Inoltre, gli hashtag #streetfashion #streetstyle mostrano un'attenzione verso la moda streetwear.
- Il cluster 5, infine, rappresenta consumatori creativi e versatili, particolarmente coinvolti nel canale dell'e-commerce, citando hashtag come #ecommerce #shopify #Etsy e

nell'utilizzo di criptovalute #NFT #XRP. Gli hashtag #Gaming #Gamer mostrano, allo stesso tempo, una passione per la pratica del gaming. Sono interessati, con un'attenzione speciale alla moda da spiaggia (#sea #switseat #bikini), agli accessori personalizzati e sembrano partecipare attivamente ai trend di moda online, quali #ootd #outfitinspiration #picoftheday. Nel complesso, sono consumatori online attivi e diversificati con una forte inclinazione verso la moda, la tecnologia e l'intrattenimento digitale.

Inoltre, abbiamo analizzato in modo più approfondito i cluster considerando l'orario di pubblicazione dei post. Questa analisi ci ha permesso di comprendere quando gli utenti e i consumatori di ciascun cluster sono più attivi sui social media, ovvero quando hanno più tempo libero e maggiori probabilità di dedicarsi a momenti di svago. Abbiamo estratto la feature "publish_hour" dalla feature "create_at", che indica il giorno e l'ora in cui è stato pubblicato ogni post e successivamente calcolato i momenti della giornata in cui c'è una maggiore attività per ciascun cluster.

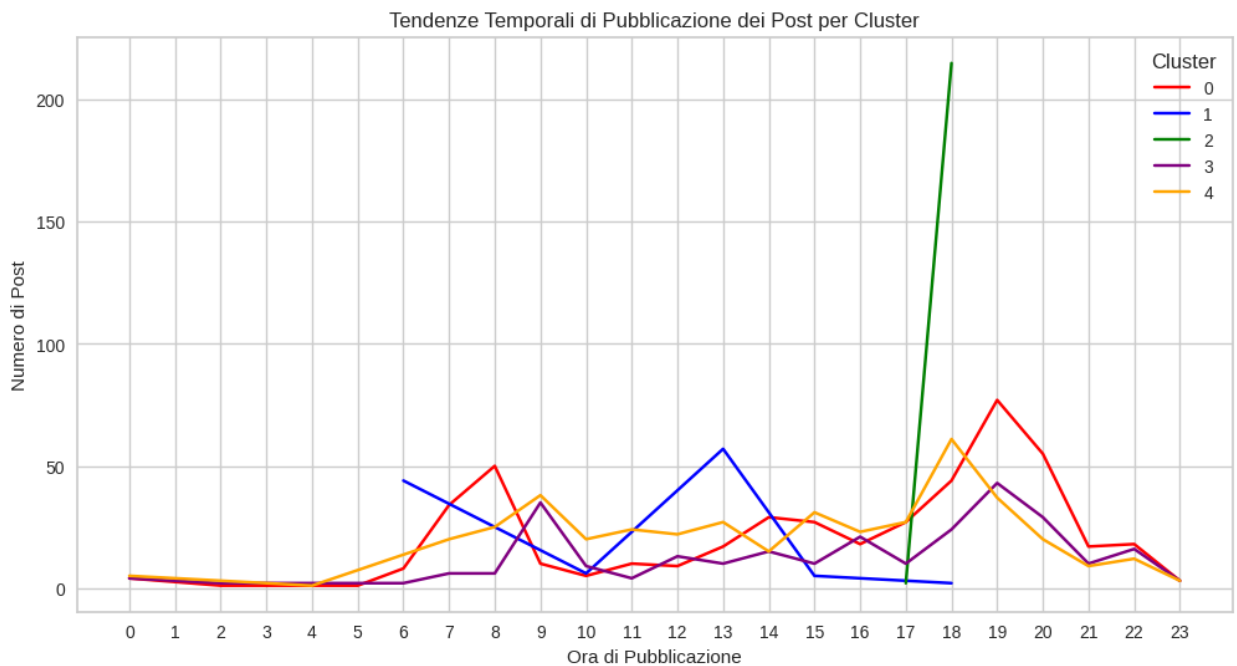


Figura 23. Grafico dell'orario di attività su Instagram dei cluster.

All'interno del Cluster 1, gli utenti sembrano concentrare principalmente le loro pubblicazioni tra le 18:00 e le 20:00, con un picco di attività intorno alle 19:00. Tuttavia, è interessante notare che mostrano anche una certa attività la mattina, soprattutto tra le 7:00 e le 9:00. Nel Cluster 2, i paesi hanno una preferenza per la pubblicazione tra le 11:00 e le 14:00, con un picco massimo alle 13:00. Nel Cluster 3, c'è una chiara preferenza per le pubblicazioni tra le 17:00 e le 18:00, con un'alta concentrazione di attività in quel momento. Il Cluster 4 mostra una maggiore attività

tra le 17:00 e le 21:00, ma anche una presenza significativa la mattina, soprattutto tra le 8:00 e le 10:00. Infine, il Cluster 5 mostra un'attività abbastanza costante per gran parte della giornata, con un picco nel pomeriggio tra le 17:00 e le 21:00.

In generale, fatta eccezione per il Cluster 1 che predilige l'intervallo di pranzo, si osserva che i paesi appartenenti agli altri cluster tendono a pubblicare più frequentemente nel tardo pomeriggio (18:00 - 20:00) seguito dalla fascia oraria tra le 7:00 e le 10:00 della mattina, che appare anche particolarmente attiva. Inoltre, è stata approfondita l'analisi per paese in ciascun cluster (Vedi Figura 24).

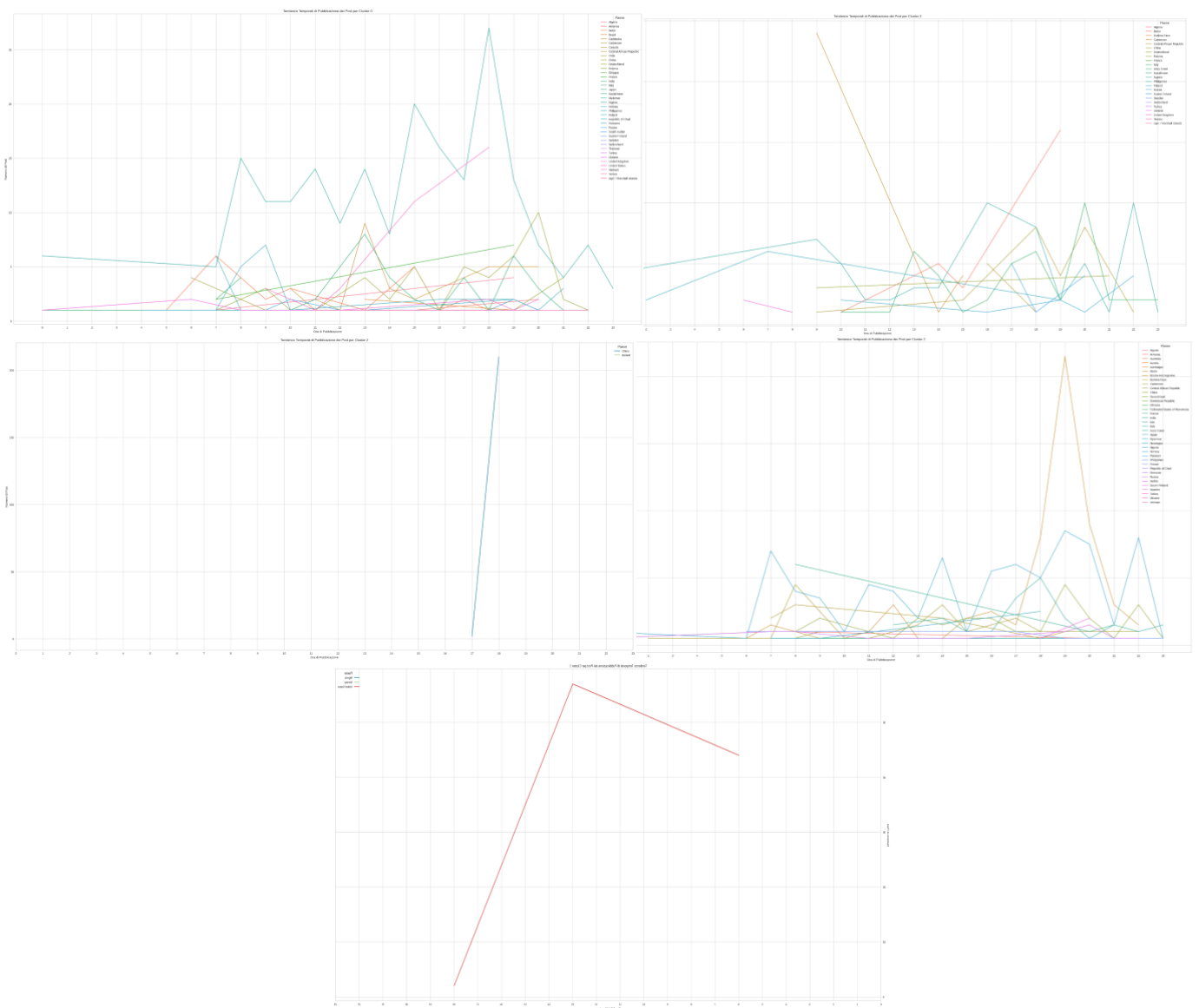


Figura 24. Grafico dettagliato dell'orario di attività su Instagram per ciascun paese all'interno dei rispettivi cluster.

Successivamente, è stata effettuata una comparazione dell'engagement presente in ciascun cluster utilizzando le feature "edge_liked_by_count" e "media_to_comment_count", che misurano le interazioni in termini di like e commenti ricevuti da ogni post.

Come possibile vedere nella Figura 25, il cluster 2 spicca come il cluster con il più alto numero di "mi piace", raggiungendo circa 250 likes. Segue il cluster 4 con quasi 150 "mi piace", mentre il cluster 3 riceve poco più di 50 likes. Infine, il cluster 1 e il cluster 5 ottengono un numero nettamente inferiore di apprezzamenti.

In merito ai commenti, il cluster 3 evidenzia il più alto livello di interazioni con più di 200 commenti, ma a discapito degli altri cluster; infatti, il cluster 2 segue con circa 65 like, mentre i cluster 1, 4 e 5 presentano un numero molto basso di commenti, appena superiori allo 0.

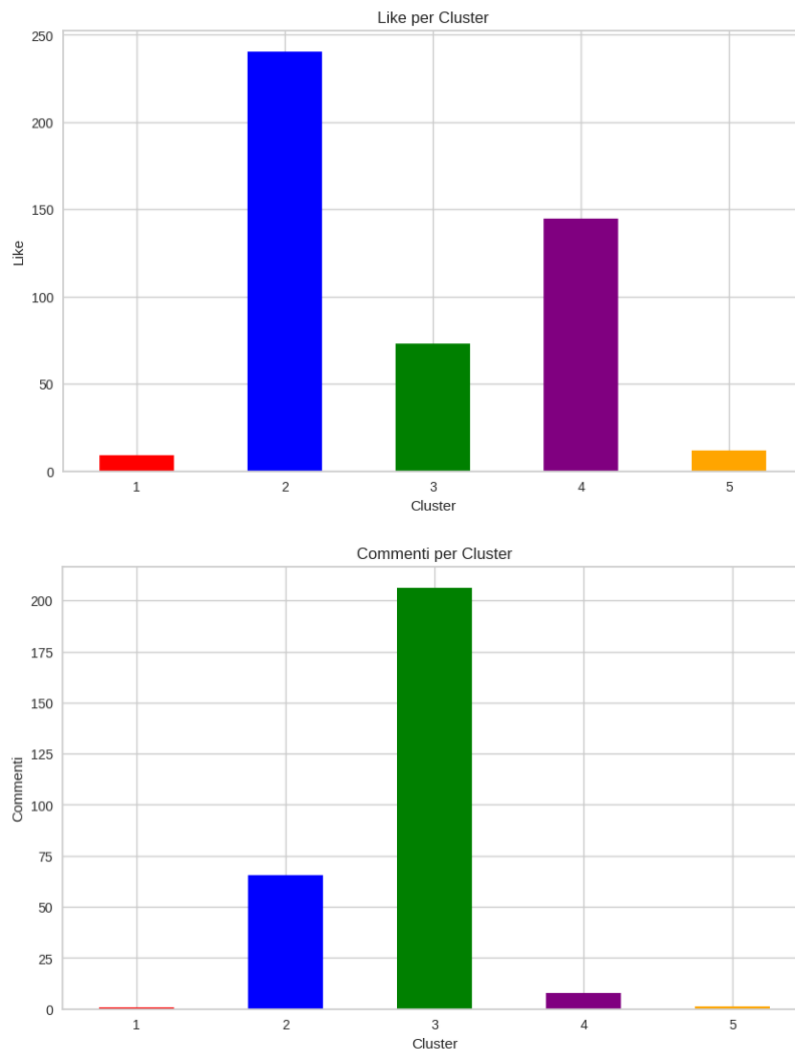


Figura 25. Grafico a barre delle interazioni basate su "mi piace" e "commenti" nei cluster.

Tuttavia, risulta necessario notare che alcuni paesi mostrano caratteristiche simili nonostante siano situati in posizioni geografiche distanti o, in generale, non siano in prossimità l'uno dell'altro. Questo può essere dovuto a fattori di diversa natura: uno di natura tecnica e si riferisce all'analisi stessa, includendo la possibile presenza di un margine di errore nell'ambito dello studio condotto, un altro di natura più ampia e concettuale legato, cioè, al fenomeno della globalizzazione.

Da un lato prettamente tecnico, l'adozione di metodologie avanzate di raccolta dati e analisi, come l'intelligenza artificiale, richiede sempre una valutazione critica della possibilità che vari fattori, come la qualità dei dati input, l'algoritmo scelto o le tecniche di validazione, possano introdurre discrepanze tra i risultati del clustering ottenuti e la "verità di riferimento" o struttura intrinseca dei dati stessi.

D'altra parte, il processo di globalizzazione ha profondamente influenzato le preferenze dei consumatori su scala mondiale. Grazie alla diffusione delle comunicazioni e del commercio internazionale, le influenze culturali e le tendenze di moda ora superano le tradizionali barriere geografiche. Questa interconnessione globale è particolarmente evidente nei social media, dove le persone possono condividere idee e stili di vita in tempo reale, oltre a essere in contatto con individui provenienti da ogni parte del mondo. Pertanto, è plausibile che questa diffusione globale di influenze abbia contribuito a creare somiglianze nelle preferenze per le borse tra nazioni distanti. Basti pensare che, secondo l'analisi condotta da HootSuite, il totale degli utenti unici sulle piattaforme social è cresciuto di 520 milioni solo nell'ultimo anno, con una crescita annua di oltre il 13%. Ciò significa che più di uno su nove degli utenti di social media di oggi ha iniziato a utilizzare piattaforme social per la prima volta negli ultimi 12 mesi (Kemp, 2022). In definitiva, la disponibilità di informazioni e l'interazione culturale stanno unificando in modo più ampio le scelte dei consumatori, attenuando le differenze geografiche nei comportamenti dei consumatori in generale, e probabilmente anche nell'acquisto di borse.

Inoltre, seppur non visibile graficamente dalla heatmap sopra riportata, è importante evidenziare che alcuni paesi come Gran Bretagna, Stati Uniti, Francia, Italia, Cina appaiono in più cluster. Questo risultato suggerisce una connessione significativa con il fenomeno della globalizzazione. Si tratta, infatti, di paesi riconosciuti come autentici centri di diffusione delle tendenze nell'ambito della moda e dell'industria dei beni di lusso, il che spiega perché le loro preferenze per le borse si manifestino in molteplici cluster. La loro influenza globale si traduce nella convergenza di tendenze e stili di moda, trasformandoli in nodi fondamentali per le analisi di comportamento dei consumatori a livello internazionale.

Questo studio di geolocalizzazione offre non solo una visione geografica di come le tendenze di moda si diffondano a livello globale, ma anche l'opportunità di acquisire preziosi insight sulle caratteristiche degli acquirenti di borse di diverse parti del mondo, fornendo alle aziende e ai marketer l'opportunità di creare prodotti, servizi e campagne pubblicitarie che rispondono in modo più preciso alle esigenze e alle preferenze dei consumatori; quindi migliorando l'efficacia complessiva delle proprie iniziative di marketing.

In particolare, questo studio permette di individuare 5 gruppi di consumatori diversi:

Profilo 1: Eco-Pratici Trend Explorers

Questi consumatori, localizzati principalmente nei paesi asiatici, sono caratterizzati da un interesse per borse pratiche e funzionali, come borse a tracolla e secchielli, riflettendo uno stile di vita dinamico. Mostrano una sensibilità alla sostenibilità e sono attenti alle promozioni e alle nuove collezioni. Questi utenti sembrano concentrare le loro attività sulle piattaforme social tra le 18:00 e le 20:00, con un picco alle 19:00, quando esplorano le tendenze di moda influenzate da celebrità e influencer. Sebbene pubblicano regolarmente, i loro post tendono a ricevere un basso numero di "mi piace" e "commenti". In generale, cercano un equilibrio tra praticità, moda e sostenibilità nelle loro scelte di abbigliamento e accessori.

Al fine di raggiungere in maniera più efficiente il target degli Eco-Pratici Trend Explorers, le aziende potrebbero focalizzare l'attenzione sul proprio impegno sostenibile. Ad esempio, sviluppando una linea di borse Eco-Friendly realizzata con materiali riciclati o biodegradabili oppure collaborando con associazioni naturalistiche/ambientaliste a cui devolvere una % del ricavato della propria linea eco-friendly. Allo stesso tempo, le aziende potrebbero comunicare il proprio valore della sostenibilità attraverso eventi, workshop o conferenze apposite. Inoltre, poiché i consumatori asiatici sono attenti alle offerte e alle promozioni, le aziende potrebbero strutturare campagne di marketing incentrate su sconti e promozioni occasionali per incentivare l'acquisto delle borse e concentrare la propria tattica di comunicazione sull'influencer marketing, collaborando quindi con talents, visto che questi consumatori sono interessati e influenzati da persone di rilievo.

Profilo 2: Travel and Quality Enthusiasts

Questo profilo si compone di consumatori localizzati principalmente nei paesi degli Stati Uniti; rappresenta una fascia di consumatori appassionati di moda, inclini a investire in borse di alta qualità e interessati a esplorare stili personali, viaggi, arte ed esperienze creative. Sono attivi su Instagram durante il giorno in particolare tra le 11:00 e le 14:00. Sono noti per ottenere un

generale un significato coinvolgimento sui loro post, sia in termini di “mi piace” che di commenti, riflettendo l'apprezzamento dei loro follower per i loro contenuti.

Dato l'interesse per borse di alta qualità, le aziende dovrebbero concentrarsi sulla promozione di prodotti che offrono eleganza e distinzione, utilizzare materiali pregiati e design sofisticati può attrarre il pubblico dei Travel and Quality enthusiasts.

L'interesse per viaggi e avventure indica un'opportunità per promuovere borse adatte a queste esperienze, ad esempio le aziende potrebbero realizzare una linea a tema viaggio, o progettate per esperienze e avventure. In aggiunta, considerando il forte interesse di questo pubblico per l'arte, le aziende potrebbero esplorare partnership con artisti contemporanei che rispecchino l'identità del brand, creando item esclusivi e particolarmente creativi.

Allo stesso modo, rispettando gli interessi dei consumatori, le aziende potrebbero offrire promozioni speciali o sconti per chi acquista una borsa e condivide una foto del prodotto in un'ambientazione di viaggio avventurosa sui social media; questo può incoraggiare i clienti a condividere le loro esperienze di viaggio e a promuovere il marchio.

Profilo 3: Craftsmen of Elegance and Adventure

Questi consumatori, presenti soprattutto in Irlanda e in Cina, sono appassionati di artigianato, avventura e moda di alta qualità, con un particolare amore per l'artigianato fatto a mano. Mostrano interesse per i viaggi e hanno una forte connessione con l'alta moda italiana, come dimostrato dall'hashtag "#madeinitaly". Preferiscono pubblicare contenuti tra le 17:00 e le 18:00, e il loro alto coinvolgimento, con oltre 200 commenti, suggerisce un pubblico molto impegnato e interessato alle marche.

Per connettersi in modo efficace con i Craftsmen of Elegance and Adventure, le aziende dovrebbero focalizzare la loro strategia di marketing sull'eccellenza della qualità e sull'artigianato. Questo consente loro non solo di differenziarsi sul mercato creando una reputazione prestigiosa ma anche di conquistare la fedeltà dei clienti. Ad esempio, possono costruire connessioni emotive con i consumatori attraverso narrazioni autentiche sul marchio e sugli artigiani. Inoltre, possono realizzare spot pubblicitari o contenuti che mostrano la creazione dei prodotti, coinvolgendo il pubblico nella loro passione per il lavoro manuale. Inoltre, potrebbero coniugare l'interesse per le avventure e in tema di contest di questo target, organizzando concorsi di foto dei clienti durante i loro viaggi con i prodotti dell'azienda.

Profilo 4: Custom Style Explorers

I consumatori, localizzati principalmente nel continente europeo, hanno uno spettro di interessi diversificato per borse di design, mostrando preferenze per stili come secchielli, bauli, sacche

e hobos. Questi individui amano personalizzare la loro moda e sono creativi. Presentano una vasta gamma di interessi, dall'artigianato culinario all'avventure all'aperto e allo stile streetwear. Preferiscono pubblicare contenuti principalmente tra le 17:00 e le 21:00, ma sono attivi anche al mattino. Sono un pubblico ideale per aziende che offrono borse versatili adatte a una varietà di stili e occasioni.

Dato che questi consumatori mostrano preferenze per una varietà di stili di borse, le aziende dovrebbero avere portfolio prodotto che includa secchielli, bauli, sacche e hobos, Questo permette loro di soddisfare una vasta gamma di gusti, ampliando il pubblico di potenziali clienti. Di conseguenza, aumenta la soddisfazione del cliente, incrementa le vendite e posiziona il marchio come versatile e adatto alle esigenze di una vasta gamma di consumatori. Inoltre, dato il bisogno di unicità e l'interesse per prodotti personalizzati, le aziende di moda potrebbero permettere ai clienti di personalizzare una borsa di collezione, facendo loro scegliere il colore, il materiale o altre caratteristiche del prodotto, migliorando il coinvolgimento e la fidelizzazione dei consumatori.

Profilo 5: Gaming Fashion Lovers

Questi consumatori, localizzati principalmente in America latina e in Africa, in Canada sono appassionati di borse spaziose e versatili, come le "sacche" ideali per le loro esigenze quotidiane. Questi individui sono altamente coinvolti nell'e-commerce e nella pratica del gaming. Sono anche amanti della moda da spiaggia e degli accessori personalizzati. Partecipano attivamente ai trend di moda online e sono diversificati nei loro interessi, spaziando dalla moda alla tecnologia e all'intrattenimento digitale. Mantengono un'attività sui social media costante durante gran parte della giornata, con un picco nel pomeriggio. Presentano un numero relativamente basso di interazione nei loro post.

Data la propensione di questi consumatori nel rivolgersi a piattaforme di e-commerce, le aziende dovrebbero potenziare e consolidare la loro presenza su questo canale di vendita e mettere in atto strategie di marketing mirate all'online shopping.

Inoltre, dato l'interesse verso il mondo del gaming, le aziende potrebbero attuare tattiche di marketing in questo ambito, volte a migliorare l'engagement dei propri clienti. Ad esempio, potrebbero collaborare con uno sviluppatore di giochi per creare un gioco online che coinvolga direttamente le borse come elemento chiave; potrebbero creare un app mobile di gioco dedicata alle borse, in cui i giocatori possono collezionare, personalizzare e sfoggiare le borse nei loro avatar virtuali. Questo può includere sfide o eventi speciali all'interno del gioco.

Questi sono solo alcuni esempi di come le aziende possono utilizzare queste informazioni come vantaggio per il proprio business. In generale, questa ricerca fornisce alle aziende di moda una serie di preziosi insights manageriali e di marketing, offrendo una panoramica dettagliata delle preferenze dei consumatori in varie regioni del mondo, nonché delle diverse caratteristiche dei loro clienti. Questo permette loro di prendere decisioni strategiche e operative informate e di prosperare nel settore delle borse e della moda sia a livello locale che globale.

4.3 Limitazioni e sviluppi futuri

Nonostante gli insight significativi emersi da questo studio, è importante considerare alcune limitazioni che potrebbero influenzare la sua applicabilità.

In primo luogo, l'analisi si basa su dati raccolti da piattaforme di social media, in particolare Instagram, il che significa che il campione scelto potrebbe non essere rappresentativo e quindi non riflettere completamente le preferenze e i comportamenti di utenti interessati alle borse.

Inoltre, potrebbe esserci una selezione non casuale dei post basata sugli hashtag utilizzati, il che potrebbe introdurre un *bias* nei dati. Questo avviene perché gli utenti utilizzano gli hashtag in modi variabile: alcuni optano per hashtag specifici, altri preferiscono quelli generici o evitano del tutto di utilizzarli. Anche la popolarità degli hashtag può influenzare la quantità di post raccolti, poiché hashtag molto popolari possono generare un numero maggiore di post rispetto a quelli meno noti, anche se potrebbero esserci post rilevanti tra questi ultimi. Inoltre, l'abuso degli hashtag da parte degli utenti che cercano di aumentare la visibilità dei propri post, anche quando non sono pertinenti allo studio, può distorcere l'analisi. Infine, l'utilizzo degli hashtag potrebbe variare in base alle differenze culturali dei paesi, ad esempio le abitudini di utilizzo dei social media possono variare da cultura a cultura, alcuni paesi potrebbero utilizzare meno hashtag o menzionare meno esplicitamente la loro posizione geografica nei post rendendo difficile la standardizzazione dei dati.

Pertanto, è essenziale considerare questi fattori per evitare conclusioni distorte nell'analisi dei comportamenti dei consumatori di borse.

Un'altra limitazione significativa riguarda la disparità nella quantità di post tra i paesi che può rendere difficile il confronto equo tra i paesi e potrebbe influenzare la rappresentatività e l'interpretazione dei risultati non riflettendo a pieno le peculiarità dei comportamenti dei consumatori di borse in quella regione.

Infine, un'ultima limitazione riguarda la metodologia di clustering utilizzata, che potrebbe essere affinata ulteriormente per migliorare la precisione nell'identificare similitudini tra regioni.

Tra i possibili sviluppi futuri che potrebbero ampliare l'efficacia di questo studio vi è l'espansione delle fonti dei dati utilizzati. Un'analisi multiplatforma, che prevede l'inclusione di dati da più piattaforme di social media potrebbe fornire una visione più completa dei comportamenti dei consumatori online. Questo potrebbe includere dati da Facebook, Twitter, TikTok e altre piattaforme popolari, consentendo una visione più completa delle preferenze dei consumatori e delle tendenze emergenti e consentendo alle aziende di adattare strategie di marketing in modo ancora più mirato, consapevole e informato.

Inoltre, per ovviare al problema di possibili *bias* presenti nei dati, si potrebbe procedere ad un maggior affinamento di questo studio di geolocalizzazione attraverso l'implementazione preliminare di un approccio di apprendimento supervisionato alle immagini nei post raccolti. Questo permetterebbe una verifica più accurata della coerenza e della rappresentatività dei dati associati alle borse. Attraverso l'uso di algoritmi di riconoscimento dell'immagine, sarebbe possibile identificare in modo più preciso se un post effettivamente riguarda borse o se potrebbe essere classificato in modo più accurato in altre categorie. Questo processo potrebbe aiutare a eliminare rumori nei dati e a garantire che l'analisi si basi su informazioni di alta qualità, migliorando la precisione delle conclusioni e delle raccomandazioni per le aziende nel settore della moda e delle borse.

Infine, uno sviluppo futuro molto interessante per il presente studio di geolocalizzazione, sarebbe esplorare l'applicazione di tecniche di deep learning. L'adozione di queste potenti e sofisticate metodologie, capaci di gestire dati complessi come quelli provenienti dai social media, potrebbe condurre a una comprensione più profonda delle tendenze geografiche nei comportamenti dei consumatori di borse.

CONCLUSIONE

In un mondo sempre più interconnesso e guidato dalla tecnologia, il settore della moda e delle borse sta attraversando una trasformazione epocale. Questa rivoluzione digitale ha radicalmente influenzato l'industria della moda, ridefinendo la concezione, l'acquisto e l'esperienza dei prodotti di moda e degli accessori per le persone. In questo contesto in continua evoluzione, l'intelligenza artificiale e la localizzazione si sono dimostrate potentissimi strumenti a disposizione delle aziende che desiderano comprendere meglio i propri clienti e offrire loro esperienze altamente personalizzate e soddisfacenti.

Nel presente elaborato, abbiamo eseguito uno studio di geolocalizzazione volto ad esaminare il profondo impatto che queste recenti tecnologie hanno sul settore, analizzando approfonditamente i comportamenti dei consumatori in varie regioni del mondo.

Attraverso un'analisi approfondita dello stato dell'arte, abbiamo notato come, spesso, le preferenze e i gusti dei consumatori siano influenzati dai fattori geografici e culturali che li caratterizzano. Nel perseguire il nostro obiettivo di esplorare e comprovare questa correlazione, abbiamo osservato che le preferenze, gli interessi, le modalità di acquisto, gli orari di attività sui social media e persino l'engagement ricevuto nei post presentano notevoli differenze da una regione all'altra. Questo ha implicazioni significative per le strategie di marketing delle aziende, che devono essere flessibili e adattabili per rispondere alle esigenze di un pubblico globale eterogeneo.

Uno dei risultati chiave di questa ricerca è stato evidenziare la diversità dei consumatori globali; tuttavia, abbiamo anche esaminato come le aziende possono sfruttare l'IA per analizzare grandi quantità di dati dai social media al fine di individuare trend emergenti. In particolare, lo studio ha identificato cinque distinti gruppi di consumatori in tutto il mondo, ognuno dei quali con interessi, gusti e caratteristiche affini, ma distinguibili dagli altri. Questo approfondimento sulle differenze o sulle sfumature che possono sussistere tra i comportamenti dei consumatori situati in diverse località del mondo offre alle aziende preziose opportunità di adattare in modo mirato le proprie strategie di marketing e di prodotto.

In aggiunta, abbiamo enfatizzato come la globalizzazione possa spiegare l'esistenza di interessi condivisi anche tra paesi geograficamente distanti. Infatti, la digitalizzazione, tra le molteplici sue sfaccettature, offre un'opportunità unica di connettere persone in tutto il mondo, superando qualsiasi barriera geografica. Questa rete globale crea un tessuto sociale e culturale in continua evoluzione, aumentando significativamente la possibilità di scambio e influenza reciproca tra

diverse comunità, con profonde implicazioni per l'industria della moda. La condivisione di idee, tendenze e esperienze avviene senza soluzione di continuità, contribuendo a plasmare in modo dinamico il panorama globale della moda.

In conclusione, l'IA e la geolocalizzazione hanno aperto nuove strade a questo settore, consentendo alle aziende di offrire esperienze più personalizzate e di adattarsi rapidamente ai cambiamenti del mercato. Tuttavia, per avere successo in questo ambiente in rapida evoluzione, le aziende devono essere agili, etiche e sempre orientate al cliente. Il futuro della moda appartiene a coloro che sanno sfruttare al meglio la potenza delle tecnologie emergenti, senza perdere di vista il valore intrinseco dell'arte e della creatività.

BIBLIOGRAFIA

- Agrawani S. (2022). "The Role of Artificial Intelligence (AI) in Personalized Marketing in Fashion and Apparel Industry."
- Alessandro Piva. (2019). "le 5v dei big data: le caratteristiche di una massa di dati."
- Al-Halah Z., Grauman K. (2020). "From Paris to Berlin: Discovering Fashion Style Influences Around the World."
- Alsaleh A. D., Elliot T. M., Fu Q. F., Thakur R. (2019). "Cross-cultural differences in the adoption of social media."
- Amaldoss, W. and Jain, S. (2005). "Pricing of conspicuous goods: a competitive analysis of social effects."
- Angevine C., Keomany J., Tomsen J. and Zemelle R. (2021). "Implementing a digital transformation at industrial companies."
- Brock J. K. -U., von Wangenheim F. (2019). "Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*."
- Calaskova K., Durana P., Adamko P. (2021). "Changes in Consumers' Purchase Patterns as a Consequence of the COVID-19 Pandemic."
- Chalmers D., Mackenzie G. N. & Carter S. (2020). "Artificial Intelligence and Entrepreneurship: Implications for Venture Creation in the Fourth Industrial Revolution."
- Davenport T., Guha A., Grewal D., Bressgott T. (2019). "How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*."
- D' Vaz - Sterling A. (2021). "Can Digitization Help Fashion Bridge Its Online And Offline Customer Relationships?"
- Deloitte. (2022). "How to leverage AI in marketing: three ways to improve consumer experience."
- Ding D., Zhang M., Pan X., Wu D., & Pu P. (2018). Geographical Feature Extraction for Entities in Location-based Social Networks.
- Dutta B. (2021). "10 Applications of Artificial Intelligence (AI) in Business."
- Forbes. (2019). "How Nike is using analytics to personalize their customer experience."
- Forbes. (2021). "How AI Is Transforming The Future Of Digital Marketing."
- Forbes. (2022). "The Top Five Ways AI Is Transforming Business."
- Forbes. (2023). "Artificial Intelligence in Fashion."
- Forbes. (2023). "24 Top AI Statistics and Trends In 2023."
- Gartner. "Digitalization."

- Gawehn E., Hiss J. A., Schneider G. (2016). “Deep learning in drug discovery. Molecular Informatics.”
- Geri C., Jain S., Zeng X., Bruniaux P. (2019). “A Detailed Review of Artificial Intelligence Applied in the Fashion and Apparel Industry.”
- Haleem A., Javaid M., Qadri A. M., Singh P. R., Suman R. (2022). “Artificial intelligence (AI) applications for marketing: A literature-based study.”
- Harvard Business School. (2021). “The Future of the Fashion Industry in a Post-COVID-19 World.”
- Harvard Business School. (2022). “The Pandemic Changed Us. Now Companies Have to Change Too.”
- Yuan S., Musibau O. H., Genç Y. S., Shaheen R., Ameen A e, Tan Z. (2021). “Digitalization of economy is the key factor behind fourth industrial revolution: How G7 countries are overcoming with the financing issues?”
- Haenlein M., Kaplan A. (2019). “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence.”
- Heriel E. K. (2021). “Digitalization: the inevitable phenomenon.”
- Immago. (2023). “How artificial intelligence is changing the fashion industry.”
- Kemp S. (2022). “L’audience pubblicitario di TikTok è + di 1 Miliardo (e altre strabilianti statistiche sui social media)”
- Luckow A., Kennedy K., Ziolkowski M., Djerekarov E., Cook M., Duffy E., Schleiss M., Vorster B., Weill E., Kulshrestha A., Smith C. M. (2018). “Artificial Intelligence and Deep Learning Applications for Automotive Manufacturing.”
- Mari Alex. (2019). “The Rise of Machine Learning in Marketing: Goal, Process, and Benefit of AI-Driven Marketing.”
- McKinsey & Company. (2020). “Fashion on Climate.”
- McKinsey & Company. (2020). “Personalizing the customer experience: Driving differentiation in retail.”
- McKinsey & Company. (2020). “How COVID-19 has pushed companies over the technology tipping point and transformed business forever.”
- McKinsey & Company. (2023). “What is AI?”
- McKinsey & Company. (2021). “The post-pandemic state of fashion.”
- McKinsey & Company. (2022). “The State of Fashion 2023: Holding onto growth as global clouds gather.”
- McKinsey & Company. (2022). “COVID-19: Implications for business.”

Mikalef P., Pappas O. I., Krogstie J., Giannakos M. (2017). "Big data analytics capabilities: a systematic literature review and research agenda."

NFT Paris. (2022). "Ai and fashion forecasting: a new approach to predicting trends."

Pwc. (2021). "Sportswear Market Outlook."

Kettydo. (2022). "Fashion Trends 2023: creatività e design vestono il consumatore ma servono anche a ingaggiarlo e fidelizzarlo."

La Repubblica (2021). "Moda nel post pandemia: i consumatori chiedono ai brand un approccio più umano."

Rathore B. (2023). "Digital Transformation 4.0: Integration of Artificial Intelligence & Metaverse in Marketing."

Rialti R., Marzi G., Caputo An, Mayah A. K. (2020). "Achieving strategic flexibility in the era of big data: The importance of knowledge management and ambidexterity."

Wired. (2008). "How much data is generated each day?"

Schwab K. (2017). "The Fourth Industrial Revolution."

Statista. (2019). "Demand share of global apparel market by region."

Statista. (2020). "Market value of artificial intelligence (AI) in marketing worldwide from 2020 to 2028."

Statista. (2023). "Fashion."

Summa Linguae. (2022). "AI and Localization: The New Way Forward."

Sukhmani A. (2022). "The Role of Artificial Intelligence (AI) in Personalized Marketing in Fashion and Apparel Industry."

The Data Analytics Report. (2022). "Importance of Big Data Analytics in Businesses."

The World Economic Forum (2019). "How much data is generated each day?"

Toppan. (2021). "The Role of Artificial Intelligence in Luxury Fashion."

James H. (2022). "Alibaba Launches AI Fashion Assistant."

Alsaleh A. D., Elliot T. M., Fu Q. F., Thakur R. (2019). "Cross-cultural differences in the adoption of social media."

Al-Halah Z., Grauman K. (2020). "From Paris to Berlin: Discovering Fashion Style Influences Around the World."

Amaldoss, W. and Jain, S. (2005). "Pricing of conspicuous goods: a competitive analysis of social effects."

Becker, M., Vignoles, VL, Owe, E., Brown, R., Smith, PB, Easterbrook, M., Herman, G., de Sauvage, I., Bourguignon, D., Torres, A., Camino, L., Lemos, FCS, Ferreira, MC, Koller, SH,

- González, R., Carrasco, D., Cadena, MP, Lay, S., Wang, Q., Yamakoğlu, N. (2012). "Culture and the distinctiveness motive: Constructing identity in individualistic and collectivistic contexts."
- Beward, C. & Gilbert D. (2006). "Fashion's world cities."
- Bain & Company. (2021). "Goods Worldwide Market Monitor 2021"
- Caprar V. D. & Neville A.B. (2012). "'Norming' and 'Conforming': Integrating Cultural and Institutional Explanations for Sustainability Adoption in Business."
- Chen F., MingFu., Li Y., Shen S., Yang J., Weng W., (2023) "A numerical tool for assessing human thermal safety and thermal comfort in cold-weather activities."
- Chi L., Lim K H., Alam, N., Butler C. J. (2016). "Geolocation Prediction in Twitter Using Location Indicative Words and Textual Features."
- Cho E., Kim-Vick J., Yi U., (2021). "Unveiling motivation for luxury fashion purchase among Gen Z consumers: need for uniqueness versus bandwagon effect."
- Christakis A. N., Fowler H. J. (2007). "The spread of obesity in a large social network over 32 years"
- Clifford, J. (1980). "Orientalism."
- Crewe, L. (1996). "Material culture: Embedded firms, organizational networks and the local economic development of a fashion quarter."
- Crewe, L. (2000). "Geographies of retailing and consumption."
- Crewe, L. (2008). "Ugly beautiful? Counting the cost of the global fashion industry."
- Crewe, L. (2010). "Wear: where? The convergent geographies of architecture and fashion."
- Crewe, L. (2016). "Placing fashion: Art, space, display and the building of luxury fashion markets through retail design."
- Crewe, L. (2017). "The geographies of fashion: Consumption, space, and value."
- Crewe L., (2017). "The Geographies of Fashion: consumption, space & value."
- Crewe, L. (2001). "The besieged body: Geographies of retailing and consumption."
- Elteren M. V. (2003). "U.S. Cultural Imperialism Theory: Only a Chimera?"
- Francalanci C., Pernici B., Scalia G., Zeug G. (2018). "Talking about Places: Considering Context for the Geolocation of Images Extracted from Tweets."
- Gregson, N., L. Crewe L., Brooks K. 2002. "Shopping, space, and practice."
- Hasan S., Ukkusuri V. S., Zhan X. (2016). "Understanding Social Influence in Activity Location Choice and Lifestyle Patterns Using Geolocation Data from Social Media."

- Hirsch M. P. (1972). "Processing Fads and Fashions: An Organization-Set Analysis of Cultural Industry Systems."
- Hofstede G., Bond H. M. (1983). "Hofstede's Culture Dimensions: An Independent Validation Using Rokeach's Value Survey."
- Incognia. (2023). "All you need to know about location technology."
- Jeong D., Chun E., Ko E. (2020). "Culture and art policy analysis in fashion capitals: New York, London, Seoul, Beijing, and Jakarta."
- Kesprov J. (2023). "Artificial Intelligence and localization: How AI is changing the landscape."
- Khan N., Ray L. R., Zhang S., Osabuohien E., (2022). "Influence of mobile phone and internet technology on income of rural farmers: Evidence from Khyber Pakhtunkhwa Province, Pakistan."
- Ioannou, L. G., Tsoutsoubi, L., Mantzios, K., Gkikas, G., Piil, J. F., Dinas, P. C., Notley, S. R., Kenny, G. P., Nybo, L., & Flouris, A. D. (2021). "The Impacts of Sun Exposure on Worker Physiology and Cognition."
- Leiserowitz A. A., Kates W. R., Parris M. T. (2006). "Sustainability Values, Attitudes, and Behaviors: A Review of Multinational and Global Trends."
- Lim S. H., Bouchacourt L., Brown-Devlin N., (2020). "Nonprofit organization advertising on social media: The role of personality, advertising appeals, and bandwagon effects."
- Liobikienė G., Mandravickaitė J., Bernatienė J. (2016). "Theory of planned behavior approach to understand the green purchasing behavior in the EU: A cross-cultural study."
- Lucesoli C. (2021). "Chi sono i KOL, i protagonisti dell'influencer marketing cinese!"
- McCann M. R., Honeycutt M. J., Keaton A. S. T. (2010). "Toward Greater Specificity in Cultural Value Analyses: The Interplay of Intrapersonal Communication Affect and Cultural Values in Japan, Thailand, and the United States."
- Cohen D., Nisbett E. R. (1997). "Field Experiments Examining the Culture of Honor: The Role of Institutions in Perpetuating Norms about Violence."
- Mertens L., Cauwenberg V. J., Ghechi A. 2015. "Does the Effect of Micro-Environmental Factors on a Street's Appeal for Adults Bicycle Transport Vary across Different Macro-Environments? An Experimental Study."
- Patchett M., Williams N. (2021). "Geographies of Fashion and Style: Setting the Scene."
- Pugno M. (2020). "Tecnologia beacon: cos'è e come funziona. L'esempio dell'Alto Adige."
- Sarker A., González-Hernandez G., Ruan Y., Perrone J. (2019). "Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter."

- Singh J. & Rane N. (2023). "Fashion Trends – Cause and Effects the Balance Between Style And Function."
- Smith, D. (1996). "Going South: Global Restructuring and Garment Production in Three East Asian Cases."
- Talen K. 2004. "How institutions evolve: the political economy of skills in Germany, Great Britain, the United States and Japan."
- TrensFormative. (2022). "L'industria del lusso e la sfida dei nuovi trend."
- Tutorialspoint. "Factors Influencing Fashion."
- "Understanding Social Influence in Activity Location Choice and Lifestyle Patterns Using Geolocation Data from Social Media."
- Verganti R. (2009). "Design driven Innovation: cambiare le regole della concorrenza innovando radicalmente il significato delle cose."
- Vermeulen J. S., Campbell M. B., e John SI Ingram S. J. (2012). "Climate Change and Food Systems."
- Weller, S. (2007). "Fashion as viscous knowledge: fashion's role in shaping trans-national garment production."
- Williams E., Grey J., Dixon B. (2017). "Improving geolocation of social media posts."
- Zheng C., Jiang Y. J., Zhou Y., Young D. S., Wang W. (2020). "Social Media User Geolocation via Hybrid Attention."

APPENDICE

DATASET IMPORT

```
from google.colab import files
uploaded = files.upload()
import io
df = pd.read_csv(io.BytesIO(uploaded[ 'Dataset_instagram.csv']), encoding='latin1', sep=';',
low_memory='False')
```

DATASET INFORMATION

```
df.head()
```

DATA PRE-PROCESSING

DATA CLEANING

#elimino valori duplicati

```
df.duplicated().sum()
df = df.drop_duplicates()
```

#elimino valori nulli e variabili con troppi valori nulli

```
df.isnull().sum()
df.dropna(axis='index', inplace=True)
```

#elimino variabili con troppi valori nulli

```
df = df.drop(['id','id_row', 'imgs', 'comments', 'display_resources_1', 'display_resources_2',
'display_resources_3', 'media_to_caption_text', 'media_to_comment_has_next_page',
'location_has_public_page', 'owner_blocked_by_viewer', 'owner_followed_by_viewer',
'owner_has_blocked_viewer', 'owner_id', 'owner_is_unpublished', 'owner_profile_pic_url',
'owner_requested_by_viewer', 'height', 'width', 'caption', 'titolo_post','owner_profile_pic_url_hd',
'edge_media_to_sponsor_user'], axis = 1) #elimino variabili non rilevanti per lo studio
```

#elimino variabili che presentano un elevato sbilanciamento

```
df = df.drop(['owner_is_private', 'owner_is_verified'], axis = 1) #elimino variabili troppo sbilanciate
```

#mi trovo il paese di origine di ciascun'osservazione dalle coordinate

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="http")
df['country'] = np.nan
```

```

for i in range(0, len(df)):
    Lat = (df.iloc[[i],[6]].values[0][0])
    print(Lat)
    Long = (df.iloc[[i],[7]].values[0][0])
    print(Long)
    Lat = ".".join(Lat.split(".", 2)[:2])
    print(Lat)
    Long = ".".join(Long.split(".", 2)[:2])
    print(Long)
    try:
        location = geolocator.reverse(Lat+","+Long)
        print(location.raw['address']['country'])
        df.iloc[i, df.columns.get_loc("country")] = location.raw['address']['country']
    except:
        print('none')
        df.iloc[i, df.columns.get_loc("country")] = np.nan
        continue

#elimino valori nulli della feature “country”
df = df.dropna(subset=["country"])

#elimino le feature ritondanti relative alla location
df = df.drop(['location_id', 'location_slug', 'location_name'], axis = 1)

#Standardizzo i valori della feature “country”
pip install pycountry
import pycountry
def standardize_country_name(country_name):
    try:
        country = pycountry.countries.search_fuzzy(country_name)[0]
        return country.name
    except LookupError:
        return country_name
df['country'] = df['country'].apply(standardize_country_name)

#Traduco in inglese i valori della feature “country”
pip install googletrans==4.0.0-rc1
from googletrans import Translator

def translate_country_name(country_name):

```

```

translator = Translator()
translated = translator.translate(country_name, src='auto', dest='en')
return translated.text

for index, row in df.iterrows():
    df.at[index, 'country'] = translate_country_name(row['country'])

#Pulisco la variabile country da alcuni refusi
df['country'] = df['country'].replace("Micronesia, Federated States of", "Federated States of
Micronesia", inplace=False)
df['country'] = df['country'].replace("Ködörösêse tî Beafrika - République Centrafricaine", "Central
African Republic", inplace=False)
df['country'] = df['country'].replace("TÃ¼rkiye", "Turkey", inplace=False)
df['country'] = df['country'].replace("á¹x82ajeá", "Majel / Marshall Islands", inplace=False)
df['country'] = df['country'].replace("Schweiz/Suisse/Svizzera/Svizra", "Switzerland", inplace=False)
df['country'] = df['country'].replace("Ireland / Ireland", "Ireland", inplace=False)
df['country'] = df['country'].replace("Tchad Chad", "Republic of Chad", inplace=False)
df['country'] = df['country'].replace("Kingdom of Cambodia", "Cambodia", inplace=False)
df['country'] = df['country'].replace("Bosna i Hercegovina / Bosnia and Herzegovina", "Bosnia-
Herzegovina", inplace=False)
df['country'] = df['country'].replace("china", "China", inplace=False)
df['country'] = df['country'].replace("Majel", "Majel - Marshall Islands", inplace=False)
df['country'] = df['country'].replace("Armenia:", "Armenia", inplace=False)
df['country'] = df['country'].replace("Algérie / ⵍⵣⵣⵓⵔ / Algeria", "Algeria", inplace=False)
df['country'] = df['country'].replace("Bosna i hercegovina / Bosnia and Herzegovina", "Bosnia and
Herzegovina", inplace=False)
df['country'] = df['country'].replace("Schweiz/Suisse/Switzerland/Swiss", "Switzerland",
inplace=False)
df['country'] = df['country'].replace("Mjajel", "Majel - Marshall Islands", inplace=False)
df['country'] = df['country'].replace("THE MOST KNOW འབྲུག་རྒྱལ་ཁབ་", "China", inplace=False)

#Copio il dataset attuale
df_orig = df.copy()
df_orig.to_csv('dataset_last.csv', index=False)
from google.colab import files
files.download('dataset_last.csv')

```

DATASET TRASFORMATION

#Trasformo le variabili categoriche in numeriche

```
label_encoder=LabelEncoder()
```

```
label_create = label_encoder.fit_transform(df["create_at"])
```

```
df.drop("create_at", axis=1, inplace= True)
```

```
df["create_at"] = label_create
```

```
label_create
```

```
label_encoder2=LabelEncoder()
```

```
label_hashtag_ricerca = label_encoder.fit_transform(df["hashtag_ricerca"])
```

```
df.drop("hashtag_ricerca", axis=1, inplace= True)
```

```
df["hashtag_ricerca"] = label_hashtag_ricerca
```

```
label_hashtag_ricerca
```

```
label_encoder3=LabelEncoder()
```

```
label_hasht_cor = label_encoder.fit_transform(df["hasht_cor"])
```

```
df.drop("hasht_cor", axis=1, inplace= True)
```

```
df["hasht_cor"] = label_hasht_cor
```

```
label_hasht_cor
```

```
label_encoder7=LabelEncoder()
```

```
label_edge_liked_by_count = label_encoder.fit_transform(df["edge_liked_by_count"])
```

```
df.drop("edge_liked_by_count", axis=1, inplace= True)
```

```
df["edge_liked_by_count"] = label_edge_liked_by_count
```

```
label_edge_liked_by_count
```

```
label_encoder8=LabelEncoder()
```

```
label_media_to_comment_count = label_encoder.fit_transform(df["media_to_comment_count"])
```

```
df.drop("media_to_comment_count", axis=1, inplace= True)
```

```
df["media_to_comment_count"] = label_media_to_comment_count
```

```
label_media_to_comment_count
```

#Eseguo la standardizzazione

```
scaler = StandardScaler()
```

```
scaler.fit(df)
```

```
scaled_df = pd.DataFrame(scaler.transform(df),columns= df.columns )
```

```
scaled_df.describe()
```

DATA EXPLORATION

STATISTIC ANALYSIS

#Statistica descrittiva

```
df.describe()
```

#Matrice di correlazione

```
sns.set(rc={"axes.facecolor":"#FFF9ED","figure.facecolor":"#FFF9ED"})  
pallet = ["#682F2F", "#9E726F", "#D6B2B1", "#B9C0C9", "#9F8A78", "#F3AB60"]  
cmap = colors.ListedColormap(["#682F2F", "#9E726F", "#D6B2B1", "#B9C0C9", "#9F8A78",  
"#F3AB60"])  
corrmat= df.corr()  
plt.figure(figsize=(6,6))  
sns.heatmap(corrmat,annot=True, chap=cmap, center=0)
```

#Verifico l'esistenza di outlier

```
sns.boxplot(df['create_at'])  
sns.boxplot(df[hashtag_ricerca])  
sns.boxplot(df[hasht_cor])  
sns.boxplot(df['country'])  
sns.boxplot(df['edge_liked_by_count'])  
sns.boxplot(df['media_to_comment_count'])
```

DATA VISUALIZATION

#Verifico quali paesi ci sono all'interno della feature "country"

```
plt.figure(1 , figsize = (14, 11))  
sns.set(style="whitegrid")  
sns.countplot(y = 'country' , data = df)  
plt.show()
```

#Verifico le tipologie di borse presenti nello studio

```
plt.figure(1 , figsize = (5, 3))  
sns.countplot(y = 'hashtag_ricerca' , data = df_orig)  
plt.show()
```

DIMENSIONALLY REDUCTION

```
pca = PCA(n_components=2)  
pca.fit(scaled_df)  
PCA_df= pd.DataFrame(pca.transform(scaled_df), columns=(["col1", "col2"]))
```



```

PCA_df.describe()
x =PCA_df["col1"]
y =PCA_df["col2"]
fig = plt.figure(figsize=(10,8))
ax = fig.add_subplot(111, projection="3d")
ax.scatter(x,y, c="maroon", marker="o" )
ax.set_title("A 3D Projection Of Data In The Reduced Dimension")
plt.show()

```

DATA MINING

NUMERO CLUSTER OTTIMALE

#Distortion score

```

from yellowbrick.cluster import KElbowVisualizer
Elbow_M = KElbowVisualizer(KMeans(), k=15)
Elbow_M.fit(pca_x)
Elbow_M.show()

```

#Wcss

```

wcss = []
for i in range(1, 11):
    km = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
    km.fit(pca_x)
    wcss.append(km.inertia_)
plt.figure(figsize=(20,8))
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method', fontsize = 20)
plt.xlabel('No. of Clusters')
plt.ylabel('wcss')

```

CLUSTERING

#Kmeans Algorithm

```

kmeans = KMeans(n_clusters = 5, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 42)
ymmeans = kmeans.fit_predict(pca_x)
plt.figure(figsize=(10, 8))
plt.title('Cluster delle Prime Due Componenti Principali (K-Means)', fontsize=18)
4 == len(np.unique(ymmeans))
num_cluster = len(np.unique(ymmeans))

```

```

colors = []
for cluster_num in range(num_cluster):
    if cluster_num == 0: # Cluster 1
        colors.append('darkred')
    else:
        colors.append(plt.cm.Red(0.2 + (0.6 / (num_cluster - 1)) * (cluster_num - 1)))
for i in range(5):
    plt.scatter(pca_x[ymeans == i, 0], pca_x[ymeans == i, 1], s=100, c=colors[i], label=f'Cluster {i + 1}')

plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], s=200, c='black',
label='Centroidi', marker='X')
plt.xlabel('PCA1', fontsize=14)
plt.ylabel('PCA2', fontsize=14)
plt.legend(fontsize=12)
plt.grid()
plt.show()

#centroidi
centri = kmeans.cluster_centers_
centri

#Silhouette score
from sklearn.metrics import silhouette_samples, silhouette_score
print('silhouette_score: %.3f' % metrics.silhouette_score(pca_x, ymeans))

#Devies-Bouldin Score
davies_bouldin_score(pca_x, ymeans)

#Calinski Harabasz Index
metrics.calinski_harabasz_score(pca_x, ymeans)

#Calcolo il conteggio dei valori in ciascun cluster
unique_clusters, cluster_counts = np.unique(ymeans, return_counts=True)
palette = sns.color_palette("Reds", len(unique_clusters))
plt.figure(figsize=(8, 6))
plt.pie(cluster_counts, labels=unique_clusters, autopct='% 1.1f%%', colors=palette)
plt.title("Distribuzione dei Cluster (K-Means)", fontsize=16)
plt.show()

```

#Verifico se ci sono sovrapposizioni tra i punti

```
from sklearn.metrics.pairwise import euclidean_distances
jitter = np.random.normal(0, 0.03, size=pca_x.shape)
pca_x_jittered = pca_x + jitter
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 1)
sns.scatterplot(x=pca_x_jittered[:, 0], y=pca_x_jittered[:, 1], hue=ymeans, palette='Set1', legend='full',
alpha=0.7, edgecolor='none')
plt.title('Grafico di Sovrapposizione dei Punti per Clustering - Kmeans')
```

#Creo una heatmap globale

```
import folium
data = pd.DataFrame({
    'Lat': df_orig['latitude'], # Utilizza 'latitude' come chiave tra parentesi quadre
    'Lon': df_orig['longitude'], # Utilizza 'longitude' come chiave tra parentesi quadre
    'Cluster': ymeans, # Assegnazione dei cluster ottenuta dall'algoritmo K-Means
    'hashtag_ricerca': df_orig['hashtag_ricerca'], # Colonna degli hashtag
    'Like': df_orig['edge_liked_by_count'], # Colonna dei like
    'Commenti': df_orig['media_to_comment_count']
})
latitudine_centrale = 0.0
longitudine_centrale = 0.0
mappa = folium.Map(location=[latitudine_centrale, longitudine_centrale], zoom_start=2) # Imposta il
livello di zoom a 2 per vedere l'intero mondo
colori_cluster = ['red', 'blue', 'green', 'purple', 'orange']
for index, row in data.iterrows():
    colore_cluster = colori_cluster[row['Cluster']] # Colore del cluster
    tooltip_text = f"Cluster: {row['Cluster']}<br>Hashtag: {row['hashtag_ricerca']}<br>Like:
{row['Like']}<br>Commenti: {row['Commenti']}"
    folium.Marker(location=[row['Lat'], row['Lon']], icon=folium.Icon(color=colore_cluster),
tooltip=tooltip_text).add_to(mappa)
mappa.save('cluster_mapKmeans.html')
```

#Agglomerative Hierarchical Algorithm

```
AC = AgglomerativeClustering(n_clusters=5, affinity='euclidean', linkage='ward')
yhat_AC = AC.fit_predict(pca_x)
plt.figure(figsize=(10, 8))
plt.title('Cluster delle Prime Due Componenti Principali (Agglomerative)', fontsize=18)
```

```

colors = ['#00008B', '#0000FF', '#1E90FF', '#87CEEB', '#ADD8E6']
for cluster_num in range(5):
    plt.scatter(pca_x[yhat_AC == cluster_num, 0], pca_x[yhat_AC == cluster_num, 1], s=100,
c=colors[cluster_num], label=f'Cluster {cluster_num}')
cluster_centers_x = [np.mean(pca_x[yhat_AC == cluster_num, 0]) for cluster_num in range(5)]
cluster_centers_y = [np.mean(pca_x[yhat_AC == cluster_num, 1]) for cluster_num in range(5)]
plt.scatter(cluster_centers_x, cluster_centers_y, s=200, c='black', marker='X', label='Centroidi')
plt.xlabel('PCA1', fontsize=14)
plt.ylabel('PCA2', fontsize=14)
plt.legend(fontsize=12)
plt.grid()
plt.show()

#Silhouette score
print("silhouette.coefficient.hierarchical: %0.3f" % metrics.silhouette_score(pca_x, yhat_AC))

#Davies Bouldin Score
davies_bouldin_score(pca_x, yhat_AC)

#Calinski Harabasz Index
metrics.calinski_harabasz_score(pca_x, yhat_AC)

#Calcolo il conteggio dei valori in ciascun cluster
unique_clusters, cluster_counts = np.unique(yhat_AC, return_counts=True)
palette = sns.color_palette("Blues", len(unique_clusters))
plt.figure(figsize=(8, 6))
plt.pie(cluster_counts, labels=unique_clusters, autopct='%1.1f%%', colors=palette)
plt.title("Distribuzione dei Cluster (Agglomerative)", fontsize=16)
plt.show()

#Verifico se ci sono sovrapposizioni tra i punti
plt.figure(figsize=(12, 5))
plt.subplot(1, 2, 2)
sns.scatterplot(x=pca_x_jittered[:, 0], y=pca_x_jittered[:, 1], hue=yhat_AC, palette='Set3',
edgecolor='none', legend='full', alpha=0.7)
plt.title('Grafico di Sovrapposizione dei Punti per Clustering - Agglomerative')

#Individuo il valore dell'iperparametro eps per l'algorithm DBSCAN
neigh = NearestNeighbors(n_neighbors=2)
nbrs = neigh.fit(pca_x)

```

```

distances, indices = nbrs.kneighbors(pca_x)
distances = np.sort(distances, axis=0)
plt.figure(figsize=(12,8))
plt.plot(distances[:,1])

#DBSCAN Algorithm
db = DBSCAN(eps=0.08, min_samples=3).fit(pca_x)
y_db = db.labels_
y_db
plt.figure(figsize=(10, 8))
plt.title('Cluster delle Prime Due Componenti Principali (DBSCAN)', fontsize=18)
num_cluster = len(np.unique(y_db))
colors = plt.cm.Oranges(np.linspace(0, 1, num_cluster))
for cluster_num in range(num_cluster):
    color = colors[cluster_num] # Assegna colori a tutti i cluster
    plt.scatter(pca_x[y_db == cluster_num, 0], pca_x[y_db == cluster_num, 1], s=100, c=color,
label=f'Cluster {cluster_num}')
plt.xlabel('PCA1', fontsize=14)
plt.ylabel('PCA2', fontsize=14)
plt.grid()
plt.show()

#Silhouette score
print("silhouette.coefficient.hierarchical: %0.3f" % metrics.silhouette_score(pca_x, y_db))

#Calinski Harabasz Index
metrics.calinski_harabasz_score(pca_x, y_db)

#Davies Bouldin Score
davies_bouldin_score(pca_x, y_db)

#Calcolo il conteggio dei valori in ciascun cluster
unique_clusters, cluster_counts = np.unique(y_db, return_counts=True)
colors = sns.color_palette("Purples", len(unique_clusters))

#Verifico se ci sono sovrapposizioni tra i punti
plt.figure(figsize=(8, 6))

plt.pie(cluster_counts, labels=unique_clusters, autopct='% 1.1f%%', colors=colors)

plt.title("Distribuzione dei Cluster (DBSCAN)", fontsize=16)

```

```
plt.show()
```

#Spectral Algorithm

```
sc = SpectralClustering(n_clusters=5,
```

```
...     assign_labels='discretize',
```

```
...     random_state=0).fit(pca_x)
```

```
y_sc=sc.labels_
```

```
plt.figure(figsize=(10, 8))
```

```
plt.title('Cluster delle Prime Due Componenti Principali (Spectral Clustering)', fontsize=18)
```

```
num_cluster = len(np.unique(sc.labels_))
```

```
colors = ['#008000', '#00FF00', '#32CD32', '#228B22', '#008B45']
```

```
for cluster_num in range(num_cluster):
```

```
    plt.scatter(pca_x[sc.labels_ == cluster_num, 0], pca_x[sc.labels_ == cluster_num, 1], s=100,  
c=colors[cluster_num], label=f'Cluster {cluster_num + 1}')
```

```
plt.xlabel('PCA1', fontsize=14)
```

```
plt.ylabel('PCA2', fontsize=14)
```

```
plt.legend(fontsize=12)
```

```
plt.grid()
```

```
plt.show()
```

#Silhouette score

```
print('silhouette_score: %.3f' % metrics.silhouette_score(pca_x, y_sc))
```

#Davies Bouldin Score

```
davies_bouldin_score(pca_x, y_sc)
```

#Calinski Harabasz Index

```
metrics.calinski_harabasz_score(pca_x, y_sc)
```

#Calcolo il conteggio dei valori in ciascun cluster

```
unique_clusters, cluster_counts = np.unique(y_sc, return_counts=True)
```

```
palette = sns.color_palette("YlGnBu", len(unique_clusters))
```

```
plt.figure(figsize=(8, 6))
```

```
plt.pie(cluster_counts, labels=unique_clusters, autopct='%1.1f%%', colors=palette)
```

```
plt.title("Distribuzione dei Cluster (Spectral)", fontsize=16)
```

```
plt.show()
```

#Verifico se ci sono sovrapposizioni tra i punti

```
from sklearn.metrics.pairwise import euclidean_distances
```

```
jitter = np.random.normal(0, 0.03, size=pca_x.shape)
```

```

pca_x_jittered = pca_x + jitter
plt.figure(figsize=(8,6))
plt.subplot(1, 2, 2)
sns.scatterplot(x=pca_x_jittered[:, 0], y=pca_x_jittered[:, 1], hue=y_sc, palette='Set2', legend='full',
alpha=0.7, edgecolor='none')
plt.title('Grafico di Sovrapposizione dei Punti per Clustering - Spectral')

```

PROFILING PROCESS

#creo un nuovo dataset aggiungendo al mio dataset originale la variabile “cluster”

```

df_pca_kmeans = pd.concat([df_orig.reset_index(drop = True), pd.DataFrame(pca_x)], axis=1)
df_pca_kmeans['cluster'] = ymeans
df_pca_kmeans.head()

```

#Individuo la distribuzione dei paesi per ciascun cluster

```

colors = ['red', 'blue', 'green', 'purple', 'orange']
cluster_labels = {0: 1, 1: 2, 2: 3, 3: 4, 4: 5}
countries_count_by_cluster = df_new.groupby('cluster')['country'].nunique()
plt.bar(countries_count_by_cluster.index.map(cluster_labels), countries_count_by_cluster.values,
color=colors)
plt.xlabel('Cluster')
plt.ylabel('Numero di paesi')
plt.title('Numero di paesi in ciascun cluster')
plt.xticks(countries_count_by_cluster.index.map(cluster_labels))
plt.show()

```

#Stampo l’elenco dei paesi per ciascun cluster

```

unique_countries_by_cluster = df_new.groupby('cluster')['country'].apply(set).to_dict()
for cluster_id, unique_countries in unique_countries_by_cluster.items():
    print(f"Cluster {cluster_id}:")
    print(unique_countries)

```

#Individuo i paesi presenti in più cluster

```

for country in df_new['country'].unique():
    num_country_clusters = df_new.loc[df_new['country'] == country, 'cluster'].nunique()
    if num_country_clusters > 1:
        countries_in_multiple_clusters.append(country)
print(countries_in_multiple_clusters)

```

#Individuo gli hashtag di ricerca presenti in ciascun cluster

```
from collections import Counter

color_map = {1: 'red', 2: 'blue', 3: 'green', 4: 'purple', 5: 'orange'}
cluster_mapping = {0: 1, 1: 2, 2: 3, 3: 4, 4: 5}
top_hashtags_per_cluster = {}

for cluster_num in df_new['cluster'].unique():
    hashtags_in_cluster = df_new[df_new['cluster'] == cluster_num]['hashtag_ricerca']
    combined_hashtags = " ".join(hashtags_in_cluster).split()
    hashtag_counts = Counter(combined_hashtags)
    top_hashtags_per_cluster[cluster_num] = hashtag_counts.most_common(5) # Estrarre i primi 5 più comuni

plt.figure(figsize=(12, 8))

for cluster_num, top_hashtags in top_hashtags_per_cluster.items():
    hashtags, counts = zip(*top_hashtags)
    cluster_color = color_map.get(cluster_mapping.get(cluster_num, 'gray')) # Se il cluster non ha un colore specifico, usa il grigio
    plt.bar(hashtags, counts, label=f'Cluster {cluster_mapping.get(cluster_num, "N/A")}',
            color=cluster_color)

plt.title("Top Hashtag di Ricerca per Cluster")
plt.xlabel("Hashtag di Ricerca")
plt.ylabel("Frequenza")
plt.xticks(rotation=45, ha='right')
plt.legend()
plt.tight_layout()
plt.show()
```

#Individuo gli hashtag correlati più frequenti per ciascun cluster

```
hashtag_freq_by_cluster = df_new.explode('hasht_cor').groupby(['cluster', 'hasht_cor']).size().reset_index(name='hashtag_count')
top_hashtags_by_cluster = hashtag_freq_by_cluster.groupby('cluster').apply(lambda x: x.nlargest(5, 'hashtag_count')).reset_index(drop=True)
top_hashtags_by_cluster = top_hashtags_by_cluster.sort_values(by=['cluster', 'hashtag_count'], ascending=[True, False])
excel_filename = "hashtag_analysis.xlsx"
top_hashtags_by_cluster.to_excel(excel_filename, index=False)
print(f"Dati esportati in {excel_filename}")
```


#Individuo le tendenze temporali per ciascun cluster

```
df_new['publish_hour'] = pd.to_datetime(df_new['create_at']).dt.hour
grouped = df_new.groupby(['cluster', 'publish_hour']).size().reset_index(name='count')
colors = ['red', 'blue', 'green', 'purple', 'orange']
plt.figure(figsize=(12, 6))
sns.lineplot(data=grouped, x='publish_hour', y='count', hue='cluster', palette=colors)
plt.title('Tendenze Temporali di Pubblicazione dei Post per Cluster')
plt.xlabel('Ora di Pubblicazione')
plt.ylabel('Numero di Post')
plt.xticks(range(24))
plt.legend(title='Cluster')
plt.show()
```

#Individuo le tendenze temporali di pubblicazione per ciascun paese

```
df_new['publish_hour'] = pd.to_datetime(df_new['create_at']).dt.hour
clusters = df_new['cluster'].unique()
for cluster in clusters:
    df_cluster = df_new[df_new['cluster'] == cluster]
    grouped = df_cluster.groupby(['country', 'publish_hour']).size().reset_index(name='count')
    plt.figure(figsize=(40, 20))
    sns.lineplot(data=grouped, x='publish_hour', y='count', hue='country')
    plt.title(f'Tendenze Temporali di Pubblicazione dei Post per Cluster {cluster}')
    plt.xlabel('Ora di Pubblicazione')
    plt.ylabel('Numero di Post')
    plt.xticks(range(24))
    plt.legend(title='Paese')
    plt.show()
```

#Calcolo la media di like per ciascun cluster

```
mean_likes_per_cluster = df_new.groupby('cluster')['edge_liked_by_count'].mean()
colors = ['red', 'blue', 'green', 'purple', 'orange']
mean_likes_per_cluster.index = [1, 2, 3, 4, 5]
plt.figure(figsize=(10, 6))
mean_likes_per_cluster.plot(kind='bar', rot=0, color=colors) # 'rot=0' evita di ruotare le etichette
sull'asse x
plt.xlabel('Cluster')
plt.ylabel('Like')
plt.title('Like per Cluster')
```

```
plt.show()
```

#Calcolo la media di commenti per ciascun cluster

```
mean_likes_per_cluster = df_new.groupby('cluster')['media_to_comment_count'].mean()
```

```
colors = ['red', 'blue', 'green', 'purple', 'orange']
```

```
mean_likes_per_cluster.index = [1, 2, 3, 4, 5]
```

```
plt.figure(figsize=(10, 6))
```

```
mean_likes_per_cluster.plot(kind='bar', rot=0, color=colors)
```

```
plt.xlabel('Cluster')
```

```
plt.ylabel('Commenti')
```

```
plt.title('Commenti per Cluster')
```

```
plt.show()
```

#Visualizzo i paesi in base al loro livello medio di like

```
grouped = df_new.groupby('country').mean()[['media_to_comment_count',  
'edge_liked_by_count']].reset_index()
```

```
sorted_likes = grouped.sort_values('edge_liked_by_count', ascending=False)
```

```
plt.figure(figsize=(12,6))
```

```
sns.barplot(x='country', y='edge_liked_by_count', data=sorted_likes)
```

```
plt.xticks(rotation=90)
```

```
plt.xlabel('Paese')
```

```
plt.ylabel('Media dei like')
```

```
plt.title('Media dei like per paese')
```

```
plt.show()
```

```
sorted_comments = grouped.sort_values('media_to_comment_count', ascending=False)
```

```
plt.figure(figsize=(12,6))
```

```
sns.barplot(x='country', y='media_to_comment_count', data=sorted_comments)
```

```
plt.xticks(rotation=90)
```

```
plt.xlabel('Paese')
```

```
plt.ylabel('Media dei commenti')
```

```
plt.title('Media dei like')
```

#Visualizzo i paesi in base al loro livello di engagement

```
df_new['Engagement'] = (df_new['edge_liked_by_count'] + df_new['media_to_comment_count']) /  
(df_new['edge_liked_by_count'] + df_new['media_to_comment_count']).count()
```

```
engagement_by_country = df_new.groupby('country')['Engagement'].mean().reset_index()
```

```
engagement_by_country = engagement_by_country.sort_values(by='Engagement', ascending=False)
```

```
plt.figure(figsize=(12, 8))
```

```
sns.barplot(x='country', y='Engagement', data=engagement_by_country, palette='pastel')
plt.title("Engagement Medio per Paese (Ordine Decrescente)")
plt.xlabel("Paese")
plt.ylabel("Engagement Medio")
plt.tight_layout()
plt.xticks(rotation=45, ha='right')
plt.show()
```